Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm

John Kingston, and Neil A. Macmillan

Citation: The Journal of the Acoustical Society of America **97**, 1261 (1995); doi: 10.1121/1.412169
View online: https://doi.org/10.1121/1.412169
View Table of Contents: https://asa.scitation.org/toc/jas/97/2
Published by the Acoustical Society of America

Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm

John Kingston[a)]
Linguistics Department, South College, University of Massachusetts, Amherst, Massachusetts 01003
Neil A. Macmillan[b)]
Psychology Department, Brooklyn College, CUNY, Brooklyn, New York 11210

(Received 22 July 1993; revised 2 September 1994; accepted 16 September 1994)

In vowel height contrasts, tongue height and soft palate height covary. A series of vowel classification experiments examined the perceptual interactions between F_1 and nasalization, the principal acoustic correlates of these articulations. Listeners classified imperfectly discriminable stimuli in the set of tasks that compose the Garner paradigm. Detection-theoretic models applied to the data led to the conclusion that vowels, whether in isolation, before oral consonants, or before nasal consonants, display integrality of F_1 and nasalization. The contrary conclusion reached by Krakow *et al.* [J. Acoust. Soc. Am. **83**, 1146–1158 (1988)] on the basis of data from a trading relations experiment reflect a limitation of that design for studying perceptual interaction. A second experiment used an array "rotated" in the stimulus space to determine whether F_1 and nasalization are privileged, perceptually primary dimensions. A new method for predicting classification performance for the rotated array without the assumption of primacy showed that they are not.

PACS numbers: 43.71.An, 43.71.Es

INTRODUCTION

Whether considered in acoustic or articulatory terms, the sounds of speech are multidimensional. That the perception of these sounds is also multidimensional is not in doubt, but the relation between stimulus and percept is far from clear. An understanding of how physical dimensions combine perceptually is of importance both theoretically, for modeling speech perception; and practically, for developing efficient coding schemes.

In this paper, we consider two related questions about the perceptual dimensionality of vowels differing in F_1 and nasalization (or, in articulatory terms, tongue and soft palate height). First, do these speech dimensions interact perceptually, so that more than one physical dimension contributes to the same psychological continuum? This is the *integrality* question. Second, of the many mathematically equivalent choices of dimensions by which our stimulus space might be characterized, are some "more equivalent" than others? This is the *primacy* question.

A. Do speech dimensions interact?

Though perceptual interaction between two physical dimensions is a distortion of the speech signal's acoustics, it is a benefit to the listener when both dimensions are attributes of the same speech sound and their interaction enhances a contrast with another speech sound. The interaction with which this article is concerned, that between nasalization and F_1, is a case in point: Diehl *et al.* (1990) have argued that for vowels in entirely oral contexts, soft palate height is directly covaried with tongue height because the resulting direct covariation of nasalization[1] with F_1 enhances vowel height contrasts.

To explain how varying soft palate height may interact with varying tongue height in influencing a listener's perception of vowel height, it is necessary to describe the acoustics of nasalized vowels. Lowering the soft palate in a nasalized vowel couples the nasal to the oral cavity; the acoustic effect of this coupling depends on the vowel's tongue height (and also its backness and rounding) and the amount of coupling (House and Stevens, 1956; Fujimura, 1960; Fujimura and Lindqvist, 1971; Stevens *et al.*, 1987; Beddor and Hawkins, 1990; Maeda, 1993. The account below follows Maeda's but accords with the other sources cited).

As the velopharyngeal port is opened in a nasalized vowel, the frequency of the zeros of the nasal cavity increases monotonically with respect to the corresponding poles of this cavity. Although there are at least two pairs of poles and zeros that may noticeably affect the perception of nasalized vowels, we will discuss only the lowest frequency pair, designated here as N_0 and N_1, because these were all that were manipulated in our stimuli.

In nonlow vowels, N_1's frequency is higher than the lowest oral pole's (F_1), so an additional spectral peak, N_1, appears above F_1 as N_0 rises away from N_1 with the velopharyngeal port's opening, and immediately above N_1 energy is sharply attenuated by N_0. In low nasalized vowels, on the other hand, N_1 is below F_1, and N_0 lies between them. In both low and nonlow vowels, F_1 also rises as the velopharyngeal port opens.

[a)]E-mail: kingston@coins.cs.umass.edu
[b)]E-mail: nam@bklyn.bitnet

0001-4966/95/97(2)/1261/25/$6.00

If N_1 and F_1 are not resolved perceptually, then the appearance of N_1 above or below F_1 increases the bandwidth of the lowest spectral prominence in a nasalized vowel. Such spectral "broadening," either locally near F_1 or more globally, has been a part of nearly all general acoustic and perceptual characterizations of vowel nasalization. As F_1's bandwidth is broader the lower the tongue, this effect of nasalization should make a nasalized vowel sound lower. F_1's rise will also make a nasalized vowel sound lower, as will N_1's emergence above F_1 in nonlow nasalized vowels, because both effects shift the lowest spectral prominence's center of gravity to higher frequencies. In low vowels, however, N_1's emergence *below* F_1 will pull this prominence's center of gravity downward, opposing the upward pull from F_1's rise. If N_1's emergence is more salient than F_1's rise, then a nasalized low vowel will sound higher rather than lower.

Wright's (1986) data show that listeners hear nasalized vowels as differing less in height than corresponding oral vowels, a result consistent with an upward shift in center of gravity for nasalized nonlow vowels but a downward shift for nasalized low vowels. Beddor and Hawkins's (1990) first experiment also shows a strong effect of center of gravity on listeners' matching of oral to nasalized vowels for height, at least for nonhigh vowels. Additionally, Beddor's (1983, 1993; also Beddor *et al.*, 1986) report of a broad crosslinguistic survey of changes in vowel height due to nasalization shows that high vowels lower and low vowels raise when nasalized. Contrastively nasalized mid vowels also lower, but the direction of height changes in contextually nasalized mid vowels depends on backness. So it would appear that N_1's emergence below F_1 is in fact more salient than F_1's rise in nasalized low vowels.

How then could the nasalization observed in low oral vowels in oral contexts enhance the height difference with high vowels, as claimed by Diehl *et al.* (1990)? It would appear that nasalization would distort the intended height of a low vowel just as much as a high one. Diehl *et al.* conjecture that the direction in which nasalization changes the perceived height of a low vowel depends on whether nasalization is heavy or only moderate. Raising occurs with heavy nasalization because N_1's prominence is great enough to shift the lowest spectral prominence's center of gravity downward. But lowering occurs with moderate nasalization because the broadening of that prominence's bandwidth and the raising of F_1 are more salient than the downward shift in center of gravity.

Our stimuli were constructed with N_1 below F_1, as in low vowels, even though the range of F_1's we employed corresponds to that of high-mid vowels. And our results all indicate that nasalization was heavy rather than moderate in our stimuli, and that nasalization lowered the center of gravity of our stimuli. This lowering was confirmed by spectral analysis. To anticipate, our results did not conform to Diehl *et al.*'s prediction, and we conjecture that if we had used more moderate nasalization we would have confirmed that prediction (an experiment is presently underway to test this conjecture).

Can the distortion of perceived height by such an interaction with nasalization also arise if a vowel is nasalized through coarticulation with an adjacent nasal consonant? Not according to the direct realist theory of speech perception (Fowler, 1986, 1990, 1991; Fowler and Smith, 1986; Krakow *et al.*, 1988). This theory claims that listeners appropriately attribute the acoustic properties of speech sounds to their articulatory sources, and thus asserts that nasalization will not integrate with F_1 and distort its perception when nasalization on the vowel results from coarticulation with an adjacent nasal consonant. Such synchronic sound changes can occur only if the source of the coarticulation is not perceived. Supporting evidence for this view comes from Kawasaki's (1978) demonstration that listeners' ratings of how nasalized a vowel is vary inversely with the amplitude of flanking nasals.

Evidence for the independent perception of tongue height and nasal coupling in vowels before nasal consonants has been provided by Krakow *et al.* (1988), who used a trading relations paradigm. Krakow *et al.*'s listeners categorized as [ɛ] or [æ] a series of articulatorily synthesized vowels varying in tongue height. The size of the synthesizer's velopharyngeal port was varied orthogonally to tongue height within blocks of trials, producing a range of amounts of nasal coupling. In the acoustic output of the synthesizer, lowering the tongue raised F_1, while increasing velar port opening (analogous to lowering the soft palate) both raised F_1 and increased the frequency separation between N_0 and N_1. Krakow *et al.* found that increasing nasal coupling shifted the boundary toward [ɛ], so that more of the continuum was heard as the lower vowel [æ]. Thus nasal coupling "traded" with tongue height in the perception of vowel height. However, this result obtained only when an oral consonant followed the vowel; when a nasal consonant followed that could serve as a coarticulatory source for the vowel's nasal coupling, no shift occurred, a contrast consistent with direct-realist predictions.

In the present experiments, we adopted an approach systematized by Garner (1974) to examine the integrality of F_1 and nasal coupling (N_0-N_1) in vowels preceding oral and nasal consonants, as well as in isolated vowels. Kingston (1991) previously used this method to show that these dimensions integrate perceptually in isolated vowels. The Garner paradigm (which we describe in more detail below) has two significant advantages over the traditional trading relations paradigm. First, it treats the two dimensions being varied symmetrically: In different tasks, listeners make judgments of both. Second, the Garner paradigm lends itself to a detection-theory analysis that unifies the outcomes of the various tasks (and can encompass trading relation experiments as well).[2]

To anticipate, our results reveal that F_1 and nasalization integrate strongly, in isolated and contextualized vowels. In isolated vowels or in an oral context, such integration is not evidence against direct realism: That theory argues that the listener has no choice but to attribute nasalization to the vowel in the absence of a coarticulatory source. But our results show that nasalization integrates with F_1 nearly as much before nasal as oral consonants, contrary to direct realism and, apparently, to Krakow *et al.*'s results. We argue in Sec. V that our results are more general than (rather than contradictory to) those of Krakow *et al.*, and that in some respects our results converge with theirs.

B. Are some speech dimensions special?

In a multidimensional stimulus space, orthogonal dimensions can be chosen in an infinity of ways. It is possible that the choice of axes is psychologically nonarbitrary, that a particular set of axes may be *perceptually primary*. The question of primacy is logically independent of the question of integrality: Even if the perceived value of a vowel for F_1 depends on its nasalization value, listeners could perceive the characteristics conveyed by F_1 and by nasalization as distinct properties of the vowel. Recently, Melara and Marks (1990) proposed a strategy for measuring primacy using "rotated" stimulus arrays in the Garner tasks. In experiment II, we apply their set of tasks to our F_1/nasalization stimulus space. The Melara and Marks analysis of the primacy question turns out to be inconclusive, but a new quantitative model applied to these data shows that the idea of primacy is unnecessary for our stimulus set.

The primacy question has not, to our knowledge, been raised in the speech domain. The idea of privileged dimensions is certainly consistent with the idea that distinctive features have invariant acoustic realizations (Stevens and Blumstein, 1978; Sussman *et al.*, 1991), and appears compatible with direct realism, but no one has specifically argued that primacy should hold for any particular speech dimensions. We view our test of primacy as an information-gathering rather than a hypothesis-testing venture.

C. Organization of the paper

Section I of this paper describes how the Garner paradigm was used in these experiments. Section II presents the results of our integrality experiments and a reanalysis of Kingston's (1991) data, and interprets them vis-a-vis the issue of perceptual integration of F_1 and nasalization according to the conventional Garner strategy. Section III reassesses the perceptual integrality of these dimensions using a detection-theoretic analysis. Section IV describes our theoretical and experimental approach to the primacy question. Section V compares our results using the Garner paradigm with those of Krakow *et al.* using the trading relations paradigm, and shows how we can accommodate both sets of results within a general model of the perceptual interaction of F_1 and nasalization. In Sec. VI we consider the implications of our results for phonetic theory.

I. THE GARNER PARADIGM: CONVERGING TASKS FOR DISTINGUISHING SEPARABILITY AND INTEGRALITY

The Garner paradigm (Garner, 1974) was devised as a test of perceptual interaction, or *integrality*. We use this term to mean that an observer's perception of one dimension of a multidimensional stimulus is influenced by the value of another, physically orthogonal dimension. To apply the paradigm, a stimulus array is constructed by varying two stimulus dimensions orthogonally, and listeners are required to classify various subsets from the array. Figure 1 shows the dimensions (F_1 and N_0-N_1) and stimulus values we used in experiment I. The stimuli are labeled by the subscripted letters A_0, B_0, C_0, and D_0. (The subscript 0 refers to the 0-deg orientation of the stimulus array with regard to the F_1 and N_0-N_1 axes; a different orientation and corresponding subscript are used in experiment II.) Values of F_1 and N_0-N_1 were chosen to make all pairs of stimuli differing on just one dimension approximately equally discriminable.

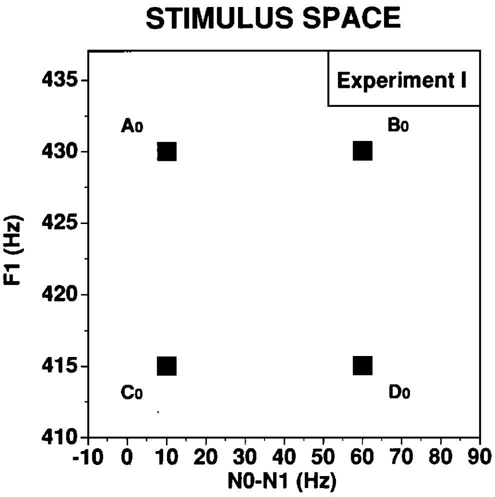

FIG. 1. The 2×2 array of stimuli used in experiment I, in a space defined by F_1 and nasalization. The subscript 0 refers to the 0-deg orientation of the stimulus array with regard to the F_1/N_0-N_1 axes.

In all of our experiments, a single stimulus was presented on each trial, and the listener chose one of two responses. The four distinct tasks are summarized in Fig. 2.

(1) In *baseline* conditions, the two possible stimuli differed along just one dimension. There were four such tasks: A_0 vs C_0 and B_0 vs D_0 for classification according to F_1 differences, and A_0 vs B_0 and C_0 vs D_0 for classification according to N_0-N_1 differences. The top left panel of Fig. 2 illustrates an N_0-N_1 classification in which stimulus A_0 (filled square) is to be classified differently than stimulus B_0 (open circle).

(2) In *correlated* tasks, listeners were presented with one of two stimuli from opposite corners of the array, B_0 vs C_0 or A_0 vs D_0. The values of the stimuli on the two dimensions could be either positively (B_0 vs C_0) or negatively (A_0 vs D_0) correlated. The top right panel in Fig. 2 illustrates the positively correlated task, B_0 (filled square) vs C_0 (open circle).

These two tasks each involved only two stimuli; the third integrality task used all four:

(3) In *selective attention* tasks, listeners classified the stimuli on the basis of N_0-N_1 differences (A_0 and C_0 vs B_0 and D_0) or F_1 differences (A_0 and B_0 vs C_0 and D_0). Because the listener tries to attend selectively to one dimension and filter out the differences between the stimuli on the other, orthogonal dimension, these are also called "filtering," or

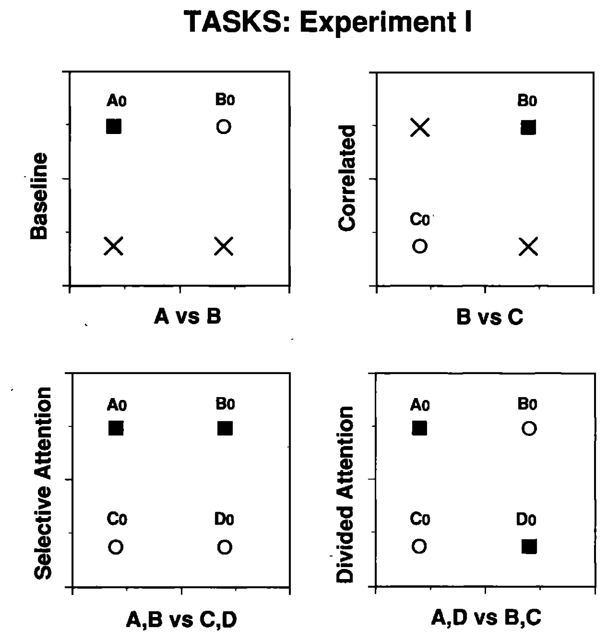

FIG. 2. Stimulus and response arrangements for the four types of classification tasks in the Garner paradigm, for the stimulus array used in experiment I. Top left: baseline, top right: correlated, bottom left: selective attention, and bottom right: divided attention. Squares and open circles indicate stimuli to which distinct responses must be given; ×'s indicate stimuli not used in the task.

"orthogonal" tasks. The bottom left panel in Fig. 2 illustrates the task in which the listener must filter out N_0-N_1 differences in order to distinguish stimuli with high F_1's (A_0 and B_0, filled squares) from those with low F_1's (C_0 and D_0, open circles).

One other four-stimulus task was run:

(4) In the *divided attention* task (Fig. 2, bottom right panel), listeners classified the positively correlated stimuli B_0 and C_0 (open circles) separately from the negatively correlated stimuli A_0 and D_0 (filled squares). Attention in this "condensation" task (Posner, 1964) is divided in the sense that listeners must attend to both dimensions.

Garner proposed that for perceptually *separable* dimensions performance on the correlated and selective attention tasks should be equivalent to that on baseline tasks. For integral dimensions, however, variation on a second dimension is expected to modify performance: Correlated performance should be better than baseline ("redundancy gain") and selective attention performance should be worse ("filtering loss"). The Garner paradigm thus provides a set of converging operations that can reveal whether the two stimulus dimensions are integrated or separable in perception.

The divided attention task is not directly relevant to assessing integrality,[3] but has been held up as a key task in measuring primacy. Further discussion will be deferred until Sec. IV, which deals with the primacy issue.

II. EXPERIMENT I

TABLE I. Klatt synthesizer parameter settings for isolated vowels.

Amplitude profile

Time (ms)	0	10	20	45	55	65
AV (dB)	0	57	60	60	57	0

F_0 and formant center frequencies and bandwidths

F_0	Formants	F_1	F_2	F_3	F_4	F_5
125	CF (Hz)	Table III	1050	2350	3300	3750
	Bandwidth (Hz)	70	110	140	250	250

Experiment I was designed to assess the integrality of F_1 and nasalization, the principal acoustic correlates of movements of the tongue body and soft palate in vowels. In most respects, the experiment is a typical realization of the Garner paradigm. The most important departure (motivated by our desire to include a detection-theory analysis of the data) was that the stimuli differed along each dimension by approximately one jnd, and the dependent variable was accuracy rather than response time.

In order to characterize the influence of context on the interaction of the stimulus dimensions, we used vowels in nasal and oral contexts, as well as in isolation. The isolated vowel conditions replicate the earlier study of Kingston (1991), who employed the traditional Garner analyses. We include a brief summary of these data, so that we can apply our detection-theory models to them as well.

A. Stimuli

Stimuli were created with Klatt's (1980) terminal analog synthesizer. One stimulus set consisted of brief, isolated vowels similar to those used by Kingston (1991). In the other sets, spectrally identical but longer vowels were placed in C_1—C_2 contexts, where C_1 was one of [b,d] and C_2 was one of [b,d,m,n]. The final consonant was thus either a voiced oral stop or a corresponding nasal. (The flanking consonants' place of articulation was varied to make it difficult for listeners to memorize the stimuli.) Settings for parameters other than F_1 and N_0-N_1 are listed in Table I (for isolated vowels) and Table II (for contextualized vowels).

The isolated vowel stimuli lasted only 65 ms from onset to offset of energy; the extremely short durations were necessary to reduce accuracy below ceiling. Because the overall amplitudes of all the vowels were matched (to within 0.2 dB), listeners could not use the ordinarily lower amplitude of nasalized vowels (House and Stevens, 1956) to detect nasalization in these stimuli. On the other hand, the reduction in F_1's amplitude by the adjacent nasal zero in the nasalized stimuli (described below) remained available as a cue to nasalization.

The formant frequencies simulated a mid, back, rounded vowel. In the isolated vowels the fundamental and formant frequencies remained constant throughout. In the contextualized vowels, these frequencies followed the time courses listed in Table II; preceding and following consonant intervals were 80 ms long; and transitions to and from the vowel lasted 30 ms. During the 140-ms steady state, the vowel was identical in all spectral properties but F_0 to the corresponding isolated vowel; F_0 followed the time course specified in Table II. Formant onset and offset frequencies specific to

TABLE II. Klatt synthesizer parameter settings for contextualized vowels.

Source amplitude profile

Time (ms)	0	80	90	270	280	360	__C
AV (dB)	45	45	60	60	45	45	b,d
AV (dB)	45	45	60	60	54	54	m,n

Fundamental frequency contour

Time (ms)	0	80	100	200	280	360
F_0 (Hz)	100	100	125	125	100	100

Formant frequency and nasal zero contours

Time (ms)	0	80	110	250	280	285	360	C
F_1 (Hz)	200	200	see Table III		200	200	200	b,d
	200	200	see Table III		480	480	480	m,n
N_0 (Hz)	[a]	[a]	see Table III		[a]	450	450	m,n
F_2 (Hz)	900	900	1050	1050	900	900	900	b__/__b
	[a]	[a]	1050	1050	900	1270	1270	__m
	1700	1700	1050	1050	1700	1700	1700	d__/__d
	[a]	[a]	1050	1050	1700	1340	1340	__n
F_3 (Hz)	2100	2100	2350	2350	2100	2100	2100	b__/__b
	[a]	[a]	2350	2350	2100	2130	2130	__m
	2600	2600	2350	2350	2600	2600	2600	d__/__d
	[a]	[a]	2350	2350	2600	2470	2470	__n

[a] Nasals did not occur in initial position.

TABLE III. Klatt synthesizer parameter settings for F_1, N_0, N_1, and N_0-N_1 for isolated and contextualized vowels used in experiment I.

Stimulus characteristic	Frequency (Hz)	
	A_0	B_0
F_1	430	430
N_0	265	315
N_1	255	255
N_0-N_1	10	60
	C_0	D_0
F_1	415	415
N_0	250	300
N_1	240	240
N_0-N_1	10	60

bilabial and alveolar consonants followed the prescriptions in Klatt (1980).

Because the nasal zero (N_0) at 450 Hz was close in frequency to F_1 at 480 Hz and F_1 had a broad bandwidth in both [m] and [n] (see Table II), the oral pole was effectively canceled in the nasals. The nasal pole (N_1) in these consonants, on the other hand, was well separated in frequency from N_0, and thus was not canceled by it.

All formant bandwidths were set to maximum values during oral stops (500 Hz for F_1 to F_4 and 700 Hz for F_5) to prevent any energy from being radiated above the fundamental. They were abruptly narrowed to values appropriate to the vowel at the boundary between consonant and vowel. During following nasals, however, bandwidths of F_2 and F_3 were allowed to remain narrower (200 Hz), as these sounds differ from oral stops in having weak formants during the interval of oral closure. Furthermore, in the nasals, F_2 and F_3 changed frequency abruptly from their value at the end of the vowel-to-consonant transition to values prescribed for bilabial or alveolar nasals by Klatt (1980), simulating the acoustic effect of shifting from an oral to nasal-plus-oral side branch resonator (Fant, 1960; Manuel, 1991).

The settings of N_0-N_1 and F_1, which are the same for isolated and contextualized vowels, are listed in Table III. The frequency of N_1 was always 175 Hz less than (and thus varied directly with) that of F_1, and N_0 was always between N_1 and F_1 in frequency. The bandwidths of N_0 and N_1 were both the default values of 90 Hz. Degree of nasalization correlated directly with N_0-N_1: the larger this separation, the less N_1's amplitude was reduced by N_0. We henceforth use N in place of N_0-N_1, as synonymous with nasalization. The 60-Hz separation in the more nasalized vowels B_0 and D_0 may be enough to make these vowels heavily rather than moderately nasalized, as N_0 is closer to the midway point between N_1 and F_1 than to N_1. The difference of 15 Hz between the F_1's of A_0 and C_0 or B_0 and D_0 is close to the three percent jnd for formant frequencies determined by Flanagan (1957), but larger than the one-to-two percent values obtained from very experienced listeners in low-uncertainty conditions by Hawks (1994) and Kewley-Port and Watson (1994).

In natural speech (or an articulatory-based synthesizer, like that used by Krakow *et al.*), increasing the size of the velopharyngeal port also increases F_1, but such a correlation is not necessary using a terminal analog synthesizer like Klatt's. We made no such adjustment in F_1, because to do so would have violated the conditions of the Garner paradigm, which requires that the physical dimensions of the stimuli be orthogonal.

B. Procedures

1. Listeners

The listeners were the experimenters and six graduate students recruited from the University of Massachusetts community. All reported normal hearing. The students were paid for their participation. Seven other potential subjects were rejected for failing to reach above-chance performance on all tasks during training.

2. Instructions

Written instructions, expanded orally by the experimenter, were used to describe the stimuli and the tasks. The listeners were told that they were to classify a set of two or four vowels according to quality and/or nasalization. In one condition, these vowels would appear in a C__C frame, while in another they would appear without any flanking consonants and would be quite brief.

For the contextualized vowel stimuli, the preceding consonants were described as [b,d] and the following consonants as [b,d,m,n]. Listeners' attention was drawn to the fact that two of the final consonants were oral and the other two nasal. The contextualized and isolated vowels were both described as varying in nasalization and quality. The diagram in (1) was shown to listeners to illustrate the orthogonality of differences in nasalization and quality (it was explained that the tilde indicated nasalization):

$$\begin{array}{cc} \text{o} & \tilde{\text{o}} \\ \text{u} & \tilde{\text{u}}. \end{array} \tag{1}$$

Nasalization was described as making a vowel sound on the one hand "buzzy" or "muffled" or on the other "fuller," "deeper," or "lower in pitch." The latter descriptions are plausible because a larger $N_0 - N_1$ difference makes N_1 more intense and thus shifts the spectral center of gravity downward in our stimuli, as when vowels are heavily rather than moderately nasalized. The difference between "o" and "u" was described as approximately that between the vowels in *daub* or *dawn* versus *bode* or *bone*, or like that between *dud* or *bun* vs *doob* or *boom* (pronounced with the vowel of *foot*, not *food*), i.e., as like that between [ɔ] vs [o] or [ʌ] vs [ʊ], but smaller.

The qualities represented by these IPA symbols cover the range of qualities imitated or named by the listeners. The vowels were always described as back, but there was variation in the description of their height or rounding. The experimenter pointed out that the listeners' impressions might not coincide with any of the descriptions, and urged them to devise their own. Listeners varied idiosyncratically in whether they reported that any of the qualities they heard was a good exemplar of any vowel category. None of these differences in the description or labeling of the stimuli correlated with listeners' relative success on particular classifications: The same classifications were relatively easy or difficult for all listeners. There were, however, clear differences between listeners in overall performance level.

Listeners were told that each classification would be binary, and that the responses Y(es) and N(o) would be arbitrarily assigned to the stimuli from task to task. For each task, the listener was shown a version of the diagram in (1) above in which each stimulus was labeled with the appropriate response. In addition, for each block of trials, the experimenter described the nature of the differences between the vowels to be classified, as well as the context (if any) in which the vowels would appear.

3. Training

Both contextualized and isolated vowel conditions began with training, in which the listeners learned to classify the vowels according to quality and nasalization. Training consisted of the following tasks, in order of what we judged to be increasing complexity. (In particular, the task in which stimuli differed on a single dimension was presented first.)

(1) one baseline task on each dimension: (a) on F_1, the low-N vowels A_0 vs C_0, and (b) on N, the low-F_1 vowels C_0 vs D_0;

(2) the negatively correlated task: low-N, high-F_1 A_0 vs high-N, low-F_1 D_0;

(3) the selective attention task in which the high-F_1 vowels A_0 and B_0 had to be classified separately from the low-F_1 vowels C_0 and D_0, with N irrelevant to the classification; and

(4) the divided attention task, in which B_0 and C_0, the vowels in which N was positively correlated with F_1, had to be classified separately from A_0 and D_0, in which the two dimensions were negatively correlated.

Listeners were instructed to assign an label, "Y(es)" or "N(o)," to each class of stimuli in each task, with the assignment varying arbitrarily from task to task. We did not find it necessary to train our listeners on the arbitrary assignment of "Y/N" labels to different stimuli in different tasks, because they were told at the outset about and trained on the arbitrary assignment, and they were told to pay careful attention to the feedback provided both in training and subsequent testing to learn what the particular assignment of labels was on each task. These instructions and experience made clear that "Y/N" had no content vis-á-vis stimulus characteristics but were merely reusable labels. Because each listener performed the tasks in a different order (determined by a balanced Latin square) and yet their performance waxed and waned with the tasks in similar ways, the labels' arbitrariness was apparently inconsequential. That it was instead task difficulty that predicted performance was shown by how closely a listener's success on a task performed early matched that on the same task repeated later.

In training on the contextualized stimuli, each vowel was presented eight times in each of the four possible combinations of framing initial and final consonants in separate final-oral and final-nasal blocks (final oral stops: [b__b, b__d, d__b, d__d] versus final nasals: [b__m, b__n, d__m, d__n]), yielding one or two blocks of 64 trials each, depending on whether two or four vowels had to be classified on that task. Within a block, both vowel and frame varied randomly. Listeners performed each vowel classification before oral and nasal consonants in paired but separate blocks; the order of oral and nasal blocks was varied pseudorandomly.

After each pair of blocks in training (and subsequent testing), the experimenter told the listener the percent correct, for each context, and asked for a description of the stimuli and any strategy the listener might have used.

4. Listening conditions and design

Listeners sat individually in an IAC sound-treated room in front of a monitor and keyboard. On each trial, a 500-ms warning message on the monitor was followed immediately by the stimulus, presented monaurally (to the ear each listener identified as "best") through a TDH-39 earphone. The listener's response (pressing the "Y" or "N" key followed by a carriage return) initiated a 500-ms message on the monitor reporting the correct answer. Thus there was a 1000-ms interval between the listener's response and the presentation of the next stimulus.

Before any vowel classification conditions (including training) were run, listeners participated in a consonant identification task using all combinations of vowels[4] and initial and final consonants. They identified the final consonant of each syllable as one of [b,d,m,n]. Each of the 64 stimuli was presented twice in separate randomizations, yielding 32 judgments of each final consonant by each listener.

In the experiment proper, all listeners performed the contextualized vowel classifications first. Each vowel stimulus was presented 10 times in each of the four consonant frames for each context block. There were thus 40 judgments of each vowel in each of the two contexts, oral and nasal, for each task. Otherwise, testing was identical to training. The total number of test trials per listener was $[(4 \text{ baseline tasks} \times 2 \text{ vowels}) + (2 \text{ correlated tasks} \times 2 \text{ vowels}) + (2 \text{ selective attention tasks} \times 4 \text{ vowels}) + (1 \text{ divided attention task} \times 4 \text{ vowels})] \times (4 \text{ C\_C frames} \times 2 \text{ contexts} \times 10 \text{ repetitions}) = 1924$.

For the isolated vowels, the same procedures were followed, but since there was no context, one or two sub-blocks of 64 (training) or 80 (testing) trials sufficed. In testing, 40 judgments were made of each isolated vowel stimulus in each classification task. The total number of trials per listener was: $[(6 \text{ tasks} \times 2 \text{ vowels}) + (3 \text{ tasks} \times 4 \text{ vowels})] \times (40 \text{ repetitions}) = 962$.

In both contextualized and isolated vowel conditions, the order in which the different classification tasks was performed was different for each listener, and was determined by a balanced Latin square.

Before testing began in each session, listeners' memories of the stimulus dimensions were refreshed by repeating the two training tasks in which the stimuli differed along just one dimension. The experiment required up to 10 h per listener to complete, distributed over 5–6 sessions lasting 1.5–2 h each. Following the consonant identification task, which lasted about 15 min, the remainder of the first session with contextualized vowels consisted of training. The same tasks were also used in training for the isolated as the contextualized vowel stimuli, but training took only half the time (about an hour) because there was no manipulation of context.

C. Kingston (1991): Stimuli and procedures

Kingston's (1991) experiment was identical to the present isolated-vowel condition, except that the F_1 difference was twice as large, and the frequency separation between F_1 and N_1 was 150 rather than 175 Hz. (One other difference with the present isolated-vowel condition is discussed in Sec. IV.) Table III lists the relevant stimulus parameters. In addition, Kingston's listeners (paid undergraduate and graduate volunteers from the Cornell University community) were neither given feedback after each trial nor aided by a diagram in keeping the stimuli straight. See Kingston (1991) for other details.

D. Results

1. Final consonant identification

Listeners were familiarized with the stimuli by having to identify the final consonant as [b,d,m,n]. Table IV is the confusion matrix for the final consonant identification, showing performance for each vowel separately. (Recall that two sets of four vowels were actually used, one from each experiment; Table IV reports only the data for syllables using the vowels of experiment I.)

TABLE IV. Proportion of final [b,d,m,n] identified as [b,d,m,n] (with standard errors), for each preceding vowel in experiment I. Each vowel-consonant combination was presented four times to each of eight listeners, so each cell represents the proportion of 32 judgments per VC combination, or the proportion of 128 judgments per final C.

		Consonant identification by vowel (0-deg vowels)			
Stim.	Vowel	"b"	"d"	"m"	"n"
b	A_0	0.875 (0.067)	0.094 (0.066)	0.031 (0.031)	0.000 (0.000)
	B_0	0.844 (0.066)	0.031 (0.031)	0.094 (0.046)	0.031 (0.031)
	C_0	0.906 (0.066)	0.063 (0.063)	0.031 (0.031)	0.000 (0.000)
	D_0	0.844 (0.066)	0.031 (0.031)	0.125 (0.047)	0.000 (0.000)
	all V	0.867 (0.032)	0.055 (0.025)	0.070 (0.020)	0.008 (0.008)
d	A_0	0.000 (0.000)	0.719 (0.100)	0.031 (0.031)	0.250 (0.106)
	B_0	0.031 (0.031)	0.531 (0.129)	0.031 (0.031)	0.406 (0.125)
	C_0	0.000 (0.000)	0.906 (0.046)	0.000 (0.000)	0.094 (0.046)
	D_0	0.000 (0.000)	0.469 (0.145)	0.000 (0.000)	0.531 (0.145)
	all V	0.008 (0.008)	0.656 (0.061)	0.016 (0.011)	0.320 (0.061)
m	A_0	0.219 (0.100)	0.063 (0.041)	0.563 (0.148)	0.156 (0.094)
	B_0	0.219 (0.120)	0.063 (0.041)	0.563 (0.148)	0.156 (0.105)
	C_0	0.125 (0.095)	0.063 (0.041)	0.688 (0.123)	0.125 (0.047)
	D_0	0.031 (0.031)	0.031 (0.031)	0.719 (0.120)	0.219 (0.100)
	all V	0.148 (0.046)	0.055 (0.019)	0.633 (0.066)	0.164 (0.043)
n	A_0	0.000 (0.000)	0.344 (0.170)	0.031 (0.031)	0.625 (0.164)
	B_0	0.000 (0.000)	0.063 (0.041)	0.000 (0.000)	0.938 (0.041)
	C_0	0.031 (0.031)	0.156 (0.081)	0.000 (0.000)	0.813 (0.092)
	D_0	0.000 (0.000)	0.125 (0.082)	0.000 (0.000)	0.875 (0.082)
	all V	0.008 (0.008)	0.172 (0.053)	0.008 (0.008)	0.813 (0.054)

For 15 of the 16 vowel/final-consonant segments, identification was correct on at least half the trials; average proportion correct was 0.74. The errors were systematic, and fell in a pattern that can be described as follows: Listeners were able to identify [b] and [n] more reliably than [d] and [m] (proportion correct=0.84 vs 0.64). Sounds having the same place of articulation were most commonly confused. A three-way repeated measures ANOVA with factors final consonant ([b,d,m,n]), response ("b", "d", "m", "n"), and preceding vowel (A_0,B_0,C_0,D_0) yielded a significant three-way interaction $[F(27,189)=2.05,\ p=0.003]$ reflecting two effects:

(1) [d] and [n] were each likely to be heard as the other consonant when the preceding vowel had a conflicting value for N: [d] was frequently heard as [n] after vowels with high N (B_0 or D_0) and [n] was heard as [d] after vowels with low N (A_0 or C_0); and

(2) the effects of conflicting N values on oral/nasal judgments were much smaller for the labial consonants [b] and [m].[5]

These results suggest that place of the articulation of the following consonant, as well as its nasality, should be included as a separate factor in analyzing the effects of context on vowel classification.

2. Analysis of vowel classification data

The preliminary analysis of both contextualized and isolated vowels has two parts. First, mean correlated and mean selective attention performance are each compared with mean baseline performance, to test for the redundancy gains and filtering losses that indicate integrality in the traditional Garner analysis. In our ANOVAs, these are planned comparisons. Second, we examine the data for asymmetries between

TABLE V. Mean d' (and standard errors) across listeners for contextualized vowels in experiment I.

		Context				
		Oral		Nasal		
Tasks	Stimuli	Labial	Coronal	Labial	Coronal	Mean
Baseline		1.16 (0.31)	1.36 (0.24)	1.31 (0.20)	1.42 (0.16)	1.32 (0.11)
N	A_0 vs B_0	1.47 (0.36)	1.33 (0.47)	1.37 (0.27)	1.67 (0.23)	1.46 (0.17)
	C_0 vs D_0	1.42 (0.44)	1.69 (0.34)	1.59 (0.30)	1.53 (0.40)	1.56 (0.18)
	N Mean	1.44 (0.37)	1.51 (0.34)	1.48 (0.23)	1.60 (0.26)	1.51 (0.15)
F_1	A_0 vs C_0	1.03 (0.38)	1.34 (0.29)	1.03 (0.27)	1.04 (0.31)	1.11 (0.15)
	B_0 vs D_0	0.75 (0.24)	1.10 (0.27)	1.26 (0.31)	1.45 (0.29)	1.14 (0.14)
	F_1 Mean	0.89 (0.30)	1.22 (0.23)	1.15 (0.29)	1.25 (0.26)	1.12 (0.13)
Correlated		1.56 (0.30)	1.82 (0.36)	1.66 (0.22)	1.77 (0.21)	1.71 (0.13)
F_1+N	B_0 vs C_0	0.56 (0.31)	1.08 (0.45)	0.68 (0.31)	0.69 (0.22)	0.75 (0.16)
F_1-N	A_0 vs D_0	2.57 (0.36)	2.56 (0.37)	2.65 (0.26)	2.85 (0.27)	2.66 (0.15)
Selective attention		1.10 (0.26)	1.25 (0.29)	1.03 (0.18)	0.98 (0.19)	1.09 (0.11)
N	A_0C_0 vs B_0D_0	1.48 (0.31)	1.63 (0.29)	1.22 (0.21)	1.17 (0.27)	1.37 (0.13)
F_1	A_0B_0 vs C_0D_0	0.73 (0.27)	0.87 (0.33)	0.84 (0.22)	0.78 (0.26)	0.81 (0.13)
Divided attention						
	A_0D_0 vs B_0C_0	0.45 (0.16)	0.66 (0.22)	0.33 (0.21)	0.46 (0.18)	0.48 (0.10)

variants of the same task type, breaking down baseline and selective attention performance into F_1 vs N and correlated tasks into positively vs negatively correlated.

Success in classifying the vowels in each condition is summarized by a mean d' calculated across all listeners. For the four-stimulus conditions, the d' calculation is heuristic, and the values cannot be interpreted as distances in a decision space. A systematic application of detection theory follows in Sec. III.

3. Vowels in context

Table V presents the results for contextualized vowels, both averaged across task types and for each variant separately. To compare the various conditions, we conducted a three-way repeated-measures ANOVA with factors task type (baseline versus correlated versus selective attention); and Nasality (oral versus nasal) and Place (labial versus coronal) of the following consonant.

The data reveal a clear redundancy gain: Mean performance on the correlated tasks ($d'=1.71$) was better than baseline ($d'=1.32$) [$F(1,7)=9.70$, $p<0.01$]. Filtering loss was smaller, but not reliably so: mean selective attention d' was 1.09 [$F(1,7)=3.27$, $p=0.09$]. This pattern was largely independent of the following consonant. Neither the nasality nor the place of the consonant affected performance on any of the three task types significantly. Overall performance was slightly better for [m] than [b] but worse for [n] than [d] [$F(1,7)=4.53, p=0.07$].

The dimension along which the vowels had to be classified strongly influenced performance. Four dimensions ean be identified in the stimulus space of Fig. 1: N, which distinguishes A_0 and C_0 from B_0 and D_0; F_1, which distinguishes A_0 and B_0 from C_0 and D_0; F_1+N, an increasing 45-deg axis that distinguishes B_0 and C_0; and F_1-N, a decreasing 45-deg axis that distinguishes A_0 and D_0. In another three-way ANOVA, the task type factor was redefined in terms of these dimensions, so that the levels were N baseline, F_1 baseline, N selective attention, F_1 selective attention, F_1+N correlated, F_1-N correlated, and divided attention.

Classification was better along the N than the F_1 dimension in baseline and corresponding selective attention tasks [average $d'=1.44$ vs 0.96, $F(1,7)=10.66, p<0.01$], indicating that we were not entirely successful in equating stimulus differences along the two dimensions. Listeners were much better at classifying vowels along the F_1-N dimension (the negatively correlated task) than along the F_1+N dimension (the positively correlated task) [$d'=2.66$ vs 0.75, $F(1,7)=84.97, p<0.0001$]. This asymmetry cannot be due to any inequality in stimulus differences and must instead reflect a genuine perceptual difference in how N and F_1 interact when negatively rather than positively correlated. The nasality and place factors followed the same pattern as in the analysis where tasks were not broken down by axis.

4. Isolated vowels: Present experiment and Kingston (1991)

Table VI presents the results for brief, isolated vowels in the present experiment and in Kingston (1991). As in the contextualized vowel condition, there is more consistent evidence of a redundancy gain than of a filtering loss. In both the present isolated vowel condition and Kingston's (1991) data, mean correlated performance exceeded baseline [present: $d'=2.04$ vs 1.30, $F(1,7)=9.38, p<0.01$; Kingston (1991): $d'=3.27$ vs 2.53, $F(1,7)=17.69, p<0.005$], but only in Kingston's (1991) data was mean selective performance reliably poorer than baseline [present: $d'=1.01$ vs 1.30, $F(1,7)=1.39$, $p=0.26$; Kingston (1991): $d'=2.13$ vs 2.53, $F(1,7)=5.07, p<0.05$].

Re-analysis by classificatory axis revealed that Kingston's (1991) listeners were substantially better at sorting isolated vowels by F_1 than N [baseline $d'=3.22$ vs

TABLE VI. Mean d' (and standard errors) across listeners for isolated vowels, experiment I and Kingston (1991).

Tasks		Present	Kingston (1991)
	Baseline	1.30 (0.26)	2.53 (0.12)
N	A_0 vs B_0	1.15 (0.33)	1.01 (0.58)
	C_0 vs D_0	1.46 (0.23)	2.65 (0.43)
	N Mean	1.31 (0.25)	1.83 (0.33)
F_1	A_0 vs C_0	1.60 (0.33)	3.47 (0.08)
	B_0 vs D_0	0.97 (0.38)	2.96 (0.22)
	F_1 Mean	1.28 (0.34)	3.22 (0.13)
	Correlated	2.04 (0.36)	3.27 (0.18)
F_1+N	B_0 vs C_0	1.52 (0.48)	2.90 (0.34)
F_1-N	A_0 vs D_0	2.56 (0.33)	3.64 (0.05)
	Selective Attention	1.01 (0.21)	2.13 (0.16)
N	A_0C_0 vs B_0D_0	1.15 (0.26)	1.56 (0.39)
F_1	A_0B_0 vs C_0D_0	0.87 (0.23)	2.69 (0.30)
	Divided attention		
	A_0D_0 vs B_0C_0	0.31 (0.16)	0.25 (0.11)

1.83 and selective attention $d'=3.64$ vs 2.90; $F(1,7)=22.23, p<0.0001$], while the listeners in the present experiment exhibited no reliable difference between the two axes [baseline F_1 $d'=1.28, N=1.31$; selective attention F_1 $d'=0.87, N=1.15$; $F(1,7)<1$]. This discrepancy, and the reversal of the effect found for contextualized vowels, probably reflects Kingston's use of a 30-Hz rather than a 15-Hz change in F_1.

Both isolated-vowel data sets reveal an advantage for classifying negatively over positively correlated stimuli, as was found with the contextualized vowels [present: $d'=2.56$ vs 1.52, $F(1,7)=11.54$, $p<0.005$; Kingston (1991): $d'=3.64$ vs 2.90, $F(1,7)=3.85$, $p=0.06$]. The more modest effect in Kingston's (1991) data may again reflect the larger F_1 difference he used.

E. Discussion

1. Garner analysis

According to the conventional interpretation of the Garner paradigm, the results of experiment I lead to the conclusion that F_1 and nasalization are perceptually integral dimensions in our stimuli: Correlated performance is uniformly better than baseline, which is better than selective attention. Not all comparisons achieve significance, but all are in the appropriate direction.[6]

Our data show one systematic effect not usually found for integral stimuli: More reliable classification of vowels differing in F_1-N than F_1+N, that is, in the negatively than the positively correlated task. Asymmetric performance has been reported by other investigators applying the Garner paradigm. Pomerantz (1981; Pomerantz and Schwaitzberg, 1975) considers unequal performance on the correlated tasks to be diagnostic of *configurality*, or the presence of emergent dimensions. Pomerantz has proposed, as a test of configurality, that performance in the divided task should exceed that in selective, a result that we did not find. Melara and O'Brien (1987), studying the interaction between "synesthetically corresponding" dimensions (pitch and visual height) found asymmetric correlated performance in the absence of high accuracy in the divided task. Our data resemble theirs, but there is no obvious "correspondence" between our dimensions.

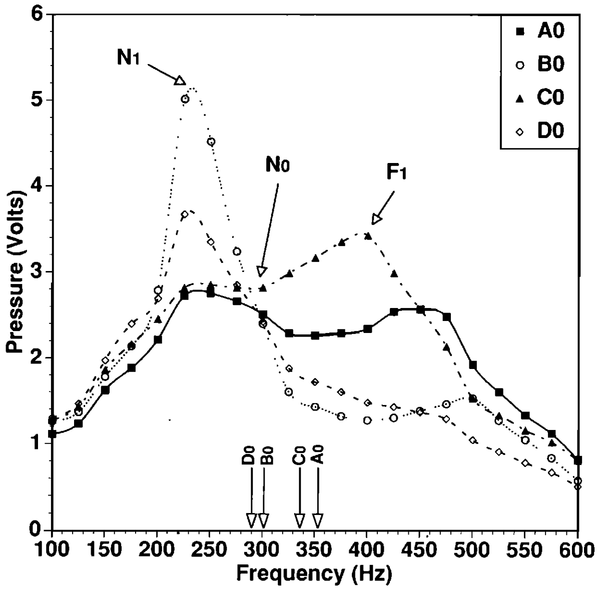

FIG. 3. Schematic spectra of the vowels used in experiment I, obtained by measuring pressure every 25 Hz between 100 and 600 Hz from a 29th-order LPC model of a 128-ms Hanning window placed over each vowel's steady state. Values plotted were obtained by subtracting mean intensity at 0 Hz from all measurements to normalize for overall differences in level, and then converting from dB to pressure (i.e., Pressure$=10^{\mathrm{dB}/20}$). Filled plotting figures are used for the unnasalized vowels A_0 and C_0 and open plotting figures for the nasalized vowels B_0 and D_0. Lines are splines fitted to the measurements. Labeled arrows indicate spectral centers of gravity for each vowel in experiment I as calculated by (2).

The consonants we used to provide a nasal or oral context for our vowels were not equally identifiable. Although no appreciable effect of context on the perceptual interaction between N and F_1 was revealed in the analyses described above, these differences led us to examine our data for each context separately in our theoretical analyses (see Sec. III).

2. Phonetic implications

Interestingly, the data are in apparent conflict with the predictions of both auditory enhancement and direct realist theories.

According to Diehl *et al.*'s (1990) hypothesis, a *positive* covariation between moderate nasalization and F_1, corresponding to the natural pattern of cooccurrence for these dimensions,[7] would be expected to enhance the vowel height contrast. In both the isolated and contextualized vowel conditions, the direction of the asymmetry between the two correlated tasks contradicts this prediction.

On the other hand, if nasalization was heavy rather than moderate in our nasalized vowels, our results are expected, because heavy nasalization lowers the center of gravity of a vowel's lowest spectral prominence rather than simply broadening its bandwidth as moderate nasalization does. Figure 3 shows how N_1's prominence and F_1's frequency combine to produce a larger difference in center of gravity between the vowels A_0 and D_0 in which F_1 and N are negatively correlated than between the positively correlated B_0 and C_0. The lowering of the center of spectral gravity brought about by an intense N_1 and a low F_1 in D_0 makes this vowel easy to distinguish from A_0, whose center of gravity is higher as a result of a weak N_1 and higher F_1. The vowels in which F_1 and N are positively correlated, B_0 and C_0, differ less in their centers of gravity because N_1 is weak when F_1 is low and vice versa.

Center of gravity was estimated from the formula:

$$\mathrm{CG}=\sum F_i P_i \Big/ \sum P_i, \qquad (2)$$

where F_i is frequency in Hz, P_i is pressure in volts, and the summations are over the frequency interval of interest. When pressure is measured every 25 Hz over the interval between 100 and 600 Hz in Fig. 3, the centers of gravity obtained for the vowels A_0, B_0, C_0, and D_0 are 352, 301, 336, and 290 Hz, respectively, values that correspond to the observed differences between listeners' success in classifying the negatively versus positively correlated stimuli.[8]

The finding that the integrality of N and F_1 is the same for contexts in which listeners could attribute nasalization to coarticulation with the following consonant, and those in which they could not, conflicts with the results of Krakow *et al.* (1988), and with the predictions of direct realism. We return to these issues in Secs. V and VI.

III. DATA ANALYSIS USING DETECTION-THEORY MODELS

A. Introduction

To this point, we have used detection theory (Green and Swets, 1966; Macmillan and Creelman, 1991) informally to summarize performance. We now use it to develop a theoretical framework for the data. An important advantage of detection theory, ideal for relating the multiple tasks of the Garner paradigm, is that it permits the estimation of the same performance measure (d') from a variety of experimental paradigms. Detection theory provides a description of our tasks in terms of a *decision space* in which distinct regions correspond to the possible responses, and distances correspond to values of d'. Recent advances in multidimensional detection theory (Ashby and Maddox, 1994; Ashby and Townsend, 1986; Graham, 1989) offer some especially versatile tools for analyzing our data.

Garner studies have, for the most part, used response time variables as dependent measures. Such experiments can also be analyzed in detection theory terms (Ashby and Maddox, 1994), but require additional assumptions that relate response times to decision-space constructs. Since we have chosen instead to use imperfectly discriminable stimuli and measure accuracy, we can apply detection theory directly.[9] Accuracy experiments were common in early research investigating dimensional integrality (e.g., Eriksen and Hake, 1955), but were usually analyzed in information-theoretic terms.

Our models assume that the observer's decision space, like the stimulus space, has two dimensions. Each of the four stimuli has an average location in the space, but various sources of noise produce trial-to-trial variability on both coordinates. Perceptual values for the stimuli thus form bivariate distributions of response likelihood, as shown in Fig. 4. The observer divides the decision space into regions corresponding to each response. Discriminability of any pair of stimuli is reflected by the distance d' between the means of the corresponding distributions, in units of their standard deviation.

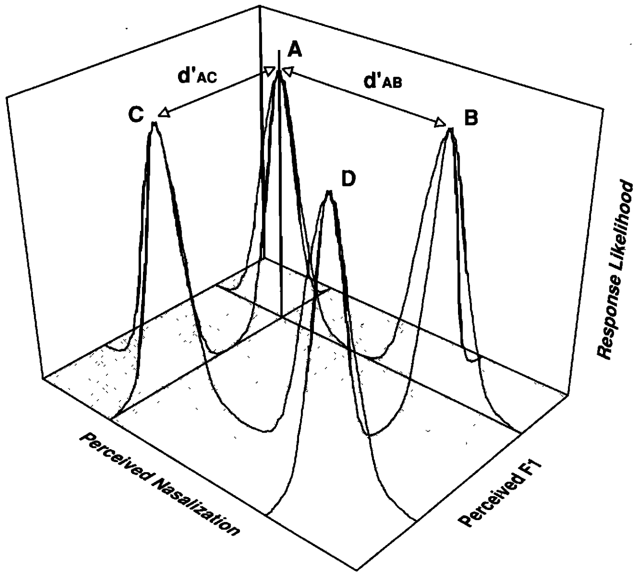

FIG. 4. Response likelihood distributions for the four stimuli of the 2×2 array, arranged in a two-dimensional decision space (the plane of the figure).

The upper panel of Figure 5 provides a two-dimensional, aerial perspective on the decision space of Fig. 4. The means of the distributions are represented as points, contours of equal likelihood as circles (the correct shape, for any value of likelihood, if the distributions are equal-variance, uncorrelated, bivariate normal, as we assume). Because the means define a rectangle in the decision space in this hypothetical case, the perceptual value of a stimulus on one dimension does not depend on its value on the other, and the dimensions are separable.

In the lower panel the means of the distributions are no longer arranged orthogonally, and the value of a stimulus on one dimension *does* depend on its value on the other. Maddox (1992) labeled this particular violation of independence "mean-shift integrality"; we adopt the shorter term "mean-integrality."

Our strategy for applying detection theory to the present data is as follows. First, we estimate the spatial arrangement of the four distributions from the two-stimulus (baseline and correlated) conditions. Second, we use the configurations obtained from the two-stimulus conditions to predict performance in the selective and divided conditions. A more detailed description of the models can be found in Macmillan and Kingston (in preparation).

B. Baseline and correlated tasks: Parallelogram model of mean integrality

The principal clue to distinguishing which of the two arrangements in Fig. 5 better describes the data is relative

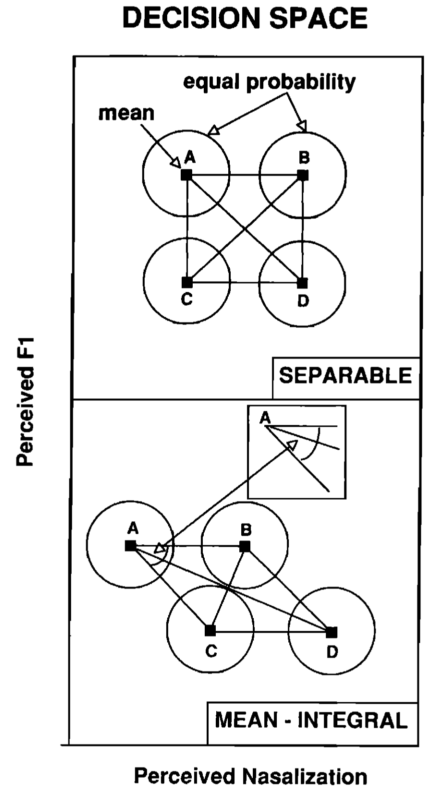

FIG. 5. Aerial view of the decision space in Fig. 4. Points represent the means of the response likelihood distributions, circles equal-likelihood contours. Top: perceptually separable stimuli; bottom: mean-integral stimuli. The value of the angle *CAB* (θ) in the inset at the bottom is the measure of mean integrality.

performance in the two correlated tasks: If these are equally difficult, the spatial arrangement of distributions is rectangular ($\theta=90°$) and the dimensions are separable; unequal accuracy, on the other hand, suggests a mean-integral arrangement, as in the bottom panel of Fig. 5.

Our interpretation of Garner paradigm results differs from the traditional one. A "redundancy gain" is observed in only one of the two correlated tasks when the stimulus dimensions are mean integral with θ is less than 60° or greater than 120°. A gain does occur for both tasks when the stimulus dimensions are separable, because the correlated stimuli are further apart in the stimulus space than the baseline stimuli (as noted by Ashby and Townsend, 1986). The size of this improvement is predictable from the Pythagorean theorem.

In order to assess mean integrality, we assumed that the mean locations in the true decision space formed a parallelogram, as in Fig. 5(b). (This simplifying assumption was not always correct; we relax it in Sec. V.) The fitting procedure was as follows: A parallelogram was constructed with length equal to d'_N (the average d' in the N baseline conditions) and height equal to d'_{F1} (the average d' in the F_1 baseline conditions). Iteration was used to find the value of θ that provided the best fit to the correlated d' values (the diagonals of the parallelogram). The magnitude of θ is our estimate of mean integrality, and the extent to which the correlated data fit the baseline-determined parallelogram is a test of the model.

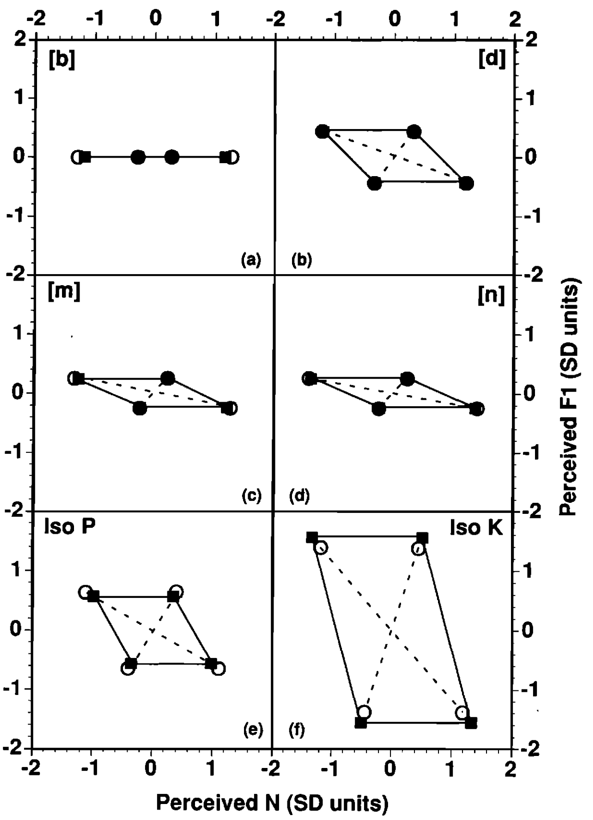

FIG. 6. Parallelogram representation for two-stimulus classification data in experiment I. The lengths of the sides equal the average of the parallel baseline d''s; the lengths of the diagonals, correlated d''s. The filled squares represent the means of perceptual distributions calculated from observed baseline performance; the open circles are means calculated from observed correlated performance. Contextualized vowels: (a) before [b], (b) before [d], (c) before [m], (d) before [n]; isolated vowels: (e) present experiment, (f) Kingston (1991).

Panels (a)–(d) of Fig. 6 show the results of this analysis for contextualized vowel data, separately for each following consonant. Mean integrality was obtained before all following consonants: Listeners always classified vowels differing in F_1-N more accurately than those differing in F_1+N. The degree of mean integrality depended on both the consonant's nasality and place: $\theta=44°$ for [d], 25° and 23° for [m] and [n], and 0° for [b]. We speculate in Sec. III E below about the relation between consonant identifiability and degree of mean integrality. The fit of the parallelograms to the correlated data was quite close for the contextualized vowel data: rms errors were 0.17 for the [b] context, 0.03 for [d], 0.06 for [m], and 0.05 for [n].

With both isolated-vowel data sets [Fig. 6, panels (e) and (f)] mean integrality was much more modest: $\theta=61°$ in the present study and 75° in Kingston (1991). The rms errors were also larger, 0.27 in the present study and 0.41 in Kingston (1991). Baseline d' was about twice as large in Kingston's experiment, which may explain the larger rms errors there.

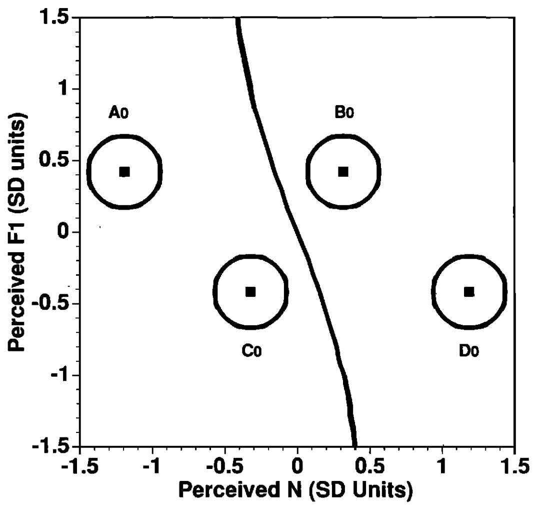

FIG. 7. Decision space for the AC vs BD selective attention task (using the results obtained before [d] in experiment I). Points are the means of distributions in the decision space, circles are equal-variance contours and the criterion curve is the locus of points for which the likelihood ratio is 1. The decision rule is to respond "A or C" for observations to the left of the curve, "B or D" otherwise.

C. Selective attention: The mean-integrality account of filtering loss

Figure 7 shows a decision space for the selective task in which stimuli A_0 and C_0 are assigned one response, stimuli B_0 and D_0 the other. The spatial arrangement is mean integral, with θ less than 90°; we use here the parallelogram for the [d] context where θ=44°. The criterion curve is the locus of points with a likelihood ratio of 1, that is, observations that are equally likely to arise from either pair of distributions.

Examination of the figure reveals that selective performance should not be quite as good as in the corresponding baseline conditions (A_0 vs B_0 and C_0 vs D_0), because of the proximity of distributions for the stimuli B_0 and C_0 that have to be classified differently on N; their proximity would also impair F_1 selective attention tasks. The amount of filtering loss can be predicted for any parallelogram. (A similar analysis, using a slightly different response criterion applied to response-time experiments, has been presented by Maddox, 1992, and Ashby and Maddox, 1994.)

No decline is predicted for the rectangular, separable arrangement. We evaluated other, mean-integral arrangements using a Monte Carlo method. As θ diverges from 90°, the magnitude of the predicted filtering loss increases; the effect is greatest when the two baseline sensitivities are similar. For example, if $d'_N = d'_{F1} = 2$ [so that baseline $p(c) = 0.84$] and θ=10°, predicted selective attention $p(c) = 0.75$; for the same θ, if $d'_N = 1$ and $d'_{F1} = 2$, $p(c)$ for the F_1 selective attention task is predicted to be 0.815.

Table VII compares average baseline performance with observed and predicted selective and divided attention performance for our data. All data are presented in terms of $p(c)_{\max}$, a measure obtained by converting d' to proportion correct under the assumption of no bias. (Because d' is a distance between two distributions, it cannot be used to summarize overall performance in the four-stimulus selective and divided attention conditions.) The analysis shows that mean integrality accounts for most of the small filtering losses in selective attention (top two-thirds of the table). The predicted decline in $p(c)_{\max}$ averages three percent for each dimension, and the observed decline averages three percent for N and five percent for F_1.

TABLE VII. Predicted versus observed percent correct [$p(c)_{\max}$] in selective and divided attention, contextualized vowels, experiment I. The column labeled "mean" represents performance after collapsing across final C's. The average differences between observed and predicted performance referred to in the text are calculated from the individual constant differences.

	Context		b	d	m	n	mean
	θ		0°	44°	25°	23°	23°
N	Baseline		77	77	77	79	78
	Selective attention	Predicted	74	75	75	76	77
		Observed	77	79	73	72	75
		Difference	−3	−4	2	4	2
F_1	Baseline		67	73	72	73	71
	Selective attention	Predicted	63	71	68	69	70
		Observed	64	67	66	65	66
		Difference	−1	4	2	4	4
	Divided attention	Predicted	62	65	65	67	65
		Observed	59	63	57	59	60
		Difference	3	2	8	8	6

Table VIII gives the same analysis for the two isolated-vowel data sets. Here, the predicted loss in selective attention is only about one percent for each dimension, and the observed losses are somewhat larger, four percent for N and six percent for F_1.

D. Divided attention: Another test of the model

Figure 8 shows a decision space for the divided task, in which stimuli A_0 and D_0 are assigned one response, stimuli

TABLE VIII. Predicted versus observed percent correct [$p(c)_{\max}$] in selective and divided attention, isolated vowels, experiment I, and Kingston (1991). For other details, see Table VII.

	Study		Present	Kingston (1991)	mean
	θ		61°	75°	68°
N	Baseline		75	82	79
	Selective attention	Predicted	74	81	78
		Observed	72	78	75
		Difference	2	3	3
F_1	Baseline		74	95	85
	Selective attention	Predicted	73	94	84
		Observed	67	91	79
		Difference	6	3	5
	Divided attention	Predicted	63	78	71
		Observed	56	55	56
		Difference	7	23	15

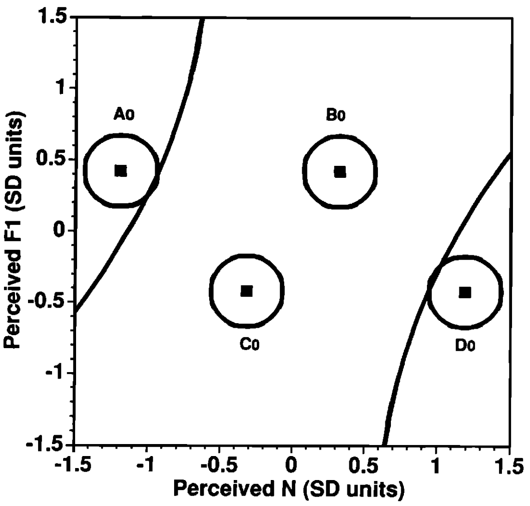

FIG. 8. Decision space for the divided attention task (using the experiment I results from the same context as in Fig. 7, i.e., before [d]). Points are the means of distributions in the decision space, and the criterion curves represent the locus of points for which the likelihood ratio is 1. The decision rule is to respond "*B* or *C*" for observations between the curves, "*A* or *D*" otherwise.

B_0 and C_0 the other. The spatial arrangement is mean integral, with θ less than 90°. The criterion curve is again the locus of points with a likelihood ratio of 1. Two special cases of this task—$\theta=0°$ and 90°, with equal baselines—are formally equivalent to models of the same-different discrimination paradigm. Compared to the one-interval task, same-different is quite difficult; predicted values can be found in Macmillan and Creelman (1991, Tables A.5.3 and A.5.4). Predictions for other cases were again made by simulation, and are compared with the data in Tables VII and VIII (bottom third).

For contextualized vowels (Table VII), our listeners' performance is worse than predicted, by five percent on average, suggesting that not all the difficulty in this task can be traced to mean-integrality effects. The discrepancy in the present isolated vowel condition (Table VIII) is about the same, seven percent. An unexpectedly huge discrepancy of 23 percent is found in Kingston's (1991) data. His listeners may have found the task hard because they received little training and no feedback, putting them at a marked disadvantage on this most complex task.

E. Conclusions from SDT analysis of experiment I

1. *The pattern of mean integrality*

The SDT analysis supports the conclusion that F_1 and N interact in our data, in the sense that they contribute to a single perceptual dimension, corresponding to F_1-N. This result conflicts with Diehl *et al.*'s (1990) predictions that positive rather than negative correlation between F_1 and N enhances a vowel height contrast, so long as N does not exceed moderate values. The discrepancy disappears if N is instead heavy in the nasalized vowels, because then the difference in their heights is reduced rather than enhanced. Ranking vowels by the center of gravity of their lowest spectral prominence is entirely analogous to ranking them by their locations along the F_1-N axis, i.e., in both: $A_0>C_0>B_0>D_0$.

This ranking of stimuli also puts vowels differing only in N (A_0 vs B_0 and C_0 vs D_0) perceptually further apart than vowels differing only in F_1 (A_0 vs C_0 and B_0 vs D_0), and apparently explains the better performance reliably observed in N than F_1 baseline tasks. However, the better performance on N baseline tasks is not a function of mean integrality but instead of an overestimate of the jnd for N (or an underestimate of the jnd for F_1), as a rectangle whose longer side corresponds to N projects in this order onto the F_1-N axis.

As Fig. 6 shows, mean integrality was strongest (actually complete: $\theta=0°$) before [b], still very strong before both [m] and [n] ($\theta=23°$, 25°), and clearly present before [d] ($\theta=44°$). Mean integrality was noticeably weaker, however, in isolated vowels ($\theta=61°$ and 75°).

We cannot completely explain this variation in the extent of mean integrality with context, but an account that makes clear predictions about its extent is Krakow *et al.*'s (1988) direct realist theory of speech perception. That theory predicts that N and F_1 will integrate in vowels only when there is no adjacent nasal consonant to which the listener could attribute any nasalization in the vowel. Our mean integrality results, as measured by the estimated θ's, appear to disconfirm direct realism's predictions strongly. Mean integrality is quite extreme before nasal as well as oral consonants, and when no context at all is present, the θ's are close enough to 90° to indicate separability rather than mean integrality.

An alternative set of direct-realist predictions uses the consonant identification data in Table IV to infer the likelihood that the following consonant is perceived as nasal, and from that likelihood rather than the intended nasality of the consonant predicts the extent of mean integrality. These predictions are as follows:

Consider first when [b] follows the vowel. Listeners reliably identified [b] as oral, even after nasalized vowels, so direct realism predicts the observed complete integration of N with F_1 before this consonant. The θ of 0° shows that our listeners heard the vowels before this consonant as differing along a single perceptual dimension corresponding to F_1-N.

What about when [d] follows? Listeners usually heard [d] as oral, but about one-third of the time they heard [d] as nasal instead. Direct realism would predict the observed compromise ($\theta=44°$) between complete integration (for trials on which the following consonant was heard as oral, predicted $\theta=0°$) and no integration (for trials on which it was heard as nasal, predicted $\theta=90°$).

Next, what about when the nasal consonants [m] and [n] follow? Listeners usually heard these consonants as nasal, but about a fifth of the time heard them as oral instead. The degree of mean-integrality predicted by direct realism is less than before [d], because these consonants were correctly identified as nasal (0.80 for [m] and 0.82 for [n]) more often than [d] was correctly identified as oral (0.66). This prediction fails, since θ is 25° and 23° before [m] and [n], compared to just 44° before [d]; that is, even though following

[m] and [n] are less ambiguously nasal than [d], mean integrality is greater.

Because direct realism claims that any nasalization in the vowel is to be attributed to its context when a nasal consonant follows, the theory also predicts that listeners should be unable to classify vowels for nasalization differences before nasal consonants, or at least should do so much less successfully. The data in Table V show, however, that our listeners were able to classify vowels by N at least as easily before nasal as oral consonants. This result differs from that obtained by Krakow and Beddor (1991), who found listeners less able to distinguish oral from nasalized vowels in nasal than oral contexts.

In contextualized vowels, mean integrality is predicted by direct realism to vary inversely with the extent to which the following consonant provides a coarticulatory source for it. In isolated vowels, where nothing provides a coarticulatory source for nasalization, integrality should be at least as extreme as in the oral context, but contrary to this prediction, we obtained the *least* mean integrality with the isolated vowels: $\theta=61°$ for the present experiment and 75° for Kingston (1991).

It is important to distinguish separability from success on single-dimension classifications; for us, the first is diagnosed by performance on the correlated tasks, whereas, the second is measured in the baseline tasks. Correlated performance in both isolated vowel experiments suggests that N and F_1 are (nearly) separable when there is no context. But we found N baseline performance to be no better in isolation than in either context (compare Table VI with Table V). The latter result is another discrepancy with Krakow and Beddor's (1991) results, which showed that listeners are better at distinguishing nasalized from oral vowels in isolation than in either context. Performance with our isolated vowel conditions may have been impaired, however, by the briefness of our isolated vowels.

2. *Variance integrality*

Another possible type of perceptual interaction is only hinted at in these data. An increase in variance accompanying an increase in the size of the stimulus set has been inferred using the trace-context theory of Durlach and Braida (1969) for intensity (Braida and Durlach, 1972) and vowels (Macmillan *et al.*, 1988). Variation along an irrelevant dimension may increase uncertainty along the relevant one, that is, increase the variance of the underlying distribution along that dimension. Such *variance integrality* would be indicated by substantial unexplained losses in the selective attention (filtering) or the divided attention task. Durlach *et al.* (1989) noted that such integrality could arise at either the sensory or context-memory level, and proposed techniques for distinguishing these outcomes by comparing discrimination and identification tasks.

In our data for both contextualized and isolated vowels, large filtering losses are not observed (see Tables VII and VIII). The decline in selective attention performance compared to baseline is only three to five percent for contextualized vowels and only four to six percent for isolated vowels, and these losses are largely accounted for by the mean integrality of the decision space.

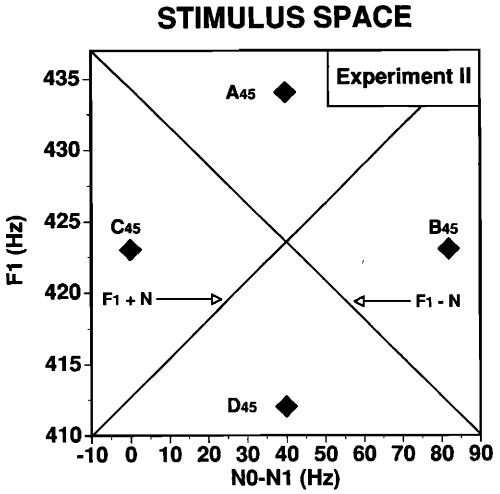

FIG. 9. The 2×2 array of stimuli used in experiment II, "rotated" 45 deg from the experiment I array in Fig. 1.

The decline in divided attention is larger. For contextualized vowels (Table VII), it is on average 15 percent, but all but five percent of it is predicted by mean integrality and the inherent difficulty of the task. For isolated vowels (Table VIII), the decline is yet larger, on average 19% in the present experiment and 34% in Kingston (1991). All but seven percent of the decline is accounted for in the present experiment, but fully 23% is unaccounted for in Kingston (1991). While the larger F_1 difference in Kingston's (1991) stimuli could have increased variance integrality, we believe his listeners simply did not master the task: They had less training, no feedback, and no diagram specifying how to classify the vowels. Kingston's data aside, variance integrality is slight in the present experiment.

IV. EXPERIMENT II: ROTATED STIMULUS ARRAYS FOR ASSESSING PRIMACY

Are F_1 and N, the dimensions we have used to describe the stimulus space of Fig. 1, uniquely psychologically accessible to the listener? Or is this choice of dimensions arbitrary? This is the primacy question, and it can be answered by comparing performance on stimulus arrays that are "rotated" in the stimulus space so as to be aligned with different sets of dimensions in the same space. Figure 9 illustrates a rotation of the original stimulus array (Fig. 1) by 45 deg with respect to the F_1 and N axes. To distinguish stimuli from the different rotations, we use subscripts: A_0 is a stimulus in the 0-deg set, A_{45} in the one rotated 45 deg. If the units on these axes are assumed to be equal, stimuli in the rotated array maintain the same Euclidean distances as at 0 deg. The 45-deg array has a particularly simple interpretation: for this array, the stimuli are parallel to axes that are the sum and difference of F_1 and N.

What might it mean for dimensions to be "primary?" Our decision-space approach suggests one answer: Perhaps distances in the psychological space must be measured along primary dimensions. Thus if F_1 and N are primary, the dis-

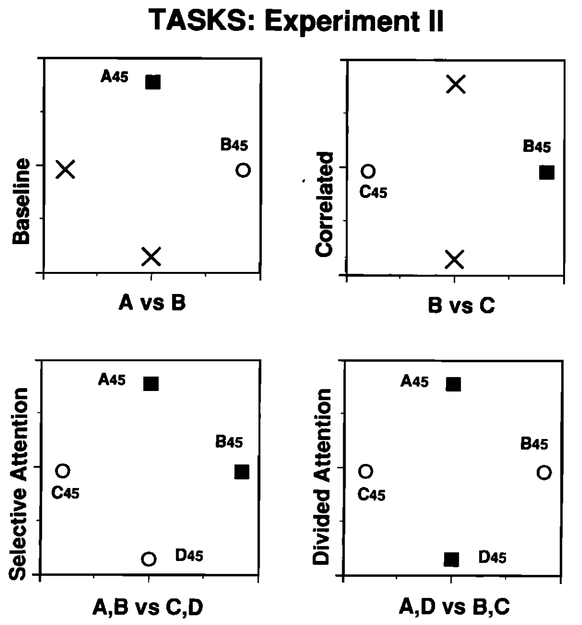

FIG. 10. Stimulus and response arrangements for the four types of classification tasks in the Garner paradigm, for the stimulus array used in experiment II. For details see Fig. 2.

TABLE IX. Klatt synthesizer parameter settings for F_1, N_0, N_1, and N_0-N_1 for isolated and contextualized vowels used in experiment II.

Stimulus characteristic	Frequency (Hz)	
	A_{45}	B_{45}
F_1	434	423
N_0	299	330
N_1	259	248
N_0-N_1	40	82
	C_{45}	D_{45}
F_1	423	412
N_0	248	277
N_1	248	237
N_0-N_1	0	40

tance between stimuli B_{45} and C_{45} is represented by a straight line along the N axis, whereas the distance between B_{45} and A_{45} is represented by the sum of two straight line segments, one parallel to each axis. In fact, these two pairs of stimuli are equally far apart by this measure. In the absence of primacy, the distances between stimuli are measured in Euclidean fashion, and B_{45} and C_{45} are farther apart than B_{45} and A_{45}. Primary dimensions, if they exist, thus impose a "city block" rather than Euclidean metric on the underlying psychological space.

The primacy question and the integrality issue are distinct, but particular models of these effects may interpret them as dependent. Adopting for the moment the city-block view of primacy, consider three possible relations between dimensions: mean integrality, variance integrality, and separability. Under mean integrality, the two stimulus dimensions are mapped onto a common psychological one. For two dimensions that are *completely* integral, the psychological space lies along a dimension that is the sum or difference of the original dimensions. City-block distances between two points in this space are systematically greater than Euclidean ones, but it appears that these metrics cannot be distinguished experimentally, so the presence of primacy cannot be determined. For separable stimuli, on the other hand, mean locations are unaffected by the second dimension, and primacy can clearly be assessed.

A method for evaluating primacy using rotated arrays in the Garner paradigm has recently been proposed by Melara and Marks (1990; see also Grau and Kemler Nelson, 1988; Smith and Kemler, 1978; Melara *et al.*, 1993a,b; Kemler Nelson, 1993). Figure 10 illustrates the various tasks of the Garner paradigm with the rotated array. The key tasks, according to Melara and Marks, are selective and divided attention. Selective attention is predicted to be poorer at 45 deg, for two reasons. First, rotation disaligns stimuli that are to be categorized together with respect to the perceptually primary dimensions. Second, among those stimuli that are categorized differently, there are stimuli that are identical on one dimension; for example A_{45} and D_{45} are identical in N, but belong to different categories in both selective attention tasks. On the other hand, performance in divided attention should improve, because rotation by 45 deg renders stimuli that are to be categorized together identical on one perceptually primary dimension. In fact, the divided attention task reduces to a classification of the two stimuli which have the same intermediate value along one of the putatively primary dimensions separately from the two that have extreme values along that dimension. Of course, if the 45-deg rotation corresponds to the perceptually primary dimensions, mirror image effects are expected, i.e., poorer selective attention and better divided attention performance at 0 than at 45 degs.

In experiment II, the same subjects who served in experiment I were tested with the 45-deg stimulus set of Fig. 9, using the tasks of experiment I (compare Fig. 10 with Fig. 2). The data allow us to apply Melara and Marks' analysis, but our detection-theory models offer a quantitative approach as well.

A. Method

1. Stimuli

Parameter values for the stimuli were identical to those used in experiment I except for the values of F_1 and N, which are given in Table IX. As in experiment I, vowels were presented both before oral and nasal consonants and in isolation. We again report the isolated vowel data of Kingston (1991), who also used a 45-deg array.[10]

2. Procedure

Procedures were the same as in experiment I, and the same subjects served as listeners. Half the listeners served in experiment I first, the other half in experiment II. Each listener completed one experiment before beginning the other.

Listeners were separately trained for each experiment. In experiment II, the training consisted of the following tasks, in order:

(1) The two correlated tasks: (a) B_{45} vs C_{45}, positively correlated stimuli differing just in N, and (b) A_{45} vs D_{45}, negatively correlated stimuli differing just in F_1;

(2) Two baseline tasks: (a) B_{45} vs D_{45}, stimuli differing in F_1+N (b) C_{45} and A_{45}, stimuli varying in F_1-N;

(3) One selective attention task: C_{45} and D_{45} vs B_{45} and A_{45}, the subsets varying in F_1+N; and

(4) The divided attention task: A_{45} and D_{45} vs C_{45} and B_{45}, subsets requiring that intermediate and extreme values of a dimension (which could be either F_1 or N) be distinguished.

The names of the tasks with the 0-deg stimulus array do not apply literally to tasks with the 45-deg array, as stimuli in the "baseline" and "selective attention" tasks can no longer be classified in terms of differences along a single dimension, whereas stimuli in the "correlated" tasks now can be, and the "divided attention" task no longer requires division of attention between dimensions. On the other hand, the Euclidean distances between the stimuli remain the same in the 45- as the 0-deg stimulus space; i.e., stimuli still differ more in the "correlated" than "baseline" tasks. The selective attention task still has irrelevant variation along an orthogonal axis. For these reasons and for ease of comparison with experiment I and previous Garner-paradigm studies, we retain the traditional terminology.

Listeners identified the consonants [b,d,m,n] following the vowels from the 45-deg array. Stimuli from the 0- and 45-deg arrays were mixed together in the consonant identification tasks, and each listener identified the consonants in this mixed stimulus set at the beginning of each experiment.

This experiment, like experiment I, required about 10 h per listener to complete, and again the first session consisted entirely of training.

TABLE X. Proportion of final [b,d,m,n] identified as [b,d,m,n] (with standard errors), for each preceding vowel in experiment II. Each vowel-consonant combination was presented four times to each of eight listeners, so each cell represents the proportion of 32 judgments per VC combination, or the proportion of 128 judgments per final C.

		Consonant identification by vowel (45-deg vowels)			
Stim.	Vowel	"b"	"d"	"m"	"n"
b	A_{45}	0.906 (0.066)	0.031 (0.031)	0.000 (0.000)	0.063 (0.063)
	B_{45}	0.875 (0.067)	0.031 (0.031)	0.094 (0.066)	0.000 (0.000)
	C_{45}	0.875 (0.067)	0.063 (0.041)	0.063 (0.041)	0.000 (0.000)
	D_{45}	0.875 (0.067)	0.000 (0.000)	0.125 (0.067)	0.000 (0.000)
	all V	0.883 (0.032)	0.031 (0.015)	0.070 (0.026)	0.016 (0.016)
d	A_{45}	0.031 (0.031)	0.500 (0.142)	0.125 (0.067)	0.344 (0.094)
	B_{45}	0.063 (0.063)	0.469 (0.110)	0.063 (0.041)	0.406 (0.125)
	C_{45}	0.031 (0.031)	0.875 (0.082)	0.031 (0.031)	0.063 (0.063)
	D_{45}	0.063 (0.063)	0.594 (0.094)	0.000 (0.000)	0.344 (0.081)
	all V	0.047 (0.024)	0.610 (0.059)	0.055 (0.022)	0.289 (0.050)
m	A_{45}	0.188 (0.103)	0.000 (0.000)	0.656 (0.115)	0.156 (0.081)
	B_{45}	0.063 (0.041)	0.063 (0.041)	0.719 (0.110)	0.156 (0.081)
	C_{45}	0.219 (0.129)	0.031 (0.031)	0.688 (0.123)	0.063 (0.063)
	D_{45}	0.250 (0.116)	0.000 (0.000)	0.719 (0.120)	0.031 (0.031)
	all V	0.179 (0.051)	0.024 (0.013)	0.695 (0.056)	0.102 (0.034)
n	A_{45}	0.031 (0.031)	0.156 (0.081)	0.000 (0.000)	0.813 (0.092)
	B_{45}	0.000 (0.000)	0.156 (0.105)	0.031 (0.031)	0.813 (0.103)
	C_{45}	0.000 (0.000)	0.156 (0.081)	0.000 (0.000)	0.844 (0.081)
	D_{45}	0.000 (0.000)	0.156 (0.094)	0.000 (0.000)	0.844 (0.094)
	all V	0.007 (0.007)	0.156 (0.043)	0.007 (0.007)	0.828 (0.044)

B. Results

1. Consonant identification

Table X summarizes our listeners' success in identifying consonants following vowels from the 45-deg stimulus array. Recall (Fig. 9) that vowels in this array have three degrees of N ($C_{45}<A_{45}=D_{45}<B_{45}$) and three degrees of F_1 ($D_{45}<C_{45}=B_{45}<A_{45}$).

As after the 0-deg vowels, [b] was reliably identified as oral, though after B_{45} and D_{45}, the vowels with highest N or lowest F_1, it was occasionally heard as nasal [m] instead. Success at identifying [d] as oral was quite a bit lower; especially after the three vowels with any amount of nasalization (A_{45},D_{45},B_{45}), [d] was quite likely to be heard as nasal [n]. The nasal [m] mirrors its oral counterpart [b], being heard as oral fairly frequently after the vowel with the lowest N (C_{45}) and also after the vowel with the lowest F_1 (D_{45}). Evidently, oral:nasal judgments of following labial consonants are difficult following this intermediate-N, low-F_1 vowel. Finally, the preceding vowel did not affect listeners' success at identifying [n] as nasal.

2. Contextualized vowels

Across all variants of each task type (Table XI), performance on the correlated tasks was better than on the baseline tasks ($d'=1.67$ vs 1.14) and selective attention performance was worse ($d'=0.87$), but only the former difference was significant [in planned contrasts: $F(1,7)=13.42, p<0.005$ for correlated vs baseline; $F(1,7)=3.41$, $p=0.09$ for selective attention versus baseline]. In the traditional interpretation, therefore, the F_1-N and F_1+N dimensions are integral for vowels in context. Correlated performance was better before labial than coronal consonants, but baseline and selective attention better before coronals, which accounted for a significant interaction between task type and place [$F(2,14)=7.60, p<0.01$].

When tasks were distinguished by the axis along which the stimuli were to be classified, striking parallels to performance with the 0-deg stimuli emerged. Planned contrasts showed that classification by differences in N (the positively correlated task) was somewhat better than by differences in F_1 (the negatively correlated task), though not significantly [$d'=1.88$ vs 1.46, $F(1,7)=1.85$, $p=0.18$], and classification by differences in F_1-N was quite a bit better than by F_1+N [$d'=1.70$ vs 0.58 for baseline and 1.40 vs 0.35 for selective attention; $F(1,7)=24.09$, $p<0.0001$].

Finally, a significant interaction between following consonant place and task type [$F(2,14)=3.04$, $p<0.02$] resulted from two effects. First, listeners classified vowels differing in N in the positively correlated task better before labial than coronal consonants [$d'=2.09$ vs 1.67]. This difference is consistent with listeners' greater success at distinguishing oral from nasal consonants when they are labial than when they are coronal. Second, listeners classified vowels differing in F_1+N in baseline and selective attention tasks better before coronal [baseline $d'=0.72$ and selective

TABLE XI. Mean d' (and standard errors) across listeners for contextualized vowels in experiment II.

		Context				
		Oral		Nasal		
Tasks	Stimuli	Labial	Coronal	Labial	Coronal	Mean
	Baseline	0.98 (0.28)	1.05 (0.26)	1.14 (0.29)	1.39 (0.28)	1.14 (0.14)
F_1-N	A_{45} vs B_{45}	1.36 (0.52)	1.47 (0.41)	1.91 (0.41)	1.95 (0.44)	1.67 (0.22)
	C_{45} vs D_{45}	1.70 (0.47)	1.65 (0.53)	1.77 (0.53)	1.80 (0.38)	1.73 (0.23)
	F_1-N Mean	1.53 (0.47)	1.56 (0.46)	1.84 (0.43)	1.88 (0.39)	1.70 (0.21)
F_1+N	A_{45} vs C_{45}	0.17 (0.10)	0.13 (0.24)	0.27 (0.22)	0.62 (0.24)	0.30 (0.11)
	B_{45} vs D_{45}	0.71 (0.23)	0.94 (0.27)	0.61 (0.32)	1.19 (0.32)	0.86 (0.14)
	F_1+N Mean	0.44 (0.15)	0.53 (0.16)	0.44 (0.23)	0.91 (0.25)	0.58 (0.10)
	Correlated	1.71 (0.38)	1.49 (0.37)	1.85 (0.46)	1.61 (0.40)	1.67 (0.19)
F_1	B_{45} vs C_{45}	2.21 (0.43)	1.65 (0.31)	1.97 (0.58)	1.70 (0.45)	1.88 (0.22)
N	A_{45} vs D_{45}	1.22 (0.49)	1.34 (0.48)	1.74 (0.49)	1.53 (0.49)	1.46 (0.23)
	Selective attention	0.77 (0.25)	1.03 (0.26)	0.81 (0.24)	0.88 (0.23)	0.87 (0.12)
F_1-N	$A_{45}C_{45}$ vs $B_{45}D_{45}$	1.39 (0.40)	1.54 (0.44)	1.34 (0.42)	1.33 (0.34)	1.40 (0.19)
F_1+N	$A_{45}B_{45}$ vs $C_{45}D_{45}$	0.16 (0.18)	0.52 (0.14)	0.29 (0.13)	0.43 (0.19)	0.35 (0.08)
	Divided attention					
	$A_{45}D_{45}$ vs $B_{45}C_{45}$	0.33 (0.13)	0.42 (0.14)	0.39 (0.15)	0.33 (0.12)	0.37 (0.06)

attention $d'=0.47$] than labial consonants [baseline $d'=0.44$ and selective attention $d'=0.22$].

3. *Isolated vowels*

The isolated vowel data from the present experiment and from Kingston (1991) are shown in Table XII. In the present experiment, both correlated ($d'=2.03$) and selective attention ($d'=1.01$) differed from baseline [$d'=1.50$; $F(1,7)=13.14$, and $F(1,7)=11.43$, both $p<0.005$], implying integrality of F_1+N and F_1-N. In the breakdown by classificatory axis, performance was significantly better on the negatively correlated F_1 than the positively correlated N axis [$d'=2.41$ vs 1.66; $F(1,7)=7.16$, $p<0.05$]. And vowels differing along the F_1-N axis were classified significantly more accurately in baseline and selective attention [$d'=1.78$ and 1.37] than those that differed along the F_1+N axis [$d'=1.22$ and 0.64; $F(1,7)=10.64$, $p<0.005$]. The first difference reverses the result obtained with contextualized vowels, which were always classified more accurately with respect to the N than F_1 axis, but the second accords with the contextualized vowel outcome.

TABLE XII. Mean d' (and standard errors) across listeners, isolated vowels, experiment II and Kingston (1991).

Stimulus set		Present	Kingston (1991)
	Baseline	1.50 (0.29)	2.39 (0.24)
F_1-N	A_{45} vs B_{45}	1.73 (0.32)	2.88 (0.37)
	C_{45} vs D_{45}	1.84 (0.29)	2.88 (0.28)
	F_1-N Mean	1.78 (0.26)	2.88 (0.17)
F_1+N	A_{45} vs C_{45}	1.04 (0.31)	1.71 (0.48)
	B_{45} vs D_{45}	1.41 (0.48)	2.09 (0.41)
	F_1+N Mean	1.23 (0.39)	1.90 (0.41)
	Correlated	2.03 (0.36)	2.77 (0.28)
N	B_{45} vs C_{45}	1.66 (0.33)	2.39 (0.51)
F_1	A_{45} vs D_{45}	2.41 (0.44)	3.15 (0.22)
	Selective attention	1.01 (0.31)	1.36 (0.25)
F_1-N	$A_{45}C_{45}$ vs $B_{45}D_{45}$	1.37 (0.34)	1.72 (0.28)
F_1+N	$A_{45}B_{45}$ vs $C_{45}D_{45}$	0.64 (0.34)	0.99 (0.25)
	Divided attention		
	$A_{45}D_{45}$ and $B_{45}C_{45}$	0.30 (0.11)	0.78 (0.25)

In Kingston's (1991) data for 45-deg isolated-vowel stimuli, correlated performance exceeded baseline, but not significantly [$d'=2.77$ vs 2.49, $F(1,7)=1.05$, $p=0.32$]; selective attention ($d'=1.36$) was significantly worse [$F(1,7)=17.61$, $p<0.001$]. In the breakdown by classificatory axis, performance was better along the negatively correlated F_1 ($d'=3.15$) than positively correlated N axis ($d'=2.39$), though not quite significantly [$F(1,7)=3.47$, $p=0.07$]. Performance on baseline and selective attention tasks was significantly better along the F_1-N than the F_1+N axis [$d'=2.88$ and 1.72 vs 1.90 and 0.99; $F(1,7)=8.73$, $p<0.01$]. Kingston's (1991) results with vowels from a 45-deg array thus accord in detail with those obtained with isolated vowels from the same rotation in the present study.

In summary, both isolated and contextualized vowels from the rotated arrays displayed integrality by the standard tests. Filtering losses were stronger than redundancy gains for the isolated vowels, while the reverse was true for contextualized vowels. We now turn to detection-theoretic modeling of our data to interpret this and the other patterns obtained with the rotated stimulus arrays.

C. Integrality

1. *Baseline and correlated tasks: Parallelogram analysis*

Parallelograms inferred from baseline and correlated performance with the contextualized vowels are shown in Fig. 11. The greatest deviation in θ's from 90° was 180°, which occurred before the consonant that was most reliably identified as oral, [b], and the least deviation (98°) was before the consonant most reliably identified as nasal, [n]. That

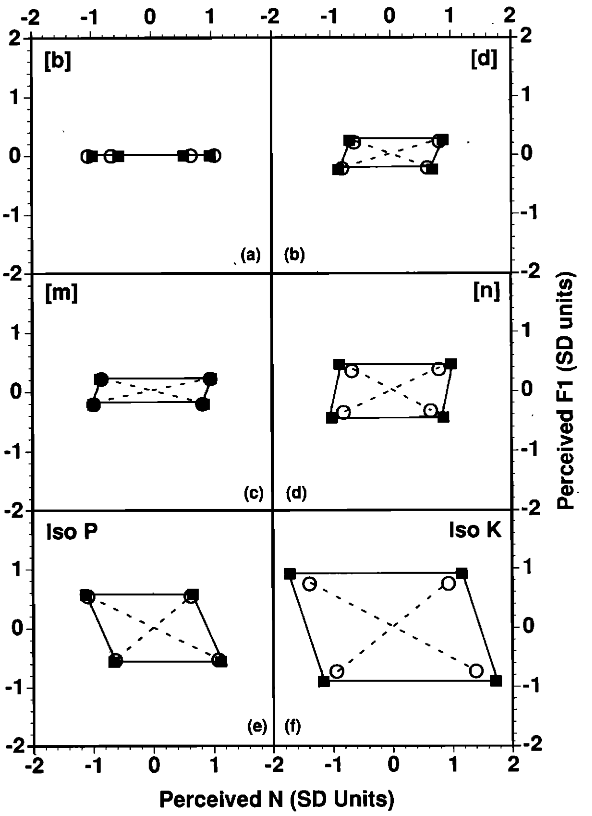

FIG. 11. Parallelogram representation for two-stimulus contextualized and isolated vowel classification data from experiment II and Kingston (1991). For details see Fig. 6.

is, the dimensions corresponding to the sum and difference of F_1 and N are completely mean-integral before [b], a context in which our listeners classified vowels as though they differed along only the N axis; and virtually separable before [n]. The θ's of 109° and 106° before [d] and [m] indicate that these dimensions are also nearly separable before these two consonants. Rms error between observed and predicted correlated performance is: 0.20 before [b], 0.15 before [d], 0.03 before [m], and 0.46 before [n].

Figure 11(e),(f) displays parallelograms inferred from the two isolated-vowel data sets. For data from the present study, rms error is modest (0.10), but for data from Kingston (1991) the discrepancy was quite large (0.66). For the present study, θ=68° and for Kingston (1991) θ=73°.

Note that θ's are uniformly (but slightly) less than 90° for the rotated arrays of isolated vowels, but greater than 90° for the rotated arrays of contextualized vowels. This result means that classification was better along the F_1 than N axis for isolated vowels, but better along N than F_1 for contextualized vowels. Better performance along F_1 than N with isolated-vowel stimuli does not depend on the size of the difference in F_1, since it was obtained in the present study as well as in Kingston (1991).

2. Selective and divided attention

For contextualized vowels, filtering losses were five percent on average for F_1-N and seven percent for F_1+N; mean integrality accounts for at most one percentage point (Table XIII, top two-thirds). The discrepancy is larger than

TABLE XIII. Predicted versus observed percent correct $[p(c)_{\max}]$ in selective and divided attention, contextualized vowels, experiment II. For other details, see Table VII.

	Context		b	d	m	n	mean
	θ		180°	110°	106°	98°	123°
F_1+N	Baseline		59	61	59	67	62
	Selective attention	Predicted	57	60	59	67	61
		Observed	53	60	56	59	55
		Difference	4	0	3	8	6
F_1-N	Baseline		78	79	82	83	81
	Selective attention	Predicted	77	78	82	83	81
		Observed	76	78	75	75	76
		Difference	1	0	7	8	5
	Divided attention	Predicted	54	56	56	61	58
		Observed	57	58	58	57	58
		Difference	−3	−2	−2	4	0

for the 0-deg stimuli, but still modest. The present isolated-vowel results are similar (Table XIV, top two-thirds), but Kingston's (1991) data again exhibit larger discrepancies.

Mean-integrality predicted essentially all of the decline in performance on divided attention for contextualized vowels (Table XIII, bottom third). For isolated vowels, on the other hand, the decline in divided attention is noticeably larger than predicted, though the discrepancy is not so large as it was for the 0-deg stimuli (Table XIV, bottom third).

D. Consistency of 45-deg and 0-deg data

In this section, we outline evidence suggesting that listeners responded to stimuli from the two rotations as though they were drawn from the same perceptual space. In the stimulus space (Figs. 1 and 9), each baseline distance at 0 deg has the same orientation as (is parallel to) a diagonal distance at 45 deg, and vice versa. If listeners treat the stimuli as belonging to a single perceptual space, we may

TABLE XIV. Predicted versus observed percent correct $[p(c)_{\max}]$ in selective and divided attention, isolated vowels, experiment II, and Kingston (1991). For other details, see Table VII.

	Study		Present	Kingston (1991)	mean
	θ		68°	73°	71°
F_1+N	Baseline		73	83	78
	Selective attention	Predicted	72	82	77
		Observed	63	69	66
		Difference	9	13	11
F_1-N	Baseline		81	93	87
	Selective attention	Predicted	81	92	87
		Observed	75	81	78
		Difference	6	11	9
	Divided attention	Predicted	65	78	72
		Observed	56	65	61
		Difference	9	13	11

TABLE XV. Correlated performance at 45 deg interpolated from averaged 0-deg baseline performance, and correlated performance at 0 deg interpolated from averaged 45-deg baseline performance. Observed is correlated performance divided by $\sqrt{2}$.

Condition/ Interpolation			Contexualized b	d	m	n	Isolated Present	Kingston (1991)
45° from 0°	N	Predicted	0.89	1.22	1.14	1.24	1.29	3.22
		Observed	0.86	0.95	1.23	1.08	1.70	2.23
	F_1	Predicted	1.45	1.51	1.48	1.60	1.33	1.83
		Observed	1.56	1.17	1.39	1.20	1.17	1.69
0° from 45°	F_1+N	Predicted	0.44	0.54	0.44	0.90	1.22	1.90
		Observed	0.40	0.76	0.48	0.49	1.07	2.25
	F_1-N	Predicted	1.53	1.56	1.84	1.87	1.78	2.88
		Observed	1.82	1.81	1.87	2.02	1.81	2.57

therefore expect that, for example, the d' from the positively correlated condition at 45 deg (B_{45} vs C_{45}) should equal the average of the two horizontal baseline conditions at 0 deg (A_0 vs B_0 and C_0 vs D_0). Correlated performance must be divided by $\sqrt{2}$ before making this comparison, to account for the greater distance between the diagonal stimuli.

Four such comparisons can be made for each stimulus context. In Table XV, correlated performance divided by $\sqrt{2}$ at one rotation ("observed") is compared to mean baseline performance at the other rotation ("predicted"). The predictions are generally successful: The average rms error is 0.27 for predicting 45-deg performance from 0 deg and just 0.15 for predicting 0-deg performance from 45 deg. Performance at the two rotations thus turns out to be highly consistent once tasks are compared in which the stimuli differ along parallel axes.[11]

This interpolation calculation does not distinguish the case where some set of axes is perceptually primary from one where none are. In the next section, we predict across rotations aspects of performance which do require taking a stand on the primacy question.

E. Primacy

In Melara and Marks's (1990) conception, nonprimary dimensions are less accessible than primary ones. Tasks that require perception in terms of nonprimary dimensions (or combinations of primary ones) should be difficult. Rotating the stimulus set away from the primary axes should therefore make the selective task more difficult (relative to baseline) and the divided task less so.

In qualitative terms, this is in fact what we found. *Unexplained* selective loss increased modestly with rotation for contextualized vowels and more noticeably for isolated vowels: A comparison of the averaged difference between predicted and observed performance for experiment I in Table VII with that for experiment II in Table XIII shows an increase from three to 5.5% for contextualized vowels, and a similar comparison of Table VIII with XIV shows an increase of 4% to 10% for isolated vowels. Rotation reduced the unexplained loss in divided attention (predicted minus observed in Tables VII vs XIII and VII vs XIV) from 6% to 0% for contextualized vowels and from 15% to 11% for isolated vowels. These changes are roughly consistent with the hypothesis that F_1 and N are primary dimensions by Melara and Marks' (1990) definition.

It is important, however to compare the data to the outcomes expected in the absence of primacy. We can use our detection-theoretic framework to predict performance quantitatively in all tasks. The first requirement is some assumption about the relation between the stimulus and perceptual spaces. The simplest possible hypothesis—that each perceptual dimension is a function of just one of the stimulus dimensions—can be ruled out because it predicts separability (θ=90°). Instead, we assume that the stimulus-perception transformation is *bilinear*, that is, that each perceptual dimension is a linear function of both stimulus dimensions:

$$x_P = a x_S + b y_S, \tag{3}$$

$$y_P = c x_S + d y_S. \tag{4}$$

In these equations, (x_S, y_S) is the location of a point in the stimulus space, and (x_P, y_P) is the corresponding distribution mean in the decision space.

The coefficients a, b, c, and d of the bilinear transformation can be inferred from the parallelogram representation for the 0-deg data. Applying the inferred transformation to the rotated array yields a predicted parallelogram for the 45-deg condition. Table XVI shows that the method provides a good account of our 45-deg data. The sides of the 45-deg parallelogram are well predicted (average rms error is 0.17 for contextualized vowels and 0.14 for isolated vowels, and the qualitative changes in θ are also accounted for (rms error is 17°).

TABLE XVI. Predicted versus observed baseline performance at 45 deg; predictions are from bilinear transformations of 0-deg data.

Condition		Contextualized b 45°	d 45°	m 45°	n 45°	Isolated Present 45°	Kingston (1991) 45°
$d'\ F_1-N$	Predicted	1.65	1.79	1.81	1.97	1.58	2.89
	Observed	1.53	1.56	1.84	1.88	1.78	2.88
$d'\ F_1+N$	Predicted	0.40	0.75	0.46	0.47	0.93	2.31
	Observed	0.44	0.53	0.44	0.91	1.23	1.90
θ	Predicted	180°	107°	122°	123°	91°	58°
	Observed	180°	110°	106°	98°	68°	73°

Nothing in the approach requires that any particular dimension have special "primary" status. With one exception, θ comes much closer to 90° with the stimuli from the 45-deg array than the 0-deg array; this is evidence that F_1+N and F_1-N are more separable than F_1 and N. However, separability does *not* implicate primacy. Primacy would be inferred for dimensions defining a rotation that produced a set of d's larger than could be predicted by a bilinear transformation of d's at another rotation. No such effect of rotation was obtained in our data. The success of the models in relating performance at different rotations makes primacy assumptions unnecessary.

V. TWO STRATEGIES FOR STUDYING PERCEPTUAL INTERACTION: THE GARNER PARADIGM AND THE TRADING RELATIONS EXPERIMENT

A. Sensitivity versus response bias

We return here to a comparison of our results with those of Krakow *et al.* (1988), and particularly to an apparent discrepancy between the two studies. In their study, judgments of vowels were strongly affected by whether the following context was oral or nasal; in our study, the effect was slight and opposite that predicted by Krakow *et al.*'s direct realist theory of speech perception. Progress toward understanding this disagreement can be made by noting two important differences between trading relation experiments like theirs and studies of perceptual integrality using the Garner paradigm.

First, in trading relation experiments, the two stimulus dimensions are treated asymmetrically, typically because one dimension is believed (*a priori*) to contribute more to perceiving the contrast under investigation than the other. Listeners in the trading relation paradigm are asked to judge the stimuli just with respect to the principal dimension, which is varied across a broad range. The perceptual value of the other dimension, which is often varied more narrowly, is assessed indirectly from its effect on judgments of the first dimension, that is, by measuring boundary shifts.

Second, such boundary shifts can arise either because the variation along the secondary dimension changes perceptual spacing along the primary dimension of judgment (the usual interpretation), or because the secondary dimension's variation changes the relative willingness of the observer to use the available responses. Common practice is to report only how variation on the secondary dimension changes the location of crossover points relative to the primary dimension, but this interpretation of trading relations results allows no distinction to be drawn between sensitivity and bias interpretations.

In the Garner paradigm, the stimulus dimensions are completely symmetric, and when accuracy is used as the dependent measure (rather than reaction time), both sensitivity and bias can be assessed. In our accuracy version of the paradigm, the range for both variables is just a jnd, the same order of magnitude as a single step along a dimension in a trading relations experiment. The two stimuli do not necessarily represent different phonetic categories, nor does their phonetic identity necessarily determine listeners' success in classifying them.

Krakow *et al.* tested for an effect of nasal coupling on the categorization of vowels differing in tongue height. They found that, in the context of following oral stops, an increase in nasal coupling shifted the category boundary between [ɛ] and [æ] toward [ɛ], that is, caused more of the tongue height continuum to be judged as [æ]. With a following nasal, on the other hand, the boundary between [ɛ] and [æ] fell in the same place as when no nasal coupling were present. In short, nasal coupling and the oral/nasal contrast in the following consonant both affected the likelihood that Krakow *et al.*'s listeners would respond "ɛ" rather than "æ." Clearly, the results of Krakow *et al.* might illustrate a pure response-bias effect. In that case, vowels labeled "ɛ" should be just as easily distinguished from vowels labeled "æ" under all degrees of nasal coupling and before nasal as well as oral consonants. If their result reflects a change in sensitivity as well as bias, however, differences in classification will appear.

Our data, on the other hand, are not susceptible to a response-bias interpretation. Listeners were generally unbiased, and what small biases they produced in making F_1 classifications were not in any way analogous to those exhibited by Krakow *et al.*'s listeners. Bias did not change with the degree of nasalization in the vowel nor whether the following consonant was nasal or oral. More important, all of our evidence regarding the integrality of F_1 and N comes from changes in estimated sensitivity between tasks, rotations, and contexts. Thus it could simply be the case that the two paradigms measure different aspects of listeners' perceptual experience of multidimensional stimuli like those used in both Krakow *et al.*'s and our experiments.

B. A common perceptual space

We can resolve some of these uncertainties and relate our results to those of Krakow *et al.* (and the predictions of direct realism) by examining a feature of the data suppressed in the standard treatment of both data sets: The perceptual spacing of vowels along the vowel height continuum. We then ask whether the following context affects this spacing.

1. Present data

In the present data, answering this question requires that we relax our assumption that performance on parallel baseline tasks (such as A_0 vs C_0 and B_0 vs D_0) is the same. (This assumption allowed us to construct parallelograms, which were then used to measure the extent of mean integrality.) If the d''s for parallel baseline tasks were different, then the corresponding spatial arrangement of underlying distributions would be affected; as shown in Fig. 12, the means might form a trapezoid rather than a square. The figure illustrates two hypothetical cases: In the upper panel, d''s for the two nasalization classifications, A_0 vs B_0 and C_0 vs D_0 are equal, so the difference $d'(AB)-d'(CD)$ is equal to 0; in the lower panel, classification of the heavily nasalized B_0 and D_0 for F_1 is substantially better than of the lightly nasalized A_0 and C_0, and as a result the difference $d'(AC)-d'(BD)$ is less than 0.

Each panel in Fig. 13 contains a trapezoid contrasting sensitivity to F_1 differences for low and high N contextualized vowels, as well as a comparable pair of trapezoids for

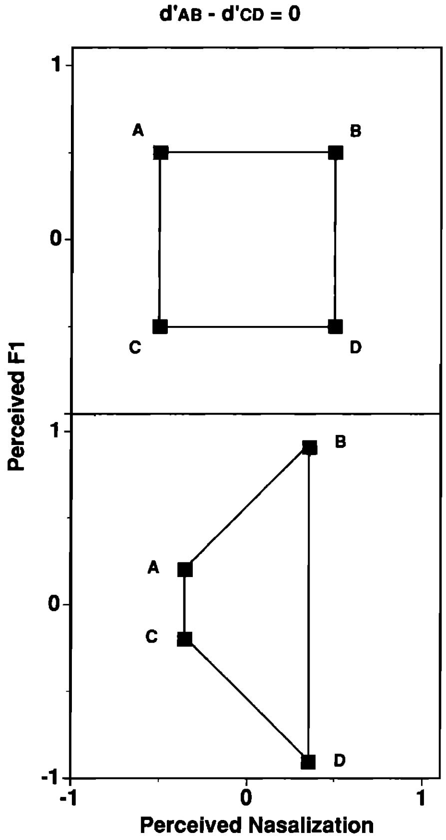

FIG. 12. Decision spaces illustrating two possible relations between d''s for parallel baseline tasks. Top: both pairs of d''s for parallel baseline tasks are equal. Bottom: d''s for the two "nasalization" baseline tasks, $d'(AB)$ and $d'(CD)$, are approximately equal, but d' for the "F_1" baseline task with heavily nasalized vowels, $d'(BD)$, is markedly larger than that for the lightly nasalized vowels, $d'(AC)$.

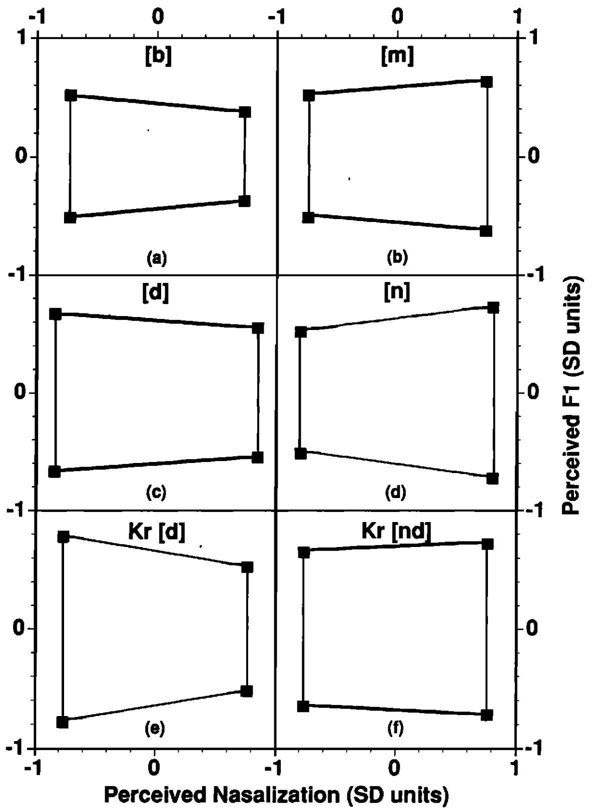

FIG. 13. Trapezoids representing inequalities between F_1 or vowel height classifications in lightly versus heavily nasalized vowels by following oral (left) versus nasal (right) context in the present experiment I (a)–(d) and in Krakow *et al.*'s (1988) data (e)–(f). Nasalization d''s for (a)–(d) are averages of $d'(AB)$ and $d'(CD)$ for the given context, while for Krakow *et al.*'s data, the averages across all contexts from experiment I are used, because Krakow *et al.* provided no data on their listeners' ability to detect nasalization differences.

Krakow *et al.*'s data.[12] When our data are plotted in this way [Fig. 13(a)–(d)], it is apparent that consonant nasality consistently affected how N influences sensitivity to F_1 differences: the high N vowels (B_0 and D_0) were more easily classified than the low N vowels (A_0 and C_0) before nasal than oral consonants, whereas in the oral context, the low N vowels are more easily classified than the high N ones.

Our data do, therefore, display context effects. The contrasting effects of the oral/nasal context on F_1 classifications for low versus high N vowels have two implications. First, the improved classification of high N vowels for F_1 differences in nasal contexts shows that N may be dis-integrated from F_1 after all, at least to the extent that it reduces perceived F_1 differences, when vowels' N values can be attributed to coarticulation with the following nasal consonant. Second, these effects show that attributing the N value to context affects sensitivity. It remains to be shown that the data of Krakow *et al.* exhibit analogous changes in sensitivity with context, along with the boundary shifts that they report.

2. Krakow et al.'s data

Krakow *et al.* reported identification functions along the tongue height continuum (which we treat as analogous to our F_1 continuum), and summarized them by their 50% points. We use these functions instead to estimate sensitivities between adjacent stimuli in the region of the [ɛ–æ] continuum where identification changed most rapidly. Assuming underlying normal distributions, we estimated d''s by calculating differences between z scores (see Macmillan and Creelman, 1991, Chap. 9). We considered the smallest amount of nasal cavity coupling used by Krakow *et al.* to be equivalent to our low N vowels (A_0 and C_0), and the largest amount of coupling to be equivalent to our high N vowels (B_0 and D_0).

To compare these inferred sensitivities with ours, we plotted analogous trapezoids in Fig. 13(e),(f). As with our data, the difference in sensitivity to vowel height differences between heavily and lightly nasalized vowels was greater before nasal than before oral consonants. Thus Krakow *et al.*'s manipulations of context produced a change in sensitivity as well as the change in response bias sought in traditional trading relations experiments. Furthermore, this effect is analogous to the asymmetry observed in our data. It is comforting that, despite rather substantial differences in method and stimulus construction, the two experiments are apparently assessing (some of) the same phenomena. That we found no bias in our results analogous to the one Krakow *et al.* obtained probably reflects the fact that our listeners were presented with stimuli differing by just a jnd and were not asked to assign them to phonetic categories.

VI. DIRECT REALISM AND AUDITORY ENHANCEMENT

How do the results and detection-theoretic interpretations presented above advance our understanding of how speech sounds are perceived? In particular, how do they accord with the predictions of the direct realist (Fowler and Smith, 1986; Krakow *et al.*, 1988) and auditory enhancement (Diehl and Kluender, 1989; also Diehl *et al.*, 1990) theories of speech perception? A complete account of our data incorporates parts of both points of view.

A. Direct realism

Direct realism allows that the acoustic properties of overlapping articulations may integrate if no information is available to the listener that the overlap is due to coarticulation of one speech sound with another. Integration is specifically predicted not to occur if such information about coarticulation is available. In accord with these predictions, Krakow *et al.* (1988) observed that increasing nasal coupling shifted tongue height judgments toward [ɛ] along an [ɛ–æ] continuum when an oral stop immediately followed the vowel, but not when a nasal did. This makes direct-realist sense, because a nasal could serve as a coarticulatory source for the nasal coupling. Our experiments have shown, however, that the acoustic correlates of varying tongue height and nasal coupling, F_1 and N, are at least as dependent, that is, mean integral, before nasal as oral consonants (and markedly more so in context than in isolation).

In Sec. V A, we distinguished between sensitivity and response-bias meanings of perceptual integration. We argued that the Garner paradigm measures integration in the form of changes in sensitivity, whereas the principal datum abstracted from the trading relations paradigm, the boundary shift, is an estimate of response bias. Thus the integration of N with F_1, as revealed by our sensitivity data, is not inconsistent with the conclusion, from Krakow *et al.*'s response-bias data, that information about coarticulation can influence the identification of vowel height. To our knowledge, direct realist theory has not made explicit predictions about sensitivity as well as bias, but the contrasting effects of oral vs nasal context on listeners' success in classifying vowels of different degrees of nasalization for F_1 in both Krakow *et al.*'s data and ours appear to accord with this theory's predictions.

Our mean integrality results make clear that integration alters sensitivity, perhaps because the acoustic consequences of tongue lowering and nasal coupling are auditorily similar. Sensitivity apparently depends on whether the two articulations work in concert or at cross purposes in changing a vowel's spectral center of gravity. Integration may also depend on whether the listener has information indicating that the overlapping articulations arise from different speech sounds and only overlap as a result of coarticulation. That such information could undo integration and prevent changes in sensitivity is one possible prediction of direct realism. Insofar as our mean-integrality measure is concerned, it is clear, however, that our listeners were not able to separate N from F_1 any more in the nasal than oral context.

We conjecture that direct realism predicts that a coarticulatory context changes how likely listeners are to assign stimuli to one category rather than another, but not how well they can tell one stimulus from another; that is, context affects identification (i.e., bias) but not classification (i.e., sensitivity). Direct realism would thus allow the same poor classification of nasalized vowels differing in F_1 in a nasal as an oral context, but a difference between contexts in whether the vowels were heard as lower or higher.

By abstracting sensitivity measures from Krakow *et al.*'s data (Sec. V B), we showed that coarticulation information *does* affect the classification of vowels differing in height as well as their identification with particular phonetic categories. The same pattern of results is present in the Krakow *et al.* study and ours: Before nasal but not oral consonants, heavily nasalized vowels differing in tongue height or its acoustic correlate F_1 are more easily distinguished than analogous lightly nasalized vowels.

Mean integrality is essentially as extreme before nasal as oral consonants in our data, but it is still possible that our listeners have not separated N from F_1 when a coarticulatory source is available. If hearing a following nasal consonant leads listeners to attribute (some of) the nasalization in the vowel to its context, they might discount some of the lowering the vowel's center of gravity caused by the prominent N_1 in the heavily nasalized vowels (Fig. 3). The consequently greater prominence of F_1 in a nasal context would raise the lowest spectral prominence's perceived center of gravity. In a vowel in which F_1 is also raised, the center of gravity would be that much higher. Therefore, a lightly nasalized vowel with a high F_1, i.e., A_0, would still differ more in center of gravity from a heavily nasalized vowel with a low F_1, i.e., D_0, than vowels with the other combinations of these dimensions, i.e., B_0 and C_0. (These inferred effects on center of gravity do not explain why a reduction in N_1's perceived prominence in a nasal context would make lightly nasalized vowels harder to classify for differences in F_1 in that context.)

In summary, both experiments reveal integration between F_1 and N in the sense that spacing along a single perceptual dimension is affected by both variables. The contrasting effects of oral and nasal contexts on listeners' success in classifying vowels of different degrees of nasalization for F_1 differences appear to accord with direct realist predictions. This comparison of results from different paradigms points up the need for theorists to make predictions about how perceptual integration in multidimensional stimuli affects both sensitivity and bias.

B. Auditory enhancement

One of our most striking results was the extent to which classification of a particular set of stimuli could be predicted from the axis along which the stimuli differed. In particular, when the axis was F_1-N, then the stimuli were invariably easier to classify than when the axis was F_1+N. If as these results, and our vowels' spectra suggest, we used heavy rather than moderate nasalization, then the direction of mean integrality that we obtained does not disconfirm Diehl *et al.*'s (1990) hypothesis that nasalization enhances a vowel height contrast when it combines with a *higher* rather than a lower F_1. Their hypothesis concerned moderate rather than heavy nasalization, which they suggest would broaden the bandwidth of the vowel's lower spectral prominence (as well as raising its F_1) but not lower its center of gravity.

C. A rapprochement

Our results also support a more basic claim of Diehl *et al.*'s auditory enhancement theory: That the acoustic correlates of independent articulations can interact perceptually with one another. The theory predicts that dimensions such as F_1 and N could be integral and their sum and difference separable, as our results have robustly shown. Furthermore, the perceptual interactions observed between F_1 and N occur in contextualized as well as isolated vowels, and before nasals as much as oral stops. These interactions occur even when higher-level information is available from context, as would be expected if they were produced by a general auditory mechanism. We hypothesize that such mechanisms respond to measurable psychoacoustic properties of the signal, here differences in the center of gravity of the vowel's lowest spectral prominence. According to Diehl *et al.* (1990; see also Diehl and Kluender, 1989; Diehl and Kingston, 1991; Kingston and Diehl, 1994, in press), it is just such mechanisms that mediate between the acoustic correlates and perceptual properties of speech sounds. Our results point to the need to separate integration as an automatic auditory process from whatever perceptual benefits or deficits may accrue from it.

When considered in light of Krakow *et al.*'s (1988) results, our results also point to the need to distinguish the automatic integration of N with F_1, which produced center of gravity differences in our stimuli, from decisional processes that may allow the listener to hear the vowel's intended height when the context provides a coarticulatory source for nasalization. These decisional processes do not undo the integration of N with F_1 that alters the vowel's center of gravity, i.e., N_1 is not heard as a distinct spectral component from F_1. Instead, the decisional processes allow the listener to discount how N_1's prominence alters the vowel's center of gravity, and they work because the listener knows how coarticulation with that context changes the vowel's center of gravity.

ACKNOWLEDGMENTS

Some of the results discussed in this paper were presented at the 125th meeting the Acoustical Society of America [Kingston and Macmillan, J. Acoust. Soc. Am. **93**, 2424–2425(A) (1993)]. This work was supported by NIH Grant No. RR07048-26 to the first author and NSF Grant No. DBS 92-12043 to the second author. We are grateful to Richard Freyman for the use of his laboratory. For very useful comments, we thank him and the other members of the University of Massachusetts Hearing Group: Rachel Clifton, Daniel McCall, Uma Balakrishnan, and Karen Helfer, as well our colleagues in the Phonetics Laboratory: Christine Bartels, José Benkí, Rachel Thorburn, and Laura Walsh. We also thank the two reviewers, Pam Beddor and Bert Schouten, whose reactions and challenges have considerably improved this paper. The contents remain our responsibility.

---

[1] Bell-Berti *et al.*'s (1979) data show clearly that the soft palate is lower in low than high vowels even in entirely oral contexts (see also Clumeck, 1976). However, because the soft palate never gets as low in oral vowels as it does in distinctively nasalized vowels or nasal consonants (Clumeck, 1976; Bell-Berti *et al.*, 1979; Cohn, 1990), even the vowel with the lowest tongue height is at most moderately nasalized.

[2] Massaro (e.g., Massaro and Cohen, 1976) and Nearey (1990) have also manipulated stimulus dimensions symmetrically in a variety of experiments.

[3] However, the divided attention task is important for diagnosing a type of perceptual interaction called "configurality" (Pomerantz and Schwaitzberg, 1975; Pomerantz, 1981). We discuss configurality in a later section.

[4] These included an additional set of four vowels used in experiment II (see Table IX). The consonant identification task thus employed 64 stimuli, 8 vowels×2 initial consonants×4 final consonants.

[5] The place confusion between the nasals [m] and [n], but not oral stops [b] and [d], also contributes to the significance of this three-way interaction.

[6] Past use of the Garner paradigm has turned up many patterns of results other than the pairing of redundancy gain with filter loss that Garner argued indicated integrality vs the joint absence of these effects that is supposed to indicate separability. Most investigators have interpreted such data as evidence of some modified form of perceptual interaction. Examples from the speech perception literature include Wood (1974), Eimas *et al.* (1981), and Green and Kuhl (1991).

[7] Reviewer P. Beddor observes that the soft palate may not rise and fall incrementally with the tongue in naturally produced vowels, but instead that high vowels may have the highest soft palate positions and low vowels the lowest, with mid vowels idiosyncratically resembling either the high or the low vowels. If the covariation in soft palate height with tongue height is thus quasicategorical rather than continuous, it would more closely resemble the other articulations that covary with the height of the tongue, e.g., rate of vocal fold vibration and degree of lip rounding (see Kingston, 1991, for discussion). But quasicategorical rather than continuous covariation does not mean that the covariation could not be controlled to produce combinations of acoustic properties that are mutually enhancing. In the particular case at hand, a positive correlation of soft palate height with tongue height would produce just such an effect.

[8] We also calculated the bandwidths of these spectral prominences 5 dB down from the most intense frequency; they are: A_0 385 Hz, B_0 87 Hz, C_0 328 Hz, and D_0 160 Hz. The resulting ranking of bandwidths: $A_0 > C_0 > D_0 > B_0$ does not predict the differences between the two correlated tasks nearly as well as the differences in center of gravity.

[9] Ordinarily when response time is the dependent measure, the stimuli differ by many jnd's on each dimension rather than little more than a single jnd, as when accuracy is the dependent measure. Despite this difference, response time and accuracy may measure the same phenomenon: in many studies (for example, those by Melara and Marks, 1990), reaction times and error rates are strongly positively correlated, i.e., faster responses are more accurate.

[10] Kingston (1991) also collected listeners' responses to a 22.5-deg array of isolated vowels; to save space, these are not discussed in this paper.

[11] Consistent performance across rotations is predicted even if listeners classify the stimuli from the two rotations in terms of primary axes, i.e., in city-block distances, rather than in Euclidean distances along whatever axes they differ. If, for example, N and F_1 are the perceptually primary axes, then the perceptual distance between stimuli along the nonprimary axis $F_1 - N$ is predicted from the city-block metric to be the sum of the distances along the F_1 and N axes. For A_0 and D_0, this distance is $\sqrt{2}$ if both baseline distances are 1, and for A_{45} and B_{45} (or C_{45} and D_{45}), this distance should be $\sqrt{[(1/\sqrt{2})^2 + (1/\sqrt{2})^2]} = 1$. That is, the ratio of correlated distances at one rotation to parallel baseline differences at the other is predicted to be $\sqrt{2}$:1 even with perceptually primary axes.

[12] To simplify the displays, we assumed for our data that sensitivity to differences in N is equal for vowels differing in F_1, and arbitrarily used the mean d' for N classifications. An arbitrary value is also used for nasalization classifications for Krakow *et al.*'s data because they provide no information on sensitivity to these differences.

Ashby, F. G., and Maddox, W. T. (**1994**). "A response time theory of separability and integrality in speeded classification," J. Math. Psychol. (in press).

Ashby, F. G., and Townsend, J. T. (**1986**). "Varieties of perceptual independence," Psychol. Rev. **93**, 154–179.

Beddor, P. S. (**1983**). "Phonological and phonetic effects of nasalization on vowel height," Indiana University Linguistics Club, Bloomington.

Beddor, P. S. (**1993**). "The perception of nasal vowels," in *Phonetics and Phonology, Vol. 5, Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic, San Diego, CA), pp. 171–196.

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., Krakow, R. A., and Goldstein, L. M. (**1986**). "Perceptual constraints and phonological change: a study of nasal vowel height," Phonol. Yearbook **3**, 197–217.

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

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

Clumeck, H. (**1976**). "Patterns of soft palate movements in six languages," J. Phon. **4**, 337–351.

Cohn, A. C. (**1990**). "Phonetic and phonological rules of nasalization," University of California, Los Angeles, Work. Pap. Phon. **76**.

Diehl, R. L., and Kingston, J. (**1991**). "Phonetic covariation as auditory enhancement: the case of the [+voice]/[−voice] distinction," Perilus **15**, 139–143.

Diehl, R. L., and Kluender, K. R. (**1989**). "On the objects of speech perception," Ecol. Psychol. **1**, 121–144.

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.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

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.

Eimas, P. D., Tartter, V. C., and Miller, J. L. (**1981**). "Dependency relations during the processing of speech," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 283–309.

Eriksen, C. W., and Hake, H. W. (**1955**). "Multidimensional stimulus differences and accuracy of discrimination," J. Exp. Psychol. **50**, 153–160.

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

Flanagan, J. L. (**1957**). "Estimates of the maximum precision necessary in quantizing certain "dimensions" of vowel sounds," J. Acoust. Soc. Am. **29**, 533–534.

Fowler, C. A. (**1986**). "An event approach to the study of speech perception from a direct-realist perspective," J. Phon. **14**, 3–28.

Fowler, C. A. (**1990**). "Sound-producing sources as objects of speech perception: Rate normalization and nonspeech perception," J. Acoust. Soc. Am. **88**, 1236–1249.

Fowler, C. A. (**1991**). "Auditory perception is not special: We see the world, we feel the world, we hear the world," J. Acoust. Soc. Am. **89**, 2910–2915.

Fowler, C. A., and Smith, M. R. (**1986**). "Speech perception as "vector analysis": An approach to the problems of invariance and segmentation," in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 123–139.

Fujimura, O. (**1960**). "Spectra of nasalized vowels," Massachusetts Institute of Technology, Research Laboratory in Electronics, Q. Prog. Rep. **58**, 214–218.

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

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

Graham, N. V. (**1989**). *Visual Pattern Analyzers* (Oxford U.P., New York).

Grau, J. W., and Kemler Nelson, D. G. (**1988**). "The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness," J. Exp. Psychol.: Gen. **117**, 347–370.

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

Green, K. P., and Kuhl, P. K. (**1991**). "Integral processing of visual place and auditory voicing information during phonetic perception," J. Exp. Psychol.: Hum. Percept. Perf. **17**, 278–288.

Hawks, J. W. (**1994**). "Difference limens for formant patterns of vowel sounds," J. Acoust. Soc. Am. **95**, 1074–1084.

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

Kawasaki, H. (**1978**). "The perceived nasality of vowels with gradual attenuation of adjacent nasal consonants," J. Acoust. Soc. Am. Suppl. I **64**, S19.

Kemler Nelson, D. G. (**1993**). "Processing integral dimensions: The whole view," J. Exp. Psychol.: Hum. Percept. Perf. **19**, 1105–1113.

Kewley-Port, D. and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

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

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Lg. **70**, 419–454.

Kingston, J., and Diehl, R. L. (**in press**). "Intermediate properties in the perception of distinctive feature values," *In Papers in Laboratory Phonology IV*, edited by A. Arvaniti and B. Connell (Cambridge U.P., Cambridge, UK).

Kingston, J., and Macmillan, N. A. (**1993**). "Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants," J. Acoust. Soc. Am. **93**, 2424–2425(A).

Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Krakow, R. A., and Beddor, P. S. (**1991**). "Coarticulation and the perception of nasality," in *Proc. 12th Intl. Cong. Phon. Sci.* (Université 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).

Macmillan, N. A., Goldberg, R. F., and Braida, L. D. (**1988**). "Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua," J. Acoust. Soc. Am. **84**, 1262–1280.

Macmillan, N. A. and Kingston, J. (**in preparation**). "Integrality, separability, and configurality: The psychophysics of the Garner paradigm."

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 *Phonetics and Phonology, Vol. 5 Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic, San Diego), pp. 147–167.

Manuel, S. Y. (**1991**). "Some phonetic bases for the relative malleability of syllable-final versus syllable-initial consonants," in *Proc. 12th Int. Cong. Phon. Sci.* (Université de Province, Aix-en-Provence), Vol. 5, pp. 118–121.

Massaro, D. W., and Cohen, M. L. (**1976**). "The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction," J. Acoust. Soc. Am. **60**, 704–717.

Melara, R. D., and Marks, L. E. (**1990**). "Perceptual primacy of dimensions: Support for a model of dimensional interaction," J. Exp. Psychol.: Hum. Percept. Perf. **16**, 398–414.

Melara, R. D., Marks, L. E., and Potts, B. C. (**1993a**). "Primacy of dimensions in color perception," J. Exp. Psychol.: Hum. Percept. Perf. **19**, 1082–1104.

Melara, R. D., Marks, L. E., and Potts, B. C. (**1993b**). "Early holistic processing or dimensional similarity?" J. Exp. Psychol.: Hum. Percept. Perf. **19**, 1114–1120.

Melara, R. D., and O'Brien, T. P. (**1987**). "Interaction between synesthetically corresponding dimensions," J. Exp. Psychol.: Gen. **116**, 323–336.

Nearey, T. M. (**1990**). "The segment as a unit of speech perception," J. Phon. **18**, 347–373.

Pomerantz, J. R. (**1981**). "Perceptual organization in information processing," in *Perceptual Organization*, edited by M. Kubovy and J. R. Pomerantz (Erlbaum, Hillsdale, NJ), pp. 141–180.

Pomerantz, J. R., and Schwaitzberg, S. D. (**1975**). "Grouping by proximity: Selective attention measures," Percept. Psychophys. **18**, 355–361.

Posner, M. I. (**1964**). "Information reduction in the analysis of sequential tasks," Psychol. Rev. **71**, 491–504.

Smith, L. B., and Kemler, D. G. (**1978**). "Levels of experienced dimensionality in children and adults," Cog. Psychol. **10**, 502–532.

Stevens, K. N., and Blumstein, S. E. (**1978**). "Invariant cues to place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Stevens, K. N., Fant, G., and Hawkins, S. (**1987**). "Some acoustic and perceptual correlates of nasal vowels," in *In Honour of Ilse Lehiste*, edited by R. Channon and L. Shockey (Foris, Dordrecht), pp. 241–254.

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. **90**, 1309–1325.

Wood, C. C. (**1974**). "Parallel processing of auditory and phonetic information in speech discrimination," Percept. Psychophys. **15**, 501–508.

Wright, J. T. (**1986**). "The behavior of nasalized vowels in the perceptual vowel space," *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, Orlando, FL), pp. 45–67.