Rhythmic Cues and Possible-Word Constraints in Japanese Speech Segmentation

James M. McQueen

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

Takashi Otake

Dokkyo University, Soka, Japan

and

Anne Cutler

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

In two word-spotting experiments, Japanese listeners detected Japanese words faster in vowel contexts (e.g., *agura*, to sit cross-legged, in *oagura*) than in consonant contexts (e.g., *tagura*). In the same experiments, however, listeners spotted words in vowel contexts (e.g., *saru*, monkey, in *sarua*) no faster than in moraic nasal contexts (e.g., *saruN*). In a third word-spotting experiment, words like *uni*, sea urchin, followed contexts consisting of a consonant-consonant-vowel mora (e.g., *gya*) plus either a moraic nasal (*gyaNuni*), a vowel (*gyaouni*) or a consonant (*gyabuni*). Listeners spotted words as easily in the first as in the second context (where in each case the target words were aligned with mora boundaries), but found it almost impossible to spot words in the third (where there was a single consonant, such as the [b] in *gyabuni*, between the beginning of the word and the nearest preceding mora boundary). Three control experiments confirmed that these effects reflected the relative ease of segmentation of the words from their contexts. We argue that the listeners showed sensitivity to the viability of sound sequences as possible Japanese words in the way that they parsed the speech into words. Since single consonants are not possible Japanese words, the listeners avoided lexical parses including single consonants and thus had difficulty recognizing words in the consonant contexts. Even though moraic nasals are also impossible words, they were not difficult segmentation contexts because, as with the vowel contexts, the mora boundaries between the contexts and the target words signaled likely word boundaries. Moraic rhythm appears to provide Japanese listeners with important segmentation cues. © 2001 Academic Press

Key Words: spoken word recognition; lexical segmentation; Japanese; mora.

Our understanding of human speech recognition has been advanced by experiments in Japan-

ese. Because the phonological structure of Japanese is different in several important respects from that of Indo-European languages such as English, experiments on spoken word recognition in Japanese offer a means to test and extend theories of speech perception in ways which would not be possible in Indo-European languages. A clear example of this can be found in the literature on lexical segmentation, the topic of this article.

A theory of segmentation developed in the late 1980s and early 1990s is that listeners use segmentation procedures which are based on the characteristic rhythm of their native language (Cutler & Mehler, 1993; Cutler, Mehler, Norris

& Seguí, 1986, 1992; Cutler & Norris, 1988). Because spoken language is continuous, spoken-word recognition entails segmentation of the speech stream into words. Language rhythm provides one means by which listeners can segment speech into discrete words. Evidence for this view came from the target-monitoring task (Mehler, Dommergues, Frauenfelder, & Seguí, 1981). French listeners detected consonant-vowel (CV) sequences like *ba* more rapidly in *balance*, where the target matches the first syllable of the word, than in *balcon*, where the target mismatches with the first syllable of the word (*bal*). When the targets were CVCs like *bal*, responses were faster to *balcon* than to *balance*. Cutler et al. (1986) found similar effects when French listeners performed target monitoring with similar materials in English. Results such as these led to the claim that French listeners use the syllable-based rhythm of French in speech segmentation.

Cutler et al. (1986) found, however, that English listeners did not show this sensitivity to syllable structure: English listeners did not perform in the same way as French listeners, neither with English nor French materials. But evidence from other tasks suggested that English listeners use the stress-based rhythm of English in speech segmentation (Cutler & Butterfield, 1992; Cutler & Norris, 1988). In a word-spotting task (Cutler & Norris, 1988), where listeners were required to spot real words embedded in nonsense words, English listeners found it harder to detect *mint*, for example, in *mintayf* (where the second syllable is strong) than in *mintef* (where the second syllable is weak). It was argued that a rhythmic segmentation procedure acts to segment a string like *mintayf* at the onset of the strong second syllable (before the [t]), making it hard to detect *mint*. But the procedure does not segment *mintef* because the second syllable is weak.

The conclusion which was drawn from these studies in French and English (and related studies in Spanish and Catalan by Pallier, Sebastián-Gallés, Felguera, Christophe, & Mehler, 1993; Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992; and in Dutch by Vroomen, van Zon, & de Gelder, 1996) was that, across the world's languages, there is a language-universal constraint on speech segmentation. This is that language rhythm is used as a segmentation cue. This constraint must of course vary from language to language depending on the characteristic rhythm of each particular language. The segmentation of syllable-timed French, Spanish, and Catalan appears to be based on the syllable; the segmentation of stress-timed English and Dutch appears to be based on strong syllables.

Japanese offered an ideal test of this proposal. Japanese rhythm is based neither on syllables nor on stress. Instead, it is based on the mora, a subsyllabic unit which can be one of five different types (see Vance, 1987, for further details): V (e.g., *a*), CV (e.g., *ta*), CCV (e.g., *gya*), a nasal coda consonant N (as in the second mora of *Honda*; *ho-N.-da*,[1] or the third mora of futon; *fu.-to-N*), and a geminate (doubled) consonant Q (e.g., the medial consonant of *Nippon* is a geminate *p*, which is part of both the coda of the first syllable and the onset of the second syllable: *ni-Q.-po-N*). It should be clear from this description that although morae are often complete syllables (e.g., the first syllable of *tanishi*, snail, is the CV mora *ta*, i.e., *ta.-ni.-shi*), there is no one-to-one correspondence between morae and syllables (e.g., the first syllable of *tanshi*, terminal, is made up of two morae, the CV mora *ta* plus the moraic nasal N, i.e., *ta-N.-shi*).

Morae form the basic rhythmic units in Japanese speech. One way in which a language's rhythm can be observed is in its verse forms. Thus, regularity in the number of syllables per line is important in French poetry, while regularity in the number of metrical stresses per line is important in English poetry. Mora-based rhythm is found in Japanese poetry: the haiku, for example, is a verse form which is defined as having three lines, consisting of five, seven and five morae, respectively. This convention reflects in part the explicit coding of morae in Japanese orthography: Each mora is coded by a single *kana* character.

Mora structure has strong influences on the acoustic-phonetic form of the speech signal. Some authors have gone so far as to argue that

---

[1] We will use a period to indicate a syllable boundary and a hyphen to indicate a mora boundary.

Japanese speech is strictly mora-timed. On one version of this view (e.g., Homma, 1981), speakers try to produce each mora in a given utterance with the same duration, and they tend to do so by lengthening or shortening the different segments within each mora to compensate for inherent differences in segment lengths. On more recent accounts, compensation also operates between segments in different morae such that the durations of larger units (e.g., words) tend to be predictable from the number of morae in those larger units (e.g., Han, 1994; Port, Dalby, & O'Dell, 1987). Although morae tend to be roughly of the same duration, there are significant exceptions to this rule (Beckman, 1982), and there is recent evidence from the analysis of spontaneous Japanese speech that speakers do not make compensatory adjustments in duration between morae within words (Warner & Arai, 2001).

Even if accounts of mora-based rhythm based on strict isochrony can thus be rejected, it remains the case that moraic structure has a central role to play in determining the rhythm of Japanese and that it is clearly coded in the speech signal. Mora structure is still the best predictor of word duration in spontaneous Japanese, at least for most speakers (Warner & Arai, 2001), and appears to determine the relative timing of vowels and consonants (Ramus, Nespor, & Mehler, 1999). Mora structure is thus coded in the signal primarily by durational differences. One striking durational difference which is particularly important to the following argument is that moraic nasal consonants tend to be considerably longer than nonmoraic nasal consonants (more than twice as long, according to the estimates of Sato, 1993, and of Otake, Yoneyama, Cutler, & van der Lugt, 1996). Otake et al. (1996), for example, found that moraic nasals which formed the medial morae in a set of 16 trimoraic words (e.g., the N in *tento*, tent, *te-N-to*) were, on average, 151.5 ms long, while nonmoraic nasal consonants which formed the onsets of the medial morae in a second set of 16 trimoraic words (e.g., the /n/ in *tenisu*, tennis, *te-ni-su*), were, on average, only 63.7 ms long.

Otake, Hatano, Cutler, and Mehler (1993) tested whether Japanese listeners use mora-rhythm to segment speech. Their predictions were that because Japanese belongs to a different rhythmic class than both French and English, Japanese listeners would show a pattern of performance which was different from that of both French and English listeners and that this pattern would reflect the rhythmic structure of Japanese. These predictions were confirmed. In a target-monitoring task, Japanese listeners found it equally easy to detect *ta*, for example, in *tanishi* and *tanshi*, presumably because *ta* is the first mora of both words. But they found it very much harder to detect *tan* in *tanishi* (with a miss rate over 60%) than in *tanshi* (with a miss rate under 10%), presumably because the mora structure of *tan* (CV-N) mismatches with the mora structure of the former word (CV-CV-CV) but not the latter (CV-N-CV). Results from phoneme detection experiments (Cutler & Otake, 1994; Otake et al., 1996) have confirmed that Japanese listeners are sensitive to the moraic structure of Japanese.

The study of Japanese thus strengthened the theory of metrical segmentation. But old habits die hard: our more recent research on lexical segmentation has reverted to the study of Indo-European languages (McQueen, 1998; McQueen, Norris & Cutler, 1994; Norris, McQueen, & Cutler, 1995; Norris, McQueen, Cutler, & Butterfield, 1997). Although these experiments were narrowly focused on only Germanic languages (English and Dutch), their scope was in another sense very broad. The goal of this research effort was to provide a more complete account of the process of spoken word segmentation and recognition in the form of an explicit computational model, the Shortlist model (Norris, 1994). An attempt was made to unify the many different results on segmentation in one processing model.

Central to Shortlist is the assumption that word recognition is based on competition between candidate words. The words most consistent with the acoustic-phonetic material in the speech input are activated, wherever they may begin, and enter a shortlist. They then compete with each other through lateral inhibitory connections. This competition process has strong empirical support (Gow & Gordon, 1995; McQueen et al., 1994;

Norris et al., 1995; Shillcock, 1990; Tabossi, Burani, & Scott, 1995; Vitevitch & Luce, 1999; Vroomen & de Gelder, 1995, 1997; Wallace, Stewart, & Malone, 1995; Wallace, Stewart, Shaffer, & Wilson, 1998; Wallace, Stewart, Sherman, & Mellor, 1995; Zwitserlood, 1989; Zwitserlood & Schriefers, 1995). The competition process alone provides a means by which the continuous input can be segmented into words. Words like *enjoy, below*, and *dull*, for example, will be activated along with the correct words given the input *enjoyable overindulgence*. But the incorrect words will lose in the competition process because together they cannot provide a complete parse of the input, with no material unaccounted for. Segmentation between *enjoyable* and *overindulgence* will occur as these candidates win the competition. This occurs without any other segmentation procedure operating.

How then can one incorporate metrical segmentation procedures in the Shortlist model? The solution which has been proposed (Norris et al., 1997) depends on two further facts. The first is the acknowledgment that lexical segmentation does not depend solely on competition and metrical structure. There are in fact many different cues which listeners can use to segment speech besides rhythmic structure: silence (Norris et al., 1997); allophonic cues, such as aspiration of word initial stops in English (Lehiste, 1960; Nakatani & Dukes, 1977); the duration of segments or syllables (Beckman & Edwards, 1990; Gow & Gordon, 1995; Klatt, 1974, 1975; Lehiste, 1972; Oller, 1973; Quené, 1992, 1993); phonotactics, including segment sequence constraints (McQueen, 1998) and vowel harmony constraints (Suomi, McQueen, & Cutler, 1997; Vroomen, Tuomainen, & de Gelder, 1998); fundamental frequency movement (Vroomen et al., 1998); and the sequential probabilities between segments (Saffran, Newport, & Aslin, 1996; van der Lugt, in press).

An important property of this multitude of cues is that they are unreliable. Not all word boundaries are marked with such cues, and the cues can be small, variable, and ambiguous (Klatt, 1976; Lehiste, 1972; Nakatani & Dukes, 1977; Quené, 1992). Nevertheless, the fact that listeners have been shown to be sensitive to all these cues suggests that when the cues are available in the speech signal, they will be used to assist in segmentation. The appropriate question to ask, therefore, is not how metrical segmentation procedures alone can be instantiated in Shortlist, but how all these cues can act to improve segmentation in a model of word recognition based on lexical competition.

The second fact on which the answer to this question depends is that listeners appear to be sensitive to the viability of stretches of speech as possible words in the speech stream (McQueen, 1998; McQueen & Cutler, 1998; Norris et al., 1997). Norris et al. (1997) showed, in two English word-spotting experiments, that listeners find it much harder to spot the real word in a nonsense word like *fapple* than to spot the real word in a nonsense word like *vuffapple*. This, they argued, was because the stretch of speech left over in *fapple* after *apple* has been found is the impossible English word *f*, while the stretch of speech left over in *vuffapple* is the possible (but nonexisting) word *vuff*. Similar findings have been obtained with Dutch listeners, who found it much harder to spot *lepel* (spoon), for example, in *blepel* than in *selepel* (McQueen & Cutler, 1998).

This sensitivity to the viability of stretches of speech as possible words has been called the Possible Word Constraint (PWC). Norris et al. (1997) implemented the PWC in Shortlist. The PWC improves segmentation in Shortlist by biasing the competition process. When the speech material between a candidate word and a likely word boundary is not a possible word, the activation of that candidate is halved. In the *fapple* case, the activation of the candidate *apple* is reduced because there is the impossible word *f* between the beginning of *apple* and the preceding silence, which is a very likely word boundary; spotting *apple* is thus more difficult than in *vuffapple*, where the PWC penalty does not apply. The PWC therefore acts so that lexical parses including impossible words are disfavored.

Under this account of segmentation, likely word boundaries can be marked in the signal by any of the wide variety of cues discussed above, including silence, as in the example above, and metrical structure. In English, for example, strong syllable onsets are assumed to be marked

in the speech input. These boundaries are then used by the PWC in the determination of whether or not particular candidate words should be penalized. In *mintayf*, for example, the strong syllable onset is marked before the [t], and the candidate *mint* will be evaluated relative to this boundary. In this case there is the single consonant [t] between the boundary and the end of *mint*. The word will thus be more difficult to spot in *mintayf* than in *mintef*, where there is no boundary cued before the [t]. There are therefore only possible words (the actual word *mint* and the possible word *ef*) between the end of *mint* and the two closest likely word boundaries (at the beginning and end of the complete string).

The aim of the present experiments was to examine this theory of lexical segmentation in Japanese. As in the case of the earlier study on metrical segmentation procedures (Otake et al., 1993), Japanese provides an opportunity to test aspects of the theory that cannot be tested in English. Japanese experiments also of course provide the opportunity to test the generality of the PWC account. This study consists of three word-spotting experiments, which examine how Japanese listeners segment continuous speech, and three control experiments.

In Experiment 1, listeners were required to spot trimoraic words in three different types of single-phoneme context, which either preceded or followed the target words. In both positions, the contexts were either single vowels, single nonmoraic consonants, or single moraic nasals (e.g., for preceding contexts, *agura*, to sit cross-legged, in *oagura, tagura* and *Nagura*; and for following contexts, *bikini*, bikini, in *bikinia, bikinip* and *bikiniN*). One goal of this experiment was to test the PWC by comparing the preceding vowel and consonant contexts. If Japanese listeners show the same sensitivity to the viability of sound sequences as possible words in the lexical parse, they, like English and Dutch listeners, should find it harder to spot *agura*, for example, in *tagura* than in *oagura*. This is because in Japanese, as in English and Dutch, single consonants are impossible words; a well-formed word must contain at least one vowel.

Note, however, that these conditions allow for a very controlled test of the PWC, one that is difficult to achieve in European languages. Each context is a single phoneme, so the segmental length of the preceding contexts is matched across conditions. In English, for example, it is hard to match contexts on number of phonemes while manipulating whether or not the contexts are possible words. If the target words were vowel-initial, glottal stops between the context and the word would tend to occur more often in the vowel context than in the consonant context (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996). If, on the other hand, the target words were consonant-initial, a consonant cluster would be formed in the consonant context but not in the vowel context. Neither of these comparisons is therefore ideal. Japanese allows for a test of the PWC with vowel-initial words and with the number of segments in the preceding context controlled since there are no glottal stops between neighboring vowels in Japanese (Pierrehumbert & Beckman, 1988; Vance, 1987). Vance (1987) argues that glottal stops in Japanese only occur either utterance-finally (after short vowels) or utterance-initially (particularly in emphatic speech). Pierrehumbert and Beckman (1988, p. 18) show an example phrase including a sequence of three vowels (spanning a word boundary) in which there were no glottal closures.

The following moraic nasal contexts provided an opportunity to extend the PWC account in a way that is not possible in European languages. There is a tension between two competing predictions for the moraic nasals. On the one hand, because they are consonants, moraic nasals may count as impossible words in Japanese and thus may cause the PWC penalty to be applied to words in moraic nasal contexts. On this account, spotting *bikini* should be harder in *bikiniN* than in *bikinia*. On the other hand, a moraic nasal may be an unproblematic context because there is a mora-boundary between the target and the context. Since Japanese listeners are sensitive to mora structure, and appear to use this as a segmentation cue (Cutler & Otake, 1994; Otake et al., 1993, 1996), the mora boundaries after *bikini* in both *bikiniN* and *bikinia* may signal that *bikini* is perfectly aligned with a likely word boundary and hence will not be penalized by the PWC in either case. Moraic nasals therefore

provide a test of the relative strength of the viability constraint and of rhythmic cues to likely word boundaries. This test cannot be performed in English, for example, where single consonants never constitute complete rhythmic units.

The preceding moraic nasal and following consonant conditions were included only so that types of contexts were balanced over position of context. Both of these contexts involve consonants in phonotactically illegal positions. Moraic nasals are coda consonants, so they may not occur in a syllable-initial position. Nonmoraic consonants do not occur in coda positions (nasal codas and geminate consonants are the only possible syllable codas; both are moraic). It is therefore difficult to interpret performance in the preceding moraic nasal and following consonant conditions. The crucial comparisons in Experiment 1 are therefore between the preceding consonant and vowel contexts and between the following moraic nasal and vowel contexts.

EXPERIMENT 1

Method

Participants. Fifty-four student volunteers took part. They were all native speakers of Japanese from the Tokyo area. They received course credits for their participation.

Materials. Fifty 3-mora target words were selected which contained as few embedded words as possible. The structure of the Japanese vocabulary is such that most of the 108 possible single morae are words or bound morphemes. It was thus impossible to avoid monomoraic embedded words. Very many bimoraic sequences are also words or bound morphemes. Such embeddings were avoided wherever possible, but embeddings which mismatched with the moraic structure of the longer word (e.g., *ari*, ant, in *atari*, hit) could not be avoided. An exhaustive search of the Japanese lexicon resulted in 24 acceptable vowel-initial words. These 24 words were then placed in three different preceding contexts: a single nonmoraic consonant (C; e.g., *agura*, to sit cross-legged, in *tagura*); a single vowel (V; e.g., *oagura*); and a moraic nasal (N; e.g., *Nagura*). A further 26 consonant-initial words were found which could be used with following

contexts, but only through the inclusion of several loan words (e.g., *bikini*). Although such words are marked in the orthography (through the use of *katakana* script), they are phonologically fully incorporated into the language. These 26 words were placed in three following contexts (consonant, e.g., *bikini*, bikini, in *bikinip*; Vowel, e.g., *bikinia*; and moraic nasal, e.g., *bikiniN*). Both the preceding and the following contexts were chosen so as to avoid creating other embedded words (again as much as was possible given the constraints of the vocabulary) and such that the complete target-bearing item was itself not a word. The 50 words, with translations, accent patterns, and in each of their three contexts, are listed in Appendix A.

Eight further trimoraic words were selected as additional targets. These words met the same constraints as the other words and were judged to be of high frequency. Four were placed in preceding syllabic (CVN) contexts (e.g., *omoshi*, weight, in *ruNomoshi*) and four were placed in following syllabic (CV) contexts (e.g., *mikan*, tangerine, in *mikanma*). The intention was that these eight targets would be very easy to spot and therefore that they would provide listeners not only with positive feedback during the experiment, but with the specific feedback that the targets to be spotted were trimoraic (thus directing the listeners' attention away from the other, unavoidable, monomoraic and bimoraic embedded words).

One hundred fillers were also constructed which contained no embedded trimoraic words and were themselves not words. As with the target-bearing items, shorter embeddings were avoided as much as possible. Forty of these fillers were matched to target-bearing items as follows: eight began with a moraic nasal followed by a vowel (matching, e.g., *Nagura*); eight began with a vowel-vowel sequence (matching, e.g., *oagura*); eight ended with a moraic nasal (matching, e.g., *bikiniN*); eight ended with a vowel-vowel sequence (matching, e.g., *bikinia*); and eight ended with a nonmoraic consonant (matching, e.g., *bikinip*). The remaining 60 fillers were all trimoraic CVCVCV sequences (matching, e.g., *tagura*, and providing additional non-target-bearing material).

Three lists were constructed. All three lists consisted of the same pseudo-random sequence of all 50 targets, 6 of the additional target-bearing items (evenly spaced through the list), and 94 of the fillers (there was always at least one filler after each target-bearing item). Targets were therefore embedded in about one third of the nonsense sequences. The only difference between the lists was in the target-bearing items, that is, in the contexts in which the targets appeared. The three preceding contexts were counterbalanced evenly across the lists, such that each list contained eight targets in each type of context. The three following contexts were counterbalanced as evenly as possible across the lists, such that each list contained nine targets in two types of context and eight targets in the third context. The remaining six fillers and two additional target-bearing items, plus two more items with embedded target words, were used to make a practice list. The construction of the lists, and of the filler materials, ensured that it was impossible for listeners to predict which items contained embedded words, where in a nonsense sequence a word might be (i.e., at the beginning or end), or what the type of context might be (i.e., moraic nasal, consonant, or vowel).

Procedure. All materials were recorded by a phonetically trained native speaker of Tokyo-dialect Japanese (the second author) in a sound-damped booth onto DAT tape, sampling at 48,000 Hz. The place of articulation of the moraic nasals was always velar, as is appropriate for moraic nasals in word-final position (as in the following context case).

In Tokyo Japanese, but not in all dialects of Japanese, words have characteristic pitch-accent patterns. Some morae, which are labeled high (H), tend to have higher pitch and greater amplitude than other morae, which are labeled low (L). The relative position of H and L morae, together with a final accent marker (^), can be used to indicate whether a word is accented or not and, if so, where the accent lies. For example, the sequence LHH indicates an unaccented word, while the sequence LHH^ indicates an accented word with the accent on the final mora. Only a limited set of possible accent patterns are permissible (see, e.g., Cutler & Otake, 1999, and Pierrehumbert & Beckman, 1988, for further details). The speaker produced each item (both target-bearing items and fillers) such that the pitch accent pattern of each complete nonsense item was well formed. He was therefore able to preserve the citation pitch accent patterns of all of the target words with following contexts (the addition of following contexts did not require any changes to the accent patterns of the words themselves, such that they sounded as they would if spoken in isolation). The citation form of the words in preceding consonantal contexts was also preserved (the addition of a preceding nonmoraic consonant also does not influence the pitch accent pattern). But for the words in preceding moraic contexts (moraic nasal and vowel), the speaker had to change the citation forms of the words so as to create phonologically legal accent patterns for the complete items. Specifically, items containing words with citation low-high-high (LHH and LHH^) patterns were produced with the initial (context) mora as low and the first mora of the target as high, thus making legal LHHH and LHHH^ sequences, respectively. LHL words plus their contexts became LHLL sequences.

It is important to emphasize that the accent patterns of words in fluent Japanese can undergo changes as a result of the phonological contexts in which they occur (Pierrehumbert & Beckman, 1988; Poser, 1984; Vance, 1987). Thus, while it is true that a given word has a particular accent pattern when spoken in isolation, pitch accents can be shifted or deleted when that word appears in continuous speech (for example, there can only be one accent within an accentual phrase). The words in preceding moraic nasal and vowel contexts therefore had accent patterns which were appropriate to those contexts. If the sequences had been meaningful, the target words could have been spoken with the accent patterns that were used. It would thus be incorrect to say that the words were wrongly accented. Thus, although speakers of Tokyo Japanese appear to use pitch accent information in spoken word recognition (Cutler & Otake, 1999), it is unlikely that they would find it harder to spot the words in these contexts simply because the

words did not have citation accent patterns. Note, however, that if the listeners did have difficulty with the words in the preceding vowel and moraic nasal contexts, this would act against the hypothesis. According to the PWC, words in preceding consonant contexts should be harder to spot than those in preceding vowel contexts (in spite of the fact that only the latter words have altered pitch accent patterns).

The materials were then transferred to computer, preserving the sampling rate of 48 kHz, and edited using Sound Designer II software to create the three lists in the correct running orders with 3 s of silence between each item. The three lists were then recorded onto the left channel of DAT tape, with a timing pulse on the right channel aligned approximately with the onset of each target-bearing item. The speech material on the left channel of the tape was presented to listeners binaurally over Audio-Technica ATH-A9 headphones; the timing pulses on the right channel could not be heard by the participants.

Listeners were tested in separate sound-attenuating carrels in a quiet room, either individually or in pairs. They were told that they would hear a list of nonsense words. They were asked to try to spot any real words, embedded at either the beginning or the end of the nonsense words. Examples of trimoraic words in all six contexts were provided in the instructions. The listeners were asked to press a button as fast as possible if they spotted a word and then to say aloud, in a low voice, what that word was. Those listeners who were tested in pairs could not hear each other's spoken responses. They were asked to press the button with their preferred hand. The spoken responses were recorded onto DAT tape. Each listener heard the practice list, and one of the three experimental lists. Eighteen participants heard each list.

The experiment was run using a Sony TCD-D10 DAT player interfaced with a personal computer running NESU experiment control software. The computer clock was started by each timing pulse and stopped by each button-press. All responses were logged on the computer. Prior to data analysis, the materials were transferred to a computer (down-sampled to 16,000 Hz, 16 bit). Each target-bearing item was examined and labeled using the Xwaves speech editor. The duration of each target word and its context was measured, as was the duration between the timing pulse and the onset of the item.

Results and Discussion

The raw reaction times (RTs) were adjusted to measure from the offset of each target word. RTs to each target in each following context were adjusted by subtracting the duration of that target, minus the duration from the timing pulse to the item onset, from the raw RTs. RTs to each target in each preceding context were adjusted by subtracting the duration of that complete item, minus the pulse to item onset difference, from the raw RTs. The spoken responses were checked to test whether each button-press was accompanied by the correct oral response. All manual responses which were accompanied by a word other than the intended target were treated as errors (only 14 responses, i.e., 0.5% of all responses). Responses slower than 2500 ms were not recorded and thus any such responses were also treated as errors. All other responses were submitted to the RT analyses. Mean RTs (measured from target offset) and mean error rates (percentage of targets missed) are shown in Table 1.

Analyses of variance (ANOVAs) were performed on both the RT and error data, with either listeners ($F1$) or items ($F2$) as the repeated measures. In the RT analysis by listeners, missing data for each listener were replaced with the overall mean of that listener; in the RT analysis

TABLE 1

Mean Reaction Times for Correct Detection (RT, in ms), Measured from Target-Word Offset, and Mean Percentage Missed Targets (Errors), in Experiment 1

Context	Consonant	Vowel	Moraic nasal
Preceding			
Example	tagura	oagura	Nagura
RT	1146	1041	984
Errors	38%	27%	41%
Following			
Example	bikinip	bikinia	bikiniN
RT	718	678	695
Errors	9%	13%	9%

by items, missing data for each item were replaced with the mean of the available RTs for that item within each context condition. Separate analyses were performed for the targets with preceding contexts and for the targets with following contexts.

The listeners' ability to spot words with preceding contexts was influenced by the nature of those contexts (RT: $F1(2,102) = 18.4$, $p < .001$; $F2(2,46) = 13.6$, $p < .001$; Errors: $F1(2,102) = 6.5$, $p < .005$; $F2(2,46) = 5.8$, $p < .01$). Planned comparisons between the vowel and consonant conditions were then carried out (all such comparisons were two-tailed t-tests, with $p < .05$, unless otherwise stated). Listeners spotted words significantly more slowly in consonantal contexts (e.g., *agura* in *tagura*) than in vocalic contexts (e.g., *agura* in *oagura*; a difference of 105 ms on average; $t1(53) = 3.4$, $t2(23) = 3.9$). Similar tests on errors showed that listeners also missed more words in the consonantal than in the vocalic contexts (11% more, on average; $t1(53) = 2.6$, $t2(23) = 2.6$). Both of these findings corroborate the previous findings in both English (Norris et al., 1997) and Dutch (McQueen, 1998; McQueen & Cutler, 1998) that listeners find it harder to spot words in impossible-word contexts (e.g., *tagura*) than in possible-word contexts (e.g., *oagura*). Japanese listeners therefore appear to use the PWC when segmenting speech.[2]

---

[2] Planned comparisons involving the preceding moraic nasal contexts were also carried out. Responses in the moraic nasal condition (e.g., *Nagura*) were reliably faster than in the consonant condition (*tagura*; by 162 ms, on average; $t1(53) = 4.5$, $t2(23) = 5.5$), and than in the vowel condition (*oagura*), though this was not significant by items (57 ms, on average; $t1(53) = 2.8$, $t2(23) = 1.4$, $p > .1$). But, although responses in the moraic nasal condition tended to be fast, they were the most errorful. More targets were missed in the moraic nasal condition than in the vowel condition (14% more errors, on average; $t1(53) = 3.7$, $t2(23) = 3.2$). There was no difference in error rates between the consonant and moraic nasal conditions. Since moraic nasals are illegal in syllable-initial position, these results are difficult to interpret. Although there were fillers which also began with moraic nasals, these morae could have provided an additional segmentation cue, resulting in the faster responses in this condition. Likewise, the absence of any lexical competitors beginning with N could have made targets in these contexts faster to spot. On the other hand, the illegality of N could have been disrupting, causing the high error rate.

One potential problem with these results is that they may reflect the operation of lexical competition, rather than of the PWC. While every attempt was made to minimize the number of embedded words in the materials, it was not possible to avoid creating some embedded words when combining the target words with their preceding contexts. These words were often very low frequency, but we counted them if they were listed in an online dictionary of Japanese (Matsumura, 1993). More embedded words were created in the consonant contexts than in the vowel contexts (all 24 items in the consonant context condition had monomoraic embeddings like *ta*, rice field, in *tagura* [this is unavoidable, since all CV morae are words], and 13 in the consonant condition had bimoraic embeddings like *tagu*, tag; 13 items in the vowel context condition had bimoraic embeddings, like *ea*, air, in *earoe*; none in the vowel condition had trimoraic embeddings). ANOVAs were carried out on the preceding consonant and vowel data (both RTs and errors), with an additional lexical embedding factor. One group of items were those matched on embedded words (either both contexts had embedded words or neither did); the other items were mismatched (an embedded word in the consonant context but not in the vowel context). Neither in the analyses of embeddings involving the first mora of the target (e.g., *ta* and *ea*) nor in those of embeddings involving the first two morae of the target (e.g., *tagu*) was there a significant interaction of the lexical embedding factor with the context (consonant vs. vowel) factor. There were also more lexical competitors consistent with the onsets of the consonant context items (125.8 competitors, on average, beginning with the first three segments, as measured in the same online dictionary; 10.7 competitors, on average, beginning with the first four segments) than were consistent with the onsets of the vowel context items (0.7 and 0.3 competitors, on average, for the first three or four segments, respectively). There were, however, no significant correlations of number of competitors with mean item RT or error rate, neither for the consonant nor the vowel contexts.

The ANOVAs on word-spotting performance in following contexts showed much weaker ef-

fects. There were no reliable differences in error rates. In RT, the effect of context was significant only by subjects: $F1(2,102) = 4.0$, $p < .05$; $F2(2,50) = 2.8$, $p = .07$. Although the error rate was on average 4% lower in the moraic nasal condition than in the vowel condition, responses were, on average, 17 ms faster in the vowel condition; neither of these differences was significant. There was therefore no evidence that target words in moraic nasal contexts were penalized by the PWC (because moraic nasals are impossible words). Instead, it would appear that the mora boundary between the end of the target word and the moraic nasal provided the listeners with a segmentation cue. Since the target was thus perfectly aligned with that boundary it was not penalized by the PWC. We suggested in the introduction that one important cue to moraic nasals, and hence to the mora boundaries before (and after) them, is provided by their duration. The following-context moraic nasals in Experiment 1 were in fact slightly longer than the following-context vowels (mean durations: moraic nasals, 209 ms; vowels, 180 ms). The moraic nasal consonants were thus as long as the syllables formed by the vowels.

An analysis using the online dictionary (Matsumura, 1993) revealed that there was again an imbalance in the number of embedded words between contexts, with more bimoraic embedded words consisting of the last mora of the target word and the context in the moraic nasal condition (22/26 like *niN*, person, in *bikiniN*) than in the vowel condition (13 like *sui*, elegant, in *garasui*). But again, in ANOVAs splitting items into those matched and those mismatched on lexical embeddings, there was no interaction of this factor with the context factor. Likewise, although there were more lexical competitors consistent with the moraic nasal contexts (281.8 competitors, on average, beginning with the last three segments) than there were in the vowel contexts (49.7 competitors, on average), competitor set size did not correlate with either mean RT or mean error rate in either the moraic nasal or the vowel conditions. A correlational analysis was also performed on the mean RTs and uniqueness points (UPs) of the following

context items. There was no systematic relationship between these measures.[3]

EXPERIMENT 2

Two principal findings have thus emerged from Experiment 1: that words with preceding consonantal contexts are harder to spot than those with preceding vocalic contexts, suggesting that Japanese listeners use the PWC in segmentation, and that words with following moraic nasals are no harder to spot than those with following vowels, suggesting that words in moraic nasal contexts do not incur a PWC penalty. But these conclusions depend on the assumption that the target words in each context were equally easy to perceive. Since each target-bearing item was a different natural utterance, there could be uncontrolled acoustic differences between the targets which could be responsible for the differences between contexts found in Experiment 1 and/or for the lack of differences between contexts. To address this concern, a go/no-go lexical decision task was used (as is common in studies employing the word-spotting task; Cutler & Norris, 1988; McQueen, 1996). In Experiment 2, each target word from Experiment 1 was excised from its context, and nonwords were created from the Experiment 1 fillers. Listeners were asked to listen to a list of these words and nonwords and to press a button whenever they heard a real word. The pattern of results in Experiment 1 could then be reinterpreted in the light of the new pattern of results.

Method

Participants. Fifty-seven student volunteers took part. They had not participated in Experiment 1. Nineteen heard each of the three lists.

---

[3] The only significant difference among the following context conditions was that responses in the consonant condition (e.g., to *bikini* in *bikinip*) were, on average, 40 ms slower than in the vowel condition (e.g., *bikinia*; $t1(53) = 2.2$, $t2(25) = 2.7$). This finding might be seen as providing additional support for the PWC: Listeners found it harder to spot words in following impossible-word contexts than in following possible-word contexts. Note, however, that syllable-final nonmoraic consonants are illegal in Japanese. The poor performance in this condition may thus have been caused by the disrupting effects of a phonotactically illegal sequence, rather than or as well as by the operation of the PWC.

They received course credits or payment and were all Japanese native speakers from the Tokyo area.

Materials and procedure. The materials were made from those used in Experiment 1. They consisted of all 60 target words (experimental words plus the additional words used as easy or practice targets), excised from their contexts; nonwords excised from the 40 fillers which were matched to the target-bearing items; and the remaining 60 unmodified nonword fillers. All items were thus trimoraic. Items were excised using the Xwaves speech editor. All cuts were made at zero-crossings such that there were no perceptible transient clicks (neither at word onsets, for words taken from preceding contexts, nor at word offsets, for words taken from following contexts). The cuts were made at the point of lowest amplitude in vowel-vowel sequences and at the offset or onset of the consonantal occlusion (for fricatives, stops, nasals, and the liquid /r/) in consonant-vowel and vowel-consonant sequences, respectively. Cuts were made at locations such that excised words sounded as natural as possible and such that the context could no longer be heard as an independent phoneme attached to the target word. There were some coarticulatory cues to the contextual segments remaining in the first vowels of words from preceding contexts and the last vowels of words from following contexts, but these cues gave the impression only that the words had been excised from longer utterances, not that the contexts were still present.

Three lists were constructed with exactly the same order of presentation of items as the three lists in Experiment 1, with each word appearing where the target-bearing item containing that word had appeared and each nonword appearing where the filler containing that nonword had appeared. These lists were played out onto DAT tape. Timing pulses were again aligned approximately with item onsets, and the durations between pulses and item onsets were again measured using Xwaves.

The testing procedure and equipment were identical to those used in Experiment 1, except that instructions for lexical decision were given. Listeners were asked to listen to a list of words and nonwords, to respond by pressing a button as fast as possible whenever they heard a real word, and to say what that word was.

Results and Discussion

Raw RTs were again adjusted to measure from the offset of each target word. All manual responses were accompanied by the correct oral response. Responses slower than 2000 ms (i.e., extreme outliers in lexical decision) were excluded and treated as errors (7 responses, 0.2%); responses slower than 2500 ms were not recorded and thus any such responses were also treated as errors. Six subjects were found to have overall error rates over 40%; they were excluded from further analysis. Furthermore, the timing pulse for one word on one of the three lists was missing, so all responses to that word in that list were lost. The item was removed completely from the analysis. All other responses were submitted to the RT analyses. Missing data were replaced in the same way as in Experiment 1. Mean RTs (measured from target offset) and mean error rates (percentage of targets missed) are shown in Table 2.

In ANOVAs on the responses to words taken from preceding contexts, there was a significant effect of context: RT: $F1(2,96) = 49.3, p < .001$; $F2(2,44) = 16.7, p < .001$; Errors: $F1(2,96) = 15.9, p < .001$; $F2(2,44) = 9.7, p < .001$. Pairwise comparisons showed that responses to words taken from consonantal contexts (e.g., *agura* taken from *tagura*) were faster than those to words taken from vocalic contexts (e.g., *agura*

TABLE 2

Mean Reaction Times for Correct Detection (RT, in ms), Measured from Word Offset, and Mean Percentage Missed Words (Errors), in Experiment 2

Original context	Consonant	Vowel	Moraic nasal
Preceding			
Example	(t)agura	(o)agura	(N)agura
RT	518	605	698
Errors	12%	14%	29%
Following			
Example	bikini(p)	bikini(a)	bikini(N)
RT	540	492	507
Errors	8%	5%	7%

taken from *oagura*), though this difference was not significant by items (a mean difference of 91 ms; $t1(50) = 4.2$, $t2(22) = 1.9$, $p = .07$). There was no reliable difference in the error rates between these two conditions. One reason why the words taken from consonantal contexts may have been responded to more rapidly than those taken from vocalic contexts is that the former words had pitch accent patterns which are appropriate for those words when spoken in isolation. The latter words had accent patterns which, though appropriate in the original contexts, are illegal in isolation (e.g., LHH words like *agura* had HHH patterns, taken from LHHH *oagura*). In line with the present results, Minematsu and Hirose (1995) have shown that Japanese words are harder to recognize when they are misaccented than when they are correctly accented. Misaccented words are analogous to misstressed English words, such as when *tyPHOON* is produced as *TYphoon*. While the pronunciation *TYphoon* is appropriate in a stress-shift environment (e.g., in *TYphoon Charlie*), it is inappropriate in isolation. It is thus interesting to note that such words, when presented in isolation to English listeners, show a small but not significant misstressing effect (Cutler & Clifton, 1984).

These results suggest that the difference in word-spotting performance (in the opposite direction) between the consonant and vowel contexts was indeed due to the segmentation process and not to acoustic differences between the targets in each context. In spite of the fact that words from consonantal contexts were responded to somewhat more rapidly in lexical decision than words from vocalic contexts, the words when they were in the consonantal contexts were harder to spot. An analysis of covariance (ANCOVA) of the RTs from the preceding vowel and consonant conditions in Experiment 1, with items as repeated factor and item RTs from the corresponding conditions in Experiment 2 as covariate, showed that the difference between these conditions remained significant ($F2(1,21) = 21.8$, $p < .001$). A similar ANCOVA on error rates also showed that the difference between impossible- and possible-word contexts remained significant ($F2(1,21) = 9.5$, $p < .01$).

Note that while the simplest result to interpret from this control experiment would have been a null result, any results which are not the same as those found in word spotting are useful control data. When a different pattern is found in lexical decision to that which was found in word spotting, as is the case here, it may indeed be due in part to acoustic confounds caused by the excision (e.g., formant transition cues signaling an absent context phoneme, or, as just mentioned, pitch accent patterns which are illegal for isolated words). But if a within-item acoustic confound were the cause of the difference in word spotting, then the effects caused by excision would have to be exactly the reverse and large enough not only to remove the original confound, but also to produce a difference in the opposite direction. This seems unlikely. There is no evidence in the control experiment to suggest the existence of an acoustic confound within the words as they were heard in the word-spotting experiment. The effect observed in word spotting therefore appears to be due to the relative ease of segmentation of the words from their contexts.

In ANOVAs on the responses to words taken from following contexts there were no significant differences between conditions in the error rates. There was, however, an effect of context in RTs, but significant only by subjects: $F1(2,96) = 8.0$, $p < .001$; $F2(2,50) = 2.3$, $p = .1$. Planned comparisons showed that there was no significant difference in lexical decision between responses to the targets taken from moraic nasal contexts and those taken from vocalic contexts. Furthermore, the nonsignificant difference in RTs between these two contexts in word-spotting remained nonsignificant in a by-items ANCOVA on those data with the lexical decision data as covariate ($F2(1,24) = 1.8$, $p > .1$). These results suggest that the failure to find a difference between the following moraic nasal and vowel conditions in word spotting was not due to an acoustic confound. It was possible that an acoustic confound could have acted to benefit words in moraic nasal contexts and thus that any difference between these conditions due to segmentation difficulty (i.e., moraic nasals, as impossible words, triggering the PWC) might have been obscured. The lexical decision data

suggest that this was not so. Instead, it appears that words in the context of following moraic nasals are just as easy to segment as words in the context of following vowels.[4]

EXPERIMENT 3

The two primary conclusions from Experiment 1 were strengthened by the results of Experiment 2. First, preceding nonmoraic consonants are difficult contexts for word spotting because those consonants, as impossible words in Japanese, cause the PWC to penalize the activation of the target words. Second, following moraic nasals are not difficult contexts for word spotting because the mora boundary between the end of the target word and the moraic nasal appears to cue listeners that there is a likely word boundary at that location. The failure to find a difference between following vocalic and

---

[4] Pairwise comparisons involving words taken from preceding moraic nasal contexts and from following consonant contexts were also carried out. Words taken from preceding moraic nasal contexts were responded to more slowly than those taken from either other preceding context (moraic nasals versus vowels: $t1(50) = 5.3$, $t2(22) = 3.8$; moraic nasals versus consonants: $t1(50) = 8.6$, $t2(22) = 6.0$) and also less accurately than in the other two contexts (moraic nasals versus vowels: $t1(50) = 4.3$, $t2(22) = 4.0$; moraic nasals versus consonants: $t1(50) = 4.4$, $t2(22) = 3.9$). In ANCOVAs on the Experiment 1 preceding vowel and moraic nasal conditions with the equivalent Experiment 2 data as covariate, the difference in RTs, as in the original analysis, was not significant ($F2(1,21) = 3.2$, $p > .05$), and the difference in errors, which was significant in the original analysis, was no longer significant ($F2(1,21) = 3.7$, $p > .05$). The high error rate on the targets in moraic nasal contexts in word-spotting can therefore at least in part be attributed to acoustic differences between the targets in the moraic nasal and vowel contexts.

Responses to words taken from following consonantal contexts were slower than those taken from following vocalic contexts (48 ms, on average, $t1(50) = 4.3$, $t2(25) = 2.1$) and those taken from following moraic nasal contexts (33 ms, on average, $t1(50) = 2.3$, but $t2(22) = 1.3$, $p > .2$). Items ANCOVAs comparing the following consonant and vowel conditions on the word-spotting RTs with the lexical decision RTs as covariate showed that the difference between these conditions was not significant ($F2(1,24) = 4.1$, $p > .05$). This suggests that the poorer performance in word spotting in the following consonant condition was due in part to acoustic differences; the targets from these contexts were harder to recognize even when they had been excised from those contexts.

moraic nasal contexts could, however, be due to a number of other factors. One is that the materials in Experiment 1 had many unavoidable lexical embeddings in addition to the intended target words. Even though the analyses of lexical embedding and of number of lexical competitors failed to reveal any significant effects of these factors, competition between the target words and these other words may have acted to mask effects of the context on word-spotting performance. Another problem is that many of the targets in the following context condition were loan words; performance on such items may not be equivalent to that on words from the core Japanese vocabulary. Finally, all of the following context words became unique on their final phoneme or earlier and thus could in principle be recognized without the need to process any of the following context. Although the correlation for the following context items between word-spotting RTs and UPs was not significant, the fast responses in these conditions relative to the preceding context conditions are consistent with the view that the following contexts were not fully processed. If so, then any differential effect between moraic nasal and vocalic contexts might therefore have been missed.

Given these problems with the following context items, it was decided to rerun Experiment 1, but with bimoraic targets. Shorter target words with late UPs may be more likely to reveal effects of following contexts. Such words were therefore used in Experiment 3. Although the most interesting comparison was between following moraic nasal and vowel contexts, the following nonmoraic consonant condition was also included for compatibility with Experiment 1. Likewise, the three preceding context conditions from Experiment 1 were included, also with bimoraic targets. The design of Experiment 3 was thus identical to that of Experiment 1; the only difference was that we used a set of shorter target words. It was predicted that the preceding context results would replicate those of Experiment 1. It was also predicted that if there was indeed a difference between following moraic nasal and vowel contexts (which Experiment 1 failed to detect), then it was more likely to be found with the shorter words used in Experiment 3.

Method

Participants. Sixty volunteer students took part. Twenty heard each of the three lists. They received course credits for their participation and were all native speakers of Tokyo Japanese. None had participated in Experiments 1 or 2.

Materials. Fifty new bimoraic targets were selected. Twenty-four were then combined with preceding contexts (consonant, vowel, or moraic nasal) and 26 with the equivalent three following contexts. As in Experiment 1, embedded words other than the targets were avoided as much as possible in the target-bearing items, but again it was impossible to avoid monomoraic embeddings. These words and their contexts are listed in Appendix B.

The other items were also analogous to those in Experiment 1: 8 additional easy target-bearing items with bimoraic targets, 40 fillers matched to each type of target-bearing item, and 60 additional bimoraic fillers. Three experimental lists with equivalent counterbalancing and ordering constraints to those in Experiment 1 were then made from these materials, together with a similar practice list.

Procedure. This was identical to that of Experiment 1, except that in the instructions listeners were given examples of bimoraic targets in each of the six contexts. Note that in the recording the citation accent patterns of the words in all three following contexts and in the preceding consonant contexts were again preserved, while those of the words in the preceding moraic nasal and vowel contexts were altered so that the complete item was a well-formed phonological word. Thus, LH and LH^ words plus their contexts became LHH and LHH^ sequences, respectively, and HL words plus context became HLL sequences.

Results and Discussion

Raw RTs were adjusted in the same way as in Experiment 1 to measure from the offset of each target word. The check of the spoken responses showed that all manual responses were accompanied by correct oral responses. Responses slower than 2500 ms were not recorded and thus any such responses were treated as errors. One subject failed to detect any target words and was excluded from the analysis. In addition, due to an extra timing pulse appearing inadvertently on one list, responses to the final target word on that list were not recorded. That item was removed from the analysis. All other responses were submitted to the RT analyses. Missing data were replaced in the same way as in the earlier experiments. Mean RTs (measured from target offset) and mean error rates (percentage of targets missed) are shown in Table 3.

As in Experiment 1, listeners' performance in the preceding context conditions depended on the nature of the contexts. ANOVAs on both RTs and errors showed significant context effects: RT, $F1(2,112) = 78.5$, $p < .001$; $F2(2,46) = 33.5$, $p < .001$; Errors, $F1(2,112) = 12.3$, $p < .001$; $F2(2,46) = 9.2$, $p < .001$. Planned comparisons on RTs among the three contexts showed that, as in Experiment 1, listeners were reliably slower to spot words in consonantal contexts (e.g., *ari*, ant, in *rari*) than in vocalic contexts (e.g., *eari*; by 112 ms on average; $t1(58) = 6.4$, $t2(23) = 3.3$). Listeners again missed more words in the consonantal than in the vocalic contexts (9% more, on average; $t1(58) = 2.6$, $t2(23) = 2.3$). In contrast to Experiment 1, it was possible to match the materials fairly well on the presence of embedded words. For 19 of the 24 targets, the context plus the first mora of the target was an embedded word in both versions (e.g., *ra*, la, and *ea*, air, in *rari* and *eari*; counts were again performed using the online dictionary, Matsumura,

TABLE 3

Mean Reaction Times for Correct Detection (RT, in ms), Measured from Target-Word Offset, and Mean Percentage Missed Targets (Errors), in Experiment 3

Context	Consonant	Vowel	Moraic nasal
Preceding			
Example	rari	eari	Nari
RT	1042	930	790
Errors	40%	31%	23%
Following			
Example	sarup	sarua	saruN
RT	839	772	781
Errors	17%	14%	19%

1993). It remained the case, however, that there were more lexical competitors in the consonant condition than in the vowel condition (there were, on average, 41.1 competitors beginning with the first three segments in the consonant condition and 2.3 such competitors, on average, in the vowel condition). Number of competitors did not correlate with mean item RT or error rate in either condition. These results thus support the conclusion drawn from the results of the equivalent conditions in Experiment 1, namely, that Japanese listeners appear to segment speech using the PWC.

In ANOVAs on the following context RTs, there was also a significant effect of context: $F1(2,112) = 10.3, p < .001; F2(2,48) = 6.0, p < .005$. This effect was due entirely to differences involving the consonantal condition. There was no difference between the moraic nasal and vowel conditions (e.g., *saru*, monkey, in *saruN* and *sarua*). There were no significant effects in the error rates in the following context conditions.[5]

For 12 of the 25 targets in moraic nasal contexts, the final mora of the target and the context formed an embedded word. For the other targets in moraic nasal contexts, and all targets in vowel contexts, the final mora of the target and the context did not form an embedded word. Any effect of lexical embedding might thus have made it harder for listeners to spot words in moraic nasal contexts than in vowel contexts. But, as in Experiment 1, ANOVAs in which items were split into those matched and those mismatched on lexical embeddings revealed no inter-

action of this factor with the context factor. There were also more lexical competitors in the moraic nasal contexts (19.3 competitors, on average, beginning with the last three segments) than in the vowel contexts (2.9 competitors, on average). But, as in all earlier experiments, there were no significant correlations of number of competitors with either mean RT or mean error rate.

The most important result from Experiment 3 is the replication, with materials which avoided the problems associated with the Experiment 1 materials, of the finding that word spotting is equally easy in moraic nasal and vocalic contexts. Moraic nasals appear to provide a segmentation cue, namely, that there is a likely word boundary at the mora boundary between the end of the target word and the moraic nasal. As in Experiment 1, the moraic nasal contexts were again slightly longer than the vowel contexts (mean durations: moraic nasals, 187 ms; vowels, 151 ms). The duration of the moraic nasal consonants thus appears to have provided an important signal to the presence of the mora boundaries at the end of the target words in this condition. Because a word such as *saru* is perfectly aligned with the mora boundary in *saruN* (it ends at that boundary), no PWC penalty is applied to the word.

For the same reasons which motivated Experiment 2, however, it was necessary to run a further control lexical decision experiment. Experiment 4 was therefore run, as an analogue to Experiment 2, but using the targets from Experiment 3. If the results of Experiment 3 were due to acoustic differences between the targets spoken in each context, then those differences should also emerge when the targets are presented in isolation, for lexical decision.

---

[5] The results for preceding moraic nasals were the same as those in Experiment 1 in RTs, but not in errors. Listeners were reliably faster to detect words in moraic nasal contexts (e.g., *Nari*) than in both the consonantal (*rari*; $t1(58) = 10.3$, $t2(23) = 10.5$) and the vocalic contexts (*eari*; $t1(58) = 6.9$, $t2(23) = 4.3$). However, in contrast to Experiment 1, where responses in the moraic nasal condition were the most errorful of the preceding context conditions, responses in this condition in Experiment 3 were the most accurate. Listeners missed fewer targets in the moraic nasal context than in either the consonantal (*t1(58) = 4.6, t2(23) = 4.1*) or the vocalic contexts (*t1(58) = 2.6, t2(23) = 2.1, p = .05*). Responses to words in the following consonant contexts (e.g., *sarup*) were slower than those in both the moraic nasal condition (*saruN*; $t1(58) = 3.1$, $t2(24) = 2.7$) and the vowel condition (*sarua*; $t1(58) = 3.6$, $t2(24) = 3.6$).

EXPERIMENT 4

Method

Participants. Forty-five student volunteers took part; fifteen heard each of the three lists. They were all native speakers of Tokyo Japanese and received course credits. None had participated in Experiments 1, 2, or 3.

Materials and procedure. These were identical to those of Experiment 2, except that the words and nonwords were excised from the

Experiment 3 materials rather than from the Experiment 1 materials.

Results and Discussion

Raw RTs were adjusted to measure from the offset of each target word. All manual responses were accompanied by the correct oral response. As in Experiment 2, responses slower than 2000 ms were excluded and treated as errors (20 responses, 0.9%); responses slower than 2500 ms were not recorded and thus any such responses were also treated as errors. For compatibility with Experiment 3, the responses to the item which had to be excluded in that experiment were also excluded here. All other responses were submitted to the RT analyses. Missing data were replaced in the same way as in the earlier experiments. Mean RTs (measured from target offset) and mean error rates (percentage of targets missed) are shown in Table 4.

In the responses to words taken from preceding contexts, there was again a significant effect of context: RT: $F1(2,84) = 19.2$, $p < .001$; $F2(2,46) = 5.2$, $p < .01$; Errors: $F1(2,84) = 21.1$, $p < .001$; $F2(2,46) = 7.6$, $p < .005$. As in Experiment 2, responses to words taken from preceding consonantal contexts (e.g., *ari* taken from *rari*) were faster than those to words taken from preceding vocalic contexts (e.g., *ari* taken from *eari*), a mean difference of 124 ms: $t1(44) = 6.8$, $t2(23) = 3.5$. Responses in the consonant condition were also, on average, 20% more accurate than in the vowel condition: $t1(44) = 7.7$,

$t2(23) = 3.4$. These differences may again in part reflect the fact that the words taken from consonantal contexts had correct pitch accent patterns while those taken from vocalic contexts had pitch accent patterns which were not appropriate for isolated words.

As in the comparison between Experiments 1 and 2, therefore, any effect of the acoustic differences between the targets in each context would appear to have acted against the effects of context observed in Experiment 3. Even though targets from consonantal contexts were easier to process than those from vowel contexts when they were presented in isolation for lexical decision, they were harder to spot when they were presented in context. Although by-items ANCOVAs on the preceding vowel and consonant error results from Experiment 3, with the Experiment 4 results as the covariate, showed that the difference between these conditions was no longer significant ($F2(1,22) = 3.5$, $p > .05$), a similar ANCOVA on the RT data showed that the RT difference in word-spotting remained significant ($F2(1,22) = 13.5$, $p < .005$). The RT effect in Experiment 3 therefore appears to reflect the fact that listeners find it easier to segment words from preceding vowel contexts than from preceding consonant contexts.[6]

---

[6] Words from preceding moraic nasal contexts (e.g., *ari* from *Nari*) were responded to more slowly ($t1(44) = 3.6$, $t2(23) = 2.1$) and less accurately ($t1(44) = 3.4$, $t2(23) = 2.6$) than words from preceding nonmoraic consonant contexts (*ari* from *rari*). These results seem to reflect the fact that the words taken from nonmoraic consonant contexts were particularly easy to detect in the lexical decision task. Words from preceding moraic nasal contexts were, however, not responded to more quickly than words from preceding vowel contexts, nor was there a fully reliable effect in the error rates for this comparison; although responses on the moraic nasal condition were 8% more accurate than those in the vowel condition, this difference was not significant by items ($t1(44) = 2.1$, $t2(23) = 1.7$, $p = .1$). Although these differences are not significant, they reflect the pattern observed in word spotting. In ANCOVAs on the word-spotting data with the lexical decision data as covariate, the difference between the preceding vowel and moraic nasal conditions remained significant in RTs ($F2(1,22) = 16.1$, $p < .005$) but not in errors ($F2(1,22) = 3.6$, $p < .05$). It therefore appears that the advantage for the preceding moraic nasal contexts in word spotting in errors (but not in RTs) was due to the better acoustic quality of the targets in the nasal contexts.

TABLE 4

Mean Reaction Times for Correct Detection (RT, in ms), Measured from Word Offset, and Mean Percentage Missed Words (Errors), in Experiment 4

Original context	Consonant	Vowel	Moraic nasal
Preceding			
Example	(r)ari	(e)ari	(N)ari
RT	648	772	733
Errors	10%	30%	22%
Following			
Example	saru(p)	saru(a)	saru(N)
RT	676	636	660
Errors	10%	7%	11%

There were no fully reliable differences in the ANOVAs on responses to words taken from following contexts. In RTs, the context effect was significant by subjects but not by items: $F1(2,84) = 3.7$, $p < .05$; $F2(2,48) = 1.1$, $p > .1$. Pairwise comparisons showed that this effect was due solely to the difference between the consonant and vowel conditions.[7] In errors, the context effect was not significant. It was only marginal by subjects: $F1(2,84) = 2.9$, $p = .06$; $F2(2,48) = 2.1$, $p > .1$. This weak effect was due only to the difference between the moraic nasal and vowel conditions: responses to words taken from vowel contexts were, on average, 4% more accurate than responses to words taken from moraic nasal contexts: $t1(44) = 2.5$, $t2(24) = 2.2$. The nonsignificant trends in word-spotting that targets in moraic nasal contexts were harder to process than those in vowel contexts may thus be due to a small acoustic confound. It seems clear, however, that there is no major difference between these two conditions and certainly no evidence that an acoustic confound (favoring the words in moraic nasal contexts) was somehow masking an effect due to the context.

EXPERIMENT 5

We can now summarize the main findings of the first four experiments. Japanese listeners found word-spotting in preceding impossible-word contexts harder than in preceding possible-word contexts, for both 2-mora targets (Experiment 3) and 3-mora targets (Experiment 1),

---

[7] Responses to words taken from following vowel contexts (e.g., *saru* from *sarua*) were, on average, 40 ms faster than responses to words taken from following consonant contexts (e.g., *saru* from *sarup*): $t1(44) = 2.6$, $t2(24) = 2.0$, $p = .06$. This trend might suggest that (as in the comparison between Experiments 1 and 2) the relatively poor performance on word spotting with following consonant contexts is due at least in part to an acoustic confound: the words with following contexts which were the hardest to spot were also detected most slowly in lexical decision. But an ANCOVA on the Experiment 3 word-spotting RTs, with the Experiment 4 RTs as covariate, comparing following consonant and vowel conditions showed that the difference between these conditions remained significant ($F2(1,23) = 9.7$, $p < .01$). The difference therefore appears to reflect how difficult listeners found it to segment words from consonant versus vowel contexts, rather than an acoustic difference between the words in the different contexts.

just like English and Dutch listeners. These results are consistent with the account of speech segmentation offered by the PWC: when the stretch of speech between the edge of a candidate word and a likely word boundary contains only consonantal material, as with, for example, *agura* in *tagura* or *ari* in *rari*, that candidate word is penalized. It is thus harder to spot than when there is vocalic material preceding the candidate word, as in *oagura* and *eari*. In contexts such as these, the candidate word is aligned with a likely word boundary (a mora boundary), and, in any case, the stretch of speech between the word and the preceding silence (the vowel context) is a possible Japanese word.

The results for following moraic nasal versus vowel contexts are also very clear. There was no evidence, either with 2- or 3-mora words, that word-spotting was harder in moraic nasal contexts than in vowel contexts. It appears that since target words were aligned with a mora boundary in both contexts, they were not penalized by the PWC and word-spotting was equally easy in each case. That is, even though single moraic nasals are impossible Japanese words, their presence in an ongoing parse of Japanese speech does not make the recognition of neighboring words harder (like *saru* in *saruN*), since in any case there will be a mora boundary between the moraic nasal and the offset (or onset) of the word. The evidence for this claim, however, comes from two null effects (the lack of a difference between the following moraic nasal versus vowel conditions in Experiment 1 and again in Experiment 3). While one can argue that the same experiments were not lacking in power (given the robust differences between the preceding consonant and vowel conditions in each experiment), it remains the case that within the set of items with following contexts, there is no condition against which to evaluate the vowel and moraic nasal conditions. Given the illegal status of the following consonants, it is impossible to say whether the vowel and moraic nasal conditions were both relatively easy conditions or both relatively hard. A more convincing demonstration that these two contexts are equally easy, therefore, would be an experiment in which there was an appropriate comparison context

(i.e., a phonotactically legal context).

Experiment 5 was an attempt to provide such a test. Words, such as *uni* (sea urchin), were placed in three preceding contexts, consisting of a CCV mora (e.g., *gya*) plus either a moraic nasal (*gyaNuni*), a vowel (*gyaouni*), or a consonant (*gyabuni*). All three of these contexts are phonotactically legal. The predictions were as follows. The results of Experiments 1 and 3 suggest that listeners should find it difficult to spot *uni* in *gyabuni*, with the mora structure *gya-bu-ni*. There is a single (nonmoraic) consonant ([b]) between the likely word boundary after *gya* and the onset of *uni*, so *uni* should be penalized by the PWC. This condition is exactly comparable to the preceding consonant conditions in the earlier experiments (e.g., *ari* in *rari*), except that there is an additional initial mora *gya* and thus that the boundary before the critical medial consonant is cued by mora structure, rather than by silence. The addition of the CCV mora makes the moraic nasal context (*gyaNuni*) legal. In this case, and in the vowel context (*gyaouni*), the target word is aligned with a mora boundary (*gya-N-u-ni*; *gya-o-u-ni*). If the results with following moraic nasals in Experiments 1 and 3 replicate, and our interpretation of them is correct, then listeners should find it just as easy to spot words in the CCVN context as in the CCVV context, and both of these contexts should be easier than the CCVC context.

Experiment 5 also provided the opportunity to deal with a lingering concern about pitch accent patterns. In the earlier experiments, the words in preceding moraic nasal and vowel contexts were produced with accent patterns which did not correspond to the accent patterns they would have if spoken in isolation. The words in preceding consonant contexts were, however, produced with their citation pitch accent patterns. As we have already argued, it is unlikely that listeners would find the noncitation forms of the words harder to recognize (since they were, after all, produced in the way that was appropriate for those particular contexts). Furthermore, to the extent that the listeners did have difficulty with the words in the preceding vowel contexts, this would work against the difference predicted by the PWC (that words in consonant

contexts should be harder to segment than those in vowel contexts). Nevertheless, we decided to control for accent patterns across contexts in Experiment 5. All target words were bimoraic and had HL accent patterns (were they to be produced in isolation). The 4-mora sequences (CCVN and CCVV, e.g., *gyaNuni* and *gyaouni*) were produced with LHHL patterns (which are themselves well-formed sequences), and the 3-mora sequences (CCVC, e.g., *gyabuni*) were produced with (legal) LHL patterns. The citation HL pattern of the target words was thus preserved in all three contexts, providing a more controlled test of the PWC than was possible in the earlier experiments.

Method

Participants. Fifty-four student volunteers received course credits for their participation. They had not taken part in any of the earlier experiments and were all native speakers of Tokyo Japanese.

Materials. Twenty-four bimoraic target words were selected following the constraints used in the earlier experiments. These words were placed in three different preceding contexts: a CCV mora plus a moraic nasal (CCVN; e.g., *uni*, sea urchin, in *gyaNuni*), the same CCV mora plus a vowel (CCVV; e.g., *gyaouni*), and the CCV mora plus a consonant (CCVC; e.g., *gyabuni*). These materials are listed in Appendix C.

Twenty-four additional bimoraic target words (taken from the list of words with following contexts used in Experiment 3) were selected and placed in following contexts. Eight were given CV contexts (e.g., *chizu*, map, in *chizuto*), eight were given vowel contexts (e.g., *moya*, mist, in *moyau*), and eight were given moraic nasal contexts (e.g., *saru* in *saruN*). Ninety-six fillers were made in which embedded words were again avoided as much as was possible given the constraints of the Japanese language. Half of the fillers were matched in phonological structure to the target bearing items. There were thus 24 bimoraic nonsense sequences with preceding contexts (8 CCVN, 8 CCVV, and 8 CCVC) and 24 bimoraic nonsense sequences with following contexts (8 CV, 8 V, and 8 N). The other 48 fillers were trimoraic (CVCVCV)

nonsense sequences. A practice set, consisting of six target-bearing items and twelve fillers with the same phonological structures as the other materials, was also made.

Three lists were constructed. As in the earlier experiments, all three lists consisted of the same pseudo-random sequence of the 48 targets and 96 fillers (there was again always at least one filler after each target-bearing item). Each list contained all 24 targets with following contexts. The only difference between the lists was in the target-bearing items with preceding contexts, that is, in the contexts in which the targets appeared. The three contexts were counterbalanced evenly across the lists, such that each list contained eight targets in each type of context. As in the earlier experiments, listeners were unable to predict which items might contain targets, whether a target would be at the beginning or the end of a nonsense sequence, or what the type of context might be.

Procedure. The materials were again recorded by the second author in a sound-damped booth onto DAT tape, sampling at 48,000 Hz. The HL accent patterns of all 24 target words were preserved in each context: the items with CCVN and CCVV contexts were produced with LHHL patterns; the items with CCVC contexts were produced with LHL patterns. The procedure was identical to that in Experiments 1 and 3. Listeners were again given examples of targets in each of the possible contexts. Eighteen subjects heard each of the three lists; all subjects heard the same practice list.

Results and Discussion

RTs were adjusted by subtraction of the measured duration of the appropriate target word from each raw RT to measure from the offset of each target word. Responses slower than 2500 ms were again not recorded and thus any such responses were treated as errors. There was no further exclusion of very fast or very slow responses. Eight manual responses (0.6%) were accompanied by an incorrect spoken response. Four of these responses were the word *ruri* (melon) to *pyururi*; this word had been overlooked during stimulus construction. All other subjects who heard *pyururi* failed to detect any

target in it. The item *uri* was therefore removed from the analysis. The other four responses accompanied by incorrect spoken responses were treated as errors. One target (*uzu* in *myueuzu*) was missed by all subjects who heard it; this item was also removed from the analysis. Four more items (*ego*, *aku*, *ane*, and *ema*) were excluded on the basis of lexical decision performance when these targets were presented to listeners without their contexts (Experiment 6). These words were particularly hard to recognize, irrespective of which type of context they had appeared in. Each of these words was missed in all three conditions (i.e., with the words taken from each of the three contexts) by more than half of the subjects who heard them. The responses to 18 words therefore remained in the analysis. Missing data for each subject (or item) were replaced with the mean of the available RTs for that subject (item) within the CCVN and CCVV context conditions. Mean RTs (measured from target offset) and mean error rates (percentage of targets missed) are shown in Table 5.

Listeners found it almost completely impossible to spot words in the CCVC contexts. There were only two words spotted, each by only one subject (*uki* in *hyupuki* and *ono* in *chuzono*). ANOVAs by both participants and items on the error rates showed that the effect of context was significant ($F1(2,102) = 382.8$, $p < .001$; $F2(2,34) = 145.0$, $p < .001$). Planned comparisons confirmed that performance in the CCVC

TABLE 5

Mean Reaction Times for Correct Detection (RT, in ms), Measured from Word Offset, and Mean Percentage Missed Words (Errors), in Experiment 5

Context	CCVN	CCVV	CCVC
Example	gyaNuni	gyaouni	gyabuni
RT	692	755	—
Errors	30%	22%	99%

Note. CCVN, consonant-consonant-vowel-moraic nasal; CCVV, consonant-consonant-vowel-vowel; CCVC, consonant-consonant-vowel-consonant. No mean latency for targets in the CCVC condition is given since no meaningful average RT could be computed in this condition.

condition was worse than in each of the other two conditions: with CCVN, $t1(53) = 21.0$, $t2(17) = 12.8$; with CCVV, $t1(53) = 29.2$, $t2(17) = 15.1$. Listeners missed more targets in CCVN contexts (e.g., *uni* in *gyaNuni*) than in CCVV contexts (e.g., *uni* in *gyaouni*) but this difference was only significant by subjects: $t1(53) = 2.4$, $t2(17) = 1.6$, $p > .1$. As in Experiments 1 and 3, analyses of lexical embedding based on the on-line dictionary (Matsumura, 1993) were carried out. The items were split into two groups, one group in which there was an embedded word in the CCVC condition, from the vowel in the first mora of the context to the end of the first mora of the target (e.g., *abu*, horsefly, in *gyabuni*) but not in either the CCVV or CCVN conditions (11 items) and one group in which there were no such embedded words in any condition (six items; one item, *ibo*, was excluded since *oai*, companion, in the CCVV context is a word but *oNi* in the CCVN context is not). In an ANOVA on the error data, there was no interaction of this lexical embedding factor with the context factor. There were very few lexical competitors beginning with the vowel of the context plus the material up to the end of the first mora of the target, but again there were more competitors (2.0, on average) in the CCVC context than in the other two contexts (means of 0.0 and 0.1 competitors in the CCVN and CCVV contexts, respectively). There were, however, no significant correlations of mean error rate with number of lexical competitors.

The RTs in the CCVN and CCVV conditions were also analyzed. Listeners spotted words in CCVN contexts faster than in CCVV contexts, but this difference was only significant by subjects: $F1(1,51) = 6.3$, $p < .05$; $F2(1,17) = 1.7$, $p > .2$. There was thus a weak speed–accuracy trade-off: Responses in CCVN contexts were faster but more errorful than those in CCVV contexts. Neither the latency nor the accuracy difference was fully reliable, however, and both the CCVN and the CCVV conditions were substantially easier than the CCVC condition. It was not necessary to perform an analysis of the RT data based on embedded words, since, apart from in the item *nyoaibo*, there were no embedded words from the vowel in the first mora of

the context to the end of the first mora of the target in either the CCVN or the CCVV contexts. There were no significant correlations of mean RT with number of lexical competitors.

The present results thus suggest, in keeping with the results of Experiments 1 and 3, that Japanese listeners find it relatively easy to spot words which are aligned with mora boundaries (whether the segment neighboring the target is a moraic nasal or a vowel) and find it very hard to spot words which are misaligned with a mora boundary (as in the CCVC condition). It remains possible, however, that these results reflect acoustic differences between the target words in different contexts rather than segmentation difficulty. As with Experiments 1 and 3, therefore, a control lexical decision experiment was run in which the target words from Experiment 5 were presented in isolation. Note that since the pitch accent patterns of the targets were the same across all three contexts, and were appropriate for the targets when spoken in isolation, this experiment provides an even better control than Experiments 2 and 4 did. Any difference found between the targets would have to be due to other acoustic differences between the targets and not to differences in pitch accent patterns.

EXPERIMENT 6

Method

Participants. Forty-two volunteers from the same population as was used in all the earlier experiments were tested. They had not taken part in any of the other experiments and received course credit for their participation.

Materials and procedure. These were the same as those of Experiments 2 and 4, except that the words and nonwords were excised from the Experiment 5 materials.

Results and Discussion

Raw RTs were adjusted as in Experiment 5 to measure from the offset of each word. Five manual responses (0.5%) were accompanied by an incorrect oral response and were therefore treated as errors. Responses slower than 2500 ms were not recorded and thus any such responses were

also treated as errors. In the other two lexical decision experiments, responses slower than 2000 ms were excluded as outliers. Since there were no such responses in this experiment, no other trimming of the RT data was required. The six words which were removed in Experiment 5 were removed here. Missing data were replaced in the same way as in Experiment 5. Mean RTs (measured from target offset) and mean error rates (percentage of targets missed) are shown in Table 6.

In ANOVAs on RTs, there was a significant effect of context: $F1(2,78) = 5.3$, $p < .01$; $F2(2,34) = 3.9$, $p < .05$. Listeners detected words taken from CCVN contexts (e.g., *uni* from *gya-Nuni*) more slowly than words taken from either CCVV contexts (e.g., *uni* from *gyaouni*; $t1(41) = 2.3$, $t2(17) = 2.2$) or CCVC contexts (e.g., *uni* from *gyabuni*; $t1(41) = 2.3$, $t2(17) = 2.1$). There was no significant RT difference between the CCVV and CCVC conditions.

In ANOVAs on errors, the context effect was significant by subjects, but not quite by items: $F1(2,78) = 13.8$, $p < .001$; $F2(2,34) = 3.2$, $p = .05$. This effect was again due to poorer performance in the CCVN condition. Listeners missed more words taken from CCVN contexts than words taken from either CCVV contexts ($t1(41) = 4.6$, $t2(17) = 2.5$) or CCVC contexts, though this latter difference was only significant by subjects ($t1(41) = 2.1$, $t2(17) = 1.4$, $p > .1$). There was no significant difference in error rates between the CCVV and CCVC conditions.

It is quite clear that listeners did not have difficulty spotting words in CCVC contexts because those words were somehow acoustically degraded relative to the words in the other contexts. In isolation, the words from CCVC contexts were detected the most rapidly, and, although they were missed more often than the words from the CCVV contexts, this difference was not significant. When an ANCOVA on the Experiment 5 error rates was carried out, with the Experiment 6 error rates as covariate, the context effect remained highly significant ($F2(2,33) = 141.1$, $p < .001$). It appears instead that spotting words in CCVC contexts was hard because listeners had particular difficulty segmenting the words from those contexts.

Experiment 6 also examined the difference between the CCVV and the CCVN contexts. Remember that in Experiment 5, although neither effect was fully reliable, responses to targets in CCVN contexts were faster but more errorful than those to targets in CCVV contexts. In Experiment 6, responses to targets taken from CCVN contexts were reliably slower and more errorful than those to targets taken from CCVV contexts. The words in CCVN contexts may therefore have been less clear tokens than those in CCVV contexts. Note that an ANCOVA on the Experiment 5 RTs in these two conditions, with the Experiment 6 RTs as covariate, showed that the difference in word-spotting latencies remained nonsignificant ($F2 < 1$), and a similar ANCOVA on errors showed that the difference in word-spotting accuracy also remained nonsignificant ($F2(1,16) = 2.6$, $p > .1$). The most important result here is that even if the words in CCVN contexts were somewhat poorer tokens, word-spotting performance in this condition remains statistically indistinguishable from that in the CCVV condition, and both of these conditions are clearly very different from the CCVC condition. As predicted, Japanese listeners find it easier to spot words which are aligned with mora boundaries than to spot words which are misaligned by a single consonant with mora boundaries.

GENERAL DISCUSSION

In Japanese, segmentation is based on the mora. This conclusion was drawn from target-monitoring (Otake et al., 1993) and phoneme

TABLE 6

Mean Reaction Times for Correct Detection (RT, in ms), Measured from Word Offset, and Mean Percentage Missed Words (Errors), in Experiment 6

Original Context	CCVN	CCVV	CCVC
Example	(gyaN)uni	(gyao)uni	(gyab)uni
RT	622	552	541
Errors	36%	20%	26%

Note. CCVN, consonant-consonant-vowel-moraic nasal; CCVV, consonant-consonant-vowel-vowel; CCVC, consonant-consonant-vowel-consonant.

detection studies (Cutler & Otake, 1994; Otake et al., 1996); the same conclusion can be drawn from the present word-spotting experiments. Words were relatively easy to spot when they were aligned with mora boundaries and hard to spot when they were misaligned with mora boundaries. Specifically, in Experiments 1 and 3, target words were aligned with mora boundaries in the preceding vowel conditions, and Japanese listeners found them easier to spot than the target words in the preceding consonant conditions, which were all misaligned with mora boundaries. Furthermore, in Experiment 5, target words were aligned with mora boundaries in the CCVV and CCVN conditions (e.g., *uni* in *gyaouni* and *gyaNuni*, respectively), and Japanese listeners again found them much easier to spot than the target words in the CCVC conditions (e.g., *uni* in *gyabuni*), which were misaligned with mora boundaries (and almost impossible to detect).

These results support the account of lexical segmentation provided by the PWC and reflect similar results in English (Norris et al., 1997) and Dutch (McQueen & Cutler, 1998). Experiments 1 and 3 are novel in that they show an effect of the possible word status of contexts which are exactly controlled for phoneme length: both the vowel (possible word) and consonant (impossible word) contexts were one phoneme long. According to the PWC account, the nearest likely word boundary to the edge of a candidate word is used in the computation of whether that candidate meets the constraint. When there is no mora boundary aligned with the beginning of the word, as with *ari* in *rari*, the next likely word boundary yet earlier in time is used in the computation, in this case the silence preceding the entire string. The stretch of speech between the beginning of the word and this likely word boundary is a single nonmoraic consonant, that is, an impossible word in Japanese. The activation of the activated candidate word *ari* is thus penalized, making word-spotting more difficult. Experiments 2 and 4 showed that the difficulty listeners had in spotting words in preceding consonant contexts was not due to an acoustic confound; indeed, words taken from these contexts were the easiest to identify.

The PWC also accounts for listeners' extremely poor performance in the CCVC condition in Experiment 5 (e.g., detecting *uni* in *gyabuni*). Again, there were no mora boundaries aligned with word onsets in this condition, so again the closest likely word boundary earlier in the sequence is used in the PWC computation. In this case, the boundary is between the first and the second mora in the sequence (e.g., before the [b] in *gyabuni*) and again an impossible Japanese word (the consonant [b] in the example) occurs between the likely boundary and the word onset. The activation of the candidate word *uni* should thus be penalized, making word spotting hard.

The effect in Experiment 5 was considerably larger than the effects in Experiments 1 and 3. One factor which may have contributed to this difference is that pitch accent patterns were controlled in Experiment 5 but not in the earlier experiments. The target words in Experiment 5 had the same accent patterns across all conditions, and this pattern was the correct pattern for the target words when spoken in isolation. In Experiments 1 and 3, the accent patterns of the words in the preceding consonant contexts were also those that are used when the targets are spoken in isolation. But in the preceding vowel contexts the accent patterns of the words were changed from those used in isolated productions (e.g., LHH words like *agura* had HHH patterns, in LHHH *oagura*). If this change in the citation accent patterns of the words in the vowel contexts made them harder to recognize than the words in the consonant contexts (e.g., *agura* kept its LHH pattern in *tagura*), this would have tended to weaken any PWC effect. Since there was no such accentual difference in Experiment 5, the PWC effect could thus have been larger.

While this remains a possible explanation, it cannot be the only one. First, note that while performance did improve (in both speed and accuracy) from the preceding vowel condition in Experiment 3 to the CCVV condition in Experiment 5 (as the accent-based explanation predicts), the major component of the increased effect size in Experiment 5 is the increase in the error rates in the CCVC condition relative to the earlier single consonant conditions (which the accent-based explanation does not predict).

Second, we think it is unlikely that the accent patterns of the words in the vowel conditions in Experiments 1 and 3 would be particularly problematic. Japanese listeners appear to use accent patterns to constrain spoken word recognition (Cutler & Otake, 1999), and misaccented Japanese words are indeed harder to recognize than correctly accented words (Minematsu & Hirose, 1995). But misaccenting in sentences has a weaker effect (Minematsu & Hirose, 1995), and, in any case, the words in the preceding vowel contexts in Experiments 1 and 3 were not misaccented: They were accented in a way that was appropriate for those contexts. If the sequences had been meaningful, the target words could have been spoken in the vowel contexts with the accent patterns which we used.

A more likely reason why the effect in Experiment 5 was larger than the effects in the earlier experiments may have been that listeners found it easier to recover from the segmentation problem caused by a string such as *rari* or *tagura* than to recover from the segmentation problem caused by a string such as *gyabuni*. If the PWC account is correct, then words in vowel contexts (either the single vowel contexts in Experiments 1 and 3 or the CCVV contexts in Experiment 5) should not be penalized by the PWC mechanism. These words should thus tend to win the competition process and be recognized in the same way as they would be in normal listening. In other words, there should be no segmentation problem to be solved for these items. But in consonant contexts (both the single consonant contexts in the earlier experiments and the CCVC contexts in Experiment 5) the target words are penalized by the PWC, and thus a segmentation problem remains for these items. Once the activation of one of these words has been halved, it will be harder for that word to win the competition process. According to the PWC account, this activation penalty will be the same irrespective of the length or complexity of the target word or of the length, position, or complexity of the context; the only thing which matters is whether or not there is a possible word between the word and the location of a possible word boundary. In other words, the same penalty is applied for all words that are misaligned with a likely word

boundary. It is possible that there was a stronger effect in Experiment 5, however, because there was stronger competition from other candidate words in Experiment 5 (where the target words were proportionally smaller parts of the complete nonsense sequence) than in Experiments 1 and 3 (where the target words formed proportionally more of the complete sequence).

In normal listening, the PWC penalty acts to suppress the activation of spurious competitors, allowing the speaker's intended words to dominate the activation pattern. Words which fail the PWC will therefore tend not be recognized. In the word-spotting task, however, listeners are encouraged to spot any embedded words, even those which fail the PWC. Listeners may therefore adopt task-specific strategies in order to find words which (through the operation of the PWC and lexical competition) would otherwise be missed. When no word is immediately recognized in a nonsense sequence, listeners may attempt some kind of search for words in that sequence (held in short-term memory). In contrast to normal recognition processes, this search process is likely to be slow and effortful. Remember that in all three word-spotting experiments, however, listeners could not predict either where the target words would be (at the beginning or end of the string) or what kind of context there would be. They could not even predict whether there would be a target or not on a given trial (and on the majority of trials there was no target). The use of search strategies is therefore likely to be limited to situations where automatic recognition processes fail. The PWC predicts that normal recognition will be very slow, or will fail, when impossible words would be part of the lexical parse, as in the consonant and CCVC contexts.

Another reason why the PWC effect in Experiment 5 was larger than the effects in Experiments 1 and 3, therefore, may be that listeners were more successful in applying a search strategy in the consonant-context conditions in the earlier experiments than in the CCVC condition in Experiment 5. Although there are no directly comparable experiments available, and certainly none in Japanese, we note that word spotting does tend to be harder for misaligned words in

longer contexts (misaligned Dutch words such as *rok*, skirt, in *fi.drok*, where the context is more than one syllable, as in the present Experiment 5 CVCC condition, were missed, on average, on 59% of trials; McQueen, 1998, Experiment 1) than for misaligned words in shorter contexts (misaligned English words such as *egg* in *fegg*, where the context is a single consonant, as in the present Experiments 1 and 3, were missed, on average, on 41% of trials; Norris et al., 1997, Experiments 1 and 2). While this comparison must be treated with caution, it is at least consistent with the present data. Spotting misaligned words (i.e., words which have had the PWC penalty applied to them) may be harder in longer contexts than in shorter contexts because of greater lexical competition from other candidates in longer contexts and/or because search strategies are less likely to succeed, within a limited response time, when more context needs to be searched. More research is required to establish exactly why the CVCC condition was so difficult. We suspect that the answer will lie in the processes which act to recover from the application of the PWC (e.g., competition, search strategies) rather than in the operation of the PWC itself.

The results for words in preceding consonant and CCVC contexts thus suggest that Japanese listeners, like English and Dutch listeners, use the PWC in segmenting continuous speech: Lexical parses of the speech input which include impossible words are disfavored. The results with moraic nasals further extend this view of lexical segmentation. They suggest that the PWC operates very locally. Only the nearest likely word boundary to the edge of the candidate word matters in the computation. Thus, even though moraic nasals are impossible words because they are single consonants, this does not cause any deactivation of the candidate word. Because a likely word boundary intervenes between the word and the nasal consonant, the PWC penalty does not apply to the word. This was found in the following context conditions in Experiments 1 and 3 (e.g., *saru* was just as easy to spot in *saruN* as in *sarua*), and in Experiment 5 (e.g., *uni* was spotted just as easily in *gyaNuni* as in *gyaouni*, and both of these contexts were

much easier environments for word spotting than was *gyabuni*).

It does not seem to matter in Japanese speech segmentation that a moraic nasal is not a possible Japanese word. The possible-word status of a moraic nasal does not seem to enter into the PWC computation because, for any word which occurs next to a moraic nasal, there must necessarily be a mora boundary (i.e., a likely word boundary) between the word and the moraic nasal. We have suggested that the boundaries after moraic nasals are likely to be signaled, at least in part, by the duration of the moraic nasals. While more work remains to be done to establish the precise acoustic correlates of moraic rhythm in the speech signal, it seems clear that Japanese listeners use moraic rhythm in their segmentation of continuous speech.

The present experiments therefore support the theory of lexical segmentation that the PWC offers. On this view, candidate words are activated by the incoming speech stream and compete with each other until a lexical parse is settled upon. Two sources of information act to bias this competition-based segmentation procedure. One is the presence in the signal of multiple cues to likely word boundaries. Although the focus of the present study has been on the cues provided by language rhythm, other cues, such as those based on allophonics and phonotactics, also appear to be used by listeners. These cues are language-specific. Note that they vary from language to language not in the sense that some languages, for example, might encourage the use of rhythmic cues in segmentation while others might encourage the use of phonotactic cues. Instead, they vary across languages in the sense that languages differ in their metrical and phonological properties. Listeners therefore use cues based on language-specific phonological structures.

Under the PWC account, these cues provide the segmentation process with probabilistic information about the location of likely word boundaries; they do not indicate where word boundaries must be. Candidate words are evaluated with respect to these likely boundaries and are penalized only if they are misaligned with the boundaries. Thus, in Japanese, recognition

of words with more than one mora (as in all the present experiments) will not be impaired, even though these words have internal mora boundaries. Such words are not misaligned with their own internal mora boundaries and thus will not be penalized by the PWC. But an embedded word which is misaligned with a word-internal mora boundary will be penalized by the PWC. For example, in *atari* (hit), there is a single non-moraic consonant, [t], between the beginning of the embedded word *ari* (ant) and the internal mora boundary between the first vowel and the [t]. The PWC penalty would therefore be applied to *ari* in *atari*. Since the activation of *ari* would therefore be reduced, *atari* would more easily win in the lexical competition process.

The other source of information used in the competition-based segmentation process is knowledge about whether or not a stretch of speech can count as a possible word in the ongoing lexical parse. The present research, along with earlier work in English (Norris et al., 1997) and Dutch (McQueen & Cutler, 1998), suggests that single consonants fail the constraint (i.e., are treated as impossible words in the ongoing lexical parse) while stretches of speech including at least a vowel pass the constraint (i.e., are treated as possible words in the lexical parse). New research on Sesotho (Cutler, Demuth, & McQueen, submitted) suggests that this may be a language-universal constraint. In Sesotho, a Bantu language spoken in southern Africa, there is a phonological restriction which requires that all content words have at least two syllables (or, more precisely, two morae). In a word-spotting experiment, however, Sesotho listeners found it as easy to spot words in monosyllabic (mono-moraic) CV contexts as to spot words in bisyllabic (bimoraic) CVCV contexts. In other words, although the single CV is not a possible word in the Sesotho language, it causes no segmentation problems; it passes the PWC.

Further evidence that the operation of the PWC does not depend on language-specific phonological constraints on the well-formedness of words comes from English. Norris, McQueen, Cutler, Butterfield, and Kearns (in press) found that English listeners could detect words like *canal* more easily in *zecanal* than in *scanal*, even though the former context [zɛ], which is an open syllable with a lax vowel, is an impossible lexical word of English. In a second experiment, Norris et al. (in press) found that English listeners could detect words like *sea* more easily in *seashəb* than in *seash*, even though the context [ʃəb] is an impossible English content word (weak syllables, i.e., syllables with the vowel schwa, cannot be content words in English). In each experiment, a syllabic context proved to be easier to segment a word from than a consonantal context.

Thus, the PWC, as a mechanism used in online speech segmentation, appears not to be sensitive to the variability between languages in what counts as a well-formed word. What counts as a possible word for the PWC is not whether a sequence of sounds constitutes a phonologically acceptable word in the native language of any one listener, but whether that sound sequence, irrespective of the listener's native language, consists of only consonantal material. The PWC therefore operates on the language-universal constraint that no chunks in the lexical parse can contain only consonants.

The PWC appears to be used for the segmentation of continuous Japanese, as it is for the segmentation of English, Dutch, and Sesotho. Japanese listeners, like native speakers of these other languages, are sensitive to the viability of sound sequences as possible words in the speech stream. Furthermore, just as the rhythmic structure of English or Dutch provides English and Dutch listeners with cues to the location of likely word boundaries (Cutler & Norris, 1988, Vroomen et al., 1996), so too does the characteristic rhythm of Japanese provide Japanese listeners with a segmentation cue. We have argued that mora boundaries signal likely word boundaries in Japanese. This means that, in Japanese, words are no harder to spot in moraic nasal contexts than in vowel contexts, even though moraic nasals are impossible words. The results with moraic nasal contexts also suggest that the computation of viability may be very local. Only the speech material between a candidate word's edge and the closest likely word boundary is evaluated by the PWC.

APPENDIX A

Target-Bearing Materials from Experiment 1 in Each of the Three Contexts

Target	Accent pattern	Translation	Consonant	Vowel	Moraic nasal
Targets with preceding contexts					
agura	LHH	(sit cross-legged)	tagura	oagura	Nagura
atari	LHH	(hit)	satari	uatari	Natari
azuki	LHH^	(azuki bean)	razuki	eazuki	Nazuki
aroe	LHH	(aloe)	karoe	earoe	Naroe
ibiki	LHH^	(snore)	nibiki	uibiki	Nibiki
ibitsu	LHH	(distortion)	ribitsu	uibitsu	Nibitsu
odeko	LHL	(forehead)	todeko	eodeko	Nodeko
odori	LHH	(dance)	rodori	eodori	Nodori
odoshi	LHH	(threat)	kodoshi	eodoshi	Nodoshi
ogori	LHH	(treat)	mogori	iogori	Nogori
ogura	LHH	(sweet azuki bean)	nogura	eogura	Nogura
okowa	LHL	(a kind of rice)	zokowa	eokowa	Nokowa
omoshi	LHH	(weight)	somoshi	eomoshi	Nomoshi
omote	LHH^	(outside)	romote	iomote	Nomote
omoyu	LHH	(rice water)	romoyu	eomoyu	Nomoyu
omutsu	LHL	(diaper)	homutsu	iomutsu	Nomutsu
onaji	LHH	(same)	monaji	eonaji	Nonaji
owari	LHH	(end)	nowari	eowari	Nowari
oyogi	LHH^	(swim)	hoyogi	ioyogi	Noyogi
ugoki	LHH^	(movement)	hugoki	eugoki	Nugoki
unaji	LHH	(nape)	kunaji	eunaji	Nunaji
uroko	LHH	(scale)	nuroko	euroko	Nuroko
uwabe	LHH	(surface)	huwabe	euwabe	Nuwabe
uwasa	LHH	(gossip)	ruwasa	euwasa	Nuwasa
Targets with following contexts					
bikini	HLL	(bikini)	bikinip	bikinia	bikiniN
garasu	LHH	(glass)	garasug	garasui	garasuN
gojira	HLL	(godzilla)	gojirab	gojirau	gojiraN
gorira	HLL	(gorilla)	gorirak	gorirai	goriraN
goruhu	HLL	(golf)	goruhub	goruhua	goruhuN
gurabu	HLL	(glove)	gurabuk	gurabua	gurabuN
gurasu	HLL	(glass)	gurasud	gurasua	gurasuN
higashi	LHH	(east)	higaship	higashia	higashiN
hotoke	LHH	(Buddha)	hotoked	hotokeu	hotokeN
karuta	HLL	(Japanese playing cards)	karutap	karutau	karutaN
medaru	LHH	(medal)	medarus	medarua	medaruN
megane	HLL	(spectacles)	meganes	meganei	meganeN
mogura	LHH	(mole)	mogurak	mogurau	moguraN
nameko	LHH	(nameko mushrooms)	namekok	namekoa	namekoN
nigari	LHH	(bittern)	nigarib	nigariu	nigariN
nigori	LHH^	(muddiness)	nigorik	nigoria	nigoriN
nobori	LHH	(up-train)	noborip	noboriu	noboriN
pedaru	LHH	(pedal)	pedarub	pedarua	pedaruN
poteto	HLL	(potato)	potetok	potetoa	potetoN
puragu	HLL	(plug)	puragus	puragua	puraguN
purasu	HLL	(plus)	purasug	purasua	purasuN
puresu	HLL	(press)	puresub	puresua	puresuN
rizumu	HLL	(rhythm)	rizumug	rizumui	rizumuN
saguri	LHH	(probe)	sagurig	saguria	saguriN
terebi	HLL	(television)	terebip	terebia	terebiN
tsuzumi	LHH	(hand drum)	tsuzumik	tsuzumia	tsuzumiN

APPENDIX B

Target-Bearing Materials from Experiment 3 in Each of the Three Contexts

Target	Accent pattern	Translation	Consonant	Vowel	Moraic nasal
		Targets with preceding contexts			
ase	HL	(sweat)	tase	iase	Nase
aki	HL	(autumn)	raki	eaki	Naki
asa	HL	(morning)	dasa	iasa	Nasa
ashi	LH^	(foot)	pashi	uashi	Nashi
ana	LH^	(hole)	zana	uana	Nana
ani	HL	(brother)	sani	eani	Nani
ane	LH	(sister)	rane	iane	Nane
ami	LH^	(net)	zami	uami	Nami
ari	LH	(ant)	rari	eari	Nari
awa	LH^	(bubble)	tawa	eawa	Nawa
isu	LH	(chair)	pisu	uisu	Nisu
ima	LH^	(living room)	nima	uima	Nima
ibo	HL	(wart)	mibo	uibo	Nibo
usu	HL	(mortar)	gusu	eusu	Nusu
uso	HL	(lie)	puso	iuso	Nuso
uta	LH^	(song)	guta	auta	Nuta
uchi	LH	(house)	zuchi	euchi	Nuchi
uni	HL	(sea urchin)	puni	iuni	Nuni
uma	LH^	(horse)	ruma	euma	Numa
uzu	HL	(whirlpool)	buzu	iuzu	Nuzu
eki	HL	(station)	keki	oeki	Neki
esa	LH^	(pet food)	resa	oesa	Nesa
eri	LH^	(collar)	reri	aeri	Neri
ebi	LH	(shrimp)	nebi	oebi	Nebi
		Targets with following contexts			
chizu	HL	(a map)	chizut	chizua	chizuN
moya	HL	(mist)	moyap	moyau	moyaN
kinu	HL	(silk)	kinup	kinua	kinuN
matsu	HL	(pine tree)	matsup	matsua	matsuN
haru	HL	(spring)	harup	harua	haruN
kuzu	HL	(trash)	kuzut	kuzua	kuzuN
naya	HL	(a shed)	nayat	nayau	nayaN
tetsu	LH	(iron)	tetsup	tetsua	tetsuN
natsu	LH^	(summer)	natsup	natsua	natsuN
fuyu	LH^	(winter)	fuyup	fuyua	fuyuN
yuzu	HL	(citron)	yuzut	yuzua	yuzuN
mizu	LH	(water)	mizup	mizua	mizuN
koya	LH^	(shed)	koyat	koyau	koyaN
hiru	LH^	(noon)	hiruk	hirua	hiruN
kaya	LH	(mosquito net)	kayap	kayau	kayaN
mitsu	HL	(honey)	mitsut	mitsua	mitsuN
saru	HL	(monkey)	sarup	sarua	saruN
netsu	LH^	(heat)	netsup	netsua	netsuN
gasu	HL	(gas)	gasup	gasua	gasuN
heya	LH^	(a room)	heyat	heyau	heyaN
tsuyu	LH^	(rainy season)	tsuyup	tsuyua	tsuyuN
zaru	LH^	(basket)	zarug	zarua	zaruN
kizu	LH	(cut)	kizup	kizua	kizuN
kutsu	LH^	(shoes)	kutsup	kutsua	kutsuN
maru	LH	(circle)	marut	marua	maruN
suzu	LH	(bell)	suzup	suzua	suzuN

APPENDIX C

Target-Bearing Materials from Experiment 5 in Each of the Three Contexts (All Words Had HL Accent Patterns)

Target	Translation	CCVN	CCVV	CCVC
ase	(sweat)	pyaNase	pyaiase	pyapase
aka	(red)	myaNaka	myaoaka	myapaka
ani	(brother)	nyaNani	nyaeani	nyabani
uki	(rainy season)	hyuNuki	hyuauki	hyupuki
uri	(melon)	pyuNuri	pyuiuri	pyururi
uzu	(whirlpool)	myuNuzu	myueuzu	myubuzu
eki	(station)	myoNeki	myoaeki	myokeki
ono	(axe)	chuNono	chuaono	chuzono
aki	(autumn)	kyaNaki	kyaeaki	kyapaki
ama	(woman diver)	byoNama	byoeama	byorama
aji	(horse mackerel)	byuNaji	byueaji	byubaji
ibo	(wart)	nyoNibo	nyoaibo	nyomibo
umi	(sea)	ryoNumi	ryoeumi	ryorumi
uso	(lie)	hyoNuso	hyoeuso	hyonuso
ego	(ego)	gyuNego	gyuaego	gyupego
oke	(tub)	syaNoke	shauoke	shasoke
asa	(morning)	nyuNasa	nyueasa	nyurasa
aku	(badness)	hyaNaku	hyauaku	hyapaku
ane	(sister)	kyuNane	kyuoane	kyupane
ato	(mark)	ryaNato	ryaeato	ryawato
usu	(mortar)	byaNusu	byaeusu	byagusu
uni	(sea urchin)	gyaNuni	gyaouni	gyabuni
ema	(votive tablet of a horse)	shoNema	shoaema	shorema
oku	(inner part)	ryuNoku	ryuaoku	ryunoku

REFERENCES

Beckman, M. E. (1982). Segment duration and the "mora" in Japanese. *Phonetica*, **39,** 113–135.

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 152–178), Cambridge, UK: Cambridge University Press.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, **31,** 218–236.

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma and D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183–196), Hillsdale, NJ: Erlbaum.

Cutler, A., Demuth, K., & McQueen, J. M. (2001) *Universality versus language-specificity in listening to running speech*. Manuscript submitted for publication.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics,* **21,** 103–108.

Cutler, A., Mehler, J., Norris, D. G., & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25,** 385–400.

Cutler, A., Mehler, J., Norris, D. G., & Seguí, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24,** 381–410.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14,** 113–121.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, **33,** 824–844.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, **105,** 1877–1888.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, **24,** 423–444.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, **21,** 344–359.

Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America*, **96,** 73–82.

Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics*, **9**, 273–281.

Klatt, D. H. (1974). The duration of [S] in English words. *Journal Speech and Hearing Research*, **17**, 51–63.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129–140.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208–1221.

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, **5**, (Suppl. 5), 1–54.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, **51**, 2018–2024.

van der Lugt, A. H. (in press) *The use of sequential probabilities in the segmentation of speech. Perception & Psychophysics.*

Matsumura, A. (1993). *Daijirin CD-ROM* (version for Windows). Tokyo: Sanseido.

McQueen, J. (1996). Word spotting. *Language and Cognitive Processes*, **11**, 695–699.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, **39**, 21–46.

McQueen, J. M., & Cutler, A. (1998). Spotting (different types of) words in (different types of) context. *Proceedings of the 5th International Conference on Spoken Language Processing*, (Vol. 6, pp. 2791–2794), Sydney, Australia.

McQueen, J. M., Norris, D. G., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 621–638.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U. H., & Seguí, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, **20**, 298–305.

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.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, **62**, 714–719.

Norris, D., McQueen, J. M., Cutler, A., Butterfield, S., & Kearns, R. K. (in press). Language-universal constraints on the speech segmentation. *Language and Cognitive Processes*,

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1209–1228.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, **34**, 191–243.

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

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, **54**, 1235–1247.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 258–278.

Otake, T., Yoneyama, K., Cutler, A., & van der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America*, **100**, 3831–3842.

Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllable structure of spoken words. *Journal of Memory and Language*, **32**, 373–389.

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.

Port, R. F, Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, **81**, 1574–1585.

Poser, W. J. (1984). *The phonetics and phonology of tone and intonation in Japanese*. Unpublished doctoral dissertation, MIT, Cambridge MA.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, **20**, 331–350.

Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America*, **94**, 2027–2035.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265–292.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, **35**, 606–621.

Sato, Y. (1993). The duration of syllable-final nasals and the mora hypothesis in Japanese. *Phonetica*, **50**, 44–67.

Sebastián-Gallés, N., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, **31**, 18–32.

Shillcock, R. C. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49), Cambridge, MA: MIT Press.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, **36**, 422–444.

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, **34**, 440–467.

Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany: State University of New York Press.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, **40**, 374–408.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition.

Journal of Experimental Psychology: Human Perception and Performance, **21,** 98–108.

Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **23,** 710–720.

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.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, **24,** 744–755.

Wallace, W. P., Stewart, M. T., & Malone, C. P. (1995). Recognition memory errors produced by implicit activation of word candidates during the processing of spoken words. *Journal of Memory and Language*, **34,** 417–439.

Wallace, W. P., Stewart, M. T., Shaffer, T. R., & Wilson, J. A. (1998). Are false recognitions influenced by prerecognition processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 299–315.

Wallace, W. P., Stewart, M. T., Sherman, H. L., & Mellor, M. (1995). False positives in recognition memory produced by cohort activation. *Cognition*, **55,** 85–113.

Warner, N., & Arai, T. (2001). *The role of the mora in the timing of spontaneous Japanese speech. Journal of the Acoustical Society of America,* **109,** in press.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, **32,** 25–64.

Zwitserlood, P., & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, **10,** 121–136.