Integrality of nasalization and F_1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision–rule interactions

Neil A. Macmillan, John Kingston, Rachel Thorburn, Laura Walsh Dickey, and Christine Bartels

Citation: The Journal of the Acoustical Society of America **106**, 2913 (1999); doi: 10.1121/1.428113
View online: https://doi.org/10.1121/1.428113
View Table of Contents: https://asa.scitation.org/toc/jas/106/5
Published by the Acoustical Society of America

Integrality of nasalization and F_1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision–rule interactions

Neil A. Macmillan[a)]
Psychology Department, Brooklyn College, City University of New York, Brooklyn, New York 11210

John Kingston, Rachel Thorburn, Laura Walsh Dickey,[b)] and Christine Bartels
Linguistics Department, South College, University of Massachusetts, Amherst, Massachusetts 01003

(Received 23 November 1998; revised 28 June 1999; accepted 16 July 1999)

In vowel perception, nasalization and height (the inverse of the first formant, F_1) interact. This paper asks whether the interaction results from a sensory process, decision mechanism, or both. Two experiments used vowels varying in height, degree of nasalization, and three other stimulus parameters: the frequency region of F_1, the location of the nasal pole/zero complex relative to F_1, and whether a consonant following the vowel was oral or nasal. A fixed-classification experiment, designed to estimate basic sensitivity between stimuli, measured accuracy for discriminating stimuli differing in F_1, in nasalization, and on both dimensions. A configuration derived by a multidimensional scaling analysis revealed a perceptual interaction that was stronger for stimuli in which the nasal pole/zero complex was below rather than above the oral pole, and that was present before both nasal and oral consonants. Phonetic identification experiments, designed to measure trading relations between the two dimensions, required listeners to identify height and nasalization in vowels varying in both. Judgments of nasalization depended on F_1 as well as on nasalization, whereas judgments of height depended primarily on F_1, and on nasalization more when the nasal complex was below than above the oral pole. This pattern was interpreted as a decision–rule interaction that is distinct from the interaction in basic sensitivity. Final consonant nasality had little effect in the classification experiment; in the identification experiment, nasal judgments were more likely when the following consonant was nasal. *© 1999 Acoustical Society of America.* [S0001-4966(99)00511-1]

PACS numbers: 43.71.An, 43.71.Es [JMH]

INTRODUCTION

A. The psychophysics of perceptual interaction

Different physical dimensions of speech sounds can contribute to the same perceptual product. In Repp's classic characterization, these physical dimensions serve as multiple cues to phonetic contrasts, and ''a change in the setting of one cue...can be offset by an opposed change in the setting of another cue so as to maintain the original phonetic percept'' (Repp, 1982, p. 87). Such changes are often described as one cue ''trading'' with the other. In this paper we consider a well-established interaction of this sort, between the height and nasalization of vowels, and ask whether it arises at a basic, sensory level, or is located instead in the listener's decision-making process. The question, and our method for reaching an answer, have implications for all trading-relation results.

To make the sensory/decision distinction explicit, consider a 2×2 stimulus set constructed by combining two values on each of two dimensions. The elements of such a stimulus set can be represented as the four corners of a rectangle in a stimulus space. If each physical dimension is transduced into an independent perceptual one, then the representation of the stimulus set can be described as a rectangle in a *perceptual space*, as in Fig. 1(a). In our application, the dimensions are the perceptual correlates of the first formant (the *Cx* stimuli are higher, with smaller values of F_1), and nasalization (the *y0* stimuli have zero nasalization, the *yM* stimuli moderate nasalization). Each point may be thought of as the mean of a bivariate distribution of perceptual effects along these dimensions, and every point in the space corresponds to a percept that may arise on a particular trial. The circles around each mean connect points of equal likelihood, and indicate the spread of the distribution.

Now suppose that a listener in a trading-relations task must sort observations arising from these stimuli into categories such as ''high'' (an appropriate response for the *Cx* stimuli) or ''mid'' for the *Dx* stimuli). Any single point may arise from more than one stimulus, and the listener does best by establishing a *decision boundary* that divides the perceptual space into regions corresponding to each response. The solid line in Fig. 1(a) shows a boundary that is perpendicular to the dimension of judgment. Applying this boundary to the rectangular representation leads to the lack of a trading relation: the value of dimension 1 (perceived nasalization) has no effect on judgment of dimension 2 (perceived F_1, or inverse vowel height).

One way in which a trading relation can arise is by use of a nonorthogonal decision boundary like the dashed line in Fig. 1(a). Using this boundary, the ''high'' versus ''low''

[a)] Electronic mail: nmacmillan@gc.cuny.edu

[b)] Now at the Max Planck Institut für Psycholinguistik, PB 130, Nijmegen, NL 6500 AH, The Netherlands.

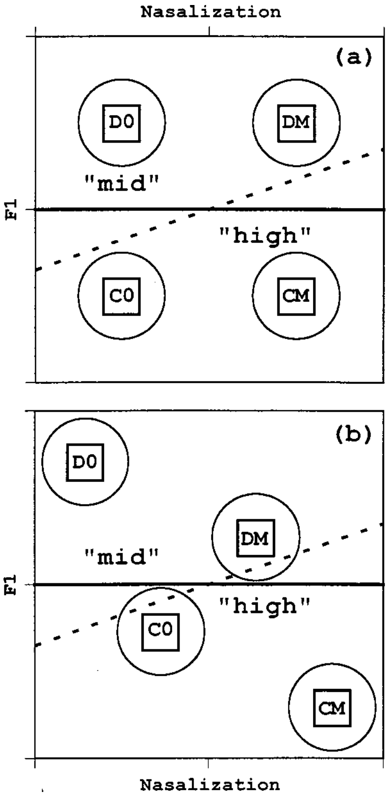


FIG. 1. Possible representations of four two-dimensional stimuli in a perceptual space. Squares represent the means of bivariate distributions. In (a), the representation is perceptually separable, because the squares form a rectangle, whereas in (b) it is perceptually integral. Solid and dashed lines are decision boundaries that might be used in phonetic identification of the vertical (F_1) dimension, the solid lines being perpendicular to perceived F_1 and the dashed ones not. Boundary shifts—discrepant positions of the boundary on the *D0-C0* and *DM-CM* segments—occur if either the boundary is nonorthogonal or the representation is perceptually integral.

decision is different for the *y0* column and the *yM* column: the *yM* stimuli are identified as ''high'' more often than the corresponding *y0* stimuli. The two stimulus dimensions provide independent information, but the listener's decision is based on a linear combination of them instead of just the relevant one; this is a trading relation whose source is *decisional.*

In Fig. 1(b), the representation itself leads to a trading relation, because the two independent stimulus dimensions map onto two nonindependent perceptual ones. To put it differently, the two stimulus dimensions contribute to a common perceptual variable that runs roughly from *D0* to *CM*. If the decision depends only on the vertical dimension, the solid-line decision boundary is used, but the result is still that the *yM* (nasalized) stimuli are identified as ''high'' more often than the corresponding *y0* (unnasalized) stimuli. This is a trading relation whose source is *sensory*. Finally, the dashed line in Fig. 1(b) shows a nonorthogonal boundary used in a non-independent space; in this case, the trading relation results from the combined sensory and decisional effects.

We report here two experiments designed to tease apart these sensory and decisional components. In experiment I, a replication and expansion of Kingston and Macmillan (1995), listeners classified stimuli from a large number of minimal, two-stimulus sets. The data allowed us to determine a sensory representation for vowels differing in F_1 and nasalization and to examine the stimulus characteristics that correspond to perceived height and nasalization. In experiment II, listeners judged F_1 while nasalization was varied irrelevantly (as in Krakow *et al.*, 1988), and in a separate condition judged nasalization while F_1 varied irrelevantly. The data provided information about both sensory and decisional aspects.

B. The trading-relations paradigm

In the trading-relations paradigm (Repp, 1982), observers provide phonetic labels for sounds drawn from a two-dimensional stimulus set. In a typical application, the data are summarized by *identification functions* that give the percentage of trials on which each value of a dimension leads to a particular response. The value of dimension 1 that is assigned on 50% of trials to each of two categories is termed a *boundary*, and the extent of trading is measured by the *boundary shift* on the stimulus axis when the value of dimension 2 is changed.

Krakow *et al.* (1988) used the trading-relations paradigm to study the interaction between the height and nasalization of vowels produced with an articulatory synthesizer. Their listeners classified vowels on continua between [æ] (a low vowel, with high F_1) and [ɛ] (a mid vowel, with lower F_1), and displayed a trading relation: The boundary between ''ɛ'' and ''æ'' shifted closer to [ɛ] with greater nasalization. Because lowering the soft palate (increasing nasal coupling) had the same perceptual effect as lowering the tongue, soft palate and tongue height can be said to integrate ''positively'' in Krakow *et al.*'s data. The same description can be applied to the acoustic consequences of these articulations, for F_1 and nasalization are inversely related to tongue and soft palate height, respectively. Krakow *et al.* obtained the boundary shift when the following consonant was oral (in context [bVd]), but not when it was nasal (in context [bVnd]). Krakow *et al.* argued that this last result occurred because listeners attributed the vowel's nasalization to coarticulation with the consonant, thereby hearing out the effect of tongue height alone.

As the comparison of Fig. 1(a) and (b) shows, such a shift in the decision boundary could arise either because the two stimulus dimensions do not map into independent perceptual dimensions, or because the decision criterion depends on the value of the orthogonal variable, or because both effects occur. The necessary independent means of determining the mapping from stimulus to perceptual dimensions is provided by fixed classification tasks.

C. The fixed-classification paradigm

In the fixed-classification paradigm, listeners' ability to distinguish between two incompletely discriminable stimuli is directly measured by asking them to assign different responses to the two stimuli. Performance is converted to an index with distance properties, such as the d' of detection theory. A geometric model of the data is constructed using data from many possible stimulus pairs.

Kingston and Macmillan (1995) examined the nasalization–F_1 interaction using this method. Stimuli were drawn from sets constructed by combining two values of F_1 with two of nasalization. Six pairs of stimuli can be drawn from each 2×2 array, a single pair differing in F_1, nasalization, or both. In any block of trials, only elements of one such pair were presented. Ability to distinguish a pair was indexed by d', and the six d' estimates were interpreted as the sides and diagonals of a parallelogram in perceptual space. Kingston and Macmillan (as well as Kingston, 1991) found that vowel pairs in which nasalization varied directly with F_1 (*DM* and *C0*) were consistently more difficult to classify than those in which they varied inversely (*D0* and *CM*), so that F_1 and nasalization integrated negatively, as in Fig. 1(b), a result that can be described as *mean-integrality*[1] (see Maddox, 1992). The effect was obtained whether the following consonant was nasal or oral.

Kingston and Macmillan's (1995) conclusions may appear to be in conflict with those reached by Krakow *et al.* (1988) using the trading-relations paradigm, but Fig. 1 makes it clear that the presence of a boundary effect and the finding of mean-integrality are, in fact, unrelated phenomena. As we have seen, a trading relation can occur whether or not dimensions are mean-integral, and vice versa. One way to understand the unrelatedness of the two measures is to see that integrality is a function of sensitivity values, whereas a boundary shift can be just a measure of response bias.

However, information about sensitivity is available from the phonetic identification task: The ability to distinguish two stimuli (differing in, say, F_1) can be estimated by subtracting the *z*-transformed proportions of ''high'' responses to them. This statistic can be interpreted as a d' value, and is related to the *slope* of the identification function. In a fixed classification study, sensitivity is measured directly as d', proportion correct $[p(c)]$, or a related variable. The major point of contact between the fixed classification and trading relations data is a comparison of their sensitivity estimates.

I. ACOUSTIC AND PSYCHOACOUSTIC DIMENSIONS CORRESPONDING TO VOWEL HEIGHT AND NASALIZATION

The dimensions of the presumed perceptual space, and their interrelation, are the consequence of a mapping from acoustic dimensions, which are themselves determined by the speaker's articulations. Both transformations are complex. Increasing the height of the tongue extends the pharyngeal cavity and narrows the vocal tract at the palate, thereby lowering F_1 and increasing perceived vowel height. Vowel nasalization is produced by lowering the soft palate, opening the velopharyngeal port and thereby acoustically coupling the nasal to the pharyngeal cavity. The two principal acoustic effects of nasal–pharyngeal coupling are adding a pole/zero pair in the low-frequency part of the spectrum, and raising F_1. Furthermore, nasalized vowels differ in the positions of the nasal pole/zero complex relative to the oral pole as a function of height: The complex lies above the low oral pole of high vowels, below the high oral pole of low vowels, and either above or below the mid-frequency oral pole of mid vowels. These effects are jointly produced in articulatorily synthesized vowels (e.g., Rubin *et al.*, 1981; Krakow *et al.*, 1988) but can also be obtained with a terminal analog synthesizer (Klatt and Klatt, 1990) by independently controlling the frequency difference between the nasal pole (N_1) and the nasal zero (N_0), and setting F_1 appropriately.

The frequency of the nasal pole/zero complex relative to F_1, and the frequency separation between the nasal pole and zero, affect both the center-of-gravity and bandwidth of the low-frequency region. The complex may consequently alter perception of the vowel's height as well as its nasalization. Perception of vowel height as well as nasalization is also altered by the F_1 raising caused by nasal–pharyngeal coupling. Finally, the listener may be uncertain about which peak to attribute to nasal–pharyngeal coupling and which to attribute to tongue height. It is not surprising, therefore, that perceptual interactions are found between height and nasalization, and that these occur whether nasalization is contrastive or not (as in English).

In this study, we factorially manipulated three stimulus variables that have been found to play a role in the perception of nasalization: consonantal context, the location of the nasal pole/zero complex relative to the oral pole, and F_1. Because of the high-dimensional nature of speech, we did not expect this design to uncover a single stimulus correlate of vowel nasalization or height, and it did not. However, the results do limit models of psychoacoustic processing, and we did not want our psychophysical conclusions to be specific to an idiosyncratic choice of stimuli.

A. Consonantal context

Our vowels were placed in CVC syllables, and the final consonant was either nasal or oral. In Kingston and Macmillan (1995) a similar manipulation had no effect on fixed classification, but in the trading relation study of Krakow *et al.* (1988) nasalized vowels were judged lower in an oral than in a nasal context. Other data also show that listeners are less likely to attribute vowel nasalization to a vowel if they perceive nasality in an adjacent, potentially coarticulating consonant. Kawasaki (1986) presented listeners with nasalized vowels between nasal consonants, and found that judged nasalization increased as the consonant's intensity was reduced. Krakow and Beddor (1991) asked listeners to match naturally produced vowels for nasalization and to judge how nasal they were. Their stimuli were oral vowels produced between oral consonants [b_d], nasalized vowels produced between nasal consonants [m_n], oral vowels cross-spliced into nasal [m_n] contexts, nasalized vowels cross-spliced into oral [b_d] contexts, and isolated oral and nasalized vowels spliced out of these contexts. Their listeners matched nasalized vowels most accurately when the vowels were in isolation, and more accurately when the nasalized vowels occurred in an oral than a nasal context. These listeners also judged a vowel to be more nasal in isolation and between oral consonants than between nasal consonants.

B. Frequencies of the nasal pole/zero and first formant

The nasal pole/zero complex may be located either above or below the oral pole: For low vowels, F_1 is high and the complex falls below it, whereas for high vowels, F_1 is low and the complex falls above it. For mid vowels, Stevens *et al.* (1987) suggest a placement below F_1, and that is what Kingston and Macmillan (1995) used. However, Beddor and Hawkins (1990) and even more directly Maeda (1993) prescribe a placement above F_1 for such vowels. The mean F_1 frequency in Kingston and Macmillan's vowels was about 400 Hz, on the cusp between the above and below cases, according to Maeda (1993). In the present experiments, both placements were used, in separate conditions. We independently manipulated the F_1 range, a supposed determinant of the proper location of the pole/zero complex, by using two sets of vowels, with F_1 centered at 480 Hz for the higher range and at 380 Hz for the lower range.

II. EXPERIMENT I: ONE-DIMENSION AND TWO-DIMENSION (CORRELATED) CLASSIFICATION

A. Methods

1. Stimuli

Both experiments used stimulus sets in which F_1 and the nasal zero-pole difference N_0-N_1 (henceforth denoted N) varied orthogonally, using the Klatt and Klatt (1990) synthesizer. The ranges of all parameters but N were identical in the two experiments, and the difference in the range of N values in experiments I and II was only 3 Hz.

Stimuli for experiment I were of the form [CVC], the initial consonant being one of [b,d] and the final consonant one of [b,d,m,n]. The values of the formants in the vowel were appropriate for a high or mid, back, rounded quality (as in our previous experiments). There were two stimulus sets, Above and Below, named for the location of the nasal pole/zero complex *re* F_1. In each set, four values of F_1 and three values of N were combined independently. There were 8 variants (2 initial consonants×4 final consonants) of each the 12 combinations, for a total of 96 distinct stimuli in the Above and in the Below conditions.

In both the Above and Below sets, F_1 could be Low (360 or 400 Hz) or High (460 or 500 Hz), straddling the frequency at which Maeda (1993) predicts that the nasal pole/zero complex crosses over from above to below F_1. The Low-F_1 stimuli are closer to those used by Kingston and Macmillan (1995) (415–430 Hz) and Kingston (1991) (400–430 Hz). The larger 40-Hz interval was used here to increase listeners' success in classifying the vowels for F_1 differences. The degree of nasalization was Zero ($N = 0$ Hz), Moderate (43 Hz), or Heavy (87 Hz).

Figure 2 shows values of F_1, N_1, and N_0 for some of the stimuli. The nasal pole N_1 was always separated from F_1 by 175 Hz. When the nasal complex was above F_1 [Fig. 2(a)], the frequencies F_1, N_1, N_0 were in that order; when the nasal complex was below F_1 [Fig. 2(b)], the frequencies N_1, N_0, F_1 were in ascending order. In this paper, we refer to each stimulus by a letter–number pair, as shown in Table I; for example *D0* is the stimulus with $F_1 = 500$ Hz and $N = 0$ Hz, *CM* is the stimulus with $F_1 = 460$ Hz and $N = 43$ Hz, etc. Other stimulus details are in Appendix A.

We used the VIIIth nerve response model of Moore and Glasberg (1987; Glasberg and Moore, 1990) to calculate, from 1024-point fast Fourier transform (FFT) spectra centered in the vowel, the sensory (equivalent rectangular bandwidth, or ERB rate) spectra of our vowels. These spectra are shown in Fig. 3. The *center of gravity* (COG) for each stimulus between 0 and 2 kHz, plotted in Fig. 4, drops with increasing nasalization and increases with F_1 both above and below.

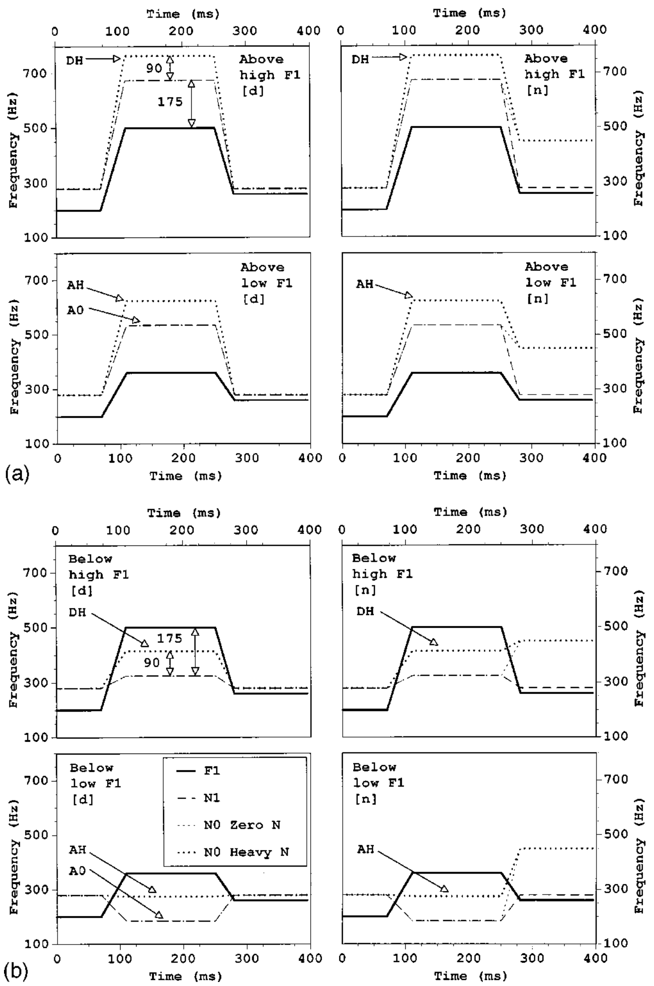


FIG. 2. Time course of oral pole, nasal pole, and nasal zero for (a) Below and (b) Above stimuli. Stimuli with the highest and lowest F_1 and highest and lowest N values and shown.

2. Procedures

Experiment I used eight listeners, paid volunteers recruited by advertisement from the undergraduate student body at the University of Massachusetts, Amherst. All spoke

TABLE I. Nomenclature for vowel stimuli in experiment I.

	First formant (F_1) in Hz	Nasalization ($N = N_0 - N_1$) in Hz		
		0 (Zero)	43 (Moderate)	87 (High)
High F_1	500	*D0*	*DM*	*DH*
	460	*C0*	*CM*	*CH*
Low F_1	400	*B0*	*BM*	*BH*
	360	*A0*	*AM*	*AH*

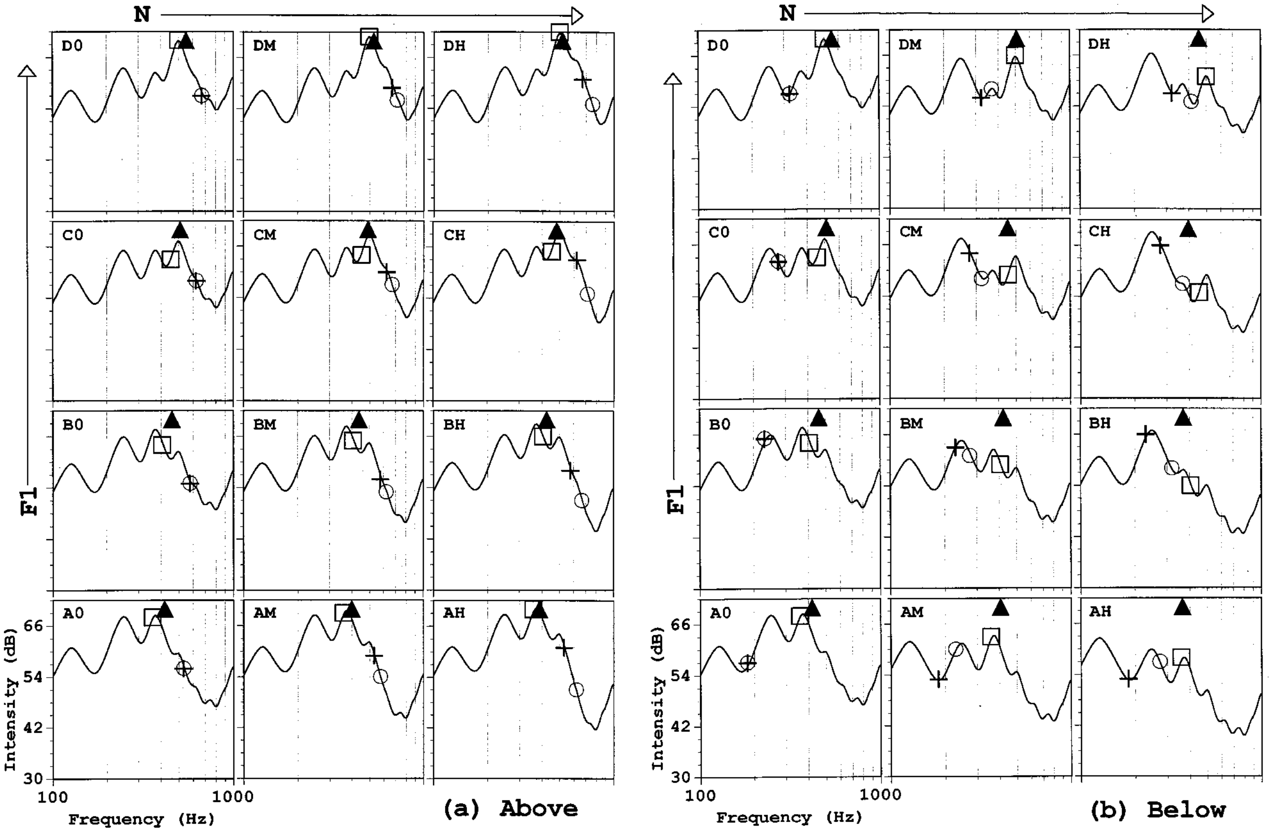


FIG. 3. The ERB-rate spectra from 0.1 to 1 kHz for selected stimuli in the (a) Above and (b) Below sets. Columns are Zero, Moderate, and Heavy nasalization; rows are $F_1 = 360$, 400, 460, and 500 Hz. Squares indicate F_1, plus signs indicate N_1, and circles indicate N_0; the triangles indicate the center-of-gravity (COG) of the 0.1- to 2-kHz interval.

English natively and none reported any speech or hearing pathology. Four were assigned to the Above condition and four to the Below condition.

Listeners heard the stimuli binaurally at self-selected comfortable listening levels over TDH-49 headphones, while sitting in semi-isolation in a sound-treated room. They performed all 18 possible one-dimension and all 12 possible two-dimension classification tasks with stimuli drawn from within the High-F_1 (stimuli *Dx* and *Cx* in Table I) and Low-F_1 (*Bx* and *Ax*) arrays. These tasks correspond to the sides (for example, *D0* vs *DM* or *B0* vs *A0*) and diagonals (for example, *D0* vs *CM* or *A0* vs *BM*) of all possible 2×2 stimulus subarrays. Listeners gave one of two responses to classify the stimulus, followed by a confidence rating on a 1–4 scale. For example, in the *D0* vs *DH* task, one button indicated that the stimulus was a ‘‘*D0*,’’ the other that it was a ‘‘*DH*.’’ The confidence judgment was prompted by a rapid tone triplet occurring 750 ms after the listener responded, and was entered, like the response, by a button press. The listener had 2000 ms to make the initial response and 1500 ms to make the confidence judgment. A 500-ms feedback light then came on over the button corresponding to the correct response, and there was an additional 1000 ms before the next trial began.

Each block consisted of 16 orientation trials in which the stimuli alternated between the two classes and between oral and nasal following consonants, and 80 randomized test trials in which the contrasting vowels occurred equally often

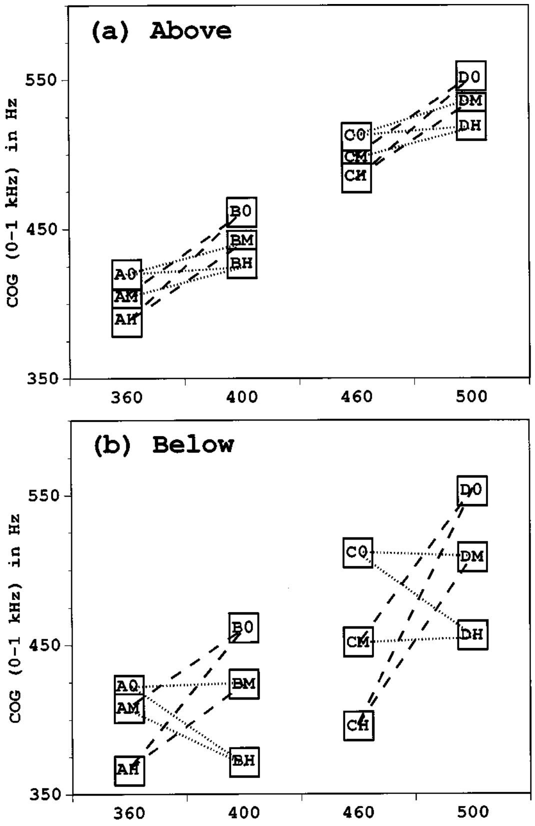


FIG. 4. Center-of-gravity for the 0- to 1-kHz region of the stimuli used in experiment I. Dashed lines connect stimuli in which F_1 and N are negatively related, dotted lines those in which they are positively related.

before oral and nasal following consonants. The listeners were taught which button to press for a particular stimulus by the orientation trials and by feedback; no other instructions were given.

Type of task alternated from day to day between single-dimension and two-dimension classification. Within a day, single-dimension tasks alternated in groups of three between the F_1 and N dimensions, and two-dimension tasks alternated from block to block between those in which the dimensions were correlated positively (for example, *C0* vs *DM*) and negatively (for example, *D0* vs *CM*). Tasks were drawn in a pseudo-randomized fashion from both the Low- and High-F_1 arrays within each day. The entire series of tasks was run twice, once early in the string of days and then again later, with the order of tasks reversed between the two runs. Task order during a day in the Above condition reversed that used in the Below condition. Between 9 and 12 blocks of trials were run each day in sessions lasting 90 min. As there were 6 different classifications for F_1, 12 different classifications for N, and 12 different combinations of F_1 and N, completing two passes through the 30 classification tasks plus the consonant identification blocks took 7–9 days, which were spread over 3–4 weeks.

To check that the final consonants were heard with the intended nasality, we asked listeners to identify the final consonants in the stimuli (as ‘‘b,’’ ‘‘d,’’ ‘‘m,’’ or ‘‘n’’) several times during the running of the experiment. Trials were presented in blocks that began with 16 orientation trials in which the stimuli cycled in a fixed pattern through the four final consonants, and continued with 96 randomized test trials in which each stimulus was presented once. All other procedures were identical to those used to collect the vowel judgments. Feedback was provided to train listeners in hearing the consonants with the intended nasality, but their success (see below) suggests that this was unnecessary. Three of the listeners in the Below condition heard four such blocks, the fourth listener five. All listeners in the Above condition heard eight blocks. As there were 24 stimuli for each final consonant, pooling responses across listeners yields 408 responses per consonant in the Below condition and 768 responses per consonant in the Above condition.

B. Results

1. Consonant identification

The overall error rate in consonant identification was 4.2%. This number is much lower than the 25.8% found by

TABLE II. Mean d_a values (standard errors) across listeners, pooled across replication and final consonant place of articulation. Stimulus parameters are F_1 (in Hz) and N [Z(ero), M(oderate), and H(eavy)]. The code for the stimuli is given in Table I.

a. Nasal complex Above F_1, vowel precedes Oral consonant

Stimulus	F_1,N	500,M	500,H	460,Z	460,M	460,H	Stimulus	F_1,N	400,M	400,H	360,Z	360,M	360,H
F_1, N	Code	*DM*	*DH*	*C0*	*CM*	*CH*	F_1, N	Code	*BM*	*BH*	*A0*	*AM*	*AH*
500, Z	*D0*	.62 (.21)	1.70 (.39)	1.23 (.38)	2.51 (.63)	3.62 (.65)	400, Z	*B0*	.96 (.42)	1.93 (.53)	2.30 (.50)	2.70 (.59)	3.49 (.25)
500, M	*DM*		.71 (.36)	.43 (.07)	.76 (.36)	1.63 (.21)	400, M	*BM*		.49 (.15)	1.93 (.35)	2.01 (.50)	3.31 (.41)
500, H	*DH*			1.63 (.21)	1.31 (.28)	.26 (.17)	400, H	*BH*			2.08 (.32)	1.90 (.26)	1.58 (.23)
460, Z	*C0*				.73 (.24)	1.47 (.52)	360, Z	*A0*				1.00 (.32)	1.79 (.47)
460, M	*CM*					1.00 (.23)	360, M	*AM*					1.18 (.27)

b. Nasal complex Above F_1, vowel precedes Nasal consonant

Stimulus	F_1,N	500,M	500,H	460,Z	460,M	460,H	Stimulus	F_1,N	400,M	400,H	360,Z	360,M	360,H
F_1, N	Code	*DM*	*DH*	*C0*	*CM*	*CH*	F_1, N	Code	*BM*	*BH*	*A0*	*AM*	*AH*
500, Z	*D0*	1.21 (.36)	1.95 (.68)	1.52 (.42)	2.66 (.56)	3.56 (.53)	400, Z	*B0*	1.14 (.36)	2.63 (.70)	1.13 (.29)	2.59 (.31)	3.84 (.14)
500, M	*DM*		.75 (.22)	.64 (.04)	.94 (.27)	2.05 (.32)	400, M	*BM*		.80 (.11)	1.45 (.44)	1.50 (.39)	2.74 (.29)
500, H	*DH*			1.60 (.43)	.59 (.13)	.54 (.31)	400, H	*BH*			1.54 (.11)	1.37 (.22)	.94 (.25)
460, Z	*C0*				1.14 (.25)	2.13 (.64)	360, Z	*A0*				1.32 (.39)	2.04 (.30)
460, M	*CM*					1.01 (.40)	360, M	*AM*					.82 (.16)

TABLE II. (*Continued.*)

c. Nasal complex Below F_1, vowel precedes Oral consonant

Stimulus	F_1,N	500,M	500,H	460,Z	460,M	460,H	Stimulus	F_1,N	400,M	400,H	360,Z	360,M	360,H
F_1, N	Code	*DM*	*DH*	*C0*	*CM*	*CH*	F_1, N	Code	*BM*	*BH*	*A0*	*AM*	*AH*
500, Z	*D0*	.43 (.26)	2.00 (.24)	1.10 (.06)	2.18 (.14)	3.81 (.39)	400, Z	*B0*	1.54 (.11)	3.78 (.30)	2.14 (.42)	3.27 (.45)	3.96 (.40)
500, M	*DM*		.67 (.16)	.62 (.24)	.87 (.15)	3.52 (.31)	400, M	*BM*		2.08 (.37)	.67 (.08)	2.32 (.17)	2.75 (.31)
500, H	*DH*			.74 (.26)	.68 (.18)	1.64 (.33)	400, H	*BH*			.35 (.23)	.17 (.14)	1.23 (.18)
460, Z	*C0*				.52 (.19)	3.17 (.47)	360, Z	*A0*				.69 (.19)	2.66 (.45)
460, M	*CM*					2.54 (.23)	360, M	*AM*					2.35 (.14)

d. Nasal complex Below F_1, vowel precedes Nasal consonant

Stimulus	F_1,N	500,M	500,H	460,Z	460,M	460,H	Stimulus	F_1,N	400,M	400,H	360,Z	360,M	360,H
F_1, N	Code	*DM*	*DH*	*C0*	*CM*	*CH*	F_1, N	Code	*BM*	*BH*	*A0*	*AM*	*AH*
500, Z	*D0*	.26 (.30)	1.75 (.15)	.89 (.10)	2.25 (.07)	3.75 (.53)	400, Z	*B0*	1.05 (.15)	3.21 (.25)	2.06 (.28)	3.51 (.40)	3.73 (.12)
500, M	*DM*		.37 (.15)	.73 (.31)	.62 (.11)	4.05 (.17)	400, M	*BM*		1.78 (.33)	.86 (.23)	2.34 (.13)	2.64 (.38)
500, H	*DH*			.15 (.28)	.28 (.17)	1.57 (.44)	400, H	*BH*			.45 (.15)	.48 (.06)	1.45 (.31)
460, Z	*C0*				.60 (.14)	3.18 (.30)	360, Z	*A0*				.68 (.17)	2.14 (.30)
460, M	*CM*					2.53 (.39)	360, M	*AM*					2.27 (.27)

Kingston and Macmillan (1995), a result of careful resynthesis of the stimuli. The largest errors were mistaking final [m] for [b] on 12.7% of trials in the Above condition, and mistaking [d] for [n] on 6.7% of trials in the Below condition. Oral:nasal confusions are thus few enough that the listeners can be considered to be hearing the vowels in consonantal contexts that are distinct on the oral:nasal dimension.

2. Vowel classification accuracy

For each pair of stimuli and subject, the data were pooled across place of articulation of the initial consonant and repetition of the task to produce a matrix in which each combination of response and confidence rating to each stimulus was separately represented. Using the method of Dorfman and Alf (1969), receiver operating characteristic (ROC) curves were fitted to the multiple estimates of hit and false-alarm proportions obtained from these matrices. A natural index of sensitivity is the area under the ROC curve, which equals optimal proportion correct in a two-alternative forced-choice task (Green and Swets, 1966; Macmillan and Creelman, 1991). We transformed this statistic to d_a, a distance measure that generalizes the better-known d' to models of unequal underlying variance (Swets and Pickett, 1982; Macmillan and Creelman, 1991). Values of d_a, averaged across subjects, with standard errors, are given in Table II. Four separate subtables are provided to describe the Above versus Below position of the nasal complex and the Oral versus Nasal feature of the following consonant.[2]

The most striking feature of the data is that the negatively correlated two-dimensional comparisons (e.g., *D0* vs *CM*) yield in every case greater accuracy than the corresponding positively correlated ones (e.g., *C0* vs *DM*). The discrepancy is quite large, averaging 2.39 d_a units. This indicates a perceptual interaction: Stimulus *C0* is hard to discriminate from *DM* because it is lower in both F_1 and *N*, and the effects of these differences cancel each other. Stimulus *CM* is easy to discriminate from *D0* because it is lower in F_1 but higher in *N*, and the effects of these differences augment each other. Lower *N* may cancel lower F_1 by undoing the lowering of COG, whereas higher *N* adds to the lowering of COG (see Figs. 3 and 4).

Geometric models of perceptual interaction (like those used in Kingston and Macmillan, 1995) take advantage of the status of d_a as a distance measure. The data allow us to test a critical distance axiom, the triangle inequality. Three sets of conditions can be examined. First, consider stimulus triples, like $\{D0,DM,DH\}$, that differ only on *N*. If these stimuli are represented as points on the same perceptual dimension, then we expect $d_a(D0,DM)+d_a(DM,DH)$

$=d_a(D0,DH)$; if not, we expect $d_a(D0,DM)+d_a(DM,DH)>d_a(D0,DH)$. Examining the 16 such triples in the table, the sum of the smaller values is 0.14 d_a units *less* than the larger one, a small violation of the triangle inequality. Second, none of the 48 triples involving the positively correlated comparisons, e.g., $\{D0,DM,C0\}$ and $\{DM,C0,CM\}$, violate the triangle inequality because d_a values for the positively correlated tasks, such as *DM* vs *C0*, are consistently small. The third relevant stimulus set is all triples, such as $\{D0,DM,CM\}$ and $\{D0,C0,CM\}$, that include a negatively correlated comparison (*D0* vs *CM*). Of 48 such comparisons, 36 fail to satisfy the triangle inequality, the average discrepancy being 0.34 d_a units.

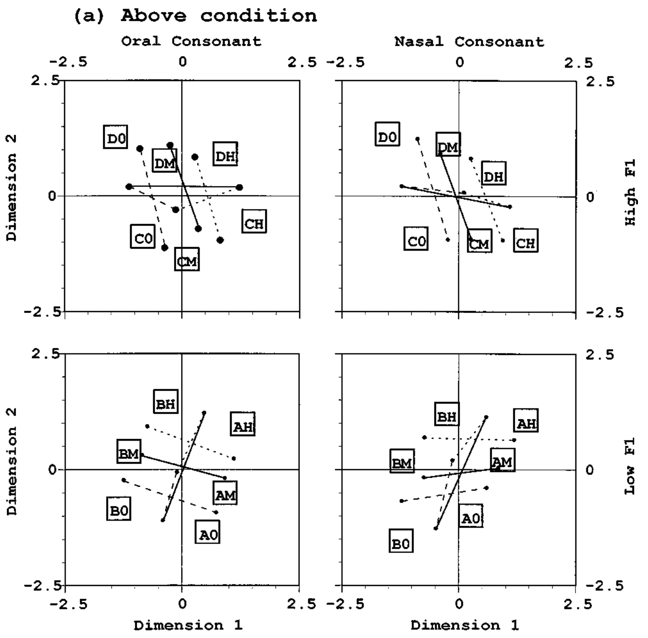
The violation of the axiom is relatively small in magnitude (equivalent to about 4 or 5 percentage points for a d_a values of 2.0), and could be interpreted as a case of ‘‘more-than-complete’’ integrality, but the effect is too systematic to be entirely due to chance. One possible culprit is the implicit assumption of our detection-theoretic analysis that all distributions have the same covariance matrix; unfortunately, the experimental design does not allow for a test. We adopt instead a data-analysis strategy that adjusts the perceptual distances between stimuli so that they *do* satisfy the axioms: multidimensional scaling.

3. Multidimensional scaling (MDS) analysis of perceptual interaction

We performed multidimensional scaling (INDSCAL) on the values of d_a; because this statistic is a distance measure, we constrained the program to fit the actual data in Table II, not an arbitrary monotonic transformation of them. We expected that two perceptual dimensions, roughly corresponding to the two stimulus dimensions, would describe the data, and considered only two-dimensional outcomes. Each section of Table II contains all pairwise d_a values for both the High-F_1 and the Low-F_1 subset (but no comparisons between these subsets), so there are eight resulting configurations, shown in Fig. 5. Stress values range from 0.12 to 0.19, averaging 0.17 (Kruskal’s stress formula 1). The proportion of variance in the data that is accounted for by distances in the configurations ranges from 0.63 to 0.91, averaging 0.77. These moderately low stress values and moderately high squared correlations imply, for metric scaling, that distances in the representation are approximately proportional to the input data, and thus justify the use of d_a as a distance measure.[3]

How is the nature of perceptual interaction captured by these representations? Let us first consider the Above/Nasal/High data [upper right panel of Fig. 5(a)]. A ‘‘centroid’’ (represented by a dot) is shown for each pair of stimuli sharing the same value of *N* (for example, *D0* and *C0*), and line segments are drawn in Fig. 5 between pairs of corresponding centroids (for example, those of {*D0,C0*} and {*DM,CM*}). Three such segments, for Zero versus Medium *N*, Zero versus Heavy, and Medium versus Heavy, indicate the contours of perceptual change as *N* is increased, keeping F_1 constant. Three analogous line segments (for example, between the centers of gravity of $\{D0,DM\}$ and $\{C0,CM\}$) show the contours of perceptual change as F_1 decreases, keeping *N* constant.

(a) Above condition



(b) Below condition

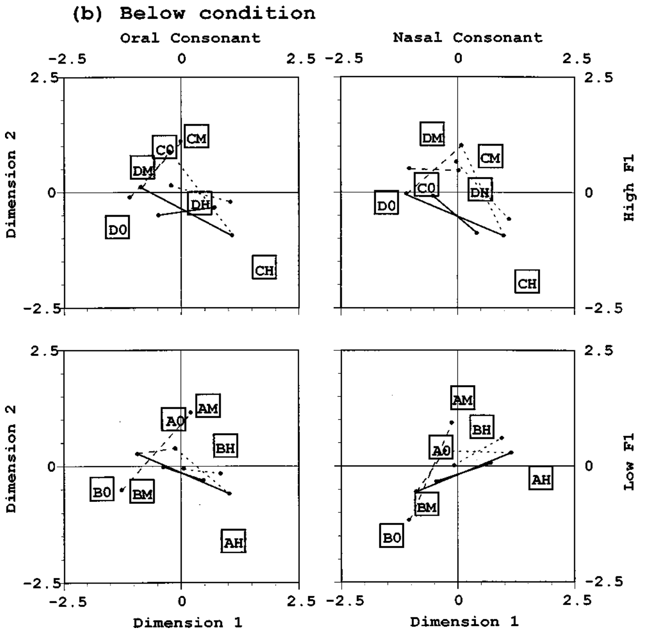


FIG. 5. The MDS configurations for all sets of six stimuli, based on the classification data of experiment I. In (a) the nasal complex is Above F_1 and in (b) it is Below. In each panel, the top row is for High-F_1 sets (stimuli *Dx* and *Cx*), the bottom row for Low-F_1 sets (stimuli *Bx* and *Ax*), the left column for vowels preceding an Oral consonant, and the right column for vowels preceding a Nasal consonant. Line segments connect midpoints of the sides of the implied quadrilateral for each pair of *N* values: solid lines for Zero: Heavy *N* (quadrilaterals {*D0,DH,C0,CH*} for High F_1 and {*B0,BH,A0,AH*} for Low F_1), dashed lines for Zero:Moderate *N* ({*D0,DM,C0,CM*}, {*B0,BM,A0,AM*}), and dotted lines for Moderate: Heavy *N* ({*DM,DH,CM,CH*}, {*BM,BH,AM,AH*}).

Degree of interaction can be measured by the angle θ at which an *N* contour and an F_1 contour intersect. An angle of 90 degrees would reflect orthogonality (noninteraction), angles of 0 or 180 degrees complete (negative or positive) interaction. For the {*D0,DM,C0,CM*} stimulus subset, these

TABLE III. Values of θ (in degrees) averaged across listeners (standard errors), representing extent of perceptual interaction between F_1 and N in various stimulus sets. All angles are less than 90 degrees, indicating a negative interaction; angles greater than 90 degrees would indicate a positive one. (Standard errors for row and column average are based on the number of measurements contributing to the mean, not the number of listeners.)

Stimulus set			Nasalization comparison			
N-complex location	Final consonant	F_1 region	Zero versus Moderate	Moderate versus Heavy	Zero versus Heavy	Average
Above F_1	Oral	High	29 (9)	87 (6)	56 (11)	57 (9)
		Low	63 (20)	74 (4)	69 (7)	68 (7)
	Nasal	High	35 (8)	51 (9)	44 (8)	43 (5)
		Low	46 (18)	48 (16)	37 (10)	44 (3)
	Average		43 (8)	65 (6)	51 (5)	53 (13)
Below F_1	Oral	High	29 (10)	32 (6)	24 (5)	28 (4)
		Low	34 (11)	28 (10)	16 (5)	26 (5)
	Nasal	High	30 (14)	20 (6)	26 (8)	25 (5)
		Low	30 (8)	35 (8)	10 (5)	25 (5)
	Average		30 (5)	28 (6)	18 (5)	26 (2)

contours intersect at an angle θ of 56 degrees, indicating that F_1 and N interact moderately in this region of the stimulus space. We used the angle θ to assess the degree of interaction in all conditions, with the results shown in Table III.

4. Effects of stimulus manipulations

Using θ as a measure, we now summarize the effects of stimulus variables on the degree of perceptual interaction.[4] To evaluate reliability, we applied the method described above to each listener's data in each condition,[5] and subjected the resulting θ values to a repeated-measures ANOVA. The independent within-subjects variables were F_1 range, N difference, and following consonant nasality; the independent between-subjects variable was the position of the nasal pole/zero complex with respect to the oral pole. There are three major results.

First, F_1 and N interacted in our listeners' perceptions: θ averaged 40 degrees (95% confidence interval$=\pm 6$ degrees), and all the values in Table III reflect negative interactions ($\theta < 90$ degrees), so that an increase in N and a decrease in F_1 had similar perceptual effects (as in Kingston and Macmillan, 1995). This conclusion paraphrases the earlier observation that highest discriminability is obtained with stimuli in which N and F_1 covary negatively.

Second, F_1 and N integrate more when the nasal pole/zero complex is below the oral pole ($\theta = 26$ degrees, ± 7 degrees) than when it is above [$\theta = 53$ degrees, ± 7 degrees; $F(1,6) = 40.6$, $p = 0.001$]. The effect in the Below condition is qualitatively different from that in the Above condition. The configurations of the Above-condition stimulus subsets are approximately parallelograms, and increases in N produce approximately linear paths through the space; whereas in the Below condition points in the perceptual space often do not form parallelograms, and increases in N produce nonlinear paths. In particular, stimuli *CH* and *AH* (heavy N and the lower value of F_1) differ from the other stimuli in a direction approximately orthogonal to the dimension along which the other stimuli vary.

Third, F_1 and N were more integral before nasal consonants ($\theta = 34$ degrees, ± 7) than oral ones [$\theta = 45$ degrees, ± 7; $F(1,6) = 6.68$, $p = 0.041$]. The difference in θ values before nasal and oral consonants was much smaller when the nasal pole/zero complex was below the oral pole (25 degrees, ± 8, versus 27 degrees, ± 10) than above (43 degrees, ± 8, versus 63 degrees, ± 10), but the interaction between these variables did not achieve significance [$F(1,6) = 4.34$, $p = 0.082$)].[6]

C. Discussion

Experiment I yielded strong evidence of mean-integrality. We first summarize this evidence and compare our MDS-psychological space assessment of interaction with our previous approach. We then examine psychoacoustic mechanisms that might be responsible for the effect.

1. Evidence of integrality

a. Perceptual-space analysis. Kingston and Macmillan's (1995) conclusion that F_1 and N integrate negatively is confirmed in these experiments. Using the MDS approach to assessing interaction, the present data display $\theta = 40$ degrees, the earlier data (when reanalyzed with MDS), $\theta = 47$ degrees (± 11). A better comparison may be with only the Below data ($\theta = 26$ degrees) or only the Below, Zero versus Heavy comparisons ($\theta = 19$ degrees), as the stimuli used to obtain the earlier data had the nasal pole/zero complex below the oral pole and a nasalization difference similar to that between Zero versus Heavy. Either comparison clearly supports the major conclusion about the direction of interaction. Under the reanalysis, the two studies also agree that the extent of integrality is greater for vowels followed by a nasal rather than oral consonant. In the present data, a nasal consonant decreases θ from 45 to 34 degrees overall, 63 to 43 degrees in the Above conditions, 27 to 25 degrees in the Below conditions; for the Kingston and Macmillan data, under the revised analysis, the shift is from 69 degrees (± 8) to 25 degrees (± 14).

The form of the interaction in the Above and Below sets is qualitatively different. For the Above data, it is clear that one dimension is related to perceived height, and perceived nasalization contributes a separate effect. As noted earlier, the second dimension for the Below data seems primarily to distinguish stimuli *CH* and *AH* from the other five stimuli in each set.

There was no interaction between High versus Low range of F_1 and the location of the nasal complex. According to Maeda's (1993) model, listeners expect a shift in the frequency of the nasal complex relative to F_1 in the range 360–500 Hz, but we obtained no evidence for such a shift.

b. Parallelogram models. In our past work we used a different method to assess degree of interaction from similar data. Kingston and Macmillan (1995) and Kingston *et al.* (1997), using only 2×2 arrays, averaged d' values for opposite sides and fit a parallelogram to points corresponding to the four stimuli. The degree of interaction was estimated as an interior angle of this figure. For noiseless data that actually *do* form a parallelogram (or a trapezoid), this method is equivalent to the one used here, so the present method can be viewed as an extension and generalization of the previous one. To compare the two techniques for real data, we applied both to data from experiment I, and data from Kingston and Macmillan (1995).

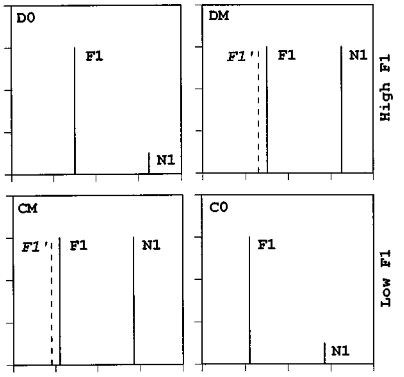
For the present data, the correlation between the two values of θ was 0.80, but values of θ obtained from parallelogram analysis were lower than those obtained from MDS by an average of 32 degrees. For the Kingston and Macmillan (1995) data, values of d' representing single-dimension and correlated task performance by each of the eight listeners were submitted to INDSCAL, and θ values were calculated from the group solutions according to the method described earlier. The values are: 8 degrees (±3) before [n], 42 degrees (±22) before [m], 79 degrees (±13) before [d], and 60 degrees (±10) before [b]. On average, these figures are 24 degrees higher than those found with the parallelogram method [10 degrees (±16), 28 degrees (±18), 32 degrees (±30), and 22 degrees (±30), respectively], although the figure for [n] context is lower. Analyzed this way, the data show that integrality is stronger before nasal ($\theta = 25$ degrees) than oral ($\theta = 69$ degrees) consonants, as in the present experiment.

The discrepancy between the two methods probably arises because the parallelogram model gives too much weight to fitting large d' values accurately. The MDS assumes no specific geometric arrangement, and is thus more general. In any case, the positive relation between values of θ obtained by the two methods means that most qualitative conclusions are unaffected by the choice of technique.[7]

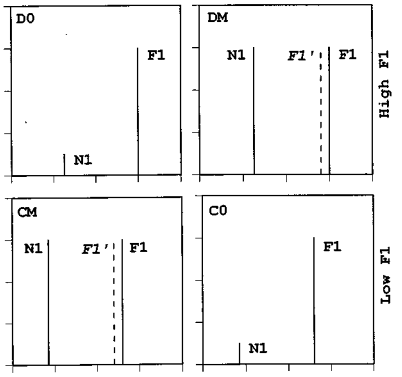
2. Psychoacoustic analysis

The major psychoacoustic questions raised by this experiment are: What aspects of the stimuli are used by the listeners in making their classifications? What type of processing led to the patterns of perceptual interaction displayed in Fig. 5 and the θ values calculated from them? We consider the related but different question of the physical correlates of speech *categories* in the discussion following Experiment II.

a. Center-of-gravity. For each pair of vowels that was discriminated in the experiment, we computed differences in the COGs of the ERB rate spectra (taken from Fig. 4). The correlations between COG differences and d_a, for the Above and Below stimulus sets and for the [d] and [n] contexts, averaged 0.72. About half the variance in d_a values is thus accounted for COG differences, suggesting that listeners rely heavily on this stimulus characteristic in classifying these vowels.



(a) ABOVE



(b) BELOW

FIG. 6. Schematic spectra of stimuli *D0, DM, C0,* and *CM*. Solid lines represent the frequency and amplitude of F_1 and N_1, the dashed line the frequency and amplitude of perceived $F_1(F_1')$, which is lowered to compensate for the raising produced by nasalization.

Although previous work on nasalization has suggested that COG over the low-frequency region plays an important role in phonetic labeling (see the discussion following Experiment II), we are aware of no previous work investigating its role in fixed classification. The COG of our stimuli increases with F_1 and decreases with *N*, so that COG differences are greater for negatively than for positively correlated-stimulus pairs. The effect is visible in Fig. 4, where the negatively correlated stimuli in each 2×2 subset are connected by dashed lines and the corresponding positively correlated stimuli by dotted lines. The difference in COG (the vertical discrepancy, not the length of the line) is uniformly greater for the negatively correlated pairs.

To see that these two effects predict the negative integrality found in our data, focus on the stimulus subarray $\{D0,DM,C0,CM\}$. The COG values for these four stimuli are 552, 531, 511, and 493 Hz in the Above condition and 552, 508, 511, and 452 Hz in the Below condition. The difference between *DM* and *C0* (20 Hz Above and 3 Hz Below) is much less than that between *D0* and *CM* (59 Hz Above and 100 Hz Below), and the discriminabilities follow the same pattern: For the Above stimuli, $d' = 2.51$ vs 0.43 before [d] and 2.66 vs 0.64 before [n]; for the Below stimuli, d_a

$=2.18$ vs 0.62 before [d] and 2.25 vs 0.73 before [n]. This asymmetry is captured, in the data analysis, by the statistic θ.

Thus there is evidence that COG is important in vowel classification. However, if decisions were based entirely on COG, the MDS representations in Fig. 5 would be one-dimensional, which they are not.

b. Perceived F_1. Diehl *et al.* (1990) suggested that judgments of both height and nasalization depend on perceived F_1, and that nasalization modifies this perception. In natural speech, the presence of nasalization increases F_1. If, as in our stimulus set, perceptible nasalization is added and F_1 stays the same, the listener may lower the perceived value of F_1 in compensation. That such a process could account for the integrality we observed is illustrated in Fig. 6, which displays schematic spectra (cf. Fig. 3) for stimuli *D0*, *DM*, *C0*, and *CM*. (The four spectra have been rearranged to highlight the key comparisons, *D0* vs *CM* and *DM* vs *C0*.) The dashed lines labeled F_1' represent the perceived location of the first formant after compensation for the raising effects of nasalization. The compensation process increases the *D0-CM* difference and reduces the *DM-C0* difference, and this effect occurs for both the Above and the Below stimuli. Baseline differences (horizontal and diagonal comparisons, in the figure) are unaffected.

The Diehl *et al.* hypothesis also makes a prediction about pairs of stimuli in which the covariation of F_1 and N does conform to listeners' expectations. With such a stimulus set, performance on the two correlated tasks should be more equal and the two stimulus dimensions should integrate less. Kingston and Macmillan (1995) used such a stimulus set. In their experiment II, stimuli were ''rotated'' 45 degrees in the stimulus space so that the A_{45}–D_{45} stimulus pair differed in F_1 but not N and the B_{45}–C_{45} stimulus pair differed in N but not F_1. Compensation for the expected effects of nasalization should have lowered perceived F_1 for B_{45} considerably relative to C_{45} because B_{45} was so heavily nasalized. On the other hand, the perceptual distance between A_{45} and D_{45} should not have been affected by compensation, as they were both equally nasalized. These effects should at least have equalized (and perhaps reversed) the differences in performance on the correlated tasks, compared to those obtained with the original unrotated stimulus array. Our MDS reanalysis of the degree of interaction in those experiments confirms these expectations: $\theta=100$ degrees (95% confidence interval ± 18) overall for the rotated data, 92 degrees (± 20) before nasal consonants and 108 degrees ($\pm$ 18) before oral consonants. These values are all close to 90 degrees, indicating that performance was nearly equal on the two correlated tasks.

Compensating for the expected raising of F_1 by nasalization is a top-down process that depends on separating nasalization from F_1 perceptually. Comparing the COG values of two stimuli, on the other hand, is an entirely bottom-up process. We reconsider the effects of COG and perceived F_1 after describing experiment II, and argue that these factors are also implicated in vowel identification.

III. EXPERIMENT II: TRADING RELATIONS

In this trading-relations experiment, listeners judged a vowel height continuum while vowel nasalization was varied (as in Krakow *et al.*, 1988), and a nasalization continuum while F_1 was varied. Both judgments were made on stimuli in which the following consonant's nasality was varied orthogonally. Two sets of results are reported here, one obtained from the same listeners who had earlier participated in experiment I and another from naive listeners. The two groups are compared to determine the effects of this prior experience.

A. Methods

1. Stimuli

The vowel array was a finer subdivision of that used in experiment I. Seven equally spaced values spanned the total ranges of F_1 and of N (see Table I). Thus, the range of F_1 was 360–500 Hz, with 23–24-Hz intervals between adjacent F_1 values, and the range of N was 0–90 Hz, with 15-Hz intervals. Otherwise, the stimuli were constructed in exactly the same way as in the previous experiment, except that the initial consonant was always [b]. Four 7×7 $F_1\times N$ arrays were constructed by varying the final consonant ([d] or [n]) and the location of the nasal complex (Above or Below F_1).

2. Subjects

Listeners who participated in experiment I remained in the same condition, Above or Below. Six additional listeners heard the Above stimuli, and six more the Below stimuli. The listeners who had participated in experiment I will be referred to henceforth as ''trained'' listeners, those who did not as ''untrained.''

3. Procedures

For the trained listeners, experimental blocks of trials began with 28 ordered orientation trials, which stepped through the relevant dimension, alternating between the lowest and highest values of the orthogonal dimension and between Oral and Nasal following consonants (7 stimulus values$\times$2 following consonants$\times$2 endpoints). The orientation trials were followed by 98 randomized test trials, one for each stimulus in the $2\times 7\times 7$ array. Listeners gave one of two responses: ''U'' versus ''O'' (henceforth ''high'' versus ''mid'') if the relevant dimension was F_1, and ''oral'' versus ''nasal'' if the relevant dimension was N. They had 2000 ms to give their response after hearing the stimulus and 1500 ms before the next trial began. Blocks alternated between F_1 and N judgments, a total of 15 blocks for each dimension in the Above condition and 17 in the Below condition. Two 2-h sessions over two days were needed. Other procedural details were as in experiment I.

A slightly different orientation procedure was used for the untrained listeners. Before having to categorize the full stimulus continua, they were presented with one or two blocks of trials in which the 28 endpoint stimuli were presented in random order. The listeners responded ''high'' versus ''mid'' or ''oral'' versus ''nasal'' and were given trial-by-trial feedback. One orientation block of this kind was

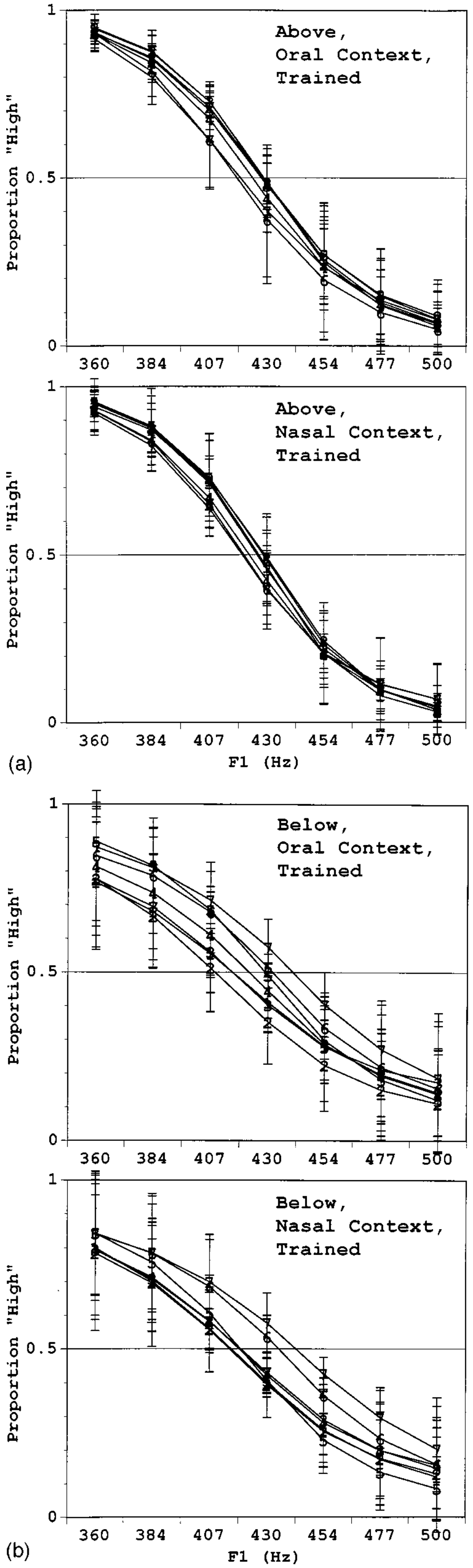


FIG. 7. Identification functions for trained listeners responding ‘‘high’’ as F_1 varies (experiment II). In (a), the nasal complex is Above F_1, in (b) Below. The parameter is *N*; the symbol 1 indicates 0 Hz, 7 is 90 Hz, and the other numerals are spaced at 15-Hz intervals between these values. In each panel, responses to vowels preceding an Oral consonant are displayed at the top, responses to vowels preceding a Nasal consonant at the bottom.

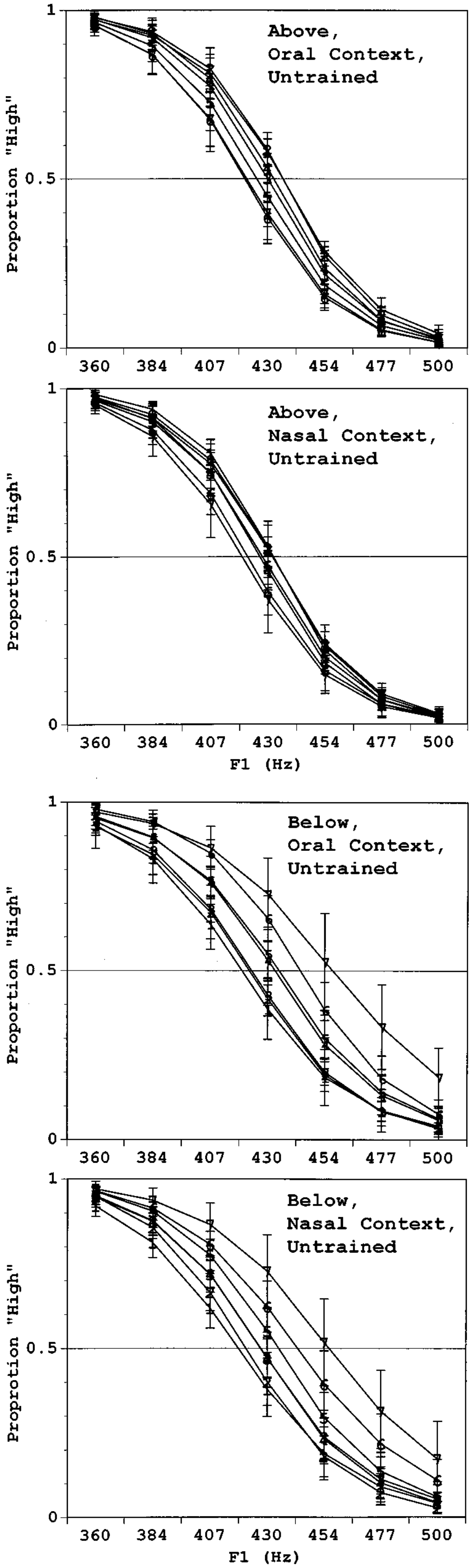


FIG. 8. Identification functions for untrained listeners responding ‘‘high’’ as F_1 varies (experiment II). See Fig. 7.

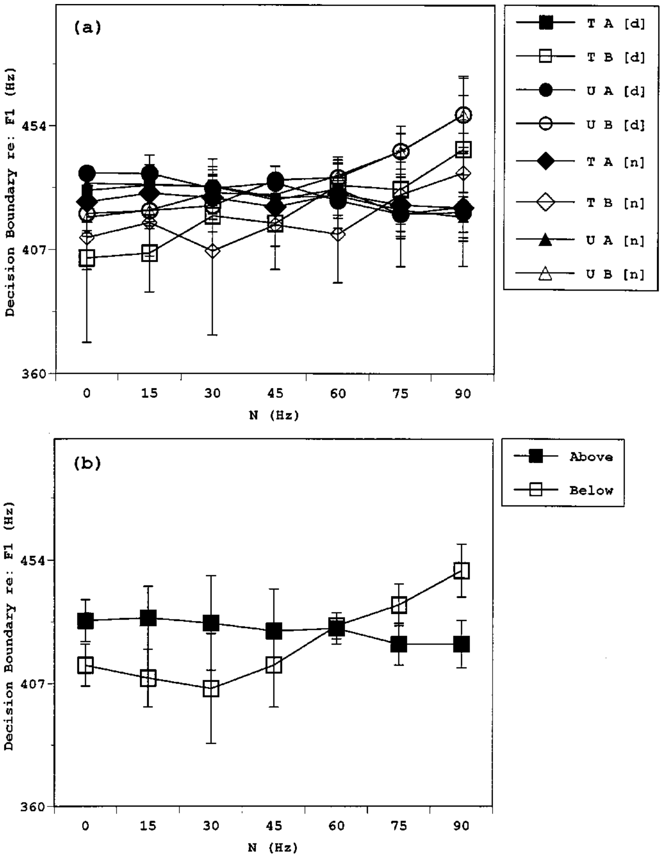


FIG. 9. Crossover (50% responding points) for ‘‘high’’: ‘‘mid’’ judgments, from the F_1 identification functions in Figs. 7 and 8. (a) Separate lines and plotting figures are used for each combination of Trained (T) and Untrained (U) listeners, Above (A) and Below (B) placement of the nasal complex, and Oral ([d]) and Nasal ([n]) consonant conditions. (b) The average result for Above and Below. Bars are 95% confidence intervals.

sufficient for ‘‘high’’ versus ‘‘mid’’ categorization, two for ‘‘oral’’ versus ‘‘nasal’’ categorization. The procedure was thus the same as for trained listeners, except that untrained listeners categorized the stimuli for ‘‘high’’ versus ‘‘mid’’ or ‘‘oral’’ versus ‘‘nasal’’ in three successive blocks before switching to the other categorization. Between 13 and 16 such blocks were run for both categorizations in 2-h time periods on different days.

Because of occasional failure to respond in the allotted time, the results reported here for trained listeners are based on an average of 14 out of the possible 15 judgments per stimulus per listener for each dimension in the Above condition and an average of 15–16 of the possible 17 judgments in the Below condition. Results for untrained listeners are based on an average of 13–16 responses in the Above condition and an average of 15 responses in the Below condition.

B. Results

1. ‘‘High’’ judgments

The average proportion of ‘‘high’’ judgments is plotted against F_1 in Fig. 7 for trained listeners and in Fig. 8 for untrained listeners. The multiple panels reflect the Above/Below distinction and Oral versus Nasal context, and the separate functions in each panel are for the seven possible values of nasalization. Logistic psychometric functions fitted to the response frequencies were used to estimate a crossover point, or category boundary—the stimulus value judged 50% of the time in each category—and a slope.

Category boundaries for ‘‘high’’ judgments are plotted in Fig. 9(a). In the Below condition, the F_1 crossover point increases as a function of *N* by 35–40 Hz, but in the Above condition it instead decreases by about 10 Hz. A repeated measure ANOVA was run in which Trained versus Untrained and Above versus Below were between-subjects variables and *N* and following consonant were within-subjects variables. The only significant main effect was *N* $[F(6,96) = 3.28,\ p = 0.006]$ and the only significant interaction was $N \times$ Above/Below $[F(6,96) = 6.84,\ p < 0.001]$. Crossover points decrease slightly and then increase noticeably in the Below condition as a function of *N*, whereas they decrease steadily in the Above condition [Fig. 9(b)].

To abstract sensitivity measures for ‘‘high’’ judgments, we examined the slopes of the psychometric functions in Figs. 7 and 8; these are plotted in Fig. 10(a). The slopes are steeper for the Untrained than the Trained listeners, and in the Above than the Below condition. In a repeated measures ANOVA using the same independent variables as in the analysis of crossover points, only the main effects of the between-subjects variables, Trained versus Untrained and Above versus Below, reached significance [Trained versus Untrained: $F(1,16) = 5.07$, $p = 0.039$; Above versus Below: $F(1,16) = 5.30$, $p = 0.035$]; see Fig. 10(b). The interaction between following consonant and Above versus Below approached significance $[F(1,16) = 3.66,\ p = 0.074]$, reflecting steeper slopes before nasal [n] Above but before oral [d] Below.

2. ‘‘Oral’’ judgments

The average proportion of ‘‘oral’’ judgments for trained and untrained listeners is plotted against *N* in Figs. 11 and 12. The multiple panels reflect the Above/Below distinction and Oral versus Nasal context, and the separate functions in each panel are for the seven possible values of F_1. Logistic psychometric functions could not always be fit reliably to the response frequencies, because in many instances (especially for the untrained listeners) the psychometric functions changed little as a function of *N*. Analysis of the ‘‘oral’’ judgments is based instead on the average proportions of ‘‘oral’’ responses across all seven *N* values for each value of F_1 and final consonant.

Mean proportions of ‘‘oral’’ judgments across *N* values for each F_1 value and following Consonant are plotted in Fig. 13(a).[8] ‘‘Oral’’ responses increase with F_1 generally, but this effect is relatively small for the untrained listeners in the Above condition. In a repeated measures ANOVA using the same independent variables as in the previous analyses, the only significant main effect was for F_1 $[F(6,96) = 63.82,\ p < 0.001]$. F_1 also interacted significantly with Above versus Below $[F(6,96) = 3.28,\ p = 0.006]$ and with the interaction between Above versus Below and Trained versus Untrained $[F(6,96) = 4.27,\ p < 0.001]$; see Fig. 13(b).

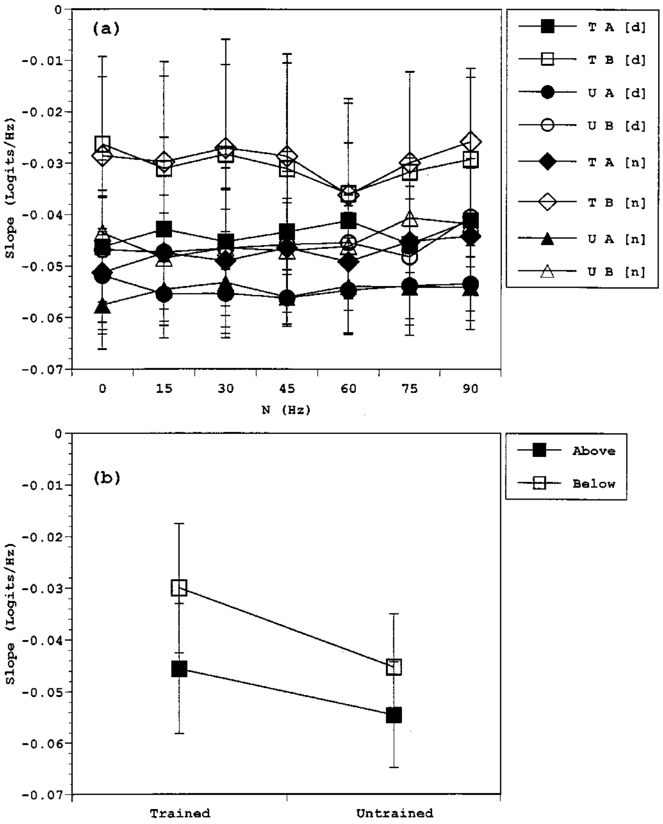


FIG. 10. Slopes for ''high'':''mid'' judgments, from the F_1 identification functions in Figs. 7 and 8. (a) Separate lines and plotting figures are used for each combination of Trained (T) and Untrained (U) listeners, Above (A) and Below (B) placement of the nasal complex, and Oral ([d]) and Nasal ([n]) consonant conditions. (b) The average result for Trained and Untrained listeners. Bars are 95% confidence intervals.

C. Discussion

The primary interest in these results is in the trading relations they display between F_1 and N, and in the effect of following-consonant nasality on these relations. A preliminary question concerns the ways in which experience in fixed classification and consonant identification (experiment I) affected these results. Comparison of the trained and untrained listeners shows two effects: (1) For ''high'' judgments, untrained listeners were *more* sensitive to F_1 differences than trained listeners, and sensitivity in the Above condition exceeded that in the Below condition more for the trained than the untrained listeners. (2) For ''oral'' judgments, untrained listeners were relatively insensitive to N differences in the Above (but not in the Below) condition. The overall pattern of results was otherwise very similar.

1. Height judgments

The likelihood of a ''high'' judgment depends on N in the Below condition, with more nasalized vowels more likely to be judged ''high'' [Figs. 7(b) and 8(b)]. On the other hand, more nasalized vowels are less likely to be judged ''high'' in the Above condition [Figs. 7(a) and 8(a)]. Increases in N produce decreases in COG (see Fig. 4) for both the Above and Below stimuli and, according to the hypothesis that COG is the mediating variable, more ''high'' judgments. The COG hypothesis is thus consistent with the Below but not the Above data.

A different hypothesis can account for the discrepancy between the Above and Below data. Suppose height judgments depend on the intensity of the harmonics on the upper skirt of the lowest spectral prominence, as well as on its COG. [This hypothesis is consistent with Assmann's (1985) finding that height judgments vary much more with the intensity of harmonics just Above F_1 than those just Below it.] These harmonics will become more intense as N increases for the Above but not the Below stimuli, because N_1 is Above rather Below F_1. This increase on the upper skirt can easily be seen by comparing panels within a row in Fig. 3(a). Above, this effect will oppose that of decreasing COG and may even lead listeners to mistake N_1 for F_1, thus reducing ''high'' responses as N increases. In the Below condition, on the other hand, decreasing COG and mistaking N_1 for F_1 cooperate, markedly increasing ''high'' responses as N increases.

Krakow and her colleagues have conducted extensive research on the nasalization–height interaction using the trading relations paradigm, but their use of an articulatory synthesizer renders their studies hard to compare with ours. For example, Krakow *et al.* (1988) used a continuum from ''mid'' [ɛ] to ''low'' [æ], and found that height judgments shifted towards ''low'' when vowels were nasalized. This is superficially the opposite of our result, but because F_1 increased in concert with N for their stimuli, the two results are quite consistent.

2. Nasalization judgments

The likelihood of an ''oral'' judgment depends strongly on F_1 in both the Above and Below conditions, with stimuli judged ''high'' more likely also to be judged ''nasal'' and stimuli judged ''mid'' more likely to be judged ''oral'' (Figs. 11 and 12). Further, sensitivity to N differences in making ''oral'' versus ''nasal'' judgments depends nonmonotonically on F_1. As a comparison with Figs. 7 and 8 shows, the effect of F_1 on ''oral'' judgments was much stronger than the effect of N on ''high'' judgments. The F_1 range is a little over 1.5 times as large as the N range, but changes response proportions by eight to nine times as much. Across the 140-Hz F_1 range, ''oral'' responses shift from 0.23 to 0.73 on average, whereas across the 90-Hz N range, ''high'' responses shift only from 0.45 to 0.50. Also unlike the effect of N on height, this interaction was equally large in the Above and Below conditions, at least for the trained listeners.

All of these asymmetries may follow from the fact that height but not nasalization is phonologically contrastive for American English vowels. In making nasalization judgments, our listeners were likely to have been judging the stimuli as much if not more in terms of their perceived vowel height as their perceived nasalization, but in making vowel height judgments, little if any mirror image effect was likely.

Speakers of Indic languages, in which nasalization is contrastive, respond more categorically to this feature than American English listeners (Beddor and Strange, 1982, for Hindi; Hawkins and Stevens, 1985, for Hindi, Gujerati, and Bangali). Hawkins and Stevens also report that Gujerati speakers showed the greatest degree of categoricalness, fol-

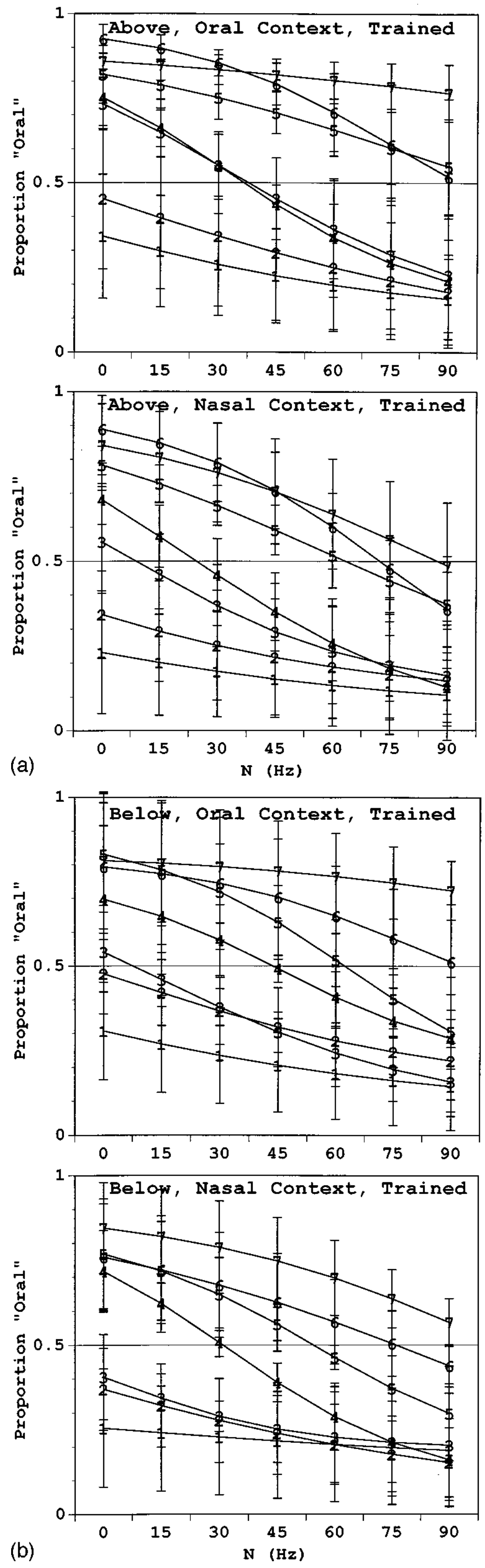


FIG. 11. Identification functions for trained listeners responding ''oral'' as N varies (experiment II). In (a), the nasal complex is Above F_1, in (b) Below. The parameter is F_1; the symbol 1 indicates 360 Hz, 7 is 500 Hz, and the other numerals are spaced at 23- to 24-Hz intervals between these values. In each panel, responses to vowels preceding an Oral consonant are displayed at the top, responses to vowels preceding a Nasal consonant at the bottom.

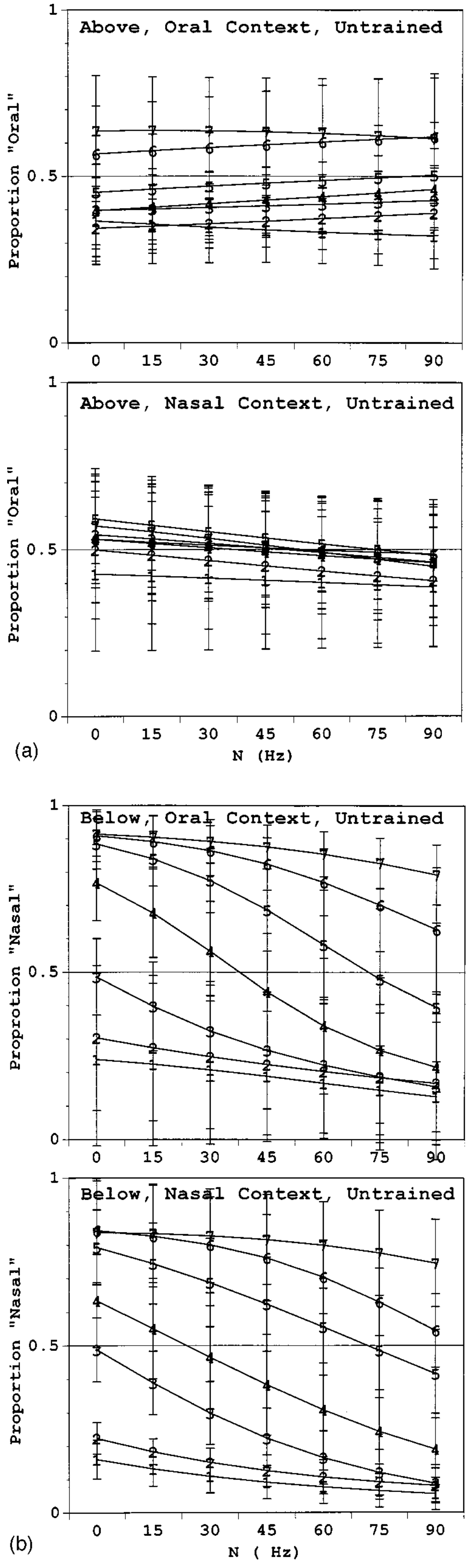


FIG. 12. Identification functions for untrained listeners responding ''oral'' as N varies (experiment II). See Fig. 11.

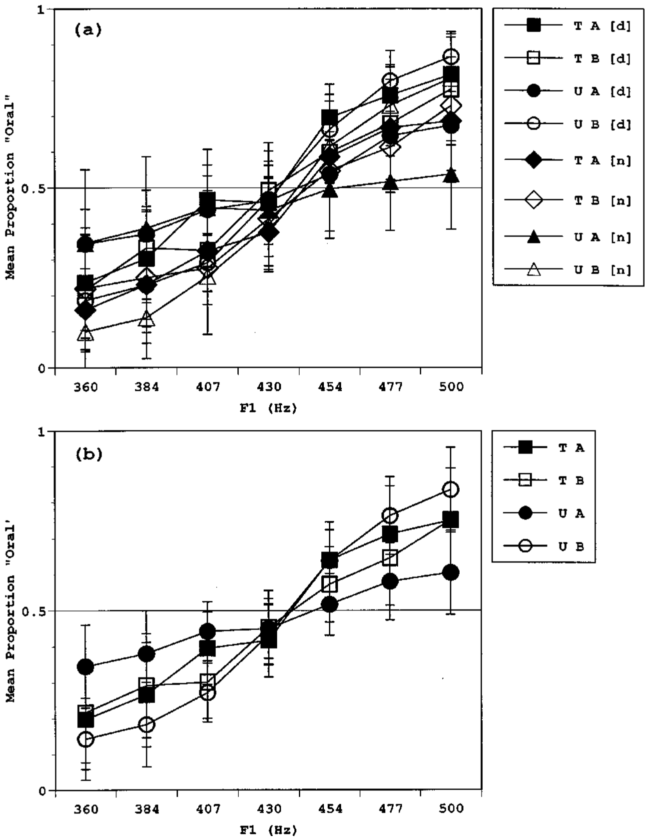


FIG. 13. Proportion of ''oral'' judgments, from the *N* identification functions in Figs. 11 and 12. (a) Separate lines and plotting figures are used for each combination of Trained (T) and Untrained (U) listeners, Above (A) and Below (B) placement of the nasal complex, and Oral ([d]) and Nasal ([n]) consonant conditions. (b) The average result for each combination of training and nasal complex placement. Bars are 95% confidence intervals.

lowed by Hindi and Bengali. Hindi and Gujerati listeners discriminated the intermediate nasalization values better than the extremes for all of [i, u, ɑ, e, o]. For American English listeners, Hawkins and Stevens found that vowels were discriminated equally well across the nasalization continuum when the vowel was [i, u, ɑ] but best with intermediate nasalization values when the vowel was [e, o]. A mid-range peak in nasalization discrimination has thus been consistently observed for mid vowels, both in our experiments (F_1 values of 360–500 Hz) and those of Hawkins and Stevens (400 Hz for [e], 430 Hz for [o]), suggesting that our listeners and theirs responded in terms of the same perceptual effects.

Hawkins and Stevens also observed that whereas the high and low vowels [i, u, ɑ] showed a discrimination peak near the crossover point for categorizing these vowels as oral versus nasal, the mid vowels [e,o] showed a peak at the point where N_1 separates from F_1 spectrally. They suggested that this separation causes listeners to hear a different vowel quality, an interpretation compatible with the conjecture that our listeners used vowel height percepts as much or more than nasalization percepts in identifying vowels as ''oral'' versus ''nasal.'' In Hawkins and Stevens's high and low vowel stimuli, [i, u, ɑ], the nasal pole/zero complex was always well separated from the oral pole, so increasing nasalization would not bring about a change in vowel quality. Further evidence that nasalization is perceived differently in mid than high or low vowels can be found in Beddor and Hawkins (1990). These authors report that in matching nasalized to oral vowels, American English listeners gave greater weight to F_1 than COG for high and low vowels, [i, u, æ, ɑ], but more or less equal weight to both F_1 and COG for mid vowels, [e, o].

3. Consonantal context

Contrary to past findings, our data do *not* reveal an interaction between perceived vowel height and the nasality of the following consonant. In Krakow *et al.* (1988), the boundary shift in height judgments with *N* did not occur when the following consonant was nasal, whereas in our data, the shift was equally strong before a nasal as an oral consonant. According to Kingston and Macmillan's (1995) reanalysis, Krakow *et al.* also found greater sensitivity when the vowel and consonant agreed in nasality: F_1 functions were steeper with higher *N* for vowels followed by a nasal consonant, but steeper with lower *N* for those followed by a oral consonant. No such slope differences are found in vowel height judgments reported here; instead, the Above data show consistently greater sensitivity to F_1 differences for all *N* values before both oral and nasal following consonants than the Below data, and this difference is greater for trained than untrained listeners [Fig. 12(a) and (b)].

There is a weak interaction between the following consonant's nasality and the nasal pole/zero complex's position relative to the oral pole: Sensitivity to F_1 is greater before [n] Above, but before [d] Below. Experiment I showed that F_1 integrates less with nasalization Above than Below. If lesser integration means that a vowel is more likely to sound nasalized for a given *N* value Above than Below, then in this respect our results correspond to those of Krakow *et al.* in showing greater sensitivity to F_1 when the vowel and consonant agree in nasality. Two caveats are, however, in order: (1) the effect observed in our data is at best marginally significant, and (2) the interaction is not between *N* and consonant nasality, but between the separability of the two dimensions and consonant nasality.

In experiments using a matching paradigm to study vowel nasalization judgments, Krakow and Beddor (1991) showed that naturally produced vowels are more reliably identified as nasalized in isolation and between oral consonants, rather than between nasal consonants, a context in which their nasalization could be coarticulatory. The psychometric functions for nasalization judgments in our data, on the other hand, do not differ between oral and nasal contexts in either the Above or Below conditions. Instead, sensitivity to *N* differences is greater for intermediate than extreme F_1 values before both oral and nasal consonants, both Above and Below (Figs. 11 and 12). Differences in how the consonant nasality contrast was implemented may account for these disparate effects of context. Krakow and Beddor compared m_n and b_d contexts, Krakow *et al.* (1988) b_nd and b_d contexts, and our experiments C_n and C_d. Again, however, the discrepancy may involve COG and perceived F_1: when F_1 was high it raised COG and perceived F_1 enough nearly to overwhelm the contrary influence of increasing *N* on these percepts, and vice versa. As a result, responses crossed over more sharply when the orthogonal variable, here F_1, had intermediate values.

IV. GENERAL DISCUSSION: SENSORY AND DECISION PROCESSES IN PHONETIC IDENTIFICATION

Fixed classification and identification provide, we have argued, rather different information about vowel perception. In particular, the mean-integrality found in experiment I is logically unrelated to the boundary shifts observed in experiment II. However, the two tasks must ultimately tap the same information. In this section we attempt to describe the primary identification results (boundary shifts and sensitivity pattern) in terms of a common perceptual space, of the sort that was derived from fixed classification data.

To start, we return to Fig. 1, which offered some alternative interpretations of Repp's (1982) concept of cue trading in perceptual-space terms. The results of experiment I showed that most 2×2 stimulus subsets were represented by quadrilaterals that were not rectangular ($\theta = 90$ degrees), but rather displayed a negative correlation between perceived F_1 and perceived N (θ averaged 53 degrees in the Above condition, 28 degrees Below). Of the possibilities outlined in the Introduction, this arrangement most resembles that of Fig. 1(c), an elaborated version of which is shown in Fig. 14 for the $\{C0,D0,CM,DM\}$ corner of the perceptual space.

The implications of this pattern for identification (experiment II) depend on the listener's decision boundary. Suppose this boundary is orthogonal to the perceived F_1 axis, as in Fig. 14(a). The boundary shift is the difference in proportion of ‘‘high’’ judgments for the nasalized and unnasalized stimuli, the vertical difference between two identification functions in graphs like Fig. 7. Expressed in z-units,

$$\text{boundary shift} = z[P(\text{``high''}|CM)] - z[P(\text{``high''}|C0)]. \tag{1}$$

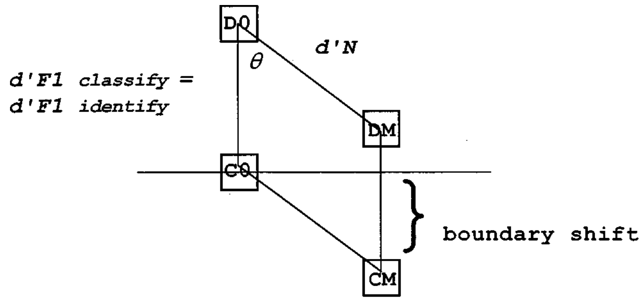
For the trained observers, we calculated the average boundary shift by this method, with the results shown in Table IV. For F_1 judgments with the Above stimuli, the negative value indicates *a reduced* tendency to say ‘‘high’’ as N increases; the positive values in the other cases indicate positive interactions between these variables.

From the geometry of Fig. 14(a), the magnitude of shift predicted from fixed classification can be expressed[9] in terms of d'_N (sensitivity to N), and θ:

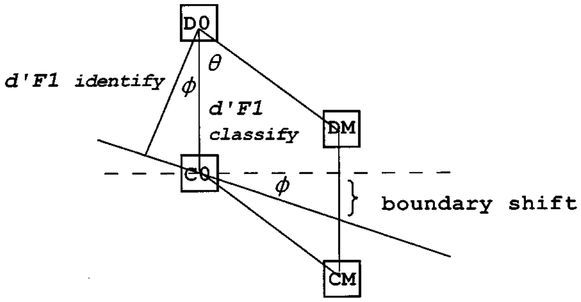
$$\text{boundary shift} = d'_N \cos(\theta). \tag{2}$$

These values are also given in Table IV, and it is clear that the observed boundary shifts are smaller than predicted for both F_1 and N judgments.

The assumption of an orthogonal decision rule also makes a prediction about sensitivity in the identification task. The observer's ability to discriminate two stimuli in this condition can be estimated, for F_1 judgments, by comparing the proportion of ‘‘high’’ responses to the stimuli:



(a) Orthogonal decision boundary



(b) Diagonal decision boundary

FIG. 14. A corner of a perceptual space that displays perceptual integrality. The vertical line from *D0* to *C0* is the optimal decision axis for a listener identifying stimuli by their F_1 values. The listener in (a) adopts a decision boundary orthogonal to this axis, obtains a sensitivity $d'(F_1)$ for discriminating these stimuli equal to that obtained in a fixed classification task, and displays a boundary shift, that is, will respond unequally to stimuli *C0* and *CM*. The listener in (b) adopts a nonorthogonal decision boundary shifted by an angle ϕ from the orthogonal one, obtains a lower sensitivity $d'(F_1)\cos(\phi)$ for discriminating *D0* and *C0*, and displays a smaller boundary shift.

$$d'_{F_1}(\text{identify}) = z[P(\text{``high''}|D0)] - z[P(\text{``high''}|C0)]. \tag{3}$$

Because the decision boundary is perpendicular to the perceived height dimension, this d' should be equal to that obtained in fixed discrimination. In fact, as Table IV shows, d' values estimated from experiment II are about half as large as the corresponding estimates from experiment I.

Both of these discrepancies between the two experiments, the prediction of too large a boundary shift and lower sensitivity in experiment II than in experiment I, can be addressed by modifying a single assumption. As shown in Fig. 14(b), let us allow the decision boundary to be a straight line that intersects the optimal boundary at a nonzero angle ϕ. The decision boundary now depends on *both* variables, a natural geometric interpretation of ‘‘cue trading.’’ Values Below the boundary are greater in both F_1 and N than corresponding values Above it. As Fig. 14(b) shows, the use of this diagonal decision rule produces a smaller boundary shift than did the orthogonal rule, namely,

TABLE IV. Observed sensitivity (experiments I and II) and boundary shifts (experiment II), together with predictions of orthogonal and nonorthogonal decision boundaries.

Dimension judged	Stimulus set	Fixed d_a	Identification d_a	θ (degrees)	Boundary shift (z-units) Observed	Boundary shift (z-units) Predicted by Eq. (2)	ϕ (degrees)	Identification d_a predicted by Eq. (5)
F_1	Above	1.23	0.92	53	-0.15	0.56	44	1.08
	Below	1.52	0.64	28	0.30	1.12	54	0.89
N	Above	0.93	0.46	53	0.46	0.74	16	0.89
	Below	1.27	0.48	28	0.45	1.34	51	0.79

$$\text{boundary shift}_1 = d'_N[\cos(\theta) - \sin(\theta)\tan(\phi)]. \tag{4}$$

This reduces to Eq. (2) when $\phi = 0$ degrees.

Perceptual distances between stimuli are now measured between points projected onto the new decision axis, which is perpendicular to the decision boundary. Under the diagonal strategy for judging F_1, sensitivity in identification drops from d'_{F_1} to

$$d'_{F_1}(\text{identify}) = d'_{F_1}\cos(\phi). \tag{5}$$

The diagonal rule model, then, predicts that boundary shifts in identification will be accompanied by lower sensitivity than in fixed classification, and thus accounts qualitatively for the important aspects of the data.

Because the model requires just one parameter to describe both results, it can be quantitatively evaluated. The parameter ϕ, estimated from Eq. (4), is given in Table IV. As Fig. 14(b) illustrates (for F_1 judgments, but the result is true for N judgments as well), the values obtained mean that listeners employ a decision boundary that depends on both F_1 and N. The last column of Table IV recalculates the predicted sensitivity in identification assuming the new decision boundary [Eq. (5)]. The average d' of 0.63 is still overpredicted—the model calculates 0.91—but the discrepancy is much less than for the orthogonal model, which predicts the same performance as in fixed classification, $d' = 1.24$. This remaining discrepancy is not entirely unexpected in view of the well-known observation that sensitivity is lower in tasks with large ranges of stimuli, like experiment II, than in fixed-classification tasks like experiment I (Braida and Durlach, 1972; Durlach *et al.*, 1989).

Let us summarize the conclusions to which these calculations have brought us. Identification functions obtained in trading-relations experiments provide two key pieces of information: a slope, which reflects sensitivity, and an intercept. If judgments of one variable actually depend on each of two variables, as is often postulated, slopes will decrease and intercepts will shift. That slopes have decreased can be determined by a converging task, fixed classification. Our data, which display this phenomenon, provide support for a psychophysical model of trading-relations effects. More generally, they show how multiple tasks, together with a model of the processes they require, can be used to explore a single perceptual model in a way that no single task can.

V. SUMMARY, CONCLUSIONS, AND A PROMISSORY NOTE

In this article, we asked whether there was a perceptual interaction between N and F_1 in vowel perception, and if so how it could be characterized. The tentative answer we have reached is that there are two distinct interactions in our data. A fixed classification experiment showed perceptual integrality of N and F_1: Increasing N or decreasing F_1 led to correlated changes in an underlying perceptual space. This is an interaction in sensitivity. A trading-relations experiment showed that judgments of N depend on both N and on F_1, and judgments of F_1 depend on largely on F_1 but to a lesser degree on N. This is an interaction in the decision process used by our observers.

We have presented a psychophysical analysis that accounts for some aspects of the data, particularly the relation between the fixed classification and identification data, and have also provided an entry point to a psychoacoustic analysis of the stimulus correlates of these perceptual outcomes. A companion paper (Kingston *et al.*, in preparation) describes yet another experiment on the interaction of N and F_1 that greatly expands the psychoacoustic analysis, and relates those results to the present ones.

ACKNOWLEDGMENTS

This work was supported by NSF Grant No. DBS 92-12043 and a grant from the PSC-CUNY award program to the first author, and NIH Grant Nos. RR07048-26 and DCO1708 to the second author. We thank our colleagues in the University of Massachusetts Phonetics Laboratory, Cecilia Kirk and Elliott Moreton, for their helpful comments on earlier drafts. Patrice Beddor, Paul Iverson, James Sawusch, and an anonymous reviewer played their roles well.

APPENDIX: STIMULUS DETAILS

In Tables AI–AIII are listed the values of the synthesis parameters (in the form of time-value pairs) other than F_1 and N in the vowel: Table AI lists source parameters, Table AII formant frequencies, and Table AIII format bandwidths and the frequencies of the nasal pole and zero in final nasal consonants. All parameters are linearly interpolated between the target values listed in theses tables.

TABLE AI. Source parameters for synthesis.

Time (ms)	F_0 (Hz)	Time (ms)	AV		Time (ms)	OQ (%)		TL (−dB at 3 kHz)	
0	100	0	45		0	75		40	
80	100	80	45		70	75		40	
					Vowel				
100	125	90	60		80	50		0	
200	125	270	60		250	50		0	
Final consonant			b,d	m,n		b,d	m,n	b,d	m,n
280	100	280	45	54	280	75	50	40	14
355	100	355	45	54	355	75	50	40	14
390	100	390	0	0	390	75	50	40	14

TABLE AII. Formant frequencies for synthesis.

Time (ms)	F_1 (Hz)				F_2 (Hz)				F_3 (Hz)				F_4 (Hz)	
0	200				900				2100				3250	
70	200				900				2100				3250	
					Vowel									
110	Target				1025				2395				3250	
250	Target				1025				2395				3250	
Final consonant	b	d	m	n	b	d	m	n	b	d	m	n	b,m	d,n
280	260	260	220	240	840	1600	840	1240	2300	2795	2300	2590	3250	3750
285	260	260	220	240	840	1600	995	1140	2300	2795	2425	2590	3250	3750
390	260	260	220	240	840	1600	995	1140	2300	2795	2425	2590	3250	3750

TABLE AIII. Formant bandwidths and N_1 and N_0 values for synthesis. Other parameters were held constant throughout the stimuli: the bandwidths of N_1 and N_0 both $=90$ Hz, $F_5=4200$ Hz, and $B_5=1500$ Hz.

Time (ms)	B_1 (Hz)			B_2 (Hz)			B_3 (Hz)			B_4 (Hz)		N_1		N_0	
0	1000			1000			1000			1000		280		280	
70	1000			1000			1000			1000		280		280	
							Vowel								
110	100			70			90			200		Target		Target	
250	100			70			90			200		Target		Target	
Final consonant	b,d	m	n	b,d	m	m	b,d	m	n	b,d	m,n	b,d	m,n	b,d	m,n
280	1000	230	120	1000	150	250	1000	250	150	1000	200	280	310	280	450
390	1000	125	120	1000	150	250	1000	250	150	1000	200	280	310	280	450

[1] Generalized Recognition Theory has introduced the terms *perceptual separability* and *perceptual integrality* (Ashby and Townsend, 1986) for (non)interactions inferred from a perceptual space, reserving unmodified *separability* and *integrality* to be used in Garner's (1974) operational sense. Perceptual integrality includes changes in variance and correlation as well as mean, so *mean-shift* (or just *mean-*) *integrality* (Maddox, 1992) is a more precise term for the kind of interaction shown in Fig. 1(b).

[2] Values of d_a did not differ significantly with place of articulation of the following consonant, but were sometimes reliably smaller in the second than the first repetition. All cases involved N differences in the Below condition: for Zero versus Moderate N, d_a was 1.58 (±0.28) in the first repetition and 1.36 (±0.19) in the second; for Moderate versus Heavy N, the values were 1.92 (±0.29) and 1.60(±0.24). Neither place of articulation nor repetition interacted significantly with any other variable in any 2×2 subset.

[3] INDSCAL provides a group solution to the collected individual listeners' distance matrices, which are composed of the d_a values for all stimulus pairs in the High- and Low-F_1 ranges. Weights are also provided for each dimension of the group solution for each listener. The stress values for our representations are higher than those often reported in the literature, but comparison is difficult because most applications use (a) more than six stimuli, and (b) nonmetric algorithms. Several aspects of the data reassure us. First, within the Above and Below stimulus conditions, the plots for High- and Low-F_1 ranges are very similar. Second, plots for the individual listeners are very similar to plots based on averages. Third, we also conducted nonmetric analyses. Nonmetric representations are very similar to metric ones (justifying the treatment of d_a as a distance measure) and have similar stress values (which may therefore be high due to the small number of stimuli).

[4] This rather ''processed'' statistic is a natural one for our geometric representations. It directly reflects the most important qualitative aspect of the data, the discrepancy between accuracy in classifying positively and negatively correlated pairs. For parallelogram-shaped integrality [as, for example, in Fig. 1(b)], it is equivalent to the measure, also called θ, used by Kingston and Macmillan (1995).

[5] Individual data were derived by applying each listener's weights for each

dimension to the group solution produced by INDSCAL.

[6] A final finding is peripheral to the major questions being investigated: The degree to which integrality was stronger in the Below condition was larger for comparisons involving the stimuli with Heavy N values. The interaction between N difference and Above:Below was not significant, but a planned comparison of the contrast between the different pairings of N values was: F_1 and N integrated more when Zero N was paired with Heavy $N(\theta = 35$ degrees, $\pm 5)$ than when moderate N was paired with heavy N $[\theta = 47$ degrees, ± 8; $F(1,6) = 6.36$, $p = 0.045]$. The θ value for the pairing of Zero with Moderate N, 37 degrees (± 13), is similar to that of Zero versus Heavy N, but does not contrast significantly with that obtained for the Moderate versus Heavy pairing.

[7] Experiment I employed a paradigm that is often used to distinguish syndromes of interaction, integrality and separability (Garner, 1974). We did not use the complete Garner paradigm, which includes selective and divided attention, and our dependent measure was accuracy rather than the more common response time. Still, the data could be used to label the type of interaction as one of a small number of previously identified categories of perceptual interaction. The most direct approach, with our data, is to compare baseline d_a for one-step N and F_1 comparisons with the one-step correlated values. The result is that correlated d_a averages 0.59 units larger, 1.83 to 1.24. This ''correlated gain'' is larger for the Above conditions (0.78) than the Below (0.40), but occurs for both, and for both consonantal contexts, a pattern that marks the dimensions F_1 and N as ''integral.'' Examination of Fig. 5, however, makes clear that focusing on a ''correlated gain'' reverses the conclusion about the conditions in which F_1 and N integrate most, for θ values deviate more from 90 degrees in the Below than in the Above condition. There is no real conflict here: The traditional analysis is operational, whereas ours depends on characteristics of an inferred perceptual space. Another essential, but less easily quantified finding in this experiment is that the perceptual spacing in the two cases follows a different pattern; the traditional taxonomy does not distinguish among interactions that differ in this way.

[8] Mean proportions across the relevant stimulus continuum estimate response variability as a function of the orthogonal stimulus variables in much the same way as category boundaries. In a repeated measure ANOVA on mean ''high'' response proportions, a significant main effect of N $[F(6,90) = 5.02$, $p < 0.001]$ and a significant interaction between N and Above versus Below $[F(6,90) = 25.51$, $p < 0.001]$ were obtained, as in the comparable analysis of category boundaries. In addition, a significant interaction was obtained between N, Above versus Below, and Trained versus Untrained $[F(6,90) = 2.93$, $p = 0.012]$, reflecting the fact that Above and Below conditions differ more as a function of N in the responses of the Untrained than the Trained listeners.

[9] The formulas are slightly different for the positive integration case (height judgments in the Above condition).

Ashby, F. G., and Townsend, J. T. (**1986**). ''Varieties of perceptual independence,'' Psychol. Rev. **93**, 154–179.

Assmann, P. F. (**1985**). ''The role of harmonics and formants in the perception of vowel quality,'' Ph.D. dissertation, University of Alberta.

Beddor, P. S., and Hawkins, S. (**1990**). ''The influence of spectral prominence on perceived vowel quality,'' J. Acoust. Soc. Am. **87**, 2684–2704.

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

Braida, L. D., and Durlach, N. I. (**1972**). ''Intensity perception. II. Resolution in one-interval paradigms,'' J. Acoust. Soc. Am. **51**, 483–502.

Diehl, R. L., Kluender, K. R., and Walsh, M. A. (**1990**). ''Some auditory bases of speech perception and production,'' in *Advances in Speech, Hearing, and Language Processing*, edited by W. A. Ainsworth (JAI, London), Vol. 1, pp. 243–267.

Dorfman, D. D., and Alf, Jr., E. (**1969**). ''Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals–Rating method data,'' J. Math. Psychol. **6**, 487–496.

Durlach, N. I., Tan, H. Z., Macmillan, N. A., Rabinowitz, W. R., and Braida, L. D. (**1989**). ''Resolution in one dimension with random variations in background dimensions,'' Percept. Psychophys. **46**, 293–296.

Garner, W. R. (**1974**). *The Processing of Information and Structure* (Erlbaum Associates, Potomac, MD).

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hawkins, S., and Stevens, K. N. (**1985**). ''Acoustic and perceptual correlates of the nonnasal-nasal distinction for vowels,'' J. Acoust. Soc. Am. **77**, 1560–1575.

Kawasaki, H. (**1986**). ''Phonetic explanation for phonological universals: The case of distinctive vowel nasalization,'' in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, Orlando, FL), Vol. 88, pp. 81–103.

Kingston, J. (**1991**). ''Integrating articulations in the perception of vowel height,'' Phonetica **48**, 149–179.

Kingston, J., and Macmillan, N. A. (**1995**). ''Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm,'' J. Acoust. Soc. Am. **97**, 1261–1285.

Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R., and Bartels, C. (**1997**). ''Integrality in the perception of tongue root position and voice quality in vowels,'' J. Acoust. Soc. Am. **101**, 1696–1709.

Kingston, J., Macmillan, N. A., Dickey, L. W., Thorburn, R., and Bartels, C. (**in preparation**). ''Integrality of nasalization and F_1. III. Multinomial modeling of two-response identification.''

Klatt, D. H., and Klatt, L. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Krakow, R. A., and Beddor, P. S. (**1991**). ''Coarticulation and the perception of nasality,'' in *Proceedings of the XII International Congress of Phonetic Sciences* (Publications de L'Universite de Provence, Aix-en-Provence), Vol. 5, pp. 38–41.

Krakow, R. A., Beddor, P. S., Goldstein, L. M., and Fowler, C. A. (**1988**). ''Coarticulatory influences on the perceived height of nasal vowels,'' J. Acoust. Soc. Am. **83**, 1146–1158.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge U. P., New York).

Maddox, W. T. (**1992**). ''Perceptual and decisional separability,'' in *Multidimensional Models of Perception and Cognition*, edited by F. G. Ashby (Erlbaum, Hillsdale, NJ), pp. 147–180.

Maeda, S. (**1993**). ''Acoustics of vowel nasalization and articulatory shifts in French nasal vowels,'' in *Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic, San Diego), Vol. 5, pp. 147–167.

Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,'' Hearing Res. **28**, 209–225.

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

Rubin, P., Baer, T., and Mermelstein, P. (**1981**). ''An articulatory synthesizer for perceptual research,'' J. Acoust. Soc. Am. **70**, 321–328.

Stevens, K. N., Fant, G., and Hawkins, S. (**1987**). ''Some acoustical and perceptual correlates of nasal vowels,'' in *Festschrift for Ilse Lehiste*, edited by R. Channon and L. Shockey (Foris, Dordrecht, The Netherlands), pp. 241–254.

Swets, J. A., and Pickett, R. M. (**1982**). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory* (Academic, New York).