



Perception of suprasegmental structure in a non-native dialect

Takashi Otake

*Faculty of Foreign Languages, Dokkyo University, 1-1 Gakuen-Cho,
Soka-Shi, Saitama 340-0042, Japan*

and

Anne Cutler

*Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen,
The Netherlands*

Received 19th April 1999 and accepted 27th September 1999

Two experiments examined the processing of Tokyo Japanese pitch-accent distinctions by native speakers of Japanese from two accentless-variety areas. In both experiments, listeners were presented with Tokyo Japanese speech materials used in an earlier study with Tokyo Japanese listeners, who clearly exploited the pitch-accent information in spoken-word recognition. In the first experiment, listeners judged from which of two words, differing in accentual structure, isolated syllables had been extracted. Both new groups were, overall, as successful at this task as Tokyo Japanese speakers had been, but their response patterns differed from those of the Tokyo Japanese, for instance in that a bias towards H judgments in the Tokyo Japanese responses was weakened in the present groups' responses. In a second experiment, listeners heard word fragments and guessed what the words were; in this task, the speakers from accentless areas again performed significantly above chance, but their responses showed less sensitivity to the information in the input, and greater bias towards vocabulary distribution frequencies, than had been observed with the Tokyo Japanese listeners. The results suggest that experience with a local accentless dialect affects the processing of accent for word recognition in Tokyo Japanese, even for listeners with extensive exposure to Tokyo Japanese. © 1999 Academic Press

1. Introduction

Listening to spoken language is tailored by one's native listening experience. Thus, phonetic contrasts which do not appear in the native language are often very difficult to perceive in a second language; cases of such language-specific listening are legion, and have been amply documented in the speech perception literature. Moreover, extensive exposure to a second language, extending even over years, is not in itself sufficient to

Correspondence to either author. E-mail addresses: otake@dokkyo.ac.jp; anne.cutler@mpi.nl.

induce sensitivity to non-native contrasts. For instance, recent studies have shown that speakers who are functionally bilingual may not make complete perceptual distinctions in one of their languages — thus Spanish-native speakers of Catalan in Barcelona, despite being exposed to Catalan from early childhood, fail to make distinctions between certain Catalan sounds which are subsumed by a single Spanish phonetic category (Pallier, Bosch & Sebastián-Gallés, 1997; Sebastián-Gallés & Soto, 1999).

Dialects of a single language may of course also differ in what phonetic contrasts they distinguish, and accordingly they can also differ in how they affect listeners' perceptual sensitivity to phonemic contrasts. Thus, speakers of standard French show little sign of using temporal information in making vowel distinctions (Gottfried & Beddor, 1988), because standard French does not use temporal cues to distinguish vowels. Swiss French does, however, and speakers of Swiss French can make use of temporal information when distinguishing vowels (Miller & Grosjean, 1997). Users of one dialectal variety of a language can, accordingly, fail to distinguish contrasts made by speakers of another variety. Labov, Karen & Miller (1991), for instance, report that users of one dialect of American English fail to discriminate vowel contrasts made in another American English variety when asked to choose between alternative words, even though separate tests show that they can perform an accurate psycho-acoustic discrimination. Similarly, speakers of American English can use the form of an intervocalic consonant as a perceptual cue to the presence or absence of a syntactic phrase boundary between two words; the final sound of *visit* is more likely to be expressed as a flap rather than as a [t] in "Try to visit India in the spring" than in "For a visit, India is great". Speakers of British English can hear the difference, but they themselves do not normally make such a distinction in production, and they do not use it in perception of American English speech either (Scott & Cutler, 1984). Neither of these results reflects simple degree of exposure, since British English speakers are routinely exposed to large amounts of American English via the broadcast media, and the Philadelphia speakers of the Labov et al. study likewise are exposed to large amounts of speech from other American varieties in which the vowel contrasts in question are distinctive. In other words, dialect forms of a single language, despite almost perfect mutual intelligibility, also show the resistance to the use of non-native contrasts, irrespective of amount of exposure, which is typical of second-language perception.

Language-specific phonetic structure can embody, besides distinctions between phonetic segments, contrasts in suprasegmental structure. Non-native suprasegmental contrasts, like non-native segmental contrasts, can present listeners with perceptual difficulty. Thus, stress contrasts in nonsense words are difficult for French speakers to perceive, but easy for Spanish speakers (Dupoux, Pallier, Sebastián-Gallés & Mehler, 1997); Spanish is a stress language, French is not. Perception of tone in Chinese is influenced for English speakers by their knowledge of pitch patterns in English (Broselow, Hurtig & Ringen, 1987). Pitch accent distinctions in Japanese can be acquired by learners who are native speakers of English or French, but their perceptual performance is not equivalent to that of native Japanese speakers, and again the performance reflects characteristics of the native system (Nishinuma, 1994; Nishinuma, Arai & Ayusawa, 1996).

Just as systematic training can enable listeners to acquire certain difficult perceptual distinctions in segmental structure (Logan, Lively & Pisoni, 1991; Lively, Pisoni, Yamada, Tohkura & Yamada, 1994), so can explicit training in perception of suprasegmental distinctions be extremely effective in improving listeners' sensitivity (Wang,

Spence, Jongman & Sereno, in press). Of course, few listeners in fact undertake such training. Moreover, we do not know of any study which has assessed training in perception of another dialect of listeners' native language. Most listening to non-native languages and non-native dialects remains influenced by constraints of native-language phonology.

In the present study, we deal with a suprasegmental difference across dialects of Japanese. Words in Standard Tokyo Japanese manifest pitch accent patterns (Sugito, 1982; Pierrehumbert & Beckman, 1988; Kubozono, 1993), and these can distinguish minimal pairs of words; thus *ame* HL means "rain", *ame* LH "candy". Tokyo Japanese words can be unaccented or accented; in unaccented words, the first mora is labelled low (L) and all subsequent morae are labelled high (H), whereas in accented words, one specific mora of the word is marked for accent and is labelled high (H). If this marked mora is the first in the word, all subsequent morae will be labelled low (L) — a two-mora word would thus have the pattern HL, a three-mora word HLL, and so on. If the marked mora is the second or a later mora in the word, the first mora will be L, all other morae between the first and the marked mora will be H, and morae after the marked mora will be L; that is, the distinguishing characteristic of a pitch accent is a fall from H, to L on anything following it. Unaccented words (LH, LHH, LHHH, etc.) are referred to as Type 0; Type 1 words (HL, HLL, HLLL, etc.) have accent on the first mora, Type 2 (LH[^], LHL, LHLL, etc.) on the second, and so on. (The carat [^] distinguishes final-accented from otherwise identical unaccented strings; thus LHHH is Type 0, unaccented, while LHHH[^] is Type 4, accented.)

Although there are further complexities in the pitch-accent system of Tokyo Japanese, there are in general only two possible ways to label the first two morae of the citation form of a word: HL- or LH-. That is to say, important pitch-accent distinctions are expressed in the initial portions of Tokyo Japanese words. Research in human spoken-word recognition has shown that listeners exploit relevant phonetic information as soon as it becomes available; any 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.

Indeed, it is clear that Tokyo Japanese listeners use pitch-accent information very efficiently and as early as it is available to them. Minematsu & Hirose (1995) found that misaccented words of Tokyo Japanese, presented in isolation, were harder to recognize than the same words with correct accent. These authors further found that Type 1 four-mora words (HLLL), presented incrementally in successively larger fragments, could be recognized on the basis of less information than Type 0 (LHHH) or Type 2 (LHLL) words. As the four-mora vocabulary of Tokyo Japanese contains less than 10% Type 1 words, this result suggests that listeners were effectively using accent to narrow down the set of potential candidate words. Cutler & Otake (1999) conducted three experiments on the use of pitch-accent information in spoken-word recognition by Tokyo Japanese listeners. In the first, they found that listeners could quite accurately judge whether an isolated syllable had been extracted from a Type 1 *vs.* a Type 0 or Type 2 word (e.g., *ka* from *baka* HL *vs.* *gaka* LH, or *naka* HL *vs.* *haka* LH[^]). Thus, acoustic cues to pitch-accent distinctions are available within a single syllable. In their second experiment, Cutler and Otake presented listeners with initial fragments of words and asked them to guess the words; 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 the first consonant and a portion of the vowel that followed it (e.g., *na-* from

nagasa HLL or *nagashi* LHH). Thus, listeners could effectively exploit the cues to accent pattern available even in the first syllable of a word. In a third experiment, Cutler and Otake found that lexical decision responses to words from a minimal pitch-accent pair (such as *ame* HL “rain” and *ame* LH “candy”) were not facilitated by prior presentation of the accentually different word; response time to decide that *ame* HL is a real word was faster if *ame* HL had been heard earlier in the experiment, but not if *ame* LH had been heard. Thus, pitch-accent information rules out activation of words which are otherwise segmentally identical. Sekiguchi & Nakajima (1999) conducted a priming study in which they presented listeners with word fragments which were segmentally ambiguous but suprasegmentally unambiguous (an example would be *naga-* from either *nagasa* HLL or *nagashi* LHH); they found that such fragments facilitated lexical decision responses to the words which they matched suprasegmentally but not to suprasegmentally mismatching words.

Together these results provide strong evidence that Tokyo Japanese listeners make early and effective use of pitch-accent information in recognizing spoken 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.

However, many varieties of Japanese differ from Tokyo Japanese in respect of accentual structure, and indeed not all varieties make use of accent to convey distinctions between words at all. The so-called “accentless” dialects are spoken in two principal areas of Japan: a large area spreading from the northern Kanto area (Tochigi and Ibaraki Prefectures) to the southern part of the Tohoku area (Fukushima and Miyagi Prefectures) on Honshu Island, and the central area (Kumamoto and Miyazaki Prefectures) of Kyushu Island. Fig. 1 is taken from Shibatani (1990) and indicates the location of these two main areas. Table I is also adapted from Shibatani (1990) and compares the accent patterns of some common words in Tokyo Japanese, representing between them the three possible accent patterns for bisyllabic words, with the same words in the varieties spoken in Mito (a city in Ibaraki) and Kumamoto (a city on Kyushu).

As Table I shows, the words in the accentless dialects are characterized by flat pitch contours, a pattern which is very different from that of the accenting variety described above. The accentless dialects of course make use of pitch contours to express other linguistic distinctions, for instance at the sentence level (Maekawa, 1997), but the point at issue here is that they do not make lexical distinctions with pitch accent. Thus for users of such dialects there would be little point in exploiting the suprasegmental features of the initial portions of spoken words in word recognition, in the way Tokyo Japanese listeners do: word recognition in accentless Japanese could not benefit from attention to suprasegmental information.

However, users of accentless Japanese are not solely surrounded by their local dialect. Precise estimates of actual dialect use in speech production are difficult to establish; there are many dialectological studies in Japanese, but the focus of these has often been on the degree to which speakers from the accentless areas can produce Tokyo Japanese pitch-accent contrasts if asked to do so (see, e.g., Inoue, 1984; Sato, 1984; Sugito, 1988). The methodology of such studies usually involves the reading aloud of word lists or short phrases, and the results suggest that apparently accurate production is common. Certainly young speakers, and university students in particular, should be able to command standard Japanese in speech production. As Jinnai (1996) points out, however, the dialectological studies which have demonstrated that deliberate production of accentual

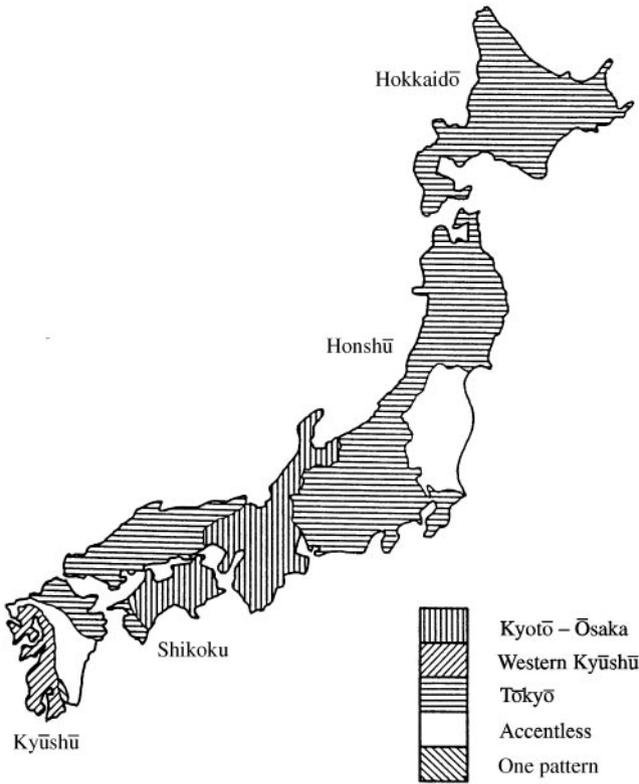


Figure 1. Map of Japan, showing the location of accented and accentless-dialect areas; from Shibatani (1990: p 211). Used with permission. ▮▮▮, Kyoto-Ōsaka; ▮▮▮, Western Kyūshū; ▮▮▮, Tōkyō; □, accentless; ▮▮▮, one pattern.

contrasts is possible do not motivate the conclusion that the local dialect has been abandoned. In his own study of dialect on Kyushu Island, Jinnai successfully elicited both dialect and standard speech from middle-aged speakers, high-school students and elementary school pupils. The factor which motivated the choice of style was the formality of the setting: all groups spoke dialect in informal settings and standard Japanese in formal situations (such as reading aloud). Thus, it is reasonable to assume that productive use of dialect is quite common (indeed, it has been reported that associated differences can even be observed in the production of foreign language by speakers of accenting *vs.* accentless varieties; Muramatsu, 1998).

The focus of the present article, however, is on perception, and here the situation is much clearer. All speakers of Japanese, including users of the accentless varieties, will be exposed daily to Standard Tokyo Japanese via the broadcast media. They will meet many speakers of Standard Tokyo Japanese in their everyday life. Thus, even speakers who never themselves make pitch-accent distinctions, *i.e.*, who use solely dialect forms in production, would have ample opportunity to practise the use of pitch-accent information in word recognition. In the present study, we examine the use of the pitch-accent information available in spoken Tokyo Japanese words by young listeners from areas where accentless Japanese is spoken.

TABLE I. Accent patterns in three sample bisyllabic words (with following particle), exhibiting the three possible patterns (LH, LH[^], HL) for such words in Tokyo Japanese. The pronunciation in Tokyo Japanese (the variety used in the stimulus materials of the present study), in the top line of the table, is compared with the pronunciation in Mito (a city in Ibaraki, thus the variety used in the area from which one of the subject groups in the present study was taken), and in Kumamoto (the variety used in the area from which the other subject group in the present study was taken). Adapted from Shibatani (1990: 190); used with permission

	<i>kaze-ga</i> “wind”	<i>hata-ga</i> “flag”	<i>sora-ga</i> “sky”
Tokyo			
Mito			
Kumamoto			

1.1. Subject groups

Participants in the present study were drawn from each of the two areas of Japan described by Shibatani (1990) as having an “accentless” local variety of Japanese — one group from the northern Kanto area, and the other from Kyushu Island. The decision to test participants from both areas was motivated by the following considerations. The first group to be tested consisted of students native to Tochigi and Ibaraki Prefectures, in the Northern Kanto area, but these subjects were all enrolled at that time in a university situated within the Standard Tokyo Japanese speech area. Although, as pointed out above, no speaker of any Japanese variety within Japan can remain isolated from exposure to Standard Tokyo Japanese, it is possible that the processing of Tokyo Japanese accent distinctions could be affected differently for those who currently live among Tokyo Japanese speakers than for those who do not. Therefore, we also tested the second group of speakers, from Kyushu Island; the area is very distant from the Greater Tokyo area, and all the subjects we tested were still resident there. As will become obvious, we do not now believe that either group is more representative of accentless-area residents than the other.

1.1.1. Tochigi and Ibaraki Japanese subjects

Tochigi and Ibaraki Prefectures are situated in the Northern Kanto area, north and northeast of Chiba and Saitama Prefectures, which in turn are on the northern periphery of the Greater Tokyo metropolitan area. They are found within the white area on the eastern side of Honshu Island in Fig. 1. The variety of Japanese spoken in these areas is characterized by a flat tone, and there are no minimal pairs of words distinguished solely by pitch (see Table I).

Thirty-two undergraduate students at Dokkyo University (Saitama Prefecture) took part in the study, of whom 20 participated in Experiment 1, and the same 20 plus a further 12 participated in Experiment 2. All had grown up in Tochigi or Ibaraki Prefectures and had received all their schooling, prior to university enrollment, in their home area. They were tested at Dokkyo University and the experimenter in both experiments was the first author, a male native speaker of Tokyo Japanese.

1.1.2. Kumamoto Japanese subjects

Kumamoto is situated on Kyushu Island in the far southwest of Japan. It is the capital city of Kumamoto Prefecture, approximately within the southern white area in Fig. 1. The variety of Japanese spoken in Kumamoto is again characterized by a flat tone and again there is no use of pitch to distinguish minimal pairs of words (see Table I).

Sixty-five undergraduate and graduate members of Kumamoto Prefectural University, Kumamoto University and Kumamoto Gakuen University took part in the study: 33 in Experiment 1, and 32 others in Experiment 2. All had been born and had received all their education in the accentless Kumamoto area, and had never lived in a Tokyo Japanese accenting area. They were tested at Kumamoto Prefectural University. Two native speakers of Kumamoto Japanese (one male, one female) served as experimental assistants; they too were selected according to the above criteria, and they used accentless Japanese throughout the testing. All listeners received their instructions from one or the other of these assistants.

2. Experiment 1

2.1. Method

2.1.1. Materials

The materials were those used by Cutler & Otake (1999; Experiment 1): 32 bisyllabic Japanese words, all with the segmental structure CVCV (where V was always short), and each containing the mora/syllable *ka*. Half of the words had initial accent (HL), half did not (LH); these patterns were checked against the Tokyo reference data given by Sugito (1995). For each pattern, in half of the words the syllable *ka* was word-initial, in half word-final. Each word was paired with another word with the contrasting accent pattern, such that the two members of a pair contained the same phonemic segment adjacent to the *ka* (e.g., *kage/kagi*; *baka/gaka*). The full set of words was: HL: *baka*, *buka*, *deka*, *huka*, *kika*, *naka*, *waka*, *yoka*, *kage*, *kagu*, *kako*, *kaku*, *kame*, *kare*, *kasa*, *kazu*; LH: *gaka*, *yuka*, *geka*, *nuka*[^], *shika*, *haka*[^], *taka*, *hoka*, *kagi*[^], *kago*, *kake*[^], *kaki*, *kami*[^], *kara*[^], *kase*[^], *kaze*. The LH words marked with [^] have final accent, while the others are unaccented.

Recordings of these words, by three female speakers of Tokyo Japanese, yielded 96 word tokens, from digitized versions of which the *ka* syllables were extracted. For each of the extracted syllables, the total syllable duration and the duration of the vowel were measured, plus, across the voiced portions of the signal only, five f_0 measures (minimum f_0 ; maximum f_0 ; f_0 range; mean f_0 ; standard deviation of f_0) and two amplitude measures (mean rms amplitude; standard deviation of rms amplitude). All these acoustic data can be found in Cutler & Otake (1999).

2.1.2. Procedure

Subjects were tested individually or in groups of up to four. They heard a tape containing the 96 *ka* tokens, in random order. The tape was played from a Sony DAT player, over Audio-Technica ATH-A9 headphones. For each *ka* token on the tape, the subjects were required to perform a two-alternative forced choice, choosing between two words from which it might have come (e.g., *kage* vs. *kagi*; *baka* vs. *gaka*). These choices were written on the response sheet in both kanji and hiragana orthography; the subjects circled their choice for each token. Note that subjects were only asked to decide between two initial syllables (one H, one L) or between two final syllables (one H, one L). The choice was, further, always between the two members of a phonetically matched pair, so that possible coarticulatory information adjacent to the *ka* boundary could not provide clues to identify the source word. Each pair occurred on the response sheet six times (corresponding to the two source words spoken by each of the three speakers), and it was given three times in each possible order, with neither source word nor speaker always having the same order.

2.2. Results and discussion

2.2.1. Perceptual judgments

The percent correct judgments were subjected to two analyses of variance, one with listeners as random factor (*F*-ratio referred to as *F*1) and the other with items as random factor (*F*2).

Tochigi/Ibaraki listeners. The overall correct response rate was very high (74.2%). Identification was more accurate for H (84.2%) than for L syllables (64.3%; $F_1 [1, 19] = 26.56, p < 0.001$; $F_2 [1, 84] = 57.14, p < 0.001$), but accuracy for initial (76%) vs. final syllables (72.4%) did not differ significantly.

Kumamoto listeners. Again the correct response rate was high (72.5%). Again identification was more accurate for H (80.1%) than for L syllables (65.0%; $F_1 [1, 32] = 34.66, p < 0.001$; $F_2 [1, 84] = 31.27, p < 0.001$), but accuracy for initial (73.4%) vs. final syllables (71.7%) did not significantly differ.

Joint analyses. A joint analysis of the two subject groups revealed no significant difference between them in accuracy, and no interactions between the groups factors and any of the other factors in the experiment.

A further joint analysis compared these two subject groups with the original subject group tested by Cutler & Otake (1999). This original group was tested at Soka, in

TABLE II. Experiment 1: percent correct assignments, overall and for each syllable type, for the original subject group tested by Cutler & Otake (1999) and for the two accentless subject groups of the present study

	Subject group		
	Tokyo (original)	Tochigi/ Ibaraki	Kumamoto
Mean	74.0%	74.2%	72.5%
Initial H	90.3%	86.5%	81.4%
Initial L	69.1%	65.6%	65.3%
Final H	84.0%	81.9%	78.7%
Final L	52.4%	62.9%	64.6%

Saitama Prefecture, and consisted exclusively of listeners from that area, i.e., native speakers of a variety with Tokyo-Japanese-style pitch accent (the horizontal-striped area in Fig. 1). Mean percent correct for each of the three groups for each type of syllable is shown in Table II. There was no main effect of the groups factor, indicating that neither of the new subject groups significantly differed in overall accuracy from the original group (mean 74%). However, two interactions did appear involving the groups factor, namely the interaction of groups with the H/L comparison ($F_1 [2, 74] = 3.27, p < 0.05$; $F_2 [2, 168] = 13.56, p < 0.001$) and the interaction of groups with the initial/final comparison ($F_1 [2, 74] = 3.37, p < 0.04$; $F_2 [2, 168] = 11.24, p < 0.001$). The original subject group showed a larger advantage for H over L syllables (87 vs. 61%, a difference of 26%, as opposed to 20 and 15% in the case of the present two groups), and the original subject group also showed an advantage of initial over final syllables (79.7 vs. 68.2%) which did not appear with either of the two new subject groups.

For completeness, it should be reported that one of the three speakers whose tokens were used in this study had a noticeably high-pitched voice, and the productions of this speaker were judged less accurately than the productions of the other two. This was true of the performance of the two new groups, just as of that of the original group; the speaker factor did not interact with the groups factor.

The perceptual judgments therefore show a clear pattern: listeners from the accentless areas can easily tell whether a CV syllable has been extracted from a Tokyo Japanese HL or LH bisyllabic word. They are, overall, as successful in performing this task as listeners from the Tokyo area. Only when different types of syllables (initial vs. final, H vs. L) are compared do small differences appear in the way the task is performed by the two groups from the accentless-variety areas on the one hand vs. the Japanese speakers from the Tokyo area on the other.

2.2.2. Correlations of perceptual judgments and acoustic measures

To obtain a uniform measure of listeners' performance across tokens, the responses were converted to percentage H judgments — the percentage of correct responses for each syllable which actually was H, and the percentage of error responses for each syllable which was actually L. Correlation coefficients were then computed comparing the mean H responses per syllable with the acoustic measures obtained for each syllable.

For the original subject group, Cutler & Otake (1999) reported that subjects were significantly more likely to decide that a syllable was H when it had high minimum f_0 , high maximum f_0 , high mean f_0 , and high mean amplitude, and were significantly less likely to decide that a syllable was H when it had a large f_0 range or when it had a large f_0 standard deviation. These four positive and two negative correlations also appeared, across the full set of 96 tokens, in the responses of each of the two new subject groups. One further correlation reached significance for both new subject groups although it had not been significant for the original subject group: H judgments were more likely when syllables had a large standard deviation of rms amplitude. Thus when averaged across all items, there was considerable similarity between the performance of the two subject groups from accentless-variety areas and that of the original subject group; only an increased sensitivity to amplitude change served to distinguish the new groups from the original.

When subsets of the materials were considered separately, however, more differences did become apparent. These can be simply summarized.

Initial syllables only. For the original subject group, all correlations which held over the whole data set also held over the subset of initial syllables. For both the new subject groups, one exception appeared: there was no significant correlation between H judgments and f_0 range.

Final syllables only. For the original subject group, there was no significant correlation between H judgments and maximum f_0 . For both the new subject groups, however, this correlation did reach significance, just as for the whole data set.

H syllables only. For the original subject group, only two correlations (of H judgments with maximum f_0 and with mean f_0) reached significance. For both the new subject groups, however, no correlations reached significance for this subset of the data.

L syllables only. For the original subject group, only four correlations (with minimum f_0 , maximum f_0 , mean f_0 and mean rms amplitude) reached significance for this subset of the data. For the Tochigi/Ibaraki subjects, no correlations reached significance for this subset, and for the Kumamoto subjects only the correlations with minimum f_0 and mean f_0 were significant.

2.2.3. Signal detection analyses

Cutler & Otake (1999) explained the significantly higher accuracy scores for H than for L syllables in the responses of the original listener group in terms of bias: listeners preferred to judge an isolated syllable as H rather than as L. Cutler & Otake also suggested that the original listeners showed a bias to judge all the productions of Speaker 1 with her high-pitched voice) as H. To assess the relative contributions of sensitivity to the information in the input *vs.* bias in the present data, measures of d' (sensitivity) and beta (bias) were calculated for each group (the original subject group as well as each of the two new groups), based on the group's summed frequencies of H and L responses to H and L syllables, respectively. Both measures were calculated according to the formulae of MacMillan and Creelman (1991): $d' = z(H) - z(F)$, $\beta = -0.5(z(H) + z(F))$, where H is the proportion of hits, F the proportion of false alarms, and z the cumulative

TABLE III. Experiment 1: *d*prime and beta values, for the original subject group tested by Cutler & Otake (1999) and for the two accentless-subject groups of the present study

	Subject group		
	Tokyo (original)	Tochigi/ Ibaraki	Kumamoto
<i>d</i> prime	1.50	1.50	1.32
Speaker 1	0.99	1.16	1.11
Speaker 2	1.73	1.65	1.33
Speaker 3	1.77	1.70	1.53
beta	0.44	0.30	0.23
Speaker 1	0.79	0.49	0.43
Speaker 2	0.25	0.11	0.07
Speaker 3	0.27	0.31	0.19

normal distribution. The results of these analyses are shown in Table III. The same measures were calculated for every subject (of each of the three groups), and these subject-based measures were then submitted to an analysis of variance.

*d*prime. There was no difference across the three groups on this measure, and no interactions reached significance.

Beta. There was a significant difference between groups on this measure: $F(2, 74) = 4.17$, $p < 0.025$. There was also a significant interaction between the inter-group comparison and the speaker variable: $F(4, 148) = 2.93$, $p < 0.025$. Separate analyses of the beta factor in the responses to the productions of each of the three speakers showed that the groups differed significantly in their responses to speaker 1 ($F[2, 74] = 7.93$, $p < 0.001$), with the original Tokyo listener group showing greater bias than either of the two new subject groups, which did not differ. There was no significant difference between the groups in responses to the productions of speaker 2 or 3.

The results are thus very clear. The two groups of subjects from the accentless-variety areas produce similar patterns of performance. Moreover, as observed above, their overall accuracy in this judgment task is virtually indistinguishable from that of a group of speakers of the standard Tokyo variety. Signal detection analyses confirm that the three groups do not differ at all on *d*prime, the measure of sensitivity to information in the input. It is reasonably easy to tell whether a syllable has been extracted from a HL or a LH word of Tokyo Japanese, and this task is no more difficult for listeners from the accentless areas than for those from the Tokyo area.

However, as seen in Table II, the two new subject groups showed worse performance than the original group on H syllables (means of 84.2 and 80.1% correct, *vs.* 87% for the original group), and better performance on L syllables (64.3 and 65% *vs.* 61%), leading to, as described above, a significantly smaller H/L effect than was observed in the original data. Signal detection analyses suggested that the responses of the original subject group were more affected by a bias towards H decisions, particularly to the productions of speaker 1, than were the responses of either of the two new subject groups. The

judgments of the two new subject groups were also less dependent on the acoustic measures for the set of H syllables and the set of L syllables considered separately. Similarly, the pattern of correlations differed for the set of initial and the set of final syllables considered separately: f_0 range was taken into account in judging initial syllables by the original subject group but not by the two new groups, and maximum f_0 was taken into account in judging final syllables by the two new groups but not by the original subject group. Although the overall accuracy of the judgments is statistically identical, there are differences in the way the new groups of listeners and the original group of listeners perform the judgment task. We will consider the implications of these differences in the general discussion.

In Experiment 2, we used a gating task to explore further the processing of initial pitch-accent patterns by the present listener groups. Experiment 1 showed that, given a forced choice between two alternatives and generous time in which to make the choice, our two new subject groups achieved a high standard of pitch-accent recognition. But Experiment 1 did not allow us to examine whether these listeners use accentual information to guide lexical hypotheses in the process of recognizing spoken words. A gating task does address this issue, since we can examine the guesses which listeners make when they are given only fragments of words, and determine whether these guesses are based on segmental information alone or on suprasegmental information as well.

3. Experiment 2

3.1. Method

3.1.1. Materials

Again we used the materials of Cutler & Otake (1999; Experiment 2). These 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 that 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) was *bakuhu/bakuchi*, *hanabi/hanawa*, *hokubu/hokuro*, *kamotsu/kamome*, *karahuru/karamatsu*, *karasu/karada*, *karuteru/karudera*, *kasegi/kasetsu*, *kokugi/kokugo*, *maguchi/maguro*, *moguri/mogura*, *mokuba/mokuji*, *nagasa/nagashi*, *namida/namiki*, *nimotsu/nimono*, *nomichi/nomiya*, *sashizu/sashiki*, *sekiri/sekiryu*, *tachiba/tachiki*, *tomato/tomari*, *wakaba/wakate*, *wakame/wakare*, *warabi/waraji*, *yomichi/yomise*. All but four LH-words (*nagashi* LHH[^], *nomiya* LHL, *karudera* LHHH[^] and *wakare* LHH[^]) were unaccented.

Twenty-four further words served as practice and warmup items. Some of these contained moraic nasals, geminate consonants, or long vowels. Twelve were three-mora words (eight LHH, four HLL) and 12 were four-mora words (six LHHH, four HLLL, and two LHHL).

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 detail in Cutler and Otake (1999).

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 contained 12 HL- and 12 LH- experimental words, and the members of any pair occurred at the same position on both tapes.

3.1.2. Procedure

There were two subject groups, of 16 listeners each, and each group heard a different one of the two tapes. The listeners were tested individually or in groups of up to eight. The tape was played, as before, over headphones from a DAT player; presentation was stopped after each fragment to allow time for the listener to record a guess as to the word's identity. The guesses were written on a response sheet in their normal Japanese orthographic forms.

Cutler & Otake (1999) concentrated their analyses on the candidate words 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-. 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 for the two new subject groups we again examine these initial guesses.

3.2. Results and discussion

3.2.1. Accent recognition

The word guesses were scored by hand and the accent pattern of each guess in Tokyo Japanese determined (using Sugito (1995) as reference). As in the study of Cutler & Otake (1999), 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.). Percent correct for each fragment was subjected to a sign test of statistical significance of the difference from chance, and analyses of variance were carried out on the accuracy scores across word types. Figs 2 and 3 show for each subject group the proportion of guesses which had the same initial accent pattern as

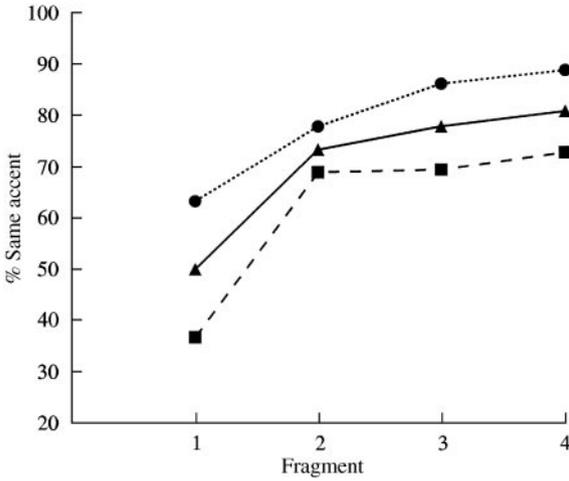


Figure 2. Experiment 2, Tochigi/Ibaraki listeners: proportion of guessed words with the same initial accent pattern as the spoken word: $\cdots\bullet\cdots$, LH-; $\text{---}\blacktriangle\text{---}$, average; $\text{---}\blacksquare\text{---}$, HL-.

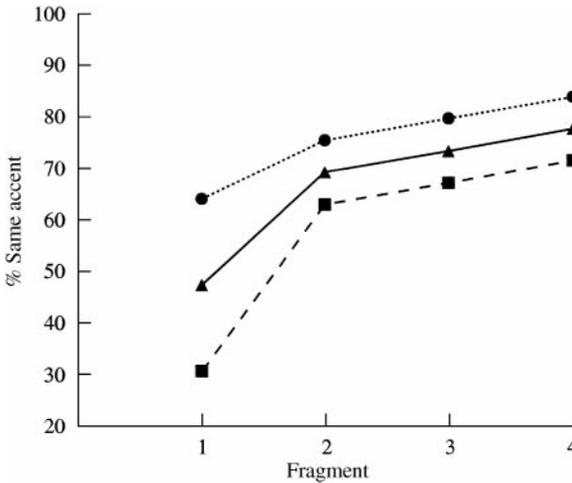


Figure 3. Experiment 2, Kumamoto listeners: proportion of guessed words with the same initial accent pattern as the spoken word: $\cdots\bullet\cdots$, LH-; $\text{---}\blacktriangle\text{---}$, average; $\text{---}\blacksquare\text{---}$, HL-.

the spoken word, for each of the first four fragments, separately for HL- and LH- words and averaged across these.

Tochigi/Ibaraki listeners. The guesses provided for fragment 1 matched the initial accent pattern of the target word in 50% of cases, which is exactly chance performance. At fragment 2, however, 73.4% of the guesses matched the initial accent pattern of the target word, which is significantly above chance performance ($z = 12.93$, $p < 0.001$), and accuracy rose further at fragments 3 and 4. Accuracy was significantly greater

for LH- words than for HL- words ($F_1 [1, 31] = 70.04, p < 0.001$; $F_2 [1, 46] = 20.85, p < 0.001$), and this difference was separately significant at every fragment. There was no significant difference between the group of 20 subjects who had also participated in Experiment 1 (mean scores for the four fragments 50.4, 72.7, 78.1, and 81.3%, respectively) and the group of 12 subjects who had not (49.3, 74.7, 77.4, and 80.2%).

Kumamoto listeners. The guesses provided for fragment 1 matched the initial accent pattern of the target word in 47.4% 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 69.3% of cases, which is significantly above chance performance ($z = 10.64, p < 0.001$), and again accuracy continued to rise across fragments 3 and 4. Accuracy was again significantly greater for LH- words than for HL- words ($F_1 [1, 31] = 55.67, p < 0.001$; $F_2 [1, 46] = 35.33, p < 0.001$), and this difference was again separately significant at every fragment.

Joint analyses. A joint analysis of the accuracy scores was also undertaken to compare the two new subject groups with the original Tokyo Japanese subject group tested by Cutler & Otake (1999); those listeners were from the same population as the original Experiment 1 subject group, but none had participated in the Experiment 1 task. The joint analysis revealed a significant difference between the groups ($F_1 [2, 97] = 16.37, p < 0.001, F_2 [2, 46] = 46.10, p < 0.001$). This difference was separately significant at fragments 2–4, but there was no difference between the groups at fragment 1. Separate comparisons between the subject groups revealed that the original subject group had higher accuracy than the Tochigi/Ibaraki group at all of fragments 2–4 (fragment 2: $F_1 [1, 66] = 4.93, p < 0.03, F_2 [1, 23] = 11.67, p < 0.005$; fragment 3: $F_1 [1, 66] = 12.55, p < 0.001, F_2 [1, 23] = 18.11, p < 0.001$; fragment 4: $F_1 [1, 66] = 10.86, p < 0.002, F_2 [1, 23] = 22.68, p < 0.001$) as well as higher accuracy than the Kumamoto group at all of fragments 2–4 (fragment 2: $F_1 [1, 66] = 13.93, p < 0.001, F_2 [1, 23] = 23.42, p < 0.001$; fragment 3: ($F_1 [1, 66] = 23.86, p < 0.001, F_2 [1, 23] = 60.37, p < 0.001$; fragment 4: $F_1 [1, 66] = 25.26, p < 0.001, F_2 [1, 23] = 32.18, p < 0.001$), whereas the two new subject groups did not differ significantly from each other at any of fragments 2, 3 or 4.

The overall analysis also revealed an interaction between the groups factor and the HL-/LH- comparison, and this interaction was significant in the comparison of the original Tokyo subject group with the Tochigi/Ibaraki group and also with the Kumamoto group. It was not significant in the comparison between the two new subject groups. The source of the interaction was that, as described above, both new subject groups showed significantly greater accuracy for LH- words than for HL- words at every fragment; the original Tokyo subject group did not show a significant difference in accuracy for the two word types at any fragment other than fragment 1.

Thus, the two groups of subjects from accentless-dialect regions can effectively make use of pitch-accent information in the recognition of spoken Tokyo Japanese words. Even half of the first vowel of a word (fragment 2) suffices to enable them to determine pitch-accent pattern significantly more accurately than chance would predict. Nonetheless, their performance does not in this case equal that of the original subject group at any of fragments 2, 3 or 4. Moreover, at each of these fragments both of the two new subject groups show an accuracy advantage for LH- over HL- words, which the original subject group did not show at fragments 2, 3 or 4.

All three subject groups show an equivalent accuracy advantage for LH- over HL- at fragment 1. For the original subject group, Cutler & Otake (1999) explained this effect as the result of simple guessing. More of the Tokyo Japanese vocabulary consists of words beginning LH- than of words beginning HL-. When no information is available about pitch accent, or indeed word length, type of word, or anything apart from the initial phoneme, subjects must guess. A reasonable guessing strategy is to choose candidates on the basis of vocabulary frequency. For pitch accent, this implies that listeners should guess more LH- words than HL- words. Since the stimulus words were equally often HL- as LH-, such guessing behavior will result in a spurious accuracy advantage for LH- over HL- words. With the original subject group, this difference between LH- and HL- words disappeared from fragment 2. With the two new subject groups, however, it was present at all fragments. This suggests that the two new subject groups had greater recourse to guessing than the original subject group did; in particular, unlike the original subject group, they still made some use of simple guessing, despite pitch-accent information available in the signal, in the later fragments. To test this suggestion, we again carried out signal detection analyses of the response patterns.

3.2.2. Signal detection analyses

Measures of *d*prime and beta, separately for responses to each fragment, were calculated for each subject in the two new groups and for each subject in the original Tokyo subject group, as well as for each group as a whole, using the same procedures as for Experiment 1. The calculations were again based on the two-way classifications of guesses as to whether their initial accent pattern (HL-, LH-) did or did not match the pattern of the word presented, separately for the HL- and LH- target words. Fig. 4 shows the group measures across fragments. The subject-based measures were again entered into an analysis of variance.

*d*prime. As Fig. 4(a) clearly shows, the *d*prime scores were higher for the original subject group than for the two new groups. There was a significant main effect of the groups factor ($F [2, 97] = 17.05, p < 0.001$), and an interaction of the groups comparison with fragments ($F [6, 291] = 2.32, p < 0.04$); separate analyses for each fragment revealed no significant difference between groups at fragment 1, but a significant difference at fragments 2–4. Comparisons of the original Tokyo subject group with each of the other two groups revealed that the *d*prime measure was higher for the original group than for the Tochigi/Ibaraki listeners, overall ($F [1, 66] = 13.41, p < 0.001$) and at each of fragments 2–4 ($F [1, 66] = 5.97, p < 0.02, F [1, 66] = 12.88, p < 0.001, F [1, 66] = 9.74, p < 0.005$, respectively), and higher for the original group than for the Kumamoto listeners, again overall ($F [1, 66] = 34.63, p < 0.001$) as well as at each of fragments 2–4 ($F [1, 66] = 14.72, p < 0.001, F [1, 66] = 25.07, p < 0.001, F [1, 66] = 25.72, p < 0.001$, respectively). However, *d*prime scores for the two new subject groups did not differ significantly at any fragment.

Beta. Fig. 4(b) shows that the beta scores for the original subject group are elevated at fragment 1 but then drop to close to zero and do not rise again for later fragments; the beta scores for the two new groups, however, begin higher than those of the original subject group, and remain elevated throughout the following fragments. For beta there was also a significant main effect of the groups factor ($F [2, 97] = 4.01, p < 0.025$),

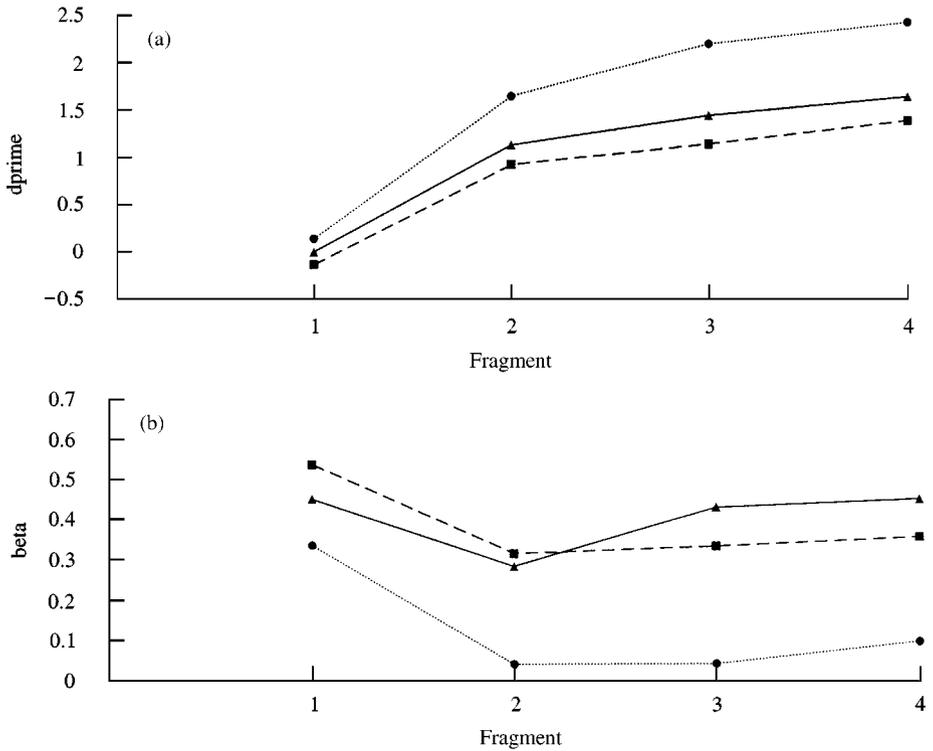


Figure 4. Experiment 2: values for dprime (a) and beta (b), for the original subject group tested by Cutler & Otake (1999) and for the two accentless-subject groups of the present study. (a) $\cdots\bullet\cdots$, Tokyo; $\text{---}\blacktriangle\text{---}$, Tochigi/Ibaraki; $\text{---}\blacksquare\text{---}$, Kumamoto; (b) $\cdots\bullet\cdots$, Tokyo; $\text{---}\blacktriangle\text{---}$, Tochigi/Ibaraki; $\text{---}\blacksquare\text{---}$, Kumamoto.

but in this case this effect did not interact with the fragments factor. Separate analyses showed that beta scores were higher for the Tochigi/Ibaraki listeners than for the original Tokyo subject group ($F[1, 66] = 6.16$, $p < 0.02$), and higher for the Kumamoto listeners than for the original Tokyo subject group ($F[1, 66] = 5.32$, $p < 0.03$), while there was no significant difference between the two new subject groups on this measure.

These analyses thus reveal that again the two new subject groups perform very similarly, and that they both differ in the same way from the original subject group. For both Tochigi/Ibaraki and Kumamoto listeners, their sensitivity to the information in the input is lower than that of the original subject group, and their bias (to respond with LH- words, corresponding to the majority distribution in the vocabulary) is higher; unlike the original subject group, the responses of both the Tochigi/Ibaraki and Kumamoto groups are still affected by bias at fragments 2–4.

4. General discussion

In both experiments, we have observed that speakers of Japanese from the accentless-dialect areas can very effectively interpret the pitch-accent information in Tokyo

Japanese words, and can use this information in the recognition of spoken words. Although inter-word differences in pitch accent are not used in the dialect of their region, these speakers experience no difficulty in distinguishing the inter-word contrasts made in Tokyo Japanese. Indeed, their overall level of performance in making the perceptual distinctions is in many respects indistinguishable from the performance of listeners from the Tokyo Japanese pitch-accent region. Thus in Experiment 1 the two-alternative forced-choice task of assigning an isolated syllable to the HL or LH word from which it had been extracted was performed just as accurately by the two groups from the accentless-dialect areas as by the original subject group tested by Cutler & Otake (1999); the level of performance of all three groups was statistically identical. Likewise, in Experiment 2 the listeners from both accentless-area groups were able to limit their word guesses to alternatives which matched the initial accent pattern of the input, with a probability which was significantly above chance, even at the second fragment (the word's initial consonant plus half of the vowel following it); this was also the pattern observed with the original listener group tested by Cutler & Otake (1999).

As we observed in the introduction, the listeners we have tested in this study will not be monodialectal users of accentless Japanese. They have certainly all been extensively exposed to Tokyo Japanese, and as university students they use it, at least in all formal situations. Our study does not address the question of whether speakers of accentless varieties who have never been exposed to pitch-accent distinctions, or have hardly ever heard them, could make effective use of the suprasegmental cues which varied in our materials. Unfortunately, this is not a question which is ever likely to find an answer, given that it would probably be impossible to find a Japanese-native subject population without at least extensive passive exposure to Tokyo Japanese.

Although the perceptual exposure was our principal focus of interest, we did in fact attempt to ascertain the extent to which our listeners used accentless Japanese in production. Informal listening in Kumamoto by the first author, the colleague who organized the testing there, and the two experimental assistants revealed considerable evidence of the use of local forms. This evidence was not available for the Tochigi/Ibaraki speakers, who were studying at a university in the Tokyo Japanese area and were presumably well accustomed to using Tokyo Japanese (especially in conversation with, as in this case, a Tokyo Japanese-speaking faculty member). Attempts to record the speech of the two listener groups were not very successful. Reading aloud from a word list by the Tochigi/Ibaraki group produces — in conformity with previous work by Sato (1984), Sugito (1988) and Jinnai (1996) — apparently acceptable Tokyo Japanese recordings. With the Kumamoto group, we attempted to record natural conversation, but a recorded interview with an unfamiliar interviewer was a formal enough situation that the speakers switched to Tokyo Japanese. Interestingly, however, analysis of the recordings suggests that these Kumamoto speakers did not always succeed in realizing their intention to use standard Japanese, since many of them used inconsistent forms and, sometimes, inaccurate accent patterns. A precise estimate of their use of dialect seems impossible; we can however conclude that these listeners from the accentless-dialect areas are at best bidialectal, and certainly not monodialectal Tokyo Japanese speakers.

What our study has demonstrated is that the speakers from these areas can make highly effective use in listening of the suprasegmental distinctions which cue lexical pitch accent, even though such distinctions are absent from the local-dialect repertoire and thus are of no use in perceptual processing of that variety. Note that it is only inter-word distinctions cued by suprasegmental structure which are at issue here; as

pointed out in the introduction, the dialects in question certainly use suprasegmental structure in many other ways, and there is no suggestion that listeners of these dialects (or of any other language or dialect) should be in principle unable to perceive distinctions of pitch. The results of our study show that they can actually make use of suprasegmental information for selecting between words. Moreover, they can use this information very early in word recognition; thus for instance a LHHH[^] word can be recognized as LH-initial and not HL-initial from only part of the first (L) syllable; it is not the case, for example, that these listeners would have to wait until the accent-carrying syllable (in a LHHH[^] word, the fourth syllable) to make the relevant distinctions.

Nonetheless, the near equivalence between these listeners' overall performance and that of the original listener group tested by Cutler & Otake (1999) does not constitute the full story. Clear and significant differences appeared in the way each of the tasks was performed, which distinguished the listeners belonging to the two accentless groups, on the one hand, from the Tokyo Japanese listeners tested by Cutler & Otake (1999) on the other. In Experiment 1, the responses of the listeners from both the Tochigi/Ibaraki group and the Kumamoto group were less affected by a bias which had characterized the responses of the original subject group, namely a preference to classify in particular the syllables produced by Speaker 1 as H. In Experiment 2, the responses of the listeners from both groups displayed significantly less sensitivity to the information in the signal than did the responses of the original subject group, and both groups also showed more reliance on guessing, as reflected in a greater tendency of their responses to reflect the pitch-accent structure of the vocabulary.

Especially striking, furthermore, is the fact that in all these respects the performance of the two groups tested here was virtually identical. Although the Tochigi/Ibaraki listeners live and work among speakers of Tokyo Japanese, their processing was, in essentially every characteristic, statistically indistinguishable from that of the Kumamoto group. Figs 2 and 3 suggest that Tochigi/Ibaraki listeners enjoyed a slight advantage of accuracy over the Kumamoto listeners in the match between the accent pattern of their word guesses and the accent pattern of the input, but this difference between the two groups was never statistically significant. We interpret this notable equivalence as an indication that exposure to Tokyo Japanese, via the broadcast media and in whatever other relevant situation, is sufficient even for the Kumamoto listeners to achieve ceiling or near-ceiling perceptual performance.

How can we interpret the differences we have observed between the dialect groups tested here and the original Tokyo Japanese group? Some aspects of the detailed results reported above could be of use in pointing to an explanation. Consider the fact that in Experiment 1 the listeners from the accentless-variety areas showed, across the whole data set, the same pattern of correlations between responses and the acoustic characteristics of the syllables as the original subject group had shown, but that the correlations did not pattern similarly across subsets of syllable types. For initial syllables only, the two new listener groups did not show the negative correlation between f_0 range and H judgments—that is, they showed no significant tendency to interpret f_0 movement in an initial *ka* syllable as an indication that the syllable came from a LH word. Such pitch movement is, however, a very good cue, since there will be a large pitch rise across the first mora of a LH word in Tokyo Japanese. This difference certainly suggests a less acute appreciation of the structure of Tokyo Japanese pitch accent patterns. Similarly, for the original subject group, mean f_0 (along with minimum and maximum f_0) was associated with a clear distinction between H and L syllables—the higher the f_0 , the more likely the

syllable was to be deemed H, and the lower the f_0 , the more likely it was to be judged L. Again, these correlations are evidence of efficient use of relevant cues, since H and L syllables showed very significant differences on all these measures. These correlations were weaker or absent in the responses of the subjects from accentless areas. Again, although the present subject groups achieved accuracy which was virtually equivalent to that of the Tokyo Japanese listeners, this accuracy does not seem to have been based on as sure a use of the acoustic cues as that available in the signal.

Of course, the Tokyo Japanese listeners' confident reliance on the f_0 cues in the Experiment 1 task did not always prove to their advantage; with speaker 1 they showed a bias to judge all her production as H, whereas the Tochigi/Ibaraki and Kumamoto listeners were able to avoid some of this bias by relying less on the use of the acoustic correlates of pitch-accent patterns despite the fact that these correlates could have proven informative to them.

In Experiment 2, the signal detection analyses revealed that the listeners from the accentless-dialect areas displayed reduced sensitivity to information in the signal, and instead a greater tendency to guess; again, although they could perform the task extremely well, they certainly did not confidently rely upon the acoustic information in the way which characterized the response of the original Tokyo Japanese subject group.

Nevertheless, in order to have achieved accurate performance, the Tochigi/Ibaraki and Kumamoto listeners must have been able to process the acoustic cues to accent which were provided in the input, and furthermore they must have been able to match these to some stored information about accent pattern in their mental representation of the stimulus words. The discussion above suggests that the way in which the acoustic cues were processed was not quite the same as the way Tokyo Japanese listeners processed the same information. We suggest that it is also quite likely that the accentless listeners' stored information about pitch accent in the mental lexicon differed from that of Tokyo Japanese listeners. Although it is hardly possible to test such a suggestion against the current data, we propose that word representations in the mental lexicon of a Tokyo Japanese speaker may contain pitch accent information as a necessary component, whereas the same information may be represented — and by implication learned — as an additional component of the lexical entry in the case of speakers who originally acquired accentless varieties. (Note that this suggestion is not unique to the present comparison. It has also been made with respect to the storage of lexical gender by, on the one hand, speakers of a language with gender, *vs.* on the other hand, fluent L2 speakers of the same language whose L1 has no gender contrasts; Guillelmon & Grosjean, 1998.)

On this account, the less confident use of acoustic correlates of pitch accent by the Tochigi/Ibaraki and Kumamoto listeners could have its basis in a realistic appreciation of a less firmly anchored status of the pitch-accent information in stored lexical representations.

Let us now consider the implications of our findings for the relationship between native and non-native phonological systems. Despite the undoubted differences in the nature of the processing of pitch accent by the Tochigi/Ibaraki and Kumamoto listeners in comparison with the original Tokyo Japanese subject group, it is important not to lose sight of the overall high level of accuracy observed in the present study. Although the non-native listeners may have reached the goal of identifying Tokyo Japanese pitch-accent patterns by a route different than that used by native listeners, reach it they did.

In this respect, the results that we have observed appear to stand in strong contrast to recent findings on phonetic processing from which the conclusion has been drawn that extensive exposure to a contrast is not sufficient to induce effective exploitation of that contrast in speech perception. The study of Pallier et al. (1997), mentioned in the introduction, clearly showed that Spanish users of Catalan in Barcelona, despite being exposed to Catalan since early childhood, and indeed despite explicit schooling in Catalan, failed to discriminate certain phonetic contrasts which are distinctive in the Catalan phonetic inventory, but not in the inventory of Spanish. Similarly, Sebastián-Gallés and Soto (1999) showed that even Spanish-dominant Catalan-Spanish bilinguals who could successfully discriminate the Catalan contrasts nevertheless needed significantly longer fragments of gated stimuli to make the distinction than Catalan-dominant listeners did. In contrast, the pitch-accent distinctions which formed the subject of the present study seem to be quite accessible to language users from a region using a variety without such contrasts.

Previous studies of the perception of suprasegmental distinctions have revealed processing differences between native and non-native listeners. Thus as Dupoux et al. (1997) showed, Spanish and French listeners differ in how accurately they can perceive stress contrasts in nonsense words spoken by a Dutch speaker; similarly, Dutch and Finnish listeners can use suprasegmental correlates of stress as a cue to word onset, while French listeners do not (Vroomen, Tuomainen & de Gelder, 1998). While the discrimination of tone contrasts is affected by word-nonword status only for native listeners and not for non-native listeners (Fox & Unkefer, 1985; Cutler & Chen, 1997), overall similarity in discriminability of tone contrasts for native and non-native listeners (Burnham, Kirkwood, Luksaneeyanawin & Pansottee, 1992; Cutler & Chen, 1997) can mask processing differences — for instance, greater interdependence of segmental and suprasegmental judgments for tone language speakers than for speakers of non-tone languages (Repp & Lin, 1990; Lee & Nusbaum, 1993), or dependence on native distinctions in making the non-native distinctions (Broselow et al., 1987; Nishinuma et al., 1996). Multidimensional scaling analyses reveal that the perceptual dimensions used by speakers of tone languages more closely resemble those used by speakers of other, unrelated tone languages than those used by speakers of non-tone languages (Gandour, 1978, 1983; Gandour & Harshman, 1978a, b).

However, non-native suprasegmental distinctions in perception can be acquired. French and American learners of Japanese can, at least at more advanced levels of study, perform quite well on tests of perceptual discrimination between pitch-accent patterns (Nishinuma, 1994; Nishinuma et al., 1996). American listeners can be taught to identify the four tones of Mandarin Chinese, and their improved sensitivity is still retained months later (Wang et al., in press). Are suprasegmental distinctions thus in principle different from segmental distinctions in their accessibility to non-native listeners?

This seems to us a highly unlikely state of affairs. Instead, we would argue that segmental distinctions are more likely than suprasegmental distinctions to be subsumed by native-language categories, especially of course in languages which make relatively less use of the suprasegmental option. In this case suprasegmental distinctions will effectively be more accessible, because they are less subject to competition from native distinctions, than segmental contrasts. This claim is based on the assumption that, as argued by many researchers in the field of cross-language speech perception (e.g., Best, 1994; Flege, 1995), listeners' sensitivity to non-native contrasts depends upon the relation of the contrasts in question to the native phonological system. A non-native contrast may map to a

listener's native system in a number of different ways; Best, McRoberts & Sithole (1988) distinguish four such relationships for segmental contrasts:

(a) The contrasting sounds are both assimilated to the same category in the speaker's native language — for instance, English [l] and [ɫ] for Japanese speakers. This “single-category assimilation”, in the Best et al. term, is the most difficult non-native contrast to perceive.

(b) The contrasting sounds may be assimilated to the native language, but to different categories. This is the case with the glottalized velar *vs.* alveolar stops of some Northwest Pacific languages, which for English speakers assimilate to the native velar and alveolar stop categories. These “opposing-category” contrasts are easy to perceive, even for listeners who have no experience with the type of contrast (as English listeners have no experience with glottalized stops), because the sounds are assimilated to categories which also contrast in the native language.

(c) One of the contrasting sounds is assimilated to a native category but the other is not. This is the case, for example, with Hindi dental *vs.* retroflex stop consonants, which English listeners assign to the same (alveolar) place of articulation category. These are “category-goodness” differences, say Best et al., i.e., they should give rise to differences in relative “goodness of fit” to the native category; they should be difficult but perceptible.

(d) Finally, it might be the case that neither contrasting sound can be assimilated to a native category. Because the world's languages have selected their phonetic stock from a relatively limited range, such “non-assimilable” cases are rare; but an example is found in the click contrasts found in Bantu languages, which for speakers of nonclick languages are unlike any of their native sounds.

Best et al. (1988) presented English-speaking listeners with Zulu click contrasts, and observed excellent discrimination performance — as good, in fact, as that of adult Zulu speakers. Likewise, Best & Strange (1992) found that Japanese-speaking adults who cannot discriminate between /r/ and /l/ can make the (very similar) distinction between English /r/ and /w/; the boundary between these two sounds is one which does not correspond to a Japanese approximant boundary, but the two sounds are also not conflated within one Japanese category in the way English /r/ and /l/ are. Thus, listeners' ability to discriminate non-native contrasts depends crucially on how the contrasts in question relate to native category distinctions.

In the case of accentless varieties of Japanese, there are no word-level categories to which the Tokyo Japanese pitch-accent contrasts can map. Thus, for speakers of only an accentless variety, the Tokyo Japanese contrasts may represent what Best et al. call unassimilable effects: contrasts which are easy to distinguish because they are orthogonal to the native phonological system.

Other inter-dialect comparisons represent other levels of the Best et al. classification. The vowel contrasts studied by Labov et al. (1991) are effectively the hardest type, namely the single-category assimilation: the distinctions made in some American varieties assimilate to one vowel category in Philadelphia English, and thus such distinctions are particularly hard to make in speech categorization tasks. The stop consonant allophones studied by Scott & Cutler (1984) are instances of the Best et al. category-goodness differences, since the distinction is between one American version which maps well onto British /t/ and another which does not; such differences should be learnable, according to Best et al., and Scott & Cutler indeed found

that British English speakers who were resident in the US could make perceptual use of the contrast.

The questions are still open whether speakers of Tokyo Japanese experience difficulty with the lack of pitch-accent contrasts in Tochigi/Ibaraki and Kumamoto Japanese, and whether pitch-accent differences between the different accenting dialects lead to perceptual confusion. Anecdotes abound which tell of misunderstandings based on inter-dialectal pitch-accent differences, but laboratory evidence is scarce. In the only such study of which we are aware, Warner (in press) presented lists of spoken Japanese words to speakers of Tokyo Japanese and of Kyoto–Osaka Japanese (see Fig. 1; Kyoto–Osaka Japanese is the second largest dialect group in Japan, and manifests many differences in pitch accent from Tokyo Japanese). The listeners heard words both in their own and the other dialect. Although the listeners were fully aware of which variety they were listening to for each word, they made many errors in identifying the non-native forms, particularly when the non-native form could be confused with a different form in the native variety. Interestingly, the Tokyo Japanese listeners had particular difficulty with contrasts in Kyoto–Osaka Japanese of a type not made in their native dialect — for instance, the ascending and descending tones which can be contrastive in the Kyoto–Osaka variety, but do not appear in Tokyo Japanese. The difference between this inter-dialect comparison and the comparison made in the present study is that the Kyoto–Osaka contrast is not orthogonal to the Tokyo Japanese phonological system in the way any pitch-accent contrast is orthogonal to the Tochigi/Ibaraki and Kumamoto systems. Ascending and descending pitch does occur in Tokyo Japanese, as a function of the location of an accent (for instance, a rise occurs in the initial mora of a LH-word, and a fall occurs on the mora following an accented mora). In Warner's experiment, however, the Tokyo Japanese listeners were required to make a distinction between these two tones on monomoraic syllables, and this proved too hard a task for them. The Kyoto–Osaka contrast appears to be particularly difficult for the Tokyo Japanese listeners because it requires them to make a categorical distinction which does not occur in their native variety, even though the perceptual distinction on which the categorization is based is presumably within their grasp. The same situation can arise with segmental distinctions; thus English listeners assimilate the aspirated and unaspirated allophones of /p/ to a single category, even though they can notice when the wrong allophone is used (Whalen, Best & Irwin, 1997). Effectively, the distinction which Warner asked of the Tokyo listeners thus fell at the Best et al. single-category level — the Kyoto–Osaka categories could be assimilated to one category of the Tokyo system.

The pitch-accent system of Japanese, and the way in which it differs across dialectal varieties, offers many further comparisons and contrasts which could be explored in future research. The present study has produced evidence that cross-dialectal perception is sensitive to the relationship between the native and the non-native phonological systems. In this respect, perceptual inequalities between dialects pattern just like perceptual inequalities between languages, and suprasegmental contrasts pattern just like segmental contrasts.

This project was financially supported by grants to the first author from the International Communication Foundation, Japan, and from the Japan Society for the Promotion of Science (Grant-in-Aid for Science Research (C), # 11610566). The project was also greatly facilitated by the help of two people to whom we express particular thanks: Kimiko Yamakawa of Kumamoto Prefectural University, whose exceptional efforts made it possible for us to test 65 Kumamoto

listeners in two days, and Roel Smits of the Max Planck Institute for Psycholinguistics, who suggested the use of signal detection analyses to us, and computed the dprime and beta measures. We are also very grateful to Ryoji Baba and Tsuyoshi Usagawa for facilitating the Kumamoto research, to Yuichiro Aramaki and Ayako Taguchi for acting as experimental assistants at Kumamoto Prefectural University, and to Keren Shatzman for technical assistance in Nijmegen. Further we thank Natasha Warner, James McQueen, Tim Vance, Terry Gottfried and an anonymous reviewer for comments on an earlier draft of the paper.

References

- Best, C. T. (1994) The emergence of native-language phonological influences in infants: a perceptual assimilation model. In *The development of speech perception* (J. C. Goodman & H. C. Nusbaum, editors), pp. 167–224. Cambridge, MA: MIT Press.
- Best, C. T., McRoberts, G. W. & Sithole, N. M. (1988) Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants, *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 345–360.
- Best, C. T. & Strange, W. (1992) Effects of phonological and phonetic factors on cross-language perception of approximants, *Journal of Phonetics*, **20**, 305–330.
- Broselow, E., Hurtig, R. & Ringen, J. (1987) The perception of second language prosody. In *Interlanguage phonology: the acquisition of a second language sound system* (G. Ioup & S. H. Weinberger, editors), pp. 350–361. Cambridge, MA: Newbury House.
- Burnham, D., Kirkwood, K., Luksaneeyanawin, S. & Pansottee, S. (1992) Perception of Central Thai tones and segments by Thai and Australian adults. Pan-Asiatic Linguistics: In *Proceedings of the 3rd International Symposium of Language and Linguistics*, pp. 546–560. Bangkok: Chulalongkorn University Press.
- Cutler, A. & Chen, H.-C. (1997) Lexical tone in Cantonese spoken-word processing, *Perception & Psychophysics*, **59**, 165–179.
- Cutler, A. & Otake, T. (1999) Pitch accent in spoken-word recognition in Japanese, *Journal of the Acoustical Society of America*, **105**, 1877–1888.
- Dupoux, E., Pallier, C., Sebastián-Gallés, N. & Mehler, J. (1997) A destressing deafness in French, *Journal of Memory and Language*, **36**, 399–421.
- Flege, J. E. (1995) Second language speech learning: theory, findings and problems. In *Speech perception and linguistic experience* (W. Strange, editor), pp. 233–277. Timonium, MD: York Press.
- Fox, R. A. & Unkefer, J. (1985) The effect of lexical status on the perception of tone, *Journal of Chinese Linguistics*, **13**, 69–90.
- Gandour, J. T. (1978) Perceived dimensions of 13 tones: a multidimensional scaling investigation, *Phonetica*, **35**, 169–179.
- Gandour, J. T. (1983) Tone perception in Far Eastern languages. *Journal of Phonetics*, **11**, 149–175.
- Gandour, J. T. & Harshman, R. (1978a) Cross-language study of tone perception. In *Linguistic variation* (D. Sankoff, editor), pp. 139–147. New York: Academic Press.
- Gandour, J. T. & Harshman, R. (1978b) Cross-language differences in tone perception: a multidimensional scaling investigation, *Language and Speech*, **21**, 1–33.
- Gottfried, T. L. & Beddor, P. S. (1988) Perception of temporal and spectral information in French vowels, *Language and Speech*, **31**, 57–75.
- Guillemon, D. & Grosjean, F. (1998) The gender-marking effect in spoken word recognition: the case of bilinguals. *Paper presented at Workshop on Processing of Grammatical Gender*, Leipzig, July.
- Inoue, F. (1984) Akusento no seisei to chikaku: Kantou ni okeru chiikisa to nenreisa (Production and perception of (Japanese pitch) accent: area and age differences in the Kanto area). In *Kindaichi Haruhiko hakase koki kinen robunshuu*. (*Papers in celebration of Dr. Haruhiko Kindaichi's 70th birthday*), Vol. 2, pp. 119–153. Tokyo: Sanseido.
- Jinnai, M. (1996) *Chihou chuukaku toshi hougen no yukue – Kyuushuu* (*Trends in the dialects of regional heartlands: Kyushu*). Tokyo: Oufuu.
- Kubozono, H. (1993) *The organization of Japanese prosody*. Tokyo: Kurosio.
- Labov, W., Karen, M. & Miller, C. (1991) Near-mergers and the suspension of phonemic contrast, *Language Variation and Change*, **3**, 33–74.
- Lee, L. & Nusbaum, H. C. (1993) Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese, *Perception & Psychophysics*, **53**, 157–165.
- Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. & Yamada, T. (1994) Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories, *Journal of the Acoustical Society of America*, **96**, 2076–2087.
- Logan, J. S., Lively, S. E. & Pisoni, D. B. (1991) Training Japanese listeners to identify English /r/ and /l/: a first report, *Journal of the Acoustical Society of America*, **89**, 874–886.

- MacMillan, N. A. & Creelman, C. D. (1991) *Detection theory: a user's guide*. Cambridge: Cambridge University Press.
- Maekawa, K. (1997) Akusento to intoneishon: Akusento no nai chiiki (Accent and intonation in accentless dialects). In *Nihongo Onsei 1: Shohougen no akusento to intoneishon (Japanese speech 1: Accent and intonation in various dialects)*, pp. 97–122. Tokyo: Sanseido.
- Miller, J. L. & Grosjean, F. (1997) Dialect effects in vowel perception: the role of temporal information in French, *Language and Speech*, **40**, 277–288.
- Minematsu, N. & Hirose, K. (1995) Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese, *Journal of the Acoustical Society of Japan*, **16**, 311–320.
- Muramatsu, M. (1998) Acoustic analysis of Japanese English prosody: comparison between Fukushima dialect speakers and Tokyo dialect speakers in declarative sentences and yes–no questions. In *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney.
- 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*, Yokohama, pp. 1739–1742.
- Nishinuma, Y., Arai, M. & Ayusawa, T. (1996) Perception of tonal accent by Americans learning Japanese. In *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, Vol. 1, pp. 646–649.
- Pallier, C., Bosch, L. & Sebastián-Gallés, N. (1997) A limit on behavioral plasticity in speech perception, *Cognition*, **64**, B9–B17.
- Pierrehumbert, J. B. & Beckman, M. E. (1988) *Japanese tone structure*. Cambridge, MA: MIT Press.
- Repp, B. H. & Lin, H.-B. (1990) Integration of segmental and tonal information in speech perception, *Journal of Phonetics*, **18**, 481–495.
- Sato, R. (1984) Mukei akusento chiiki ni okeru akusento no kyoutsuuka: Utsunomiyashi ni okeru shouchousa kara (Standardization of (Japanese pitch) accent in accentless areas: from a report in Utsunomiya City) in *Hirayama Teruo hakase koki kinenkai hen. Gendai hougengaku no kadai (Issues in today's dialectology, Papers in honour of Dr. Teruo Hirayama's 70th birthday)*, Vol. 2, pp. 211–234. Tokyo: Meijishoin.
- Scott, D. R. & Cutler, A. (1984) Segmental phonology and the perception of syntactic structure, *Journal of Verbal Learning and Verbal Behavior*, **23**, 450–466.
- Sebastián-Gallés, N. & Soto, S. (1999) On-line processing of native and non-native phonemic contrasts in early bilinguals, *Cognition* **72**, 111–123.
- Sekiguchi, T. & Nakajima, Y. (1999) The use of lexical prosody for lexical access of the Japanese language, *Journal of Psycholinguistic Research*, **28**, 439–454.
- Shibatani, M. (1990) *The languages of Japan*. Cambridge: Cambridge University Press.
- Sugito, M. (1982) *Nihongo akusento no kenkyuu (Studies on Japanese accent)*. Tokyo: Sanseido.
- Sugito, M. (1988) Roudoku onsei ni okeru akusento no chiikiki, nenrei, seibetsu ni yoru sai (Accentual differences between areas, age and sex in narration), *Kokugogaku (Studies in the Japanese language)*, **154**, 132–144.
- Sugito, M. (1995) *Akusento onsei jitsen (Accent pronunciation dictionary)* CD ROM. Tokyo: Maruzen.
- Vroomen, J., Tuomainen, J. & de Gelder, B. (1998) The roles of word stress and vowel harmony in speech segmentation, *Journal of Memory and Language*, **38**, 133–149.
- Wang, Y., Spence, M., Jongman, J. & Sereno, J. A. (in press) Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*.
- Warner, N. (in press) Recognition of accent patterns across dialects in Japanese. In *Proceedings of the 23rd Annual Meeting*. Berkeley Linguistics Society.
- Whalen, D. H., Best, C. T., & Irwin, J. R. (1997) Lexical effects in the perception and production of American English /p/ allophones, *Journal of Phonetics*, **25**, 501–528.