The representation of Japanese moraic nasals

Takashi Otake and Kiyoko Yoneyama
Dokkyo University, Soka, Japan

Anne Cutler[a] and Arie van der Lugt
Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 21 November 1995; revised 25 June 1996; accepted 23 July 1996)

Nasal consonants in syllabic coda position in Japanese assimilate to the place of articulation of a following consonant. The resulting forms may be perceived as different realizations of a single underlying unit, and indeed the kana orthographies represent them with a single character. In the present study, Japanese listeners' response time to detect nasal consonants was measured. Nasals in coda position, i.e., moraic nasals, were detected faster and more accurately than nonmoraic nasals, as reported in previous studies. The place of articulation with which moraic nasals were realized affected neither response time nor accuracy. Non-native subjects who knew no Japanese, given the same materials with the same instructions, simply failed to respond to moraic nasals which were realized bilabially. When the nasals were cross-spliced across place of articulation contexts the Japanese listeners still showed no significant place of articulation effects, although responses were faster and more accurate to unspliced than to cross-spliced nasals. When asked to detect the phoneme following the (cross-spliced) moraic nasal, Japanese listeners showed effects of mismatch between nasal and context, but non-native listeners did not. Together, these results suggest that Japanese listeners are capable of very rapid abstraction from phonetic realization to a unitary representation of moraic nasals; but they can also use the phonetic realization of a moraic nasal effectively to obtain anticipatory information about following phonemes.

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

INTRODUCTION

Language-specific phonological structure influences the way in which listeners process speech. This is most obvious in the difficulty of hearing phonetic contrasts in foreign languages if the contrasts are not present in the native language. But aspects of phonological structure also differ in their salience to listeners as a result of the way languages are structured. Thus, although the syllable is a construct indispensable for universal phonological theory, syllables as structural units are more salient to some language users than to others. Speakers of French, Spanish, and Catalan find it relatively easy to extract syllabic structure from speech input in on-line listening tasks (Mehler *et al.*, 1981; Sebastián-Gallés *et al.*, 1992; Bradley *et al.*, 1993), but speakers of English and Japanese do not (Cutler *et al.*, 1986; Bradley *et al.*, 1993; Otake *et al.*, 1993). On the other hand, speakers of English are sensitive to foot structure (Cutler and Norris, 1988; Cutler and Butterfield, 1992) and speakers of Japanese are sensitive to mora structure (Otake *et al.*, 1993; Cutler and Otake, 1994).

The present investigation concerns Japanese listeners' sensitivity to mora structure. The mora is a component of syllable structure; essentially, light open syllables consist of one mora, while syllables with complex vowels or a coda consist of two. Japanese has a very restricted phonological inventory, which includes only five types of mora: CV, CCV, V, nasal coda (which we can represent as N), and geminate consonant (which we can represent as Q). The mora structure of a name such as *Asahi* is V-CV-CV, *Kirin* CV-CV-N, *Sapporo* CV-Q-CV-CV, *Suntory* (with a long final vowel) CV-N-CV-CV-V.

The mora is the basis of traditional verse forms in Japanese (a haiku, for example, consists of 17 morae in groups of 5, 7, 5). It also plays a central role in Japanese orthography. Japanese has a mixed orthographic system: most content words are written in *kanji* (Chinese characters), but function words and inflectional affixes are written in *hiragana*, and loan words from foreign languages are written in *katakana*; both these latter (*kana*) orthographies directly represent morae.

A recent series of studies has revealed that Japanese listeners can exploit the moraic structure of words in on-line listening tasks. Otake *et al.* (1993) presented Japanese listeners with lists of naturally spoken words and asked them to press a response key as soon as they detected a word beginning with, for example, the sequence TA. (The target was specified either as Roman characters or as a spoken string, never in Japanese orthography.) Listeners responded equally rapidly to such a target in *tanishi* and in *tanshi*, both of which begin with the mora *ta*-. The target TAN, on the other hand, could be detected in *tanshi* (where the mora structure is CV-N-CV: *ta-n-shi*, so that the target corresponds exactly to two morae) but not in *tanishi* (where the mora structure is CV-CV-CV: *ta-ni-shi*, so that the target corresponds to the whole of one mora and part of another). In other words, listeners did not detect occurrences of targets which did not respect mora boundaries. Similarly, Cutler and Otake (1994)

[a]Electronic mail: anne.cutler@mpi.nl

found that Japanese listeners detected phoneme targets which were in themselves moraic more rapidly than targets which formed part of a CV mora; thus /n/ was detected more rapidly in *inka* and *kanko* than in *inori* or *kanojo*, /o/ was detected more rapidly in *aokabi* and *taoru* than in *kokage* and *etoku*.

In both studies non-native listeners produced a different result with the Japanese materials. Otake *et al.* found that French listeners and English listeners, unlike the native Japanese listeners, did respond to the target TAN in *tanishi*; they were apparently undisturbed by the absence of correlation to mora boundaries. The French listeners, however, further showed a characteristic syllable-based response pattern by finding TA more difficult to detect in the closed first syllable of *tanshi* than as the open first syllable of *tanishi*; this, as we saw, was a pattern which the Japanese listeners did not show. Cutler and Otake found that English listeners did not respond faster to the moraic than to the nonmoraic phoneme targets in their Japanese materials; what these listeners did show was faster response times to consonant than to vowel targets, which is consistent with other evidence from English that vowel targets are difficult to detect (Cutler *et al.*, 1996; van Ooijen, 1994), and, again, is a pattern which Japanese listeners did not show.

Moreover, Cutler and Otake observed that Japanese listeners even respond to English nasals in syllabic coda position as if they were equivalent to moraic nasals in Japanese: /n/ targets were detected more rapidly in English words like *inlet*, *endear*, *concur*, *fender* than in words like *aniseed*, *enamel*, *sanity*, *canoe*. Again, no difference between detection times for /n/ targets in these two word types was observed with native English listeners.

The moraic nasal as phoneme target in a detection task presents an interesting challenge. How abstract is listeners' representation of such a target? In one sense the moraic nasal in Japanese is unquestionably a unitary abstract entity to literate language users; the orthography abstracts away from the phonetic realization of moraic nasals in that there is only one nasal coda *kana* character. But nevertheless the realization of the moraic nasal in speech is phonetically highly variable: there is complete homorganic assimilation (Vance, 1987). Thus a nasal coda would be realized as IPA [n] before an alveolar consonant, it would be realized as [m] before a bilabial, as [n̪] before a dental and as [ŋ] before a velar. Japanese contains the bilabial stop consonants [b] and [p], the velar stops [g] and [k], lamino-dental stops [t] and [d], and [r] which is realized as an alveolar tap (Vance, 1987).[1] We would expect that Japanese listeners would have no problem carrying out a task in which they were required to respond to ''the'' moraic nasal irrespective of how it was realized. Otake and Yoneyama (1995) report a cross-splicing experiment in which listeners' transcriptions of words containing nasals respected the syllabic position of the nasal (syllable-final, i.e., moraic, versus syllable-initial, i.e., nonmoraic) but not its original phonetic realization. Nonetheless, listeners might show effects of the considerable degree of allophonic variation, if we chose a task that was sensitive enough to pick up whether some realizations were in some sense more canonical than others.

Reaction time (RT) methodology in general, and the phoneme detection task in particular, can be sensitive enough to address this issue. RT in phonetic decision (Whalen, 1984), repetition (Whalen, 1991) and lexical decision (Streeter and Nigro, 1979; Whalen, 1991; Marslen-Wilson and Warren, 1994) tasks is sensitive to the goodness of fit between a phoneme and its immediate context; minor mismatches lead to slower RTs. The same effect has been demonstrated in phoneme detection (Martin and Bunnell, 1981, 1982); phoneme detection is also slowed by minor temporal adjustments to the speech signal preceding the target sound (Meltzer *et al.*, 1976; Martin, 1979; Martin *et al.*, 1980; Buxton, 1983). Further, phoneme detection is speeded if the identity of a following phoneme is fully predictable (Mills, 1980; Swinney and Prather, 1980). No phoneme detection study has directly addressed the issue of canonicality of phoneme realization, but relevant evidence comes from a study by Cutler *et al.* (1987) of the effects of how the phoneme target is specified. Typically, targets will be modeled via examples (/b/ as in *boy*), although the experimental words in which the targets are to be detected may match the model to varying extents. Cutler *et al.* found that match versus mismatch between model and experimental words in the immediate phonetic context of the target phoneme had undetectable effects on responses: subjects were just as fast and as accurate to detect a target modelled as ''/b/ as in *Ben*'' in *benches*, *beggar*, and *billiards*. More abstract phonological structure did, however, have an effect; subjects given targets modeled as belonging to a cluster (/b/ as in *blue*) were faster to detect targets in words with such a cluster onset (*blot*, *blank*) than in words with a singleton onset (*best*, *bind*), while with the same materials the pattern reversed for subjects given singleton models (/b/ as in *bowl*). This result, and the effects of contextual predictability, suggest that phoneme detection RT should show effects of subjects' expectations about how moraic nasals are realized, if such expectations exist. In experiment 1, therefore, we asked Japanese listeners to detect moraic nasals in different phonetic contexts. We predict that these listeners will be able to perform the task with ease—that is, miss rates will be low; and we predict that they will show the usual ''mora effect,'' i.e., they will be faster and more accurate detecting these moraic targets than nonmoraic other targets. But our principal initial question is whether their RTs to the moraic targets will differ across contexts, i.e., whether phonetically different realizations will be responded to equally rapidly and accurately as instances of the specified target, or whether some realizations will be responded to more rapidly and/or accurately than others. To this end we also chose to present the target specification not as the *kana* character for a moraic nasal (which might have maximized listeners' tendency to abstract away from the phonetically variable realizations), but rather as the Roman character N. The major motivation for this choice was in fact to achieve comparability with control experiments with non-Japanese listeners. But previous work has shown that Japanese undergraduates have no difficulty with this form of target presentation in a phoneme detection task (Cutler and Otake, 1994); they are highly familiar with the Roman alphabet, and can readily propose Roman transliterations of Japa-

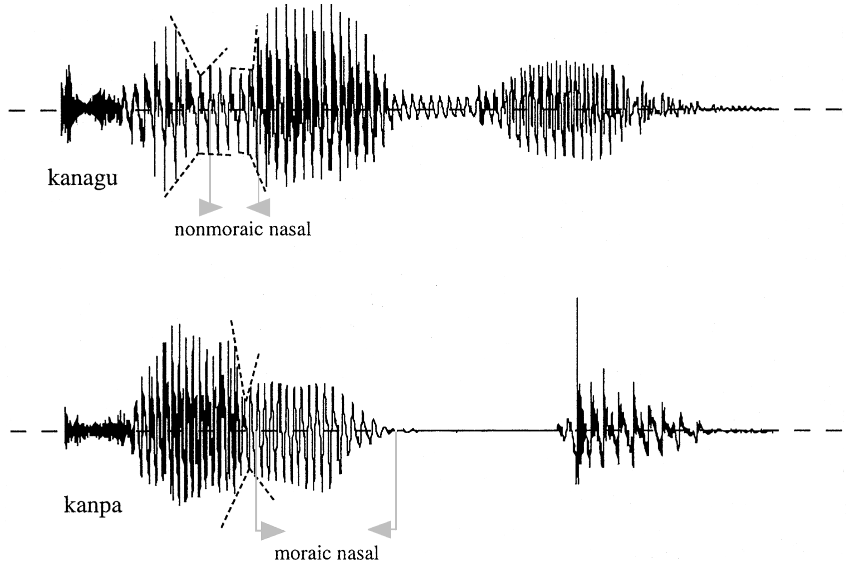

FIG. 1. Waveform of stimulus words *kanagu* and *kanpa*, showing onset and offset decisions for nasal duration measurement.

nese words. Thus we could in effect maximize the chance of finding effects of context if there are any.

I. EXPERIMENT 1

A. Materials

Thirty-two Japanese words were chosen as target-bearing items. Sixteen of these contained moraic nasals, a further 16 contained one of the four CV morae *na, ni, no, nu*. All the words were trimoraic, with the nasal target sound occurring as or in the medial mora. The 16 words with moraic nasal targets had the structure CVCCV, and the 16 words with nonmoraic targets the structure CVCVCV. Thus the target sound was always the third phoneme of the word.

The 16 words with moraic nasal targets were: *kanpa, denpa, kinba, tonbo, bando, tento, shindo, konto, danro, kenri, kinri, konro, tanka, denki, ringo, rongo*. As can be seen, the first four precede bilabial stops, the next four precede dental stops, the following four precede the alveolar tap [r], and the last four precede velar stops; the nasal assimilates to each place of articulation (thus the IPA transcription of the nasal would be [m], [n̪], [n], and [ŋ], respectively). Within each such set of four words, one word has each of the four vowels a, e, i, o preceding the target.

The 16 words with nonmoraic nasal targets were: *kanagu, senaka, tonari, sunaba, tanishi, tenisu, hiniku, yonige, senobi, hinode, kinoko, sunoko, tanuki, tenuki, jinushi, hunuke*. Four words contained each of the morae *na, ni, no, nu*. The preceding vowel varied across each set of four.

Sixteen further dummy target items were chosen, in which a nasal sound occurred in a position other than in the experimental items (i.e., with one exception, not as third phoneme in a trimoraic word); these were: *naifu, nira, nidai, nedoko, mehana, wani, sune, guinomi, enpitsu, danchi, ondo, bijinga, ken, imon, chikusan, tsuushin*. The further words *sanso, kengaku, kazan, rokuon, namida, nekku, inishie* and *anago* served as practice items. Finally, an additional 245 words containing no alveolar, dental, or velar nasals served as filler items. (Some filler words contained bilabial nasals, for reasons connected with experiment 2b, below.) The words were arranged into 12 practice sequences, of which eight contained an occurrence of the target and four did not, and 64 sequences in the experiment proper, of which 48 contained a target and 16 contained no occurrence of the target. These sequences were recorded onto DAT tape by the first author, a speaker of Standard Tokyo Japanese.

To ascertain the duration of target phonemes, and in order that RT could be determined exactly from target phoneme onset, the stimuli were digitized at the Max-Planck-Institute with a sampling frequency of 16 kHz, and the onset and offset of each target phoneme was established. This was done by a combination of visual detection of spectral shifts in the waveform, and auditory verification. If one imagines a line connecting the peaks of the waveform during the vowel and the following nasal, a clear shift may be observed in this line's slope; the onset of the nasal was marked at the first positive-going zero crossing after this change in the pattern of the waveform. Figure 1 shows the resulting decision points for a nonmoraic and a moraic nasal (in *kanagu* and *kanpa*, respectively). All decisions were checked by repeated listening to the utterance on both sides of the visually determined point.

B. Subjects

Forty undergraduate students of Dokkyo University, all speakers of Tokyo Japanese, participated in the experiment for a small payment. The majority of the subjects had English as a major subject.

C. Procedure

The subjects were tested in pairs in a quiet room. They were instructed to listen for any word containing a nasal sound which could be represented with the letter N in a Roman transcription of the word, and to press the response key as soon as they had detected any occurrence of this target. Given that phoneme monitoring is a difficult task for Japa-

TABLE I. Mean response time (ms) and mean number of missed responses per condition in experiment 1 (Japanese subjects).

Moraic nasals			
bilabial (*tonbo*)	dental (*konto*)	alveolar (*konro*)	velar (*rongo*)
699 (9.7%)	695 (16.9%)	691 (12.9%)	692 (14.5%)
Nonmoraic nasals			
na (*senaka*)	ni (*tenisu*)	no (*senobi*)	nu (*tenuki*)
738 (37.9%)	735 (27.4%)	750 (37.1%)	727 (40.3%)

nese listeners, there were two separate practice sessions. First, subjects were asked to listen with eyes closed as the experimenter read out a number of words, and to raise their hand whenever a word contained a nasal sound as specified. If correct performance confirmed that the subjects had understood the nature of the task, they were presented with the 12 practice sequences. Only if subjects then also responded to a majority of the eight target items in the practice sequences were they tested in the experiment proper. Only if subjects responded to more than half of the targets in the experiment proper within a response window of one second was their data analyzed. 31 of the 40 subjects met this criterion.

The sequences were presented via 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 Sona Graph 5500 to ascertain reaction times from target onset.

D. Results and discussion

Analyses of variance were conducted on the mean rates of missing data in each condition and the mean response times in each condition, separately across subjects (F_1) and across items (F_2) as random factors. In the response time analyses, missing values were replaced by the mean for that subject in that condition (for the analysis with subjects as random factor), or by the mean for that item in that condition (for the analysis across items). The condition means are given in Table I.

The analysis of response times showed that moraic targets were detected significantly more rapidly (694 ms) than nonmoraic targets (737 ms; $F_1[1,30]=22.77$, $p<0.001$; $F_2[1,24]=4.3$, $p<0.05$). There was no difference among the moraic or the nonmoraic phonetic context conditions. The analysis of the miss rates showed the same pattern: Moraic targets were detected significantly more accurately (13.5% missed) than nonmoraic targets (35.7%; $F_1[1,30]=30.84$, $p<0.001$; $F_2[1,24]=21.43$, $p<0.001$), but there were no other significant differences between conditions.

The advantage for moraic targets in this task replicates the results of earlier studies. We extended the earlier results by carrying out a further analysis in which we examined possible effects of morphological structure. In ten of the moraic-nasal words (e.g., *kan-pa*), a morpheme boundary followed the nasal (and the word's representation in kanji characters reflected this); this was (necessarily) never the case in the nonmoraic-nasal words. Conversely, in eight of the nonmoraic-nasal words (e.g., *te-nuki*), a morpheme boundary preceded the target sound, which was never the case in the moraic-nasal words. For both moraic- and nonmoraic-nasal item sets we compared responses to the items containing such a boundary with responses to items in which the target occurred at a point which was not a morpheme boundary (as in *konto, sunoko, kana-gu*). There were no effects of this morphological factor on either miss rate (all F's<1) or response time ($F<1$ for moraic, $F=1.4$ $p>0.2$ for nonmoraic). This is in agreement with an unpublished study by Hatano and Otake which compared detection of targets consisting of either the initial vowel, or the initial two-vowel sequence, in Japanese words such as *ieji* (morphemically *ie-ji*) and *ieki* (morphemically *i-eki*); responses to both types of targets were essentially identical in both types of word. Thus the moraic effect observed in earlier work does not seem to be contaminated by morphological effects.

New in the current results is the absence of any effect of phonetic context. Neither in RTs nor in miss rates was any effect to be seen of the fact that the phonetic realization of the target in the moraic condition was highly variable: [m], [n], [n̪], [ŋ]. Thus these results seem to suggest that Japanese listeners are able to abstract rapidly from the allophonically variable realization of a moraic nasal to a unitary target representation. Nevertheless, to test alternative possibilities we carried out some additional analyses. First we compared the measured durations of the phoneme targets. Moraic phonemes in Japanese are known to be longer than nonmoraic (Han, 1994; Sato, 1993), and our target phonemes were no exception to this rule: The mean duration of moraic targets was 151.5 ms, of nonmoraic targets 63.7 ms,[2] this difference was significant ($F[1,24]=70.66$, $p<0.001$). (There was no difference between the four phonetic realization contexts for the moraic targets, however: $F<1$.) This difference raises the possibility that the observed differences in RT and miss rate could have been due to durational differences between the targets in coda versus prevocalic position, rather than to these targets' differing moraic status *per se*. Accordingly a correlation analysis was carried out (separately for moraic and nonmoraic targets, given that significant differences exist between these two types of target on all the measures involved), in order to check whether RT and/or miss rate were indeed negatively related to duration of the target. No significant correlation was observed ($r[15]=0.07$, 0.28, -0.14 and -0.05, respectively for RT to moraic targets, miss rate to moraic targets, RT to nonmoraic targets, miss rate to nonmoraic targets; all $p>0.25$; note that the correlation coefficients for the moraic targets were, in the event, positive). This finding in fact replicates results from previous phoneme-monitoring studies which have reported negative correlations between target duration and both RT and miss rate for vowel targets (van Ooijen, 1994; Cutler *et al.*, 1996) but never for consonant targets, including nasal consonants (van Ooijen, 1994; Norris *et al.*, 1992).

Next, in experiment 2, we obtained comparison data on the same materials from listeners who do not know the language. This enables us to ascertain whether, for example, moraic nasals may be detected without reference to their place of articulation, e.g., by detection of nasalization during the preceding vowel. If so, then non-native listeners might be able to show the same pattern of results. As a non-native control group we chose speakers of Dutch. This enables an extreme test of such a suggestion, since for Dutch listeners, instructions to detect the target /n/ should produce no response with a prebilabial moraic nasal, as the realization of the nasal in this context should rather be perceived as the phoneme /m/. Moreover, Dutch listeners can certainly exploit cues to consonant identity in a preceding vowel (Pols and Schouten, 1978). Thus if a significant proportion of Dutch subjects respond to the bilabial nasal given /n/ as target, and if there is in general no difference between phonetic context conditions, then it may be the case that the apparent abstraction from phonetic context by subjects in experiment 1 is based on functional equivalence of anticipatory information across the phonetic contexts.

II. EXPERIMENT 2

A. Materials and procedure

Experiment 2 was carried out in two parts. The first part was a direct replication of experiment 1, with /n/ as specified target phoneme. If subjects fail to respond to bilabial nasals (as expected), then it is reasonable to expect that this is because these nasals are perceived to be instances of a phoneme other than the specified target. Therefore in the second part of the experiment the bilabial nasals were presented with /m/ as target phoneme, in the expectation that this would elicit response from Dutch subjects.

For experiment 2a the experimental materials were exactly as for experiment 1. For experiment 2b, a subset of these materials was used, in which the four experimental words with prebilabial moraic nasals, plus seven words with nonmoraic occurrences of the phoneme /m/, served as targets; thus there were 11 experimental sequences, to which were added a further five filler sequences and four practice sequences.

Subjects were tested three at a time in separate sound-attenuated booths. They were given written instructions. The stimuli were presented over Sennheiser headphones. Since phoneme detection is not a particularly difficult task for Dutch listeners, no initial practice session was conducted to test whether the task was understood. However, there was a short break after the twelve practice trials during which, if necessary, subjects were reinstructed by the experimenter.

Response times were recorded by a Hermac AT computer, from signals aligned with the onset of each sequence of words. This enabled the detection of false alarms. Response times were corrected for the distance from signal to target phoneme onset, to give measurements from the latter point.

TABLE II. Mean response time (ms) and mean number of missed responses per condition in experiments 2a and 2b (Dutch subjects).

Experiment 2a			
Moraic nasals			
bilabial (*tonbo*)	dental (*konto*)	alveolar (*konro*)	velar (*rongo*)
⋯ (91.3%)	630 (16.3%)	564 (6.5%)	584 (20.7%)
Nonmoraic nasals			
na (*senaka*)	ni (*tenisu*)	no (*senobi*)	nu (*tenuki*)
618 (27.2%)	642 (30.4%)	650 (26.1%)	674 (31.5%)
Experiment 2b			
moraic bilabials (e.g., *tonbo*)	nonmoraic bilabials (e.g., *yomogi*)		
506 (3.1%)	530 (6.5%)		

B. Subjects

Forty-eight subjects (24 in each experiment), all native speakers of Dutch in the age of 18 to 30, were recruited from the Max Planck Institute subject pool. They were paid Hfl. 8.50 for their participation. The same criterion for acceptance was applied as for experiment 1. Twenty three of 24 subjects met this criterion in experiment 2a, and all 24 in experiment 2b.

C. Results and discussion

The means for each condition of each experiment are shown in Table II. Analyses of variance were carried out as for experiment 1.

In experiment 2a, the largest effect was that subjects did not respond to bilabial moraic nasals as occurrences of the target sound /n/. 91.3% of instances elicited no response. Thus it was clear that Dutch subjects at least were not responding indiscriminately to all nasals in coda position. Further analyses were carried out excluding the bilabial moraic nasals. Moraic targets (dental, alveolar, velar) were detected more rapidly (593 ms) and more accurately (14.5% missed) than nonmoraic targets (RT 646 ms, miss rate 28.8%: RT: $F_1[1,22]=11.34$, $p<0.005$, $F_2[1,26]=7.27$, $p<0.02$; miss rate: $F_1[1,22]=16.86$, $p<0.001$, $F_2[1,26]=5.27$, $p<0.05$). Among the nonmoraic targets, no condition effects were observed either in RTs or in miss rate. Among the moraic targets, however, there was a significant difference in RTs ($F_1[2,44]=4.17$, $p<0.025$, $F_2[2,9]=5.1$, $p<0.05$); Newmann–Keuls *post hoc* analyses showed that targets in alveolar and velar contexts were detected significantly more rapidly than targets in dental contexts. In the analysis of miss rates, the phonetic context condition effect was significant only in the analysis by subjects ($F_1[2,44]=4.67$, $p<0.02$; $F_2[2,9]=1.21$); Newman–Keuls *post hoc* analyses showed that targets in alveolar contexts were detected significantly more accurately than targets in dental and velar contexts.

In experiment 2b, there was no difference between moraic and nonmoraic nasals in RT or in miss rate, in analyses either by subjects or by items; as Table II shows, the mean miss rate for the bilabial moraic nasals given /m/ as target was low and the mean RT fast, indicating that these phonemes were clearly perceived as appropriate examples of /m/ by Dutch listeners.

As for experiment 1, a correlation analysis was carried out involving the performance measures and the measured durations of the target phonemes. For the bilabial moraic targets, both performance measures were the relevant item means from experiment 2b; for all other targets, the item means were from experiment 2a. Again, no significant correlation was observed ($r[15]=0.04$, -0.22, 0.14, and -0.25 for RT to moraic targets, miss rate to moraic targets, RT to nonmoraic targets and miss rate to nonmoraic targets respectively, all $p>0.35$).

The results of this control experiment show that Dutch listeners do not respond to bilabial nasals given /n/ as a target; instead, they respond, very appropriately, to such targets as instances of /m/. Furthermore, for Dutch listeners the three remaining contexts do not seem to provide equally good tokens of /n/: nasals in alveolar contexts seem to be easiest to detect, followed by nasals in velar contexts, with dental contexts producing the hardest realizations. Thus for these listeners the task could definitely not be performed without reference to nasal place of articulation context. This finding is consistent with the results of previous experiments on assimilated phonetic sequences in Dutch by Koster (1987), in which listeners were significantly slower detecting word-final [n] in sequences such as *groen boek* when the target phoneme had undergone (optional) bilabial assimilation than when it had not. Note that it does not appear to be the case that the Dutch listeners simply responded to later cues than the Japanese listeners; the responses of the Dutch listeners (experienced members of the MPI subject pool, to be sure, with considerable practice in RT experiments) were in fact about 100 ms faster than those of the Japanese listeners. Experiment 2 therefore suggests that the performance we observed by listeners in experiment 1, abstracting away from specific realization of moraic nasals to a single unitary representation, is not an artifact of any simple acoustically based strategy which would be equally available to any listener group irrespective of native language.

The replication of the moraic/nonmoraic difference from experiment 1 might not have been expected with these subjects, given that the English listeners tested by Cutler and Otake (1994) consistently detected moraic and nonmoraic nasals equally rapidly in Japanese words. In Cutler and Otake's materials, however, all moraic nasals were realized in dental context, and it is precisely the dental context which, as we saw, proved to be most difficult for Dutch listeners (as Table II shows, the mean for dental moraic targets differs from the mean for nonmoraic targets by only 16 ms, whereas the average moraic-nonmoraic difference was 53 ms).

In further experiments, we put the ability of Japanese listeners to represent varying realizations of moraic nasals in a single unitary manner to a further test, by constructing realizations which fail to match the constraints of their phonetic context. This was done using the materials of experiment 1, but cross-splicing nasals across phonetic contexts. The following section describes the construction of these materials and experiment 3, in which they were rated for naturalness by native Japanese listeners.

III. EXPERIMENT 3

A. Materials and procedure

The materials were generated from the recording used for experiment 1. For each moraic-nasal-target word three new versions were constructed, in which the original nasal was replaced by a nasal from one of the other three target words having the same pretarget vowel. Thus the set *tonbo, konto, konro, rongo* (in which we can refer to the four original nasals by the designations M, ND, NR, NG, respectively) contained, after cross-splicing, 16 members: *toMbo, toNDbo, toNRbo, toNGbo, koMto, koNDto, koNRto, koNGto, koMro, koNDro, koNRro, koNGro, roMgo, roNDgo, roNRgo, roNGgo*.

The cross-splicing was performed, using the XWAVES/ESPS software at the Max-Planck-Institute, as follows. First each moraic nasal target phoneme was edited out of its original context, cutting at the onset of the nasal and at the start of the closure for the stop, i.e., at the previously ascertained onset and offset points of the nasal. After editing, each nasal was reassembled with three different contexts, i.e., the contexts with the same vowel preceding the nasal as the original and with a different place of articulation following it, which resulted in 48 cross-spliced versions of the original 16 targets. All 64 items, original and cross-spliced, plus all fillers, were then filtered through a low-pass filter, cutting off frequencies above 4 kHz, to eliminate possible artefactual noise resulting from the splicing.

A tape was prepared containing all these 64 words, with an additional 60 filler words, of which 40 were unaltered naturally spoken words from the recording used for experiments 1 and 2, and 20 were cross-spliced words which were created from further words from the tape. The cross-splicing in this last group was designed to produce an unnatural-sounding result; these filler items therefore served to define an effective range for subjects' responses. The last group was deliberately kept small, however, in order to avoid a contrast effect according to which the cross-spliced experimental items might sound relatively acceptable.

The items were recorded on the tape in random order; each item was however recorded twice in succession.

B. Subjects

Twenty undergraduate and graduate students from the same population as experiment 1 participated in the experiment for a small payment. Most subjects also participated in experiment 4 or 5, below.

C. Procedure

The subjects were tested in a single group in a language laboratory at Dokkyo University. They were given written instructions, which were based on those used by Fear *et al.*

TABLE III. Mean naturalness ratings (1=best, 5=worst) in experiment 3 for words from experiments 4, 5, and 6 containing moraic nasal targets.

	Nasal realization			
Context	bilabial	dental	alveolar	velar
bilabial (*tonbo*)	1.26	2.42	2.54	2.2
dental (*konto*)	2.48	1.21	2.16	2.83
alveolar (*konro*)	2.5	2.38	1.18	1.83
velar (*rongo*)	1.86	2.66	1.89	1.3

(1995), to rate the naturalness of each stimulus word on the tape. The instructions stated that they should listen to the stimulus, and then circle on the scale on their response sheet one of the numbers 1 to 5, with a rating of 1 signifying that the pronunciation of the word on the tape was appropriate and exactly as would be expected for the perceived word, and a rating of 5 signifying that the pronunciation could not be recognized as an appropriate word. The stimuli were presented over headphones from a DAT recorder.

D. Results and discussion

Table III presents the mean naturalness rating for each version of each place of articulation type. It can be seen that the original, unspliced words (on the diagonal from top left to bottom right) receive, as expected, the best ratings. Thus subjects were clearly sensitive to the effects of the splicing manipulation. Nevertheless, the overall rating for the whole set of experimental words was 2.0 (''the pronunciation of the word is fairly natural although not quite perfect''), and the range for the 48 cross-spliced words was from 1.2 (near ''exactly as would be expected'') to 3.65 (between ''unusual'' and ''very strange''); that is to say, the experimental items were not in general perceived to be unacceptable renditions of the target words.

An analysis of variance revealed no main effect of context place of articulation and no main effect of nasal place of articulation, but an interaction between these factors ($F1[9,171]=66.72$, $p<0.001$; $F2[9,48]=4.32$, $p<0.001$). Newman–Keuls *post-hoc* tests revealed that the ratings for the original words did not differ among themselves, but all were significantly better than all but the three highest rated of the cross spliced word types. Within the 12 cross-spliced word types, few significant differences emerged. High ratings were received by three types: Velars in alveolar context, bilabials in velar context, and alveolars in velar context; ratings for these three did not differ among themselves and did not differ from ratings for the original words, and they were also significantly better than the word type which received the lowest ratings, namely velars in dental context. There were no other significant differences within the cross-spliced set. Thus no particular nasal type, and no particular context, emerged as particularly amenable or resistant to the cross-splicing manipulation; ratings depended on the particular pairing of nasal type and context. However, it is quite clear that the mismatch between nasal type and context affects Japanese listeners' judgments of naturalness in the cross-spliced words.

An important factor which could contribute to unnaturalness in our materials is the effect of pitch accent pattern. Recall that all our stimulus words were trimoraic. There are in Tokyo Japanese three possible pitch accent patterns for uninflected trimoraic words: HLL (high-low-low), LHH, or LHL (in fact, all our experimental words had either HLL or LHH patterns). The middle accent value is realized, at least in principle, on the middle mora (in fact, a pitch change associated with a second mora can begin during the vowel of an initial CV mora, although this is most likely to occur in fast speech; Nagano-Madsen, 1991, 1992). Thus taking a nasal mora from a HLL word and splicing it into a LHH word, or vice versa, would create an impossible accent pattern: LLH or HHL. In our materials, 22 of the 48 cross-spliced items suffered such an accent mismatch. To assess the effects of accent placement, the 48 cross-spliced words were divided into two sets—the 22 involving an accent mismatch between source and context word, and the 26 in which source and context word had the same accent pattern. The mean rating for words constructed from mismatched source and context was 2.76, for words constructed from matched source and context 1.94, a highly significant difference ($F_1[1,19]=151.32$, $p<0.001$; $F_2[1,46]=20.81$, $p<0.001$); the accent factor therefore did indeed appear to play a substantial role in the naturalness ratings.

Having established that a mismatch between nasal place of articulation and context place of articulation results in words which are not unacceptable to Japanese listeners, but nevertheless are perceptibly different from the unspliced original words, we tested in the following experiments the effects of this mismatch on phoneme detection performance.

IV. EXPERIMENT 4

A. Materials and procedure

The original and cross-spliced items from experiment 3 all served as moraic targets in this experiment; the nonmoraic targets from experiments 1 and 2 were also used. Four tapes were made, each as for experiment 1 except that context place of articulation and nasal place of articulation for the moraic targets were counterbalanced across tapes. The procedure was as in experiment 1.

B. Subjects

Forty undergraduate students from the same population as experiment 1 participated in the experiment for a small payment. Ten subjects heard each of the four manipulated experimental tapes. None had taken part in experiment 1. The field of study of the majority of the subjects was English. The results of six subjects were rejected for failing to meet the criterion (all missed more than 60% of targets in total); the remaining subjects were distributed such that the four tape groups had, respectively, nine, ten, eight, and seven members.

TABLE IV. Mean response time (ms) and mean number of missed responses per condition in experiment 4 (Japanese subjects).

Moraic nasals				
	Nasal realization			
Context	bilabial	dental	alveolar	velar
bilabial (*tonbo*)	708 (15.9%)	744 (22.0%)	733 (17.3%)	725 (29.2%)
dental (*konto*)	760 (27.3%)	703 (9.5%)	733 (12.2%)	759 (16.7%)
alveolar (*konro*)	740 (28.9%)	714 (12.6%)	721 (22.8%)	720 (19.7%)
velar (*rongo*)	737 (13.8%)	736 (9.5%)	723 (26.6%)	695 (24.6%)
Nonmoraic nasals				
na (*senaka*)	ni (*tenisu*)	no (*senobi*)	nu (*tenuki*)	
783 (39%)	766 (27.2%)	778 (44.1%)	749 (40.4%)	

C. Results and discussion

The relevant means are shown in Table IV. Analyses were performed as for experiments 1 and 2.

An initial analysis compared responses to moraic versus nonmoraic targets. As in all earlier studies with Japanese listeners, the moraic targets were detected more rapidly (737 ms) and more accurately (19.1% missed) than the nonmoraic (RT 769 ms, miss rate 37.7%; RT: $F_1[1,33]=12.65$, $p<0.005$; $F_2[1,96]=14.26$, $p<0.001$; miss rate: $F_1[1,33]=30.47$, $p<0.001$; $F_2[1,96]=31.1$, $p<0.001$).

Subsequent analyses compared the conditions among the moraic targets. There was no main effect of context place of articulation and no main effect of nasal place of articulation either in RTs or in miss rates. Although the results in Table IV certainly suggest that some substitutions were more acceptable than others (e.g., before an [r], a pre-dental [n̪] was responded to slightly faster and more accurately than the original [n], while a pre-bilabial [m] was responded to more slowly and less accurately), the interaction between context place of articulation and nasal place of articulation was also insignificant in both RTs and miss rates. Thus again the Japanese listeners seemed to be responding with equal facility to moraic nasals irrespective of their precise realization; [m], [n̪], [n], and [ŋ] were equally rapidly and accurately responded to as instances of the designated target.

A more sensitive test to determine whether listeners were affected by the cross-splicing manipulation was undertaken by comparing nasals in their original contexts versus the mean of their cross-spliced versions. The mean miss rate for nasals in their original contexts was 18.4%, the mean miss rate for cross-spliced nasals 19.4%, a statistically insignificant difference (both F_1 and $F_2<1$). The mean RT to nasals in their original contexts was 705 ms, to cross-spliced nasals 735 ms, and this difference was significant ($F_1[1,33]=11.54$, $p<0.005$; $F_2[1,24]=4.81$, $p<0.05$). Thus although mismatch between context and realization did not prevent subjects from responding to the targets, it did cause some detectable delay in their responses, indicating that the mismatch was of perceptual significance.

A correlation analysis was carried out, as for experiments 1 and 2, between the duration of the actual nasal target and mean RT and mean miss rate for each of the 64 original and cross-spliced moraic target items; as in the earlier experiments, neither correlation was significant and both correlation coefficients were in this instance positive ($r[63]=0.24$ for RT, 0.04 for miss rate). A further correlation analysis, however, established that there was a positive relationship between RT and rated naturalness (the lower, i.e., better, the rating received by an item in experiment 3, the lower, i.e., faster, the RT in the present experiment): $r[63]=0.29$, $p<0.025$. Rated naturalness was not, however, significantly correlated with miss rate.

Our final experiments further explored possible effects of such perceived mismatch between nasal place of articulation and context place of articulation. For instance, match versus mismatch effects might be detectable in processing of the following consonant.[3] It is possible that listeners exploit homorganic assimilation of a nasal in word perception to obtain anticipatory information about the place of articulation of the consonant following the nasal. Thus listeners' performance of a phoneme detection task on the post-nasal consonant in our materials might be more difficult in the cross-spliced than in the original words. Such a task also offers us the opportunity to compare the effects of mismatch between context and realization of the moraic nasal in native and non-native listener populations. Monitoring by non-native listeners from our control group for cross-spliced nasal targets would be difficult to interpret, given that for these listeners some of the targets fall into the class /n/ and some into the class /m/; the post-nasal consonants, however, should present no particular difficulty as monitoring targets for Dutch listeners.

V. EXPERIMENT 5

A. Materials and procedure

The materials were the four tapes constructed for experiment 4. Target specifications were prepared separately for each sequence. For the sequences containing an experimental item with a moraic nasal, the target was the consonant following the nasal. One item, *tento,* had to be omitted since it contained two occurrences of the same consonant. Thus there were 15 experimental target items on each tape, and the target consonants were: p(2), b(2), d(2), t(1), r(4), k(2), g(2). For filler sequences (both experimental and practice) which originally contained no occurrence of the original nasal target, one of these seven targets was chosen such that the sequence still contained no occurrence of the target. The remaining sequences (both experimental and practice), including those which had originally contained nonmoraic experimental targets, were assigned one of the seven targets such that there was one occurrence of the target in the sequence. The seven consonant targets occurred with roughly equal frequency. In a few cases filler items were rearranged across sequences to avoid multiple occurrences of a target within a

TABLE V. Mean response time (ms) and mean number of missed responses per condition in experiment 5 (Japanese subjects), as a function of preceding moraic nasal.

	Nasal realization			
Target	bilabial	dental	alveolar	velar
bilabial (*tonbo*)	504 (0%)	596 (7.5%)	600 (2.5%)	572 (27.5%)
dental (*konto*)	583 (20%)	497 (6.7%)	558 (3.3%)	658 (10%)
alveolar (*konro*)	605 (17.5%)	537 (12.5%)	591 (5%)	559 (10%)
velar (*rongo*)	645 (10%)	595 (10%)	575 (10%)	563 (5%)

TABLE VI. Mean response time (ms) and mean number of missed responses per condition in experiment 6 (Dutch subjects), as a function of preceding moraic nasal.

	Nasal realization			
Target	bilabial	dental	alveolar	velar
bilabial (*tonbo*)	502 (2.5%)	523 (7.5%)	548 (12.5%)	541 (30%)
dental (*konto*)	567 (30%)	516 (20%)	547 (20%)	558 (20%)
alveolar (*konro*)	453 (36.7%)	408 (33.3%)	441 (50%)	447 (36.7%)
velar (*rongo*)	544 (12.5%)	513 (15%)	482 (22.5%)	449 (25%)

sequence. This was carried out at the Max-Planck-Institute from the same computer files that had created the tapes for experiment 4.

The procedure was as in experiments 1 and 4 except that the target for each sequence was presented visually immediately prior to presentation of the sequence. Again, the targets were presented as Roman letters. Response times to the stop consonant targets were computed from the onset of the closure.

B. Subjects

Forty graduate and undergraduate students from the same population as for experiment 1 participated in the experiment for a small payment. None had taken part in experiment 1 or 4. Most of the subjects had English as their major field of study. All 40 subjects met the selection criterion.

C. Results and discussion

Table V shows the means for each condition. Again, the results were analyzed as in the previous experiments.

All targets in this experiment were in nonmoraic position, therefore the analyses considered only the effects of context place of articulation and nasal place of articulation. As in experiment 3, there was no main effect of either of these factors, either in RTs or in miss rates. (Note that an effect of context place of articulation in this case would effectively have indicated that the various phoneme targets elicited different response patterns; the absence of such an effect is consistent with previous reports from stop consonant detection studies in English: Martin, 1977.) Again, we compared responses to consonants following nasals in their original contexts versus their cross-spliced versions. The mean RT to consonants following nasals in their original contexts was 544 ms, to consonants following cross-spliced nasals 592 ms, and this difference was significant ($F_1[1,39]=25.95$, $p<0.001$; $F_2[1,52]=7.11$, $p<0.02$). The mean miss rate for consonants following nasals in their original contexts was 4%, while the mean miss rate for consonants following cross-spliced nasals was 11.8%, and this difference was also significant (although marginally so in the analysis across items: $F_1[1,39]=16.62$, $p<0.001$; $F_2[1,52]=3.73$, $p<0.06$). Thus a mismatch between context and realization of the moraic nasal phoneme exercised a clear effect on responses to the following consonant.

A correlation analysis again confirmed a positive relationship between RT and rated naturalness: $r[59]=0.29$, $p<0.025$. For this experiment, there was also just such a positive relationship with miss rate (the lower, i.e., better, the rating in experiment 3, the lower the miss rate in the present study): $r[59]=0.37$, $p<0.005$.

In our last experiment we presented the materials of experiment 5 to listeners from our Dutch-speaking control group.

VI. EXPERIMENT 6

A. Materials and procedure

The materials were as for experiment 5. However, the target /r/ was replaced by the target /d/ as the alveolar /r/ is perceived most frequently as /d/ by Dutch listeners. For this reason one further target-bearing item, *danro*, had to be dropped because it now contained two occurrences of the specified target. After the experiment had been completed it was discovered that one sequence with the target /d/ (for *shindo*) had contained an occurrence of /r/ in an earlier word, which had elicited false-alarm responses from all subjects, so that it was necessary to drop that target item also.

The procedure was as in experiment 2 except that the target for each sequence was presented visually immediately prior to presentation of the sequence. The instructions further emphasized that the target G referred to the sound /g/ (which does not occur in native Dutch words, but is known to Dutch listeners in loan words such as *goal*).

B. Subjects

Forty-three subjects from the same population as for experiment 2 participated in the experiment. The data of three subjects who failed to meet the criterion for inclusion were not analyzed. Of the remaining 40 subjects, ten heard each of the four stimulus tapes.

C. Results and discussion

The means for each condition are presented in Table VI. The same analyses were performed as for experiment 5.

The pattern of results was the same for RTs and for miss rates. The effect of nasal place of articulation was never significant. The effect of phonetic context was always significant (RT: $F_1[3,117]=10.16$, $p<0.001$, $F_2[3,36]=5.79$, $p<0.005$; Miss rate: $F_1[3,117]=9.17$, $p<0.001$, $F_2[3,36]=3.35$, $p<0.05$); *post-hoc* analyses showed that this effect was due to faster RTs, and more missed responses, to alveolar targets than to targets in the other three places of articulation. Again, we compared responses to consonants following nasals in their original contexts versus their cross-spliced versions. The mean RT to consonants following nasals in their original contexts was 481 ms, to consonants following cross-spliced nasals 513 ms; the mean miss rate for consonants following nasals in their original contexts was 22.3%, and for consonants following cross-spliced nasals 21.9%. Neither difference reached significance. Thus for this group of subjects responses to the post-nasal consonants were unaffected by whether the realization of the immediately preceding nasal was contextually matched or mismatched.

We also carried out a correlation analysis between the naturalness rating as assigned to each item by the Japanese listeners of experiment 3 and the RTs and miss rates for the same items by the Dutch listeners of the present study. As might be expected, no relationship was found.

Experiments 5 and 6 showed, therefore, very different patterns of results; native listeners make use of place of articulation cues in moraic nasals in the processing of a following consonant, but non-native listeners do not.

VII. GENERAL DISCUSSION

Our experimental exploration of the processing of Japanese moraic nasals via the phoneme detection task has led to a confirmation of previous findings and to two new conclusions. As in previous on-line listening studies, the Japanese listeners consistently showed sensitivity to mora structure; responses in experiments 1 and 4 were faster to the moraic target phonemes than to nonmoraic targets. Correlation analyses showed that this effect could not be ascribed to durational differences between moraic and nonmoraic nasals.

Further, we observed in experiment 1 that Japanese listeners responded equally rapidly and accurately to moraic nasal phonemes irrespective of whether they were realized as [m], [n], [n̪] or [ŋ]. In experiment 4, they again responded equally rapidly and accurately to the four place of articulation realizations although they were sensitive to whether or not the place of articulation of the nasal matched the phonetic context in which it was heard. Thus our first conclusion on the basis of the present findings is that the phonological equivalence of the moraic nasal in its various allophonically conditioned realizations is exploitable in on-line listening tasks by Japanese language users: no one place of articulation is more readily recognized as an instance of the target specification than others. The contextual variation is nonetheless real, as the control experiment 2 with Dutch listeners attested; these listeners responded differently to the different places of articulation, in line with the categories of their native phonology. For the Japanese listeners, however, each moraic nasal is actually just as good as each other one irrespective of phonetic context, as the lack of difference in RTs in experiments 1 and 4 showed. For these listeners, therefore, the representation of moraic nasals involves a level of abstraction away from the actual realization. The representation which they have is in effect equivalent to an archiphoneme which admits of different contextually determined forms. But recall that their instructions in the monitoring task did not actually require them to abstract away from surface form in this way; they were merely asked to respond to any sound which could be represented with the Roman letter N. That they responded under these instructions both to the nonmoraic [n] in *na, ni, no, nu*, and to the moraic nasals whatever their realization suggests that their own representation of the latter is an abstract one.

Second, the sensitivity of Japanese listeners to the contextual appropriateness of the place of articulation with which a nasal is realized, demonstrated in experiments 3 and 4, was further found in experiment 5 to carry over to the processing of the following consonant. Detection of the post-nasal consonant was faster when it was preceded by its original homorganically matched nasal than when it was preceded by a nonmatched nasal. (Note that the significant difference in RT between original and cross-spliced nasals clearly shows that the Japanese listeners' responses were sensitive to phonetic detail in the signal; the absence of difference between place of articulation realizations of the moraic nasal cannot be ascribed to failure to perceive variation in the phonetic realization.) This was not a simple effect of the splicing manipulation *per se*, since Dutch listeners' responses to the post-nasal consonant in the control experiment 6 were unaffected by contextual mismatch. The effects of mismatch between nasal and context on Japanese listeners' processing of the post-nasal consonant therefore prompt our second conclusion, namely that these listeners can use place of articulation cues in a moraic nasal to obtain information about place of articulation of a following phoneme.

Such use of anticipatory information is fully consistent with previous reports in the speech perception literature. Studies with the gating task, for example, have shown that listeners can use anticipatory nasalization in a vowel to derive information about an upcoming nasal or stop consonant (Ellis *et al.*, 1971; Warren and Marslen-Wilson, 1987, 1988; Lahiri and Marslen-Wilson, 1991). Where anticipatory information in fact fails to match the actual consonant which follows, processing is adversely affected in many speech perception tasks (Streeter and Nigro, 1979; Whalen, 1984, 1991; Warren and Marslen-Wilson, 1987; Marslen-Wilson and Warren, 1994). The experiments by Martin and Bunnell (1981, 1982) show very clearly that anticipatory information is available earlier than immediately prior to the phoneme in question; detection of a vowel in the second syllable of nonsense CVCV sequences such as *kuti* or *kota* was slower when the sequences had been constructed by exchanging the initial CV syllables from different sequences (*koti*, *kuta*).

Whalen (1991) argued that effects of mismatch between anticipatory cues and segment identity were operative at a prelexical level of processing, since he found significant mismatch effects in the processing of nonsense words. We found, in experiment 6, no effects of mismatching coarticulatory information in the processing of what were effectively nonsense words, namely Japanese words for Dutch listeners. However, we do not interpret this finding as evidence against Whalen's (1991) conclusion, though it certainly suggests that the use of coarticulatory information in a nasal to obtain information about a following consonant should not be seen as a low-level acoustic effect.

Instead, we assume that the fact that Japanese listeners make use of nasal-stop coarticulation, but Dutch listeners, presented with exactly the same input, do not, reflects the relative informativeness of such information in the two languages. In Japanese, homorganic assimilation of a nasal and a following consonant across syllable boundaries but within a word is obligatory (Vance, 1987), and indeed such assimilation can also occur across word boundaries. However, the word-level phonology of Dutch differs from that of Japanese in relevant respects. First, Dutch assimilation phenomena principally concern voicing rather than place (Slis, 1985). Second, although Dutch does exhibit widespread place assimilation in nasal-stop sequences; there are many obligatory or optional exceptions, which can occur across word boundaries (*om te eten; een kat; den Bosch*), across syllable boundaries within words (especially in prefixed and compound words: *omkeer, omdraaien, renbaan, ovenplaat, bloemkool*; but also *vreemdeling, ruimte*), and even within syllables (*hemd, komt*). Thus (at least with foreign input) Dutch listeners apparently ignore place of articulation information in a preceding nasal when processing a stop consonant. The experiments of Koster (1987) showed furthermore that assimilated nasals before stops cause processing difficulty for Dutch listeners; as described above, listeners were significantly slower detecting word-final [n] before a bilabial stop when assimilation had applied than when it had not. Koster also found this result when the Dutch listeners were presented with English sequences such as *chain broken*; in other words, the listeners applied the same processing criteria to foreign as to native input.

Place of articulation in a nasal is a reliable source of information about a following stop, for Japanese listeners, and they make use of it; it is less reliable for Dutch listeners, and it is not used. Just as the Bengali and English listeners tested by Lahiri and Marslen-Wilson (1991) responded differently to similar coarticulatory information in Bengali and English CV[n] syllables in accord with the possibilities of their native phonology, so too have the Japanese and Dutch listeners in the present study made different use of the same information in the Japanese words with moraic nasals. Thus our results illuminate Japanese language users' exploitation of their native phonology in listening in two ways. First, in accord with phonological accounts which stress the underlying unitary nature of the moraic nasal phoneme, our listeners were clearly able to abstract away from the variability in the phonetic realization to a single unified representation. Second, despite this efficient abstraction from the surface variability, they were able to exploit the same variability in the light of the phonological constraints of their language to obtain anticipatory cues about upcoming phonetic segments.

ACKNOWLEDGMENTS

This collaboration was supported by a grant from the Human Frontier Science Program, and the first author further acknowledges support from the Japanese Ministry of Education (Grant Number 06610475) and the Max Planck Society. We are grateful to Mary Beckman, Dianne Bradley, and Fumio Inoue for discussions which stimulated these studies. The second author is now at the Department of Linguistics, Ohio State University. Address correspondence to Dr. Anne Cutler, Max-Planck-Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands (E-mail: anne.cutler@mpi.nl).

[1]We are grateful to Mary Beckman for extensive advice on place of articulation in Japanese consonants.

[2]This is a ratio of moraic to nonmoraic nasal duration of 2.378:1, remarkably in accord with the ratio of 2.39:1 reported by Sato (1993) for averages across multiple tokens of moraic and nonmoraic nasals from eight speakers.

[3]We thank Dianne Bradley for suggesting experiment 5.

Bradley, D. C., Sánchez-Casas, R. M., and García-Albea, J. E. (**1993**). ''The status of the syllable in the perception of Spanish and English,'' Language Cogn. Process. **8**, 197–233.

Buxton, H. (**1983**). ''Temporal predictability in the perception of English speech,'' in *Prosody: Models and Measurements*, edited by A. Cutler and D. R. Ladd (Springer-Verlag, Heidelberg).

Cutler, A., and Butterfield, S. (**1992**). ''Rhythmic cues to speech segmentation: Evidence from juncture misperception,'' J. Mem. Language **31**, 218–236.

Cutler, A., and Norris, D. G. (**1988**). ''The role of strong syllables in segmentation for lexical access,'' J. Exp. Psychol. Human Percept. Perform. **14**, 113–121.

Cutler, A., and Otake, T. (**1994**). ''Mora or phoneme? Further evidence for language-specific listening,'' J. Mem. Language **33**, 824–844.

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

Cutler, A., Butterfield, S., and Williams, J. N. (**1987**). ''The perceptual integrity of syllabic onsets,'' J. Mem. Language **26**, 406–418.

Cutler, A., van Ooijen, B., Norris, D., and Sánchez-Casas, R. (**1996**). ''Speeded detection of vowels: A cross-linguistic study,'' Percept. Psychophys. **58**, 807–822.

Ellis, L., Derbyshire, A. J., and Joseph, M. E. (**1971**). ''Perception of electronically gated speech,'' Language Speech **14**, 229–240.

Fear, B. D., Cutler, A., and Butterfield, S. (**1995**). ''The strong/weak syllable distinction in English,'' J. Acoust. Soc. Am. **97**, 1893–1904.

Han, M. S. (**1994**). ''Acoustic manifestations of mora timing in Japanese,'' J. Acoust. Soc. Am. **96**, 73–82.

Koster, C. J. (**1987**). *Word Recognition in Foreign and Native Language* (Foris, Dordrecht, The Netherlands).

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

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

Martin, J. G. (**1979**). ''Rhythmic and segmental perception are not independent,'' J. Acoust. Soc. Am. **65**, 1286–1297.

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

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

Martin, J. G., Mills, C. B., Meltzer, R. H., and Shields, J. L. (**1980**). ''Anticipatory coarticulation and reaction time to phoneme targets in spontaneous speech,'' Phonetica **37**, 159–168.

Martin, M. A. (**1977**). ''Reading while listening: A linear model of selective attention,'' J. Verbal Learn. Verbal Behav. **16**, 453–463.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U., and Seguí, J. (**1981**). ''The syllable's role in speech segmentation,'' J. Verbal Learn. Verbal Behav. **20**, 298–305.

Meltzer, R. H., Martin, J. G., Mills, C. B., Imhoff, D. L., and Zohar, D. (**1976**). ''Reaction time to temporally displaced phoneme targets in continuous speech,'' J. Exp. Psychol.: Human Percept. Perform. **2**, 277–290.

Mills, C. B. (**1980**). ''Effects of context on reaction time to phonemes,'' J. Verbal Learn. Verbal Behav. **19**, 75–83.

Nagano-Madsen, Y. (**1991**). ''Moraic nasal and tonal manifestation in Osaka Japanese: Implications for the representation of mora,'' Proceedings of the XIIth International Congress of Phonetic Sciences, Aix, Vol. 2, 210–213.

Nagano-Madsen, Y. (**1992**). *Mora and Prosodic Coordination* (Lund U. P., Lund).

Norris, D. G., van Ooijen, B., and Cutler, A. (**1992**). ''Speeded detection of vowels and steady-state consonants,'' Proceedings of the Second International Conference on Spoken Language Processing, Banff, Canada, Vol. 2, 1055–1058.

Ooijen, B. van (**1994**). ''Processing of vowels and consonants,'' PhD. dissertation, University of Leiden.

Otake, T., and Yoneyama, K. (**1995**). ''Moraic status and syllable structure in speech perception,'' Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Vol. 2, 686–689.

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

Pols, L. C. W., and Schouten, M. E. H. (**1978**). ''Identification of deleted consonants,'' J. Acoust. Soc. Am. **64**, 1333–1337.

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

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

Slis, I. H. (**1985**). ''The voiced-voiceless distinction and assimilation of voice in Dutch,'' PhD. dissertation, University of Nijmegen.

Streeter, L. A., and Nigro, G. N. (**1979**). ''The role of medial consonant transitions in word perception,'' J. Acoust. Soc. Am. **65**, 1533–1541.

Swinney, D. A., and Prather, P. (**1980**). ''Phonemic identification in a phoneme monitoring experiment: the variable role of uncertainty about vowel contexts,'' Percept. Psychophys. **27**, 104–110.

Vance, T. J. (**1987**). *An Introduction to Japanese Phonology* (SUNY, Albany).

Warren, P., and Marslen-Wilson, W. D. (**1987**). ''Continuous uptake of acoustic cues in spoken word recognition,'' Percept. Psychophys. **41**, 262–275.

Warren, P., and Marslen-Wilson, W. D. (**1988**). ''Cues to lexical choice: Discriminating place and voice,'' Percept. Psychophys. **43**, 21–30.

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

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