619

Perception of the [m]–[n] distinction in VC syllables

Bruno H. Repp
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

Katyanee Svastikula
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695 and Department of Linguistics, University of Connecticut, Storrs, Connecticut 06268

(Received 10 March 1987; accepted for publication 29 September 1987)

This study complements earlier experiments on the perception of the [m]–[n] distinction in CV syllables [B. H. Repp, J. Acoust. Soc. Am. **79**, 1987–1999 (1986); B. H. Repp, J. Acoust. Soc. Am. **82**, 1525–1538 (1987)]. Six talkers produced VC syllables consisting of [m] or [n] preceded by [i,ɑ,u]. In listening experiments, these syllables were truncated from the beginning and/or from the end, or waveform portions surrounding the point of closure were replaced with noise, so as to map out the distribution of the place of articulation information for consonant perception. These manipulations revealed that the vocalic formant transitions alone conveyed about as much place of articulation information as did the nasal murmur alone, and both signal portions were about as informative in VC as in CV syllables. Nevertheless, full VC syllables were less accurately identified than full CV syllables, especially in female speech. The reason for this was hypothesized to be the relative absence of a salient spectral change between the vowel and the murmur in VC syllables. This hypothesis was supported by the relative ineffectiveness of two additional manipulations meant to disrupt the perception of relational spectral information (channel separation or temporal separation of vowel and murmur) and by subjects' poor identification scores for brief excerpts including the point of maximal spectral change. While, in CV syllables, the abrupt spectral change from the murmur to the vowel provides important additional place of articulation information, for VC syllables it seems as if the formant transitions in the vowel and the murmur spectrum functioned as independent cues.

PACS numbers: 43.71.Es

INTRODUCTION

In recent studies, Kurowski and Blumstein (1984) and Repp (1986, 1987) have investigated the perception of the [m]–[n] distinction in natural CV syllables. Their results have shown that, for prevocalic nasal consonants, place of articulation information is generally contained both in the spectrum of the nasal murmur and in the vocalic formant transitions following the point of release. In addition to combining these two separate sources of information, however, listeners derive information from the spectral relationship between the two signal portions, which appears to be a crucial cue for the [m]–[n] distinction in the context of front vowels such as [i]. Kurowski and Blumstein (1984, 1987) hypothesized the existence of a single auditory property for place of articulation representing the spectral change from the murmur into the vowel, and both they and Repp (1986) speculated about the possible role of auditory short-term adaptation caused by the murmur in establishing or enhancing this distinctive auditory property at vowel onset. The most recent perceptual results (Repp, 1987), however, suggest that the perception of this spectral relationship does not depend on peripheral auditory enhancement, at least not under favorable listening conditions.

The present study complements the CV syllable experiments of Repp (1986, 1987) by examining the perception of the [m]–[n] distinction in VC syllables using similar methods. (To facilitate comparisons with the earlier data, the nasal consonant [ŋ], which occurs only in postvocalic position in English, was not included.) As in prevocalic nasals, the place of articulation of postvocalic nasals is conveyed by the vocalic formant transitions and the nasal murmur, occurring in reverse order. There are several important differences, however, which make a comparison interesting.

First, final nasal consonants may be released, and if so, the release transient (which may continue into a brief neutral vowel) contains salient additional place of articulation cues. Clearly, to compare the perception of initial and final nasals, only final nasals without releases should be considered. Nevertheless, the omission of an additional (however, optional) piece of information may entail some loss in intelligibility.

Second, there are two reasons for expecting the spectral relationship between vowel and murmur to be less salient perceptually in VC than in CV syllables. One reason is that, because the murmur *follows* the vowel containing the formant transitions ("vowel" is used here to denote the signal portion preceding the point of closure in a VC syllable), it cannot have any auditory adaptation effect on the vowel. Although adaptation caused by the vowel may modify the auditory representation of the murmur, it seems unlikely that this peripheral interaction would enhance place of articulation information, since it would only attenuate the already weak higher formants of the murmur, which are continuous with the formants at vowel offset. Thus one process hypothesized to establish relational spectral information for

 0001-4966/88/010237-11$00.80

place of articulation (Kurowski and Blumstein, 1984) presumably does not operate here. A second reason is that the transition between vowel and murmur in VC syllables is not as abrupt as the murmur–vowel transition in CV syllables (cf. Kurowski and Blumstein, 1987). Vowels preceding nasal consonants are commonly nasalized, more so than following vowels (see, for example, Ali *et al.*, 1971; Ostreicher and Sharf, 1976), and this anticipatory opening of the velar port reduces the spectral contrast between the vowel and the murmur. Also, Schouten and Pols (1979) have suggested that, while prevocalic nasal murmurs contain no formant transitions, formant movements may extend from a vowel into a following murmur, suggesting that articulatory adjustments continue after oral closure. This would also contribute to making the spectral change less abrupt.

Third, the perceptual contributions of the vowel and murmur components themselves may also differ between CV and VC syllables. The vocalic formant transitions of final nasals are not mirror images of those of initial nasals, and, in fact, may be more distinctive (Broad and Fertig, 1970). Perceptual experiments with truncated CV and VC syllables containing stop consonants have suggested that VC transitions provide stronger place of articulation cues than CV transitions (Ohde and Sharf, 1977, 1981; Pols and Schouten, 1978, 1981). It remains to be seen whether this is also true for nasal consonants. Final nasals also have longer murmurs than initial ones (Malécot, 1956). Although murmur duration as such does not seem to have much of an effect on intelligibility (Repp, 1987), the possible presence of formant transitions in the murmur and its terminal position in the utterance may give it greater salience in VC than in CV syllables.

These considerations lead to the prediction that, although identification accuracy may be lower for (unreleased) final than for initial nasals in full syllables, the vowel and murmur components by themselves should be at least as informative in VC as in CV syllables. This paradoxical situation could arise because the spectral change between vowel and murmur is less important perceptually in VC syllables, so that the place of articulation information derives from two independent cues, as it were, without any additional "relational term" in the perceptual equation.

The only previous study in the readily accessible literature that compared the perception of nasal consonants in natural CV and VC syllables was conducted by Malécot (1956). It included [m, n, ŋ] in the context of a single vowel, [æ], apparently produced by a single talker. Stimuli were constructed by cross-splicing murmurs and vowels; murmurs, but not vowels, were also presented in isolation. The results suggested that murmur cues were more salient in final than in initial position, but they did not permit any conclusions about the relative contribution of the vocalic formant transitions.

I. EXPERIMENT 1

Repp's (1986) waveform-editing study with CV syllables included five conditions, four of which were replicated here with VC syllables: progressive truncation from the beginning, progressive truncation from the end, presentation of brief excerpts from the vicinity of the point of closure (corresponding to the point of release in CV syllables), and replacement of the same brief segments in the intact syllables with signal-correlated (i.e., envelope-matched) noise. The first two conditions served to determine the relative informativeness of the vowel and murmur portions in isolation, and the extent to which place of articulation information in each is located near the closure point. The other two conditions assessed the perceptual importance of the relationship between vowel and murmur spectra, and the extent to which that relational information rests on the availability of the point of maximal spectral change (the closure point). If that point is perceptually important, brief excerpts straddling the closure point should yield higher identification performance than excerpts from either side of that point, and replacement of these segments with noise should lead to lower identification scores than replacement of segments from within the vowel or the murmur.

A. Methods

1. Talkers and recording procedure

Six native speakers of American English served as talkers, three males (AA, GK, and JS) and three females (CG, SN, and BT). Five were researchers or graduate students under 40 years of age; one (AA) was an experienced phonetician in his early sixties. Three of the talkers (AA, CG, BT) had also served in the earlier study on CV syllables (Repp, 1986); the other three talkers of that study were no longer available and had to be replaced.

The talkers were asked to produce the syllables [ɑm, im, um, ɑn, in, un] as naturally as possible. The recording was done in a sound-insulated booth using high-quality equipment.

2. Stimuli and test sequences

The basic set of stimuli included 36 syllables (6 talkers × 6 syllables). These syllables were low-pass filtered at 4.9 kHz, digitized at a 10-kHz sampling rate, and stored in separate computer files. The waveforms were then inspected to determine whether any syllables had a final release. Of the 36 tokens, 14 were found to be released (all utterances of female talkers CG and BT, and one each of JS and SN). In each of these tokens, the portion of the waveform including and following the final release was removed, so as to ensure homogeneity of the stimulus set and to facilitate comparisons with syllable-initial consonants. It was assumed that these tokens would be equivalent to originally unreleased ones; however, see experiment 2 below for a detailed investigation of this issue.

Subsequently, a waveform editor was used to place seven markers ("cutpoints") in each file, as illustrated in Fig. 1. The marker labeled "0" was placed at the onset of what was taken to be the first glottal cycle of the nasal murmur. This point (the closure point) was defined as a visible amplitude drop and/or a decrease in high-frequency oscillations in the waveform. For reasons having to do with the ease of locating exact zero crossings (see Repp, 1986), the marker was placed at a downgoing zero crossing in male waveforms (Fig. 1, upper panel), but at an upgoing zero crossing in

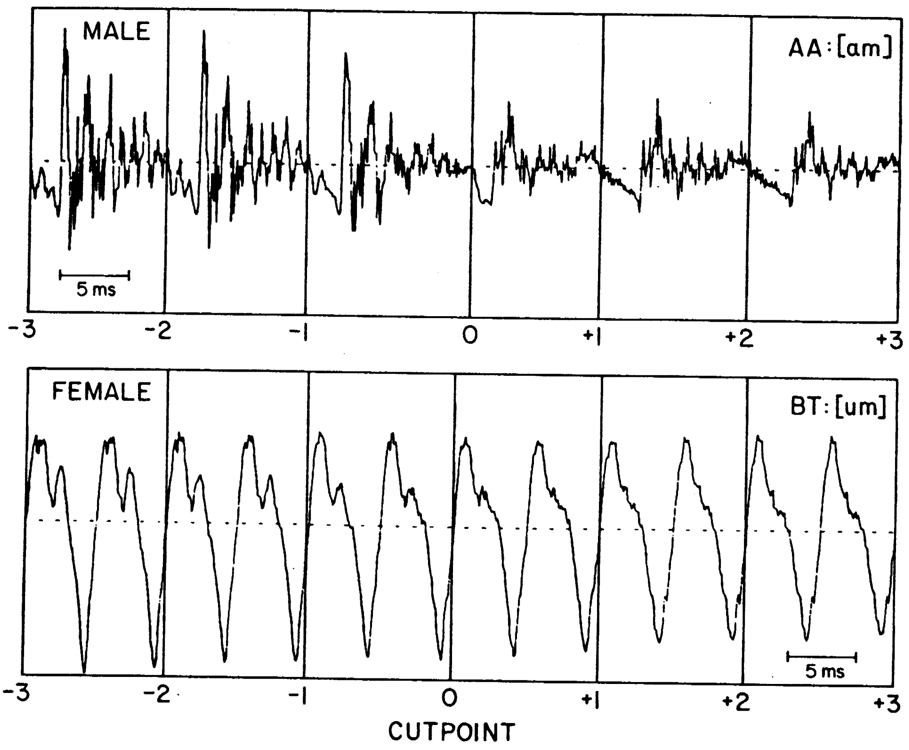

FIG. 1. Oscillograms of the waveforms in the vicinity of the presumed point of closure for a male [ɑm] and a female [um], with cutpoint markers in place.

female waveforms (lower panel). No perceptual consequences of this procedural difference were expected. The closure point could be determined with some confidence in [ɑ –] and [i –] syllables (see upper panel), but it was almost impossible to find in [u –] syllables (see lower panel). In these syllables, therefore, we made an "educated guess" based on the slope of the amplitude envelope, on the expected intrinsic duration of [u] as compared to [i] and [ɑ] (Peterson and Lehiste, 1960), and on listening carefully to the gated stimulus portions. The markers labeled – 3, – 2, – 1, + 1, + 2, and + 3 were placed at the onsets of glottal cycles preceding and following the 0 marker. The intermarker intervals will be referred to as "segments." There was one glottal cycle per segment for males (except for talker JS, whose relatively high fundamental frequency in three syllables suggested having two glottal cycles per segment in those tokens) and two for females. The average duration of the intermarker intervals, calculated over the – 3 to + 3 range, and the corresponding average fundamental frequencies in the vicinity of the closure point for the six talkers were as follows: 10.3 ms, 97 Hz (AA); 9 ms, 112 Hz (GK); 10 ms, 146 Hz (JS); 9.2 ms, 217 Hz (CG); 11 ms, 182 Hz (SN); 11.3 ms, 176 Hz (BT). A nominal segment duration of 10 ms will be assumed in discussing the results.

The stimuli, with cutpoint markers in place, were used to prepare four test tapes corresponding to four experimental conditions. Each test tape contained 7–8 test sequences. Each test sequence consisted of the 36 individual syllables, modified as described below, in random order. The interstimulus interval was 3 s.

(a) Truncation from the beginning ("murmurs"). There were eight test sequences in this condition. The first sequence contained the full syllables. The remaining seven sequences presented the stimuli starting at cutpoints – 3, – 2, – 1, 0, + 1, + 2, and + 3, respectively.

(b) Truncation from the end ("vowels"). Eight test sequences were prepared for this condition. The first sequence contained the full syllables. The remaining sequences presented the stimuli ending at cutpoints + 3, + 2, + 1, 0, – 1, – 2, – 3, respectively.

(c) Extraction of brief segments ("excerpts"). This tape contained seven test sequences containing the following excerpts: – 3/ + 3 (i.e., from cutpoint – 3 to cutpoint + 3), – 2/ + 2, – 1/ + 1, – 2/0, 0/ + 2, – 3/ – 1, and + 1/ + 3. Therefore, the duration of the stimuli was about 60 ms in the first sequence, 40 ms in the second sequence, and 20 ms in the remaining sequences. The segments in sequences 1–3 straddled the closure point, whereas those in sequences 4 and 6 came from within the vowel, and those in sequences 5 and 7 came from within the murmur.

(d) Replacement of segments with signal-correlated noise ("SCN"). There were seven test sequences in this condition, each containing full syllables in which the + 1/ + 3, – 3/ – 1, 0/ + 2, – 2/0, – 1/ + 1, –2/ + 2, and – 3/ + 3 segments had been replaced with signal-correlated noise. (The order is reversed with respect to the excerpt tape.) The duration of the noise in the syllables thus was about 20 ms in sequences 1–5, 40 ms in sequence 6, and 60 ms in sequence 7. The noise was generated by a computer program from specified segments within each waveform by randomly reversing the polarity of digital sampling points with a probability of 0.5 (Schroeder, 1968). By this method, the amplitude envelope of the original signal was maintained, but the spectral information was destroyed.

3. Subjects and procedure

Twelve native speakers of American English served as listeners. They were paid student volunteers with reportedly normal hearing.

The test tapes were played to the subjects binaurally at a comfortable volume over TDH-39 earphones in a quiet room. There were one or two subjects per session, which lasted about 90 minutes. The excerpts tape was always presented last, since it was considered to be the most difficult test due to the short duration of the stimuli. The other three conditions were presented in all six possible orders, with two subjects for each. The order of test sequences within each condition was fixed roughly according to progressive difficulty, as described above.

The subjects were asked to judge whether a stimulus was derived from a syllable ending with [m] or [n], by writing down /m/ or /n/ for each stimulus. If no nasal consonant was heard, a guess was to be made. Subjects were told that there were several talkers, and that there was an equal number of [– m] and [– n] syllables. One subject was replaced because his identification of the unaltered syllables was at chance level.

4. Data analysis

The (untransformed) data were analyzed in two types of analysis of variance (ANOVA): across subjects (averaged over talkers) and across talkers (averaged over subjects). Therefore, two F values will be reported for each effect tested, and only effects for which both F values are significant will be reported. Differences among individual syllables were assessed by including consonant and vowel as factors in the ANOVAs. In the analysis across talkers, talker sex was an additional factor.

B. Results and discussion

1. Murmurs

The results of the murmurs condition are shown in Fig. 2 as the open circles. It may be noted, first, that the full syllables were not perfectly identified: The average score was only 88% correct, which contrasts with the near-perfect identification of unaltered CV syllables (Repp, 1986). Closer inspection of the data revealed that the female tokens were much more poorly identified (77% correct) than the male tokens (98% correct). Four female tokens were especially conspicuous in that they tended to be identified at or below chance accuracy. Three of them (CG's [ɑn] and [in], and BT's [um]) had been originally released, and listening to the original stimuli confirmed that they had been correctly articulated by the talkers. Thus the removal of the final release may have impaired the intelligibility of these syllables (however, see experiment 2 below). The fourth "bad" token (SN's [um]) may have been poorly articulated. Even with these tokens omitted, however, the score for female utterances was only 90% correct.

As the vowel was cut back, performance declined to somewhat below 70% correct. The leveling off of identification performance in the vicinity of the 0 cutpoint confirms that this marker had been placed with reasonable accuracy.

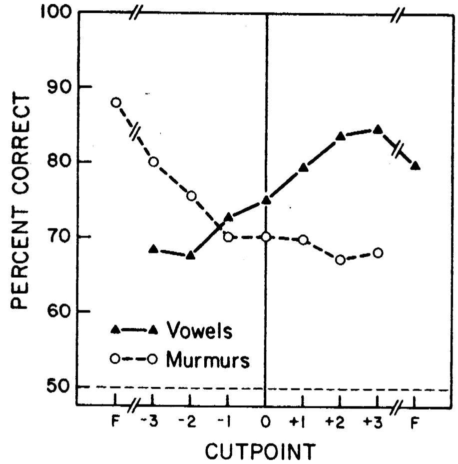

FIG. 2. Percent correct identification for syllables in the murmurs and vowels conditions of experiment 1 as a function of truncation (F = full syllable).

Further cutback of the murmur itself did not lead to any decline in identification scores. This is not surprising, since the murmurs were rather long in duration (236 ms on the average, ranging from 119 to 416 ms), so that removal of the initial 30 ms hardly made any difference. If there were any formant movements during this portion (as observed by Schouten and Pols, 1979), they were not perceptually salient. Identification of the isolated VC murmurs was quite comparable in accuracy to that of isolated CV murmurs (Repp, 1986), even though the latter were of much shorter duration. Since murmur duration has relatively little influence on identifiability, within limits (Repp, 1986, 1987), it may be concluded that VC and CV murmurs convey about the same amount of place of articulation information. Malécot's (1956) observation that final murmurs are perceptually more salient than initial ones may hold only when conflicting transitions and murmurs are spliced together.

The scores for individual syllables, averaged over talkers, are shown in Fig. 3. It is evident that [– m] syllables suffered much less from elimination of the vocalic formant transitions than did [– n] syllables. That is, isolated [m] murmurs were identified more accurately than [n] murmurs. It is not clear whether this should be considered a response bias or a consequence of labial place of articulation information somehow being conveyed more strongly in murmurs (see also Malécot, 1956; Repp, 1987). It cannot have been entirely due to a response bias, however, because it depended on the original vocalic context: [(u)m] murmurs were less well identified than [(ɑ)m] and [(i)m] murmurs (though there may have been one bad token of [um]), but [(ɑ)n] murmurs were less well identified than [(u)n] and [(i)n] murmurs. Although most of these differences parallel those found with CV syllables (Repp, 1986), there is one striking difference: While [m(i)] murmurs were identified at chance level, [(i)m] murmurs were identified quite well (82% correct).

The ANOVAs showed the expected significant main effect of cutback [$F(7,77) = 14.02$, $p < 0.0001$; $F(7,28) = 12.17$, $p < 0.0001$]. The consonant by cutback interaction

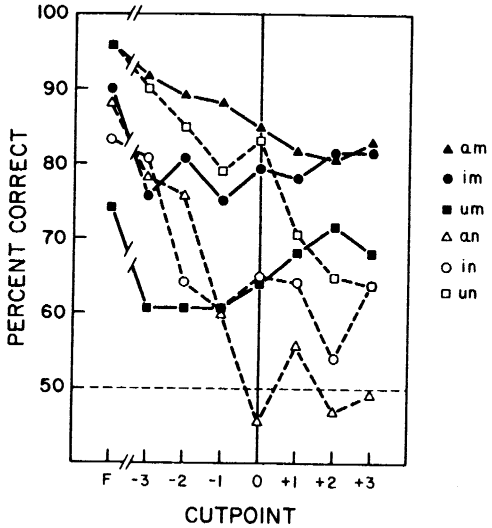

FIG. 3. Identification scores for individual syllables in the murmurs condition (experiment 1).

was also significant [$F(7,77) = 4.63$, $p = 0.0002$; $F(7,28) = 3.88$, $p = 0.0045$], which confirms the trend of [− n] syllables to be harmed more by truncation than [− m] syllables. Furthermore, there was a significant vowel by consonant interaction [$F(2,22) = 30.40$, $p < 0.0001$; $F(2,8) = 5.96$, $p = 0.0260$], reflecting the fact that [m] syllables showed the opposite effects of vowel context ([ɑ] > [i] > [u]) than did [n] syllables ([u] > [i] > [ɑ]). In the talker analysis, there was also a significant talker sex by cutback interaction [$F(7,28) = 2.74$, $p = 0.0268$], reflecting a reduction of the talker sex effect as the syllables were progressively truncated. A separate analysis of the isolated murmurs only (cutpoints + 1, + 2, + 3) showed neither the consonant main effect nor the consonant by vowel interaction to be reliable, due to large talker variability.

2. Vowels

The average results of the vowels condition are plotted in Fig. 2 as the filled triangles. The score for full syllables is on the right-hand side here, and it is even lower than that for the identical stimuli in the murmurs condition (80% correct). This poor accuracy was, in part, due to two subjects who, for unexplained reasons, performed at chance level with the full syllables, even though they did all right in the subsequent stimulus blocks of the vowels condition. If their data are omitted, the score rises to 85% correct.

Reading the graph in Fig. 2 from right to left, we see that elimination of all but 20 ms of the murmur (+ 2 cutpoint) left intelligibility unaffected, as it did also in the CV syllable experiment (Repp, 1986). Further truncation reduced identification performance gradually to 68% correct when the last 30 ms of the vowel were removed. Although performance seems to level off there, it presumably would have declined further, had the vowel been cut back more. The intelligibility score for truncated vowels (− 3 cutpoint) is comparable to that for isolated murmurs, and also to that for truncated vowels and isolated murmurs of CV syllables (Repp, 1986). Although it is difficult to compare results across experiments, the prediction that the formant transitions of VC syllables would be relatively more informative than those of CV syllables is not supported.

The data for individual syllables are shown in Fig. 4. Two syllables are clearly separated from the others here: Identification of [in] and [um] was poor to begin with and went to chance after elimination of the murmur. Not surprisingly, these are the syllables with minimal formant movements, due to the closeness of the places of articulation of vowel and consonant (cf. Repp, 1987). Truncated [ɑ] vowels yielded the highest consonant identification scores, presumably because they have the most pronounced formant transitions. The pattern is quite different from that for CV syllables (Repp, 1986): Both [u(m)] and [u(n)] were much more poorly identified than their CV counterparts, whereas [i(m)] was identified better than [(m)i].

The statistical analysis revealed significant main effects of cutpoint [$F(7,77) = 8.35$, $p < 0.0001$; $F(7,28) = 8.70$, $p < 0.0001$] and of vowel [$F(2,22) = 98.93$, $p < 0.0001$; $F(2,8) = 20.97$, $p = 0.0007$], reflecting the higher performance for [ɑ −] syllables, as well as an interaction between these two factors [$F(14,154) = 2.71$, $p = 0.0014$; $F(14,56) = 2.65$, $p = 0.0050$], due to the increase in the vowel effect as stimulus duration decreased. There was also a vowel by consonant interaction [$F(2,22) = 56.91$, $p < 0.0001$; $F(2,8) = 15.26$, $p = 0.0019$], reflecting the fact that [um] < [un] but [im] > [in], and a three-way interaction of these two factors with talker sex [$F(2,8) = 10.08$, $p = 0.0065$], since the vowel by consonant interaction was mainly due to the female talkers. With the murmurs of all stimuli cut back, it is difficult to attribute any remaining sex differences to the removal of final releases in female tokens, so they must have a different origin (see experiment 2).

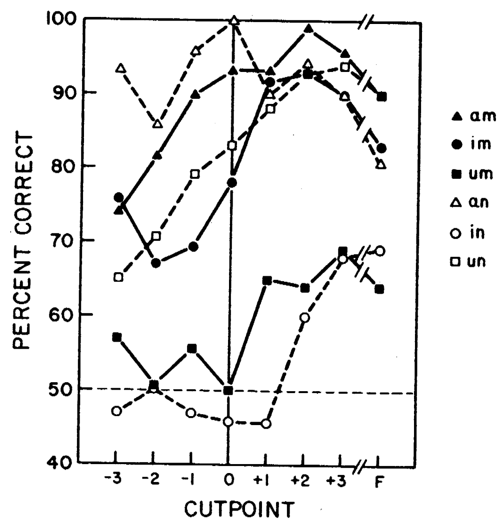

FIG. 4. Identification scores for individual syllables in the vowels condition (experiment 1).

3. Excerpts

The overall results for the excerpts condition are shown as the open triangles in Fig. 5. The left panel shows the effect of reducing the duration of excerpts from about 60 to 20 ms; the right panel shows the effect of varying the location of a 20-ms excerpt. As can be seen, performance for 60-ms excerpts was already poor but was reduced further when excerpt duration was shortened. Performance for 20-ms excerpts was only slightly above chance and apparently did not depend on location. These results contrast with those for CV syllable excerpts of the same duration, for which performance was not only generally higher, but also showed an advantage for short excerpts straddling the murmur–vowel boundary. The local spectral change across the closure point does not seem to be perceptually important in VC syllables.

Results for individual syllables are shown in Fig. 6. Scores for [un] and [in] were lower than those for the other syllables. Otherwise, the general conclusions hold for individual syllables. Only one syllable, [ɑn], showed a tendency for a peak in the center of the right-hand graph. Note that the pattern for vowel excerpts ($-3/-1$) approximates that for truncated vowels (Fig. 3), whereas that for murmur excerpts ($+1/+3$) resembles that for isolated murmurs (Fig. 4), with the characteristic advantage for [m] murmurs.

The ANOVAs on the three conditions of varying duration showed the expected main effect of duration [$F(2,22) = 7.67$, $p = 0.0030$; $F(2,8) = 11.01$, $p = 0.0050$] as well as a vowel main effect, [ɑ] > [i,u], [$F(2,22) = 25.98$, $p < 0.0001$; $F(2,8) = 10.15$, $p = 0.0064$]. The analysis of the five 20-ms excerpt conditions varying in location yielded main effects of consonant [$F(1,11) = 5.95$, $p = 0.0329$; $F(1,4) = 16.82$, $p = 0.0148$] and vowel [$F(2,22) = 6.48$, $p = 0.0061$; $F(2,8) = 7.00$, $p = 0.0175$], but no effect of location.

4. SCN

The overall results of the SCN condition are shown in Fig. 5 as the filled circles. The conditions in the left panel vary in SCN duration, whereas those in the right panel vary in SCN location. Neither variable, however, seemed to have much effect on performance, which was similar to that for full, unaltered syllables. The results for individual syllables are shown in Fig. 7. These functions, too, are rather flat, and the differences among syllables are similar to those found among full syllables (see Figs. 3 and 4). These results contrast with those for CV syllables, where perception of [mi] and [ni] was strongly affected by SCN.

ANOVAs on the duration data revealed no significant effect of that factor, only a vowel by consonant interaction [$F(2,22) = 45.56$, $p < 0.0001$; $F(2,8) = 6.63$, $p = 0.0200$], due to [ɑm] > [ɑn], [im] > [in], but [um] < [un]. Likewise, in the ANOVAs on the location data, there was no main effect of location but the same vowel by consonant interaction [$F(2,22) = 28.69$, $p < 0.0001$; $F(2,8) = 11.38$, $p = 0.0046$], plus a three-way interaction of these two factors with talker sex [$F(2,8) = 5.90$, $p = 0.0266$]; as usual, the differences were more pronounced for female talkers.

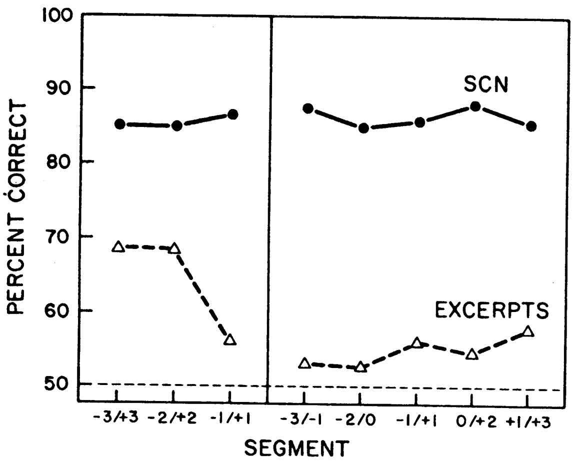

FIG. 5. Percent correct identification for syllables in the SCN and excerpts conditions (experiment 1) as a function of SCN or excerpt duration (left panel) and location (right panel).

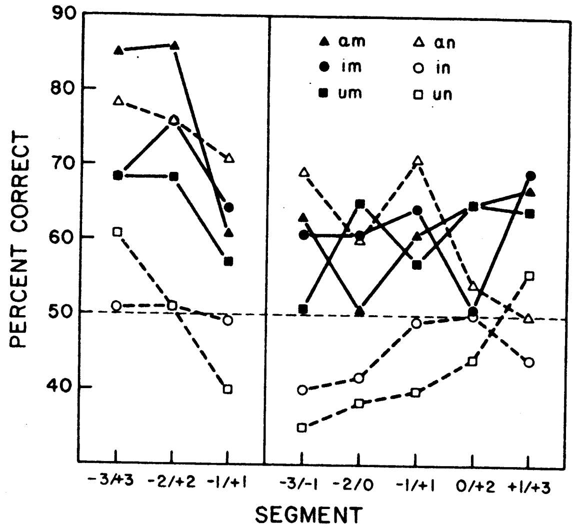

FIG. 6. Identification scores for individual syllables in the excerpts condition (experiment 1).

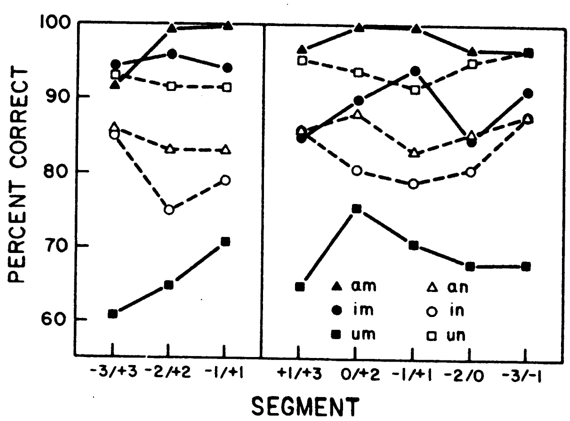

FIG. 7. Identification scores for individual syllables in the SCN condition (experiment 1).

C. Summary and conclusions

Even though full VC syllables (with releases removed) were less intelligible than CV syllables, identification scores were about the same for their respective components (murmur and vowel) presented in isolation. If the stimulus components were equally informative, why were full VC syllables less well identified than full CV syllables? The reason may be that relational spectral information is less salient in VC stimuli, because of the relatively more gradual transition from the vowel to the murmur. Listeners then have to rely mainly on integrating the information provided independently by vowel and murmur in VC syllables, whereas, in CV syllables, the relatively abrupt change in spectral structure across the release serves as an additional salient cue. Several observations are consistent with this suggestion. First, there was no perceptual advantage for brief excerpts that straddled the point of closure; thus, in contrast to CV syllables, having the point of maximal spectral change available did not aid identification. Second, substituting SCN for as much as 60 ms of waveform surrounding the closure point left performance virtually unaffected, suggesting that listeners continued to integrate information provided by (truncated) vowel and murmur across the noise. In CV syllables, by contrast, identification of the [mi]–[ni] distinction was severely affected by interpolated SCN. Third, the [im]–[in] distinction was generally perceived more accurately than the [mi]–[ni] distinction, due to much better identification of the labial consonant in VC syllables. There was reason to believe that relational spectral information is especially crucial for [mi] identification, whereas it seems to be much less important for [im] identification. (Why the *isolated* signal components are more informative for [im] than for [mi], however, remains a mystery.) Finally, Repp (1986) proposed a simple mathematical model of cue integration that predicts exactly the obtained overall score of 85% correct for full syllables, given two independent cues that lead, by themselves, to 70% and 75% correct identification, respectively (cf. Fig. 2). Thus, overall at least, VC identification performance seems more consistent with the hypothesis of independent cues than does CV identification performance.

The hypothesis that relational spectral information for place of articulation in nasals is less important in VC than in CV syllables will be pursued further in experiment 4.

II. EXPERIMENT 2

One potential problem with experiment 1 (and with experiments 3 and 4, which used the same stimulus materials) is the inclusion of both unreleased and originally released tokens, with their releases removed. The latter tokens were produced almost entirely by female talkers. Female speech turned out to be less intelligible than male speech. This could have been a genuine effect of talker sex—an interesting ancillary finding. Alternatively, however, the removal of the final release may have made originally released tokens less intelligible than originally unreleased ones, either because there are articulatory and acoustic differences between released and unreleased tokens preceding the release, or because of some waveform-editing artifact (e.g., abrupt offset of the murmur). This would constitute a methodological flaw.

Experiment 2 was a control study conducted with newly recorded materials after experiments 1, 3, and 4 had been completed. Its purpose was to determine (1) whether removal of final releases reduces intelligibility (as it most likely will), (2) whether the resulting releaseless tokens are less intelligible than unreleased tokens produced by the same talkers, and (3) whether female tokens are less intelligible than male tokens.

A. Methods

Ten new talkers, five males and five females, were recorded producing the same VC syllables as in experiment 1. Each talker was instructed to produce the syllables first with a final release (as demonstrated by the experimenter) and then without a release. All syllables were digitized, and their waveforms were inspected and carefully listened to, to make sure all tokens had been produced as intended. (The recordings of several additional talkers were rejected because they seemed to contain ambiguous tokens; this was especially evident for one female.) The closure point was marked in all syllables, following the same procedures as in experiment 1. The final release portion of the released (R +) tokens was located by eye and ear and removed to generate a set of releaseless (R −) tokens, to be compared with the originally unreleased (UR) syllables.

Three randomized test sequences were recorded. The first contained the full syllables (10 talkers $\times$ 6 syllables $\times$ 3 versions = 180 stimuli), the second the vowels only (120 stimuli, since R + and R − tokens were identical here), and the third the murmurs only (180 stimuli), with interstimulus intervals of 2.5 s. Ten subjects identified the stimuli as /m/ or /n/.

B. Results and discussion

The results are displayed in Table I. Percent correct scores are broken down by test (full syllables, vowels, murmurs), talker sex (male, female), release type (R +, R −, UR), and syllables. The right-most column lists the scores for all syllables combined.

The average scores for full syllables show that removal of the releases from released tokens resulted in an intelligibility decrease of about 3% for male and 10% for female talkers. (Because of ceiling effects, the decrement was not tested for significance.) Importantly, however, the R − tokens were no less intelligible than the UR tokens; in fact, they received slightly higher scores in male productions (a nonsignificant difference). In addition, male syllables were more intelligible than female syllables, though this difference was not quite significant across talkers [$F(1,9) = 32.01$, $p = 0.0003$; $F(1,9) = 4.45$, $p = 0.0679$]. Inspection of scores for individual talkers revealed that three female talkers were much less intelligible than the remaining seven talkers. Thus the conclusion is warranted that some, but not all, female-produced syllables were more ambiguous than male-produced syllables. Moreover, Table I shows that the sex difference for full syllables rests almost entirely on [u −] syllables.

Similar results were obtained for the isolated vowel portions. There was no significant difference between released

TABLE I. Percent correct identification of individual syllables in experiment 2 (R+ = released, R− = releaseless, UR = unreleased).

Sex	Type	[im]	[in]	[ɑm]	[ɑn]	[um]	[un]	Average
				Full syllables				
Male	R+	100	100	100	98	100	100	100
	R−	98	98	100	98	90	100	97
	UR	98	78	100	96	96	92	93
Female	R+	90	96	100	98	96	100	97
	R−	94	92	100	90	60	88	87
	UR	94	92	100	100	56	84	88
				Vowels				
Male	R	78	60	94	96	62	94	81
	UR	86	48	92	100	56	94	79
Female	R	66	50	92	98	46	74	71
	UR	70	50	70	96	38	76	67
				Murmurs				
Male	R+	98	86	98	70	100	76	88
	R−	88	74	78	70	80	70	77
	UR	66	48	90	58	62	50	62
Female	R+	82	100	98	92	100	96	95
	R−	58	86	78	68	50	76	69
	UR	60	82	70	80	42	86	70

and unreleased syllables; if anything, the former were slightly more intelligible. However, there was a significant talker sex effect in favor of male speech [$F(1,9) = 36.04$, $p = 0.0002$; $F(1,9) = 11.85$, $p = 0.0088$]. This difference was more or less present for all six syllable types.

Removal of the releases from isolated released murmurs resulted in a substantial intelligibility decrement [$F(1,9) = 100.83$, $p < 0.0001$; $F(1,9) = 26.19$, $p = 0.0009$], especially in the female tokens, although the two-way interaction was not significant across talkers. As to the comparison between R− and UR tokens, the former were the more intelligible in male speech, though the corresponding main effect and interaction were not significant across talkers. There was no significant talker sex effect for isolated murmurs.

The average scores for R− and UR tokens combined across all talkers were very similar to those obtained with analogous stimuli in experiment 1: 91% correct here versus 88% there for full syllables; 74% versus 75% for vowels; 70% versus 70% for murmurs.

In summary, these results vindicate the procedures of experiment 1. Although removal of the release from released final nasal consonants does impair their intelligibility, the resulting releaseless tokens are *not* less intelligible than tokens produced without any release. Rather, there seems to be a genuine difference in the intelligibility of male and (some) female talkers, which derives from the vowel rather than the murmur portion of the stimuli, as already suggested in experiment 1. The reason for this difference is unknown; one possibility is that anticipatory vowel nasalization is especially strong in some female talkers, and that this makes the formant transitions less salient. We may now proceed to experiments 3 and 4, which used the materials of experiment 1.

III. EXPERIMENT 3

The purpose of experiment 3 was to examine one particular explanation of the perceptual contribution of the nasal murmur to place of articulation perception (cf. Repp, 1987: experiment 5). In addition to providing direct spectral cues to place of articulation, the murmur serves as an important, and possibly essential, manner cue. Failure to perceive the correct nasal manner may interfere with place of articulation perception, since the precise acoustic structure of the vocalic formant transitions may deviate from that typically associated with oral stop consonants. Initial nasal consonants are often perceived as nonnasal stops when the murmur is removed (Kurowski and Blumstein, 1984; Repp, 1987), but for a stimulus of ambiguous manner, place of articulation identification is no better when manner is identified correctly than when it is not (Repp, 1987). This suggests that accuracy of place perception does not depend on oral/nasal manner perception in CV syllables. Experiment 3 examined the same issue in VC syllables. Because of the stronger nasalization of the vowel in this context, it was expected that removal of the murmur would interfere less with perception of nasal manner than it does in CV syllables.

A. Methods

Experiment 3 was run in two different versions, which will be referred to as 3a and 3b. The vowels tape of experiment 1 was reused in experiment 3a. That is, the subjects heard eight sequences of 36 stimuli each, and the syllables were truncated progressively from the end. The only difference was in the instructions: Whereas previously the subjects had to make a forced choice between /m/ and /n/, they were now encouraged to write down /m,n,b,d/ or any other consonant they heard, or a dash if they did not hear any consonant. A new group of 12 paid student volunteers was recruited.

Experiment 3b was motivated by the suspicion that the subjects in experiment 3a may have been biased against stop consonant responses by hearing full nasal syllables first. The instructions of experiment 3b emphasized even more that responses should reflect what was heard, not inferences about deleted consonants. Only stimuli truncated at points 0, −1, −2, and −3 (i.e., vowels without murmurs) were used in a completely randomized sequence. A new group of eight subjects served as listeners.

B. Results and discussion

The overall results are shown in Fig. 8 in terms of three response measures: the percentage of consonant responses, $p(C)$; the percentage of correct place of articulation identifications given that a consonant was heard, $p_c(P|C)$; and the percentage of nasal consonant responses given that a consonant was heard, $p(N|C)$. Consonant responses in experiment 3a dropped from 100% to about 80% as the syllables were cut back. In experiment 3b, fewer consonant responses were given, and truncation of the vowel seemed to have a stronger effect than in experiment 3a. The place identification scores were rather similar in the two experiments, falling from an initial 90% correct in full syllables (experiment

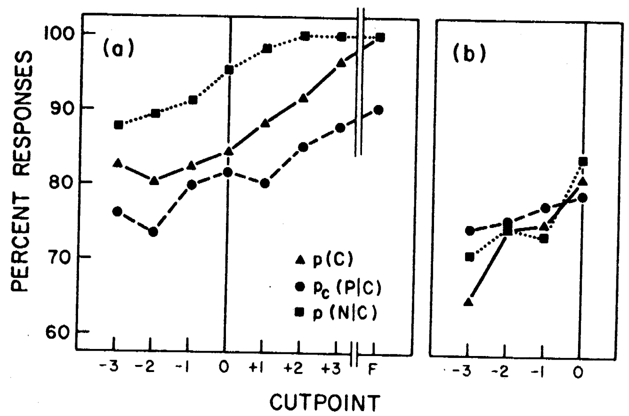

FIG. 8. Percentage of consonant responses (triangles), of correct place of articulation identifications given that a consonant was heard (circles), and of nasal responses given that a consonant was heard (squares). Panels (a) and (b) show results of experiments 3a and 3b, respectively.

3a) to about 75% correct at cutpoint − 3. (Considering that, in a two-alternative forced-choice task, random guesses on the remaining trials would increase the score to about 88% correct, the present subjects' performance was much better than that of the subjects in experiment 1, for reasons that are not clear.) Nasal consonant responses dropped only slightly with truncation in experiment 3a, to about 88%, but more sharply in experiment 3b, to 71%. This difference was probably a consequence of the changes in instruction and stimulus sequence.

Of the differences among individual syllables, the following are worth mentioning: (1) Consonant responses were less frequent for [in] and [um] than for the other syllables. A similar difference was observed for CV syllables by Repp (1987). It reflects the shallow formant transitions in these syllables, which have similar articulatory configurations for vowel and consonant. (2) Perception of nasal manner was poorest in [im] and [in], just as in CV syllables (Repp, 1987), probably because high vowels are generally less nasalized. Nasal consonant responses dropped by a substantial amount for these two syllables in experiment 3b. Even so, each VC syllable in experiment 3b received more nasal responses than the corresponding CV syllable in the analogous experiment of Repp (1987: experiment 5), which almost certainly reflects the greater nasalization of vowels preceding rather than following a nasal consonant.

It is evident from these results that removal of the nasal murmur interfered with the perception of nasality, but not very much. Therefore, only a small part of the improvement in place of articulation perception consequent upon the reinstatement of a deleted murmur could be due to the full restoration of nasal manner cues. An additional response measure that bears on this issue is the correct place of articulation score contingent on perceived nasal or non-nasal manner. If correct place perception depended on correct manner perception, then place identification should be more accurate in nasal than in oral consonant responses to the same stimuli. These percentages, computed from the average scores for each syllable at cutbacks 0, − 1, − 2, and − 3, and subsequently averaged over syllables and truncation conditions, were 76.2% (nasal) and 81.8% (oral) in experiment 3a and 74.5% (nasal) and 76.0% (oral) in experiment 3b. The latter values are more reliable because of the larger proportion of non-nasal responses in experiment 3b. Neither, however, gives any indication that perception of nasal manner aided place of articulation identification. A similar conclusion was reached for CV syllables (Repp, 1987).

In both parts of the experiment, listeners gave more labial responses when identifying stimuli as non-nasal rather than nasal, but this difference (5% and 10%, respectively, in experiments 3a and 3b) was much smaller than that obtained for CV syllables (Repp, 1987). For CV syllables, this criterion shift was attributed to the absence of stop consonant release bursts in prevocalic nasals, but this consideration does not apply to postvocalic nasals. Part of the bias, therefore, seems to have a different origin, perhaps deriving from differences between oral and nasal stop consonants in formant trajectories and/or vowel amplitude envelope.

In summary, this experiment suggests that the contribution of the murmur to place of articulation perception derives directly from spectral information in the murmur (i.e., its formants above 1 kHz), not indirectly from nasal manner cues. This outcome is consistent with the hypothesis that, in VC syllables, vowel and murmur constitute independent sources of place of articulation information.

IV. EXPERIMENT 4

Experiment 4 had two parts that were run in a single experimental session that included also experiment 3a. Their purpose was to obtain further evidence on the relative (un)importance of relational spectral information in the perception of postvocalic nasals. Each part resembled an earlier experiment with CV syllables (Repp, 1987).

A. Experiment 4a

This experiment corresponds to experiment 2 in Repp (1987). In that study, a decrement in CV identification performance was found when the murmur and vowel components were presented successively to opposite ears. This decrement was attributed to subjects' increased difficulty in utilizing relational spectral information across different channels. Since it has been hypothesized that such relational information is less important for the perception of VC syllables, the prediction is that channel separation will prove relatively less harmful to VC syllable identification. That is, even when vowel and murmur occur in different ears, listeners should be able to integrate these two sources of information centrally, and this may be all that is needed to identify final nasal consonants.

1. Methods

This short test consisted of 108 stimuli resulting from the randomization of three conditions with 36 syllables each. In one condition (V + M), vowel and murmur immediately followed each other on the same channel. In the second condition (V/M), the vowel occurred on one channel and the murmur on the other. In the third condition, the isolated vowel was presented without the murmur. All vowel portions occurred on the same channel, which was presented to the left ear for half of the subjects and to the right ear for the other half. The subjects made a forced choice between /m/ and /n/. Since no ear differences were apparent, the data of all subjects were combined.

Although it was not strictly necessary, a procedure used in the corresponding CV syllable experiment to avoid ceiling effects was followed here also: Both vowel and murmur were truncated by 30 ms (i.e., at cutpoints − 3 and + 3, respectively). Actually, this manipulation worked against the hypothesis under test: The truncation introduced a more abrupt spectral change between vowel and murmur than occurs normally in VC syllables.

2. Results and discussion

The results are shown in Table II. The overall score for the V + M syllables was 81% correct, and that for isolated vowels was 67% correct, in agreement with experiment 1. Performance in the split (V/M) condition was 77% correct, only slightly lower than in the V + M condition. In fact, this difference was not significant, whereas that between the V/M and V conditions was [$F(1,11) = 17.03$, $p = 0.0017$; $F(1,5) = 8.33$, $p = 0.0344$]. The pattern of differences among individual syllables is consistent with earlier results (see Fig. 4), except for the poor identification of [im] in isolated vowels.

These results are consistent with the prediction that channel separation of vowel and murmur has little effect on VC intelligibility. For CV syllables, by contrast, performance was significantly (8%) lower in the M/V than in the M + V condition (Repp 1987). The results thus support the hypothesis that the identification of final nasal consonants rests mainly on the integration of independent cues, with no significant contribution of relational spectral information.

B. Experiment 4b

This experiment corresponds to experiment 3 in Repp (1987) on CV syllables. It was intended to address the same hypothesis as the preceding experiment, viz., that spectral change information is relatively unimportant in VC syllables, but used a different technique to separate the murmur from the vowel—intervals of silence. The prediction was that temporal separation of the two signal components, within limits, should have little effect on intelligibility.

1. Methods

The vowel and murmur components were the same as in the preceding experiment; that is, they had been truncated by 30 ms to increase the error rate. Silent intervals of five durations were inserted between the two components: 0, 30, 60, 120, and 240 ms. Thus there were $5 \times 36 = 180$ stimuli, which were all randomized together. The subjects made a forced choice between /m/ and /n/ for each stimulus.

2. Results and discussion

The results are shown in Table III. It is evident that the effect of separating vowel and murmur by varying amounts of silence was minimal, as predicted. The effect of silence duration was nonsignificant in both ANOVAs. Although the effect of a similar manipulation on CV syllables (Repp, 1987) was smaller than expected, it was at least twice as large as the present effect and significant across subjects. Thus the results lend further support to the hypothesis that relational spectral information is less important in VC than in CV syllables.

V. SUMMARY AND CONCLUSION

The present series of experiments provides a variety of evidence in support of the hypothesis that relational spectral information is less important for place of articulation identification of nasal consonants in VC than in CV syllables. This evidence includes (1) lower identification scores for full VC than CV syllables despite comparable intelligibility of the isolated components (experiments 1 and 2), (2) absence of a peak in identification performance for brief excerpts straddling the closure point, and absence of a dip in performance when the same signal portions were replaced with SCN (experiment 1), (3) absence of a significant intelligibility decrement in split-channel presentation of vowel and murmur (experiment 4a), and (4) absence of a significant effect of temporal separation (of up to 240 ms) of vowel and murmur (experiment 4b). Although these results are mostly negative, they contrast with the stronger effects of similar manipulations in CV syllables. Since the spectral relationship between vowel and murmur seems to be unimportant in the perception of postvocalic nasals, and since the murmur does not appear to make its contribution to place of articulation perception *via* its function as a manner cue (experiment 2), the vowel and murmur essentially provide two independent sources of spectral information that are presumably integrated by the listener at a cognitive level (see Massaro and Oden, 1980).

Why are spectral relationships less important in VC than in CV syllables? Two factors may play a role. First, as already mentioned in the Introduction, spectral change is simply less pronounced and occurs at a slower rate in VC syllables, because of greater anticipatory nasalization, possible articulatory motion beyond the point of closure, and per-

TABLE II. Percent correct identification of individual syllables in experiment 4a (V = vowel, M = murmur, / = split channels).

Condition	[im]	[in]	[ɑm]	[ɑn]	[um]	[un]	Average
V + M	86	75	96	86	53	89	81
V/M	83	64	90	86	65	76	77
V	56	47	81	94	57	64	67

TABLE III. Percent correct identification of individual syllables in experiment 4b.

Silence	[im]	[in]	[ɑm]	[ɑn]	[um]	[un]	Average
0 ms	88	69	99	86	64	94	83
30 ms	85	81	94	92	60	92	84
60 ms	85	85	94	90	56	93	84
120 ms	85	75	90	86	65	92	82
240 ms	89	75	89	90	54	88	81

haps also slower closing than opening gestures. Second, there may be an inherent asymmetry in auditory sensitivity to direction of a change in level. The spectral change across the release in CV syllables is not only more abrupt but entails *increases* in the levels of most frequencies in the spectrum. The change in VC syllables across the point of closure, on the other hand, consists of a level *reduction* across most of the spectrum. Psychoacoustic research has shown that intensity decrements in pure tones are poorly detected by human infants and monkeys, even though adult humans detect them as well as intensity increments (Sinnott *et al.*, 1985; Sinnott and Aslin, 1985). One might speculate that critical phonetic learning occurs in humans before the auditory capacities necessary for intensity decrement detection are acquired. Moreover, in a relatively continuous sound, auditory short-term adaptation may reduce a listener's sensitivity to relatively smooth intensity decrements. Paradoxically, then, adaptation may not be responsible for the relative importance of relational spectral information in CV syllables (as hypothesized by Kurowski and Blumstein, 1984, but called into question by Repp, 1987) but rather for its lack of importance in VC syllables.

ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. We are grateful to Diana Matson for extensive assistance with experiment 2, to Hwei-Bing Lin for assistance with experiment 3b, and to all colleagues at Haskins Laboratories who served as talkers.

Ali, L., Gallagher, T., Goldstein, J., and Daniloff, R. (1971). "Perception of coarticulated nasality," J. Acoust. Soc. Am. 49, 538–540.

Broad, D. J., and Fertig, R. H. (1970). "Formant-frequency trajectories in selected CVC-syllable nuclei," J. Acoust. Soc. Am. 47, 1572–1582.

Kurowski, K., and Blumstein, S. E. (1984). "Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants," J. Acoust. Soc. Am. 76, 383–390.

Kurowski, K., and Blumstein, S. E. (1987). "Acoustic properties for place of articulation in nasal consonants," J. Acoust. Soc. Am. 81, 1917–1927.

Malécot, A. (1956). "Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique," Language 32, 274–284.

Massaro, D. W., and Oden, G. C. (1980). "Evaluation and integration of acoustic features in speech perception," J. Acoust. Soc. Am. 67, 996–1013.

Ohde, R. N., and Sharf, D. J. (1977). "Order effect of acoustic segments of VC and CV syllables on stop and vowel identification," J. Speech Hear. Res. 20, 543–554.

Ohde, R. N., and Sharf, D. J. (1981). "Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow-up study," J. Acoust. Soc. Am. 69, 297–300.

Ostreicher, H. J., and Sharf, D. J. (1976). "Effects of coarticulation on the identification of deleted consonant and vowel sounds," J. Phon. 4, 285–301.

Peterson, G. E., and Lehiste, I. (1960). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. 32, 693–703.

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

Pols, L. C. W., and Schouten, M. E. H. (1981). "Identification of deleted plosives: The effect of adding noise or applying a time window (A reply to Ohde and Sharf)," J. Acoust. Soc. Am. 69, 301–303.

Repp, B. H. (1986). "Perception of the [m]–[n] distinction in CV syllables," J. Acoust. Soc. Am. 79, 1987–1999.

Repp, B. H. (1987). "On the possible role of auditory short-term adaptation in perception of the prevocalic [m]–[n] contrast," J. Acoust. Soc. Am. 82, 1525–1538.

Schouten, M. E. H., and Pols, L. C. W. (1979). "CV- and VC-transitions: a spectral study of coarticulation—Part II," J. Phon. 7, 205–224.

Schroeder, M. R. (1968). "Reference signal for signal quality studies," J. Acoust. Soc. Am. 44, 1735–1736.

Sinnott, J. M., and Aslin, R. N. (1985). "Frequency and intensity discrimination in human infants and adults," J. Acoust. Soc. Am. 78, 1986–1992.

Sinnott, J. M., Petersen, M. R., and Hopp, S. L. (1985). "Frequency and intensity discrimination in humans and monkeys," J. Acoust. Soc. Am. 78, 1977–1985.