CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Exploring variation in nonnative Japanese learners' perception of lexical pitch accent: The roles of processing resources and learning context

Seth Goss*

Emory University
*Corresponding author. E-mail: sethjgoss@emory.edu

Abstract

This article reports findings on the effects of processing resources and learning context on the perceptual learning of lexical pitch accent in beginning nonnative Japanese learners. Native English speakers in at-home and study-abroad contexts were tested twice during a semester of Japanese study on their ability to judge the correctness of and categorize nouns by their pitch pattern. Regression analyses indicated that the ability to store nonnative-like sound sequences in phonological short-term memory (PSTM), as well as auditory processing ability, predicted a significant degree of perceptual gains made over a 12-week interval. However, these predictors were task specific in that PSTM capacity predicted correctness judgment gains, while auditory processing accounted for variation in categorization. Furthermore, despite learners in the at-home context performing slightly better overall, processing resources adhered to the same predictive pattern when context was taken into account. The results suggest that (a) neither increased input during study-abroad nor targeted instruction is sufficient for most learners to acquire lexical accent; (b) processing resources support the acquisition of lexical prosody, but these may depend on how learning is assessed; and (c) PSTM operates across learning contexts, suggesting it to be a domain-general capacity in early-stage nonnative language acquisition.

Early-stage learning of nonnative language (L2) suprasegmentals—features such as tone and pitch accent—can be characterized at its most basic level as a process of mapping fundamental frequency (F0) parameters onto a word's phonetic segments (So & Best, 2010; Wayland & Guion, 2004). Yet perceiving this phonetic information is a frequent source of difficulty for L2 learners, especially those from non-tonal language backgrounds (e.g., Burnham & Mattock, 2007; Wong & Perrachione, 2007). Japanese lexical pitch accent is an example of a suprasegmental that is acquired with widely differential success across the L2 proficiency continuum (e.g., Ayusawa, 2003;

Shibata & Hurtig, 2008; Taylor, 2011). Although native (L1) English learners of Japanese, for instance, exhibit clear development in the perception of phonemic length contrasts absent in their L1 (e.g., Hirata, 2004; Shibata & Hurtig, 2008), a similar trajectory is unattested with lexical pitch accent. Several reasons for this difficulty have been posited, including pitch's low perceptual prominence (Schaefer & Darcy, 2014; Tsurutani, 2011), its limited functional load in disambiguating minimal pairs (Kitahara, 2001; Tamaoka, 2014), variation in cross-linguistic speech categories (So & Best, 2010), and individual differences in speech processing resources (Bowles, Chang, & Karuzis, 2016; Goss & Tamaoka, 2019; Shport, 2016). This latter claim has yet to be adequately characterized for L2 Japanese learners. Given evidence that lexical accent is low prominence and carries a comparatively low functional load, those with a greater capacity for extracting relevant phonetic information from speech input may be more successful learners.

Lexical accent learning in beginning learners of L2 Japanese may be partially constrained by phonological short-term memory (PSTM) capacity and auditory processing ability, as learners lack substantial experience with the acoustic cues to lexical accent and possess a small lexical store (Martin & Ellis, 2012; Sunderman & Kroll, 2009). Exploring these capacities is vital because they may account for some of the previously observed variation in L2 accent acquisition (e.g., Nishinuma, Arai, & Ayusawa, 1996; Taylor, 2011). Although the involvement of domain-general capacities in L2 suprasegmental acquisition has been examined before (e.g., Bowles et al., 2016; Goss & Tamaoka, 2019; Shport, 2015; Wong & Perrachione, 2007), this relationship needs further exploration in beginning learners of a pitch-accent language, as these capacities may play a greater role in early-stage L2 learning (Hummel, 2009). Furthermore, because learners with greater speech processing resources may be better equipped to utilize the increased input in the L2 environment (Sunderman & Kroll, 2009), the current study compared learning contexts that afford different amounts of speech input (i.e., at home and study abroad).

The current study contributes to the understanding of how differences in processing resources influence the development of prosodic perception over a semester of instructed L2 Japanese acquisition in two learning contexts. Forty beginning L2 Japanese learners (L1 English speakers) in two learning contexts—at home ($n = 20$) and study abroad ($n = 20$)—were tested twice during a semester of language study on two perception tasks: judging lexical accent pattern correctness (a task that requires word recognition) and categorizing words by pattern (a task that is acoustically biased). PSTM capacity and auditory processing ability, along with their interaction with learning context, were then used as predictors in regression analyses on accuracy gains on the perception tasks over a semester of study.

Lexical pitch accent in L2 Japanese

In Japanese, F0, of which pitch is the perceptual correlate, is used by L1 listeners to distinguish words in minimal pairs (Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999), for compound segmentation (Hirose & Mazuka, 2015), and in the parsing of phrases (Ito & Speer, 2008). Japanese has thus been classified typologically as a pitch-accent language (Shibatani, 1990). The primary acoustic cue to accent location is a fall in F0 over a two-mora unit (H*+L tone; Beckman & Pierrehumbert, 1986).

In accent-bearing words, these patterns are lexically specified, meaning that the H* tone is linked to a single mora, and followed by a steep fall in pitch (L tone) on the next mora or syllable. There is also a large class of unaccented words in Japanese that do not contain an F0 fall (Ito & Mester, 2016). L1 Japanese listeners, and even highly proficient L2 learners with large vocabularies (Goss & Tamaoka, 2019), are assumed to represent these accent patterns as lexically relevant information in their mental lexicon (Cutler & Otake, 1999).

The current study examined the perception of three accent types found in Tokyo Japanese: (a) initial mora accent, (b) medial accent, and (c) unaccented. The corresponding F0 contours for each of these can be described as "falling (H*L)," "rising then falling (LH*)," and "rising (LH)," with examples of trimoraic words of each pattern being *MEgane* "glasses" (H*LL), *taMAgo* "egg" (LH*L), and *saKANA* "fish" (LHH).

L1 English listeners find Japanese lexical accent difficult because F0 variations are used postlexically in English to indicate pragmatic focus and speaker identity (Ladd, 2008). Previous research has shown low perceptual accuracy for lexical accent across the proficiency spectrum (Shibata & Hurtig, 2008), and on a range of perception tasks (see Ayusawa, 2003, for review). Yet the ability to perceive word-level pitch contrasts varies widely among individuals, and it is this variation that is the current study's primary focus. For example, Shport (2016) found a large degree of individual variation among Japanese-naïve L1 English speakers in their weighting of F0 height and direction cues to lexical accent, which may have to an extent been attributable to general cognitive ability or attentional biases. Accent labeling tasks using real words have also elicited marked variation, with one study reporting a 30% difference in mean accuracy between low- and high-performing groups of L1 English listeners (Nishinuma et al., 1996). Production studies have noted low accuracy but relatively little variation, leading some to claim that L1 English speakers are unable to acquire stable representations of lexical accent patterns in long-term memory (Taylor, 2011). However, because pitch accent production involves articulatory demands absent in perception and may inadequately reflect long-term knowledge—given that most beginning Japanese learners are equally poor at accent production—the current study takes perception as a starting point for its investigation of variation.

Perception task type

Task type may influence the resources that listeners employ in speech perception. For instance, some tasks, such as same–different discrimination, primarily involve an acoustic mode of perception (Bent, Bradlow, & Wright, 2006), while others, such as lexical decision tasks and correctness judgments, are lexical as they require the comparison of a perceived stimulus to a representation in long-term memory (Goss & Tamaoka, 2015). Discrimination and categorization tasks are common in L2 perception studies and have been used to answer questions about how speech categories are established cross-linguistically (e.g., Shport, 2016; Wong & Perrachione, 2007). However, these tasks can potentially be performed without knowledge of the target stimuli, and listeners with a sharp ear for pitch, such as trained musicians, often perform well on these for the very reason that discrimination and categorization likely involve an acoustic mode of perception (Cooper & Wang, 2012). Variation even among L1 Japanese listeners has been observed on acoustically biased tasks,

suggesting that performance is more closely tied to a listener's general perceptual resources than to lexical knowledge (Goss & Tamaoka, 2015; Shport, 2016).

At the opposite end of the perceptual spectrum are correctness judgment tasks, which require listeners to access a word's representation from memory, and can thus be used to examine lexical knowledge more directly than discrimination or categorization tasks. Goss and Tamaoka (2015) reported that L1 Japanese speakers from Tokyo-type accent regions were able to judge the correctness of high-frequency words at a mean accuracy of 93%, well above averages reported in previous categorization tasks (Sakamoto, 2010; Shport, 2008). In the same study using shared stimuli, L1 listeners' accuracy on a task that involved categorization (4AFC) of spoken accent patterns using visual representations of pitch contours was 61%, a difference suggesting that separate perceptual modes are invoked by these tasks (Strange & Shafer, 2008). If we consider this markedly lower performance and the fact that categorization can be performed by target-language naïve listeners based on a comparison of the visual pitch contour with the aural stimulus, then the task can be considered acoustically biased relative to a correctness judgment (Goss & Tamaoka, 2015).

The current study used both a correctness judgment and a categorization task to examine not only differences in task performance but also whether accuracy on the two tasks differs as a function of listeners' processing resources. Examining both acoustic and lexically biased perception tasks in the same individuals can provide insight into L2 speech processing. Namely, because these tasks likely implicate different perceptual modes, they enable the investigation of the effect of disparate processing resources, which are described next, on L2 speech perception mechanisms.

Processing resources and L2 learning
Phonological STM
The working/short-term memory (WM/STM) systems have been extensively studied over the past four decades in both L1 and L2 acquisition research (Baddeley & Hitch, 1974; see Wen, 2014, for review). The present study focuses on the phonological loop of the STM system, and the phonological storage subcomponent in particular. This mechanism is responsible for the brief maintenance of sound information and the transfer of these memory traces into the more durable long-term store (Baddeley, 2003). Research on the phonological memory system has linked STM function to many facets of child and adult L2 learning including grammar (Martin & Ellis, 2012; O'Brien, Segalowitz, Freed, & Collentine, 2006; Williams & Lovatt, 2003), vocabulary (Hu, 2003; Speciale et al., 2004), speech fluency (O'Brien, Segalowitz, Freed, & Collentine, 2007), and pronunciation (Nagle, 2013). It has thus been characterized as a support mechanism, not only for input processing but for language learning in general (Baddeley, Gathercole, & Papagno, 1998). One central assumption for a phonological STM-language acquisition link is that L2 learners with a greater capacity for holding speech input in the phonological store may ultimately be more successful at establishing long-term phonological representations of lexical forms based on this input (Baddeley et al., 1998).

Yet phonological store capacity may not be uniformly involved throughout the L2 proficiency continuum. Previous L2 studies have indicated that PSTM is

primarily implicated in early-stage learning (Cheung, 1996; Kormos & Safar, 2008; O'Brien et al., 2006), but that its role diminishes as one acquires a larger lexical store, enabling a more efficient perceptual system (Hu, 2003; Martin & Ellis, 2012; cf. Hummel, 2009). For example, Goss and Tamaoka (2019) found that in advanced L2 Japanese learners, PSTM capacity failed to account for variation in pitch accent perception, with task performance being primarily a function of vocabulary size and tone language knowledge. It has thus been proposed that PSTM functions as an independent learning mechanism in the early stages of L2 phonological sequence learning but becomes inextricably tied to long-term lexical representations in more experienced learners (Juffs & Harrington, 2011).

Recent accounts have claimed that STM performance is closely linked to language experience in L2 listeners (e.g., Kaushanskaya, 2012; Martin & Ellis, 2012). For instance, in a study on listening comprehension in L1 and L2 Dutch speakers, Andringa, Olsthoorn, van Beuningen, Schoonen, and Hulstijn (2012) found that verbal STM, defined as a general, or single, resource capacity, did not predict listening comprehension in the L2. Rather, they concluded that L2 listening ability is mostly a function of language knowledge. Why did a verbal STM task not predict listening comprehension in nonnatives? The researchers conjectured that their nonword repetition task did not reflect experience with Dutch, the language in which listening comprehension was measured. Martin and Ellis (2012) echoed this assumption in stating that "PSTM tasks are better predictors of foreign language vocabulary learning when they are more word-like in the foreign language than the native language" (p. 406), suggesting that PSTM capacity is to an extent a proxy for L2 experience. It is thus possible that performance on an STM task that closely resembles target language phonotactics, including pitch accent patterns, may be predictive of L2 suprasegmental learning.

Considering the numerous findings on PSTM's involvement in L2 word learning, how might the phonological store be involved in the early-stage learning of lexical pitch accent? L2 learners with a higher PSTM capacity may be more successful at extracting relevant phonetic information from speech input, including the prosodic cues (i.e., F0 direction and peak) to accent patterns. It may be that greater memory capacity enables learners to process different types of phonetic information simultaneously (Sunderman & Kroll, 2009). PSTM has been found to be predictive of both L2 segmental production (Nagle, 2013) and perception (MacKay, Meador, & Flege, 2001). Because pitch patterns are lexically specified in Tokyo Japanese, and are present in the input to learners, it is therefore possible that higher span learners are better able to process, and subsequently acquire, accent patterns than lower PSTM capacity learners. In other words, learners with a higher capacity for holding L2-like sequences in STM are perhaps also better at processing L2 lexical forms, including accent patterns, in the spoken input. In turn, these higher capacity learners may also establish long-term representations of accent patterns more readily, and thus display greater gains on tasks that measure the consolidation of lexical form, such as judging accent correctness.

Auditory processing ability

An expanding body of research has uncovered a connection between general auditory processing mechanisms and the ability to process linguistic cues (see

Antoniou & Chin, 2018; Asaridou & McQueen, 2013, for reviews). This bottom-up view of perception suggests that domain-general auditory mechanisms may support the learning of suprasegmentals, such that highly pitch-sensitive listeners are initially able to perceive tones more accurately than those with less acuity for discriminating F0 variations (e.g., Cooper & Wang, 2012; Perrachione, Lee, Ha, & Wong, 2011; Wayland, Herrera, & Kaan, 2010; Wong & Perrachione, 2007). For example, learners' ability to perceive nonlexical F0 contrasts and musical experience have been shown to strongly predict the ability to learn Mandarin tones (Bowles et al., 2016; Wong & Perrachione, 2007). L1 English-speaking musicians are also better than nonmusicians at Cantonese tone identification (Cooper & Wang, 2012). However, following a period of training with tones, the involvement of domain-general auditory resources may lessen, given equivalent posttraining accuracy for musicians and nonmusicians (Wayland et al., 2010).

In the case of Japanese lexical accent perception, Shport (2016) found moderate correlations ($r = .35$ and $r = .38$) between indices of musicianship (years of experience and instruments played, respectively) and posttraining gains on an accent pattern identification task by L1 English listeners. Goss and Tamaoka (2015) found that F0 discrimination ability accounted for a significant amount of variation in L1 Japanese listeners' perception of accent patterns, suggesting that even native listeners, with their well-established lexical store, rely on a domain-general perceptual capacity to perform certain listening tasks. However, at the advanced L2 proficiency levels, F0 discrimination was no longer a significant predictor of accent perception or was a least masked by the stronger predictor of Japanese lexical knowledge (Goss & Tamaoka, 2019). This contradictory pattern of findings—that musicianship and acoustic sensitivity relate to perceptual ability in both inexperienced learners and L1 listeners, but not in advanced learners—calls for further data on the role of this capacity in L2 lexical accent learning. It may be that low-proficiency learners are perceiving accent contrasts phonetically (i.e., as uncategorizable F0 variations), and are thus situated at the lower end of a phonetic-to-lexical continuum (Wong & Perrachione, 2007).

Processing resources, learning context, and input

The differential effects that at-home (AH) versus study-abroad (SA) contexts have on lexical acquisition have been extensively documented (e.g., Collentine, 2004; Freed, 1995; Grey, Cox, Serafini, & Sanz, 2015; Sunderman & Kroll, 2009). Yet the influence of context on L2 suprasegmental acquisition remains conspicuously understudied. Immersion in the target-language environment affords increased *access* to L2 input and interaction, and thus greater exposure to a language's phonological patterning (Collentine & Freed, 2004). For Japanese lexical accent, this entails that increased exposure to accent patterns in the speech environment would potentially aid in the establishment of this word-level cue. Along these lines, we can further assume that learners in the SA context could build these long-term representations more rapidly and effectively than those in a lower input AH context.

However, previous research has indicated little benefit of increased input in L2 lexical accent acquisition, despite gains reported in other areas, such as speech fluency (Collentine, 2004; Huensch & Tracy-Ventura, 2017; O'Brien et al., 2007). For example, Lee, Murashima, and Shirai (2006) examined changes in production ability of

lexical accent over a 2-year period by L1 Cantonese speakers studying partly in Japan and partly in their home country. They found no evidence of production gains, despite learners showing clearly improved overall Japanese ability. However, the small sample size ($n = 3$) makes it difficult to generalize their findings to the larger population of L2 Japanese learners. Likewise, studies on L1 English speakers in Japan have shown low accuracy on pitch accent identification (Nishinuma et al., 1996), which does not improve on subsequent testings (Arai, 1997).

Some accounts of L2 acquisition during SA have attempted to explain the lack of benefit of target-language immersion (e.g., Grey et al., 2015; Sunderman & Kroll, 2009). For instance, Sunderman and Kroll (2009) considered the role of SA on L2 acquisition in light of the observation that not infrequently, learners return with little or no improvement in their receptive or productive vocabulary skills. They proposed a "resource threshold hypothesis" in which insufficient WM capacity may hinder a learner's ability to benefit from the increased spoken input available in the L2 setting. Their findings suggest that learners with WM capacities below a certain threshold are unable to utilize the speech input to a comparable degree as learners above this cutoff. Others have found aptitude measures to be better predictors of L2 acquisition than learning context (e.g., O'Brien et al., 2007). The fact that access to increased speech input is not universally beneficial to L2 acquisition necessitates the investigation of whether domain-general capacities interact with learning context in such a way that higher capacity learners are better able to utilize the available input in pitch accent learning. Because PSTM capacity may operate independently of language experience at the beginning level, it may be that high-capacity learners attain better perceptual accuracy regardless of context.

In addition to learning context, we must also consider the effect of pedagogical input on pitch accent learning as its treatment varies greatly by language curriculum. Some Japanese language curricula provide learners only a cursory introduction to the accentual system (Shport, 2008), while others implement detailed instruction and correction of accent patterns (see Goss, 2018, for review). It is unknown whether learners who receive classroom input on lexical accent develop any differently from those who do not. We thus take differences in pedagogical input into account when comparing learners from different Japanese curricula.

In the current study, we operationalized learning context as higher input (SA) and lower input (AH) contexts. We measured the development over a semester of language study in these two contexts with a focus on how speech processing resources relate to gains in perception accuracy. However, it must be noted that the SA learners did not receive any pedagogical input on pitch accent, while the AH learners had received, and continued to receive during the study, classroom instruction on pitch accent. As such, the current study's design was not orthogonal, and we cannot disentangle the effects of higher input and pedagogical intervention on pitch accent learning. Therefore, this manipulation is exploratory, and it is essentially a comparison of the effect of high input with that of classroom instruction.

Research Questions

This study examined five research questions on L1 English-speaking learners' perceptual learning of Japanese lexical pitch accent. The first two questions address

gains in perceptual accuracy over a semester and the role of learning context in learners' perception of lexical accent. Questions 3–5 are of primary interest and focus on the involvement of processing resources in accent perception and their interaction with context.

1. Do L2 Japanese learners in both SA and AH contexts make gains in their ability to perceive lexical accent patterns over a semester of study?
2. After accounting for differences in initial performance, does increased input in a SA context result in higher perceptual accuracy at the end of a semester of language study relative to an AH group who received instruction on accent?
3. Is PSTM capacity predictive of gains in perceptual accuracy after controlling for other variables?
4. Are learners with higher PSTM capacity better able to take advantage of increased input in the SA context?
5. Is auditory processing ability predictive of gains in perceptual accuracy after controlling for other variables?

Extrapolating from previous studies on lexical accent learning, we first predicted perceptual gains to be small but highly variable across individual learners. We also predicted little benefit from SA (i.e., higher input/usage) on accent acquisition relative to the AH group. For the processing resources, we predicted that learners with greater PSTM capacity will be more efficient at extracting pitch cues from speech input and will thus make larger gains over the semester. Furthermore, because PSTM may function as a domain-general resource at the beginning level, those with a higher PSTM capacity should show greater gains regardless of differences in amount of input afforded by learning context. Finally, learners with greater auditory processing resources will also display greater gains on the perception tasks.

Method

Participants

Forty-four L1 English speakers (mean age = 21.4 years, SD = 3.3; 18 female) were recruited for this study: 23 undergraduates at a university in the United States (hereafter, AH group) and 21 undergraduate study-abroad students (SA group) in central Japan. None of the participants in the AH group had spent time in Japan, and all had studied Japanese under the same language curriculum. The SA group included learners from several Japanese language curricula in their home countries (US and UK) who were enrolled in a year-long study-abroad program at the time of the research. Four participants were excluded from the analysis: 1 SA participant who indicated unfamiliarity with several test stimuli on a postexperiment questionnaire, and 3 AH participants who did not attend both test sessions. Thus, data from 20 AH learners and 20 SA learners were included in the analyses. Participant characteristics are shown in Table 1.

Learners in both contexts were enrolled in a second-year Japanese class at the time of the experiment. We selected this level because of the requisite level of

Table 1. Learning-experience measures for SA (n = 20) and AH (n = 20) groups

	SA group *M* (*SD*)	AH group *M* (*SD*)
Age (*years*)	21.9 (3.7)	21.1 (2.2)
LoS (*hrs*)	233 (34.8)	225 (16.6)
Usage per week (*hrs*)	23.2 (3.2)	14.9 (2.7)
Classroom use	10	6
Listening to music	3.8 (1.7)	3.2 (1.1)
Watching TV/videos	5.1 (2.5)	4.7 (1.8)
Speaking w/friends	4.3 (1.2)	1.0 (0.3)
Self-rated proficiency	3.8 (0.7)	3.3 (0.4)

Note: LoS represents length of study at time of first test session. Usage represents time spent using Japanese in these activities.

receptive vocabulary knowledge to perform the real-word perception tasks, and because they possessed some degree of knowledge of the lexical accent system. All participants in the SA group indicated that they were generally aware of the presence of pitch accent but had not received perception or production training. In contrast, the AH group was enrolled in a language curriculum that did practice explicit, in-class instruction and correction of pitch accent, which entailed instructors regularly pointing out learners' pitch production errors, providing normlike models, and giving one lecture on the function of the pitch accent system. Note that none of the stimuli used in this study, nor the individual participants themselves, were targeted for classroom correction.

Classroom contact hours and self-reported usage and proficiency measures were controlled as follows. Participants had a comparable mean length of instruction at the start of the study, t (38) = 1.34, p = .187, d = 0.42. Self-rated proficiency on a 1–10 scale (1 = *no ability*; 10 = *highly proficient*) was similar for the two groups, t (38) = 1.14, p = .261, d = 0.36. However, the groups differed in their usage of Japanese per week, t (38) = 3.27, p < .01, d = 1.03, which was assessed as the number of classroom hours, plus extracurricular time spent speaking and listening to Japanese. Thus, the SA group was considered the higher usage/–correction group, and the AH group lower usage/+correction. In other words, the participants were at approximately the same proficiency level at the onset of study abroad in the SA group, but differed primarily on accent instruction in their prior Japanese study. No participant reported having extensive music training (i.e., >5 years), and none had studied a tone language. All participants reported normal hearing. A small payment was given to participants at the end of each test session.

Instruments

F0 discrimination
Sensitivity to differences in F0 was measured with a test of just noticeable difference (JND). In this task, participants heard two pure tones: the first tone of the pair was a

500 Hz tone that remained constant, while the second tone differed by set intervals of hertz (Mandel, 2009). The between-stimulus frequency difference either increased or decreased depending on the accuracy of participants' responses. For example, at the 6 Hz interval, the first tone was 500 Hz and the second tone 506 Hz. Participants judged whether the second tone was of higher or lower pitch than the first tone. Each tone was 250 ms in length, and the two tones were separated by a 500-ms pause. The resulting score represented the JND in hertz at which a listener could discriminate the paired tones, with a lower JND indicating greater sensitivity to F0 differences. Stimuli were presented through headphones at a loudness comfortable for normally hearing participants.

Nonword pitch discrimination

A nonword discrimination task was used as the second measure of auditory processing ability. In same–different discrimination tasks, difficulty is typically a function of the degree of acoustic similarity of a stimulus pair, as it involves no comparison with established lexical knowledge (Bent et al., 2006). Three accent patterns were elicited from a female native speaker of Tokyo Japanese on two trimoraic carrier nonwords: *NEmate* (H*LL), *neMAte* (LH*L), *neMATE* (LHH); and *NAmugi* (H*LL), *naMUgi* (LH*L), *naMUGI* (LHH). The mean F0 fall (which signals accent location) in the H*LL pattern, measured at the vowel midpoints, between the initial accented (H*) and medial (L) moras, was 72 Hz, and 60 Hz between the medial accented (H*) and final (L) in the LH*L pattern. The unaccented LHH pattern featured an F0 rise from the initial to the medial mora, but no F0 fall. Analysis of mean F0, amplitude, and vowel duration showed that the patterns acoustically resembled those patterns used in nonword stimuli in previous studies (e.g., Cutler & Otake, 1999; Shport, 2016), and were presented without phonetic manipulation. The stimulus structure was 3 accent patterns × 4 repetitions × 3 orders × 2 nonwords, which yielded 72 pairs, half of which were same and half different trials. The interstimulus interval was set to 750 ms, a length that is within the optimal range (500–1000 ms) for phonetic discrimination performance (Gerrits & Schouten, 2004). Trial order was randomized for each listener. Although the use of nonwords in this task biased listeners' decisions toward a phonetic mode of perception, it is more wordlike than the F0 discrimination measure as the nonwords carried pitch patterns analogous to the three accent types in the lexical perception tasks.

Serial nonword recognition

PSTM capacity was measured with a serial nonword recognition task (SNWR). This task was used in previous studies on pitch accent perception (Goss & Tamaoka, 2015, 2019), and its design is based on a syllable-based measure used by O'Brien et al. (2006). In contrast with nonword repetition tasks, SNWR requires no vocalization of the stimuli, thus eliminating the articulatory demands inherent in speech production. The task was presented aurally to participants and was composed of Japanese mora-based nonwords with a consonant–vowel–consonant–vowel segmental structure. All nonwords in this task were spoken with

a low-high pitch accent. This was done because our aim was to measure the relationship between STM capacity and the processing of lexically accented words, on the assumption that stimuli phonotactically resembling Japanese words, including F0 information, would have greater predictive power for real-word perception tasks (e.g., Kaushanskaya, 2012; Martin & Ellis, 2012). Participants heard paired lists of nonwords of varying length and were tasked with deciding if both lists were in the same or different order, requiring them to monitor the serial order of the nonwords to make this decision. The interstimulus interval between the nonwords in each list was 750 ms and a 1.5-s pause separated the lists. Note that participants were not asked to use pitch accent to make their judgments. Refer to Goss and Tamaoka (2019) for a more detailed description of the task.

Correctness judgment and categorization tasks

Two tasks measured lexical accent perception ability. First was a correctness judgment task (PitchID), in which participants decided whether a word's accent pattern was correct or incorrect. This task necessitated access to long-term knowledge of word form and was considered to closely reflect a linguistic mode of perception. The second was a categorization task (PitchCAT), where participants assigned a schematized F0 contour to a word's accent pattern. It differs from the first task in that it likely involves long-term knowledge to a lesser degree, as it can be performed by a Japanese-naïve listener who is highly sensitive to pitch variations and capable of assigning a visual pitch contour to the perceived pattern.

Thirty-six nouns were selected as stimuli for the perception tasks from an introductory Japanese textbook (Jorden & Noda, 1987). Given the limited learning experience of the participants, as well as findings that accent patterns in low-frequency words are perceptually difficult for L1 Japanese (Shport, 2015), we selected test stimuli that the learners were certain to have encountered. Receptive knowledge of the words was confirmed on a postexperiment translation task. Target words were embedded in the initial position of carrier sentences that were set to a length of 6–7 moras for 3-mora target nouns as in 手紙を書く /tegami o kaku/ "(I) write a letter" and 7–8 moras for 4-mora targets as in 飛行機に乗る /hikooki ni noru/ "(I) take an airplane." Including target nouns, carrier sentences had a mean length of 6.8 moras ($SD = 0.71$). Half of the stimuli ($n = 18$) were 3-mora target nouns, and the other half 4-mora nouns.

Three lexical accent patterns were used in this experiment. For 3-mora nouns: Pattern 1: H*LL, Pattern 2: LH*L, and Pattern 3: LHH; and for 4-mora nouns: H*LLL, LH*LL, and LHHH.[1] All stimuli were produced by a male native speaker of Tokyo Japanese.[2] The stimuli were sampled at 44.1 kHz and presented to participants without phonetic manipulation.

Lexical frequency was examined for the accent patterns using a corpus of approximately 300 million Japanese words (Amano & Kondo, 1999, 2000). The mean frequency for all test stimuli was 58.6 tokens per million ($SD = 73$), and by pattern as follows: Pattern 1 ($M = 55.5$, $SD = 71.6$), Pattern 2 ($M = 39.2$; $SD = 67.5$), and Pattern 3 ($M = 81.1$, $SD = 73.6$). Although Pattern 2 was lower in frequency than the other two patterns, a one-way analysis of variance confirmed that this difference
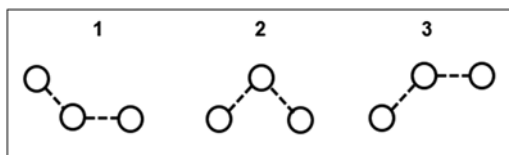
Figure 1. In the PitchCAT task, listeners categorized a sentence-initial noun (e.g., /*tegami o kaku*/ "(I) write a letter") into one of three pitch contours by selecting a visual representation of the pitch pattern (three-mora word display shown).

was not significant, $F(2, 33) = 0.977, p = .387$, which was verified on a post hoc test. Refer to Appendix A for a complete list of test stimuli.

In addition, half of the noun stimuli ($n = 18$) were spoken with the correct pitch accent and half with an incorrect accent pattern. The number of correctly and incorrectly accented items was balanced to control for potential response bias. The PitchID and PitchCAT tasks were combined into one task using shared stimuli as follows. For the correctly accented items only, immediately following the PitchID judgment, participants were required to categorize only the target noun into one of three pitch patterns by selecting an image that matched the pitch contour (Figure 1). Listeners heard each stimulus once and were never required to categorize a word that was spoken with an incorrect accent pattern. We consider the combined tasks as separately analyzable because they elicited different accuracy rates from listeners of both L1 and advanced L2 Japanese backgrounds in previous studies (Goss & Tamaoka, 2015, 2019), and thus likely involved different processing strategies despite being performed in succession. The aural–visual aspect of the PitchCAT task, which required listeners to process two forms of input, may have further served to differentiate the tasks to participants.

Procedures

Data for the PitchID task were structured as 2 word lengths (3 and 4 moras) × 3 accent patterns × 6 words × 40 listeners ($n = 1,440$), with half the number of words (i.e., correctly accented items only) for the PitchCAT task ($n = 720$). Participants in both the SA and AH groups were tested in two sessions approximately 12 weeks apart during a semester of Japanese study. Early in the semester (Time 1), each participant was administered, in order, the F0 discrimination task, nonword discrimination task, and SWNR test, followed by the PitchID and PitchCAT tasks. In the second session (Time 2), learners again performed the PitchID and PitchCAT tasks, with test stimuli randomized to mitigate sequence memorization. Participants performed all tasks with headphones at a comfortable volume in a quiet room. They responded on a standard keyboard, and these were recorded via Superlab 5.0 software. Including practice trials for each task and a short break, the first test session lasted 45 min, and the second session 20 min. After completing the first test session, participants filled out a language-learning background questionnaire.

Table 2. Descriptive statistics for processing resource measures separated by group

	SA (*n* = 20)	AH (*n* = 20)			
Variable	*M* (*SD*)	*M* (*SD*)	Max possible score	Max	Min
F0 discrimination	2.73 (1.88)	2.62 (1.44)	—	−0.4	5
Nonword discrimination	67.25 (3.35)	68.03 (3.38)	72	72	58
PSTM capacity	82.1 (17.22)	83.60 (16.40)	144	125	44

Note: Mean scores for pitch sensitivity were log-transformed from Hz; a lower score indicates a lower just noticeable difference for F0.

Results

Predictors

Descriptive statistics for the three processing resource measures were calculated for learners in the two contexts and are presented in Table 2. The mean F0 discrimination for the SA group was 14 Hz (log-converted 2.73), and 11 Hz (log-converted 2.62) for the AH group. On a perceptual scale for pitch, this difference is approximately one semitone apart. Scores were similar between groups, $t (38) = 0.197$, $p = .845$, $d = 0.06$, but participants exhibited a wide range of variation (1.5 Hz–48 Hz). Nonword discrimination accuracy was 93% (SA) and 94% (AH), and was also equivalent between groups, $t (38) = 1.47$, $p = .149$, $d = 0.46$. Mean scores for PSTM did not differ between the two groups, $t (38) = 0.057$, $p = .955$, $d = 0.02$, although there was again a wide degree of individual variation (31%–87% correct). These comparisons established the homogeneity of participants in both learning contexts on the processing measures, enabling their use as predictors in the subsequent regression models.

Lexical accent perception tasks

Accuracy on the two lexical accent perception tasks was next examined. On the PitchID task, the AH group had a mean accuracy of 51.7% (18.6 words) at Time 1 and 58% (20.9) at Time 2. The SA group had a mean accuracy of 46.5% (16.75) at Time 1 and 52.3% (18.85) at Time 2 on the same task. PitchCAT data showed that the AH group's mean accuracy was 53.8% (9.70 words) at Time 1 and 55.8% (10.05) at Time 2. By comparison, the SA group had an accuracy of 45% (8.1) at Time 1 and 49.7% (8.95) at Time 2. Results are shown in Table 3. Data for these tasks were normally distributed and contained no outliers, so we conducted paired-samples *t* tests to examine whether the accuracy gains made over the semester were significant for each group (Research Question 1). Both the AH group, $t (19) = 4.23$, $p < .01$, $d = 1.42$, 95% confidence interval (CI) [0.150, 3.35], and the SA group, $t (19) = 3.46$, $p < .01$, $d = 1.14$, 95% CI [0.803, 3.37], made significant gains at judging accent pattern correctness over the 12-week interval. However, on the visual categorization task neither the AH, $t (19) = 0.571$, $p > .05$, $d = 0.12$, 95% CI [−0.932, 1.63], nor the SA group, $t (19) = 1.65$, $p > .05$, $d = 0.35$, 95% CI [−0.227, 1.92], significantly improved.

Table 3. Group comparison of mean accuracies and gains on accent perception tasks

Variable	Maximum possible	SA ($n = 20$)		AH ($n = 20$)	
		Time 1	Time 2	Time 1	Time 2
PitchID	36	16.75 (1.74)	18.85 (1.93)	18.6 (1.09)	20.90 (2.02)
PitchCAT	18	8.1 (2.43)	8.95 (2.31)	9.70 (3.01)	10.05 (2.41)
PitchID gain	—	—	2.01 (2.77)	—	2.34 (2.41)
PitchCAT gain	—	—	0.85 (2.28)	—	0.36 (2.7)

Note: Gain scores = (T2 score – T1 score). *SD* shown in parenthesis.

Although the groups were matched on all processing resource predictors and learning experience, the heterogeneous learning backgrounds on lexical accent of the two groups necessitated the control of Time 1 performance when assessing the effect of context on Time 2 performance (Research Question 2). As an exploratory analysis, we constructed a mixed-effects logistic regression model of the log odds of correct identification on the PitchID task in R using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014). The model included fixed effects for pretest scores and context (treatment coded), and random intercepts for participants and items. The model ($n = 1{,}440$, log-likelihood $= -969$) showed a significant effect of pretest score ($\beta = 0.65$, $z = 6.10$, $p < .001$) and a marginally significant effect of context ($\beta = 0.21$, $z = 1.92$, $p = .056$) on Time 2 performance. A second model ($n = 1{,}440$, log-likelihood $= -971$) with only a fixed-effect term for pretest accuracy was then built. The fit of these two models was compared on a likelihood ratio test, which revealed that the model including context provided only a marginally better fit, χ^2 (1) $= 3.67$, $p = .055$, than the reduced model. This analysis indicates that after accounting for the variance in Time 1 scores, the context variable, defined as either increased usage (SA) or extensive correction (AH), had little effect on accent correctness judgments at Time 2.

For the PitchCAT task, a model was constructed with the same fixed and random effects structure. The model ($n = 720$, log-likelihood $= -487$) showed a significant effect of pretest score ($\beta = 0.62$, $z = 4.08$, $p < .001$) on Time 2 performance. Context was not significant ($\beta = 0.19$, $z = 1.25$, $p = .21$). A second model ($n = 720$, log-likelihood $= -488$), which included only pretest accuracy ($\beta = 0.63$, $z = 4.16$, $p < .001$) was compared to the full model. Model fit was again compared on a likelihood ratio test, which revealed no significant difference between the models, χ^2 (1) $= 1.55$, $p = 0.21$, indicating that accuracy at Time 2 did not differ between the AH and SA contexts.

Processing resource measures

Research Questions 3 and 5 asked what amount of variation in lexical accent gains over a semester of study could be predicted by the processing resource measures. Question 4 concerned the interaction between PSTM and context. To answer these questions, two generalized linear models were constructed to examine the variance accounted for by four predictors on gains in the two lexical perception tasks:

learning context, nonword discrimination ability, and PSTM capacity, and the interaction between PSTM and learning context. Because the two measures of auditory processing ability (F0 discrimination and nonword discrimination) were highly correlated ($r = -.539$) and thus measured the same construct, only the nonword discrimination task was used in the models.

Residualized change scores for accuracy were used as the dependent variable in the models, because the raw gain scores were correlated with pretest scores on both the PitchID ($r = -.491$) and PitchCAT ($r = -.588$) tasks (Dalecki & Willits, 1991). These were generated by regressing posttest scores onto pretest scores, and then finding the difference between the observed and predicted posttest scores. They represent the portion of the posttest score that is not predictable from the pretest score (Cronbach & Furby, 1970). The residualized scores were uncorrelated with pretest scores on both tasks and were used as the measure of change in the regression models. Assumptions for regression were met: residuals were normally distributed and variances were homogenous. In addition, given the comparatively small sample size ($n = 40$) for both perception tasks, we considered statistical power and effect size with regard to the number of predictors (four) in both models. Based on previous calculations, our sample met the requisite size of $n > 30$ to detect a medium to large effect size of $R^2 = .25$ (Cohen, Cohen, West, & Aiken, 2003). Furthermore, to make the models as robust as possible, we used a bootstrapping procedure with regression, which randomly resampled the original data 1,000 times to create new samples on which confidence intervals for the perceptual gains are based. This method may yield better estimates of the population values in small samples (Hinneburg, Mannila, Kaislaniemi, Nevalainen, & Raumolin-Bruberg, 2007).

Predictors were added sequentially to the model for PitchID in the following order. Learning context was entered first to control for differences in group performance and accounted for 14.3% ($\Delta R^2 = .143$; $\beta = .378$; $p = .016$) of the variance. The addition of the next two predictors was informed by the strength of correlation with the dependent measure. Nonword discrimination was added to the model but did not reach significance as a predictor ($\Delta R^2 = .011$; $\beta = .106$; $p = .487$). PSTM capacity was added next and accounted for 19.5% ($\Delta R^2 = .195$; $\beta = .443$; $p = .002$) of the variance. This indicates that, after controlling for differences in pretest performance, learning context, and auditory processing ability, a significant proportion of the variation in learners' accuracy gains on correctness judgments can be explained by differences in PSTM capacity. That is, higher PSTM capacity learners tended to improve more on correctness judgments over the test interval. Finally, the interaction term of PSTM and context was added, but failed to attain significance as a predictor ($\Delta R^2 = .019$; $\beta = .173$; $p = .315$). We can interpret this as showing no effect of differences in PSTM capacity on the ability to process pitch accent in contexts where input differs. In other words, it appears that learners with higher PSTM capacity make larger gains whether they study at home or abroad. In total, the regression model for PitchID with all four predictors entered accounted for 36.8% ($R^2 = .368$); $F (4, 35) = 5.10$; $p = .002$, of the total variance in change scores. The full model for PitchID gain with bootstrapped confidence intervals is shown in Table 4.

A second regression model was constructed for PitchCAT gains with similar parameters as the previous model. As with that model, residualized change score

Table 4. Multiple linear regression model for residualized change scores of PitchID

Predictor	B	SE	df	t	p	95% CI
Intercept	−5.32	1.42		−3.39	<.01	[−7.91, −3.58]
Context	−1.46	2.99	1,38	−0.47	.54	[−7.78, 3.64]
Nonword discrimination	0.08	0.20	1,37	0.47	.65	[−0.32, 0.50]
PSTM capacity	0.06	0.02	1,36	3.39	<.01	[0.03, 0.12]
PSTM × Context	0.03	0.03	1,35	1.01	.23	[−0.04, 0.11]

Table 5. Multiple linear regression model for residualized change scores of PitchCAT

Predictor	B	SE	df	t	p	95% CI
Intercept	−3.49	2.50		−1.75	.08	[−8.62, 0.83]
Context	2.94	4.85	1,38	0.75	.45	[−6.99, 12.37]
PSTM capacity	0.02	0.02	1,36	1.29	.20	[−0.29, 0.09]
Nonword discrimination	0.62	0.21	1,37	2.77	<.01	[0.24, 1.11]
PSTM × Context	−0.03	0.04	1,35	−0.86	.39	[−0.14, 0.07]

was the dependent variable. First, learning context was not a significant contributor to the model ($\Delta R^2 = .007$; $\beta = -.081$; $p = .619$). The addition of PSTM capacity did not account for any significant variation ($\Delta R^2 = .054$; $\beta = .211$; $p = .162$). Next, nonword discrimination score was added and explained 15.8% ($\Delta R^2 = .158$; $\beta = .398$; $p = .011$) of the variance in gain scores. Finally, the interaction of PSTM and context failed to account for any significant variance in the model ($\Delta R^2 = .016$; $\beta = .131$; $p = .396$). The finding that nonword discrimination ability accounted for a significant amount of variance in PitchCAT gains suggests that the processing of F0 variations is tied to learners' ability to assign visual category labels to the spoken stimuli. Altogether, the four predictors in the final regression model for PitchCAT explained 23.4% ($R^2 = .234$); $F (4, 35) = 2.67$; $p = .04$, of the variance in learners' gains in categorizing words by pitch accent pattern. The full model for PitchCAT gain is shown in Table 5.

Discussion

The goal of this study was to provide an account of the previously reported variation in the perceptual learning of Japanese lexical accent by L1 English-speaking learners. At-home and study-abroad learners performed two perception tasks twice during a semester of Japanese study. The main findings were as follows: (a) learners in each context made only small gains in perceptual accuracy over a semester; (b) when initial differences at the pretest were controlled for, Time 2 scores differed only marginally by learning context on the correctness judgment (PitchID) task, but not on the categorization (PitchCAT) task; (c) PSTM predicted gains on the PitchID task;

(d) there was no interaction between PSTM capacity and context; and (e) auditory processing ability predicted gains on the PitchCAT task. Each of these findings is discussed below.

The first research question addressed whether L2 Japanese learners' perception of lexical accent improves over a semester of instruction. The results indicated that, on average, learners made only small gains on the perception tasks over the 12-week learning period. Although numerical gains were found on both the PitchID and PitchCAT tasks, these attained statistical significance only in the PitchID task. This finding mirrors previous L2 research that has found that for most Japanese learners from non-tonal language backgrounds, the integration of pitch as a lexical property occurs, if it ever does, after a greater length of study than possessed by the present sample of learners (e.g., Lee et al., 2006; Nishinuma et al., 1996; Shibata & Hurtig, 2008; Taylor, 2011).

Although L1 Japanese speakers utilize pitch cues in lexical access and the parsing of larger prosodic groupings (Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999), it carries a comparatively lower functional load in Japanese than does tone in Mandarin Chinese (Schaefer & Darcy, 2014; Shibata & Shibata, 1990). Furthermore, the acoustic correlate of pitch accent, a fall in F0 over a two-mora unit, is known to be difficult for L1 English speakers to use reliably in word identification (Shport, 2016), because this cue is not particularly informative at the word level in their L1 (Ladd, 2008). For these reasons, L1 English speakers have difficulty reweighting their perceptual system.

Yet, when looking at learners in both contexts, variation in accuracy gains on both the PitchID (min/max gains as percent correct: –9% to 22%; overall accuracy range: 38%–69%) and PitchCAT tasks (min/max gains as percent correct: –22% to 33%; overall accuracy range: 22%– 78%) showed that some learners had begun the process of internalizing these accent patterns over the short learning interval. Previous studies have noted similar variation in accuracy gains, but these have primarily used training paradigms (Hirata, 1999; Shport, 2016), leaving the question of the effects of classroom instruction unanswered.

The second research question asked whether accuracy differed by learning context after a semester of study, given that the SA group reported greater Japanese input and usage. Because the effects of increased Japanese usage and pedagogical input cannot be disentangled with the current design, this comparison is exploratory, but merits consideration nonetheless. It was found that the AH group improved by an average of 6.5% on the PitchID task, while the SA group gained 5.6%. The logistic regression model for correctness judgments with pretest score and context as predictors revealed only a marginal effect of context on posttest performance. Although we cannot pinpoint the source of the effect with the present design, neither study abroad nor pedagogical input appeared to make much difference in lexical accent acquisition. After controlling for variance in pretest scores at the onset of study abroad, given the AH group's (+correction) better pretest performance, it was found that gains over 12 weeks of language study were of approximately the same magnitude for both groups. Similarly, for the PitchCAT task, a comparison of the models with and without context revealed no significant difference. Prior studies have shown the idiosyncratic effects of the SA environment on L2 acquisition, with gains reported in speech fluency and vocabulary use

(e.g., O'Brien et al., 2006, 2007), but not specifically for lexical accent production or perception (e.g., Lee et al., 2006; Taylor, 2011). The current results suggest that without having their attention directed to lexical accent through lab-based, high-variability training (Hirata, 1999; Shport, 2016), most learners do not improve much in their ability to perceive accent patterns.

The next three questions were of primary interest to this study. Two regression models explored the effects of learning context, processing resources and their inter-action with context, on perceptual gains over time. Research Question 3 examined the amount of variance in perceptual gains that PSTM capacity predicted over the learning period. The results showed that PSTM, as measured on a nonword recognition task, accounted for 19.5% of the unique variance (i.e., after partialling out context and other predictors) in gains on the PitchID task.

Learners with a higher capacity to temporarily store nonwords that adhered to Japanese phonotactics in PSTM showed greater development over the semester in their ability to judge the correctness of pitch accent patterns in real words. Previous research on the STM system and L2 acquisition suggests that high-span learners are better at the consolidation of L2 phonological form (e.g., Hummel, 2009; MacKay et al., 2001; Martin & Ellis, 2012; Sunderman & Kroll, 2009), a relationship which was obtained with explicit judgments on accent-pattern correctness in the current study. PSTM has been previously linked to multiple domains of L2 phonological acquisition. For example, Hummel (2009) found that PSTM capacity accounted for a significant proportion ($r^2 = .20$) of the variation in L2 English learners' vocabulary production in a lower proficiency group, but not in a higher proficiency group. It has also been found to be a significant predictor of the acquisition of L2 phonological structure (Nagle, 2013). Furthermore, MacKay et al. (2001) reported that nonword repetition scores accounted for 15% and 8% of the variance in error rates on identifying English word-final and word-initial consonants, respectively.

To accurately judge the correctness of a word's accent pattern in the current study, learners had to compare the stimuli with stored representations in their long-term memory. The ability to do this required accurate representations of the stimuli in memory. Thus, it appears that learners with a higher PSTM capacity were better at internalizing these accent patterns from spoken input, which enabled more accurate posttest judgments. The phonological short-term store appears to be invoked in the retention of not only segmental features but also linguistically relevant F0 information. The fact that the nonwords in the SNWR task phonotactically resembled the target-language stimuli suggests that performance on this task may have been conditioned by learners' knowledge of Japanese phonological patterning, which included lexical accent patterns. This would suggest that SNWR performance is a proxy for learning experience, which aligns with recent notions that PSTM is closely tied to long-term lexical and phonotactic knowledge (e.g., Andringa et al., 2012; Kaushanskaya, 2012; Speciale et al., 2004). It may be that learners with a higher PSTM capacity had more experience with Japanese, or a larger vocabulary, than those with lower capacity. Yet, this may not have been the case for the proficiency-matched learners in this study. Although lexical knowledge was not directly assessed, we found that PSTM capacity was uncorrelated with length of study ($r = .165$), self-rated proficiency ($r = .049$), or usage measures ($r = .089$),

suggesting that broadly defined Japanese experience was unrelated to phonological store capacity in this sample of learners.

The current finding provides a good comparison with Goss and Tamaoka's (2019) study on pitch accent perception in advanced L2 Japanese learners. In their study, PSTM was no longer a predictor of accuracy on perception tasks of similar design to the present study in advanced level learners. L2 vocabulary size, by extension a greater number of exemplars of word form in the long-term store, was a significant predictor of perception accuracy on both correctness judgment and categorization tasks. The fact that PSTM predicted the ability to judge pattern correctness in beginners in the current study, but not at the advanced level, suggests that this capacity operates as a domain-general processing mechanism in early-stage learning, but its role diminishes once a sufficient store of word forms is established through learning experience (Hummel, 2009; Martin & Ellis, 2012). Listening ability for this specific lexical cue was to a degree a function of the ability to process target-like stimuli in PSTM.

In addition, the fact that no interaction was found between this capacity and context (Research Question 4), suggests that high-capacity learners were better regardless of whether they had access to increased input or correction. A significant interaction would indicate that PSTM is connected to language experience, as learners with access to greater input in the SA context would likely have improved more than those in the AH context. However, it may be that higher capacity learners can better utilize any input, amount or quality notwithstanding. We cannot conclude from the current findings that PSTM is an independent processing mechanism in L2 acquisition, and this was not our aim. However, we can claim that this index of language-processing ability was related to gains on a lexical perception task and is thus involved in the consolidation of suprasegmental information in the long-term lexical store. Yet, we must point out that the short confidence interval suggests that the effect is comparatively small, and we caution against its overstatement.

Research Question 5 considered the construct of auditory processing ability as a predictor of gains in pitch accent perception. It was found that performance on a nonword discrimination task explained 15.8% of learners' pattern categorization development. That is, learners who were better at discriminating pitch contrasts in nonwords improved more on a task that required assigning visual category labels to the accent patterns of Japanese words. Despite a widely assumed dissociation of speech and nonspeech perception (Strange & Shafer, 2008), this finding suggests that a nonlinguistic perceptual capacity is involved to a degree in the categorization of lexically relevant F0 variations into discrete patterns. It also aligns with previous research showing broadly defined pitch perception ability to be implicated in the learning of tonal categories (e.g., Bent et al., 2006; Bowles et al., 2016; Wong & Perrachione, 2007). The current study extends this connection to the perception of lexical pitch accent.

Some evidence exists for overlapping auditory and language-specific pitch processing in Japanese. For instance, L1 Japanese speakers' sensitivity to F0 variations accounted for variation in real-word accent perception (Goss & Tamaoka, 2015). Musicianship may also be related to both pitch accent and lexical tone acquisition in L2 learners (Shport, 2016; Wong & Perrachione, 2007). The link between pitch discrimination in nonwords and lexical accent perception observed in the current

study can potentially be attributed to two factors. First, in a delayed categorization task, listeners may have based their decision on a trace of the stimulus' pitch contour in auditory memory (Fujisaki & Kawashima, 1969). Considering that the listeners were all beginners and may not have well-established representations of the pitch patterns in their long-term memory in the first place, those with greater auditory processing capacity could be expected to perform well on this task. The fact that this capacity was not operative in the PitchID task reinforces the differential roles of speech processing resources. Second, learners with greater auditory processing ability may have also been better at assigning visual labels, not entirely dissimilar to musical notations or accent pattern labels used in language textbooks, to the F0 variations of real words (Asaridou & McQueen, 2013). The H and L pitch sequences found in Japanese words were perhaps easier for those listeners who were better at equating the bitonal (H-L/L-H) pitch notations used in the PitchCAT task with the test stimuli, which suggests the intriguing possibility that nonlinguistic pitch cues can be exploited in pedagogical input to language learners.

Conclusion

Some limitations must be mentioned in interpreting the above findings. The relatively small sample size restricts its generalizability, and larger numbers would allow for more complex regression modeling. Next, because the same stimuli were used on the pre- and posttests, it is possible that learning was to an extent driven by these shared stimuli, although randomization of the stimuli and the 12-week test interval served to mitigate this issue. Relatedly, learners with a higher PSTM capacity may have been better at remembering specific stimuli to later learn their accent patterns, although none reported attempting to do so on a posttest questionnaire. However, assuming that gains on the perception tasks did evince learning, future studies should allow for generalization of learning to new stimuli or talkers, factors that add to the complexity of accent pattern learning for L1 English speakers (Hirata, 2015; Shport, 2016). In addition, a more robust test of the effects of learning context and pedagogical input should employ a fully crossed design by including +/–correction groups in both the SA and AH contexts. Testing the same predictors in a production task is also an important avenue to explore in future studies, given the effects of PSTM and auditory processing that were found.

Limitations notwithstanding, the present study extends previous findings on the role of individual difference measures to include links between PSTM capacity and auditory processing ability and the L2 acquisition of a pitch accent language. These factors may account for some of the previously reported variation in Japanese lexical accent learning (e.g., Nishinuma et al., 1996; Taylor, 2011). They also build upon theoretical accounts of cross-linguistic prosodic perception grounded in phonological features (Hallé, Chang, & Best, 2004), by adding processing resource constraints. Moreover, the task-specific effect of the predictors demonstrates the need for the inclusion of multiple measures in gauging L2 listening ability. Specifically, PSTM related to the consolidation of word form, as indexed by the PitchID task, while auditory ability predicted PitchCAT, which involved the abstraction of pitch patterns to visual schema. As learning to associate F0 variations

with a word's form is particularly challenging for L2 learners from nontonal L1s, techniques to supplement aural learning must be considered for low-PSTM learners (Hummel & French, 2010). This is not intended as evidence for the futility of incorporating suprasegmental instruction into the L2 classroom. To the contrary, acknowledging the fact that learners bring different capacities to the learning task can aid instructors in designing individualized training methods.

Notes

1. We selected the three most frequently occurring accent patterns in three- and four-mora words (Tanaka & Kubozono, 2012), although the total number of possible patterns is four and five (i.e., number of syllables $+$ 1), respectively.

2. F0 contours of the productions were checked in Praat and were found to adhere acoustically to the intended pattern (Boersma & Weenink, 2018). A previous study using the same speaker and words of similar frequency, length, and sentential contexts reported that L1 Japanese listeners could identify the accent patterns at a mean accuracy of 93% (Goss & Tamaoka, 2015).

References

Amano, N., & Kondo, K. (1999). *Lexical properties of Japanese: Vol. 1*. Tokyo: Sanseido.

Amano, N., & Kondo, K. (2000). *Lexical properties of Japanese: Vol. 7*. Tokyo: Sanseido.

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, **62** (Suppl.), 49–78.

Antoniou, M., & Chin, J. (2018). What can lexical tone training studies in adults tell us about tone processing in children? *Frontiers in Psychology*, **9**. doi: 10.3389/fpsyg.2018.00001

Asaridou, S., & McQueen, J. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Frontiers in Psychology*, **4**, 1–14.

Arai, M. (1997). The results of a longitudinal survey on the perception of the Tokyo accent: American learners living in Kyoto. *Spoken Japanese language education: Looking to the 21st century* (pp. 73–79). Tokyo: National Language Research Institute.

Ayusawa, T. (2003). Acquisition of Japanese accent and intonation by foreign learners. *Journal of the Phonetic Society of Japan*, **7**, 47–58.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, **36**, 189–208.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. **8**, pp. 47–90). New York: Academic Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, **105**, 158–173.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Version 1.1.–7. Retrieved from http://CRAN.R-project.org/package=lme4

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, **3**, 255–309.

Bent, T., Bradlow, A., & Wright, B. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, **32**, 97–103.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer, Version 6.0.39. Retrieved from http://www.praat.org/

Bowles, A., Chang, C., & Karuzis, V. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, **66**, 774–808.

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 259–280). Amsterdam: Benjamins.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology*, **32**, 867–873.

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Routledge.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, **26**, 227–248.

Collentine, J., & Freed, B. (2004). Learning context and its effects on second language acquisition. *Studies in Second Language Acquisition*, **26**, 153–171.

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *Journal of the Acoustical Society of America*, **131**, 4756–4769.

Cronbach, L., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin*, **74**, 68–80.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, **105**, 1877–1888.

Dalecki, M., & Willits, F. (1991). Examining change using regression analysis: Three approaches compared. *Sociological Spectrum*, **11**, 127–145.

Freed, B. (1995). Language learning and study abroad. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 3–33). Amsterdam: Benjamins.

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*, **28**, 67–73.

Gerrits, E., & Schouten, M. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, **66**, 363–376.

Goss, S. (2018). A critical pedagogy of lexical accent in L2 Japanese: Insights into research and practice. *Japanese Language and Literature*, **52**, 1–26.

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**, 143–154.

Goss, S., & Tamaoka, K. (2019). Lexical accent perception in highly-proficient L2 Japanese learners: The roles of language-specific experience and domain-general resources. *Second Language Research*, **35**, 351–376.

Grey, S., Cox, J., Serafini, E., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *Modern Language Journal*, **99**, 137–157.

Hallé, P., Chang, Y., & Best, C. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, **32**, 395–421.

Hinneburg, A., Mannila, H., Kaislaniemi, S., Nevalainen, T., & Raumolin-Bruberg, H. (2007). How to handle small samples: Bootstrap and Bayesian methods in the analysis of linguistic change. *Literary and Linguistic Computing*, **22**, 137–150.

Hirata, Y. (1999). *Acquisition of Japanese rhythm and pitch accent by English native speakers* (Unpublished doctoral dissertation, University of Chicago).

Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, **32**, 565–589.

Hirata, Y. (2015). L2 Phonetics and Phonology. In H. Kubozono, (Ed.), *Phonetics & Phonology Volume: The Handbook of Japanese Language and Linguistics*, 719–762. Berlin: De Gruyter Mouton.

Hu, C. (2003). Phonological memory, phonological awareness, and foreign language word learning. *Language Learning*, **53**, 429–462.

Hirose, Y., & Mazuka, R. (2015). Predictive processing of novel compounds: Evidence from Japanese. *Cognition*, **136**, 350–358.

Huensch, A., & Tracy-Ventura, N. (2017). Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, **38**, 755–785.

Hummel, K. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics*, **30**, 225–249.

Hummel, K., & French, L. (2010). Phonological memory and implications for the second language classroom. *Canadian Modern Language Review*, **66**, 371–391.

Ito, J., & Mester, A. (2016). Unaccentedness in Japanese. *Linguistic Inquiry*, **47**, 471–526.

Ito, K., & Speer, S. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, **58**, 541–573.

Jorden, E., & Noda, M. (1987). *Japanese: The spoken language, Part 1*. New Haven, CT: Yale University Press.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, **44**, 137–166.

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition*, **15**, 470–489.

Kitahara, M. (2001). *Category structure and function of pitch accent in Tokyo Japanese* (Unpublished doctoral dissertation, Indiana University, Bloomington).

Kormos, J., & Safar, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, **11**, 261–271.

Ladd, B. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.

Lee, W., Murashima, K., & Shirai, Y. (2006). Prosodic development in the acquisition of Japanese: A longitudinal study of three native speakers of Cantonese. *Japan Journal*, **10**, 38–51.

MacKay, I., Meador, D., & Flege, J. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, **58**, 103–125.

Mandel, J. (2009). Adaptive pitch test. Available from http://tonometric.com/adaptivepitch/

Martin, K., & Ellis, N. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, **34**, 379–413.

Nagle, C. (2013). A reexamination of ultimate attainment in L2 phonology: Length of immersion, motivation, and phonological short-term memory. In E. Voss, S.-J. D. Tai, & Z. Li (Eds.), *Selected proceedings of the 2011 Second Language Research Forum* (pp. 148–161). Somerville, MA: Cascadilla Proceedings Project.

Nishinuma, Y., Arai, M., & Ayusawa, T. (1996). *Perception of tonal accent by Americans learning Japanese*. Paper presented at the 4th International Conference on Spoken Language Processing, Philadelphia, PA.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, **27**, 377–402.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains. *Studies in Second Language Acquisition*, **29**, 557–581.

Perrachione, T., Lee, J., Ha, L., & Wong, P. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America*, **130**, 461–472.

Sakamoto, E. (2010). *An investigation of factors behind foreign accent in the L2 acquisition of Japanese lexical pitch accent by adult English speakers* (Unpublished doctoral dissertation, University of Edinburgh).

Schaefer, V., & Darcy, I. (2014). Linguistic prominence of pitch within the native language determines accuracy of tone processing. In R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. M. Miguel, A. M. Tseng, . . . D. Walter (Eds.), *Selected proceedings of the 2012 Second Language Research Forum* (pp. 1–14). Somerville, MA: Cascadilla Proceedings Project.

Sekiguchi, T., & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, **28**, 439–454.

Shibata, T., & Hurtig, R. (2008). Prosody acquisition by Japanese learners. In Z. Han. (Ed.), *Understanding second language process* (pp. 176–204). Clevedon: Multilingual Matters.

Shibata, T., & Shibata, R. (1990) Akusento wa doo'ongo o donoteido benbetsu shiuruno ka? Nihongo, eigo, chuugokugo no baai [How much can accent distinguish homophones? Cases of Japanese, English and Chinese]. *Mathematical Linguistics*, **17**, 317–327.

Shibatani, M. (1990). *The languages of Japan.* Cambridge: Cambridge University Press.

Shport, I. (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). München: Iudicium Verlag.

Shport, I. (2015). Perception of acoustic cues to Tokyo Japanese pitch-accent contrasts in native Japanese and naïve English listeners. *Journal of the Acoustical Society of America*, **138**, 307–318.

Shport, I. (2016). Training English listeners to identify pitch accent patterns in Tokyo Japanese. *Studies in Second Language Acquisition*, **38**, 739–769.

So, C., & Best, C. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, **53**, 273–293.

Speciale, G., Ellis, N., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, **25**(2), 293–321.

Strange, W., & Shafer, V. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–191). Philadelphia: Benjamins.

Sunderman, G., & Kroll, J. (2009). When study-abroad experience fails to deliver: The internal resource threshold effect. *Applied Psycholinguistics*, **30**, 79–99.

Tamaoka, K. (2014). The Japanese writing system and lexical understanding. *Japanese Language and Literature*, **48**, 431–471.

Tanaka, S., & Kubozono, H. (2012). *Nihongo no hatsuon kyooshitsu: Riron to renshuu [Introduction to Japanese pronunciation: Theory and practice].* Tokyo: Kurosio.

Taylor, B. (2011). Variability and systematicity in individual learners' Japanese accent. *Pozanan Studies in Contemporary Linguistics*, **47**, 146–158.

Tsurutani, C. (2011). L2 intonation: A study of L1 transfer in Japanese intonation by English-speaking learners. *Second Language*, **10**, 79–102.

Wayland, R., & Guion, S. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, **54**, 681–712.

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, **38**, 654–662.

Wen, Z. (2014). Theorizing and measuring working memory in first and second language research. *Language Teaching*, **47**, 174–190.

Williams, J., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, **53**, 67–121.

Wong, P., & Perrachione, T. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, **28**, 565–585.

Appendix A

Target words (underlined) with carrier sentences used in the PitchID/PitchCAT tasks

あちらに行く	*atira ni iku*	"(I) go over there."
タクシーに乗る	*takusii ni noru*	"(I) take a taxi."
仕事をする	*sigoto o suru*	"(I) go to work."
花屋に行く	*hanaya ni iku*	"(I) go to the flower shop."
土曜日に行く	*doyoobi ni iku*	"(I) go on Saturday."
本屋に行く	*honya ni iku*	"(I) go to the bookstore."
あさってからする	*asatte kara suru*	"(I) will do it the day after tomorrow."
家族に言う	*kazoku ni iu*	"(I) tell my family."
飛行機に乗る	*hikooki ni noru*	"(I) ride an airplane."

(Continued)

(Continued)

ビールを飲む	biiru o nomu	"(I) drink beer."
あなたがいる	anata ga iru	"(You) are here."
デパートで買う	depaato de kau	"(I) buy it at the department store."
新聞を読む	sinbun o yomu	"(I) read the newspaper."
八時に来る	hatizi ni kuru	"(I) come at 8 o'clock."
二十歳になる	hatati ni naru	"(I) turn twenty."
今晩までいる	konban made iru	"(I) will stay until tonight."
フランスに行く	furansu ni iku	"(I) go to France."
手前で止まる	temae de tomaru	"(I) stop in front (of someplace)."
ベーコンを食べる	beekon o taberu	"(I) eat bacon."
砂糖を使う	satoo o tukau	"(I) use sugar."
手紙を書く	tegami o kaku	"(I) write a letter."
赤ちゃんがいる	akatyan ga iru	"(I) have a baby."
たまごを食べる	tamago o taberu	"(I) eat eggs."
グリーンもある	guriin mo aru	"There are green (ones) also."
ホテルに泊まる	hoteru ni tomaru	"(I) stay in a hotel."
外国に行く	gaikoku ni iku	"(I) go abroad."
名前を呼ぶ	namae o yobu	"(I) call (someone's) name."
図書館に行く	tosyokan ni iku	"(I) go to the library."
彼女がいる	kanojo ga iru	"(She) is here/(I) have a girlfriend."
グレーもある	guree mo aru	"There are grey (ones) also."
両親に会う	ryoosin ni au	"(I) meet my parents."
左へ曲がる	hidari e magaru	"(I) turn left."
弁当を買う	bentoo o kau	"(I) buy a boxed lunch."
毎日を楽しむ	mainiti o tanosimu	"(I) enjoy each day."
勉強をする	benkyoo o suru	"(I) study."
地下鉄に乗る	tikatetu ni noru	"(I) ride the subway."