

Muradás-Taylor, Becky ORCID

logoORCID: <https://orcid.org/0000-0001-7275-6016> (2021)

Accuracy and Stability in English Speakers' Production of Japanese Pitch Accent. *Language and Speech*, 65 (2). pp. 377-403.

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/5391/>

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version:

<https://journals.sagepub.com/doi/full/10.1177/00238309211022376>

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. [Institutional Repository Policy Statement](#)

# RaY

Research at the University of York St John

For more information please contact RaY at [ray@yorks.ac.uk](mailto:ray@yorks.ac.uk)

# Accuracy and Stability in English Speakers' Production of Japanese Pitch Accent

Language and Speech

1–27

© The Author(s) 2021



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/00238309211022376

[journals.sagepub.com/home/las](https://journals.sagepub.com/home/las)**Becky Muradás-Taylor** 

York St John University, UK

## Abstract

Standard Japanese uses pitch accent to distinguish words such as initially-accented *hashi* “chopsticks” and finally-accented *hashi* “bridge.” Research on the second language acquisition of pitch accent shows considerable variation: in accuracy scores in identification, in different dominant accent types in production, and in the unstable accent types of repeated words. This study investigates pitch accent production in English-speaking learners of Japanese, asking how accuracy and stability vary (a) with amount of Japanese experience and (b) between learners. Two groups of learners (13 less experienced; 8 more experienced) produced 180 words in three contexts (e.g., *ame* “rain,” *ame da* “it’s rain,” and *ame ga furu* “rain falls”). Three Japanese phoneticians identified the accent types of the words that the learners produced. The results showed no difference in accuracy or stability between the two groups and little inter-learner variation in accuracy: all had low accuracy. Although some learners had relatively high stability, they did not maintain accent type contrasts across contexts. These results suggest that first language English speakers do not encode pitch accent in long-term memory, raising questions for future research and language teaching.

## Keywords

prosody, suprasegmental, acquisition, experience, individual difference

## Introduction

Second language (L2) phonology research tells us that adult users of an L2 acquire aspects of sound systems that differ from their first language (L1): more-experienced L2 users differ from less-experienced users in their production of vowel quality and quantity, voice onset time (VOT), syllable structure, and the use of phonological rules such as flapping of alveolar stops (Zampini, 2008). However, certain aspects of L2 sound systems, such as the contrast between English /ɹ/ and /l/ for L1 Japanese speakers, are difficult to acquire without intensive training (Bradlow et al.,

### Corresponding author:

Becky Muradás-Taylor, York St John University, Lord Mayor's Walk, York, YO31 7EX, UK.

Email: [b.muradas-taylor@yorksj.ac.uk](mailto:b.muradas-taylor@yorksj.ac.uk)

1999) or years of residence in an L2-speaking country (Flege et al., 1995). For particularly difficult contrasts, speakers with more L2 experience do not always differ from speakers with less experience: see, for example, Larson-Hall (2006) for /ɹ/ and /l/; Levy and Strange (2008) for /u/ and /y/; Pallier et al. (1997) for /e/ and /ɛ/; and Dupoux et al. (2008) for lexical stress. This paper considers the L2 production of Standard Japanese pitch accent by L1 English-speaking learners of Japanese—anecdotally difficult to acquire—comparing learners with more/less Japanese experience.

L2 phonology research often reports large individual differences. These are well-studied regarding global foreign accent (Piske et al., 2001) and perception (e.g., Goss, 2015), but less well-studied regarding the production of particular segments or prosody. Instead, findings are often reported as group means, giving an impression of “average behaviour which belies [. . .] important inter-learner differences” (Munro & Derwing, 2015, p. 31). Previous research on English speakers’ L2 acquisition of Japanese pitch accent has reported individual differences for perception, with some learners able to identify accent types with high accuracy (Hirano-Cook, 2011; Nishinuma et al., 1996; Shport, 2011), and production, with some learners frequently accenting the penultimate syllable, some the antepenultimate syllable, and others producing most words unaccented (Kuno, 1998; Taylor, 2011b). However, it is not known whether some learners’ accent types are more accurate than others, conforming more closely to Standard Japanese norms.

Previous research also suggests that L1 English speakers’ L2 Japanese accent types are unstable, with Yamada (1994, p. 118) reporting that one learner produced the word *yappari* “as I thought” with three different accent types in one conversation. This aligns with a key difference between the prosody of Japanese and English: Japanese has pitch accent which is realized as a pitch fall (Vance, 2008), whereas English has stress, and stressed syllables can have a variety of pitch shapes depending on the phrase level intonation (Pierrehumbert, 1980). Acquiring Japanese pitch accent involves not just acquiring words’ pitch accent types, but producing these stably in different contexts. It is not known how stable learners’ accent types are, or how stability varies with experience or between learners.

This study investigates the accuracy and stability of accent types produced by L1 English-speaking learners of L2 Japanese, comparing a less-experienced group ( $n = 13$ ) and a more-experienced group ( $n = 8$ ), differing in hours of Japanese instruction and time spent in Japan. Learners read aloud Japanese words ( $n = 180$ ) 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 types that the learners produced.

The accuracy and stability of both the less-experienced and more-experienced groups, and of individual learners, are analyzed, showing that L1 English speakers’ L2 Japanese pitch accent is inaccurate and unstable for all learners, even for those with more Japanese experience. Previous research has shown both large differences in individuals’ ability to identify different accent types (Goss, 2015; Hirano-Cook, 2011; Nishinuma et al., 1996; Shport, 2011) and large differences in the accent types that learners use in production (Kuno, 1998; Taylor, 2011b); this study shows that no learners produce accurate and stable accent types. The data presented in this paper illustrates how pitch accent manifests in individual learners’ speech, and suggests that L1 English-speaking learners of L2 Japanese do not encode pitch accent in long-term memory.

This study complements research showing that pitch accent is a robust predictor of degree of perceived foreign accent in L2 Japanese speech (Idemaru et al., 2019). It also offers a different perspective on the acquisition of L2 prosody from research on stress or tone. Research on stress has investigated learners’ knowledge of stress rules (Archibald, 1992; Pater, 1997) or statistical regularities in stress placement (Guion, 2005; Guion et al.; Wayland et al., 2006), mainly eliciting nonce words, presumably because real words show ceiling effects. Research on Chinese tone has investigated learners’ production patterns when either mimicking an L1 model or reading aloud

**Table 1.** Accent types of two-syllable and three-syllable words.

Syllable number	Initial accent	Medial accent	Final accent	Unaccented
2	<i>ame da</i> high–low–low “it’s rain”		<i>inu da</i> low–high–low “it’s a dog”	<i>hoshi da</i> low–high–high “it’s a star”
3	<i>hanabi da</i> high–low–low–low “it’s a firework”	<i>wagashi da</i> low–high–low–low “it’s a traditional sweet”	<i>sashimi da</i> low–high–high–low “it’s raw fish”	<i>sakana da</i> low–high–high–high “it’s a fish”

words with annotated tone labels (Hao, 2012). In contrast, this study investigates the accuracy and stability of pitch accent in learners’ production of real words without a model or pitch accent markings.

Note that “accurate” is used in this paper as shorthand for a correspondence with Standard Japanese norms. It is not intended to imply that L2 users of Japanese *should* produce Japanese a particular way. Instead, Standard Japanese is used as a reference to describe how the learners use pitch in their spoken Japanese. In addition, “stable” is used to refer to accent types that are unaffected by context, whether or not they are accurate.

### 1.1 Pitch in Japanese and English

Pitch accent varies with dialect in Japanese (Hirayama, 1998; Kubozono, 2012). Standard Japanese, which is based on the Tokyo variety (Kubozono, 2012), is referred to as “Japanese” in this section. “English” is used here as shorthand for General American or Standard Southern British English, which share properties regarding pitch and stress.

Japanese accent types are differentiated by the presence or absence of an accent and, where there is one, its position. This paper describes accent types as “initial accent,” “medial accent,” “final accent,” and “unaccented.” Words of  $n$  syllables can have  $n + 1$  possible accent types: initial, final, and unaccented for two-syllable words, and initial, medial, final, and unaccented for three-syllable words (Kindaichi & Akinaga, 2001, appendix, p. 10; Kubozono, 2008, p. 168; Vance, 2008, p. 154). The accent, where there is one, is realized by a sharp pitch fall. Initially-accented and medially-accented words are differentiated by the position of the pitch fall—compare the high–low–low pitch pattern of initially-accented *hanabi* “firework” with the low–high–low pitch pattern of medially-accented *wagashi* “traditional sweet.” The difference between final accent and unaccented is neutralized in isolation: final accent is only realized when a word is followed by a copula (e.g., *da*) or function word (e.g., the topic marker *wa* or subject marker *ga*; Vance, 2008, pp. 144–145). Thus, finally-accented *sashimi* “raw fish” and unaccented *sakana* “fish” both have the pitch pattern low–high–high with no pitch fall in isolation, but *sashimi da* “it’s raw fish” has a pitch fall on *da* (i.e., low–high–high–low) and *sakana da* “it’s fish” does not (i.e., low–high–high–high). All possible accent types for two-syllable and three-syllable words followed by *da* are listed in Table 1.

Japanese pitch accent shares some properties with English stress. Both can distinguish words, as in the Japanese pair *hashi* “chopsticks” and *hashi* “bridge,” and the English pair (*an*) *object*, which has stress on the first syllable, and (*to*) *object*, which has stress on the second. Unlike languages such as Polish which have fixed stress, the accented syllable in Japanese and the stressed syllable in English both vary between words: Japanese has initially-accented *hanabi* “firework,” medially-accented *wagashi* “traditional sweet,” and finally-accented *sashimi* “raw fish”; English has initially-stressed *Canada*, medially-stressed *banana*, and finally-stressed *kangaroo*.

However, there are differences between Japanese pitch accent and English stress. First, there are words in Japanese that are unaccented (Vance, 2008, pp. 144–145). Unaccented words have no sharp pitch fall, such as the word *sakana* “fish” with the pitch pattern low–high–high. In contrast, no multisyllabic word in English is lexically “unstressed”: there is no equivalent to unaccented words such as *sakana* in Japanese.

Second, Japanese pitch accent and English stress have different acoustic correlates. Japanese pitch accent is realized only with F0, perceived as a sharp fall in pitch from the accented syllable (Vance, 2008, p. 143). For example, the initially-accented word *hashi* “chopsticks” has the pitch pattern high–low. In contrast, English stress has multiple acoustic correlates: stressed and unstressed syllables differ in duration and in “spectral balance,” which correlates well with listeners’ impression of loudness (Sluijter & van Heuven, 1996a, 1996b); vowels in unstressed syllables are reduced (e.g., the different vowel quality in the first syllable of the noun *object* and the verb *object*); and voiceless stops in stressed syllables are more aspirated than in unstressed syllables (Beckman, 1986, p. 162). In addition, when words in English receive phrase-level prominence (also called “focus” or, confusingly, “accent” or “pitch accent”) they are realized with a pitch movement and a boost in intensity on the stressed syllable (Sluijter & van Heuven, 1996a, 1996b).

Third, pitch in Japanese, but not in English, has lexical function (Beckman, 1986; Beckman & Pierrehumbert, 1986; Sluijter & van Heuven, 1996a, 1996b). This can be illustrated by taking the English word *happy* and pronouncing it with a rise, making the question *happy?*; in contrast, if you take the Japanese word *hashi* “chopsticks,” which has a high–low pitch pattern, and pronounce it with rising intonation starting from the first syllable, you get *hashi* “bridge” (low–high). Although stressed syllables can have a pitch movement in English, this is not always a fall. In the examples given by Pierrehumbert (1980, pp. 7–8), stressed syllables can be associated with a high tone, a fall from a high to a low tone, a rise from a low to a high tone, or a low tone; combined with “phrase accents” and “boundary tones” these could be used, respectively, as the answer to a question, when calling out to someone, to convey incredulosity, and as a question (Pierrehumbert, 1980). This is why producing the word *hashi* with falling pitch gives “chopsticks” and rising pitch gives “bridge,” but producing the word *happy* with rising pitch just changes the intonation.

Pitch is sometimes described as a correlate of English stress, usually with reference to Fry (1958). This has been described as a “common misunderstanding” (Beckman & Edwards, 1994, p. 13). Instead, stressed syllables in English are “docking sites” for phrase-level prominence (Sluijter & van Heuven, 1996b, p. 2471). Although pitch is not a correlate of stress, English-speaking listeners do use pitch as a cue to stress where that information is available, for example, in distinguishing the noun *object* and the verb *object* in citation form (Fry, 1958). And there is evidence that the F0 of stressed syllables is slightly different from that of unstressed syllables even in the absence of phrase-level prominence (Plag et al., 2011). However, this does not change the fact that pitch is an unreliable correlate of stress in English—there is no one-to-one mapping between pitch shape and stress position, in the way that accented syllables have a pitch fall in Japanese.

The different function of pitch in Japanese and English is key to this study. Japanese accent types are stable: initially-accented words have a pitch fall on the initial syllable, medially-accented words have a pitch fall on the medial syllable, finally-accented words have a pitch fall on the final syllable (except in isolation where they have no pitch fall), and unaccented words have no pitch fall. This contrasts with English, where, although stress position is generally stable, pitch shape is not, because it is determined post-lexically whether stressed syllables have high, low, falling or rising pitch.

## 1.2 English speakers' perception of Japanese pitch accent

L1 English speakers' perception of Japanese pitch accent is task dependent. They have high accuracy on discrimination tasks, such as AX tasks, where participants judge whether the second word has the same accent type as the first (Hirano-Cook, 2011), and perform as well as L1 Japanese speakers on ABX tasks, where participants judge whether the third word has the same accent type as the first or second (Sakamoto, 2011). However, they are less accurate than L1 Japanese speakers on identification tasks, where participants identify words' accent types (Hirano-Cook, 2011; Nishinuma et al., 1996; Sakamoto, 2011; Shport, 2011), or on tasks where participants judge whether words' accent types are "correct" (Goss, 2015; Shibata & Hurtig, 2008).

L1 English speakers' ability to perceive Japanese pitch accent does not improve consistently with increased Japanese experience. Goss (2015, p. 35) argues that "accent perception does not develop in parallel with proficiency level." Hirano-Cook (2011) reports an increase in accuracy rate with proficiency on an identification task, but of the five participant groups in her study, only the first and the third had accuracy that differed significantly (Goss, 2015, p. 37; Hirano-Cook, 2011, p. 34–35). Sakamoto (2011) reports a difference between more-experienced and less-experienced learners on both a categorical perception task and an identification task; however, the majority (10 out of 16) of her less-experienced learners perceived pitch accent categorically (Sakamoto, 2011; see also Goss, 2015, p. 38). On a task where participants judged whether words' accent types were "correct," Shibata and Hurtig (2008) found no difference between novice, intermediate, and advanced learners, with even advanced learners performing no better than chance.

In summary, English-speakers' perception of Japanese pitch accent (a) depends on the task and (b) does not show a strong correlation with Japanese experience. Importantly, considerable individual variation is observed in learners' identification of Japanese accent types. Mean accuracy rates for low- and high-scoring groups varied as much as 36%–79% in Hirano-Cook (2011, p. 46), 46%–75% in Shport (2011, p. 176), and 42%–73% in Nishinuma et al. (1996, p. 647). Even among L1 English speakers with no experience of Japanese, accuracy on an identification task ranged from 27% to 90%, with some people outperforming the highest scoring L1 Japanese listeners (Shport, 2016, p. 23). Clearly, some L1 English speakers can identify Japanese accent types with high accuracy.

## 1.3 English speakers' production of Japanese pitch accent

Research on English speakers' production of Japanese pitch accent has investigated L1 speakers of American English (Kuno, 1998; Toki, 1980; Ueyama, 2012), Australian English (Yamada, 1994; Yoshimitsu, 1981), and Standard Southern British English (Taylor, 2011a, 2011b, 2012a, 2012b). These findings are reported here as examples of English-speaking learners of Japanese.

Conflicting findings have been reported regarding accuracy. Yoshimitsu (1981) reported high accuracy: 84%–100%, 83%–95%, and 68%–93% for three learners across lexical classes. However, Yoshimitsu (1981, p. 67) pointed out that the two learners with the highest accuracy for verbs (94% and 99%) predominantly used the *masu* form which has the same accent type for all verbs. At the other end of the scale, Yamada (1994) reported that her participants only acquired between 2% and 4.5% of words' accent types, defining "acquired" as used frequently with the Standard Japanese accent type, not affected by "interlanguage strategies," and unaffected by a model accent. Toki (1980) did not report a percentage accuracy, but few of his transcriptions of learners' pitch patterns correspond to Standard Japanese accent types.

Yoshimitsu (1981), Yamada (1994), and Toki (1980) collected data as connected speech, which means that accent deletion or compression (Beckman & Pierrehumbert, 1986, pp. 264/272;



Venditti, 2005, p. 175) may have obscured some words' accent types. Kuno (1998) avoided this by having learners read isolated words but did not report accuracy. Taylor (2011a) analyzed a subset of the data reported in the current study (13%), focusing on isolated pure nouns. Accuracy ranged from 33% to 66% depending on syllable number and accent type, with initially-accented two-syllable nouns showing the highest accuracy. However, this was due to chance matches between the learners' production and Standard Japanese—two-syllable nouns that are initially-accented in Standard Japanese were not produced with more initial accent than two-syllable nouns that are unaccented in Standard Japanese. For three-syllable nouns there was some evidence of acquisition of initial accent, medial accent and unaccented for less-experienced learners, and also for final accent for more-experienced learners. Taylor (2011a) suggested that the acquisition of pitch accent for three-syllable but not two-syllable nouns could be because three-syllable nouns have a contrast in accent position (initial vs. medial) whereas isolated two-syllable nouns only contrast in accent-edness (initial vs. unaccented). The latter has no parallel in English since no English words are lexically specified to have no stress. However, accuracy was still low for three-syllable words, with only initially-accented words exceeding 50%. Overall, we do not have a clear picture of how accuracy varies with experience or between learners.

Instability of the learners' accent types has also been reported. As well as Yamada (1994, p. 118) who described a learner producing the word *yappari* with three different accent types, Toki (1980, pp. 86–87) reports that, when reading aloud a text containing two instances of the phrase *ome ni kakaru* “meet,” all 20 learners used a different pitch pattern the second time. Toki (1980, p. 87) suggested that this was a context effect, with accent type affected by surrounding words, but did not investigate further.

There are some suggestions of context effects in previous research. The first is in Yamada (1994). She categorized learners' accent types as A, B, C, or O, where A, B, and C refer to accent position, and O is unaccented; she also used the combination categories OA, OB, OC, etc. for phrases containing two content words. This use of combination categories may imply that when two content words are produced together, the first is unaccented and the second accented. For example, in the phrase *hotondo no gakusei* “most students,” *hotondo no* “most” would be unaccented and *gakusei* “students” would be accented (Yamada, 1994, p. 113). However, it is not clear whether there is evidence for this “unaccented first word” pattern; we do not know whether words before another content word are more likely to be unaccented than other words in L1 English speakers' Japanese.

Another potential context effect was observed in a study by the current author on L1 English speakers' production of final accent (Taylor, 2012a). Since final accent in Standard Japanese is not produced in isolation, the finally-accented word *otoko* “man” is unaccented in isolation but finally accented before *da* (i.e., *otoKO da* “it's a man”). However, the learners produced combinations of accent types not seen in Standard Japanese, such as unaccented in isolation but medially accented before *da* (e.g., *otoko* “man,” *oTOko da* “it's a man”), and medially accented in isolation but finally accented before *da* (e.g., *oTOko, otoKO da*). It is not known whether this phenomenon is limited to words with final accent in Standard Japanese, or whether appending a function word affects the accent type of learners' words more generally.

Most importantly, we do not know how common this instability is—to what extent do L1 English-speaking L2 learners of Japanese maintain different accent types stably in different contexts? This is of interest because it aligns with a key difference between Japanese pitch accent and English stress, namely that Japanese pitch accent is always realized with a pitch fall (Vance, 2008), whereas stressed syllables in English can have a variety of pitch shapes (Pierrehumbert, 1980).

Individual variation has also been reported in previous research. Kuno (1998) showed that some learners most frequently accent the penultimate syllable, some the antepenultimate syllable, and

some most frequently produce words unaccented. Yoshimitsu (1981) reported that two learners tended to accent unaccented words and one learner to de-accent accented words. Taylor (2011b) analyzed a subset (49%) of the current data—pure nouns and verbs in isolation and before a function word—and showed that some learners produce more initially-accented words, some produce more unaccented words, and some produce a mixture with no dominant accent type. In addition, Taylor (2011b) also claimed that the relation between accent type and other factors (word length, lexical class, context) varied between learners. However, it is not known whether learners vary in accuracy or stability.

Nor do we have a clear sense of whether accuracy or stability increase with experience. None of the above production studies (i.e., Kuno, 1998; Toki, 1980; Yamada, 1994; Yoshimitsu, 1981) compared groups of learners with different amounts of Japanese experience. The only exception is Taylor (2011a), which considered pure nouns in isolation, and found a small difference between more-experienced and less-experienced learners in three-syllable but not two-syllable nouns.

There are two further factors that influence stress placement in English, and therefore may influence English-speaking learners' accent types in Japanese: lexical class and syllable structure. Horiguchi (1973) predicted where English speakers might stress Japanese words according to English stress rules, comparing learners' productions with her predictions. The learners' productions sometimes matched and sometimes did not match the predictions, with no clear pattern emerging. We now know that English stress placement is affected by multiple, competing, probabilistic factors, including the stress patterns of phonologically similar words (Guion et al., 2003). The effect of lexical class and syllable structure is beyond the scope of the current study—different lexical classes are included only to elicit the full range of possible accent types, and the effect of syllable structure is minimized by considering only words containing light (CV) syllables.

Lastly, let us turn to research which investigates learners' phonetic realization of Japanese accent types (Kondo, 2007; Sakamoto, 2011; Ueyama, 2012). In Sakamoto (2011), L1 English learners of Japanese imitated non-words produced by L1 Japanese speakers with different accent types, and the learners' accent types were identified by L1 Japanese speakers. The percentage of words identified as intended was higher for the more-experienced learners, who had had more Japanese instruction and spent more time in Japan. An acoustic analysis showed differences between L1 and L2 Japanese speakers in pitch peak location and degree of pitch fall that were more pronounced for the less-experienced learners. In Kondo (2007) and Ueyama (2012), English-speaking learners of Japanese produced disyllabic words with either the first or second syllable accented. Some learners were found to lengthen the accented syllable, similar to a stressed syllable in English, and others did not, like the L1 Japanese speakers; however this did not correlate with the learners' proficiency (Ueyama, 2012, pp. 51–53).

In summary, individual variation has been reported in which accent types learners use, and experience effects have been observed in learners' phonetic realization of Japanese accent types when imitating an L1 Japanese model. However, it is not known how accuracy and stability vary between learners or with experience in learners' production of real words.

## *1.4 Research questions*

This paper aims to answer the following questions regarding the pitch accent of L1 English-speaking learners of L2 Japanese:

- (1) What percentage of words are produced with accurate accent types according to Standard Japanese norms?



- (2) What percentage of words are produced with stable accent types across different contexts, specifically, in isolation, before a function word, and before a content word?
- (3) What percentage of words are produced with accurate and stable accent types?
- (4) How do the above percentages vary with less/more Japanese experience, as measured by hours of Japanese instruction plus time spent in Japan?
- (5) How do the above percentages vary between learners?

## 2 Method

Two groups of English-speaking learners of Japanese (less experienced  $n = 13$ ; more experienced  $n = 8$ ) read aloud 180 Japanese words in three contexts. Japanese phoneticians ( $n = 3$ ) identified the accent type of each word that they produced.

### 2.1 Participants

The participants were L1 speakers of Standard Southern British English. They were either enrolled on, or had graduated from, an undergraduate degree course in Japanese, consisting of three years of study at a British university and one year at a university in Japan.

The less-experienced group ( $n = 13$ ) had completed one or two years of their course. They had received an average of 250 hours of Japanese instruction (minimum 70 hours, maximum 430 hours, standard deviation  $SD$  90). Seven of the less-experienced participants had never been to Japan and none had stayed more than three months.

The more-experienced group ( $n = 8$ ) had received more Japanese instruction, averaging 970 hours (minimum 640 hours, maximum 1400 hours,  $SD$  320). They had spent at least a year in Japan. One had also spent 10 months in Japan before going to university. Most were about to graduate at the time of the study; however, one had graduated and lived in Japan for three years.

All the participants will have been exposed to Standard Japanese, which is used in formal situations throughout Japan. The less-experienced participants will have predominantly heard Standard Japanese in audio-visual teaching materials, but will also have been exposed to others' L2 Japanese and perhaps also Japanese dialects. The more-experienced participants are likely to have heard both Standard Japanese and other dialects. The areas of Japan that the more-experienced participants had lived in were Tokyo ( $n = 3$ ), Nagoya ( $n = 3$ ), both Tokyo and Kagoshima ( $n = 1$ ), and Kumamoto ( $n = 1$ ). The accentual system of the dialects spoken in Tokyo and Nagoya are that of Standard Japanese, with some lexical exceptions (Hirayama, 1998, p. 129); the dialects spoken in Kagoshima and Kumamoto have different accentual systems to Standard Japanese (Hirayama, 1998, p.133). When asked if they considered their Japanese accent to be standard, seven of the more-experienced participants answered "yes" (including the two participants who had lived in Kagoshima and Kumamoto), and one participant answered "don't know," but had lived in Tokyo, so is likely to have been exposed predominantly to Standard Japanese.

The participants will have received little instruction on pitch accent. Textbooks aimed at English speakers do not usually mark words' accent types (Shport, 2008) and accent instruction is usually limited to the occasional pointing out of homonyms (Shport, 2008, p. 166; see also Goss, 2018).

### 2.2 Words

The words varied in length (two or three syllables), lexical class (pure noun, derived noun, or verb), and Standard Japanese accent type (initial accent, medial accent, final accent, or unaccented). "Derived noun" refers to a noun derived from a verb, such as *hanashi* "speech" from *hanasu* "to speak"; "pure noun" refers to all other nouns.

**Table 2.** The 15 word types.

Syllable number	Initial accent	Medial accent	Final accent	Unaccented
2	Pure nouns Verbs		Pure nouns Derived nouns	Pure nouns Derived nouns Verbs
3	Pure nouns	Pure nouns Verbs	Pure nouns Derived nouns	Pure nouns Derived nouns Verbs

The three lexical classes were chosen because they have different accent types in Standard Japanese. Pure nouns have no restrictions on accent type: two-syllable pure nouns can be initially accented, finally accented, or unaccented, and three-syllable pure nouns can be initially accented, medially accented, finally accented, or unaccented. In general, derived nouns are finally accented or unaccented (Kindaichi & Akinaga, 2001, appendix, p. 12) and verbs are penultimately accented or unaccented (Kindaichi & Akinaga, 2001, appendix, p. 49). Although exceptions exist, words with other accent types were not included in this study. In total, 12 words of 15 different word types (e.g., initially-accented two-syllable pure nouns, unaccented three-syllable derived nouns etc.) were used, as shown in Table 2. The full list of words ( $n = 180$ ) is in the Appendix.

Words were selected using the following criteria. They were restricted to those listed in the New Meikai accent dictionary (Kindaichi & Akinaga, 2001) as having a single accent type. Loan words from English were excluded. Words containing heavy (bimoraic) syllables were excluded: long vowels, consecutive vowels, geminate consonants, and moraic nasals. Words with a high vowel (/i/ or /u/) between voiceless consonants (/k/, /s/, /ʃ/, /t/, /tʃ/, /ts/ etc.) were excluded to avoid vowel devoicing (Vance, 2008, p. 206). This was because words such as *ashita* /aʃi̥ta/ “tomorrow” might be perceived by English listeners as containing a consonant cluster, giving the impression of a heavy syllable, and also because any vowel devoicing would affect accent identification.

Most words (73%) were selected from the beginners’ textbooks *Minna no Nihongo* 1 and 2 (3A Network, 1998). For some word types, additional words were needed because fewer than 12 words from the textbooks met all the criteria. This affected medially-accented and finally-accented three-syllable pure nouns, which are less common than initially accented and unaccented ones (Sugito & Tahara, 1989), and derived nouns. Effort was made to choose words that the learners would know, such as *itoko* “cousin” and *yogore* “dirt,” but less common words such as *shimi* “stain” and *aseri* “haste” were also included.

The words’ accent types are assumed not to be predictable (Vance, 2008, p. 155; see also Kubozono, 2008). In a few cases, the Standard Japanese accent type could potentially be deduced from the words themselves—for example, verbs ending in *bu* (e.g., *tobu* “fly”) tend to be unaccented in Standard Japanese, and those ending in *tsu* (e.g., *motsu* “hold”) tend to be accented (Kindaichi & Akinaga, 2001, appendix, p. 52). However, such trends are rare and not widely known.

## 2.3 Contexts

Table 3 illustrates the three contexts for each of the lexical classes.

The “before a function word” context was formed as follows. For pure and derived nouns, the function word was the copula *da* “is.” For verbs, two function words were used: *node* “because” and *hodo* “as far as” or “as much as.” In Standard Japanese, unaccented verbs take final accent before certain function words, such as *node*, and are unaccented before others, such as *hodo*

**Table 3.** The three lexical classes in the three contexts.

Lexical class	In isolation	Before a function word	Before a content word
Pure noun (e.g., <i>ame</i> “rain”)	<i>ame</i> “rain”	<i>ame da</i> “it’s rain”	<i>ame ga furu</i> “rain falls”
Derived noun (e.g., <i>tsugi</i> “next”)	<i>tsugi</i> “next”	<i>tsugi da</i> “it’s next”	<i>tsugi ni iku</i> “I go next”
Verb (e.g., <i>iku</i> “go”)	<i>iku</i> “I go”	<i>iku node</i> “because I go” <i>iku hodo</i> “as far as I go”	<i>iku noga sukida</i> “I like to go” <i>iku kotoga dekiru</i> “I can go”

(Kindaichi & Akinaga, 2001, appendix, pp. 74–75; Vance, 2008, pp. 154–173). Both *node* and *hodo* were therefore used to elicit all possible Standard Japanese accent types for verbs: final accent for unaccented verbs before *node*, unaccented for unaccented verbs before *hodo*, initial accent for two-syllable accented verbs before *node* or *hodo*, and medial accent for three-syllable accented verbs before *node* or *hodo*. The function word following the nouns was one-syllable long (*da*) and those following the verbs were two-syllables long (*node/hodo*); this is true of many utterances in Japanese and could not be avoided in this study.

The “before a content word” context was formed as follows. For pure and derived nouns, the content words were chosen for meaning; they were between two and four syllables long and contained no heavy syllables. The function words (*ga* and *ni* in the examples) are required by the grammar. For verbs, it was necessary to append two different expressions—*noga sukida* “like” and *kotoga dekiru* “can”—to elicit all Standard Japanese accent types. Unaccented verbs take final accent before *noga suki da* “like” and are unaccented before *kotoga dekiru* “can.” Here, *suki* “like” and *dekiru* “can” are the content words, and *noga* and *kotoga* are function words necessary to nominalize the verb.

In Standard Japanese accent types are stable in these contexts: initially-accented words, for example, have initial accent in isolation, before a function word, and before a content word. As final accent is not realized in isolation, “stable” final accent is unaccented in isolation and finally accented before a function word or content word.

## 2.4 Data collection

The participants read aloud the words and phrases from the cards shown in Figure 1. At the top was the word or phrase in Standard Japanese orthography: a mixture of *kanji* (Chinese characters) and *kana* (Japanese syllabary). Beneath that the word or phrase was written in *kana* only. At the bottom was an English translation of the word.

A total of 13,356 utterances were elicited. Since two versions of the “before a function word” and “before a content word” context were needed for verbs, 636 words and phrases were elicited from each of the 21 participants: 180 isolated words, 228 words before a function word, and 228 words before a content word.

The recording (bit depth 16 bits, sample rate 44.1 kHz) was carried out in phonetics laboratories of UK universities. Each participant read the cards in a different order. The cards were divided into eight sets. Each took about five minutes to read aloud, and the participants took a break between sets. The recording lasted approximately 90 minutes per participant including breaks. No financial reward was given for taking part.

<p>雨</p> <p>あめ</p> <p>雨=rain</p>	<p>雨だ</p> <p>あめだ</p> <p>雨=rain</p>	<p>雨が降る</p> <p>あめがふる</p> <p>雨=rain</p>
----------------------------------	------------------------------------	----------------------------------------

**Figure 1.** Sample cards.

Precautions were taken to elicit intonation that was as neutral as possible. Dummy cards were inserted into each set: two at the beginning to avoid the effect of any initial hesitation, and three at the end to avoid any final intonation. Each card appeared only once, although the participants were free to repeat words if they hesitated or misread them. The researcher, who was present, occasionally requested that a word be repeated if the vowels/consonants produced did not match those on the card. If the participants misread a word and repeated it with emphasis on the corrected portion, the researcher asked them to say it again.

Some utterances (147 out of the total of 13,356) were discarded because they contained a pronunciation error, making them unrecognizable as the target word, or were missing from the recording. The remaining 99% ( $n = 13209$ ) were successfully elicited.

## 2.5 Accent type identification

The learners' accent types were identified by phonetically-trained Japanese speakers. Given that the participants have English as their L1, they may manipulate acoustic correlates other than pitch. Duration, for example, is known to affect L1 Japanese listeners' perception of pitch accent in L2 users' speech (Kondo, 2007, p. 1651). Since the cumulative effect of pitch peak location, pitch slope, pitch range, duration, and any other correlates is unknown, L1 Japanese judges were used. Phonetically-trained judges were necessary because, although untrained listeners are able to recognize whether or not words' accent types are accurate, they are unable to identify accent types (Goss & Tamaoka, 2015). Accent identification took place in 20 one-hour sessions. The accent type of each word was identified by three judges. Due to the time commitment involved, six judges were used in rotation.

The judges heard each utterance twice, and either indicated that the word was unaccented, or marked the position of the accent. They were provided with a hiragana transcription of the word plus any immediately following function word—with *ame ga furu* "rain falls," for example, transcribed as *ame ga*—but heard the whole of the utterance, including all function and content words. They marked the accent with the symbol ー. In the "before a content word" context, such as *ame ga furu* "rain falls," the accent type of the content word (i.e., *furu* "falls"), may undergo deletion or compression (Beckman & Pierrehumbert, 1986; Venditti, 2005); this does not affect the results as it was only the accent type of the target word, (i.e., *ame* "rain"), that was judged.

The judges were instructed to listen for the lexically relevant information—a sharp pitch fall—and to ignore any word-initial or word-final boundary tones. Japanese has a rise at the beginning of each "accentual phrase" (Venditti, 2005, p. 175). For example, the word *wagashi* "traditional sweet," spoken in isolation, has the pitch pattern low–high–low with an initial rise followed by fall. However, the phrase *kono wagashi* "this traditional sweet" has the pitch pattern low–high high–high–low; *wagashi* here has no initial rise. This is because the rise is part of the accentual phrase; only the fall is part of the lexically-specified accent type. For the current study, words with a

low–high–low pitch pattern and a high–high–low pitch pattern were both judged as having medial accent, as were words with a rise to a following function word, resulting in a low–high–low–high pitch pattern (e.g., for *wagashi da*). However, where the pitch pattern of a three-syllable word before a function word was high–low–high–low with two falls, the judges marked two accents. Twelve practice tokens were used to reach a consensus on how to implement these instructions before commencing.

Each word was considered to have the accent type that two out of three, or all three, judges had identified. Some words were excluded because each judge identified a different accent type ( $n = 402$ ), or they were judged to have two accents ( $n = 5$ ). The accent types of a total of 12,802 words were identified; this was 97% of the 13,209 words that were elicited (and 96% of the total 13,356 words).

Considering that the words were produced by L2 speakers, known to use acoustic correlates that differ from L1 speakers (Kondo, 2007; Sakamoto, 2011; Ueyama, 2012), the inter-rater reliability was high. In total, 69% of words showed agreement between all three judges, and 97% showed agreement between at least two judges. Krippendorff's alpha was calculated as a measure of inter-rater reliability as it is suitable for nominal judgements (i.e., the categories "initial," "medial," "final," and "unaccented"), can handle missing data, and can handle a design which is not fully crossed, with different words rated by different subsets of judges (Hallgren, 2012). In addition, the percentage agreement and Cohen's kappa (Hallgren, 2012) were calculated for each pair of judges. Krippendorff's alpha was 0.69 overall, and the pairs of judged ranged from 71% agreement (Cohen's kappa 0.61) to 82% agreement (Cohen's kappa 0.75). Given that L1 Japanese listeners have been found to identify L1 Japanese speakers' accent types with only 61% accuracy (Goss & Tamaoka, 2015), the inter-rater reliability is high, presumably because the judges were phonetically trained.

## 2.6 Coding for accuracy and stability

The data was coded as follows. The target word in each utterance was coded as having an accent type which did or did not correspond to Standard Japanese. The construct definition of "accuracy" used in this paper is therefore a phonological one; it does not necessarily imply a syllable-by-syllable match with Standard Japanese including the initial rise.

Each target word was coded as having a "stable" or "unstable" accent type, irrespective of accuracy. Words with initial accent in isolation, before a function word, and before a content word were coded as stable, as were words with medial accent in all three contexts, or unaccented in all three contexts. Words that were unaccented in isolation and finally accented before a function or content word were also coded as having stable accent, because final accent is not realized in isolation. Words with any other combination of accent types (e.g., initial accent in isolation, unaccented before a function word, and finally accented before a content word) were coded as unstable.

Each target word was coded as either being, or not being, accurate *and* stable, that is, conforming to Standard Japanese across all three contexts.

# 3 Results

## 3.1 Accuracy and stability by group

The percentage of utterances where the target word was produced with accurate Standard Japanese accent types was 43% for the 13 less-experienced learners. It was also 43% for the 8 more-experienced learners. A generalized linear mixed effect model was fitted to the data using the package lme4 (Bates et al., 2015) in R (R Core Team, 2019) to investigate the relationship

between accuracy and experience. The model included random intercepts to account for variation in learners' and words' accuracy (Westfall et al., 2014; Winter, 2020). Including a fixed effect for experience did not improve the model fit according to a likelihood ratio test,  $\chi^2(1) = 0.0003$ ,  $p = 0.99$ , telling us that the more-experienced learners did not produce more words with accurate accent types than the less-experienced learners.

The percentage of words produced with stable accent types (i.e., the same in all three contexts), was 40% for the less-experienced learners and 41% for the more-experienced learners. A generalized linear mixed effect model, with random intercepts for learners and words, was fitted to the data to investigate the relationship between stability and experience. Including a fixed effect for experience did not improve the model fit according to a likelihood ratio test,  $\chi^2(1) = 0.21$ ,  $p = 0.65$ , telling us that the more-experienced learners did not produce more words with stable accent types than the less-experienced learners.

The percentage of words produced with accurate and stable accent types (i.e., Standard Japanese accent type in all contexts), was 18% for the less-experienced learners and 19% for the more-experienced learners. Generalized linear mixed effect modelling showed that the more-experienced learners did not produce more words with accurate and stable accent types than the less-experienced learners,  $\chi^2(1) = 0.22$ ,  $p = 0.64$ .

As no difference was found between the groups, the percentages are reported here for the whole group of 21 learners: 43% of words were produced with accurate accent types, 40% with stable accent types, and 18% with accurate and stable accent types.

Post hoc analysis was carried out to explore whether the low accuracy and stability were caused by fatigue during data collection, where the learners read aloud eight sets of words. Generalized linear mixed effect modelling showed that there was no effect of set number on accuracy,  $\chi^2(1) = 0.30$ ,  $p = 0.58$ : later sets did not have lower accuracy than earlier sets. In addition, words which appeared in all three contexts late in the recording ( $n = 512$ ) were no more unstable (59.8%) than words which appeared in all three contexts early in the recording ( $n = 528$ , 59.7%). The low accuracy and stability are properties of English speakers' Japanese pitch accent, not an effect of fatigue.

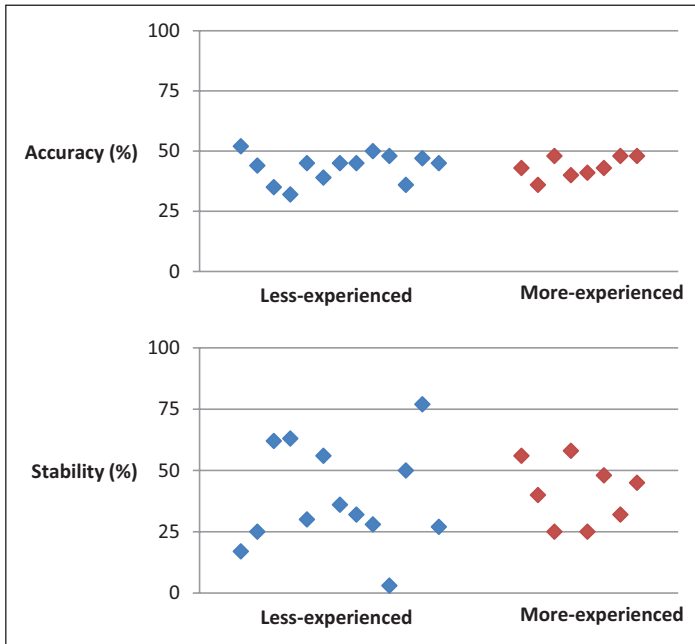
Further post-hoc analysis was also carried out to investigate the effect of including words not in beginners' textbooks. Generalized linear mixed effect modelling showed that finally-accented and unaccented words had lower accuracy than initially-accented and medially-accented words, but including a fixed effect for "textbook" (i.e., whether or not words were in the beginners' textbooks) did not improve the model fit,  $\chi^2(1) = 1.30$ ,  $p = 0.25$ . This tells us that the learners' low accuracy and stability is not due to the inclusion of words that do not appear in beginners' textbooks.

Finally, a reviewer of this paper suggested that the use of human judges could have led to lexical bias: "*ame* in isolation triggers the perception of both 'rain' (high-low accent) and 'candy' (unaccented) [. . .] because there is no context, but listeners will be more biased to hear 'rain' in the context with the verb *furu* 'fall.' This would not be the case when there is no counterpart for a word, such as *hima* 'free time' (unaccented): because there is no *hima* with an accent, [. . .] listeners will be biased to hear the word correctly both in isolation and in context." Post-hoc generalized linear effect mixed modeling was carried out to compare the stability of two-syllable pure nouns with and without homophones of different accent types. Including a fixed effect for "homophone" (i.e., whether or not words have an accent-opposed homophone), did not improve the model fit,  $\chi^2(1) = 0.49$ ,  $p = 0.48$ , telling us that the learners' low stability is not due to lexical bias.

### 3.2 Accuracy and stability by learner

The percentage of utterances where the target word was produced with an accurate accent type showed little variation between learners, ranging from 32% to 52% (mean 43%,  $SD$  5). The percentage of words produced with stable accent types varied between learners, ranging from 3% to





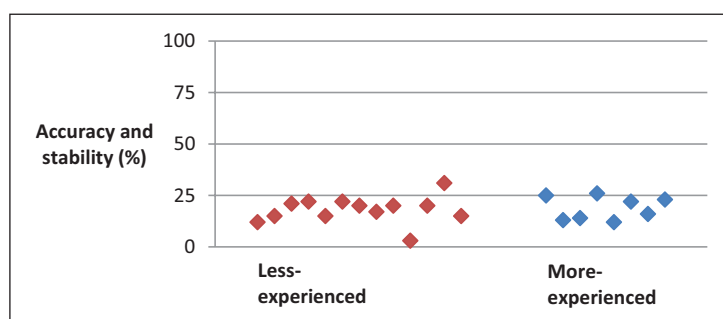
**Figure 2.** Inter-learner variation for accuracy and stability.

77% (mean 40%,  $SD$  18). The percentage of words produced with accurate and stable accent types ranged from 3% to 31% (mean 18%,  $SD$  6). Figure 2 shows the large variation between learners for stability and small variation between learners for accuracy.

Two thirds of the learners have stability below 50%, producing more words with unstable accent types than they do stable accent types. Of the 21 learners, 20 produced at least one word (minimum 4, maximum 20, mean 9,  $SD$  5), with a different accent type in each context. Learner ME03 produced the noun *hako* “box” with initial accent in isolation (*Hako*), unaccented before a function word (*hako da* “it’s a box”), and final accent before a content word (*haKO wo akeru* “open the box”). Learner LE02 produced the verb *nomu* “drink” with initial accent in isolation (*NOMu*), final accent before a function word (*noMU hodo* “as much as I drink”), and unaccented before a content word (*nomu koto ga dekiru* “I can drink”). The take-home message regarding stability is that all the learners had unstable accent types, ranging from one in four words unstable (23%) to nearly all words (97%).

Figure 3 illustrates the small variation between learners in the percentage of words that are both accurate and stable. It is low for all learners: no learner produced more than a third of all words with accurate and stable accent types (max. 31%). Although this was included as a separate research question, the variation between learners can be attributed almost entirely to the difference in stability, since the correlation between the percentage of stable words and the percentage of accurate and stable words is strong ( $r = 0.9$ ) and the correlation between the percentage of accurate words and the percentage of accurate and stable words is weak ( $r = -0.25$ ).

Table 4 details the accuracy, stability, and accent types of each learner. It shows that, for three-syllable words, medial accent was the most frequent accent type for many learners; unaccented was the most frequent for some, and no learner had initial or final accent as their most frequent accent type. For two-syllable words, initial accent was the most frequent accent type for many learners,



**Figure 3.** Inter-learner variation for the percentage of words that are both accurate and stable.

**Table 4.** Accuracy, stability, and accent types of each learner.

Learner	Accuracy (%)	Stability (%)	Percentage of each accent type (%)	
			Two-syllable words	Three-syllable words
LE01	52	17	un 48 in 45 fin 7	un 49 med 29 in 19 fin 3
LE02	44	25	in 46 fin 41 un 13	med 68 fin 20 un 7 in 5
LE03	35	62	in 83 un 12 fin 5	med 69 in 26 fin 2 un 2
LE04	32	63	in 93 fin 6 un 1	med 49 in 45 fin 3 un 3
LE05	45	30	un 52 in 42 fin 6	un 62 med 30 in 5 fin 2
LE06	39	56	in 95 un 3 fin 2	med 52 in 26 un 18 fin 5
LE07	45	36	in 73 un 21 fin 6	med 42 un 30 in 22 fin 5
LE08	45	32	in 72 un 17 fin 11	med 51 un 29 in 15 fin 5
LE09	50	28	in 62 un 33 fin 5	un 49 med 31 in 11 fin 9
LE10	48	3	un 45 in 43 fin 12	med 55 un 42 in 1 fin 1
LE11	36	50	in 67 fin 27 un 7	med 83 in 8 fin 7 un 2
LE12	47	77	un 88 in 10 fin 2	un 94 med 5 fin 1 in 0
LE13	45	27	in 68 un 25 fin 7	med 55 un 24 in 18 fin 3
ME01	43	56	in 91 un 8 fin 1	med 54 un 23 in 22 fin 2
ME02	36	40	in 82 un 11 fin 8	med 54 in 31 un 12 fin 3
ME03	48	25	un 43 in 32 fin 26	med 48 un 36 fin 9 in 7
ME04	40	58	in 86 un 10 fin 4	med 54 in 35 un 8 fin 2
ME05	41	25	in 65 un 31 fin 4	med 50 un 27 in 22 fin 2
ME06	43	48	in 82 un 15 fin 3	med 56 un 24 in 15 fin 5
ME07	48	32	in 72 un 21 fin 7	med 35 un 32 in 26 fin 7
ME08	48	45	un 73 in 17 fin 11	un 70 med 19 fin 8 in 4

in = initial accent; med = medial accent; fin = final accent; un = unaccented.

unaccented for some, and none had final accent as their most frequent accent type. A tendency to accent words on the penultimate syllable—the medial syllable for three-syllable words and initial syllable for two-syllable words—therefore emerges from the data.

Table 5 reanalyzes the learners' accent types as antepenultimate accent, penultimate accent, final accent, and unaccented. This illustrates that some learners have a dominant accent type that they use more frequently than other accent types—penultimate accent for fourteen learners and

**Table 5.** Dominant accent types.

Dominant penultimate accent		Dominant unaccented	
Learner	Percentage of each accent type (%)	Learner	Percentage of each accent type (%)
LE02	pen 58 fin 30 un 10 ante 3	LE01	un 48 pen 37 ante 10 fin 5
LE03	pen 76 ante 14 un 7 fin 3	LE05	un 58 pen 36 fin 4 ante 3
LE04	pen 70 ante 24 fin 4 un 2	LE12	un 91 pen 7 fin 1 ante 0
LE06	pen 72 ante 13 un 11 fin 3	ME08	un 71 pen 18 fin 9 ante 2
LE07	pen 57 un 26 ante 11 fin 6	No dominant accent type	
LE08	pen 61 un 23 ante 8 fin 8		
LE11	pen 76 fin 16 ante 4 un 4	Learner	Percentage of each accent type (%)
LE13	pen 61 un 24 ante 9 fin 5	LE09	pen 45 un 42 fin 7 ante 6
ME01	pen 72 un 16 ante 11 fin 1	LE10	pen 50 un 43 fin 6 ante 1
ME02	pen 67 ante 17 un 11 fin 5	ME03	pen 40 un 39 fin 17 ante 4
ME04	pen 69 ante 19 un 9 fin 3		
ME05	pen 57 un 29 ante 12 fin 3		
ME06	pen 68 un 20 ante 8 fin 4		
ME07	pen 53 un 26 ante 13 fin 7		

ante = antepenultimate accent; pen = penultimate accent; fin = final accent, un = unaccented.

unaccented for four (three learners are considered not to have a dominant accent type because the most frequent accent type is within 10 percentage points of the next).

Table 5 shows considerable variation between learners in the frequency of each accent type. Learners who share a dominant accent type do not necessarily have a second most frequent accent type in common—some learners with dominant penultimate accent, for example, have unaccented as their next most frequent accent type, some have antepenultimate accent, and two (LE02, LE11) have final. The proportion of unaccented varied between learners from as low as 2% (LE04) to as high as 91% (LE12); the proportion of penultimate accent varied from as low as 7% (LE12) to as high as 76% (LE03, LE11). Looking back at Table 4 we can see that the proportion of three-syllable words with initial (antepenultimate) accent varied from less than 2% (LE10, LE12) to as high as 45% (LE04). Additional analysis showed that the proportion of words before a function word or content word having final accent varied from as low as 2% (LE12, ME01) to as high as 42% (LE02).

Although the variation in accuracy between learners was small (minimum 32%, maximum 52%, mean 43%, *SD* 5), a comparison of Tables 4 and 5 shows that learners who have dominant unaccented (LE01, LE05, LE12, ME08) or no dominant accent type (LE09, LE10, ME03) all have above average accuracy (45% or over). This is likely to be an artefact of the data set elicited, because the Standard Japanese accent types are not evenly distributed, with more unaccented (42%) than other accent types (23% final, 21% initial, 15% medial). This distribution occurred because derived nouns and verbs only have two accent types, one of which is unaccented. The take-home message regarding accuracy remains, therefore, that inter-learner variation is minimal, with all learners having low accuracy.

The data in Tables 4 and 5 can be analyzed further to shed light on the inter-learner variation in stability. The learner with the highest stability of 77% (LE12) produces 91% of words unaccented. In fact, there is a strong positive correlation ( $r = 0.87$ ) between stability (Table 4) and the percentage of the most frequent accent type for each learner (Table 5), showing that learners with high stability have a strongly dominant accent type. In contrast, there is a negative correlation ( $r = -0.56$ ) between stability and accuracy. It will be argued in the discussion that the strongly dominant accent

type and slightly lower accuracy in learners with high stability imply that no learners can be considered to have acquired the ability to maintain different accent types stably across different contexts.

## 4 Discussion

### 4.1 Accuracy and stability

The first research question asked what percentage of words were produced with accurate accent types, conforming to Standard Japanese norms. The results showed that 43% were accurate; over half (57%) were produced with accent types not corresponding to Standard Japanese. The learner with the highest accuracy only reached 52%. The learners' accuracy is lower than that reported by Yoshimitsu (1981), which ranged from 68% to 100% across lexical classes. This is probably because some accent types in Yoshimitsu (1981) were obscured by accent deletion or compression in connected speech (Beckman & Pierrehumbert, 1986; Venditti, 2005), and because the participants in Yoshimitsu (1981) predominantly used the *masu* form of verbs which all have the same accent type.

The second research question asked what percentage of words were produced with stable accent types across three different contexts: in isolation, before a function word, and before a content word. The results showed that 40% of the words' accent types were stable; the majority of words (60%) showed at least one accent type change. Even for the learner with the highest stability (77%), nearly a quarter of all words were unstable. This paper's focus on stability was motivated by the description of a learner who produced the word *yappari* with three different accent types in one conversation (Yamada, 1994, p. 118). Such instability is clearly prevalent in L1 English-speaking learners, a finding consistent with our understanding of the post-lexical function of pitch in English.

The third research question asked what percentage of words were produced with both accurate and stable accent types. The result was 18%. This is higher than 2%–4.5%, which was the percentage of words that Yamada (1994) considered her participants to have acquired, defining "acquired" as used frequently with the Standard Japanese accent type, not affected by "interlanguage strategies," and unaffected by a model accent. The figure of 18% includes chance matches with Standard Japanese—producing most words unaccented or with penultimate accent leads to accurate accent type for some words. Because of the uneven distribution of accent types in the data set, it is not possible to compare the percentage of accurate and stable words to chance. However, it is clear that this study supports Yamada (1994) in showing little acquisition of Standard Japanese accent types.

To acquire Standard Japanese pitch accent, you must learn which words take which accent type and maintain these in different contexts. As acquisition occurs you might expect learners to become both more accurate and more stable. However, learners with higher stability had slightly less accurate accent types. Instead, stability correlated with the percentage of the most frequent accent type ( $r = 0.87$ ). This means that learners with high stability had a dominant accent type that they produced irrespective not only of context, but also of the word's accent type in Standard Japanese. This implies that even those learners with relatively high accuracy and stability (and even the highest was only 31%) cannot be considered to have acquired the ability to maintain stable and accurate accent type contrasts across contexts.

A weakness of the current study is that the participants were not asked whether they knew the words they produced. However, a comparison of words that do and do not appear in beginners' textbooks showed no effect on accuracy. This lack of difference is interesting. When investigating the L2 acquisition of English stress by Spanish, Korean and Thai learners, Guion and colleagues used nonce words; with real words they observed high accuracy (98%, 98%, and 90% respectively for typically stressed English words and 96%, 95%, and 72% respectively for atypically stressed

ones; Guion et al., 2004 for Spanish; Guion, 2005 for Korean; Wayland et al., 2006 for Thai). The finding that words in beginners' textbooks, such as *tsuri* "fishing" and *hanashi* "speech," had accuracy as low as words not in beginners' textbooks, including low frequency words such as *shimi* "stain" and *aseri* "haste," is a striking illustration of how difficult Japanese pitch accent is for L1 English speakers.

## 4.2 Experience

The fourth research question asked how accuracy and stability vary with Japanese experience. The more-experienced learners, despite having had more Japanese instruction than the less-experienced learners (an average of 970 hours compared with 250 hours) and spending a year in Japan, did not produce a higher percentage of accurate accent types than the less-experienced learners (both 43%) or a higher percentage of stable accent types (both 40%). These results imply that additional experience does not contribute to increased accuracy or stability in the pitch accent production of real words by English speakers.

This mirrors the claim by Goss (2015, p. 35) that "accent perception does not develop in parallel with proficiency level." Previous research is less clear regarding production. Taylor (2011a) analyzed a subset of the data of the current study and showed a small difference between the more-experienced and less-experienced group. It is possible that a different methodology, that controlled for chance matches between the learners' productions and Standard Japanese (i.e., with a data set balanced across accent types rather than across lexical classes and syllable number) might reveal a difference between more and less-experienced learners. However, the current study shows that all learners have low accuracy. If there is a difference between more and less-experienced learners that is not visible in the current findings, it must be small. In terms of acoustic correlates of pitch accent, Kondo (2007) and Ueyama (2012) showed no correlation between syllable length and proficiency. But Sakamoto (2011) showed that more-experienced learners produce pitch accent with acoustic correlates closer to L1 Japanese speakers when imitating an L1 model. It appears, therefore, that for production there is a difference between more- and less-experienced learners, but in the controlling of acoustic cues when imitating a model, not in the production of real words whose accent types must be encoded in long-term memory.

We know from L2 phonology research that adult users of an L2 can acquire aspects of sound systems that differ from their L1, such as vowel quality and VOT (Zampini, 2008). The current study suggests that even experienced L1 English speakers acquire little Standard Japanese pitch accent. This makes it similar to other particularly difficult aspects of L2 sound systems, such as the contrast between English /ɹ/ and /l/ for L1 Japanese speakers, which show no difference between more-experienced L2 speakers and speakers with little or no L2 experience (Larson-Hall, 2006).

## 4.3 Individual differences

The final research question asked how accuracy and stability vary between learners. Little variation was observed between learners in accuracy (minimum 32%, maximum 52%, mean 43%, *SD* 5). This contrasts with the larger variation in stability (minimum 3%, maximum 77%, mean 40%, *SD* 18). It also contrasts with the large variation in how frequently the learners use each accent type, whether they have a dominant accent type and, if so, which accent type that is. Lastly, it contrasts with the consistent finding in previous research of large individual differences in identification (Hirano-Cook, 2011; Nishinuma et al., 1996; Shport, 2011), even among speakers with no Japanese experience (Shport, 2016). Given these substantial differences in other aspects of both perception and production, the lack of variation in accuracy is striking.

To the author's knowledge, the finding of universally low accuracy in production, despite markedly different accent type stability and frequency of each accent type has not been reported on elsewhere. The inter-learner variation seen here differs in type to individual differences usually discussed in L2 acquisition research (see, for example, Dewaele, 2009), where some people are "better" (i.e., more accurate, fluent, etc.) than others. By responding to the call for production research that analyzes individual learners' productions as well as group means (Munro & Derwing, 2015, p. 31), this study highlights an unusual pattern of individual differences: different in behavior but not accuracy.

Several learners had penultimate accent as a dominant accent type. One might anticipate that, for three-syllable words, initial stress would transfer from English, not medial stress: Cutler and Carter (1987) show that 90% of content words in English have strong initial syllables, and the metrical theory of English stress assigns antepenultimate stress to word with light penults (Hayes, 1982). However, it is penultimate accent that is observed. Further research is needed to explore this apparent contradiction between theory and data.

Some learners had unaccented as their dominant accent type. Previous research has either claimed that English-speaking learners of Japanese cannot produce words unaccented (Toki, 1980, p. 96), or that learners use unaccented when "hesitant" (Yamada, 1994, p. 113) or when producing words whose accent type is "difficult" (Yoshimitsu, 1981, p. 72). Unaccented, rather than being impossible or exceptional, is clearly the preferred accent type for some learners—an interesting finding, since no multisyllabic words in English have no stress.

In summary, some learners had dominant penultimate accent, some dominant unaccented, some no dominant type. However, despite differences in stability and in which accent types were produced, accuracy was low for all learners.

#### *4.4 Why is pitch accent so difficult to acquire?*

Recent research by the current author described a trilingual speaker of English and the Nigerian languages Nupe and Hausa who produced Japanese pitch accent that was highly accurate and stable, despite not having lived in Japan or received explicit instruction on Japanese pitch accent (Muradás-Taylor, 2018). It is likely that her other L1s, Nupe and Hausa, which are both tonal, helped her acquire Japanese pitch accent. So why, for monolingual English speakers, is Japanese pitch accent so difficult to acquire? In this section, the findings of the current study will be compared to previous research to argue that L1 English speakers' difficulty acquiring L2 Japanese pitch accent lies in encoding pitch accent in long-term memory.

Production—specifically, the acoustic realization of pitch accent—does not appear to be the main source of difficulty. The majority of words (60%) were produced with different accent types in different contexts, which could suggest difficulty controlling pitch production. However, we know from Sakamoto (2011) that more-experienced learners can imitate accent types with acoustic correlates that are closer to Standard Japanese norms than less-experienced learners. If production were the main source of difficulty, we would expect more-experienced learners to produce a greater number of accurate and stable accent types than less-experienced learners. But this is not what was observed.

Perception is also unlikely to be the main source of difficulty. We know from previous research that L1 English speakers can discriminate accent types with accuracy as high as L1 Japanese speakers (Hirano-Cook, 2011; Sakamoto, 2011), but that there are individual differences in how well L1 English speakers can identify accent types (Hirano-Cook, 2011; Nishinuma et al., 1996; Shport, 2011, 2016). L1 English-speakers might be expected to acquire pitch accent easily (because they can discriminate different accent types) or, alternatively, for only some learners to acquire it



(because some learners can also identify accent types). Neither of these are what is observed: all learners produce inaccurate and unstable accent types. Goss (2015) shows that variation in accuracy in perception can be explained by individual differences in phonological short-term memory capacity, acoustic pitch sensitivity (Goss, 2015, pp. 138–140) and the size of the L2 lexicon (Goss, 2015, pp. 96–97). But these factors must have limited relevance here, as producing real words with accurate and stable accent types is difficult for all learners.

An inability to categorize pitch differences according to the Japanese linguistic system (Hirata, 2015, p. 736; see also Goss, 2018, p. 8) also cannot explain L1 English speakers' difficulty acquiring pitch accent. An inability to categorize explains why some learners can discriminate but not identify accent types. However, if categorizing accent types were the cause of L1 English speakers' difficulty acquiring pitch accent, you would expect to see large individual differences, as are seen for identification, not the universally low accuracy seen here.

Instead, it appears that L1 English-speaking learners of L2 Japanese do not encode pitch in words' representations in long-term memory, even if they can identify them. This would explain why, despite the large individual differences consistently seen in identification studies (Hirano-Cook, 2011; Nishinuma et al., 1996; Shport, 2011, 2016), all 21 learners in the current study showed low accuracy in production. And it is consistent with previous research showing that even advanced learners are unable to judge whether words' accent types are "correct" (Shibata & Hurtig, 2008), because this requires knowledge of which words have which accent types in Standard Japanese.

Little is known about the representation of suprasegmental contrasts in the lexicon of L2 speakers (Braun et al., 2014). Being able to perceive pitch accent but not encode it lexically is consistent with our understanding of the function of pitch in English. Although L1 English-speaking adults (Cutler, 1986) and infants (Curtin, 2010) encode stress in the lexical representation, pitch is unlikely to be encoded because it is an unreliable cue to stress. And perceiving but not encoding a suprasegmental contrast mirrors the stress "deafness" that has been observed in L1 French learners of L2 stress languages (Dupoux et al., 1997), even in experienced learners (Dupoux et al., 2008).

#### **4.5 Future research avenues**

This study raises five main questions for future research, as follows.

First, is it possible that learners do not acquire pitch accent because they do not need to? Japanese pitch accent not only varies with dialect (Kubozono, 2012), but also has low functional load in Standard Japanese: according to Kitahara (2001, p. 4), only 13% of one- to four-syllable words contrast only in accent type. We need to compare English speakers' acquisition of Japanese pitch accent with their acquisition of Chinese tone, which has less dialectal variation and a higher functional load. Research into the acquisition of Chinese tone has investigated English speakers' production of tones either on mimicking an L1 model, or on reading aloud words annotated with tone labels (Hao, 2012). It must be left to future research to investigate learners' production of real words whose tones must be encoded in long-term memory.

Second, can learners acquire Standard Japanese pitch accent with more years of exposure or training, as has been shown for Japanese speakers' acquisition of English /r/ and /l/ (Bradlow et al., 1999; Flege et al., 1995)? Most of the more-experienced learners had only spent one year in Japan, with the longest length of residence being four years; this is dwarfed by the average length of residence of 21 years in Flege et al. (1995) who observed successful acquisition of /r/ and /l/ in L1 Japanese speakers. Research has shown an effect of training on English speakers' identification of accent types (Shport, 2016). And a "Web-based prosodic reading tutor," which trains students using a combination of audio, accent type notation, and visual pitch contours, has been shown to

improve “naturalness” as rated by Japanese teachers (Minematsu et al., 2016). However, it is not known whether training can lead to longer-term changes in L1 English speakers’ ability to encode the accent types of new words in the lexical representation.

Third, what factors affect words’ accent types? Why do several learners have penultimate accent as their dominant accent type, when English stress rules would seem to predict antepenultimate? Are the accent types of words with three or more syllables, which have a contrast in accent position, easier to acquire than those of two-syllable words, which only contrast in accentedness in isolation, as suggested by Taylor (2011a)? What is the relation between context and accent type? Are words before another content word more likely to be unaccented than other words, as could be implied by the accent categorization system used by Yamada (1994)? More work is needed in this area.

Fourth, why do individual learners’ accent systems develop so differently? Why do some learners have dominant penultimate accent, others dominant unaccented, and others no dominant accent type? Does the relation between accent type and other factors (word length, lexical class, context) vary between learners, as claimed by Taylor (2011b)? It is important to remember that this is not the sort of individual difference that we are used to—with some learners being better than others—but is a difference in behavior that does not correlate with accuracy.

Fifth, how does non-Standard pitch accent affect comprehensibility and intelligibility (Munro & Derwing, 1995)? This question has implications for the Japanese language classroom. In the US and UK, pitch accent is not normally marked in textbooks or explicitly taught (Shport, 2008). The low accuracy observed in the current study implies that L1 English speakers do not acquire Standard Japanese pitch accent through implicit learning alone. This could be taken as a “call-to-action” to find effective training methods (Goss, 2018, p. 11). An alternative approach would be to accept that L1 English speakers do not normally acquire Standard Japanese pitch accent and, instead, incorporate into the Japanese language classroom teaching materials featuring speakers who use a variety of “Japaneses.” This would mirror the call by Walker (2010) and others for a variety of Englishes to be used in the English language classroom. The following benefit could follow for learners: rather than feeling that Standard Japanese pitch accent is “correct” and that their own pitch accents are deficient, learners would understand that there are many varieties of Japanese, gaining confidence in their own variety. Research into comprehensibility and intelligibility is needed to explore this question further.

## 5 Conclusion

This paper investigates L1 English speakers’ L2 acquisition of Standard Japanese pitch accent, taking the novel approach of having learners produce words in three contexts (e.g., *ame* “rain,” *ame da* “it’s rain,” and *ame ga furu* “rain falls”) to investigate the accuracy and stability of the learners’ accent types. The accent types produced by two learner groups, with differing amounts of Japanese experience, were analyzed, showing that L1 English L2 Japanese speakers produce inaccurate and unstable accent types irrespective of amount of Japanese experience. Analysis of individual learners showed that those learners who produced relatively stable accent types had a dominant accent type that they used irrespective of a word’s accent type in Standard Japanese, not an acquired ability to maintain accent type contrasts across contexts. Neither more-experienced learners nor any individual learners had high accuracy in their pitch accent production, suggesting that pitch accent is as difficult as /ɹ/ and /l/ for L1 Japanese speakers. The learners’ low accuracy in production contrasts with the large individual differences in identification consistently found in previous research (Hirano-Cook, 2011; Nishinuma et al., 1996; Shport, 2011, 2016). It is proposed that L1 English listeners do not encode Japanese pitch accent in long-term memory. This has implications for future research and Japanese language teaching.

## Acknowledgements

This paper has been a long time in the making. Heartfelt thanks to my PhD supervisor, Professor Tanomu Kashima, the participants, the University of Nagoya judges, colleagues from both the University of York and York St John University, presentation audiences at J-SLA, EuroSLA, EPIP, BAAL, ISMBS and New Sounds, and the reviewers.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Nagoya University Scholarships for Outstanding Graduate Students, a Japan Student Services Organisation Honors Scholarship for Privately Financed International Students, a Daiko Foundation Ryugaku Gakugei Scholarship, and the Great Britain Sasakawa Foundation [grant number 4181].

## ORCID iD

Becky Muradás-Taylor  <https://orcid.org/0000-0001-7275-6016>

## References

- 3A Network (1998). *Minna no nihongo: shokyuu I/II honsatsu* [Minna no nihongo: Beginner 1/2 main textbook]. Tokyo: 3A Network. [In Japanese.]
- Archibald, J. (1992). Transfer of L1 parameter settings: Some empirical evidence from Polish metrics. *Canadian Journal of Linguistics*, 37(3), 301–339. <https://doi.org/10.1017/S0008413100019903>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beckman, M. E. (1986). *Stress and non-stress accent*. Foris.
- 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 University Press.
- Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, 3, 255–309. <https://doi.org/10.1017/S095267570000066X>
- Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977–985. <https://doi.org/10.3758/BF03206911>
- Braun, B., Galts, T., & Kabak, B. (2014). Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. *Second Language Research*, 30(3), 323–350. <https://doi.org/10.1177/0267658313510926>
- Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*, 105(4), 376–385. <https://doi.org/10.1016/j.jecp.2009.12.004>
- Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29(3), 201–220. <https://doi.org/10.1177/002383098602900302>
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2(3–4), 133–142. [https://doi.org/10.1016/0885-2308\(87\)90004-0](https://doi.org/10.1016/0885-2308(87)90004-0)
- Dewaele, J.-M. (2009). Individual differences in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 623–646). Emerald.
- Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing ‘deafness’ in French? *Journal of Memory and Language*, 36(3), 406–421. <https://doi.org/10.1006/jmla.1996.2500>
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress ‘deafness’: The case of French learners of Spanish. *Cognition*, 106(2), 682–706. <https://doi.org/10.1016/j.cognition.2007.04.001>
- Fllege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech*, 38(1), 25–55. <https://doi.org/10.1177/002383099503800102>
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(2), 126–152. <https://doi.org/10.1177/002383095800100207>

- Goss, S. J. (2015). *The effects of internal and experience-based factors on the perception of lexical pitch accent by native and nonnative Japanese listeners* (Doctoral dissertation, Ohio State University).
- Goss, S. (2018). A critical pedagogy of lexical accent in L2 Japanese: Insights into research and practice. *Japanese Language and Literature*, 52(1), 1–24.
- 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. <https://doi.org/10.1111/jpr.12076>
- 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. <https://doi.org/10.1017/S0272263105050230>
- 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. <https://doi.org/10.1177/00238309030460040301>
- 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(3), 207–226. <https://doi.org/10.1017/S1366728904001592>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269–279. <https://doi.org/10.1016/j.wocn.2011.11.001>
- Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry*, 13(2), 227–276. <https://linguistics.ucla.edu/people/hayes/papers/Hayes1982ExtrametricalityAndEnglishStress.pdf>
- Hirano-Cook, E. (2011). Japanese pitch accent acquisition by learners of Japanese: Effects of training on Japanese accent instruction, perception, and production (Doctoral dissertation, University of Kansas). <https://core.ac.uk/download/pdf/213395046.pdf>
- Hirata, Y. (2015). L2 phonetics and phonology. In H. Kubozono (Ed.), *The handbook of Japanese language and linguistics: Phonetics and phonology* (pp. 719–762). Mouton de Gruyter.
- Hirayama, T. (1998). Zen nihon no hatsuon to akusento [Pronunciation and accent across Japan]. In NHK Broadcasting Culture Research Institute (Ed.), *Nihongo hatsuon akusento jiten* [Japanese pronunciation accent dictionary] (appendix pp. 123–173). NHK. [In Japanese.]
- Horiguchi, S. (1973). Eigo kokumin ni yoru nihongo no yon-onsetu meishi no akusento no yosoku to sono jissai [English speakers' predicted and actual accent types for Japanese 4-syllable nouns]. *Journal of Japanese Language Teaching*, 19, 97–112. [In Japanese.]
- Idemaru, K., Wei, P., & Gubbins, L. (2019). Acoustic sources of accent in second language Japanese speech. *Language and Speech*, 62(2), 333–357.
- Kindaichi, H., & Akinaga, K. (Eds.) (2001). *Shin Meikai nihongo akusento jiten* [New Meikai Japanese accent dictionary]. Sanseido. [In Japanese.]
- Kitahara, M. (2001). Category structure and function of pitch accent in Tokyo Japanese (Doctoral dissertation, Indiana University). <http://www.f.waseda.jp/kitahara/Paper/thesis-dist.pdf>
- Kondo, M. (2007). Acoustic realization of lexical accent and its effects on phrase intonation in English speakers' Japanese. In Proceedings of the 16th International Conference on Phonetic Sciences, Saarbrücken, Germany, 6–10 August 2007 (pp. 1649–1652). International Phonetic Association. <http://www.icphs2007.de/conference/Papers/1306/1306.pdf>
- Kubozono, H. (2008). Japanese accent. In S. Miyagawa & M. Saito (Eds.), *The Oxford handbook of Japanese linguistics* (pp. 165–191). Oxford University Press.
- Kubozono, H. (2012). Varieties of pitch accent systems in Japanese. *Lingua*, 122(13), 1395–1414. <https://doi.org/10.1016/j.lingua.2012.08.001>
- Kuno, S. (1998). Bei-eigowasha ni okeru nihongo akusento no seisei [The production of Japanese accent by American English speakers]. *Phonological Studies*, 1, 83–90. [In Japanese.]
- Larson-Hall, J. (2006). What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /ɪ/ and /I/ by Japanese speakers. *Language and Speech*, 49(4), 521–548. <https://doi.org/10.1177/00238309060490040401>

- Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36(1), 141–157. <https://doi.org/10.1016/j.wocn.2007.03.001>
- Minematsu, N., Hirano, H., Nakamura, N., & Oikawa, K. (2016). Improvement of naturalness of learners' spoken Japanese by practicing with the web-based prosodic reading tutor, Suzuki-kun. *Speech Prosody*, 257–261. [https://www.isca-speech.org/archive/SpeechProsody\\_2016/pdfs/64.pdf](https://www.isca-speech.org/archive/SpeechProsody_2016/pdfs/64.pdf)
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(Supplement 1), 285–310. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Munro, M. J., & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, 1(1), 11–42. <https://doi.org/10.1075/jslp.1.1.01mun>
- Muradás-Taylor, B. (2018). Japanese pitch accent production in an English/Nupe/Hausa trilingual: Accuracy, stability and F0 realisation. In B. Elena (Ed.), *Crosslinguistic research in monolingual and bilingual speech* (pp. 164–180). Institute of Monolingual and Bilingual Speech.
- Nishinuma, Y., Arai, M., & Ayusawa, T. (1996). Perception of tonal accent by Americans learning Japanese. In *Proceedings of the Fourth International Conference on Spoken Language Processing, Philadelphia, Pennsylvania, USA*, October 3–6, 1996 (pp. 646–649). Institute of Electrical and Electronics Engineers.
- Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioural plasticity in speech perception. *Cognition*, 64(3), B9–B17. [https://doi.org/10.1016/S0010-0277\(97\)00030-9](https://doi.org/10.1016/S0010-0277(97)00030-9)
- Pater, J. V. (1997). Metrical parameter missetting in second language acquisition. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 235–261). Benjamins.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology).
- Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191–215. <https://doi.org/10.1006/jpho.2001.0134>
- Plag, I., Kunter, G., & Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics*, 39(3), 362–374. <https://doi.org/10.1016/j.wocn.2011.03.004>
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Sakamoto, E. (2011). An investigation of factors behind foreign accent in the L2 acquisition of Japanese lexical pitch accent by adult English speakers (Doctoral dissertation, University of Edinburgh). <https://era.ed.ac.uk/bitstream/handle/1842/5692/Sakamoto2011.pdf?sequence=1&isAllowed=y>
- Shibata, T., & Hurtig, R. R. (2008). Prosody acquisition by Japanese learners. In Z. Han & E. S. Park (Eds.), *Understanding second language process* (pp. 176–204). Multilingual Matters.
- Shport, I. A. (2008). Acquisition of Japanese pitch accent by American learners. In P. Heinrich & Y. Sugita (Eds.), *Japanese as foreign language in the age of globalization* (pp. 165–187). Iudicium Verlag.
- Shport, I. A. (2011). Cross-linguistic perception and learning of Japanese lexical prosody by English listeners (Doctoral dissertation, University of Oregon). <https://core.ac.uk/download/pdf/36686736.pdf>
- Shport, I. A. (2016). Training English listeners to identify pitch-accent patterns in Tokyo Japanese. *Studies in Second Language Acquisition*, 38(4), 739–769. <https://doi.org/10.1017/S027226311500039X>
- Sluijter, A. M. C., & van Heuven, V. J. (1996a). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100(4), 2471–2485. <https://doi.org/10.1121/1.417955>
- Sluijter, A. M. C., & van Heuven, V. J. (1996a). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the Fourth International Conference on Spoken Language Processing, Philadelphia, Pennsylvania, USA*, October 3–6, 1996 (pp. 630–633). Institute of Electrical and Electronics Engineers.
- 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. <https://doi.org/10.1121/1.417955>
- Sugito, M., & Tahara, H. (1989). Toukeiteki kanten kara mita Oosaka akusento: Toukyou to no hikaku wo chuushin ni [A quantitative analysis of word-accent in the Osaka dialect]. *Studies in Phonetics and Speech Communication*, 3, 143–165. [In Japanese.]



Taylor, B. (2011a). Do English learners of Japanese produce isolated nouns with Standard Japanese lexical accent? *Second Language*, 10, 15–31. [https://doi.org/10.11431/secondlanguage.10.0\\_15](https://doi.org/10.11431/secondlanguage.10.0_15)

Taylor, B. (2011b). Variability and systematicity in individual learners’ Japanese lexical accent. *Poznan Studies in Contemporary Linguistics*, 47(1), 146–158. <https://doi.org/10.2478/psicl-2011-0012>

Taylor, B. (2012a). Eigo wo bogo to suru nihongo gakushuusha ni yoru gomatsu akusento no seisei [The production of word-final accent by English-speaking learners of Japanese]. *Issues in Language and Culture*, 13, 77–94. [In Japanese.]

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

Toki, T. (1980). Eigo wo bogo to suru gakushuusha ni okeru akusento no keikou [Accent trends in learners with English as a first language]. *Inter-University Center Journal*, 3, 78–96. [In Japanese.]

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

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

Walker, R. (2010). *Teaching the pronunciation of English as a lingua franca*. Oxford 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(3), 285–304. <https://doi.org/10.1007/s10936-006-9016-9>

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5), 2020–2045. <https://doi.org/10.1037/xge0000014>

Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge.

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.

Yoshimitsu, K. (1981). Gaikokujin gakushuusha no akusento [Foreign learners’ accent]. *Journal of Japanese Language Teaching*, 45, 63–75. [In Japanese.]

Zampini, M. L. (2008). L2 speech production research. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 219–250). John Benjamins Publishing.

# Appendix. Full list of words (n=180).

## 2 syllable pure nouns

Initial accent			Final accent			Unaccented		
雨	ame	rain	部屋	heya	room	箱	hako	box
地図	chizu	map	犬	inu	dog	暇	hima	free time
駅	eki	station	鍵	kagi	key	紐	himo	string
船	fune	boat	怪我	kega	injury	星	hoshi	star
傘	kasa	umbrella	町	machi	town	医者	isha	doctor
窓	mado	window	夏	natsu	summer	椅子	isu	chair
猫	neko	cat	音	oto	sound	壁	kabe	wall
奥	oku	inside	島	shima	island	道	michi	road
空	sora	sky	裏	ura	back	右	migi	right
外	soto	outside	山	yama	mountain	水	mizu	water
海	umi	sea	雪	yuki	snow	庭	niwa	garden
夜	yoru	night	夢	yume	dream	横	yoko	sideways

(continued)



(continued)

## 3 syllable pure nouns

Initial accent			Medial accent		
出口	deguchi	exit	あなた	anata	you
技術	gijutsu	technique	八時	hachiji	8 o'clock
五月	gogatsu	May	花屋	hanaya	flower shop
花火	hanabi	fireworks	一部	ichibu	a part
医学	igaku	medicine	いとこ	itoko	cousin
彼女	kanojo	girlfriend	事務所	jimusho	office
家族	kazoku	family	中身	nakami	contents
眼鏡	megane	glasses	二か所	naname	diagonally
緑	midori	greenery	斜め	nikasho	two places
涙	namida	tears	おもちゃ	omocha	toy
荷物	nimotsu	luggage	おしゃれ	oshare	fashionable
姿	sugata	figure	和菓子	wagashi	japanese sweets
Final accent			Unaccented		
昼間	hiruma	day time	左	hidari	left
地獄	jigoku	hell	日にち	hinichi	date
鏡	kagami	mirror	田舎	inaka	countryside
言葉	kotoba	words	子供	kodomo	child
娘	musume	daughter	車	kuruma	car
仲間	nakama	companion	昔	mukashi	the past
表	omote	front	同じ	onaji	the same
男	otoko	man	大人	otona	adult
刺身	sashimi	raw fish	魚	sakana	fish
宝	takara	treasure	仕事	shigoto	work
手元	temoto	handy	手紙	tegami	letter
夜中	yonaka	middle of the night	後ろ	ushiro	behind

## 2 syllable derived nouns

Final accent			Unaccented		
飽き	aki	boredom	空き	aki	opening
振り	furi	pretence	減り	heri	loss
晴れ	hare	sun	貸し	kashi	something to lend
褒め	home	praise	負け	make	defeat
勝ち	kachi	victory	真似	mane	imitation
賭け	kake	bet	乗り	nori	upbeat feeling
漏れ	more	leak	抜け	nuke	gap
慣れ	nare	practice	染み	shimi	stain
伸び	nobi	growth	釣り	tsuri	fishing
次	tsugi	next	売り	uri	the market
膿	umi	pus	詫び	wabi	apology
読み	yomi	reading	揺れ	yure	vibration

(continued)

(continued)

## 3 syllable derived nouns

Final accent			Unaccented		
歩き	aruki	walking	始め	hajime	beginning
焦り	aseri	hurry	代わり	kawari	replacement
話	hanashi	speech	飾り	kazari	decoration
痛み	itami	pain	煙	kemuri	smoke
曇り	kumori	cloud	周り	mawari	vicinity
流れ	nagare	flow	並び	narabi	ranking
悩み	nayami	worry	踊り	odori	dance
残り	nokori	leftovers	遅れ	okure	delay
恐れ	osore	fear	終わり	owari	end
動き	ugoki	movement	泊まり	tomari	stopover
別れ	wakare	separation	続き	tsuzuki	continuation
休み	yasumi	holiday	汚れ	yogore	dirt

## 2 syllable verbs

Initial accent			Unaccented		
出す	dasu	to put out	踏む	fumu	to tread
出る	deru	to leave	貼る	haru	to stick
書く	kaku	to write	行く	iku	to go
噛む	kamu	to bit	いる	iru	to be
切る	kiru	to cut	消す	kesu	to delete
見る	miru	to see	着る	kiru	to wear
持つ	motsu	to carry	泣く	naku	to cry
飲む	nomu	to drink	寝る	neru	to sleep
脱ぐ	nugu	to take off	乗る	noru	to ride
住む	sumu	to live	押す	osu	to push
取る	toru	to take	する	suru	to do
打つ	utsu	to hit	飛ぶ	tobu	to fly

## 3 syllable verbs

Medial accent			Unaccented		
歩く	aruku	to walk	浴びる	abiru	to soak up
選ぶ	erabu	to choose	遊ぶ	asobu	to play
話す	hanasu	to speak	入れる	ireru	to put in
走る	hashiru	to run	磨く	migaku	to polish
被る	kaburu	to wear	並ぶ	narabu	to line up
壊す	kowasu	to break	登る	noboru	to climb
戻る	modoru	to go back	送る	okuru	to send
起きる	okiru	to get up	探す	sagasu	to search
泳ぐ	oyogu	to swim	泊まる	tomaru	to stay
食べる	taberu	to eat	続く	tsuzuku	to continue
届く	todoku	to deliver	痩せる	yaseru	to lose weight
休む	yasumu	to rest	汚す	yogosu	to dirty