536

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

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

(Received 5 August 1985; accepted for publication 18 February 1986)

The contribution of the nasal murmur and the vocalic formant transitions to perception of the [m]–[n] distinction in utterance-initial position preceding [i,ɑ,u] was investigated, extending the recent work of Kurowski and Blumstein [J. Acoust. Soc. Am. **76**, 383–390 (1984)]. A variety of waveform-editing procedures were applied to syllables produced by six different talkers. Listeners' judgments of the edited stimuli confirmed that the nasal murmur makes a significant contribution to place of articulation perception. Murmur and transition information appeared to be integrated at a genuinely perceptual, not an abstract cognitive, level. This was particularly evident in [– i] context, where only the simultaneous presence of murmur and transition components permitted accurate place of articulation identification. The perceptual information seemed to be purely relational in this case. It also seemed to be context specific, since the spectral change from the murmur to the vowel onset did not follow an invariant pattern across front and back vowels.

PACS numbers: 43.71.Es

INTRODUCTION

In a recent study on the perceptual integration of nasal murmur and vocalic formant transition cues to place of articulation of nasal consonants, Kurowski and Blumstein (1984), henceforth K&B, showed that not only did both cues contribute to the perception of the [m]–[n] distinction, but also that their contributions were nearly equal. Their materials were 50 CV syllables uttered by a male speaker of American English, five tokens each of [m,n] followed by [i,e,a,o,u]. Portions of these syllables were presented to listeners as follows: (1) the full murmur (up to the point of consonantal release); (2) the full vowel[1] (i.e., the stimulus portion following the release, which included initial formant transitions); (3) the last six pitch pulses of the murmur; (4) the first six pitch pulses of the vowel; and (5) the last three pulses of the murmur followed by the first three pulses of the vowel (i.e., the six pulses surrounding the release). The principal findings were that (a) the full murmur and the full vowel were about equally informative when presented separately (about 80% correct place of articulation identification); (b) shortening of these stimulus portions to only six pitch pulses led to a nonsignificant decrease in identification scores (about 77% correct); and (c) scores were highest for stimuli that included both the end of the murmur and the beginning of the vowel (89% correct).[2]

Although it was known from earlier studies that the vocalic formant transitions are strong cues to the place of articulation in nasal consonants (e.g., Larkey *et al.*, 1978; Liberman *et al.*, 1954) and also that nasal murmurs in isolation can be identified at levels better than chance (Malécot, 1956; Nakata, 1959), K&B were the first to systematically compare identification of the two stimulus components in isolation and in combination. Their study contrasts with previous work by Malécot (1956), Nord (1976), and Recasens (1983), who used various combinations of conflicting murmurs and transitions to assess their relative contributions. With such stimuli, the transitions almost always emerge as the dominant place of articulation cue. K&B point out that this result could be due to artificial spectral discontinuities occurring at the splicing point, although the mechanism that would lead to perceptual dominance of the transitions over the murmur in such a situation has not been defined. (See Tartter *et al.*, 1983, for a similar argument concerning the perception of stop consonant place of articulation in VCV stimuli.) In any case, K&B avoided this possible problem by combining only murmurs and transitions deriving from the same utterance. This, however, resulted in an ambiguity of their results which they acknowledge: The murmur and the transitions could act as independent cues that are combined at some higher level of processing (cf. Massaro and Oden, 1980; Repp, 1982), or the murmur and the transitions might be integrated at an early perceptual level and thus might constitute a single effective cue. This second possibility was favored by K&B on grounds of parsimony and because it is more compatible with the search for invariant properties that Blumstein and her associates are engaged in (e.g., Blumstein and Stevens, 1979, 1980; Lahiri *et al.*, 1984). These two hypotheses may be called the multiple-cue (or late integration) and single-cue (or early integration) hypotheses, respectively.

The present experiment addressed several issues relevant to these hypotheses, as applied to nasal consonant perception, thereby extending the work of K&B. Although the study was mainly an attempt to replicate the results of K&B using a larger variety of test utterances and conditions, some of the conditions were novel and explored the nature of the perceptual integration process and the role of dynamic stimulus information.

Although K&B's study was carefully conducted, and incorporated five different vowel contexts, it had two methodological limitations. One is the use of a single talker: The surprisingly high identification scores for isolated murmurs could have reflected a peculiarity of his articulation. The other feature is that the subjects were permitted to respond with "b" and "d" (rather than "m" and "n") to the isolated vowel portions. While these stimuli indeed lacked nasal manner cues, the use of different response categories introduced a confounding factor. If it were the case that listeners applied slightly different criteria in place of articulation decisions for oral and nasal stop consonants (see Miller, 1977), then the scores for isolated vowel stimuli, containing acoustic information appropriate for nasal stops but being labeled as oral stops, may have been artificially depressed. It seemed important to rule out both of these possibilities, for they endanger the principal results and conclusions of K&B. The present study achieved this (1) by using six different talkers, at the price of sacrificing the assessment of within-talker variability and of using only three vowel contexts, and (2) by requiring a forced choice between "m" and "n" for all stimuli, at the price of creating a more restricted response situation.

In addition to these methodological changes, the present study expanded the range of techniques employed to assess the nature and distribution of the place of articulation information for nasal consonants. Five different waveform editing techniques were used, each with a number of gradations: (a) progressive truncation from the beginning of the syllable; (b) progressive truncation from the end; (c) extraction of brief segments from the vicinity of the consonantal release; (d) replacement of corresponding segments in the intact syllable with noise; and (e) elimination of dynamic spectral variation in short excerpts.

These techniques complemented each other in mapping out the temporal distribution of the acoustic cues that enable listeners to distinguish [m] and [n] in utterance-initial position. In particular, they provided additional information about the relative importance of perceiving the spectral change from the murmur into the vowel. Although K&B did not emphasize this point, it is clear from their approach that they considered spectral change as the basis for an invariant property associated with place of articulation (cf. Lahiri *et al.*, 1984). The gradual truncation conditions (a) and (b) assessed how much of the murmur or the vowel is needed to maintain accurate perception, and whether there is an abrupt drop in performance when one of these portions is removed altogether. The extraction condition (c) tested whether performance would be better for brief excerpts straddling the release (the point of maximal spectral change) than for excerpts of the same duration from within the murmur or vowel, thus partially replicating K&B. Conversely, the replacement condition (d) asked the same question by selectively replacing acoustic segments from within the syllable with noise, the prediction being that performance would be hurt most when the replaced segment included the point of release. An additional question of interest in that condition concerned subjects' ability to integrate murmur and vowel information across an intervening noise, allowing for the possibility of some form of perceptual restoration of the missing acoustic information (cf. Samuel, 1981; Warren, 1970, 1984; Whalen and Samuel, 1985). The final condition (e) explored the role of dynamic spectral change in the murmur and the vowel by concatenating steady-state murmur and vowel segments. The perceptual data were supplemented by an acoustic analysis of the stimuli, to determine any invariant correlate of the [m]–[n] contrast.

I. METHODS

A. Talkers and recording procedure

Six talkers participated, three males (AA, TG, SS) and three females (CG, SM, BT); all are native speakers of American English. AA is an experienced phonetician in his late fifties; the others are investigators or graduate students under 40 years of age.

The talkers were asked to produce the syllables [mɑ, mi, mu, nɑ, ni, nu] twice in that order, with similar intonation for all syllables. The recording session was deliberately informal and permitted a variety of speaking styles. The syllables were recorded using a Sennheiser microphone, placed approximately 10 in. from the talker's mouth, and a high-quality tape recorder.

B. Stimuli and test sequences

One good token of each syllable was selected from each talker's productions. The basic stimulus set thus consisted of 36 syllables (6 talkers × 6 utterances). These syllables were low-pass filtered at 4.9 kHz, digitized at a 10-kHz sampling rate, and stored in separate computer files. Using a waveform editing program, seven markers ("cutpoints") were subsequently placed in each file, as illustrated in Fig. 1. The marker labeled "0" was placed at the onset of the first pitch pulse following the point of release. This point was defined as a visible increase in high-frequency components in the oscillogram, as is clearly illustrated in Fig. 1; it could be located without difficulty in all tokens. In some syllables, it fell within a glottal cycle, as illustrated in the lower panel of Fig. 1. (This occasional contamination of what was, by definition, the last pitch pulse of the murmur must be kept in mind when interpreting the data.) Owing to the necessity of placing the markers at zero crossings, different criteria for the onset of a pitch period were used for male and female utterances, as shown in Fig. 1: In male waveforms, the marker was placed at a downgoing zero crossing, but in female waveforms, where the downgoing slope was often very steep, it was placed at the preceding upgoing zero crossing. No per-

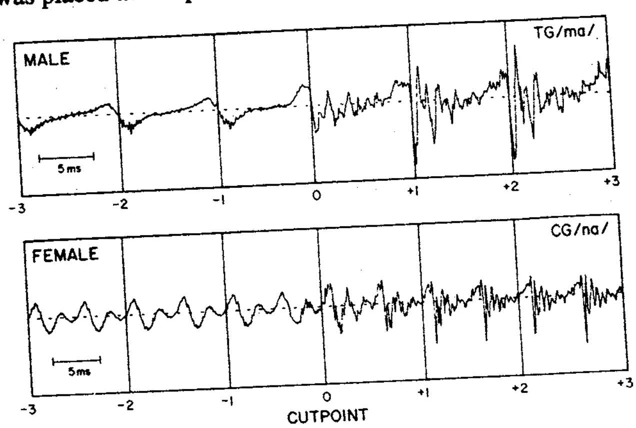

FIG. 1. Central portions of the waveforms of [mɑ] produced by a male talker (TG) and of [nɑ] produced by a female talker (CG). The figure illustrates the placement of cutpoint markers.

ceptual consequences of this difference were expected.[3]

The other six markers, labeled -3, -2, -1, $+1$, $+2$, and $+3$, were placed at corresponding locations at the onsets of the three preceding and following pitch periods in male utterances. In female utterances, with their higher fundamental frequencies, the pitch periods were treated in pairs, as illustrated in Fig. 1. (Thus the -3 marker, for example, was placed six pitch periods before the release.) The average durations of the intermarker intervals, calculated over the -2 to $+2$ range, and the corresponding fundamental frequencies for the six talkers were as follows: 10.3 ms, 97 Hz (AA); 8.9 ms, 112 Hz (TG); 10.4 ms, 97 Hz (SS); 10.4 ms, 193 Hz (CG); 10.9 ms, 183 Hz (SM); 10.5 ms, 190 Hz (BT). In the following discussion, the intermarker interval (also referred to as "segment duration") will be assigned a nominal duration of 10 ms.[4]

The set of 36 waveforms, with cutpoint markers in place, was used to generate a variety of test sequences. There were five test tapes, corresponding to the five parts of the experiment (a)–(e). Each tape contained between five and eight test sequences. Each test sequence consisted of a single randomization of the 36 syllables, with various modifications as described below. The interstimulus interval was 3 s; there were longer pauses between test sequences.

(*a*) *Truncation from the beginning* ("*Vowels*"). This tape contained eight test sequences. The first sequence contained the unaltered syllables, and the subsequent sequences presented the stimuli starting at cutpoints -3, -2, -1, 0, $+1$, $+2$, and $+3$, in that order.

(*b*) *Truncation from the end* ("*Murmurs*"). This tape also contained eight test sequences. The first sequence contained the unaltered syllables, and the subsequent sequences presented the stimuli up to cutpoints $+3$, $+2$, $+1$, 0, -1, -2, and -3, in that order. It should be noted here that the murmur portions varied widely in duration, ranging from 46–223 ms, with an average duration of 103 ms.[5] Thus there was little left of some murmurs in the most extreme truncation condition.

(*c*) *Extraction of brief segments* ("*Excerpts*"). This tape contained seven test sequences presenting the following excerpts: $-3/+3$ (i.e., from cutpoint -3 to cutpoint $+3$), $-2/+2$, $-1/+1$, $-2/0$, $0/+2$, $-3/-1$, and $+1/+3$. Thus 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 release, whereas those in sequences 4–7 came from within the murmur (4,6) or the vowel (5,7).

(*d*) *Replacement of segments with signal-correlated noise* ("*SCN*"). This tape contained seven test sequences, with the replaced excerpts being $+1/+3$, $-3/-1$, $0/+2$, $-2/0$, $-1/+1$, $-2/+2$, and $-3/+3$ (the reverse order of the Excerpts tape). Thus, the stimuli in sequences 1–5 contained 20 ms of noise, those in sequence 6 contained 40 ms, and those in sequence 7 contained 60 ms of noise. A computer program was used to generate signal-correlated noise (SCN) frcm specified segments within a waveform by randomly reversing the polarity of digital sampling points with a probability of 0.5. This results in noise which retains the amplitude envelope of the original signal but is

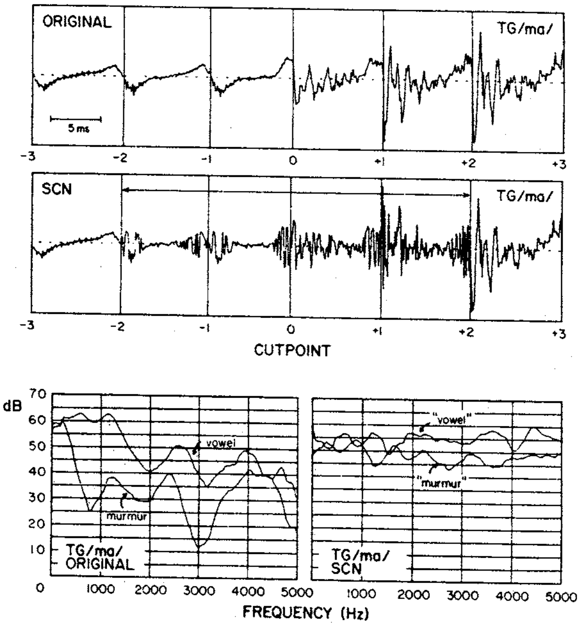

FIG. 2. Central portion of the waveform of [mɑ] produced by a male talker (TG) in its original form (top panel) and after the four glottal periods between cutpoints -2 and $+2$ were replaced with signal-correlated noise (SCN) (center panel). The bottom panels show smoothed Fourier spectra of the "murmur" ($-2/0$) and "vowel" ($0/+2$) portions before and after replacement with SCN.

spectrally uniform (Schroeder, 1968). An example is shown in Fig. 2. The top panels compare the waveforms of the central portions of a male [mɑ] in its original form and after the $-2/+2$ segment was replaced with SCN (as in test sequence 6). Below on the left are the smoothed Fourier spectra of the $-2/0$ (murmur) and $0/+2$ (vowel onset) segments. Note the pronounced spectral peaks and the differences between murmur and vowel spectra. On the bottom right are the spectra of the corresponding SCN segments. It is evident that the spectral difference between murmur and vowel is erased; both the murmur-derived and the vowel-derived SCN have flat spectra with random fluctuations due to the short time window. Only the difference in absolute amplitude remains, though it is reduced due to the conversion of low-frequency into wide-band energy, especially in the murmur segment.

(*e*) *Elimination of dynamic spectral variation* ("*Static Excerpts*"). This final part of the experiment was exploratory in nature and included five test sequences. Artificial steady-state murmurs and vowels (i.e., prolonged vowel onsets) were constructed by iterating the penultimate segment ($-2/-1$) of the murmur and the first segment of the vowel ($0/+1$), respectively.[6] In the first test sequence, three repetitions of the murmur segment (i.e., three male or six female pitch pulses) were followed by three repetitions of the vowel segment. In sequences 2 (murmurs) and 3 (vowels), these 30-ms components were presented in isolation; and in sequences 4 (murmurs) and 5 (vowels), the static murmurs and vowel onsets were extended to 60 ms (i.e., six iterated segments). The artificial vowel segments, being prolonged onsets, had phonetic qualities different from the original [i, ɑ, u].

C. Subjects and procedure

The subjects were 12 paid volunteers, mostly Yale undergraduates. Because of time constraints, two subjects could not listen to the last test tape (e). Ten of the subjects were native speakers of American English; the remaining two were native speakers of Russian and Chinese, respectively, but fluent in English. Their results did not differ systematically from those of the other subjects.

The tapes were played back at a comfortable intensity over TDH-39 earphones in a quiet room. Each subject listened to all tapes (with the two exceptions just noted) in a single session lasting about 100 min. The order of the vowel, murmur, and SCN conditions was counterbalanced across subjects. The excerpts always followed these three conditions, and the static excerpts were last. This was done because the excerpts conditions were considered the most difficult. There were short rest periods between test tapes.

Within each condition, the test sequences were presented in the order in which they had been recorded, as described above. This order generally proceeded from easy to difficult, so the earlier sequences provided practice for the later ones.[7]

The subjects' task was to label in writing each stimulus as beginning with m or n; or, if the stimulus did not sound like it contained a nasal consonant, to guess whether it was derived from a [m-] or [n-] syllable. In no case was identification of the vowel required. The subjects were told that there were a number of different talkers, that there was an equal number of [m-]-derived and [n-]-derived stimuli in each test sequence, and that all stimuli had been constructed from a single basic set. In the vowels condition, the subjects were alerted to the fact that the stimuli in the later sequences might be perceived as beginning with an oral stop or with no consonant at all. (The correspondence of b and m, and of d and n, was explained.) In the murmurs condition the subjects were warned about the short duration of some stimuli in the later sequences. Preceding the presentation of each test tape, the stimulus manipulation was explained in nontechnical terms.

D. Statistical analysis

The data of each condition (or a subset thereof) were subjected to two kinds of repeated-measures analysis of variance (ANOVA): In one ("across subjects"), correct responses were added up over the six talkers, and subjects constituted the random factor, with consonant, vowel, and segment duration and/or location as fixed factors. In the other analysis ("across talkers"), correct responses were added up over the 12 (or 10) subjects, and talkers constituted the random factor, with talker sex as an additional fixed factor. Results from both analyses will be reported, since a genuine effect should generalize to both listener and talker populations. Of the two F values reported for each effect, the first is across subjects and the second is across talkers.

E. Acoustic analysis

To track spectral peaks over time and from the murmur into the vowel, a standard LPC analysis (ILS package distributed by Signal Technology, Inc.) was performed on all syllables, using 14 coefficients and a 20-ms analysis window moving in 10-ms steps. The ILS peak-picking routine was used to estimate formant frequencies. In addition, Fourier spectra of precisely specified time intervals were computed using another ILS program.

II. RESULTS AND DISCUSSION

A. Vowels

The overall results for the vowels condition (truncation from the beginning) are shown as the solid function in Fig. 3. It can be seen that identification of the full, unaltered syllables (F) was nearly perfect (99% correct). Elimination of the murmur (cut at 0) reduced performance to 85% correct, and truncation of the vowel onset reduced scores even more. However, performance was still significantly above chance when the first 30 ms of the vowel were excised (cut at $+3$); the remainders of the formant transitions thus still contained some usable place of articulation cues. Two aspects of these data deserve comment.

First, elimination of all but the last 20 ms of the murmur (cut at -2) reduced scores only slightly (to 96% correct); and the presence of only 10 ms of murmur (cut at -1) produced significantly better performance ($p < 0.001$, sign test across subjects) than no murmur at all (cut at 0). Although the identifiability of 10-ms murmur segments in isolation was not tested and may conceivably be better than chance, their significant contribution is more plausibly attributed to an enhancement of transition perception than to any independent cue value of the murmur segment itself. This interpretation is consistent with K&B's hypothesis of a single integrated auditory property for nasal place of articulation. However, the advantage could also be attributed to the availability of sufficient nasal manner cues: In the author's informal judgment, the majority of the syllables cut at 0 sounded as if they began with oral stops (see also K&B's Table IV), whereas all syllables cut at -1 were perceived as beginning with nasal stops. Perception of the correct manner may have enhanced perception of the place of articulation cues.

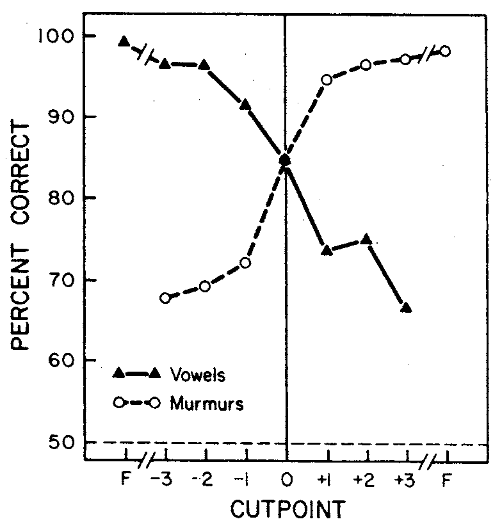

FIG. 3. Percent correct identification scores as a function of stimulus duration in the vowels and murmurs conditions. F stands for "full syllable."

Second, the score of 85% correct for isolated full vowels (cut at 0) is not unlike that obtained by K&B in their "long transitions" condition (80% correct), which confirms that the formant transitions provide strong but not entirely sufficient cues to place of articulation. The use of nasal rather than oral consonant responses in the present study did not seem to make a substantial difference.

These overall results need to be qualified in view of large differences among individual syllables, which are shown in Fig. 4. It is evident that identification of nasal consonants was much poorer in [i] context than in [ɑ] and [u] contexts, as also observed by K&B. Identification of [mi] and especially [ni] suffered much more than the other syllables from truncation of the murmur, and at cutpoints beyond + 1 the two syllables could not be discriminated at all. Thus the formant transitions, especially beyond the first pitch pulse of the vowel, did not provide salient place cues in [i] context. The syllable [ni], in addition, seemed to require at least 20 ms of murmur to be identifiable. The data also replicate K&B's finding that [n] was identified more accurately than [m] from transitions in back vowels, while the reverse was true for the front vowel [i]. The difference in back vowel contexts can be explained in terms of transition length, reflecting distances traversed by the tongue in moving from the occlusion to the anticipated vowel configuration.

To avoid ceiling effects, only the data for cutpoints 0 and beyond (i.e., for isolated vowel stimuli) were entered into the ANOVAs, which yielded four significant effects: A main effect of duration [$F(3,33) = 18.23$, $p<0.0001$; $F(3,12) = 13.45$, $p = 0.0004$], reflecting the decline in performance with increasing vowel truncation; a main effect of vowel [$F(2,22) = 67.83$, $p<0.0001$; $F(2,8) = 58.79$, $p<0.0001$], reflecting mainly the poorer scores for [i]; a consonant by duration interaction [$F(3,33) = 4.88$, $p = 0.0065$; $F(3,12) = 6.91$, $p = 0.0059$], indicating that [m] identification was hurt more by vowel truncation than was [n] identification;

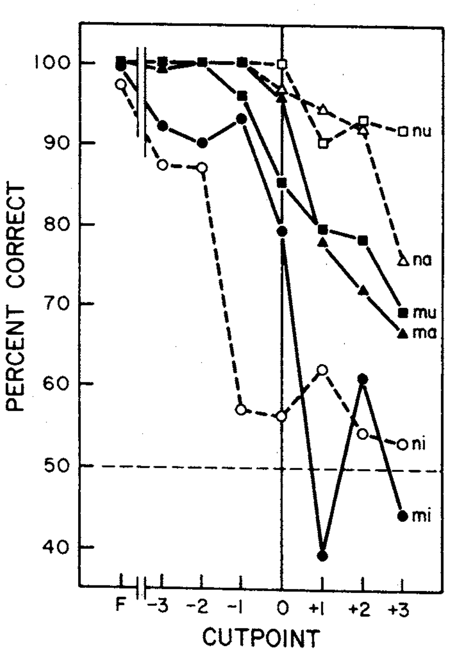

FIG. 4. Individual syllable scores in the vowels condition.

and a consonant by vowel by duration interaction [$F(6,66) = 4.41$, $p = 0.0008$; $F(6,24) = 2.82$, $p = 0.0320$], mainly due to the large advantage of [mi] over [ni] in the "0" cutpoint condition, where the consonant by vowel interaction described above (though it was not significant overall) was most pronounced.

Acoustic analysis of the vocalic stimulus portions revealed patterns that matched the perceptual findings. The syllables [mɑ] and [nɑ] were consistently distinguished by the second formant ($F2$), whose onset was 400–600 Hz higher in [nɑ] than in [mɑ]. The syllables [mu] and [nu] showed even larger differences in $F2$ onset, although $F2$ peaks could not be located reliably in three talkers' tokens of [mu]. In both [ɑ] and [u] vowels, the $F2$ differences persisted well beyond the first 50 ms following the release, which explains the above-chance identification of truncated vowels. The syllables [mi] and [ni], by contrast, were only minimally distinct at vowel onset. There were no indications of any difference in $F2$; instead, $F3$ and $F4$ onsets appeared to be somewhat higher for [ni] than for [mi]. These small differences, moreover, tended to disappear soon after the release, which explains the vulnerability of [i] vowels to truncation. All these observations are consistent with those on formant transitions in initial [b] and [d] preceding [i, ɑ, u] (Fant, 1973; Kewley-Port, 1982).

B. Murmurs

The overall results for the murmurs condition (truncation from the end) are represented by the dashed line in Fig. 3. Reading the graph from right to left, it is evident, first, that reduction of the vowel to its initial 10 ms (cut at + 1) had little effect on identifiability of the consonant (94% correct). (Indeed, to the author these stimuli sound remarkably natural, like released nasal consonants.) This confirms that significant place-of-articulation information is located at the very onset of the vowel, immediately following the release, as has also been observed in connection with oral stop consonants (Blumstein and Stevens, 1980; Kewley-Port *et al.*, 1983).

Complete elimination of the vowel portion (cut at 0) resulted in a clear drop in performance to 85% correct, the same score as for isolated vowels, and only slightly higher than K&B's score of 81% correct for their "long murmurs." At first blush, therefore, the results seem to replicate K&B's finding that, on the whole, isolated murmurs and vowels carry about the same amount of place of articulation information. It must be kept in mind, however, that the last pitch pulse of the murmur was "contaminated" with incipient high-frequency energy in some syllables. Indeed, elimination of the final 10-ms segment of the murmur (cut at − 1) led to a further substantial reduction in performance, to 72% correct. By contrast, when K&B eliminated the final pitch pulses of their isolated murmurs in a control study, performance stayed the same, which suggests that their stimuli had uncontaminated offsets. (For a possible reason see footnote 3.) Therefore, the score of 72% correct is a better estimate of the intelligibility of the full isolated murmurs in the present study. Unless it is argued that the first pitch pulses of the vowel contained extra place cues due to residual nasalization and therefore should be excluded also, the conclusion must be that, overall, isolated vowels were more informative than isolated murmurs ($p<0.001$, sign test across subjects). Nevertheless, identification scores for isolated murmurs were clearly above chance, which confirms K&B's general observation that these signal portions contain useful place of articulation information, probably throughout their duration.

There were large differences among individual syllables, however, which are shown in Fig. 5. As in the vowels condition, scores for [mi] and [ni] were generally lower than those for other syllables. Thus it is not the case that the nondistinctive formant transitions in [i] are compensated for by more intelligible murmurs. Regarding the intelligibility of isolated murmurs (cut at − 1, − 2, − 3), it seems that the differences were almost exclusively among [m] murmurs, with [m(ɑ)] best and [m(i)] worst, whereas [n] murmurs from different vocalic contexts were identified about equally well. (K&B also found that [m(i)] murmurs were much more poorly identified than [m(ɑ)] and [m(u)] murmurs, and that [n(ɑ)] and [n(u)] scores were the same; in other respects their results were different.) Interestingly, the pattern found here is consistent with considerations from the acoustic theory of speech production: First, because of the fixation of the tongue tip during alveolar but not labial closure; lingual anticipation of the following vowel will be more evident in [m] murmurs than in [n] murmurs (see Hecker, 1962). Second, the acoustic effect of the oral shunt on the nasal murmur spectrum will be greater when the tongue body is low (as in [m(ɑ)]) than when it is high (as in [m(i)]), in proportion to the degree of coupling of the oral and nasal-pharyngeal cavities (see Kitazawa and Doshita, 1984). For these reasons, [m(ɑ)] may be expected to contain the most salient place of articulation cues, followed by [m(u)] and [n] murmurs, while the elevated tongue body during [m(i)] may in fact make this murmur more [n]-like than the [n] murmurs.

The data for uncontaminated isolated murmurs (cut at − 1, − 2, − 3) were submitted to ANOVAs, which yielded three significant effects: A main effect of vowel [$F(2,22)=36.83, p<0.0001$; $F(2,8)=6.92, p=0.0180$], reflecting mainly the lower scores for [-(i)] murmurs; a consonant by vowel interaction [$F(2,22)=13.45$, $p=0.0002$; $F(2,8)=4.76, p=0.0435$], reflecting the presence of a vowel effect for [m] but not for [n] murmurs; and a consonant by duration interaction [$F(2,22)=6.31$, $p=0.0068$; $F(2,8)=5.00$, $p<0.0389$], which apparently derives from the fact that [n] murmurs, but not [m] murmurs, suffered from the excision of the penultimate pitch pulse (cut at − 1 vs − 2).[8] The lower F values in the ANOVA across talkers indicate considerable talker variability in nasal murmur spectra, a well-known phenomenon often commented on in the literature (e.g., Fant, 1960; Fujimura, 1962; Glenn and Kleiner, 1968). The unpredictable nature of that variability, as compared to the somewhat more regular scaling differences for oral resonances, may also have been responsible for the overall difference in scores between isolated murmurs and vowels in the present mixed-talker design. The subjects of K&B, of course, had to cope only with a single talker's utterances.[9]

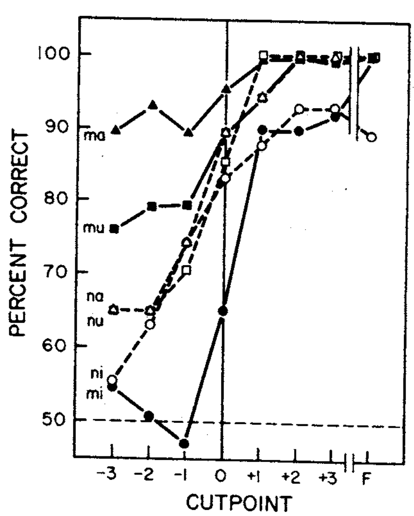

FIG. 5. Individual syllable scores in the murmurs condition.

Acoustic analysis of the nasal murmurs revealed that, in [mɑ] and [nɑ], the $F2$ differences observed at vowel onset were contiguous with similar $F2$ differences in the murmur. In other words, murmurs preceding [ɑ] generally showed distinct spectral peaks between 1 and 2 kHz, which were at least 600 Hz higher for [n] than for [m]. Although K&B did not report such a difference for their talker's [− a] murmurs, it is consistent with the acoustic theory of speech production, which predicts a lower oral resonance for [m] than for [n] (Fant, 1960; see also Saito and Itakura, 1984). Similar differences in $F2$ frequency tended to be present in [mu] and [nu] murmurs, though less clearly and less consistently. (See also K&B.) Differences in [mi] and [ni] murmurs were least systematic and showed large individual differences. These observations agree well with the perceptual data and the articulatory considerations presented above.

C. Excerpts

We turn next to the excerpts condition, which partially replicates the study of K&B. The overall results are shown as the open triangles in Fig. 6. The data have been divided into two parts. On the left we see the effect of reducing the length of excerpts centered on the release from 60 to 20 ms. It can be seen that performance was quite accurate for 60- and 40-ms durations (which replicates K&B), but reduction to 20 ms resulted in a substantial decline in performance, though scores remained far better than chance. On the right in Fig. 6 we see the effect of moving the location of a 20-ms excerpt across the release; the data point for − 1/ + 1 segments is duplicated here. There was a clear peak in performance for the − 1/ + 1 excerpts, which enclosed the release. The results thus replicate K&B's finding that identification of "mixed" excerpts is more accurate than that of equal-duration murmur or vowel ("transition") excerpts, even though the present excerpts were shorter than K&B's. Performance for 20-ms murmur excerpts (− 3/ − 1, − 2/0) was only slightly below that for vowel excerpts (0/ + 2, + 1/ + 3), which is also consistent with K&B's findings.

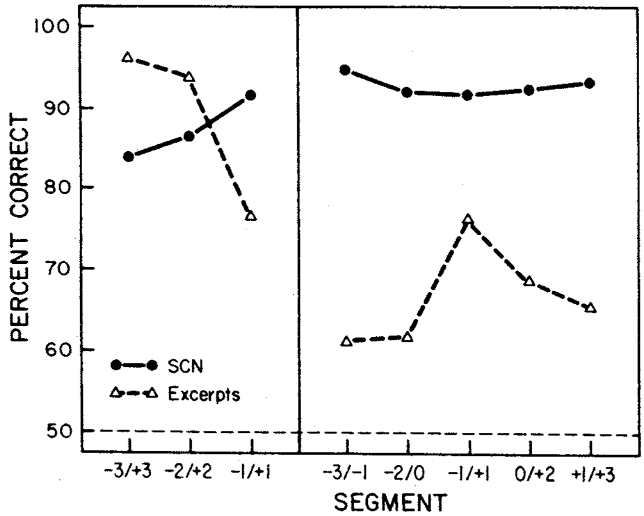

FIG. 6. Percent correct identification scores in the excerpts and SCN conditions. The left panel shows the effect of (excerpted or replaced) segment duration; the right panel shows the effect of moving a segment of constant duration across the point of release. The $-1/+1$ data points are duplicated in the two panels.

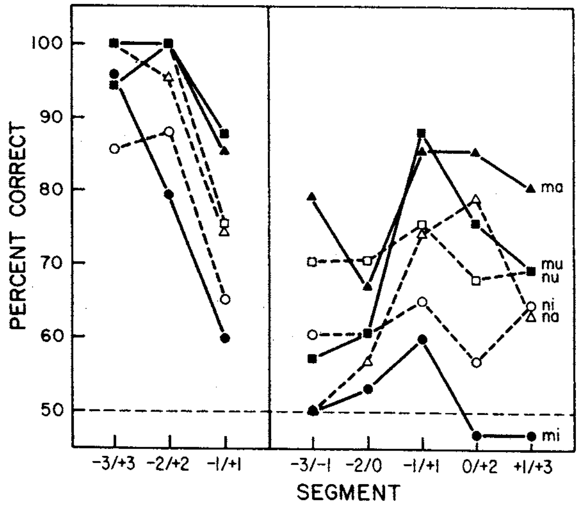

FIG. 7. Individual syllable scores in the excerpts condition.

The results for individual syllables are shown in Fig. 7. Syllables including [u] and [i] all showed a tendency for 20-ms excerpt scores to peak at $-1/+1$; for [mɑ] and [nɑ], equivalent scores were obtained for $-1/+1$ and $0/+2$ (vowel onset) excerpts. The rank ordering of the different syllables as vowel excerpts ($0/+2$, $+1/+3$) was not very similar to that of full isolated vowels (Fig. 4: 0, +1 cutpoints), which suggests a role of the transitions beyond the initial 30 ms. The pattern for murmur excerpts ($-3/-1$, $-2/0$) was more similar to that for full isolated murmurs (Fig. 5: -1, 0 cutpoints), especially for [m] murmurs.

The data for 20-ms excerpts were submitted to ANOVAs, which yielded three significant effects: A main effect of vowel [$F(2,22) = 20.07$, $p < 0.0001$; $F(2,8) = 25.05$, $p = 0.0004$], due to the poor performance for [– i] syllables; a main effect of location [$F(4,44) = 4.98, p = 0.0021$; $F(4,16) = 5.10, p = 0.0076$], which confirms the better performance for segments straddling the release; and a consonant by vowel interaction [$F(2,22) = 21.66$, $p < 0.0001$; $F(2,8) = 6.54$, $p = 0.0207$], reflecting the different vowel effects for [m-] and [n-] syllables. The vowel by location interaction alluded to above (in connection with the equivalence of $-1/+1$ and $0/+2$ scores for [– ɑ] syllables only) was marginally significant across subjects [$F(8,88) = 2.12, p = 0.0420$] but not across talkers.

To gain some insight into the nature of the spectral information that enabled listeners to identify place of articulation in brief excerpts straddling the release, the patterns of *spectral change* from the murmur into the vowel were examined, in the hope that they would reveal distinctive and context-insensitive patterns for [m] and [n] (cf. Lahiri *et al.*, 1984). To quantify the change in the whole spectrum across the release, the difference between the raw Fourier spectra of the end of the murmur ($-2/0$) and of the onset of the vowel ($0/+2$) was computed for each syllable. These difference spectra are shown in Fig. 8, separately for the six syllables, with the six talkers' curves superimposed. Despite considerable talker variability, fairly typical patterns of spectral change can be seen, particularly in the region between 1–3 kHz. For [mɑ] and [mu], there is less relative energy increase from the murmur into the vowel around 2–2.5 kHz than at 1 kHz, leading to a negative slope of the difference spectrum in that region, whereas [nɑ] and [nu] difference spectra tend to have flat or rising slopes in the same region. Thus [m] and [n] in these back vowel contexts have distinctive patterns of spectral change across the release, which largely reflect the different $F2$ onset frequencies and the concomitant amplitude increase in the vowel. The difference spectra for [ni], with generally rising slopes between 1–3 kHz, also fit this pattern; those for [mi], however, besides being highly variable, are quite different, having the most steeply rising slopes of all. The difference spectra for [mi] and [ni] differ somewhat in their slopes, which may provide a (rather unreliable) context-dependent cue for this contrast. There is no indication in these data, however, of any

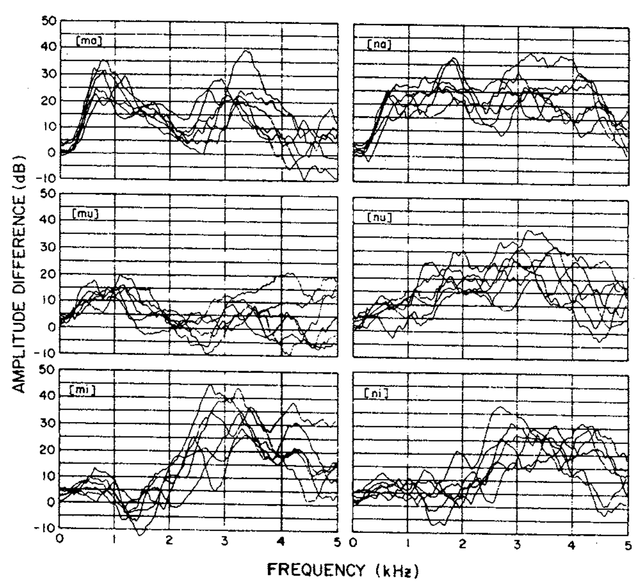

FIG. 8. Raw Fourier difference spectra of 20-ms vowel ($0/+2$) and murmur ($-2/0$) segments of six syllables produced by six talkers. The difference spectra of the six talkers are superimposed in each panel.

invariant spectral change property distinguishing [m] and [n] across all vocalic contexts.

D. Summary of vowels, murmurs, and excerpts results

The results from the three conditions discussed so far essentially confirm the findings of K&B, and they dispel any reservations about their generality across different talker populations and testing procedures. K&B's main findings, that murmurs and transitions both contribute to place of articulation identification (except perhaps in [– i] context) and that performance is best when both components are represented in a stimulus, were replicated. Their observation that murmurs and transitions in isolation are about equally identifiable was confirmed for brief excerpts, although in longer stimuli there seemed to be a certain advantage for the transitions, particularly when the vowel was [ɑ]. More significantly, perhaps, the intelligibility rank order of individual syllables was quite different for isolated murmurs and vowels, in a way that could be related to acoustic properties of the stimuli. The very poor intelligibility of both stimulus components in [– i] syllables was noted, although these syllables were identified quite well when both components were present. The spectral change across the release does not seem to provide an invariant correlate of place of articulation, though it may serve as a context-dependent cue.

E. Signal-correlated noise (SCN)

In this condition, it will be recalled, brief segments of the waveform in the vicinity of the release (corresponding to those presented in the excerpts condition) were replaced with SCN, thus rendering these segments spectrally uninformative. Figure 6 shows the overall results (filled circles). Consider first the right-hand panel, where the effect of removing various 20-ms segments is shown. The question of interest here was whether replacement of the 20-ms segment straddling the release (– 1/ + 1) would have a more detrimental effect than replacement of a 20-ms segment from within the murmur or the vowel. It can be seen that, compared to the near-perfect scores for intact syllables (Fig. 3), performance was somewhat reduced in all SCN conditions, but there was no clear tendency for scores to be lowest in the – 1/ + 1 condition. This contrasts with the clear peak obtained for the excerpts. In the left-hand panel of the figure, which should be read from right to left for the SCN data, the effect of extending the SCN segment from 20–60 ms is shown. This manipulation resulted in a moderate decline in performance, but scores were still surprisingly high in the 60-ms SCN (– 3/ + 3) condition (84% correct).

The scores for individual syllables are shown in Fig. 9. Some striking differences are evident: [mɑ] and [nɑ] were not affected at all by SCN, not even in the most extreme condition, and [mu] and [nu] were affected only slightly in the 60-ms condition. The [mi] and [ni] syllables supplied virtually all the errors. Both of these syllables were substantially affected even by 20-ms segments of SCN, but while identification of [ni] remained above chance when the SCN segment was extended to 60 ms, identification of [mi] went to chance. There was also a difference in pattern for the two syllables: [mi], but not [ni], showed a tendency for performance to be lowest when the 20-ms SCN segment straddled the release.

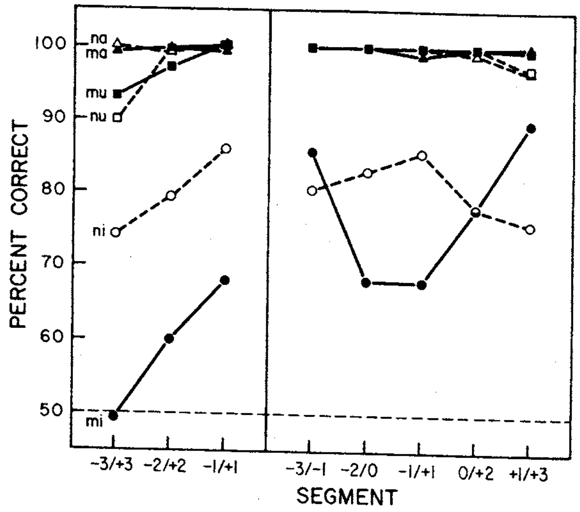

FIG. 9. Individual syllable scores in the SCN condition.

Only the 20-ms data for the [mi] and [ni] syllables were submitted to ANOVAs, which yielded one significant effect: The consonant by location interaction just described [$F(4,44) = 4.85, p = 0.0025$; $F(4,16) = 6.28, p = 0.0031$]. In the ANOVA across talkers, there was also a marginally significant effect of talker sex [$F(1,4) = 8.14, p = 0.0463$], due to higher error rates for female speech.

F. A simple model of "late" information integration

The remarkably high performance for [– ɑ] and [– u] syllables in all SCN conditions, as well as the absence of a specific drop in performance when the 20-ms segment straddling the release was replaced with SCN (except for [mi]), raise some interesting questions about the nature of perceptual integration in these stimuli. When the murmur is immediately followed by the transitions, listeners have the opportunity to establish the single auditory property that, according to K&B's early integration hypothesis, underlies place of articulation perception. Since such auditory integration processes are likely to have a relatively short time window (a few tens of milliseconds, see Blumstein and Stevens, 1980), they should not operate across intervening noise whose duration exceeds the integration span and which, moreover, may enter into and distort the product of integration. The excellent recognition of [mɑ] and [nɑ] when as much as 60 ms of SCN was present, therefore, cannot have been due to a very early integration process.

That some form of integration nevertheless took place is clear from a comparison of SCN identification scores with those for the murmur and vowel portions preceding and following the noise, obtained in the murmurs and vowels conditions of the experiment. For example, the average score for [nɑ] in the 60-ms (– 3/ + 3) SCN condition was 100% correct, whereas that for the isolated murmur component (cut at – 3) was 65% correct, and that for the isolated vowel component (cut at + 3) was 76% correct. Clearly, the listeners cannot have relied on one or the other component alone; they must have combined information from the two sources in the SCN condition. (See Whalen and Samuel, 1985, for a similar result.)

It is conceivable that this integration occurred at a rather late stage in perception. Such a late integration process might evaluate each source of information separately and then combine the results according to some probabilistic rule, much as proposed by Massaro and Oden (1980). The well-known model of these authors, however, is formulated for designs in which two or more cues are varied factorially; it cannot be applied directly to experiments in which two cues are presented separately and in combination. A very simple "late integration" model may be devised for this situation, however, based on the following assumptions: (a) A stimulus component either provides "correct" information for the phonetic segment intended by the talker, with a certain probability, or it provides none at all, in which case the listener makes a random guess (i.e., we exclude the possibility that a cue reverses polarity due to some manipulation); (b) when two components are present, a listener will respond correctly when either component provides correct information (i.e., it is not necessary that both of them do). This second assumption is conservative and predicts a maximal benefit from the presence of two independent sources of information, thus counteracting the hypothesis to be tested shortly, viz., that actual performance is even better than predicted by this model.

Expressed in more formal terms, the probability of giving a correct response to an isolated murmur component is assumed to be

$$P_m = p_m + 0.5(1 - p_m), \quad (1a)$$

and similarly for an isolated vowel component,

$$P_v = p_v + 0.5(1 - p_v), \quad (1b)$$

where P_m and P_v are the observed response proportions, while p_m and p_v are probabilities reflecting the information content of each component. We wish to predict from P_m and P_v the correct response proportion when both components are present, P_{mv}. Since an incorrect response will result only when neither component is informative, and then only in half of the instances because of random guessing between two alternatives, we find that

$$P_{mv} = 1 - 0.5(1 - p_m)(1 - p_v). \quad (2)$$

From Eqs. (1a) and (1b) we can derive that $p_m = 2P_m - 1$, and $p_v = 2P_v - 1$, which may be substituted into Eq. (2). After some simplification, this yields

$$\hat{P}_{mv} = 1 - 2(1 - P_m)(1 - P_v), \quad (3)$$

which is the sought-after prediction formula.

We can now attempt to predict the results for murmur-vowel stimuli from the results for isolated murmur and vowel components (even though averaging of scores over subjects and talkers may introduce some distortion in the calculations). If the obtained scores, P_{mv}, match the predicted scores, $\hat{P}_{mv}$, we may conclude that integration of murmur and vowel information took place at a late stage. If $\hat{P}_{mv}$ scores exceed P_{mv} scores, on the other hand, some more direct, more "perceptual" kind of integration would be indicated.

TABLE I. Percentage differences between obtained scores P_{mv} and predicted scores $\hat{P}_{mv}$ for individual syllables in four conditions.

	Syllables					
Conditions	[mi]	[ni]	[mɑ]	[nɑ]	[mu]	[nu]
Full	14	8	0	1	3	0
SCN (− 1/ + 1)	33	6	4	3	9	5
SCN (− 2/ + 2)	− 2	13	4	5	6	4
SCN (− 3/ + 3)	1	15	6	17	8	− 4

Table I presents the difference scores, $P_{mv} - \hat{P}_{mv}$, for individual syllables in four conditions: Full syllables (scores averaged over the replications of this test in the murmurs and vowels conditions) and SCN syllables with 20 ms, 40 ms, and 60 ms of noise centered over the release (− 1/ + 1, − 2/ + 2, − 3/ + 3). The P_m and P_v scores for the predictions come from the murmurs (0, − 1, − 2, − 3) and vowels (0, + 1, + 2, + 3) conditions, respectively. A positive difference score thus means that the obtained score exceeded the predicted one. It is evident from Table I that the difference scores are mostly positive and quite large in some instances. (Exceptions are full [− ɑ] and [− u] syllables, for which predicted scores were very high, and [mi] in the SCN conditions, for which all scores were very low. The large difference score for [mi] in the − 1/ + 1 condition may be an abnormality, since below-chance performance was predicted.) Moreover, there is no clear trend for difference scores to decrease as the SCN increased in duration. This leads to the tentative conclusion that some form of early perceptual integration did occur, not only when murmur and vowel followed immediately upon each other (as hypothesized by K&B), but also when as much as 60 ms of noise intervened.

What could account for this perceptual integration across such a relatively wide interval? One possibility is that the murmur spectrum somehow survives in auditory memory, not being masked by the following noise, so that auditory integration still occurs when the vowel begins. Another possibility is that the acoustic information replaced by the noise is somehow reconstituted in the listener's perceptual system from long-term knowledge of acoustic-phonetic properties of speech, in a manner akin to the "phonemic restoration" phenomenon (see Samuel, 1981; Warren, 1970, 1984; Whalen and Samuel, 1985) so that perceptual integration of the filled-in information with the actual input becomes possible. Yet other possibilities, of course, are that the simple model applied in this section is based on faulty assumptions, or that isolation of stimulus components changes their acoustic properties in ways that make predictions of the sort attempted here inappropriate. We will return to this last issue in Sec. III.

G. Static excerpts

The final condition of the experiment, it will be recalled, examined the contribution of dynamic spectral change within the murmur and particularly within the vowel (the formant transitions) by presenting steady-state signal components generated by iterating one (male) or two (female)

TABLE II. Percent correct scores for the static excerpts condition. M = murmur segment (− 2/ − 1), V = vowel segment (0/ + 1).

Conditions	Syllables [mi]	[ni]	[mɑ]	[nɑ]	[mu]	[nu]	Average
3M	62	63	70	58	65	68	64
3V	38	67	73	80	80	72	68
6M	52	47	68	58	58	47	55
6V	52	55	68	67	67	60	62
3M + 3V	50	92	100	98	100	95	89
$P_{mv} - \hat{P}_{mv}$	− 3	16	16	15	14	13	

pitch periods. At the same time, the design of the static excerpts condition replicated rather closely the conditions employed by K&B. The questions of interest were whether concatenation of a static murmur and a static vowel onset would enable listeners to identify the nasal consonants accurately, and how scores in that condition would compare with those for stimuli containing dynamic changes and those for isolated static murmurs and vowels.

The results are presented in Table II. Looking first at the 3M + 3V results, we see that the average score for these 60-ms murmur–vowel stimuli (89% correct) was only slightly lower than that for the corresponding dynamic (− 3/ + 3) stimuli in the excerpts condition (96% correct). Moreover, it is immediately evident that this reduction was entirely due to the syllable [mi], which could not be identified at all in static excerpts. Identification of the other five syllables was basically unaffected by removal of dynamic information. This result indicates that the formant transitions, at least during the first 30 ms of the vowel, made no important contribution to perception of the [m]–[n] distinction. Rather, the onset spectrum of the vowel seemed to convey the distinctive information.

The poor intelligibility of [mi] in static excerpts is puzzling because the formant transition cues for that syllable seemed to be ineffective to begin with. However, the abrupt decline of [mi] scores consequent upon truncation of the first vowel segment in the vowels condition (see Fig. 4) does indicate a perceptual role of a very short-term spectral change cue. Specifically, the vowel onset may contain a spectral transient due to the parting of the lips, whose relationship to the following vowel spectrum is perceptually important in the case of [mi]. This would also be consistent with the sensitivity of [mi] to replacement of pitch periods in the vicinity of the release with SCN, even though replacement of the − 2/0 segment was even more detrimental than replacement of the 0/ + 2 segment (see Fig. 8). Finally, the result is also consistent with the reciprocal relation of the perceptual salience of release bursts and formant transitions noted in stop consonants (Dorman *et al.*, 1977): The very ineffectiveness of the [mi] formant transitions may make even a very weak transient perceptually useful.

Turning now to the remaining four static excerpts tests in Table II, it is clear that performance for these isolated steady-state murmur and vowel onset stimuli was rather poor. Scores were somewhat higher for vowel than for murmur stimuli, and scores surprisingly declined as segment durations increased from 30–60 ms. This latter effect may have been due to the artificial spectral homogeneity of the stimuli, which may have become increasingly apparent to listeners as duration increased.

The data for these four tests were entered into ANOVAs with segment duration and location as crossed factors, which yielded two significant effects: A main effect of vowel [$F(2,18) = 15.20, p = 0.0001$; $F(2,8) = 8.21, p = 0.0115$], due to poorer performance for [− i] syllables, and a main effect of duration [$F(1,9) = 6.22$, $p = 0.0342$; $F(1,4) = 16.66, p = 0.0151$]. The main effect of location [$F(1,9) = 3.40$, $p = 0.0982$; $F(1,4) = 12.79$, $p = 0.0232$], which compared murmur and vowel stimuli, was significant only across talkers. In the talker analysis, there was also a significant talker sex by vowel interaction [$F(2,8) = 4.96$, $p = 0.0398$]: Overall, female speech accounted for more errors in [− i] and [− u] contexts and for fewer errors in [− ɑ] context than male speech.

Finally, let us compare in Table II the scores for isolated static components of 30-ms duration (3M, 3V) with the scores obtained when these components were concatenated (3M + 3V). This comparison is analogous to that conducted by K&B, and it is clear that performance benefited enormously from the presence of both components, except in the case of [mi]. The bottom row in Table II shows that the increase was considerably larger than predicted by the late integration formula derived in the preceding section (except for [mi]), which suggests that perceptual integration, perhaps of the kind discussed by K&B, did indeed occur in these artificial stimuli.

H. Summary of SCN and static excerpts results

These conditions yielded some interesting findings which add to those of the first three conditions and of K&B. The SCN conditions and their analysis by means of a simple late integration model suggested that genuinely perceptual integration occurs not only when the murmur and vowel components are contiguous, but also when they are separated by as much as 60 ms of noise. While this supports K&B's general notion of a single perceptual cue, it casts doubt on their specific hypothesis that the perceptual integration takes place at an early auditory level. The static excerpts results showed that, although dynamic spectral change beyond the vowel onset, such as formant movements, may contribute place-of-articulation information, this information is generally not necessary for correct identification. The syllable [mi] followed a different pattern, however, and both [mi] and [ni] were much more vulnerable to SCN than the other syllables, which suggests that the place-of-articulation information in [− i] context is of a different kind than that in [− ɑ] and [− u] contexts.

III. GENERAL DISCUSSION

The present experiment was stimulated by the recent findings of K&B that the nasal murmur and the vocalic formant transitions make about equal contributions to the perception of the [m]–[n] distinction in CV syllables. In their study K&B used a single talker and permitted stop consonant responses when nasal manner cues were absent in the stimuli. The present study, which used six talkers and required a forced choice between m and n responses throughout, essentially confirmed the findings of K&B, although place of articulation information in the murmur seemed somewhat less salient than that in the formant transitions.

It was K&B's hypothesis that murmur and transitions constitute a single integrated property in the auditory system, which may provide invariant perceptual information about place of articulation.[10] As to the invariant nature of this property, the present study does suggest that formant movements contribute relatively little to perception of the [m]–[n] distinction, which paves the way for an invariant measure of spectral change from the murmur to the vowel onset. Such a simple measure, however, proved to be invariant (if at all) only across the two back vowel contexts, [ɑ] and [u]; a very different criterion seems to be required to distinguish [m] and [n] in [– i] context. Indeed, it may be that spectral change cues are really important only in that context, where neither component suffices by itself.[11] It remains to be seen whether more sophisticated indices of spectral change can be found that remain more nearly invariant across different vocalic contexts.

The K&B hypothesis of a single integrated auditory property for the place of articulation was supported by the present findings in so far as they suggested that the integration process does not (exclusively) take place at an abstract level of information integration. However, the listeners' apparent ability to perform such truly perceptual integration across an intervening noise (cf. Whalen and Samuel, 1985) makes it difficult to conceive of the process as a purely auditory one. At the very least, an auditory memory for spectral information must be invoked, together with an ability to reject or "listen through" noninformative noise. Although it is auditory information that is perceptually integrated, the integrative function itself should perhaps not be characterized as being auditory in nature. Indeed, it may well be specific to speech perception (Repp, 1982; see also footnote 10).

One strictly auditory process that probably does play a role in the perception of nasal consonants is short-term neural adaptation (see, e.g., Harris and Dallos, 1979). Kurowski and Blumstein (also, Blumstein and Stevens, 1979) specifically refer to Delgutte's (1980; Delgutte and Kiang, 1984) neurophysiological studies of cats, which show that a nasal murmur adapts auditory neurons in the low-frequency range, so that the response of these neurons to the onset of a following vowel is reduced. Although there is little reason to doubt that such internal high-pass filtering of the vowel onset does occur in human listeners, it seems unlikely that this process can account fully for the perceptual integration observed. First, although short-term adaptation may extend over 100 ms or more (Harris and Dallos, 1979; Delgutte, 1980), it may not be sufficiently strong after a 60-ms intervening noise to have much of an effect on the auditory representation of the vowel onset. Second, and more importantly, the subtraction of murmur from vowel onset spectra (Fig. 8) essentially approximates (perhaps overestimates) the high-pass filtering caused by auditory adaptation; as we have seen, no invariant property emerged from this exercise. The role of auditory adaptation nevertheless deserves continued attention: Neither K&B nor the present author took this effect into account when presenting vowel portions in isolation. One may well argue that the intelligibility of these stimulus components was reduced because not only the preceding murmur but also its auditory aftereffect had been removed. Perhaps, if the aftereffect were simulated by high-pass filtering the onsets of isolated vowels, their intelligibility would improve so much that the scores for concatenated murmur and vowel components would no longer exceed the predictions of a late integration model, or might even equal those for isolated vowels. This possibility is currently under investigation.

There are two reasons why high-pass filtering of vowel onsets may improve the identification of place of articulation. First, a number of studies have shown that the first formant transition may interfere somewhat with the accurate registration of higher formant transitions, so that a benefit may accrue from attenuation of $F1$ (e.g., Danaher and Pickett, 1975; Hannley and Dorman, 1983). Second, reduction of $F1$ energy may also lead to increased perception of nasal manner (e.g., Delattre, 1954), which in turn may enhance the identification of nasal consonant place of articulation. Indeed, although K&B considered place of articulation perception apart from manner perception, an important confounding factor in their study as well as in the present one was that isolated vowel stimuli were generally perceived as beginning with oral, not nasal stops. Even if the perceptual criteria pertaining to spectral correlates of place of articulation in the vowel were the same for oral and nasal stops (and they are at least very similar; see Miller, 1977), the periodic stimulus portion following a nasal stop release lacks the abrupt onset and release burst characteristics of oral stop consonants (except perhaps in [mi]). Thus even though it may be perceived as beginning with an oral stop in isolation, it is not a "good" oral stop, and this may affect identification of place of articulation. Addition of the murmur restores perception of the correct manner class, which in itself may be responsible for at least part of the improvement in identification scores. It would be useful to dissociate manner and place perception in future research, not only by simulating low-frequency auditory adaptation but also perhaps by examining nasal consonants in the context of nasal vowels.

To conclude, while this study represents a significant extension of the work of K&B, it by no means settles all the issues raised by their work. To gain a better understanding of nasal consonant perception, future studies will have to take into account models of peripheral auditory processing, consider possible interactions of manner and place perception, and conduct a more extensive search for invariant acoustic properties.

ACKNOWLEDGMENTS

This research was supported by NICHD grant HD-01994 and BRS grant RR-05596 to Haskins Laboratories. Some results were reported at the 108th meeting of the Acoustical Society of America in Minneapolis, MN, October 1984.

[1]K&B used the term long transitions for this stimulus portion. That formant transitions often extend beyond the initial 60 ms or so is illustrated by K&B's footnote 1, which reports [a] second-formant frequencies almost 300 Hz higher following [n] than [m] "around the center of the vowel well past the formant transitions" (K&B, p. 389). See also Kewley-Port (1982) for analogous observations on stop consonants.

[2]The study did not include a condition in which the full, unaltered syllables were presented for identification. By using truncated murmurs and vowels, K&B (who did not motivate this choice) presumably wanted to emphasize the concentration of place-of-articulation information around the release. However, a comparison of identification scores for full murmurs and vowels (about 80% correct) with those for full syllables (surely better than 90% correct) would have led to very similar conclusions.

[3]Apparently K&B even placed their markers in the middle of glottal cycles (see their Fig. 1, left-hand panel).

[4]A repeated-measures ANOVA was conducted on the intermarker intervals in the -2 to $+2$ range, with the factors before/after release, consonant, and vowel. There were no significant effects in this analysis, showing in particular that (1) $F0$ did not change abruptly at the release, and (2) $F0$ did not differentiate [m] and [n].

[5]A repeated-measures ANOVA was conducted on the murmur durations, with the factors consonant and vowel. There were no significant effects. Individual differences among talkers were considerable, however: Average murmur durations ranged from 70 to 152 ms, and standard deviations ranged from 10 to 43 ms.

[6]As pointed out earlier, the last murmur segment ($-1/0$) sometimes contained incipient high-frequency energy from the release; this is why the preceding murmur segment was used for iteration. The iteration of two pitch pulses in the female tokens did not result in noticeable fluctuations of timbre.

[7]This arrangement differs from that employed by K&B, who presented diverse stimuli in a single randomized sequence. The present design, with homogeneous blocks of stimuli graded according to difficulty, favored the most difficult conditions, thus working against the perceptual integration advantage resulting from the simultaneous availability of murmur and transition cues. Such an advantage was nevertheless obtained, which suggests that practice effects were negligible. Another important departure from K&B's design is the use of multiple talkers, which may have increased the difficulty of all identification tasks.

[8]An unexpected difference between male and female talkers was noted in the 0 and $+1$ truncation conditions, which were not included in the ANOVAs: The average scores of both conditions were 98, 98, and 94% correct for the three male talkers, and 90, 90, and 87% correct for the three female talkers. The cause of this difference is unknown. Note that there were no effects of talker sex for either isolated murmurs or isolated vowels.

[9]Another possibility considered was that the rather short durations of some of the murmurs employed here were responsible for the lower murmur identification scores. The average murmur duration (103 ms) was only slightly less than that in the K&B study (117 ms), but variability was much larger. However, inspection of the data revealed that, although the shortest murmurs did not receive very high scores, many long murmurs yielded scores that were equal or even poorer. Murmur duration was entered as a covariate into an analysis of covariance, which yielded results similar to the ANOVA together with a pooled regression coefficient of -0.01, indicating that murmur duration did not account for any significant variation in the data.

[10]When K&B say that "the auditory system does not treat transitions separately from the murmur" (p. 389), do they mean to imply that listeners would not be able to discriminate a stimulus with initial murmur frcm one in which the murmur has been deleted and the vowel onset has been modified acoustically (by some kind of high-pass filtering) to simulate the effect of the murmur on the auditory response at vowel onset? This prediction should be easy to disconfirm, for the murmur is easily detectable as a separate auditory event. If their statement is to be interpreted as meaning that, *as a cue to place of articulation*, the murmur and the transitions form a single integrated property, then they must mean that the integration is a speech-specific, not a general auditory function.

[11]In a perceptual study with synthetic speech, Carlson *et al.* (1972) found that the frequency of the second nasal formant during the murmur was critical for the [mi]–[ni] distinction. The present data offer little support for this observation.

Blumstein, S. E., and Stevens, K. N. (**1979**). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. **66**, 1001–1017.

Blumstein, S. E., and Stevens, K. N. (**1980**). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**, 648–662.

Carlson, R., Granström, B., and Pauli, S. (**1972**). "Perceptive evaluation of segmental cues," Speech Trans. Lab. (Stockholm) **QPSR 1**, 18–24.

Danaher, E. M., and Pickett, J. M. (**1975**). "Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss," J. Speech Hear. Res. **18**, 261–271.

Delattre, P. (**1954**). "Les attributs acoustiques de la nasalité vocalique et consonantique," Stud. Ling. **8**, 103–108.

Delgutte, B. (**1980**). "Representation of speech-like sounds in the discharge patterns of auditory nerve-fibers," J. Acoust. Soc. Am. **68**, 843–857.

Delgutte, B., and Kiang, N. Y. S. (**1984**). "Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics," J. Acoust. Soc. Am. **75**, 897–907.

Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (**1977**). "Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues," Percept. Psychophys. **22**, 109–122.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, G. (**1973**). "Stops in CV–syllables," in *Speech Sounds and Features* (MIT, Cambridge, MA), pp. 110–139.

Fujimura, O. (**1962**). "Analysis of nasal consonants," J. Acoust. Soc. Am. **34**, 1865–1875.

Glenn, J. W., and Kleiner, N. (**1968**). "Speaker identification based on nasal phonation," J. Acoust. Soc. Am. **43**, 368–372.

Hannley, M., and Dorman, M. F. (**1983**). "Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **74**, 40–51.

Harris, D. M., and Dallos, P. (**1979**). "Forward masking of auditory nerve fiber responses," J. Neurophysiol. **42**, 1083–1107.

Hecker, M. H. (**1962**). "Studies of nasal consonants with an articulatory speech synthesizer," J. Acoust. Soc. Am. **34**, 179–188.

Kewley-Port, D. (**1982**). "Measurement of formant transitions in naturally produced stop consonant–vowel syllables," J. Acoust. Soc. Am. **72**, 379–389.

Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (**1983**). "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants," J. Acoust. Soc. Am. **73**, 1779–1793.

Kitazawa, S., and Doshita, S. (**1984**). "Nasal consonant discrimination by vowel independent features," Stud. Phon. **18**, 46–58.

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.

Lahiri, A., Gewirth, L., and Blumstein, S. E. (**1984**). "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study," J. Acoust. Soc. Am. **76**, 391–404.

Larkey, L., Wald, J., and Strange, W. (**1978**). "Perception of synthetic nasal consonants in initial and final syllable position," Percept. Psychophys. **23**, 299–312.

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (**1954**). "The role of consonant–vowel transitions in the perception of the stop and nasal consonants," Psychol. Mono. **68**.

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.

Miller, J. L. (**1977**). "Nonindependence of feature processing in initial consonants," J. Speech Hear. Res. **20**, 519–528.

Nakata, K. (**1959**). "Synthesis and perception of nasal consonants," J. Acoust. Soc. Am. **31**, 661–666.

Nord, L. (**1976**). "Perceptual experiments with nasals," Speech Transmission Lab. (Stockholm) **QPSR 2-3**, 5–8.

Recasens, D. (**1983**). "Place cues for nasal consonants with special reference to Catalan," J. Acoust. Soc. Am. **73**, 1346–1353.

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.

Saito, S., and Itakura, F. (**1984**). "Further study on the frequency spectrum deviations between speakers," Ann. Bull. RILP **18**, 107–112.

Samuel, A. G. (**1981**). "The role of bottom-up confirmation in the phonemic restoration illusion," J. Exp. Psychol. **7**, 1124–1131.

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

Tartter, V. C., Kat, D., Samuel, A. G., and Repp, B. H. (**1983**). "Perception of intervocalic stop consonants: The contributions of closure duration and formant transitions," J. Acoust. Soc. Am. **74**, 715–725.

Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**, 392–393.

Warren, R. M. (**1984**). "Perceptual restoration of obliterated sounds," Psychol. Bull. **96**, 371–383.

Whalen, D. H., and Samuel, A. G. (**1985**). "Phonetic information is integrated across intervening nonlinguistic sounds," Percept. Psychophys. **37**, 579–587.