Mora or Phoneme? Further Evidence for Language-Specific Listening

ANNE CUTLER

MRC Applied Psychology Unit, Cambridge, United Kingdom, and Max-Planck-Institut fur Psycholinguistik,

Nijmegen, The Netherlands

AND

TAKASHI OTAKE

Dokkyo University, Soka, Japan

Japanese listeners detect speech sound targets which correspond precisely to a mora (a phonological unit which is the unit of rhythm in Japanese) more easily than targets which do not. English listeners detect medial vowel targets more slowly than consonants. Six phoneme detection experiments investigated these effects in both subject populations, presented with native- and foreign-language input. Japanese listeners produced faster and more accurate responses to moraic than to nonmoraic targets both in Japanese and, where possible, in English; English listeners responded differently. The detection disadvantage for medial vowels appeared with English listeners both in English and in Japanese; again, Japanese listeners responded differently. Some processing operations which listeners apply to speech input are language-specific; these language-specific procedures, appropriate for listening to input in the native language, may be applied to foreign-language input irrespective of whether they remain appropriate. © 1994 Academic Press, inc.

Listening to speech is in large part language-specific. This would be a trivial observation if it only concerned words; clearly, listeners can only understand spoken words in a language they know. But language-specificity in listening extends well beyond the lexicon; for instance, prelexical processing of speech is highly language-specific. The most obvious manifestation of this is in the available phonological repertoire in terms of which listeners interpret speech input; speech contrasts which are not part of the native repertoire are notoriously difficult to perceive. Thus, for example, English-speakers who do not have a distinction between retroflex and dental stops in their language cannot perceive this contrast in Hindi input (Werker & Tees, 1984); Japanese-speakers, who do not have a contrast between /r/ and /l/, have trouble hearing this distinction in English (Goto, 1971).

These limitations in making phonologically relevant distinctions do not simply arise because the relevant low-level perceptual skills have atrophied in the absence of suitable input; Best, McRoberts, and Sit-

hole (1988) effectively showed that perceptual skills remain intact for distinctions which are not present in the native language but are also not preempted by any native language contrast (in their case, Zulu clicks, to which English contains no similar contrast, were easily discriminated by English-speakers). Thus the limitations are located not in initial percepts but in linguistic processing: part of efficient speech recognition is rapid assignment of speech sound percepts to native phonological categories, and a side effect is that other categories are excluded from consideration.

Similarly, there are efficient language-specific procedures for segmenting continuous speech input. Segmentation is a problem for the listener, because spoken utterances really are continuous, in that there are only rarely gaps between words, or other cues to word boundary location. Nevertheless, speech recognition must be a process of identifying individual words in the input, because it is individual words which a listener already knows and has stored in lexical memory; whole utterances will usually never have been heard before. This aspect of speech recognition is also extremely efficient, in that we rarely experience any conscious difficulty with segmentation, just as we rarely experience any conscious difficulty with phonetic categorization.

In English, such efficiency is achieved by exploiting the language's characteristic stress rhythm, in which strong syllables (containing full vowels) contrast with weak syllables (containing reduced vowels); English-speakers effectively use the stress rhythm to segment speech by assuming that strong syllables are word-initial. Evidence for this comes from spontaneous and experimentally induced word boundary misperceptions, which in English systematically involve boundaries hypothesized at strong syllable onsets but overlooked at weak syllable onsets (Cutler & Butterfield, 1992). Further evidence arises from word-spotting experiments in which listeners find it hard to spot a real word embedded in nonsense if the word spans a strong syllable boundary, i.e., a segmentation point (Cutler & Norris, 1988); similarly, listeners find monosyllabic words easier to spot when they are preceded by weak syllables than when they are followed by weak syllables (McQueen, Norris, & Cutler, 1994). A procedure of assuming strong syllables to be word-initial is highly efficient in English because it corresponds to reality—most strong syllables in English *are* word-initial (Cutler & Carter, 1987).

However, a stress-based segmentation procedure must of necessity be language-specific, since many languages do not have the opposition between strong and weak syllables so characteristic of English. Indeed, evidence from French suggests that listeners employ a syllabic segmentation procedure; target detection experiments show that targets which correspond exactly to syllables are detected faster than targets which do not (Mehler, Dommergues, Frauenfelder, & Segui, 1981; Segui, Frauenfelder, & Mehler, 1981).

All languages can be described in terms of syllables, so syllabic segmentation is not necessarily language-specific in the sense that stress-based segmentation is; nevertheless, the procedure characteristic of French listeners does appear to be language-specific, since experiments parallel to those of Mehler et al. (1981) show that syllabic segmentation is used only in restricted circumstances by speakers of Spanish and Catalan (Bradley, Sanchez-Casas, & Garcia-Albea, 1993; Sebastian-Gall6s, Dupoux, Segui, & Mehler, 1992) and is not used at all by English speakers (Cutler, Mehler, Norris, & Segui, 1986) or by Japanese speakers (Otake, Hatano, Cutler, & Mehler, 1993).

The use of syllabic segmentation by French listeners and stress-based segmentation by English listeners suggests that language rhythm may be the key to predicting the basis of speech segmentation: just as stress is the basis of speech rhythm in En-

glish, so is the rhythm of French syllabically based. In Japanese, in contrast, the unit of language rhythm is the mora; Japanese poetic forms like the *haiku* are defined in terms of number of morae per line. The mora is a subsyllabic unit which can be a vocalic nucleus (V), a nucleus plus syllabic onset (CV or CCV), or a consonant, usually a nasal consonant (N), in syllabic coda position. To illustrate moraic structure with some well-known names, *Sega* has two morae (CV-CV: *Se-ga), Honda* has three (CV-N-CV: *Ho-n-da),* and *Tokyo* has four (CV-V-CCV-V: *To-o-kyo-o).* The extension of the rhythmic segmentation hypothesis from French and English to Japanese led Otake *et al.* (1993) to predict moraic segmentation by Japanese listeners. In their study listeners were presented with words like *tanishi* (CV-CV-CV) and *tanshi* (CV-N-CV) and were asked to detect targets corresponding to the first two or three phonemes. Note that in both examples the first two phonemes constitute exactly the first mora; the first three phonemes correspond to one-and-a-bit morae in *tanishi* and two morae in *tanshi.* Otake *et al.* found that two-phoneme targets were detected equally rapidly and accurately in both word types, as the moraic hypothesis predicted, since they correspond exactly to the initial mora in both word types. Three-phoneme targets, however, were detected accurately when they corresponded exactly to two morae (e.g., in *tanshi),* but simply were not responded to when they corresponded to one-and-a-bit morae (e.g., in *tanishi);* again, this is as the moraic hypothesis predicted.

Moreover, moraic segmentation is another language-specific effect: Otake *et al.* found that neither English listeners nor French listeners produced the response pattern characteristic of Japanese listeners when presented with the same materials.

What the French listeners in Otake et al.'s experiment did show, however, was a characteristically *syllabic* response pattern with the Japanese input. French listeners

also showed syllabic responding with English input (Cutler et al., 1986). In other words, these listeners were applying to foreign-language input the segmentation procedures which work efficiently with their native language—irrespective of whether or not the procedures would suit the different input. This suggests that segmentation procedures may indeed be highly similar to phonological categorization procedures: they effectively aid processing of the native language, but they may reduce processing efficiency for input in a foreign language. The potential importance of this conclusion warrants further investigation of the topic, and the present study was designed to achieve this. French listeners have been shown to use their characteristic syllabic segmentation with both English and Japanese input; the next step is therefore to investigate whether the other language-specific segmentation procedures are also applied to foreign-language input by speakers who use them for their native language. Stress-based segmentation by English listeners has not been tested with foreign-language input because the languages with which English has chiefly been compared do not have English-like stress contrasts and therefore do not allow a simple test of the hypothesis; future extensions to other languages will be necessary to achieve this test. The mora, however, is a phonological concept which, at least in certain respects, can be applied to the description of many languages. This it is relatively easy to test the moraic hypothesis on non-Japanese as well as on Japanese materials.

This study therefore aims to investigate whether moraic segmentation by Japanese listeners, like syllabic segmentation by French listeners, also appears with foreign-language input. In addition, it has another aim, namely, a further test of the mora hypothesis. This is called for because, on the one hand, only the one study (Otake *et al.,* 1993) has so far tested it, and converging evidence from differing tasks is always desirable; on the other hand, it is also the case

that the morae involved in Otake et al.'s study were primarily CV, which is the most common mora structure (Otake, 1990). In the present study we therefore used single-phoneme morae, which are less common, in an attempt to ascertain the generality of moraic segmentation.

The task used here is phoneme detection. This allows us to add a further dimension to the present study, because there appear to be language-specific effects (albeit not segmentation effects) in the phoneme detection task. In this task, listeners' response time and accuracy in detecting single-phoneme targets in spoken input (usually lists of words) are measured. Recent studies have shown that English listeners detect vowel targets both more slowly and less accurately than stop consonant targets, even when the former are highly distinct ([a], [i]) while the latter are confusable ([p], [t]; van Ooyen, 1994; van Ooyen, Cutler, & Norris, 1991); vowels are also detected more slowly than fricatives (Norris, van Ooyen & Cutler, 1992). The disadvantage for vowels in comparison to stop consonants or fricative targets did *not* appear, however, with Spanish subjects listening to Spanish words in an unpublished study by van Ooyen and Sanchez-Casas (1993). The vowel repertoires of English and Spanish are, of course, very different: English has many, highly confusable vowels, while Spanish has only five vowels, which occupy distinct positions in vowel space. Also, dialect distinctions are signaled primarily by vowel quality in English but not in Spanish. In both respects Japanese resembles Spanish more closely than it resembles English: Japanese, too, has just five, highly distinct, vowels, and dialect distinctions are signalled less by vowel quality than they are in English. It is therefore of interest to determine whether vowel/consonant differences in phoneme detection appear in Japanese.

The present study was designed as a cross-linguistic series of experiments, in which listeners were presented with native-and foreign-language materials. To begin with, in Experiment 1 Japanese listeners were presented with Japanese materials. They were asked to detect vowel and consonant targets, which could be moraic or nonmoraic. The moraic hypothesis predicts that moraic targets will be easier to detect than nonmoraic. The experiment also allows a test of whether phoneme detection in Japanese will be differentially difficult for vowels versus consonants.

EXPERIMENT 1

Method

Materials. Two phoneme targets were chosen, one vowel and one consonant. These will be signified as O and N, respectively. Japanese has only five vowels, and one of these is described in Maddieson (1984) as the lower-mid-back rounded vowel [o]; this is the sound represented here by 0. Japanese has, again according to Maddieson (1984), 21 distinct consonant sounds, including three nasals; N here represents a sound that, in our materials, was realized phonetically as a dental-alveolar voiced nasal in syllable-initial position, but as a palato-velar voiced nasal in syllable-final position.

Thirty-two meaningful Japanese words, covering a wide range of frequency of occurrence, were chosen as stimulus words. In half of these the designated target phoneme was O, in half it was N. Counterbalanced with this factor were two further factors: in half of the words the target phoneme occurred as second phoneme in the word, in half as third phoneme, and half of the target phonemes were single-phoneme morae, while half formed part of CV morae. The phonetic contexts in which the target phonemes occurred were held constant as far as possible in order to ensure uniformity of realization across target occurrences. In Roman transcription the stimulus words with O target were *aoki, aokabi, aokusa, aokasu, kaori, taoru, haori, saori* (all moraic) and *kokeshi, kokage, tokai, tokage,*

atokuchi, itoku, utoku, etoku (nonmoraic), and the stimulus words with N target were *inka, unka, enka, anka, kinko, kinka, kanko, kanki* (moraic) and *inori, unome, enoki, anoyo, kinori, kinoko, kanoko, kanojo* (nonmoraic).

A further 245 words were chosen and arranged into 64 sequences varying in length from two to six words. In the 32 experimental sequences the target words occurred in the second, third, fourth, or fifth position. Of the 32 sequences which did not contain one of the experimental target words, half contained no occurrence of the specified target, while the other half contained a dummy target. The dummy target words contained the phoneme target in phonetic contexts and/or positions in the word which differed from those in the experimental target words; for O, the dummy target words were *okuba, sora, sugao, hiroba, hikishio, oiru, gohan, kimochi;* for N they were *rokuon, wani, ringo, nomi, botan, tanuki, nedoko, hanki.* Also, the dummy target words occurred in positions in the sequence different from those of the experimental targets, including first and last position.

Twelve practice sequences were also constructed. These too varied in length from two to six words, and four of them had no occurrence of the specified target.

The experimental and practice sequences were recorded on digital audio tape by a male native speaker of Standard Tokyo Japanese. Each sequence was preceded by its number. The words were spoken at a normal rate with approximately two seconds between words and approximately 5 s between sequences.

Subjects. Forty undergraduate members of Dokkyo University took part in the experiment for a small payment. None were students of language-related subjects.

Procedure. The subjects were tested in pairs in a quiet room. They were instructed to listen for a word containing the sound represented by the Roman character (either O or N), which was specified as target for each sequence, and to press a response key as soon as they had detected any occurrence of this target. (Because targets could occur anywhere is a word, this is the "generalized phoneme-monitoring" procedure; Frauenfelder & Segui, 1989.) The target for each sequence was presented visually on a 15 x 17-cm card, immediately prior to the beginning of the sequence.

The sequences were presented over headphones from a DAT recorder. The output from this recorder was also fed via a mixer to a second DAT recorder which via the same mixer also recorded a pulse triggered by the subject's response.

The intervals between onset of the target phoneme and response pulse were measured individually for each subject on a Kay SonaGraph 5500 to ascertain reaction times from target onset.

Results and Discussion

Mean number of missed responses and mean response times (RTs) were determined for each subject and each item, and separate analyses of variance were conducted on each measure with subjects and with items as random factors. In this and in all following experiments, we report as significant only those effects which reached our c riterion of significance in both subjects and items analyses.

The mean RTs and miss rates are shown in Table 1. Analyses of variance showed that the only effect significant in both RT analyses was the difference between moraic targets (mean RT 642 ms) and nonmoraic targets (mean RT 681 ms; $Fl[1,39] = 27.08$, $p < .001$; $F2[1,24] = 7.73$, $p < .01$). The main effects of vowel versus consonant

TABLE 1
MEAN RESPONSE TIME (MS) AND MISS RATE AS A FUNCTION OF MORAIC STRUCTURE FOR VOWEL AND CONSONANT TARGETS, RESPECTIVELY: EXPERIMENT 1 (JAPANESE LISTENERS, JAPANESE WORDS)

Target phoneme	O	N
Moraic	640(5.3%)	645(10%)
Nonmoraic	679 (26.2%)	685 (24.7%)

target and of target position in the word were not significant, and no interactions reached significance. Exactly the same pattern appeared in the analysis of missed responses: only the main effect of mora structure (7.7% of moraic targets versus 25.5% of nonmoraic targets missed) reached significance (Fl[1,39] = 56.29, $p < .001$; F2[l,24] = 13.42, $p < .005$).[1] Thus it would appear that Japanese listeners' faster response to targets which correspond to morae as opposed to targets which do not correspond to morae is not restricted to CV morae; exactly the same pattern is found with single-phoneme targets. The present result thus further strengthens the conclusion of Otake et al. (1993): mora structure is crucially involved in the process by which Japanese listeners convert spoken input into lexically accessible representations. The mora can form the basis of speech segmentation for native speakers of this language.

No effect of the vowel/consonant manipulation appeared, either in response times or in miss rates. This failure of Japanese listeners to show a differences in vowel versus consonant detection further suggests that such differences may indeed, as van Ooyen and Sanchez-Casas' (1993) preliminary comparison of English with Spanish had suggested, be language-specific.

As with Otake et al.'s (1993) experiments, however, the present series is designed as a cross-linguistic investigation, and one purpose of the comparison across listener groups is to rule out possible alternative explanations of the predicted result. For instance, it could be the case that acoustic differences between moraic and nonmoraic realizations of our targets O and N in Japanese speech might have rendered some targets intrinsically easier to perceive than others; that is, the faster detection

times for moraic targets might simply result from greater acoustic clarity. One form which such articulatory differences might take is that moraic phonemes might have longer duration than nonmoraic; certainly, this is the case with N (Sato, 1993). For vowels at least, longer duration leads to faster response times in the phoneme-monitoring performance of both English and Spanish listeners (Cutler, van Ooyen, Norris, & Sanchez-Casas, 1994). If such acoustic effects were indeed responsible for the results of Experiment 1, then moraic targets should be more rapidly and accurately detected by any listener, native speaker of Japanese or not. A test for this possibility was provided by Experiment 2, in which the same materials were presented to a nonnative listener group.

EXPERIMENT 2

Method

Materials. The materials were the same as those in Experiment 1. However, the tape was copied and in the process the numerals at the start of each sequence were removed, since these would essentially be nonwords to English listeners and hence would merely function to lengthen each sequence by one item. A timing mark, aligned roughly with the onset of each target item, was placed on the second channel of the tape (where it would be inaudible to the subjects).

Subjects. Twenty-four students at Sussex University took part in the experiment for a small payment. None had any knowledge of Japanese.

Procedure. The subjects were tested individually in a sound-dampened room. They were presented with instructions which described the task and illustrated the materials with Japanese words and names which would be known to English subjects, such as *Tokyo, Sony, sushi.*

The sequences were presented over headphones from a DAT recorder. The targets were displayed in uppercase on a VDU

---

[1] Experiment 1 was in fact run twice. In a separate run, with 37 subjects, an error in target assignment meant that one item had to be dropped from the analysis. The pattern of results, however, was identical to that reported here as Experiment 1.

screen. Target presentation, timing, and response collection were under the control of a Zenith microcomputer running the TSCOP experimental software (Norris, 1984).

The intervals between target onset and timing mark were measured and the response times adjusted by these amounts to give responses from exact target onset.

Results and Discussion

Mean number of missed responses and mean RTs were determined for each subject and each item, and separate analyses of variance were conducted with subjects and with items as random factors. Mean RTs and mean number of missed targets for each condition are shown in Table 2.

The moraic effect found in Experiment 1 was not replicated with these nonnative listeners; the effect of mora structure was nonsignificant both in the analysis of RTs and in the analysis of miss rates. The main effect of target position in the word was also nonsignificant in both. The main effect of vowel versus consonant target was, however, significant; consonants (mean RT 687 ms, mean miss rate 12.5%) were detected both faster and more accurately than vowels (mean RT 731 ms, mean miss rate 30.2%; $Fl[l,23] = 4.79$, $p < .04$, $F2[1,24] = 4.5, p < .05$ for RT, $Fl[l,23] = 28.17$, $p < .001, F2[1,24] = 21.15, p < .001$ for miss rates). Exactly one significant interaction was observed, between target position and moraic structure in the miss rate analysis: in second position moraic targets were detected more accurately than nonmoraic, but in third position nonmoraic targets were de-

tected more accurately than moraic ($Fl[l,23] = 7.96$, p< .01;$F2[1,24] = 4.68$, $p < .05$). We can see no obvious explanation of this interaction. As Table 2 shows, RTs to moraic vowels were faster than to nonmoraic (with a difference in the opposite direction for consonants); however, the interaction between moraic structure and vowel versus consonant targets did not reach our criterion of significance, since it was nonsignificant across items (although significant at .02 across subjects). Separate *t* tests showed that the effect of moraic structure was in fact nonsignificant for both vowels and consonants.

These results do not at all resemble the results produced by the native listener group. Thus Experiment 2 has ruled out the possibility that the mora effect we observed with Japanese listeners in Experiment 1 might be simply due to heightened acoustic clarity of moraic phonemes.

The results do not simply constitute a null-result control finding, however. The one effect which was significant in both RTs and miss rates—the detection advantage for consonantal over vowel targets-is itself suggestive of language-specific responding on the part of the present subject group. It will be recalled that this effect was not observed with Japanese listeners in Experiment 1, but that other studies have suggested that it is typical of native English listeners detecting phoneme targets in medial position in a word (van Ooyen, 1994; van Ooyen et al., 1991; Norris et al., 1992). Here the effect has appeared even with input which is acoustically less ambiguous than that in which the effect first appeared. In contrast to English, with its many, confusable vowels, Japanese has only five vowels and these are relatively distinct. Thus the present result suggests that for English listeners, detection of vowels is more difficult than detection of consonants irrespective of the complexity of the vowel repertoire present in the input.

As described in the introduction, French listeners have been shown to apply syllabic

TABLE 2

MEAN RESPONSE TIME (MS) AND MISS RATE AS A FUNCTION OF MORAIC STRUCTURE FOR VOWEL AND CONSONANT TARGETS, RESPECTIVELY: EXPERIMENT 2 (ENGLISH LISTENERS, JAPANESE WORDS)

Target phoneme	O	N
Moraic	705(31.8%)	696(13%)
Nonmoraic	756 (28.6%)	677 (12%)

segmentation not only to French-language input, but also to input in English (Cutler et al., 1986) and Japanese (Otake et al., 1993). It now appears that application of native-language procedures to foreign-language input may extend beyond this initial case. Thus Experiment 2 offers further evidence in support of an important finding which has previously been confined to one experimental effect only: procedures which listeners apply to input in their native language may also be applied to input in a foreign language.

However, the present result, in which O was detected both less accurately *and* more slowly than N, actually contrasts in part with the Norris et al. (1992) study, since in that study detection of English [a] and [i] was found to be *not* significantly worse than detection of the English nasal consonants [m] and [n]. A significant disadvantage for vowels appeared in the miss rate for targets in word-initial position, but not in word-medial position (the position analogous to that of targets in the present experiment). Two obvious explanations of this difference present themselves: it could simply be that the present O targets are harder for English listeners than the vowels tested by Norris et al., or it could be that the English listeners' difficulty with vowel targets in the present study reflects a mismatch between the input and expectations based on the native vowel repertoire. Such expectations certainly do affect vowel perception in foreign language input (Bohn & Flege, 1990). Of course, there is also some mismatch between the consonants of English and Japanese; but is fair to say that the N targets in the Japanese input might be closer than the O targets to English listeners' expectations based on their native phonemic repertoire. Experiment 3, conceived as a direct English analogue of Experiment 1, allows us to shed further light on this issue, as well as providing the next step in our investigation of the language-specificity of the mora effect shown by Japanese listeners.

Method

Materials. Thirty-two meaningful English words were chosen as stimulus words. In half of these the designated target phoneme was O, in half it was N. The vowel target was always realized as the low-back-rounded [D], as in British English pronunciations of *hot, bother.* The consonant target was, as in the Japanese materials, realized somewhat differently in syllable-initial (voiced alveolar-nasal) and syllable-final position (voiced alveolar or palatal-nasal). As in the preceding experiments, two further factors were counterbalanced with the vowel/consonant factor. First, in half of the words the target phoneme occurred as second phoneme in the word and in half as third phoneme. Second, an attempt was made to match the English materials to the moraic structure contrast in the Japanese materials. This was done in the following way: half of the target phonemes were preceded by a vowel (in the case of O) or formed a syllable code preceding a consonantal onset (in the case of N). These items (e.g., O in *eon, geography;* N in *inlet, conceal)* were therefore roughly matched to the Japanese moraic targets. The other half occurred in alternating sequences of consonants and vowels (e.g., O in *horrify, abolish;* N in *enamel, sanity)* and hence were roughly matched to the Japanese nonmoraic targets. In fact, it proved extremely difficult to achieve this goal of matching to the Japanese materials, and in both cases the matching is far from perfect. In particular, vowel-vowel sequences are rare in English. Very often, one of the vowels in such a sequence becomes reduced; we were careful to ensure that all the vowel targets in our stimuli were realized as [D]. Additionally, such sequences would usually be spoken in British English with an interpolated intervocalic glide; a glide plus vowel (as for example in the initial syllable of *Yokohama)* is actually a CV mora in Japanese. However, the matching represents the best that we

could achieve within the phonological repertoire of English.

The stimulus words with O target were *ion, eon, iodic, ionic, kiosk, neon, dioxide, geography, cockerel, horrify, torrid, dogged, atomic, illogical, aloft, abolish,* and the stimulus words with N target were *inlet, onward, incur, endear, fender, candy, conceal, confide, aniseed, enervate, enamel, anarchic, canopy, sanity, denial, canoe.* The range of frequency of occurrence was analogous to that in Experiment 1. As can also be seen, where possible half the stimulus words in each group had primary stress on the syllable containing the target, and half did not.

Exactly as in Experiment 1, a further 245 words were chosen and arranged into 64 sequences. The sequences varied in length from two to six words, and the target words occurred in second, third, fourth or fifth position. Of the 32 sequences which did not contain one of the experimental target words, half contained no occurrence of the specified target, while the other half contained dummy targets varying from the experimental targets in the same ways as in Experiment 1. The dummy target words for O were *officer, shallot, embryonic, adoptive, idiotic, occupy, padlock, follicle;* for N they were *panic, balloon, pencil, nervous, green, monastic, northerly, tender.*

Twelve practice sequences were also constructed. These too varied in length from two to six words, and four of them had no occurrence of the specified target.

The experimental and practice sequences were recorded on digital audio tape by a male native speaker of Standard Southern British English. The rate was matched to the Experiment 1 tape. A timing mark was placed on the second channel of the tape aligned roughly with the onset of each target item.

Subjects and procedure. Twenty-four Sussex University students took part in the experiment for a small payment. None had taken part in Experiment 2. The procedure was the same as that for Experiment 2, except that no special instructions with regard to language were required.

Results and Discussion

Mean number of missed responses and mean RTs were determined for each subject and each item, and separate analyses of variance were conducted on each measure with subjects and with items as random factors. The mean RTs and miss rates are shown in Table 3. Analyses of variance showed that no effect at all reached significance in the RT analysis; in the miss rate analysis there was a significant advantage for consonant (mean miss rate 4.2%) over vowel targets (11.2%; $F\backslash [1,23] = 13.8$, $p = .001$; $F2[1,24] = 7.12$, $p < .02$) and for targets in second phoneme position (mean miss rate 4.4%) over targets in third position (10.9%; $Fl[1,23] = 11.75$, $p < .005$; $F2[1,24] = 6.11, p < .03$).

It will be recalled that a stress contrast had been built into the present materials set where possible. (This was in half of the materials—the O targets in post-vocalic position such as *dioxide, kiosk* and the N targets in pre-consonantal position such as *fender, conceal.* These were the targets matched to the moraic phonemes in Experiments 1 and 2. The O targets in interconsonantal position, such as *dogged, aloft,* could not be unstressed because then they would have become reduced, i.e., changed their vowel quality, and the N targets in intervocalic position, such as *aniseed, enamel,* differed on an extra dimension as a function of stress: they were ambisyllabic when preceded by stress but not when followed by stres5. Thus neither of these types of target

TABLE 3
MEAN RESPONSE TIME (MS) AND MISS RATE AS A
FUNCTION OF "MORAIC STRUCTURE" FOR VOWEL
AND CONSONANT TARGETS, RESPECTIVELY:
EXPERIMENT 3 (ENGLISH LISTENERS,
ENGLISH WORDS)

Target phoneme	O	N
"Moraic"	698 (9.4%)	672 (4.7%)
"Nonmoraic"	683 (13%)	680 (3.6%)

lent themselves to an analysis of stress effects.) A subanalysis of the relevant half of the materials showed no significant effect of stress for the English listeners. Mean RT to targets in stressed syllables (e.g., O in *ionic, geography;* N in *inlet, fender)* was 655 ms, and mean miss rate was 6.8%; for targets in unstressed syllables (e.g., O in *ion, kiosk;* N in *incur, conceal),* the corresponding values were 716 ms and 7.3%. Neither difference reached our criterion of significance (although in RTs the effect was significant across subjects), and the stress contrast did not interact with any other factor. These findings are in line with previous failures to find significant phoneme detection differences between targets in stressed versus unstressed position with English listeners and laboratory-read speech (e.g., Mehta & Cutler, 1988).

The present results with English listeners and English resemble the results of Experiment 2 with English listeners and Japanese words; they resemble even more closely the results of Norris *et al.* (1992) with English listeners and English words containing vowel and nasal consonant targets, since both in that and in the present study only a limited advantage for nasal consonants over vowels appeared, and only in miss rates, not in RTs. Thus the results of Experiment 2 may include effects of inappropriate expectations generated by the native phonemic repertoire. We will return to this question below.

In Experiment 4 Japanese listeners were presented with English input. This constitutes a further extension of the previous investigations of the mora effect with Japanese listeners in Experiment 1 and in the study by Otake et al. (1993); in that earlier study, all materials were in Japanese, for both Japanese and nonnative listeners. The present experiment thus allows the first test of whether Japanese listeners apply their characteristic listening procedures used with input in their native language to input in a foreign language; that is, it is the first test of whether moraic segmentation by

Japanese listeners is independent of language of input, just as previous studies have shown syllabic segmentation to be for French listeners (Cutler et al., 1986; Otake et ah, 1993).

Experiment 4 also allows us to test whether the mismatch between the English and Japanese phonemic repertoire played a role in the results of Experiment 2. If English listeners listening to Japanese found O targets harder than N targets because the relevant vowel sounds in English and Japanese are (in whatever dimension) less well matched than the relevant consonant sounds, then the same should be true of Japanese listeners hearing English. In other words, if repertoire mismatch effects play a role, we would predict a vowel/consonant effect—a detection disadvantage for vowels in comparison with consonants—in this experiment, even though no such effect appeared with Japanese listeners in Experiment 1.

<center>EXPERIMENT 4</center>

Method

The materials were the same as those in Experiment 3. The subjects were 40 students at Dokkyo University who took part in the experiment for a small payment. None had taken part in Experiment 1. All had knowledge of English, having taken courses in English and spent at least 10 months in English-speaking countries. The procedure was the same as that in Experiment 1.

Results and Discussion

Mean number of missed responses and mean RTs were determined for each subject and each item, and separate analyses of variance were conducted with subjects and with items as random factors. Mean RTs and mean number of missed targets for each condition are shown in Table 4.

The most striking effect was evident in the vowel/consonant contrast and particularly in the miss rates: vowel targets (mean

TABLE 4

MEAN RESPONSE TIME (MS) AND MISS RATE AS A
FUNCTION OF "MORAIC STRUCTURE" FOR VOWEL
AND CONSONANT TARGETS, RESPECTIVELY:
EXPERIMENT 4 (JAPANESE LISTENERS,
ENGLISH WORDS)

Target phoneme	O	N
"Moraic"	743(69.1%)	699(17.8%)
"Nonmoraic"	733 (43.4%)	726 (33.8%)

miss rate 56.2%) were very much harder for these subjects to detect than consonantal targets (mean miss rate 25.8%; Fl[1,39] = 101.75, p < .001, F2[1,24] = 26.18, p < .001). Although this high miss rate for vowels renders analysis of the vowel RT data somewhat questionable, the vowel/consonant comparison in the RTs was in the same direction: vowels (mean RT 738 ms) were harder to detect than consonants (mean RT 712 ms); Fl[1,39] = 23.34, p < .001, F2[1,24] = 28.16, p < .001.

The contrast between the targets matched to moraic versus nonmoraic positions did not reach significance in either analysis. The vowel/consonant contrast interacted significantly with the "moraic/ nonmoraic" contrast: Fl(1,39) = 13.84, p < .001,F2(1,24) = 13.19, p< .005 for RT; Fl(1,39) = 60.56, p < .001, F2(1,24) = 12.18, p < .005 for miss rate. As can be seen from Table 4, the source of this interaction was an advantage of "moraic" over "nonmoraic" consonantal targets, but a difference in the opposite direction for vowels.

A further subanalysis investigated whether the Japanese listeners were sensitive to the stress contrast built into half the English materials. The mean response times to stressed versus unstressed targets were 402 and 401 ms, respectively, and the mean miss rates were 43.8 and 43.1%; neither difference was significant and the stress effect did not interact with any other factor.

Thus in this experiment the Japanese listeners showed a vowel/consonant difference, in the same direction as that shown by English listeners, even though they had

shown no such effect with their native language. This finding supports the proposal that mismatch between the input and the native language phonemic repertoire plays a role in phoneme detection in a foreign language.

Moreover, the main effect of "moraic" structure did not itself reach significance; instead, it interacted with the vowel/ consonant contrast. We examined the components of this interaction separately. For the vowels, there was no significant difference in RT to "moraic" versus "nonmoraic" targets (p > .1); miss rate, however, was significantly greater to "moraic" than to "nonmoraic" targets (t[39] = 6.25, p < .001). For the consonants, the reverse was true: "moraic" targets showed an advantage over "nonmoraic" both in response times (t[39] = 3.44, p < .001) and miss rate (t[39| = 4.35, p < Ml).[2]

The fact that the overall miss rate for vowels was greater than 50% clearly indicates that the English vowels presented severe detection difficulty to the Japanese listeners. In fact, these listeners had even greater difficulty than the English listeners of Experiment 2 had experienced with Japanese vowels (in that case, the overall miss rate for vowels was around 30%). Certainly our subjects were in no doubt about the difficulty of the vowel targets; many of them spontaneously commented on it, and a repeated comment concerned the difficulty of distinguishing British English [D] (as in Southern British English *hot, boss)* from [a] (as in Southern British English *heart, bath).*

[2] Experiment 4 was also run twice. In an initial run, subjects who had minimal experience of spoken English were chosen: they were students of economics or law and had never spent time in English-speaking countries. For these subjects the phoneme-monitoring task proved extremely difficult. A very large number of responses were missed or were too long to be recordable by the experimental setup. Only 20 of the 40 subjects even produced more than one-third of responses (12 of 32). For these 20 subjects, however, an analysis of variance showed exactly the same pattern of responses as those with the second subject group reported here as Experiment 4.

Such comments had not been made by the subjects of Experiment 2. With miss rates above 50%, there is in fact little point in interpreting RTs (as Otake et al. [1993] also argued). Nor do we feel that the difference in miss rate between "moraic" and "nonmoraic" vowels actually reflects moraic structure. Although the present English materials were matched to the Japanese materials as well as we could achieve, there is no doubt that the matching did not actually amount to a contrast in moraic structure, and as we pointed out in the introduction to Experiment 3, the matching was worse for the vowel targets than for the consonants.

In an attempt to gain further insight into subjects' perceptions of our stimulus materials, we presented the words containing vowel targets in post vocalic position *(ion, geography* etc.) to 10 new Japanese subjects (with virtually no knowledge of English) and asked them to write the words in *katakana* script *(katakana* is the mora-based script usually used to represent foreign loan words). Although all the target words containing vowel-vowel sequences *(eon, geography)* are, as mentioned above, spoken in British English with an interpolated glide, the subjects made almost no use of the character representing glide plus vowel (as in the first syllable of *Yokohama).* Instead, their preferred representation of the English glide was as an inserted vowel, and the moraic O was represented with the character for the corresponding moraic vowel in Japanese (thus *ion,* for example, was represented by all 10 subjects with four single-phoneme morae). It therefore appears doubtful that the glide and vowel were actually perceived as a single moraic unit by subjects in the detection experiment. The fact that the miss rate was even higher for "moraic" than for "nonmoraic" vowels may nevertheless indicate that perception of O was more difficult when it was preceded by a vowel/glide.

As we further pointed out above, the matching in phonological structure to the original Japanese materials is rather closer

for the consonant targets we used. Certainly, subjects did not make spontaneous comments about the difficulty of detecting N. The miss rate for consonants (even averaged over "moraic" and "nonmoraic" cases) was less than half that for vowels. Therefore the consonant targets alone offer a reasonable response-time database, and a chance to test for effects of at least roughly matched moraic structure. We believe that the results for N targets—a significant advantage for "moraic" over "nonmoraic" N in both RTs and miss rates—constitute a true indication of moraic responding by the Japanese listeners in this experiment. Note that no difference between the two types of N target appeared with the English listeners in Experiment 3. No obvious acoustic confound suggests itself as an alternative explanation. The difference is exactly as predicted by the moraic hypothesis. Just as French listeners show syllabic segmentation in both native- and foreign-language input, therefore, Japanese listeners, given the opportunity, show moraic responding irrespective of whether they are listening to their own or another language.

In our next two experiments, the language-specific effects that we have observed in the phoneme detection task in Experiments 1-4 are explored in a more varied range of contexts than those used so far. There is reason to believe that the vowel/consonant effect, at least, is not constant across word-initial versus word-medial positions (van Ooyen, 1994; van Ooyen et al., 1991; Norris et al., 1992). In the experiments of Otake et al. (1993), the target sequences which were responded to differently as a function of whether or not they corresponded to moraic structure all occurred in word-initial position. In Experiments 1-4, however, all targets occurred in word-medial position. No effects of second- versus third-phoneme position in the word were observed with Japanese listeners; however, word-initial position was not tested. According to our interpretation of the findings so far, we would predict that

the detection advantage for moraic over nonmoraic phonemes would appear in any word position. This prediction is tested in Experiment 5.

It is not possible, however, given the constraints of Japanese phonology, to vary moraic structure and initial versus medial word position *and* the vowel/consonant factor independently. There is no such thing in Japanese as a word-initial moraic consonant; as described in the introduction, moraic consonants are always syllabic codas. Nor is there such a thing as a word-initial nonmoraic vowel; any word-initial vowel is a mora by itself. Experiment 5 was designed to test two of these factors independently: moraic structure and position in the word.[3] Targets therefore included initial moraic phonemes, noninitial moraic phonemes, initial nonmoraic phonemes and noninitial nonmoraic phonemes. By definition, these comparisons were therefore confounded with the vowel-consonant factor: the initial moraic phonemes were always vowels and the initial nonmoraic phonemes were always consonants. To even up the distribution of vowels and consonants across these factors, the noninitial moraic phonemes were always consonants, and the noninitial nonmoraic phonemes were always vowels. Since Experiment 1 showed no sign of a vowel/consonant effect for Japanese subjects listening to Japanese, this feature of the present experimental design should not introduce confounding in the results.

EXPERIMENT 5

Method

Materials. Thirty-two meaningful Japanese words were chosen as stimulus words, eight each with target phonemes which we will refer to as A, O, N, and K. The O and N targets were as in Experiments 1-2; A represents a low central unrounded vowel,

[a], and K a voiceless velar plosive, [k]. Within each set of eight, four targets were single-phoneme morae and four were part of CV morae. For the vowels, the moraic targets occurred in word-initial position while the nonmoraic targets occurred as the fourth phoneme in the word; for the consonants, the nonmoraic targets occurred in word-initial position while the moraic targets occurred as the third phoneme in the word. In Roman transcription the stimulus words were A: *akui, aizu, asobi, aite, gohan, sekai, kurashi, kotai;* O: *okuba, ochiba, omake, okazu, kamotsu, midori, kamoku, kimochi;* N: *kinjyo, genshi, kinshi, danchi, namida, naifu, namae, naisho;* K: *rokku, dekki, sikki, sakka, kaigi, kamera, kaeri, kazari* (in each case the first four examples are moraic; the next four nonmoraic)

A set of filler words was chosen and experimental and practice sequences were constructed exactly as for the previous experiments. The experimental and practice sequences were recorded on digital audiotape by a male native speaker of Standard Tokyo Japanese. Each sequence was preceded by its number. Rate of speech was again as in the previous experiments.

Subjects and procedure. Twenty undergraduate members of Dokkyo University took part in the experiment for a small payment. None had been subjects in Experiments 1 or 4. The procedure was as in Experiment 1. A question arises with the measurement of stop consonant targets such as /k/, namely whether response time should be measured from the onset of closure or from the onset of the release burst. In word-initial position only the burst comes into consideration, since when words are spoken in isolation the onset of a closure is indistinguishable from preceding silence. In word-medial position, however, one has the choice. Given that the geminate consonant is signalled by closure duration, it could be argued that it is more appropriate to measure from closure onset; on the other hand, there is strong reason to doubt that the place of articulation of the stop (and

---

[3] In fact, Experiment 5 was designed before the results of van Ooyen (1994) and her colleagues had established the vowel/consonant effect in phoneme detection by English listeners.

hence, the presence versus absence of a match to the specified target) can be unambiguously determined prior to the release burst, since the primary cues to place are in the spectrum of the release burst and the formant transitions into the following vowel (Dorman, Studdert-Kennedy, & Raphael, 1977). In particular, although for stops in intervocalic position there is a small amount of information in the preclosure transition, the strongest cue is in the transition into the following vowel (Dorman & Raphael, 1980). Moreover, measuring from the burst preserves compatibility with the word-initial measurements. However, for completeness we measured both points and conducted RT analyses on both sets of measurements.

Results and Discussion

Mean number of missed responses and mean RT were determined for each subject and each item, and separate analyses of variance were conducted on each measure with subjects and with items as random factors and with position in the word and moraic structure as independent variables. Table 5 presents the mean RTs and miss rates. RTs for moraic K are timed from the burst onset.

The miss rate analysis is independent of measurement considerations. Analyses of variance showed that the only significant effect in the miss rates was one of target position ($Fl[1,19] = 20.12$, $p < .001$, $F2[1,24] = 21.84$, $p < .001$): targets were detected more accurately in the word-initial than in the word-medial position.

In the RT analyses, there was again a significant advantage for initial over medial targets ($Fl[1,19] = 7.84$, $p < .02$; $F2[1,24] = 7.88$, $p < .01$), as well as a significant overall advantage for moraic over nonmoraic targets ($Fl[1,19] = 72.82$, $p < .001$, $F2[1,24] = 37.3 1$, $p<.001$). There was also an interaction between the two factors: $Fl[1,19] = 12.55$, $p<.01$, $F2[1,24] = 4.86$, $p < .04$). This interaction presumably reflects the fact that the moraic effect was much larger in medial position; however, it was also significant in initial position (t [19] $= 6.52$, $p < .001$).

Thus Experiment 5 has demonstrated a new effect in Japanese phoneme-monitoring: targets in word-initial position are detected faster than targets in word-medial position. The effect of moraic structure, however, remains significant in initial position, as we predicted. Experiment 5 also differed from the preceding experiments in including four rather than two phoneme targets. The t tests (one-tailed) revealed that the predicted mora effect was significant for all four phonemes: $t[19] = 5.54$, $p < .001$ for A; $t[19] = 4.44$, $p < .001$ for O; $/[19] = 1.9$, $p < .04$ for N; $t[19] = 2.71$, $p < .01$ for K.

The above response time analyses are based on measurements for moraic K taken from the onset of the release burst. If responses are, instead, measured from the closure onset, the mean RT to this target rises from 445 to 714 ms, longer than to all other moraic or nonmoraic targets. This in itself suggests that the closure did not in fact convey the acoustic information necessary for identification of a velar target. However, the main effect of mora structure remained significant when the analysis was repeated with responses to moraic K timed from closure onset: $Fl[1,19] = 18.39$, $p < .001$, $F2[1,24] = 9.57$, $p < .01$.

In Experiment 6, the materials of Experiment 5 were presented to English listeners with no knowledge of Japanese. For English listeners, the informative value of these materials is of course different. Experiments 2 and 3 showed that English lis-

TABLE 5
MEAN RESPONSE TIME (MS) AND MISS RATE AS A FUNCTION OF MORAIC STRUCTURE AND POSITION IN THE WORD: EXPERIMENT 5 (JAPANESE LISTENERS, JAPANESE WORDS)

Position	Initial	Medial
Moraic	534(1.3%)	550(10%)
Nonmoraic	618 (.6%)	737 (14.4%)

teners performing phoneme detection exhibit no effect of mora structure either in Japanese or in English. Therefore there is little point in a further demonstration of this null effect with new materials. English listeners do, however, in general show a detection advantage for consonants over vowels. The materials of Experiment 5 include two additional phoneme targets, A and K, both of which, as described by Maddieson (1984), are close to equivalent sounds in English, as described by Ladefoged (1982) and Wells (1982); at least, they are as close as the N targets are. Thus Experiment 6 will allow further examination of the effects of phoneme repertoire mismatch: if the disadvantage for O in comparison to N targets shown by English listeners in Experiment 2 was solely due to phoneme repertoire mismatch, then we should observe a difference between A and O targets in the present experiment (with A targets patterning similarly to N and K targets). On the other hand, if some part of the vowel/consonant effect observed in Experiment 2 represents the general difficulty experienced by English listeners with vowel versus consonant targets, appearing with foreign as well as native language input, then we should expect to observe difficulty for both vowel targets.

Moreover, the word position factor will prove informative. Previous results for English suggest that in general the vowel/consonant effect is stronger in word-medial than in word-initial position when target words are presented in isolation (van Ooyen, 1994). The proposed explanation for this is that variability of vowel realization is perceived by listeners to be greater when (as for medial vowels) there is preceding phonemic context than when (as for initial vowels) the preceding context is silence. In Experiment 6 we can also test whether such an interaction of the vowel/consonant difference with position in the word appears with foreign as well as with native input.

Therefore the moraic factor was ignored

in analyzing the results of these materials for English listeners; RTs and miss rates were compared for the two factors of vowel versus consonant targets and position in the word. The inevitable confounding with moraic structure (initial vowels and medial consonants were always moraic, medial vowels and initial consonants were always nonmoraic) should, given the consistent failure to find moraic effects with English listeners, be irrelevant.

EXPERIMENT 6

Method

The materials were the same as those in Experiment 5. However, as for Experiment 2 the tape was copied and the numerals at the start of each sequence were removed. A timing mark was placed on the second channel of the tape aligned roughly with the onset of each target item. The subjects were 23 members of the Applied Psychology Unit subject panel who took part in the experiment for a small payment. None had any knowledge of Japanese. The data for three subjects who missed more than 50% of all targets were not analyzed. Although this experiment was conducted at the Applied Psychology Unit rather than at Sussex University, the procedure (including the portable microcomputer for experimental control) was exactly as for Experiment 2.

Results and Discussion

Mean number of missed responses and mean RT were determined for each subject and each item, and separate analyses of variance were conducted with subjects and with items as random factors and with position in the word and vowel versus consonant targets as independent variables. Mean RT and mean number of missed targets for each condition are shown in Table 6.

The miss rates showed significant main effects of both the vowel/consonant factor (vowels were missed significantly more often than consonants: $F1[\ 1,19] = 17.01$, $p<$

TABLE 6

MEAN RESPONSE TIME (MS) AND MISS RATE AS A FUNCTION OF POSITION IN THE WORD FOR VOWEL AND CONSONANT TARGETS SEPARATELY: EXPERIMENT 6 (ENGLISH LISTENERS, JAPANESE WORDS)

Position	Initial	Medial
Vowels	655 (13.8%)	882 (38.7%)
Consonants	684 (3.8%)	655 (20%)

.001; $F2[1,24] = 10.31$, $p < .005$) and of position (word-initial targets were detected more accurately than word-medial; $Fl[1,19] = 17.09$, $p < .001$; $F2[1,24] = 21.21$, $p < .001$). The two effects did not interact.

In the response time analyses (with K responses timed from the burst), there was again a significant overall advantage for consonant over vowel targets ($Fl[1,19] = 22.23$, $p < .001$; $F2[1,24] = 8.77$, $p < .02$). The effect of word position was nonsignificant, but it did interact with the vowel/consonant factor $CFl[1,19] = 54.73$, $p < .001$; $F2[1,24] = 12.39$, $p < .005$). The source of this interaction is clear from Table 6: consonant targets were detected more rapidly than vowel targets in medial position but not in initial position. The t tests showed that there was no significant difference between initial vowels and consonants, but significantly faster responses to medial consonants than to medial vowels ($/[19] = 7.16$, $p < .001$). (Note that if K responses are timed from the onset of closure, however, the mean RT to medial consonants rises from 655 to 803 ms; the overall advantage for consonant targets disappears, and the main effect of word position becomes significant. The interaction between the two factors, however, remains significant.)

A subanalysis was conducted to assess whether there were detectability differences between the two vowel targets. (The problems with measuring RT to K make it difficult to compare the consonants, at least

on RT; the miss rates for N, 14.4%, and for K, 9.4%, were not significantly different.) Responses to A and O were indeed significantly different, but not in the predicted direction: A (mean miss rate, 36.9%; mean RT, 781 ms) was harder to detect than O (mean miss rate, 15.6%; mean RT, 680 ms) on both measures (miss rate: $Fl[1,19] = 27.32$, $p < .001$, $F2[1,12] = 6.54$, $p < .03$; RT: $F1[1,19] = 15.63$, $p < .001$, $F2[1,12] = 5.92$, $p < .04$). This result casts at least some doubt on the explanation of the results of Experiment 2 in terms of phoneme repertoire mismatch alone. (As a check, the results of Experiment 5 for the two vowels alone were also analyzed; detection of A was in fact poorer than detection of O in that study also, but the differences were not as great and failed to reach our criterion of significance either in miss rates or RT.) The results are consistent with at least some portion of English listeners' difficulty with medial vowel versus consonant targets in Japanese input being due to a more general effect of phonemic repertoire, that which causes difficulty in the native language as well. We will return to this question in the general discussion.

Experiments 5 and 6 have thus usefully extended the findings from Experiments 1-4. The moraic effect which Japanese listeners show in phoneme detection in their native language appears in word-initial as well as in word-medial position and with a variety of phoneme targets. The vowel/consonant differences characteristic of English listeners' phoneme detection appear again in Japanese; moreover, these effects are dependent upon position in the word in input presented in a foreign language in the same way as they are in input presented in the native language.

GENERAL DISCUSSION

Two major conclusions emerge from this study. One is that the importance of moraic structure for prelexical processing of Japanese by Japanese listeners is reconfirmed.

For these listeners, morae form the natural basis of segmentation; single-phoneme morae function equivalently to CV morae in this respect.

The second conclusion concerns the extent of language-specific listening. Just as French listeners apply syllabic segmentation to speech input not only in their native language but also in English and in Japanese, so, the present study tells us, do Japanese listeners apply moraic segmentation to foreign-language input, as far as the structure of the input allows it. Moraic segmentation is a language-specific effect in exactly the same way that French listeners' syllabic segmentation and English listeners' stress-based segmentation are language-specific. Japanese listeners show moraic segmentation; non-Japanese listeners presented with exactly the same input do not show it. Japanese listeners apply moraic segmentation not only to input in their native language, but to any input in which the opportunity for moraic segmentation presents itself.

Similarly, English listeners, who find the phoneme detection task harder for medial vowels than for consonants in their native language, also display a detection disadvantage for medial vowels in foreign-language input. In part, this may be due to the fact that the phonemic repertoires of Japanese and English provide closer matches for the nasal consonant target we used in Experiments 1-4 than for the vowel target of those experiments; the Japanese listeners in Experiment 4 found the English vowel targets harder to detect than the English consonant targets, just as the English listeners found the Japanese vowel targets harder than the Japanese consonants. But we feel that this is not the whole story. First, it does not explain why, in Experiment 6, both Japanese medial vowel targets proved more difficult for English listeners to detect than both medial consonant targets, although the A vowel target should have been no more mismatched to the English phonemic reper-

toire than the consonant targets. But second and more importantly, it does not explain why English listeners should consistently show a detection disadvantage for medial vowel over consonant targets, either in response times or miss rates or both, in their *native* language, while neither Japanese listeners, as we have observed, nor Spanish listeners, as van Ooyen and Sanchez-Casas [1993] reported, show such an effect. In one of the experiments reported by Cutler et al. (1994), English listeners also failed to profit at all from restriction of the vowel inventory within the experimental context; that is, medial vowel detection was unaffected by whether the experimental materials as a whole contained any and all the vowels of English, or were restricted to five highly distinct vowels.

We believe that English listeners' difficulty with medial vowel targets in the phoneme: detection task is indeed an effect of phonemic repertoire, but one that manifests itself also in native language processing: vowel detection, especially in word-medial position where coarticulatory effects are strongest, is difficult for speakers of any language with a densely populated vowel repertoire. English is such a language; so is, for example, French, so that we would predict that French listeners would also find detection of vowel targets more difficult than detection of consonant targets. (Some support for this prediction is available from results in the literature: the response times for medial vowel detection reported by Mehler et al. [1981] are around 150% of the response times for unprimed medial stop consonant detection reported by Frauenfelder and Segui [1989] with a similar French subject population.) In languages like English and French, the task of matching an input to a prespecified vowel target is one for which subjects set their response criterion very high; their slower responses and higher miss rates result from this caution. The high response criterion is specific

to the phoneme detection task. Van Ooyen (1994) has argued, on the basis of findings from a different experimental task, that English listeners treat vowels in normal speech input as unreliable, indeed mutable, sources of phonemic information. In the phoneme detection task, in which listeners are required to decide that the input exactly matches the target specification, mutability is highly undesirable; a high response criterion in the detection task, particularly with respect to detection in medial position, is the listeners' solution to this problem.

On the other hand, in languages like Japanese and Spanish, with just a few highly distinct vowels, vowel information is reliable in normal speech processing, and no need arises for a high response criterion in the detection task, in any position in the word. That is, we believe that the vowel/consonant effect in phoneme detection is another language-specific processing effect, albeit one that is confined to a particular experimental situation and may not necessarily play a large role in normal listening. The present results suggest that listeners who exhibit such an effect with input in their native language will, as with the other language-specific effects we have discussed, also exhibit it with input in a foreign language.

Finally, a third effect which manifested itself in the present results was an effect of position of the phoneme target in the word. Targets in word-initial position were, overall, easier to detect than targets in word-medial position. This effect appeared both with Japanese and with English listeners in Experiments 5 and 6. In principle, position effects allow for explanation at different levels. On the one hand, there is an obvious explanation in terms of a low-level perceptual effect, namely that any phoneme target preceded by silence is easier to detect than the same phoneme target preceded by speech. On the other hand, an advantage for word-initial targets could also appear because word onsets are particularly important for lexical access (as argued, for instance, by Cutler, Hawkins and Gilligan [1975]).[4] These competing explanations are of course easily testable, since the low-level explanation predicts that the word position effect should only appear with input consisting, as here, of lists of words; it should not appear when the input consists of continuous speech. Likewise, this explanation predicts that the effect should appear with nonwords as well as with real words. The present results seem to us to support a low-level interpretation, since the position effect appeared when English listeners were presented with Japanese input, i.e. effectively with nonwords. However, a test of the effects of target position in the word using continuous speech input would be more decisive.

In conclusion, then, the present study has confirmed and extended previous evidence of language-specific response patterns in prelexical processing. Moreover, it has established that French listeners are not unique in applying their language-specific syllabic segmentation procedures (perhaps inappropriately) to input in other languages; Japanese listeners likewise apply their moraic pattern of processing to input which is in a language other than Japanese, and English listeners experience difficulty detecting vowels even when the input is in a language with a small and highly distinct vowel repertoire. We believe that this finding has potentially important implications for understanding the processes of acquisition of a second language. In this series of experiments listeners have been shown to engage in language-specific processing where it is, in essence, inappropriate: the procedures appropriate for the native language do not lend themselves well to the foreign language input to which they are here applied. It must be considered to what extent such behaviors persist in listeners attempting to acquire, as adults, a sec-

[4] This explanation was suggested to us by Kari Suomi.

ond language which differs from the native language in the pertinent respects.

Here it is particularly relevant to consider studies of bilingual listeners carried out by Cutler, Mehler, Norris, and Segui (1992). Cutler et al.'s French-English bilinguals were highly efficient, indeed performing to all appearances at native competence, in both languages. Nevertheless their performance in Cutler et al.'s experiments showed that they commanded only one rhythmically based segmentation procedure: either syllabic segmentation, as characteristic of French listeners, or stress-based segmentation, as characteristic of English listeners, but not both. Cutler et al. hypothesized that segmentation procedures might arise, not through accumulation of experience by mature language users, but from the prelinguistic infant's attempts to solve the initial segmentation problem posed by continuous speech input in the absence of an existing lexicon. In order to begin to construct a lexicon, infants have to find words in the input without any prior knowledge of what the words might be like. The rhythmic structure of the language to which they are exposed may be sufficiently salient that it enables them to construct a viable segmentation procedure. Solving the initial segmentation problem may, then, only need to be done once; this in turn would have as a consequence the persistence of only one segmentation procedure into mature language use.

Such an explanation may seem to offer gloomy news for second-language learners; acquiring more than one language-specific segmentation procedure may be impossible. However, there are other aspects of Cutler et al.'s (1992) study which lead to a more optimistic conclusion. First, the subjects in Cutler et al.'s study were not apparently inhibited by the availability of only one segmentation procedure; whichever procedure they commanded, they were all performing at apparently native levels of production and comprehension competence in both their languages. Thus rhythm-

based segmentation procedures, though apparently developed by all listeners as part of the normal process of acquiring a (first) language, do not appear to be strictly necessary for competent language use. Cutler et al. (1992) argued that rhythm-based segmentation procedures are, instead, heuristic devices which listeners exploit because they render speech segmentation more efficient. In the absence of such shortcuts, segmentation of continuous speech is still possible, and indeed may in principle be achieved by a variety of methods such as competition between candidate words, sequential recognition of known words, or intermediate computation of a prelexical representation (for further discussion of this range of possibilities, and for evidence that spoken word recognition by English listeners involves both rhythm-based segmentation *and* interword competition, see McQueen, Norris, & Cutler, 1994).

Second, the bilingual listeners were an exception to the pattern of results shown in the present study and in the studies of French monolingual listeners by Cutler et al. (1986) and Otake et al. (1993), in that the segmentation procedure which they commanded was not applied to input in the language for which it was inappropriate. That is, those bilinguals who showed syllabic segmentation when they were listening to French did not show it when they were listening to English—although in the studies with monolingual listeners, syllabic segmentation was applied both to English and to Japanese input. Thus the French-dominant bilinguals' competence in English had enabled them, presumably through experience with the inefficiency of syllabic segmentation when applied to English, to inhibit this procedure in one language while continuing to use it in the other. Note that monolinguals could not achieve such inhibition; as Cutler et al. (1992) pointed out, inefficiency of a segmentation procedure will only be experienced if it interferes with successful word recognition, and for monolinguals, who do not know the words of the

foreign language, there is no recognition process to be interfered with.

Thus it seems highly likely that adult second-language learners should, with increasing competence in the new language, be able to inhibit the monolingual-like application of native-language rhythm-based heuristic segmentation procedures to nonnative input, and rely instead on the alternative means of achieving segmentation without recourse to potential shortcuts. Indeed, recent studies have shown that syllabic segmentation of English does not appear with Spanish native speakers who speak good English but learned English as adults, even under the task conditions which produce syllabic effects in Spanish (Bradley et al., 1993), or with French native speakers in the process of learning English (Kearns, 1994). The question of the degree of perceptual competence necessary for abandonment of inappropriate segmentation procedures is of course as yet unexplored. A further topic for potentially fruitful research is the effect of explicit attention to lexical segmentation and to the relationship between speech rhythm and word boundaries in teaching the comprehension of a second language.

REFERENCES

BEST, C. T., MCROBERTS, G. W., & SITHOLE, N. M. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance,* 14, 345-360.

BOHN, O.-S., & FLEGE, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics,* 11, 303-328.

BRADLEY, D. C, SANCHEZ-CASAS, R. M., & GARCIA-ALBEA, J. E. (1993). The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes,* 8, 197-233.

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

CUTLER, A., & CARTER, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language,* 2, 133-142.

CUTLER, A., HAWKINS, J. A., & GILLIGAN, G. (1985). The suffixing preference: A processing explanation. *Linguistics,* 23, 723-758.

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., MEHLER, J., NORRIS D. G., & SEGUI, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology,* 24, 381—410.

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., VAN OOYEN, B., NORRIS, D. G., & SANCHEZ-CASAS, R. (1994). *Speeded detection of vowels: A cross-linguistic study.* Submitted for publication.

DORMAN, M. F., & RAPHAEL, L. J. (1980). Distribution of acoustic cues for stop consonant place of articulation in VCV syllables. *Journal of the Acoustical Society of America,* 67, 1333-1335.

DORMAN, M. F., STUDDERT-KENNEDY, M., & RAPHAEL, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics,* 22, 109-122.

FRAUENFELDER, U. H., & SEGUI, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition,* 17, 134-140.

GOTO, H. (1971). Auditory perception by normal Japanese adults of the sounds "1" and "r." *Neuropsychologia,* 9, 317-323.

KEARNS, R. K. (1994). *Prelexical speech processing in mono- & bilinguals.* Ph.D. thesis, University of Cambridge.

LADEFOGED, P. (1982). *A course in phonetics* (2nd ed.). New York: Harcourt Brace.

MADDIESON, I. (1984). *Patterns of sounds.* Cambridge: Cambridge University Press.

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

MEHLER, J., DOMMERGUES, J.-Y., FRAUENFELDER, U., & SEGUI, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior,* 20, 298-305.

MEHTA, G., & CUTLER, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech,* 31, 135-156.

NORRIS, D. G. (1984). A computer-based programmable tachistoscope for non-programmers. *Behavior*

Research Methods, Instrumentation & Computers, 16, 25-27.

NORRIS, D. G., VAN OOYEN, B., & CUTLER, A. (1992). Speeded detection of vowels and steady-state consonants. *Proceedings of the 2nd International Conference on Spoken Language Processing,* (Vol. 2, pp. 1055-1058). Banff, Canada.

VAN OOYEN, B. (1994). *Processing of vowels and consonants.* Ph.D. Dissertation, University of Leiden.

VAN OOYEN, B., CUTLER, A., & NORRIS, D. G. (1991). Detection times for vowels versus consonants. *Proceeding of EUROSPEECH 91* (Vol. 3, pp. 1451-1454). Genoa.

VAN OOYEN, B., & SANCHEZ-CASAS, R. (1993). *A cross-linguistic difference in phoneme detection.* Paper presented to the Experimental Psychology Society, Cambridge, UK.

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

OTAKE, T., HATANO, G., CUTLER, A., & MEHLER, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language,* 32, 253-278.

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

SEBASTIAN-GALLES, N., DUPOUX, E., SEGUI, J., & MEHLER, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language,* 31, 18-32.

SEGUL J., FRAUENFELDER, U., & MEHLER, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology,* 72, 471-477.

WELLS, J. C. (1982). *Accents of English. 2. The British Isles.* Cambridge: Cambridge University Press.

WERKER, J. F., & TEES, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America,* 75, 1866-1878.