

## The perception of non-native lexical pitch accent by speakers of ‘accentless’ Japanese dialects

Akira Utsugi<sup>1</sup>, Masatoshi Koizumi<sup>2</sup>, Reiko Mazuka<sup>1,3</sup>

<sup>1</sup> Laboratory for Language Development, Brain Science Institute, RIKEN, Japan

<sup>2</sup> Department of Linguistics, Tohoku University, Japan

<sup>3</sup> Department of Psychology and Neuroscience, Duke University, USA

autsugi@brain.riken.jp, koizumi@sal.tohoku.ac.jp, mazuka@brain.riken.jp

### Abstract

While Standard (Tokyo) Japanese has a lexical tonal system known as a system of ‘lexical pitch accent’, there are some varieties of Japanese, called ‘accentless’ dialects, which do not have any lexical tonal phenomena. We investigated how the speakers of those dialects perceive Standard Japanese accent, which is nonexistent in their native dialect’s phonology. The results of the Sequence Recall task showed that their scores were lower than those of control (Standard Japanese) participants. We also found a large variance in the results of ‘accentless’ participants, which was probably caused by their exposure to Standard Japanese.

**Index Terms:** non-native speech perception, pitch accent ‘deafness’, sequence recall task

### 1. Introduction

Perception of non-native sounds has been a good empirical source to construct models of speech perception. It is well known that a sound contrast in second or foreign languages is often difficult to perceive if it does not exist in the native language. A famous example is the segmental contrast of English /r/ and /l/ for Japanese listeners, whose native language does not have such a contrast in its phonology [1-2]. A similar phenomenon is found in suprasegmental contrasts as well. For example, Mandarin tone contrasts are difficult for English listeners [3-4]. French listeners have a difficulty in distinguishing stress contrasts in Spanish, a phenomenon known as stress ‘deafness’ [5]. Such empirical studies have provided the basis for a theoretical model [6] and for the development of neuroscientific approaches to speech perception [7-8].

A cross-dialectal study of lexical pitch accent in Japanese is potentially a significant case study of speech perception. Firstly, relatively little has been known about the perception of lexical pitch accent compared to other lexical suprasegmentals such as tone and stress. Secondly, the advantage of a cross-dialectal study compared to a cross-linguistic study is that it allows a better control of experimental influence factors. For languages differ in many ways such as syntax, segmental phonology, prosody, etc., it is difficult to determine which factors affect experimental results. In the present study, we compare two groups of Japanese varieties which have a clear difference in the presence or absence of lexical pitch accent while having similar syntax and segmental phonology.

One group we deal with is Standard (Tokyo) Japanese. Its pitch lexically contrasts in whether a word has a local pitch fall or not, and, if it has, on which moraic boundary the fall is aligned. This pitch fall has been called an ‘accent kernel’, an ‘accent’, or a ‘lexical pitch accent’. Unlike so-called tone

languages such as Mandarin, in which tones are specified on every syllable, it has been argued that Standard Japanese requires sparse tonal specification [9-10]. The other group we deal with is speakers of so-called ‘accentless’ dialects. This group of dialects does not employ pitch at the lexical level at all [11], although it employs pitch at the sentence level [12-13].

It has often been said that speakers of ‘accentless’ dialects have difficulties acquiring the lexical pitch accent of Standard Japanese [14]. This is reminiscent of stress ‘deafness’ in French. However, beyond episodic stories, there has been little evidence for their difficulties in perception of standard lexical pitch accent. Thus, one of the aims of this study is to show clear evidence of their difficulty in perception. Previous studies showed that subjects from areas of ‘accentless’ dialects clearly exploited information of Standard Japanese lexical pitch accent in perception experiments [15-16]. However, it is possible that these results were due to a non-phonological factor. Since the pitch accent contrast is phonetically large in Standard Japanese, participants might have exploited acoustic cues in the experimental tasks.

The present study focuses on perception at the phonological rather than the acoustic level. We applied a task called ‘high phonetic variability sequence recall task (hereafter, ‘sequence recall task’), which was developed in stress ‘deafness’ studies [17-18]. This task has a high memory load so that the acoustic level is not accessible and, therefore, phonological representations are highlighted.

One potential problem in studies of the ‘accentless’ dialects is a possible influence of Standard Japanese. Many dialectological studies have reported that a change from ‘accentless’ to pitch accent dialects is in progress, probably because of the exposure to Standard Japanese [14, 19-21]. We expect that the present approach will work as a tool to sort out ‘genuine’ speakers of the ‘accentless’ dialects as well. Thus, another aim of this study is to see how robust this approach is to sort out such speakers. This is an important process before moving on to further steps such as neuroscientific approaches to the perception of lexical pitch accent, in which genuine accentless speakers will be needed.

### 2. Method

#### 2.1. Subjects

Twenty four subjects, aged between 18 and 25, participated in the experiment. They were classified into two groups: ‘accentless’ and ‘standard’, the latter being included as a control group. The accentless group consisted of participants from a part of the area of the ‘accentless’ dialects: Southern Miyagi, Southern Yamagata, and a large part of Fukushima

prefecture. The exact area was determined based on literature on Japanese dialectology [22]. The standard group consisted of participants from the area of Standard Japanese: Tokyo, Saitama, Kanagawa, and Chiba prefectures. Each group consisted of twelve subjects (6 females and 6 males).

## 2.2. Stimuli

Several tokens of the two non-words /manu/ and /menu/ were recorded by six native speakers of Standard Japanese, or varieties of Japanese with the same tonal system as Standard Japanese. The recorded tokens included both initial-mora-accented and unaccented pitch patterns.

One token of each non-word of each speaker was chosen as a source sound for pitch resynthesis. Three speakers' source sounds were chosen from initial-mora-accented pitch patterns, while the other three speakers' source sounds were from unaccented pitch patterns. Tokens with glottalized final mora were not chosen as source sounds because they sounded unnatural after pitch resynthesis.

The selected source sounds were resynthesized by the PSOLA algorithm on Praat, software developed by P. Boersma and D. Weenink (University of Amsterdam), to manipulate fundamental frequencies ( $f_0$ ). Three versions of /manu/ (high-fall, rise-fall, and rise-high) and one version of /menu/ (rise-fall) were created for each speaker, as shown in Figure 1. In each resynthesized stimulus, a turning point was fixed on the boundary between the first and second morae. The maximal and minimal  $f_0$  values in each stimulus were set to the same values as in the corresponding source stimulus. Peak intensity was equalized for all the stimuli.

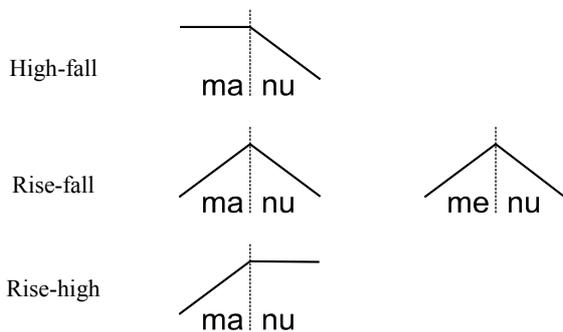


Figure 1:  $F_0$  shapes of re-synthesized stimuli.

High-fall and rise-high were expected to be identified as first-mora-accented and second-mora-accented (or unaccented words), respectively, for the standard group. However, high-fall is not a typical pitch shape of a Standard Japanese first-mora-accented word. It is known that, in fact, it begins with a low pitch and has an  $f_0$  peak within or often slightly after the first mora [23-25], which is closer to the rise-fall pattern in Figure 1. Thus, rise-fall was expected to be identified as the same category as high-fall, i.e. the first-mora-accented category.

Three contrasts were created based on these stimuli, as shown in Table 1. Each contrast consisted of two items and each item consisted of six stimuli (i.e. stimuli based on six speakers' pronunciations).

The first contrast was expected to be difficult to discriminate for both groups, since the two stimuli were to be identified as the same category in 'accentless' as well as Standard varieties of Japanese. The second contrast, which is lexical in Standard Japanese, was expected to be easy to

discriminate for the standard group, while it was expected to be difficult to discriminate for the accentless group, who has no lexical pitch contrast. The third contrast, i.e. segmental, was expected to be easy to discriminate for both groups.

Table 1. *Contrasts used in the experiment.*

Contrast	First item	Second item
i. Non-lexical pitch	/manu/ with rise-fall	/manu/ with high-fall
ii. Lexical pitch	/manu/ with rise-fall	/manu/ with rise-high
iii. Segmental	/manu/ with rise-fall	/menu/ with rise-fall

## 2.3. Procedure

The experiment was conducted on a laptop computer, on which the experimental program (created with E-Prime 2.0) was run. Participants listened to the experimental stimuli through headphones. The experiment consisted of three parts, each corresponding to each contrast. The order of the parts was counterbalanced. Each part consisted of a learning phase, a warm-up phase, and an experimental phase.

In the learning phase, subjects were first told that they would learn words in a foreign language, and were asked to press the key '1' on the laptop keyboard, which initiated the replay of all the six stimuli of the first item (e.g., in the case of contrast (ii) in Table 1, it was /manu/ with rise-fall). Then, they were asked to press the key '2', which initiated the replay of all the six stimuli of the second item (e.g., in the case of the contrast (iii), it was /menu/ with rise-fall). After that, they were asked to press the keys '1' and '2' as often as they wanted to associate the keys with the items; by pressing each key, one of the stimuli of the corresponding item was played. This was followed by a training block, in which stimuli were played subsequently and subjects were asked to respond whether it was '1' or '2' by pressing keys after each stimulus. This block continued either until they answered correctly for seven times in a row, or until the number of trials reached twenty.

The warm-up phase consisted of one block, in which the subjects listened to a sequence of two stimuli and were asked to type associated keys. For example, in the contrast (iii), when they heard a sequence of /manu/ and /menu/ in this order, the answer was 12. All four possible sequences (11, 12, 21, 22) were randomly presented without repetition; thus, the block consisted of four trials. Within each trial, each word was chosen from a different speaker. A trial included 80 ms of ISI and was followed by the word 'hai' (yes). A feedback was shown on the screen after each trial.

The experimental phase consisted of three blocks, in which two-word, three-word, and four-word sequences, respectively, were presented. The blocks had the same structure as the warm-up block, but without including feedback. Another difference was the number of repetitions and trials. In the two-word-sequence block, each of the four possible sequences was used four times. In the three-word-sequence block, each of the eight possible sequences was used two times. In the four-word-sequence block, each the sixteen possible sequences was used once. Thus, every experimental block contained sixteen trials.

The experiments for the accentless group were conducted at Tohoku University, which is located in the area of the accentless dialects, and those for the standard group were

conducted at RIKEN, which is located in the area of Standard Japanese. At most two subjects were tested at the same time in a quiet room.

## 2.4. Data analysis

Outliers were eliminated from analysis by the following procedure. First, an accuracy rate was calculated for each block (i.e. the two-, three-, and four-sequence blocks) for each contrast in each subject, and the rates of the three blocks were averaged. Then, these averaged scores were transformed into z-scores for each group's contrast. If a subject had a score beyond  $\pm 2.246$  in at least one of the contrasts, that subject was treated as an outlier. The criterion was based on Van Selst and Jolicoeur [26].

Data for all the subjects other than the outliers were subjected to analysis. The analysis was based on accuracy rates for each block. Statistic tests were conducted as shown below in detail.

## 3. Results

Figures 2 and 3 show the accuracy rates for the accentless and standard groups, respectively. In both groups, the rates for non-lexical pitch contrast were lower and those for segmental contrast were higher. A between-group difference was found in lexical pitch contrast, where scores for the accentless group were lower than those for the standard group.

The accuracy rates for all participants were subjected to an ANOVA with one between-subject factor, Group (accentless vs. standard), and two within-subject factors, Contrast (non-lexical pitch vs. lexical pitch vs. segmental) and Sequence Length (two vs. three vs. four). We found a significant interaction between Group and Contrast ( $F(2,38) = 3.5875, p = 0.0374$ ). We also found significant main effects for Contrast ( $F(2,38) = 115.5619, p < 0.001$ ) and Sequence Length ( $F(2,38) = 54.8247, p < 0.001$ ), and a significant interaction between Contrast and Sequence length ( $F(4,76) = 7.2242, p < 0.001$ ). The following effects were statistically not significant: the main factor of Group ( $F(1,19) = 0.1888, p = 0.6688$ ), the interaction between Group and Sequence length ( $F(2,38) = 0.4663, p = 0.6309$ ), and the interaction of all the three factors ( $F(4,76) = 0.6278, p = 0.6441$ ).

As post-hoc tests, one-way ANOVAs (factor: Group) were conducted for each contrast. The significance levels were set to 0.016 (Bonferroni correction). The results revealed that there was a significant effect of lexical pitch contrast ( $F(1,61) = 6.7556, p = 0.0117$ ), while there were no significant effects of non-lexical pitch contrast ( $F(1,61) = 2.2273, p = 0.1407$ ) and segmental contrast ( $F(1,61) = 2.8099, p = 0.0988$ ).

We also conducted a Bartlett Test of Homogeneity of Variances with Group as an independent factor for each contrast. The significance levels were set to 0.016 (Bonferroni correction). The results revealed that the difference of variances was significant for the lexical pitch contrast ( $p < 0.001$ ), suggesting that the accentless group had a larger variance than the standard group. A significant difference of variances was also found for the segmental contrast ( $p < 0.001$ ). No significant difference was found for the non-lexical pitch contrast ( $p = 0.763$ ).

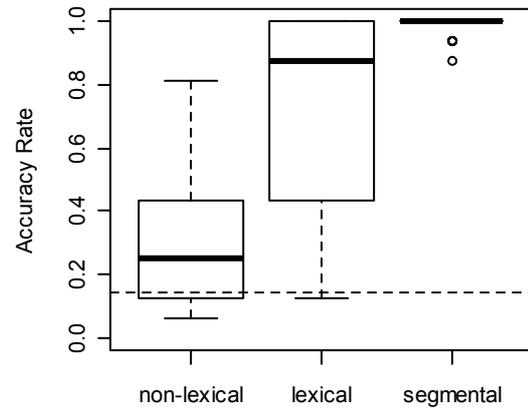


Figure 2: Accuracy rates for the accentless group. Non-lexical: non-lexical pitch contrast, lexical: lexical pitch contrast, segmental: segmental contrast. Horizontal dashed line indicates the chance level (0.14583).

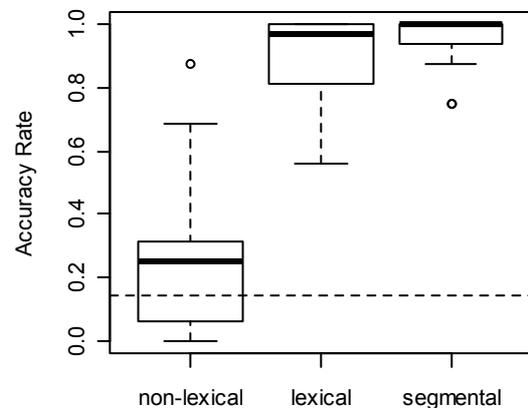


Figure 3: Accuracy rates for the standard group. Non-lexical: non-lexical pitch contrast, lexical: lexical pitch contrast, segmental: segmental contrast. Horizontal dashed line indicates the chance level (0.14583).

## 4. Discussion

As stated in the previous section, the interaction between Group and Contrast was significant. According to the post-hoc test, the two groups differed only in their discrimination of the lexical pitch contrast. This result agrees with our prediction; since there is no contrast of lexical pitch accent in traditional 'accentless' dialects, the discrimination of lexical pitch accent

should be difficult, and thus the scores were lower for the accentless group.

It should also be noted that the mean score of the lexical pitch contrast was clearly above chance level (0.14583) even for the accentless group. As suggested by the larger variance in lexical pitch contrast for this group, some subjects in the accentless group had very high scores, while some had low scores close to chance level. We consider that this large variance is a reflection of the various degrees of standardization in each subject in the accentless group. Subjects with higher scores would be affected more by Standard Japanese, while those with lower scores would be less affected. This suggests that our experiment can be a good tool to sort out 'genuine' accentless speakers. Alternatively, one might consider that this was rather a reflection of subjects' sensitivity to non-native contrasts. However, if higher scores were due to subjects' high sensitivity, such subjects should have scored high in the non-lexical pitch contrast as well. Thus, the sensitivity view is not supported.

We also found a clear main effect of Contrast. As can be seen from Figures 2 and 3, the non-lexical pitch contrast showed lower scores and the segmental contrast showed higher scores. These results agree with our prediction. Interestingly, the scores for the lexical pitch contrast were lower than those of the segmental contrast even in the results of the standard group. This differs from the results of the stress-deafness study by Dupoux and his colleagues [5, 17-18]. In their studies, the scores for the stress contrast were as high as those for the segmental contrast for Spanish speakers, who have a lexical stress contrast. The question arises whether Spanish stress contrast and Japanese pitch accent contrast differ in some way in speech perception even though they are equally lexical suprasegmental contrasts in a phonological sense. This remains an open question.

## 5. Conclusion

Our experiment showed a clear difference in the perception of lexical pitch accent between Standard Japanese and 'accentless' dialect speakers. Some (but not all) subjects from the area of the 'accentless' dialects had lower scores in the high phonetic variability sequence recall task. This suggests that, when acoustic cues are not sufficiently accessible, such subjects have difficulties in the discrimination of contrasts of Standard Japanese lexical pitch accent. In other words, we found pitch accent 'deafness', which parallels stress 'deafness' in the sense of Dupoux et al. [5, 17-18]. We also found high variance in the 'accentless' group, suggesting some degree of standardization. The present task worked as a good tool to sort out 'genuine' accentless speakers.

The future direction will be to investigate how such cross-dialectal perceptual differences are reflected in brain activities.

## 6. Acknowledgements

We would like to thank Yuko Sogabe, Mihoko Hasegawa, and Satoshi Imamura for helping our experiment, and Sho Tsuji and Kuniyoshi Tanaka for their invaluable comments on the manuscript. This work was supported by KAKENHI (No. 21820079).

## 7. References

[1] H. Goto, "Auditory perception by normal Japanese adults of the sounds "L" and "R", " *Neuropsychologia*, 9, 317-323, 1971.

- [2] K. Miyawaki, *et al.*, "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Perception and Psychophysics*, 18, 331-340, 1975.
- [3] C. Kiriloff, "On the auditory perception of tones in Mandarin," *Phonetica*, 20, 63-67, 1969.
- [4] Y. Wang, *et al.*, "Training American listeners to perceive Mandarin tones," *JASA*, 106, 3649-3658, 1999.
- [5] E. Dupoux, *et al.*, "A Deafness "Deafness" in French?," *Journal of Memory and Language*, 36, 406-421, 1997.
- [6] C. Best, "Learning to perceive the sound pattern of English," in *Advances in Infancy Research*, C. Rovee-Collier and L. Lipsitt, Eds., Hillsdale: Ablex, 1995, 217-304.
- [7] D. Callan, *et al.*, "Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models," *Neuroimage*, 22, 1182-1194, 2004.
- [8] B. Chandrasekaran, *et al.*, "Relative influence of musical and linguistic experience on early cortical processing of pitch contours," *Brain and Language*, 108, 1-9, 2009.
- [9] S. Kawakami, "On the relationship between word-tone and phrase-tone in Japanese," *Onsei no Kenkyuu*, 169-177, 1961.
- [10] J. Pierrehumbert and M. Beckman, *Japanese tone structure*. Cambridge, MA: MIT Press, 1988.
- [11] Z. Uwano, "Classification of Japanese accent systems," in *Cross-linguistic studies of tonal phenomena: Tonogenesis, typology, and related topics*, S. Kaji, Ed., Tokyo: ILCAA, 1999, 151-178.
- [12] Y. Yamaguchi, *Nihongo hoogen ikkei akusento no kenkyuu*. Tokyo: Hituzi, 1998.
- [13] Y. Igarashi, "Typology of intonational phrasing in Japanese dialects," in *Prosodic typology*, 2, S.-A. Jun, Ed., New York: Oxford University Press, forthcoming.
- [14] Y. Mase, "Gengo keisei ni oyobosu terebi oyobi toshi no gengo no eikyoo," *Kokugogaku*, 125, 1-19, 1981.
- [15] T. Otake and A. Cutler, "Perception of suprasegmental structure in a non-native dialect," *Journal of Phonetics*, 27, 229-253, 1999.
- [16] T. Otake and M. Higuchi, "A role of pitch accent in spoken-word recognition in accentless Japanese dialects: Evidence from Fukushima listeners," in *Speech Prosody 2004*, Nara, Japan, 2004, 41-44.
- [17] E. Dupoux, *et al.*, "A robust method to study stress 'deafness'," *JASA*, 110, 1606-1618, 2001.
- [18] E. Dupoux, *et al.*, "Persistent stress 'deafness': The case of French learners of Spanish," *Cognition*, 106, 682-706, 2008.
- [19] R. Sato, "Mukey akusento chiiki ni okeru akusento no kyoutsuuka: Utsunomiyashi ni okeru shouchou kara," in *Gendai hougengaku no kadai* 2, Tokyo: Meiji Shoin, 1984, 211-234.
- [20] F. Inoue, "Akusento no seisei to chikaku: Kantou ni okeru chiikisa to nenreisa," in *Kindaichi Haruhiko hakase koki kinen robunshu*, 2, Tokyo: Sanseido, 1984, 119-153.
- [21] T. Onishi, "Hoogen akusento no genzai: Sendaishi hoogen ni okeru akusento no kakutoku wo chuusin ni," *Nihongogaku*, 11, 98-113, 1992.
- [22] T. Hirayama, *Nihongo onchoo no kenkyuu*. Tokyo: Meiji Shoin, 1957.
- [23] J. V. Neustupný, "Nihongo no akusento wa kootei akusento ka," *Onseigakkai Kaihoo*, 121, 1-7, 1966.
- [24] M. Sugito, *Nihongo akusento no kenkyuu*. Tokyo: Sanseido, 1982.
- [25] T. Ishihara, "Tonal alignment in Tokyo Japanese," PhD, The University of Edinburgh, 2006.
- [26] M. Van Selst and P. Jolicœur, "A Solution to the Effect of Sample Size on Outlier Elimination," *The Quarterly Journal of Experimental Psychology*, 47A, 631-650, 1994.