Journal of Memory and Language **46,** 296–322 (2002)
doi:10.1006/jmla.2001.2814, available online at http://www.academicpress.com on IDEAL®

Rhythmic Categories in Spoken-Word Recognition

Anne Cutler

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

and

Takashi Otake

Dokkyo University, Saitama, Japan

Rhythmic categories such as morae in Japanese or stress units in English play a role in the perception of spoken language. We examined this role in Japanese, since recent evidence suggests that morae may intervene as structural units in word recognition. First, we found that traditional puns more often substituted part of a mora than a whole mora. Second, when listeners reconstructed distorted words, e.g. *panorama* from *panozema*, responses were faster and more accurate when only a phoneme was distorted (*panozama*, *panorema*) than when a whole CV mora was distorted (*panozema*). Third, lexical decisions on the same nonwords were better predicted by duration and number of phonemes from nonword uniqueness point to word end than by number of morae. Our results indicate no role for morae in early spoken-word processing; we propose that rhythmic categories constrain not initial lexical activation but subsequent processes of speech segmentation and selection among word candidates.

Key Words: rhythmic categories; spoken-word recognition; Japanese; mora; phonemes.

Cross-linguistic studies of spoken-language perception have shown that speakers of different languages are sensitive to differing levels of structure in speech. English listeners are sensitive to the boundaries between stress units (Cutler & Butterfield, 1992; Cutler & Norris, 1988), French listeners are sensitive to the boundaries between syllables (Cutler, Mehler, Norris, & Segui, 1986; Mehler, Dommergues, Frauenfelder, & Segui, 1981), and Japanese listeners are sensitive to the boundaries between morae (subsyllabic components; Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993; Otake, Hatano, & Yoneyama, 1996; Otake, Yonayama, Cutler, & Van der Lugt, 1996).

This body of work has shown that each of these units plays a role in the way listeners of the languages in question segment spoken input in order to find the words in a continuous speech stream as rapidly and as efficiently as possible. The units are also relevant for many other aspects of language processing; thus work on English, French, and Japanese has shown the importance of stress units, syllables, and morae in language acquisition (Echols, 1996; Inagaki, Hatano, & Otake, 2000; Mehler, Dupoux, & Segui, 1990) and in language production (Cutler & Young, 1994; Ferrand, Segui, & Grainger, 1996; Kubozono, 1989, 1990). Each of the units listed also corresponds to the unit most relevant for describing speech rhythms (and the rhythms of verse forms) in the language in question; from this it is concluded that language-specific

This research was financially supported by grants to the second author from the International Communication Foundation, Japan, and from the Japan Society for the Promotion of Science (Grant-in-Aid for Science Research (C) 11610566). We are grateful to Makiko Komada for help with Part I and to Risa Komaba for assistance with the construction of the materials for Experiments 1–4. We further thank James McQueen, Roel Smits, and Arthur Samuel for helpful comments on the text; Colin Brown for suggesting Experiment 5; and Harald Baayen for statistical assistance. A partial report of Part I was presented to the EUROSPEECH meeting in Aalborg, Denmark, in September 2001, and a brief report of Experiments 1–3 in Part II was presented to the 138th meeting of the Acoustical Society of America, Columbus, OH in November 1999.

Address correspondence and reprint requests to Anne Cutler, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. E-mail: anne.cutler@mpi.nl. Fax: +31 24 3521213.

0749-596X/01 $35.00

rhythmic categories play an important role in the processing of spoken language.

The language-specificity of this role is underlined by the fact that listeners are *not* sensitive to units relevant for other languages but irrelevant for their own; thus the work on speech segmentation has made clear that neither English nor Japanese listeners are sensitive to syllabic units (Cutler et al., 1986; Otake et al., 1993), and both French and English listeners are insensitive to morae (Cutler & Otake, 1994; Otake et al., 1993). However listeners *are* sensitive to their native categories in foreign-language input—French listeners to syllables in Japanese or English input (Cutler et al., 1986; Otake et al., 1993), Japanese listeners to morae in English and French input (Cutler & Otake, 1994; Otake et al., 1996).

Older models of spoken-language processing envisaged that such rhythmic categories might be the indivisible structures used to represent speech input and to access lexical forms (e.g., Mehler, 1981). Current models, however, propose a structure within which an early level of representation in terms of rhythmic categories has a less obvious role. Computational models of spoken-word recognition (e.g., Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986; Norris, 1994) assume that spoken input automatically activates lexical candidates in a continuous manner, without the necessary intervention of such representations as mora, syllable, or stress unit. There is abundant evidence for concurrent activation of multiple lexical candidates (e.g., Connine, Blasko, & Wang, 1994; McQueen, Norris, & Cutler, 1994; Tabossi, Burani, & Scott, 1995; Zwitserlood, 1989), for activation of word candidates on the basis of partial and imperfect evidence (e.g., Connine et al., 1994; Goldinger, Luce, Pisoni, & Marcario, 1992; Radeau, Morais, & Segui, 1995; Slowiaczek, Nusbaum, & Pisoni, 1987), and for continuous modulation of activation by early evidence of upcoming phonetic information (Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999).

A framework combining automatic continuous activation of word candidates with a role for rhythmic categories in the process of modulating activation by interword competition has been proposed by Norris, McQueen, Cutler, and Butterfield (1997), who provided both supporting experimental evidence and a computational implementation of the framework within the Shortlist model of spoken-word recognition (Norris, 1994). In Norris et al.'s proposal, activated lexical candidates are tested for the viability of their candidacy within their surrounding speech context. Candidates fail the test, and are consequently reduced in activation, if they would leave adjacent speech input which could not be itself a word. For instance, the spoken input *bring* might activate the word candidate *ring*, but accepting *ring* would leave [b] stranded, and [b] cannot by itself be a word. This Possible-Word Constraint is proposed by Norris et al. to apply for the context immediately adjacent to a word candidate, up to the nearest indicator of a relevant boundary. Boundaries relevant to the possible-word calculation can be, for instance, silence, or language-specific constraints on phonetic sequences within words and syllables, or, importantly for our present considerations, the boundaries of the rhythmic categories known to affect speech segmentation in the language in question.

For the English data on the role of stress units in speech segmentation, Norris et al. (1997) demonstrated that simulations with this framework could accurately capture the pattern of the empirical findings. Subsequent studies in Japanese by McQueen, Otake, and Cutler (2001) lent further support to the proposal in two ways: first, by showing that the Possible-Word Constraint held for the recognition of Japanese words, as it did for the recognition of English words in Norris et al.'s experiments; and second, by showing that boundaries of morae were relevant in assessing the size of the effect in Japanese. Specifically, a single-phoneme context which was in its own right a mora did not reduce activation of an adjacent word as drastically as did a single-phoneme context which was nonmoraic. It has also recently been claimed by Content, Kearns, and Frauenfelder (2001; see also Content, Meunier, Kearns, & Frauenfelder, 2001) that the role of the syllable in the recognition of spoken French accords with Norris et al.'s proposal; the syllable does not constitute an intermediate level of representation on which lexical access is based, but syllable boundaries are nevertheless relevant to lexical activation in that they constitute likely word boundaries.

In this framework all the rhythmic categories which have been extensively studied—stress unit, syllable, and mora—play the same role: their boundaries function as alignment points for lexical segmentation. In no language is a rhythmic category as such proposed to participate in any other way in the spoken-language recognition process.

These three rhythmic categories are, however, different in many ways. First, they differ in (average) size. A stress unit can consist of several syllables; consider the English phrase *government of a dominion,* which has the rhythm SWWWWWSW. Unless a speaker made one of the weak syllables strong, or deleted one or more weak syllables, the first stress unit in this phrase would contain six syllables. Thus the stress unit is larger in size than the syllable. The mora, on the other hand, is smaller in size than the syllable. In Japanese, a mora can be a consonant-vowel (CV) or consonant-glide-vowel string (e.g., *ka*, *ho*, *kyo*), a single vowel, or a syllable coda; *Honda*, for example, has three morae (*ho-n-da*). As this last example shows, the mora is a subsyllabic unit in that a syllable can contain more than one mora. A bimoraic syllable can have a coda, such as *hon* in *Honda* or *nik* in *Nikko*, or a long vowel (such as the two morae of each syllable of *Tokyo*, which has long vowels and is actually *to-o-kyo-o*).

Further, the rhythmic categories differ in the degree to which language users develop awareness of them as a result of conventions of language use. Because they play an important role in poetic forms, it is clear that poets, and others for whom poetic forms are important, will be able to manipulate rhythmic categories explicitly. But poetry aside, there are extensive differences in the availability of stress units, syllables, and morae to language users. In English and Dutch, stress units are not easily made explicit; there are no stress marks in the orthography, for instance, and no punctuation conventions invoke stress units. In French, syllables are only marginally more accessible; there is, for example, an orthographic convention by which words broken across handwritten lines are supposed to be broken at syllable boundaries, but it is questionable whether in an age of word processors this rule actually leads to widespread awareness of syllable boundaries.

In contrast, morae in Japanese are very accessible to language users. The kana orthographies explicitly encode mora structure, and many words (e.g., function words, loan words, and foreign names) are written in kana in Japanese text. Japanese language games are mora-based, e.g., the children's game of shiritori, in which players have to produce a word beginning with the final mora of the word given by the previous player. A player who produces a word ending with a mora which cannot be word-initial, e.g., *gohan*, "rice," ending with a moraic nasal, loses the game.

This difference in accessibility is arguably responsible for the differing role of rhythmic categories across languages in some metalinguistic tasks. For instance, explicit segmentations of words are based on morae in Japanese (Otake, Davis, & Cutler, 1995; Otake & Yamamoto, 1997), but are not based on stress units in English (Otake et al., 1995; Otake, Yoneyama & Maki, 1998; Otake, 2000).

The accessibility of morae in Japanese is, furthermore, accompanied by an inaccessibility of submoraic structure. Phonemic structure of words is (to a large extent, at least) accessible to speakers of languages with an alphabetic script by virtue of their literacy. Japanese is not written with an alphabet, and literate Japanese have no need to manipulate phonemes as such. Explicit judgements about phonemic sequence are hard for Japanese listeners if the sequences do not mirror mora structure; Kashino, Van Wieringen, and Pols (1992) and Kakehi, Kato, and Kashino (1996) found that truncated CV sequences were perceived equally well by Japanese and Dutch listeners, but truncated VC sequences and the VC portion of VCCV sequences were perceived significantly better by Dutch than by Japanese listeners. Kakehi et al. (1996) argued on the basis of these and other results that phonemic sequence information is not available to Japanese listeners except as a consequence of moraic sequence information; that is, “the basic speech perception unit in Japanese is larger than a phoneme, and [. . .] the integration of phoneme cues is constrained by these [. . .] perception units” (1996, p. 141).

Subsequent studies of vowel epenthesis (intrusion) in the perception of Japanese by Dupoux et al. (1999) and Dehaene-Lambertz, Dupoux, and Gout (2000) produced similar cross-language asymmetries. Dupoux et al. (1999) showed that French listeners could easily distinguish between VCCV and VCVCV strings even when the medial vowel in the latter was very short, while Japanese listeners systematically found such distinctions more difficult. Dupoux et al. argued that the Japanese listeners effectively perceived the VCCV strings as VCVCV, inserting a vowel in accordance with the moraic structure which their prelexical perceptual processing imposed; only the structure with the inserted vowel was accessible, i.e., the submoraic structure was reinterpreted in moraic terms. Dupoux et al. concluded that their results argued against models of spoken-word recognition in which only phonemic or subphonemic processing preceded lexical access and instead supported the existence of higher order structures which they variously described as “an extra layer of processing” or “categories that span larger chunks of signal” (i.e., larger than single phonemes). Dehaene-Lambertz et al. (2000) supported Dupoux et al.’s behavioral data with electrophysiological evidence and similarly argued that the data supported early “coarse coding” of the input signal into “large processing units.”

These claims in effect assign to the mora a more central role in processing than that enjoyed by other rhythmic categories. In Norris et al.’s (1997) framework, rhythmic categories function via their boundaries in the alignment of lexical candidates, but they do not participate as a structural unit in the speech perception process, irrespective of their size or their role in metalinguistic manipulation. The mora, however, either because it is the smallest rhythmic category that has been proposed or because it is the only one to be explicitly coded in orthography and manipulated in language games and other metalinguistic operations, has been accorded in these recent proposals a special status. In the present study we examine the foundation for this special status. In Part 1 we examine evidence from Japanese puns, specifically a formal language game involving mora substitution. We ask whether the substitution is more often than would be expected by chance in effect phoneme substitution, i.e., whether structural features below the level of the mora play a role in determining acceptable puns. In Part 2 we report four experiments in which listeners reconstructed words of which a single mora had been distorted; in these experiments we compared the effects of distorting only one or both phonemes of a CV mora. In Part 3 we describe a lexical decision experiment in which the time to reject a nonword was analyzed as a function of phonemic versus moraic structure. All of the tasks we describe involve access of word forms given input consisting of similar words (Part 1) or nonwords (Parts 2 and 3). The moraic proposal of Kakehi et al. (1996) or Dupoux et al. (1999) predicts that the moraic structure of the input will constrain access more effectively than the submoraic structure, whereas the universal framework proposed by Norris et al. (1997) predicts the reverse: the greater the actual overlap between input and target form, in morae but also where applicable in submoraic structure, the easier the access will be.

PART 1: EVIDENCE FROM GOROAWASE

Naturally occurring language games, secret languages, and word play have attracted considerable attention from linguists, since these forms of language use offer evidence of awareness of linguistic units; though such awareness may not be overtly expressed, it can be inferred from the language users’ ability to manipulate linguistic units independently. Many language games show evidence that language users can manipulate phonemes, even language users without knowledge of an alphabet (Yip, 1982; Mann, 1991).

Language games in Japanese generally call on manipulation of morae (Katada, 1990). But most of the morae of Japanese have the structure CV, and CV is also by far the most common mora structure in natural Japanese speech (Otake, 1990). It is thus possible to examine the patterns of Japanese language play to determine whether there is evidence that language users in effect separately manipulate the submoraic units (consonants and vowels) of CV morae, by choosing mora substitutions which are more similar (overlap with the target mora in either the consonant or vowel) rather than less similar (overlap in neither).

Goroawase is a standardized form of punning in Japanese (Suzuki, 1959). It has a traditional, more cultivated form, in which the game is to make a comment (presumably relevant to an ongoing conversation) which is at the same time a subtle distortion of a quotation from poetry or literature or of a proverb. The speaker thus shows off both linguistic facility and depth of cultural learning. However, it also has a simpler form consisting of transformation of any word or name or saying, not substantially different from ordinary punning (dajare; Otake & Cutler, 2001; Suzuki, 1959). In both cases, the speaker's task is to choose words which are similar to the target words which listeners are supposed to reconstruct; thus goroawase offers an opportunity to examine the metric of perceived similarity between words.

Goroawase has a long history and has appeared in Japanese literature of many periods. It was particularly popular in the Edo period (19th century). That period also saw the flowering of the Japanese art of the woodcut, and one very well-known artist of the time, Utagawa Kuniyoshi, produced two series of woodcuts consisting exclusively of goroawase jokes. Both series are based on the names of the stations (coaching inns) along the main Tokyo-to-Kyoto highway. The first series is made up entirely of pictures of cats performing some action and contains all 55 station names, while the second series, being entirely depictions of facial expressions, contains 30 of the names (see Inagaki, 1985).

For the first woodcut in the first series, for example, the associated place name is *Nihonbashi*. The woodcut depicts a cat playing with two pieces of dried fish.[1] The cat is a constant across the series; the relevant concept which the perceiver is here intended to reconstruct is *nihon*, "two cylindrical objects" *dashi*, "soup ingredients." To construct the substitution, one mora of the target word has been altered: *ba* has been replaced by *da*. However, the target and substituting morae share the vowel and differ only in the consonant. Thus the substitution in this case is very similar to the target name, overlapping with it in all but a single phoneme. In our study of the two Utagawa Kuniyoshi woodcut series, we evaluated the degree of similarity of each target–substitution pair and compared the observed degree of similarity with the degree of similarity which might be expected by chance. If the metric of perceived similarity is based solely on the mora, we would expect to find that observed similarity is approximately as would be expected by chance sampling from the range of phonologically permissible mora substitutions. If, however, the metric of perceived similarity is based on amount of actual overlap, i.e., also on submoraic overlap, we would expect to find that substituting morae are more likely than would be expected by chance to preserve part of the target morae.

Method

We tabulated all the mora substitutions in the two series of Utagawa Kuniyoshi woodcuts. The two cases in which the puns were fully homophonous with the target name were excluded. Many of the puns are more complicated than the one described above, involving substitution of more than one mora. Two cases involved deletion of a mora as well as substitution; although it was arguably possible to determine which mora had been deleted, we decided to ignore these two cases also. In calculating the frequencies we further took into account obligatory phonological phenomena such as assimilation of place in nasal-obstruent sequences; thus although the medial nasal mora in *Nihonbashi* would be pro-

[1] The example described here, along with other examples from the two series of Utagawa Kuniyoshi woodcuts, can be viewed at http://www.mpi.nl/world/persons/private/anne/cat.html.

nounced /m/, whereas in *nihondashi* it would be pronounced /n/, as can be seen above this was not counted as a separate substitution since it follows automatically from the substitution of *da* for *ba*. Likewise, alteration of *Nissaka* to *kuttaka* was counted as two substitutions (of the first mora, *ni*, becoming *ku* and of the third, *sa*, becoming *ta*); the geminate coda was not additionally counted as changing from /s/ to /t/ because the geminate always takes on the value of the onset of a following CV. After the exclusions, 52 puns remained in the first series, involving a total of 97 morae altered, and 29 in the second series, involving 42 mora alterations; thus the total of altered morae was 139.

To make it possible to compare the observed degree of similarity with what would be expected by chance, we calculated the range of substitutions possible for every altered mora in the two sets, assessing for each case the number of possible mora changes which would have preserved a consonant (V changes), preserved a vowel (C changes), or preserved neither (M changes). Japanese has five vowels and 25 possible prevocalic onsets; also all five vowels can stand alone as morae. Not all onset–vowel combinations occur in the language, however, so that there are only 103 possible morae. Two of these are coda morae (the nasal coda and the geminate coda), and one creates a long vowel by duplicating a preceding vowel. In calculating the possibilities we again took into account the obligatory phonological phenomena as well as other occurrence restrictions on morae (e.g., the nasal coda cannot occur word-initially, and the geminate coda cannot occur word-initially or word-finally).

As an example, the calculation for the alteration of *ba* in *Nihonbashi* was as follows. The number of possible morae is 103, or *ba* plus 102 others. All other vowels than /a/ could also occur after /b/, so that 4 changes would be possible in which the consonant was preserved but the vowel was changed. Also, any of the other onsets could occur with /a/, or the /b/ could be removed to let /a/ stand alone, giving 25 possible changes in which the vowel was preserved but the consonant was altered. The two coda morae and the long vowel mora could, however, not possibly occur after *Nihon*-, so that three substitutions were illegitimate. The number of legitimate mora substitutions which did not preserve either the C or the V was then the total number minus those that preserved some part and those that were illegitimate, i.e., $102 - (4 + 25 + 3) = 70$. Some other calculations involved much smaller sets of possibilities. Thus substitutions involving the mora *yu* (/ju/) allowed only two possibilities in which the consonant was preserved, since only three morae have /j/ onset; likewise, substitutions involving any mora with the vowels /i/ or /e/ allowed maximally 11 or 12 possibilities in which the vowel was preserved, these vowels occurring in only 12 or 13 morae respectively. The appropriate numbers were calculated for every substitution individually.

Note that this calculation of course does not take into account what possibilities are or are not allowed by the vocabulary of the language. However, a vocabulary-based calculation is not feasible for several reasons. First, the goroawase puns involve substitution of almost any number of morae in the original word and, as we saw, deletion of morae. Under these circumstances it becomes impossible to know where to draw the line between a replacement word which might be considered and one which would not be considered without making assumptions which prejudice the very question at issue. And, second, the vocabulary in question is that of the Edo period, which is not necessarily approximated by 21st-century counts. Under these circumstances a calculation of phonologically permissible changes seemed the only option.

Results and Discussion

Of the 139 substitutions, both vowel and consonant were altered in 44 cases, only the vowel was altered in 22 cases, and only the consonant was altered in 73 cases. That is, there are more than twice as many cases preserving either the vowel or consonant of the original mora as cases in which both vowel and consonant are replaced.

The calculation of expected similarity by chance showed that across the 139 altered morae, the mean number of possible substitutions involving only a consonant change was

19.6, the mean number involving only a vowel change was 3.4, and the mean number involving both a vowel and consonant change was 69.3. (This implies that on average 9.7 morae were impermissible substitutions.) Converted to percentages and expressed as those percentages of 139, these counts suggest that random selection among the phonologically legitimate options would have yielded 34.62 cases involving change only in either consonant or vowel (29.51 consonant change, 5.11 vowel change) and 104.38 cases involving substitution of both consonant and vowel. A comparison of the actual figures for partial preservation versus complete substitution (95 to 44) with the above expected figures (34.62 versus 104.38) reveals a highly significant difference [$\chi^2(1) = 140.24$, $p < .001$]. Thus, the puns are significantly more likely to preserve part of the substituted mora than would be expected by chance. A further comparison of the actual figures for substitution of consonant versus vowel (73 to 22) with expected figures (80.95 versus 14.05), derived by applying the ratio of 29.51:5.11 to the partial preservation total of 95, also reveals a significant difference [$\chi^2(1) = 5.27$, $p < .05$], suggesting that the C:V substitution ratio is somewhat less unbalanced than would be expected by chance. Thus mora substitution in a pun involves preservation of some part of the original mora, and though many more phonologically possible options involve vowel preservation (i.e., consonant substitution) than the reverse, either type of substitution is acceptable.

In summary, the evidence from goroawase (and, indeed, from modern dajare: Otake & Cutler, 2001) suggests that mora substitution is more often than would be expected by chance in effect phoneme substitution because two words which overlap in all but a single consonant or vowel form a better target–pun pair than two words which overlap in all but a CV mora. Morae are undeniably accessible to Japanese language users' awareness—perhaps especially in the Edo period, during which mora-based poetic forms such as haiku flourished. Phonemes, in contrast, are much less accessible. Yet these goroawase jokes suggest a metric of similarity between target and pun which would be better measured in terms of actual overlap, including submoraic structure, than in terms of morae. Since puns are not successful if they render the target inaccessible, this in turn suggests that, as proposed in the spoken-word recognition framework of Norris et al. (1997), half a mora preserved is better than complete substitution because the greater the actual overlap between input and target form, the easier lexical access will be.

PART 2: EVIDENCE FROM WORD RECONSTRUCTION

In this section we report a series of experiments in which we distorted individual CV morae in Japanese words, and asked whether distortions which left part of the mora intact (replacing only the C or only the V) were more or less harmful to word recognition than distortions which left nothing of the mora intact (replacing both the C and the V).

For these experiments we adapted a relatively new experimental paradigm in which listeners are presented with spoken input differing in some small respect from a real word, and their task is to find the real word which has been distorted. The task was invented independently in somewhat different forms by Van Ooijen (1996) and Boelte (1997). Van Ooijen named the task Word Reconstruction, and in her version, English listeners heard nonwords such as *eltimate* or *maleen*, each of which could be reconstructed into two possible existing words, by substitution of a single vowel (*ultimate* and *malign*) or consonant (*estimate* and *marine*) respectively. Van Ooijen was interested in comparing the contributions of vowel and consonant information in spoken-word processing (and her results showed that substitution of vowels was easier than substitution of consonants, a result since replicated with the same task in Spanish and Dutch: Cutler, Sebastian, Soler, & Van Ooijen, 2000). Boelte called the task Word Correction; in his version, listeners heard a list of words and nonwords and responded to the latter by supplying the nearest word. Boelte was interested in phonetic distance between nonword and base word (e.g., *bishop* distorted as *bishot* or *bishos*; the latter was harder to correct); note that his form of the task, and the questions he addressed, also resemble the literature on Mispronunciation Monitoring (e.g., Cole, 1973), in which listeners hear sentences and respond whenever they hear a word which is mispronounced.

Although Word Reconstruction (to use Van Ooijen's term) requires an explicit decision, it cannot be one in which listeners consciously check all possible options; for that the response times are too short. In both forms of the task, listeners have no idea where in the word a substitution should be made, yet the phenomenal experience is that the word "springs to mind" without the necessity of explicit canvassing of options. Mean response times (to press a button once the word is found; after responding in this way the subject is then required to produce the reconstructed word) average just over 2 s from stimulus onset in the harder form of the task, in which there are two options for response (Van Ooijen, 1996; Cutler et al., 2000), and just over 1 s, again from stimulus onset, in the one-option form (Boelte, 1997). In each form of the task responses below a second are not unusual. It is reasonable to assume therefore that the input has at least partially activated the real word(s) which it almost matches, in the same fashion as has been observed in the many experimental demonstrations of lexical activation by incomplete input. This account explains the fact that Van Ooijen's form of the task, with two possible response options, proves harder: When two words are activated at once, they compete with one another, with the result that each is less activated than would have been the case without the competitor. That concurrent multiple activation produces interword competition and ensuing reduction of activation for the competing words has also been repeatedly demonstrated in the laboratory (Goldinger, Luce, & Pisoni, 1989; Goldinger et al., 1992; McQueen et al., 1994).

In our version of the task each stimulus was a nonword (there were no real-word distractors as in Boelte's experiments) with putatively one real-word neighbor (not two as in Van Ooijen's studies). Furthermore, we rendered the task easier by making listeners aware of where in the word the substitution should be made. The subjects were not told the substitution position outright; instead, we induced the information via a set of practice examples. If, in an analogous English experiment, listeners were given examples like *telephoce*, *rewarb*, *familoo*, and *stampeen*, then the correct induction would be that the substitution position should be the final phoneme, producing *telephone*, *reward*, *family*, and *stampede*. In our four experiments with this task the substitution position varied (initial mora in Experiment 1, final mora in Experiment 2, and a specified medial mora in Experiment 3 and 4), and thus the practice examples varied across experiments. However, the correct induction from the examples was always that the relevant substitution unit was the mora; in this way, we hoped to construct the optimum opportunity for mora-based effects on early word processing should these exist.

Suppose that, as argued by Dupoux and colleagues (e.g., 1999), the mora is the earliest available unit for recognizing Japanese speech. Then, we should expect response times (and miss rates if these are still considerable in this easier form of the task) to be equivalent for substitution of a whole mora (CV, e.g., *ka* in *kamera* replaced by *ni*, giving *nimera*) or a partial mora (C or V, e.g., *ka* in *kamera* replaced by *na* or *ki*, to give *namera* or *kimera*); that is, *kamera* should be equally easy to reconstruct from *namera*, *kimera*, or *nimera*. Conversely, if the mora has no role at all at this early stage, we should expect responses to be always faster (and more accurate) for substitution of a partial mora (C or V) than of a whole mora (CV), making *kamera* harder to find in *nimera* than in either *namera* or *kimera*. An intermediate pattern could result if the position of the distortion in the word exercised strong effects on activation. For instance, it could be the case that intact initial portions of a word suffice for activation, so that distortions later in the word (*kamera* replaced by *kamena*, *kamere*, or *kamene*) have more graded effects, while in contrast any initial distortion severely disrupts activation (so that *namera*, *kimera*, and *nimera* have equally drastic effects). Cross-experiment comparisons allow us to test for differences in the type of effects at different substitution positions within the word.

Experiment 1

Materials

Sixty Japanese words, judged to be of quite high familiarity to undergraduate listeners, were chosen; the most important selection criterion for the words was that the planned alteration would yield a unique reconstruction solution, i.e., that no other word of the same length shared the unaltered remainder of the word. Needless to say, the relatively simple phonological structure of Japanese rendered this criterion hard to meet. For some words there was therefore indeed an alternative solution, but one of much lower familiarity than the target word (e.g., *bokashi*, "color gradation in painting," instead of *mukashi*, "long ago"). A number of the words met the uniqueness criterion by virtue of being loan words or contracted expressions (the latter being a common word formation process in Japanese); such words were, however, always in extremely common use (an example loan word is *kamera*, "camera," and an example contraction is *pasokon*, "personal computer").

Thirty of the words were three morae in length, and the remaining 30 had four morae. The experimental stimuli consisted of three versions of each word, in each of which the initial CV mora had been replaced by another CV mora. The three versions were constructed such that one replacement mora shared the consonant with the original mora but not the vowel; another shared the vowel but not the consonant; the third shared neither the vowel nor the consonant. The replacing consonant and vowel in the first two versions were always also the consonant and vowel of the whole-mora replacement. Thus the three versions of *pasokon* were *pesokon* (*pa* replaced by *pe*), *tasokon* (*pa* replaced by *ta*), and *tesokon* (*pa* replaced by *te*). We refer to these conditions as V, C, and M replacements respectively. The complete materials (for this and all experiments) can be found in the Appendix.

Three presentation lists were constructed, each containing one version of each of the 60 items. Replacement condition (vowel, consonant, or whole mora replaced) was counterbalanced across presentation list. The lists were recorded by a male native speaker of Standard Tokyo Japanese onto Digital Audio Tape, sampling at 48,000 Hz. The pitch accent pattern of the base word was preserved in each transformed version. The rate of presentation was adjusted to one word every 10 sec, and timing marks coincident with the onset of each item were added to the second channel of the tape (where they would be inaudible to the listeners). The duration of each item was measured using Sound Designer II software in order to allow recorded response times to be adjusted to given times from item offset.

Familiarity ratings for the base words were ascertained from Amano and Kondo (1999). On a scale from 1 (*unfamiliar*) to 7 (*extremely familiar*) the experimental base words received ratings in the mid to upper range of the scale; the overall mean rating was 6.23.

Participants

Forty-five undergraduates of Dokkyo University took part in the study in return for a small payment. All were native speakers of Japanese with no reported hearing impairment. Fifteen received each of the three presentation lists.

Procedure

The participants were tested one or two at a time in separate sound-attenuating carrels in a quiet room. They heard the items binaurally over Audio-Technica ATH-A9 headphones. The experiment began with a practice session. The listeners were instructed to listen to each item and to change a part of it to produce a known word. The instructions did not specify which part should be changed; the 12 practice items (a random ordering of four cases each of the three types of replacement used in the experiment) were used to draw listeners' attention to the necessary change. Listeners were given a chance to hear the practice set a second or third time if desired. This indirect method of instruction was successful; all listeners produced correct responses, involving substitution of the initial mora, by the end of the practice session.

Listeners were instructed to press a response key as soon as they had identified a real word and then to speak this word into a microphone in front of them. This method of response collection was used by Van Ooijen (1996) and Boelte (1997) because it avoids the necessity of additional controls on materials construction; if spoken responses were used to trigger a voice key, it would be necessary to control word-initial phonetic structure within subsets of the materials, which would render the already very difficult task of finding materials for the word reconstruction paradigm effectively insuperable. The spoken responses were nevertheless recorded and checked for conformity with the intended response. The collection and storage of the response times (RTs) was controlled by a personal computer running NESU experimental control software.

Results and Discussion

The listeners' spoken responses were analyzed, and incorrect responses, failures to produce a spoken response, and responses after timeout were treated, as were missing keypress responses, as missing data. An unintended alternative response was produced to certain items by a few participants, and these items were discarded; this resulted in the loss of nine items (including *mukashi*, for which six participants produced *bokashi*, five of them in the M-replaced condition in which the input was *pokashi*).

The overall mean RT (from item offset) was 1125 ms and the overall mean rate of missed responses 12.85%. The mean RTs and miss rates for each condition are displayed in Fig. 1. As can be seen, the pattern of results was the same for both measures: responses were easier (RTs were fastest and error rates lowest) when only V or C was replaced than when the whole mora (C plus V) was replaced.

Analyses of variance were carried out, separately across participants and across items, for both RTs and miss rates. The main effect of replacement condition was significant, beyond the .001 level, across participants and across items, for both RTs [$F1(2,84) = 61.89$, $F2(2,100) = 22.6$] and miss rates [$F1(2,84) = 54.02$, $F2(2,100) = 22.61$]. The significance of the differences between conditions was tested via *t*-tests which compared the intermediate condition (C replaced) with the other two. Response

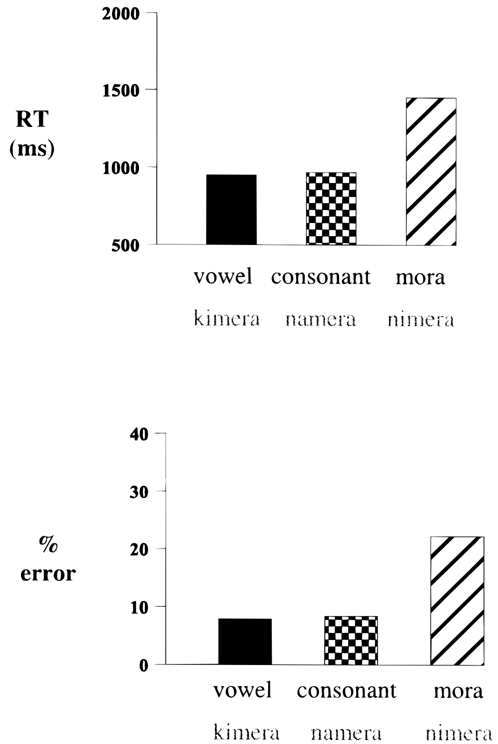

FIG. 1. Experiment 1: Mean response time (in milliseconds) and percentage error for word reconstruction as a function of whether replacement of the first mora of the target word involved the vowel, the consonant, or the whole mora.

times were faster [$t1(44) = 10.53$, $p < .001$; $t2(50) = 4.95$, $p < .001$] and miss rates were lower [$t1(44) = 8.13$, $p < .001$; $t2(50) = 5.44$, $p < .001$] in the C-replaced condition than in the M-replaced condition. However, neither RTs nor miss rates differed significantly in the C-replaced condition and the V-replaced condition.

Four-mora words were easier (responded to faster and more accurately) than three-mora words [$F1(1,42) = 22.64$, $p < .001$, $F2(1,49) = 6.05$, $p < .02$ for RTs; $F1(1,42) = 41.39$, $p < .001$, $F2(1,49) = 5.42$, $p < .025$ for miss rates]; this was expected given that the task is harder in three-mora words (in which the replacing mora constitutes 33% of the word) than in four-mora words (in which the replacing mora constitutes only 25% of the word). The three versus four-mora comparison also interacted significantly with the main effect of replacement condition in the RTs [$F1(2,84) = 9.38$, $p < .001$, $F2(2,98) =$

6.4, $p < .005$; this interaction did not reach significance in the miss rates]. The source of the interaction was a reversal of the difference between the V-replaced and C-replaced conditions for three-mora words (for which C-replacement was somewhat easier) versus four-mora words (C-replacement harder); however, *t* tests showed that there was no significant difference between these two conditions for either word length, while for each word length both V-replaced and C-replaced conditions were significantly faster than M-replaced.

This experiment has clearly demonstrated that partial information about a mora (either the C or the V) can be informative to listeners; word reconstruction was significantly easier when the initial mora had been replaced by another mora sharing with it either C or V. However, there was no significant difference in advantage for retention of either the C or V.

In our next experiment, in contrast, the replaced mora was always the word-final mora. In this case the Norris et al. (1997) proposal leads to some additional predictions. Recall that our materials were constructed such that in most cases there should be only one possible response and in all cases only one possible high-frequency response. Although listeners cannot know until the item is finished how long it will be, so that (for instance) a three-mora item can receive competition from four-mora and longer words, at item offset there should be only one active candidate of the appropriate length available. When replacement has occurred on the final mora, the beginning of the word remains intact and retrieval should be easier than when (as in Experiment 1) it is the final portion that has remained intact. This implies that RTs should be faster overall in Experiment 2 than in Experiment 1. But in addition, more of the intended candidate is available intact in the V-replaced than in the C-replaced condition. Norris et al.'s framework, in which it is assumed that this incoming information can be utilized continuously, thus predicts that V-replacement will be the condition from which retrieval of the intended word is easiest. Therefore in Experiment 2 we expect to observe that the M-replaced condition is again hardest, because it provides no partial information about the replacement mora, but we expect also that the C-replaced condition should prove harder than the V-replaced condition.

Experiment 2

Materials

A further 60 items were constructed, in the same manner as for Experiment 1, except that all substitutions were made on the final morae. An example target word is *tempura*, of which the three versions were *tempuri* (*ra* replaced by *ri*), *tempuka* (*ra* replaced by *ka*), and *tempuki* (*ra* replaced by *ki*). Familiarity scores were again ascertained; the overall mean score was 6.17. The materials were recorded and measured as for Experiment 1.

Participants and Procedure

Forty-five undergraduates of Dokkyo University took part in the study in return for a small payment. All were native speakers of Japanese with no reported hearing impairment, and none had taken part in Experiment 1. Again 15 received each of the three presentation orders. The procedure was as in Experiment 1, except that the practice induced attention to the final rather than to the initial mora of the stimuli.

Results and Discussion

The responses were analyzed in the same way as for Experiment 1. Unintended alternative responses caused us to discard six items. The overall mean RT was 740 ms and the overall mean rate of missed responses 15.99%; condition means are shown in Fig. 2.

The main effect of condition was again always significant beyond the .001 level for both RTs [$F1(2,84) = 21.66$, $F2(2,106) = 15.77$] and miss rates [$F1(2,84) = 67.73$, $F2(2,106) = 15.32$]. RTs were again faster [$t1(44) = 3.22$, $p < .002$; $t2(53) = 2.39$, $p < .02$] and miss rates were lower [$t1(44) = 5.08$, $p < .001$; $t2(53) = 3.42$, $p < .001$] in the C-replaced condition than in the M-replaced condition. Here, however, RTs were significantly slower [$t1(44) = 3.76$, $p < .001$; $t2(53) = 3.11$, $p < .003$] and miss rates significantly higher

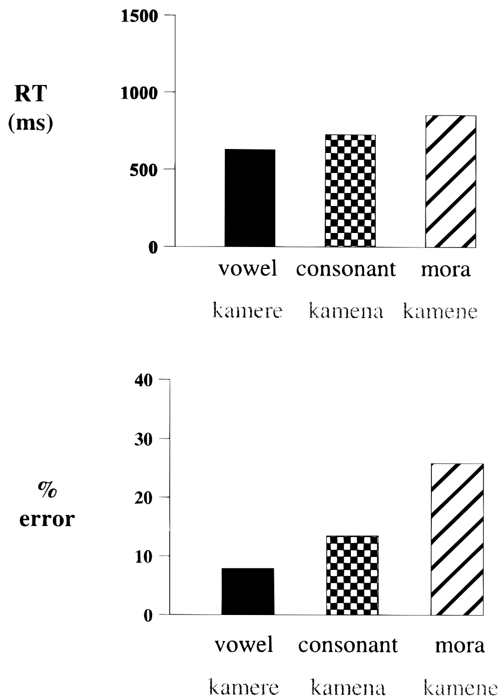

FIG. 2. Experiment 2: Mean response time (in milliseconds) and percentage error for word reconstruction as a function of whether replacement of the final mora of the target word involved the vowel, the consonant, or the whole mora.

[$t1(44) = 3.57$, $p < .001$; $t2(53) = 2.11$, $p < .05$] in the C-replaced condition than in the V-replaced condition. Responses to four-mora words were again faster and more accurate than responses to three-mora words [$F1(1,42) = 36.23$, $p < .001$, $F2(1,52) = 6.69$, $p < .02$ for RTs; $F1(1,42) = 102.62$, $p < .001$, $F2(1,52) = 8.39$, $p < .01$ for miss rates] but there was no interaction of this factor with the comparison of replacement conditions; in both word lengths the three conditions showed the pattern of Fig. 2.

Thus this experiment has produced the pattern of results predicted by the proposal of Norris et al. (1997): a significant advantage of the V-replacement condition over the C-replacement condition as well as the significant advantage of both of these over the M-replacement condition. The overall mean RT in this experiment was also noticeably faster, reflecting the fact that most of the information relevant to identifying the word was available before the distortion was presented; the V versus C-replacement difference further suggests that this information can be exploited continuously rather than only on a mora-by-mora basis.

In our third experiment we substituted a word-medial mora. We expect that this kind of substitution will prove harder for listeners, since the distortion will result in two intact but separated fragments rather than an intact two or three-mora sequence constituting the end (Experiment 1) or the beginning (Experiment 2) of a word. Nevertheless, we expect that listeners will be able to perform the task, and the medial-mora manipulation further enables us to compare the contribution of consonant preservation (in V-replacement) and vowel preservation (in C-replacement) in a difficult form of word reconstruction.

Experiment 3

Materials

Again 60 items were constructed in the same manner as for the preceding experiments, except that all substitutions were made on the second mora. An example item is *kodomo* (“child”), of which the three versions were *kodemo* (*do* replaced by *de*), *konomo* (*do* replaced by *no*), and *konemo* (*do* replaced by *ne*). The materials were recorded and measured as for the preceding experiments; the overall mean familiarity score was here 6.08.

Participants and Procedure

Forty-five undergraduates of Dokkyo University took part in the study in return for a small payment. All were native speakers of Japanese with no reported hearing impairment, and none had taken part in Experiments 1 or 2. Again 15 received each of the three presentation orders. The procedure was as again in Experiment 1, except that the practice session induced attention to the second mora of each stimulus item.

Results and Discussion

The responses were analyzed in the same way as for Experiments 1 and 2. Unintended alternative responses caused us to discard three items. The overall mean RT was 1569 ms and the overall mean rate of missed responses 27.12%. Thus the task was, as predicted, harder with this type of substitution. As Fig. 3 shows, however, the patterning of the three conditions strongly resembles that observed in Experiment 2.

The main effect of replacement condition was once more significant, at beyond the .001 level, for both RTs [$F1(2,84) = 63.68$; $F2(2,112) = 53.47$] and miss rates [$F1(2,84) = 160.61$; $F2(2,112) = 39.89$]. RTs were faster [$t1(44) = 5.9$, $p < .001$; $t2(56) = 6.31$, $p < .001$] and miss rates were lower [$t1(44) = 8.11$, $p < .001$; $t2(56) = 6.48$, $p < .001$] in the C-replaced condition than in the M-replaced condition. As in Experiment 2, RTs were significantly slower [$t1(44) = 3.32$, $p < .001$; $t2(56) = 4.0$, $p < .001$] in the C-replaced condition than in the V-replaced condition. Miss rates were significantly higher in the C-replaced condition than in the V-replaced condition in the analysis across participants [$t1(44) = 3.64$, $p < .001$], but the difference was only marginally significant in the analysis across items [$t2(56) = 1.89$, $p < .065$]. Four-mora words were easier to reconstruct than three-mora words [$F1(1,42) = 89.15$, $p < .001$, $F2(1,55) = 16.35$, $p < .001$, for RTs; $F1(1,42) = 193.91$, $p < .001$, $F2(1,55) = 11.25$, $p < .02$, for miss rates], but there was again no interaction of this factor with replacement condition. In all analyses both three and four-mora words produced the same ordering of conditions as illustrated in Fig. 3.

A further analysis was prompted by the discovery that the RTs in this experiment were on average much longer than those in Experiments 1 and 2, but that the RTs in fact spanned quite a large range. This raised the possibility of, for instance, different response strategies applying at different processing stages, giving a different pattern of results in fast and slow responses. To test this, we split the RT distribution into the faster 50% of RTs (mean 706 ms) and the slower 50% (mean 3063 ms) and analyzed each set separately. The pattern of results for each set was identical to that of the overall analysis, leading us to conclude that this pattern was not affected by different response strategies.

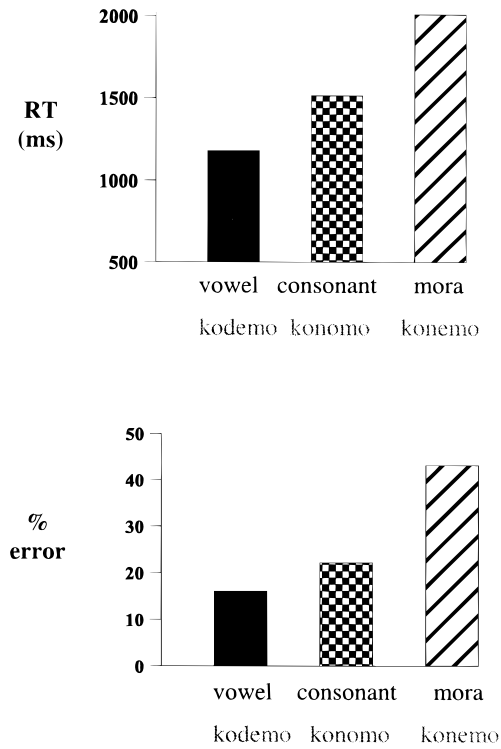

FIG. 3. Experiment 3: Mean response time (in milliseconds) and percentage error for word reconstruction as a function of whether replacement of the second mora of the target word involved the vowel, the consonant, or the whole mora.

The results of Experiment 3 further confirm the advantage of partial information within the mora for word identification. However, the fact that consonant preservation (i.e., the V-replacement condition) led to easier word reconstruction than vowel preservation (in the C-replacement condition) here cannot be explained with reference to the size of the intact portion of the target word (the explanation offered above for the difference in Experiment 2). In this experiment, a larger intact portion to the left of the replacement in V-replacement is compensated by a larger intact portion to the right of the replacement in C-replacement; for example, *kodemo* preserves three phonemes of *kodomo* from onset and two from offset, while *konomo* preserves two phonemes from onset and three from offset. However, we also argued with respect to Experiments 1 and 2 that information from onset (as in Experiment 2) constrains word retrieval more effectively than information from offset (Experiment 1); it could therefore be that this effect is also operative in fragments of one and the same word.

We explored this issue further in a final experiment in which we made the listeners' task harder still; substitutions were word-medial, but some were in the second and some in the third mora (of four-mora words). If information about word beginnings is more constraining than information about word endings, and the greater contribution of maintaining the C of a CV mora arises simply because the C precedes the V, then we should observe a graded effect across the four replacement conditions in this experiment, from C-replaced in the second mora (hardest) to V-replaced in the third mora (easiest). Only if V-replacement in general makes word reconstruction easier than C-replacement will we here observe a difference between the V and C conditions (and then in both positions).

Experiment 4

Materials

The 30 four-mora items of Experiment 3 were supplemented with a further 30 four-mora items in which substitutions were made on the third mora. An example of these additional items is *panorama*, for which the three versions were *panorema* (*ra* replaced by *re*), *panozama* (*ra* replaced by *za*), and *panozema* (*ra* replaced by *ze*). The materials were again recorded and measured as for the preceding experiments, and familiarity scores for the 30 new items were ascertained; the overall mean score for the set of 60 items was 6.03.

Participants and Procedure

Forty-eight further undergraduates from the same population took part, in return for a small payment. All were native speakers of Japanese with no reported hearing impairment, and none had taken part in Experiments 1 to 3. Sixteen received each of the three presentation lists (with eight of these receiving each of the two possible orders of the second-mora and third-mora blocks). The procedure was as in Experiment 3, except that the experiment contained two blocks (change in second and change in third mora); a new practice session, with examples redirecting attention to the new substitution location, preceded the second block.

Results and Discussion

The responses were analyzed in the same way as for Experiments 1 to 3. Unintended alternative responses caused us to discard four items. The overall mean RT was 1249 ms and the overall mean rate of missed responses 23.52%. As Fig. 4 shows, the patterning of the three conditions again resembles that observed in Experiments 2 and 3.

The analysis of this experiment contained two additional comparisons, of mora position (i.e., changes in the second versus the third mora) and order (i.e., comparing the groups which had had the two mora position conditions in one versus

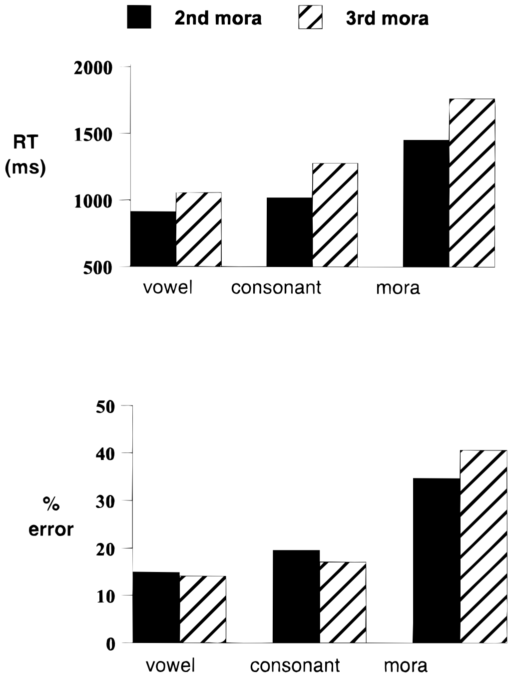

FIG. 4. Experiment 4: Mean response time (in milliseconds) and percentage error for word reconstruction as a function of whether replacement of a medial mora of the target word involved the vowel, the consonant or the whole mora, and as a function of whether second or third mora was replaced.

the other order). Although RTs were faster for changes in second versus third mora position, the difference was not statistically reliable; there was no such effect in the miss rates, and on neither measure did the mora position comparison interact with the effect of replacement condition or the order effect. In particular, there was no indication of a graded effect across the word: RT was slower for third- than for second-mora replacement, and though it decreased from C to V in the second mora, it then increased again for the C of the third mora.

The order effect never interacted with the effect of replacement condition, and it was not itself significant in the RTs. However, it affected the miss rates, which were higher for the listeners who changed second morae then moved to third-mora changes (26.1%) than for those who had the two conditions in the reverse order [21%; $F1(1, 42) = 5.15$, $p < .03$; $F2(1, 54) = 16.74$, $p < .001$]. The difference between these groups of participants was apparent in the blocks they heard first (27% versus 22.5% missed items) as well as in the blocks they heard second (25% versus 19.5%) and thus appears to be a difference between these two groups in facility with the task.

The main effect of replacement condition was as always significant, beyond the .001 level, for RTs [$F1(2,84) = 72.37$, $F2(2,108) = 29.77$] and for miss rates [$F1(2,84) = 94.11$, $F2(2,108) = 32.77$]. RTs were again faster [$t1(47) = 7.27$, $p < .001$; $t2(55) = 5.11$, $p < .001$] and miss rates were lower [$t1(47) = 8.71$, $p < .001$; $t2(55) = 6.96$, $p < .001$] in the C-replaced condition than in the M-replaced condition. As in Experiment 3, RTs were significantly slower [$t1(47) = 3.93$, $p < .001$; $t2(55) = 2.28$, $p < .03$] in the C-replaced condition than in the V-replaced condition, but miss rates were only significantly higher in the C-replaced condition than in the V-replaced condition in the analysis across participants [$t1(47) = 2.3$, $p < .03$]; in the analysis across items the difference was not significant [$t2(53) = 1.51$].

Experiment 4 has thus confirmed the pattern of results of the preceding experiment. Although both the V-replacement and C-replacement conditions proved easier than M-replacement, there was also a difference between the first two: Replacement of a vowel proved easier than replacement of a consonant. This effect could not be explained in terms of the amount of available information from the onset of the word because C-replacement in the third mora (four phonemes from onset preserved) proved harder than V-replacement in the second mora (three phonemes from onset preserved); it could thus reflect a vowel/consonant difference of the kind already reported for English, Dutch, and Spanish (Cutler et al., 2000; Van Ooijen, 1996). The principal result of our series of four word reconstruction experiments, however, remains the consistent contribution of submoraic segmental information to performance in the word reconstruction task: replacing a mora was always easier when part of the mora was preserved.

PART 3: EVIDENCE FROM LEXICAL DECISION

To strengthen further the conclusion that information at a level below the mora plays a role in lexical processing, we decided to conduct a lexical decision experiment using the materials from the word reconstruction experiments. The logic behind this final experiment was based on a study by Marslen-Wilson (1984) addressing the notion of word uniqueness point (UP; the point at which phonemic information distinguishes a word from all other words beginning in the same way). Marslen-Wilson found that "NO" responses to spoken nonwords in lexical decision were roughly constant from the point at which a nonword deviated from all existing words; looked at another way, responses measured from nonword offset were faster, the more phonemes there were between the nonword uniqueness point (NUP) and the measuring point.

We therefore computed the NUP for each nonword in Experiments 1–4, namely the point (phoneme or mora) at which it deviated from the existing Japanese word which was its nearest onset-overlapping competitor. For instance, consider the three nonwords formed from *tsumiki*: *tsuniki*, *tsumeki*, and *tsuneki*. The first becomes a nonword at *tsuni-*, at which point it deviates from *tsunami* and so on; the second becomes a nonword at the end because there are several words beginning *tsumek*-; the third becomes a nonword at *tsunek*-, where it deviates from *tsuneni* ("always") and other words. We then calculated the number of phonemes and the number of morae between the NUP and the end of the nonword. When the NUP is considered as a phoneme position, these three nonwords have respectively two, zero, and one phonemes post-NUP. When the NUP is considered as a mora position, the first nonword has one mora post-NUP and the other two nonwords have none. Note, incidentally, that since a considerable number of the base words, and hence the nonwords, involved single-phoneme morae (vowels or coda consonants, as in *ka-se-t-to*, *su-ka-a-to*, *ra-ji-o*, *ta-ku-shi-i*, and *za-bu-to-n*), there is no simple mapping between the post-NUP count in morae and the post-NUP count in phonemes.

This calculation revealed that, across all experiments, the NUP in the V-replaced nonwords was very slightly later than in the C- and M-replaced nonwords (with most of this difference of course due to Experiment 2 in which the final mora was replaced). Although the differences across replacement conditions within a single item were small, however, the differences across items were substantially larger, with post-NUP count in phonemes roughly split four ways between zero, one, two, and three or more. This suggested that there would be sufficient range in the experimental materials for a comparison of the effect of NUP in lexical decision; in Experiment 5 we put such a comparison to the test.

Experiment 5

Materials

One hundred twenty sets of 3 nonword versions were chosen from the materials of Experiment 1–4; 30 sets were taken from each experiment. From Experiment 4, all 30 items in which the third mora had been replaced were chosen, and from Experiments 1–3 half of each materials set, in each case 15 three-mora and 15 four-mora items. No base word was used in more than one chosen item. None of the items dropped from the earlier analyses were used except for two items in the Experiment 4 set.

One hundred thirty-three real words and 15 further nonwords were chosen, all of which, like the experimental nonwords, were three or four morae in length. Ten each of these words and nonwords were used as a practice set, and 5 each were warmup items at the beginning of the experimental lists. The remaining 118 real words were interspersed among the 120 experimental nonwords to form three lists. Each list contained all the words and filler nonwords and one version of each of the experimental nonwords. Original replacement condition (vowel, consonant, or whole mora replaced) was counterbalanced across presentation list. The lists were newly recorded by the same speaker as for Experiments 1–4 (the previous recordings were not used since it was necessary to match audio quality across words and nonwords). Timing signals were aligned with the onset of each item. For each item we measured the total duration and the duration from NUP to the end of the item.

Participants

Fifty-one undergraduate members of Dokkyo University took part in the experiment in return for a small payment. All were native speakers of Japanese with no reported hearing impairment, and none of them had participated in Experiments 1 to 4. Seventeen received each of the three presentation lists.

Procedure

The testing set-up was as for Experiments 1 to 4. Participants were, however, in this case given two response keys, labeled YES and NO, and were instructed to decide for each spoken item whether it was a Japanese word and to signal their decision by pressing the appropriate response key as quickly as possible.

Results and Discussion

The response times for each item were adjusted for item duration so as to give RTs from item offset. The mean NO RT for the experimental nonwords was 503 ms. The error rate was less than 1%.

A correlation analysis across items examined the relationship between duration from NUP to item end and mean correct NO RT, using mean item log(RT) instead of mean item RT to correct for nonnormality in RT distributions. The two were significantly correlated both across all items [$r(359) = -.36$, $p < .001$] and for each condition separately [V-replaced: $r(119) = -.22$, $p < .02$; C-replaced: $r(119) = -.49$, $p < .001$; M-replaced: $r(119) = -.33$, $p < .001$]. In other words, the earlier the item became a nonword, the faster listeners could reject it.

A linear-regression analysis was then carried out comparing the factors Duration (from NUP to item end), Phonemes (number of phonemes not shared with nearest onset-overlapping competitor), and Morae (number of morae not shared with competitor) as determinants of log(RT). A mild degree of collinearity was established (condition index 12.10); that is, the factors under test do co-vary, but not to the extent that all separate contributions are excluded. The regression analysis in fact showed significant effects of Duration [$F(1, 356) = 53.68$, $p < .001$] and Phonemes [$F(1, 356) = 4.83$, $p < .03$], but the effect of Morae was insignificant ($F < 1$). The factor Duration accounted for 11.17% of the variance and Phonemes for an additional 2.5%.

This experiment thus clearly supports our interpretation of listeners' processing in the word reconstruction experiments: Word retrieval is effected by continuous exploitation of the incoming acoustic information. Because processing is effectively continuous, phonemic information, i.e., information below the level of the mora, contributes to word recognition, and its contribution is not dependent upon prior availablity of mora units.

GENERAL DISCUSSION

Our findings do not support the suggestion that the mora functions as an early structural unit in Japanese listeners' recognition of spoken words. Partial information about a mora constitutes useful information for listeners: it consistently and significantly speeds their reconstruction of slightly distorted words, whether these are puns in a natural setting or stimulus items in an experimental setting. Similarly, the point at which a nonword can definitively be rejected is predicted better by the phoneme boundary at which it diverges from the nearest word than by the mora boundary at which divergence occurs. Thus although morae are more accessible to Japanese listeners' awareness than phonemes are, morae do not appear to play a direct role in Japanese spoken-word recognition.

In Japanese puns, such as the traditional goroawase examples which we examined in part 1 of our study, substitutions involve single phonemes more often than could be predicted by chance sampling from the phonological possibilities. Word reconstruction, as described in part 2, is not driven by mora substitution even under conditions, such as we constructed here, in which listeners' attention is focused at the mora level and mora-based substitution strategies would always work. As the results clearly indicate, the words to be reconstructed were even under these conditions more easily accessible when more of their phonetic material was available; the C- and V-replacement conditions provided an additional phoneme of information over and above that provided in the M-replacement condition, and listeners responded in all four reconstruction experiments more rapidly and more accurately when this extra information was on hand. In lexical decision, as we demonstrated in part 3, the pattern of results equally clearly showed that listeners can reject a nonword when the next incoming phoneme is incompatible with continuation as a word; the NO response does not have to await completion of the mora.

Across the reconstruction experiments we compared substitution of initial, medial, and final morae. This allowed the opportunity for several potential intermediate patterns of results to appear, whereby morae might have been observed to function more as integral units in some positions than others. Several such hypotheses might be proposed, but all can be scotched given the robust and unequivocal pattern in our data: at any position, replacement of an entire CV mora is always harder than replacement when partial submoraic information is available.

Recall that our listeners in Experiments 1–4 were never made aware that in some trials they were effectively replacing single phonemes. As we discussed in the introduction, Japanese is not written with an alphabetic orthography, and explicit manipulation of phonemes is not an easy task for Japanese listeners. In monitoring tasks, detection of morae is consistently easier than detection of nonmoraic phonemes (Cutler & Otake, 1994; Otake et al., 1996). The instructions of Experiments 1–4 focused listeners' attention at the level of the moraic unit, which is one which Japanese listeners are used to manipulating, inter alia in popular language games (Katada, 1990). As far as listeners knew, they were substituting morae in all trials. Nonetheless, their performance was significantly better when part of the mora they had to find was preserved in the input stimulus.

Further, the difference between the M-replaced and the C- and V-replaced word reconstruction conditions cannot be due to competition between a greater number of potentially available word options in the M-replaced case; our stimuli were chosen such that in all three conditions there was exactly one solution (and the few items for which this turned out not to be true were discarded from the results). Instead, our results suggest that the activation of word candidates in Japanese listeners' spoken-language processing is a process of continuous refinement of lexical options on the basis of any and all incoming phonetic information, just as indicated for English, Dutch, and other languages by the evidence summarized in the introduction.

Likewise, the lexical decision evidence reported in part 3 produced for Japanese the same pattern shown by Marslen-Wilson (1984) for English: the greater the distance between the nonword uniqueness point and the offset of a nonword, the faster are responses measured from that latter point. If the relevant unit of computation for lexical activation in Japanese were the mora, then one would expect that the relationship between response time and distance would be stronger when distance was measured in morae rather than phonemes; instead, it was weaker. From the point at which the stimulus item became a nonword, NO response was highly correlated with the distance to the end of the item—that is, the earlier the item became a nonword, the faster it could be rejected. Duration between nonword uniqueness point and item offset is of course fully determined by how much speech material—how many phonemes, how many morae—remains to be uttered post-NUP. However, the number of phonemes made a separate contribution to the response time measure, while the number of morae did not. This is because (by definition) the minimal quantum of information which distinguishes between two words, or a nonword and a word, is a phoneme. It does not mean that phonemes are a necessary level of representation in Japanese word recognition any more than morae are. Phonemes may be coarticulated so that information about, for instance, the place of articulation of a stop consonant may be available in the preceding phoneme; the recognition process is thus able to proceed in a truly continuous manner. Our results are thus in agreement with other recent findings on the contribution of internal structure to the perception of larger phonological units in speech. Nearey (in press) argues on the basis of an extensive review of speech perception work in phonetics that the perception of larger units such as syllables in English is entirely predictable from the perception of segmental parts of those units. Allen (1994) reaches the same conclusion on the basis of a review of psychoacoustic studies of the processing of speech signals, again mostly in English. Recent psychoacoustic experiments in Japanese by Kato (1999) have also suggested that perception of temporal units in speech input is predictable from perception of lower level (submoraic) information.

Our results thus make it clear that the relevant metric of similarity—allowing listeners to succeed in identifying a distorted word in a pun or to reconstruct a word from a nonword in a word reconstruction experiment—is not subject to limitations imposed by the intervention of "large processing units" in Japanese word recognition. Recognition of spoken Japanese words is a continuous process by which listeners evaluate the incoming signal as it comes in rather than via an initial "coarse coding" into morae. In short, our results do not support the claims of Dupoux et al. (1999) and others that the mora is the earliest available unit in spoken-word recognition by Japanese listeners.

Instead, the results support the predictions based on the universal framework proposed by Norris et al. (1997) for the interpretation of the role of rhythmic categories in speech recognition. Rhythmic categories do not intervene as structural units in word recognition; no intermediate level of representation in terms of morae is required to explain the recognition of Japanese any more than intermediate syllabic representations are necessary for the recognition of French or intermediate representations in terms of stress units are required to explain the recognition of English. Spoken-word recognition in all these languages is a continuous process of optimal evaluation of the incoming signal leading to activation and deactivation of word candidates. Note that although the uniqueness point (or, in our Experiment 5, the NUP) has a role to play in making distinctions between word candidates, it does not interrupt the continuous evaluation process; experiments in English have shown that information after the uniqueness point is also relevant in processing both words (Taft & Hambly, 1986) and nonwords (Goodman & Huttenlocher, 1988). This is indeed necessary for recovery from early mispronunciation or perceptual error, and it also of course makes it possible for listeners to appreciate puns and perform the word reconstruction task.

The details of our word reconstruction results further support our universalist interpretation. We found that preservation of consonantal information (in V-replacement) proved of significantly more use to listeners in several of our experiments than preservation of vowel information (in C-replacement). This effect was predicted in final position, in Experiment 2, simply on the basis that more of the intended word uniqueness point remained intact in the V-replaced than in the C-replaced condition. However, the effect also appeared in word-medial position, and Experiment 4 showed that simple position in the word—such that later phonemes could always be more easy to substitute—did not explain the difference. Instead, the results suggest a possibly greater constraint on lexical identity exercised by consonants than by vowels. Other word reconstruction experiments (Van Ooijen, 1996; Cutler et al., 2000) have shown that single-phoneme substitution is easier for vowels than for consonants in both Dutch and English (both of which have many vowels and many consonants) as well as in Spanish (with only five vowels and about four times that number of consonants). Thus the effect is insensitive to the relative number of vowels and consonants in a language. Cutler et al. (2000) argued that listeners have experience that vowels are more likely to be altered by surrounding phonetic context than consonants are. This has in turn generated repeated experience of necessary alteration of initial vowel hypotheses when these do not succeed in producing a viable lexical parse of the input. The fact that a vowel replacement advantage appears not only when listeners are explicitly selecting phoneme-sized substitutions (as is the case in the European experiments), but also when listeners are unaware that phonemic units are playing a role (as in the present study), suggests that the effect is operative at the level at which word candidates are activated and evaluated rather than as part of a process of conscious manipulation of phonetic structure.

The one aspect of our results which this suggestion cannot immediately account for is the failure of an advantage for vowel replacement to appear in Experiment 1, in which reconstruction was required given a distorted initial mora. In that experiment, RTs were 19 ms faster, and miss rates half-a-percentage lower, for vowel replacement than for consonant replacement, but neither of these effects reached significance. However, note that in that experiment more of the intended word remained continuously intact in the C-replaced (*namera*) than in the V-replaced condition (*kimera*). If, as we proposed for Experiment 2, reconstruction is easier the more of the word remains intact, it could be the case that in Experiment 1 this target intactness effect canceled out the (independent) vowel/consonant effect. Such an explanation would also account for why the vowel/consonant difference was rather stronger (robust across both RTs and miss rates) in Experiment 2 than in Experiments 3 and 4 (robust only across RTs): in Experiment 2 the target intactness effect reinforced the vowel/consonant effect.

In conclusion, then, we argue that the contribution of rhythmic categories in word recognition is the same for all languages. Although the mora is the smallest of the rhythmic categories proposed to play a role in the perception of spoken language, it no more participates as an integral unit of representation for input processing and lexical access than larger units such as the syllable or the stress unit do. Instead, the role of the mora, like that of the rhythmic categories relevant for the processing of languages other than Japanese, is in selecting between alternative possible parses of a continuous speech input. Speech input triggers automatic activation of multiple word candidates with which it is compatible; a competition process between these activated candidate words then ensues. This competition process is sensitive not only to match and mismatch between activated candidates and further arriving input, but also to the implications of each candidate for parsing the surrounding speech context. As argued by Norris et al. (1997), candidate words which leave adjacent speech input which could not itself be a word should be disfavored, i.e., reduced in activation. It is this possible-word computation in which rhythmic categories play a role. The boundaries of rhythmic units can function as one type of boundary within which speech context can be tested for viability as a potential word.

Norris et al. (1997) produced evidence that such possible-word computation affects word recognition in English and showed in computer simulations that their Possible-Word Constraint framework correctly captured available results on effects of stress units—the rhythmic categories relevant to English and Dutch—in spoken language processing in both those languages. McQueen et al. (2001) have extended Norris et al.'s finding of a PWC effect to Japanese and have shown, moreover, that morae are directly involved in assessing the size of the effect in Japanese. The findings of McQueen et al. (2001) thus foreshadow the present work; if morae are similar to other rhythmic categories in the way they participate in the computation of segmentation, then it is reasonable to expect that they are similar to other rhythmic categories in the way they participate in spoken-word recognition also. The present findings close the circle: Although morae are relevant to the process of speech segmentation for lexical recognition, they are not relevant to the process of initial lexical activation, despite recent claims that they play a necessary structural role in this process.

As we noted in the introduction, there are differences between the rhythmic categories. To our knowledge, it has never been claimed that stress units play a direct intermediate role in lexical activation in English - the only proposals concerning stress units have emphasised the importance of stress unit boundaries for segmentation, as currently captured by Norris et al.'s (1997) PWC proposal. It is only the smaller units - the syllable and the mora - for which a direct structural role has been proposed. Indeed, Dupoux et al. (1999) and Dehaene-Lambertz et al. (2000), while reporting data on the mora, couch their explanation in terms of a proposal originally formulated for the syllable (Mehler, Dupoux & Segui, 1990). However, note that recent evidence on the role of the syllable in processing spoken French (Content, Kearns et al., 2001; Content, Meunier et al., 2001) is also entirely consistent with our present universalist perspective: rhythmic boundaries are relevant for segmentation, but rhythmic units do not intervene as independent structural units. Content and colleagues report several converging lines of experimental evidence from French, from phoneme monitoring, fragment monitoring, word spotting, and explicit segmentation judgements and conclude that syllables do not constitute indivisible units which participate in the perceptual processing of spoken French. Instead, they argue, the important role of the syllable in speech segmentation for French listeners is centred on the function of syllable onsets as alignment points for word onsets. In other words, they place syllables in French in the same role as is proposed here and by McQueen et al. (2001) for morae in Japanese, and as Norris et al. (1997) proposed for stress units in English and Dutch. Rhythmic categories are not "units of perception," but modulate the activation and competition process so as to render spoken-word recognition as efficient as we listeners know it to be.

Finally, however, we return to another issue raised in the introduction to point out one way in which morae do differ from the other rhythmic categories. Morae, we argue, clearly play a greater role in listeners' awareness in Japanese than do stress units in English and Dutch or syllables in French. They are relevant for poetic forms, for orthography, and for many language games. Thus when Japanese speakers pun, or perform word reconstruction, they use conscious moraic representations, and show no awareness of universal effects involving submoraic similarity. The discrimination tasks used by Kakehi et al. (1996), Dupoux et al. (1999), and Dehaene-Lambertz et al. (2000) would, on our interpretation, have called on processing involving this level of awareness. However, we do not believe that the greater accessibility of morae than of other rhythmic categories to awareness should motivate claims for a special role in the early stages of word recognition. The recognition process is continuous and the role played in it by all rhythmic categories is alike.

APPENDIX

Materials Used in the Experiments

Each real word is listed with its English gloss and the replacement mora in the M nonword; e.g., for Experiment 1: *kamera* "camera" (*ni*). The C nonword is always constructed by replacing the target mora with the consonant of the M nonword plus the vowel in the original word, and the V nonword is always constructed by replacing the target mora with the vowel of the M nonword plus the consonant in the original word. For *kamera* the M nonword is thus *nimera*, the C nonword *namera*, and the V nonword *kimera*. An asterisk after the replacement mora, e.g., in Experiment 1: *kasetto* "cassette" (*te*)*, indicates that the item in question was also used, in all three of its versions, in Experiment 5.

Japanese word	English word	Mora
	Experiment 1 (word-initial CV morae replaced)	
dejitaru	"digital"	(*ta*)
hagaki	"postcard"	(*ru*)
kasetto	"cassette"	(*te*)*
sukaato	"skirt"	(*ne*)*
moderu	"model"	(*su*)
tabako	"cigarette"	(*ro*)
kamera	"camera"	(*ni*)
gasorin	"gasoline"	(*pi*)
rimokon	"remote control"	(*nu*)
tonneru	"tunnel"	(*me*)
suzume	"sparrow"	(*ro*)
kimuchi	"kimchi"	(*na*)*
zabuton	"cushion"	(*ku*)*
rajio	"radio"	(*gu*)*
taoru	"towel"	(*se*)*
mahuraa	"muffler"	(*se*)
karada	"body"	(*pi*)
takushii	"taxi"	(*no*)*
kemuri	"smoke"	(*hi*)*
masukomi	"mass communication"	(*se*)
supiido	"speed"	(*ha*)*
tomato	"tomato"	(*sa*)*
yubiwa	"ring"	(*ma*)*
kakezan	"multiplication"	(*ru*)*
yakusoku	"promise"	(*ko*)
tenpura	"tempura"	(*ko*)*
sebiro	"suit"	(*ko*)*
kurabu	"club"	(*mi*)

sarada	"salad"	(*mu*)
monitaa	"monitor"	(*se*)*
hoteru	"hotel"	(*se*)
kazari	"ornament"	(*te*)*
pantsu	"pants"	(*te*)
kanemochi	"richness"	(*ru*)*
misoshiru	"miso soup"	(*ra*)
takoyaki	"octopus fritter"	(*pe*)
bideo	"video"	(*na*)
nagagutsu	"rubber boots"	(*ku*)*
kodomo	"child"	(*se*)
yononaka	"world"	(*ha*)*
kanzume	"canned food"	(*hi*)*
namida	"tear"	(*te*)*
karaoke	"karaoke"	(*pi*)
hirune	"nap"	(*ge*)*
mayuge	"eyebrow"	(*to*)
pasokon	"personal computer"	(*te*)
tomodachi	"friend"	(*sa*)
sakana	"fish"	(*ro*)*
poteto	"potato"	(*ki*)*
hasami	"scissors"	(*ne*)*
kuchiburu	"lips"	(*so*)*
tamago	"egg"	(*so*)*
risutora	"restructure"	(*ko*)
sutamina	"stamina"	(*ko*)
hanabi	"fireworks"	(*ke*)*
katsudon	"rice with cutlet"	(*ri*)*
kutsushita	"socks"	(*ri*)
terebi	"television"	(*ka*)
mukashi	"old times"	(*po*)
supootsu	"sports"	(*na*)*

Experiment 2 (word-final CV morae replaced)

dejitaru	"digital"	(*ki*)
hagaki	"postcard"	(*me*)*
tebukuro	"gloves"	(*na*)*
manaita	"chopping board"	(*ze*)*
moderu	"model"	(*ko*)
tabako	"cigarette"	(*me*)
kamera	"camera"	(*ne*)*
gomibako	"waste basket"	(*ni*)
mayonaka	"midnight"	(*re*)*
purikura	"print club"	(*hu*)
wasabi	"horseradish"	(*no*)*
posuto	"mailbox"	(*ke*)*
hamigaki	"toothpaste"	(*po*)
jigoku	"hell"	(*na*)*
makura	"pillow"	(*ji*)*
kakizome	"first calligraphy on New Years Day"	(*to*)*
hayari	"fashion"	(*pa*)
natsumero	"old-time songs"	(*ba*)
hirune	"nap"	(*pa*)
masukomi	"mass communication"	(*no*)
garakuta	"junk"	(*ne*)*
tomato	"tomato"	(*be*)
nezumi	"mouse"	(*ro*)*
tempura	"tempura"	(*ki*)

yakusoku	“promise”	(*me*)
kamaboko	“boiled fish paste”	(*pa*)*
sebiro	“suit”	(*ka*)
yasumi	“recess”	(*na*)
sarada	“salad”	(*no*)*
kaminari	“thunder”	(*bo*)*
hoteru	“hotel”	(*ka*)
kazari	“ornament”	(*bo*)
nomiya	“bar”	(*ro*)*
nagasode	“long sleeve”	(*pa*)*
misoshiru	“miso soup”	(*ga*)
takoyaki	“octopus fritter”	(*pa*)
goruhu	“golf”	(*mo*)*
yakiniku	“barbecue”	(*mo*)*
kodomo	“child”	(*se*)*
yononaka	“world”	(*bo*)
teeburu	“table”	(*ke*)
namida	“tear”	(*ko*)
tengoku	“heaven”	(*ri*)*
miyage	“souvenir”	(*sa*)
hasami	“scissors”	(*go*)
mikazuki	“crescent moon”	(*no*)*
toshiyori	“old person”	(*ko*)*
sakana	“fish”	(*re*)
poteto	“potato”	(*sa*)
mayuge	“eyebrow”	(*so*)*
kuchibiru	“lips”	(*ko*)
tamago	“egg”	(*hi*)
risutora	“restructure”	(*be*)
sutamina	“stamina”	(*pe*)
chikoku	“delay”	(*ni*)*
kudamono	“fruit”	(*ke*)*
kutsushita	“socks”	(*po*)*
terebi	“television”	(*mo*)*
kisoku	“rules”	(*me*)*
kuchibeni	“lipstick”	(*ke*)*

Experiment 3 (word-medial CV morae replaced)

kakezan	“multiplication”	(*pu*)
gihuto	“gift”	(*ke*)
gasorin	“gasoline”	(*me*)
kanazuchi	“hammer”	(*ho*)*
kurabu	“club”	(*go*)
wakaba	“young leaves”	(*ho*)*
miruku	“milk”	(*be*)
guratan	“gratin”	(*ki*)*
zabuton	“cushion”	(*ke*)
sugoroku	“a board game”	(*ma*)*
yubiwa	“ring”	(*ne*)
kanojo	“she”	(*ma*)*
takushii	“taxi”	(*mo*)
kimono	“kimono”	(*te*)*
gurahu	“graph”	(*mo*)
takenoko	“bamboo shoot”	(*pu*)*
wagomu	“rubber band”	(*ne*)*
nagagutsu	“rubber boots”	(*ho*)
tsumiki	“brick”	(*ne*)*
masukomi	“mass communication”	(*pa*)

pasokon	"personal computer"	(*ma*)
tomato	"tomato"	(*ke*)
remon	"lemon"	(*gi*)*
horumon	"hormone"	(*ka*)*
yakusoku	"promise"	(*po*)*
mahuraa	"muffler"	(*ro*)
sebiro	"suit"	(*no*)
rizumu	"rhythm"	(*ga*)*
sarada	"salad"	(*mo*)
monitaa	"monitor"	(*pa*)
megane	"eyeglasses"	(*so*)*
beruto	"belt"	(*ko*)*
puraza	"plaza"	(*ku*)
rakurosu	"lacrosse"	(*pe*)*
misoshiru	"miso soup"	(*ra*)
takoyaki	"octopus fritter"	(*mi*)
gorufu	"golf"	(*pe*)
rimokon	"remote control"	(*se*)
kodomo	"child"	(*ne*)
yononaka	"world"	(*hi*)
pokemon	"pocket monster"	(*na*)*
burashi	"brash"	(*me*)*
katarogu	"catalog"	(*mo*)*
bikini	"bikini"	(*pa*)
pasuta	"pasta"	(*ko*)
karaoke	"karaoke"	(*hu*)
kamisori	"razor"	(*ka*)*
baketsu	"bucket"	(*ma*)*
poteto	"potato"	(*na*)
hanage	"nose hair"	(*so*)*
wagamama	"selfish"	(*re*)*
tamago	"egg"	(*ku*)
risutora	"restructure"	(*po*)
roketto	"rocket"	(*sa*)*
nekoze	"hunchback"	(*me*)*
rakudai	"fail"	(*mo*)*
yakezake	"drink in despair"	(*ha*)*
nanatsu	"seven"	(*ho*)*
dorama	"drama"	(*ke*)*
yakimochi	"jealousy"	(*po*)*

Experiment 4 (word-medial CV morae replaced)

Second mora		
yakezake	"drink in despair"	(*ha*)
kanazuchi	"hammer"	(*ho*)
risutora	"restructure"	(*po*)
yakusoku	"promise"	(*po*)
nagagutsu	"rubber boots"	(*ho*)
katarogu	"catalogue"	(*mo*)
misoshiru	"miso soup"	(*ra*)
takenoko	"bamboo shoot"	(*pu*)
mahuraa	"muffler"	(*ro*)
zabuton	"cushion"	(*ke*)
roketto	"rocket"	(*sa*)
wagamama	"selfish"	(*re*)
pasokon	"personal computer"	(*ma*)
karaoke	"karaoke"	(*hu*)
rakudai	"fail"	(*mo*)

rimokon	"remote control"	(*se*)
horumon	"hormone"	(*ka*)
monitaa	"monitor"	(*pa*)
gasorin	"gasoline"	(*me*)
rakurosu	"lacrosse"	(*pe*)
kakezan	"multiplication"	(*pu*)
yononaka	"world"	(*hi*)
masukomi	"mass communication"	(*pa*)
sugoroku	"a board game"	(*ma*)
guratan	"gratin"	(*ki*)
takushii	"taxi"	(*ro*)
pokemon	"pocket monster"	(*na*)
kamisori	"razor"	(*ka*)
takoyaki	"octopus fritter"	(*mi*)
yakimochi	"jealousy"	(*po*)
Third mora		
bakudan	"bomb"	(*po*)*
kosumosu	"cosmos"	(*za*)*
chirigami	"tissue paper"	(*te*)*
hanamizu	"runny nose"	(*ra*)*
nekutai	"necktie"	(*po*)*
tomodachi	"friend"	(*so*)*
gomibako	"waste basket"	(*nu*)*
harusame	"bean-starch vermicelli"	(*re*)*
kurubushi	"ankle"	(*ga*)*
manekin	"mannequin"	(*mo*)*
hamigaki	"toothpaste"	(*zu*)*
rajikon	"radio control"	(*me*)*
chikamichi	"shortcut"	(*re*)*
tokudane	"scoop"	(*ro*)*
murasaki	"purple"	(*ne*)*
kakegoe	"shout"	(*ri*)*
harapeko	"hungry"	(*sa*)*
hachimitsu	"honey"	(*pe*)*
pokeberu	"pocket bell"	(*zo*)*
mizutama	"polka-dot"	(*re*)*
mezamashi	"alarm clock"	(*ki*)*
masayume	"prophetic dream"	(*ka*)*
gokiburi	"cockroach"	(*re*)*
tsukemono	"pickles"	(*ka*)*
nakimushi	"crybaby"	(*ro*)*
panorama	"panorama"	(*ze*)*
basutabu	"bathtub"	(*ge*)*
haramaki	"stomach band"	(*ho*)*
rajikase	"radio cassette"	(*to*)*
monosashi	"measure"	(*mo*)*

REFERENCES

Allen, J. B. (1994). How do humans process and recognize speech? *IEEE Transactions on Speech and Audio Processing*, **2,** 567–577.

Amano, S., & Kondo, K. (1999). *NTT Deetabeesu Shiriizu Nihongo no Goitokusei* (Vol. 1). Sanseido: Tokyo.

Boelte, J. (1997). *The role of mismatching information in spoken word recognition*. Hamburg: Verlag Dr. Kovac.

Cole, R. A. (1973). Listening for mispronunciations, a measure of what we hear during speech. *Perception & Psychophysics*, **11,** 153–156.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics*, **56,** 624–636.

Content, A., Meunier, C., Kearns, R. K., & Frauenfelder, U. H. (2001). Sequence detection in pseudowords: Where is the syllable effect? *Language and Cognitive Processes*, **16**.

Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*, **45**, 177–199.

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

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

Cutler, A., & Norris, D. G. (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., Sebastian-Gallés, N., Soler Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, **28,** 746–755.

Cutler, A., & Young, D. (1994). Rhythmic structure of word blends in English. In *Proceedings of the Third International Conference on Spoken Language Processing, Yokohama*, (Vol. 3, pp. 1407–1410). Tokyo: Acoustical Society of Japan.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlation of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, **12,** 635–647.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, **25,** 1568–1578.

Echols, C. H. (1996). A role for stress in early speech segmentation. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 151–170). Hillsdale, NJ: Erlbaum.

Ferrand, L. Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, **35,** 708–723.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning, A distributed model of speech perception. *Language and Cognitive Processes*, **12,** 613–656.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: effects of competition and inhibition. *Journal of Memory and Language*, **28,** 501–518.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18,** 1211–1238.

Goodman, J. C., & Huttenlocher, J. (1988). Do we know how people identify spoken words? *Journal of Memory and Language*, **27,** 84–698.

Inagaki, K., Hatano, G., & Otake, T. (2000). The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of Experimental Child Psychology*, **75,** 70–91.

Inagaki, S. (1985). *Edo no Asobie*, Tokyo: Tookyooshoseki.

Kakehi, K., Kato, K., & Kashino, M. (1996). Phoneme/syllable perception and the temporal structure of speech. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 125–143). Berlin: Mouton de Gruyter.

Kashino, M., van Wieringen, A., & Pols, L. (1992). Cross-language differences in the identification of intervocalic stop consonants by Japanese and Dutch listeners. *Proceedings of the Third International Conference on Spoken Language Processing, Yokohama* (Vol. 3, pp. 1079–1082). Tokyo: Acoustical Society of Japan.

Katada, F. (1990). On the representation of moras: Evidence from a language game. *Linguistic Inquiry*, **21,** 641–646.

Kato, H. (1999). *Perceptual characteristics of temporal structures in speech: Towards objective assessment of synthesis rules*. Ph.D. thesis, Kobe University.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language & Speech*, **32,** 249–278.

Kubozono, H. (1990). The role of the mora in speech production of Japanese. *Proceedings of the International Conference on Spoken Language Processing, Kobe, Japan* (Vol. 1, pp. 501–504). Tokyo: Acoustical Society of Japan.

Mann, V. A. (1991). Are we taking too narrow a view of the conditions for development of phonological awareness? In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman* (pp. 55–64). Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 125–150). Hillsdale, NJ, Erlbaum.

Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, **101,** 653–675.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18,** 1–86.

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.

McQueen, J. M., Norris, D. G., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, **25,** 1363–1389.

McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*, **45**, 103–132.

Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society London B* **295,** 333–352.

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

Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 236–262). Cambridge, MA: MIT Press.

Nearey, T. M. (in press). On the factorability of phonological units in speech perception. In J. Local & R. Ogden (Eds), *Papers in laboratory phonology VI*. Cambridge, UK: Cambridge Univ. Press.

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

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

van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, **24,** 573–583.

Otake, T. (1990). Rhythmic structure of Japanese and syllable structure. *IEICE Technical Report*, **89,** 55–61.

Otake, T. (2000). *Listeners' representations of within-word structure: A cross-linguistic study of Japanese and English*. Paper presented at the Seventh Conference on Laboratory Phonology, Nijmegen.

Otake, T., & Cutler, A. (2001). Recognition of (almost) spoken words: Evidence from word play in Japanese. *Proceedings of EUROSPEECH 2001, Aalborg*, September, 2001.

Otake, T., Davis, S., & Cutler, A. (1995). Listeners' representations of within-word structure: A cross-linguistic and cross-dialectal investigation. In *Proceedings of EUROSPEECH 95, Madrid* (Vol. 3, pp. 1703–1706).

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32,** 358–378.

Otake, T., Hatano, G., & Yoneyama, K. (1996). Speech segmentation by Japanese listeners. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies* (pp. 183–201). Berlin: Mouton de Gruyter.

Otake, T., & Yamamoto, K. (1997). *Listeners' representations of within-word structure for bilingual Japanese*. Paper presented at the 134th meeting Acoustical Society of America, San Diego.

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.

Otake, T., Yoneyama, K., & Maki, H. (1998). Non-native listeners' representations of within-word structure. In *Proceedings of ICA/ASA* (Vol. 3, pp. 2067–2068).

Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, **21,** 1297–1311.

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Acoustic-phonetic priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13,** 64–75.

Suzuki, T. (1959). *Kotoba Asobi Jiten*, Tokyo: Tookyoodoo.

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

Taft, M., & Hambly, G. (1986). Exploring the Cohort Model of spoken word recognition. *Cognition*, **22,** 259–282.

Yip, M. (1982). Reduplication and C-V skeleta in Chinese secret languages. *Linguistic Inquiry*, **13,** 637–661.

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

(Received December 23, 1999)
(Revision received March 14, 2001)
(Published online December 13, 2001)