Pitch Accent Constrains Lexical Activation in Japanese Spoken Word Recognition:
A semantic priming study

Terumichi ARIGA

Abstract

The present study investigated whether Japanese listeners utilize pitch accent for lexical activation using two cross-modal semantic priming experiments. Segmentally identical but suprasegmentally different words were presented as auditory primes, followed by visual targets that were semantically congruent or incongruent with the primes. Participants made lexical decision on visual targets while listening to auditory primes. In Experiment 1, targets were presented at the offset of the primes. In Experiment 2, targets were presented in the middle of the primes. Based on the reaction time, a semantic priming effect was observed in the semantically congruent condition (e.g., *jidoo* HLL "children"–*gakkoo* "school" and *jidoo* LHH "automatic"–*kikai* "machine") but not in the incongruent condition (e.g., *jidoo* HLL "children"–*kikai* "machine" and *jidoo* LHH "automatic"–*gakkoo* "school") in both experiments. The results demonstrated that the auditory primes activate only words whose pitch accent was consistent with the auditory input in the listener's mental lexicon, providing further evidence that suprasegmental information constrains lexical activation in Japanese spoken word recognition.

Key Words: psycholinguistics, spoken word recognition, lexical activation, Japanese pitch accent, semantic priming

1. Introduction

Spoken word recognition is the process by which humans map speech signals onto their mental representation of words. When listeners recognize spoken words, they activate words that correspond to the auditory information in the mental lexicon. With the development of, and access to, models of spoken word recognition (e.g., Marslen-Wilson 1987; McClelland & Elman 1986; see Cutler 1995 for review), it is still unclear whether suprasegmental information such as lexical accent plays a role in the process of lexical activation.

In Japanese, word accent is lexically specified. For instance, *ringo* "apple" contains an unaccented LHH pitch pattern and is unacceptable when pronounced using other accented pitch patterns, such as HLL and LHL. Additionally, the Japanese accent discriminates certain words that segmental information cannot distinguish. For instance, *jidoo* "children" is pronounced with accented HLL, whereas *jidoo* "automatic" is pronounced with unaccented LHH, although both are segmentally identical. Thus, pitch accent in Japanese forms a part of lexical information stored in listeners' lexical memory and is potentially used for identifying words in addition to segmental information.

Previous studies have demonstrated that pitch accent information plays a role in Japanese spoken word recognition. If words are presented with an incorrect accent (e.g., *raion* *HLLL; *raion* LHHH for "lion" is correct), listeners take longer time to repeat words orally with the correct accent than if words are presented with the correct accent (Minematsu & Hirose 1995). In the gating paradigm, in which initial fragments of words from onset are presented incrementally, listeners can infer words based on fragmented pitch accent information (Cutler & Otake 1999). Many event-related potential studies suggest that the brain detects pitch accent information and compares it with lexical memory (Hirose et al. in press; Koso & Hagiwara 2009; Koso et al. 2011). These studies imply that pitch accent may help identify words that are being spoken.

However, listeners can recognize words even if spoken with the incorrect accent. The dialectical variety of Japanese pitch accent is an example. For instance, *haru* "spring" is pronounced with the HL pattern in Tokyo Japanese whereas the LH pattern is followed in Kinki Japanese; LH for *haru* "spring" is an inappropriate pitch pattern and becomes altogether another expression in Tokyo Japanese (*haru* LH would mean "to stick"). In such cases, listeners experience minimal difficulty in understanding words spoken with the nonnative dialectical accent (Shibata 1961). Pitch accent information at the lexical level may also vary by compound accent rule (Kubozono 2008). Although *jidoo* "children" is pronounced with the accented HLL pitch pattern in isolation, it is deaccented in its compound forms, such as *jidoo-fukushi* LHH-HLL "child welfare." In cases like these, a pitch accent may disturb the understanding of spoken words; hence, listeners of Japanese should activate words without pitch accent information.

Although previous studies offer ample evidence that native Japanese listeners can perceive pitch accent, the question remains on when and how pitch accent information is utilized in Japanese spoken word recognition. Spoken word recognition consists of two sequential processes, namely, lexical activation and lexical selection. Two possibilities exist for the role of suprasegmental information (Cutler et al. 1997). The first is that suprasegmental information constrains lexical activation. This possibility may result in the activation of words consistent with segmental, as well as suprasegmental, information. The second is that suprasegmental information does not constrain lexical activation but plays a role during lexical selection. This notion hypothesizes that all words consistent with segmental information are activated once regardless of suprasegmental information, and that subsequently, suprasegmental information screens out inappropriate words.

Cutler (1986) found that suprasegmental lexical stress does not constrain lexical activation in English spoken word recognition. In Cutler's experiment, participants listened to segmentally identical but suprasegmentally different words, such as *FORbear* and *forBEAR* (capitals indicate stressed syllable), and then words semantically related to them were presented visually on the display. Participants were asked to make a lexical decision on the visually presented targets as rapidly as possible. Based on the analysis of reaction time data, Cutler reported that the auditory prime *FORbear* exerted a semantic priming effect not only on the semantically related target *ancestor* but also on the semantically unrelated target *tolerate*, which was originally semantically related to *forBEAR*. This result suggests that auditory *FORbear* activates *FORbear* and *forBEAR* in the mental lexicon of listeners. Cutler concluded that words in the mental lexicon may be activated regardless of the suprasegmental information in English.

In Japanese, however, Sekiguchi and Nakajima (1999) reported that pitch accent constrains Japanese spoken word recognition. In their experiment, which was similar to Cutler's (1986), either one of Japanese minimal accent pairs, such as *jidoo* HLL and *jidoo* LHH, were presented auditorily, and then the identical or the other words appeared on the display in kanji characters. Participants made lexical decisions on visual target words by pressing the button and their reaction time was measured. They discovered that the lexical decision of visually presented *jidoo* HLL (児童 in kanji characters) was facilitated if followed by auditory prime *jidoo* HLL but not if followed by auditory *jidoo* LHH. Based on these results, Sekiguchi and Nakajima concluded that *jidoo* HLL is activated by auditory *jidoo* HLL but not by *jidoo* LHH. Thereby, pitch accent is relevant to lexical activation.

The two previous studies reported contrasting results on the lexical activation process of segmentally identical but suprasegmentally different words in English and Japanese. It may suggest that English and Japanese have different strategies for utilizing suprasegmental information in lexical activation.

However, another possibility exists that their opposite opinions resulted from the use of different priming paradigms. In the Japanese study, Sekiguchi and Nakajima (1999) used the repetition priming paradigm, where the same lexical items are used as primes and targets. Repetition priming denotes a facilitatory effect due to the repetition of the same lexical items within a short time and does not necessarily require lexical activation from the prime stimuli (Forster & Davis 1984). In Sekiguchi and Nakajima's experiment, they claimed that repetition priming effects were not observed on accentually opposed targets (e.g., *jidoo* LHH–*jidoo* HLL) because the prime *jidoo* LHH did not activated *jidoo* HLL. However, in their paradigm, pitch pattern commonality between primes and targets would simply affect the repetition priming effects. Even if words are activated regardless of the pitch pattern of the auditory primes, that is, auditory prime *jidoo* HLL activated both *jidoo* HLL and *jidoo* LHH in the mental lexicon, repetition priming effect can occur only on the target *jidoo* HLL because the pitch pattern of the auditory prime is only congruent with the target *jidoo* HLL. Therefore, a repetition priming effect obtained in the lexical decision of the visual target *jidoo* HLL following the auditory prime *jidoo* HLL does not necessarily mean listeners activated *jidoo* HLL but not *jidoo* LHH in their mental lexicon.

In contrast, the semantic priming paradigm used by Cutler (1986) occurs through spreading activation from the prime stimuli. Words with similar or associated meanings are supposedly stored in proximity in the mental lexicon, which form a network. When a word is processed, a listener activates not only the word itself but also words semantically related to it (Meyer & Schvaneveldt 1971). In Cutler (1986), the lexical decision of *ancestor* preceded by a prime *FORbear* was facilitated as listeners activated *ancestor* in their mental lexicon, which was triggered by the spreading activation of *FORbear*. However, the lexical decision of *tolerate* preceded by a prime *FORbear* was also facilitated because listeners unconsciously activated *tolerate*, which was triggered by the spreading activation of *forBEAR*. Hence, listeners activated *FORbear* but also *forBEAR* through the prime word *FORbear*. This semantic priming effect is strong evidence of the lexical activation process that is not constrained by suprasegmental information in English.

The present study speculates the possibility of the influence of the difference in priming paradigm. In Cutler (1986) and Sekiguchi and Nakajima (1999), the difference of the priming paradigm used in the experiments confounded the observing language and opposing results. Although suprasegmental information does not seem to constrain lexical activation in English, recognition of words can be influenced by the English stress itself if the task is lexical categorization (Cutler & Clifton 1984) or repetition priming (Cooper et al. 2002). In addition, studies in other languages that maintain that suprasegmental information is utilized during lexical activation (van Donselaar et al. 2005; Soto-Faraco et al. 2001) were conducted using repetition priming paradigm as well. Thus, a possibility exists that semantic priming effects were observed in semantically congruent and incongruent targets, whereas repetition priming effects were observed only in accentually congruent targets. The focus of the investigation is whether accent information plays a role in lexical activation. In other words, the study aims to determine whether listeners activate words with or without consideration of the accent of the auditory input. To apply this context in Japanese, a study on semantic priming that utilizes spreading activation in mental lexicon is necessary.

Therefore, the present study investigates whether pitch accent information constrains lexical activation in Japanese spoken word recognition using a cross-modal semantic priming paradigm. Minimal accent pairs are used as auditory primes and their semantically related words are used as visual targets. If pitch accent does not play a role in lexical activation in Japanese spoken word recognition, homophonic words that are contrasted sorely by pitch accent should be equally activated regardless of the pitch pattern of the auditory prime. That is, the semantic priming effects of the accentually contrasted primes should occur on semantically congruent and incongruent targets. In contrast, if pitch accent plays a role, the accentually contrasted homophones should not be activated concurrently by the auditory prime. That is, the semantic priming effects of the accentually contrasted primes should occur only on semantically congruent targets.

2. Experiment 1

2.1 Materials

Eighteen minimal accent pairs were prepared as auditory prime items, which consisted of three or four moras written in two Japanese kanji characters, and were familiar words according to the word familiarity survey conducted by Amano and Kondo (1999) (mean rating = 5.70, *SD* = 0.36).[1] Pitch pattern was accented (on the first mora: e.g., HLL) or unaccented (LHH). Referring to an accent dictionary (NHK 1998), the study verified that each word is pronounced with only one pitch pattern in Tokyo Japanese. In order to avoid the interference of homophones, the study conducted a pretest on 20 Japanese native speakers, who were instructed to listen to words and to write them down in kanji characters. The items for which more than 80 % of the participants gave the same interpretation were selected for the present study.

Visual target items were then prepared. For each prime minimal accent pair, the target items were selected, such that they were semantically related to one prime item but not to the other. All target words were written in two kanji characters and were familiar words according to Amano and Kondo (1999) (mean rating = 6.06, *SD* = 0.28).

The design was a 3 × 2 within-subject condition, as shown in Table 1. The condition of the prime–target relationship (Congruency) was either congruent (e.g., *jidoo* HLL "children" is semantically related to *gakkoo* 学校 "school," whereas *jidoo* LHH "automatic" is related to *kikai* 機械 "machine"), incongruent (*jidoo* LHH "automatic" is not semantically related to *gakkoo* 学校 "school," whereas *jidoo* HLL "children" is not related to *kikai* 機械 "machine"), or baseline.

Table 1 Example of materials used in the experiment

Prime–target relationship (Congruency)	Accent pattern of target-related words (Accent)			
	accented		unaccented	
	Prime	Target	Prime	Target
congruent	*jidoo* HLL	学校 "school"	*jidoo* LHH	機械 "machine"
incongruent	*jidoo* LHH	学校	*jidoo* HLL	機械
baseline	*tegami*	学校	*tegami*	機械

The baseline condition is the control, where the prime words are not semantically related to either target words (*tegami* "letter" is not semantically related to *gakkoo* or *kikai*). This condition is necessary for measuring the semantic priming effect of prime items on the lexical decision of the target items. The condition of the accent pattern of target-related words (Accent) was either accented or unaccented, which refers to the pitch pattern of the prime word in the congruent condition.[2]

To confirm the validity of the semantic relationship between the prime and target of each item, a rating pretest was conducted on 20 Japanese native speakers. Participants rated the semantic relatedness for each prime and target word from 1 (lowest) to 6 (highest) points. The results indicated that the mean ratings were 5.35, 1.94, and 1.84 for the congruent, incongruent, and baseline conditions, respectively. One-way ANOVA demonstrated the significant effect of the prime–target relationship ($F(2, 105) = 479.180$, $p < 0.001$). Tukey's multiple comparisons illustrated that the ratings for the congruent condition were significantly higher than those for the incongruent or baseline condition (ps < 0.001). No significant difference was observed between the ratings for the incongruent and baseline conditions ($p = 0.735$).

To obtain sufficient data points, two conditions for each test item were presented to each participant. The two conditions for each test item did not contain the same prime word or target word. The overall test trials were 36 trials. Conditions presented to participants were counterbalanced. In addition to the test items, 108 filler items were added. Thirty-six filler items contained word targets, whereas the other 72 contained nonword targets. By summing up the test items and the filler items, the nonword proportion of the lexical decision of the targets was 50 %. The total 144 trials were presented in random order.

The audio files used for the auditory primes were obtained from the CD-ROM data of Amano and Kondo (1999). The recordings were that of the natural pronunciation of words in isolation spoken by a female native speaker of Tokyo Japanese. The experimenter verified that all auditory materials were pronounced with proper pitch accent pattern. They were manipulated in Praat 6.1.16 (Boersma & Weenink 2020) to remove silence and change the duration to 600 ms (for words with three moras) or 800 ms (for words with four moras).

2.2 Participants

Sixty-three undergraduate or graduate students (mean age = 20.08, *SD* = 1.30), who were native speakers of Tokyo Japanese, participated in the experiment. None participated in any preliminary surveys. They received a monetary compensation for their participation.

2.3 Procedure

The experiment was conducted online through PCIbex (Zehr & Schwartz 2018). The participants individually accessed special websites on a personal computer in a silent room in their homes. They listened to auditory materials through headphone or earphone.

The experiment consisted of three tasks: a lexical decision task, a dummy sound memory task, and a pronunciation task. Figure 1 illustrates the procedure for the lexical decision task. At the beginning of each trial, a fixation point (+) was displayed at the center of the display. After 1000 ms, the auditory prime was presented. The duration of the prime was either 600 ms (words with three moras) or 800 ms (words with four moras). At the offset of the auditory prime, the fixation point disappeared, and subsequently, the visual target was presented in kanji characters at the center of the display (stimulus onset asynchrony (SOA) was 600 ms or 800 ms). The participants made a lexical decision about the target and pressed the F key for a word response or the J key for a nonword response on the keyboard as rapidly as possible. The visual target remained on the display until participants pressed the F or J key or until 3000 ms passed after the display. Reaction time from the onset of targets to the participant's lexical decision and response accuracy were recorded. The interval between trials was 1000 ms.

To make the participants pay attention to the auditory prime, the dummy sound memory task was inserted after every 36 lexical decision trials. In the task, a word list was presented on the display and the participants selected words they had listened to as auditory primes in the previous lexical decision session. After the lexical decision task, the pronunciation task was conducted, where participants pronounced all the words used as auditory primes. This was to confirm the participants' interpretation of the pitch accent pattern of the prime. Participants took optional breaks between tasks. The experiment lasted for approximately 1 h.

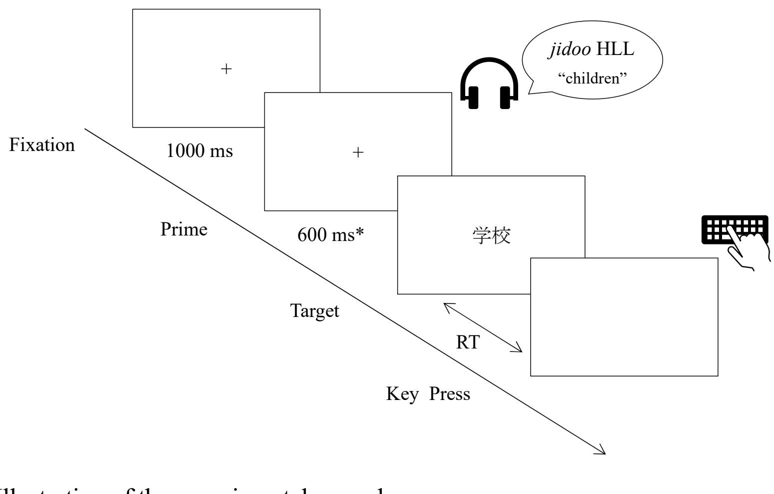

Figure 1 Illustration of the experimental procedure

* The duration of the prime that consists of four moras was 800 ms.

2.4 Results

Statistical analysis was conducted in R 3.6.1 (R Core Team 2019). Data from three participants were excluded as they made incorrect responses in more than 30 % of all trials. The data of the remaining 60 participants were used for analysis.

The overall correct response rate was near ceiling (98.33 %) and the linear mixed effects (LME) model analysis did not indicate any significant difference between conditions (*p*s > 0.1), which suggests that the task itself was not difficult to complete accurately.

The reaction time data in the trials where participants made incorrect responses, the trials using the primes which participants mispronounced in the pronunciation task, and data with reaction time of less than 300 ms or more than 1400 ms were excluded as outliers. Data for the log-transformed reaction time were submitted to the LME models. Accent (accented/unaccented), Congruency (congruent/incongruent/baseline), and their interaction were set as fixed effects, whereas individual differences among participants and items were set as random effects. After an optimal model was selected, data points of the reaction time that exceeded 2.0 *SD* more than that of the residuals from the reaction times estimated by the model were excluded (cf., Arai & Roland 2016). In total, 12.04 % of all data points were excluded.

Figure 2 illustrates the by-subject mean reaction times for each condition, and Table 2 provides a summary of the LME analysis. The mean reaction time for the congruent condition (*jidoo* HLL–*gakkoo* and *jidoo* LHH–*kikai*) was significantly faster than that for the baseline condition (p = 0.005), which suggests that there was a semantic priming effect in the congruent condition. However, the mean reaction time for the incongruent condition (*jidoo* LHH–*gakkoo* and *jidoo* HLL–*kikai*) was not significantly different from that for the baseline condition (p = 0.757), which suggests that no semantic priming effect was observed in the incongruent condition. If pitch accent

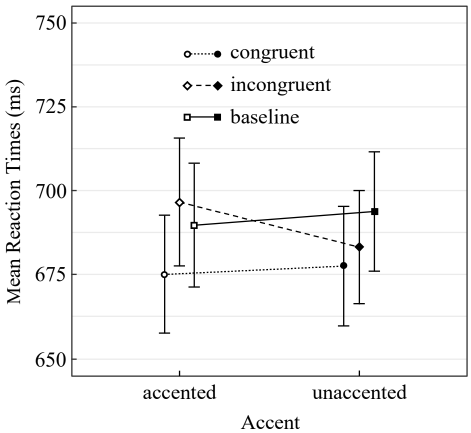

Figure 2 By-subject mean reaction times for each condition (Experiment 1). Error bars indicate the standard error (*SE*) of the mean.

Table 2 Summary of the LME analysis (Experiment 1)

	β	*SE*	*df*	*t*	*p*	
(Intercept)	6.510	0.026	72.824	253.423	< 0.001	***
Accent	0.006	0.019	27.817	0.312	0.757	
congruent	–0.022	0.008	1803.262	–2.810	0.005	**
incongruent	–0.002	0.008	1803.863	–0.309	0.757	
Accent × congruent	–0.005	0.016	1803.353	–0.298	0.765	
Accent × incongruent	–0.025	0.016	1803.788	–1.575	0.115	

Significant codes: *** $p < 0.001$, ** $p < 0.01$

was not utilized in lexical activation, then accented and unaccented primes would not be distinguishable, and the semantic priming effect would occur even in the incongruent condition. However, the semantic priming effect was observed in the congruent condition and not in the incongruent condition. This suggested that only the word the accent pattern of which is consistent with a prime word was activated in the mental lexicon of the listeners. In addition, the main effect of Accent, the interactions between Accent and the congruent condition as well as Accent and the incongruent condition were not significant (*p*s > 0.1). The semantic priming effect was observed only in the congruent condition across both Accent conditions. Thus, such a strategy of utilizing pitch accent for lexical activation was regular regardless of the pitch pattern of the auditory input.

2.5 Discussion

The results suggested that the auditory prime word activated only the word in the mental lexicon that was consistent with its pitch accent pattern. As such, *jidoo* HLL activated only *jidoo* HLL, whereas *jidoo* LHH activated only *jidoo* LHH. This supports the hypothesis that pitch accent constrains lexical activation in the process of spoken word recognition in Japanese.

However, another interpretation of this experimental result is possible, as mentioned by Sekiguchi and Nakajima (1999). The timing of the presentation of the target words may have been inappropriate for the objective of observing lexical activation in the process of spoken word recognition. In this experiment, the visual targets were presented at the offset of the auditory primes. When the lexical decision was made, the participants finished listening to the whole input of prime words. Therefore, listeners may initially activate words regardless of perceived pitch accent, then screened out inappropriate candidates before the target lexical decision.

Lexical activation in spoken word recognition is incremental. If the lexical activation and utilization of pitch accent are simultaneous, then candidate words that are inconsistent with the perceived pitch accent may be excluded immediately after the detection of the deviation. When a listener processes the auditory *jidoo* HLL, accentually opposing *jidoo* LHH is screened out from candidate words at least at the disambiguation point of the pitch pattern, specifically, at the offset of the second mora (*jido* HL). In contrast, if the pitch accent is not utilized, then *jidoo* HLL and *jidoo* LHH may be equally activated.

Experiment 2 investigated this possibility by changing the position of the presentation of the visual targets, which were presented at the disambiguating point of the pitch pattern of the prime words. If pitch accent information constrains lexical activation, then Experiment 2 should obtain the same results. The present experiment differed from that of Sekiguchi and Nakajima (1999) in that the gated stimuli (Grosjean 1980) were not used as auditory primes. As listeners can compute even after presenting fragments of word initial segments, the priming effect by such gated stimuli may not fully reflect the listeners' temporal processing while listening to auditory primes. To address this concern, the present study presented the auditory prime words from onset to offset, which is similar to Experiment 1, and imposed the target lexical decision while listeners were listening to the auditory primes.

3. Experiment 2

3.1 Materials

The items and design were the same as those used in Experiment 1 (Table 1). The difference between Experiments 1 and 2 is the timing of the presentation of the visual targets. In Experiment 2, the target was presented at the offset of the second mora of the prime (e.g., *jido* HL/*jido* LH from *jidoo* HLL/*jidoo* LHH), thus stimulus onset asynchrony (SOA) was the duration of the primes' first to second mora.[3] SOA was set for each prime item, including the baseline condition and filler items. The mean SOA was 381 ms (*SD* = 36). The recordings used in Experiment 2 were the same as those used in Experiment 1.

3.2 Participants

Twenty-eight undergraduate or graduate students (mean age = 22.64, *SD* = 6.12), who were native speakers of Tokyo Japanese, participated in the experiment. None participated in any preliminary surveys or Experiment 1. They received a monetary compensation for their participation.

3.3 Procedure

The experiment was conducted online through PCIbex (Zehr & Schwartz 2018). The procedure was the same as that of Experiment 1. The participants made a lexical decision as rapidly as possible on a visual target that was presented at the offset of the auditory prime's second mora.

Reaction time from the onset of targets to the participant's lexical decision and response accuracy were recorded. The prime was played in full even when the participants made a lexical decision before it ended.

3.4 Results

All participants made correct responses in more than 80 % of all trials. The overall response rate was near ceiling (98.51 %) and the LME analysis did not find any significant difference between conditions (*p*s > 0.1).

The reaction time data in the trials where participants made incorrect responses, the primes of which participants misinterpreted in the pronunciation task, and data with reaction time of less than 300 ms or more than 1400 ms were excluded as outliers. Data for the log-transformed reaction time were submitted to the LME models. Accent (accented/unaccented), Congruency (congruent/incongruent/baseline), and their interaction were set as fixed effects, whereas individual differences among participants and items were set as random effects. After an optimal model was selected, data points of the reaction time that exceeded 2.0 *SD* more than that of the residuals from the reaction times estimated by the model were excluded. In total, 13.19 % of all data points were excluded.

Figure 3 illustrates the by-subject mean reaction times for each condition, and Table 3 provides a summary of the LME analysis. The mean reaction time for the congruent condition (*jidoo* HLL–*gakkoo* and *jidoo* LHH–*kikai*) was significantly faster than that for the baseline condition (*p* = 0.031), which suggests that there was a semantic priming effect in the congruent condition. However, the mean reaction time for the incongruent condition (*jidoo* LHH–*gakkoo* and *jidoo*

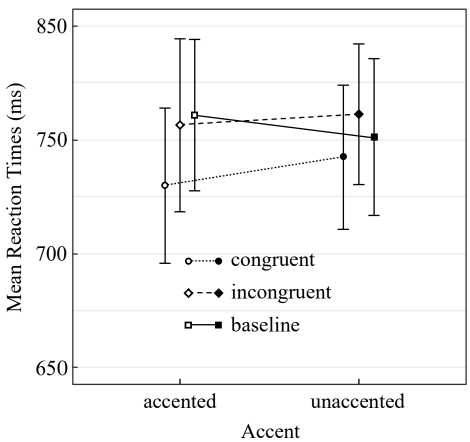

Figure 3 By-subject mean reaction times for each condition (Experiment 2). Error bars indicate the *SE* of the mean.

Table 3 Summary of the LME analysis (Experiment 2)

	β	*SE*	*df*	*t*	*p*	
(Intercept)	6.594	0.041	30.910	161.587	< 0.001	***
Accent	–0.009	0.018	827.087	–0.497	0.620	
congruent	–0.029	0.013	826.16	–2.167	0.031	*
incongruent	0.005	0.013	826.726	0.342	0.732	
Accent × congruent	0.034	0.026	827.042	1.273	0.203	
Accent × incongruent	0.020	0.027	827.123	0.755	0.450	

Significant codes: *** $p < 0.001$, * $p < 0.05$

HLL–*kikai*) was not significantly different from that for the baseline condition ($p = 0.732$), which suggests that no semantic priming effect was observed in the incongruent condition. In addition, the main effect of Accent, the interactions between Accent and the congruent condition as well as Accent and the incongruent condition were not significant (*p*s > 0.1). The semantic priming effect was observed only in the congruent condition across both Accent conditions. These results are similar to those of Experiment 1, suggesting that prime words were accentually distinguished and that only the word the accent pattern of which is consistent with a prime word was activated in the mental lexicon of the listeners.

3.5 Discussion

The results of Experiment 2, where visual targets were presented at the offset of the second mora of the prime words, were the same as those of Experiment 1. This finding indicated that even in the middle of processing auditory prime words (i.e., incomplete auditory input), accentually opposing words were screened out from the candidates. In other words, the listeners activated *jidoo* HLL but not *jidoo* LHH from the fragmented auditory *jido* HL input and activated *jidoo* LHH but not *jidoo* HLL from the fragmented auditory *jido* LH.

The mean reaction times for all conditions in Experiment 2 were approximately 50 ms longer than those in Experiment 1. This slight increase of the reaction time indicated that the task in Experiment 2 was more difficult than that of Experiment 1. This may be because the target overlapped with the prime. In Experiment 2, targets overlapped with primes for at least 200 ms (the duration of primes was either 600 ms or 800 ms while the SOA was 381 ms on average). However, the reaction time increase was far shorter than that overlapping duration. Thus, participants in Experiment 2 did not merely wait for the full presentation of the auditory prime input. Instead, they made a lexical decision on the targets based on the input available before they finished listening to the prime. Therefore, the results of Experiment 2 supported the hypothesis that listeners utilize pitch accent information during lexical activation.

4. General Discussion

The present study aimed to investigate whether pitch accent information plays a role in lexical activation in Japanese spoken word recognition by two cross-modal semantic priming experiments. Segmentally identical but suprasegmentally different words were presented as auditory primes, followed by visual targets that were semantically congruent or incongruent with the primes. Participants made lexical decisions on visual targets and their reaction time was analyzed. The experiments produced the following results. First, accented primes facilitated processing targets semantically related to accented words (e.g., *jidoo* HLL–*gakkoo*), whereas unaccented primes facilitated processing targets semantically related to unaccented words (e.g., *jidoo* LHH–*kikai*). However, the primes did not facilitate processing incongruent targets (e.g., *jidoo* HLL–*kikai* and *jidoo* LHH–*gakkoo*). Second, such semantic priming effects were observed when the targets were presented at the offset of the primes and in the middle of processing the primes.

The results are consistent with those of Sekiguchi and Nakajima (1999) and suggest that pitch accent information constrains lexical activation in Japanese spoken word recognition. The minimal accent pairs used in the experiments were semantically unrelated to each other. Hence, they are stored separately in the semantic network of the mental lexicon. Accented prime words activated accented words but did not activate unaccented words. The same is true for unaccented prime words. Thus, a facilitatory priming effect observed in the congruent condition but not the incongruent condition indicates that only words whose pitch accent is consistent with that of the auditory input are activated. Moreover, the trend of the priming effect did not differ regardless of whether target presentation was at the offset or in the middle of priming effect. This indicates that accentually opposed words are excluded at the stage of lexical activation, rather than lexical selection.

The findings of the present study and Cutler's (1986) study lead to a suggestion that the constraint placed by suprasegmental information on lexical activation is language-specific. Suprasegmental information itself is perceptible in Japanese (Cutler & Otake 1999; Minematsu & Hirose 1995) and in English (Cutler & Clifton 1984). Cutler (1986) noted although English stress does not constrain lexical activation, it is utilized at the lexical selection process that follows lexical activation. In contrast, the present study suggests that in Japanese, lexical pitch accent is utilized earlier at the level of lexical activation.

The present study proposes that the role of accent information in lexical activation may be dependent on the role of accent in distinguishing homophonic words in that language. The distinctive function of accent, which is the notion of the importance of accent in distinguishing words, varies across languages. For instance, Shibata and Shibata (1990) reported that the number of lexical items that are segmentally identical but suprasegmentally different was low in English but high in Japanese. In English, as the change of the stress typically involves an alternation of vowel quality, homophones distinguished only by stress accent are rare. In contrast, Japanese, which has a limited set of phonemes, contains many homophones that are otherwise indistinguishable without pitch accent. Moreover, in most of the cases in English, segmentally identical but suprasegmentally different words belong to different parts of speech (i.e., *FORbear* is a noun, whereas *forBEAR* a verb). In contrast, segmentally identical but suprasegmentally different words in Japanese frequently belong to the same part of speech. Such examples, among which are *denki* HLL "electric light" and *denki* LHH "biography" (both are nouns), may frequently be confusing without accent information or context (e.g., *kore wa <u>denki</u> desu*. "This is <u>an electric light/a biography</u>."). Thus, suprasegmental information is more important for identifying words in Japanese than it is in English.

Therefore, listeners should utilize pitch accent information as well as segmental information in Japanese spoken word recognition, whereas listeners need not utilize stress accent information in English spoken word recognition. The process of spoken word recognition is rapid and highly reliable. To achieve reliable processing, available information should be exploited as a cue for searching the appropriate words that correspond to the auditory information in the mental lexicon. In Japanese, where suprasegmental information plays a crucial role in distinguishing spoken words, listeners use this information for lexical activation in spoken word recognition.

Acknowledgements

This paper is based on the author's graduation thesis and the oral presentation at *The 162nd Meeting of the Linguistic Society of Japan* (June, 2021). The author is grateful to Professor Yuki Hirose for supervising this work. The author also thanks two anonymous reviewers for their valuable comments on the earlier version of the manuscript.

Appendix: List of materials (Experiment 1, 2)

Prime			Target	
accented	unaccented	baseline	accented-related	unaccented-related
engi HLL	*engi* LHH	*nendo*	*geinoo* 芸能	*jinja* 神社
kageki HLL	*kageki* LHH	*donabe*	*butai* 舞台	*kyokutan* 極端
koogo HLL	*koogo* LHH	*suiso*	*jumban* 順番	*nichijoo* 日常
sadoo HLL	*sadoo* LHH	*suika*	*wafuu* 和風	*soochi* 装置
jidoo HLL	*jidoo* LHH	*tegami*	*gakkoo* 学校	*kikai* 機械
tsuuka HLL	*tsuuka* LHH	*shinwa*	*kinsen* 金銭	*fumikiri* 踏切

higan HLL	*higan* LHH	*genri*	*tassei* 達成	*kisetsu* 季節
yooi HLL	*yooi* LHH	*hindo*	*jumbi* 準備	*kantan* 簡単
yooki HLL	*yooki* LHH	*rokuga*	*shokuhin* 食品	*kaisei* 快晴
anzan HLLL	*anzan* LHHH	*temboo*	*nimpu* 妊婦	*sansuu* 算数
enchoo HLLL	*enchoo* LHHH	*kaisui*	*sensei* 先生	*yotei* 予定
keihoo HLLL	*keihoo* LHHH	*ennichi*	*saiban* 裁判	*kinkyuu* 緊急
goosei HLLL	*goosei* LHHH	*taibatsu*	*zeitaku* 贅沢	*kagaku* 科学
shoogai HLLL	*shoogai* LHHH	*kakunin*	*jumyoo* 寿命	*konnan* 困難
suihei HLLL	*suihei* LHHH	*mooten*	*sensoo* 戦争	*chokusen* 直線
sonchoo HLLL	*sonchoo* LHHH	*densetsu*	*senkyo* 選挙	*taisetsu* 大切
toorui HLLL	*toorui* LHHH	*unsei*	*kashi* 菓子	*yakyuu* 野球
doojoo HLLL	*doojoo* LHHH	*zeikin*	*bujutsu* 武術	*kyookan* 共感

Notes

[1] In Amano and Kondo (1999), subjective familiarity was rated from 1 (least familiar) to 7 (most familiar). The words used in the experiments were supposed to have higher rates of familiarity.

[2] The pitch patterns of target items were not controlled, as semantic priming is supposed to occur through semantic similarity and not phonological (i.e., accentual) similarity. However, as a reviewer pointed out, an accentual relationship should be considered to exclude any possibility for phonological interference.

[3] SOA of items which consisted of 1 mora + 1 syllable (e.g., *jidoo*) was the median of the offset of the first and last moras.

References

Amano, S., & Kondo, T. (1999). *NTT deetabeesu shiriizu: Nihongo no goitokusei [NTT database series: Lexical properties of Japanese]*. Tokyo: Sanseidoo.

Arai, M., & Roland, D. (2016). Statistical analysis of eye-movement data and reading time data in language comprehension research. *Proceedings of the Institute of Statistical Mathematics*, *64*(2), 201–231.

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer. https://www.praat.org/

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, *45*(3), 207–228.

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma, & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183–196). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*(3), 201–220.

Cutler, A. (1995). Spoken word recognition and production. In J. L. Miller, & P. D. Eimas (Eds.), *Speech, language and communication* (pp. 97–136). San Diego: Academic Press.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustical Society of America*, *105*(3), 1877–1888.

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, *58A*(2), 251–273.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680–698.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267–283.

Hirose, Y., Kobayashi, Y., Chen, T., Ito, A., & Ito., T. (in press). ERP responses to different types of pitch accent violation in Tokyo Japanese: Rule application or lexical memory? In H. Jeon (Ed.), *Japanese/Korean linguistics vol. 28*. Stanford: CSLI Publications.

Koso, A., & Hagiwara, H. (2009). Event-related potential evidence of processing lexical pitch-accent in auditory Japanese sentences. *Neuroreport*, *20*(14), 1270–1274.

Koso, A., Ojima, S., & Hagiwara, H. (2011). An event-related potential investigation of lexical pitch-accent processing in auditory Japanese. *Brain research*, *1385*, 217–228.

Kubozono, H. (2008). Japanese accent. In S. Miyagawa, & M. Saito (Eds.), *The Oxford handbook of Japanese linguistics* (pp. 165–191). Oxford: Oxford University Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234.

Minematsu, N., & Hirose, K. (1995). Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese. *Journal of the Acoustical Society of Japan (E)*, *16*(5), 311–320.

NHK (1998). *NHK nihongo hatsuon akusento jiten [Japanese pronunciation accent dictionary]*. Tokyo: Nippon Hoosoo Kyookai.

R Core Team (2019). R: A language and environment for statistical computing. https://www.R-project.org/

Sekiguchi, T., & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, *28*(4), 439–454.

Shibata, T. (1961). Nihongo no akusento [Japanese Accent]. *Gengo Seikatsu*, *117*, 14–20.

Shibata, T., & Shibata, R. (1990). Akusento wa dooongo wo donoteido bembetsu shiuru ka: Nihongo, eigo, chuugokugo no baai [How significant is word accent in differentiating homonyms in Japanese, English, and Chinese?]. *Keiryoo Kokugogaku*, *17*(7), 317–327.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://www.pcibex.net/