# Acoustic and perceptual correlates of the non-nasal–nasal distinction for vowels

495

Sarah Hawkins
*Haskins Laboratories, New Haven, Connecticut 06511*

Kenneth N. Stevens
*Department of Electrical Engineering and Computer Science, and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

(Received 19 June 1984; accepted for publication 28 December 1984)

For each of five vowels [i e ɑ o u] following [t̪], a continuum from non-nasal to nasal was synthesized. Nasalization was introduced by inserting a pole-zero pair in the vicinity of the first formant in an all-pole transfer function. The frequencies and spacing of the pole and zero were systematically varied to change the degree of nasalization. The selection of stimulus parameters was determined from acoustic theory and the results of pilot experiments. The stimuli were presented for identification and discrimination to listeners whose language included a non-nasal–nasal vowel opposition (Gujarati, Hindi, and Bengali) and to American listeners. There were no significant differences between language groups in the 50% crossover points of the identification functions. Some vowels were more influenced by range and context effects than were others. The language groups showed some differences in the shape of the discrimination functions for some vowels. On the basis of the results, it is postulated that (1) there is a basic acoustic property of nasality, independent of the vowel, to which the auditory system responds in a distinctive way regardless of language background; and (2) there are one or more additional acoustic properties that may be used to various degrees in different languages to enhance the contrast between a nasal vowel and its non-nasal congener. A proposed candidate for the basic acoustic property is a measure of the degree of prominence of the spectral peak in the vicinity of the first formant. Additional secondary properties include shifts in the center of gravity of the low-frequency spectral prominence, leading to a change in perceived vowel height, and changes in overall spectral balance.

PACS numbers: 43.71.Es, 43.70.Fq, 43.70.Hs, 43.71.An

## INTRODUCTION

### A. Review of the acoustics of nasal vowels

The phonemic opposition of oral and nasal vowels is widespread among languages, but the acoustic property underlying this opposition is not well understood. A nasal vowel is produced by introducing acoustic coupling between the oral and nasal cavities at a point that is about halfway along the vocal tract between the glottis and the lips. The effect of this acoustic coupling is to shift the natural frequencies of the vocal tract (i.e., the formant frequencies for the non-nasal vowel), and to add pole-zero pairs to the vocal-tract transfer function (Fant, 1960; Fujimura, 1960, 1961; Fujimura and Lindqvist, 1971; House and Stevens, 1956; Mrayati, 1976; Lonchamp, 1979). The principal and most consistent consequence on the acoustic spectrum of the vowel appears to be at low frequencies, in the vicinity of the first formant. When the cross-sectional area of the velopharyngeal opening is gradually increased, there is usually a shift in the first formant frequency, and there is an increase in its bandwidth. A pole-zero pair is introduced in the vicinity of the first formant and the spacing between the pole and zero increases as the velopharyngeal opening increases, with the result that the additional pole shows increased spectral prominence with larger openings. These observations are based on calculation of the transfer function of a system of coupled tubes, and are confirmed by data from the acoustic spectra of naturally spoken nasal vowels and from studies of the behavior of articulatory models.

Sometimes more than one additional resonance in the vicinity of the first formant can be observed in acoustic spectra of nasal vowels. These additional resonances must arise because the impedance of the nasal cavity has more than one low-frequency resonance below about 1.5 kHz. These resonances are presumably a consequence of the sinuses, which are often of appreciable volume, and which are coupled into the nasal cavity proper through relatively narrow openings (Maeda, 1982a; Lindqvist and Sundberg, 1972). In any event, the end result is that the introduction of acoustic coupling to the nasal cavity modifies the spectrum of a vowel in the vicinity of the first formant, so that the narrow spectral prominence associated with the first formant is replaced by a pole-zero-pole combination (possibly with additional pole-zero pairs). The additional pole-zero pairs can result in a "filling in" of the valleys in the spectrum above or below the frequency of the original first formant. Consequently there is a broader frequency region in the vicinity of the first formant over which the spectral energy is distributed for a nasal vowel, as opposed to a relatively narrow spectral prominence for a non-nasal vowel. This type of modification of the spectrum was first noted by Hattori *et al.* (1958).

0001-4966/85/041560-16$00.80

In addition to modification of the spectral shape in the vicinity of the first formant, nasalization in natural speech can also give rise to changes in the spectrum at higher frequencies. There may be shifts in the frequencies of higher formants, modifications in the amplitudes of the spectral peaks, and introduction of additional spectral peaks. These changes in the acoustic spectrum at high frequencies do not seem to be as consistent across different speakers and vowels as those in the vicinity of the first formant. This lack of consistency in the effects of nasalization at higher frequencies is predictable in view of the substantial individual differences in the anatomy of the nasal cavities.

Spectrograms and spectra contrasting nasal and non-nasal vowels produced by a male speaker of Gujarati (a language spoken in India which has a nasal–non-nasal phonemic opposition for each of its vowels) are illustrated in Fig. 1. The frequency resolution of the analyzing system does not always permit clear observation of a zero in the spectrum, but at least one additional pole and consequent broadening of the low-frequency spectral prominence can be seen in each of the nasal vowels. In all the nasal spectra the first formant peak is rendered less prominent by the filling in of the valley in the spectrum either above or below the first formant.

## B. Perception of nasal vowels

Experiments indicate that nasal vowels can function perceptually as a class in distinction from oral vowels, even for speakers of languages lacking the non-nasal–nasal opposition as a phonemic contrast (Butcher, 1976; Wright, 1980). Several studies have attempted to determine the acoustic and articulatory attributes that give rise to listeners' perception of nasal vowels as a class and as distinct from their non-nasal counterparts. These studies fall roughly into two categories depending upon how the stimuli are manipulated: studies searching for the acoustic prerequisites for the perception of nasality by direct manipulation of the spectrum of the sound, and studies with articulatory synthesizers in which the perception of different degrees of nasalization is examined by varying the area of the velopharyngeal port.

Studies involving direct manipulation of the spectrum indicate that vowels are judged to be nasal when the spectrum is modified in the low-frequency range and sometimes when the amplitude of the third formant is increased. Using the Haskins pattern-playback system, Delattre (1954) produced nasal vowels by adding energy in the vicinity of the fundamental frequency. Takeuchi *et al.* (1974, 1975) added pole-zero pairs to spectra of natural speech at various frequencies and found that the greatest number of nasal judgments tended to occur when a pole-zero pair was introduced in the vicinity of the first formant. Raising the amplitude of the third formant further increased judgments of nasality for some vowels. For most vowels nasal judgments were in the range 40%–60%, although they were higher for [u] and [ʉ].

In studies with articulatory synthesizers, House and Stevens (1956) and Maeda (1982a) produced several steady-state vowels with various degrees of opening of the velopharyngeal port. Their results showed that a greater velopharyngeal opening was needed to produce a given level of nasality (as judged by listeners) for the low vowels than for the high

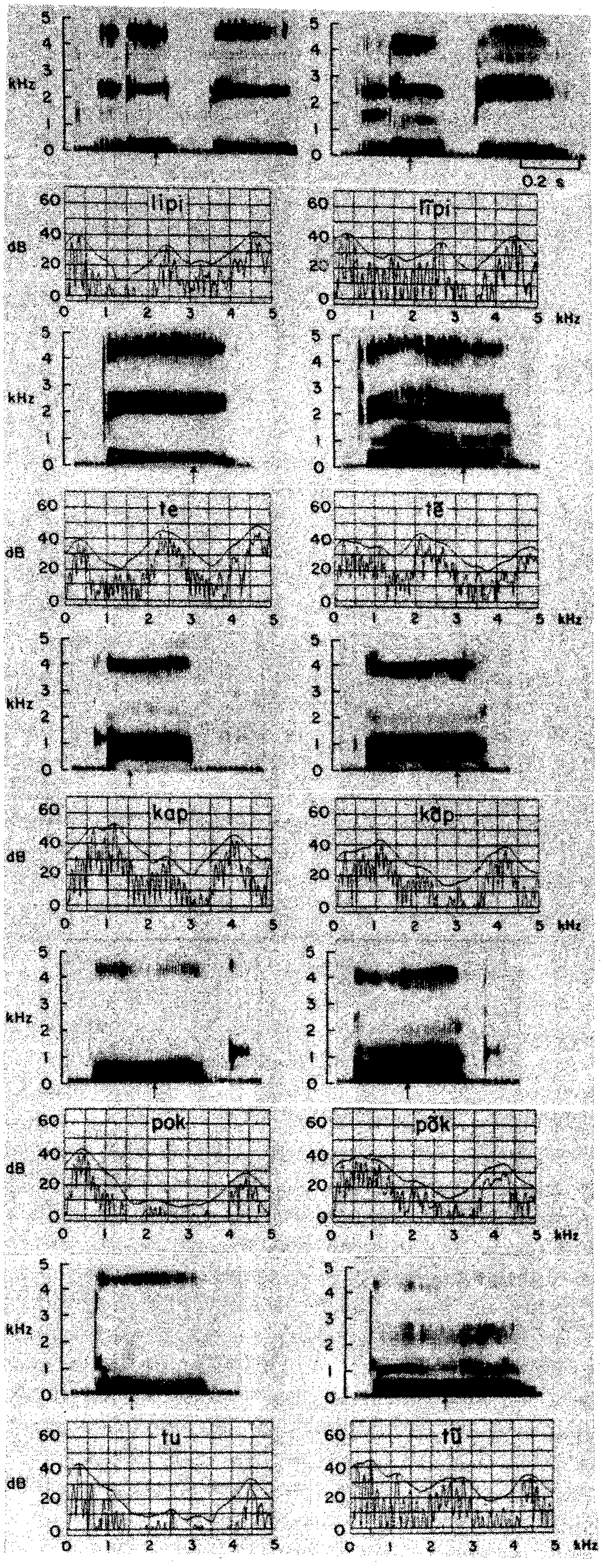

FIG. 1. Spectrograms of words illustrating the non-nasal–nasal distinction for vowels in Gujarati, together with spectra sampled within the vowels. The non-nasal vowels are in the left column, with the spectra immediately below the corresponding spectrograms, and the nasal vowels are in the right column. The time at which the spectrum is sampled is indicated by the arrow. Two spectra are shown in each panel, each representing a spectrum sampled over a 25.6-ms Hamming window. The lower curve is the discrete Fourier transform, and the upper smooth curve is a "pseudo-spectrum" for which each point is obtained by performing a weighted average of groups of points on the DFT over a bandwidth of 300 Hz. Speaker VS (male).

vowels. Observations of the articulators using cineradiography (Delattre, 1968), fiberoptics (Benguerel and Lafargue, 1981), and fiberoptics with electromyography (Henderson, 1984) also show that there are in fact systematic differences in velopharyngeal port area depending upon vowel height. In both House and Stevens' (1956) and Maeda's (1982b) studies, the spectra of stimuli that were judged to be nasal were observed to be flatter than non-nasal spectra in the vicinity of the first formant. That is, the prominence of the spectral peak for the formant was reduced, with additional peaks sometimes present in the vicinity of the formant. These observations support the findings of studies that manipulated the spectrum directly.

Beddor and Strange (1982) examined the identification and discrimination of stimuli varying along a continuum from [bɑ] to [bɑ̃]. These stimuli were produced by systematically varying the velopharyngeal port area in articulatory synthesis. Identification functions were steeper for Hindi listeners, for whom the non-nasal–nasal vowel contrast is phonemic, than for American English listeners, for whom it is not. The two language groups also differed slightly in their placement of the 50% crossover boundary. Discrimination functions were similar for the two groups of listeners when the difference between stimuli was small: responses were around chance at the endpoints and peaked around the category boundary. In a simplified discrimination task, however, the American but not the Hindi listeners appeared better able to distinguish between non-nasal than between nasal vowels. These data indicate that linguistic experience may affect the perception of vowel nasality in relatively complex ways.

The literature suggests, then, that vowels will be perceived as nasal if the acoustic spectrum of a non-nasal vowel is modified so that the prominence in the vicinity of the first formant is broadened and flattened. For some vowels, additional consequences of the low-frequency spectral modification, which is the primary and most consistent modification, are changes in the amplitude of higher-frequency components, and/or changes in the perceived height of the vowel due to a shift in the center of gravity of the low-frequency prominence. Beddor and Strange's (1982) results indicate that responses to stimuli varying along a non-nasal–nasal continuum are not strongly dependent on whether the listener's native language distinguishes phonemically between nasal and non-nasal vowels. There are, however, behavioral differences that do appear to be dependent upon linguistic experience.

A possible conclusion from these data is that nasality in vowels involves some acoustic property that is common to all nasal vowels and that may be perceived as nasality regardless of linguistic experience, together with other properties that may differ across vowels and across languages. The purpose of this paper is to attempt to identify and specify the parameters of such a universal nasal property, using synthetic speech in order to minimize the occurrence of the more language- or vowel-specific concomitant properties. The first experiment described here sought to produce good exemplars of nasal vowels by inserting an additional pole-zero pair in the spectrum of non-nasal vowels, in accordance with predictions from acoustic theory and analysis of natural speech. Subsequent experiments used the "best" exemplar for each nasal vowel as endpoints in acoustic continua. These continua were designed to test whether the presence of a low-frequency pole-zero pair produces similar responses from listeners with different linguistic backgrounds.

## I. GENERAL DESCRIPTION OF STIMULI

In all the experiments the stimuli were consonant–vowel syllables beginning with a voiceless unaspirated dental stop, [t̪], and the vowel was one of [i e ɑ o u] or their nasal counterparts, [ĩ ẽ ɑ̃ õ ũ]. Parameter values for these stimuli were selected in accordance with measurements of words produced by a male native speaker of Gujarati. The syllables were generated on a Klatt synthesizer (Klatt, 1980) with the formant resonators connected in cascade. The duration of the vowel was 325 ms, and the fundamental frequency was given an appropriate rising–falling contour. The initial stop burst was identical for all stimuli, and the first 40 ms after voicing onset were identical for both non-nasal and nasal vowels of the same vowel quality.

A spectrogram of a typical nasal stimulus is shown in Fig. 2, and trajectories of various parameters for this stimulus are displayed in Fig. 3. The time from stop consonant release to onset of voicing was 10 ms. Immediately following the onset of voicing, there were appropriate formant transitions from the dental consonant to the steady-state vowel with which each syllable terminated. (The example in Figs. 2 and 3 for the vowel [ẽ] is slightly atypical, in that there was some diphthongization of the vowel, as evidenced by the slowly rising $F2$, in accordance with measurements on the Gujarati tokens. There was no such diphthongization for the other vowels.) Detailed characteristics of the initial transitions and the non-nasal vowels are listed in Table I. Slight modifications of the measured formant frequencies for the natural Gujarati tokens were made in order to obtain a best match of vowel quality (based on informal listening) between the synthesized vowels and the naturally produced vowels.

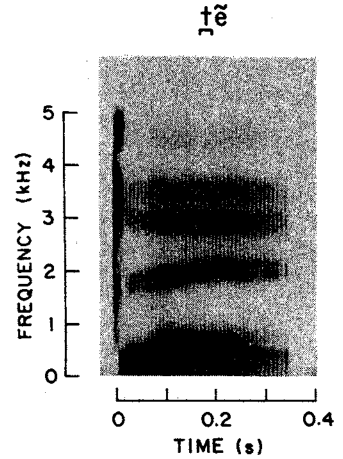

FIG. 2. Spectrogram of a typical synthetic consonant–vowel stimulus [t̪ẽ]. Introduction of a pole-zero pair to simulate nasalization begins at 40 ms, and nasalization is complete at 80 ms. For this vowel, some diphthongization is introduced through the rising second formant in the initial part of the vowel.

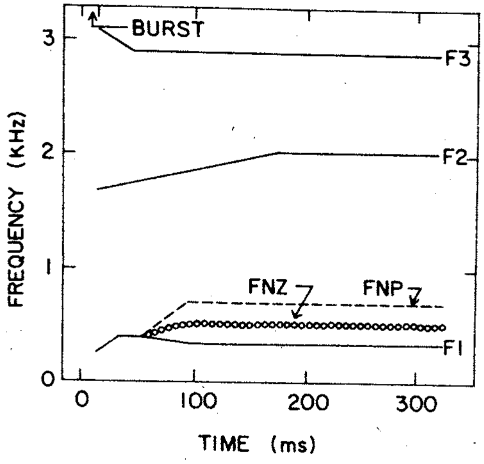

FIG. 3. Trajectories of the first three formant frequencies and of the frequencies of the additional pole (FNP) and zero (FNZ) for the synthetic nasal stimulus [t̪ẽ], whose spectrogram is shown in Fig. 2. Time is measured from the onset of the burst, whose duration was 15 ms.

For all vowels, the fourth and fifth formant frequencies were 3500 and 4500 Hz (except $F4 = 3700$ Hz for [i]), and bandwidths B3, B4, and B5 for the higher formants were 150, 170, and 250 Hz, respectively.

Each synthetic nasal vowel differed from its non-nasal congener in either one or two ways: all nasal vowels contained an additional pole-zero pair, and in some cases the frequency of $F1$ differed from that of the equivalent non-nasal vowel. The additional pole-zero pair for a nasal vowel began to separate 40 ms after voicing onset, and during the next 40 ms the frequencies of the pole and zero underwent piecewise-linear motions toward values appropriate for the intended nasal vowel. The beginning values of the frequency of the pole (FNP) and the frequency of the zero (FNZ) at 40 ms after voicing onset were always 400 Hz. The bandwidths of FNP and FNZ were always 100 Hz. In addition to introducing a local perturbation of the spectrum in the vicinity of FNP and FNZ, the pole-zero pair also has a small influence on the amplitudes of the formant peaks at higher frequencies. For the vowels in which the frequency of the first formant ($F1$) in the nasal vowel was different from $F1$ for the corresponding non-nasal vowel, $F1$ underwent a piecewise-linear trajectory from the non-nasal to the nasal value during the time interval from 40–80 ms after voicing onset. In the example in Figs. 2 and 3, the non-nasal $F1$ is 400 Hz and the nasal $F1$ is 350 Hz.

TABLE I. Characteristics of the non-nasal vowels that were used as starting points in developing the nasal vowels and the non-nasal–nasal continua. $F1$ to $F3$ represent the steady-state vowel formant frequencies; $F1_i$ to $F3_i$ are the starting frequencies of the first three formants for the transitions following the release of the burst for the initial [t]; and $B1$ and $B2$ are the bandwidths of the first two formants. All frequencies are in Hz. Other parameters of the stimuli are described in the text.

| | i | e | ɑ | o | u |
|---|---|---|---|---|---|
| $F1$ | 270 | 400 | 700 | 430 | 270 |
| $F1_i$ | 200 | 270 | 350 | 270 | 200 |
| $F2$ | 2300 | 2025 | 1150 | 850 | 850 |
| $F2_i$ | 1800 | 1700 | 1500 | 1350 | 1350 |
| $F3$ | 2900 | 2915 | 2500 | 2500 | 2500 |
| $F3_i$ | 3000 | 3100 | 2800 | 2800 | 2800 |
| $B1$ | 60 | 60 | 80 | 80 | 80 |
| $B2$ | 80 | 80 | 100 | 100 | 100 |

The selection of 40 ms as the duration of the initial non-nasal time interval following the obstruent consonant before the onset of nasalization was based on informal observation of this time interval for a number of CV utterances in several languages. There is considerable variability in this time interval, but 40 ms is within the observed range. The selection of 400 Hz as the frequency at which the pole and zero begin to separate represents an estimate of the natural frequency of the nasal cavity with a closed velopharyngeal port. This value is consistent with estimates reported in the literature (House and Stevens, 1956; Fant, 1960; Maeda, 1982a; Lindqvist and Sundberg, 1972) although, again, there is presumably considerable interspeaker variability, and there are no satisfactory measurements of this frequency for a number of talkers.

## II. PRELIMINARY EXPERIMENT

The aim of the preliminary experiment was to establish a configuration of $F1$, FNP, and FNZ that would lead to acceptable versions of each of the nasal vowels [ĩ ẽ ɑ̃ õ ũ] for native speakers of Gujarati, which has all these vowels in its phonemic inventory. Our strategy was to generate a number of possible versions of each of the nasal vowels by manipulating $F1$, FNP, and FNZ, and to ask listeners to rate these stimuli in terms of their acceptability as nasal vowels.

### A. Procedure

Informal observations, as well as theoretical considerations, led us to expect that the best versions of nasal vowels would be obtained when the frequency of the nasal zero was placed approximately midway between the original first formant frequency and the frequency of the nasal pole. The absolute difference in frequency between the two poles was in question, however. We were also aware that nasalization of the vowel might cause some shift in $F1$ relative to its frequency in the non-nasal vowel, especially for nonhigh vowels. For each vowel, then, we generated 12–24 stimuli with various values of FNP and FNZ (with FNZ always between FNP and $F1$) and, in some cases, with more than one value of $F1$.

The parameter values for the constellations of test stimuli for each vowel are shown as the points plotted in Fig. 4. In this figure we plot the values of FNP and FNZ for each stimulus. Figure 4 also displays the $F1$ values for the stimuli (horizontal lines) except for [ɑ̃], where $F1$ is too high to be displayed, and gives loci of points corresponding to FNZ midway between $F1$ and FNP (diagonal lines), i.e., $\text{FNZ} = (\text{FNP} + F1)/2$.

There were 86 stimuli in all. Three repetitions of each stimulus were generated, and the resulting stimuli were arranged in random order and recorded on four test tapes. The test tapes were presented to three groups of listeners: three native speakers of Gujarati, six native speakers of American English, and five native speakers of other languages of northern India (Punjabi, Hindi, and Bengali) which have a nasal–non-nasal opposition for vowels. Listeners were given an answer sheet on which the intended nasal vowel was indicated

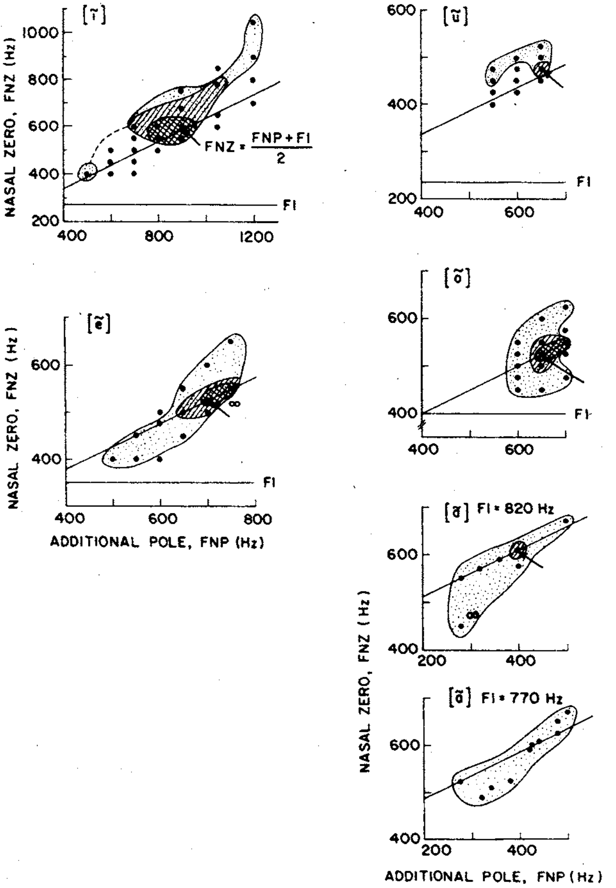

FIG. 4. Stimuli used in the preliminary experiment to select a good nasal vowel for each vowel quality. Each stimulus is represented in terms of the frequency of the additional pole (FNP) and additional zero (FNZ). Note that the axes for [ĩ] are half the scale of those of the other vowels. The frequency of $F1$ is represented as a horizontal line (except for [ɑ̃], where it is given numerically). The diagonal line on each graph is the frequency at which the nasal zero is exactly halfway between the frequencies of $F1$ and the nasal pole [$\mathrm{FNZ} = (\mathrm{FNP} + F1)/2$]. The arrows and the dotted, hatched, and cross-hatched regions identify stimuli whose responses fell into particular categories as follows. Stimuli outside the dotted areas were excluded from consideration due to inconsistencies in the responses, or due to poor vowel quality. Stimuli within the hatched areas were those which subjects judged most nasal. Stimuli within the cross-hatched areas were those most preferred within this last (most nasal) group. (See text for details.) The arrows point to the stimuli chosen as the nasal endpoint of a non-nasal–nasal continuum for each vowel. Within a given vowel set, all stimulus parameters other than FNP and FNZ remained the same, with these exceptions (indicated by open circles): for [e], of the two stimuli with FNP = 750 Hz and FNZ = 525 Hz, $F1$ was 400 Hz in one case and 430 Hz in the other; for [ɑ] with $F1 = 820$ Hz, of two stimuli with FNP = 300 Hz and FNZ = 470 Hz, $F2$ was 1150 Hz (as for all the other stimuli on this graph) in one case, and uniquely set at 1300 Hz in the other case.

for each stimulus item. For each stimulus, the listeners were asked to respond to two questions: (1) indicate if the vowel quality is unlike that of the intended vowel; (2) indicate the adequacy of the nasal quality of the vowel on a four-point scale ranging from "poor" to "good."

## B. Results

Analysis of the judgments of vowel quality indicated that the American listeners judged the vowel quality to be inadequate more often than did the Indian listeners, and were also more consistent in their judgments. For this reason, a stimulus was excluded from further consideration if it accrued a total of four or more "poor" judgments from the American listeners. (Stimuli excluded on this criterion were as follows: 13 tokens of [ĩ]; two of [ẽ]; and six of [ũ]. These numbers reflect the difficulty we had in synthesizing a range of high nasal vowels of good quality.) It was assumed that the remaining stimuli were all adequate in vowel quality.

Some of the remaining stimuli elicited inconsistent judgments of degree of nasality for certain subjects. An inconsistent judgment was defined as one where the range of ratings was greater than one (on the four-point scale) for the three repetitions of each stimulus. For example, a rating by a single subject of 2, 3, 2 for a given stimulus was considered to be a "consistent" response; ratings of 2, 3, 1 or 2, 4, 4, however, would be considered inconsistent responses. Stimuli eliciting three or more such inconsistent sets of ratings from the American listeners, or four or more from the Indian listeners, were excluded from the analysis. (These stimuli were two tokens of [ẽ] and two of [õ].) In all, a total of 25 stimuli were excluded on the basis of poor quality or inconsistent response. The following numbers of stimuli were left in each category: 11 tokens of [ĩ], 12 of [ẽ], 19 of [ɑ̃], 13 of [õ], and six of [ũ]. These are those stimuli within the dotted areas of Fig. 4.

The remaining stimuli were then ranked according to their nasality ratings. The range of ratings was small for all language groups, especially for the Indian listeners. The maximum possible range was 3.0 points. For each vowel set, the Indian listeners' average range of ratings was always less than one point, whereas the American listeners' average range of ratings varied from just less than one point (for [ɑ̃]) to almost two points (for [ĩ]). That is, the Indian listeners were not discriminating between the stimuli in terms of nasality, as reinforced by various statistical analyses. For example, Kendall coefficients of concordance between ratings for each vowel by the three Gujarati listeners fell far short of significance. In contrast, Spearman rank-order correlations between the American listeners, randomly divided into two groups of three, were significant at the 0.05 level of probability or better for all vowels except [ɑ̃], for which there were a large proportion of tied ranks. (Spearman correlations were used with the American listeners since a Kendall coefficient of concordance on six individuals is clumsy to compute.)

The fact that the Indian listeners appeared to differentiate very little between the stimuli in terms of nasality is consistent with the results of discrimination experiments reported later in this paper as well as by Beddor and Strange (1982). It did not help us to select the "ideal nasal vowels" for our subsequent experiments, however. It was decided to identify the most acceptable nasal vowels from the rankings of the American listeners, as long as those vowels selected were *among* those most favored by the Gujaratis. That is, the wider range of rankings given by the Americans was taken as evidence that their rankings probably gave the most reliable measure of adequacy, while the validity of the chosen tokens as acceptable Gujarati nasal vowels was ensured by requiring these tokens to be among the three or four most preferred by the Gujaratis. The most acceptable pole-zero-pole combinations as defined in this way are those falling within the hatched areas in Fig. 4. For all the stimuli, the average rating by American listeners was at least 2.9 on the four-point scale from 1 (poor) to 4 (good). For the vowels [ĩ], [ẽ], and [õ], certain pole-zero-pole combinations stood out as most preferred within a wider range of preferred combinations. These stimuli are identified by the cross hatching in Fig. 4.

It can be seen from Fig. 4 that all of the preferred pole-zero combinations of each vowel lie on or close to the line labeled $\mathrm{FNZ} = (\mathrm{FNP} + F1)/2$, and there is at least 100 Hz between the zero and each pole. This finding is consistent with predictions from acoustic theory and provides a systematic basis for selecting parameters for a set of "ideal" nasal vowels.

Parameter values of the stimulus that received the highest nasality rating within the preferred group for each of the five vowel sets were selected as the nasal endpoints of the five continua used in the identification and discrimination experiments which follow. The stimuli are identified with an arrow in Fig. 4. The characteristics of these most preferred nasal vowels are summarized in Fig. 5. Figure 5 shows again that the best nasal vowels were produced with a zero that is about midway between the first formant and the additional resonance. It indicates the shift in the first formant for the nasal vowel in relation to that for the corresponding non-nasal vowel. It also shows that there is not much variation in the location of the low-frequency pole-zero-pole combination from one nasal vowel to another compared with the variation in $F1$ for non-nasal vowels. This configuration of two poles and a zero is similar to that predicted on the basis of acoustic theory (Stevens *et al.*, in press).

## III. IDENTIFICATION TEST

Having established an appropriate set of synthetic nasal vowels (in a [tṼ] context) that were acceptable to a group of listeners, we proceeded next to produce, for each of the five vowels, a series of stimuli ranging from a non-nasal to a nasal extreme. Vowels in these series were identified as nasal or non-nasal by listeners whose language included a non-nasal–nasal distinction for vowels and by listeners whose language did not use that distinction. The principal aims of these experiments were to determine: (1) whether listeners yield similar identification responses to the stimuli independent of their experience with the non-nasal–nasal distinction, and (2) whether some acoustic property or processing mechanism can be defined, independent of the vowel, that characterizes vowels identified as nasal as opposed to vowels identified as non-nasal.

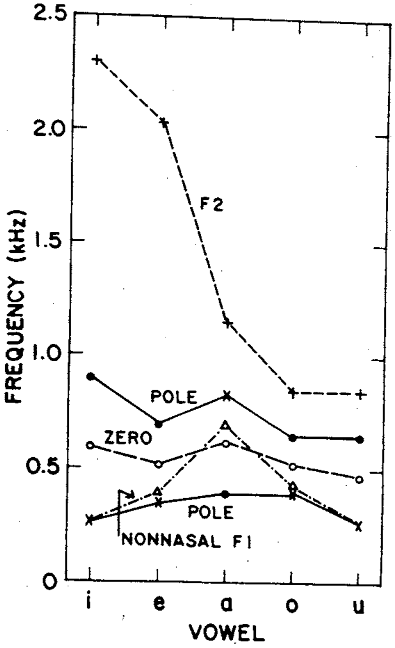

FIG. 5. Frequencies of poles and zero corresponding to the nasal endpoint stimulus for each vowel set, identified by the arrows in Fig. 4. Also shown (△) is the frequency of the first formant for each non-nasal vowel. The crosses (×) represent the (shifted) first-formant frequency ($F1$), solid circles (●) represent the frequency of the additional pole (FNP), and open circles (○) represent the frequency of the zero (FNZ).

### A. Stimuli

For each of the five series of stimuli, the vowel at the non-nasal end of the continuum had the formant frequencies listed in Table I, and the pole-zero pair was located at 400 Hz (i.e., there was no effect of the additional pole and zero). For the vowel at the nasal end of the continuum, values for $F1$, FNZ, and FNP defined stimuli that were rated as the most acceptable nasal vowels in the preliminary experiment, as summarized in Fig. 5. Intermediate stimuli on the continuum were obtained by interpolating in equal steps between values of $F1$, FNZ, and FNP for the non-nasal and nasal extremes. There were nine items in the continuum for the vowels [i e ɑ o] and eight for [u]. Figure 6 shows values of $F1$, FNZ, and FNP for the continua for each of the vowels. For the high vowels [i] and [u] there is no shift in $F1$ along the continuum, whereas there is a change in $F1$ for the vowels [e ɑ o].

In Fig. 7 we show spectra of three of the stimuli in each continuum: the two endpoint stimuli and an intermediate stimulus. While the stimulus description in terms of pole and zero locations in Fig. 6 indicates that there is a regular progression in the stimulus properties for each continuum, examination of Fig. 7 suggests that there may be other ways of describing the stimuli. Stimulus changes might be specified in terms of amplitudes of particular spectral components or the ratio of the amplitude of a spectral peak to the amplitude of a valley, or in terms of the center frequency or width of a gross spectral prominence that may consist of two or three closely spaced peaks.

For example, in the case of the stimuli on the [i–ĩ] continuum, the spectrum is changing in several ways that could potentially be used by a listener either to discriminate between stimuli or to identify nasality. These include: (1) the frequency of the additional pole, (2) the relative amplitude of the spectral peak arising from this pole, and (3) the relative spectral amplitude at high frequencies, which changes gradually over the entire continuum. Similar changes in the amplitudes of spectral peaks occur for stimuli on the other continua. For example, there is an increase of 8 to 12 dB in the relative amplitude of the high-frequency peaks (amplitude of $F3$ peak and above in relation to $F1$ peak) from the non-nasal end of the continuum to the nasal end for [i e o u], for which $\mathrm{FNZ} < \mathrm{FNP}$. For the vowel [ɑ], for which $\mathrm{FNZ} > \mathrm{FNP}$, there is a decrease of about 5 dB in the absolute high-frequency amplitude over the continuum, but the

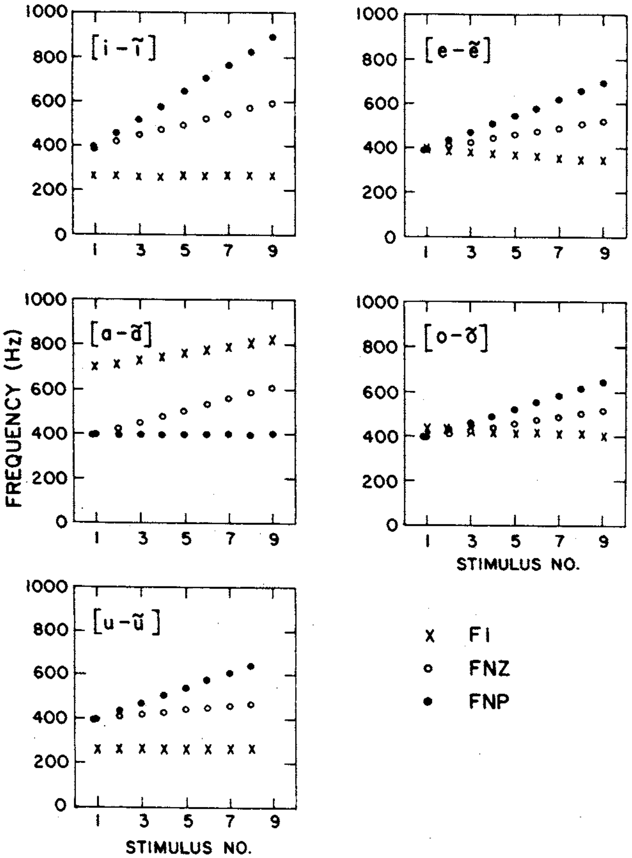

FIG. 6. Frequencies of the two poles (× and ●) and zero (○) for the stimuli comprising the non-nasal–nasal continua for each vowel. For the non-nasal vowel in each case (stimulus 1) the additional pole and zero were set at 400 Hz, and canceled each other.

change in relative amplitude follows the same pattern as for the other vowels, the increase being just 4 dB between stimuli 1 and 9.

For all of the continua there is a spectral minimum, arising from the zero, and this minimum can be seen in the spectrum of the most nasal stimulus and sometimes in the intermediate stimulus. It is probable, however, that a change in the depth of this minimum contributes little to a listener's identification or discrimination of the stimuli (Malme, 1959).

The interaction of the pole-zero-pole combination of $F1$, FNZ, and FNP gives rise to a broad low-frequency spectral prominence with different shapes depending on the vowel. In the case of the back vowels, $F2$ also contributes toward shaping this low-frequency prominence. For example, in the [o–õ] continuum, as the spacing between FNP and FNZ increases, the amplitude of the spectral peak corresponding to $F1$ decreases while the amplitude of the $F2$ peak increases. This change in balance between the two peaks can lead to an increase in the frequency of the center of gravity of the low-frequency spectral prominence relative to that for the non-nasal stimulus (Chistovich and Lublinskaya, 1979; Chistovich *et al.*, 1979). A possible consequence is a shift to a more open vowel quality, as will be discussed later when we examine the results. A similar but smaller effect can be expected for the [e–ẽ] continuum, for which there is also a symmetrically spaced pole-zero-pole combination in the $F1$ region. In the case of the non-nasal–nasal continua for the vowels [i ɑ u], the frequency at which the additional pole-zero pair is introduced (400 Hz) early in each continuum is well separated from the original first-formant frequency. For these three vowels, if the additional spectral peak due to FNP causes a shift in the low-frequency center of gravity, this effect is expected to be small. It would be in the direction of a higher effective $F1$ or more open vowel quality for the high vowels [i] and [u], and a lower effective $F1$ or more close quality for the low vowel [ɑ]. It should be noted that, if the spacing of the pole-zero pair were increased beyond certain critical limits the quality of the vowel is likely to change dramatically. For example in the [ɑ–ɑ̃] continuum, vowels with a wider pole-zero spacing than that of stimulus 9 begin to sound like [ɜ].

In summary, then, there are several acoustic attributes that change as the pole-zero pair is manipulated in the various stimulus continua; these have potential consequences in the way the stimuli are identified or discriminated by listeners. The shape of the low-frequency spectral prominence is modified by the addition of a subsidiary peak that is above or below the frequency of the $F1$ peak or that splits the $F1$ peak into two nearby peaks of about the same amplitude. There may be a shift in the perceived vowel height, and there may be a shift in overall spectral balance of high-frequency amplitude in relation to low-frequency amplitude that could be interpreted as a change in voice quality (of the type that arises from laryngeal manipulations). These changes in the stimuli are automatic consequences of inserting the additional pole-zero pair, and as such they cannot be manipulated independently in a speech synthesizer without introducing other side effects into the stimulus characteristics.

### B. Procedure

An identification test was prepared for each vowel continuum. Each stimulus was repeated nine times (eight times for ([u–ũ]) and order of presentation was counterbalanced so that a given stimulus item followed each other stimulus item (including itself) exactly once. Otherwise, the ordering of the stimuli was random. With this construction of the tests, we were able to counterbalance any effect the previous stimulus might have on the identification of any particular stimulus item. An item consisted of hearing a particular stimulus twice, followed by a 4-s-pause during which the subject identified the vowel in the syllable as either nasal or non-nasal by marking the appropriate column on a sheet of paper. A practice test consisting of six items from each continuum (including the extreme stimuli and some intermediate stimuli) was heard before the test proper. Rests were given as requested between separate vowel tests, but not within a test. Stimuli were presented free field in a sound-treated listening room. The subjects, who were tested in groups of from one to five, had no difficulty understanding what was required of them after non-nasal and nasal vowels were demonstrated for them by the experimenter—in minimal pairs in their own language for Indian subjects, and as isolated vowels or preceded by [t̪] for American subjects.

Table II lists the native languages and the number of subjects for this experiment and the later discrimination experiment. There were native speakers of Gujarati, Hindi,

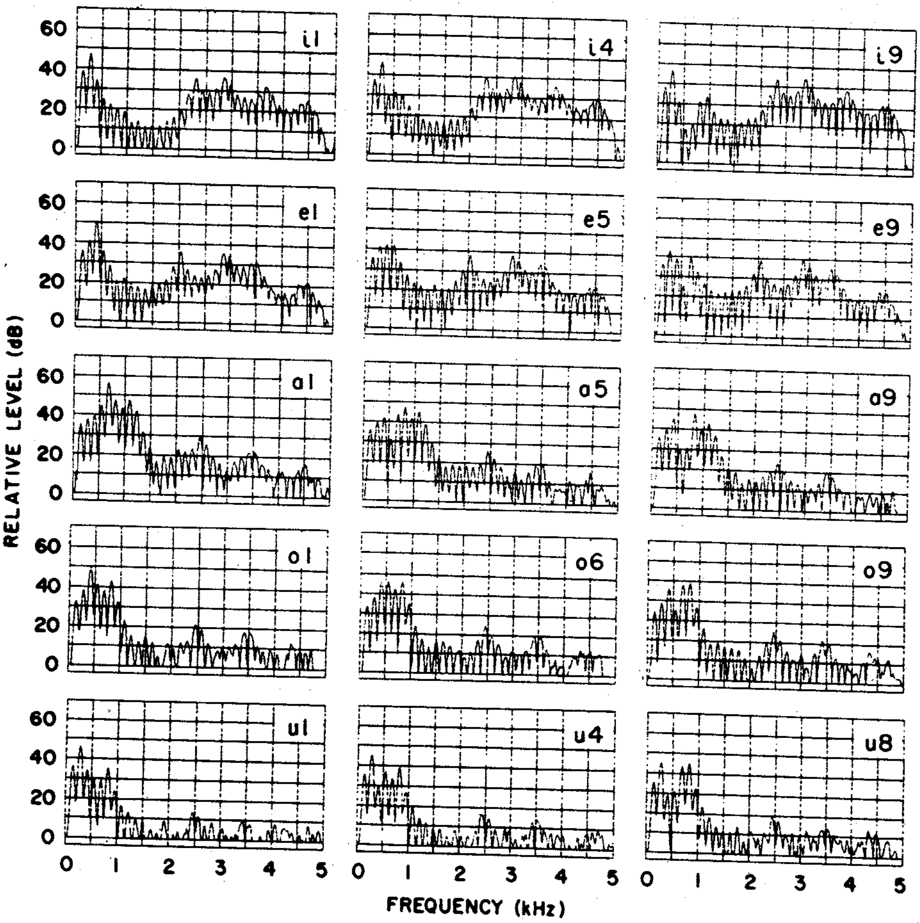

FIG. 7. Spectra of a number of the stimuli from the continua used in identification tests. Each row gives examples of spectra from one of the five vowels as shown. The left panel in a row represents the stimulus at the non-nasal end of the continuum, and the right panel represents the most nasal stimulus. The middle panel is the spectrum of the intermediate stimulus that is closest to the average 50% crossover point in the identification function for that vowel. The stimulus number is indicated in each panel. Spectra are discrete Fourier transforms calculated from the waveform weighted with a Hamming window of duration 26 ms.

Bengali, and English. The Indian subjects were faculty and students at universities in the Boston area, or their spouses, all of whom were highly educated and held or were qualified to hold positions in the professions. The "naive English speakers" were students at MIT who were untrained in phonetics, unused to hearing synthetic speech, and knew no language other than English. Except for their monolingualism, these subjects were similar to the Indian subjects in their linguistic sophistication. The "non-naive English speakers" were trained phoneticians and/or members of the Speech Communication Group at MIT, whose native language was English. They were included to help assess the effect of experience or training on perception of nasal vowels.

English is considered not to possess a non-nasal–nasal opposition for vowels, although contextually determined vowel nasalization is strong in some dialects of American English. Indeed, speakers from the southeastern United States may produce a sequence of vowel and nasal consonant as a nasal vowel (without a nasal murmur) as in ['mɜ̃ki] for *monkey*. Malécot (1960) has argued that all American listeners need to distinguish nasal from non-nasal vowels in some phonetic environments, as in the pair *cat* versus *can't*. The occurrence of such minimal pairs is much less systematic and widespread in American English than in the Indian languages involved in these experiments, however. None of our "naive English speakers" spoke a dialect employing strong vowel nasalization, but two of our "non-naive English speakers" did; their experimental responses fell solidly in the middle of those of the rest of their group.

Hindi and Bengali, like Gujarati, possess the nasal–non-nasal opposition for vowels. Hindi vowels are relatively similar to those of Gujarati in quality and distributional characteristics. The Hindi speakers' responses were thus expected to be similar to those of the Gujarati speakers. However, there is some question as to the status of the non-nasal–nasal distinction in Bengali (see, for example, Ferguson and Chowdhury, 1960). Although usually cited as having an equal number (seven) of oral and nasal vowels, the oral vowels are much more frequent in standard colloquial Bengali than are the nasal vowels. Ferguson and Chowdhury state that the degree of nasality "is relatively weak and at times may even be a kind of breathiness rather than nasality in the strict sense" (1960, p. 37). In addition to these factors clouding the status of distinctively nasal vowels in Bengali, there appear to be many dialects that are even less consistent than standard colloquial Bengali in their separation of oral and nasal vowels. These factors could cause the Bengali speakers to respond to identification and discrimination tests differently from Gujarati and Hindi speakers, although not necessarily in the same way as American listeners.

TABLE II. Native languages of subjects used in identification and discrimination experiments, and number of subjects tested in each group.

| Language group | Number |
|---|---|
| Gujarati | 10 |
| Hindi | 5 |
| Bengali | 5 |
| Naive American English | 10 |
| Non-naive American English | 10 |

The characteristics of the five groups of subjects made it reasonable to analyze the data of this and subsequent experiments in two ways: comparing both American groups with the Gujarati and Hindi groups (excluding the Bengalis due to the questionable status of vowel nasality in their language), and comparing all five language groups separately.

## C. Results

Figure 8 shows identification functions for all five groups of listeners for all the vowels. All of the listener groups identified the nasal end of the continuum with essentially unanimous responses, except for [ĩ], for which the Bengali and Hindi listeners gave less than 90% "nasal" responses.

Regression lines were calculated on the two points above and the two points below the 50% crossover of the identification function for each subject, that is, on the most linear portions of the functions. Slopes of the regression lines were calculated on these same portions and were used to determine precise 50% crossover points. In this and all subsequent statistical analyses, the data for [u–ũ] were transformed to make them comparable to the other (nine stimulus) continua. Figures are plotted with the untransformed data.

The 50% response points for each subject were subjected to a repeated-measures analysis of variance with two language families (American English versus Gujarati and Hindi) × 5 vowels × 20 and 15 subjects per group. The only significant effect was for vowels [$F(4,132) = 31.82$, $p \ll 0.0001$], with the difference between the American and Indian listeners falling far short of significance [$F(1,33) = 0.28, p < 0.6$]. The difference across vowels in the 50% crossover point was expected, since the effect of the additional pole-zero pair began to be perceptible at different points in each stimulus continuum, depending largely upon the step size between stimuli.

We carried out a second analysis of variance that was identical to the first except that the two-level contrast between Indians and Americans was replaced by a five-level contrast between each language group separately. (Hence there were 5 languages × 5 vowels × 10 and 5 subjects per group.) The vowel × language group interaction was significant [$F(16,140) = 2.32, p < 0.005$]. Although there were no consistent trends that characterized one group as being different from another among vowels, the Bengalis tended to be more variable among vowels in their crossover points relative to the other language groups. This observation is borne out in other tests not reported here. The main effect of vowels was again highly significant [$F(4,140) = 33.15$, $p \ll 0.0001$], and there was again no significant main effect of language group [$F(4,35) = 1.01, p < 0.4$]. As Fig. 8 shows, the spread between language groups' crossovers was largest for the high vowels [i] and [u] and smallest for [ɑ] and [o]. Tests for simple main effects of language within the vowels [i], [e], and [u] separately were not significant, indicating that the significant vowel × language interaction arose from differences between vowels and small changes in the order of languages within each vowel. These analyses confirm, then, that the 50% response points were quite similar for all language groups.

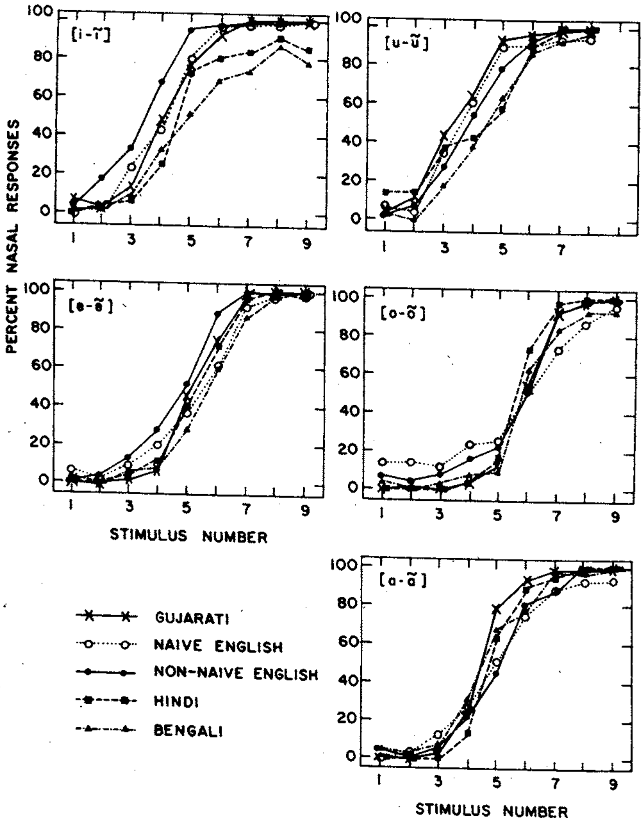

FIG. 8. Average identification functions for each of the stimulus continua for the five groups of listeners as indicated. The number of subjects in each group is given in Table II.

The sharpness of the identification boundary and the consistency of responses within a group is reflected in the slope of the regression line for each identification function. We would expect the identification functions for subjects whose language includes the non-nasal–nasal vowel distinction to have steeper slopes than those for subjects whose language lacks the distinction. A repeated-measures analysis of variance on 2 language families × 5 vowels × 20 and 15 subjects per group confirmed this expectation. Neither the interaction between vowels and language family nor the main effect of vowels was significant, but the Americans' slopes were significantly less steep than the Indians' (Gujarati and Hindi speakers). [For the interaction, $F(4,132) = 0.79, p < 0.5$; for vowels, $F(4,132) = 0.71, p < 0.6$; and for language family $F(1,33) = 10.03, p < 0.003$.] The second analysis of variance, on all five language groups, corroborated these findings. There was no significant interaction between vowel and language group [$F(16,140) = 1.10$,

$p < 0.4$] nor did the slopes of the vowels differ among themselves [$F(4,140) = 1.80$, $p < 0.1$]. The language groups did differ among themselves however [$F(4,35) = 3.09$, $p < 0.03$]. Protected $t$-tests (Winer, 1971, *p*. 199) between pairs of individual language groups indicated that the only significant difference was between the two extremes, the Gujarati and the naive American groups [$t(35) = 1.94$, $p < 0.05$, one-tailed]. However, the mean steepness for the Hindi group (32.4 % per stimulus step) was closer to that of the Gujaratis (33.4) than it was to any other group, while the Bengali group (28.9) was similar to the naive American group (28.5) and the American phoneticians (30.1).

These analyses of the 50% crossover points and the slopes of the functions together suggest that experience with the non-nasal–nasal vowel distinction can affect the sharpness of the identification boundary and the consistency of responses between listeners, but does not influence the placement of the boundary between non-nasal and nasal categories.

As already noted, the variation in crossover point from one vowel to another is not unexpected in view of the different step sizes in FNP and FNZ for the different continua. The magnitude of the frequency difference FNP–FNZ for a given stimulus does, however, provide a measure of the difference between the spectrum of that stimulus and the spectrum of the stimulus at the non-nasal end of the continuum. In Fig. 9 we show the average crossover point for each vowel, expressed in terms of the pole-zero spacing. Also displayed are the ranges of crossover points across the five language groups and the pole-zero spacings for the endpoint stimuli. The average pole-zero spacing at crossover is in the range 75–110 Hz for the different vowels. The spacing for the vowel [i] at crossover is somewhat greater than that for the other vowels. Further indication of the amount by which the spectrum must be modified to give rise to a nasal response is given in Fig. 7, in which the "intermediate" spectrum is the spectrum of the vowel stimulus located closest to the average crossover point.

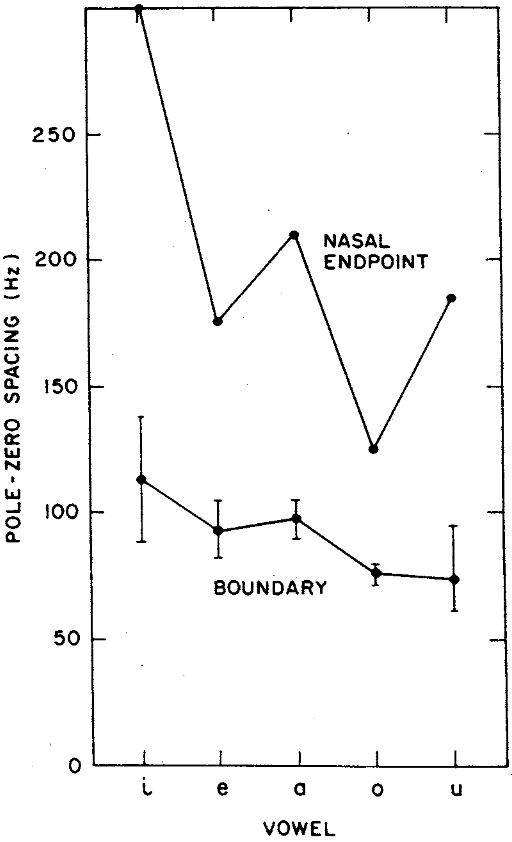

FIG. 9. Pole-zero spacing corresponding to the endpoint stimulus (top curve) and to the average 50% crossover point (bottom curve) for each continuum, as obtained from the identification experiment. The vertical bars on the bottom curve represent the range of crossover points over the five language groups.

## D. Stability of the identification boundary

A number of factors other than experience may affect the subjects' placement of the identification boundary. Lack of stability of the boundary could be caused by two types of contextual effects: an immediate context effect and a range effect. In the former case, the categorization of a given stimulus may be influenced by the quality of the stimuli preceding it. So, for example, a vowel of moderate nasality may be classified as non-nasal when it is immediately preceded by one or more strongly nasal vowels, but as nasal when it is preceded by strongly non-nasal vowels. This effect will add noise to the classification of intermediate stimuli on a continuum. The range effect refers to the tendency for subjects to place a binary category boundary approximately in the middle of a continuum of stimuli (Parducci, 1974; Rosen, 1979). Extending the range of the continuum at only one end will tend to shift the category boundary toward that end, whereas truncating one end of the continuum will shift the boundary in the opposite direction. Certain types of tasks, including those in which stimuli are presented in random order as in these experiments, are particularly prone to this type of range effect (Macmillan, in press).

The influence of *immediate context* was assessed by determining, for each stimulus, the number of "nasal" responses when directly preceded by stimulus 1 or 2, the most non-nasal stimuli, compared with when it was directly preceded by stimulus 8 or 9, the most nasal stimuli, for the Gujarati and naive American groups only.

A repeated-measures analysis of variance (2 language groups × 2 contexts × 5 vowels × 9 stimuli) was performed on the total number of "nasal" responses to each stimulus preceded by each (non-nasal or nasal) context. There was the predicted shift towards more "non-nasal" responses after a strongly nasal vowel, and vice versa [$F(1,14) = 6.58$, $p < 0.02$], but the size of the shift was not the same for all vowels, as illustrated in Fig. 10, which shows the interaction between vowel and immediate context [$F(4,56) = 3.32$, $p < 0.02$]. The identification of front vowels was more susceptible to the nasality of the immediately preceding stimulus than was the identification of back vowels. This shift in boundary becomes increasingly small through [i e ɑ o u], with no shift at all for [u]. Inspection of the data suggests that the shift was due in large part to the naive Americans rather than to the Gujaratis, but the difference between the two language groups was insufficient to achieve significance either in the main effect of language group or in any of its interactions.

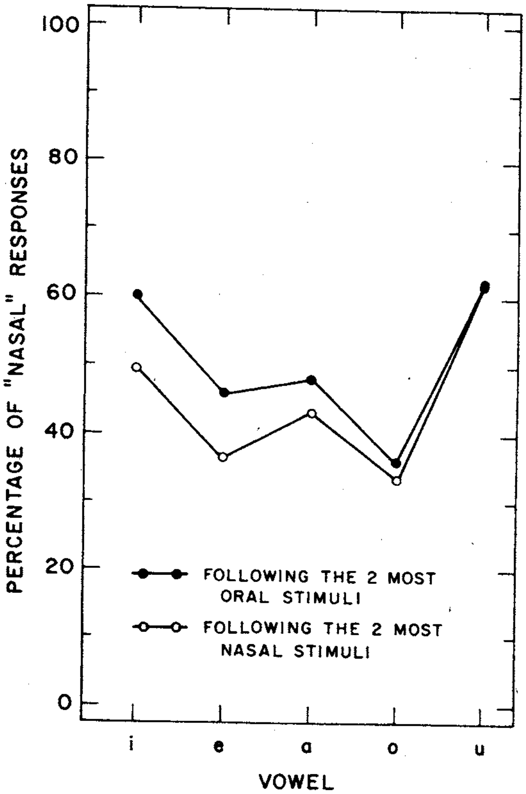

FIG. 10. Influence of the nasality of an immediately preceding vowel on the identification of a vowel as nasal or non-nasal. ●: "nasal" responses to stimuli which immediately followed either of the two most oral stimuli in the identification experiment (stimuli 1 or 2). O: "nasal" responses to stimuli which immediately followed either of the two most nasal stimuli in the experiment (stimuli 8 and 9, or 7 and 8 for [u–ũ]). The interaction between vowels and nasality of the preceding stimulus was significant in an analysis of variance (see text). Subjects were 10 Gujaratis and 10 naive Americans.

The presence and size of a *range effect* was assessed by repeating the identification experiment with all instances of the two most nasal stimuli excised from the tape. If stimulus range influenced the subjects' judgments, the identification boundary should shift one stimulus towards the non-nasal end in this truncated continuum. Five of the naive American subjects and five of the Gujaratis served in both experiments.

Figure 11 shows the crossover points for each vowel, with two language groups pooled. The shift was generally in the expected direction—towards the lower numbered stimuli—but its size varied among vowels. The boundary shifted by less than one step for all continua except [o–õ]. A repeated-measures analysis of variance [2 language groups × 2 continua (truncated or complete) × 5 vowels × 5 subjects in each group] confirmed these observations, since the interaction of vowels with continuum, shown in Fig. 11, was significant $[F(4,32) = 6.03, p < 0.001]$, while the main effect of continuum was not $[F(1,8) = 2.86, p < 0.1]$. One-tailed protected $t$ tests proved only the shift for [o–õ] to be significant $[t(8) = 2.53, p < 0.025]$, although that for [ɑ–ɑ̃] was also relatively large.

There was no overall difference between the two language groups in the extent of the range effect $[F(1,8) = 0.36, p < 0.57]$, but there were differences between language groups dependent upon the vowel $[F(4,32) = 3.09, p < 0.03]$. No interpretation is placed on the significance of this interaction, since it depends on a difference among vowels in the Gujaratis' and Americans' responses to the two continua, and this difference did not itself achieve significance.

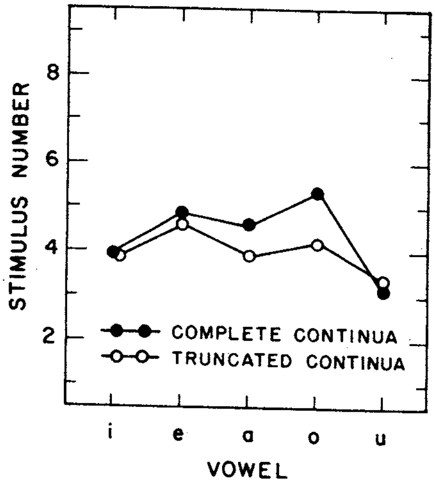

FIG. 11. Effect of the range of the stimulus continua on judgments of nasality. Fifty-percent crossover points in the identification functions are shown for the complete continua (●), with nine stimuli per vowel continuum (eight for [u–ũ]), and for continua truncated by removing the two most nasal stimuli (O). The interaction of vowel with complete or truncated continuum was significant in an analysis of variance (see text). Subjects were 5 Gujaratis and 5 naive Americans.

In summary, the data were subject to both immediate context and range effects, but the effects tended to be small, and differences between language groups were not statistically significant. The two effects appear to be independent of each other in that the nasality of the immediately preceding stimulus affected identification of front vowels but not back vowels, while the range of the stimulus continuum only significantly affected [o–õ], and did not affect front vowels at all. Conclusions concerning the range effect are necessarily tentative, since there were only five subjects in each group.

## IV. DISCRIMINATION TEST

Having established that listeners from different language backgrounds apparently base their identification of the non-nasal–nasal distinction on similar criteria, we turned next to an examination of the behavior of the listeners when they are asked to discriminate between stimuli on the continua. In particular, we sought answers to two questions. (1) Given that the items on any one of these stimulus continua are equally spaced, at least on one set of physical measures, do listeners discriminate between stimuli better over one part of a continuum than another? More specifically, are the stimuli that are identified consistently as being within the non-nasal or the nasal region less well discriminated than those located near the identification boundary? Such a result could indicate a nonlinear relation between the simple physical parameters describing the stimulus and the auditory representation of the stimulus. (2) Is the discrimination behavior different for the different language groups? A difference between language groups would suggest that linguistic experience influences those attributes of the stimuli the listeners are able to attend to when they make a discrimination. Results of Beddor and Strange (1982) suggest that this could be the case.

To answer these questions, we prepared for each vowel series a 4IAX discrimination test for pairs consisting of stimuli that are one and two steps apart on the continuum. Pairs were randomized, and each pair was repeated a total of eight times. The interval between stimuli in a pair was 0.26 s, and between pairs it was 0.42 s. The response time between trials was 2.3 s. The subjects were the same as those used in the identification tests. The discrimination test was always done after the identification test, with at least one week between the two. Stimuli were presented over high quality headphones in the same sound-treated listening room as for the identification experiment. Subjects were tested in groups of from one to four, with rests given as necessary, but not within individual vowel tests.

The results for two-step discrimination are summarized in Fig. 12. For clarity of presentation, and consistent with the analysis in terms of two language families, the figure shows discrimination functions for the two American groups together, and for the Gujarati and Hindi groups together. The Bengali data are omitted from Fig. 12, but they were roughly similar to those of the Gujarati and Hindi groups, although a little more noisy. The ticks on the curves represent mean identification boundaries for these groups from 50% crossover points determined from the data in Fig. 8.

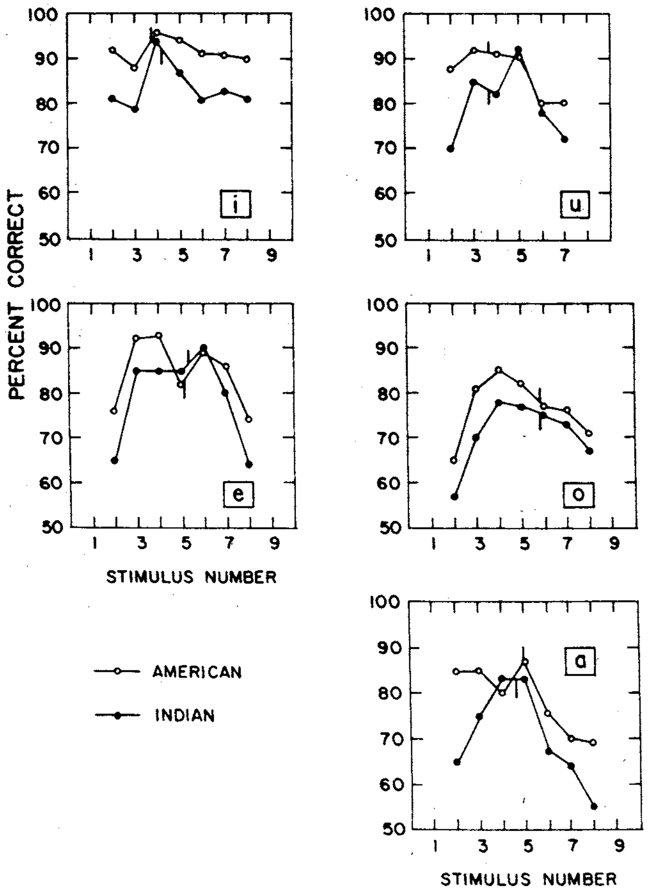

FIG. 12. Average two-step discrimination functions for two groups of listeners. Data from the Gujarati and Hindi listeners were combined to form one set of functions (●), and data from the naive and non-naive American groups were combined to form the other set (○). The short vertical lines in each function indicate the 50% crossover boundaries of the corresponding identification functions.

A repeated-measures analysis of variance was conducted on these data, with 2 language families ×5 vowels ×7 pairs of stimuli and 20 and 15 subjects within groups. [The analyses of variance using five and four language groups, i.e., with and without the Bengalis, gave essentially the same results as the analysis in terms of two language families. Inclusion of the Bengalis tended to increase the significance of terms, but did not change terms from significant to nonsignificant, or vice versa. The only major difference between the analyses in terms of separate language groups compared with two language families was that the pair × group interaction was not significant in the former, whereas the (equivalent) pair × language family interaction was significant in the latter.]

In the analysis using two language families, there were significant main effects of language family $[F(1,33) = 4.05, p<0.05]$, of vowel $[F(4,132) = 24.63, p<0.0001]$, and of stimulus pair $[F(6,198) = 23.84, p<0.0001]$, as well as significant interactions between stimulus pair and language family $[F(6,198) = 2.72, p<0.01]$, and between vowel and stimulus pair $[F(24,792) = 4.20, p<0.0001]$. The stimulus pair effect reflects a general tendency for discrimination to be worse at the extremes of the continua. To clarify the interactions, simple main effects of stimulus pair within each vowel were calculated for American and Indian groups separately. They were all significant at $p<0.01$ or better, except for the Americans' [i–ĩ] and [u–ũ] continua, which did not contain significant differences. The Indian subjects consistently discriminated pairs in the vicinity of the phonetic boundary more accurately than pairs from the extremes of each continuum—that is, more accurately than pairs clearly within the non-nasal or the nasal categories. The American subjects showed the same pattern for the midvowels, [e–ẽ] and [o–õ], but differed from the Indians for [i–ĩ], [ɑ–ɑ̃], and [u–ũ]. The Americans' consistently good discrimination in the [i–ĩ] continuum may be attributable to a ceiling effect that could obscure a pattern similar to the Indians'. Their discrimination of [ɑ–ɑ̃] and [u–ũ] contrasted sharply with the Indians' pattern, however. Pairs within the non-nasal categories of [ɑ–ɑ̃] and [u–ũ] were discriminated better than those within the nasal category, and there was little difference in accuracy of discrimination between non-nasal pairs and pairs spanning the phonetic boundary. That Beddor and Strange (1982), using a completely different method of synthesis, found this same difference between American and Hindi listeners' discrimination of their [ɑ–ɑ̃] continuum, suggests that the pattern of discrimination for [ɑ–ɑ̃] and [u–ũ] deserves some attention. As in Beddor and Strange's data, Dunn (Bonferroni) tests [Keppel (1982), p. 146 ff.] proved that only pair 1,3 (the most non-nasal) in the [ɑ–ɑ̃] and [u–ũ] continua differed significantly between the Indian and American listeners ($p<0.01$ for both continua). However, these observations on differences in the overall shapes of the discrimination functions are confirmed by tests for linear and quadratic trends for the language families and vowels separately.

These analyses confirm, then, that the American and Indian listeners differed in some respects in their discrimination between pairs of stimuli. Apparently listeners may use different strategies for discrimination depending upon the particular vowel and upon whether their native language includes the non-nasal–nasal opposition for vowels.

## V. DISCUSSION

### A. An acoustic correlate of the feature [ + nasal]?

Our experiments confirm that reliable percepts of nasal vowels can result when a pole-zero pair is added to the spectrum of an oral vowel in the vicinity of the first formant. Informal comments by speakers of Gujarati and by American researchers in speech indicated that the majority of our vowels sounded very natural. The nasal vowels whose naturalness was least satisfactory tended to be the high vowels [ĩ] and [ũ]. It is possible that the quality of the synthesized nasal vowels could be improved further by other changes, such as shifts in the frequencies or amplitudes of the higher formants or the introduction of additional pole-zero pairs at higher frequencies or in the vicinity of $F1$.

The identification functions for the different non-nasal–nasal continua show a substantial degree of agreement for the groups of listeners with different language backgrounds. The agreement is particularly notable in view of the fact that some listener groups (particularly the Americans) are not normally considered to use the nasal feature distinctively for vowels in their language, whereas others do have phonemic nasal vowels. One is led to conclude that, whatever acoustic property or properties give rise to the identification responses, all listeners use roughly the same properties irrespective of whether their native language includes a phonemic non-nasal–nasal vowel distinction (cf. Wright, 1980).

The question of whether there is a common acoustic property, independent of the vowel, that gives rise to identification of the feature [ + nasal] is more difficult to answer. The fact that listeners with different language experience show similar crossover points in their identification functions for each vowel could lead one to conclude that they are responding to the same acoustic property for each vowel. A stronger hypothesis would be that there is a common acoustic property for all nasal vowels, and hence a common attribute in the pattern of auditory response. If a different property were to signal the presence of the feature [ + nasal] for each vowel, it is unlikely that listeners would respond in essentially the same way to the different vowels independent of whether their language contained the non-nasal–nasal distinction.

Can we identify such a common property for all nasal vowels? Or, equivalently, can we postulate an auditory processing mechanism that shows a pattern of response with a common attribute for any vowel that is identified as [ + nasal]? One way of describing the vowels in the present study is in terms of the spacing between the additional pole and zero. In Fig. 9 we have shown that the pole-zero spacing corresponding to the 50% crossover points does not show a great deal of variation from one vowel to another (75 to 110 Hz). Associated with a given pole-zero spacing there is a particular maximum deviation (in dB) of the spectrum from the shape for the non-nasal vowel. It is to be expected that a given maximum deviation relative to the non-nasal spectrum would be perceptually more salient if it occurred in a frequency region near a spectral peak. In fact, there are perceptual data that support this expectation (Klatt, 1982). Figures 6 and 7 show that at the crossover point the pole-zero pair is in a frequency region that is somewhat removed from a spectral peak for the vowel [i], and, to some extent, for [ɑ], but is closer to the $F1$ spectral peak for the remaining vowels. Consequently one might expect that the pole-zero spacing required to elicit a [ + nasal] response would be greater for [i] (and possibly for [ɑ]) than for the other vowels. This trend can be seen in the data in Fig. 9, both for the stimuli at the crossover and, more strongly, for the endpoint stimuli. Thus at least one measure that seems to relate closely to the listeners' responses is the maximum deviation of the spectrum from the original non-nasal spectrum in the $F1$ region, with some weighting applied depending on the frequency location of this deviation in relation to the original spectral peaks and valleys.

Although a measure of this type may describe the data well, we do not consider it particularly satisfactory as a hypothesis about the perception of nasal vowels. A major objection is that it requires the listener to compare the heard spectrum of a potential nasal vowel with the memory of the spectrum of a non-nasal vowel. Even though nasal vowels are linguistically marked, it seems unlikely that the listener uses such a cumbersome procedure to identify nasality. We therefore search for alternatives.

Another way of describing the stimuli is in terms of the degree of low-frequency prominence in the spectrum. (Maeda, 1982b, has attempted to define such a measure.) As we introduce the pole-zero pair with an increased spacing, we are reducing the degree of prominence of the $F1$ peak in the spectrum, so that a single narrow spectral peak no longer dominates the low-frequency range. This reduced prominence is achieved by creating an additional spectral peak near $F1$ or by splitting or broadening the $F1$ peak.

We do not know how the pattern of auditory response changes as the degree of prominence of a spectral peak is manipulated. It is likely that the physiological processes involved in making a phonetic distinction occur over several stages in the auditory pathway. Something *is* known, however, about responses of the auditory nerve to vowel-like stimuli. We might ask whether the pattern of auditory-nerve responses could begin to reflect a grouping into two classes corresponding to nasal and non-nasal vowels. Such a tendency would be provocative, but clearly could not provide the entire basis on which the phonetic distinction is made. In this spirit, we consider a possible pattern of auditory-nerve responses for the feature [ + nasal]. Specifically, it may be appropriate to interpret the concept of prominence with reference to the synchrony of firings of primary auditory neurons rather than to attributes of the spectrum—an approach consistent with some theories of pitch perception (cf. Srulovicz and Goldstein, 1983).

It is known that a stimulus with a prominent low-frequency spectral peak leads to responses of auditory neurons that tend to be synchronous to the frequency of this peak (Sachs and Young, 1980; Delgutte and Kiang, 1984). This synchrony extends over auditory neurons that cover a range of characteristic frequencies in the vicinity of the stimulus frequency if the spectral prominence is sufficiently narrow and well separated from other prominences. If this low-frequency spectral peak is broadened, or if an additional spectral peak is introduced nearby, there will be a reduction in the range of characteristic frequencies over which the auditory neurons fire synchronously with the $F1$ prominence. Introduction of a pole-zero pair with a gradual increase in spacing will monotonically increase the amplitude of the additional spectral peak. It is to be expected, then, that at a particular spacing the additional peak will be of sufficient amplitude that it will extinguish the synchrony of firings to $F1$ for auditory neurons in its frequency range. That is, there will be an abrupt reduction in the range of neurons that respond synchronously to $F1$.

As we have seen, the pole-zero spacing necessary to accomplish the change in identification response that may follow a change in auditory response is in the range 75–110 Hz. This critical spacing is somewhat greater for the vowel [i] than for the other vowels presumably because, for a given pole-zero spacing, the FNP peak due to the additional pole is weaker for this vowel than for the others, as discussed above. Consequently a greater pole-zero spacing is needed for [i] if the FNP peak is to eliminate the synchrony to $F1$.

An alternative way of achieving a reduction of the $F1$ prominence similar to that produced by the additional pole-zero pair would be to broaden the bandwidth of the first formant, without adding a pole-zero pair. There are indeed data to indicate that broadening $F1$ leads to the perception of nasality (Delattre, 1968; Hawkins and Stevens, 1983). In natural speech, the bandwidth of $F1$ is almost certainly increased by the additional acoustic losses introduced by coupling to the nasal cavity; this would contribute additional reduction of prominence over that resulting from the pole-zero pair.

In summary, then, there may be some basis for supposing that the pattern of response at the level of the auditory nerve is such that a rather simple detection procedure could be used to identify a vowel as a member of the class [ + nasal]. Further studies of auditory responses to these types of sounds are clearly needed, however, to support this assertion and to provide a more quantitative specification of the acoustic correlate of this feature.

### B. Language- and vowel-dependent effects

While there is some evidence that listeners respond in a distinctive way to a simple acoustic property of nasality in the identification tests, there are aspects of the identification and discrimination data suggesting that other factors are influencing the responses. We turn now to an examination of these factors.

A consistent difference between language groups was that both American groups showed a pattern of discrimination that differed from that of the Indian groups. That is, the Americans discriminated better between stimuli at the non-nasal end of the continua for [ɑ u] (and possibly for [i]) than at the nasal end. There were also differences in the way listeners responded to particular vowel continua. For example, there appeared to be some differences between vowels in the extent to which they were subject to range effects: the placement of the identification boundary was particularly dependent on the range of nasality for the vowel [o–õ], and particularly insensitive to the range of nasality for [u–ũ]. Also, the effect of immediate context was greatest for the front vowels and was small or nonexistent for the back vowels.

In order to interpret these findings, we propose that the basic nasality property is accompanied by one or more additional acoustic properties, perhaps different for different vowels, that change within the various stimulus continua, and can influence the reponses of the listeners. Among the most important of these additional properties, we suggest, are changes in vowel quality, caused by shifts in the center of gravity of low-frequency spectral prominences, and changes in overall spectral balance caused by the presence of the pole-zero pair (as discussed in connection with Fig. 7).

We consider first a possible explanation for the influence of the preceding stimulus on the identification of nasality for the front vowels (Fig. 10). The principal acoustic attribute distinguishing front vowels from back vowels is the relatively large amplitude of the spectrum at high frequencies, in the range of $F2$ and $F3$. Thus for the front vowels, we might expect that the changes in high-frequency spectral amplitude that occur through the continua are perceptually more salient than they are for the back vowels. This change in spectral balance could lead to the perception of a change in voice quality that listeners might associate with a change in nasality, particularly for ambiguous stimuli. Perhaps listeners' assessment of these aspects of voice quality is more prone to contextual effects than is their assessment of the primary property of nasality.

Another example of the influence of changes in vowel quality on the perception of nasality is in the range effect, which was largest for the continuum [o–õ] for both groups of subjects. Several phoneticians judged that there was a distinct change in quality of the vowel at or near stimulus 4, where [o] became [ɔ] (a finding that is consistent with the description of this stimulus continuum in connection with Fig. 7). Apparently, in the absence of any unambiguously nasal vowels in the truncated continuum, subjects listening to stimuli in this set adjusted their criterion to one of vowel quality. That is, the listeners seemed to be willing to make judgments of vowel nasality based largely upon differences in perceived vowel height. In the nine-stimulus continuum, which included unambiguously nasal stimuli, they presumably heard the shift in vowel height, but placed more weight on the percept associated with the wider pole-zero-pole separation around stimuli 5 and 6. Further support for the idea that vowel height can influence judgments of nasality is found in the data of Lonchamp (1978) for French.

The performance of listeners in the discrimination experiment is also consistent with the hypothesis that listeners' responses are determined by additional secondary properties as well as by a primary property of nasality. In general, the Indian and American listeners give similar discrimination functions for the vowels [e] and [o]. Both groups show reduced discrimination at the two ends of the continuum, with a rather narrow peak for [o] and a broader plateau or peak in the middle range for [e]. For these midheight vowels, the introduction of the pole-zero pair results in a splitting of the $F1$ prominence. When the two component peaks are close together for the first three or four stimuli in the continuum, discrimination between stimuli is relatively poor. Discrimination approaches a maximum when the spacing between these peaks is about 100 Hz for [e] (stimulus 3-4) and about 80 Hz for [o] (stimulus 4-5). These stimulus ranges do not, however, correspond to the boundaries of the untruncated non-nasal–nasal identification functions. Apparently the listeners are distinguishing here a change in vowel quality rather than a change in nasality. We note, however, that the identification boundaries for the truncated [o–õ] continuum are in the range of the discrimination peak (near stimulus 4, from Fig. 12). This correspondence provides some support for our hypothesis that, with the truncated [o–õ] continuum, the listeners are basing their identification judgments more on vowel quality than on a nasality property.

For the high vowels [i u] and the low vowel [ɑ], the introduction of the pole-zero pair is at a frequency that is well separated from $F1$: above $F1$ for [i] and [u] and below $F1$ for [ɑ]. The discrimination functions for these vowels have different shapes for the American and Indian listeners. Both groups show a decrease in accuracy of discrimination as the nasal end of the continuum is approached (although this trend does not apply to [i–ĩ] for the Americans). This decrease in discriminability may be a consequence of the fact that, for a two-step change, the change in the maximum amplitude deviation in the spectrum, relative to that for a non-nasal vowel, becomes smaller as the pole-zero separation becomes larger. The Indian listeners, but not the Americans, also show a decrease in accuracy of discrimination for stimuli at the non-nasal end of the continuum for these three vowels, with a peak in the middle range. One can surmise that the experience of the Indian listeners with the non-nasal–nasal distinction orients them to focus on the primary nasality property. This conclusion is supported by the fact that their discrimination peaks tend to be close to their identification boundaries for these vowels. The American listeners, on the other hand, tend not to show reduced discrimination for the stimuli at the non-nasal end of the continuum for these three vowels. The Americans, not possessing a ready way of categorizing these vowels in their language, may be more analytical in their judgments of the stimuli, and may focus on detailed attributes such as a change in amplitude at high frequencies, or shifts in the frequencies or amplitudes of particular spectral peaks. This strategy could result in improved discrimination at the lower end of the continuum.

### C. Implications and further questions

The hypothesis that there is a primary nasality property and accompanying secondary properties raises some additional questions. For example, the shift in vowel height with nasalization, observed for some vowels in this experiment, is a secondary property that is found in some languages. This change in vowel height occurs in only a minority of languages, but when it does occur the tendency is for high vowels to lower and low vowels to raise (Beddor, 1983). Shifts in height for the midheight vowels are less consistent, but there is a tendency for these vowels to lower when nasalized. Why is it that some languages seem to introduce additional properties, such as changes in vowel quality, to complement or enhance the basic property of nasality? We have noted in Fig. 5 that the spectral modifications associated with the addition of a pole-zero pair in the vicinity of the first formant tend to produce a low-frequency pole-zero-pole combination that varies very little from one vowel to another. The perceptual result is a loss of discriminability between nasal vowels along the high–low dimension, a situation favoring the development of additional enhancing properties. This loss of discriminability for nasal vowels has been verified experimentally by Wright (1980).

Languages vary in how they deal with this reduction in discriminability. Some languages have the same number of nasal as non-nasal vowels, with no reported differences in quality between the two sets. In a substantial minority of languages that contrast nasal and non-nasal vowels, there is a reduced number of nasal vowels (Ferguson, 1963, 1975; Ruhlen, 1975). Most commonly it is the midvowels that are missing in these imbalanced systems (Wright, 1980). The problem of reduced discriminability is thereby avoided in that only those vowels with the most distinctive values of $F1$ are retained.

A final question is concerned with the physiological and psychophysical basis for the perception of the primary nasal property itself. On the basis of physiological data, we have speculated that there is a distinctive change in the auditory response resulting from modification of the spectral prominence in the vicinity of the first formant. Further physiological data are needed, however, for stimuli with spectral prominences of the type found in nasal vowels. Of relevance in the psychophysical domain is the work of Chistovich *et al.*, (1979). In a task in which one-formant vowels were matched against two-formant vowels with various spacings, they found a center-of-gravity effect when the formant spacing was 3–3.5 Bark or less. Beddor and Hawkins (1984) found a similar effect for the pole-zero-pole combinations of nasal vowels, although the center of gravity of the prominence could not be simply described in terms of spectral measures. Psychophysical studies are clearly needed in order to understand further the perception of sounds characterized by spectral peaks with various degrees and shapes of prominence.

## ACKNOWLEDGMENTS

We would like to thank our subjects for their cooperation, and Patrice Beddor, Gunnar Fant, Len Katz, Christopher Darwin, Rena Krakow, and Suzanne Boyce for helpful discussions. This work was supported in part by NIH grants NS-04332 (to the Massachusetts Institute of Technology) and NS-07237 and RR-05596 (to Haskins Laboratories), and by grant MCS-87-12899 from the National Science Foundation.

Beddor, P. S. (**1983**). *Phonological and Phonetic Effects of Nasalization on Vowel Height* (Indiana University Linguistics Club, Bloomington, IN).

Beddor, P. S., and Hawkins, S. (**1984**). "The 'center of gravity' and perceived vowel height," J. Acoust. Soc. Am. Suppl. 1 **75**, S86.

Beddor, P. S., and Strange, W. (**1982**). "Cross-language study of perception of the oral–nasal distinction," J. Acoust. Soc. Am. **71**, 1551–1561.

Benguerel, A. -P., and Lafargue, A. (**1981**). "Perception of vowel nasalization in French," J. Phonet. **9**, 309–321.

Butcher, A. (**1976**). "The influence of the native language on the perception of vowel quality," Institut für Phonet., Universität Kiel, Arbeitsberichte **6**, 1–137.

Chistovich, L. A., and Lublinskaya, V. V. (**1979**). "The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli," Hear. Res. **1**, 185–195.

Chistovich, L. A., Sheikin, R., and Lublinskaya, V. (**1979**). "'Centers of gravity' and spectral peaks as determinants of vowel quality," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, New York), pp. 143–157.

Delattre, P. (**1954**). "Les attributs acoustiques de la nasalité vocalique et consonantique," Studia Linguist. **8**, 103–109.

Delattre, P. (**1968**). "Divergences entre la nasalité vocalique et consonantique en français," Word **24**, 64–72.

Delgutte, B., and Kiang, N. Y. S. (**1984**). "Speech coding in the auditory nerve: I. Vowel-like sounds," J. Acoust. Soc. Am. **75**, 866–878.

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

Ferguson, C. A. (**1963**). "Assumptions about nasals: A sample study in phonological universals," in *Universals of Language*, edited by J. J. Greenberg (MIT, Cambridge, MA), pp. 53–60.

Ferguson, C. A. (**1975**). "Universal tendencies and 'normal' nasality," in *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, edited by C. A. Ferguson, L.M. Hyman, and J. A. Ohala (Language Universals Project, Stanford University, Stanford, CA), pp. 175–196.

Ferguson, C. A., and Chowdhury, M. (**1960**). "The phonemes of Bengali," Language **36**, 22–59.

Fujimura, O. (**1960**). "Spectra of nasalized vowels," Res. Lab. Electron. Q. Prog. Rep. no. **58**, MIT, 214–218 (July 15, 1960).

Fujimura, O. (**1961**). "Analysis of nasalized vowels," Res. Lab. Electron. Q. Prog. Rep. no. **62**, MIT, 191–192 (July 15, 1961).

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

Hattori, S., Yamamoto, K., and Fujimura, O. (**1958**). "Nasalization of vowels in relation to nasals," J. Acoust. Soc. Am. **30**, 267–274.

Hawkins, S., and Stevens, K. N. (**1983**). "A cross-language study of the perception of nasal vowels," J. Acoust. Soc. Am. Suppl. 1 **73**, S54.

Henderson, J. B. (**1984**). "Velopharyngeal Function in Oral and Nasal Vowels: A Cross-Language Study," unpublished doctoral dissertation, University of Connecticut.

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

Keppel, G. (**1982**). *Design and Analysis: A Researcher's Handbook* (Prentice-Hall, Englewood Cliffs, NJ), 2nd ed.

Klatt, D.H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 979–995.

Klatt, D.H. (**1982**). "Prediction of perceived phonetic distance from critical-band spectra: A first step," Proc. ICASSP-82, IEEE Cat. no. 82CH1746-7, 1278–1281.

Lindqvist, J., and Sundberg, J. (**1972**). "Acoustic properties of the nasal tract," Phonetica **33**, 161–168.

Lonchamp, F. (**1978**). "Recherches sur les indices perceptifs des voyelles orales et nasales: application à la structure du système vocalique français et de diverses autres langues," Thèse 3$^{ème}$ cycle, l'Université de Nancy.

Lonchamp, F. (**1979**). "Analyse acoustique des voyelles nasales françaises," Verbum: Rev. de Linguist. Publiée Univ. Nancy II, **II**, 9–54.

Macmillan, N. (**1985**). "Beyond the categorical/continuous distinction: A psychophysical approach to processing modes," in *Categorical Perception*, edited by S. Harnad (Cambridge U. P., New York) (in press).

Maeda, S. (**1982a**). "The role of the sinus cavities in the production of vowels," Proc. ICASSP-82, Paris, 911–914.

Maeda, S. (**1982b**). "Acoustic correlates of vowel nasalization: A simulation study," J. Acoust. Soc. Am. Suppl. 1 **72**, S102.

Malécot, A. (**1960**). "Vowel nasality as a distinctive feature in American English," Language **36**, 222–229.

Malme, C. (**1959**). "Detectability of small irregularities in a broadband noise spectrum," Res. Lab. Electron. Q. Prog. Rep. no. **52**, MIT, 139–142.

Mrayati, M. (**1976**). "Contribution aux études sur la production de la parole," Thèse de doctorat, Grenoble, France.

Parducci, A. (**1974**). "Contextual effects: A range-frequency analysis," in *Handbook of Perception: Vol. 2, Psychophysical Judgment and Measurement*, edited by E. Carterette and M. P. Friedman (Academic, New York), pp. 127-141.

Rosen, S. M. (**1979**). "Range and frequency effects in consonant categorization," J. Phonet. **7**, 393–402.

Ruhlen, M. (**1975**). "Patterning of nasal vowels," in *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, edited by C. A. Ferguson, L. M. Hyman, and J. A. Ohala (Language Universals Project, Stanford University, Stanford, CA), pp. 175–196.

Sachs, M. B., and Young, E. D. (**1980**). "Effects of nonlinearities on speech encoding in the auditory nerve," J. Acoust. Soc. Am. **68**, 858–875.

Srulovicz, P., and Goldstein, J. L. (**1983**). "A central spectrum model: A synthesis of auditory-nerve timing and place cues in nonaural communication of frequency spectrum," J. Acoust. Soc. Am. **73**, 1266–1267.

Stevens, K. N., Fant, G., and Hawkins, S. (**1985**). "Some acoustical and perceptual correlates of nasal vowels," in *Festschrift for Ilse Lehiste*, edited by R. Channon and L. Shockey (Foris, Dordrecht, Holland).

Takeuchi, S., Kasuya, H., and Kido, K. (**1974**). "An active model for extraction of nasality and its perceptual evaluation," in *Speech Communication: Vol. 3, Speech Perception and Automatic Recognition,* edited by G. Fant (Almqvist and Wiksell, Stockholm), pp. 141–148.

Takeuchi, S., Kasuya, H., and Kido, K. (**1975**). "On the acoustic correlate of nasality," J. Acoust. Soc. Jpn. **31**, 298–309.

Winer, B. J. (**1971**). *Statistical Principles in Experimental Design* (McGraw-Hill, New York), 2nd ed.

Wright, J. (**1980**). "The behavior of nasalized vowels in the perceptual vowel space," in *Report of the Phonology Laboratory, 5* (University of California, Berkeley), pp. 127–163.