

Perceptual parsing of acoustic consequences of velum lowering from information for vowels

CAROL A. FOWLER and JULIE M. BROWN
Haskins Laboratories, New Haven, Connecticut
and University of Connecticut, Storrs, Connecticut

Three experiments were designed to investigate how listeners to coarticulated speech use the acoustic speech signal during a vowel to extract information about a forthcoming oral or nasal consonant. A first experiment showed that listeners use evidence of nasalization in a vowel as information for a forthcoming nasal consonant. A second and third experiment attempted to distinguish two accounts of their ability to do so. According to one account, listeners hear nasalization in the vowel as such and use it to predict that a forthcoming nasal consonant is nasal. According to a second, they perceive speech gestures and hear nasalization in the acoustic domain of a vowel as the onset of a nasal consonant. Therefore, they *parse* nasal information from a vowel and hear the vowel as oral. In Experiment 2, evidence in favor of the parsing hypothesis was found. Experiment 3 showed, however, that parsing is incomplete.

Theories of speech perception differ in their characterization of phonological categories and of the stimulus information that listeners are presumed to use to detect the phonological properties of spoken utterances. In some theoretical accounts, phonological categories are cognitive categories in the mind; listeners identify intended phonological properties of an utterance from associated acoustic and, sometimes, optical cues. The results of initial stages of auditory (visual) analysis of a speech stimulus are mapped onto the mental phonological categories. In other accounts, phonological categories are gestural in nature; listeners identify the phonological properties of utterances by extracting acoustic and, sometimes, optical information about them that is available because gestural causes have specifying acoustic and, sometimes, optical consequences.

As distinct as these views are conceptually, they have proven difficult to distinguish experimentally. Generally, researchers agree on what serves as acoustic information for the phonological properties of words. They disagree on what the acoustic information serves as information for: abstract, mental categories or phonological actions of the vocal tract.

One approach that we have taken to try to distinguish the views experimentally has been to look closely at listeners' modes of attention to the acoustic speech signal. Although researchers generally agree on what serves as acoustic information, there may be some cases in which listeners who are attempting to extract information about gestures should differ in their mode of attention to the

acoustic speech signal from listeners who are attempting to find cues for abstract phonological categories.

In particular, for gesture perceivers, the whole array of (detectable) acoustic consequences of a gesture should jointly specify its source. Because, owing to coarticulation, different gestures have converging effects on common acoustic dimensions, listeners must *vector analyze* (Fowler & Smith, 1986) or *parse* (Fowler, 1996) the signal. That is, they must track the acoustic speech signal along gestural lines, parsing acoustic information for one gesture from information for another, even when the gestures have converging effects on common acoustic dimensions. Listeners who undertake initial stages of speech-independent auditory analysis of the acoustic speech signal and then map the consequences onto abstract phonological categories are not expected to parse the signal along gestural lines.

To date, gestural parsing has been shown to occur in two kinds of domains: segment-segment coarticulation (Fowler, 1984; Fowler & Smith, 1986) and segment-prosody coarticulation (Fowler & Brown, 1997; Pardo & Fowler, 1997). Parsing is shown jointly by two kinds of findings. First, listeners use information about a gesture *x* that, owing to coarticulation, occurs in the domain of acoustic information predominantly about another gesture *y* as information for *x*. Second, their ability to do so does not depend on their hearing gesture *y* as the context-sensitive version, *y_c*. To the contrary, *y* is shown not to sound context sensitive to the listener.

Two examples may make these kinds of findings clear. In utterances /ibəbi/, /abəbi/, /ibəba/, and /abəba/, the portion of the acoustic signal in which the schwa vowel predominates is, nonetheless, affected by acoustic consequences of the surrounding gestures, including those of the flanking vowels. Listeners have been shown to use acoustic information for a coarticulating vowel in the domain of schwa as information for the coarticulating vowel (Fowler & Smith, 1986). This has been shown by cross-

The research was supported in part by NICHD Grant HD-01994 to Haskins Laboratories. Correspondence concerning this article should be addressed to C. A. Fowler, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511 (e-mail: fowler@haskins.yale.edu).

—Accepted by previous editor, Myron L. Braunstein

splicing /bə/ from one vocalic context into another and asking listeners to identify the final vowel of each trisyllable as quickly as they can. Response times to identify a final /i/ or /a/ were slowed and errors marginally increased when the acoustic signal during schwa contained misleading information about the forthcoming vowel.

These results could come about in either of two ways. The schwa vowel might sound context sensitive (i.e., it might sound to listeners like an /i/- or /a/-colored vowel), and listeners might use the vowel coloring to predict the forthcoming vowel. Alternatively, listeners may use information for /i/ or /a/ during schwa as information for the onset of the /i/ or /a/ gesture. Discrimination experiments suggest the latter interpretation (Fowler, 1981; Fowler & Smith, 1986). In that research, listeners judged the schwa vowel in *i₁bə₁bi* (in which subscripts show the original context in which the /bə/ syllable had been produced) as being more like the schwa in *a₁bə₁ba* than like that in *a₁bə₁ba*, even though the schwas in *i₁bə₁bi* and *a₁bə₁ba* were acoustically identical. This is consistent with listeners extracting from the domain of schwa the coarticulatory effects that the flanking vowels should have had on it and hearing the residuals as schwa.

Similar findings are obtained when listeners are presented with fundamental frequency contours in which two gestures have had converging effects. For example, there is a well-known, universal tendency (see, e.g., Whalen & Levitt, 1995) for high vowels, such as /i/ or /u/, to be associated with higher fundamental frequencies, other things being equal, than are low vowels, such as /a/. Although the reason for the association between vowel height and fundamental frequency height is unknown, the observation that the effect occurs in infant vocalic sounds (Whalen, Levitt, Hsaio, & Smorodinsky, 1995) and in languages independently of the size of the vowel inventory (Whalen & Levitt, 1995) suggests that the effect is an automatic consequence of high or low vowel production and not, as has been suggested, an enhancement of the acoustic distinctiveness of high and low vowels (e.g., Diehl & Kluender, 1989). Listeners also behave as if the fundamental frequency *perturbation* that is due to vowel height is part of the vowel, not part of the intonation contour. They use a high fundamental frequency as information that an ambiguous vowel is higher rather than lower (Reinholt Peterson, 1986), and they hear /i/ and /a/ vowels that are equal in fundamental frequency as being different in pitch. /i/ sounds lower than /a/ when the fundamental frequency is the same (Fowler & Brown, 1997; Silverman, 1987).

Where estimates of parsing accuracy have been made, the estimates differ. For example, Silverman (1987) found near-perfect extraction of fundamental frequency perturbations that were due to vowel height from perception of intonational accent; that is, listeners ascribed close to the measured 10–15 Hz /i/–/a/ difference in “intrinsic” fundamental frequency to /i/’s vowel height (and did not hear it as part of the intonational accent). However, in different assessments, Fowler and Brown (1997) found

accurate parsing, excessive parsing, and insufficient parsing of the effects of vowel height differences on pitch judgments.

All of these outcomes may reflect perception outside the laboratory under some conditions. Ohala (1981) has suggested that both insufficient parsing and excessive parsing may underlie sound changes that commonly occur. For example, in many languages (e.g., French), phonologically nasal vowels developed historically in the context of nasal consonants that followed them, which weakened and were lost. Perhaps, as the consonants weakened, they became inaudible, and listeners could not ascribe the (still audible) nasalization of the vowels to any source other than the vowels themselves (cf. Ohala, 1981). As for excessive parsing, Ohala provides an example from the language Shona, in which a labiovelar glide became velar in the context of following a labial consonant.

Parsing of coarticulatory nasalization in a vowel is the focus of our study. It is a new kind of example for the investigation of gestural parsing in that relevant sequences, vowels followed by nasal consonants, involve at least three gestures—the vocalic gesture, the consonant’s oral constriction gesture, and the velum-lowering gesture, which overlaps temporally with both of the other gestures. An interesting question, then, is whether the perceived nasality is ascribed both to the vowel and to the consonant, in accord with the temporal overlap, or to the consonant only, in accord with each segment’s phonological or gestural properties. (A third option appears to be ruled out by phenomenal experience and by spelling conventions. The nasal gesture is not heard as a third, distinct segment located serially between an oral vowel and an oral consonant.) The production of vowel–nasal sequences has been studied, and the gestural composition of such sequences is understood fairly well. From the perspective of a theory that listeners perceive gestures, these findings clearly suggest the prediction that listeners will parse the nasal gesture from the vowel and will ascribe it to a following nasal consonant. This is because (excluding from consideration any lowering of the velum that may occur at vowel onset, if the vowel is preceded by an oral consonant; see, e.g., Bell-Berti, 1980) there is just one velum-lowering gesture in such sequences, and it appears to be phased with respect to the oral constriction for the consonant, not with respect to the vowel (see, e.g., Bell-Berti & Krakow, 1991; Krakow, 1989). Accordingly, sensitive gesture perceivers should hear nasalization in association with the consonant, and not in association with the vowel.

However, the occurrence of sound changes involving nasalization suggests that underparsing may sometimes occur. First, as was noted, phonological nasalization of vowels has occurred in some languages in association with the weakening and ultimate loss of a following nasal consonant. During some phase of the sound change, it is likely that members of such language communities heard the coarticulatorily nasalized vowels as nasal. Second, however, synchronically, in languages in which vowels

are present in phonologically oral–nasal pairs, members of a pair often exhibit height differences consistent with underparsing. Beddor (1983) showed that, overall, the nasal resonance in the vicinity of F_1 raises the centroid of spectral energy in that region in low vowels and lowers it in high vowels. Consistent with that effect, Beddor found, in a survey of 75 relevant languages, that, quite consistently, where there was a height difference between oral and nasal vowels, high nasal vowels were lower and low nasal vowels were higher than their oral counterparts. Compatible, too, are findings by Wright (1986) of an overall shrinking of the perceptual space of isolated nasalized vowels, as compared with isolated oral vowels.

The research on the perception of vowel–nasal sequences that is relevant to the issue of the accuracy of nasal gesture parsing provides a variety of outcomes. In NVN (nasal consonant–nasalized vowel–nasal consonant) sequences, Kawasaki (1986) found that listeners (but only those who were phonetically untrained) judged vowels to be more nasalized as the flanking nasal consonants were increasingly attenuated in amplitude. This is consistent with the interpretation that at least some parsing occurs. Accordingly, when the nasal consonants are clearly audible, listeners ascribe the nasality during the vowel to the consonants and do not hear the vowels as (strongly) nasalized. Only when there is no audible source for the nasality in the vowel does parsing necessarily fail to occur. Findings by Krakow, Beddor, Goldstein, and Fowler (1988) suggest wholly accurate parsing. They found vowel identification differences consistent with perceived height differences between nasalized and oral vowels in a b_d context, but no quality changes at all for the nasalized vowels presented in b_nd contexts.¹

In contrast to these findings, in an ABX task with trials of the sort (among others)

A	B	X
NVN	NVN	V

Krakow and Beddor's (1991) participants were quite accurate in matching V to NVN. However, the prediction, based on the theory that gestures are perceived and that gestural parsing is accurate, is that participants would be systematically wrong on trials such as these, because, with nasality parsed from \tilde{V} in NVN, it should sound like V. Furthermore, parsing the effects that nasalization should have had from V in NVN should make it sound different from V.

The present study looks at nasal parsing, using techniques previously used to examine segment–segment coarticulatory parsing (Fowler & Smith, 1986). Thus, we ask two questions. First, do listeners use coarticulatory nasal information in a vowel as information that a nasal consonant is forthcoming? Second, if they do, is it because they hear a context-sensitive vowel (specifically, a nasalized vowel) or because they parse the nasalization from the vowel and hear it as the start of the consonant?

With respect to the first question, using an off-line task (gating), Marslen-Wilson and colleagues (Lahiri &

Marslen-Wilson, 1991; Warren & Marslen-Wilson, 1987) provided a positive answer. English listeners were more likely to guess a word ending with a nasal consonant when given a CV \tilde{V} fragment from a CVN than when given a CV fragment from a CVC. However, listeners were not striking in their accuracy. Lahiri and Marslen-Wilson reported that 83.4% of the responses were CVCs, given a CV fragment; 16.6% of the responses were CVN. (Lahiri and Marslen-Wilson pointed out that this breakdown not only corresponds well with the acoustic evidence that listeners were given, it also mirrors the relative frequency of CVCs and CVNs in English.) When given a CV \tilde{V} fragment, the proportion of CVC responses fell, but only to 59.3% of responses. Listeners still guessed a CVC most of the time.

In Experiment 1, the same question is asked, but using an on-line, rather than an off-line measure. When listeners must classify a consonant as oral or nasal as quickly as they can, are classification response times and accuracies affected by information in a preceding vowel that a forthcoming consonant will be nasal or oral?

EXPERIMENT 1

Method

Participants. The participants were 18 native speakers of English, who reported having normal hearing. They received course credit for their participation.

Stimulus materials. Utterances used in the experiment were natural productions spoken by a male native speaker of English. In an initial study, 6 talkers, 3 male and 3 female, all native speakers of American English, produced 10 tokens each of the nonsense disyllables /bamə/, /banə/, /babə/, and /badə/. Acoustic analysis of /a/ (near the acoustically defined midpoint) revealed several kinds of differences between nasalized and oral vowels. Most consistently across talkers, F_1 's bandwidth was broader (numerically for all 6 talkers, significantly for 4), and its amplitude was lower in the nasalized vowels (numerically for all 6; significantly for 5). Some talkers also showed a shift in F_1 frequency, but the direction of the effect was not consistent across talkers. Four talkers showed numerical raising of F_1 (3 significant); 2 showed lowering (1 significant). Statistically significant effects on F_2 frequency (2 talkers showing raising), bandwidth (1 widening, 1 narrowing) and amplitude (1 increase in amplitude) were considerably less common than significant effects on F_1 . The talker selected to provide disyllables for the present experiments showed no effect of nasalization on the frequency of F_1 (numerically 10 Hz lowering), but significant broadening of its bandwidth and marginal lowering of its amplitude ($p = .06$). He showed no significant effects of nasalization on F_2 . The talker was not chosen on the basis of the acoustic measurements; he was the first male talker recorded; in isolation, his nasal vowels sounded nasalized.

For Experiment 1, two tokens of each of the chosen talker's utterances of /bamə/, /banə/, /babə/, and /badə/ were selected. Except for /banə/, they were the first two tokens of the disyllables that the talker had produced. In the case of /banə/, we chose the first and third tokens, because we found the second token difficult to distinguish from /bamə/. The disyllables were filtered at 11 KHz and sampled at 22 KHz (using SoundScope, GW Instruments, Somerville, MA).

The first and second syllables of the disyllables were then spliced or cross-spliced to create 16 new disyllables. In spliced disyllables, /ba/ from one token of a disyllable type was spliced onto Nə or C(oral consonant)ə from the other token of the same type. For ex-

ample, /ba/ from the first token of /bamə/ (henceforth, bam1) was spliced onto /mə/ from the second token of /bamə/ (henceforth, mə2) to make bam1mə2. Splicing cuts were made at the onset of consonant closure, as judged from waveform and spectrographic displays of the disyllables. In the same way, bam2mə1 and analogous versions of disyllables involving the other three consonants were created. In cross-spliced tokens, the first syllable from a disyllable having a second syllable nasal consonant was spliced onto the second syllable of a disyllable having a same-place oral consonant in the second syllable. Using the notation above, for example, we created bam1bə1, bam2bə2, ban1də1, and ban2də2. Opposite cross-splicings were created as well: bab1mə1, bab2mə2, bad1nə1, and bad2nə2. In these disyllables, the perception was of a CVCV or CVNV, with the second C or the N determined by the consonant in the second syllable. In spliced disyllables, the first vowel provided accurate information about the nasality (and place) of the forthcoming consonant; in cross-spliced disyllables, place information was accurate, but nasality was not.

Each of the 16 spliced or cross-spliced disyllables appeared once in each of 10 blocks of trials; the order of the disyllables was randomized differently for each block.

Procedure. The participants sat facing a computer screen with their left and right forefingers resting on two buttons of a button box. The left button was labeled "M or N" and the right button was labeled "B or D." On each trial, the screen presented a trial number. Shortly thereafter, listeners heard (via a 12-bit D/A converter) one of the disyllables over headphones, and their task was to hit either the left or the right button to signal whether the consonant they heard was M or N (left hand) or B or D (right hand). (Response and responding hand were not counterbalanced, because we were not interested in response time differences as a function of whether the second consonant was nasal or oral.) The participants were instructed to respond as quickly and accurately as they could. Following their response, the computer screen showed the participants their response time in milliseconds.

Results

Response times were measured from the splice point (the acoustically defined onset of the second syllable). For each participant, response times were excluded from analyses if they were inaccurate or if they differed from the participant's condition mean by more than 2.5 stan-

dard deviations. Response times are presented in Figure 1. The main questions of interest were whether listeners would respond more slowly and less accurately when information in the vowel misled them (cross-spliced disyllables) as to the oral or nasal character of the following consonant. As Figure 1 shows, response times to cross-spliced disyllables were slower (by 53 msec) than those to spliced disyllables. In an analysis of variance (ANOVA), with splicing condition (spliced, cross-spliced) and vowel nasality (oral, nasalized) as factors, the effect of splicing condition was highly significant [$F(1,17) = 76.35$, $MS_e = 49,637$, $p < .0001$]. The effect of vowel nasality was also significant, but much smaller [$F(1,17) = 8.07$, $MS_e = 2,279$, $p = .011$]; responses following nasalized vowels were faster (by 11 msec) than those following oral vowels. The interaction was marginal [$F(1,17) = 3.63$, $MS_e = 4221$, $p = .074$]. Response times slowed by 37 msec when a nasalized vowel was followed by an oral rather than a nasal consonant [$F(1,17) = 10.71$, $MS_e = 12,450$, $p = .0045$], but they slowed by 68 msec when an oral vowel was followed by a nasal, as compared with an oral, consonant [$F(1,17) = 35.60$, $MS_e = 41,397$, $p < .0001$].

Accuracy was very high, averaging .98 across conditions; in fact, it averaged .98 in each of the four conditions. Accordingly, there were no differences in accuracy between responses to spliced and those to cross-spliced disyllables.

Discussion

The results are clear in showing that listeners use information in a vowel that a forthcoming consonant is or is not nasal. As we discussed in the introduction, this can have either of two interpretations. The coarticulatorily nasalized vowel may sound nasalized to the listener, who uses the context sensitivity in the vowel to predict the context. Alternatively, the listener may have parsed the nasal gesture from the vowel and may hear it as the onset of the consonant. Experiments 2 and 3 were designed to distinguish these interpretations.

Before turning to those experiments, however, we consider an unexpected (albeit only marginally significant) outcome of this experiment. Listeners were more slowed by a nasal consonant that followed an oral vowel than by an oral consonant that followed a nasal vowel. Possibly, this pattern is similar to that found by Lahiri and Marslen-Wilson (1991), using gating. When given a CV fragment, listeners guessed a CVC word approximately 83% of the time and a CVN 17%; when they were given a C \bar{V} fragment, the percentages were 59% and 41%. If these response percentages reflect expectancies, nasal consonants are very much unexpected following an oral vowel. One interpretation of this outcome is that two factors underlie guesses. One is information in the vowel; the other is lexical frequency, with CVCs being considerably more frequent than CVNs (or, in the present experiment using nonsense words, an effect of lexical neighborhood frequency).

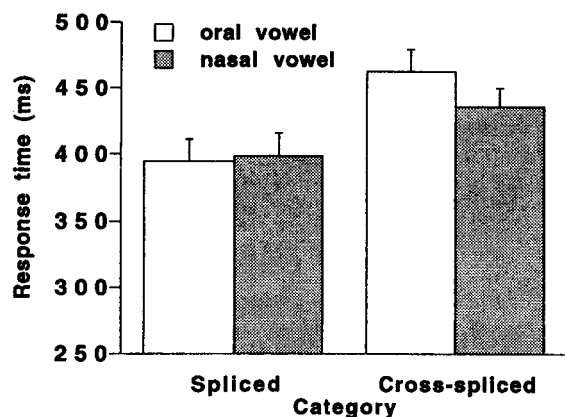


Figure 1. Response times (with standard error bars) to spliced and cross-spliced disyllables of Experiment 1.

Accordingly, given a CV as a first syllable, listeners develop a strong expectation for a following oral consonant that is met in the spliced condition but violated in the cross-spliced conditions. Expectations are weaker given C \tilde{V} as a first syllable, and the effects of cross-splicing are correspondingly weaker.

This explanation may not be wholly satisfactory, in part because the results of Lahiri and Marslen-Wilson (1991) in fact imply that C \tilde{V} s give rise to weak expectations of a following oral, not nasal, vowel and because we do not see response times favoring CVCs in the spliced condition.

A different factor that Lahiri and Marslen-Wilson (1991) considered in order to explain their gating data, on both English and Bengali listeners (Bengali being a language with phonologically nasalized as well as contextually nasalized vowels), is *underspecification*. In phonological theory (e.g., Archangeli, 1988), because nasality in the vowels of English is predictable, vowels in lexical entries for words do not include either + or – nasal as a feature. Accordingly, when a CV fragment is presented to the lexicon, it is wholly consistent with the complete word's being a C \tilde{V} N as well as with its being a CVC. Similarly, when a C \tilde{V} fragment comes in, although it has acoustic properties that are not associated with any of the specified features of a vowel in a lexical entry, the fragment does not mismatch entry specifications of CVC words, because vowels in the lexicon have no specifications for nasality for the acoustic information to mismatch.

It is unlikely that this factor explains the response pattern of Experiment 1 either (as perhaps it should not; stimuli in the experiment do not have lexical entries). The slowest response times were obtained for stimuli in which an oral vowel was followed by a nasal consonant, but this is a juxtaposition that should be as good a match as an oral vowel followed by an oral consonant in underspecified lexical entries.

For the present, the finding is unexplained (and its repeatability both across tokens and across talkers is unknown).

EXPERIMENT 2

The purpose of the next experiment was to address the question of the basis on which responses to cross-spliced items might have been slowed. One possibility is that nasal vowels, audibly nasalized to listeners, led them to predict that a following consonant would be nasal, and correspondingly, audibly oral vowels led them to predict that a forthcoming consonant would be oral. Alternatively, in a parsing account, listeners may not hear the vowel as nasal but, rather, may use nasality in the domain of the vowel as information for the beginning of the production of a nasal consonant. This misleads listeners when the consonant turns out to be oral. A third possibility is that both of these accounts may be accurate, because parsing is partial.

In a first effort to test for parsing, we used synthetic speech so that we could vary information for nasality along a continuum and, thereby, provide a sensitive test.

We used an AXB paradigm in which A and B were /amə/ and X was /abə/ or vice versa. The initial /a/ vowels varied in the magnitude of information for a nasal gesture co-occurring with vowel production. Our prediction was that, on average, listeners would consider /amə/ in the X position in a triad as being a closer match to flanking /abə/s having /a/s of lower nasality than itself. That is, if listeners parse nasality from the vowel and ascribe it to the nasal consonant, they should hear the vowel as less nasal than it is. (They might even hear it as oral, if parsing is perfect.) Accordingly, listeners should hear X as a better match to an /abə/ of lower nasality. Findings should be reversed if X is /abə/; that is, /abə/ should be found to match an /amə/ of greater nasality than itself, because listeners will parse nasality from the flanking /amə/s.

Method

Participants. The participants were 26 students at the University of Connecticut. Five were graduate students, who volunteered and were uncompensated; the rest were undergraduates, who received course credit. All were native speakers of English, who reported having normal hearing. Thirteen of the participants took part in the main experiment. Thirteen participated in a pretest of the adequacy of the acoustic stimuli. Data from 1 undergraduate participant in the main experiment were excluded from analyses, because the participant had misunderstood the instructions.

Stimulus Materials. In order to create the synthetic /abə/ and /amə/, we used as a guideline the measurements from the natural speech of the male speaker whose speech we had used in Experiment 1. As we described in the Method section of Experiment 1, second and third formants of the vowels /a/ and /ə/ were measured at the vowel midpoints; in addition, formant transitions were measured for the intervocalic consonant /b/ or /m/. Next, durations of the all the phonemes except the initial /b/ were taken. Finally, the fundamental frequency (*f*₀) variation for five tokens of /babə/ were measured.

To create a continuum of nasality for /a/, we varied the bandwidth of the *F*₁ in seven equidistant steps from 100 to 400 Hz. The amplitude of the nasal resonance (having a frequency of 600 Hz) was also varied in seven steps (from 0 to 35 dB in synthesizer parameter values). Therefore, the most oral token of /a/ had an *F*₁ bandwidth of 100 Hz and a nasal resonance amplitude of 0 dB. The most nasal token of /a/ had an *F*₁ bandwidth of 400 Hz, a nasal resonance frequency of 600 Hz, and a nasal resonance amplitude of 35 dB. /a/ vowels were 180 msec in duration.

The pretest was conducted on the seven synthetic /a/ tokens. We ran the test to determine whether the continuum in fact varied in perceived nasality and, if so, whether perceived nasality varied monotonically, or even linearly, with changes in acoustically specified nasality. To answer these questions, we created a 70-item test order in which each /a/ vowel appeared 10 times in random order. The test sequence was preceded by three repetitions of each endpoint of the continuum, to demonstrate the range of variation in the vowels to listeners, whose task was to rate the nasality of the vowel on a scale of 1 to 7.

For the main experiment, the seven tokens of /a/ were spliced onto either /bə/ or /mə/. The tokens of /abə/ and /amə/ were 360 msec in duration. /a/ was 180 msec in duration, and /bə/ and /mə/ were

each 180 msec in duration, including a 60 msec bilabial closure. $F1$ was 703 Hz, $F2$ was 1260 Hz, and $F3$ was 2875 Hz at vowel midpoint for /a/. $F1$ was 600 Hz, $F2$ was 1283 Hz, and $F3$ was 2850 Hz at vowel midpoint for /ə/. There was a falling and rising formant transition surrounding the bilabial closure ending or starting at 560 Hz for $F1$, 1200 Hz for $F2$, and 2750 Hz for $F3$. The bilabial closure was silent for the /b/. Nasal resonance in the closure of /m/ was achieved by increasing the amplitude of the nasal resonance from 0 dB for /b/ to 35 dB. In addition, the /a/s in /abə/ and /amə/ were given a rising $f0$ contour that began at 115 Hz, peaked during /a/ at 130 Hz, and fell to 115 Hz; during /ə/ (roughly following the measured $f0$ contour of our speaker), $f0$ began at 115 Hz, fell to 98 Hz, and then rose to a final value of 115 Hz.

From the two continua, we constructed AXB trials. In an AXB trial, three stimuli are presented in succession. For convenience, the first is called A, the second X, and the third B. The participants' task is to decide whether X is more like A or more like B. In the present study, the AXB trials were either of two types. In one trial type, A and B were versions of /abə/, and X was /amə/. In the other type, A and B were versions of /amə/, and X was /abə/. We selected triads in which the nasality of /a/ in X fell between that of the /a/s in A and B and in which A and B were three or four steps apart on the nasality continuum. For example, labeling the least nasalized /a/ vowel as /a₁/ and the most nasalized /a/ as /a₇/, a sample triad was /a₆bə/, /a₃mə/, /a₂bə/. All the triads meeting the criteria above were used. There were 17 distinct triads (listed in the Appendix) in which /amə/ was X, and 17 in which /abə/ was X. In the test, each occurred four times, in two of which the identities of A and B were switched, relative to those in the other two; this made 136 randomized trials in all.

Procedure. On the pretest, the participants were tested in small groups. They were told that they would be hearing /a/ vowels that varied in nasality. Their task was to rate the degree of nasality on a scale of 1 to 7. Before taking the test, they heard three examples of each of the continuum endpoints, to give them an idea of the full range of variability that they would be hearing.

On the main test, the participants were also tested in small groups. They were told that, on each of 136 trials, they would hear a triad of disyllables. Their task was to focus on the first vowel of each disyllable and to decide whether that in X was more like A or more

like B, writing A or B on their answer sheets. They were instructed to guess if necessary.

On both tests, there were 3.5 sec between trials.

Results

Figure 2 shows the results of the pretest. Although the participants found the task difficult, they were able to do it. Ratings increased monotonically across the continuum, averaging 2.81 at the least nasalized extreme and 5.2 at the most. In an ANOVA, with continuum member as the repeated measures factor, the differences in ratings are highly significant [$F(6,72) = 12.70, p < .0001$]. In a regression analysis in which $F1$ bandwidth was used to predict average ratings, the R^2 was .968. Accordingly, we can conclude that our continuum adequately conveys information for differential nasalization to listeners and that judged nasality varies linearly across the continuum.

In an ANOVA on judgments in the main experiment, the dependent measure was the proportion of trials on which the participants chose the less nasalized of A or B as matching X. If parsing occurs, this proportion should be high if X is /amə/ and low if X is /abə/. Independent variables were identity of X (/amə/ or /abə/) and step size. The variable step size reflected the distance of X (along the nasality continuum) from A and B. Specifically, there were five levels to this variable, each reflecting a pair of distances: the distance of X from the less nasalized of A or B and the distance of X from the more nasalized of A or B. Step sizes were 3-1 (i.e., X was three step sizes more nasalized than the less nasalized of A or B, and one step size less nasalized than the more nasalized of the two), 2-1, 2-2, 1-2, and 1-3. (The Appendix provides the results separately for each trial type.)

The idea was to find out whether, in the array of step sizes we used, we could see a point at which preference

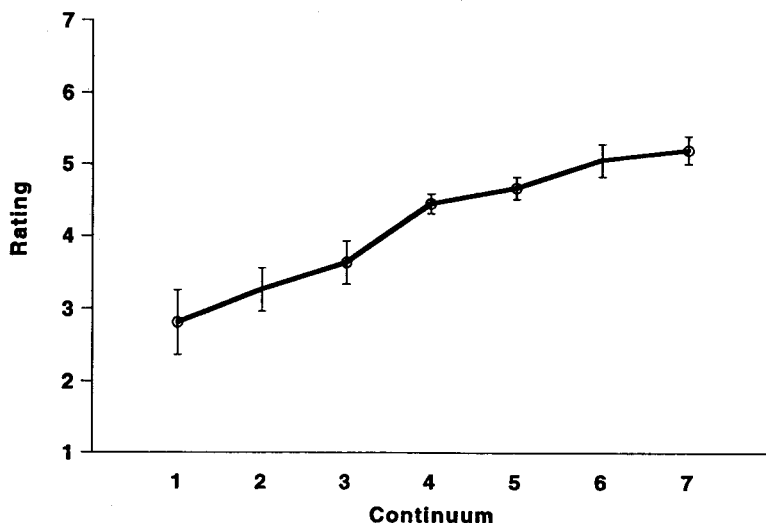


Figure 2. Average ratings of isolated synthetic vowels that vary in the magnitude of information for nasality. The data are from the pretest of Experiment 2.

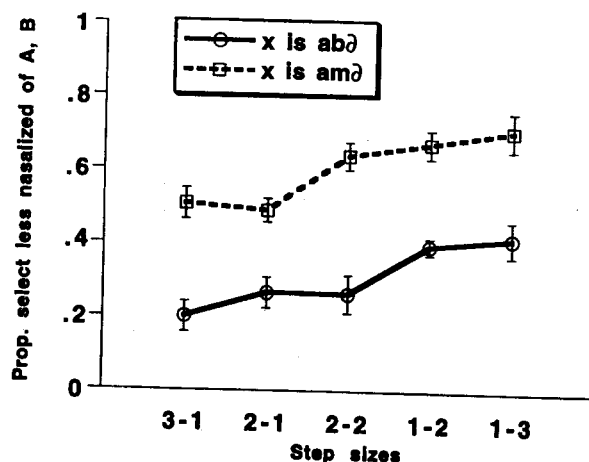


Figure 3. Proportion of trials (with standard error bars) on which listeners selected the less nasalized disyllable of A or B as matching X, as a function of the identity of X (/abə/ or /amə/) and of the continuum step size between X and the less nasalized of A or B and X and the more nasalized of A or B. The data are from Experiment 2.

for the less nasalized of A or B when X was /amə/ shifted to a preference for the more nasalized, because the distance from the less nasalized was large and that from the more nasalized was sufficiently smaller. Likewise, we looked for an opposite finding when X was /abə/. The findings are presented in Figure 3. The figure shows that listeners chose the less nasalized disyllable of A or B as better matching X more than half the time (.60) when X was /amə/ and less than half the time (.31) when X was /abə/. This difference, the main effect of X identity, was highly significant [$F(1,11) = 46.38, p < .0001$]. Furthermore, in t tests against the chance value of .50, both values differed significantly from chance [/amə/: $t(11) = 4.47, p = .0009$; /abə/: $t(11) = -6.83, p < .0001$].

Figure 3 also shows that, as X shifts from being closer to the more nasalized of A or B to being closer to the less nasalized of the pair, listeners increase their likelihood of choosing the less nasalized of A or B as more like X. This increase, the effect of step, was also significant [$F(4,44) = 11.30, p < .001$]. Although, at every step size, preference for the less nasalized of A or B fell below .5 when X was /abə/, at 1-3, the proportion selection of the less nasalized (.417) did not differ from chance [$t(11) = 1.73$]. When X was /amə/, preferences fell to chance at steps 2-1 and 3-1. In neither case did preferences shift over to the opposite category. Perhaps, had the steps 1-4 and 4-1 been included, we would have seen a shift in preference.

In the ANOVA, the interaction between identity of X and step size was not significant.

Discussion

The experiment revealed, as we had predicted, a highly significant tendency to respond as the parsing hypothesis predicts. That is, listeners matched /a/ of a given de-

gree of nasality differently when it was adjacent to /mə/ than when it was adjacent to /bə/, and the difference suggested that when /a/ was adjacent to /mə/, it sounded less nasal to listeners than when it was adjacent to /bə/. This is consistent with Kawasaki's (1986) findings that nasality ratings of a vowel increase as flanking nasal consonants are attenuated in amplitude.

A weakness of the present study is that it did not allow us to estimate the completeness of parsing. This is, in part, because our set of stimuli did not include any in which we could see a crossover from a preference for the less nasalized of A or B to one for the more nasalized when X was /amə/, and vice versa when X was /abə/. In Experiment 3, we show, using the natural speech of Experiment 1, that parsing must be far from complete.

EXPERIMENT 3

In Experiment 3, we used the natural speech of Experiment 1 and some new disyllables. An advantage of synthetic speech continua of the sort we used in Experiment 2 is that it allows us to see parsing when it is incomplete. That is, even if parsing does not change /a/ to perceived /a/₁, it may change it to /a/₆ or some other value lower than 7. By shifting back to natural speech, we require parsing, if it is to show up, to have an effect on the endpoints, because endpoints are all that natural speech offers.

We selected a vowel matching task. On each trial, listeners heard a sequence of two disyllables in which the first vowels either matched (were both /a/, as in the disyllables of Experiments 1 and 2, or were both /ʌ/) or mismatched (one was /a/ and one /ʌ/). The participants' task was to indicate by buttonpress, as quickly as possible, whether the vowels matched or mismatched.

Critical manipulations varied whether the vowels also matched or mismatched in nasality and whether following consonants were oral or nasal. There were three critical trial types, each represented by one or two examples in Table 1. Looking only at trials in which the first vowels of a disyllable pair matched (in the examples, both are /a/), in one pair type, either both vowels were nasalized and both were followed by nasal consonants (e.g., *bam1mə2 bam2mə1*), or both vowels were oral and were followed by oral consonants (e.g., *bab1bə2 bab2bə1*). From any theoretical perspective, vowel matches on these trial types should be easy to detect. In a second trial

Table 1
Sample Trial Types in Experiment 3.

- | |
|----------------------------|
| 1a. <i>bam1mə2 bam2mə1</i> |
| 1b. <i>bab1bə2 bab2bə1</i> |
| 2a. <i>bam1mə2 bam2bə2</i> |
| 2b. <i>bab1mə2 bab2bə2</i> |
| 3. <i>bam1mə2 bab1bə2</i> |

Note—numbers refer to the token, 1 or 2, from which the syllable was taken. The consonant following /ba/ in each initial syllable was the consonant that had followed it in its original context. Initial syllables all are open syllables, however.

type, either both vowels were nasalized, but now one was followed by a nasal consonant and one by an oral consonant (e.g., *bam1mə2 bam2bə2*), or the vowel was oral in both members of a pair, but one was followed by a nasal and one by an oral consonant (e.g., *bab1mə1 bab2bə1*). Listeners who parse along gestural lines and parse completely or nearly so (henceforth, *parsers*) should find these pairs more difficult to match than pairs of the first type, for two reasons. First, in a pair such as *bam1mə2 bam2bə2*, parsers will hear the first vowel of *bam1mə2* with the effects of nasalization partialled out, but they will hear the nasality in the first vowel of *bam2bə2*. This will not change the vowel from /a/ to something else, but it should slow the match decision and, perhaps, make it less accurate. In a pair such as *bab1mə1 bab2bə1*, listeners should attempt to pull from /a/ in the first disyllable the effects that nasalization should have had on the vowel spectrum. This may change its quality and make it more difficult to match to the /a/ in the second disyllable. There is a second reason why trials of Type 2 may be difficult. In contrast to trials of Type 1, when vowels are the same in trials of Type 2, consonants are different. Because there is some evidence (e.g., Tomiak, Mullenix, & Sawusch, 1987) that vowels and consonants are processed integrally (in the sense of Garner, 1974), this may slow response times and decrease accuracy. As for listeners who do not parse, they should experience the second source of difficulty only on trials of Type 2. For nonparsers or listeners who parse just a little (henceforth, *nonparsers*), trials of Types 1 and 2 are similar in that vowels in both disyllable members are either both oral or both nasal. This should make the matching fairly easy. However, nonparsers (like parsers) may show slowing on trials of Type 2, relative to Type 1, because of the different consonants in the former, but not in the latter trials. The final trial type had initial vowels that differed in nasality but were followed by a consonant that matched the vowel in nasality (e.g., *bab1bə2 bam1mə2*). For parsers, these trials should be almost as easy as trials of Type 1. They should be easy, because parsing will eliminate nasalization from the nasalized vowel; they may be harder than trials of Type 1 only because the second consonants of Type 3 disyllables mismatch. For nonparsers, these should be the most difficult trials of all. For these listeners, trials of Type 3 are the only trials in which matching vowels differ in nasalization; furthermore, as for trials of Type 2, consonants within a pair mismatch. In short, for parsers, the ordering of response times on match trials should be $1 < 3 < 2$. For nonparsers, the ordering should be $1 < 2 < 3$.

Method

Participants. The participants were 18 students, who volunteered for course credit. All were native speakers of English who reported normal hearing.

Stimulus materials. The speaker who had produced the disyllables used in Experiment 1 also produced the same disyllables, sub-

stituting the vowel /ʌ/ for /a/, yielding /bʌmʌ/, /bʌnʌ/, /bʌbʌ/, and /bʌdʌ/. The talker produced 10 tokens of each disyllable in random order. These were measured in the same way as the tokens from the other experiments. This talker showed significant raising of F1 in the context of a nasal consonant. This made his nasalized /a/ and /ʌ/ vowels more similar in F1 (698 vs. 637 Hz) than were his oral /a/ and /ʌ/ (707 vs. 598 Hz). F1's bandwidth was marginally broadened in nasalized vowels ($p = .055$), but its amplitude was the same in nasalized as in oral vowels. The talker showed a significant, approximately 30-Hz, lowering of F2 in nasalized vowels, but no change in bandwidth or amplitude.

We selected the first two tokens each of the four new disyllables and spliced and cross-spliced them as we had for the disyllables of Experiment 1.

In trials of Type 1a (both vowels nasalized, both followed by nasal consonants), there were eight distinct vowel-matching trials and eight mismatching trials. Matching trials were *bam1mə2 bam2mə1*, *ban1nə2 ban2nə1*, and versions in which both initial vowels were /ʌ/. Each of these four trials had a counterpart in which the disyllables in a trial were reversed in order. Mismatching trials were like matching trials, except that one stressed vowel (equally often the first or the second) was /a/ and one was /ʌ/. Trials of Type 1b (both vowels oral, both followed by oral consonants) were analogous to those of the first type, but with oral vowels and oral consonants. For trials of Type 2a (nasal vowels, one followed by an oral and one by a nasal consonant), pairs involving matching /a/ were *bam1mə2 bam2bə2*, *bam2mə1 bam1bə1*, *ban1nə2 ban2də2*, and *ban2nə1 ban1də1*. These appeared with disyllables in the reverse order as well. In addition, analogous trials with matching /ʌ/ vowels and with mismatching /a/ and /ʌ/ vowels were used. Trials of Type 2b (oral vowels, one followed by an oral and one by a nasal consonant) were analogous, except that first syllables had oral vowels. Finally, in trials of Type 3 (one vowel oral, one nasal, each followed by a consonant of the appropriate nasality), trials involving matching /a/ were *bab1bə2 bam1mə2*, *bab2bə1 bam2mə1*, *bad1də2 ban1nə2*, and *bad2də1 ban2nə2*. The same disyllable pairs were also presented in the opposite order, and there were counterpart trials with matching /ʌ/ and with mismatching /a/ and /ʌ/.

There were 16 trials of each type. There were equal numbers of mismatching trials, for a total of 128 test trials. In addition, there were 4 practice trials.

Procedure. The participants listened over headphones. They were seated facing a computer screen. Left and right forefingers were placed respectively on response buttons labeled DIFF (for *different vowels*) and SAME. The participants were told that they would hear a sequence of two disyllables on each trial. The first vowel of a disyllable might have the vowel /a/, as in *bomber* or *body*, or the vowel /ʌ/, as in *bummer* or *buddy*. Within a trial, the first vowels of the disyllables might be the same or different. The participants' task was to indicate as quickly as possible whether the vowels were the same (both /a/ or both /ʌ/) or different by hitting the appropriate response key. To help them make the discrimination, we presented four practice trials, in which we told them the correct responses in advance.

On each trial, the computer provided feedback for the participants by printing the trial number and the participant's reaction time in milliseconds.

Results and Discussion

Same trials. Predictions are most straightforward for trials on which the correct response was *same*. Accordingly, we present the findings on these trials first. Figure 4A presents the mean response times for each of the three trial types. The response pattern matches the non-

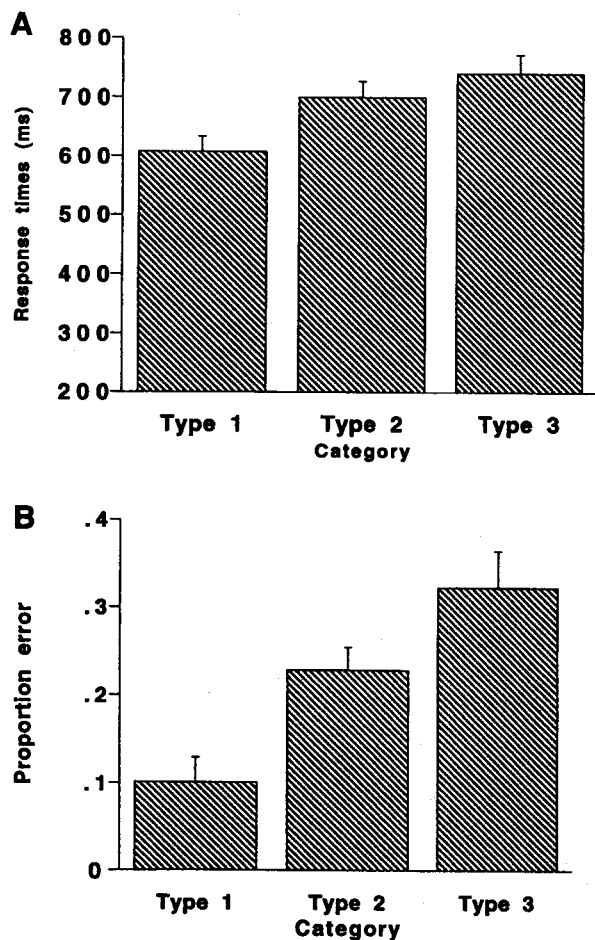


Figure 4. Response times and error rates (with standard error bars) on the *same* trials of Experiment 3. (See the text for a description of the trial types.)

parsing (or incomplete parsing) hypothesis considerably more closely than it does the parsing (complete or nearly complete) hypothesis. From the perspective of either hypothesis, response times should be fastest for trials of Type 1, in which vowels across a disyllable are either both nasal (Type 1a) or both oral (Type 1b) and in which consonants are appropriate to the vowel in nasality. The critically different predictions from the two hypotheses were that parsers should show response times to trials of Type 3 (vowels mismatched in nasality, but consonants appropriate to their vowels in nasality) that were as fast or just a little slower than responses to trials of Type 1. However, nonparsers should find these trials most difficult of all. The latter was the finding.

In an ANOVA on response times, with the independent variable trial type, the effect of trial type was highly significant [$F(2,34) = 26.29, p < .0001$]. Response times were faster to trials of Type 1 than to trials of Type 2 [$F(1,34) = 24.13, p < .0001$] and to Type 2 than to Type 3 trials [$F(1,34) = 4.68, p = .04$]. Error rates showed a

compatible pattern, as Figure 4B shows. Error rates were lowest on Type 1 trials, next on Type 2, and highest on Type 3 trials. In an ANOVA, there was a significant effect of trial type on error rates [$F(2,34) = 15.30, p < .0001$]. In tests pairing means, as in the tests of response times, the error rate on trials of Type 1 differed from that on trials of Type 2 [$F(1,34) = 10.01, p < .003$], and the error rate on trials of Type 2 differed from that on Type 3 trials [$F(1,34) = 5.51, p = .025$].

Different trials. The responses to trials on which the vowels were different (one was /a/ and one /ʌ/) might be expected to show a pattern like that for the responses on *same* trials, but in reverse. That is, factors that make vowels sound alike should speed response times and make decisions more accurate when the vowels are the same but slow them or make them less accurate when they are different. Factors that should make two vowels sound different will slow response times and reduce accuracy when the vowels are, in fact, the same but speed responses and make them more accurate when they are different.

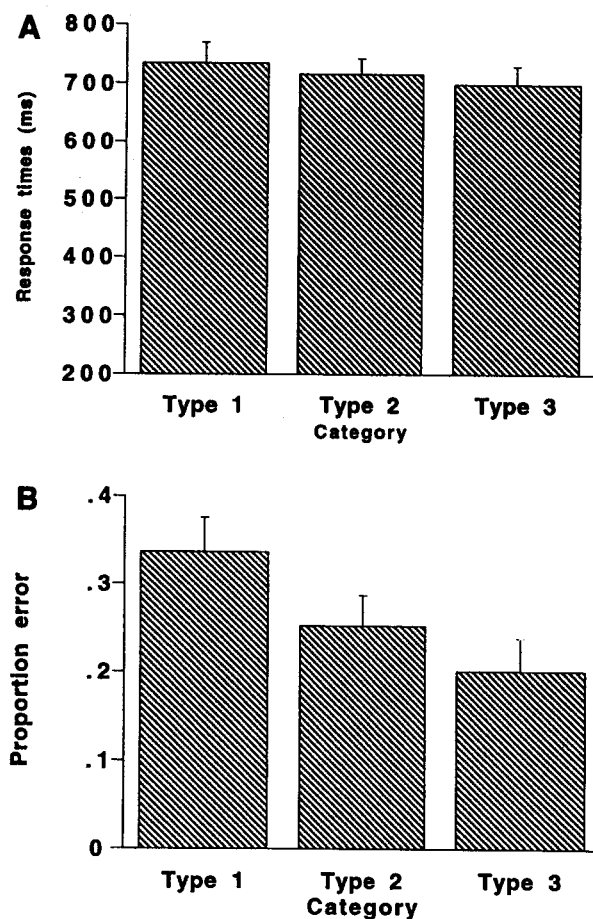


Figure 5. Response times and error rates (with standard error bars) on the *different* trials of Experiment 3. (See the text for a description of the trial types.)

To a close approximation, that was the result. Figures 5A and 5B display response times and errors, respectively, to different trials.

An ANOVA on response times showed a nonsignificant effect of trial type, with, however, response times patterning numerically as expected. Error rates on *different* trials, like response times, showed a pattern opposite to that on *same* trials. The ANOVA on errors showed a significant effect of condition [$F(2,34) = 10.87, p < .0002$]. Trials of Type 1 were associated with a higher error rate than were those of Type 2 [$F(1,34) = 9.26, p = .005$]; the 4% difference in accuracy between trials of Types 2 and 3 did not reach significance.

In Experiment 3, we asked whether parsing is sufficient to make a nasalized vowel in the context of a nasal consonant sound more like an oral vowel in the context of an oral consonant than like itself in the context of an oral vowel. The answer is a clear no. In conjunction with the findings of Experiment 2, which showed evidence of parsing, the findings suggest that parsing of the acoustic consequences of a nasal gesture for a nasal consonant in the domain of a vowel is partial. In the General Discussion section, the implications of this finding will be considered.

GENERAL DISCUSSION

In Experiment 1, we found, consistent with other findings of listeners' use of coarticulatory information (e.g., Fowler, 1981, 1984; Fowler & Smith, 1986; Martin & Bunnell, 1981; Whalen, 1984; but see Lehiste & Shockey, 1972), that listeners use the acoustic consequences of articulatory gestures that begin during production of a preceding gesture as information for the forthcoming one. As we explained in the introduction, this may be for either or both of two reasons. Anticipatory coarticulatory information may make a perceived segment sound context sensitive. In the case of our stimuli, it may make a vowel adjacent to a nasal consonant sound nasalized. The nasalized vowel, then, implies a forthcoming nasal consonant, and that may be the basis on which listeners develop the expectations that research reveals. A different possibility is that listeners may hear the onset of a velum-lowering gesture as information for the beginning of a nasal consonant, and their expectations may derive from that. Finally, a third possibility is that the parsing described by the second account is partial, leaving a vowel preceding a nasal consonant still sounding nasalized after parsing.

The first two accounts make different predictions about how listeners should hear nasalized vowels in their proper contexts. The first account implies that vowels should sound nasalized; the second implies that they should not. Experiments 2 and 3 provided evidence in favor of both accounts, strongly suggesting that the gestural parsing assumed by the second account is only partial.

Experiment 2 found clear evidence that a nasalized vowel next to a nasal consonant is a closer match to a less nasalized vowel next to an oral consonant than to a more nasalized vowel next to an oral consonant. Compatibly,

a nasalized vowel next to an oral consonant is a better match to a more than to a less nasalized vowel next to a nasal consonant. This is consistent with the parsing hypothesis. However, Experiment 3 showed that the parsing is insufficient to make an *endpoint* nasalized vowel sound oral. In the context of a nasal consonant, a naturally produced nasalized vowel sounds more like a nasalized vowel in the context of an oral consonant than like an oral vowel in an oral context.

These findings are consistent with a variety of suggestions of underparsing in the literature. As we summarized in the introduction, there is diachronic evidence (see, e.g., Beddor, 1983; Ohala, 1981) that vowels may become phonologically nasalized when adjacent nasal consonants weaken and are lost. The weakening of the nasal consonants may reduce the salience of acoustic evidence of them under conditions in which nasalization of the vowels is still apparent. If other evidence of the nasal consonant is inaudible, nasalization in a vowel cannot be ascribed to the consonant and must, instead, be ascribed to the vowel. This underparsing, unlike that apparent in Experiment 3 of the present series, occurs under conditions in which the nasal consonant is inaudible or difficult to hear.

A second relevant finding, reviewed by Beddor (1983), is evidence that, synchronically, in languages having oral-nasal vowel pairs, high oral vowels have somewhat lower nasal counterparts, and low oral vowels have higher nasal counterparts. This is consistent with underparsing, because an acoustic effect of nasalizing high vowels is to raise the spectrum in the vicinity of F_1 , and an acoustic effect of nasalizing low vowels is to lower it (Beddor, 1983).

A final relevant finding is the perceptual evidence suggesting that nasalization of isolated vowels shrinks the perceptual vowel space (again, with high vowels lowering and low vowels raising; Wright, 1986).

Despite all of this evidence, there are findings by Krakow et al. (1988) that parsing of nasality can be highly accurate (and, then, evidence by Krakow & Beddor, 1991, that it is partial). Evidence from another domain, parsing of intrinsic f_0 effects from vowels, shows a similar variety of findings on the accuracy of parsing (Fowler & Brown, 1997; Silverman, 1987).

Some of these differences may well reflect differences in experimental paradigm. Notably, Krakow et al. (1988) presented nasalized or oral vowels in real words (bed, bad, bend, band); a single stimulus item was presented on each trial, and participants decided what word had occurred. In contrast, in Experiment 3 here and in research by Krakow and Beddor (1991) in which parsing was partial, two or three nonword (Experiment 3) or real-word (Krakow & Beddor, 1991) items were presented on a trial, and listeners decided whether vowels matched in some way. Although it is not entirely clear why a word identification task should give rise to complete parsing and a matching task to partial parsing, the tasks do direct listeners' attention differentially to information in the

stimuli. Possibly, a distinction made by Gaver (1993) about kinds of listening is relevant here. Gaver proposed that, in *everyday* listening, listeners use structure in acoustic signals as information for its causal source in the environment. Accordingly, listeners literally hear a door slam; they do not hear the acoustic consequences of a door slam and infer the slamming event. In everyday speech perception, listeners extract information about gestures. However, particularly in the laboratory, everyday listening can give way to a special mode of listening in which listeners attend to the acoustic signal itself or to the sensations it causes. The everyday task of word identification may promote everyday listening even in the laboratory. In contrast, tasks that direct listeners to attend to vowel sounds may promote attention to sensations.

Recall, however, that there is evidence that underparsing of nasalization in vowels may occur outside the laboratory (e.g., in the diachronic and synchronic evidence described by Ohala, 1981, and Beddor, 1983). Accordingly, other variables must be important in addition to or instead of differences in experimental paradigms.

In the introduction, we pointed out that vowel–nasal consonant sequences comprise at least three overlapping gestures—a vocalic gesture, a consonantal constriction gesture, and a velum-lowering gesture that overlaps with both but that appears to be phased systematically with the consonantal constriction, and not with the vocalic gesture. Despite the articulatory asymmetry in phasing, perhaps the velum gesture perceptually belongs, to a degree, to both the vocalic and the consonantal gesture.

If so, a second case that might tend to yield underparsing is the anticipatory coarticulatory effect of a lip-rounding gesture for a vowel on a preceding consonant. This is a gesture, like the velum-lowering gesture, that overlaps minimally with two others—in this case, the oral constriction gesture for the preceding consonant and the vocalic gesture. Analogously to the velum-lowering gesture, the rounding gesture appears to be phased systematically with its vocalic gesture, not with the consonantal gesture (see, e.g., Boyce, Krakow, Bell-Berti, & Gelfer, 1990). Possibly, in this type of case too, under the conditions of many kinds of perceptual test, parsing would be found to be partial. (Perhaps, however, as we speculated above, under conditions that foster everyday listening, such as those of Krakow et al., 1988, parsing would tend to be complete.)

Another variable relevant to the accuracy of parsing outside the laboratory may be speaking style—particularly, speaking rate. Acoustic evidence of nasalization in the vowel may be difficult to parse from other properties of the vowel, because the nasal resonance may lie so close to $F1$ as to be difficult for perceivers to separate the two gestural sources of acoustic energy. Research by Bell-Berti and Krakow (1991) suggests a variable that might make parsing easier, however. They show that, at rapid rates of articulation or at normal rates with no vocalic segments intervening between the vowel of interest and the nasal

consonant, the velum-lowering gesture for a consonant may occur at vowel onset. However, at slower rates or with intervening vocalic segments, it may be delayed. Possibly, parsing the nasal gesture from a vocalic gesture is facilitated if acoustic evidence for their distinctness is fostered by such a delay. This would lead to a prediction of more complete parsing for vowels in which the onset of nasalization is delayed.

In short, a number of variables—including variables, such as task, that are special to experimental procedures adopted in the laboratory and others that are relevant to perception outside the laboratory—are likely to have an impact on the accuracy of parsing. Additional research is required to establish which are effective and the conditions under which they have their effect.

REFERENCES

- ARCHANGELI, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183-207.
- BEDDOR, P. (1983). *Phonological and phonetic effects of nasalization on vowel height*. Bloomington: Indiana University Linguistics Club.
- BELL-BERTI, F. (1980). Velopharyngeal function: A spatio-temporal model. In N. Lass (Eds.), *Speech and language: Advances in basic research and practice* (pp. 291-316). New York: Academic Press.
- BELL-BERTI, F., & KRAKOW, R. (1991). Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America*, 90, 112-123.
- BOYCE, S., KRAKOW, R., BELL-BERTI, R., & GELFER, C. (1990). Converging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics*, 18, 173-188.
- DIEHL, R., & KLUENDER, K. (1989). On the objects of speech perception. *Ecological Psychology*, 1, 121-144.
- FOWLER, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech & Hearing Research*, 24, 127-139.
- FOWLER, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, 36, 359-368.
- FOWLER, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1730-1741.
- FOWLER, C. A., & BROWN, J. M. (1997). Intrinsic f_0 differences in spoken and sung vowels and their perception by listeners. *Perception & Psychophysics*, 59, 729-738.
- FOWLER, C. A., & SMITH, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 123-136). Hillsdale, NJ: Erlbaum.
- GARNER, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.
- GAVAR, W. W. (1993). How do we hear in the world?: Explorations in ecological acoustics. *Ecological Psychology*, 5, 285-314.
- KAWASAKI, H. (1986). Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 81-103). Orlando, FL: Academic Press.
- KRAKOW, R. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. Unpublished doctoral dissertation, Yale University, New Haven.
- KRAKOW, R., & BEDDOR, P. (1991). Coarticulation and the perception of nasality. In *XIIIth International Congress of Phonetic Sciences* (pp. 38-41). Aix en Provence: Publications de L'Université de Provence.
- KRAKOW, R., BEDDOR, P., GOLDSTEIN, L., & FOWLER, C. (1988). Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America*, 83, 1146-1158.
- LAHIRI, A., & MARSLER-WILSON, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245-294.

- LEHISTE, I., & SHOCKEY, L. (1972). On the perception of coarticulation effects in English VCV syllables. *Journal of Speech & Hearing Research*, 15, 500-506.
- MARTIN, J., & BUNNELL, H. T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America*, 69, 559-567.
- OHALA, J. (1981). The listener as a source of sound change. In C. Masek, R. Hendrick, R. Miller, & M. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178-203). Chicago: Chicago Linguistics Society.
- PARDO, J. S., & FOWLER, C. A. (1997). Perceiving the causes of coarticulatory acoustic variation: Consonant voicing and vowel pitch. *Perception & Psychophysics*, 59, 1141-1152.
- REINHOLD PETERSON, N. (1986). Perceptual compensation for segmentally-conditioned fundamental-frequency perturbations. *Phonetica*, 43, 31-42.
- SILVERMAN, K. (1987). *The structure and processing of fundamental frequency contours*. Unpublished doctoral dissertation, Cambridge University.
- TOMIAK, G., MULLENIX, J., & SAWUSCH, J. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, 81, 755-764.
- WARREN, P., & MARSLER-WILSON, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 41, 262-275.
- WHALEN, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, 35, 49-64.
- WHALEN, D. H., & LEVITT, A. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, 23, 349-366.
- WHALEN, D. H., LEVITT, A., HSAIO, P., & SMORODINSKY, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9-, and 12-month old French- and English-learning infants. *Journal of the Acoustical Society of America*, 97, 2533-2539.
- WRIGHT, J. (1986). The behavior of nasalized vowels in the perceptual vowel space. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 45-67). Orlando, FL: Academic Press.

NOTE

1. This was the result if vowel durations were the same in the b_d and b_nd contexts. When the vowels in b_nd were shortened to mirror natural durational differences in the two contexts, listeners identified nasalized vowels in the b_nd context as /ε/ more frequently than they identified oral vowels in the b_d context as /ε/.

APPENDIX

Proportion of AXB Trials of Experiment 2 on Which Participants Selected the Less Nasalized of A or B as a Better Match for X

Data are presented separately for the different trial types of Experiment 2. The numbers in the column headings represent the nasalization of /a/. For example, 1-2-4 represents an AXB trial in which the least nasalized vowel appeared as the first vowel of A (B), the second least nasalized vowel appeared as the first vowel of X, and the intermediately nasalized vowel appeared as the first vowel of B (A).

	1-2-4	1-2-5	1-3-4	2-3-5	1-3-5	2-3-6
X is abə	.52	.40	.31	.44	.23	.48
X is amə	.75	.77	.40	.69	.56	.60
	2-4-5	2-4-6	1-4-5	3-4-6	3-4-7	
X is abə	.33	.33	.19	.25	.38	
X is amə	.52	.60	.50	.75	.75	
	3-5-6	4-5-7	2-5-6	3-5-7	4-6-7	3-6-7
X is abə	.13	.29	.25	.31	.29	.17
X is amə	.50	.65	.48	.60	.54	.54

(Manuscript received October 20, 1997;
revision accepted for publication September 29, 1998.)