Vowel height and the perception of consonantal nasality[a)]

Arthur S. Abramson,[b)] Patrick W. Nye, Janette B. Henderson,[b)] and Charles W. Marshall

Haskins Laboratories, New Haven, Connecticut 06510
(Received 13 March 1979; accepted for publication 12 May 1981)

By means of an articulatory synthesizer, the preception of the oral–nasal distinction in consonants was explored experimentally. This distinction was chosen because it is achieved by a very simple articulatory maneuver and because it is phonologically relevant in virtually every language. Lowering the velum in equal increments provided continua of CV syllables varying in size of velopharyngeal port which were divided perceptually into /d/ and /n/ categories by American English listeners. To test the hypothesis that the coarticulation of these nasal consonants with lower (more open) vowels requires a larger area of velopharyngeal coupling to give a nasal consonant percept, three oral–nasal continua incorporating the vowels /i/, /ʌ/, and /ɑ/, respectively, were presented for identification. The results were compared with those of A. S. House and K. N. Stevens [J. Speech Hear. Disord. **21**, 218–232 (1956)] and A. S. House [J. Speech Hear. Disord. **22**, 190–204 (1957)] obtained with steady-state vowels and consonantal murmurs and with those of M. H. L. Hecker [J. Acoust. Soc. Am. **34**, 179–188 (1962)]. Three conclusions emerged. First, the relationship between vowel height and the amount of velopharyngeal coupling needed for a nasal percept occurs in conditions where subjects are required to make linguistically relevant judgments. Second, the relationship can arise in conditions where vocalic coarticulation is present. Third, the relationship is not confined to vowels but can also be observed in the case of dynamically articulated consonants. One of the continua was also used for discrimination experiments, which yielded the classical pattern of high discriminability at the category boundary.

PACS numbers: 43.70.Dn, 43.70.Jt, 43.70.Ve

INTRODUCTION

The communicative function of nasality in speech continues to engage the attention of phoneticians and general linguists. In virtually every language, nasality is phonologically relevant for consonants; in some languages, it is also relevant for vowels. In addition, in comparison with many other distinctive phonetic features, its execution requires a rather simple articulatory maneuver.

We report here on certain perceptual effects arising from control of the velopharyngeal mechanism. Specifically, we investigated a behavioral link between production and perception that determines the position of the perceptual boundary between oral and nasal categories of consonants. We also gave some attention to psychophysical aspects of discrimination behavior at the category boundary. This research was done with the aid of a computer-implemented model of the supraglottal vocal tract designed by Mermelstein (1973) and improved by Rubin and Baer (1978) and Rubin *et al.* (1979).

Our interest in *consonantal* nasality sprang out of the well-known correlation between the size of the velopharyngeal port during *vocalic* nasality and, in traditional terms, the height of the vowel: the higher the vowel, the higher the velum. Numerous observations made over more than 40 years ago (Nusbaum *et al.*, 1935; Harrington, 1944; and Bloomer, 1953; all cited in House and Stevens, 1956) have attested to the vowel-height/velar-height relationship as a normal feature of speech production. In 1951, McDonald and Baker (p. 11) suggested that the correlation might be due to the speaker's efforts to maintain a "characteristic balance or ratio between oral and nasal resonance." This resonance ratio presumably depends on the relative sizes of the velopharyngeal port and the posterior opening into the oral tract. Hence, when the speaker intends to produce no audible nasal output, a lower velum is tolerated for an open vowel than for a close vowel. Conversely, to achieve nasal excitation for a more open vowel, the speaker must lower the velum more than he would for a close vowel. Thus the origin of this effect is usually presumed to be perceptual. However, it may also have consequences that span a broader context. Since coarticulation effects allied with perceptual phenomena operating both forward and backward in time are known to cause interactions between adjacent phones, it is possible that a similar interaction could also be observed in consonant–vowel syllables. For example, Bell-Berti *et al.* (1979) have shown that the effects of vowel height on velar height extend into adjacent consonants. Against this background, then, we set out to discover whether a correlation might also exist between the size of the velopharyngeal port required for *nasal consonant* perception and the height of the following vowel.

A. Velar independence of tongue and jaw motion

The question of whether velar motion can be caused by motion of the tongue and jaw has been the subject of speculation and research by a number of investigators. A relationship between tongue height and velar height could be mediated by connections of the palatoglossus muscles between the tongue and the soft palate (Moll, 1962). Alternatively, as Ackerman (1935) and Podvinec (1952) have suggested, movements of the larynx and pharynx may determine velar position through connections provided by the palatopharyngeus muscles. Moll and Shriner (1967) performed an analysis of cinefluorographic data and concluded that the de-

---

[a)]This is a revised and expanded version of an oral paper presented at the 95th Meeting of the Acoustical Society of America, 16–19 May 1978, Providence, RI.

[b)]Also University of Connecticut, Storrs, CT.

 0001-4966/81/080329-11$00.80

gree of velar elevation in speech is affected in part by the time available for movement and in part by mechanical restraints imposed by the tongue through the palatoglossus muscles. However, Lubker (1968), using a combination of electromyographic and cinefluorographic techniques, obtained evidence that contradicted the conclusions of Moll and Shriner. Lubker found increased levator palatini activity under conditions in which, if the tongue were responsible for velar elevation, such activity would not have been expected. He concluded that "greater palatal elevation may accompany vowels with high tongue position simply because such elevation is needed to prevent the vowel from being detected as nasal in quality." Later evidence reported by Bell-Berti (1976, 1980) is consistent with Lubker's conclusion. Bell-Berti showed that the magnitude of electromyographic activity in the levator palatini muscle is independent of the place of articulation of the consonant. Thus, in circumstances where the tongue body is bunched, as for a velar consonant, the activity of the levator palatini required to achieve a given degree of velar closure is not reduced in the way in which an assumed link between tongue body and velum would lead one to expect.

Thus the weight of evidence favors the conclusion that velar motion is not influenced by any mechanical linkage with the tongue and jaw and variations in velar height do not bear any *direct* relationship to vocal-tract openness. We must suppose, therefore, that variations in velar height are the result of acquired habits, and hence, only *indirectly* related to other articulatory events—possibly through auditory feedback of the acoustic signal. Thus, if auditory monitoring is involved, we may hypothesize that the velopharyngeal port size corresponding to the boundary between an oral and a nasal *percept*, should be smaller for a close vowel than for an open vowel. A speech synthesizer based upon a model of the vocal tract offers the most efficient way of generating the utterances needed to test an hypothesis of this kind, since precise incremental control of the simulated velum can be maintained throughout.

B. Nasal perception: previous vocal tract studies

Several perceptual studies of nasal phenomena employing vocal-tract synthesis techniques have been previously made (House and Stevens, 1956; House, 1957; Hecker, 1962). The electrical-analog synthesizers used in these studies were not governed, however, by built-in vocal-tract constraints. Instead, the models were controlled by parameters that *independently* specified the electrical characteristics of each successive section along the tract and considerable care had to be taken to obtain vocal-tract configurations that were in anatomical and perceptual agreement—particularly under dynamic conditions (Hecker, 1962).

The topic of oral-nasal boundary phenomena was first addressed by House and Stevens (1956). They used an electrical analog to obtain perceptual data on the relationship between the boundary-defined velopharyngeal port size and vowel openness and showed this relationship to be consistent with observations of nasal articulations (Harrington, 1944; Bloomer, 1953). In addition, their analysis tended to support an hypothesis which, drawing on the resonance-ratio concept of McDonald and Baker (1951), holds that to achieve either an oral or nasal percept, a suitable ratio of acoustic *impedances* of the nasal tract and the oral tract is necessary. However, House and Stevens' results depended on nonlinguistic evaluations of *vowel* nasality and involved unnaturally large amounts of oral–nasal coupling. The velopharyngeal port area required to achieve close to 100% nasal responses in House and Stevens' study was nearly 4 cm^2. Similar port areas were employed (with later versions of the electrical analog) by House (1957) for his best nasal vowels and apparently for his static consonants /m/, /n/, and /ŋ/ since he wrote that the two tracts were "coupled maximally." Also, Hecker (1962) appears to have employed similar amounts of nasal coupling in his study which was based on the same nasal consonants coarticulated with the vowels /i/, /ɪ/, /æ/, /ʊ/, and /u/. On the other hand, the observations of several investigators (Passavant, 1863; Björk, 1961; Nylén, 1961; Warren, 1967; Isshiki *et al.*, 1968) converge on the opinion that the *linguistically* useful region of velopharyngeal control lies in the range from zero to little more than 1 cm^2. Since we were intrigued by the apparent relationship between velopharyngeal port size, vowel openness, and the perceived oral–nasal boundary demonstrated by House and Stevens and we were concerned about the degree of nasal coupling, we felt that their result could be accepted with more confidence if new data were obtained with utterances based on articulatory specifications that were physiologically more plausible and in a situation requiring linguistic identifications rather than quality judgments.

C. Is nasal consonant perception affected by vowel height?

Using our articulatory synthesizer, we might have designed an experiment paralleling that of House and Stevens but have drawn our utterances from a language, such as Hindi or Portuguese, that has oral–nasal contrasts in both high and low vowels. However, because we were interested in extending observations of the velar-height/vowel-height relationship to coarticulated consonants and because we had an unlimited number of English-speaking subjects available, we chose to work with English consonants. Our approach involved examining an extension of House and Stevens' hypothesis, namely, that speakers and listeners also adopt as their *consonant* category boundary a particular ratio of oral and nasal impedances. Our hypothesis was that for a nasal consonant coarticulated with an open ("low") vowel, the speaker would have to adopt a lower position for his velum to ensure that the balance of resonances was appropriate for a nasal percept. Conversely, the listener's decision as to whether a particular segment is nasal would depend on whether he perceives that the ratio between velar opening and oral opening within the speaker's vocal apparatus has matched the criterion. To examine this hypothesis, we synthesized a sequence of consonant–vowel stimuli that through a progressive

increase in the size of the velar port, in the judgment of listeners varied from [dɑ] to [nɑ].

D. Are articulatorily synthesized nasal consonants perceived categorically?

With the set of stimuli already synthesized for an exploration of the foregoing question, vowel height and the perception of consonantal nasality, we realized that the series could be used to try to shed light on the underlying basis for arguments pertaining to categorical perception (Liberman *et al.*, 1957). These arguments rest upon the finding that listeners are largely unable to discriminate speech sounds that they would not normally label as different from one another. Thus variants taken from a consonantal continuum can be sorted into phonological categories by listeners, but discrimination tests yield high levels of performance only in the region of the perceptual boundary on the stimulus dimension. The discrimination of variants within a single category is typically not much better than chance. We wondered whether the phenomenon of categorical speech perception might be a direct by-product of the fact that the intervals between successive synthesized samples from a continuum are customarily defined with reference to *spectral* structure rather than to the structure of articulation. It could be argued that continua based on incremental adjustments in formant frequencies might give rise to a discrimination peak at the category boundary because a nonlinear relationship may exist between formant frequency loci in the frequency–time domain and displacements of the articulators. If the scale on which the successive increments were calculated was to be defined in articulatory terms (and was thus based on a metric conceivably shared by the mechanisms of perception), discrimination functions without a peak might be observed. We felt that our articulatory synthesizer now gave us the opportunity to test the hypothesis that there is a nonlinear relationship between the metric of acoustic frequency and the metric of articulatory movement. The most likely conditions in which such a nonlinearity might be found were clearly in a case in which the specifications for articulatory production were relatively simple but the resulting output was spectrally quite complex and, therefore, difficult to synthesize accurately with any terminal analog synthesizer. The oral–nasal distinction met this criterion[1] since it is mainly achieved in nature by lowering the velum and is reproduced with the articulatory synthesizer by manipulating the parameter that controls the area of the velopharyngeal port. Perhaps, if our hypothesis was correct, we would find that utterances differing by equal articulatory displacements would be equally discriminable and that our listeners would fail to display any discrimination peaks at the oral–nasal boundary.

I. THE SYNTHESIS OF SPEECH BY ARTICULATORY MODEL

The utterances that we created to explore these two issues were generated by a digital-computer model of the vocal tract (Rubin and Baer, 1978; Rubin *et al.*, 1979). Control of the model was achieved by manipulating parameters that described the moment-by-moment positions in the midsagittal plane of the tongue, lips, jaw, velum, and hyoid bone. The model imposes a number of constraints upon the control parameters in order to make the configurations that it adopts conform, with certain limits, to normal vocal physiology. For example, two such basic features built into the model are (1) the tongue is attached to the jaw so that jaw lowering also lowers the tongue, and (2) enlargement of the velar port narrows the dimensions of the oral tract in the velopharyngeal region.

The model divides the overall tract into three branches: pharyngeal, oral, and nasal. Once the midsagittal outline has been determined, a grid system is imposed upon it and used to compute a center line, the overall tract length, and the dimensions of the tract in the sagittal plane at intervals of 0.875 cm along its length. Then, using scaling coefficients derived from anatomical measurements, cross-sectional areas of the vocal tract are computed from the sagittal-plane dimensions. Finally, these areas are used to form a sequence of 0.875 cm tubes that serve as an approximation of the vocal tract.

The aerodynamic behavior of the tract is modeled as a plane wave that suffers attenuation within each tube section, and transmission and reflection at the boundaries between sections. This model is described by a Kelly–Lochbaum system of equations (Kelly and Lochbaum, 1962; Mermelstein, 1972) used to calculate the vocal-tract transfer function in z space. The transfer function is then realized by a direct-form digital filter which is updated every pitch pulse and excited by the glottal pulse (represented in the time domain). When the velar port is opened, the overall output of the synthesizer is the sum of the oral and nasal branches. The coupling of the nasal tract results in the shifting of formant frequencies and nasal pole-zero pairs in the output spectrum.

On the other hand, some constraints are the result of simplifications designed to speed up calculation of the speech output. For example, each vocal-tract configuration is derived as a series of between 18 and 22 short cylinders of various cross-sectional areas connected end-to-end. To obtain the sound output, this compound tube is excited at its distal end by pulses representing glottal vibration and, when appropriate, by noise introduced at the "glottis" or any point of constriction along the tube. In addition, further simplifications involve holding each vocal-tract configuration constant during each "glottal" period and assuming the velar aperture to be circular in shape.[2]

A. Articulatory specifications of the stimuli

Construction of the test stimuli began with a determination of the articulatory configurations required to achieve the desired steady-state vowels ([ɑ], [ʌ], and [i]) and a satisfactory closure articulation. The durations of the transitions from the articulatory closure to the vowel-target were specified in such a way as to produce the syllables [dɑ], [dʌ], and [di]. All three syllables used essentially the same [d] closure gesture,

murmur, and transition timing controls. To produce a nasal consonant, the synthesizer specifications called for the velopharyngeal port to open (with glottal pulsing present and the oral tract occluded) for a period of 50 ms prior to tongue-tip release. With the addition of the vowel segments, the utterances always reached a total length of 340 ms.[3] For experiments 1, 2, and 3, the nasal coupling was maintained throughout the syllable whereas, for the discrimination study, the nasal coupling was confined to the nasal closure and movement toward the steady-state vowel. The fundamental frequency contour (F_0) was designed to begin at 100 Hz and to fall to 70 Hz at the end of the utterance. Irrespective of which vowel or velar port size specification was delivered to the synthesizer, both the F_0 and *input* excitation energy contours remained the same. Thus, as a consequence of their higher radiation efficiency, the peak *output* energy delivered during the production of utterances incorporating open vowels was always higher than that observed in the utterances containing close vowels. Furthermore, the peak output level tended to be reduced by up to 3 dB as the nasal coupling increased. Figure 1 displays an acoustic spectrogram, a stylized spectrogram predicted by the synthesizer program, and an F_0 contour for the syllable [ni] that illustrates the relative durations of the nasal murmur (a), the consonant–vowel transition (b), and the vowel segment (c). Following these specifications, the members of a sequence of stimuli were created that, starting with a closed velar port, differed from one another by an increasing number of fixed increments in the radius of the velar port aperture. The maximum velar radius used was 2.47 mm (corresponding to a port area of 19.2 mm^2) because an opening of this size was consistent with the anatomical evidence and also produced an acoustic output that pilot experimental subjects reliably identified as nasal.

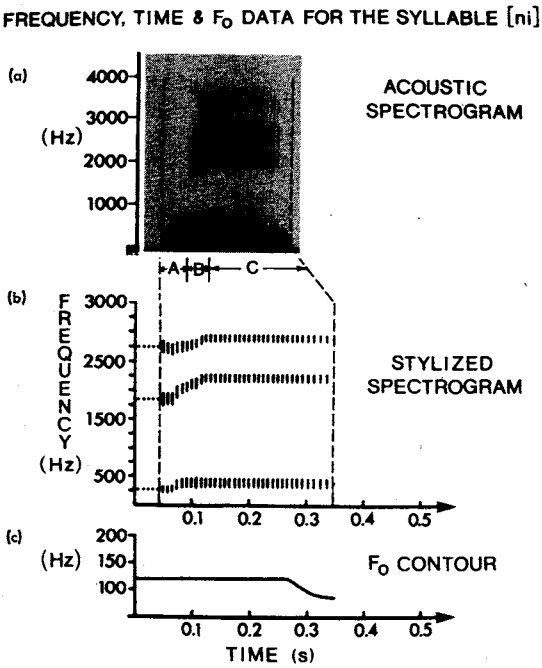

FIG. 1. Three displays of the syllable [ni]. (a) Acoustic spectrogram. (b) Stylized spectrogram predicted by the synthesizer program. (c) F_0 contour. The three segments of the time axis labeled A, B, and C represent the durations of the nasal murmur, the consonant–vowel transition region, and the vowel portion of the stimulus, respectively. The stylized spectrogram indicates the formant peaks by vertical bars located at successive 12.8 ms intervals. The lengths of these vertical bars are inversely proportional to the formant bandwidths.

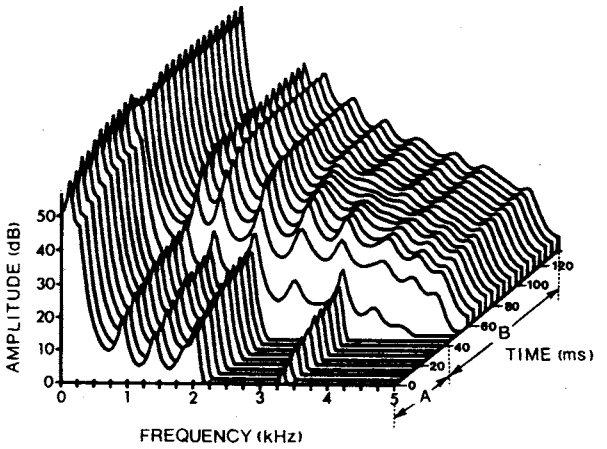

FIG. 2. A plot in three dimensions of the CV transition in the acoustic output of a synthesized [ni]. The synthesizer computed the speech at a 20 kHz sample rate. The spectral cross sections were obtained every 5 ms by computing 26 LPC coefficients using a window size of 512 samples. The time segments A and B are described in Fig. 1.

B. Acoustic characteristics of the stimuli

The acoustic consequences of imposing an opening gesture of the velar port during a [d]-vowel movement are complex. The first effect is that a nasal murmur is introduced during the initial period of oral closure and is followed by nasal coupling effects that are seen throughout the transition toward and during the following vowel. These effects are illustrated in Fig. 2 which shows a three-dimensional acoustic spectrogram of the consonant–vowel transition in the syllable [ni] as derived by a linear predictive coding (LPC) analysis of the synthesizer's digital waveform output.

The most obvious effect of the increasing velar opening is an increase in amplitude of the nasal murmur; however, in addition, shifts in the spectral cross sections occur that are caused by changes in the location of the oral zero and changes in the geometry of the velar aperture as it enlarges. Figure 3 shows the relative amplitude of the murmur as a function of the port area. The articulatory model leads to the formulation of a complex filter (typically around 39 pole terms and 19 zero terms) which must be simplified in order to extract formant information for display or other purposes. Analysis of the nasal murmur using root solving of the LPC spectrum, for example, resolves the voice bar as two poles. One pole is located at 200 Hz with a bandwidth of 80 Hz, and the other is at 500 Hz with a bandwidth of 200 Hz. When the resolution of the LPC analysis is reduced so as to lump these poles together to form one broad formant, an effective pole at 291 Hz with a

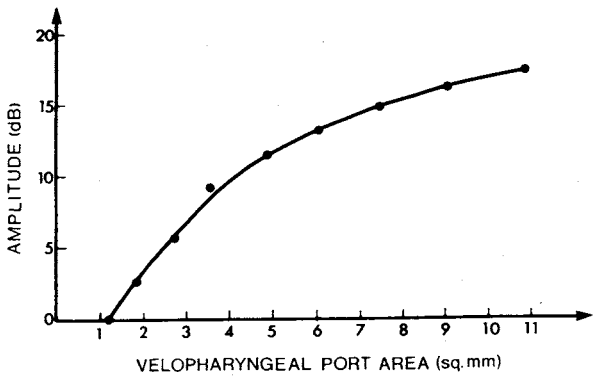

FIG. 3. Amplitude of the nasal murmur (see time segment A of Fig. 1) measured in decibels for the syllable [ni] as a function of velopharyngeal port area.

bandwidth of around 200 Hz is produced. The narrow resonance with a broad resonance on its right shoulder also produces a deep valley in the spectrum at 1500 Hz which is further enhanced by a zero. Region (A) of Fig. 2 shows the spectral cross sections of the nasal murmur for [ni] with the higher poles shown. The figure also shows that the first zero in the nasal murmur appears at 2500 Hz.

The spectral characteristics of the three vowels are presented in Table I which contains a listing of their formants and bandwidths when the areas of the velar port aperture were set at 1.2 or 10.8 mm^2. For reasons that will be explained later, these values of velar port opening were adopted as the extreme oral and nasal ends of most of the stimulus sequences used in this study. Also in Table I is the characteristic nasal pole/zero pair that emerges at about 400 Hz and splits apart as it moves upwards in frequency with increasing nasal coupling. At corresponding velar port openings, it appears that the effects of the nasal coupling upon different vowels result in different locations for the pole/zero pair. A prominant effect, well illustrated in the case of the vowel /ʌ/, is that F_1 becomes a cluster of two poles and a zero. For example, in [ʌ] as the zero moves closer to F_1 and becomes separated from its pole, a splitting of F_1 is achieved. The overall effect of the many different cluster configurations that occur is the production of an F_1 whose bandwidth has been effectively broadened—a broadening that is further enhanced by higher losses within the walls of the nasal cavity.

TABLE I. Frequencies of the first nasal pole and nasal zero together with the frequencies and bandwidths of the first three oral formants. Measurements made in Hertz for velar port apertures of 1.2 and 10.8 mm^2. In the case of the vowel ʌ(10.8) note that F_1 splits in two.

Vowel	NP	NZ	F1	BW	F2	BW	F3	BW
i(1.2)	403	411	323	107	2181	100	2892	159
i(10.8)	597	662	375	117	2208	146	2882	164
ʌ(1.2)	447	409	584	81	1187	109	2444	144
ʌ(10.8)	650	490	483/669	130/137	1182	102	2517	150
ɑ(1.2)	426	410	824	100	1180	163	3082	298
ɑ(10.8)	496	453	844	131	1208	142	3092	284

Using the graphical analysis method adopted by Fujimura (1962) and Fujimura and Lindqvist (1971), we also investigated the motion of the pole/zero pair produced by nasal coupling within the transition and steady-state vowel regions. Parts (a), (b), and (c) of Fig. 4 show the frequency of F_1, F_2, and the nasal pole-zero pair as a function of nasal coupling for the vowels [ɑ], [ʌ], and [i]. The range of coupling shown is the same as that examined by Fujimura but goes well beyond the range employed in our study. For the vowel [ɑ], our graph shows a pattern that is essentially the same as that found by Fujimura for the same vowel. The frequency ordering is a nasal pole, a nasal zero, the pole for F_1, and then the pole corresponding to F_2. For the vowel [i], since F_1 is lower than the frequency at which the nasal pole/zero pair emerges, the pattern consists of the pole of F_1, a nasal pole, a nasal zero, and then the pole of F_2.

II. EXPERIMENTS ON VOWEL HEIGHT

A. Tests

Our study of velar height/vowel height effects employed the oral–nasal contrast in the English consonants /d/ and /n/ coarticulated with the vowels /ɑ/, /ʌ/, and /i/. We chose articulatory configurations for the vowels [i] and [ɑ] as end points and then a configuration

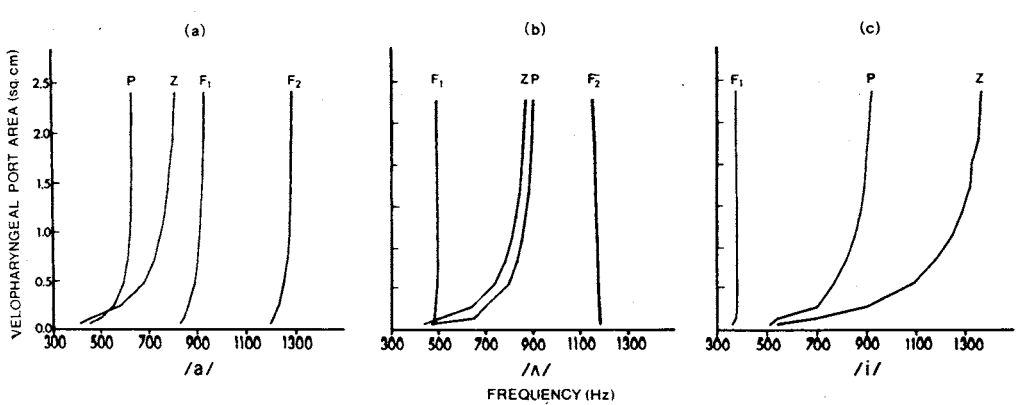

FIG. 4. Parts (a), (b), and (c) show the first two formant frequencies (F_1 and F_2) and the first nasal pole/zero pair (P and Z) as a function of velopharyngeal port area for the vowels /ɑ/, /ʌ/, and /i/. The F_1, F_2, P, and Z trajectories were obtained by tracing the peaks from a Fourier transform of the vocal-tract transfer function and proceed beyond the velopharyngeal port areas employed in this study.

midway between these points for an acceptable [ʌ]. Taking 2.47 mm as the maximum velar port radius and 0 mm as the minimum, 17 equal increments in velar port radius were computed. Then, on the basis of pilot work aimed at reducing the number of stimuli from 17 to 9, we eliminated four stimuli from the oral end that could always be identified with 100% accuracy and four stimuli from the extreme nasal end that could be identified with similar accuracy. Hence, the velar port parameter at the oral end of the series was set at a velar port radius of 0.618 mm (a port area of 1.2 mm^2) and the extreme nasal end of the continuum was bounded by an utterance produced with a velar port radius of 1.85 mm (a port area of 10.8 mm^2). These end points plus the seven intermediate values constituted the nine velar port magnitudes employed in synthesizing the three continua based on the vowels /ɑ/, /ʌ/, and /i/. The resulting utterances were recorded on magnetic tape under three different experimental conditions and presented to native speakers of English for identification as the consonants /d/ or /n/.

The synthesized utterances were delivered to panels of listeners over headphones (type TDH-39) in a quiet room at 80 dB SPL measured at the peak of the vowel /ɑ/. In some instances noted below, the levels of the close vowels were amplified to achieve a peak sound pressure level equal to that reached by the vowel /ɑ/.

The different experimental conditions were employed in an effort to ensure that none of the observed boundary shifts could be a product of the method of stimulus presentation or of masking effects within the stimuli. Hence, in experiment 1, all three vowels (with their associated consonantal segments) appeared in randomized order on each of three tapes, each vowel being generated at its natural output level relative to the 80 dB SPL standard level defined by the peak amplitude of [ɑ]. The three different randomizations each contained 135 stimuli. In the second experimental condition, individual randomizations of the three vowels were recorded on three separate tapes each of which contained 90 exemplars of the chosen vowel. Before presenting each tape, the output level of the recorder was adjusted to maintain an 80 dB SPL peak output for the vowel represented. The third experiment employed syllables for which the rms amplitude contours of [i] and [ʌ] had been computationally equated with their counterparts incorporating [ɑ]. The amplitude profiles of the utterances were checked for equality with an rms meter prior to recording on audio tape. Hence, to reduce backward masking of the consonantal segment by the vowel, the relative energies of the consonant and vowel segments (for a given degree of nasal coupling) were maintained constant across all vowel types.

B. Results

The first of our experiments on the topic of the velar-height/vowel-height relationship employed mixed vowels at their *natural* output levels relative to [ɑ]. A group of 24 listeners heard three randomizations of 135 utterances (9 values of port size ×3 vowels ×5 repetitions per randomization) and wrote down their consonant

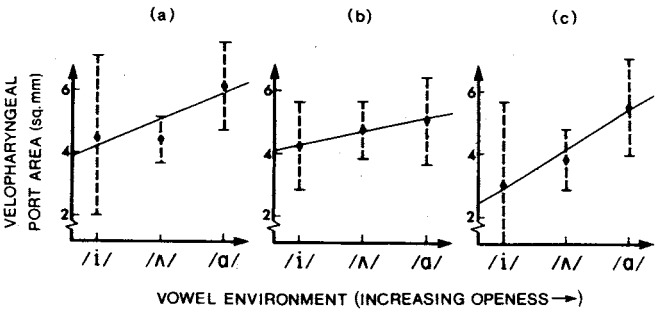

FIG. 5. Results from three studies of the effects of vowel openness on the oral–nasal consonant boundary position. Average velopharyngeal port areas at category boundaries are plotted against the associated vowel environment. Vowels appear from left to right in order of increasing openness. Straight lines linking the category boundaries have been computed by linear regression. Broken vertical lines indicate standard deviations. Parts (a), (b), and (c) were obtained from experiments 1, 2, and 3, respectively (see text).

classifications. This group of listeners had some prior experience listening to synthetic speech, although very few had received much practice making repetitive quasicontext-free linguistic judgments. The judgments provided by these listeners were tabulated as a function of velopharyngeal port size and vowel class, then expressed as percentages. Finally, Probit analysis was applied to find the intercepts of 50% threshold lines with the data.

As an illustration of the method used in computing the 50% threshold, one should look ahead to Fig. 6. The figure shows the identification data obtained for a series of utterances extending from [dɑ] to [nɑ]. The Probit calculations (Finney, 1971) led to the fitting of a smooth sigmoidal curve to the percentage data whereupon the intercept of a 50% threshold line with the fitted curve could be found.

In Fig. 5(a) are plotted averages of the 24 intercepts (in units of velar port area: mm^2) for each of the three vowels in order of increasing openness. The graph indicates that the consonant category boundary does systematically move nasalward as the vowels become more open and that the largest boundary shift occurs between both the /i/ and /ʌ/ environments and the environment of the vowel /ɑ/. This observation is confirmed by an analysis of variance which shows that the probability of such a shift occuring by chance is $p < 0.01$ [$F(2, 46) = 12.15$, $MS_e = 1.79$]. Moreover, the Newman–Keuls test of significance applied to the differences between the boundary means shows the differences between the /i/ and /ɑ/ environments and the /ʌ/ and /ɑ/ environments to be significant at the $p < 0.01$ level.

Experiment 1, however, contained two presentation features that might have influenced the results. First, since all the different utterances based on the three vowels were mixed together, they may conceivably have permitted the listener to make relative judgments of the stimuli and be influenced by superficial stimulus features unrelated to those conveying the oral/nasal distinction. Such features might have been correlated

with, but not actually dependent on, the degree of nasality in the stimulus. Removing the opportunity for making relative judgments among the utterances would be expected to sharply reduce such an interaction effect. Second, since the natural (and different) output levels determined for the three vowels by the physical characteristics of the model were employed in experiment 1, it was possible to hypothesize that the shift in oral/nasal boundary could be due to variations in the effectiveness of the different vowels as backward maskers of the nasal consonants. As the vowels increase in loudness with increasing openness, a larger degree of nasal coupling would be required to achieve a nasal resonance that was sufficiently prominent to overcome backward masking effects. To determine whether either of these differences in presentation might give rise to a boundary shift, experiments 2 and 3 were conducted.

In experiment 2, the recordings of the utterances were grouped under the three vowels and presented in three separate sessions to avoid relative judgment effects. Each recording contained ten repetitions at each of the nine velar port sizes in randomized sequences. A panel of 24 listeners, drawn from the same source as before, was employed to identify the consonants in writing. The responses obtained from these listeners, having first been subjected to Probit analysis as described earlier, yielded the graph shown in Fig. 5(b). Once again, a consonant boundary shift occurs in the direction indicating a greater tolerance of nasality in open vowels. While this boundary shift is somewhat smaller than the previous observation, the statistical evidence proves in fact to be stronger because the mean-square error is smaller. In this case, the probability of the shift occurring by chance is $p < 0.01$ [$F(2, 46) = 7.66$, $MS_e = 0.519$], and the Newman–Keuls test shows the boundary difference between the /i/ and /ɑ/ environment to be significant at the $p < 0.01$ level and the boundary difference between the /i/ and /ʌ/ environment to meet the $p < 0.05$ criterion. The /ʌ/ and /ɑ/ boundary difference, on the other hand, does not reach the $p < 0.05$ criterion of significance.

Experiment 3 employed utterances in which the vowel and consonantal output amplitudes bore a constant relationship to one another across all vowels for each degree of nasal coupling. Again 24 listeners heard three tapes in different orders. Each tape contained four repetitions of each of the three vowel types at each of the nine velar port sizes presented randomly. The category boundary data for the vowels shown in Fig. 5(c) are once again similar to the two earlier findings. Analysis of these results shows that the boundary shift is significant at the $p < 0.01$ level [$F(2, 46) = 13.84$, $MS_e = 2.99$], while the Newman–Keuls test indicates that the boundary differences between the /i/ and /ɑ/ environments and the /ʌ/ and /ɑ/ environments are significant at the $p < 0.01$ level.

Hence, the data from all three experiments support the conclusion that there is a genuine perceptual compensation made for the velar port size differences that are frequently observed in naturally produced vowels. The data also indicate that, while the absolute boundary locations are influenced to some degree by different experimental conditions, the increase in velar port size marking the category boundary with increasing vowel openness remains intact. Moreover, there appears to be no evidence that the increase is brought about by the backward masking of consonantal segments by vowel segments since it still survives when the relative amplitudes are held constant. It is also apparent that House and Stevens' boundary shift in the perception of *vowel* nasality, albeit more dramatic than ours which is based on *consonant* nasality, was not an artifact of the unnaturally large amount of nasal coupling that they employed.

III. EXPERIMENT ON CATEGORICAL PERCEPTION

A. Tests

Turning to the discrimination study foreshadowed in the Introduction, we chose as our stimulus set a series of nine "partially coarticulated" synthetic syllables containing the vowel [ɑ]. That is, in the production of these synthetic syllables, the velopharyngeal port was kept open up to the end of the articulatory movement at the onset of the final fixed vocal-tract shape for the vowel. We chose this set rather than the "fully coarticulated" set used in the vowel-height study, because we felt that for arguments on categorical perception of consonantal features we should confine the increments of nasality to the segments conventionally attributed to the consonant, namely the simulated closure and the transition toward the steady-state vowel. The port was completely closed at the oral end of the continuum and opened to 19.2 mm^2 at the extreme nasal end. Discrimination tests were then prepared for one-, two-, and three-step differences. The utterances were presented in AXB triads to groups of listeners such that either the first or the third member of the triad was physically identical with the second member; the listener's task was to say which utterance, the first or the third, was the same as the middle one. Each triad was repeated ten times. For the one-step experiment, 25 native speakers of English participated, while for the two- and three-step experiments, 18 listeners were employed of whom 12 were a subset of the original 25. Identification tests were prepared by randomizing ten repetitions of the series of variants and presenting them to the 25 English speakers for labeling as /d/ or /n/.

B. Results

Evidence of categorical perceptual behavior is indicated when subjects find stimuli that fall *within* categories almost indistinguishable, while stimuli that straddle the category boundary can be discriminated with considerably greater frequency. The issue lay in whether the results of our identification and discrimination experiments performed with articulatorily synthesized oral and nasal consonants showed the familiar categorical behavior or whether they exhibited a flat unchanging discrimination behavior across the category boundary which would indicate a nonlinear re-

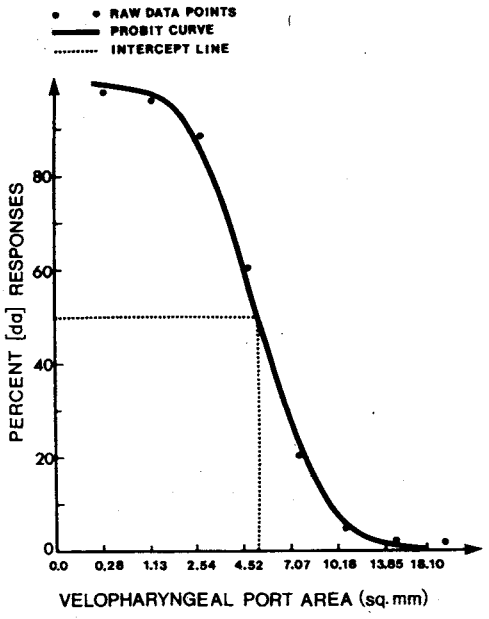

FIG. 6. Identification data points, each representing the average score obtained from 25 listeners employed in experiment 4, are plotted for a continuum of utterances from [da] to [na] incremented along the radius scale. A fitted continuous curve has been computed by Probit analysis. A 50% threshold line intercepts the computed curve at a velopharyngeal port area 5.23 mm^2.

lationship between the articulatory and the spectral frequency domains.

The identification data obtained with the [da]–[na] series are shown in Fig. 6. The nine utterances are ordered along the abscissa from a closed velopharyngeal port on the left to the largest opening (a port radius of 2.47 mm or an area of 19.2 mm^2) on the right. The Probit curve is intercepted by the 50% threshold to give the category boundary at a velopharyngeal port radius of 1.29 mm (an area of 5.2 mm^2). Figure 7 shows the companion discrimination data for the same series of utterances. The ordinate displays percentages of correct discrimination for the one-, two-, and three-step differences as indicated by the coded lines. For the one-step difference, the listeners' performances hover around chance, but, for two- and three-step differences, greater discrimination acuity emerges as a peak in the region of the category boundary. Thus the discrimination functions of the listeners do not vary monotonically and the results do not appear to be in conflict with the classical observations. Moreover, the results are fully consistent with the nasal discrimination data obtained by other investigators who have used continua of spectrally specified increments in studies of oral–nasal perception (see Miller and Eimas, 1977).

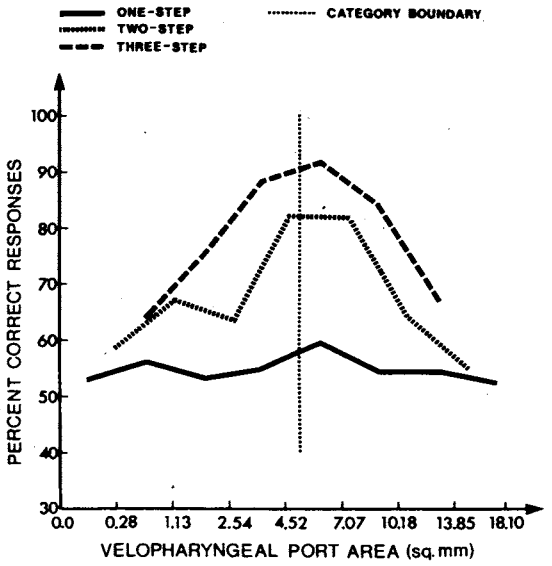

FIG. 7. Discrimination data for a continuum of utterances extending from [da] to [na] incremented by units defined on the radius scale. One-step data are the averages for 25 listeners, while two- and three-step data are the averages for 18 listeners. The vertical line marks the oral–nasal category boundary obtained from an identification test (see Fig. 6).

IV. DISCUSSION

Our newly available techniques of articulatory synthesis of speech have been shown to be powerful enough to yield continua of variants that native speakers of English have no difficulty in dividing into categories of oral and nasal consonants. That is, incrementally increasing the size of the velopharyngeal port of our model brought about suitable acoustic coupling between the oral and nasal cavities which caused listeners to shift their labeling behavior from a well-established oral category through a brief zone of ambiguity to a well-established nasal category.

Our results clearly show that, under a variety of conditions, the openness of the following vowel does influence a listener's perceptual judgments of the oral/nasal boundary in initial consonants. The results are robust and consistent with numerous observations of articulation. What then, might be the physical basis for the listeners' behavior?

McDonald and Baker (1951) hypothesized that the behavior might arise because listeners unconsciously adopt as their boundary criterion a characteristic balance between oral and nasal resonance. However, *balance* and *resonance* as metrical concepts suffer from an unfortunate ambiguity that effectively prevents any effort to verify the hypothesis by objective measurement. House and Stevens (1956) attempted to avoid this dilemma by interpreting the resonance ratio as a ratio of oral and nasal impedances—parameters that could be directly measured from their electrical analog of the vocal tract.

Despite the accessibility of the impedances in a vocal-tract model, however, the House and Stevens ratio is still not easy to employ. The reason lies in the fact that oral and nasal impedances are complex functions of frequency that vary with the degree of nasal coupling. Thus, after forming their ratio (also a complex function of frequency), there remains the question of whether the whole utterance or certain frequency bands

should be considered to be chiefly involved in the perception of the oral/nasal distinctions.

House and Stevens approached this problem by arguing that the changes in the vocal output are most likely to occur in those spectral regions where the difference between the oral and nasal impedances is greatest (i.e., where the ratio of oral impedance to nasal impedance is a maximum) and where output signal energy is usually highest (i.e., in the region of the first formant, F_1). This argument suggests that for a given amount of nasal coupling, the F_1 output of a close vowel should be modified more than the output of an open vowel like [ɑ]. Data gathered by these authors on the relationship between velar port area and the relative amplitude of F_1 did indeed reveal that the spectrum for close vowels underwent larger changes as a result of nasal coupling than the spectrum for open vowels. Moreover, when these data were combined with perceptual data on the relationship between velar port area and vowel nasality for vowels of varying openness, House and Stevens were able to provide indirect[4] support for the impedance ratio hypothesis by showing that the relative amplitude of F_1 at the category boundary remains essentially constant as a function of vowel openness.

The argument of House and Stevens, however, has not received sufficient support from studies of the perceptual cues for nasals nor, for that matter, has their argument been proved fallacious. It remains merely plausible and highly speculative. Indeed, employing a parallel argument, based on measurements of admittances at the entrances of the oral and nasal branches of our synthesizer, we can also demonstrate that the category boundary should shift in the direction indicated by our listeners' perceptual data. But, without a more detailed acoustic model for nasality, better data on the perceptual cues for nasals and a better understanding of the performance of the ear, it appears that very little that is new can be added at this time about the mechanism underlying our listeners' perceptual behavior.

A reassuring finding of our study is that linguistically relevant judgments support a correlation between vowel height and the amount of velopharyngeal coupling needed for a nasal percept. The speakers of American English employed by House and Stevens (1956) were asked to judge vowels for nasal quality, much as speech therapists do in auditorily assessing the success of cleft palate repair. This nonlinguistic judgment seemed worrisome not only because vowels as such are not phonemically differentiated in English by the feature of nasality, but also because House and Stevens used an unnaturally large amount of nasal coupling. In our experiments, the listeners simply had to label each stimulus as beginning with a /d/ or /n/, a perfectly straightforward distinction in English phonology.

We have thus extended the observations on the category-boundary effects of vowel openness in the perception of oral–nasal distinctions. We have moved from work based exclusively on vowels (House and Stevens, 1956) to consonants preceding coarticulated vowels of three heights. Our results show that a significant shift of category boundary occurs for judgments of oral versus nasal consonants. Traditional phonetics, incidentally, would lead us to expect some somewhat audible assimilation of a vowel to a *following* nasal consonant; indeed, Ali *et al.* (1971) have demonstrated that such assimilated nasality can be quite perceptible. Our finding of the effect of vowel height on the port size required for the identification of a *preceding* consonant as nasal is probably less predictable from their work and traditional phonetics.

A third conclusion to be drawn from our data is that the correlation between port size and vowel height is perceptually important, not just for steady-state vowels but also for dynamically articulated consonants. That is, in our articulatory synthesis we have stimulated both the static and dynamic aspects of articulation to yield acceptable CV syllables. Indeed, the closure of the tongue tip against the alveolar ridge, which is the most static part of the "segment" being labeled in the experiments, is the same for all three vowel heights. The combination of articulatory movement and, of course, the steady state of the following vowel is what differs across the three vocalic environments.[5]

It seems to us that the last two of the foregoing conclusions of our study, namely those on coarticulation itself and the dynamics of CV articulation, bear on arguments concerning the status of the phonetic segment as a unit of production and perception (Studdert-Kennedy, 1980, pp. 48–50). The coordinated control of vocalic gestures and velopharyngeal port needed for the oral–nasal distinction should be taken into account in arguments concerning models of speech production (cf., Kent and Minifie, 1977; Fowler, 1980).

Our foray into research on categorical perception was meant to examine the possibility that the classical findings of high acuity of discrimination in the region of the boundary between phonetic categories were artifactual because of nonlinearities between the acoustic continua used and articulation. Using our articulatory synthesizer, however, to make a continuum from [dɑ] to [nɑ] in English we obtained the normal results in such discrimination tests. While not wishing to enter into any controversy concerning the reasons for categoricalness in speech perception and its possible links to articulatory control, we can at least conclude that letting the output spectrum of our synthetic syllables vary with increments of articulatory change i.e., changes in the size of the velopharyngeal port, yields results that are in agreement with recent experiments using a terminal analog synthesizer on the same phonetic distinction.

We plan to extend our research on the oral–nasal contrasts from consonants to vowels, but, unlike House and Stevens, we intend to experiment on vowels from a language in which the oral–nasal contrast is phonologically relevant. The experiment which we wish to carry out will require that the chosen language possess pairs of vowels ranging from high to low. This excludes, for example, French in which distinctive nasality is restricted to two vowel heights. A suitable language, we

believe, is Hindi. We are now preparing the groundwork for a study based on that language as a sequel to the present paper.

ACKNOWLEDGMENTS

We wish to thank Thomas Baer for his thoughtful comments on this paper and Philip Rubin for his unstinting assistance in mastering the speech synthesizer. We have also benefitted from discussion with Frederika Bell-Berti and Alvin M. Liberman. This research was supported under NSF Grant BNS-76-82023 and BRSG Grant RR-05596 to Haskins Laboratories.

[1]The acoustic analyses of nasal consonants, reported by Fant (1960), Fujimura (1961), and Fujimura and Lindqvist (1971) have revealed a number of salient features. The first formant typically occurs around 300 Hz and is well separated from the upper formant structure, due to the combined effects of the pharyngeal and nasal cavities determining the fundamental resonance. The increased length of the direct acoustic transmission path causes a reduction in the concomitant average formant spacing and, therefore, a high density of formants in the middle-frequency range. The higher rate of energy loss in the nasal cavity causes greater damping of the resonances with a characteristic increase in the widths of some of the formant peaks; in the case of /n/ the relatively high absorption of sound energy in the termination of the oral cavity shunt causes greater damping of the antiresonance and, consequently, an increase in its bandwidth.

[2]In nature, the velopharyngeal port is not circular. Björk's (1961) tomographic and cineradiographic study concludes that the velopharyngeal port area is a linear function of the port's sagittal minor axis and that the constant of proportionality is 10 mm. Hence, the anatomical structure of the port is more nearly rectangular. However, for the purpose of computing the degree of oral–nasal coupling, only the total area of the port need be considered and a circular approximation to the velopharyngeal port introduces no significant error.

[3]A modification of this specification, termed "partial coarticulation," was used in a pilot study of oral–nasal category boundary movement and in the experiment on categorical perception reported in this paper. The modification involved closing the velum at the end of the tongue movement when it had reached its position for the steady-state vowel. The boundary movement data obtained with these utterances were essentially the same as, although somewhat less robust than, the results obtained with "fully coarticulated" vowels (i.e., when the velum remained lowered throughout the vowel).

[4]The evidence was indirect because it rested on the implicit assumption that the relative F_1 amplitude and the impedance ratio at the F_1 peak frequency are directly related. In practice, this relationship depends critically on the damping characteristics of the nasal tract and becomes less direct as the damping factor is reduced and the opportunity for nasal resonance is enhanced. The nasal tract of House and Stevens' articulatory analog appears to have been heavily damped and, therefore, they could assume a direct relationship. Since the nasal resonance characteristics of different speakers appear to differ widely and since there are no definitive acoustic data on the subject of nasal damping, it is difficult to make any strong claims for behavioral accuracy from the nasal section of any articulatory model. This is a point made by House and Stevens in their paper and one to which we must also subscribe.

[5]It is appropriate at this point to comment on the velar port areas that have marked the category boundaries in our experiments. Some early observations made by Passavant (1863) and a co-worker, who controlled a speaker's minimum velar port area by means of rubber tubes of differing internal diameters, have been cited by Fritzell (1969). These observations appear to suggest that the degree of velar port openness which is able to influence consonant production may be significantly greater than that revealed by the present data. Similar, more recent, work by Isshiki *et al.* (1968) seems to lead to a comparable conclusion. However, it should be noted that neither of these studies employed groups of listeners who were tested on repeated productions of isolated consonant–vowel syllables. The purpose of the tubes was to find the point at which nasal distortion became apparent in *continuous* speech production. Moreover, informal methods of judging that distortion were deemed to be satisfactory. Reviewing the results of these perceptual judgments together with electromyographic and other data, Bell-Berti (1980) notes that "oral consonants are distorted at smaller port areas than are close vowels." The experiments described here suggest that, because only a binary consonantal judgment was required and coarticulation was present, listeners may have been satisfied by a relatively "weak" acoustic cue to mark the perceptual boundary between oral and nasal categories.

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

Ackerman, E. L. (1935). "Action of the velum palatinum on the velar sounds /k/and /g/," Vox **31**, 2–9.

Bell-Berti, F. (1976). "An electromyographic study of velopharyngeal function in speech," J. Speech Hear. Res. **19**, 225–240.

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

Bell-Berti, F. (1980). "Velopharyngeal function: a spatial-temporal model," in *Speech and Language: Advances in Basic Research and Practice*, Vol. IV, edited by Norman J. Lass (Academic, New York).

Björk, L. (1961). "Velopharyngeal function in connected speech," Acta Radiol. Suppl. 202.

Bloomer, H. (1953). "Observations on palato-pharyngeal movements in speech and deglutition," J. Speech Hear. Dis. **18**, 230–246.

Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, 's-Gravenhage).

Finney, D. J. (1971). *Probit Analysis* (Cambridge U. P., Cambridge).

Fowler, C. A. (1980). "Coarticulation and theories of extrinsic timing," J. Phonetics **8**, 113–133.

Fritzell, B. (1969). "The velopharyngeal muscles in speech: An electromyographic and cineradiographic study," Acta Oto-laryngol. Suppl. 250.

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

Fujimura, O., and Lindqvist, J. (1971). "Sweep-tone measurements of vocal-tract characteristics," J. Acoust. Soc. Am. **49**, 541–558.

Harrington, R. A. (1944). "A study of the mechanism of velopharyngeal closure," J. Speech Dis. **9**, 325–345.

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

House, A. S. (1957). "Analog studies of nasal consonants," J. Speech Hear. Dis. **22**, 190–204.

House, A. S., and Stevens, K. N. (1956). "Analog studies of the nasalization of vowels," J. Speech Hear. Dis. **21**, 218–232.

Isshiki, N., Honjow, I., and Morimoto, M. (1968). "Effects of velopharyngeal incompetence upon speech," Cleft Palate J. **5**, 297–310.

Kelly, J. L., and Lochbaum, C. (1962). "Speech synthesis," Proc. Fourth Int. Congr. Acoust. G42, 1–4.

Kent, R. D., and Minifie, F. D. (1977). "Coarticulation in recent speech production models," J. Phonetics 5, 115–117.
Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (1957). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. 54, 358–368.
Lubker, J. (1968). "An electromyographic-cinefluorographic investigation of velar function during normal speech production," Cleft Palate J. 5, 1–18.
McDonald, E. T., and Baker, H. K. (1951). "Cleft palate speech: an integration of research and clinical observation," J. Speech Hear. Dis. 16, 9–20.
Mermelstein, P. (1972). "Speech synthesis with the aid of a recursive filter approximating the transfer function of the nasalized vocal tract," Proc. IEEE Conf. Speech Commun. Process., Newton, MA, 152–155.
Mermelstein, P. (1973). "Articulatory model for the study of speech production," J. Acoust. Soc. Am. 53, 1070–1082.
Miller, J. L., and Eimas, P. D. (1977). "Studies on the perception of place and manner of articulation: a comparison of the labial–alveolar and nasal–stop distinctions," J. Acoust. Soc. Am. 61, 835–845.
Moll, K. L. (1962). "Velopharyngeal closure on vowels," J. Speech Hear. Res. 5, 30–37.
Moll, K. L., and Shriner, T. H. (1967). "Preliminary investigation of a new concept of velar activity during speech," Cleft Palate J. 4, 58–69.
Nusbaum, E. A., Foley, L., and Wells, C. (1935). "Experimental studies of the firmness of velar-pharyngeal occlusion during the production of English vowels," Speech Monographs 2, 71–80.
Nylén, B. O. (1961). "Cleft palate and speech," Acta Radiol. Suppl. 203.
Passavant, G. (1863). *Über die Verschliessung des Schlundes beim Sprechen* (J. D. Sauerländer, Frankfurt a. M.).
Podvinec, S. (1952). "The physiology and pathology of the soft palate," J. Laryngol. Otol. 66, 452–461.
Rubin, P., and Baer, T. (1978). "An articulatory synthesizer for perceptual research," J. Acoust. Soc. Am. Suppl. 1, 63, S45.
Rubin, P., Baer, T., and Mermelstein, P. (1979). "An articulatory synthesizer for perceptual research," Haskins Labs. Status Rep. Speech Res. SR–57, 1–15.
Studdert–Kennedy, M. (1980). "Speech perception," Lang. Speech 23, 45–65.
Warren, D. (1967). "Nasal emission of air and velopharyngeal function," Cleft Palate J. 4, 148–156.