Japanese Psychological Research | THE JAPANESE PSYCHOLOGICAL ASSOCIATION

*Japanese Psychological Research*
2015, Volume 57, No. 2, 143–154

doi: 10.1111/jpr.12076

# Predicting lexical accent perception in native Japanese speakers: An investigation of acoustic pitch sensitivity and working memory[1]

SETH GOSS[2]* *Ohio State University*

KATSUO TAMAOKA *Nagoya University*

**Abstract:** Spoken language perception may be constrained by a listener's cognitive resources, including verbal working memory (WM) capacity and basic auditory perception mechanisms. For Japanese listeners, it is unknown how, or even if, these resources are involved in the processing of pitch accent at the word level. The present study examined the extent to which native Japanese speakers could make correctness judgments on and categorize spoken Japanese words by pitch accent pattern, and how verbal WM capacity and acoustic pitch sensitivity related to perception ability. Results showed that Japanese listeners were highly accurate at judging pitch accent correctness ($M = 93\%$), but that the more cognitively demanding accent categorization task yielded notably lower performance ($M = 61\%$). Of chief interest was the finding that acoustic pitch sensitivity significantly predicted accuracy scores on both perception tasks, while verbal WM had a predictive role only for the categorization of a specific accent pattern. These results indicate first, that task demands greatly influence accuracy and second, that basic cognitive capacities continue to support perception of lexical prosody even in adult listeners.

**Key words:** speech perception, lexical accent, pitch sensitivity, verbal working memory.

Perception of spoken language is a complex task that may be supported by a listener's cognitive resources. Two such resources purportedly involved in the processing of speech input are nonlinguistic, or acoustic, pitch sensitivity and verbal working memory (WM). The assumption is that those who are more sensitive to subtle acoustic variations (e.g., Deutsch, Dooley, Henthorn, & Head, 2009; Wayland, Herrera, & Kaan, 2010) or who possess a greater capacity to temporarily maintain aural input in WM (e.g., Baddeley, Gathercole, & Papagno, 1998; Kaushanskaya & Yoo, 2013) are capable of utilizing these abilities when processing spoken language as well. In other words, measures of basic, nonlinguistic capacities may predict how accurate one is in processing certain features of spoken language. In the case of Japanese, which is a lexically accented language, these cognitive abilities may continue to play a role in native

*Correspondence concerning this article should be sent to: Seth Goss, Department of East Asian Languages and Literatures, Ohio State University, Columbus, OH 43210, USA. (Email: goss.33@osu.edu)

[1]The first author was supported financially by a Japan Foundation Dissertation Fellowship while conducting this research.

[2]We are grateful to Nobuhiro Saito for recording test stimuli and Sachiko Kiyama for suggesting the adaptive pitch test. We also thank Kyoko Hayakawa and Rinus Verdonschot for their assistance with participant recruitment and software expertise.

(L1) listeners' ability to perceive spoken word-level accents. Despite the general assumption that speech perception in the L1 is robust, automatic, and places little demand on cognitive resources (e.g., Strange & Shafer, 2008), variation in lexical accent perception ability among Tokyo-standard Japanese speakers has been reported in previous research (e.g., Hirano-Cook, 2011; Shport, 2008). Some of this variation is likely a consequence of the use of experimental tasks that vary in the demands placed on the listener and the presence of low-frequency stimuli, both of which have been noted to influence speech perception (Strange & Shafer, 2008). However, differences in individual listeners' cognitive capacities may underlie a significant portion of the variation in perception accuracy. However, to our knowledge no link has yet been established in the literature between basic cognitive resources and the perception of lexically accented speech. To address this gap, the present study first compared L1 Japanese listeners' performance on a two-phase lexical accent perception measure (a correct or incorrect judgment followed by a categorization task) neither of which have been measured together in the same listener population. Furthermore, lexical frequency was controlled in order to minimize perception difficulty arising from low-frequency stimuli. We then measured participants' verbal WM and acoustic pitch sensitivity to explore how basic cognitive abilities relate to perception ability on the two accent perception tasks.

In Japanese, pitch accent is specified in the lexicon as a part of word form (Beckman & Pierrehumbert, 1986). Accent location is marked by a single phonetic cue, which acoustic models of pitch accent indicate is a lexically linked High (H) tone followed by a drop to Low (L) pitch (Venditti, 2005). For example, a three-mora Japanese noun can have n + 1 pitch patterns, as shown in the words below (examples from Vance, 2008).[3]

1 *makura wa* (MAkura-wa) "pillow + TOPIC marker"
2 *tamago wa* (taMAgo-wa) "egg + TOPIC marker"
3 *takara wa* (taKARA-wa) "treasure + TOPIC marker"
4 *sakana wa* (saKANA-WA) "fish + TOPIC marker"

These lexically linked accents are thought to be stored in a native speaker's mental lexicon in combination with a word's segmental form (e.g., Beckman & Pierrehumbert, 1986; Otake & Cutler, 1999). Accordingly, when native Japanese speakers process spoken language, pitch accents are likely activated simultaneously with lexical form. It follows then that L1 speakers should be able to accurately perceive whether aurally presented words were spoken with the correct or incorrect accent pattern in the same way that they can make decisions on features such as vowels and consonants. In fact, Shibata and Hurtig (2008) reported that adult L1 Japanese speakers made correctness judgments on high-frequency spoken words with a 97% accuracy rate. Ueno et al. (2014) also found that L1 Japanese could perceive, then imitate aloud, spoken words containing both correct and incorrect accents with near-ceiling accuracy. However, perception tasks that require listeners to identify accent location or categorize words by accent type have produced notable performance variation. For instance, Shport (2008) noted that Japanese listeners only attained a mean accuracy of 59% in identifying pitch accent location on an AXY perception task. The task required participants to listen to a set of three words, one of which (either X or Y) differed in pitch accent from the other two, and decide if the odd-one-out was the X or Y stimulus. In a similar type of prominence identification task, Hirano-Cook (2011) also showed a lower than expected accuracy rate of 70% in L1 listeners' ability to mark the accent location of spoken words on Japanese-script transcriptions. Despite both researchers acknowledging the inherent task difficulty, this poor performance on categorization-type tasks led them to

[3]Mora pronounced with a high pitch are indicated in capital letters. Accent location is perceived by Japanese speakers on the high-pitched mora immediately preceding the fall to low pitch.

conclude that pitch accent is a relatively minor aspect of lexical form.

Methodologically, both the studies by Shport (2008) and Hirano-Cook (2011) used low-frequency words, and their perception tasks, which required listeners to identify accent location or compare two or more pitch patterns, were likely more cognitively demanding than a correctness judgment or repetition task. Strange and Shafer (2008) noted that experimental variables greatly influence perception performance, in that L1 listeners can make correctness judgments based on existing knowledge of phonological categories, while categorization tasks, which require explicit acoustic comparison or metalinguistic knowledge of the features being tested, may involve additional processing resources. Therefore, in the present study both a correctness judgment and a categorization task were conducted with the same group of participants, with the aim of first exploring how cognitive resources are implicated in these disparate tasks types, and second how this two-step decision process influences accuracy.

Recent research has attempted to shed light on basic cognitive abilities that support spoken language processing (e.g., Wayland et al., 2010; Wong & Perrachione, 2007). Next, we will elaborate on two such abilities that may be involved in the processing of lexical pitch accent: verbal WM and acoustic pitch sensitivity.

Verbal WM and its correlate the phonological loop have been shown to be implicated in a range of language tasks including vocabulary acquisition (Gathercole, Willis, Emslie, & Baddeley, 1992), reading (Jincho, Namiki, & Mazuka, 2008), and spoken language comprehension (Adams, Bourke, & Willis, 1999). The phonological loop is the subcomponent of WM that is responsible for the short term storage and maintenance of phonological information derived from both spoken and visual (i.e., text or namable images) input (Baddeley et al., 1998). One fundamental role is to therefore maintain speech input for a short duration, so that more in-depth processing can be performed. In the case of Japanese, this entails that pitch accent information is held in the short-term phonological store, which may be implicated in lexical processing tasks. Furthermore, listeners with a large store capacity are likely capable of accurately retaining more sound information when making subsequent lexical judgments involving pitch accent. While much research has examined the role of verbal WM in the processing of vocabulary and syntax, to our knowledge no studies have sought to demonstrate the possible involvement of WM in the maintenance of lexical-level accent. Despite the absence of research, Chan, Ho, and Cheung (1998) demonstrated that pitch sensitivity gained through musical training improved adult Chinese speakers' verbal memory for native vocabulary. In their study, participants with music training scored 16% higher than nonmusician controls on a lexical recall (i.e., verbal memory) task featuring orally presented words. Their finding suggests shared memory resources for nonlinguistic pitch and tone-carrying words in a tone language. Semal, Demany, Ueda, and Halle (1996) found that on a speech memory task, listeners processed pitch variations in speech and nonspeech sounds in a similar manner, and concluded that both types of sound are handled similarly by the short-term memory system.

In Japanese, lexical pitch accent may play a facilitative role in both word and nonword retention in verbal WM. On a verbal WM task, Yuzawa and Saito (2006) reported that 3- to 4-year-old Japanese children showed superior recall ability for nonwords that were spoken with an existing pitch pattern as opposed to those produced without pitch variation. It is important to note here that all words in Tokyo-standard Japanese feature some type of pitch variation, even the so-called unaccented pattern (example 4, above). Prosody has also been shown to facilitate recall in real-word stimuli in young children, but not in older children (e.g., Roy & Chiat, 2004; Yuzawa, 2002).

Although previous research has suggested a facilitative effect of lexical accent on recall in memory tasks, the opposite connection, namely, verbal WM supporting the processing of accented words, has yet to be established. Considering the earlier discussion of pitch accent as an inherent part of word form in Japanese, in the present study we assumed that verbal WM capacity will be predictive of performance on a cognitively demanding perception task such as lexical categorization, which requires an explicit phonetic analysis of the target while maintaining pitch information in memory. In addition, the use of lexically accented nonwords in the WM task is likely a critical factor if we are to compare serial recall capacity with performance on real-word perception tasks, as the recall of accented nonwords may relate to the perception of accented speech (Yuzawa & Saito, 2006).

The second cognitive capacity of interest is acoustic pitch perception. Research from the field of cognitive science has suggested a link between a listener's sensitivity to acoustic pitch and the capacity to accurately perceive speech sounds by adult L1 listeners. Liu, Patel, Fourcin, and Stewart (2010) have reported evidence of shared structures for music and language perception in the brain. In their study, native speakers of English who claimed to be "tone-deaf" (congenital amusica) lagged behind matched normal participants in their ability to correctly identify intonation patterns in their native language, despite all participants reporting normal communicative ability in their L1. Nan, Sun, and Peretz (2010) added support to these findings by demonstrating that even native speakers of a tone language (Mandarin Chinese) who reported tone deafness also displayed impaired processing of L1 tone patterns, again in the presence of normal production ability. Further, in a normal adult population, Bent, Bradlow, and Wright (2006) demonstrated that L1 Chinese speakers consistently made perception errors on nonspeech pitch contours that were phonetically similar to L1 tone patterns, indicating that these listeners processed speech and nonspeech pitch with shared perceptual resources.

However, a contrary viewpoint posits that perception for speech and nonspeech features is separated very early in life, as a result of the tuning-in to one's native language sound categories, which in turn facilitates acquisition of L1 phonological regularities (e.g., Werker & Curtin, 2005). Models of speech perception such as the native language magnet (NLM) theory state that as we gain experience with our first language, our perceptual space is molded to the sound categories of that language (Iverson & Kuhl, 1996). Accordingly, this would predict dissociation between the processing of nonspeech sounds and performance on L1 speech perception tasks by adult listeners. In other words, listeners may perceive unknown phonological categories like nonspeech or non-native tones acoustically, while perceiving speech sounds linguistically. Yet the relationship between the type of acoustic sensitivity measure and the speech feature under investigation (i.e., lexical pitch) may hinge on the phonetic similarity between the stimuli in both tasks. Namely, an acoustic sensitivity measure resembling the bitonal (Low-High/High-Low) pitch accent structure of Japanese may be predictive of performance on word perception tasks. Therefore, our third aim is to examine the relationship between acoustic pitch sensitivity and the processing of real-word stimuli by L1 Japanese listeners.

In sum, the present study aims to investigate the relationship between basic cognitive abilities and the perception of lexically accented words in two speech perception tasks by adult L1 Japanese speakers. We first predicted that adult listeners can reliably use pitch accent to judge the correctness of spoken word accents, but that accent categorization, which involves greater processing demands, will yield more individual variation. Secondly, as Japanese pitch accent fundamentally consists of contrasting high and low tones, listeners who possess greater acoustic pitch sensitivity will perform better on both lexical perception tasks. Lastly, verbal WM span will be implicated in the memory-resource demanding accent categorization task, but not in the speeded correctness judgment task.

## Method

### *Participants*

Thirty native Japanese speakers (19 female and 11 male) with a mean age of 19.2 years

($SD = 0.78$) at a research university in Japan participated in the experiment. The participants indicated on a questionnaire that they were from the Tokai region in central Japan, which is classified as a Tokyo-standard accent region (Shibatani, 1990). Participants reported normal hearing, and none had an extended length of musical training (≤3 years) or fluency in a tone language. All participants were volunteers and were compensated for their participation.

### Materials

*Verbal WM.* The serial nonword recognition (SNWR) task was first used by O'Brien, Segalowitz, Collentine and Freed (2006) as a measure of phonological short-term memory, and provides an alternative to the conventional nonword repetition (NWR) tasks that require vocalization of nonword stimuli. Eliminating the production aspect may enable a more accurate measure of verbal WM capacity. In contrast with the English-based nonwords used by O'Brien et al. (2006), in the current study stimuli for the SNWR task were nonwords based on the Japanese mora system, of the structure /(C)VCV/. Nonword stimuli were recorded by a speaker of Tokyo-standard Japanese and were spoken with an unaccented Low-High pitch pattern (see example 4 in Introduction). We chose to record the nonword stimuli with an existent accent pattern because our aim was to measure the relationship between WM capacity and the processing of lexically accented words, on the assumption that pitch-carrying nonwords will better predict real-word task performance. The task design is as follows. Two sets consisting of the same number of nonwords were played to participants with a 1500 ms pause between sets. The sets increased from four items in the practice phase to five, six, and seven nonwords for the test phase. The interstimulus interval (ISI) between the nonwords in each list was 750 ms. Half of the sets at each length consisted of "same" trials and the other half of "different" trials. In the different trials, the order of two of the nonwords was switched. However, the first and last items of the list were never switched. The participants' task was to decide if both sets were in the same or different order, requiring them to keep track of the serial order of the nonwords in order to make this decision.

Sample nonword sets (5 nonwords in length). Switched items are in italics.

| | | | | | |
|---|---|---|---|---|---|
| (Same) | gohe | zuka | imyo | heji | baro |
| | | | | | <pause> |
| | gohe | zuka | imyo | heji | baro |
| (Different) | chida | *ruge* | *hami* | jare | kebu |
| | | | | | <pause> |
| | chida | *hami* | *ruge* | jare | kebu |

*Pitch sensitivity.* An adaptive pitch test that increased in difficulty based on participants' performance was used as the measure of acoustic pitch sensitivity. This task is web-based and is similar in format to commonly used AX discrimination tasks, in that the first pitch stimulus (A) remained constant, while the second stimulus (X) varied by a predetermined parameter (Mandel, 2009). In this task specifically, the first tone of the two-item pair was a 500 Hz pure tone, and the second tone differed by set intervals of hertz (96, 48, 24, 12, 6, 3, 1.5 Hz, etc.), with the between-stimulus pitch difference either increasing or decreasing based on participants' responses. For example, at the 12 Hz interval, the first tone was 500 Hz and the second tone 512 Hz. Each tone was 500 ms in length and tones were separated by a 250 ms pause. Loudness was adjusted to an adequate level for each participant prior to the task. The resulting score represented the degree to which a listener could reliably distinguish the paired tones, with greater sensitivity indicated by a lower score.

*Pitch accent perception.* Two perception tasks using Japanese sentences were the dependent variables. First, participants judged if a noun followed by a postposition was spoken with the correct or incorrect accent pattern (below, PitchID). Second, for the correct items only, they categorized the noun-plus-postposition into one of four pitch contours in a four-alternative forced choice (AFC) task (PitchCAT). A total of 32 test sentences were created, all of which were of the structure

(N + postposition + V, e.g., 一人で行く, *Hitori de iku*, "I go alone"). Half of these stimuli ($N = 16$) were spoken with the correct accent and the other half with an incorrect accent. Stimuli were composed of a mixture of three- and four-mora nouns featuring one of four accent patterns (for 3-mora words: HLL-*L*, LHL-*L*, LHH-*L* and LHH-*H*; 4-mora words: HLLL-*L*, LHLL-*L*, LHHH-*L*, LHHH-*H*).[4] Words were controlled by frequency across each of the four accent types based on the NTT database of approximately 300 million words (Amano & Kondo, 2000). Total mean normative frequency for all stimuli was 36.6 per million ($SD = 52.6$) and by pattern as follows: Pattern 1 ($M = 35.7$, $SD = 54.0$), Pattern 2 ($M = 15.1$; $SD = 15.1$), Pattern 3 ($M = 47.7$, $SD = 67.5$), Pattern 4 ($M = 46.3$; $SD = 61.4$). Although the normative frequencies appeared to vary between patterns, this difference was not significant, $F(3, 28) = 0.633$, $p = .600$. Accent validity ratings were also checked in the same database, which revealed that three words failed to reach 100% inter-rater agreement among Tokyo-dialect speakers, indicating the presence of a variant accent pattern for each of these items (Amano & Kondo, 1999). We therefore removed these words from our analysis of the PitchCAT task, yielding 13 items, as listeners may have categorized targets based not on the accent of the spoken stimuli, but on their own lexical representation.[5] Lastly, all carrier sentences were controlled to a mean length of 6.7 mora ($SD = 0.52$). Refer to Appendix for a complete list of test words, including the excluded items.

Stimuli were sampled at 44 100 Hz by a native speaker of Tokyo-standard Japanese with extensive training in accent production. No acoustic modification was made to the test stimuli, although, pitch waveforms were checked in Praat sound-analysis software to confirm that accent location matched the intended pattern (Boersma & Weenink, 2008). Participants were instructed to focus on the sentence-initial (N + postposition) in making their decisions. After a correctly accented word was presented and the correctness judgment made, four pitch graphs corresponding to the four possible pitch contours were immediately displayed onscreen. Figure 1 shows an example of the categorization graphs. The participants' task was to select the pitch graph that they felt most closely fit the accent pattern of the spoken target.

### *Procedure*

Participants completed four tasks in the following order: (a) acoustic pitch sensitivity test, (b) SNWR task (verbal WM), (c) the combined lexical accent correctness judgment (PitchID),

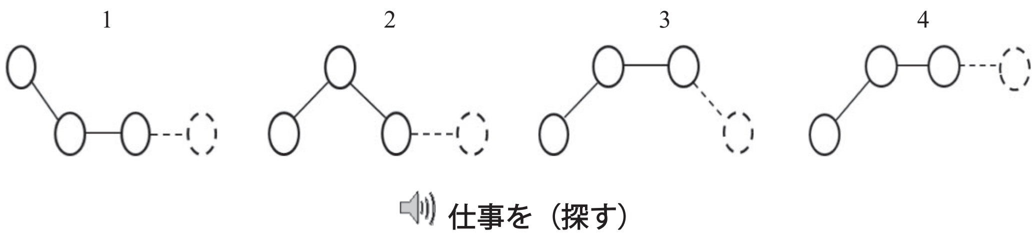


**Figure 1** In the PitchCAT task, listeners categorized the noun-plus-postposition (/shigoto o sagasu/ "(I) search for a job") into one of four pitch contours by selecting a visual representation of the pitch pattern. Only the target noun-plus-preposition *shigoto o* was presented visually.

[4]Note that for 4-mora words, a fifth pattern (LHHL-L) exists in Tokyo-standard Japanese, but was not used in order to maintain the structure of the 4-response PitchCAT task.

[5]We thank an anonymous reviewer for confirming the accent validity ratings. Note that all 32 words were retained for the PitchID task, as their removal did not affect perception accuracy.

**Table 1** Descriptive statistics for all variables

| Variable | Max | *M* | *SD* | Range | |
|---|---|---|---|---|---|
| Pitch sensitivity | – | 2.49 | 1.44 | 5.00 | –0.91 |
| Verbal working memory | 144 | 84.07 | 17.34 | 50.00 | 117.00 |
| PitchID | 32 | 29.77 | 1.63 | 26.00 | 32.00 |
| PitchCAT | 13 | 7.93 | 2.97 | 3.00 | 13.00 |
| PitchID reaction time (ms) | – | 2,215.00 | 233.00 | 1816.00 | 2683.00 |

*Note*. Three items were removed from the original 16 words in the PitchCAT analysis.

and (d) categorization (PitchCAT) tasks. Besides the web-based pitch sensitivity test, the stimuli were presented in E-Prime 2.0 experimental software. Responses were made via a response box, with the WM and PitchID tasks requiring a two-button Yes/No answer, and the PitchCAT task using a four-button response. Reaction time (RT) was measured on the PitchID task, and participants were instructed to respond as rapidly as possible after the spoken stimulus ended. Tasks were presented aurally with headphones in a quiet room. All tasks were preceded by a short practice phase to familiarize participants with the procedures, with further explanation provided by the researcher if participants were unclear about the task procedures. Due to the complexity of the PitchCAT task in particular, participants were familiarized with sample pitch contour graphs (Figure 1) similar to those in the main categorization task during the practice phase. All participants stated that they were clear with the categorization process prior to beginning the main experiment. Participants took approximately 20 min to complete all experimental tasks.

## Results

Descriptive results for all tasks in the order of presentation are shown in Table 1. Pitch sensitivity scores reflected the average threshold at which participants were able to distinguish two pure tones. Thus, the lower the numerical score, the more sensitive the listener was to pitch variations. Scores were $log_2$-converted for analysis in order to regularize the Hz intervals and obtain a more normal distribution. Verbal WM was scored by a weighted measurement, assigning points to correct responses based on the length of the sets, with the longer sets (i.e., those containing more nonwords) yielding a higher point total (see O'Brien et al., 2006 for the scoring procedure).

Participants successfully judged 93% of the spoken noun-plus-postposition targets as having either the correct or incorrect accent pattern (PitchID). However, only 61% of the correctly accented words were accurately categorized by visual pattern on the four-choice categorization task (PitchCAT). The RT in milliseconds from the onset of the audio in the PitchID task was also measured as a predictor for PitchCAT performance.

Correlation coefficients for the predictor variables and the PitchID and PitchCAT tasks are presented in Table 2. Significant negative correlations were found between pitch sensitivity in both the PitchID ($r = -.42$, $p < .05$) and PitchCAT tasks ($r = -.42$, $p < .05$), indicating that those with a lower acoustic sensitivity threshold were more accurate on the spoken word perception tasks. However, no significant correlations were found for verbal WM capacity on either task. The RT for PitchID was also negatively correlated with participants' performance on the PitchCAT task, suggesting that listeners who made faster correctness judgments were in turn more accurate at categorizing accent patterns. Note also that PitchID accuracy was not correlated with PitchCAT score, a finding which will be discussed later.

A multiple regression analysis was next conducted with pitch sensitivity and verbal WM entered as predictors for PitchID performance. As shown in Table 3, these two variables accounted for 27% of the variance in participants' accent correctness judgments, with pitch

**Table 2** Correlation matrix for all test variables

| Variable | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. Verbal working memory | – | | | | |
| 2. Pitch sensitivity | .28 | – | | | |
| 3. PitchID | .18 | –.42* | – | | |
| 4. PitchCAT | –.11 | –.42* | .01 | – | |
| 5. PitchID reaction time | .04 | –.20 | –.06 | –.39* | – |

*$p < .05$.

**Table 3** Multiple regression results for the PitchID and PitchCAT tasks

| Task | Predictors | $R$ | $R^2$ | $\beta$ | $t$ |
|---|---|---|---|---|---|
| PitchID | | .52 | .27* | | |
| | Pitch sensitivity | | | –.51 | –2.99** |
| | Verbal working memory | | | .32 | 1.87 |
| PitchCAT | | .53 | .28* | | |
| | Pitch sensitivity | | | –.35 | –1.99* |
| | PitchID reaction time | | | –.33 | –1.91 |
| | Verbal working memory | | | .03 | 0.16 |

**$p < .01$. *$p < .05$.

sensitivity being identified as a highly significant predictor ($\beta = -.51$, $p < .001$), and verbal WM capacity approaching significance ($\beta = .32$, $p = .073$). Likewise, we created a separate regression model with three predictors for PitchCAT: pitch sensitivity, verbal WM, and PitchID RT. Table 3 shows that these three variables accounted for 28% of the variance in accent categorization, with participants' pitch sensitivity score again a significant predictor ($\beta = -.35$, $p < .05$) and RT approaching significance ($\beta = -.33$, $p = .067$). However, verbal WM failed to attain significance as a predictor ($\beta = .03$, $p = .875$).

We next wished to explore listeners' error patterns on the perception tasks, as specific accent types may have influenced overall performance. Table 4 provides accuracy scores by accent pattern for both the PitchID and PitchCAT tasks. On the PitchID task, accuracy scores on Pattern 3 words (82.5%) were markedly lower than for the other three accent patterns. For the four-choice categorization task, initial-mora accented Pattern 1 (HLL-L) words were categorized with 72.5% accuracy, but noticeably dropped on the remaining patterns.

Although the predicted correlation between verbal WM capacity and the more demanding PitchCAT task was not found, we next wanted to determine whether WM related to accuracy on any specific accent pattern in this task. When analyzed by pattern, a significant correlation ($r = .476$, $p < .01$) was found only for Pattern 4 (unaccented, or LHH-H) words, suggesting that memory span for unaccented nonwords related to categorization ability of real words of the same pitch pattern. To further explore performance on the unaccented stimuli in the PitchCAT and PitchID tasks, we then separated participants into three even groups ($n = 10$) according to the highest ($M = 102.9$, $SD = 9.43$), middle ($M = 83.8$; $SD = 5.26$), and lowest ($M = 65.5$; $SD = 8.68$) verbal WM scores, $F(2, 27) = 54.61$, $p < .001$. A one-way ANOVA revealed that the WM groups differed significantly in their categorization of Pattern 4 words, $F(2, 27) = 4.049$, $p < .05$, with the High-span group ($p < .05$) and Mid-span group ($p = .064$) outperforming the Low-span group. By contrast, no group difference was found for Pattern 4 words on the PitchID task, $F(2, 27) = 1.00$, *ns*, likely due to a performance

**Table 4** Overall responses by accent pattern on both accent perception tasks

| Task | Target | Correct responses | | Incorrect responses | | Total |
|---|---|---|---|---|---|---|
| | | (*n*) | % | (*n*) | % | |
| PitchID | Pattern 1 | 220 | 91.67 | 20 | 8.33 | 240 |
| | Pattern 2 | 235 | 97.92 | 5 | 2.08 | 240 |
| | Pattern 3 | 198 | 82.50 | 42 | 17.50 | 240 |
| | Pattern 4 | 239 | 99.58 | 1 | 0.42 | 240 |
| PitchCAT | Pattern 1 | 87 | 72.50 | 33 | 27.50 | 120 |
| | Pattern 2 | 49 | 54.44 | 41 | 45.55 | 90 |
| | Pattern 3 | 59 | 61.66 | 61 | 38.33 | 60 |
| | Pattern 4 | 67 | 55.83 | 53 | 44.17 | 120 |

*Note*. PitchID consisted of 120 correctly accented and 120 incorrectly accented items. PitchCAT featured only the correctly accented items, yielding 120 total items. One Pattern 2 item and two Pattern 3 items were removed from the analysis.

ceiling effect for this pattern (99.5% mean accuracy). Thus, the group analysis suggested that serial recognition of unaccented nonwords on the WM task was predictive of categorization performance for Pattern 4 words only, which were spoken with the same Low-High unaccented pitch pattern.

To summarize the results overall, L1 Japanese speakers were highly accurate at making correctness judgments on the accent patterns of spoken words, while greater individual variation was observed in the categorization task. Of particular interest was that listeners with greater sensitivity to nonlinguistic pitch contrasts on the adaptive pitch test were superior at both judging and categorizing lexical accents. Although verbal WM was not predictive of overall accuracy, when stimuli were analyzed by pattern, listeners with a higher verbal WM span for unaccented nonwords were better at categorizing real words with the same accent pattern (i.e., Pattern 4).

## Discussion

The present study examined three predictions regarding Japanese speakers' ability to perceive lexical pitch accent. We first considered the effect of task type on perception accuracy scores for lexically accented words. Then we measured acoustic pitch sensitivity and verbal WM capacity as two possible predictors of perception task accuracy.

First, native Japanese speakers were accurate ($M = 93\%$) at determining whether high-frequency noun-plus-postposition targets were spoken with the correct or incorrect pitch accent (PitchID). This finding corroborates with previous perception research which indicates that pitch and tone are accessed along with word form by native listeners (Otake & Cutler, 1999; Shibata & Hurtig, 2008). Contrast this with the categorization task (PitchCAT), on which listeners were only able to accurately categorize 61% of the targets. Despite participants reporting familiarity with task instructions, accuracy rates were similar to those reported in previous studies using accent-pattern comparison or categorization tasks (e.g., Hirano-Cook, 2011; Shport, 2008). Although the task design makes direct comparison of the results problematic (1-of-2 correctness judgment vs. 1-of-4 categorization), we can interpret the performance discrepancy and resulting lack of correlation between the two tasks as suggestive of the differences in processing demands placed on L1 listeners by the two tasks, despite having well-developed native lexical stores. In other words, accent correctness judgments were intuitively possible for

Japanese speakers, but categorization, in which participants' held a spoken stimulus in WM while selecting a visual representation of the accent pattern, likely placed a high processing load even on native listeners. By comparison, Ueno et al. (2014) recently found that L1 Japanese speakers could repeat aloud single, lexically accented words at near-perfect accuracy on an immediate imitation task, echoing our participants' high scores on the speeded PitchID task. Thus, the present findings suggest that the low categorization accuracy and wide variation (ranging from 23 to 100%) were potentially a byproduct of this two-step decision process itself, rather than a reflection of participants' reduced perceptual ability per se. Accordingly, conclusions that pitch accent is not a salient feature in spoken Japanese based solely on categorization tasks are unwarranted.

Second, the multiple regression analysis showed that acoustic pitch sensitivity significantly predicted accuracy for the PitchID and PitchCAT tasks. In other words, listeners who were highly sensitive to subtle, nonlinguistic pitch variations performed better on both lexical accent perception measures, regardless of the differences in task demands pointed out above. The fact that a nonlinguistic capacity predicted lexical perception ability is in line with speech perception models which suggest that speech is processed in a bottom-up fashion, whereby spoken words are processed by listeners starting with an acoustic analysis of the input, prior to the availability of higher level lexical or contextual information (e.g., Norris, McQueen, & Cutler, 2000; Warrier & Zatorre, 2004). This finding also provides evidence for shared perceptual resources for pitch and tone in both speech and nonspeech sounds. Although all participants reported normal hearing and no extensive music training, both of which may relate to pitch perception (e.g., Deutsch et al., 2009), a significant degree of variation was found that could be explained by differences in general acoustic sensitivity. This finding runs contrary to models which assume that adult perception of native language phonological categories is the product of early-age L1 experience and unrelated to basic auditory capacities (e.g., Iverson & Kuhl, 1996; Strange & Shafer, 2008). Although we do not dispute that language experience in infancy is the primary source of adult perceptual acuity in the L1, in the case of lexical accent, the phonetic similarity between the tones in the acoustic sensitivity measure and accent perception tasks in the present study may have contributed to the predictive power of this nonlinguistic capacity. Further, because pitch accent is considered to have a relatively low psychoacoustic salience, in that its phonetic cue of pitch height is not as prominent as, for example, the length contrasts found in Japanese vowels and consonants, a nonlinguistic capacity may enhance perception of this relatively subtle feature (Bohn, 1995).

Lastly, the role of verbal WM in accent perception was negligible in terms of overall task performance. That our serial recognition task required participants to focus on the linear order of the nonword stimuli, rather than explicitly on their sound properties, may account for this finding. Yet, we did find that WM capacity significantly predicted performance on Pattern 4 (LHH-H pitch) words, which were identical in accent pattern with that used in the nonword recognition task. Interestingly, this echoes the connection between pitch accent contributing to recall ability that has been noted in previous research (e.g., Yuzawa & Saito, 2006), albeit in the direction that memory for nonwords is predictive of word perception only for targets of the same accent pattern, suggesting a possible pattern-specific effect for pitch memory. Future studies should employ verbal WM measures featuring a variety of accent patterns to confirm if indeed pitch memory is selectively predictive of lexical accent pattern. In addition, further exploration of the role of verbal WM in tone languages like Mandarin, in the form of tasks which specifically target pitch memory, is needed to shed light on this issue.

In conclusion, the present study showed that L1 Japanese listeners can reliably make correctness decisions on lexically accented spoken stimuli. However, categorizing words by their accent pattern in a task requiring phonetic comparison with visual pitch contours proved to be more difficult. Importantly, acoustic pitch sensitivity, a general cognitive ability, significantly predicted performance variation on both experimental tasks. In addition, verbal WM capacity predicted perception ability only in an accent pattern-specific manner on the memory-intensive categorization task. This pattern of findings suggests that even among adult L1 Japanese listeners, task demands contribute to performance variation, and domain-general cognitive resources continue to support the perception of speech sounds involving lexical-level accent.

## References

Adams, A. M., Bourke, L., & Willis, C. (1999). Working memory and spoken language comprehension in young children. *International Journal of Psychology*, *34*, 364–373.

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.

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

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. (2008). *Praat: Doing phonetics by computer*. Retrieved from <http://www.fon.hum.uva.nl/praat/>

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279–304). Baltimore, MD: York Press.

Chan, A., Ho, Y., & Cheung, M. (1998). Music training improves verbal memory. *Nature*, *396*, 128.

Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in an American conservatory: Association with tone language fluency. *Journal of the Acoustical Society of America*, *125*, 2398–2403.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*, 887–898.

Hirano-Cook, E. (2011). Japanese pitch accent acquisition by learners of Japanese: Effects of training on Japanese accent instruction, perception and production. (Unpublished doctoral dissertation.) University of Kansas, Lawrence, KS.

Iverson, P., & Kuhl, P. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1130–1140.

Jincho, N., Namiki, H., & Mazuka, R. (2008). Effects of verbal working memory and cumulative linguistic knowledge on reading comprehension. *Japanese Psychological Research*, *50*, 12–23.

Kaushanskaya, M., & Yoo, J. (2013). Phonological short-term and working memory in bilinguals' native and second language. *Applied Psycholinguistics*, *34*, 1005–1037.

Liu, F., Patel, A., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, *133*, 1682–1693.

Mandel, J. (2009). *Adaptive pitch test*. Retrieved from <http://tonometric.com/adaptivepitch/>.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain*, *133*, 2635–2642.

Norris, D., McQueen, J., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–370.

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.

Otake, T., & Cutler, A. (1999). Perception of suprasegmental structure in a nonnative dialect. *Journal of Phonetics*, *27*, 229–253.

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, *47*, 223–234.

Semal, C., Demany, L., Ueda, K., & Halle, P. A. (1996). Speech versus nonspeech in pitch memory. *Journal of the Acoustic Society of America*, *100*, 1132–1140.

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

Shibatani, M. (1990). *The languages of Japan*. Cambridge, UK: Cambridge University Press.
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). Munich: Iudicium Verlag.
Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hanson Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193–218). Amsterdam: John Benjamins.
Ueno, T., Saito, S., Saito, A., Tanida, Y., Patterson, K., & Lambon Ralph, M. (2014). Not lost in translation: Generalization of the primary systems hypothesis to Japanese-specific language processes. *Journal of Cognitive Neuroscience*, *26*, 433–446.
Vance, T. (2008). *The sounds of Japanese*. Cambridge, UK: Cambridge University Press.
Venditti, J. (2005). The J_ToBI Model of Japanese Intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology and intonation of phrasing* (pp. 172–200). Oxford, UK: Oxford University Press.
Warrier, C., & Zatorre, R. (2004). Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain*, *127*, 1616–1625.
Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, *38*, 654–662.
Werker, J., & Curtin, S. (2005). PRIMIR: A developmental model of speech processing. *Language Learning and Development*, *1*, 197–234.
Wong, P., & Perrachione, T. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585.
Yuzawa, M. (2002). Effects of pitch accent on short-term retention of words in young children. *Japanese Journal of Psychology*, *73*, 258–263. (In Japanese with English abstract.)
Yuzawa, M., & Saito, S. (2006). The role of prosody and long-term phonological knowledge in Japanese children's nonword repetition performance. *Cognitive Development*, *21*, 146–157.

(Received February 17, 2014; accepted September 6, 2014)

# Appendix

## Target words (N = 32) and postpositions with carrier sentences used in the PitchID and PitchCAT tasks

| | |
|---|---|
| 医学を学ぶ *igaku o manabu* "(I) study medicine" | 家内に話す *kanai ni hanasu* "(I) speak to my wife" |
| めがねをかける *megane o kakeru* "(I) put on glasses" | 技術をみがく *gijutsu o migaku* "(I) polish my skills" |
| フォークで食べる *fooku de taberu* "(I) eat with a fork" | 荷物を運ぶ *nimotsu o hakobu* "(I) carry luggage" |
| おもちゃで遊ぶ *omocha de asobu* "(I) play with a toy" | あなたがいる *anata ga iru* "You are there" |
| 八時に起きる *hachiji ni okiru* '(I) wake up at 8 o'clock' | *ハサミがある *hasami ga aru* "There are scissors" |
| 中身を見る *nakami o miru* "(I) look at the contents" | *刺身を食べる *sashimi o taberu* "(I) eat sashimi" |
| ななめにする *naname ni suru* "(I) turn it sideways" | 昼間に行く *hiruma ni iku* "(I) go in the afternoon" |
| 言葉にする *kotoba ni suru* "(I) put it into words" | 娘がいる *musume ga iru* "(I) have a daughter" |
| 夜中に起きる *yonaka ni okiru* "(I) wake up at midnight" | 男がいる *otoko ga iru* "There is a man" |
| いなかに住む *inaka ni sumu* "(I) live in the countryside" | 左に曲がる *hidari ni magaru* "(I) turn left" |
| ハガキを出す *hagaki o dasu* "(I) send a postcard" | 昔に戻る *mukashi ni modoru* "(I) return to the past" |
| 手前に引く *temae ni hiku* "(I) pull it toward me" | 仕事を探す *shigoto o sagasu* "(I) search for a job" |
| 来月に行く *raigetsu ni iku* "(I) go next month" | タクシーに乗る *takushii ni noru* "(I) get in a taxi" |
| あさってまで待つ *asatte made matsu* "(I) wait until the day after tomorrow" | 飛行機に乗る *hikooki ni noru* "(I) ride in an airplane" |
| *九日に行く *kokonoka ni iku* "(I) go on the 9th" | 妹がいる *imooto ga iru* "(I) have a younger sister" |
| 夕方になる *yuugata ni naru* "It becomes evening" | 地下鉄に乗る *chikatetsu ni noru* "(I) ride the subway" |

*Note.* Items preceded by an asterisk were removed from the PitchCAT analysis.