Japanese pitch accent production in an English/Nupe/Hausa trilingual: accuracy, stability and F0 realisation

Becky Muradás-Taylor
York St John University

Introduction

This is a case study of a person with three L1s – English and the Nigerian languages Nupe and Hausa – who started studying Japanese as an additional language in the UK at the age of 30. The study investigates the participant's production of pitch accent in their spoken Japanese, focusing on its accuracy, i.e. adherence to Standard Japanese norms, stability, i.e. the extent to which repeated words have the same accent type, and F0 realisation, i.e. the F0 peak location and rate of F0 fall. These are compared to the accuracy, stability, and F0 realisation of the accent types produced by 21 monolingual English learners of Japanese (Muradás-Taylor, in progress; Taylor, 2012). Unlike the monolingual English learners, the English/Nupe/Hausa trilingual learner is shown to produce pitch accent that is highly accurate and stable. In addition, the acoustic data indicates that F0 peak location and rate of F0 fall could also be consistent with Standard Japanese norms.

Although the participant is a trilingual learner of another language, this is not L3/Ln phonology research (Cabrelli Amaro, 2012). It is beyond the scope of this paper to consider which of the trilingual participant's L1s most affects their Japanese, and how this relates to factors such as typological distance or language status. No attempt is made to identify whether any of the learners L1s are more dominant than any other, nor how they interact with one another. Instead, this paper's significance lies in the fact that it demonstrates that it is, in fact, possible to acquire accurate and stable Standard Japanese pitch accent. This has implications for research on monolingual English learners of Japanese, who produce accent types that are inaccurate and unstable, even after four years of Japanese study, including a year studying abroad in a university in Japan (Muradás-Taylor, in progress; Taylor, 2011a; Taylor, 2011b; Taylor, 2012). English speakers' difficulty acquiring pitch accent has been argued by the current author (Muradás-Taylor, in progress; Taylor, 2011a; Taylor, 2011b; Taylor, 2012) to be due to pitch not having lexical function in English, combined with the effect of pitch accent having low functional load in Standard Japanese (Kitahara, 2001), and showing considerable dialectal variation (Kubozono, 2012). However, alternative explanations could be: insufficient Standard Japanese input (see Flege, 2009) or lack of explicit instruction (see Thomson & Derwing, 2015). This study on the English/Nupe/Hausa trilingual learner - who has never lived in Japan, studied Japanese with a L1-English speaking tutor in the UK, and did not receive explicit instruction on pitch accent - allows us to be more confident in attributing the monolingual English learners' difficulty to a linguistic cause: the difference between the English/Nupe/Hausa trilingual learner and the monolingual English learners is their L1(s), not input or instruction. And unlike English, which does not use pitch lexically, Nupe and Hausa, both of which are tone languages, do.

This paper expands on Muradás-Taylor (2017) which reported the accuracy and stability of the English/Nupe/Hausa trilingual learner's accent types in a three-minute audio recording. A second judge has been used to identify the speaker's accent types, and clearer criteria were developed to decide what to count as a word, and how to deal with downstepped or deleted accents (Venditti, 2005:175). More detail about the speaker's pitch accent patterns have been added, and the comparison of accuracy and stability with the 21 monolingual English learners has been expanded. In addition, analysis on F0 realisation is reported – this is new for both the English/Nupe/Hausa trilingual learner and the monolingual English learners. The acoustic data obtained from the recordings is limited in scope, but it is included in the paper because of the questions it raises for future research.

Pitch in Japanese, Nupe, Hausa and English

Pitch in Japanese has lexical function, i.e. it is used to distinguish words. For example in Standard Japanese *hashi* 'chopsticks' has initial accent and *hashi* 'bridge' has final accent. What distinguishes different accent types is the presence or absence of an accent and its position (Vance, 2008). Initially-accented words have an accent on the first syllable; finally-accented words have an accent on the final syllable; unaccented words have no accent. Words longer than two syllables can have an accent on the second, third, fourth, etc. syllable: for simplicity, they are often referred to as having medial accent.

The accent in Standard Japanese is realised as a sharp fall in pitch (Vance, 2008). Sakamoto (2011) argues, based on both a literature review and her own data, that the acoustic correlates of pitch accent in Standard Japanese are F0 peak position, rate of F0 fall, and F0 range. Of these, F0 range does not appear to correlate consistently with accent type in the L1 Japanese-speaker data in Sakamoto (2011). But, there does seem to be agreement in the literature that accented words in Standard Japanese have a F0 peak, which is located around the end of the accented syllable, followed by a steep fall.

Pitch in Nupe and Hausa also has lexical function; both are tonal. Nupe has five tones: high, mid, low, rise, fall; Hausa has three: high, low, and a fall derived from a sequence of high and low (Yip, 2002). All syllables are specified for tone, unlike pitch accent where one syllable per accented word (and no syllables in unaccented words) have an accent.

Pitch in English does not have lexical function. Instead, English uses pitch post-lexically, so that, for example, a question may have rising pitch, and a statement, falling. English stress has multiple acoustic correlates: duration, spectral tilt (which listeners perceive as loudness), and, when focused, a pitch movement and a boost in intensity (Sluijter & van Heuven, 1996a, 1996b). The pitch movement on focused syllables is known as a 'pitch accent'. Despite using the same terminology, a Japanese pitch accent and an English one are different: Japanese is lexical and always a fall; English is post-lexical, and can take different shapes depending on whether the utterance is a statement, question, expression of surprise, etc.

L1 English listeners are able to use pitch as a cue to the difference between the noun subject and the verb subject (Fry, 1958). This is the likely source of the 'common misunderstanding' (Beckman & Edwards, 1994:13) that pitch is an acoustic correlate of stress but has arisen due to the confound between utterance-level post-lexical pitch and stress. In fact, pitch is an unreliable cue to word meaning in English: unlike the Japanese words *hashi* 'chopsticks' and *hashi* 'bridge', you can say the English word *happy*, for

example, with falling pitch, rising pitch, falling-rising pitch etc., and you still have the same word.

Accuracy and stability of Japanese pitch accent by L1 English speakers

Research investigating the pitch accent production of three Australian English learners of Japanese reported that one learner produced the word *yappari* 'as I thought' with three different accent types in one conversation: initial accent, medial accent, and unaccented (Yamada, 1994). Drawing on this, a study was designed to investigate the accuracy and stability of accent types produced by L1 English-speaking L2 Japanese learners (Muradás-Taylor, in progress; Taylor, 2011a; Taylor, 2011b; Taylor, 2012). It is the English/Nupe/Hausa trilingual learner's different behaviour to these monolingual English learners that is of interest in the current study. The participants were L1 speakers of Standard Southern British English: a less-experienced group (n=13) who had studied Japanese for one or two years (mean 250 hours), and a more-experienced group (n=8) who had completed four years of a Japanese degree, including a year in Japan. They read aloud Japanese words in three contexts: in isolation (e.g., *ame* 'rain'); before a function word (e.g., *ame da* 'it's rain') and before a content word (e.g., *ame ga furu* 'rain falls'). Japanese phoneticians identified the accent type of each word that the participants produced.

Muradás-Taylor (in progress) showed that these participants produced pitch accent that is neither accurate nor stable. The accuracy, i.e. percentage match with Standard Japanese, was 43% for the less-experienced group (lowest participant 32%, highest participant 52%, SD=6) and also 43% for the more-experienced group (lowest participant 36%, highest 48%, SD=5). The learners' accent types showed considerable instability: only 18% of the words produced by the less-experienced group, and 19% of the words produced by the more-experienced group had accent types that were produced accurately and stably across the three contexts. In sum, all learners' accent types were inaccurate and unstable – neither the more-experienced group, nor any individual learner, had accurate and stable accent types.

Acoustic realisation of Japanese pitch accent by L1 English speakers

To measure the acoustic realisation of accent types produced by English-speaking learners of Japanese, Sakamoto (2011) used a method in which the learners imitated native speaker productions. Sakamoto (2011:48) explains that this was necessary since English-speaking learners were found to be unable to produce accent types from written prompts, even with accent markings. For unaccented words, it was found that almost all words were identified by native speaker judges as being unaccented; the learners were similar to native speakers in that they didn't have a sharp pitch fall in their productions. For accented words, some words were misidentified by the listeners as having different accent types. This was reflected in the acoustic analysis with some words showing an earlier peak, some a later peak, and some no sharp pitch fall when one was expected. However, this study included all productions by the learners, whether or not their accent types were accurate. This means that it is not clear whether the early and late F0 peaks that were observed were due to learners producing an inaccurate accent type, or due to an inaccurate phonetic realisation of a correct accent type. It is difficult to draw conclusions, therefore, as to whether English speakers' L2

Japanese pitch accent realisations (F0 peak position, rate of F0 fall) are within Standard Japanese norms in words with accurate pitch accent.

Other research has measured the ratio between F0 of the first and second vowel (Kondo, 2012; Ueyama, 2012). This measure will not be used in the current study, following Sakamoto (2011:41-42), who argues that it is less useful than F0 peak location and rate of F0 fall, since L1 speakers of Japanese do not use it to distinguish between different accent types. In addition to pitch, individual variation is also observed in duration, which is not a correlate of pitch accent in Standard Japanese. Some, but not all, learners lengthen the accented syllable (Kondo, 2012; Ueyama, 2012).

Research questions

1) What percentage of words does an English/Nupe/Hausa trilingual learner of Japanese produce with accurate accent types, i.e. accent types that conform to Standard Japanese norms?
2) Are repeated words produced with the same or different accent types? What is the percentage of words that have stable accent types? And what is the percentage of words that have both accurate and stable accent types?
3) How do F0 peak location and rate of F0 fall compare to Standard Japanese norms for the English/Nupe/Hausa trilingual learner?
4) How do (1), (2) and (3) compare to monolingual English learners of Japanese?
5) What does (4) tell us about L2 acquisition of pitch accent?

Method

Data collection

The data for the English/Nupe/Hausa trilingual learner of Japanese was a three-minute audio recording of a presentation on a topic chosen by the participant. The presentation was part of an assessment; informed consent for it to be used for research was obtained subsequently. The data for the monolingual English learners of Japanese was 180 words which were read aloud in three contexts (e.g., *ame* 'rain', *ame da* 'it's rain' and *ame ga furu* 'rain falls'). Full details are in Taylor (2012) and Muradás-Taylor (in progress).

Participants

The English/Nupe/Hausa trilingual learner is the same as that in Muradás-Taylor (2017). The participant is described with the gender neutral pronoun 'they' in order to preserve their anonymity. They were raised in Nigeria and the UK. They were born in Nigeria, lived in the UK from age one to age five, and Nigeria from age six to age 23. They were raised trilingually: English and the Nigerian languages Nupe and Hausa were all used in the home from birth. They were educated in English both in the UK and in Nigeria, and studied Hausa at school from age nine. The participant started a beginner-level Japanese class at a UK university at the age of 30. The course was delivered by a British English tutor for two hours a week for 12 weeks. They then took the intermediate-level course with the same tutor which was three hours a week for a further 12 weeks. They did not receive any explicit pitch accent instruction on these courses. At the time that the data was collected, they had never been to Japan.

The key facts regarding the participant's language background are firstly that, as well as English, they have two further L1s, Nupe and Hausa, which are tonal and therefore have lexical pitch; and secondly, that the participant had not received explicit instruction on pitch accent and had not been to Japan. However, the participant's language background is actually more complex: as well as their three L1s (English/Nupe/Hausa), they also learned Arabic from the age of three for Quranic recitation, and Hindi from the age of six through interaction with Hindi-speaking neighbours. They also spent time in Malaysia and India between the ages of 23 and 27. Additionally, before entering the beginner-level Japanese class, they had a long-standing interest in Japanese poetry, role-playing games and anime.

The monolingual English learners are the same participants as in Taylor (2012), i.e. 21 speakers of Standard Southern British English (SSBE) divided into a less-experienced group (n=13) or a more-experienced group (n=8) based on their amount of Japanese language instruction, and whether or not they had completed a study abroad year in Japan. The less-experienced group had completed one or two years of an undergraduate degree course in Japanese and had not yet spent the year in Japan. They had received an average of 250 hours of Japanese language instruction (min 70 hours, max 430 hours, SD 90). The more-experienced group had all completed an undergraduate degree course in Japanese, receiving an average of 970 hours of Japanese language instruction (min 640 hours, max 1400 hours, SD 320). Seven of the less-experienced participants had never been to Japan; six had visited, but none had stayed more than three months. The more-experienced group (n=8) had all spent a year in Japan; one had spent an additional 10 months in Japan before going to university, and one had lived in Japan for three years after graduating.

Both the English/Nupe/Hausa trilingual learner and the 21 monolingual English learners have been exposed to Standard Japanese. The English/Nupe/Hausa trilingual learner and the less-experienced monolingual English learners have predominantly heard Standard Japanese in teaching materials, but have also been exposed to the L2 Japanese of other English-speakers, and perhaps also other Japanese dialects through friends or the media. The more-experienced participants are likely to have been exposed to both Standard Japanese and other dialects.

Segmentation into words

Before the accent types of the words produced by the English/Nupe/Hausa trilingual learner were identified, two decisions had to be made: a) what to count as a word, and b) whether to exclude any words due to downstep or deletion of their pitch accent.

For the purpose of this study, a word was defined as follows. Since final accent is not realised in isolated words, it was decided to judge words together with the function word that follows them, where there was one e.g., *mazu wa* 'first + TOPIC MARKER', *hanashi desu* 'speech + COPULA', *romanchikku na* 'romantic + ADJECTIVAL MARKER'. Words were judged with the following function word even on the three occasions when there was a pause between the word and the function word e.g., *rekishi...ga* 'history + SUBJECT MARKER'. In three cases, words could not be judged with the following function word. Firstly, *machimachi* 'diverse' was judged alone due to the presence of a filler, i.e. *machimachi...er...no* 'diverse + er + ADVERBIAL MARKER'. Secondly, *hitotsu desu* 'one + COPULA' was judged as two words as the copula was focussed, i.e. had its

own pitch accent. Thirdly, for the six examples where two function words followed a content word e.g., *itta toki ni* 'went + when + at' and *furansu kara no* 'France + from + POSSESSIVE MARKER', the content word was judged alone, and the two function words judged together. This is because the first function word can have its own accent in Standard Japanese – in the examples above both *toki* 'when' and *kara* 'from' would be finally accented. The following were judged as one word: compound nouns e.g., *kankou spotto* 'tourist spot' (n=5); verbs made up of noun plus *suru* 'to do' (n=3) e.g., *shoukai suru* 'to introduce'; however, phrasal verbs (n=2) e.g., *yatte(i)mashita* 'was doing' were judged as two separate words. Five words were excluded: three were one-syllable function words produced separately from other words, one was unclear, and one was a loanword, which was not produced with Japanese syllable structure (Taj Mahal).

Next it was decided whether to exclude any words due to their accents being deleted or compressed. In Standard Japanese, two accented words can be grouped together in one 'accentual' phrase, with the second accent deleted, or in one 'intonational' phrase, in which case the second accent is retained but compressed or 'downstepped' (Venditti, 2005:175). In Muradás-Taylor (2017) an attempt was made to exclude such contexts systematically but did not prove possible. Instead, words were included as long as it was possible to identify their accent type. However, returning to this data for the second time, some of the words which had been excluded also seemed to have audible accents. It was, therefore, decided to have the second judge listen to the words that had previously been excluded. To mitigate for any downstepping or deletion, the judges listened to words both in isolation but also in context. In three utterances, the accent type of the second component was deleted, but this was judged as 'accurate', as it is consistent with Standard Japanese: *sukina hitoni* 'to people who like', *mitanoha* 'saw + NOMINALISER + TOPIC MARKER', *ookina mononi* 'into a big thing'. A further 10 words which were excluded in Muradás-Taylor (2017) were included in the analysis, giving a total of 13 additional words. This had the advantage of not requiring additional criteria for exclusion on the basis of downstep/deletion. In principle, the inclusion of these additional words could have led to an over-estimation of accuracy, but a comparison of the results presented in this paper (section 3.1.1) and those in Muradás-Taylor (2017) shows that this has not occurred.

Note that segmenting into words and controlling for accent deletion or downstep was not necessary for the data from the monolingual English learners. Words were recorded in three contexts (e.g., *ame* 'rain', *ame da* 'it's rain' and *ame ga furu* 'rain falls') but judged as whole utterances. Downstep was not relevant as only the first word in each utterance was identified. (i.e. *ame* 'rain' in the examples above).

Accent type identification

Next the English/Nupe/Hausa trilingual learner's accent types were compared to Standard Japanese norms and coded as accurate or inaccurate, and identified as initially accented, finally accented, medially accented (including accent position) or unaccented. In total, the accent types of 129 words were identified (tokens, not types, as the stability of the accent types was also of interest). Fifteen words appeared more than once.

In Muradás-Taylor (2017), the learner's accent types were identified by only one judge, the L1 English L2 Japanese-speaking author of this paper. For this paper, to increase methodological rigour, the accent types were also identified by a second judge. The

second judge is from Tokyo, where Standard Japanese is spoken, and is a trained linguist, although not a phonetician. The second judge could identify the learner's accent types as being accurate or inaccurate quickly and with confidence in most cases. The two judges agreed on 121 words out of 129 and disagreed on 8 (percentage agreement 94%, Cohen's Kappa 0.66). Disagreements were resolved through discussion and seemed to be caused by non-Standard rhythm, an intonational rise obscuring the accent type, or uncertainty over which words in an inaccurate utterance were non-Standard. In one case the second judge considered a word to be inaccurate, but it was coded as accurate because the New Meikai Japanese accent dictionary (Kindaichi & Akinaga, 2001) contained an alternative form, which corresponded to the accent type that the learner used.

Although the second judge could confidently identify accent types as being accurate or inaccurate, they were not able to categorise the accent types that were produced as initially accented, medially accented, finally accented or unaccented. This is usual for L1 Japanese listeners (Goss & Tamaoka, 2015). Since the majority of words were produced with accurate Standard Japanese accent types (see Results), it was clear which accent types the learner produced. For the small number of words that were produced with inaccurate accent types (only 13 out of 129 words, see Results), the first judge (L1 English, L2 Japanese; phonetically-trained author of this paper) said aloud all possible versions of each word (e.g., *TAzuneta* or *taZUneta* or *tazuNEta* or *tazuneta*) and the two judges identified the learner's accent types in collaboration. Note that this discussion was only needed for a small proportion of words, and that it is only the accuracy/inaccuracy judgement, for which there is good inter-rater agreement, that is needed to calculate the percentage accuracy and stability, that are the focus of this paper.

Results

Accuracy and stability

The participant produced 129 words in total, of which 116 were produced with accurate Standard Japanese accent types and 13 were not, giving a percentage accuracy of 90%. Note that despite more consistent criteria for word segmentation, the inclusion of words whose accent types may have been downstepped/deleted, and the addition of a second judge, the percentage accuracy is consistent with Muradás-Taylor (2017) where it was reported as 91%.

Although not the main focus of the paper, a brief analysis was carried on the accent types of the accurate and inaccurate words. Of the 116 words that were produced accurately, 44 were unaccented, 33 were initially accented, 29 were medially accented, and 10 were finally accented. Of the inaccurate words, one involved accenting an unaccented word, four involved a change of accent position, and eight involved de-accenting accented words. No more specific pattern (e.g., finally-accented words consistently being produced unaccented, for example) was observed. Nine of the 11 inaccurate words contained a heavy syllable (moraic nasal, long vowel or long consonant) which means that the inaccurate words were more likely to contain a heavy syllable (69%) than the accurate words (43%). No trends were observed in the part of speech or mora number of the inaccurate words. The inaccurate words tended to cluster together – three utterances contained three inaccurate words e.g., in *ookina*

mono ni natteku no wa tanoshikatta 'it was fun that it became big', *ookina* 'big', *natte* 'become' and *tanoshikatta* 'it was fun' all had inaccurate accent types.

Sixteen words appeared more than once in the recording. Of these, 12 (75%) had the same accent type on each occasion. The accent types of these 12 words were all accurate. The percentage of words produced with accurate and stable accent types was, therefore, also 75%. The words produced with accurate and stable accent types included: initially-accented *indo* 'India' (n=3) and *ato* 'after' (n=3); medially-accented *tatemono* 'building' (n=2) and *ookii* 'big' (n=2); unaccented *sono* 'that' (n=4) and *itta* 'went' (n=2).

The use of a second judge and more consistent criteria for word segmentation and inclusion has led to the percentage of stable accent types being found to be lower than that reported in Muradás-Taylor (2017), which was 93% (previously, 1 unstable word out of 15; now, 4 unstable words out of 16). Note, however, that three of the words with unstable accent types were also repeated accurately: *nanka* 'like', was produced with accurate initial accent five times but was unaccented once; *suggoku* 'extremely' was produced with accurate medial accent twice but was unaccented once; and *tanoshikatta* 'it was fun' was accurately accented on *no* twice but on *ka* once.

Figure 1 compares the accuracy of the English/Nupe/Hausa trilingual learner (TRI) to the 21 monolingual English learners in Muradás-Taylor (in progress), where LE is the less-experienced group (n=13) and ME is the more-experienced group (n=8). From Figure 1, it is clear that the monolingual English learners all have low accuracy, with little variation between learners, and the English/Nupe/Hausa trilingual learner has high accuracy.

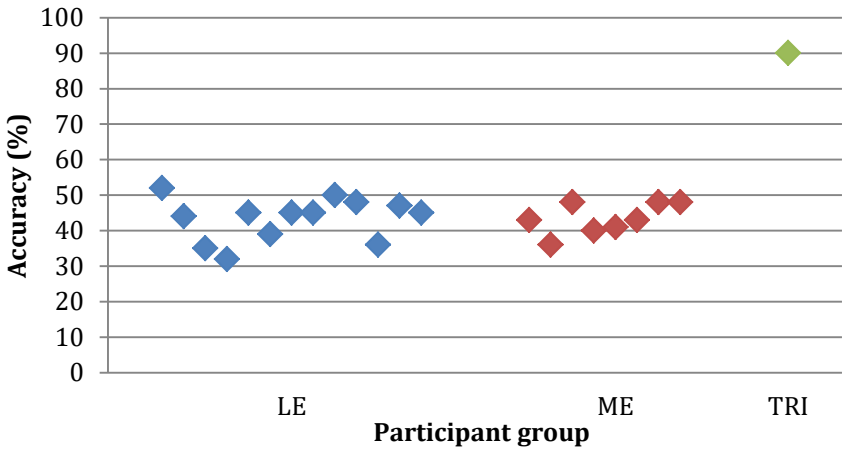

Figure 1. Accent type accuracy of the English/Nupe/Hausa trilingual learner compared to the monolingual English learners

Figure 2 compares the stability of the English/Nupe/Hausa trilingual learner to the 21 monolingual English learners in Muradás-Taylor (in progress). It can be seen that there is variation between the monolingual English learners, but that the English/Nupe/Hausa trilingual learner has the second highest stability of all the speakers.

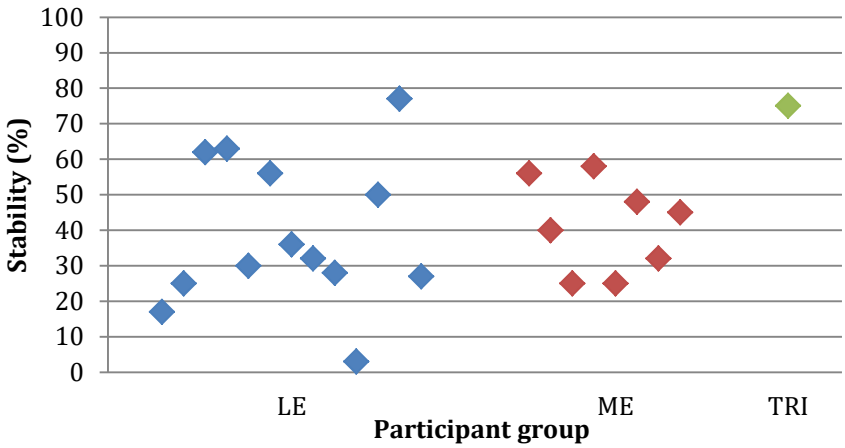

Figure 2. Accent type stability of the English/Nupe/Hausa trilingual learner compared to the monolingual English learners

Figure 3 compares the percentage of words produced with both accurate and stable accent types for the English/Nupe/Hausa trilingual learner compared to the 21 monolingual English learners in Muradás-Taylor (in progress). It is clear that the monolingual English learners have a low percentage of words produced with both accurate and stable accent types, and the English/Nupe/Hausa trilingual learner has a high percentage.

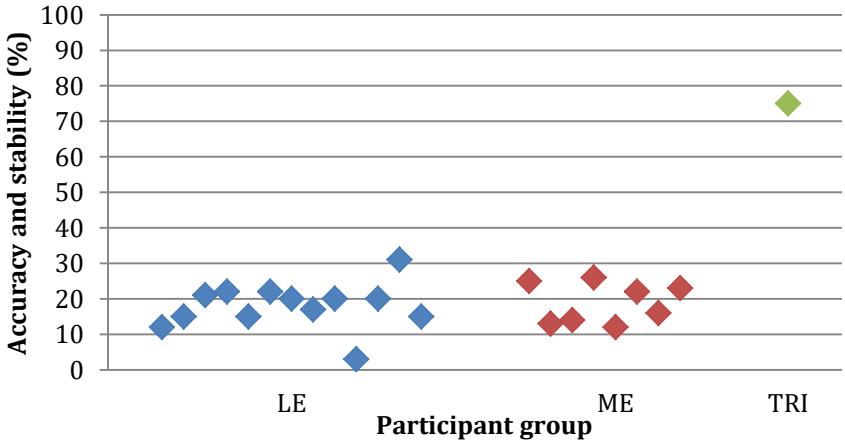

Figure 3. Accent type accuracy and stability of the English/Nupe/Hausa trilingual learner compared to the monolingual English learners

In sum, the English/Nupe/Hausa trilingual learner produced 90% of words with accurate accent types. Of the 16 words that appeared more than once in the recording, the majority (12, i.e. 75%) were repeated with accurate accent types. In fact, three of the four words that showed accent type instability were repeated accurately on other occasions. Overall, therefore, the English/Nupe/Hausa trilingual learner produced accent types that are highly accurate and stable. The percentage of accurate accent types contrasts strikingly with the monolingual English learners, whose accuracy ranged from 32% to 52% (mean 43%). The percentage of stable accent types for the monolingual English learners ranged from 3% to 77% (mean 40%). The

English/Nupe/Hausa trilingual learner's stability of 75% is toward the top of this range. However, it is qualitatively different. The monolingual English learners with high stability produce a single accent type – usually the penult – for most words in most contexts (Muradás-Taylor, in progress). In contrast, the English/Nupe/Hausa trilingual learner produces unaccented, initially-accented, medially-accented, and finally-accented words accurately and is mostly able to maintain the production of these accent types accurately in different contexts. This is reflected in the percentage of accurate and stable accent types, which is 75% for the English/Nupe/Hausa trilingual learner, but ranges from 3% to only 31% (mean 18%) for the monolingual English learners. It can be said, therefore, that the English/Nupe/Hausa trilingual learner produces Standard Japanese accent types that are accurate and stable, but none of the monolingual English learners do, even those who have completed four years of a Japanese degree including a year studying in Japan.

Acoustic analysis

Utterances were selected from data recorded in the previous studies (Muradás-Taylor, 2017 for the English/Nupe/Hausa trilingual learner; Taylor, 2012, for the monolingual English learners) and the phonetic correlates of pitch accent (F0 peak location, rate of F0 fall) were measured and compared to Standard Japanese norms. This study differs from previous research on the acoustic realisation of L1 English-speakers' L2 Japanese (Sakamoto, 2011) in selecting words for acoustic analysis that have already been identified as having a particular accent type. The scope of the acoustic analysis in the current study is, therefore, to determine whether words that have been identified as having certain accent types show a pitch shape (F0 peak position, rate of F0 fall) that is within Standard Japanese norms.

For the English/Nupe/Hausa trilingual learner it was decided to measure the F0 of two-syllable initially-accented, finally-accented, and unaccented words. In order to measure F0, words were restricted to those with a following particle (since the difference between final accent and unaccented is neutralised in isolation, Vance, 2008:144-145). Some words were not suitable for analysis because a pitch trace was not observable on Praat (Boersma & David, 2005). Only one example of each accent type remained: initially-accented *mazu wa* 'firstly', finally-accented *tsugi wa* 'next', and unaccented *sore ni* 'and also'.

For the monolingual English learners, only pure nouns were analysed, as the other lexical classes (verbs, derived nouns) do not have all three accent types in two-syllable words. Only words that were judged by all three judges as having correct Standard Japanese accent types were used. Of the 21 participants, only two speakers (one female, one male) produced at least one two-syllable initially-accented word, finally-accented word and unaccented word, which was judged by all three judges as having a correct Standard Japanese accent type. This is despite the study design being to elicit 12 words with initial accent in Standard Japanese, 12 with final, and 12 unaccented. The female speaker produced *chizu da* 'it's a map' with accurate initial accent, *kagi da* 'it's a key' with accurate final accent, and *kabe da* 'it's a wall' accurately unaccented. The male speaker produced *chizu da* 'it's a map' with accurate initial accent, *oto da* 'it's a sound' with accurate final accent, and *migi da* 'it's right' accurately unaccented. These tokens were therefore used in the analysis. None of the remaining 19 speakers produced three

words which were judged by all three judges as having different accent types, and therefore were excluded from further analysis.

Figure 4 shows initial accent produced by the English/Nupe/Hausa trilingual learner and the monolingual English learners. All show an F0 peak within the first vowel. For the English/Nupe/Hausa trilingual learner this occurs two thirds through the first vowel. For the English monolingual learners this occurs earlier: at the onset of the first vowel. All three speakers show a fall from this peak. The F0 of English/Nupe/Hausa trilingual learner falls 80 Hz in 180 ms to a low within the second vowel. The rate of fall is 0.44 Hz per ms. The F0 of the female English monolingual learner (centre) falls 80 Hz in 287 ms to a low within the third vowel. The rate of fall is 0.29 Hz per ms. The F0 of the male English monolingual learner (right) falls 27 Hz in 311 ms to a low within the third vowel. The rate of fall is 0.09 Hz per ms.

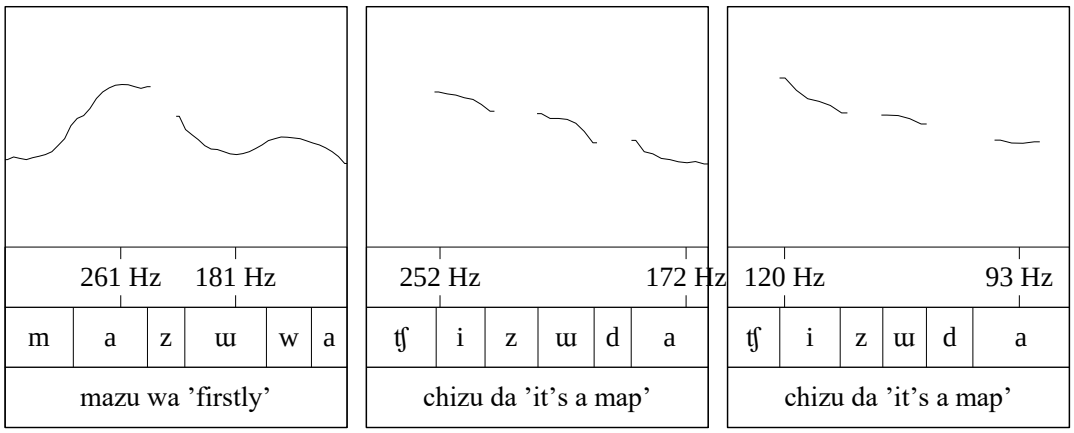

Figure 4. F0 of initial accent by the English/Nupe/Hausa trilingual learner (left) and monolingual English learners (centre and right)

Figure 5 shows final accent produced by the English/Nupe/Hausa trilingual learner and the monolingual English learners. The English/Nupe/Hausa trilingual shows an F0 peak just after the second vowel, i.e. within the labial velar approximant /w/ which follows the vowel. The monolingual English learners show an F0 peak within the second vowel: about a third from the end for the female speaker (centre) and a sixth from the end for the male speaker (right). All three speakers show a fall from this peak. The F0 of the English/Nupe/Hausa trilingual learner falls 133 Hz in 109 ms to a low within the third vowel. The rate of fall is 1.22 Hz per ms. The F0 of the female monolingual English learner (centre) falls 54 Hz in 116 ms to a low within the third vowel. The rate of fall is 0.46 Hz per ms. The F0 of the male monolingual English learner (right) falls 46 Hz in 144 ms to a low within the third vowel. The rate of fall is 0.32 Hz per ms.

Figure 6 shows unaccented words produced by the English/Nupe/Hausa trilingual learner and the monolingual English learners. The English/Nupe/Hausa trilingual learner shows a fall of 14 Hz from the first vowel to the alveolar flap /ɾ/ followed by a gradual rise of the same amount to the final vowel. The female monolingual English learner (centre) shows an F0 peak at the beginning of the first vowel which falls 27 Hz in 303 ms to a low in the third vowel. The rate of fall is 0.09 Hz per ms. The male monolingual English learner (right) shows an F0 peak at the beginning of the first vowel

which falls 28 Hz in 197 ms to a low after the end of the second vowel i.e. during the voiced alveolar stop /d/. The rate of fall is 0.14 Hz per ms.

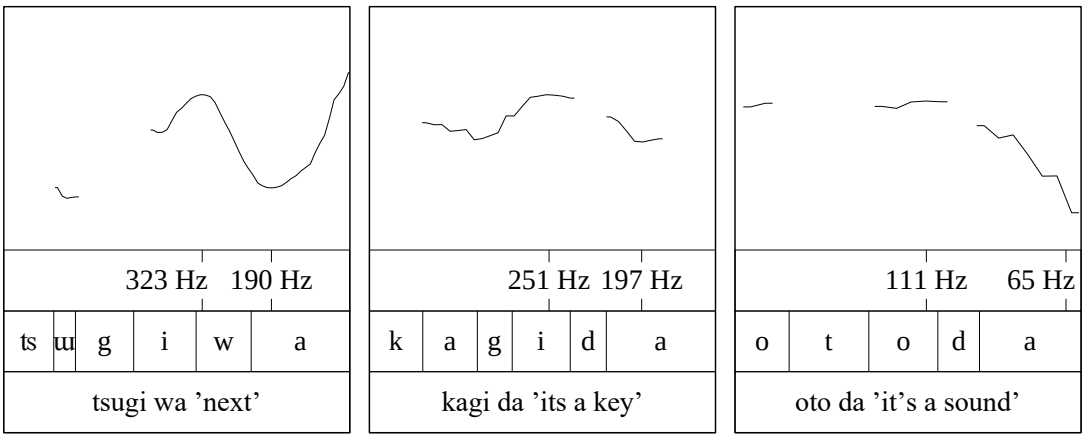

Figure 5. F0 of final accent by the English/Nupe/Hausa trilingual learner (left) and the monolingual English learners (centre and right)

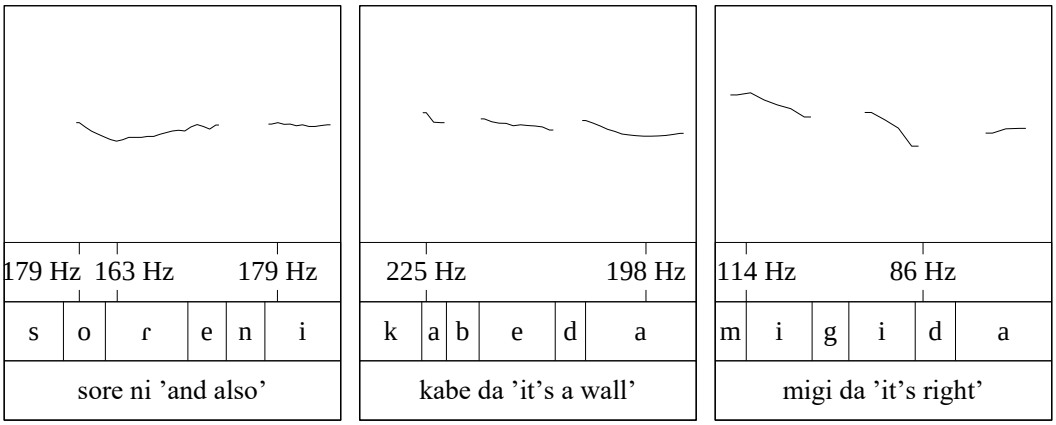

Figure 6. F0 of unaccented by the English/Nupe/Hausa trilingual learner (left) and the monolingual English learners (centre and right)

To what extent are these productions consistent with descriptions in the literature of Standard Japanese? For the limited data available, i.e. one token per accent type per speaker, it would appear that the English/Nupe/Hausa trilingual learner's productions are consistent with descriptions of Standard Japanese pitch accent, but the monolingual English learners' productions are not, at least for accented words.

For the English/Nupe/Hausa trilingual learner, F0 peak location and rate of F0 fall for the initially-accented and finally-accented words appear to be consistent with Standard Japanese. The initially-accented word has a F0 peak position two thirds through the first vowel, and the finally-accented word has an F0 peak position just after the second vowel. Standard Japanese is described as having a fall around the end of the accented syllable (Sakamoto, 2011). The rate of F0 fall is 0.44 Hz per ms for initial accent and 1.22 Hz per ms for final accent. This is considerably steeper than the monolingual English learners and therefore likely to be consistent with Standard Japanese which is

described as having a 'steep' slope. The unaccented word had no F0 fall word except for a 14 Hz dip around the alveolar flap /ɾ/. An aural impression of the pitch is of a rise from the first vowel to the second, so it seems likely that the initial 14 Hz fall in the first vowel is in fact an artefact of the preceding voiceless fricative /s/. Research with a greater number of tokens and L1 Japanese-speaking controls is needed to confirm this, but no obvious differences to descriptions of Standard Japanese norms are observable in the data available.

For the monolingual English learners, the F0 peak for the initially-accented words is at the onset of the first vowel. This is likely to be earlier than Standard Japanese, which is described as having a fall around the end of the accented syllable (Sakamoto, 2011). The rate of slope was 0.29 Hz per ms for initial accent and 0.46 Hz per ms for final accent for the female speaker; and 0.09 Hz per ms for initial accent and 0.32 Hz per ms for final accent for the male speaker. It is unclear from previous research if this is within Standard Japanese norms, but combined with the early peak it would seem likely that the slope is less steep than Standard Japanese norms. As for unaccented words, both speakers show a fall from the first vowel. For the female speaker this is less steep than her accented words (0.09 Hz per ms cf 0.29 for initial and 0.46 Hz per ms for final). Since unaccented words in Standard Japanese can also have a slight fall, but not a steep one (Sakamoto, 2011), this does not appear to be inconsistent with Standard Japanese norms. For the male speaker, there is a greater fall on the unaccented word (0.14, compared to 0.09 on initial accent). It is not clear why the judges in Taylor (2012) unanimously said that *chizu da* was initial and *migi da* was unaccented. Perhaps the fall on the middle vowel is partly an artefact of the voiced alveolar stop /d/ and not, in fact, representative of the pitch of the word as perceived aurally.

In summary, the F0 peak for the English/Nupe/Hausa trilingual learner is later than the monolingual English learners for initial accent, and the slope is steeper for initial and final accent. Since Standard Japanese has been described as having a late F0 peak and steep slope, it seems likely that the English/Nupe/Hausa trilingual learner, but not the monolingual English learners, are consistent with Standard Japanese norms for accented words. For unaccented words the picture is less clear: both the English/Nupe/Hausa trilingual learner and the female monolingual English learner appear to be within Standard Japanese norms. The male monolingual English learner is less clear, perhaps due to the effect of the voiced alveolar stop on the F0 of the second vowel.

The amount of data that could be analysed in this study is limited, and therefore these conclusions are necessarily tentative. To confirm that the English/Nupe/Hausa trilingual learner's F0 peak location and rate of F0 fall are within Standard Japanese norms, more data, and L1 Japanese-speaking controls are needed. For the monolingual English learners, perhaps the most important finding is that only 2 out of 21 speakers produced tokens that were judged as having three different accent types. This has implications for future research and will be returned to in the discussion.

Discussion

The English/Nupe/Hausa trilingual learner of Japanese produced Standard Japanese accent types that were highly accurate and stable. All possible accent types (initial, final,

medial, unaccented) were produced. Ninety percent of words had accurate accent types, and of the words that appeared more than once in the recording, 75% had accent types that were both accurate and stable. In addition, although only limited acoustic data could be obtained from the recording, the phonetic realisation of F0 peak location and rate of F0 fall also seemed to be consistent with Standard Japanese norms, with a late F0 peak and steep F0 fall for accented words. These findings are in striking contrast to the monolingual English learners of Japanese (Muradás-Taylor, in progress; Taylor, 2011a, b; Taylor, 2012) who only produced an average of 43% of words with accurate accent types (min 32% max 52%), 18% of words with accurate and stable accent types (min 3% max 31%), and, from the limited data available for acoustic analysis in the current study, may produce accented words with an earlier F0 peak and less steep slope than Standard Japanese norms.

This paper is significant because it demonstrates that it is, in fact, possible to acquire accurate and stable pitch accent. The English/Nupe/Hausa trilingual learner has acquired this level of Standard Japanese pitch accent despite never visiting Japan, only receiving 60 hours of Japanese language instruction over 24 weeks, studying Japanese in the UK with an English-speaking tutor, and not receiving any explicit instruction on pitch accent. The difficulty monolingual English learners have acquiring Standard Japanese pitch accent cannot, therefore, be easily dismissed as due to lack of instruction or input. Instead it requires a linguistic explanation, i.e. the influence of the English/Nupe/Hausa trilingual learner's other L1s Nupe and/or Hausa.

A possible counter-argument to this claim is that the English/Nupe/Hausa trilingual learner appears to be a particularly good language learner, who, for example, picked up Hindi from neighbours. The English/Nupe/Hausa trilingual learner also had considerable exposure to Japanese culture before starting formal study, perhaps casting doubt on the claim that their Standard Japanese input is limited. However, many of the monolingual English learners in Taylor (2012) also expressed a keen interest in Japanese culture, and yet the highest accuracy of any learner was only 52%. This returns us to the original argument that it is most likely to be their differing L1s that explains the difference between the English/Nupe/Hausa trilingual learner and the monolingual English learners.

What property of the English/Nupe/Hausa trilingual learner's L1s has aided this acquisition? Nupe and Hausa are both tonal. Nupe has five tones: high, mid, low, rise, fall; Hausa has three: high, low, and a fall derived from a sequence of high and low (Yip, 2002). Neither have a pitch accent system like Japanese where only one (or no) syllables per word have an accent. Further research is needed to determine whether it is the influence of the Nupe, the Hausa, the combination of the two, or even the combination of one or both with English that has enabled this speaker to acquire Japanese pitch accent so easily. Further research is needed also to determine whether the speaker has acquired a pitch accent system with one syllable specified for accent per word, or whether all syllables in the speaker's Japanese are specified for tone, and whether this is observable in the speaker's Japanese. It is this question which should drive future acoustic research, as well as verifying, with more data and L1 Japanese controls, whether F0 peak location and rate of F0 fall are within Standard Japanese norms.

The most significant finding regarding the acoustic analysis for the monolingual English learners was how little data was available for analysis. Sakamoto (2011) reported that

some words produced by English-speaking learners had an early peak, some a late peak, and some no pitch fall. But it was not clear if these realisations were due to learners producing an inaccurate accent type, or inaccurate phonetic realisation of the correct accent type. To avoid this problem, the decision was made in this study to only analyse data from words that had been judged by all three judges in Taylor (2012) as having a particular accent type. This meant that as many as 19 of the 21 learners had to be excluded from analysis because they did not produce at least one token of each of initial accent, final accent and unaccented, despite reading aloud 12 words of each accent type during data collection. The acoustic analysis on the remaining two learners suggested that there was some overlap with Standard Japanese norms, albeit with perhaps an earlier F0 peak and less steep slope. However, this overlap cannot be taken as evidence that pitch accent has been acquired by these speakers, since none of the monolingual English learners produced more than 31% of words with accurate and stable accent types. With hindsight, there is little meaning in asking how monolingual English learners of Japanese realise the categories of a pitch accent system that they clearly do not have. Instead of pursuing further research with more data and L1 Japanese speaking controls, it would seem that a different approach is needed. Anecdotally, L1 English speakers' L2 Japanese production seems to show less influence of stress over time. Is it possible that monolingual English learners of Japanese do not acquire a pitch accent system, but do learn to suppress their English stress cues? Acoustic research has shown individual variation in the duration of accented syllables, with some but not all, learners lengthening the accented syllable (Kondo, 2012; Ueyama, 2012). Perhaps a longitudinal study investigating the possible suppression of cues to stress other than pitch might be a more meaningful avenue for research in the future.

This paper shows that L2 Japanese learners' Standard Japanese pitch accent accuracy and stability varies considerably depending on their L1. Accuracy and stability can be very low, like monolingual English learners, or very high, like this English/Nupe/Hausa trilingual. This is noticeably different from the L2 acquisition of English stress, where learners produce the stress patterns of real words with high accuracy, even when analysis of nonce word production shows us that their knowledge of the relation between stress and syllable number and lexical class differs from L1 speakers. This has been shown for L1 speakers of Spanish (Guion, Harada, & Clark, 2004) which has stress, but also for L1 speakers of Korean and Thai (Guion, 2005; Wayland, Guion, Landfair, & Li, 2006), which do not. Muradás-Taylor (in progress) argues that monolingual English learners do not learn Standard Japanese pitch accent even after years of Japanese study because of the relatively low functional load of Standard Japanese pitch accent (Kitahara, 2001) and the dialectal variation in words' accent types and accentual systems (Kubozono, 2012). It would seem that learners acquire Japanese pitch accent if their L1 facilitates this (which Nupe and/or Hausa appear to), but not if their L1 doesn't, like English.

To summarise, this paper demonstrates, through a case study of an English/Nupe/Hausa trilingual learner of Japanese, that the L2 acquisition of Standard Japanese pitch accent is possible. In comparison, it is concluded that monolingual English learners of Japanese do not acquire Standard Japanese, not because of limited input or instruction, but because of cross-linguistic influence from their L1. Further research is needed into Nupe and/or Hausa speakers' acquisition of Japanese pitch accent. For English monolingual learners of Japanese it is argued that, since they do not

have a pitch accent system, it is not meaningful to ask how each category of this system is realised; instead, an alternative avenue of enquiry is suggested regarding the possible inhibition of English stress cues.

References

Beckman, M.E. & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P.A. Keating (ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology III* (pp. 7-33). Cambridge, UK: Cambridge University Press.

Boersma, P. & David, W. (2005). *Praat: doing phonetics by computer* [Computer program]. Retrieved from http://www.praat.org/

Cabrelli Amaro, J. (2012). L3 phonology: An underestimated domain. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (eds.), *Third language acquisition in adulthood* (pp. 33-60). Amsterdam, Netherlands: John Benjamins.

Flege, J. (2009). Give input a chance! In T. Piske & M. Young-Scholten (eds), *Input matters in SLA* (pp. 175-190). Bristol, UK: Multilingual Matters.

Fry, D.B. (1958). Experiments in the perception of stress. *Language and speech, 1*(2), 126-151.

Goss, S. & Tamaoka, K. (2015). Predicting lexical accent perception in native Japanese speakers: An investigation of acoustic pitch sensitivity and working memory. *Japanese Psychological Research, 57*(2), 143-154.

Guion, S.G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition, 27*, 503-533.

Guion, S.G., Clark, J.J., Harada, T., & Wayland, R.P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words*. Language and Speech, 46*(4), 403-426.

Guion, S.G., Harada, T., & Clark, J.J. (2004). Early and late Spanish–English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition, 7,* 207-226.

Kindaichi, H. & Akinaga, K. (eds.) (2001). *Shin Meikai nihongo akusento jiten [New Meikai Japanese accent dictionary].* Tokyo: Sanseido.

Kitahara, M. (2001). *Category structure and function of pitch accent in Tokyo Japanese*. PhD Thesis, Indiana University.

Kondo, M. (2007). Acoustic realization of lexical accent and its effects on phrase intonation in English speakers' Japanese. *Proceedings of the 16th International Conference on Phonetic Sciences* (pp. 1649-1652).

Kubozono, H. (2012). Varieties of pitch accent systems in Japanese. *Lingua*, *122*(13), 1395-1414.

Muradás-Taylor, B. (2017). Japanese pitch accent in an English/Nupe/Hausa trilingual. In E. Babatsouli (ed.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2017* (pp. 199-205). ISBN: 978-618-82351-1-3. URL: http://ismbs.eu/publications-2017

Muradás-Taylor, B. (in progress). Accuracy and stability in English speakers' production of L2 Japanese pitch accent.

Sakamoto, E. (2011). *Investigation of factors behind foreign accent in the L2 acquisition of Japanese lexical pitch accent by adult English speakers*. PhD Thesis, University of Edinburgh.

Sluijter, A.M.C. & van Heuven, V.J. (1996a). Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of the Fourth International Conference on Spoken Language Processing*, *2*, 630-633.

Sluijter, A.M.C. & van Heuven, V.J. (1996b). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*(4), 2471-2485.

Taylor, B. (2011a). Do English learners of Japanese produce isolated nouns with Standard Japanese lexical accent? *Second Language, 10,* 15-31.

Taylor, B. (2011b). Variability and systematicity in individual learners' Japanese lexical accent. *Poznan Studies in Contemporary Linguistics,*,*47*(1), 146-158.

Taylor, R.L. (2012). *Eigo washa ni yoru nihongo no go akusento no shuutoku [The acquisition of Japanese lexical accent by English speakers]*. PhD Thesis, Nagoya University.

Thomson, R.I. & Derwing, T.M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*, 326-344.

Ueyama, M. (2012). *Prosodic transfer: An acoustic study of L2 English and L2 Japanese.* Bologna: Bononia University Press.

Wayland, R., Guion, S.G., Landfair, D., & Li, B. (2006). Native Thai speakers' acquisition of English word stress patterns*. Journal of Psycholinguistic Research*, *35*, 285-304.

Yamada, N. (1994). Nihongo akusento shuutoku no ichidankai — gaikokujin gakushuusha no baai [A stage in acquiring Japanese accent: The case of foreign learners]. *Journal of Japanese Language Teaching*, *83*, 108-120.

Vance, T. (2008). *The sounds of Japanese.* Cambridge, UK: Cambridge University Press.

Venditti, J. J. (2005). The J_TOBI model of Japanese intonation. In S. Jun (ed.), *Prosodic typology and transcription: A unified approach* (pp. 172-200). Oxford, UK: Oxford University Press.

Yip, M. (2002). *Tone.* Cambridge, UK: Cambridge University Press.