Integrating Articulations in the Perception of Vowel Height

John Kingston
Department of Linguistics, South College, University of Massachusetts, Amherst, Mass., USA

Abstract. In vowels contrasting for height, a large number of articulations covary with tongue height, which is supposed to be the principal bearer of the contrast. However, attempts to link these covarying articulations to tongue movement physiologically have been largely unsuccessful, and the particular pattern of covariation appears to make more sense as a concerted effort to influence the perceived height of $F_1$. The experiments reported here used the Garner interference paradigm, modified to assess the perceptual primacy of stimulus dimensions, to show that the acoustic effects of two of these covarying articulations, velum height (nasalization) and rate of vocal fold vibration (pitch) are integrated perceptually with the acoustic effects of varying tongue height. This perceptual integration suggests that the different articulations are not independently perceived, contrary to the predictions of direct realist theories of speech perception, that articulatory events covary so as to enhance each other's perceptual effects, and that the surface phonological or initial phonetic representation of vowels might be quite richly specified (contrary to claims of phonetic underspecification).

1. Introduction

Many articulatory settings covary directly with the height of the tongue in vowels: the higher the vowel, the more advanced the tongue root [Lindau, 1975, 1978; 1979; Jackson, 1988], the higher the jaw [Lindblom and Sundberg, 1971], the more protruded the lips in back vowels and the more retracted the lips in front vowels [Linker, 1982], the higher the velum [Moll and Shriner, 1967; Lubker, 1968; Bell-Berti, 1976; Bell-Berti et al., 1979], and the higher the rate of vocal fold vibration [Peterson and Barney, 1952; House and Fairbanks, 1953; Lehiste and Peterson, 1961; Honda, 1987; Steele, 1986; Ohala and Eukel, 1987; Silverman, 1987]. Duration, on the other hand, varies inversely with tongue height [Lindblom, 1967; Lehiste, 1970; Westbury and Keating, 1980; Keating, 1985]. The conventional view is that these other articulatory settings are secondary to the height of the tongue in distinguishing one vowel from
another, and may even be mechanical consequences of moving the tongue [for an argument of this type, see Ohala and Eukel, 1987]. Here, it will be argued that vowel 'height' is instead an integration of all these settings, with no priority, for articulatory control nor acoustic goals, given to the height of the tongue. The basis of this claim is that, aside from the duration difference, and the labial articulations, which affect \( F_3 \) and thus the percept of vowel backness most strongly, all these articulatory settings converge on a single effect, the manipulation of the perceived \( F_1 \). (Perceived \( F_1 \) is used as shorthand in this paper for the listener's response to the distribution of energy in the lower part of the spectrum, which contains the first resonance, as well as the acoustic effects of a variety of other articulations.) It is to bring about this convergence that all these articulatory events covary as they do, and not some purely physiological dependency between them [see DiehI and Kluender, 1989, and DiehI, this volume, for similar views].

The available evidence indicates in fact that the articulatory settings that covary with tongue height are independently controlled. Many languages of East and West Africa exhibit contrasts for the position of the tongue root as well as for tongue height [Hall et al., 1974; Lindau, 1975, 1978, 1979; Jackson, 1988]; the velum is actively raised in higher vowels [Lubker, 1968]; \( F_0 \) is actively elevated through the contraction of the cricothyroid in higher vowels [Vilkman et al., 1989; Honda and Fujimura, 1989] and \( F_0 \) differs in the same direction between vowels of different heights in esophageal as well as laryngeal speech [Gandour and Weinberg, 1980; Fox and Trudeau, 1988]; and speakers differ in how much of the vertical position of the tongue is achieved by jaw as opposed to tongue raising or lowering [Ladefoged, 1990]. These demonstrations of independent control undermine any attempt to explain the observed covariation which yields the other articulations physiologically to the movement of the tongue. Explanation of a different sort must be sought, like the one argued for here that unites these articulations in their effects on the perceived frequency of the lowest resonance of the vowel.

The convergence of the acoustic effects of these articulations could indicate that the distinctive features that represent vowel height contrasts, [high] and [low], of Clements's [1989] feature [open], do not despite their names, refer simply to tongue height or degree of constriction, but are instead labels for acoustic or perceptual values [for similar views, see Ladefoged, 1980; Disner, 1983]. In a weak sense of this view, vowel height, i.e., perceived \( F_1 \), is always and primarily a function of tongue height, even though the other articulations that covary with tongue height enhance the perceived difference in \( F_1 \) of vowels produced with different tongue heights. In the strong sense of this view, on the other hand, vowel height contrasts do not necessarily imply differences in tongue height; instead the acoustic effects of each of the covarying articulations contribute equally to the perceived \( F_1 \). Height then is an abstract scale representing the covariation among these articulations jointly. Evidence from the behavior of vowels in phonological inventories, rules, and changes in support of this stronger view will be outlined in the last section of this paper.

If vowel height is an intonations in the phonetics, of all these articulatory events, of of direct perception of, and acoustic effect also undern of speech gestures does a possibility that during any in the signal it may be un an acoustic property is the gesture that produces it is segment or because of with gestures of adjacent however, this theory does argu will always be resolve a is evaluated that is long to reveal any coarticulatory's assume articulations listed above articulatory origins, and if they have the same acoustic effect cannot be attributed, ex to just one of them. That all other than tongue height acoustic properties of the vo e.g., its pitch, nasality, etc., unless the listener can re them perceptually (in any height itself affects more the tual separation would ind acoustic effects of the articu vary with tongue height m different tongue heights edgish simply by adding to t between them. (Such a theor course, provide any explain these other articulations of
the vertical position of the jaw as opposed to that of lowering [Ladefoged, demonstrations of independence any attempt to ex- ered coarticulation which yokes articulations physiologically to the tongue. Explanations sort must be sought, like the one here that unifies these antithetical effects on the perceived lowest resonance of the
tiveness of the acoustic effects not the features that represent (high) and (low) of [voice] features [open], do not names, refer simply to tongue degrees of constriction, but are for acoustic or perceptual similar views, see Ladefoged, 1983]. In a weak sense of vel height, i.e., perceived as a function of even though the other articulations with tongue height perceived difference in F1, used with different tongue the strong sense of this view, hand, vowel height contras- sarily imply differences in instead the acoustic effects covarying articulations covary to the perceived F1. Height
dictate scale representing the among these articulations- from the behavior of onological inventories, rules, in support of this stronger outline in the last section of

tently in the way they do.) The experiments reported in this paper are a first attempt to test whether the acoustic effects of these covarying articulations are instead integrated in the perception of vowel height, and the extent to which listeners can separate their articulatory origins.

Looking for this sort of integration is a test of a strong version of the auditory enhancement theory argued for by Diehl and his colleagues [Kluender et al., 1988; Diehl and Kluender, 1989; Diehl and Walsh, 1989, Diehl, this volume]. This theory claims that articulations covary in speech because the perceptual effect of one articulation exaggerates the perceptual value of another so that that speech sound becomes more distinct from one in which the two articulations covary in the opposite direction. For example, vowels are lengthened before [+voice] stop closures to exaggerate the perceived shortness of what is already a shorter closure than in [−voice] stops [Kluender et al., 1988]. The perceived duration of the [−voice] stop closure would of course be lengthened by the shorter vowel that precedes it. The perceptual effect of vowel duration on judgments of following stop closure duration is attributed to a general auditory mechanism Kluender et al. [1988] call 'durational contrast', and proponents of this theory argue that many cases of enhancement can be attributed to such mechanisms. This paper does not address the issue of whether enhancement of one perceptual effect by another arises out of such general auditory mechanisms, but instead the more general prediction of enhancement theory that the perceptual effects of covarying articulations merge into a single, more distinct perceptual object.
2. Methods and Stimuli

2.1. What Kind of Experimental Design Is Needed?

The hypothesis that all the articulations which covary with tongue height do so because their perceptual effects are the same suggests that listeners do not actually hear them as separable articulations, or at least not as articulations associated with any other contrast than vowel height. Accordingly, what is needed is an experimental design which allows a test of the extent to which the acoustic effects of these various articulations are perceptually integrated. More precisely, what is needed is a test of whether the perception of the value of a vowel has along an acoustic dimension reflecting one articulation is influenced by its value along an acoustic dimension determined by another, covarying articulation. Without a demonstration that a vowel's perceptual value along one acoustic dimension depends on its acoustic value along another (vice versa), then the only motivation for covarying articulations is that differences along more than one dimension at a time make vowels more distinct. Of interest here is the stronger claim that articulations covary because their perceptual effects exceed their sum and actually converge on a single, exaggerated perceptual value, for F1.

At the same time, since each of these articulations participates in phonemic contrasts in its own right, their acoustic effects could also define orthogonal dimensions by which listeners sort vowels. For example, nasalization may make a low vowel sound lower, as a result of perceptual integration of the acoustic effects of nasalization with those of tongue height, and also simply nasalized, since that is an acoustic dimension orthogonal to vowel quality. To sum up, needed are tests of perceptual integration of orthogonal acoustic dimensions and tests of whether listeners can also classify stimuli along these dimensions individually. These two tests are provided by the Garner [1974] paradigm, adapted as suggested by Melara and Marks [1990], which is described in the next section.

2.2. Integrability vs. Separability: The Garner Paradigm

The integrability of apparently orthogonal stimulus dimensions has frequently been examined using a paradigm developed by Garner [1974]. In this paradigm, subjects are required to classify, as quickly as possible, stimuli selected from a two-by-two array defined by the dimensions whose integrability is being investigated (table 1 indicates the dimensions that were investigated in the experiments reported here).

The stimuli are supposed to be separated along each dimension so as to be equally discriminable. Stimuli are selected from this array to be presented to subjects for classification in a number of different ways (fig. 1):

1. In baseline classification, stimuli are selected from just one side of the array and they differ along just one dimension: i.e., A vs. C or B vs. D on the Tongue Height dimension, and A vs. B or C vs. D on the Nasalization or Pitch dimensions.

2. In correlated classification, stimuli are selected from opposite corners of the array, and they differ along both dimensions at once. Their values on these dimensions are either positively, B (both high) vs. C (both low), or negatively, A (low, high) vs. D (high, low), correlated.

3. In selective attention, all four stimuli in the array are selected, but the perceiver must attend selectively to just one dimension while ignoring the other. In classification for Tongue Height, A and B are contrasted with C and D, while for Nasalization or Pitch, A and C are contrasted with B and D. The orthogonal difference along the other dimension is supposed to be irrelevant to the classification.

4. In divided attention, a two positively correlated stimulus is denoted by A and D, the two orthogonal stimuli. Both dimensions are the nature of the correlational test.

If the two dimensions of the array are integrated, as they are in the correlated and divided attention tasks, then performance on these two tasks and the baseline task differ along just one dimension.
of perceptual integration of the nasalization with those of tongue
imply nasalized, since that is an orthogonal to vowel quality. To
make a test of perceptual integration across dimensions and a test of
individuals, these two tests are also classified stimuli along
orthogonally with vowel quality. These two tests are
paradigm, adapted a
ra and Marks [1990], which is de-

**ility vs. Separability:**

**Paradigm**

ity of apparently orthogonal stimuli frequently been examined
veloped by Garner [1974]. In this
test is required to classify, as e.g., stimuli selected from a two-by-
by the dimensions whose inter-
engaged (Table 1 indicates the di-
eaks investigated in the experiment
not supposed to be separated along
as to be equally discriminable,
ed from this array to be presented
ification in a number of different
classification, stimuli are selected
of the array and they thus differ in
tension: i.e., A vs. C or B vs. D in
a single dimension, and A vs. B or C
on or Pitch dimensions.
ked classification, stimuli are at
in the corners of the array, and they
dimensions at once. Their values
are either positively, B (both low), or negatively, A (low, high)
correlated.
attention, all four stimuli in theut the perceiver must divide at
one dimension while ignoring the
for Tongue Height, A and B in
d, while for Nasalization are
or D, while for Nasalization are
separable, then performance will differ between
these tasks and the baseline task, in which stimuli differ along just one dimension at a time. Integration
of dimensions means that the value a stimulus
has along one dimension changes is perceptual value along the other dimension. For example, a
high value for Nasalization could make a sound lower than its value for Tongue Height alone
lead one to expect.

Integration of dimensions means that the value of a stimulus
has along one dimension changes its perceptual value along the other dimension. For example, a
high value for Nasalization could make a vowel sound lower than its value for Tongue Height alone
lead one to expect.

Integration causes the stimuli to be classified more rapidly and accurately in the correlated than
the baseline task, because the change in the stimuli's
dimensional values along one dimension by their
valu

Fig. 1. Stimulus arrays at 0°
(left) and 45° (right), illustrating
class membership in instances of
the various tasks at the two
rotations; the stimulus arrays at 22.5°
would be intermediate. The
vertical axis is the (inverse of) Tongue
Height (F1) and the horizontal axis
is either Nasalization (N1−N2) or
Pitch (F0). Circles enclose stimuli
which belong to a class on a given
block of trials, squares enclose
stimuli that do not. At 0°, the illus-
trated Baseline and Selective
Attention tasks require judgments
of Tongue Height: the Correlated
task contrasts positively correlated
stimuli; and the Divided Attention
task contrasts positively with nega-
tively correlated stimuli.

(4) In divided attention, all four stimuli are
gain selected, but the perceiver must divide atten-
tion between the two dimensions, since B and C, the
two positively correlated stimuli, have to be disting-
ished from A and D, the two negatively corre-
lated stimuli. Both dimensions are always relevant,
as is the nature of the correlation between their
values.

If the two dimensions of the stimuli are varied at
once, as they are in the correlated, selective atten-
tion, and divided attention tasks, and if these
dimensions are integrated perceptually rather than
separable, then performance will differ between
these tasks and the baseline task, in which stimuli
differ along just one dimension at a time. Integration
of dimensions means that the value a stimulus
has along one dimension changes its perceptual value along the other dimension. For example, a
high value for Nasalization could make a vowel sound lower than its value for Tongue Height alone
lead one to expect.
stimuli. Demonstrating integration therefore requires that performance also be worse on the selective attention task compared to baseline. With perceptually separable dimensions, variation on the unattended dimension would not affect accuracy in classifying the stimuli along the attended dimension in the selective attention task. Integration reduces performance in selective attention because the supposedly irrelevant differences along the unattended dimension change the stimuli's perceived value on the attended dimension. This increases the variability among the members of each class and with it the perceivers' uncertainty about whether a stimulus belongs to a class.

Having to attend to both dimensions at once would by itself reduce performance in the divided attention task, but integration should add to this difficulty. Integration should degrade performance even more in divided attention because the mutual changes in a stimulus's value along one dimension by its value along the other will make it more difficult to detect that the stimulus's values were positively or negatively correlated. With perceptually separable dimensions, on the other hand, the stimulus's value along each dimension can be independently measured and the polarity of the correlation between the values on the two dimensions would therefore be more easily determined.

The natural pattern of covariation between Nasalization or Pitch and Tongue Height could also influence performance in these classification tasks if listeners found the vowels in which the dimensions varied in natural speech more distinct than those in which the covariation went in the opposite direction. If this is so, then vowels in which Pitch and Tongue Height are positively correlated or in which Nasalization and Tongue Height are negatively correlated should be more easily distinguished (and less easily grouped together) than vowels in which these correlations have the opposite polarity. Thus performance should be better on the positively than negatively correlated task when vowels vary in Pitch and Tongue Height, but better for the negatively than positively correlated task when they vary in Nasalization and Tongue Height. Conversely, performance should be even poorer in the divided attention task since vowels which have been made more distinct by natural covariation will have to be classed together. Performance on the baseline and selective attention tasks should not be affected by any greater distinctiveness that may arise out of natural covariation of dimensions in each of the baseline tasks, a distinction has to be drawn between a vowel which is more distinct because it exhibits natural covariation and one which is less distinct because it does not. In the selective attention tasks, each class contains only one vowel and one less distinct vowel. In both kinds of tasks, then, the effect of an additional distinctiveness may arise out of natural covariation of dimensions which is evenly distributed among the subtasks.

In summary, by altering the perceptual value of a stimulus along one dimension through its value along the other, integration should facilitate fast, accurate classification in the correlated tasks compared to the baseline task, but interfere with it in the selective and divided attention tasks. If integration also renders some of the stimuli more distinct than the others, perhaps those in which stimulus's value vary as they do naturally, then further effects will be observed in the correlated and divided attention tasks. Dimensions that are perceptually separable, neither facilitation nor interference is expected since the value a stimulus has along one dimension does not alter its perceptual value along the other.

2.3. Perceptual Primacy

2.3.1. Rotation of the Array

Melara and Marks [1990] argue that the perceptual integration of dimensions does not preclude independent evaluation of a stimulus along perceptually primary dimensions: these are the psychologically real dimensions in the stimulus space along which any stimulus is evaluated. They suggest that even if two dimensions are integrated in the perception of the stimulus as a whole, subjects may still extract the stimulus's attributes, i.e. its values along perceptually primary dimensions. Following Smith and Kemler [1978] and Grau and Kemler Nelson [1988], Melara and Marks [1990] show how rotating the array in the plane alters the perceptual interaction between its members and thereby reveals the orientation in the stimulus space corresponding to perceptually primary dimensions.

If the original orientation (0°) corresponds to perceptually primary but integrable dimensions, then rotation (fig. 1) increases the intraclass variability on the attended dimension in distracting variability on the other.

Rotation should thus cause more the selective attention task. Further rotation is another 30°, stimuli which are sorted into different classes, i.e. A vs. D, are actually identical on one selecting selective attention even more.

On the other hand, rotation causes faster and more accurate attention task. As in the selective a four stimuli from the array are provided attention task, but the subception between the two dimensions positively correlated or congruent from the negatively correlated or independent attention task arises from the attended differences, and by a pair of each class more closely along the primary perceptual dimensions. If to 22.5°, the congruent stimuli differ more on the horizontal than on the vertical, while for the incongruent the vertical difference expands a bit shrinks. With further rotation, the stimuli differ only along dimension and congruent ones a vertical. As Melara and Marks [1] this rotation the divided attention a same-different decision along and should thus be faster and more the original orientation.

When dimensions are integrally primary, the difference in response time or accuracy should be greater because perceptual primary between divided attention tasks is more consistent.

2.3.2. Congruence

Melara and Marks [1990] define perceptual congruity, the vs. incongruent variation 10° or at issue is whether perceivers can...
Integrating Articulations

on the attended dimension in addition to the
displacing variability on the other dimension.

Rotation should thus cause more interference in
the selective attention task. Furthermore, when the
array is rotated 45°, stimuli which are supposed to
be sorted into different classes, i.e. A vs. D and B
vs. C, are actually identical on one dimension, mak-
ing selective attention even more difficult.

On the other hand, rotation should make re-


2.4. The Stimuli

The perceptual integrability of the acoustic ef-


The effect on formant frequencies of different
degrees of tongue height, lip, and larynx position
were determined through the use of a software
model, Vocal 2.1 (kindly provided by Peter Ladef-
gedt), which allows simulation of vocal tract
shapes, transforms them to area functions, and then
calculates the formant frequencies such a shape
would produce [Ladefoged, n.d.; see also Lindau,
1975, 1978, 1979; Jackson, 1988]. (Henceforth,
Tongue Height will refer simply to the position of
the tongue in the mouth, or to its principal acoustic
correlate, F1. Vowel height, on the other hand, will
be reserved to refer to the phonological contrast be-


al orientation (0°) corresponds to
inary but integrable dimensions, e.g., 1) increases the intraclass vari-
Table 2. First three formant frequencies (in Hz) resulting from simulations of three degrees of Tongue Height

<table>
<thead>
<tr>
<th>Tongue Height</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1 mm</td>
<td>385</td>
<td>1,030</td>
<td>2,360</td>
</tr>
<tr>
<td>0 mm</td>
<td>430</td>
<td>1,080</td>
<td>2,320</td>
</tr>
<tr>
<td>-1 mm</td>
<td>450</td>
<td>1,070</td>
<td>2,320</td>
</tr>
</tbody>
</table>

range from 1 mm below the neutral position to 1 mm above it, with the further assumption of 2 mm of lip protrusion, 2 mm of larynx depression, and a lip aperture 13 mm wide [see Lindblom and Sandberg, 1971, Riordan, 1977, and Linker, 1982, for discussion of these articulations and their acoustic effects.] With each millimeter of tongue raising, the tongue was advanced a third of a millimeter to simulate the more forward position of higher back vowels. The resulting values (rounded to the nearest 5 Hz) for the first three formants are given in Table 2.

Although Tongue Height clearly affects F1 and F3 as well as F0, for simplicity’s sake it was decided to employ constant values for these two formants in the stimuli; the values used were 1,050 and 2,350 Hz, respectively. Variation in F1, alone of the 65 Hz that a 2 mm difference in Tongue Height produces is more than sufficient to produce a different vowel quality. These formant frequencies were then used to synthesize vowels through the cascade branch of a Klatt synthesizer. Nasalization was synthesized by placing a nasal zero (N2) between a nasal (N1) and oral pole (F1) (the former was always 150 Hz lower than the latter); the further the nasal zero is from either of the two flanking poles, the more nasalized the resulting vowel sounds [Stevens et al., 1987]. Although nasalized vowels are typically weaker than oral ones [House and Stevens, 1956], the peak amplitude of all vowels was normalized to within 0.2 dB of one another, which would largely prevent listeners from using an amplitude difference in the nasalization by Tongue Height experiment. This was done so that performance would be determined by the effects on the spectrum of separating the nasal zero from the nasal pole by various amounts. Any reduction in intensity in the region of the first formant in the nasalized vowels is of course not affected by this normalization, so the lower intensity in that region of the spectrum which is found in natural nasalized vowels [House and Stevens, 1956] could have been used by the listeners in the Nasalization by Tongue Height experiment. Finally, F0 was simply varied directly.

Pilot experiments in which listeners discriminated between members of stimulus pairs differed by varying amounts along the dimensions of Tongue Height (the inverse of F1), Nasalization and Pitch were then run to determine rough judgments as to the detectable differences along these various dimensions. For F1, a difference of about 25 Hz was reliably detected; for Nasalization, an N2–N1 difference of 10 Hz was distinguishable from one of 50 Hz; and for Pitch, just a 5 Hz difference was detectable. Since differences are easiest to detect when the stimuli are presented side by side for comparison, and the stimuli were to be presented individually in the experiments reported here, somewhat larger differences than the minimum ones detectable were employed in constructing the stimulus arrays.

Finally, it is important to note that varying Nasalization and Pitch does not directly affect the measured frequency of F1 in these stimuli; only Tongue Height differences do. However, it is predicted that the acoustic effects of these articulations will affect the perceived frequency of F1. It is this prediction in fact that these experiments are designed to test.

The stimuli for the two sets of arrays, at various rotations, are given in Table 3. Henceforth in reference to these arrays, a distinction will be drawn between Tongue Height which increases top to bottom and F1 which increases from bottom to top. Nasalization and Pitch increase from left to right (fig. 1).

The differences in F1, the inverse of Tongue Height, are the same for the two experiments. It was set to the intermediate value of 125 Hz in all the stimuli in the Nasalization by Tongue Height experiment, while the difference between the nasal formant and the nasal zero was set to 0 in the Pitch by Tongue Height experiment, eliminating any nasalization.

In terms of the acoustic measures themselves, F1, N2–N1, and F0, the congruent stimuli in both sets of arrays would be B and C since it is these stimuli in which values along each pair of acoustic dimensions are either both high or both low, while in A and D they have opposite values in which pair of values in a way best corresponds to the given height contrasts. If more Nasal vowel sound lower for a given Tongue Height sound higher, then A and B are the distinct in vowel sound; higher. Pitch makes a vowel sound higher, then A and C are the distinct in vowel sound for the Pitch by Tongue Height arrays, and distinctness does not a. Since the contrasts of the arrays, B and C are the distinct stimuli and A are the ones in all future discussion.

Since error rate rather than recognition of performance in these vowels were made quite brief, h
Integrating Articulations

Table 3. Stimm for Nasalization by Height and Pitch by Height arrays

<table>
<thead>
<tr>
<th>Rotation</th>
<th>0°</th>
<th>22.5°</th>
<th>45°</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Nasalization by Tongue Height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F₁</td>
<td>460</td>
<td>460</td>
<td>465</td>
</tr>
<tr>
<td>N₂</td>
<td>320</td>
<td>370</td>
<td>337</td>
</tr>
<tr>
<td>N₁</td>
<td>310</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>N₂–N₁</td>
<td>10</td>
<td>60</td>
<td>22</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td>D</td>
</tr>
</tbody>
</table>

| Pitch by Tongue Height |    |       |    |      |    |      |
| F₁       | 430 | 430   | 437 | 425 | 445 | 424 |
| F₀       | 290 | 290   | 289 | 324 | 295 | 309 |
| N₂       | 280 | 280   | 287 | 275 | 295 | 274 |
| N₁       | 10  | 60    | 2   | 49  | 0   | 35  |
| F₁       |     |       | A  |     | A  |     |
| F₀       |     |       | B  |     | B  |     |

For the two sets of arrays, at the vari-

given in Table 3. Henceforth in two

distinction will be drawn between

differed in F₁, which increases from bottom on and Pitch increase from left to


cases in F₁, the inverse of Tongue


game for the two experiments.


termediate value of 125 Hz in all the

pression by Tongue Height experi-

difference between the nasal for-

he acoustic measures themselves,

ence both would be B and C since it is the

values along each pair of acoustic

either high or both low, while

from the onset of energy to its offset, in order to reduce accuracy overall. Amplitude rose steeply from 0 to 27 dB in the first 10 ms and then more gradually to 60 dB in the next 10 ms, remained at 60 dB for 25 ms, and then fell back to 0 dB in mirror image of its rise (recall that the peak amplitude in the output waveform was also normalized for all stimuli). The frequencies of the first oral and nasal formants, the nasal zero, and the fundamental were constant throughout each stimulus, and the frequencies of formants above the first as well as all formant bandwidths were constant across stimulus.

2.5. Connections among Design Components

If integration of dimensions occurs, performance on the correlated task should be faster and
more accurate than that on the baseline tasks, while performance of the selective and divided attention tasks should be slower and less accurate. If integration occurs in these experiments, it would show that the perception of vowel quality integrates differences in the acoustic signal due to velar or laryngeal articulations with those due to movements of the tongue.

Even if integration occurs, listeners may still sort the stimuli by their values along the individual dimensions. For example, a vowel may still be perceived as having a high Tongue Height when Nasalization is also high, even though its perceived Tongue Height is lower than it would be if Nasalization were low. Conversely, a vowel may still be perceived as having a low degree of Nasalization when Tongue Height is high, even though its perceived Nasalization is higher than it would be if Tongue Height were low. If listeners can still judge a stimulus’s value on these dimensions individually, even if that value is distorted by the stimulus’s value on the other dimension, then the dimensions are perceptually primary, in accord with their traditional labeling with different features, [high] and [low] Tongue Height, [nasal], and [High] (Pitch). As a result of integration, the listener’s percept of the vowel’s height is more than just an additive composite of its values for each of these dimensions, but at the same time the vowel is recognizable as being in large part composed of settings along each of these dimensions, because they are also perceptually primary.

Perceptual primacy is assessed in two ways. First, performance on selective and divided attention compared to the baseline tasks should change in opposite ways with rotation of the stimulus array away from its orientation to the original stimulus dimensions. If that orientation aligns with perceptually primary dimensions, then performance on a selective attention task should decline, since intraclass variability is increased with rotation. Performance on the divided attention task should, on the other hand, improve, since rotation reduces the intraclass variability with respect to the primary dimensions. Second, with perceptually primary dimensions, congruent stimuli should be perceived more reliably than incongruent ones, since in the congruent stimuli both stimulus dimensions vary in the same direction, while in the incongruent ones, they vary in opposite directions. If it should turn out that vowels which exhibit natural covariation for these dimensions are more distinct from one another than those which do not, then performance on tasks in which the incongruent stimuli must be distinguished should actually be better in the Pitch by Tongue Height experiment than in the Nasalization by Tongue Height experiment because congruence composite with the natural pattern of covariation of those dimensions.

2.6. Method of Presentation

The stimuli were presented to subjects seated in a sound-treated room, at comfortable levels over headphones to one ear, the right for all subjects but 1, who claimed his left ear was better. The subject ran the experiment themselves from a video display terminal through which they entered their response and controlled their progress from block to block. Each block of trials began with a series of practice trials in which the stimulus that belonged to the category were presented in alternation with those that did not. Whether a stimulus belonged to the category was indicated by the words ‘yes’ or ‘no’ on the computer screen immediately before the subject heard the stimulus. The message remained on the screen for 500 ms, and then the subject heard the stimulus. They were then to respond by striking labeled ‘yes’ and ‘no’ keys (these were the ‘1’ and ‘2’ keys of the keypad and the subjects struck them with the index fingers of their left and right hands, respectively). The next trial began 250 ms after their response. After the subjects had heard 8 repetitions of each stimulus in training, the test trials began. In each block of test trials, they heard 24 instances of each stimulus, in random order. That work had shown that to prevent subjects from responding at chance on the more difficult tasks, they needed to be reminded of what sounds belonged to the category during the test trials; accordingly, 8 of the 24 test trials were preceded by ‘yes’ or ‘no’, as in the practice trials, while the remaining 16 were preceded by ‘guess’ (which trials were preceded by ‘yes’, ‘no’ or ‘guess’ was also random). Only their performance on the 16 trials where they had to guess was included in the results. The subjects were instructed to respond as quickly as possible to both the practice and test trials, and if anything to sacrifice accuracy to speed. They were also told to give their immediate impression of which belonged to the category and not when the decision was difficult.

There were two sessions, one for the Tongue Height arrays and a second for the Pitch by Tongue Heights arrays. Half of the subjects in each session, there were 27 blocks of trials 9 tasks by 3 rotations. The 9 tasks were A vs. B, C vs. D, A vs. C, 2 correlated tasks (B vs. C and A vs. D), and 1 divided attention tasks (A and B vs. C and D). The 3 rotations were 45°. Subjects heard all 9 tasks with both, but the order in which they were determined by a Latin square order in which the 9 tasks were pre-rotation. Expect that the first task to perform on any rotation was a 45° restriction ensured that they started finding out what one of its ‘criterion’, whether a given stimulus or to the category on a given task was across subjects.

2.7. Subjects and the Practice of the Experiments

Subjects in this experiment we volunteers from the Cornell University who answered posted advertisements. 5 undergraduate and 3 grad had some background in linguistics beginning their 2nd year of graduate studies, but all were entirely ignorant of the experiment. Six of the subjects were women and the other 2 men. Each in both sessions of the experiment took approximately an hour and a half, the two sessions were days: for the first the first session was the second in the same day. Subjects paced themselves and were told that they could take as much time as they wished between blocks.
distinct from one another than those on performance tasks in which stimuli must be distinguished better in the Pitch by Tongue movement in the Nasalization by Tongue movement acoustic congruence consistent pattern of covariation of those distingui

od of Presentation

were presented to subjects seated in a room, at comfortable levels on one ear, the right for all subjects but 1 whose left ear was better. The subjects themselves from a video display which they entered their responses: their progress from block to block. Trials began with a series of practice the stimuli that belonged to the set 2 in alternation with those that were a stimulus belonged to the set 3 by presenting the words 'yes' or 'no' screen immediately before the stimulus. The message remained or 500 ms, and then the subject heard they were then to respond by striking the 'yes' or 'no' keys (these keys were the 1 through the keypad and the subjects struck the index fingers of their left and right...)

The next trial began 500 ms. After the subjects heard each stimulus in the training, the last block of test trials, they heard each stimulus, in random order. Known that to prevent subjects from 4 attention on the more difficult tasks; they mimicked of what sounds belonged to the test trials; accordingly, 2 of 4 were presented by 'yes' or 'no', and 2 of the remaining were presented by 'yes' or 'no', while the remaining 16 were presented by 'yes' in the 16 trials where they had to decide. The subjects were asked as quickly as possible to both 3 test trials, and if anything to such a speed. They were also told to give

their immediate impression of whether a stimulus belonged to the category and not to reflect even when the decision was difficult.

There were two sessions, one for the Nasalization by Tongue Height arrays and the other for the Pitch by Tongue Heights arrays. Half the subjects did Nasalization by Tongue Height first and the other half Pitch by Tongue Height first. Within each session, there were 27 blocks of trials, consisting of 9 tasks by 3 rotations. The 9 tasks were 4 baseline tasks (A vs. B, C vs. D, A vs. C, and B vs. D), 2 correlated tasks (B vs. C and A vs. D), 2 selective attention tasks (A and B vs. C and D and A and C vs. B and D), and 1 divided attention task (B and C vs. A and D), and the 3 rotations were 0, 22.5, and 45°. Subjects heard all 9 tasks within a rotation together, but the order in which they heard each rotation was determined by a Latin square, as was the order in which the 9 tasks were presented within a rotation. Expectation of the first task that a subject had to perform on any rotation was a baseline task. This restriction ensured that they started each rotation by finding out what one of its 'dimensions' was. Finally, whether a given stimulus or stimuli belonged to the category on a given task was counterbalanced across subjects.

2.7. Subjects and the Practical Conditions of the Experiments

Subjects in this experiment were 8 paid ($5/h) volunteers from the Cornell University community who answered posted advertisements: all 8 were students, 5 undergraduate and 3 graduate. Five of 8 had some background in linguistics, in fact 2 were beginning their 2nd year of graduate study of linguistics, but all were entirely ignorant of the purpose of the experiment. Six of the subjects were women and the other 2 men. Each subject took part in both sessions of the experiment, each of which took approximately an hour and a half. For all but 1 subject, the two sessions were run on different days; for that subject the first session was run in the morning and the second in the afternoon of the same day. Subjects paced themselves in a session, and were told that they could take breaks whenever they wished between blocks.

Since each stimulus was presented 16 times and there were 8 subjects, the results are based on 128 responses to each stimulus in each of the 9 tasks at each rotation. Given that there were 16 repetitions, of 4 stimuli, in 9 tasks, at 3 rotations, in 2 experiments, each subject produced 3,456 responses.

3. Results

3.1. Introduction

All subjects reported somehow that vowel quality was manipulated in both experiments. All also noticed that Pitch was manipulated in the Pitch by Tongue Height experiment, but the only subjects who described the orthogonal variable as Nasalization in the other experiment either had some linguistic training or experience with a language such as French in which nasalization is contrastive on vowels. The other subjects described some of the vowels in the Nasalization by Tongue Height experiment as muffled or otherwise obscured. These reports suggest that the subjects were able to detect the principal acoustic manipulations of these experiments, but they do not require that Nasalization, Pitch, and Tongue Height be perceptually primary dimensions. All the subjects also reported far more difficulty with the selective and divided attention blocks in which two stimuli belonged to each class rather than just one. This difficulty could reflect interference from perceptual integration, or just the greater difficulty in forming a class of two rather than just one member. Both of these ambiguities are resolved by the analysis of the subjects' performance.

The dependent measure in these experiments was error rate as a proportion of the
Table 4. Mean error rates (with standard errors), as a proportion of total responses for each combination of Task Type and Rotation: Nasalization by Tongue Height and Pitch by Tongue Height

<table>
<thead>
<tr>
<th>Task</th>
<th>Rotation</th>
<th>0°</th>
<th>22.5°</th>
<th>45°</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>SE</td>
<td>mean</td>
<td>SE</td>
<td>mean</td>
</tr>
<tr>
<td>Nasalization by Tongue Height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.148</td>
<td>0.018</td>
<td>0.160</td>
<td>0.029</td>
<td>0.151</td>
</tr>
<tr>
<td>Negative correlation</td>
<td>0.008</td>
<td>0.005</td>
<td>0.020</td>
<td>0.012</td>
<td>0.025</td>
</tr>
<tr>
<td>Positive correlation</td>
<td>0.099</td>
<td>0.042</td>
<td>0.215</td>
<td>0.071</td>
<td>0.156</td>
</tr>
<tr>
<td>Selective</td>
<td>0.190</td>
<td>0.022</td>
<td>0.287</td>
<td>0.051</td>
<td>0.282</td>
</tr>
<tr>
<td>Divided</td>
<td>0.441</td>
<td>0.027</td>
<td>0.398</td>
<td>0.030</td>
<td>0.379</td>
</tr>
<tr>
<td>mean</td>
<td>0.175</td>
<td>0.026</td>
<td>0.236</td>
<td>0.031</td>
<td>0.205</td>
</tr>
<tr>
<td>Pitch by Tongue Height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0.074</td>
<td>0.017</td>
<td>0.057</td>
<td>0.018</td>
<td>0.064</td>
</tr>
<tr>
<td>Negative correlation</td>
<td>0.035</td>
<td>0.015</td>
<td>0.020</td>
<td>0.008</td>
<td>0.031</td>
</tr>
<tr>
<td>Positive correlation</td>
<td>0.027</td>
<td>0.007</td>
<td>0.039</td>
<td>0.018</td>
<td>0.023</td>
</tr>
<tr>
<td>Selective</td>
<td>0.191</td>
<td>0.034</td>
<td>0.235</td>
<td>0.027</td>
<td>0.278</td>
</tr>
<tr>
<td>Divided</td>
<td>0.339</td>
<td>0.049</td>
<td>0.289</td>
<td>0.057</td>
<td>0.339</td>
</tr>
<tr>
<td>mean</td>
<td>0.134</td>
<td>0.023</td>
<td>0.128</td>
<td>0.022</td>
<td>0.147</td>
</tr>
</tbody>
</table>

Fig. 2. Mean error rates (with standard errors), expressed as the proportion of the responses in a given task for each subject: the effects of Rotation in the Nasalization vs. Pitch by Tongue Height experiments (a) and the effects of Task Type in the Nasalization vs. Pitch by Tongue Height experiments (b).

3.2. Assessing Perceptual Integration: Overall Error Rates

Mean error rates (with standard errors) obtained in the two experiments for each combination of Task Type and Rotation are presented in Table 4 (positive correlation refers to the task where the stimuli either had high or low values on both acoustic dimensions, B and C, while negative correlation refers to the task where the stimuli had a high value on one acoustic dimension and a low value on the other, A and D).

Differences in overall error rate were assessed in terms of a three-way ANOVA with Experiment (Nasalization or Pitch), Rotation (0, 22.5, and 45°), and Task Type (baseline, negatively correlated, positively correlated, selective attention, or divided attention) as independent variables. The main effect of differences in Experiment was quite significant, $F_{1,203} = 34.987$, $p < 0.0001$, with substantially more errors overall (51%) in the Nasalization by Tongue Height than the Pitch by Tongue Height experiment. Rotation was highly significant, $F_{2,203} = 2.000$, $p = 0.1$, and post hoc tests showed that all pairwise comparisons between Task Types, except those with positively correlated and 1 were significant at 0.05: these were significantly lower correlated than baseline tasks (but only higher for selective at higher for divided attention).
low values on both acoustic dimension A and C, while negative corre-
sponded to the task where the stimuli varied on one acoustic dimension and on the other, A and D.

Fig. 2. Mean error rates (with standard errors), expressed as the propor-
tion of the responses in a given task for each subject: the effects of Rotation in the Nasalization vs. Pitch by Tongue Height experiments (a) and the effects of Task Type in the Nasalization vs. Pitch by Tongue Height experiments (b).

Interpreting the different effects of negative vs. positive correlation requires looking at the interactions in this analysis, one of which reached significance, and another which approached it. The interaction between Experimen
t and Rotation, $F_{(2, 20)} = 2.887, p = 0.058$ (Fig. 2a), was almost significant because rotation affects error rates more in the Nasalization by Tongue Height than Pitch by Tongue Height experiment, and because more errors occurred at 22.5° than at other rotations. The differences be-
Table 5. Mean differences (with standard errors) between selective or divided attention and baseline error rates by Rotation: Nasalization by Tongue Height and Pitch by Tongue Height

<table>
<thead>
<tr>
<th>Task</th>
<th>Rotation</th>
<th>0°</th>
<th>22.5°</th>
<th>45°</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
<td>SE</td>
<td>mean</td>
<td>SE</td>
</tr>
<tr>
<td>Nasalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>by Tongue Height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selective - Baseline</td>
<td></td>
<td>0.042</td>
<td>0.022</td>
<td>0.127</td>
<td>0.038</td>
</tr>
<tr>
<td>Divided - Baseline</td>
<td></td>
<td>0.293</td>
<td>0.029</td>
<td>0.338</td>
<td>0.051</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td>0.168</td>
<td>0.017</td>
<td>0.232</td>
<td>0.041</td>
</tr>
<tr>
<td>Pitch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>by Tongue Height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selective - Baseline</td>
<td></td>
<td>0.118</td>
<td>0.020</td>
<td>0.179</td>
<td>0.023</td>
</tr>
<tr>
<td>Divided - Baseline</td>
<td></td>
<td>0.266</td>
<td>0.049</td>
<td>0.232</td>
<td>0.043</td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td>0.191</td>
<td>0.034</td>
<td>0.201</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Between Experiments in the effects of Task Type reflected in the second significant interaction, F(6,200) = 4.818, p = 0.001 (Fig. 2b), include a drop in error rate for both the positively and negatively correlated tasks compared to baseline performance in the Pitch by Tongue Height experiment, but only for the negatively correlated task in the Nasal by Tongue Height experiment vs. substantial increases in error rate for both experiments in the selective and divided attention tasks. The interference indicated by the poor performance on these last two tasks indicates integration of both Nasalization and Pitch with Tongue Height, but integration also predicts facilitation on both the positively and negatively correlated tasks. The absence of such facilitation in the positively correlated task in the Nasalization by Tongue Height experiment thus requires explanation. This difference between the two experiments in the behavior of positively vs. negatively correlated stimuli is shown to arise from differences in perceptual primacy between Nasalization and Pitch relative to Tongue Height in the evaluation of the effects of congruence in section 3.4.

3.3. Assessing Perceptual Primacy

Perceptual primacy of the dimensions of Tongue Height, Nasalization, or Pitch was assessed first by looking at whether the difference between selective attention and baseline performance gets larger with rotation from 0 to 45°, while the differences between divided attention and baseline performance gets smaller. If Rotation interacts with Task Type in this way, then the dimensions corresponding to the original orientation are perceptually primary. More generally, the interaction of rotation with these two kinds of differences all critically primary dimensions to they are those dimensions a rotation at which the selective baseline differences are sr divided attention vs. baseline largest. These two kinds of differences calculated for each task, rotation and serve as the dependant three-way ANOVA in which the dependent variables are Rotatio
Integrating Articulations

Two kinds of differences allows the perceptually primary dimensions to be discovered. These are those dimensions aligned with the rotation at which the selective attention vs. baseline differences are smallest and the divided attention vs. baseline differences largest.

These two kinds of differences were calculated for each task, rotation, and subject and serve as the dependent measure in a three-way ANOVA in which the independent variables are Rotation, Experiment, and Attention (subset of Task Types; table 5).

Neither Rotation nor Experiment was significant, though Attention was, $F_{1,77} = 44.321$, $p < 0.0001$, reflecting the poorer performance overall in divided compared to selective attention. The interaction of Experiment by Attention reached significance, $F_{1,77} = 5.805$, $p = 0.0184$, because in the Nasalization by Tongue Height experiment the divided attention error rate differs more from the selective attention error rate than...
in the Pitch by Tongue Height experiment. As predicted by the hypothesis that the dimensions at one of the orientations are perceptually primary, the interaction of Rotation with Attention did approach significance, $F_{(2,77)} = 2.932, p = 0.0593$, reflecting progressive reduction of the difference in performance on the divided vs. selective attention tasks with rotation (fig. 3a).

However, figures 3b, c, which break out the effect of Rotation by Experiment, show that while rotation away from $0^\circ$ uniformly increases the error rate in selective attention tasks in both experiments, its effect on error rates in the divided attention tasks is not uniform across or within experiments. Inexplicably, the divided attention error rate is highest at $22.5^\circ$ for the Nasalization by Tongue Height experiment (fig. 3b), while that error rate drops with rotation to $22.5^\circ$ and then with rotation to $45^\circ$ rises again in the Pitch by Tongue Height experiment (fig. 3c). While the increase in error rate for selective attention with rotation away from $0^\circ$ does support interpreting the dimensions aligned with the array at $0^\circ$ as the perceptually primary ones, the changes in divided attention error rates do not. The assessment of perceptual primacy is therefore pursued along a different path in the evaluation of congruence in the next section.

### 3.4. Assessing Perceptual Primacy through Congruence

In both experiments, there is a contrast between stimuli in which dimensions vary congruently vs. those where they vary incongruently, but the two experiments differ in whether the congruent stimuli are also those which exhibit the naturally occurring covariation of dimensions. For the Nasalization by Tongue Height experiment, the congruent stimuli, B and C, are also those which exhibit the natural covariation of dimensions, a large $N_1-N_2$ difference with high $F_1$ and vice versa. For the Pitch by Tongue Height experiment, on the other hand, the congruent stimuli exhibit the opposite pattern of covariation, high $F_0$ with high $F_1$ and vice versa. Of interest here then, is whether the congruent or the naturally covarying vowels are perceived more accurately in both experiments.

For each experiment, subject, rotation and task type (baseline, correlated, selective attention, and divided attention), the errors were separately tabulated for each of A, B, C, and D, and then expressed as a proportion of the total responses for that subject, experiment, rotation and task type. The error proportions for the congruent stimuli B and C, were then added together, as were the error proportions for the incongruent stimuli A and D, and a congruence score was calculated by subtracting the congruent sum (B+C) from the incongruent one (A+D). These scores would tend to be positive if fewer errors were made for the putatively congruent B and C than the putatively incongruent A and D. Mean congruence scores (with standard errors) are presented in table 6 (note that errors are pooled within a task type).

<table>
<thead>
<tr>
<th>Task</th>
<th>Rotation</th>
<th>0°</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasalization by Tongue Height</td>
<td>Baseline</td>
<td>-0.012</td>
<td>-0.012</td>
</tr>
<tr>
<td>Correlated</td>
<td>-0.041</td>
<td>-0.041</td>
<td></td>
</tr>
<tr>
<td>Selective</td>
<td>-0.097</td>
<td>-0.097</td>
<td></td>
</tr>
<tr>
<td>Divided</td>
<td>-0.016</td>
<td>-0.016</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>-0.041</td>
<td>-0.041</td>
<td></td>
</tr>
</tbody>
</table>

| Pitch by Tongue Height | Baseline | -0.010 | -0.010 |
| Correlated | 0.004 | 0.004 |
| Selective | 0.023 | 0.023 |
| Divided | -0.025 | -0.025 |
| mean | -0.002 | -0.002 |

but near zero or weakly positive for Tongue Height experiment and C were clearly perceived for the Nasalization by Tongue Height experiment than incongruent / there was little difference in between congruent and incongruent the Pitch by Tongue Height Task Type was also significant for divided attention at no polarity for baseline tasks. not significant as a main effect.

Two interactions achieved Experiment by Task Type, $F_{(6,10)} = 3.379$, $p = 0.0132$ (fig. 4a), and Rot Type, $F_{(6,10)} = 3.379$, $p = 0$.
Integrating Articulations

<table>
<thead>
<tr>
<th>Task</th>
<th>Rotation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0°</td>
</tr>
<tr>
<td></td>
<td>mean</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nasalization by Tongue Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>Correlated</td>
</tr>
<tr>
<td>Selective</td>
</tr>
<tr>
<td>Divided</td>
</tr>
<tr>
<td>mean</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pitch by Tongue Height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
</tr>
<tr>
<td>Correlated</td>
</tr>
<tr>
<td>Selective</td>
</tr>
<tr>
<td>Divided</td>
</tr>
<tr>
<td>mean</td>
</tr>
</tbody>
</table>

The Experiment by Task Type interaction leads to reinterpretation of the two significant main effects, since congruence scores for correlated and selective attention tasks are only markedly negative in the Nasalization by Tongue Height experiment; in the Pitch by Tongue Height experiment, the congruence scores in these two tasks are much closer to zero, as are congruence scores overall for this experiment. This disparity is reversed for divided attention where congruence scores are more strongly positive in the Pitch than Nasalization by Tongue Height experiment. In correlated and selective attention tasks, the two members of either the congruent and incongruent pairs must always be assigned to different classes. In the Nasalization by Tongue
Height experiment, achieving this result is clearly much more difficult for B and C, the stimuli in which the two dimensions covary as in natural vowels, than A and D, in which they covary inversely, but only a much smaller disparity can be observed in the Pitch by Tongue Height experiment, and that is due to negative congruence scores at 22.5° alone. In the divided attention task, where the congruent stimuli are to be distinguished from the incongruent, there is an advantage for the congruent over the incongruent stimuli. This advantage is greater in the Pitch by Tongue Height experiment, where it increases with rotation away from 0°, than in the Nasalization by Tongue Height experiment, where congruence scores are only noticeably positive at 45°. Interpretation of both results is aided by figure 1. If the A-B and C-D sides of the stimulus array were lengthened perceptually with rotation away from 0°, performance should improve for the incongruent A and D stimuli over the congruent B and C stimuli for the two tasks. Members of each pair are assigned classes, the correlated attention tasks, but this advantage reversed for the task where stimuli are assigned to the divided attention task. What is why the first effect pronounced in the Nasalization Height experiment, but the second pronounced in the Pitch by Tongue Height experiment.

A series of independent tests were performed for each task. Task Type and Rotation in the experiment to see whether the congruence differed from zero; their results obtained from the A-D Nasalization by Tongue Height experiment were significantly negative results (p < 0.05). Equally negative results (0.05 > p > 0.10) negative trials, and positive values for the line task at 22.5° (t(7) = 2.66; and the divided attention task; 2246. p = 0.0596). The negative congruence scores putatively incongruent pair A and D, were usually classified more reliably than the putatively congruent A and C, in this experiment. The result does not support the Nasalization and Tongue Height experiment. A particularly primary dimension not depend on the listeners' independent measurements along the two dimensions, in which the dimensions are apparently more readily. This result is a generalization error rates observed for A and D
C stimuli for the two tasks in which the members of each pair are assigned to different classes, the correlated and selective attention tasks, but this advantage should be reversed for the task where the congruent stimuli are assigned to the same class, the divided attention task. What remains puzzling is why the first effect is more pronounced in the Nasalization by Tongue Height experiment, but the second is more pronounced in the Pitch by Tongue Height experiment.

A series of independent two-tailed t tests were performed for each combination of Task Type and Rotation in each Experiment to see whether the congruence scores differed from zero; their results resemble those obtained from the ANOVA. For the Nasalization by Tongue Height experiment, significantly negative results were obtained in 3 (of 12) tests (p < 0.05), almost significantly negative results in 2 others (0.05 < p < 0.10), negative trends in 5 others, and positive values for only 2: the baseline task at 22.5° (t(7) = 2.634, p = 0.0337) and the divided attention task at 45° (t(7) = 2.246, p = 0.0596). The preponderance of negative congruence scores shows that the putatively incongruent pair of stimuli, A and D, were usually classified more accurately than the putatively congruent pair, B and C, in this experiment. However, this result does not support the conclusion that Nasalization and Tongue Height are perceptually primary dimensions, since it does not depend on the listeners' ability to make independent measurements of the stimuli along the two dimensions. Instead, vowels in which the dimensions covary unnaturally are apparently more readily distinguished. This result is a generalization of the lower error rates observed for A and D in the negatively correlated task compared to B and C in the positively correlated one (see end of section 3.2). The t tests for the Pitch by Tongue Height experiment yielded quite different results. Congruence scores were significantly positive in just one case, the divided attention task at 45° (t(7) = 3.300, p = 0.0131) and approached positive significance in one other, divided attention at 22.5° (t(7) = 2.218, p = 0.062); there was 1 case where the test approached negative significance, selective attention at 22.5° (t(7) = −2.350, p = 0.051); otherwise, the congruence scores did not differ from zero and no trends emerged (5 were weakly positive and 4 weakly negative). The overall failure of the congruence scores to differ significantly from zero in the Pitch by Tongue Height experiment suggests that listeners do not reliably classify vowels along these two dimensions independently of perceived vowel height – this result also generalizes the similarity in error rates between the positively and negatively correlated tasks in the Pitch by Tongue Height experiment.

3.5. Summary

That correlated variation facilitated classification, while orthogonal variation interfered with it in both these experiments suggests that both Nasalization and Pitch are integrated with Tongue Height in the perception of vowels. The perceptual primacy of any of these dimensions remains in doubt, however, since although selective attention performance did decline with rotation away from 0°, divided attention performance did not improve uniformly. Nasalization appears from these results to be more of an independent attribute of vowels than Pitch does, although the analysis of the congruence scores suggests a different interpretation.

There it was found that stimuli which combine Nasalization with Tongue Height in ways not ob-
served in natural speech were classified more accurately than those that did, across tasks and rotations. Those stimuli in which these dimensions covary naturally are classified less accurately as a result of more thorough integration of the two dimensions in them, which makes it more difficult to independently evaluate their values along each dimension. It is harder to detect whether a vowel has a high amount of Nasalization when its Tongue Height is Low (= high F₁) and vice versa, than when these two dimensions covary in the opposite direction. With covariation in opposite directions (low F₁ with high N₁ – N₂ and vice versa), two relatively different dimensions of difference can be more readily detected, and as a result those vowels with unnatural covariation are classified more accurately. The perceptual primacy of Nasalization and Tongue Height can thus be blocked by their integration into a percept of vowel height. On the other hand, by either test, Pitch does not appear to be a perceptually primary attribute of vowels in these stimuli.

4. Discussion

4.1. Implications for Theories of Speech Perception

These experiments address the question of whether the effects of independent articulations are separated in perception as predicted by direct realist theories, at least in their strong form. Proponents of these theories have argued that such separation is possible so long as a coarticulatory source is available for nasalization and tongue height see Beddor et al., 1986, and Krakow et al., 1988; for similar results regarding pitch and tongue height see Silverman, 1987, and see Fowler and Smith, 1986], but they would argue that when no coarticulatory source exists, as with these stimuli, then integration may take place. But this does not explain why these patterns of covariation occur in natural speech outside contexts in which they could arise via articulation: low vowels have intrinsically more nasalization and lower pitches than high vowels. In fact, if it is the height of the tongue that is the target of perception, then these patterns of covariation are just as much noise, since in a direct realist theory they convey the occurrence of irrelevant articulations.

On the other hand, from the perspective that all the speaker’s articulatory behavior is intended to influence the perceived height of F₁, and from that to produce a particular percept of vowel height, then the observed covariation AND the ready integration across dimensions become intelligible. The objects of speech perception are not articulatory, but acoustic, and the many-to-one mapping of the one onto the other ceases to be a problem and becomes a positive advantage for the listener. The integration between Nasalization or Pitch and Tongue Height demonstrated in these experiments supports the strong version of the auditory enhancement theory [Diehl and Kluender, 1989; Diehl, this volume], although these experiments do not point to precisely what auditory mechanism is responsible [for a likely mechanism, see Syrdal and Gopal, 1986].

4.2. Implications for the Representation of Vowel Height

4.2.1. Mutual Dependencies between the Covarying Articulations and Their Categorical Representation

The results of the two experiments also undermine the traditional phonological representation of vowel height as a single articulatory property. At least from the listener’s point of view, vowel height can integrated rather than additive the effects of tongue height, and rate of vocal fold vibration may integrate the effects of varying supralaryngeal articular movements. The speaker does not achieve covariation because raising the particular height in the mouth is a goal in itself. The linked context of the observed covariation is that the speaker is trying to convey a particular vowel height, and through it height of the vowel.

This covariation raises problem for the representational height in phonological contrasts of these covarying articulators independently in the stated logical constraints or rules, the six differing in height also b the features [high] and [low] (nasal), [High] (pitch), as w advanced tongue root, [rou] and the synchronic and diachronic patterns of these other articulatory values so constrained as to ren phonemic: For example:

(1) ATR. Differences in position have brought about ch height in Mon-Khmer languages, 1976; Huffman, 1976). In languages which exhibit vow tongue root position in W Africa, the low vowel does instead takes a single value, [Hall et al., 1974]. Somewhat, the high vowels al
which they could arise via coarticulation, and lower pitches than in fact, if it is the height of the target of perception, then covariation are just science in a direct realist theory, the occurrence of irrelevant articulators, too. The speaker does not achieve the observed covariation because raising the tongue to a particular height in the mouth mechanically perturbs the other articulators, but by controlling each of these articulatory events independently. The linked control of each articulation, the observed covariation, does have a single goal, but it is acoustic or more properly perceptual, the manipulation of perceived F<sub>1</sub> and through it the perceived height of the vowel.

This covariation raises an interesting problem for the representation of vowel height in phonological contrasts. Since each of these covarying articulators is referred to independently in the statement of phonological contrasts or rules, then should vowels differing in height also be specified for the features [high] and [low] (tongue height), [nasal], [High] (pitch), as well perhaps as [advanced tongue root], [round], etc.? Both the synchronic and diachronic behavior of vowels differing in height suggest that their values for these other articulators are not always so constrained as to remain safely subphonemic. For example:

1 ATR. Differences in tongue root position have brought about changes in vowel height in Mon-Khmer languages (Gregersen, 1976; Huffman, 1976). In many of the languages which exhibit vowel harmony for tongue root position in West and East Africa, the low vowel does not contrast, but instead takes a single value, usually [+ATR] (Hall et al., 1974). Somewhat less frequently, the high vowels also do not contrast, typically taking [+ATR] values [Hall et al., 1974].

2 Rounding. Contrasts for lip rounding are similarly dependent on tongue height, where in a great many languages, only nonlow back vowels are rounded. In languages with front rounded vowels, the occurrences of such a vowel at a given height implies the occurrence of all front rounded vowels of greater height, though not vice versa [Crothers, 1978; Disner, 1983, 1984, 1985]. Hayes [1990] describes diphthongization in Quebec French, the Lund dialect of Swedish, and Eastern Finnish in which a monophthong breaks into higher and lower components. In each of these cases, if a nonlow back vowel becomes low, it loses distinctive rounding, and if a low back vowel becomes nonlow it acquires it. The case of Quebec French is particularly interesting since a contrast for backness is preserved between the low vowels derived by diphthongization despite the loss of rounding from the back one, which shows that rounding is linked to height in this process.

3 Nasal. The susceptibility of vowels to nasalization depends on their height: low vowels are more likely to be distinctively nasalized than nonlow ones, and high vowels more than mid ones [Ohala, 1974, 1975; Kuhlken, 1978]; nasalization also frequently changes vowel height, both diachronically and synchronically, lowering high and mid vowels and raising low ones [Kuhlken, 1978; Wright, 1980; Beddor, 1983; Beddor et al., 1986].

1 Neither favoring nasalized high vowels over nasalized mid vowels nor raising nasalized low vowels is predicted by the uniform (and inverse) covariation of nasalization with vowel height. However, the acoustic effects, and their likely perceptual con-
(4) Pitch. Differences in rate of fundamental frequency between high and low vowels have led in one case to tone splitting, in the Northern Mon-Khmer language U [Svantesson, 1989], although this is very rare [Hombert, 1977a], while in Foochow and Lahu, vowels have been raised by high and rising tones, respectively [Hombert, 1977a]. Andersen [1986] suggests that [+ATR] vowels may have raised high tones historically in the Chari-Nile language, Lugbara, so perhaps vowel height may influence the phonological use of F0 through another of its covarying articulations.

Many of these phonological consequences are the result of what Ohala [1985] has described as the misperception of the speaker's intent by the listener. For example, nasalization changes vowel height because its acoustic effects lead the listener to hear a vowel of different height than the speaker actually produced, and the listener's (or the language learner's) own production is determined henceforth by the misperception. But it is also important to distinguish this sort of case from others listed above, where it is just the inherent covariation among articulations, and their convergent perceptual effects, rather than the effects of one distinctive feature on the perception of another, that influence phonological patterning. Inherent covariation among articulations acts more as a constraint on phonological inventories and processes than as a source of change. Even so, the splitting of U’s tones as a result of inherent difference in F0 between high and low vowels shows that change, though rare, can come from this inherent covariation.

Representing these articulations, or their perceptual correlates, in terms of individual features does not require that they be separable, however, since the value for one feature is still predictable from the value for another, and this covariation could be achieved by means of bidirectional redundancy rules of the form [a high] → [-a-nasal], etc. The bidirectionality of these rules expresses the perceptual integration of the acoustic consequences of these articulations, whose independence, strictly from the speaker's point of view, is expressed by the use of a separate feature for each one. (Stevens et al. [1986] and Stevens and Keyser [1989] give similar arguments for using distinctive features for enhancement purposes, but Stevens and Keyser distinguish between what they call 'primary' and 'secondary' features, whereas these contrasts, here, I argue distinction, at least for vowels.

Two further pieces of evidence can be categorized representation of these articulations. First are the splitting in U [Svantesson, 1979], inherent difference in pitch and low vowels has split an already high and lower reflects it in the language. The pitch clearly becomes categorical, so by itself, even if it were a split tone. (The neutralizing contrasts in high and low vowel languages [Hall et al., 1974] evidence of the categorical behavior of these articulations.) So it is that some of these are not only in step with tongued vowels above some arbitrary categorical behavior is apparent in the tongue height continuum of the value of the features below it. Categoricality is apparent in data on lip rounding and is a characteristic of pitch differences as well. The partitioning of language by the context of the covarying articulations, their use by the speaker to enhance the characteristics of the language, and the categoricality from phonological descriptions (see Pierrehumbert, 1989) evidence that phonetic
but Stevens and Keyser distinguish, for consonants, between what they call 'primary' features, those employed to bear a contrast, and 'secondary' features, which enhance these contrasts. Here, I argue against this distinction, at least for vowels.)

Two further pieces of evidence support categorical representation of these covarying articulations. First are cases like tone splitting in U [Svanhasson, 1989], where the inherent difference in pitch between high and low vowels has split earlier tones into higher and lower reflexes in the present-day language. The pitch difference has clearly become categorical, and has done so by itself, even if it were not categorical originally. (The neutralization of [ATR] contrasts in high and low vowels in many languages [Hall et al., 1974] is also evidence of the categorical behavior of these covarying articulations.) Second is evidence that some of these articulations do not covary in step with tongue height, but instead show extreme 'high' values for all vowels above some arbitrary dividing point in the tongue height continuum and extreme 'low' values below it. This incipient categoricity is apparent in Linker's [1982] data on lip rounding and is also characteristic of pitch differences associated with tongue height [see Silverman, 1987, for discussion]. The partitioning of the tongue height continuum by the extreme values of the covarying articulations originates in their use by the speaker to exaggerate contrasts, or more generally the perceptual motivations for their use, but can also be viewed, from the perspective of the grammar of the language, as the persistence of categoricity from phonological representations [see Pierrehumbert, 1980, for other evidence that phonetic implementation needs to refer directly to the categorical properties of speech sounds].

4.2.2. Does Covariation Persist When an Articulation Is Used Contrastively?

However, representing all these covarying articulations in terms of bidirectional redundancy rules referring to categorical features also overstates the magnitude of the differences in these articulations between high and low vowels, and even worse incorrectly predicts that these features may not contrast phonemically in vowels. For example, if low vowels are inherently [+ nasal], then no contrast should be possible between a low oral and low nasal vowel; for that matter, if each height has an associated specification for nasality, then no vowel, at any height, should contrast for nasality. What is needed here is a phonetic representation that allows specification of nasality differences among what are phonemically all [-nasal] vowels.

The solution to this dilemma is to recognize that vowel height contrasts are abstract in the phonology: the point of this paper has been in fact to argue that they are so abstract in that component of the grammar as not to have any necessary association with the height of the tongue per se (nor with degree of constriction). (Whether other phonemic contrasts are abstract in the same sense remains to be determined, but I su-

---

3 John McCarthy [personal commun.] notes that this view is very like the traditional view of stress in English, as an abstract coding of prominence, which is realized in terms of a variety of different articulations, variously selected according to context [Beckman, 1966]. Stress, too, may exhibit integration of different dimensions, for example of amplitude and duration, as suggested by Beckman [1986] in her notion of 'total amplitude'.

spect they are.) Once these abstract height contrasts are implemented in the phonetics, however, they are spelled out in terms of a set of at least quasi-categorical articulations involving the velum, larynx, jaw, lips, and tongue. If another phonemic contrast also gets spelled out in terms of any of these articulators, then in some cases their inherent covariation in spelling out vowel height will be subordinate to, and in physical terms smaller than the articulatory events and acoustic consequences that spell out the phonemic contrast. That superordinate contrast may even wipe out the inherent covariation, as for example if phonemically nasalized vowels of different heights (or those which are nasalized by an adjacent nasal) no longer differ in nasalization, as phonemically oral vowels do.

The data show, however, that differences in nasalization between vowels of different heights are frequently preserved when they’re contextually nasalized, and in some cases even when they are distinctively nasalized. Clumeck [1976] shows that relative to the duration of the vowel, the higher the velum begins to lower in oral vowels before nasal consonants in English, French, Amoy, and Hindi, and he finds the same dependency in Swedish long but not short vowels, which are so short that the whole vowel is nasalized. The magnitude of velar lowering in oral vowels before nasal consonants was not entirely predictable from height for all the speakers, but even so all the significant differences between vowels rank them in terms of magnitude of velar lowering as: low > mid > high. Moll [1962] and Kuehn [1976] also report larger velar opening for low than high vowels before nasals. In addition, Clumec [1976] observed that low vowels were produced by English speakers with velar lowering in oral environments, but this was not observed for the other languages.[This difference in vowel height between high and low vowels outside of nasal contexts was also observed by Moll, 1962, and Kuehn 1976, for English, and by Kuenzel, 1977, for isolated German vowels]. Finally, Clumec [1976] reports that the magnitude of velar lowering did not differ for distinctively nasalized vowels of different heights in Hindi or French, but for Amoy and Portuguese it did, with less velar lowering for higher vowels in both languages. The data of Bell-Bent et al. [1979] from a single speaker of English show consistently lower velar position in lower than high vowels in oral as well as nasal contexts in disyllabic utterances of the form [VCmVp] or [VmCVp], where the C is an oral or nasal consonant and the Vs were either both [i] or both [a]. The velum was actually lower in all the segments, consonants as well as vowels, oral as well as nasal consonants, when the vowel was [a]. Differences in nasalization associated with vowel height must not only survive contextual nasalization but also influence its contrastive realization. Finally, contrary to all these observations, Vaissiere [1988] found no differences in degree of velar lowering between vowels of different heights in English in connected speech.

Even an independent phonemic contrast on a vowel for an articulator will always constrain its use in context. Rise in height contrasts, and in the English language, exaggerates its effect. Leman [1984] has shown that vowels contrasting in vowel height do not [similar results are presented by Steele, 1986]. Focus implemented in part by small context containing a high tone in question, so F0 is being used in this case (and also in the data). Another possible cause of the pitch increase in the pitch of the vowel is the local expansion of the pitch which contributes substantially to the F0 differences between vowels in this context [Beck, 1986, and Piner, 1986, and 1988]. Consonants [1987] has shown that the expected effects of vowel height on when judging whether a vowel is high or low F0 on a high context to the vowel’s height, this is not the case for the vowel to be high, while a relatively low F0 is sufficient for it to be high. Clearly, F0 can at once convey both of two vowel heights.

F0 can be used to convey neither because the F0 value in terms of the vowel’s height is relatively high, depending on F0 values else contour, and perhaps even contour. Its local value in
ort vowels, which are so short a time as to make it appear that the vowel is nasalized. The nasal articulation in oral vowels being consonants was not entirely pre-

dominant for all the speakers, but all the significant differences rank them in terms of nasal lowering as: low > mid > high [1962] and Kuehn [1976] also regarded nasal opening for low than high or nasals. In addition, Clum- 

ieved that low vowels were pro-


glised speakers with velar lowering environments, but this was not the case in velar lowering environments. The data of Moll, 1962, and Kuehn, English, and by Kuenzel, 1977, for 


mam vowel. Finally, Clum- 


events that the magnitude of velar lowering did not differ for distinctively high vowels of different heights in Hindi but for Amoy and Portuguese it is lower than for higher vowels. The data of Bell-Ben from a single speaker of English 


tently lower velar position in high vowels in oral as well as in disyllabic utterances of the Vp [1967] and [CVCp], where the con-


struent and the V were either a or [a]. The velum was actually the segments, consonants, and en-


tals as nasal consonants and the vowel was [a]. Differences on associated with vowel height 


ily survive contextual nasalization influence its contrastive reall
d, contrary to all these obser-


ssions [1988] found no differ-


g of velar lowering between 


vowels of different heights for 2 speakers of English in connected speech.


Evan an independent phonemic contrast on a vowel for an articulation which other- 


wise covaries with tongue height does not always constrain its use in conveying vowel 


height contrasts, and in the case of F0 actually exaggerates its effect. Ladd and Silver-


man [1984] have shown that when German 


vowels contrasting in vowel height are in focus, the difference in F0 between high and 


low vowels is far larger than when they are not [similar results are presented for Eng-


lish by Steele, 1986]. Focus was apparently implemented in part by mapping a pitch accent containing a High tone onto the vowels in question, so F0 is being used contrastively in this case (and also in Steele's English data). Another possible aspect of focus, a 


local expansion of the pitch range, may also contribute substantially to the increase in 


the F0 differences between high and low vowels in this context [Beckman and Pierre-


humbert, 1986, and Pierrehumbert and Beckman, 1988]. Conversely, Silverman 


[1987] has shown that listeners factor out the expected effects of vowel height on F0 when judging whether a vowel is in focus: a relatively high F0 on a high vowel is attributed to the vowel's height and is insufficient for the vowel to be judged in focus, while a relatively low F0 on a low vowel is sufficient for it to be judged in focus. Clearly, F0 can at once contribute to conveying both vowel height and intonation.


F0 can probably be used for two purposes at once because the interpretation of an F0 value in terms of the tones that represent intonational contrasts in the phonology depends on F0 values elsewhere in the F0 contour, and perhaps even on the entire contour. Its local value in isolation of this larger context is only relevant to judgments of vowel height (and perhaps obstruent voicing). Nasalization, on the other hand, may be used for only one purpose at a time because its contrastive use and its use in connection with vowel height are confined to the same local domain in most languages. This account predicts that F0 would be constrained in its use in conveying vowel height contrasts in tone languages where F0 contrasts in each vowel. This prediction is falsified by the partial maintenance of height differences in F0 in the tone languages Yoruba [Hombert, 1977b] and Taiwanese [Zee, 1980]. In both languages, F0 is substantially higher for high than low vowels produced on a high tone, but on a low tone the F0 differences are negligible. This partial preservation probably reflects differences in the greater expandability of the high as compared to low end of the pitch range which have been observed elsewhere [Pierrehumbert and Beckman, 1988] and suggests that Ladd and Silverman's [1984] results should be reinterpreted as an effect of High tone rather than focus. In any case, even partial maintenance of height-related differences in F0 when it is used contrastively is sufficient to show that contrastive use of an articulation on the individual vowel itself does not preclude its use in conveying vowel height.


However, despite the fact that the domain of phonological contrast for tone is as local in these tone languages as the domain of nasalization in most languages, the interpretation of F0 values as particular tones probably still requires global reference to the F0 contour, just as in languages where the domain of tonal contrasts is the intonational phrase. This hypothesis is supported by the fact that the phonetic implementa-
tion of tonal contrasts, however they arise in the phonology of different languages, is achieved by essentially the same mechanisms [Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988, chapter 8]. Interpretation of $F_0$ values in terms of tones, whether lexical or intonational, must therefore depend on the same global reference to the larger $F_0$ contour. That $F_0$ can be used in tonal as well as vowel height contrasts at the same time, while nasalization cannot, is because the PHONETIC domain of $F_0$ differences is far more global than that of nasalization; the phonological origin of the $F_0$ differences does not matter.

4.2.3. Phonetic Underspecification

The image of vowel articulations described above, in which all articulators contribute to conveying a contrast in vowel height through their influence on perceived $F_0$, differs markedly from the underspecified phonetic representations argued for by Keating [1988, 1990]. In conveying vowel height, no articulator lacks specification, although as just described, the realization of other contrasts by means of the same articulators may alter their contribution to conveying vowel height contrasts. A strong interpretation of Keating's [1988, 1990] proposal predicts coarticulation will arise only when an adjacent segment is unspecified for an articulation and allows that articulation in its neighbor to be anticipated or to carry over. This kind of coarticulation onto unspecified segments will only be possible if that articulation is not already used in the target segment to enhance its distinctiveness. However, the use of an articulation in conveying vowel height does not block an adjacent segment from imposing its value for that articulation on a vowel, as nasals impose [+ nasal] on preceding vowels, regardless of their height (any more than the use of nasalization to enhance vowel height contrasts blocks nasal contrast in vowels). The persistence of differences in nasalization between vowels of different heights under contextual if not contrastive nasalization suggests instead that coarticulation can take place even when the target segment is specified for that articulation, and that that specification alters the extent in both time and space of the coarticulation [see Stevens, 1990, for similar arguments].

4.2.4. The Phonological Equivalence of Covarying Articulations: [ATR] and [high]

The extremely rich representation of vowel height advocated here, which combines tongue height with both laryngeal and supralaryngeal articulations, is anticipated by Hyman's [1988] argument that [+ high, - low, - ATR] vowels must be phonetically equivalent to [- high, - low, + ATR] vowels if their behavior in vowel height harmony in Esimbi is to be accounted for. The results of Khabanyane's [1989] study of height assimilation in Sesotho [Southern Sotho, see also Doke and Mofokeng, 1957, and Clements, 1989] also become more intelligible from this perspective. In this language, which distinguishes apparently between five degrees of height (table 7), the mid vowels /e/ and /o/ are raised to /i/ and /u/ if the vowel in the following syllable is higher, i.e., by any of /e, o, i, u, o, u/; and /i/ and /u/ are raised, but not quite to /i/ and /u/, by either of /i, u/ (a following velar nasal /ŋ/ which derives historically from *ni, also induces both kinds of raising). The change in /e, o/ to /i, u/ is apparently of [- ATR] to [+ ATR], but only some of the conditioning vowels [+ ATR], while the others [+ ATR] are [- ATR]. (The change in [+ high, - low, - ATR], but does not apply.) The apparent lack of [ATR] on the conditioning environment once it is recognized that tongue height, in addition to a distinctive feature in this language, /i, u/ will have relatively high height in terms of the conditioning environment because vowels should trigger [+ ATR] in /e, o/ as just as the [+ high] appears in /e, o/ in Esimbi, [+ high] appears in /e, o/ in Esimbi, [+ high] appears in /e, o/ in Esimbi, [+ high] appears in /e, o/ in Esimbi, [+ high] appears in /e, o/ in Esimbi, [+ high] appears in /e, o/ in Esimbi, [+ high] appears in /e, o/ in Esimbi, [+ high] appears in /e, o/ in Esimbi. This is so because [ATR] is only triggered with the assimilation operator [ATR] is also suggested by its application to Sesotho, where what the rule manipulates should be raised toward /i, u/ if the higher vowel [Clements, 1987a, b] the raising of the mid /e, o/ to /i, u/ collapses a of the high ones does not have the dental consequence of a reassignment of the occurrence of the contexts in which raising does not occur, and v therefore phonemically distinguishable.

However, stating the rule from [- ATR] to [+ ATR] on feet on the high [- ATR] vowel not collapse with /i, u/. B therefore be seen as a phone that advances the tongue root of the following vowels root is phonetically more whether because of phonological condition as [+ ATR] or because the Advancing the tongue root of...
of the conditioning vowels /e, o, i, u/ are [+ ATR], while the others /i, o, i, u/ are [+ high]. (The change in [i, u] is in the direction of [+ ATR], but does not go all the way.) The apparent lack of uniformity in the conditioning environment disappears once it is recognized that [ATR] covaries with tongue height, in addition to serving as a distinctive feature in this language. Thus, /i, u/ will have relatively high values phonetically for [ATR] because they are high vowels and should trigger the change to [+ ATR] in /e, o/ just as the contrastively [+ ATR] but nonhigh /e, o/ do. Just as in Embi, [+ high] appears to be equivalent to [+ ATR]; this is so because [ATR] ordinarily covaries directly with tongue height. That the assimilation operates in terms of [ATR] is also suggested by its discontinuous application to Sesotho vowels. If height were what the rule manipulated, then /e, o/ should be raised toward /i, u/ by a following higher vowel [Clements, 1989]. The fact that the raising of the mid [- ATR] vowel /e, o/ to /e, o/ collapses a contrast, while that of the high ones does not, is an accidental consequence of a recent, limited generalization of the occurrence of /e, o/ to contexts in which raising does not condition their occurrence, and where they are therefore phonemically distinct from /e, o/.

However, stating the rule as a change from [- ATR] to [+ ATR] overstates its effect on the high [- ATR] vowels, which do not collapse with /i, u/. Raising should therefore be seen as a phonetic process advancing the tongue root of vowels when they are followed by vowels whose tongue root is phonetically more advanced, whether because of phonological specification as [+ ATR] or because they are higher. Advancing the tongue root causes the non-

| Table 7. Sesotho vowels, after Khabanyane [1989] and Clements [1989] |
|------------------|-------|----|----|
| i    | u    | +  | -  |
| e    | o    | -  | +  |
| a    | -    | -  | +  |

4.3. Conclusion

The surface representation of vowel height, at the point of entry into phonetic implementation, would be in terms of the array of features whose settings will influ-
ence $F_1$, specifying the full panoply of articulations that covary in conveying this contrast. It is probable that no distinction can be drawn in this representation between distinctive and redundant features, all are on equal footing because they all contribute to conveying the same contrast. This rich specification is motivated perceptually, by the listener's demands for differences in $F_1$ enhanced by these covarying articulations. Finally, phonetic representations resemble phonological ones in using categorical specifications of articulatory events.

Acknowledgments

I am grateful to Carol Kruminis and Paul Iversen for timely and useful discussions of method, to Lyn Frazier, John McCarthy, and Elizabeth Selkirk for their insights into the issues raised here and their requests for clarification, and to Randy Diehl and Keith Klouedorf for provoking these experiments in the first place.

References


Phonetica: Phonetic and phonological rules of Italian.

UCLA Working Papers Phonet. 76.

Typology and universals of vocal space: Greenberg, Universals of human language.


Kluender, K. R.: The objects of speech.


Walsh, M. A.: An auditory basis for the length effect in the perception of stops.

J. acoust. Soc. Am. 82: 2154-2164.


Gasser, W. R.: The processing of information and structure.


Heyer, B.: Diphthongization and coalescence.


J. Phonet. 5: 5-16 (1977a).


Hendi, K.: Relationship between pitch control and vowel articulation.

Phonet. 5: 5-16 (1977a).


House, A. S.; Stevens, K. N.: Analogue studies of the assimilation of vowels.

J. Speech Hear. Disord. 21: 218-228 (1956).


Hyman, L.: Underspecification and vowel height transfer in Eswatini.


Jackson, M. T.: Phonetic theory and cross-linguistic variation in vowel articulation.


In: Fromkin, Phonetics.


Ladefoged, P.: What are linguistic sounds made of?


Ladefoged, P.: On dividing phonetics and phonology; Comments on the papers by Clements and by Browman and Goldstein.


Prof. John Kingston
Department of Linguistics
South College
University of Massachusetts
Amherst, MA 01003 (USA)