

On the possible role of auditory short-term adaptation in perception of the prevocalic [m]-[n] contrast

Bruno H. Repp

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

(Received 23 February 1987; accepted for publication 8 July 1987)

Acoustic information about the place of articulation of a prevocalic nasal consonant is distributed over two distinct signal portions, the nasal murmur and the onset of the following vowel. The spectral properties of these signal portions are perceptually important, as is their relationship (the pattern of spectral change). A series of experiments was conducted to investigate to what extent relational place of articulation information derives from a peripheral auditory interaction, viz., short-term adaptation caused by the murmur. Experimental manipulations intended to disrupt the effects of such adaptation included separation of the murmur and the vowel by intervals of silence, presentation to different ears, and reversal of order. Other tests of the possible role of adaptation included manipulation of murmur duration, murmur-vowel cross splicing, and high-pass filtering of the excised vowel onset. While the results of several experiments were compatible with the peripheral adaptation hypothesis, others did not support it. An alternative hypothesis, that the manner cues provided by the murmur are crucial for accurate place judgments, was also discredited. It was concluded that, at least under good listening conditions, the perception of spectral relationships does not depend on peripheral auditory enhancement and probably rests on a central comparison process.

PACS numbers: 43.71.Es, 43.71.Cq

INTRODUCTION

The present study continues recent research on the perceptual integration of nasal murmur and vowel onset cues to the [m]-[n] distinction in CV syllables (Kurowski and Blumstein, 1984; Repp, 1986). Kurowski and Blumstein showed that each of these signal portions may carry considerable place of articulation information, and that subjects' identification performance is better when both are present (as they normally are) than when only one is present. They suggested that the two cues may function as a "single auditory property." However, their data also seemed consistent with the alternative possibility that the two cues are processed separately and combined at a later, evaluative stage in perception (see, e.g., Massaro and Oden, 1980). Repp referred to these two hypotheses as single cue (or early integration) and multiple cue (or late integration), respectively.

In addition to replicating and extending Kurowski and Blumstein's findings using a multitalker stimulus set, Repp made a preliminary attempt to address these two integration hypotheses. He formulated a simple probabilistic model of late information integration that predicted identification accuracy when two cues are available from identification performance for each cue presented in isolation. The predictions of the model generally fell short of the obtained identification scores, which was taken to mean that perceptual integration did occur at a relatively early stage, as hypothesized by Kurowski and Blumstein. However, the model may well have been too simple to represent the processes of

cognitive information integration. Another relevant piece of information obtained in Repp's study was that murmur and vowel onset cues still appeared to be integrated better than predicted by the model (or, in any case, permitted surprisingly high identification scores) when as much as 60 ms of the waveform surrounding the point of articulatory release was replaced with noise. This finding casts doubt on the role of a peripheral integration mechanism, since such a mechanism presumably should have been more sensitive to disruption of physical continuity. However, the noise may have enabled listeners to "restore" the missing acoustic information (cf. Warren, 1984). Clearly, Repp's data were not sufficient to decide between the early and late integration hypotheses, and further research was called for.

The concept of late integration needs little justification, since separate sources of information can always be combined in cognitive decision making as long as they are available at the same time (see, e.g., Massaro and Oden, 1980). The concept of early integration is more controversial, however. According to Kurowski and Blumstein's hypothesis, murmur and vowel onset "are not represented as separate cues, but are integrated by the auditory system into one unitary representation" (p. 389, emphasis added). As support for this claim, they cite the physiological studies of Delgutte (1980; Delgutte and Kiang, 1984) who found in cats that the neural response to a vowel onset was altered by a preceding nasal murmur, due to short-term adaptation of auditory nerve fibers. Kurowski and Blumstein conclude from this finding that "the auditory system does not treat transitions [i.e., the vowel onset] separately from the murmur" (p.

389). However, while Delgutte's results suggest that the auditory representation of the vowel onset is not *independent* of the preceding murmur, it does not follow that the two signal components, therefore, form an auditory unit. That is, one must distinguish between early *integration*, which yields a single auditory property, and early *interaction* among stimulus portions, which may modify their auditory representations while preserving them as separate sources of information that could be integrated by a later, cognitive process. Auditory adaptation would seem to be a likely source of early stimulus component interaction, but it is not clear how it ever could merge two signal portions of very different spectral structure and considerable temporal extent. Indeed, adaptation serves to enhance spectral changes in the signal (Summerfield *et al.*, 1984) and thus is a mechanism of *differentiation*, not of integration. Thus early integration of the kind envisioned by Kurowski and Blumstein seems unlikely as a general auditory function. Rather, the concept seems to reflect the axiomatic belief that single auditory properties underlie phonetic distinctions. This assumption is intended to relieve the listener's perceptual system from a computational burden, which instead falls upon the investigator trying to define the critical properties (Repp, 1987b).

Instead of the early integration hypothesis, therefore, the present series of experiments is concerned mainly with the perceptual consequences of early auditory interaction—henceforth, the (auditory short-term) *adaptation hypothesis*. Auditory short-term adaptation has been amply demonstrated not only in animals' auditory nerves (see also, e.g., Smith, 1977; Harris and Dallos, 1979; Abbas and Gorga, 1981; Eggermont, 1985) but also behaviorally in humans in the form of forward masking, decay of sensation, and auditory aftereffects (e.g., Zwislocki *et al.*, 1959; Plomp, 1964; Wilson, 1970; Widin and Viemeister, 1979; Viemeister and Bacon, 1981), including tasks involving phonetic judgments (Summerfield *et al.*, 1984; Summerfield and Assmann, 1987), even though adaptation may not be the only factor contributing to these phenomena. For all we know, then, auditory adaptation occurs continuously as we listen to speech. The question is: Does it help speech perception? Summerfield *et al.* (1984) and Summerfield and Assmann (1987) have argued that adaptation serves to enhance regions of spectral change, and that this may increase the intelligibility of speech, especially in noisy environments. In the specific case that concerns us here, *viz.*, prevocalic nasal consonants, significant spectral change occurs at the point of release, where the nasal murmur changes into the vowel (and also, beyond that point, during the formant transitions in the vowel). The murmur thus presumably has an adapting effect on the vowel onset that is proportional to the murmur spectrum, resulting mainly in attenuation of frequencies below 1000 Hz, where the murmur has most of its energy. Since distinctive place of articulation information is located at higher frequencies, some enhancement of vowel onset cues may result from the suppression of irrelevant spectral components (cf. Danaher and Pickett, 1975; Hannley and Dorman, 1983). The transitions of the second and third formants following vowel onset may also be enhanced somewhat by the (weak) presence of these formants in the mur-

mur. More generally, the negative aftereffect of the murmur results in a direct auditory representation of the differences in spectral amplitude between the murmur and the onset of the vowel. This direct spectral difference information may be perceptually valuable, especially for the labile [mi]–[ni] distinction (Repp, 1986).

It could be that such relational spectral information is the critical cue for place of articulation distinctions (see Lahiri *et al.*, 1984). This need not be so, however, since the murmur, as well as the later portions of the vowel, provides additional spectral (and temporal) information that may feed into a central integration process. Repp's (1986) preliminary acoustic analyses suggest that spectral difference information alone is not sufficient to distinguish [m] and [n] across all vowel contexts, at least not in an invariant fashion. It also seems to vary in perceptual importance depending on the vowel, being more essential in [–i] than in [–a] context, for example. Thus it may be only one of several ingredients that enter into phonetic decisions. This means that the inputs to the central decision process probably include the murmur spectrum, the spectral relationship between the murmur and the vowel onset, and the continuing pattern of spectral change during the vowel.

The present series of experiments was designed to test the adaptation hypothesis in a variety of ways. To repeat, that hypothesis states that adaptation by the nasal murmur modifies the internal representation of the vowel onset spectrum and thus makes spectral difference information directly available to the auditory system, which is important for the correct perception of place of articulation. Therefore, identification scores should drop if the effect of adaptation is reduced or eliminated. It was assumed that auditory adaptation, being a peripheral process, would be sensitive to disruptions of the physical continuity of murmur and vowel, so experiments 1–3 introduced manipulations such as order reversal, spatial separation, and temporal separation of murmur and vowel components. If such disruptions reduced identification performance substantially, a role of peripheral adaptation in providing place of articulation information would be suggested. If they had no effect, the auditory adaptation hypothesis could be rejected. A potential problem with this approach is that it is quite possible that spectral difference information, if it is not available as the direct consequence of peripheral auditory adaptation (and even if it is), is computed at a higher level in the perceptual system with the help of auditory memory (see Summerfield *et al.*, 1984; Summerfield and Assmann, 1987), as suggested, for example, by research on auditory profile analysis (see Green, 1983). Such a central comparison process may also be sensitive to disruptions of physical continuity, and unless such disruptions turn out to be ineffective, the outcome of the experiments will be consistent with both peripheral and central explanations. To distinguish further between these accounts, experiments 4, 6, and 7 examined several predictions thought to be specific to peripheral adaptation, concerning the effects on intelligibility of murmur duration, murmur/vowel mismatches, and simulated spectral enhancement at vowel onset. Experiment 5 addressed two alternative hypotheses, which will be introduced at that point.

I. GENERAL METHODS

A. Subjects

Three different groups of 12 or 13 student volunteers served as paid subjects, each in a single session including several experiments. All subjects were native speakers of American English and considered themselves to be free of hearing problems.

B. Stimuli

The same basic stimulus set as in Repp (1986) was used, and the earlier article may be consulted for details. Briefly, the stimuli were [ma, mi, mu, na, ni, nu] produced by three male and three female talkers, 36 syllables in all. The syllables were low-pass filtered at 4.9 kHz, digitized at 10 kHz, and modified as required. The onsets of three pitch periods (or pairs of pitch periods, in female tokens) preceding and following the point of release were marked to serve as cut-points in waveform editing. The temporal distance between these markers was approximately 10 ms.

C. Procedure

The subjects listened in a quiet room over TDH-39 earphones at a comfortable intensity. Unless mentioned otherwise, all stimulus presentations were binaural with interstimulus intervals of 3 s. The subjects in experiments 1-4 made a forced choice between /m/ and /n/ for each stimulus, guessing when no nasal consonant was perceived. The subjects in experiments 5-7 used a free response set including /m, n, b, d/ and /-/ (no consonant) as explicit choices. The first group of subjects participated in experiments 1 and 4; the second group in experiments 3 and 2; and the third group in experiments 5-7, in fixed order. (The experiments were renumbered for expository reasons in this article.)

D. Data analysis

Analyses of variance were performed on overall identification scores both across subjects (averaged over talkers) and across talkers (averaged over subjects). Therefore, two *F* values will be reported for each effect tested.¹ Differences among individual syllables will be discussed in a qualitative fashion.

II. EXPERIMENT 1

Experiment 1 tested the auditory adaptation hypothesis in a drastic fashion by reversing the order of the murmur and vowel components. Clearly, this manipulation eliminates any adapting effect the murmur might have on the vowel onset. Therefore, if adaptation enhances place of articulation cues, performance in the reversed condition should be much worse than when the murmur immediately precedes the vowel. On the other hand, if most of the place of articulation information results from processing the two sources of information separately and coding them in a more permanent form before central integration (e.g., as vectors of likelihoods of category membership; see Chistovich, 1985; Mas-saro and Oden, 1980), then their order might be less important. However, if important spectral relationships are

extracted centrally, that process may well be sensitive to order also. Thus it was perhaps unlikely that no decline in performance would result from an order reversal; nevertheless, the fact that this result would provide conclusive evidence against the auditory adaptation hypothesis justified the experiment.

A. Methods

The experiment included five conditions, each represented by a test sequence consisting of one randomization of the 36 stimuli. The first sequence contained the original, unaltered syllables and served as warmup. The second sequence contained the same syllables, but with about 60 ms of the waveform surrounding the point of release excised. In other words, approximately the last 30 ms of the murmur and the first 30 ms of the vowel (each corresponding to three male or six female pitch pulses) were removed, and the two truncated stimulus components were joined together. This excision was done to increase the number of errors and thus to reduce ceiling effects. The relatively abrupt change from the murmur to the vowel was thought to enhance the effect of adaptation on the remaining place of articulation cues in the vowel, or at least not to decrease it. That the truncated vowel portions, as well as the truncated murmurs, still contained considerable place of articulation information was clear from earlier data (Repp, 1986). To confirm this, and to illustrate to the subjects the nature of the separate stimulus components, the third and fourth test sequences contained the truncated murmurs and vowels, respectively, in isolation. The critical fifth sequence contained the truncated vowels *followed* by the truncated murmurs after a 300-ms silent interval. This interval was inserted to prevent the perception of postvocalic nasal consonants.

B. Results and discussion

The results, averaged over subjects and talkers, are summarized in Table I. Performance for the unaltered syllables was 95% correct; nearly all errors occurred with [ni]. Excision of 60 ms surrounding the release caused a 10% drop in the average score, although identification of [ma] and [na] remained unaffected. Scores for isolated truncated murmurs and vowels were 56% and 61% correct, respectively. From these scores, Repp's (1986) simple late integration formula predicts an overall performance of 66% correct for murmurs and vowels combined, without any relational information

TABLE I. Percent correct scores for individual syllables in the five conditions of experiment 1 (M = murmur, V = vowel).

Conditions	Syllables						Average
	[mi]	[ni]	[ma]	[na]	[mu]	[nu]	
Full syllable	97	74	100	99	100	100	95
M + V	68	64	99	99	89	92	85
M	56	47	65	49	61	58	56
V	51	49	58	71	57	81	61
V + (300 ms) + M	57	47	82	76	68	71	67

added. Clearly, however, such relational information played a role when murmur and vowel were concatenated (condition 2): Scores were much higher than predicted. In condition 5, on the other hand, where the murmur *followed* the vowel, performance was 67% correct. This is close to the predictions of the model, and, while it is marginally better than identification of isolated vowels [$F(1,11) = 4.25$, $p = 0.0636$; $F(1,5) = 9.50$, $p = 0.0274$], it is substantially lower than the 85% correct obtained in the second condition [$F(1,11) = 49.35$, $p < 0.0001$; $F(1,5) = 48.75$, $p = 0.0009$]. As Table I shows, this latter difference was obtained for all individual syllables, even though they differed markedly in their vulnerability to truncation.

These results confirm the important perceptual role of spectral difference information. When this information is directly available, speech intelligibility is much higher than when listeners can rely only on the cognitive integration of independent sources of information. Models of speech perception that assume the integration of independent cues (e.g., Massaro and Oden, 1980) are incomplete in this respect. The results are thus consistent with the adaptation hypothesis, but they cannot be taken as direct support for it. Relational information could also be derived by a nonperipheral spectral comparison process sensitive to temporal order and/or temporal separation.

III. EXPERIMENT 2

Before turning to finer parametric stimulus variations, the results of a second gross manipulation shall be reported. The rationale for experiment 2 was that, if adaptation takes place in the peripheral auditory system, it should be sufficient to present the stimulus components to different ears to eliminate it. Summerfield *et al.* (1984) found that an auditory aftereffect believed to rest on adaptation disappeared when the adapting and test stimuli were presented to opposite ears. However, any central processes that extract spectral relationships might operate on inputs from different ears. As in experiment 1, it was perhaps unlikely that the segregation of murmur and vowel would have no effect at all on intelligibility, but the strong implications such an outcome would have for the adaptation hypothesis made the experiment worthwhile.

A. Methods

The same truncated murmurs and vowels as in experiment 1 were used. There were three conditions, each consisting of one presentation of the 36 stimuli. In contrast to experiment 1, however, the three conditions were randomized together. Two conditions were identical with conditions 2 (truncated murmur immediately followed by truncated vowel) and 4 (isolated truncated vowels) of experiment 1, except that presentation was monaural. In the third, "split" condition, the truncated murmur occurred on the opposite channel, immediately preceding the truncated vowel, which was on the same channel as the other stimuli. Half the subjects received the vowel portions in the left ear, and half in the right ear. No ear differences were apparent, so the data were pooled over this variable.

B. Results and discussion

Performance for the monaural murmur-vowel stimuli was 86% correct, which is similar to the score obtained (with different subjects) in experiment 1. Performance for isolated vowels (67% correct) was somewhat higher than in experiment 1 but matches the score obtained by Repp (1986). Performance in the novel split condition was 78% correct, significantly higher than for isolated vowels [$F(1,11) = 17.47$, $p = 0.0015$; $F(1,5) = 9.08$, $p = 0.0297$] but lower than for monaural murmur-vowel stimuli [$F(1,11) = 23.93$, $p = 0.0005$; $F(1,5) = 8.07$, $p = 0.0362$].

Differences among individual syllables may be examined in Table II. It appears that [m-] syllables gained more from the addition of a contralateral murmur to the isolated vowel than did [n-] syllables. This is surprising in the case of [mi], whose murmur by itself conveyed very little reliable information, whereas the murmurs of [ma] and [mu] yielded the highest scores in isolation (see Table I; also, Repp, 1986) and, therefore, were expected to make a large contribution. In the case of [na] and [nu], the negligible gain may have been due to the fact that the isolated vowels were identified almost as well as the monaural murmur-vowel stimuli. The possibility of response biases cannot be ruled out.² If the task is considered as one of [m]-[n] discrimination within each vocalic context (e.g., if percent correct scores are computed for [m]-[n] pairs) all inconsistencies disappear, and performance in the split condition is intermediate between the other two conditions in all three vocalic contexts.

The results suggest, then, that channel separation of murmur and vowel disrupts the extraction of spectral difference information. This is consistent with the adaptation hypothesis, but it could also be that there is a central process of spectral comparison which is sensitive to spatial separation of sound sources. The scores in the split condition seem fairly close to what one should expect on the basis of late integration of independent sources of information, so the central process responsible for that integration presumably was not affected. While the results of experiment 2, like those of experiment 1, do not permit rejection of any specific hypothesis, they do suggest that spatiotemporal contiguity of signal components is required for the effective detection of relational spectral cues.

IV. EXPERIMENT 3

The obvious next step was to determine how close in time the two signal components must be for listeners to reap

TABLE II. Percent correct scores for individual syllables in the three conditions of experiment 2 (M = murmur, V = vowel, / = split between ears).

Conditions	Syllables						Average
	[mi]	[ni]	[ma]	[na]	[mu]	[nu]	
M + V	72	75	100	83	92	93	86
V	49	53	79	81	57	86	67
M/V	76	43	97	76	83	90	78

the benefits of spectral difference information. One of the more striking findings of Repp (1986) was that substitution of signal-correlated noise (SCN) for the 60 ms of waveform surrounding the consonantal release resulted only in a relatively small decrement in overall identification performance; the syllables [mi] and [ni] supplied virtually all the errors. Repp concluded that murmur and residual vowel onset cues were perceptually integrated across the intervening noise; that is, it appeared that spectral difference information remained largely intact.³

This result is not necessarily damaging to the adaptation hypothesis. Short-term adaptation may last for 150 ms or more (Delgutte, 1980; Summerfield *et al.*, 1984), and a brief broadband noise may dilute but not eliminate the effect, just as would decay of adaptation during a 60-ms silent interval. However, if this interval were extended, a substantial decrement in adaptation should be observed.

To test these predictions, experiment 3 assessed identification performance for stimuli whose murmur and vowel components were separated by silent intervals of up to 240-ms duration. The use of silence rather than noise was justified by the results of another experiment, not reported in detail here, which showed that intervening signal-correlated noise, broadband noise, and silence had statistically equivalent effects.⁴

A. Methods

The truncated murmur and vowel components were used again, separated by 0, 30, 60, 120, or 240 ms of silence. All five conditions were randomized together and recorded in five blocks of 36 syllables each.

B. Results and discussion

The results are summarized in Table III. There was no decline in performance over the first 60 ms of separation. Only at the longer intervals was there a small reduction in performance. Overall, the effect of temporal separation was significant across subjects [$F(4,44) = 3.70, p = 0.0111$] but not across talkers. With regard to individual syllables, it can be seen that identifiability declined with silence duration for [n-] but not for [m-] syllables. This may once again have been due either to an /m/ response bias which emerged as the murmur was separated from the vowel, or it may indicate that labial place of articulation was perceptually more stable under these conditions.

TABLE III. Percent correct scores for individual syllables in the five conditions of experiment 3.

Silence duration	Syllables						Average
	[mi]	[ni]	[ma]	[na]	[mu]	[nu]	
0 ms	71	67	100	89	92	94	85
30 ms	81	57	99	94	93	96	87
60 ms	78	51	100	90	93	94	85
120 ms	74	53	99	86	99	81	82
240 ms	76	54	97	83	92	76	80

These results are not so easy to reconcile with the adaptation hypothesis. First, the decline in performance was small and did not occur with all syllables and talkers. Second, there seemed to be no decline at all over the first 60 ms of separation, although auditory adaptation, which decays exponentially (Harris and Dallos, 1979; Eggermont, 1985), should have decreased significantly in that interval. Since the truncated murmur and vowel components were in their original temporal relationship when separated by 60 ms, a perceptual advantage resulting from this fact may conceivably have counteracted any decline due to decay of adaptation at short intervals. Apparently, however, listeners still had spectral difference information available with 240 ms of temporal separation, and this suggests that they used auditory memory for the murmur to determine its spectral relationship to the vowel onset. Whether this was a compensatory perceptual strategy or whether it reflects what occurs in intact syllables is not clear.

V. EXPERIMENT 4

Experiment 4 addressed two further predictions of the adaptation hypothesis, which contrasted with predictions arising from the alternate hypothesis that murmur and vowel onset function as independent cues that are integrated at a late stage (e.g., Massaro and Oden, 1980). One prediction concerned the effect of murmur duration. Physiological studies have shown that auditory adaptation in animals increases with adaptor duration up to about 100 ms (Harris and Dallos, 1979; Westerman and Smith, 1984). Even though the temporal parameters may not be exactly the same in the human auditory system, to the extent that auditory adaptation by the murmur enhances the spectral structure at vowel onset, there should be a beneficial effect of increasing murmur duration (up to about 100 ms) on identification of murmur-vowel stimuli. In isolated murmurs, however, there can be no such enhancing effect of adaptation; therefore, increasing murmur duration beyond some minimum should have little influence on intelligibility. This was already suggested by Repp's (1986) analysis of the effect of natural variations in murmur duration; in addition, he found that the intelligibility of truly steady-state isolated murmurs *decreased* as their duration was increased, perhaps because their artificial quality became more apparent as they got longer. Thus a statistical interaction of the effect of murmur duration with the factor of presence versus absence of a following vowel is predicted. A contrasting prediction emerges from the late integration of independent cues hypothesis: Whether increasing murmur duration increases or decreases the informational value of the murmur, it should do so regardless of the context in which the murmur occurs.

A second prediction examined by experiment 4 was this: If auditory adaptation caused by the murmur improves perception of higher formants at vowel onset, then a beneficial effect of prefixing an isolated vowel portion with a murmur should be obtained regardless of whether or not the murmur derives from the same utterance. The reason is that all murmurs are spectrally rather similar below 1000 Hz, where most of their energy is concentrated. And although [m] and [n] murmurs differ in the frequencies of their higher for-

mants, which are continuous with the formants at vowel onset, it may be argued that the spectral change at vowel onset would be enhanced even more if the murmur formants were different from those at vowel onset. The paradoxical prediction is, therefore, that addition of an inappropriate murmur to an isolated vowel may *improve* identification, relative to the isolated vowel baseline. The opposite result is predicted by the independent cues hypothesis: The introduction of a conflicting cue cannot possibly improve performance. (Late integration of murmur and vowel onset cues may occur following an early auditory interaction, in which case two opposing tendencies may cancel in the data.) To test these predictions, the experiment included both compatible and conflicting murmur-vowel combinations. Thus it was also possible to compare directly two types of conditions for nasal consonants that previously have been employed in separate studies (Kurowski and Blumstein, 1984; Malécot, 1956) or with other place of articulation contrasts (Recasens, 1983).

A. Methods

The experiment included one long randomized test sequence composed of $9 \times 36 = 324$ stimuli, and a shorter sequence of $3 \times 36 = 108$ stimuli. The stimulus components were steady-state murmurs generated by reiterating a single 10-ms segment of the original murmurs, taken from the vicinity of the release (see the Static Excerpts condition of Repp, 1986) and vowel portions whose initial 10 ms (one male or two female pitch pulses) had been removed.⁵ The first test sequence contained the vowel portions in isolation and immediately preceded by 1, 3, 6, or 12 segments of matched or mismatched murmur. The murmur durations thus were in the vicinity of 10, 30, 60, and 120 ms. The mismatched murmurs came from the syllable with the same vowel but a different consonant, produced by the same speaker. The second, shorter test sequence contained only isolated murmurs of 30-, 60-, and 120-ms duration. (The 10-ms murmurs were omitted because they were easily missed in listening.)

B. Results and discussion

The overall results are shown in Fig. 1. The figure plots percent correct scores as a function of murmur duration for isolated murmurs and for murmur-vowel stimuli with matched and with mismatched components. (In the case of mismatched components, "correct" responses are defined with respect to the *vowel* portion.) The data point on the ordinate, corresponding to zero murmur duration, represents the score for isolated vowels (72% correct). The results indicate that addition of a 10-ms matched or mismatched murmur to the vowel changed identification performance little, whereas addition of a murmur 30 ms long or longer resulted in an improvement, but only if the murmur matched the vowel. Mismatched murmurs neither improved nor hindered identification. Isolated murmurs of 30- and 60-ms duration were identified at levels above chance, but 120-ms murmurs could not be reliably identified. This last finding (which may have been a consequence of the artificial steady-state nature of the murmurs; cf. Repp, 1986) contrasts with the differential effect of 120-ms matched and mis-

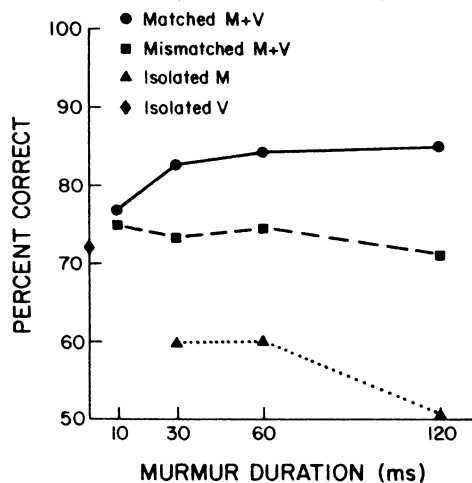


FIG. 1. Results of experiment 4: percent correct identification of isolated vowels (V), isolated murmurs (M), and murmur-vowel stimuli (M + V) with matched and mismatched components as a function of murmur duration.

matched murmurs when they preceded a vowel.

A two-way analysis of variance of the scores for the murmur-vowel stimuli yielded a significant effect of match/mismatch [$F(1,11) = 19.01, p = 0.0011$; $F(1,4) = 13.31, p = 0.0218$] and a significant interaction with murmur duration [$F(3,33) = 6.34, p = 0.0016$; $F(3,12) = 4.28, p = 0.0285$], obviously due to the shortest murmur duration, whereas the main effect of murmur duration was not significant. A separate analysis of the isolated murmurs showed a significant effect of murmur duration [$F(2,22) = 3.98, p = 0.0335$; $F(2,8) = 8.85, p = 0.0094$], suggesting that the performance decrease for the longest murmurs was real.

These overall results cannot be given much weight, however, in view of very striking dependencies on vocalic context. The pattern of results for individual syllables is shown in Fig. 2. Each panel shows data for one vocalic con-

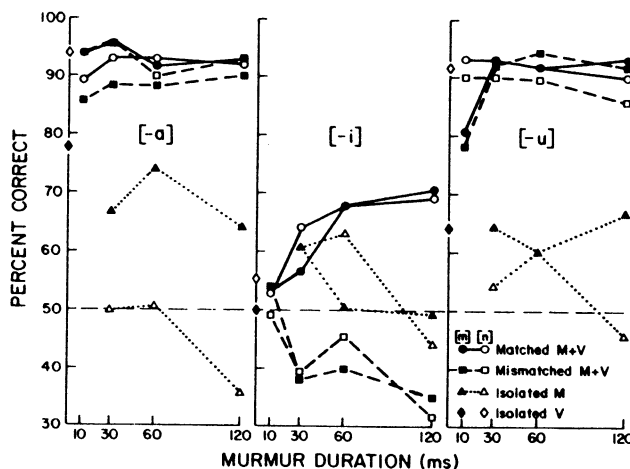


FIG. 2. Results for individual syllables in experiment 4.

text, with solid and open symbols representing [m-] and [n-] syllables, respectively. Consider, first, the [-a] and [-u] syllables (left and right panels). The isolated vowels of [na] and [nu] were identified much more accurately than those of [ma] and [mu], which replicates earlier findings (experiments 1 and 2, Repp, 1986) and probably reflects the greater perceptual salience of alveolar than labial formant transitions (or onset spectra). Because of this pattern, a [(m)a] or [(m)u] vowel benefited from addition of a murmur (even a 10-ms one), while a [(n)a] or [(n)u] vowel did not. Identification performance was uniformly high for all murmur-vowel stimuli in [-a] and [-u] context. Moreover, there was very little difference between scores for stimuli with matched and mismatched components. Identification of [(m)a] and [(m)u] vowels was *improved* by addition of a mismatched murmur almost as much as by addition of a matched murmur, and identification of [(n)a] and [(n)u] vowels was at least not hampered by addition of a mismatched murmur.

This part of the data is consistent with the adaptation hypothesis. As to the predicted effects of murmur duration, they are smaller than expected but are also compatible with the hypothesis. The results are inconsistent with the independent cues hypothesis, according to which performance should have decreased in the mismatched conditions.

The pattern for [mi] and [ni] stimuli (center panel of Fig. 2) is very different from the results just described. Identification of isolated vowels and isolated murmurs was extremely poor, in agreement with earlier results. Addition of a 10-ms murmur to the vowel had no effect, but addition of a murmur 30 ms or more in duration elicited responses that reflected the nature of the *murmur*. Thus there was a large effect of match versus mismatch, which accounts for the average effect shown in Fig. 1. Since the murmurs were barely discriminable in isolation, especially when they were 120 ms long, listeners cannot have relied on them directly to identify the consonant in these syllables. The data thus support the earlier conclusion (Repp, 1986) that [-i] syllables are special in that place of articulation information lies almost entirely in the *relationship* between the murmur and the vowel, that is, in the pattern of spectral change. A possible implication is that there are differences between [m(i)] and [(n)i] murmurs that are difficult to detect in isolation but that become perceptually salient when the murmur is followed by a vowel. Such a retroactive enhancement effect would refute the adaptation hypothesis. Yet, there is a way in which it could arise through adaptation: Different murmurs might impose their inverse spectrum on the vowel onset, thereby creating a place of articulation cue following the release. On the other hand, the independent cues hypothesis, unless it is extended to include relational information, cannot explain how murmurs that are uninformative in isolation convey phonetic differences in context.

In summary, while the results of experiment 4 argue very clearly against the independent cues hypothesis and thereby affirm the importance of relational spectral information, they are perhaps still compatible with a peripheral account of spectral difference detection.

VI. EXPERIMENT 5

Prior to experiments 6 and 7, which attempted to test the adaptation hypothesis in yet another way, experiment 5 examined two alternative explanations of how a preceding murmur might enhance the perception of vowel onset cues. One hypothesis (Repp, 1986) takes account of the fact that the murmur is the major carrier of nasal manner information. If it were the case that place of articulation perception is not independent of manner perception (see Miller, 1977; Carden *et al.*, 1981), then hearing the correct manner may enhance the accuracy of place identification. Kurowski and Blumstein (1984) reported that their CV syllables were identified as beginning with oral stop consonants when the nasal murmur was excised. Their subjects chose from the response set /m,n,b,d/ and gave about 84% /b,d/ responses to murmurless stimuli but only about 12% to stimuli with an initial murmur. Thus removal of the nasal murmur clearly changed manner of articulation perception and perhaps affected place of articulation perception as well, particularly since the isolated vowel portions of nasals lack the release bursts commonly associated with oral stop consonants. In Repp's (1986) experiments, and in experiments 1-4, subjects always were required to make a forced choice between /m/ and /n/, regardless of whether they perceived the correct manner or indeed any consonant at all. One purpose of experiment 5 was to determine first whether the present stimuli resembled those of Kurowski and Blumstein (1984) in that removal of the murmur resulted in the almost complete loss of nasal manner cues, and then whether correct perception of place was contingent on correct perception of manner.

A second hypothesis addressed by experiment 5 derives from observations by Pols and Schouten (1978, 1981) on the perception of truncated stop-consonant-vowel syllables. These authors argued that the relatively abrupt stimulus onset following truncation causes spectral splatter (a "click sensation") that interferes with the perception of place of articulation cues. Identification scores improved substantially when the truncated syllables were preceded by noise bursts that masked the abrupt onset (Pols and Schouten, 1978). Ohde and Sharf (1981) applied a smoothing function to the onsets of truncated CV syllables, apparently with similar results (see Pols and Schouten, 1981). It is possible that part of the intelligibility decrement for isolated vowel portions in experiments 1-4 was caused by abrupt stimulus onsets. To check on this, a smoothing function similar to that used by Ohde and Sharf (1981) was applied to the stimulus onsets on half the trials in this experiment.

A. Methods

The experimental tape contained $8 \times 36 = 288$ isolated vowel stimuli in random order. Each vowel was truncated approximately 0, 10, 20, and 30 ms after the release (see Repp, 1986); thus none of them contained any nasal murmur. (It was quite clear from informal listening that inclusion of even a very brief murmur resulted in the perception of a nasal consonant.) Each truncated stimulus occurred in

two versions, one unaltered and the other with a linear amplitude ramp, rising from near-zero to full intensity in 10 ms, applied to the onset of the digitized waveform. The subjects' task was to report for each stimulus the initial consonant they heard, choosing from the set /m,n,b,d/, and to write down a dash when no consonant was heard.

B. Results and discussion

The overall results, averaged over the ramped and un-ramped stimulus versions, are shown in Fig. 3. Three measures were derived from the data. The first, $p(C)$, was the percentage of trials on which a consonant was reported. Not surprisingly, it declined with progressive truncation [$F(3,36) = 47.05$, $p < 0.0001$; $F(3,15) = 94.55$, $p < 0.0001$], although the vowel portions were still heard as containing initial consonants on about half the trials even after their initial 30 ms had been deleted. The other two measures were conditional on a consonant being reported. The percentage of correct place identifications, $p_c(P|C)$, declined only very slightly with truncation [$F(3,36) = 2.17$, $p = 0.1081$; $F(3,15) = 6.45$, $p = 0.0051$], suggesting that the decrease in two-alternative forced-choice identification scores with progressive truncation (Repp, 1986) was caused more by the total loss of consonantal cues than by misleading residual cues. Most interestingly, the percentage of nasal consonant responses, $p(N|C)$, did not decline at all with progressive truncation, but actually showed an initial increase [$F(3,36) = 9.72$, $p = 0.0001$; $F(3,15) = 10.22$, $p = 0.0006$]. Regardless of how much consonantal information was available, about half of the consonants reported were nasals. This percentage is much higher than that reported by Kurowski and Blumstein (1984), even though removal of the nasal murmur undoubtedly caused a significant loss of nasal manner information. Presumably, the talker used by Kurowski and Blumstein closed his velum more rapidly after the consonantal release than did the present talkers, who tended to nasalize the vowel onset.

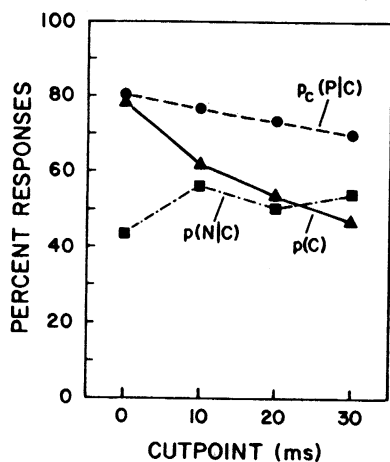


FIG. 3. Results of experiment 5: percentages of consonant responses, $p(C)$, of correct place of articulation identifications given a consonant response, $p_c(P|C)$, and of nasal consonant responses given a consonant response, $p(N|C)$, as a function of cutpoint location.

Differences among individual syllables are shown in Fig. 4. With regard to the percentage of consonant responses (left panel), it can be seen that [mu] and [ni] were affected much more by excision of the murmur (0-ms cutpoint) than the other syllables. This probably reflects the weak formant transitions in these stimuli, which have similar articulatory configurations for consonant and vowel. Further truncation had especially strong effects on [ma] and [mi], indicating the loss of rapid labial transients at stimulus onset. Perception of the consonants in [na] and [nu], which have relatively long vocalic formant transitions, was most resistant to vowel truncation.

The most striking difference in correct place of articulation identification scores (center panel) was between [ni] and all other syllables. Without the murmur, [ni] tended to be misidentified as labial, which indicates that the vowel did not contain any useful formant transition information. The same may well be true for [mu], and the 70%–80% labial responses to both of these syllables may represent a bias to respond with labial consonants in the absence of clear place of articulation cues. Only [ma] and [mi] were affected by vowel truncation.

The percentage of nasal responses (right panel) was lower for [mi] and [ni] than for the other syllables. The difference between [-i] and [-a] syllables may be explained by the fact that the velum is raised faster for high than for low vowels following a nasal consonant (Bell-Berti *et al.*, 1979), making the former less nasalized. It is not clear, however, why the [-u] syllables resembled more the [-a] than the [-i] syllables in degree of perceived nasality, or why perception did not fully compensate for the expected differences in velar elevation for vowels of different heights (see Abramson *et al.*, 1981).

The principal hypothesis addressed by this experiment concerned the possible dependence of place perception on manner perception. Since only about half of the initial consonants perceived in truncated syllables were nasal, it is indeed possible that place of articulation perception suffered because of insufficient manner cues. If so, place identification contingent on correct perception of nasal manner should have been more accurate than place identification contin-

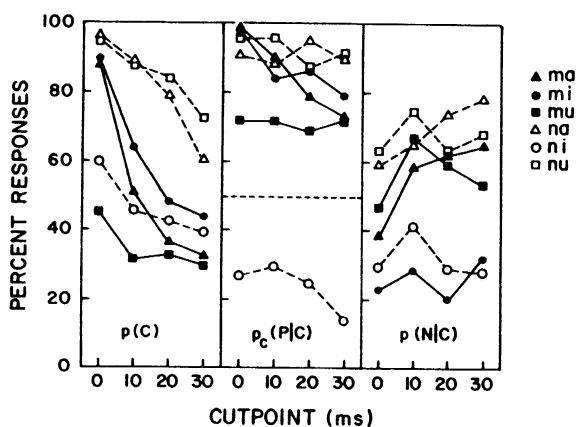


FIG. 4. Results for individual syllables in experiment 5.

gent on perception of non-nasality. Examination of these percentages (computed from the syllable averages), however, revealed only a small difference (2% on the average) in the predicted direction. This difference, moreover, derived entirely from the stimuli with tapered onsets (5.5% average difference); for the others, there was a 1.6% difference in the opposite direction. Although the effect of amplitude tapering deserves attention (see below), all stimuli in earlier experiments were, of course, untapered. For those stimuli, then, there is no evidence that incorrect perception of manner impaired place of articulation identification, so the perceptual enhancement of place cues when a vowel is prefixed with a murmur cannot be explained on that basis.

It is noteworthy, however, that there were very large differences among individual syllables. The differences between correct place identification scores contingent on perceived nasal and non-nasal manner were: - 2.8% for [ma], - 18.3% for [mi], - 28% for [mu], 12.5% for [na], 42% for [ni], and 6.3% for [nu]. It thus appears that, when a consonant was perceived as non-nasal, there was a strong shift in favor of labial responses; the differences in absolute magnitude of this shift among the six syllables probably derived largely from ceiling effects. Thus there *was* a dependency of place of articulation identification on manner, though in terms of criterion rather than accuracy. This effect is in agreement with earlier findings (Miller, 1977; Larkey *et al.*, 1978) of a relative shift in the category boundary on synthetic /ba/-/da/ and /ma/-/na/ continua. One likely cause for this is the absence of release bursts in both the synthetic stop-consonant-vowel stimuli used previously and in the present vowel portions. In real speech, alveolar oral stops have stronger release bursts than do labial oral stops, so the absence of bursts promotes the perception of labials, provided that a stop consonant is perceived.

Turning finally to the effect of amplitude tapering, there were small but consistent effects on two of the three overall performance measures. The percentage of consonant responses was reduced by about 7% at all stages of truncation [$F(1,12) = 16.19, p = 0.0017; F(1,5) = 6.12, p = 0.0563$], which suggests a loss of general manner cues at stimulus onset. Given that a consonant was heard, however, place of articulation identification was improved by about 5% overall [$F(1,12) = 7.60, p = 0.0174; F(1,5) = 11.17, p = 0.0205$]. This effect is in agreement with the observations of Pols and Schouten (1981) on the interfering effect of abrupt stimulus onsets, although the size of the present effect was rather small—certainly much smaller than the improvement obtained by Pols and Schouten (1978) with a noise prefix. Actually, the present improvement derived solely from those trials on which nasal consonants were perceived (cf. the interaction reported above); when nasality was not perceived, there was no effect of tapering. This is less in agreement with Pols and Schouten. Onset tapering had no systematic overall effect on nasal manner perception.

In summary, the results of this experiment do not support the hypothesis that, when a vowel is preceded by its original murmur, part of the improvement in place of articulation identification derives from the restoration of correct manner identification. Perception of nasal manner does not

seem to enhance perception of place, at least not in untapered stimuli as used previously; it only shifts the response criterion in favor of alveolar responses. The second hypothesis, that elimination of abrupt onsets improves place perception, receives some limited support from the present results. Though the effect is rather small, it may add to the contribution of a preceding murmur. However, it cannot explain correct perception of the intact syllable [ni], or of [mi] with truncated vowel, for which the murmur and the vowel in isolation are equally uninformative. The concept of relational information is still required, and so we must return to the adaptation hypothesis.

VII. EXPERIMENT 6

The final two experiments in this series provided perhaps the most direct test of the adaptation hypothesis. If peripheral adaptation by the murmur enhances spectral information at vowel onset, then it should be possible to simulate this enhancement by filtering the vowel onset in the absence of a preceding murmur. Such artificial enhancement then should result in improved place of articulation identification from isolated vowel components. Confirmation of this prediction would not only provide strong support for the adaptation hypothesis, but it would also lead to a reevaluation of earlier conclusions based on place of articulation identification from isolated vowel portions (experiments 1, 2, and 5; Kurowski and Blumstein, 1984; Repp, 1986), which did not consider that removal of a murmur also eliminates its adaptive aftereffect.

In choosing an appropriate filtering function, decisions had to be made concerning its shape, depth, and decay over time. Acoustical analysis of the nasal murmurs indicated that most of their energy was below 1000 Hz, and that the peak corresponding to the first formant was about 30 dB higher, on the average, than the peaks of the higher formants above 1000 Hz. Only the higher formants, however, varied with place of articulation. Ideally, the spectral shape of the filter should initially mirror that of the natural murmur and then wane over time, simulating decay of auditory adaptation. These objectives were difficult to achieve simultaneously with the facilities available. In experiment 6, therefore, it was decided to use a simple high-pass filter with a cutoff frequency of 1000 Hz, which permitted variable stop band attenuation to simulate decay. The experiment thus tested one specific version of the adaptation hypothesis, viz., that enhancement of place cues in higher formant transitions at vowel onset results from suppression of energy in the region of the first formant. As to the decay time, it was assumed that it would be rather short during stimulation by the vowel itself. (Most estimates of decay times in the literature derive from observations during silent intervals.) Even if the range chosen (up to 30 ms) seems too short, it became clear during stimulus preparation that more extensive filtering led to very unnatural-sounding stimuli.

A. Methods

The basic stimuli were the complete vowel portions of the original 36 syllables. Even though ceiling effects in per-

formance were expected to limit the sensitivity of the experiment to beneficial (but not detrimental) effects of filtering, no truncation was performed on the vowels in this study and the next, so as to preserve the original acoustic properties of the vowel onsets. Three degrees of high-pass filtering were imposed on initial pitch-pulse segments, leaving the rest of the waveform intact: (1) The initial 10-ms segment only, with 10-dB stop band attenuation; (2) the initial segment with 20 dB, and the following segment with 10-dB stop band attenuation; and (3) the initial segment with 30 dB, the following segment with 20 dB, and the final segment with 10-dB stop band attenuation. Thus three degrees of adaptation with three decay times were crudely simulated. The filtering was performed digitally, using an eighth-order elliptic filter with a fixed cutoff frequency of 1000 Hz and variable attenuation, constructed by the EFI subroutine of the ILS package (version 4.0, Signal Technology, Inc.). The boundaries of the pitch pulse(s) to be filtered in each pass through the routine were specified precisely in tenths of milliseconds, according to Repp's (1986) cutpoint markers. The result was verified through inspection of waveforms and acoustic analysis. The four series of 36 stimuli (3 filtered, 1 unaltered) were randomized together. Subjects were instructed to identify each stimulus as beginning with /m,n,b,d/ or /-/ (no consonant).

B. Results and discussion

The overall results are shown in Fig. 5 in terms of the three performance measures introduced in experiment 5. Looking first at the $p(C)$ scores, it can be seen that, in agreement with the results of experiment 5, the unaltered syllables elicited close to 80% consonant responses. This percentage declined to 65% with progressive filtering [$F(3,36) = 18.47$, $p < 0.0001$; $F(3,15) = 14.49$, $p = 0.0001$], suggesting that the first formant contributed general consonant manner information. A decline with re-

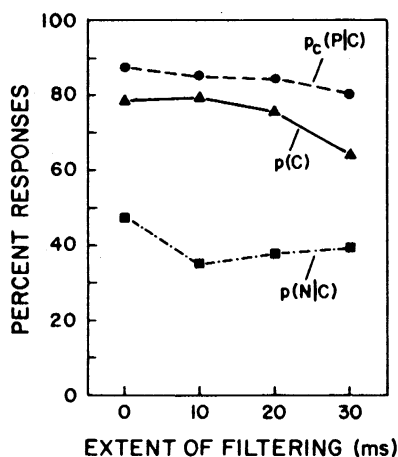


FIG. 5. Results of experiment 6: Three performance measures as a function of temporal extent of high-pass filtering.

spect to the unaltered stimuli was also observed in the conditional percentage of nasal consonant responses, $p(N|C)$ [$F(3,36) = 5.33$, $p = 0.0038$; $F(3,15) = 6.86$, $p = 0.0039$], although it did not seem to depend on the extent of filtering. Most importantly, the conditional percentage of correct place of articulation identifications, $p_c(P|C)$, also declined rather than increased, with increasing extent of filtering. Although absence of an increase in performance could be blamed on ceiling effects, and although the decline is rather small and nonsignificant, these data offer no support for the hypothesis that attenuation of irrelevant low-frequency energy enhances place of articulation cues at higher frequencies.

Figure 6 shows the results for individual syllables. In the left panel, it can be seen that consonant responses decreased most strongly for [ma] and [mi], whereas [mu] actually showed an increase with filtering. Place perception suffered in all syllables but the poorly identified [ni], for which there was an increase with filtering. Since identification of this syllable never exceeded chance level, the increase is probably a criterion effect. Perception of nasality suffered in all syllables but [ma], which showed an increase with filtering. These interactions are curious, but they do not change the general conclusions.

VIII. EXPERIMENT 7

The results of experiment 6 lend no support to the specific hypothesis that auditory adaptation enhances place of articulation perception through elimination of irrelevant low-frequency spectral energy. It is still possible, however, that a beneficial effect of adaptation occurs at higher frequencies, where the important place of articulation cues reside. To test this version of the adaptation hypothesis, it was necessary to use a filter that preserves the detailed spectral shape of the murmur, with some loss of flexibility in other respects.

A. Methods

From each of the 36 original murmurs, a 14-coefficient LPC spectrum was computed using a 25.6-ms Hamming

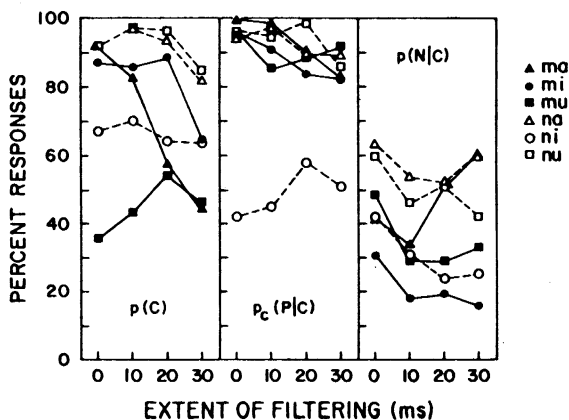


FIG. 6. Results for individual syllables in experiment 6.

window ending about 10 ms before the point of release (ANA program of the ILS package). Each of these spectra was subsequently used as an inverse filter on the complete vowel portion of each syllable (FLT program). Degree of attenuation could not be varied easily in this procedure. To vary temporal extent in synchrony with pitch pulses, which could not be done directly, the initial one, two, or three pitch-pulse segments of the filtered vowel (about 10, 20, and 30 ms long) were concatenated with the remainder of the unfiltered vowel, using a waveform editing program. The success of the filtering procedure was verified by acoustic analysis. The resulting 4×36 stimuli (including the unaltered versions) were recorded in a randomized sequence. The subjects' instructions were the same as in experiment 6. Two additional sequences of 36 stimuli each were recorded afterwards, each containing the excerpted initial 30-ms segments of the vowels, first unfiltered and then filtered. The purpose of this was to assess to what extent any perceptual effects of filtering depended on the following unfiltered vowel or were artifacts of the abrupt amplitude change between filtered and unfiltered waveform segments. In responding to these final two sequences, subjects had to make a forced choice between /m/ and /n/ for each stimulus.

B. Results and discussion

Figure 7 shows the overall results for the main test. It can be seen that the pattern was rather similar to that obtained with high-pass filtering (Fig. 5). Consonant responses increased slightly initially but then decreased with increasing filtering [$F(3,36) = 13.98$, $p < 0.0001$; $F(3,15) = 8.91$, $p = 0.0012$]. Nasal consonant responses dropped considerably with minimal filtering and then recovered partially as filtering increased [$F(3,36) = 26.79$, $p < 0.0001$; $F(3,15) = 16.45$, $p = 0.0001$]. Correct place of articulation responses were not significantly affected, but certainly showed no tendency to increase. The results for the isolated 30-ms segments likewise showed no advantageous effects of filtering: Forced-choice identification scores were

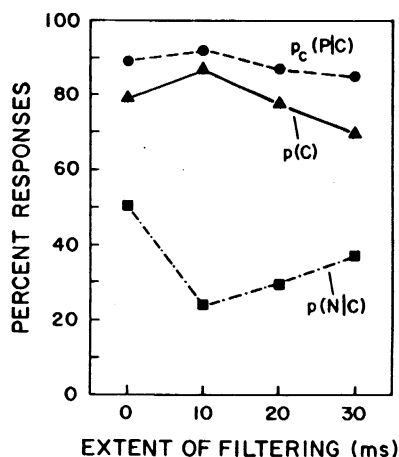


FIG. 7. Results of experiment 7: three performance measures as a function of temporal extent of inverse filtering.

66.5% and 64.3%, respectively, for unfiltered and filtered excerpts—a nonsignificant difference.

Scores for individual syllables are shown in Fig. 8. It can be seen that consonant responses increased initially for [mu] and [ni], suggesting that an initial amplitude discontinuity provides a general consonant manner cue. With more extensive filtering, however, the cue lost its effectiveness, and consonant scores declined for all syllables. Place of articulation was strikingly improved by filtering for one syllable, [ni], but it decreased for [mi] and [mu]. The opposite effects of filtering on [mi] and [ni] suggest that, rather than improving place of articulation perception, the filtering introduced a bias to perceive /n/. No striking differences among individual syllables were observed with regard to perception of nasal manner.

In summary, these results do not support the adaptation hypothesis. It is possible, of course, that perceptual benefits of spectral enhancement are obtained only when a murmur is physically present. If so, however, the implication would be that the crucial spectral relationships are computed at a higher level, rather than being directly available in the auditory system.

IX. SUMMARY AND CONCLUSIONS

As was already clear from earlier research, the murmur and vowel portions of nasal-consonant-vowel syllables do not make independent contributions to place of articulation perception; their relationship also plays a role. (For a recent convincing demonstration of the general importance of spectral change information in speech perception, see Furui, 1986.) This finding, which is strongly supported by the present results, argues against models of perceptual integration based on spectrographically defined cues, which do not take relational information into account. Such models have, more or less explicitly, formed the basis of much past research on speech perception (e.g., Massaro and Oden, 1980; Repp, 1982). While they may be accurate when the cues represent different (e.g., spectral versus temporal) aspects of the speech signal, they need to be augmented by a relational term when both cues are from the same physical dimension.

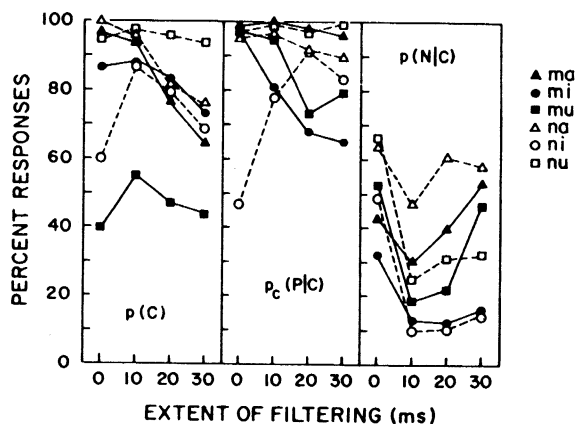


FIG. 8. Results for individual syllables in experiment 7.

The focus of the present series of experiments was the question of how listeners extract spectral relationships from the acoustic signal. That the auditory system computes some kind of running Fourier transform of the input has been an unquestioned underlying assumption. Given this assumption, there are two ways in which a listener may derive relational spectral information: directly, through auditory transforms caused by peripheral adaptation, or indirectly, through a central comparison of the spectra of successive signal portions. These two processes are not mutually exclusive: Although central comparisons seem superfluous after peripheral processes have done the work, they may substitute for peripheral processes that are artificially disrupted, and they may also serve to compute higher-order patterns of change (e.g., the second derivative of the input). The effect of adaptation in nasal-consonant-vowel syllables would be to enhance the spectral change at vowel onset and beyond. According to the strong version of the adaptation hypothesis espoused by Kurowski and Blumstein (1984), the resulting direct auditory representation of the spectral relationship would be the one and only place of articulation cue, making any further integration higher up in the system unnecessary. According to a weaker version of the hypothesis, the information obtained from the modified vowel onset is combined with cues obtained independently from preceding and following signal portions. The weaker version was considered more realistic because human listeners clearly have the ability to combine multiple sources of information and will make use of that ability whenever multiple sources are available. Peripheral auditory processes do not seem to have the integrative power to combine temporally distributed phonetic information. On the contrary, it was argued that adaptation helps differentiate the signal into contrasting auditory components.

From a review of the physiological and psychoacoustic literature, it was concluded that short-term adaptation almost certainly does take place in the human auditory system during speech perception. The internal representation of the auditory signal from which phonetic information is derived, particularly at points following rapid spectral change, is, therefore, different from the one visible in a spectrogram or oscillogram. However, does adaptation have any consequences for the intelligibility of speech? Summerfield *et al.* (1984) have pointed out some putative general advantages, such as improvement of the signal to noise ratio, but such advantages exist only relative to a hypothetical auditory system or speech recognition device in which no adaptation occurs. The former may not exist, since adaptation may well be a general design feature of neural systems. As to the latter, it should be noted that adaptation can only enhance existing spectral change, not create it. Its perceptual effect is thus comparable to a lowering of the threshold for spectral change detection on an arbitrary scale, which a machine can easily emulate, and whose net effect is zero. Thus there is perhaps no real "advantage" to be had from adaptation and spectral enhancement, except perhaps when the spectral change is right at the detection threshold. Similar conclusions have been drawn from studies of the effects of bandwidth narrowing and spectral enhancement on speech intel-

ligibility in the hearing impaired (Summerfield *et al.*, 1985; Leek *et al.*, 1987).

It is still meaningful, however, to ask whether any perceptual *disadvantage* results from a reduction of adaptation, achieved by stimulus manipulations in the laboratory. The problem here is that such manipulations may have repercussions at all levels of the system, so it is not clear whether a performance decrement results specifically from the absence of peripheral spectral enhancement or from interference with a more central process of spectral comparison or integration. This problem beset experiments 1-3, in which auditory short-term adaptation was interfered with and identification performance decreased accordingly. Had it not decreased at all, this would have been evidence that adaptation plays no role in the perception of prevocalic nasal consonants. As it was, the only indication that adaptation is perhaps unimportant was the rather small decrease in intelligibility consequent upon temporal separation of murmur and vowel portions (experiment 3).

Experiment 4 added two other relevant findings. Reduction of murmur duration, which presumably diminished the degree of adaptation, caused a performance decrement, but only at the very shortest duration. Although a ceiling effect may have imposed some limits, this finding is somewhat unfavorable to the adaptation hypothesis. The other finding was that mismatched murmurs did not lead to a performance decrement in [-a] and [-u] syllables, which confirmed a prediction of the adaptation hypothesis. A very different result was obtained with [-i] syllables, however, which was more difficult (but not impossible) to reconcile with the adaptation hypothesis. All in all, the hypothesis emerged relatively unscathed from experiments 1-4.

Experiment 5 considered two alternative hypotheses, neither of which received much support. First, place of articulation perception was no more accurate for stimuli whose nasal manner was correctly perceived. Second, smoothing the abrupt stimulus onset caused by removal of the murmur engendered only a small improvement in identification performance—not enough to account for the high intelligibility of combined murmur and vowel onset cues.

The adaptation hypothesis was still viable at this point. Experiments 6 and 7, however, yielded results that were clearly contrary to its predictions: A simulation of spectral enhancement at the onset of isolated vowel portions generally harmed, rather than improved, place of articulation identification. It may be argued that the situation was too artificial, and that spectral change information can be utilized only when the signal portion preceding the change (the murmur) is physically present. This objection, however, would be tantamount to saying that spectral change information is obtained by a more central computational process, rather than by peripheral adaptation. Or, in other words, it is the spectral change itself that is perceptually important, and not its auditory transformation through adaptation.

To compute the relationship between two stimulus components, it seems necessary that relatively analog representations of these components be available to the central nervous system. Once the murmur has been processed separately and encoded as a vector of categorical possibilities (Chistovich,

1985; Massaro and Oden, 1980), there is no way of recovering spectral relationships during processing of the vowel. This consideration points to *auditory memory* as a mediator in the central perceptual integration of stimulus components. That is, listeners may be able to hold on to a relatively faithful auditory representation of the nasal murmur even across a stretch of intervening noise or silence, and to compare that memory trace to the vowel onset spectrum. Moreover, even though the temporal separations employed in the present experiments are within the range of short-term auditory storage (Cowan, 1984), it seems likely that listeners rely on long-term auditory storage in making spectral comparisons, one reason being that the vowel would tend to "overwrite" the murmur in a sensory buffer (Cowan, 1984). Long-term auditory storage may last for a number of seconds, depending on the amount of detail to be retained. Even a life span of 1 s would be more than sufficient to account for the findings of the present study. This explanation is consistent with the very gradual decline in performance as a function of temporal separation.

Why are the murmur and vowel components integrated at all? The auditory adaptation hypothesis advanced by Kurowski and Blumstein (1984) was an attempt to provide a low-level explanation: Integration is assumed to occur because of general principles of auditory processing, and the speech perceiver merely needs to "pick up" the neatly parceled, unitary auditory properties to arrive at phonetic judgments. It seems, however, that auditory operations alone are insufficient to account for the perceptual integration of speech components. Indeed, it is not the signal portions themselves that are integrated (i.e., they remain audible as separate auditory events; this is even more obvious in the case of fricative-vowel syllables, for example) but the *information* they convey. The information, to deserve that name, must inform the listener about some event he or she has learned (or was born) to recognize. The rationale for information integration thus must be sought in the listener's mental representations of common speech patterns, which, in turn, reflect the regular occurrences of acoustic (and articulatory) events in speech production (see also Repp, 1987a,b). That is, the cues provided by the nasal murmur and by the following vowel are "integrated" because they, and their relationship (i.e., the pattern of spectral change reflecting articulatory movement), all contribute information about place of articulation of prevocalic consonants, and because listeners *know* this from long experience with speech as individuals and as members of the human species. In other words, the perceptual integration of the articulatory information conveyed by auditorily distinct speech components is a centrally guided, not a peripheral phenomenon. It reflects the listener's knowledge of the way speech is patterned, not principles governing the operation of the auditory system.

ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD-01994 and by BRS Grant RR-05596 to Haskins Laboratories. I am grateful to Katyanee Svastikula and Hwei-Bing

Lin for extensive help with stimulus preparation and running subjects. Critical comments by Quentin Summerfield and an anonymous reviewer contributed substantially to the improvement of the manuscript. Results from experiment 4 were reported at the Twenty-Sixth Annual Meeting of the Psychonomic Society in Boston, MA, in November 1985.

¹Because of frequent perfect scores, an arcsine transformation of proportions was not used. It is believed that none of the conclusions would have changed, had such a transformation been applied.

²Although it seemed at times as if isolated murmurs elicited a response bias in favor of /m/ (cf. also, Malécot, 1956), this tendency may indicate that labial place of articulation is more effectively conveyed by the murmur spectrum than is alveolar place. It also depends on the original vocalic context in a way that can be rationalized by reference to speech production (Repp, 1986). It is not clear, therefore, whether a meaningful distinction between discriminability and response bias can be made.

³A related result has been obtained by Whalen and Samuel (1985), who substituted a nonspeech noise for the initial 60 ms of the vowel in fricative-vowel syllables and found that classification reaction time was slowed when the fricative noise had been cross spliced from a different vocalic context. That is, listeners detected subtle phonetic mismatches between fricative noise and vowel across a 60-ms intervening noise, just as they did when no noise was present. The detection of such mismatches may rest on the extraction of spectral difference information from the speech signal.

"Signal-correlated noise is spectrally uniform (Schroeder, 1968) but preserves the amplitude envelope of the replaced signal, which may aid listeners in "restoring" missing phonetic information (see Warren, 1984; Whalen and Samuel, 1985); if anything, however, the noise interfered more with consonant identification than did silence. In a recent study using similar methods, Parker and Diehl (1984) likewise found no difference between the effects of intervening noise and silence on vowel identification performance in "centerless" CVC syllables, and Whalen (1984) also found effects of fricative-vowel mismatches across an intervening 60-ms silent interval, just as he did across an intervening noise (Whalen and Samuel, 1985).

⁴The artificial murmurs were used to have better control over murmur duration and amplitude contour, and slightly truncated vowels were employed to avoid ceiling effects in performance. The truncation was less than in experiments 1-3, but for no stringent reason; as before, it was assumed that truncation would merely reduce the information available without changing basic auditory and perceptual processes.

Abbas, P. J., and Gorga, M. P. (1981). "AP responses in forward-masking paradigms and their relationship to responses of auditory-nerve fibers," *J. Acoust. Soc. Am.* **69**, 492-499.

Abramson, A. S., Nye, P. W., Henderson, J. B., and Marshall, C. W. (1981). "Vowel height and the perception of consonantal nasality," *J. Acoust. Soc. Am.* **70**, 329-339.

Bell-Berti, F., Baer, T., Harris, K. S., and Niimi, S. (1979). "Coarticulatory effects of vowel quality on velar function," *Phonetica* **36**, 187-193.

Carden, G., Levitt, A., Jusczyk, P. W., and Walley, A. (1981). "Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place," *Percept. Psychophys.* **29**, 26-36.

Chistovich, L. A. (1985). "Central auditory processing of peripheral vowel spectra," *J. Acoust. Soc. Am.* **77**, 789-805.

Cowan, N. (1984). "On short and long auditory stores," *Psychol. Bull.* **96**, 341-370.

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.

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.

Eggermont, J. J. (1985). "Peripheral auditory adaptation and fatigue: A model oriented review," *Hear. Res.* **18**, 57-71.

Furui, S. (1986). "On the role of spectral transition for speech perception," *J. Acoust. Soc. Am.* **80**, 1016-1025.

- Green, D. M. (1983). "Profile analysis: A different view of auditory intensity discrimination," *Am. Psychol.* **38**, 133-142.
- 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.
- 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," *J. Acoust. Soc. Am.* **76**, 391-404.
- Larkey, L. S., Wald, J., and Strange, W. (1978). "Perception of synthetic nasal consonants in initial and final syllable position," *Percept. Psychophys.* **23**, 299-312.
- Leek, M. R., Dorman, M. F., and Summerfield, Q. (1987). "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **81**, 148-154.
- 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.
- 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.
- Parker, E. M., and Diehl, R. L. (1984). "Identifying vowels in CVC syllables: Effects of inserting silence and noise," *Percept. Psychophys.* **36**, 369-380.
- Plomp, R. (1964). "Rate of decay of auditory sensation," *J. Acoust. Soc. Am.* **36**, 277-282.
- 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.
- 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.
- Repp, B. H. (1986). "Perception of the [m]-[n] distinction in CV syllables," *J. Acoust. Soc. Am.* **79**, 1987-1999.
- Repp, B. H. (1987a). "The role of psychophysics in understanding speech perception," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht, The Netherlands).
- Repp, B. H. (1987b). "Integration and segregation in speech perception," in *Proceedings of the Twelfth International Congress of Phonetic Sciences*, Vol. 2 (Tallinn, Estonia, USSR).
- Schroeder, M. R. (1968). "Reference signal for signal quality studies," *J. Acoust. Soc. Am.* **44**, 1735-1736.
- Smith, R. L. (1979). "Adaptation, saturation, and physiological masking in single auditory nerve fibers," *J. Acoust. Soc. Am.* **65**, 166-178.
- Summerfield, Q., and Assmann, P. (1987). "Auditory enhancement in speech perception," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht, The Netherlands).
- Summerfield, Q., Foster, J., and Tyler, R. (1985). "Influences of formant bandwidth and auditory frequency selectivity on identification of place of articulation in stop consonants," *Speech Comm.* **4**, 213-229.
- Summerfield, Q., Haggard, M., Foster, J., and Gray, S. (1984). "Perceiving vowels from uniform spectra: Phonetic exploration of an auditory after-effect," *Percept. Psychophys.* **35**, 203-213.
- Viemeister, N. F., and Bacon, S. P. (1982). "Forward masking by enhanced components in harmonic complexes," *J. Acoust. Soc. Am.* **71**, 1502-1507.
- Warren, R. M. (1984). "Perceptual restoration of obliterated sounds," *Psychol. Bull.* **96**, 371-383.
- Westerman, L. A., and Smith, R. L. (1984). "Rapid and short-term adaptation in auditory nerve responses," *Hear. Res.* **15**, 249-260.
- Whalen, D. H. (1984). "Subcategorical phonetic mismatches slow phonetic judgments," *Percept. Psychophys.* **35**, 49-64.
- Whalen, D. H., and Samuel, A. G. (1985). "Phonetic information is integrated across intervening nonlinguistic sounds," *Percept. Psychophys.* **37**, 579-587.
- Widin, G. P., and Viemeister, N. F. (1979). "Intensive and temporal effects in pure-tone forward masking," *J. Acoust. Soc. Am.* **66**, 388-395.
- Wilson, J. P. (1970). "An auditory after image," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).
- Zwislocki, J., Pirodda, E., and Rubin, H. (1959). "On some poststimulatory effects at the threshold of audibility," *J. Acoust. Soc. Am.* **31**, 9-14.