Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

A robust method to detect dialectal differences in the perception of lexical pitch accent

Akira Utsugi (1,2), Masatoshi Koizumi (3) and Reiko Mazuka (1,4)

(1) Laboratory for Language Development, Brain Science Institute, RIKEN, Wako, Japan
(2) JSPS Postdoctoral Fellow, Japan
(3) Department of Linguistics, Tohoku University, Sendai, Japan
(4) Department of Psychology and Neuroscience, Duke University, Durham, USA

PACS: 43.71.-k, 43.71.Hw

ABSTRACT

While Standard (Tokyo) Japanese has a lexical tonal system known as 'lexical pitch accent', there are some varieties of Japanese, called 'accentless' dialects, which do not have any lexical tonal phenomena. We investigated the differences in the perception of lexical pitch accent between the speakers of the accentless dialect and those of Standard Japanese, and the robustness of two approaches to investigate such dialectal differences. We conducted two experiments: categorical perception and sequence recall experiments. The former is an approach that has been traditionally employed to study the perception of phonological contrasts. The latter is a more recent method employed in studies of 'stress-deafness' in French by Dupoux and his colleagues, in which participants listen to sequences of several nonsense words and answer the order of the words. The results of the categorical perception experiment showed no clear dialectal differences. On the other hand, the results of the sequence recall task showed that the scores of the 'accentless' group were clearly lower than those of control (Standard Japanese) participants in the discrimination of nonsense words whose pitch accent differences corresponded to lexical differences in Standard Japanese phonology. Thus, it is concluded that the latter experimental approach is more robust to study dialectal differences in pitch accent perception than the former.

INTRODUCTION

This paper reports the results of experiments that comprise part of our project aiming at investigating how language experience affects the perception of pitch. Specifically, the project focuses on a cross-dialectal comparison of lexical pitch accent in Japanese. We consider that this approach is a significant case study of pitch perception for the following reasons. Firstly, although lexical tone has been studied extensively [1-3], relatively little is known about the perception of lexical pitch accent, which is another type of lexical system that employs pitch. While in lexical tonal systems like Mandarin, tone is lexically specified on most syllables within a word, in the lexical pitch accent system accent is lexically specified on only one mora or syllable within a word [4-5]. 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, Standard Japanese and so-called 'accentless' dialects, which have a clear difference in the presence or absence of lexical pitch accent while having similar syntax and segmental phonology.

'Accentless' dialects are defined by the lack of pitch specification at the word level in Japanese dialectology [6]. It should be noted that these dialects may have some pitch specification at the sentence level [7-8].

Although the definition of 'accentless' dialects is based on their phonological structure, peculiarities of these dialects have also been found in studies of speech perception. Japanese dialectologists have pointed out that speakers of these dialects have difficulties acquiring the lexical pitch accent of Standard Japanese [9]. This is reminiscent of stress 'deafness' in French [10-11]. Inoue conducted a lexical judgement task using minimal-pair words which differ only in the specification of lexical pitch accent in Standard Japanese with participants from Standard Japanese and accentless dialect regions [12]. He found that the accentless dialect speakers tended to confuse the minimal words. Otake and his colleagues conducted a judgement task using isolated syllables extracted from words and a gating task to examine cross-dialectal differences of pitch perception between Standard Japanese and accentless dialect speakers [13-14]. According to their experiments, although overall patterns were similar among groups, the accentless-region group performed slightly less sensitive in pitch accent perception. However, these previous experiments were based on lexical items. As Otake and Cutler suggest, it is likely that their results reflected between-group differences in how pitch accent information is stored in the mental lexicon [13]. Experiments based on lexical items as in the previous studies cannot answer the question of whether there is a between-group difference in pitch perception independent from the mental lexicon.

The present paper has two research questions. The first question is whether there is a cross-dialectal difference in pitch perception. To avoid influences of the mental lexicon, we adopted experimental paradigms based on nonsense words. The second question is, in case there is a cross-dialectal difference, what kind of approach robustly identifies it. Identifying a cross-dialectal difference in a robust behavioural experiment is a significant process before moving on to further steps such as neuro-scientific approaches.

It should also be noted that there could be an influence of Standard Japanese on accentless dialects. 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 [9, 12, 15-16]. This is another reason why we need a robust approach; if the approach is robust enough, it is expected that cross-dialectal differences are identified in such a complicated sociolinguistic situation.

We conducted two experiments, one on categorical perception (hereafter, 'CP'), and one employing the high phonetic variability sequence recall task (hereafter, 'sequence recall task').

EXPERIMENT 1

We conducted a CP experiment consisting of two tasks: identification and discrimination tasks. It is known that typical CP results have two features: (i) results of identification tasks show an S-shaped pattern, the abrupt rise of which indicates a category boundary, and (ii) results of discrimination tasks show a sharp peak around the category boundary estimated from the results of the identification task [17]. The aim of the present experiment is to test whether this experimental paradigm works for a cross-dialectal study of the perception of lexical pitch accent.

Methods

Subjects

Twenty-four subjects, aged between 19 and 27, 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 [18]. 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).

Stimuli

Several tokens were recorded by native speakers of Standard Japanese, or varieties of Japanese with the same tonal system as Standard Japanese, for both Experiment 1 and 2. Experiment 1 used two tokens of the non-word /manu/ spoken by a female speaker, in which one is initial-mora-accented and the other is unaccented. Fundamental frequency (f0) values were measured for these tokens, using the autocorrelation algorithm on Praat, a software developed by P. Boersma and D. Weenink (University of Amsterdam). The following values were measured;

- $f0_{HLmax}$: f0 maximum in the initial-mora-accented token
- $f0_{HLmin}$: f0 minimum within the second mora in the initial-mora-accented token
- $f0_{LHmax}$: f0 maximum in the unaccented token
- $f0_{LHmin}$: f0 minimum within the first mora in the unaccented token

Two sets of stimulus continua were created by manipulating an f0 contour of the recorded sound. The source sound for this manipulation was the initianl-mora-accented token among the two tokens above. PSOLA algorighm on Praat was used for the manipulation.

Set 1 is a continuum consisting of ten steps ranging from a rise-fall to rise-high contour, as in Figure 1. The turning point was fixed on the boundary between the first and second mora. The maximal f0 value was determined by averaging $f0_{HLmax}$ and $f0_{LHmax}$, and the minimal f0 value was by $f0_{HLmin}$ and $f0_{LHmin}$. They were 285 Hz and 160 Hz respectively. Steps were created to have equal intervals in semitones.

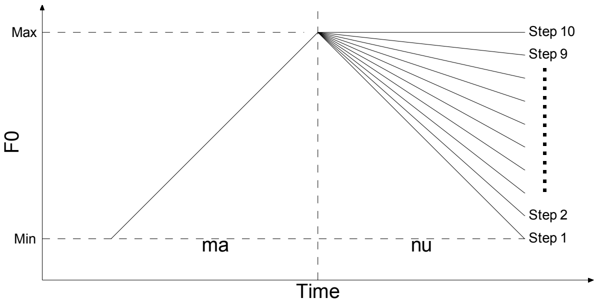

Figure 1. Stimulus continuum of Set 1

Set 2 is a continuum consisting of ten stimuli between a rise-fall and high-fall, as shown in Figure 2. The turning point, f0 maximum, f0 minimum, and step interval settings were the same as in Set 1.

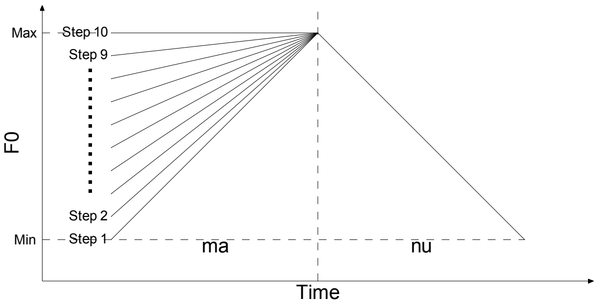

Figure 2. Stimulus continuum of Set 2

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 f0 peak within or often slightly after the first mora [19-21], which is closer to the rise-fall pattern in our stimuli. Thus, rise-fall was expected to be identified as the same category as high-fall, i.e. the first-mora-accented category, and to be categorically distinguished from rise-high, by Standard Japanese speakers.

Our predictions were as follows. The standard goup was expected to show the typical result of CP for Set 1 but not for Set 2. The accentless group was expected to show non-categorical results for both Set 1 and 2 since they do not have lexical pitch contrast in their dialect.

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 two tasks, an identification and a discrimination task. The identification task was follwed by the discrimination task. The participants took a break for fifteen minutes between the tasks.

The identification task was conducted in the ABX format, i.e. three stimuli were played in a row, in which the third stimulus (X) was the same as either the first (A) or second (B) stimulus. A and B were extreme stimuli in each stimulus continuum. The participants conducted both of the two sets, the order of sets being counterbalanced. The number of trials was 2 sets × 10 stimuli × 5 repetitions = 100. Short breaks were set after every 25 trials. ISI was set to 300 ms.

The discrimination task was conducted in the AX format, i.e. two simuli were played in a row, in which the second stimulus (X) was either the same as or different from the first stimulus (A). The participants were asked to answer whether X was the same as A or not. A and X were one step apart in stimulus continua since the two-step-apart pair was too easy to discriminate in the pilot experiment. We created 36 pairs for each set: 9 'different' pairs × 2 orders (AB or BA) + 18 'same' pairs. The participants conducted both of the two sets, the order of sets being counterbalanced.. The number of trials was 2 sets × 36 pairs × 5 repetitions = 360. Short breaks were set after every 20 trials. ISI was set to 300 ms.

In each task, the experimental block was preceded by a practice block, which consisted of four 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.

Results

Figures 3-6 show the results of the identification task. As can be seen, clear S-shapes were not observed in any set and group.

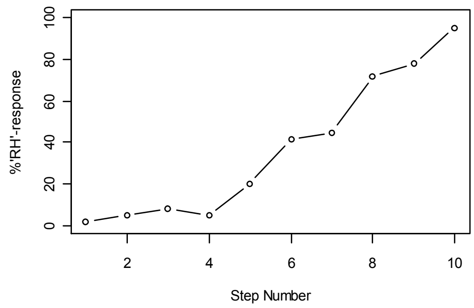

Figure 3. Result of identification task for the accentless group (Set 1)

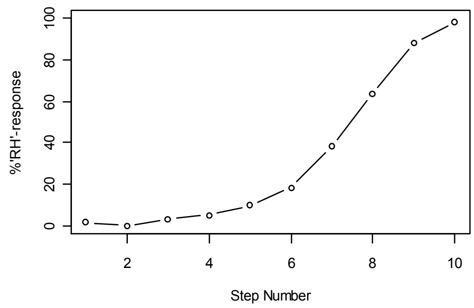

Figure 4. Result of identification task for the standard group (Set 1)

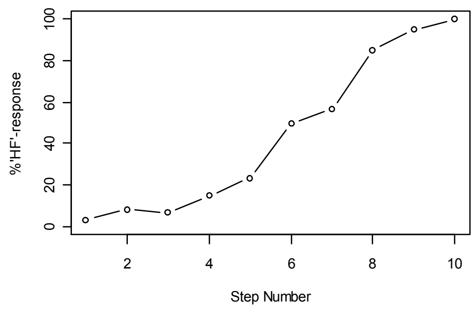

Figure 5. Result of identification task for the accentless group (Set 2).

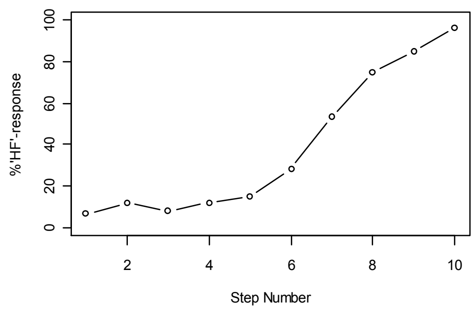

Figure 6. Result of identification task for the standard group (Set 2).

We fitted logistic curves shown in (1) to these results to estimate category boundaries, following previous literature on CP [22-24]. θ_c in this formula indicates the category boundary.

$$P(\theta) = \frac{1}{1 + e^{-k(\theta - \theta_c)}} \qquad (1)$$

Table 1 shows estimated values of the category boundaries (θ_c).

Table 1. Estimated category boundaries

	Accentless	Standard
Set 1	6.23	6.85
Set 2	6.90	7.39

Figures 7-10 show the results of the identification task. As can be seen, sharp discrimination peaks were not found. The rates of 'different'- responses increased as the step number increased. The maxima of the response curves did not agree with the estimated values of threshold in the identification task.

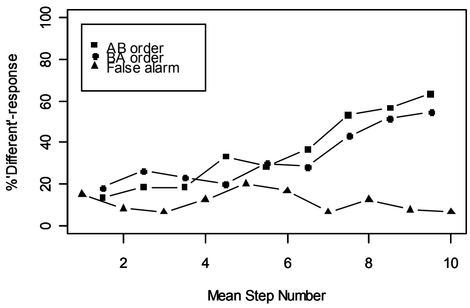

Figure 7. Result of discrimination task for the accentless group (Set 1)

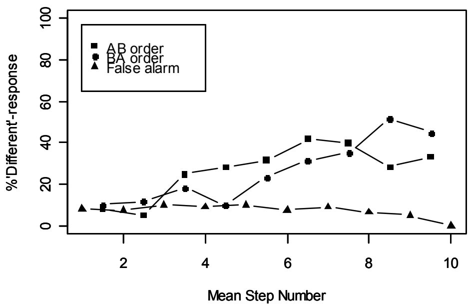

Figure 8. Result of discrimination task for the standard group (Set 1)

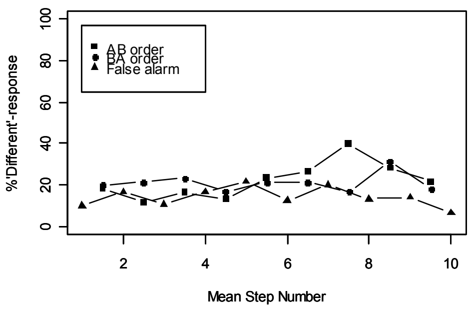

Figure 9. Result of discrimination task for the accentless group (Set 2)

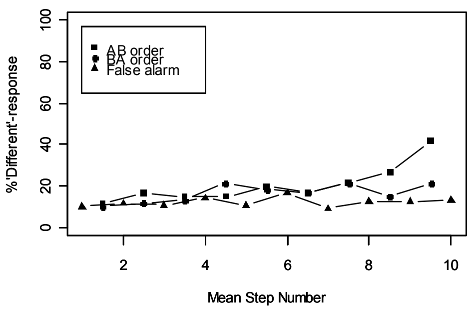

Figure 10. Result of discrimination task for the stnadard group (Set 2)

Discussion

As stated earlier, CP is characterized by an S-shaped pattern in an identification task and a sharp peak around a category boundary in a discrimination task. The results of the present experiment showed none of those features. However, this should not be interpreted as indicating that neither of the stimulus sets was phonological in Japanese. We rather consider that Set 1 was phonological for the standard group but did not lead to the typical CP results. It has often been reported that typical CP results are not obtained in some phonological contrasts (e.g. vowels [25] and lexical tones [26-27]).

Cross-dialectal differences were not found. There are three possible explanations for this result. The first possibility is that cross-dialectal tonal difference has no effect on pitch perception. The second possibility is that, although cross-dialectal tonal difference might affect pitch perception, participants from accentless regions perfectly acquired lexical pitch accent through longstanding exposure to Standard Japanese, and thus were not 'genuine' accentless speakers. The third possibility is that, although participants from accentless regions perceive pitch differently from Standard Japanese speakers, such a perceptual difference was not reflected in the experimental results.

The fact that no typical CP features were identified in the standard group's Set 1 suggests that the results might reflect perception at the acoustic level rather than at the phonological level, supporting the third possibility above. The next section reports another experiment that was conducted to test this suggestion.

It should also be discussed why the rates of 'different'- responses increased as the step number increased in the discrimination task. One possible explanation would be that participants heard stimuli in an f0 scale closer to a linear than to a semitone scale. As step intervals were set equal in semitones, the larger the step number was, the larger the interval was in the linear scale. If this explanation is true, it supports the view that the results of the discrimination task reflect perception at the acoustic level.

This explanation raises another question, though. If participants' perception was purely acoustic, similar results to Set 1 would be expected in Set 2. However, this was not the case. The accuracy rate in Set 2 was clearly lower than that in Set 1. One possible explanation for this is that our results reflect an interaction between acoustic and categorical perception. Alternatively, there might be an asymmetry in acoustic pitch perception. There were two differences between the sets: fall vs. rise and utterance-final vs. initial. Either or both of them might affect the results. Although this is a significant issue in pitch perception, a detailed examination of CP is beyond the scope of the present paper.

EXPERIMENT 2

We conducted the sequence recall task, which was developed in stress 'deafness' studies [10-11]. This task has a high memory load so that the acoustic level is not accessible and, therefore, phonological representations are highlighted. Therefore, this task was expected to show cross-dialectal perceptual differences, which had not been obtained in Experiment 1.

Methods

Subjects

Twenty-four subjects, who did not participate in Experiment 1, participated in the experiment. They were aged between 18 and 25, and were classified into two groups: 'accentless' and 'standard', as in Experiment 1. Each group consisted of twelve subjects (6 females and 6 males).

Stimuli

In Experiment 2, we used tokens of the two non-words /manu/ and /menu/ by six speakers. The tokens included both initial-mora-accented and unaccented pitch patterns. F0 values were measured for these tokens as in Experiment 1.

Among these tokens, one token of each speaker was chosen for the source sound for the f0 manipulation. 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 to manipulate f0. 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 11. Pitch contours were the same as the extreme stimuli in Experiment 1. The maximal and minimal f0 values were determined based on measured f0 values as in Experiment 1. Peak intensity was equalized for all stimuli.

High-fall

Rise-fall

Rise-high

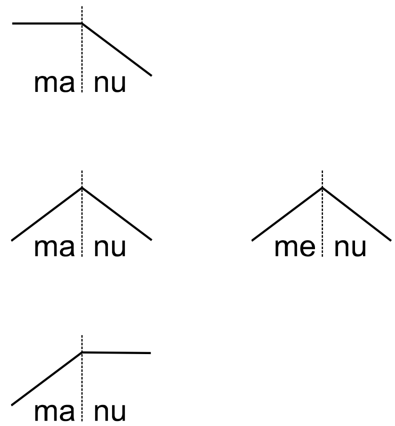

Figure 11. F0 shapes of re-synthesized stimuli

Three contrasts were created based on these stimuli, as shown in Table 2. 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, whose members have no lexical pitch contrast. The third contrast, i.e. segmental, was expected to be easy to discriminate for both groups.

Table 2. 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

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 (iii) in Table 2, 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.

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 +/- 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 [28].

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.

Results

Figures 12 and 13 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 ($p = 0.763$).

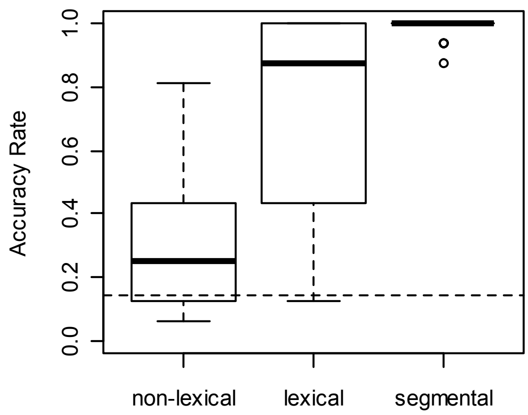

Figure 12. Accuracy rates for the accentless group. Non-lexical: non-lexical pitch contrast, lexical: lexical pitch contrast, segmental: segmental contrast. Horizontal dashed line indicates the chance level (0.14583).

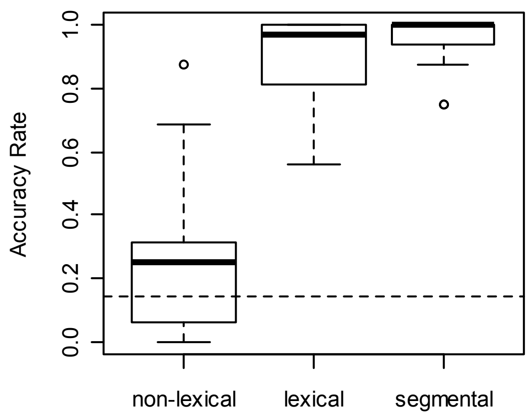

Figure 13. Accuracy rates for the standard group. Non-lexical: non-lexical pitch contrast, lexical: lexical pitch contrast, segmental: segmental contrast. Horizontal dashed line indicates the chance level (0.14583).

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 suggests that the participants from the accentless dialect region tend to have difficulties in discriminating contrasts of lexical pitch accent in Standard Japanese.

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. 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 12 and 13, 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 [10-11, 29]. 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.

GENERAL DISCUSSION

So far, we have looked at two experiments, CP and the sequence recall task. In CP, a between-group difference was not found. On the other hand, in the sequence recall task, scores of the accentless group were lower than those of the standard group in a contrast in which the location of lexical pitch accent differs. These results suggest that (i) speakers from accentless dialect regions tend to have difficulties in the perception of Standard Japanese lexical pitch accent and (ii) this perceptual difference was not reflected in the results of CP. Therefore, it is concluded that the sequence recall task is more robust to investigate dialectal differences in the perception of lexical pitch accent.

The results of the sequence recall task further suggest that there is a high variance in the perception of lexical pitch contrast for the accentless dialect speakers, probably due to the various degrees of exposure to Standard Japanese. That means that only a part of our participants in the accentless group had difficulties in the perception of lexical pitch accent.

One might wonder why the CP experiment did not show robust results. This is probably due to the well-known fact that phonological contrast does not always lead to typical CP results. As stated in the discussion of Experiment 1, vowels and lexical tones are known as such exceptional cases. The results of Experiment 1 suggest that this is also the case for lexical pitch accent. However, it is still unclear whether the results of Experiment 1 purely reflect acoustic perception, as discussed in the discussion of Experiment 1. Although this issue diverges from the aims of our project, it may be worth examining in detail to contribute to CP studies.

The next step of our project will be to investigate how the cross-dialectal perceptual differences found in the present study are reflected in brain activities.

ACKNOWLEDGEMENTS

Experiment 2 in this paper is based on a paper presented at Speech Prosody 2010, Chicago. We would like to thank for the invaluable comments on our presentation at the conference. We are also grateful to Yuko Sogabe, Mihoko Hasegawa, and Satoshi Imamura for helping with our experiment, Sho Tsuji and Kuniyoshi Tanaka for their invaluable comments on the manuscript, and Takayuki Kagomiya for his advices on data analysis. This work was supported by KAKENHI (No. 21820079 in 2009 and No. 22-4891 in 2010).

REFERENCES

1 C. Kiriloff, "On the auditory perception of tones in Mandarin," *Phonetica,* vol. 20, pp. 63-67, 1969.
2 Y. Wang*, et al.*, "Training American listeners to perceive Mandarin tones," *JASA,* vol. 106, pp. 3649-3658, 1999.
3 Y. Xu*, et al.*, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," *JASA,* vol. 120, p. 1063, 2006.
4 S. Hattori, "On'inron kara mita kokugo no akusento," *Kokugogaku Kenkyuu,* vol. 2, 1954.
5 J. Pierrehumbert and M. Beckman, *Japanese tone structure*. Cambridge, MA: MIT Press, 1988.
6 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, pp. 151-178.
7 Y. Yamaguchi, *Nihongo hoogen ikkei akusento no kenkyuu*. Tokyo: Hituzi, 1998.
8 Y. Igarashi, "Typology of intonational phrasing in Japanese dialects," in *Prosodic typology*. vol. 2, S.-A. Jun, Ed., New York: Oxford University Press, forthcoming.
9 Y. Mase, "Gengo keisei ni oyobosu terebi oyobi toshi no gengo no eikyoo," *Kokugogaku,* vol. 125, pp. 1-19, 1981.
10 E. Dupoux*, et al.*, "A robust method to study stress "deafness"," *JASA,* vol. 110, pp. 1606-1618, 2001.
11 E. Dupoux*, et al.*, "Persistent stress `deafness': The case of French learners of Spanish," *Cognition,* vol. 106, pp. 682-706, 2008.
12 F. Inoue, "Akusento no seisei to chikaku: Kantou ni okeru chiikisa to nenreisa," in *Kindaichi Haruhiko hakase koki kinen robunshuu.* vol. 2, Tokyo: Sanseido, 1984, pp. 119-153.
13 T. Otake and A. Cutler, "Perception of suprasegmental structure in a non-native dialect," *Journal of Phonetics,* vol. 27, pp. 229-253, 1999.
14 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, pp. 41-44.
15 R. Sato, "Mukei akusento chiiki ni okeru akusento no kyoutsuuka: Utsunomiyashi ni okeru shouchousa kara," in *Gendai hougengaku no kadai 2*, Tokyo: Meiji Shoin, 1984, pp. 211-234.
16 T. Onishi, "Hoogen akusento no genzai: Sendaishi hoogen ni okeru akusento no kakutoku wo chuusin ni," *Nihongogaku,* vol. 11, pp. 98-113, 1992.
17 A. Liberman*, et al.*, "The discrimination of speech sounds within and across phoneme boundaries," *Journal of Experimental Psychology,* vol. 54, pp. 358-368, 1957.
18 T. Hirayama, *Nihongo onchoo no kenkyuu*. Tokyo: Meiji Shoin, 1957.
19 J. V. Neustupńy, "Nihongo no akusento wa kootei akusento ka*,* vol. 121, pp. 1-7, 1966.
20 M. Sugito, *Nihongo akusento no kenkyuu*. Tokyo: Sanseido, 1982.
21 T. Ishihara, "Tonal alignment in Tokyo Japanese," PhD, The University of Edinburgh, 2006.
22 B. McMurray and M. Spivey, "The categorical perception of consonants: the interaction of learning and processing," in *Chicago Linguistic Society*, 1999, pp. 205-219.
23 H. Kato*, et al.*, "Bogowasha to hibogowasha ni yoru nihongo tokushuhaku onso no chikaku," *Technical Report of IEICE,* vol. SP2004-36, pp. 43-48, 2004.
24 T. Kagomiya and S. Nakagawa, "Kotsudo choonpa chokaku ni yoru sokuon choon goiakusento no chikaku," in *2008 Autumn Meeting of the Acoustical Society of Japan*, 2008, pp. 459-460.
25 D. Fry*, et al.*, "The identification and discrimination of synthetic vowels," *Language and Speech,* vol. 5, pp. 1-189, 1962.
26 A. Abramson, "The noncategorical perception of tone categories in Thai," *Frontiers of speech communication research,* pp. 127-134, 1979.
27 A. Francis*, et al.*, "On the (non) categorical perception of lexical tones," *Perception & Psychophysics,* vol. 65, p. 1029, 2003.
28 M. Van Selst and P. Jolicoeur, "A Solution to the Effect of Sample Size on Outlier Elimination," *The Quarterly Journal of Experimental Psychology,* vol. 47A, pp. 631-650, 1994.
29 E. Dupoux*, et al.*, "A Destressing "Deafness" in French?," *Journal of Memory and Language,* vol. 36, pp. 406-421, 1997.