interspeech 2008 incorporating SST 08

The Role of Japanese Pitch Accent in Spoken-Word Recognition: Evidence from Middle-aged Accentless Dialect Listeners

Takashi Otake [1] , *Marii Higuchi* [2]

[1]E-Listening Laboratory, Basic Research and Development, Tokorozawa, Japan
[2]Graduate School of Foreign Languages, Dokkyo University, Soka, Japan

otake@e-listeninglab.com, g4678009@dokkyo.ac.jp

Abstract

This paper investigates the role of pitch accent information in spoken-word recognition in listeners in Fukushima, Japan, whose dialect is accentless. Previous research revealed that accentless listeners were less sensitive to pitch accent than Tokyo Japanese listeners. The present study asked whether middle-aged listeners' use of accent information would differ from that of young listeners. 40 middle-aged Fukushima listeners were presented with Tokyo Japanese materials used in the earlier study, employing a gating task. Results show that middle-aged Fukushima accentless listeners are even less sensitive to the pitch accent information than the younger accentless listeners. The findings suggest that the exploitation of pitch information by younger listeners reflects adaptation to the standard Tokyo Japanese dialect, e.g., via the media, whereas the older listeners are less influenced by the standard.

10.21437/Interspeech.2008-337

1. Introduction

Studies in spoken-word recognition have suggested that human listeners exploit various sources of phonetic information in order to identify target words, as soon as each type of information becomes available [1,2]. The use of word-initial phonetic information is particularly important for spoken-word recognition because it results in a drastic reduction in the set of possible word candidates. If a listener perceives a voiceless bilabial stop [p], for example, words starting with other speech sounds are excluded from the set of candidates for recognition. In this respect it can be said that the principal function of word-initial phonetic information is to single out possible word candidates [2]. The mechanism underlying human spoken-word recognition involves multiple, concurrent availability of potential candidates, together with rapid and continuous updating of the candidate set in response to the changing information in the input. It is the rapidity of the integration of phonetic information on which the speed an effortlessness of human word recognition rests.

Phonetic information can be suprasegmental as well as segmental. Thus, it is reasonable to ask whether word-initial suprasegmental information has an equivalent function to discriminate word candidates [3]. This issue has been investigated to date in several languages including English, Dutch, Chinese and Japanese, each of which uses different type of suprasegmental cues (e.g., stress for English and Dutch, tone for Chinese and pitch for Japanese) [4,5,6,8]. Tokyo Japanese is an ideal language to test this problem because the pitch accent system begins with either a high pitch or a low pitch at the word-initial position [7]. Inter-word distinctions in the initial portions of words could therefore be of great use to listeners, enabling drastic reductions in the set of possible candidate words for recognition just as segmental information does.

A recent study indeed has revealed that Tokyo Japanese listeners make early and effective use of Japanese pitch accent information in recognizing spoken words [8]. One of the experimental tasks employed in the study was a gating task, where listeners were presented with initial fragments of words and asked to guess the words. The results showed that the listeners' guesses overwhelmingly had the same initial accent structure (HL- or LH-) as the presented word even when all that the listener had heard of the word was half of the first syllable, indicating that the word-initial pitch accent is indeed involved in the selection of word candidates. Thus, Tokyo Japanese listeners seem to effectively exploit the cues to accent pattern available even in the first syllable of a word. In other words, the suprasegmental structure of Tokyo Japanese words is, like the segmental structure, of use in narrowing down the set of potential candidates for lexical recognition.

Tokyo Japanese is the dominant dialect and is regarded as the standard language in Japan. If only this variety were spoken by Japanese, the findings might suggest that the function of word-initial suprasegmental information is a necessary part of Japanese spoken-word recognition. However, the reality is far more diversified. Many varieties of Japanese dialects differ from Tokyo Japanese in respect of accent structure [9]. Among the Japanese dialects one type of Japanese dialects is particularly interesting with respect to the problem at hand, because the words in the dialects are characterized by flat pitch contours [10]. It is reported that speakers of the dialects experience a hard time distinguishing the word accents in the standard variety. As a result, *hashi* (HL) 'chopsticks' and *hashi* (LH) 'bridge' can be homophonous for these dialect speakers. These dialects are called "accentless dialects," and are spoken in two principal areas of Japan: a large area spreading from the northern Kanto (area) (Tochigi/Ibaraki prefectures) to the southern part of the Tohku area (Fukushima/Miyagi prefectures) on Honshu Island, and the central area (Kumamoto/Miyazaki prefectures) of Kyushu Island [9]. Listeners of such dialects have no reason to attend to accent information in their home variety; so that when listening to the standard variety they may not benefit from suprasegmental information to the extent possible for Tokyo Japanese listeners.

Recent studies have investigated accentless dialect listeners' use of suprasegmental information, employing the same experimental tasks as in [8]. Interestingly, the results revealed that listeners from all the dialect groups tested (Tochigi/Ibaraki, Kumamoto and Fukushima) showed sensitivity to the pitch information, though these listeners were less sensitive than Tokyo listeners had been [11, 12]. One conclusion drawn from these studies was that the pitch accent information in accentless listeners' mental lexicon may

Accepted after peer review of full paper

not be clearly represented; due to the constant exposure to Tokyo Japanese which is available to all residents of Japan via TV or radio media, they may have been well adapted to perceiving it, but still did not use it as efficiently as Tokyo listeners do.

This explanation then raises another interesting question. All accentless dialect listeners are exposed to Tokyo Japanese, but is there an age difference affecting the exploitation of word-initial pitch information to single out word candidates? Suppose, for example, that accentless listeners who have been living in the same linguistic community for more than 40 years are presented with Tokyo Japanese materials. We can propose two possible outcomes. One is that the longer the listener's experience with a non-native dialect, the better the processing becomes. On this hypothesis, older listeners could show more efficient use of accent information than the listeners in [11, 12] did. The other possibility is that the extent of exposure to the media will be much greater for the younger listeners that it was for the older listeners at the equivalent time of their life. On this hypothesis, we might expect the older accentless listeners to be even less sensitive to accent information than the young listeners in the earlier study. Note that in a recent study of L2 segmental distinctions such as /r/-/l/, younger Japanese listeners' performance was always better than that of older listeners [13].

The present study thus extended this series of experiments on accent to middle-aged accentless Fukushima listeners. Since young listeners' data were already reported in the previous study [12], we were able to make a direct comparison between the middle-aged and the young accentless listeners, both from the same linguistic community (the experiments of both subject groups were carried out in 2003 and 2004, respectively).

2. Experiment

2.1. Method

2.1.1. Materials

The materials were those used by Cutler and Otake (1999; Experiment 2)[8]. They consisted of 24 pairs of Japanese words, with each pair having a common initial bimoraic CVCV sequence, but differing in segmental structure from the fifth segment on. The accent pattern of the two words in each pair differed, in one word the initial CVCV sequence was HL, in the other LH. Thus, *nagasa* and *nagashi* formed a pair; both begin *naga-*; the accent pattern of *nagasa* is HLL, while *nagashi* is LHH. Twenty-two pairs had three morae, the remaining two pairs four morae, and no words contained moraic nasals, geminate consonants or long vowels. The complete set of pairs (in HL-/LH-order) is shown in Table 1.

Table 1. Stimuli used in the experiment

bakuhu/bakuchi hanabi/hanawa hokubu.hokuro kamotsu/kamome karasu/karada karahuru/.karamatsu karuteru/karudera, kasegi/kasetsu, kokugi/kokuge, maguchi/maguro moguri/mogura mokuba/mokuji, nagasa/nagashi namida/namiki, nimotsu/nimono nomichi/nomiya, sashizu/sashiki, sekiri/sekiyu tachiba.tachiki tomato/tomari wakaba/wakate warabi/waraji yomichi/yomise

Twenty-four further words served as practice and warm up items. Some of these contained moraic nasals, geminate consonants, or long vowels. Twelve were three-mora words and 12 were four-mora words.

All experimental and filler words, each preceded by a short carrier phrase *Sore wa …* ("It is…") appeared in gated form on a tape made from an original recording by a male native speaker of Tokyo Japanese. In the gated version, each word, always preceded by the carrier phrase, was presented in increasingly large fragments, incrementing in each case by a portion of the word up to the midpoint of the following phoneme. This was achieved by determining as near as possible on the basis of visual and auditory information, the boundaries of each phonetic segment, following which a marker was placed at the midpoint of each such demarcated region. The first fragment then presented the carrier plus the target word up to the first marker, i.e., presented no more than part of the word's initial consonant; the next fragment added further material up to the second marker, i.e., the midpoint of the first vowel, and so on. The cuts producing the fragments were made in such a way as to avoid abrupt amplitude changes which might lead to the perception of illusory clicks. The gating procedure, and acoustic analyses of the materials, are described in greater details in Cutler and Otake (1999)[8].

There were two experimental tapes, each containing all filler words and one member of each experimental pair. Accent pattern was counterbalanced across tapes; each tape had 12 HL- and 12 LH- experimental words, and the members of any pair occurred at the same position on each tape.

2.1.2. Subjects

Forty middle-aged volunteers (average age: 47.8 years; 27 females and 13 males) in Hobara town in Fukushima, took part in the experiment. All had grown up in Fukushima prefecture and had received all their schooling in their home town area. They received a small payment for participating. They were tested as a group at a town hall in Hobara town. A native speaker of Fukushima Japanese, the second author of this paper, conducted the experiment using this dialect.

2.1.3. Procedure

There were two subject groups, of 20 listeners each, and each group heard a different one of the two tapes. The subjects were tested in groups of up to six, and heard the tape over headphones from a DAT recorder. Presentation was halted after each fragment to allow time for the subjects to record a guess as to the word's identity. The guesses were written on a response sheet in their normal Japanese orthographic form.

2.2. Results

2.2.1 Analysis of middle-aged accentless dialect listeners

The earlier studies concentrated their analyses on the candidate word guessed at fragments 1-4, and in particular on the comparison of the accent pattern of these candidate words for target words beginning HL- and LH- [8]. From fragment 5 onwards, segmental information could distinguish between the members of the word pairs; for the first four fragments, however, the only distinguishing information was accentual. The word guesses for the first four fragments thus most directly address the question of interest: could listeners make use of the accentual cues available in the initial bimoraic portion of each stimulus pair (such as *naga- in nagasa* HLL and *nagashi* LHH), a portion which was segmentally matched

but accentually different? In the present analyses we examine these initial guesses.

The word guesses were scored by hand and the accent pattern of each guess in Tokyo Japanese determined. As in the study of Cutler and Otake (1999) [8], only the first two morae of the guessed words were taken into consideration, which effectively resulted in a two-way classification of accent patterns into HL- (Type 1 accent) vs. LH- (Types 0, 2, 3, etc.).

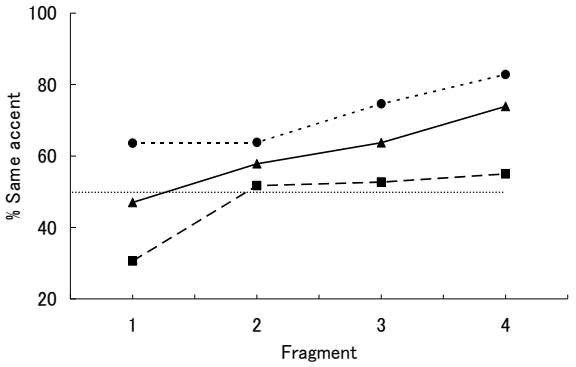

Fig. 1. Middle-aged Fukushima accentless listeners: proportion of guessed words with the same initial accent pattern (solid line with triangle, Average; dotted line with square, HL- words; dotted line with circle, LH- words).

Fig. 1 shows the proportion of guesses which had the same initial accent pattern as the spoken word, for each of the first four fragments, separately for HL- and LH- words and averaged across these. Percent correct for each fragment was subjected to a sign test of statistical significance of the difference from chance, and analyses of variances were carried out on the accuracy scores across word types.

The guesses provided for fragment 1 matched the initial accent pattern of the target word in 47.1% of cases, which is not significantly different from chance. The guesses provided at fragment 2 matched the initial accent pattern of the target word in 57.8% of cases, which is significantly above chance performance ($z=3.57$, $p<0.001$), and again accuracy continued to rise across fragments 3 and 4, each of which is significantly above chance performance. As can be seen from Fig. 1, accuracy was always greater for LH- words than for HL- words (subject analysis only) ($F1\ [1,78]=9.216$, $p<0.003$; $F2\ [1,46=2.37$, $p=0.133$), and this difference was again separately significant at every fragment. The overall tendency was very similar to that in previous studies. However, there were two points which differed considerably from the previous studies. First, the whole accuracy pattern for LH-, HL- and average was much lower, and closer to chance level. Second, the accuracy pattern for the HL- words stayed almost at chance level (statistically not different from chance for any position). This pattern was not observed either in Tokyo Japanese or young accentless dialect listeners [8,11,12].

2.2.2 Joint analysis with young accentless dialect listeners

In order to make a direct comparison between middle-aged Fukushima listeners with Fukushima young listeners, a joint analysis of the accuracy scores was undertaken. Fig. 2 shows the mean accuracy scores for Fukushima high school listeners, adapted from the previous study [12]. As can be seen from this figure, the mean scores at the four positions for young listeners were always higher than that of the middle-aged listeners. The joint analysis revealed that there was significant difference between the two groups except at the first position (fragment 1: $F1[1,158] = 0.003$, $p=0.95$, $F2[1,94] = 0.134$, $p=0.715$; fragment 2: $F1[1,158] = 27.9$, $p<0.001$, $F2[1,94]=44.709$, $p<0.001$; fragment 3: $F1\ [1,158] = 10.58$, $p<0.001$, $F2[1,94]=1-.373$, $p<0.002$; fragment 4: $F1\ [1,158] = 8.225$, $p<0.02$, $F2[1,94]=2.531$, $p=0.115$).

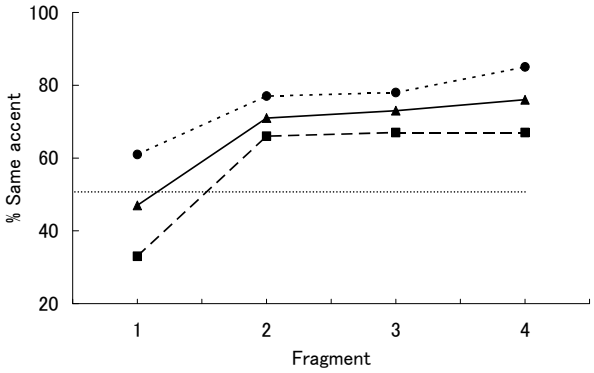

Fig. 2. Young Fukushima accentless listeners: proportion of guessed words with the same initial accent pattern (solid line with triangle, Average; dotted line with square, HL- words; dotted line with circle, LH- words).

.

2.2.2.1 Analysis of HL-words

The joint analysis of the accuracy scores for HL-words between middle-aged listeners and the young listeners show a significant difference between the two groups ($F\ [1,75] = 89.11$, $p<0.001$). Separate comparisons between the subject groups across positions revealed that Fukushima high school listeners had higher accuracy than Fukushima middle-aged listeners at three positions (fragment 2: $F1\ [1,74] = 13.16$, $p<0.001$, $F2[1,46]=8.328$, $p<0.006$; fragment 3: $F1\ [1,74] = 17.34$, $p<0.001$, $F2[1,46]=6.398$, $p<0.05$; fragment 4: $F1\ [1,74] = 8.62$, $p<0.01$, $F2[1,46]=2.136$, $p=0.151$). These results suggest the middle-aged listeners were less sensitive to Tokyo Japanese pitch accent in these words than the young listeners. The important finding for HL- words was the lack of significance at any position for the older listeners.

2.2.2.2 Analysis of LH-words

The joint analysis of the accuracy scores between the young listeners and the middle-aged listeners show a significant difference between the two groups ($F\ [1, 75] = 141.83$, $p<0.001$). Separate comparisons between the subject groups again revealed that the young listeners had higher accuracy than the middle-aged listeners at all of fragments 2-4 (fragment 2: $F1[1,74] = 42.32$, $p<0.001$, $F2[1,46]=12.032$, $p<0.001$; fragment 3: $F1[1,74] = 71.27$, $p<0.001$, $F2[1,46]=0.706$, $p=0.405$; fragment 4: $F1[1,74] = 68.69$, $p<0.001$, $F2[1,46]=0.984$, $p=0.326$]. These results suggest that both subject groups were sensitive to the pitch accent pattern, although the middle-aged listeners were less sensitive to Tokyo Japanese pitch accent than the young listeners.

3. General Discussion

This study has investigated whether middle-aged listeners whose native Fukushima dialect is accentless can use word-

initial pitch accent information in Tokyo Japanese to reduce possible word candidates. We were able to make a direct comparison with these listeners' younger counterparts. The analysis has revealed that the middle-aged Fukushima accentless listeners showed far less sensitivity to the word-initial pitch accent information of Tokyo Japanese than the young listeners. Thus, these results clearly suggest that the middle-aged Fukushima accentless listeners were insensitive to the Tokyo pitch accent. The results raise several interesting implications.

Most importantly, the greatest difference was found in the HL- word subset. This is clear evidence that the present listener group was guessing. In the Japanese vocabulary, LH-words greatly outnumber HL- words (60% for LH-words) [8]. Accordingly, a guessing strategy in the gating task will result, in the present binary analysis, in an inflated score for LH-words and a reduced score for HL- words. This is exactly the pattern we observed [8].

To date, three groups of young accentless listeners (Tochigi/Ibaraki, Kumamoto and Fukushima) had been tested; they showed a very similar response pattern. The common observation was that although their performance level did not reach the level of Tokyo Japanese listeners, it was quite high. We now see that the present older listener group perform significantly worse than the previously tested groups.

This may suggest that pitch accent information in non-native dialects may be more difficult to deal with for the older than the younger listeners. However, since these listeners were not yet in the age range confronted with age-related hearing loss, it is more likely that their exposure to the standard dialect was, as we suggested above, less effective at the time at which their speech perception skills were not yet fixed. The younger listeners have simply had much more exposure to the media than their elders had at the same age.

In this study our attention was mainly focused on how accentless dialect listeners perceive the word-initial Tokyo Japanese pitch accent with respect to spoken-word recognition. Our assumption here is that sensitivity to the pitch information should help access words. The more one is sensitive to the Tokyo Japanese pitch accent information, the better words spoken in that variety will be discriminated.

However, assuming that all listeners whose dialects have clear pitch accent information store pitch patterns in their mental lexicon, the situation is in fact more complicated. For example, it is often said that the Kansai dialect, which used to be the dominant dialect in Japan, has accentual patterns nearly opposite to those of Tokyo Japanese. Thus, if a word like *hashi,* described earlier, is spoken in Tokyo Japanese but heard by a Kansai dialect listener, he or she may interpret the opposite meaning [14]. If the same word is spoken by a Kansai dialect speaker, Tokyo Japanese listeners in turn will experience the reverse word activation. Thus, the use of word-initial pitch accent information is helpful only among listeners who share the same accentual system. Otherwise, it may even cause serious confusion among some dialect listeners. Obviously, more investigation is needed of this factor too.

4. Conclusions

This paper examined the role of Japanese pitch accent information in spoken-word recognition by middle-aged Fukushima accentless listeners. The results have shown that middle-aged accentless listeners were considerably less sensitive to word-initial pitch accent information than the young accentless listeners. This finding suggests that access of the target words by the middle-aged listeners disregards potentially useful information, and hence when listening to the standard dialect their processing may be less effective than that of younger listeners.

5. Acknowledgements

This research was supported by a grant from the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (C) 11610566). Thanks to all participants in Hobara town in Fukushima prefecture and to Anne Cutler and anonymous reviewers for valuable comments and suggestions.

6. References

[1] Martin, J. G. and Bunnell, H. T., 1981. Perception of anticipatory coarticulation effects. *J. Acoust. Soc. Am*. 69, 559-567.

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

[3] Cutler, A., Dahan, D., and van Donselaar, W., 1997. Prosody in the comprehension of spoken language: A literature review. *Lang. and Speech* 40-2, 141-201.

[4] Cooper, N., Cutler, A., and Wales, R., 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Lang. and Speech* 45, 207-228.

[5] Donselaar, W. van, Koster, M., Cutler, A., 2005. Exploring the role of lexical stress in lexical recognition. *Quart. J. Exp. Psychol.* 58, 251-273.

[6] Cutler, A. and Chen, H.-C. 1997., Lexical tone in Cantonese spoken-word processing. *Perc. Psychophys.* 59, 165-179.

[7] Haraguchi, S., 1977. *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusha.

[8] Cutler, A. and Otake, T., 1999. Pitch accent in spoken-word recognition in Japanese. *J. Acoust. Soc. Am*. 105, 1877-1888.

[9] Shibatani, M., 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.

[10] Maekawa, K., 1997. Akusento to intoneishon: Akusento no nai chiiki (Accent and intonation in accentless dialects). In *Nihonogo Onsei 1: Shohougen no akusento to intoneishon (Japanese speech 1: Accent and intonation in various dialects)*, 97-122. Tokyo: Sanseido.

[11] Otake, T. and Cutler, A., 1999. Perception of suprasegmental structure in non-native dialect. *J. Phon.* 27, 229-253.

[12] Otake, T. and Higuchi, M., 2004. A role of pitch accent in spoken-word recognition in accentless Japanese dialects: Evidence from Fukushima listeners. *Proc. Speech Prosody 2004*, CD-ROM, 41-44.

[13] Aoyama, K., Flege, J.F., Guion, S.G., Akahane-Yamada, R., and Yamada, T. 2003. Foreign accent in English words produced by Japanese children and adults. *Proc. 15th ICPhS.* 3201-3234.

[14] Warner, N., 1997. Recognition of accent patterns across dialects in Japanese. *Proc. 23rd Ann. Meeting. Berkeley Linguistic Society*, 14-17.