The Use of Pitch Information in L1 and L2 Japanese Word Recognition

MARIKO HONDA
University of Reading
mariko.honda@reading.ac.uk

ABSTRACT. This study examines the role of pitch accent in the perception of spoken Japanese by native speakers of Japanese with and without pitch accent in their variety and by students learning Japanese with and without pitch accent in their L1. A gating paradigm using Japanese pitch accented (minimal pair) words was employed to determine how native pitch accent patterns influence word recognition. The results suggest a facilitatory effect of L1 pitch accent background, and of Japanese L1 in word recognition.

1 Introduction

While tone languages like Chinese use pitch variation on most syllables/words, a few languages have restricted pitch accent contrasts only on certain words (Roach, 2001). Japanese is regarded as one such pitch accent language, though there are interesting differences among varieties of Japanese in this respect – most varieties of Japanese have regional pitch accent patterns and a few varieties do not use pitch accent to distinguish words. A number of Scandinavian languages also fall into this category.

It has been reported that for selection of words in the mental lexicon, not only phonetic segment information but other articulatory phonetic information is used (Lahiri & Marslen-Wilson, 1991), which can include accentual information. Our study follows up previous research findings on the use of pitch accent in Japanese word recognition. A study of the use of pitch accent in Japanese word recognition (Otake & Cutler, 1999) found that pitch accent information was used in lexical selection by both Tokyo pitch accent speakers and accentless speakers of Japanese; however, contact with a local accentless dialect influenced the processing of pitch accent for word recognition in standard (Tokyo) Japanese. Another study revealed that both pitch accent and non-pitch accent speakers used pitch information in word activation; however pitch accent speakers demonstrated higher sensitivity in their performance than non-pitch accent speakers (Otake, 2002). A Japanese pitch accent identification test (Hirata, Ayusawa, Nakagawa, & Odaka, 1997) revealed that L1 subjects' performance differed depending on regional accentual background. It also showed such perceptual differences among L2 speakers of Japanese, depending on the subjects' L1.

Japanese pitch accent can be acquired by learners with different suprasegmental backgrounds (Nishinuma, 1994; Nishinuma, Arai, and Ayusawa, 1996). However, a pitch accent distinction test conducted on native speakers of English and native speakers of French learning Japanese found that their performance was significantly different from that of native Japanese speakers, reflecting perceptual

strategies more appropriate to their L1 (Nishinuma, 1994; Nishinuma, Arai, and Ayusawa, 1996). The study reported here has as subjects L2 learners of Japanese whose L1 is English, and whose L1 is Norwegian. English is a stress accent language, and Norwegian is, like Japanese, a pitch accent language. Currently, no studies on the use of pitch accent contrast in Japanese by Norwegian learners appear to have been carried out. This study investigates the following questions and related hypotheses:

Q1. Is the perception of pitch accent in word identification significantly different among native Japanese speakers with different accent backgrounds? If so, in what ways does it differ?
H1. There will be statistically greater sensitivity in the use of pitch information in word identification by native Japanese speakers with a pitch accent background.

Q2. Is pitch accent information involved in recognizing Japanese words by L2 speakers of Japanese with different L1 accent backgrounds?
H2. There will be statistically greater dependence on the use of F0 contour in Japanese word identification by L2 speakers with a L1 pitch accent background.

It was envisaged that we could also investigate whether L1 performance differs from L2 performance, and if so, to what extent.

To test these hypotheses, a gating task was used. This is a part of larger study, including a word discrimination task which has been described in more detail elsewhere (Honda, 2007).

2 Methodology

The gating paradigm addresses the question of how much acoustic-phonetic information is needed for subjects to identify a presented utterance (Grosjean, 1996), and allows precise control over the segmental and prosodic cues presented to subjects in a spoken word recognition task.

A recorded spoken word stimulus was altered so that only a small amount at the beginning of the test word was heard in the first instance. The rest was then presented in segments (‘gates’) of increasing duration of 25 msec, the longest (base) stimulus corresponding to the entire utterance, which was in the range between 200 and 300 msec. Subjects were asked to indicate what they heard after the presentation of each stimulus; hence the answer to each stimulus indicated how the subjects’ percept developed as more information became available. Subjects were asked to respond to what they heard in written form on an answer sheet. For each test word, the proportion of correct word identification responses was calculated for each gate that was presented, and comparisons were made between the subject groups.

2.1 Subjects

Four groups of subjects participated in this study: Group 1 consisted of 17 Japanese L1 speakers with the Tokyo pitch-accent dialect (JL1_PA); Group 2 had 17 Japanese L1 speakers with a non-pitch accent dialect (JL1_NPA); Group 3 was made up of 17 intermediate/advanced-level learners of Japanese as a foreign language who had English as their L1 (JL2_E); and Group 4 had 25 intermediate/advanced learners of Japanese with Norwegian as their L1 (JL2_N). The subject groups were homogeneous in that they were all university students, except for three English subjects.

2.2 Materials

Six words (three minimal pairs) were selected to be used in this experiment: '*kaki* (LH *persimmon*)', '*kaki* (HL *oyster*)', '*ame* (LH *sweets*)', '*ame* (HL *rain*)', '*hashi* (LH *bridge*)', and '*hashi* (HL *chopsticks*)'. These are all two-mora words and have pitch increase/declination after the three different medial consonants, [k], [m], and [ʃ]. The words were gated and built into frame sentences, including neutral sentences and semantically biased sentences for '*ame*'; for instance, '*korewa* X *desu*. *This is X,'*or 'X *wo tabemasu. I eat X,' etc.* In this way, the gated stimulus of each word, always preceded or followed by the carrier phrase, was presented in increasingly larger fragments, incrementing in each case by 25msec. This resulted in the stimuli having acoustic information after the gate, which deviates from the usual gating procedure. However, in order to produce a natural effect in the Japanese sentences, we decided to adopt this procedure. The carrier frame had a function word immediately following the gate, and the transition between the fragment or full word and the function word was held constant.

The twelve test sentences presented are as follows: *Sorewa* ***kaki*** *desuka*. (Is that a persimmon?); ***Kaki*** *wo tabemasu*. (I eat the persimmon.); *Sorewa* ***kaki*** *desuka*. (Is that an oyster?); ***Kaki*** *wo tabemasu*. (I eat the oyster.); *Sorewa* ***ame*** *desu*. (That is sweet.); ***Ame*** *wo tabeteimasu*. (I'm eating a sweet.); *Arewa* ***ame*** *desu*. (That is rain.); ***Ame*** *ga futteimasu*. (It is raining.); *Arewa* ***hashi*** *desu*. (That is a bridge.); ***Hashi*** *ga miemasu.* (You can see the bridge.); *Korewa* ***hashi*** *desu*. (These are chopsticks.); ***Hashi*** *ga arimasu*. (The chopsticks are there.).

2.3 Procedure

The recording took place in a sound treated room in the Speech Research Laboratory at the University of Reading. The sentences were elicited from a Japanese L1 speaker with the standard pitch accent and digitally recorded on the speech analysis programme, *Audition*, onto the computer connected by a microphone (Pioneer Dynamic Microphone DM-21).

Accurate segmentation was achieved by using the digital speech edit programme, *SFS* (Speech Filing System). The pitch contour (in blue) with waveform and spectrogram of each fragment of the first word, '*kaki*' (accent type 0) are presented as follows.

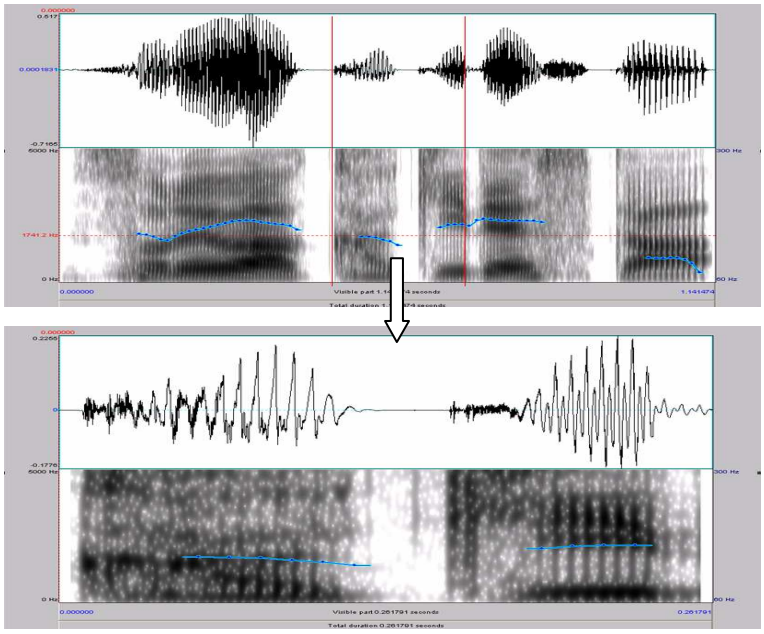


Fig. 1 Pitch contour with waveform and spectrogram of the stimulus: full sentence version 「*Sorewa kaki desuka*. (Is that a persimmon?)」 and extended version of the full gate word (the part between red lines above) 「*kaki* (persimmon)」 (275msec)

Subjects were tested individually using a laptop computer from which they listened to the stimuli bi-aurally over headphones (Synthesiser EH1430). After they had signed the consent form and filled in the language background questionnaire, 13 copies of the answer sheet were provided: one for the practice test and 12 for the official test. They were first asked to write down their name, which was later used to identify each subject. Most of the subjects found the gating task straightforward and the experimenter observed that no difficulty was encountered.

2.4 Analysis

Subsequent to scoring the word decision point (msec) of each subject by hand, comparisons among the subject groups were conducted. The percentage of correct answers for each fragment was subjected to a sign test of statistical significance of the difference from chance. Analyses of variance (ANOVA) were carried out to compare the mean of word decision times among groups. Although the gating stimuli are incremental, ANOVA can still give an overall sense of what

occurred among the group comparisons, and it was decided to refer to it. We labeled the point where subjects identified the correct answer for the first time, as the 'decision point'.

3 Results

For each target word, group mean decision points are presented with p-values in ANOVA and plots which show the confidence intervals for the respective groups. The expected number of subjects is doubled because the same word was presented twice in separate frame sentences.

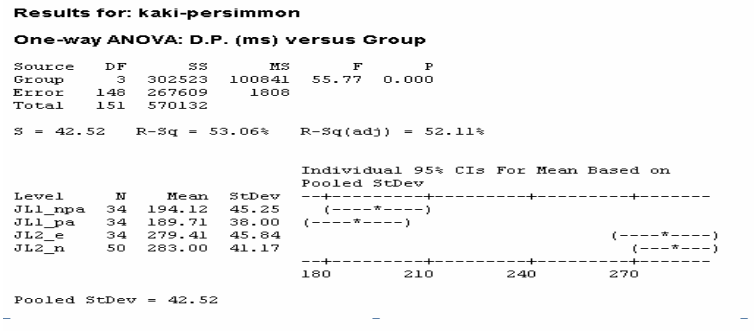


Fig. 2 ANOVA of word decision point: independent variable was subject group (*kaki*, persimmon)

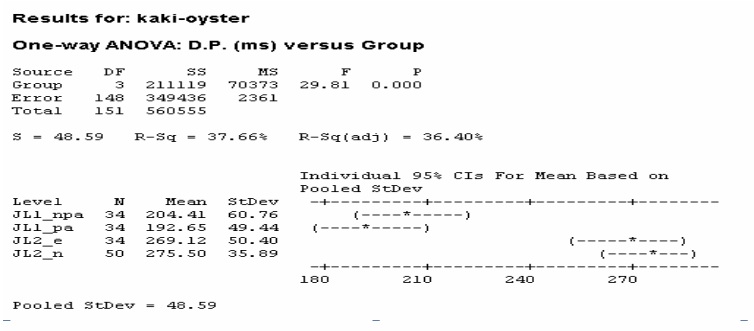


Fig. 3 ANOVA of word decision point: independent variable was subject group (*kaki*, oyster)

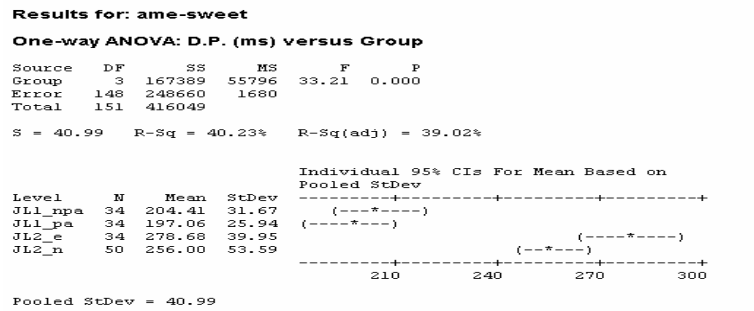
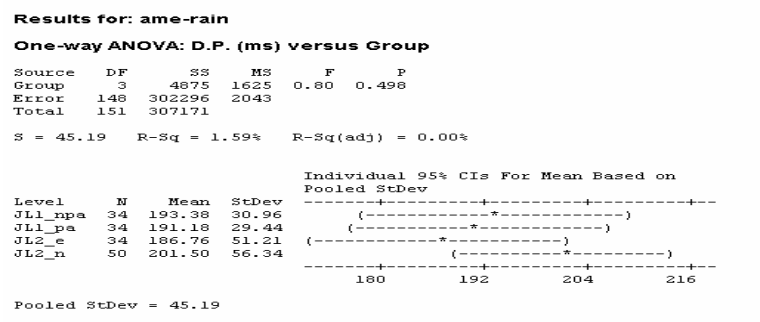


Fig. 4 ANOVA of word decision point: independent variable was subject group (*ame*, sweets)



Results for: ame-rain

One-way ANOVA: D.P. (ms) versus Group

Source	DF	SS	MS	F	P
Group	3	4875	1625	0.80	0.498
Error	148	302296	2043		
Total	151	307171			

S = 45.19 R-Sq = 1.59% R-Sq(adj) = 0.00%

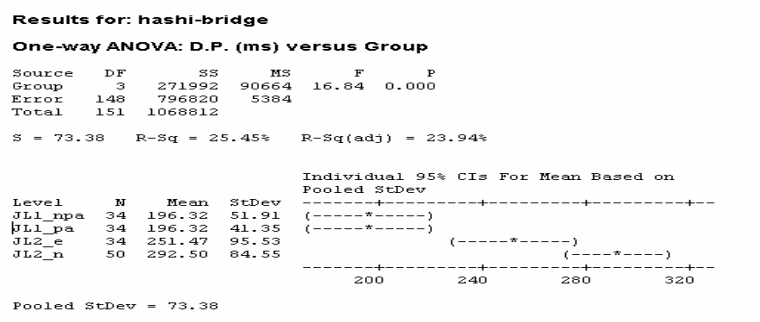
Level	N	Mean	StDev
JL1_npa	34	193.38	30.96
JL1_pa	34	191.18	29.44
JL2_e	34	186.76	51.21
JL2_n	50	201.50	56.34

Individual 95% CIs For Mean Based on Pooled StDev

180 192 204 216

Pooled StDev = 45.19

Fig. 5 ANOVA of word decision point: independent variable was subject group (*ame*, rain)



Results for: hashi-bridge

One-way ANOVA: D.P. (ms) versus Group

Source	DF	SS	MS	F	P
Group	3	271992	90664	16.84	0.000
Error	148	796820	5384		
Total	151	1068812			

S = 73.38 R-Sq = 25.45% R-Sq(adj) = 23.94%

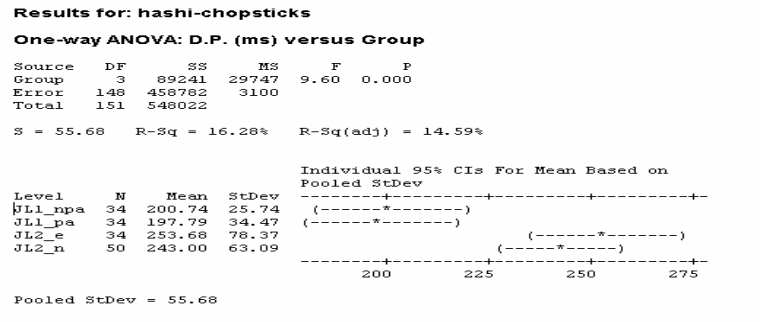
Level	N	Mean	StDev
JL1_npa	34	196.32	51.91
JL1_pa	34	196.32	41.35
JL2_e	34	251.47	95.53
JL2_n	50	292.50	84.55

Individual 95% CIs For Mean Based on Pooled StDev

200 240 280 320

Pooled StDev = 73.38

Fig. 6 ANOVA of word decision point: independent variable was subject group (*hashi*, bridge)



Results for: hashi-chopsticks

One-way ANOVA: D.P. (ms) versus Group

Source	DF	SS	MS	F	P
Group	3	89241	29747	9.60	0.000
Error	148	458782	3100		
Total	151	548022			

S = 55.68 R-Sq = 16.28% R-Sq(adj) = 14.59%

Level	N	Mean	StDev
JL1_npa	34	200.74	25.74
JL1_pa	34	197.79	34.47
JL2_e	34	253.68	78.37
JL2_n	50	243.00	63.09

Individual 95% CIs For Mean Based on Pooled StDev

200 225 250 275

Pooled StDev = 55.68

Fig. 7 ANOVA of word decision point: independent variable was subject group (*hashi*, chopsticks)

There are statistically significant differences among the four groups in the identification point of the word (p=0.000) on 5 out of 6 words presented. The JL1_PA group shows the earliest mean decision point among the 4 groups on all the words apart from *ame*, rain and *hashi*, bridge. The JL1_NPA group has

longer mean decision points than the JL1_PA, consistent with H1, but this difference is statistically insignificant. The JL2_E group does not always have greater means than the JL2_N group, which is inconsistent with H2.

The poorer performance of the JL2 groups included a great number of zero responses, suggesting to us that JL2 speakers, even advanced learners of Japanese, are not proficient enough to be tested by using the gating technique, and in the end full-scale comparisons were made only between L1 speaker groups.

A further question, whether the accent pattern of the candidate words guessed by subjects matches that of target words, is also of interest and requires supplementary analysis.

4 Discussion

A significant difference in sensitivity and accuracy between the L1 and L2 groups was observed. L1 Japanese speakers with a non-pitch accent background tended to take longer to identify the word presented, but the mean time difference is too small to be significant. However, no significant difference among the four subject groups was found on the semantically biased sentence; all groups performed equally well, and this may support a previous study which found that less advanced learners rely on the available semantic information while advanced learners and L1 speakers rely on phonetic information in their pre-perceptual processing. While highly significant differences were observed between the L1 and L2 groups, the Norwegian group demonstrated tendencies similar to those of the English group, contrary to expectation. This result might be dependent on the methodology employed; the non-native groups showed a greater level of internal inconsistency in test performance, and a few subjects who showed a total lack of awareness of the Japanese standard pitch might be skewing the results. One of the factors influencing this was the kind of Japanese they had been exposed to: some learners from the Kansai area, where different pitch systems from standard Japanese are employed, performed poorly in both the perception and production tests. Further research would be required to investigate degree of contact with standard Japanese and other sociolinguistic factors which might influence pitch contrast perception. Another pedagogical factor may also be very important: the role of pronunciation, training, and especially training in suprasegmental aspects of speech in Japanese second and foreign language teaching.

References

Grosjean, F. (1996). Gating. *Language & Cognitive Processes* 11.6: 597–604.

Hirata, E., T. Ayusawa, C. Nakagawa, and K. Odaka. (1997). Perception test of Tokyo accent in Japanese. *Study Report 1997, The Ministry of Education, Japan,* pp.61–72.

Honda, M. (2007) The Role of Prosody in Japanese: the use of pitch information in spoken word recognition by L1 and L2 speakers. In *Proceedings of the 5th Cambridge Postgraduate Conference in Language Research*, Cambridge (forthcoming).

Lahiri, A. and W. D. Marslen-Wilson. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38, 245–294.

Nishinuma, Y. (1994). How do the French perceive tonal accent in Japanese?: Experimental evidence. In *Proceedings of the 3rd International Conference on Spoken language Processing*, pp. 1739–1742. Yokohama.

Nishinuma, Y., M. Arai, and T. Ayusawa. (1996). Perception of tonal accent by American learning Japanese. In *Proceedings of the 4th International Conference on Spoken language Processing*, pp. 646-649. Philadelphia.

Otake, T. (2002). A role of prosody in spoken-word recognition. *Journal of the Phonetics Society of Japan* 6(2), 56–6.

Otake, T. and A. Cutler. (1999). Perception of suprasegmental structure in a nonnative dialect. *Journal of Phonetics* 27(3), 229–253.

Roach, P (2001) *Phonetics*. Oxford: Oxford University Press.