Pitch accent in spoken-word recognition in Japanese

Anne Cutler[a)]
Max-Planck-Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

Takashi Otake[b)]
Faculty of Foreign Languages, Dokkyo University, 1-1 Gakuen-cho Soka, Saitama 340, Japan

(Received 17 January 1998; revised 6 November 1998; accepted 13 November 1998)

Three experiments addressed the question of whether pitch-accent information may be exploited in the process of recognizing spoken words in Tokyo Japanese. In a two-choice classification task, listeners judged from which of two words, differing in accentual structure, isolated syllables had been extracted (e.g., *ka* from *baka* HL or *gaka* LH); most judgments were correct, and listeners' decisions were correlated with the fundamental frequency characteristics of the syllables. In a gating experiment, listeners heard initial fragments of words and guessed what the words were; their guesses overwhelmingly had the same initial accent structure as the gated word even when only the beginning CV of the stimulus (e.g., *na*- from *nagasa* HLL or *nagashi* LHH) was presented. In addition, listeners were more confident in guesses with the same initial accent structure as the stimulus than in guesses with different accent. In a lexical decision experiment, responses to spoken words (e.g., *ame* HL) were speeded by previous presentation of the same word (e.g., *ame* HL) but not by previous presentation of a word differing only in accent (e.g., *ame* LH). Together these findings provide strong evidence that accentual information constrains the activation and selection of candidates for spoken-word recognition. © *1999 Acoustical Society of America.* [S0001-4966(99)03003-9]

PACS numbers: 43.71.Es, 43.71.Hw [WS]

INTRODUCTION

A Japanese word spoken in isolation has a characteristic prosodic pattern: its pitch-accent pattern. In Tokyo Japanese, words can be accented or unaccented. In accented words, one mora of the word is marked as carrying accent and is assigned the accent label high (H). If the marked mora is the first in the word, subsequent morae will be labeled low (L): the pattern will therefore be HL for a two-mora word, HLL for a three-mora word, and so on. If the marked mora is the second or a later mora in the word, the first mora will be low, all other morae between the first and the accented mora will be high, and all morae after the accented mora will be low. Thus *Toyota* is a three-mora word (*to-yo-ta*) in which the first mora is accented: HLL; *Mitsubishi* has four morae (*mi-tsu-bi-shi*) with accent falling on the second mora: LHLL. In unaccented words the first mora is labeled L and all subsequent morae are labeled H; the pattern LHH can therefore describe both an unaccented word and an accented word with accent on the final mora. Unaccented words are refered to as type 0; type 1 words have accent on the first mora, type 2 on the second, and so on. Thus HLL is type 1; LHLL is type 2, and LHH is type 3 or 0.

In fact, the pitch accent system of Japanese is yet more complex than the above description suggests, and there is a large and lively literature on the question of how best to capture its regularities (e.g., Haraguchi, 1977, 1988; McCawley, 1977; Pierrehumbert and Beckman, 1988; Sugito, 1982). Particular controversies concern, for instance, the behavior of accent in words beginning with a long vowel. However, none of the research reported in the present paper used these controversial cases, and the chief characteristic of the Tokyo Japanese accent system which is important for our study is unaffected by the phonological disputes. Words differ in pitch accent, and at least in CVCV-initial words the system described above implies that the first two morae cannot both be assigned the same pitch accent label. There are only two possible ways to label the initial two morae of such a word: HL- or LH-. Our research addresses the role of this word-initial accent distinction in the recognition of Tokyo Japanese.

Any distinction in word-initial position is potentially informative for our understanding of how human listeners recognize spoken words. Human word recognition is a highly efficient process. Relevant information about segment identity is exploited as soon as it becomes available, and it may become available as much as a whole syllable in advance (Martin and Bunnell, 1981, 1982). Coarticulatory information can lead to earlier identification of upcoming segments (Lahiri and Marslen-Wilson, 1991), and mismatching coarticulatory information can hamper recognition (Whalen, 1984, 1991; Marslen-Wilson and Warren, 1994; McQueen *et al.*, in press). We ask in this paper whether pitch-accent information in the initial portions of Japanese words can also be used in recognition as soon as it becomes available.

Psycholinguists usually conceive of the process by which spoken words are recognized as consisting of separable subparts. In the initial stage candidate words are activated by information in the signal; this process of activation is usually held to be bottom-up and not open to influence from higher levels of processing (Norris, 1994; Marslen-

[a)] Electronic mail: anne.cutler@mpi.nl
[b)] Electronic mail: otake@dokkyo.ac.jp

Wilson, 1987; but see McClelland and Elman, 1986). Activated word candidates then compete for recognition and a winning candidate emerges at this later competitive stage. Thus the input may activate many candidate words which are not actually present in the signal, and these will compete with and potentially slow the recognition of the actually spoken word (McQueen *et al.,* 1994); the more efficient the exploitation of different sources of information in the signal, therefore, the fewer such spurious competitors will be activated.

To be sure, listeners sometimes do not exploit potentially relevant information, particularly in the earliest stages of word recognition. The initial activation of word candidates in English does not appear to draw on stress information, for instance; word pairs which are distinguished by stress where this does not involve a vowel-quality difference, e.g., *FORbear* and *forBEAR*, are both initially activated irrespective of which one was spoken (Cutler, 1986). Since most stress differences in English in fact do involve vowel-quality differences—in the case of pairs such as *SUBject* and *subJECT*, or *REfuse* and *reFUSE*—it is apparently efficient enough for such vowel differences to be exploited in initial activation, enabling *SUBject* to be distinguished from *subJECT* in much the same way as *batter* from *better*, or *marine* from *maroon*, without additional exploitation of the suprasegmental cues which distinguish *FORbear* from *forBEAR*. Fewer than 20 minimal pairs in English are distinguished solely by suprasegmental structure, so that the failure to incorporate into the initial activation process any means of distinguishing them results in only a trivial increase in the already extensive number of homophonic word pairs in English (e.g., *match, angle, career*). Thus stress information may fail to be exploited in word recognition in English because it does not produce a significant and reliable effect on the number of activated candidate words.

In the present study we consider the question of whether initial pitch accent patterns play a role in the recognition of spoken words in Tokyo Japanese. The general efficiency of the word-recognition process, and all the evidence of early use of relevant information in the signal to constrain activation of candidate words, lead to the supposition that distinctive pitch-accent information in the initial portions of words may be exploited by listeners. And yet, it has been claimed that pitch accent is unimportant for recognition of Japanese utterances. Thus Shibata (1961, p. 19) writes: ''The reason why the dialectal differences [in accent] are so great, I believe, is that accent plays no very important role in communication;'' and Vance (1987, p. 107) maintains: ''There is little doubt that the functional load of accent distinctions in standard Japanese is very low... accent is probably the most difficult aspect of standard pronunciation for non-standard speakers to master, but incorrect accent patterns very seldom cause any confusion for listeners.''

One reason why this may be true is that pitch accent in Japanese certainly provides information other than that relevant to word recognition. One of the most salient characteristics of the pitch accent system is that it is dialectally variable. The two major dialect groups of Japan, Tokyo Japanese and Osaka Japanese, differ noticeably in pitch accent patterns. Thus pitch accent patterns constitute a major cue to a speaker's dialectal background, and listeners will be accustomed to exploiting pitch accent to gain such information about speakers. Pitch accent could, in consequence, be less important for word recognition due to the fact that it is usefully providing another sort of information. [''The primary importance of accent patterns is social rather than linguistic. Incorrect patterns mark a speaker as a nonnative of the Tokyo area'' (Vance, 1987, p. 107).] Note that Scott and Cutler (1984) showed that perceptual exploitation of a phonetic effect as a correlate of syntactic structure was not manifested by listeners for whom that same phonetic effect was a marker of sociolect.

Further, because pitch accent differences are signalled by $F0$, the information which they provide may only relatively slowly become available, so that activation of words may occur without reference to pitch accent information; Cutler and Chen (1997) showed that some tonal distinctions in Cantonese were perceptually available later than the segmental information distinguishing the vowels of the same syllable. A study by Walsh Dickey (1996) indeed suggests that pitch-accent processing is slow. In her experiment Japanese listeners were asked to make same–different judgments on pairs of CVCV words or nonwords; when members of a pair differed, it could be either on one of the four phonetic segments or in pitch accent. ''Different'' judgments were significantly slower for pairs differing in pitch accent than for pairs which differed segmentally, irrespective of the position of the segmental difference. Walsh Dickey argued that perception of the pitch accent could not be accurately determined until the second syllable since it could best be achieved by comparison of the two syllables. Note that Cutler and Chen's (1997) study of the perception of Cantonese tone also included a same–different judgment experiment, and also found that pairs differing in tone were judged more slowly than pairs differing in any segment.

However, the claim that pitch-accent information is not important to listeners in spoken-word recognition has hardly been put to direct experimental test. Nishinuma (1994; Nishinuma *et al.*, 1996) studied the classification of pitch-accent patterns by nonnative adult learners of Japanese; this task, like Walsh Dickey's (1996) same–different method, does not actually require word recognition. Otake *et al.* (1993) varied initial accent pattern in experiments in which listeners detected CV targets; responses to initial targets were equally rapid and accurate whether the word began with a HL- (e.g., *monaka*) or LH- (e.g., *kinori*) accent pattern, but, again, this tells us only that neither initial pattern causes listeners perceptual difficulty. In an earlier study we observed that Japanese listeners find cross-spliced words with a correct segmental sequence but an impossible accent pattern (one which could not occur in the language) hard to process (Otake *et al.*, 1996). The only study we could find which directly addressed the role of pitch accent in word recognition was by Minematsu and Hirose (1995). In two of the three experiments they report, native listeners were presented with misaccented speech. Misaccented words in isolation proved harder to recognize than their correctly accented counterparts; however, misaccenting had less effect in context than in isolation. Their other study used the gating task, in which words are presented in successively larger fragments. Minematsu and Hirose do not state the actual stimuli used in this experiment, only that they were four-mora words with accent types 0, 1, and 2. They found that HL- words were recognized on the basis of less information than LH- words. As the four-mora vocabulary contains only about 7.5% HL- words (NHK, 1985), this result suggests that listeners were effectively using accent to rule out candidate words: HL- portions rule out more competitors than LH- portions do.

Experiment 2 below also uses the gating task, but in a way specifically designed to assess the contribution of pitch accent in spoken-word recognition. Experiment 3 addresses the same issue via another standard psycholinguistic lexical-processing task: lexical decision. Before describing those experiments, however, we report an initial study in which we investigated the domain of available accentual information. Cutler (1986) argued that English stress, where it involves no segmental correlate in vowel quality, can hardly be computed without comparison across syllables—thus English listeners can only tell whether the syllable *for-* comes from *FORbear* or *forBEAR* when they hear the word's second syllable and compare the relative stress levels of the two syllables. Walsh Dickey (1996), as described above, made exactly the same claim about the perception of pitch-accent patterns. If such cross-syllable comparison is indeed necessary, it could reduce the relative usefulness of pitch accent in constraining lexical activation. In experiment 1, therefore, we used a two-choice classification procedure to ask: Does a single CV syllable extracted from either syllable of a bimoraic/bisyllabic word contain sufficient accentual information to enable the accent pattern of the whole word to be accurately identified by listeners?

I. EXPERIMENT 1

A. Materials

Thirty-two words were chosen, 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); in this and in the later experiments, accent assignment was checked against the Tokyo Japanese 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, nukaˆ, shika, hakaˆ, taka, hoka, kagiˆ, kago, kakeˆ, kaki, kamiˆ, karaˆ, kaseˆ, kaze*. The LH words marked with ˆ have final accent, the others are unaccented.

All words were recorded by three female speakers of Tokyo Japanese, who were naive as to the purpose of the experiment. The 96 resulting productions were digitized, using the ESPS speech editing system with WAVES+, and the *ka* syllables were extracted from each production. These 96 *ka* tokens were recorded, in random order, onto digital audio tape. Note that vowel-final short syllables produced in isolation are typically closed with a glottal stop, and this was the case in all 48 *ka*-final tokens; this glottal stop was included in the tokens on the tape.

The following nine acoustic measures were computed, using ESPS, for each syllable: minimum fundamental frequency ($F0$); maximum $F0$; $F0$ range; mean $F0$; standard deviation of $F0$; total syllable duration; vowel duration; mean rms amplitude; and standard deviation of rms amplitude. The $F0$ and amplitude measures were computed across the voiced portions of the signal only; aspiration following the /k/, and creak, if any, preceding the glottal stop, were not included in these measures, nor in the vowel duration measure.

B. Subjects

Twenty-four undergraduates of Dokkyo University participated in the experiment. All were native speakers of Japanese, from the Kanto area (Tokyo and environs, but excluding Ibaraki and Tochigi prefectures where dialectal differences from Tokyo Japanese can be observed in accent patterns). They received a small payment for participating.

C. Procedure

Subjects were tested individually or in pairs. They heard the tape containing the *ka* tokens from a JVC Victor DAT player over Audio-Technical ATH-A9 headphones, and were required to choose for each token between two words from which it might have come (e.g., *kage* HL versus *kagi* LH; *baka* HL versus *gaka* LH). These choices were written on the response sheet, in both kanji and hiragana orthography, and the subjects circled their choice for each token. Note that subjects were never asked to decide whether a syllable was word-initial or word-final; each choice was 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, minimizing the possibility that coarticulatory information adjacent to the *ka* boundary could 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 keeping the same order.

D. Results

1. Perceptual judgments

The overall correct response rate was high (74%). Responses were more accurate for H (87%) than L syllables (61%; $F1[1,23]=72.75$, $p<0.001$; $F2[1,84]=97.63$, $p<0.001$), and for initial (80%) than final syllables (68%; $F1[1,23]=23.92$, $p<0.001$; $F2[1,84]=18.41$, $p<0.001$). There was no significant difference in response rate to final H syllables which were accented (80% correct) versus unaccented (85%).

There was, however, a significant effect of speaker, with speaker 1 receiving lower correct-identification scores (64%) than speakers 2 and 3 (78%, 79%; $F1[2,46]=17.51$, p

TABLE I. Mean values on eight acoustic measures (note: the ninth measure referred to in the text, $F0$ range, is the minimum-maximum $F0$ difference), and mean percent correct responses, for *H* versus *L ka* syllables in initial versus final position, for each speaker (S1, S2, S3).

	Minimum	Maximum	Mean	s. d.			Total	Vowel	Percent
	$F0$	$F0$	$F0$	$F0$	Mean rms	s.d. rms	duration	duration	correct
	(Hz)	(Hz)	(Hz)	(Hz)	amplitude	amplitude	(s)	(s)	responses
Initial syllables									
H									
S1	270	288	282	5.9	1326	274	1.34	0.83	96.4
S2	223	248	237	7.8	1004	283	1.29	0.87	83.9
S3	233	262	254	9.7	931	218	1.26	0.76	90.6
Mean	242	266	258	7.8	1087	258	1.30	0.82	90.3
L									
S1	200	231	215	8.4	1000	298	1.57	1.00	47.9
S2	176	209	188	10.5	694	169	1.31	0.76	80.2
S3	164	195	181	10.9	647	193	1.47	0.75	79.2
Mean	180	212	195	10.0	780	220	1.45	0.84	69.1
Final syllables									
H									
S1	221	254	239	9.0	1214	373	1.36	0.98	82.8
S2	184	212	193	8.0	882	143	1.34	0.88	86.5
S3	186	216	200	7.8	716	229	1.84	1.41	82.8
Mean	197	227	211	8.3	937	248	1.51	1.09	84.0
L									
S1	183	257	210	22.6	1003	353	1.31	0.84	30.2
S2	138	189	162	18.2	799	234	1.11	0.62	64.1
S3	159	241	187	24.0	569	312	1.88	1.52	63.0
Mean	160	229	186	21.6	790	299	1.43	0.99	52.4

<0.001; $F2[2,84]=13.02$, $p<0.001$). An analysis of the results excluding speaker 1 revealed that both the main effect of H/L (H 86%, L 72%) and the main effect of position (initial 84%, final 74%) remained statistically significant across the productions of speakers 2 and 3 ($F1[1,23]=18.24$, $p<0.001$, $F2[1,56]=20.01$, $p<0.001$ for H/L, $F1[1,23]=11.83$, $p<0.005$, $F2[1,56]=8.57$, $p<0.005$, for position).

Fifteen of the 96 items received scores below chance; all were L syllables mistakenly judged by the majority of subjects as H. Eleven of those were spoken by speaker 1. Of the fifteen items, eight had scores significantly below chance (9/24 or less); six of these (*ka* from *kago, baka, naka, buka, deka, yoka*) were spoken by speaker 1, and five of these were final L syllables. Thus this speaker (who, as shown below, had a notably high voice) systematically failed to produce clearly final-L syllables (not one of her eight final-L items was identified with accuracy significantly above chance). The other low-scoring items were from *kami, kasa, kaze, kase, kika* (speaker 1), *naka, deka, yoka* (speaker 2), and *yoka* (speaker 3).

Subjects scored less well in the first quartile of the experiment (66% correct) than in the following quartiles (76%, 75%, 79%); an analysis of variance showed a significant effect of quartile ($F1[3,59]=10.07$, $p<0.001$) and *t*-tests showed performance in the first quartile to be significantly worse than in each of the later quartiles, which did not significantly differ.

2. Acoustic analyses

Table I shows the mean value on each of the nine measures, separately for the four syllable types, for each speaker and averaged across speakers. Analyses of variance across the tokens were computed for each measure. The main focus of interest here is where acoustic differences between H and L syllables are to be observed, since the listeners' task in this experiment was in effect the H/L categorization.

Pitch: The five measures which we made of the pitch characteristics of the syllables revealed a simple and consistent pattern. The minimum, maximum, and mean $F0$ values for the syllables tended to pattern together: if one of these measures showed a significant difference between H syllables and L syllables, so did the others. Likewise, the two remaining measures, $F0$ range and standard deviation of $F0$ (both of which provide crude estimates of the amount of pitch movement across a syllable), also pattern together, and separately from the other set.

The minimum, maximum, and (therefore also the) mean $F0$ were all significantly higher in H syllables than in L syllables ($F0$ min: $F[1,28]=259.33$, $p<0.001$; $F0$ max: $F[1,28]=56.43$, $p<0.001$; $F0$ mean: $F[1,28]=310.78$, $p<0.001$), and were also significantly higher in initial than in final syllables ($F0$ min: $F[1,28]=107.75$, $p<0.001$; $F0$ max: $F[1,28]=9.08$, $p<0.01$; $F0$ mean: $F[1,28]=126.45$, $p<0.001$). On each measure there was also a significant interaction between H/L and position, whereby the H/L difference was greater in initial than in final syllables ($F0$ min: $F[1,28]=16.28$, $p<0.001$; $F0$ max: $F[1,28]$

$=64.34$, $p<0.001$; $F0$ mean: $F[1,28]=58.92$, $p<0.001$).

All three of these measures also showed a significant effect of speaker ($F0$ min: $F[2,56]=79.23$, $p<0.001$; $F0$ max: $F[2,56]=48.53$, $p<0.001$; $F0$ mean: $F[2,56]=104.49$, $p<0.001$). The source of this effect was that speaker 1 had a significantly higher voice, approximately 35 Hz higher on each $F0$ measure, than the other two. An analysis of the results for only the syllables of speakers 2 and 3 showed that all the main effects and interactions remained significant as reported above (for H/L: $F0$ min: $F[1,28]=121.39$, $p<0.001$; $F0$ max: $F[1,28]=22.34$, $p<0.001$; $F0$mean: $F[1,28]=128.6$, $p<0.001$; for position: ($F0$ min: $F[1,28]=54.15$, $p<0.001$; $F0$ max: $F[1,28]=6.57$, $p<0.02$; $F0$ mean: $F[1,28]=66.41$, $p<0.001$; for the interaction: ($F0$ min: $F[1,28]=6.35$, $p<0.02$; $F0$ max: $F[1,28]=24.33$, $p<0.001$; $F0$ mean: $F[1,28]=27.23$, $p<0.001$).

Both the $F0$ range and the standard deviation of $F0$ were significantly greater for L than for H syllables ($F0$ range: $F[1,28]=52.12$, $p<0.001$; $F0$sd: $F[1,28]=59.61$, $p<0.001$), and significantly greater in final than in initial syllables ($F0$ range: $F[1,28]=43.82$, $p<0.001$; $F0$sd: $F[1,28]=36.37$, $p<0.001$). The interaction was also significant (greater H/L differences in final than in initial syllables; $F0$ range: $F[1,28]=22.99$, $p<0.001$; $F0$sd: $F[1,28]=30.58$, $p<0.001$). On neither of these two measures was there a significant effect of speaker.

Duration: Neither durational measure showed significant H/L differences. Final syllables were longer than initial (overall: $F2[1,28]=4.9$, $p<0.05$; vowel: $F2[1,28]=29.8$, $p<0.001$).

Amplitude: H syllables had significantly greater mean amplitude than L syllables ($F[1,28]=10.85$, $p<0.005$). There was no difference between initial and final syllables, or interaction between H/L and syllable position. The standard deviation of amplitude showed no main effect of either variable. However, there was again an effect of speaker on both amplitude measures (mean: $F[2,56]=64.61$, $p<0.001$; s.d.: $F[2,56]=31.92$, $p<0.001$), and again, this was due to deviance of the productions of speaker 1, who spoke significantly louder than the others.

3. Correlations

To obtain a uniform measure of listeners' performances, the responses were converted to percentage H judgments, that is, the percentage of correct responses for syllables which actually were H, and the percentage of error responses for those which actually were L. Correlation coefficients were then computed across mean H responses per item and the acoustic measures obtained for each item.

Over all 96 tokens, there were significant positive correlations between H responses and four of the nine acoustic measures: subjects were more likely to decide that a syllable was H when it had high minimum $F0$ ($r[95]=0.66$, $p<0.001$), high maximum $F0$ ($r[95]=0.52$, $p<0.001$), high mean $F0$ ($r[95]=0.67$, $p<0.001$), and high mean amplitude ($r[95]=0.38$, $p<0.001$). There were significant negative correlations with two other measures: subjects were more likely to decide that a syllable was H when it had low $F0$ range ($r[95]=-0.32$, $p<0.002$) and low $F0$ standard deviation ($r[95]=-0.38$, $p<0.001$). These correlations suggest that high absolute $F0$ and high amplitude signaled a H syllable; pitch movement signaled a L syllable.

Responses to initial syllables showed the same pattern of relationship to $F0$ and amplitude as displayed in the overall correlations ($F0$ min: $r[47]=0.86$, $p<0.001$; $F0$ max: $r[47]=0.89$, $p<0.001$; $F0$ mean: $r[47]=0.91$, $p<0.001$; $F0$ range: $r[47]=-0.30$, $p<0.04$; $F0$ sd: $r[47]=-0.29$, $p<0.05$; rms-mean: $r[47]=0.50$, $p<0.001$), while only four of the six significant correlations in the overall analysis were significant for final syllables ($F0$ min: $r[47]=0.66$, $p<0.001$; $F0$ mean: $r[47]=0.52$, $p<0.001$; $F0$ range: $r[47]=-0.55$, $p<0.001$; $F0$ sd: $r[47]=-0.66$, $p<0.001$). The pattern of correlation was furthermore not the same for each speaker. Responses to all three speakers' productions correlated in the same way with the $F0$ measures, but only the responses to the productions of speaker 2 showed a statistically significant relationship to amplitude.

Nor was the pattern the same for H versus L syllables separately. The likelihood of H responses to syllables which actually were H correlated only with the maximum and the mean $F0$, and only relatively weakly: $F0$ max: $r[47]=0.29$, $p<0.05$; $F0$ mean: $r[47]=0.32$, $p<0.05$. In contrast, the likelihood of H responses to syllables which actually were L correlated with minimum $F0$ ($r[47]=0.37$, $p<0.01$), with maximum $F0$ ($r[47]=0.43$, $p<0.002$) and with mean $F0$ ($r[47]=0.44$, $p<0.002$), plus a marginal correlation with mean amplitude ($r[47]=0.25$, $p<0.09$).

E. Discussion

It is clear that Japanese listeners can determine with a high degree of success from which of two accentually different bisyllabic words a single syllable has been extracted. Overall, there was a higher percentage of correct responses for H than for L syllables, and there were somewhat lower correlations of responses to H syllables with acoustic factors; these two aspects of the results may reflect a bias towards treating the single syllables as monomoraic isolates marked H [note that of Japanese monomoraic words, 70% have type 1 accent (NHK, 1985)].

We expected that listeners' judgments would principally be based on $F0$ values, and the pattern of correlations is certainly consistent with such an interpretation: syllables with high absolute $F0$ were judged H, syllables with $F0$ movement were judged L. Listeners can also make some use of the amplitude. Durational factors seem to play little role in signaling whether a syllable is H or L.

However, not all speakers are equally successful at conveying the H/L difference. Our speaker 1 produced these two syllable types in a less differentiated way than speakers 2 and 3, and correspondingly she received a lower mean percentage of correct responses from the listeners. Recall also that scores were lower at the beginning of the experiment than at the end, i.e., listeners seemed to have been learning the task. It could be that part of this involved learning about the characteristics of the particular speakers' voices. Speaker 1 spoke with a higher-pitched voice than the other two speakers did,

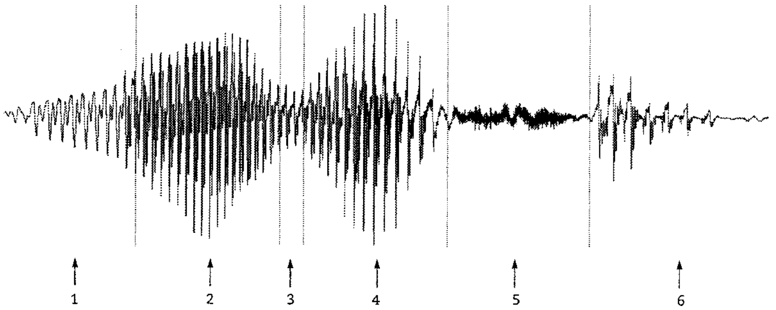

FIG. 1. How the gating fragments were prepared in experiment 2, illustrated on the word *nagasa*. The boundaries of the six phonetic segments were marked, then a marker was placed at the midpoint of each segment. Fragment 1 included the carrier plus the word up to the midpoint 1, fragment 2 the carrier plus the word up to the midpoint 2, and so on. ''Fragment'' 7 consisted of the carrier plus the entire word.

and this too, may have influenced the success with which her productions were judged; the response rates may have been more uniform had all speakers had comparable $F0$. Certainly the inconsistency among speakers which we observed suggests that listeners cannot rely on clear information being immediately available from all speakers.

Perhaps the most interesting aspect of the present results, however, is that cues to the H/L distinction are better conveyed in initial than in final syllables. The acoustic measures showed greater H/L differentiation in initial than in final syllables; the overall percentage correct was higher for initial than for final syllables; and the correlations between responses and acoustic factors were stronger in initial than in final syllables. This suggests that pitch accent information may already be available to listeners in just the position where it would be of most use to them in distinguishing spoken words. The question of whether listeners are, in fact, able to exploit the available cues to this purpose is addressed in experiment 2 via a gating task, in which we can examine the lexical hypotheses which listeners entertain when they are presented with fragmentary information about the initial portions of a word.

II. EXPERIMENT 2

A. Materials

Twenty-four pairs of Japanese words were selected. Within each pair, the two words began with the same initial bimoraic CVCV sequence, again with V always short, but differed in segmental structure from the fifth segment (third mora) on. The accent pattern of the two words also differed; in one word the initial CVCV sequence was HL, in the other LH. In this way we made sure that the initial segmental information alone could not determine listeners' word guesses. Thus *nagasa* and *nagashi* formed a pair; both begin *naga-*; the accent pattern of *nagasa* is HLL, while *nagashi* is LHHˆ. There were 22 pairs with three morae, and 2 with four morae; no words contained moraic nasals, geminate consonants, or long vowels. The complete set of pairs (in HL-/LH- order) was: *bakufu/bakuchi, hanabi/hanawa, hokubu/hokuro, kamotsu/kamome, karafuru/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/sekiyu, tachiba/tachiki, tomato/tomari, wakaba/wakate, wakame/wakare, warabi/waraji, yomichi/yomise*. All but four LH- words (*nagashi* LHHˆ, *nomiya* LHL, *karudera* LHHHˆ, *wakare* LHHˆ) were unaccented.

A further 24 words were selected to serve as practice and warmup items. Some of these fillers contained moraic nasals, geminate consonants, or long vowels. Twelve were three-mora words (eight LHH, four HLL), and twelve four-mora (six LHHH, four HLLL, two LHHL).

All words were recorded by a male native speaker of Tokyo Japanese, in a short carrier phrase *Sore wa*... (''It is...''). The speaker avoided fully devoicing potentially devoiced vowels in the first two morae of the words, so that $F0$ measures could be undertaken. A gated version of each word was made, in which the word was presented in increasingly large fragments. The carrier phrase was always included [as preceding context greatly facilitates recognition of the segmental identity of very short fragments of Japanese speech (Kuwabara, 1982)], and the word fragments incremented by phoneme transitions. To achieve this, the segments of each word were labeled such that the portion of the signal carrying information about each phoneme was demarcated as closely as could be ascertained; this was achieved by a combination of visual inspection of the waveform and auditory judgment. A marker was then placed at, as near as could be determined, the midpoint of each such demarcated region. Each additional fragment then added a portion of the word up to the next marker. (Most cuts were made on a zero crossing; otherwise, the offset of the signal of a fragment was ramped to avoid abrupt amplitude changes which might lead to the perception of illusory clicks. The ramping was achieved by multiplying the fragment's final frame with a mask consisting of a linear ramp from 1.0 to 0.0.) Thus the word *nagasa* was presented in seven fragments; fragment 1 contained the carrier plus transition into the initial phoneme, fragment 2 continued into the first vowel, fragment 3 into the second consonant, fragment 4 into the second vowel, fragment 5 into the third consonant, fragment 6 into the third vowel, and fragment 7 contained the whole word. The advantage of this procedure, over a procedure in which successive fragments are incremented by a constant temporal interval, is that each fragment is guaranteed to contain more relevant phonetic information than the preceding fragment, and that the (perceptually informative) transitions from one segment to the next are minimally disrupted. Figure 1 illustrates the gating procedure.

Two experimental tapes were made, 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.

B. Subjects

Thirty-six undergraduate members of Dokkyo University participated in the experiment, in return for a small payment. All were native speakers of Tokyo Japanese from the Tokyo metropolitan area or Kanegawa, Saitama, or Chiba prefecture. None had taken part in experiment 1. Eighteen participants heard each experimental tape.

C. Procedure

The listeners were tested individually or in groups of two to five. The words were again presented over headphones from a DAT player; the tape was stopped after presentation of each fragment to allow time for the listener to record a guess as to the word's identity, along with a confidence rating for that guess. The guesses were written on a response form in normal Japanese orthography, and the confidence ratings were recorded by circling a number on a scale of 1 to 5, with 1 representing no confidence and 5 representing complete certainty.

An important determinant of guessing responses in a task such as this is word frequency, or familiarity. Thus we wished to compare the relative familiarity of the actual target words and the words guessed by the listeners. However, we could find no published frequency norms for Japanese containing all the relevant words. Accordingly we collected relative familiarity judgments for all guessed words and targets from a separate group of 45 subjects, none of whom had participated in the listening tasks. These subjects judged for 1033 pairs of items (1033 separate word guesses collected in the experiment below, with in each case the actual word that was being presented when the guess was produced) which member of the pair was the more familiar word to them. The average ratings computed across subjects for each item pair allowed us to make the requisite comparisons.

When the gating task is used to study the word recognition process, three dependent variables may be evaluated (Grosjean, 1996): the point at which the spoken word is definitively recognized, the confidence ratings assigned to correct guesses as a function of amount of information available, and the nature of the candidate words proposed at each point in the stimulus presentation. In the present study, the recognition of the spoken word was not the focus of interest; the first of these dependent variables was therefore not relevant. Instead, we used the task to assess listeners' recognition of accent pattern; in particular, we wished to know whether listeners made effective 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. This question is most directly addressed by analyzing the candidate word guesses produced by listeners at fragments 1, 2, 3, and 4, and in particular by comparing the accent pattern of these candidate words for target words beginning HL- and LH-. From fragment 5 onwards, segmental information could distinguish the members of the word pairs; for the first four fragments, however, the only distinguishing information was accentual.

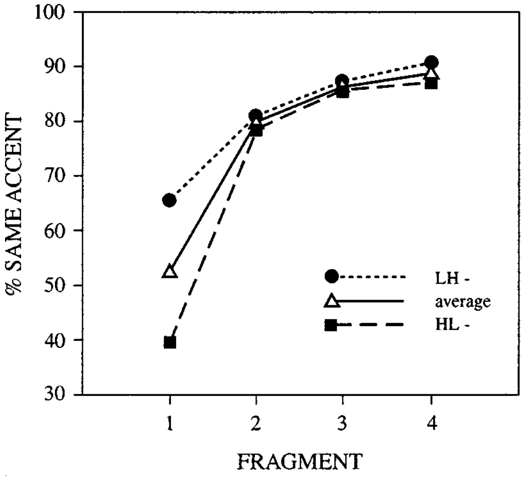

FIG. 2. Experiment 2: Proportion of guessed words with same initial accent pattern as the spoken word.

Acoustic analyses of the initial bimoraic portions were carried out in the same manner as in experiment 1, in order to explore what cues listeners would use to guide their word guesses (in particular, should there prove to be considerable interitem variability in the proportion of accentually correct word guesses).

The confidence ratings of accentually correct versus incorrect guesses were also compared, as well as the rated familiarity of accentually correct versus incorrect guesses in comparison to the target word.

D. Results

1. Accent recognition

The word guesses were scored by hand and the accent pattern of each guess ascertained. Since the initial bimoraic (segmentally ambiguous) portion was the crucial focus of interest, only the corresponding initial portion of the accent pattern of each guess was considered. In effect, this resulted in a two-way classification of alternative ''initial accent patterns:'' type 1 (HL-) versus all other (LH-, including types 0, 2, and 3) patterns. Thus for *nagashi* LHHˆ, guesses such as *nagai* LHL and *namae* LHH (unaccented) were scored as accentually correct. Figure 2 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.

At fragment 1, which contained information only about the initial consonant of the word, 52.66% of guessed words had a correct initial accent pattern. This number was not significantly different from chance, which is 50% for this two-way classification ($z=1.53$, $p>0.05$). It can be seen that in fact LH- words produce more guesses with the correct accent pattern than HL- words at this point. This is presumably the expected lexical type frequency effect: the vocabulary contains approximately 60% LH- to 40% HL- words (NHK, 1985).

At fragment 2, which contained information as to the initial CV mora, 79.63% of guessed words had a correct initial accent pattern, and this was significantly higher than would be predicted by chance ($z = 17.38$, $p < 0.001$). It is thus obviously the case that as soon as any vocalic information was available, subjects were able to use it to extract accent information. Note that virtually never were the word guesses which subjects produced on the basis of such minimal information actually correct; but they did manifest the correct initial accent pattern. Thus the 18 listeners given *nagasa* HLL guessed for the second fragment (*na-*) 16 different words, all different from *nagasa: nabe nagashi naifu naito naka nakai nama namida nanzan napukin Nara Narita Naruse NASA nasu Natoo*. The initial accent pattern of 14 of these guesses is HL-; only two (*nagashi nakai*) begin LH-. It is clear from this list that segmental information was well perceived; indeed, for fragment 2, for example, 94.09% of guesses began with the correct consonant and 98.27% had the correct vowel, both types of segmental information being significantly better represented in the guessed words than the accentual information ($t[47] = 5.31$ for consonants, 8.65 for vowels, both $p < 0.001$).

The proportion of guessed words with correct accent pattern continued to show further small increments across fragments 3 and 4: 86.57% and 88.88% correct, respectively. The arrival of distinctive segmental information in fragment 5 considerably narrowed the range of subjects' guesses, and nearly all words were recognised by all listeners by the sixth fragment. Note that some pairs, e.g., *karuteru karudera*, involved more fragments than in Fig. 1.

2. Acoustic analyses and correlations

The analysis of the $F0$ characteristics of the initial bimoraic portions of the words showed, as in experiment 1, a significant difference between H and L syllables. The H syllables in word-initial position had a mean $F0$ of 190.7 Hz with a standard deviation of 15.1 Hz, while L word-initial syllables had a mean $F0$ of 116.3 Hz with a standard deviation of 6.8 Hz (recall that the speaker in this case was a male). The average minimum, maximum, and range measures were also significantly higher for H than L syllables ($F0$ min: $t[46] = 7.79$, $p < 0.001$; $F0$ max: $t[46] = 24.43$, $p < 0.001$; $F0$ range: $t[46] = 3.3$, $p < 0.002$; $F0$ mean: $t[46] = 18.53$, $p < 0.001$; $F0$ sd: $t[46] = 3.89$, $p < 0.001$).

There was, however, no significant correlation between acoustic measures and the accent patterns of guesses at fragment 2, at which these differences in the first syllable are first available. This probably reflects the fact that the proportion of responses with correct accent pattern was already so high that there was no scope for interitem variation through which effects of the acoustic information could be observed. One correlation with the fragment 3 responses reached our criterion (0.05) of statistical significance: for words beginning LH-, the greater the standard deviation of $F0$ across syllable 1, the higher the proportion of responses with LH-accent patterns ($r[23] = 0.44$, $p < 0.04$). This is in line with the results of experiment 1 in which pitch movement was associated with L judgments to isolated syllables.

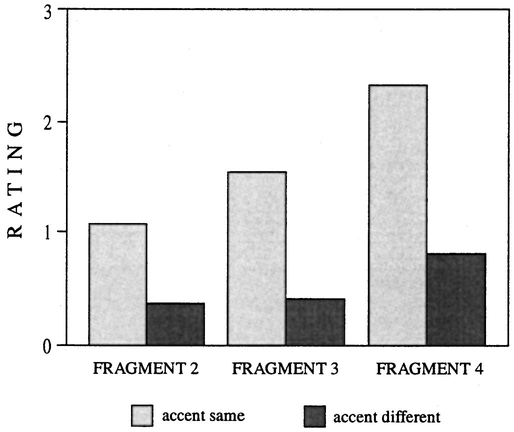

FIG. 3. Experiment 2: Confidence ratings for incorrect word guesses with same initial accent pattern as spoken word versus differing pattern.

3. Confidence ratings

Figure 3 shows listeners' confidence ratings for incorrect word guesses at fragments 2, 3, and 4. Although listeners were well aware that these fragments do not suffice to discriminate words, so that their confidence in their guesses was in general very low, they were nevertheless significantly more confident in guesses with the same initial accent pattern as the spoken word (mean rating 1.67) than in guesses with different accent (mean 0.5; $t[95] = 7.71$, $p < 0.001$).

4. Familiarity

The relative familiarity judgments for the guessed words and targets were analyzed. Here 63% of guesses were rated higher in familiarity than their targets. However, the strength of this familiarity effect was not significantly different for words which had the same initial accent pattern as their targets (61.5% more familiar) than for words which had different accent than their targets (65% more familiar).

E. Discussion

Experiment 2 has shown that even partial presentation of the first vowel of a word is sufficient for listeners to ascertain the initial accent pattern of a spoken Japanese word in nonconstraining context, and to use this information to restrict the word candidates which they consider possible continuations of this fragment.

In our third experiment we use a response-time measure to address the question of whether activation of word candidates in spoken-word recognition is constrained by pitch-accent information. The task which we use is auditory lexical decision (Goldinger, 1996), in which listeners' response time to decide whether a spoken string is a real word is measured. In this task, repeated presentation of a word leads to accelerated responses on the second presentation (''repetition priming''). As described in the introduction, minimal stress pairs in English (*FORbear–forBEAR*) are both activated when either one is heard. Minimal pitch-accent pairs in Japanese also exist, such as *ame* which with HL accent means ''rain'' and with LH accent ''candy.'' If either form activates both lexical representations, as in English, then presentation of one member of a pair (e.g., *ame* HL) should produce repetition priming for a subsequent presentation of the other member (e.g., *ame* LH). However, if pitch-accent information fully constrains lexical activation, repetition priming should be affected only by the same word, not by its minimal pair. Experiment 3 tested this issue.

III. EXPERIMENT 3

A. Materials

Twenty-four minimal accent-pairs of bimoraic bisyllabic words were chosen. The complete set was: *ame, chiri, kaki, kiji, michi, mushi, sake, shiro, washi; aka, asa, ashi, hashi, ichi, ima, ishi, kame, kami, kata, mesu, seki, sumi, toshi, umi*. One member of each accent pair was initially accented (HL), while the other was either LH unaccented (the first 9 pairs in the list) or LH accented on the second mora (the remaining-15 pairs).

One hundred other words were chosen to serve as control and filler words. Twenty-four of these were bimoraic/bisyllabic and were used as control words; the remaining words could be two, three, or four morae in length, and some contained long vowels, moraic nasals, or geminate consonants. One hundred and eight nonwords were also constructed; these were constructed to resemble the filler words in phonological structure.

All words and nonwords were recorded by a female native speaker of Tokyo Japanese, and digitized. Six running orders (tapes) were constructed; each contained all filler words and all nonwords, in the same order. For each accent-pair, one member served as target on three tapes and the other on the other three tapes; each tape contained 12 HL and 12 LH experimental target words. Each set of three tapes on which a given target word occurred differed in the nature of the prime word which preceded the target; on one tape the prime was (a different token of) the same word as the target (e.g., *ame* HL was preceded by *ame* HL), on another tape it was the accent pair (e.g., *ame* HL was preceded by *ame* LH), and on the third tape it was the control word (e.g., *ame* HL was preceded by *eki* HL). The nature of the prime was counterbalanced across tapes for each experimental word accent pattern. The prime preceded the target immediately or with one, two, or three intervening items; this factor was also counterbalanced with the other variables. Each running order was copied to DAT tape and timing marks were placed at the onset of each item.

B. Subjects

Ninety undergraduate members of Dokkyo University participated in the experiment, some for course credit and some for a small payment. Again, all were native speakers of Tokyo Japanese from the Tokyo metropolitan area or Kanegawa, Saitama, or Chiba prefecture, and none had participated in experiments 1 or 2. Fifteen subjects heard each of the six running orders.

C. Procedure

Subjects were tested individually or in pairs. The stimuli were presented over headphones from a DAT recorder as before. Subjects were seated in front of two response keys in a quiet room. They were instructed to decide for each item as quickly as possible whether or not it was a real word of Japanese, and to press the YES response key to signify a positive decision, the NO response key for a negative decision. Only YES responses were recorded. Timing and data collection were controlled by a Toshiba computer running the NESU experimental control software.

TABLE II. Experiment 3: Mean response times (ms) to decide that target is a word, as a function of preceding presentation of same word, accent pair, or control word.

	Prime type		
	Same word	Accent pair	Control
HL words	751	795	781
LH words	761	821	781
Mean	756	808	781

D. Results and discussion

The response times were subjected to separate analyses of variance with subjects and with items as random factors. Missing responses were replaced by the mean for the relevant subject in the relevant condition. Miss rates were not analyzed because the proportion of missed data was very low (4.5% of the total including responses which were lost due to equipment malfunction as well as failures to respond and erroneous decisions). The mean RTs across items and subjects are presented in Table II.

The statistical analyses revealed a main effect of prime type ($F1[2,168]=14.78$, $p<0.001$; $F2[2,46]=10.04$, $p<0.001$). *Post hoc* analyses showed that decisions to target words when the target had been preceded by itself as prime were significantly faster than when it had been preceded by a control word ($t1[89]=2.27$, $p<0.03$; $t2[47]=2.25$, $p<0.03$) or when it had been preceded by its accent pair ($t1[89]=4.56$, $p<0.001$; $t2[47]=3.71$, $p<0.001$); the difference between the accent-pair prime condition and the control condition was significant across subjects but not across items.

There was no significant effect of the lag between prime and target, and no significant effect of the target word's pitch-accent pattern, nor did either of these factors interact with the prime-type-effect. There was also no significant effect of the tapes variable, which was included in the analysis by subjects. An additional unequal-N analysis across items compared the HL/LH-accented pairs with the HL/LH- unaccented pairs; this accent-type factor had no effect itself and did not interact with any other factor, including prime type.

The experiment thus revealed no facilitation of lexical-decision responses as a result of prior presentation of a word's minimal accent pair. Only prior presentation of the word itself produced repetition priming. This result indicates that HL/LH minimal accent pairs do not facilitate recognition of one another's lexical representations, which in turn suggests that pitch accent information constrains lexical activation.

IV. GENERAL DISCUSSION

Results from three experiments have clearly demonstrated that the pitch-accent information available in spoken Tokyo Japanese words can be, and is, exploited by listeners in the process of word recognition. In experiment 1 listeners were easily able to assign a syllable to one of two word choices on the basis of accentual structure. In experiment 2, listeners used accentual information to guide their guesses of completions of partial word fragments, and they were more confident in guesses with the same accentual structure as the word from which the fragment actually came in than in guesses with different accentual structure. In experiment 3, repetition priming for a spoken word was exercised by its own representation but not by that of its minimal accent pair. Accent information constrains the activation and selection of word candidates in the process of spoken-language recognition by human listeners.

This contradicts the suggestion, referred to in the Introduction, that Japanese pitch accent has little importance for word recognition, and instead confirms the suggestion, based on the earlier findings of Minematsu and Hirose (1995), that listeners actively use this information. It also demonstrates that the recognition of Japanese words is sensitive to nonsegmental information in a way that the recognition of English words apparently is not. As described in the introduction, lexical stress in English appears to play no role in word activation (Cutler, 1986). However, recent results demonstrating that lexical processing is constrained by stress in a language otherwise closely similar to English, namely Dutch (Koster and Cutler, 1997; van Donselaar and Cutler, 1997), suggest that the situation of English is rather unusual [probably because of the very strong correlation of English stress with vowel quality (see Cutler *et al.*, 1997)]. Thus Japanese is allied with many other languages in that word activation can draw on nonsegmental information. The correlations between listener performance and the various $F0$ measures which we report above in experiment 1 suggest that it is the suprasegmental factors of pitch level and pitch movement that listeners are drawing on to derive accentual constraints on word identity. The results of experiment 2 show that this information is exploited very early: short CV syllables, truncated midway through the vowel, provide sufficient cues to the distinction between words beginning HL- versus LH-. This is consistent with research on the perception of tone in Norwegian and in Chinese showing similarly that suprasegmental information can be exploited even in very short speech fragments. Thus Efremova *et al.* (1963) found that Norwegian listeners needed only part of a syllable to distinguish, in a forced-choice task, between two forms of the same verb differing in tonic accent (signaled in Norwegian via fundamental frequency variation). Similarly, Tseng (1990) found that tones on isolated Mandarin Chinese vowels could be correctly identified in fragments comprising only the initial 25% of the vowels; and even though the Cantonese listeners in Cutler and Chen's (1997) experiment responded more rapidly to segmental distinctions than to tonal distinctions, they still made use of the tonal information as early as it became available. In experiment 2, likewise, word-initial segmental structure was perceived even more accurately than accentual structure, but a very short fragment of speech still provided significant information about accent.

As we pointed out in the Introduction, little research has previously addressed the role of Japanese accent in the process of word recognition. Walsh Dickey's (1996) study used a same–different judgment task, and thus did not actually require lexical processing. Hirose *et al.* (1993) report, as well as the pilot version of work reported by Minematsu and Hirose (1995), a gating experiment in which synthesized versions of four-mora real words and nonwords with various accent patterns were presented to listeners; the smallest fragment included the boundary of the third and fourth morae, and the fragments increased in duration outwards from that point. The listeners in their experiment were not asked to name the words, however; they were required to identify the accentual pattern (and were able to do so, usually without needing to hear the entire word). Similarly, Nishinuma's (1994; Nishinuma *et al.*, 1996) experiments with non-native listeners required explicit identification of the accent pattern. In our experiments, however, the listeners never suggested that they were aware that our research was centered on the role of accentual information. All the tasks which we used involved decisions about words, and the participants in our experiments simply engaged, as instructed, in lexical processing. Insofar as our study can be compared directly with that of Minematsu and Hirose (1995), our results are in accord with theirs; their finding that words in isolation were significantly harder to identify when misaccented also suggests that accentual patterns constrain word activation.

The experimental materials in our study were deliberately confined to simple cases in which the manifestations of pitch accent could be easily observed. In experiment 1 we used only words with the structure CVCV in which all vowels were short; the overlapping parts of the experimental pairs in experiment 2 again had just this structure; and the minimal pairs of experiment 3 were either CVCV or VCV, once more with only short vowels. None of our experimental items contained nasal morae, geminates, or long vowels; in future work it will clearly be interesting to extend our investigations to these other phonological structures. In our experiments we also avoided as far as possible the occurrence of devoiced syllables, in which the manifestation of accent has been the subject of considerable attention (see e.g., Maekawa, 1990). Note, however, that by examining only the clear cases we have nevertheless produced a finding which can generalize to the majority of Japanese utterances. Of all the possible Japanese morae (defined in terms of separate IPA transcriptions) 60% in fact have the structure CV; Otake (1990) computed that CV morae accounted for over 70% of mora tokens in actual speech samples.

Limits on the generality of our findings may arise, of course, from other factors. As mentioned in the Introduction, accent patterns vary across dialect. Some dialects, in fact (especially those spoken in the northern part of Japan, such as the Ibaraki and Tochigi provinces), do not manifest accent variation. Speakers of these dialects would thus presumably not behave exactly as the listeners in the present study. Whether speakers of nonaccentual dialects would display sensitivity to accentual information in recognizing words in Tokyo Japanese and other accentual dialects is as yet an open question. In the present study we were careful to confine our materials, and the dialects spoken by our listeners, to one variety: Tokyo Japanese. We see no reason in principle why our results should not generalize to other accentual varieties, however.

A substantial minority of (especially) longer words [about 10% (Shibata, 1961)] can have more than one possible accent pattern (although a given speaker will tend to use only one of them, just as a speaker of English may choose between stress patterns as in *CONtroversy* versus *conTROversy*). Such factors could, again, reduce the value of accentual information for lexical access, making it (like, in fact, most information in the speech signal) not fully deterministic, but probabilistic in nature.

Additional reservations which must be maintained pending further investigations include the practical usefulness of accentual information in natural continuous-speech contexts. For instance, preceding context can cause assimilation of accent patterns, as when the accent of the initial mora of *kodomo* LHH ‘‘child’’ can be raised from L to H in *kono kodomo* LH HHH ‘‘this child’’ (Hattori, 1960). Pitch accent patterns of words can also interact with following speech context. In particular, the distinction between final-accented and unaccented words becomes realised in context, in that the two accent types exercise differing effects on a following phonologically weak element such as a particle: an accented final mora will force L accent on the following element but an unaccented final mora will not. This final-accented versus unaccented distinction plays a central role in the challenge which Pierrehumbert and Beckman’s (1988) autosegmental/metrical account of the pitch-accent system poses to earlier views. Pierrehumbert and Beckman support their model with extensive phonetic data, and there is phonetic evidence (e.g., Maekawa, 1990, 1995; Kubozono, 1993; Warner, 1997) supporting their analysis [but see Vance (1995) for phonetic support for the alternative view].

Our findings, however, crucially concern word-initial accent effects, and although there is disagreement on how these should be described—as successions of different markings (HL versus LH) or as different associations of tones (L%-HL-L versus L%-H-HL)—there is no controversy as to the existence of pitch-accent differences in the initial portions of Japanese words. Moreover, the robustness of the pitch-accent effects which we have observed in a variety of word processing tasks does not suggest exploitation of accentual information to be a resource with only very limited application in language recognition.

One as yet unanswered question concerns the fine detail of our general demonstration that listeners are able to exploit pitch-accent patterns in word recognition. We have in the present experiments addressed only the crudest distinction which it is possible to draw on the basis of initial pitch-accent patterns: a division between type 1 accent [accent on the first mora, comprising about 40% of lexical types (NHK, 1985)] and all other accent types (the remaining 60% of lexical types). The present results do not shed light on the question of whether listeners can continue to narrow their word-recognition choices on the basis of later-arising distinctions (LHH versus LHL; LHHH versus LHHL, etc.). Our supposition based in part on the conclusions drawn from research on other languages, e.g., the use of stress information in lexical processing in English versus Dutch, would be that pitch-accent information will be used to the extent that it exercises a useful degree of constraint on the population of potential word candidates. The crude bipartite division which is made possible by exploitation of the initial HL- versus LH- distinction is obviously a highly effective means of cohort reduction, but we suspect that the added value of later-arising distinctions may be very much less. Whether listeners can usefully distinguish final-accented versus unaccented words is also an empirical issue which remains to be addressed; Sugito (1998) argues that speakers do distinguish these accentual structures even in isolation, but our results, concerning as they do exclusively the initial portions of words, do not answer the question of whether listeners can exploit this distinction. (For instance, would minimal pairs with these two accent patterns prime one another’s representations? Consider *hashi*, one of our minimal pairs in experiment 3. We compared *hashi* HL ‘‘chopsticks’’ with *hashi* LH ‘‘bridge;’’ but there is also *hashi* LH unaccented ‘‘edge.’’ The results of experiment 3 suggest that *hashi* HL and *hashi* LH, do not activate one another’s representations. But is this also true of *hashi* LH, and *hashi* LH unaccented?)

Machine recognition of spoken Japanese, as recent research (Hirose, 1997; Hirose and Iwano, 1997) has established, can be rendered more efficient by explicit analysis of $F0$ contours and comparison of the result with stored information on the accentual patterns of words. This is exactly as would be expected, given that, as our series of experiments has shown, human listeners engaging in the processing of spoken words find it effective to do the same.

ACKNOWLEDGMENTS

This research was supported by the Human Frontier Scientific Program project ‘‘Processing Implications of Contrasting Language Phonologies,’’ by Grant No. 06610475 from the Japanese Ministry of Education, Science, Sports and Culture to the second author, and by a grant to the second author from the Max Planck Society. Both authors thank the Department of Linguistics, Ohio State University, for hospitality during the preparation of the paper. We further thank Mary Beckman for extensive discussions on the structure of Japanese accent, M. Sugito for permission to use the database of Japanese accent patterns (Sugito, 1995), Mary Beckman, Carlos Gussenhoven, and James McQueen for comments on the text, and Takaaki Nakazawa and Daisuke Ito (Dokkyo University), Kiyoko Yoneyama, Aki Arai, Yoko Kuroda, and Maki Yoshida (Ohio State University), and Mark Scholten, Peter van de Pol, Femke van de Weide, Mattijn Morren, Bram de Kruijff, and Daan Broeder (MPI) for technical assistance. The results of experiment 1 were reported in part to the 6th Australian International Conference on Speech Science and Technology, Perth, in December 1996, and the results of experiment 2 were reported in part to the 134th meeting of the Acoustical Society of America, San Diego, December 1997.

Cutler, A. (**1986**). ''Forbear is a homophone: lexical prosody does not constrain lexical access,'' Language and Speech **29**, 201–220.

Cutler, A., and Chen, H.-C. (**1997**). ''Lexical tone in Cantonese spoken-word processing,'' Percept. Psychophys. **59**, 165–179.

Cutler, A., Dahan, D., and van Donselaar, W. (**1997**). ''Prosody in the comprehension of spoken language: A literature review,'' Language and Speech **40**, 141–201.

Efremova, I. B., Fintoft, K., and Ormestad, H. (**1963**). ''Intelligibility of tonic accents,'' Phonetica **10**, 203–212.

Goldinger, S. (**1996**). ''Auditory lexical decision,'' Language and Cognitive Processes **11**, 559–567.

Grosjean, F. (**1996**). ''Gating,'' Language and Cognitive Processes **11**, 597–604.

Haraguchi, S. (**1977**). *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology* (Kaitakusha, Tokyo).

Haraguchi, S. (**1988**). ''Pitch accent and intonation in Japanese,'' in *Autosegmental Studies on Pitch Accent*, edited by H. van der Hulst and N. Smith (Foris, Dordrecht), pp. 123–150.

Hattori, S. (**1960**). ''On'inron kara mita nihongo no akusento (Japanese accent from the point of view of phonology),'' in *Gengogaku no Hoohoo*, by S. Hattori (Iwanami, Tokyo), pp. 240–272.

Hirose, K. (**1997**). ''Disambiguating recognition results by prosodic features,'' in *Computing Prosody*, edited by Y. Sagisaka, N. Campbell, and N. Higuchi (Springer, Heidelberg), pp. 327–342.

Hirose, K., and Iwano, K. (**1997**). ''A method of representing fundamental frequency contours of Japanese using statistical models of moraic transition,'' in Proceedings of EUROSPEECH 97, Rhodes, pp. 311–314.

Hirose, K., Minematsu, N., and Ito, M. (**1993**). ''Experimental study on the role of prosodic features in the human process of spoken word perception,'' in Proceedings of the ESCA Workshop on Prosody, pp. 200–203.

Koster, M., and Cutler, A. (**1997**). ''Segmental and suprasegmental contributions to spoken-word recognition in Dutch,'' in Proceedings of EUROSPEECH 97, Rhodes, pp. 2167–2170.

Kubozono, H. (**1993**). *The Organization of Japanese Prosody* (Kurosio, Tokyo).

Kuwabara, H. (**1982**). ''Perception of CV-syllables isolated from Japanese connected speech,'' Language and Speech **25**, 175–183.

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

Maekawa, K. (**1990**). ''Production and perception of the accent in the consecutively devoiced syllables in Tokyo Japanese,'' in Proceedings of the First International Conference on Spoken Language Processing, Kobe, Vol. 1, pp. 517–520.

Maekawa, K. (**1995**). ''Is there 'dephrasing' of the accentual phrase in Japanese?'' in *Papers from the Linguistics Laboratory* (Working Papers in Linguistics, Ohio State University), Vol. 44, pp. 146–165.

Marslen-Wilson, W. D. (**1987**). ''Parallel processing in spoken word recognition,'' Cognition **25**, 17–102.

Marslen-Wilson, W., and Warren, P. (**1994**). ''Levels of perceptual representation and process in lexical access: words, phonemes, and features,'' Psychol. Rev. **101**, 653–675.

Martin, J. G., and Bunnell, H. T. (**1981**). ''Perception of anticipatory coarticulation effects,'' J. Acoust. Soc. Am. **69**, 559–567.

Martin, J. G., and Bunnell, H. T. (**1982**). ''Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences,'' J. Exp. Psychol. Hum. Percept. Perf. **8**, 473–488.

McCawley, J. (**1977**). ''Accent in Japanese,'' in *Studies in Stress and Accent*, edited by L. Hyman (Univ. of Southern California, Los Angeles), pp. 261–302.

McClelland, J. L., and Elman, J. L. (**1986**). ''The TRACE model of speech perception,'' Cogn. Psychol. **18**, 1–86.

McQueen, J. M., Norris, D. G., and Cutler, A. (**1994**). ''Competition in spoken word recognition: Spotting words in other words,'' J. Exp. Psychol. Learn. Mem. Cog. **20**, 621–638.

McQueen, J. M., Norris, D. G., and Cutler, A. (**in press**). ''Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches,'' J. Exp. Psychol. Hum. Percept. Perf.

Minematsu, N., and Hirose, K. (**1995**). ''Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese,'' J. Acoust. Soc. Jpn. **16**, 311–320.

NHK (Nihon Hoosoo Kyookai) (**1985**). Nihongo Hatsuon Akusento Jiten (NHK's Dictionary of Japanese Accent and Pronunciation) (NHK, Tokyo).

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

Nishinuma, Y., Arai, M., and Ayusawa, T. (**1996**). ''Perception of tonal accent by Americans learning Japanese,'' in Proceedings of the Fourth International Conference on Spoken Language Processing, Philadelphia, Vol. 1, pp. 646–649.

Norris, D. G. (**1994**). ''Shortlist: A connectionist model of continuous speech recognition,'' Cognition **52**, 189–234.

Otake, T. (**1990**). ''Rhythmic structure of Japanese and syllable structure,'' IEICE Technical Report **89**, 55–61.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (**1993**). ''Mora or syllable? Speech segmentation in Japanese,'' Journal of Memory and Language **32**, 358–378.

Otake, T., Yoneyama, K., Cutler, A., and van der Lugt, A. (**1996**). ''The representation of Japanese moraic nasals,'' J. Acoust. Soc. Am. **100**, 3831–3842.

Pierrehumbert, J. B., and Beckman, M. E. (**1988**). *Japanese Tone Structure* (MIT, Cambridge, MA).

Scott, D. R., and Cutler, A. (**1984**). ''Segmental phonology and the perception of syntactic structure,'' Journal of Verbal Learning and Verbal Behavior **23**, 450–466.

Shibata, T. (**1961**). ''Nihongo no akusento'' (Japanese accent) Gengo Seikatsu **117**, 14–20.

Sugito, M. (**1982**). *Nihongo Akusento no Kenkyuu* (Studies on Japanese accent) (Sanseido, Tokyo).

Sugito, M. (**1995**). *Osaka-Tokyo Akusento Onsei Jiten* (Osaka-Tokyo Accent Pronunciation Dictionary), CD-ROM (Maruzen, Tokyo).

Sugito, M. (**1998**). ''Hana to hana no seisei to chikaku'' (Production and perception of *hana* and *hana*), in *Hana to Hana: Nihongo Onsei no Kenkyuu* (Hana and Hana: Studies on Japanese Speech) (Izumishoin, Osaka).

Tseng, C.-Y. (**1990**). *An Acoustic Phonetic Study on Tones in Mandarin Chinese* (Academia Sinica, Taipei).

van Donselaar, W., and Cutler, A. (**1997**). ''Exploitation of stress information in spoken-word recognition in Dutch,'' J. Acoust. Soc. Am. **102**, 3136(A).

Vance, T. J. (**1987**). *An Introduction to Japanese Phonology* (State Univ. of New York, Albany).

Vance, T. J. (**1995**). ''Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese,'' Journal of Phonetics **23**, 487–499.

Walsh Dickey, L. (**1996**). ''Limiting-domains in lexical access: Processing of lexical prosody,'' in *Linguistics in the Laboratory*, edited by M. Dickey and S. Tunstall, University of Massachusetts Occasional Papers in Linguistics (Univ. of Massachusetts, Amherst), Vol. 19, pp. 133–155.

Warner, N. (**1997**). ''Japanese final-accented and unaccented phrases,'' Journal of Phonetics **25**, 43–60.

Whalen, D. H. (**1984**). ''Subcategorical mismatches slow phonetic judgments,'' Percept. Psychophys. **35**, 49–64.

Whalen, D. H. (**1991**). ''Subcategorical phonetic mismatches and lexical access,'' Percept. Psychophys. **50**, 351–360.