Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 36 (2008) 28–54
Journal of Phonetics
www.elsevier.com/locate/phonetics

On the internal perceptual structure of distinctive features: The [voice] contrast

John Kingston[a,*], Randy L. Diehl[b], Cecilia J. Kirk[c], Wendy A. Castleman[b]

[a]*Department of Linguistics, South College 226, University of Massachusetts, Amherst, MA 01003-9274, USA*
[b]*Department of Psychology and Center for Perceptual Systems, 1 University Station A8000, University of Texas, Austin, TX 78712, USA*
[c]*Department of Communication Disorders, University of Canterbury, Private Bag 4800, Christchurch 8020, New Zealand*

Received 19 November 2005; received in revised form 27 January 2007; accepted 12 February 2007

Abstract

Several fixed classification experiments test the hypothesis that F_1, f_0, and closure voicing covary between intervocalic stops contrasting for [voice] because they integrate perceptually. The perceptual property produced by the integration of these acoustic properties was at first predicted to be the presence of low-frequency energy in the vicinity of the stop, which is considerable in [+voice] stops but slight in [−voice] stops. Both F_1 and f_0 at the edges of vowels flanking the stop were found to integrate perceptually with the continuation of voicing into the stop, but not to integrate with one another. These results indicate that the perceptually relevant property is instead the continuation of low-frequency energy across the vowel-consonant border and not merely the amount of low-frequency energy present near the stop. Other experiments establish that neither F_1 nor f_0 at vowel edge integrate perceptually with closure duration, which shows that only auditorily similar properties integrate and not any two properties that reliably covary. Finally, the experiments show that these acoustic properties integrate perceptually (or fail to) in the same way in non-speech analogues as in the original speech. This result indicates that integration arises from the auditory similarity of certain acoustic correlates of the [voice] contrast.

0. Introduction

On examining how phonological feature distinctions are realized phonetically, one is struck by the sheer number of distinct articulatory and acoustic correlates of any minimal contrast. It has been easy to show that many of the acoustic correlates contribute individually and jointly to the listener's recognition of the feature's value. The [voice] contrast between intervocalic stops (and obstruents generally) is particularly rich in this respect (Lisker, 1986). Compared to [−voice] stops, [+voice] stops have shorter closures, vocal fold vibration that lasts longer into closures, and lower F_1 and f_0 at the edges of flanking vowels (among many other differences).

How do the multiple acoustic correlates of a distinctive feature value give rise to a coherent speech percept? This question has been answered in three competing ways. First, listeners have learned that the acoustic

*Corresponding author. Tel.: +1 413 545 6837; fax: +1 413 545 2792.
E-mail addresses: jkingston@linguist.umass.edu (J. Kingston), diehl@psy.utexas.edu (R.L. Diehl), cecilia.kirk@canterbury.ac.nz (C.J. Kirk), wendycastleman@yahoo.com (W.A. Castleman).

0095-4470/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2007.02.001

properties covary reliably (Holt, Lotto, & Kluender, 2001; Kluender, 1994; Nearey, 1990, 1997); second, the acoustic properties result from a single gesture (direct realism: Fowler, 1986, 1990, 1991, 1992, 1996; motor theory: Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985); or third, the acoustic properties produce similar auditory effects (Diehl & Kingston, 1991; Kingston & Diehl, 1994, 1995; Kingston, Diehl, Kluender, & Parker, 1990; Parker, Diehl, & Kluender, 1986). We will call these alternatives the "associative," "gestural," and "auditory" explanations, respectively. The present study was designed to test the auditory explanation; however, as noted in Section 4, the results also bear on the validity of the associative and gestural explanations.

In [+voice] stops, speakers keep the vocal folds vibrating and slacken them to lower F_1 and f_0 as well as to sustain low-frequency periodic energy into the closure. All three acoustic effects concentrate energy at low frequencies in and near the [+voice] stop closure. In [−voice] stops, speakers instead cut off vibration and do not slacken the folds to keep F_1 and f_0 high as well as to prevent low-frequency energy from continuing into the closure. Energy is thereby raised in frequency and dispersed across frequencies near [−voice] stop closures. According to the auditory explanation, these articulations are deliberately produced because their acoustic consequences are similar enough to one another that they integrate and enhance the perceptual difference between minimally contrasting sounds. We refer to the perceptual effect produced by integrating auditorily similar acoustic properties as an "intermediate perceptual property" (IPP). It is intermediate because it lies between the raw, measurable acoustic properties of the speech signal and the distinctive feature values, because an acoustic property may contribute to more than one IPP, and because each distinctive feature value is typically determined by the values of more than one IPP.

IPPs are not a new level of neural or mental representation that is separate from the level at which the acoustic properties themselves are represented perceptually, nor are they produced by any special perceptual mechanism or cognitive activity. As used here, the notion of "integration" just means that two or more cues may contribute to a more distinctive percept because they are auditorily similar and therefore mutually reinforcing. Consider the following visual analogy. An observer must try to discriminate between two small disk-shaped stimuli—one blue and one red—presented very briefly and at low contrast. Assume that performance is not much above chance. Now a blue-green ring is added to the blue disk and reddish-orange ring is added to the red disk, enlarging their circumferences. Immediately, discrimination performance improves because the hue differences between the two disks are displayed over a larger region. The fact that the rings have hues similar to those of the smaller disks to which they were joined is a crucial factor in their discrimination-enhancing effect. Had the rings been assigned in the opposite manner, discrimination performance is not likely to have improved and may even have gotten worse. The point is that cue integration may involve little more than a summing of similar effects at the level at which the primary cues are represented.

In an entirely parallel way, an IPP is the integrated perceptual effect of auditorily similar acoustic properties. Others have also proposed that acoustic properties cohere in this way. For example, the compact versus diffuse and rising versus falling properties that Stevens and Blumstein (Blumstein & Stevens, 1979, 1980; Stevens & Blumstein, 1978) identified in short-term spectra calculated at the borders between consonants and vowels are IPPs. Whether a short-term spectrum is compact or diffuse, rising or falling does not depend on a single acoustic property having a particular value nor even on a specific array of acoustic properties having particular values. By "acoustic property," we refer here to those properties that previous phonetic investigations have shown to differ between contrasting categories and whose differences influence listeners' responses in perceptual tests. In this paper, they are properties that Lisker (1986) has shown to be characteristic of the [voice] contrast. Instead, a short-term spectrum may have one of these qualities when enough of an array of acoustic properties have values within the right ranges. A large number of different combinations of acoustic properties with broad ranges of values will produce these qualities, and are therefore, by hypothesis, perceptually equivalent. The description above of how a variety of acoustic properties each contribute to the percept of low-frequency energy shows that this percept may also be multiply determined by the acoustic properties of particular tokens of [+voice] and [−voice] stops. We refer to this IPP as the "low-frequency property." Besides trying to explain why stops contrasting for [voice] differ in the ways that they do, our other intention here is the same as Stevens and Blumstein's, namely, to explain how listeners can perceive the same qualities in what appear to be acoustically different stimuli.

Here we wish to evaluate the claim that voicing during the consonant closure interval and low values of f_0 and F_1 adjacent to the closure interval all contribute to a single IPP, the low-frequency property. This requires a test of perceptual coherence. In a series of studies (Kingston & Diehl, 1995; Kingston & Macmillan, 1995; Kingston, Macmillan, Walsh Dickey, Thorburn, & Bartels, 1997; Macmillan, Kingston, Thorburn, Walsh Dickey, & Bartels, 1999), we have developed such a test by adapting an experimental paradigm using the fixed classification tasks first introduced by Garner (1974).

Although we expect some instances of covarying acoustic properties to be learned associations rather than IPPs, we predict that these will not cohere perceptually in the test we have devised. Specifically, we begin by testing whether differences in intervocalic closure duration cohere perceptually with differences in F_1 or f_0 at the edges of flanking vowels. In the auditory explanation, differences in closure duration have no effect on the value of the low-frequency property. Therefore, closure duration differences should not cohere perceptually with F_1 and f_0 differences.

The auditory explanation predicts that a pair of acoustic properties that cohere perceptually in stimuli that are recognizable as speech should also cohere in non-speech analogues of those speech stimuli. But the associative and gestural hypotheses do not—at least without additional assumptions.

In this paper, we report the results of three sets of fixed classification experiments using the adapted version of Garner's paradigm, which evaluates the perceptual effects of combining pairs of acoustic properties that covary in intervocalic stops that contrast for [voice]. (Results of the first set of experiments were originally reported by Diehl, Kingston, & Castleman, 1995.) These properties are how much F_1 and f_0 frequencies are lowered at the edges of the flanking vowels, how long voicing continues into the stop closure from the preceding vowel, and how long the stop closure itself lasts.

These fixed classification experiments may measure a very different kind of perceptual interaction between acoustic properties than has been measured in what are called "trading relations" experiments. In trading relations experiments, listeners' categorization of stimuli varying in their value for one acoustic property has often been shown to depend on the value of a second acoustic property that the experimenter can vary orthogonally. The second property is described as "trading" with the first when the category boundary with respect to the first property shifts as a function of the second's value (Repp, 1982).

In categorization experiments of this kind, Lisker (1986) has shown that a [+voice] response is more likely when F_1 or f_0 falls considerably, when voicing continues far into the closure, and when the closure itself does not last long. Lisker's listeners categorized stimuli as [+voice] versus [−voice] that varied orthogonally along two of these acoustic properties at once. Lisker showed that these properties do trade; for example, his listeners were more likely to give a [+voice] response to a stimulus with a long closure duration if voicing continued for awhile into that closure.

Rather than measuring trading relations between acoustic properties in categorization, our fixed classification experiments instead test whether stimuli varying orthogonally in two acoustic properties are more discriminable when they combine these properties' values in the way they are combined in speech. For example, are stimuli that pit a low F_1 and long voicing continuation against a high F_1 and short voicing continuation more discriminable than those that pit the opposite combinations of values for these acoustic properties against one another? Physically, both pairs of stimuli are equally different, so the fixed classification experiments test whether these acoustic properties covary in one direction but not the other so as to increase the perceptual distance between [+voice] and [−voice] stops. These fixed classification experiments thus compare the effects of competing combinations on sensitivity to stimulus differences, rather than comparing response biases, as in the trading relations experiments.[1]

At first glance, the experiments conducted here resemble those of Fitch, Halwes, Erickson, and Liberman (1980) and Best, Morrongiello, and Robson (1981), both of which showed that stimuli were more discriminable when two acoustic properties "cooperated" (i.e., when they covaried in the same direction as in natural utterances) than when they "conflicted" (i.e., when the direction of covariation was opposite to that of natural utterances). Both Fitch et al. and Best et al. found that the cooperating stimuli were more discriminable than the conflicting stimuli only when they fell on opposite sides of the category boundary.

[1]Macmillan et al. (1999) show how results obtained with the trading relations and fixed classification paradigms can be related quantitatively to one another. We do not attempt to do that here because we lack the necessary trading relations data.

In the present study, stimulus values were selected such that all stimuli were within a phonological category. This choice was made because we test perceptual coherence by comparing the discriminability of stimuli that combine acoustic properties in different ways, and listeners must make some errors for their responses to be interpretable.

To summarize, these fixed classification experiments are designed to test the hypothesis that perceptual distance increases because acoustic properties integrate into IPPs. They test the specific hypothesis that only those acoustic properties that produce similar auditory effects integrate into an IPP. The particular IPP investigated by these experiments is the low-frequency property. The more F_1 and f_0 fall at the edges of vowels flanking an intervocalic stop closure and the longer voicing continues into that closure, the more low-frequency energy is present in the vicinity of the stop closure. Therefore, all pairs of F_1, f_0, and voicing continuation are predicted to integrate with one another into the low-frequency property.

In Section 1, we describe how the Garner paradigm has been adapted to the task of discovering IPPs. In Section 2, we then outline the stimulus manipulations in each experiment and list the features that are the same in all of them. The experiments themselves are reported in Section 3, which is followed by general discussion in Section 4.

1. The adapted Garner paradigm

In the experiments reported here as well as in earlier work (Kingston & Macmillan, 1995), we use an experimental paradigm adapted from Garner (1974), which is designed to measure the perceptual (in)dependence of two stimulus dimensions. Four stimuli are constructed by combining two values of each of the two dimensions, as in Fig. 1. In different blocks of trials, observers sort pairs of stimuli from the 2×2 array. In "single-dimension" tasks, observers sort two stimuli drawn from one of the sides of the array, for example, *MM* versus *PM* with respect to Dimension 1 or *MM* versus *MP* with respect to Dimension 2. In "correlated" tasks, observers sort two stimuli at the ends of one or the other diagonal through the array, i.e., the positively correlated *MM* versus *PP* or the negatively correlated *MP* versus *PM*. Stimuli are chosen so as to differ by about a just noticeable difference (jnd), and listeners' accuracy in classifying the stimuli is determined. Following detection theory nomenclature (Macmillan & Creelman, 1991, 2005), both the single-dimension and correlated tasks are "fixed classification" tasks in which listeners classify each stimulus pair member as belonging to one of two arbitrary response categories. In general, the listeners are unable to classify the stimuli correctly on every trial, and thus their imperfect success measures how discriminable the stimuli are in each block of trials.

When on a particular trial a listener correctly classifies the stimulus by giving the response assigned to that stimulus, the response is scored as a "hit," but if the listener gives that response to the other stimulus, it is scored as a "false alarm." A good measure of the listeners' ability to classify the stimulus correctly is the

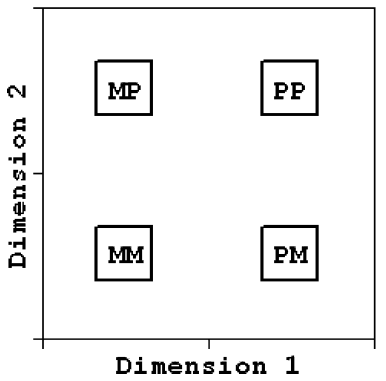

Fig. 1. Schematic 2×2 stimulus array varying along two arbitrary dimensions. Positively correlated *MM* and *PP* stimuli correspond to the pattern of covariation of these dimensions in [−voice] and [+voice] stops, respectively. "*M*" and "*P*" stand for "minus" and "plus," respectively. Negatively correlated *PM* and *MP* stimuli represent combinations of stimulus values not observed in nature.

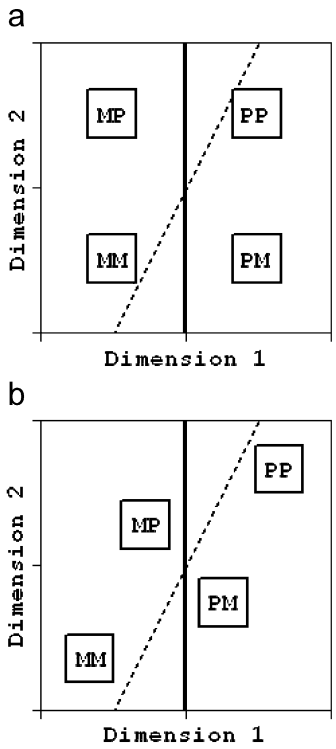

Fig. 2. Alternative mappings of the 2×2 stimulus array in Fig. 1 onto the corresponding perceptual space: (a) perceptually separable dimensions versus (b) perceptually integral dimensions. In (a), the slanting decision bound produces a trading relation, i.e, more "*M*" responses with respect to D1 when D2 is *P* than when it is *M*, and the vertical one does not. A trading relation in this case reflects decisional integrality but perceptual separability. In (b), the vertical decision bound produces a trading relation whereas the slanting one does not. A trading relation in this case instead reflects decisional separability but perceptual integrality.

proportion of hits, discounted by the proportion of false alarms.[2] For classification of the stimulus pairs, *MM* versus *MP*, detection theory prescribes that the discriminability measure is d', which is the difference between the *z* transforms of the hit and false alarm proportions, i.e, the difference between the *z* transforms of the proportion of "*MM*" responses to the *MM* and *PM* stimuli:

$$d' = z(P(Hits) - z(P(FA)).$$

The same prescription is used to calculate the d' values for classification of the other five pairs of stimuli (see Fig. 1): Dimension 1: *MP* versus *PP*, in addition to *MM* versus *PM*; Dimension 2: *MM* versus *MP* and *PM* versus *PP*; positively correlated: *MM* versus *PP*; and negatively correlated: *MP* versus *PM*.

The accuracy measure d' can readily be interpreted as the perceptual *distance* between the two stimuli to be classified in a particular task, and the six d' values obtained from four single-dimension and two correlated tasks can then be used to determine how the stimuli are mapped from the two orthogonal dimensions of the stimulus space onto the corresponding perceptual space (Ashby & Townsend, 1986; Kingston & Macmillan, 1995; Macmillan & Creelman, 1991; Macmillan & Kingston, 1995; Maddox, 1992). Fig. 2 shows two alternative mappings to the perceptual space. In the rectangular arrangement in Fig. 2a, the perceptual dimensions are just as orthogonal as the original stimulus dimensions. This is the mapping that we suggest corresponds to perceptually independent, or *separable*, dimensions. This geometry predicts that performance on the correlated tasks will actually be better than on the single-dimension tasks for separable dimensions, by $\sqrt{2}$ if performance on the orthogonal single-dimension tasks is equal or by $\sqrt{d'^2_1 + d'^2_2}$ if not. However, when

[2]Because there are only two stimuli and two responses in a block of trials, the other responses are the complements of hits and false alarms.

performance on one correlated task is better than performance on the other, the two dimensions can be viewed as perceptually non-independent, or *integral*. Fig. 2b illustrates such an outcome, where performance on the positively correlated task is far more accurate than on the negatively correlated task. In the figure, the stimuli are mapped onto a non-rectangular arrangement in the perceptual space where the diagonal corresponding to the better correlated task is much longer than one corresponding to the poorer one. This asymmetry shows that listeners in fact hear stimuli differing along these two dimensions as though they instead differed along a dimension parallel to the longer diagonal, here a dimension we would refer to as [Dimension 1 + Dimension 2]. If the other diagonal were the longer one, the perceptual dimension would instead be [Dimension 1−Dimension 2]. Because two stimulus dimensions have collapsed to some significant degree onto a single perceptual dimension in such a case, a marked difference in performance between the correlated tasks clearly diagnoses perceptual integrality.

The two panels in Fig. 2 also show alternative criteria or perceptual boundaries (the solid vertical and dashed slanting lines) for classifying the stimuli in a selective attention task with respect to Dimension 1. In this task, all four stimuli are presented in a block of trials and the listener must classify them according to their differences along one dimension, here Dimension 1, while ignoring their differences along the other—hence the task is called "selective attention." A listener who uses the vertical criterion crosses over from hearing the stimuli as "*M*" to hearing them as "*P*" for this dimension at the same value for Dimension 1 regardless of the value of Dimension 2. A listener who uses the slanting criterion instead crosses over at lower value for Dimension 1 when Dimension 2 is *M* than when it is *P*.

Both kinds of criteria are shown in Fig. 2 to demonstrate that the listener may or may not take the value of Dimension 2 into account when deciding how to classify the four stimuli with respect to Dimension 1 quite independently of whether the dimensions are perceptually separable (Fig. 2a) or integral (Fig. 2b). That is, *decisional* separability—the vertical criterion—is independent of *perceptual* separability—the rectangular arrangement of the stimuli (Ashby & Maddox, 1990; Ashby & Townsend, 1986; Macmillan et al., 1999; Maddox, 1992). For this reason, accuracy in the selective attention task is uninformative about whether the two dimensions integrate perceptually. We therefore did not run it in the experiments reported below. Even so, selective attention is what is tested in trading relations experiments, with the attended dimension reduced to just two values, and decisional separability reflects trading relations' measure of perceptual interaction, response bias.

2. Stimulus manipulations and general procedures

2.1. Stimuli

The first of three sets of fixed classification experiments paired F_1 or f_0 with closure duration (Experiments 1a, b). Because differences in closure duration are assumed not to contribute to the low-frequency property, the auditory hypothesis predicts that both F_1 and f_0 will be perceptually separable with respect to closure duration, even though these properties covary reliably.[3]

The second set of experiments used stimuli composed from two of the three possible pairs of the properties predicted by the auditory hypothesis to integrate into the low-frequency property: $F_1 \times$ Voicing Continuation (VC) in the stop closure and $f_0 \times$ Voicing Continuation (Experiments 2a, b). The third set used stimuli composed from the third pair: $f_0 \times F_1$ (Experiments 3a–c). Experiments 2a, b, and 3a used two versions of the stimuli: a speech version in which the vowels had three or five formants and the stimuli sounded like vowel-stop-vowel sequences and a non-speech analogue version in which only the first formant was present in the vowels and the stimuli sounded like a pair of muted trumpet notes separated by a gap. The non-speech analogue versions permit a test of the hypothesis that the low-frequency property is produced by general auditory integrative mechanisms.

In predicting that the results for the non-speech analogues should pattern in the same way as for the corresponding speech stimuli, we mean that the direction of effects observed should be the same, not that the

[3]Closure duration is, however, predicted to integrate with voicing continuation and F_1 into another IPP, the consonant: vowel duration ratio; see Section 4.

magnitudes of those effects should necessarily be equal. Discriminability of acoustic differences may be assumed to vary as a function of other properties of the stimuli, for example, the presence or absence of higher formants. (In general, we expect that the greater complexity of speech stimuli would tend to make them less discriminable than the corresponding non-speech stimuli inasmuch as the additional formants of the former may act as a kind of perceptual noise in the present tasks.) Thus, an ordinal rather than an interval scale is appropriate for comparing the speech and non-speech effects.

Experiments 3b, c studied further the perceptual integration of F_1 and f_0. In addition to the offset–onset values of F_1 and f_0, these experiments manipulated the f_0 trajectory into and out of the stop closure. In Experiment 3b (as well as Experiment 3a), the f_0 trajectory out of the closure into the following vowel was the mirror image of its trajectory into the closure from the preceding vowel. That is, at the offset of the first vowel or vowel analogue, f_0 either remained level or fell into the closure, and at the onset of the following vowel or analogue, f_0 symmetrically remained level or rose out of the closure. In Experiment 3c, on the other hand, the f_0 trajectories into and out of the closure were asymmetric: f_0 was level or fell into the closure from the preceding vowel as in Experiments 3a, b, but it fell from a higher versus lower starting point into the following vowel. This manipulation tests the claim advanced by Silverman (1986) that it is low f_0 values at the edges of a stop closure and not a falling–rising f_0 trajectory that makes [+voice] responses more likely.[4]

2.2. General procedures

2.2.1. Stimulus arrays and tasks

The speech stimuli in all the experiments consisted of a vowel-stop-vowel sequence, the non-speech stimuli of a vowel analogue-gap-vowel analogue sequence. Both were synthesized through the parallel branch of the Klatt synthesizer. All stimulus values reported in the descriptions of the stimuli are synthesis parameters, not measurements of the stimuli. The original version (Klatt, 1980) was used to synthesize the stimuli for the experiments run in Texas, and the Sensimetrics implementation of KLSYN88 (Klatt & Klatt, 1990) was used to synthesize them for the experiments run in Massachusetts. Vowels had three or five formants in the speech stimuli, and vowel analogues in the non-speech stimuli always had just a single formant.

All the experiments used a 2×2 stimulus array, constructed by varying synthesis parameters orthogonally along two dimensions, illustrated schematically in Fig. 1, where *M* (minus) and *P* (plus) represent values of each dimension that would produce low and high values, respectively, for the low-frequency property. In each stimulus set, the combination of two *P* values in the upper right corner of the array should produce the strongest low-frequency value, and the combination of two *M* values in the lower left corner should produce the weakest. Either of the other two combinations of *M* and *P* values should, on the other hand, produce a percept that is ambiguous as to its low-frequency value. If the pattern of acoustic covariation observed in [+voice] and [−voice] stops enhances the perceptual contrast between them, then listeners should more accurately classify stimuli in which the properties are positively rather than negatively correlated. Stimulus values in the two experiments (1a, b) that varied pairs of properties that are not expected to integrate into the low-frequency property, f_0 or F_1 and closure duration, may also be classified as *M* and *P* because they, too, covary between stops contrasting for [voice]: *MM* stimuli have a high f_0 or F_1 at vowel edge and a long closure duration, *PP* stimuli, a low f_0 or F_1 at vowel edge and a short closure duration.

In separate blocks of trials, listeners classified the members of all six pairs of stimuli drawn from these arrays; these pairs break down into two single-dimension classifications for each dimension: Dimension 1, *MM* versus *PM* and *MP* versus *PP*, and Dimension 2, *MM* versus *MP* and *PM* versus *PP*, and two correlated classifications, positively correlated *MM* versus *PP* and negatively correlated *PM* versus *MP*. Hereafter, we will continue to use *M* to refer to stimulus values normally associated with [−voice] stops (i.e., brief voicing continuation, high f_0 or F_1 at vowel edge, or long closure duration) and *P* to refer to stimulus values normally associated with [+voice] stops (i.e., long voicing continuation, low f_0 or F_1 at vowel edge, or short closure duration.

[4]There is actually a similar uncertainty about whether it is F_1's offset–onset frequency or transition duration that influences [voice] judgments; see Stevens and Klatt (1974), Fischer and Ohde (1990), Kluender (1991), and Benkí (2001) for relevant data.

2.2.2. Other general procedures

2.2.2.1. Listeners. In all the experiments, listeners were adult native speakers of English who reported no speech or hearing pathology. They were recruited from the communities of the University of Texas at Austin or the University of Massachusetts at Amherst, and participated in the experiments in exchange either for course credit or payment. No listener participated in more than one experiment, nor did any listener participate in both the speech and non-speech conditions in the second and third sets of experiments.

2.2.2.2. Stimulus presentation and response collection. The stimuli were output at 10 kHz from a PC and low-passed before being amplified and presented binaurally to listeners through headphones. In the experiments run in Texas, the low-pass filter's cutoff frequency was 4.9 kHz, and for those run in Massachusetts, it was 4.133 kHz. Because all stimulus manipulations were below 1 kHz, the difference in filter cutoff frequencies in no way affects the manipulated stimulus properties. The stimuli were presented at 72 dB SPL through Beyer DT-100 headphones in the experiments run in Texas, and at self-selected comfortable levels through TDH-49 headphones in the experiments run in Massachusetts. Experiments in both locations were run in sound-treated, quiet rooms, and up to four listeners at a time were run in a single session.

Except where otherwise noted, each block of trials began with 32 training or orientation trials in which the two stimuli were presented 16 times each in random order, followed by 80 or 96 test trials in which they were presented 40 or 48 times each in random order. Only the test trials were scored. Listeners were trained at the beginning of each block because the difference(s) between the stimuli changed from block to block. There was no gap between the training and test trials. In all trials, listeners were given up to 2 s to respond, followed by a feedback interval in which a light came on above the button or key they should have pressed, and a pause before the next stimulus was presented. The feedback interval lasted 0.25 s and the pause 0.5 s in the experiments run in Texas, but these intervals were twice as long, 0.5 and 1 s, respectively, in the experiments run in Massachusetts. Listeners pressed one of two buttons or response keys to indicate which of the two stimuli they had heard.

At the beginning of the experiment, listeners were told that on each trial in a block of trials they would hear one of two stimuli, and that their task was to push one button when they heard one stimulus and the other button when they heard the other stimulus. They were also told that they would be taught which button to press in response to each stimulus during training trials at the beginning of each block, and that they would learn which button to press by paying attention to the correspondence between the stimulus they had just heard and which feedback light came on. The listeners were warned that the stimuli would be hard to tell apart and that they were not expected to respond correctly on every trial, and also reassured that the feedback lights would continue to come on over the correct button during the test trials following the training trials. Finally, they were told that the stimuli would be classified according to different criteria on each block of trials, which they would have to learn from the training trials at the beginning of the block. They were not told that one or the other stimulus in a block of trials would reappear in other blocks of trials, or that all six blocks of trials used just four different stimuli drawn from a 2×2 array.

The six fixed classification tasks in each experiment were run in counterbalanced order across sessions.

The auditory hypothesis predicts that d' values will be higher for classification of positively correlated *MM* versus *PP* than negatively correlated *MP* versus *PM* stimuli, for non-speech as well as speech stimuli, for dimensions that integrate into the low-frequency property: F_1, f_0, and Voicing Continuation but not Closure Duration. The single-dimension classifications are run to test whether listeners rely more on one dimension than the other in classifying the stimuli. Listeners' performance on these tasks also helps us interpret their performance on the correlated tasks in cases where they are much better at one single-dimension task than the other. Above, we showed that listeners' performance on the correlated tasks with perceptually separable dimensions is predicted to be $\sqrt{d'^2_1 + d'^2_2}$ in this case. But if listeners are much better at classifying the stimuli for one of these dimensions than the other, say Dimension 2, then correlated performance will then be close to $\sqrt{d'^2_2}$, or generally, close to performance on the better single-dimension task. This result should furthermore be obtained for both correlated tasks, as the stimuli differ by the same amount for the dimension to which listeners are more sensitive. We instead find, in some cases, that the d' for one correlated task is larger than the

d' for the better single-dimension task, and d' for the other correlated task is smaller. This result can only arise if the dimensions integrate rather than remaining perceptually separable, so this ranking of task performance is another indication of integrality.

Macmillan et al. (1999) show how the d' values in these six tasks can be used to construct a geometrical model of perceptual integration that quantifies the extent to which the two acoustic dimensions have integrated (see also Kingston & Diehl, 1995; Kingston & Macmillan, 1995, for earlier versions of this model). Such geometric models, which are displayed in Fig. 2 and described in the accompanying discussion, are the first step in quantifying the extent to which the two dimensions integrate. We do not develop this kind of model for any of the results reported here because the question that concerns us in each of these experiments is whether the two manipulated dimensions have integrated, and this question is more easily answered by showing that listeners do or do not discriminate stimuli in one correlated task better than the other. The reader interested in developing such geometric models using the procedures presented in Macmillan et al. (1999) will find average performance on the single-dimension tasks listed in Appendix A.

Another question, which we do not address here, is whether listeners perceive a difference in a particular acoustic dimension in the same way in all experiments where that dimension is manipulated. For example, in Experiments 1a, b, listeners discriminate stimuli differing in closure duration by 40 ms. The other dimensions manipulated in Experiments 1a and b are F_1 and f_0, respectively. In each experiment, two versions of each stimulus pair were presented, one in which the other dimension, f_0 in Experiment 1a and F_1 in Experiment 1b, was held constant at its low value and the other in which it was held constant at its high value. A look at the d' values for discrimination of stimuli differing just in closure duration in Appendix A shows that at times the value of the dimensions held constant in the single-dimension tasks influenced listeners' success at discriminating the stimuli, sometimes profoundly. For example, in Experiment 1b, stimuli differing by 40 ms in closure duration were nearly equally discriminable regardless of whether f_0 was high or low when F_1 was low (average $d' = 1.884 \pm 0.579$ vs. 1.975 ± 0.776), but the stimuli with high f_0 were far less discriminable than those with low f_0 when F_1 was high (0.135 ± 0.197 vs. 3.192 ± 0.573). While these differences in performance on the single-dimension tasks are potentially interesting, they do not detract from finding or failing to find evidence of integration in (un)equal performance on the two correlated tasks.

The next three sections report the results of the (Section 3.1) F_1 and f_0 by Closure Duration experiments (1a, b), (Section 3.2) F_1 and f_0 by Voicing Continuation experiments (2a, b), and (Section 3.3) F_1 by f_0 experiments (3a–c), respectively.

3. The experiments

3.1. Experiment 1: Separability of closure duration from f_0 (Experiment 1a) and F_1 (Experiment 1b)

The hypothesis that only auditorily similar acoustic properties will integrate perceptually makes negative as well as positive predictions. In particular, it predicts that two properties which are not auditorily similar, such as closure duration and F_1 or f_0 at the edges of flanking vowels, will not integrate. Closure duration cannot contribute to the presence or absence of low-frequency energy in the vicinity of a stop closure and therefore should not integrate with acoustic properties that do, even though closure duration reliably covaries with F_1 and f_0 at flanking vowel edges.

Closure duration is a reliable correlate of the [voice] contrast in intervocalic stops (Lisker, 1957, 1986), and thus covaries with F_1, f_0, and voicing continuation into the stop closure. Closure duration is short when F_1 and f_0 offset–onset frequencies are low, and when voicing continues for awhile into the stop from the preceding vowel, but long when these other properties are at the other extreme of their range of values. However, a short closure duration does not in any obvious way add low-frequency energy in the vicinity of the stop closure, so the auditory hypothesis predicts no enhancement from these combinations of closure duration values with F_1 or f_0 values.

3.1.1. Stimuli and procedures

3.1.1.1. Stimuli. The stimuli in these experiments consisted of two 205 ms long, five formant vowels separated by a closure lasting either 30 or 70 ms. Figs. 3a and b show how Closure Duration (CD) and the

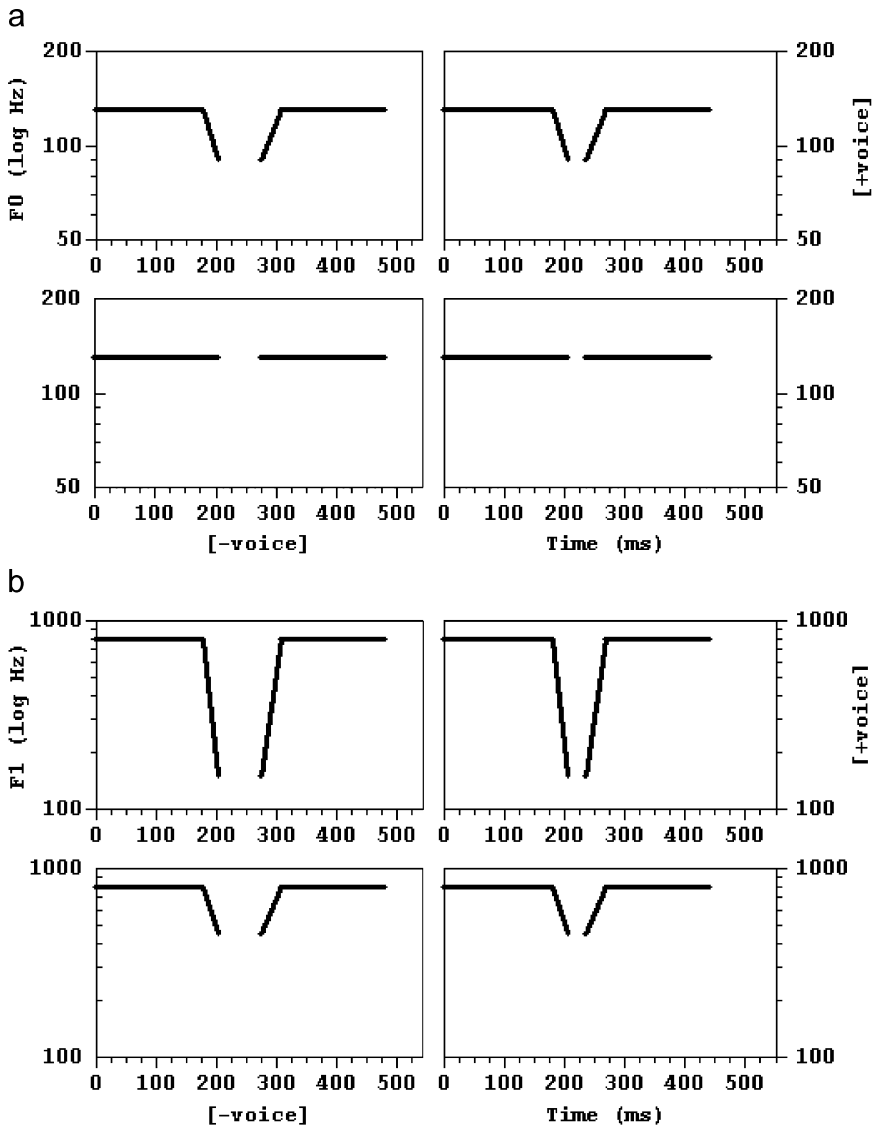

Fig. 3. 2×2 stimulus arrays showing the trajectories of the synthesis parameters in the F_1 and $f_0 \times$ Closure Duration experiments (Experiments 1b, a). Arrangement of stimuli corresponds to that in Fig. 2, in that positively correlated stimuli display the observed combinations of parameter values in [+voice] (upper right) and [−voice] (lower left) stops, and negatively correlated stimuli do not. The horizontal axis in each panel is time in ms, and the vertical axis is frequency in Hz (logarithmic scale). (a) $F_1 \times$ Closure Duration: left = long closure duration of 70 ms, right = short closure duration of 30 ms, top = low F_1; offset–onset frequency of 150 Hz (650 Hz fall), bottom = high F_1 offset–onset frequency of 450 Hz (350 Hz fall), (b) $f_0 \times$ closure duration: left and right same as (a), top = low f_0 offset–onset frequency of 90 Hz (40 Hz fall), bottom = high f_0 offset–onset frequency of 130 Hz (0 Hz fall).

other manipulated dimensions were orthogonally varied. The 70 ms stop closure is shown in the left column in these figures, the 30 ms closure in the right. Four distinct 2×2 stimulus arrays were constructed by combining these Closure Duration values with two values of f_0 or F_1 offset–onset frequency and one value of the remaining property, F_1 or f_0 offset–onset frequency. We measured the effects of pairing f_0 with Closure Duration with one group of listeners in Experiment 1a, and that of F_1 with Closure Duration with another group of listeners in Experiment 1b. In Experiment 1a, f_0 either remained constant at 130 Hz through both of the vowels (Fig. 3b, bottom row) or fell to 90 Hz at vowel offset–onset (Fig. 3b, top row, respectively). The high and low $f_0 \times$ Closure Duration conditions combined the 2×2 array of f_0 and Closure Duration values with the high and low F_1 offset–onset values of 450 and 150 Hz, respectively, used in Experiment 1b.

In Experiment 1b, F_1 either fell 650 Hz to 150 Hz (Fig. 3b, top row) or 350 Hz to 450 Hz (Fig. 3b, bottom row) from its 800 Hz steady-state value. The high and low $F_1 \times$ Closure Duration conditions combined those 2×2 arrays with the high and low f_0 offset–onset values of 130 and 90 Hz, respectively, used in Experiment 1a. Thus, the perceptual interaction between closure duration and F_1 or f_0 is tested with both values of the other spectral property. No voicing occurred during the closure. The steady-state frequencies of F_2 and F_3 were 1150 and 2400 Hz, and at vowel offset and onset they fell to 800 and 1750 Hz, respectively, conveying a bilabial place of articulation. These transitions lasted 40 ms. The values of F_4 and F_5 were 3300 and 3850 Hz, respectively, throughout the stimuli. The *M* or [−voice] values for these dimensions are high f_0 or high F_1 and long Closure Duration, and the *P* or [+voice] values are low f_0 or low F_1 and short Closure Duration.

No non-speech conditions were run in this experiment, because the hypothesis under test does not depend on differences or similarities in listeners' responses to speech stimuli and their non-speech analogues.

3.1.1.2. Procedures. At the beginning of each block of trials, listeners heard 32 randomized training trials with feedback after their responses, followed by 96 randomized test trials, also with feedback, yielding 48 responses per stimulus per task. Two different groups of eleven listeners each performed the six classification tasks on both the high and low F_1 variants of the $f_0 \times$ Closure Duration array (Experiment 1a) or on both the high and low f_0 variants of the $F_1 \times$ Closure Duration array (Experiment 1b); stimuli from the high and low variants were presented in separate blocks of trials. The order in which tasks and variants were presented was counterbalanced across sessions. Both experiments were run in Massachusetts.

3.1.1.3. Results. Figs. 4a and b show d' values in the correlated tasks averaged across listeners in the high and low variants of the f_0 and $F_1 \times$ Closure Duration conditions, respectively. The values of the 95% confidence intervals for these means are displayed as error bars in the figure. (Confidence intervals were calculated using the formula $1.96\sigma/\sqrt{n}$, where n is the number of listeners.) In Experiment 1a, positively and negatively correlated stimuli were equally easy to discriminate, regardless of whether F_1 was low or high, but in Experiment 1b, discriminability of the correlated stimuli depended on whether f_0 was high or low: Positively correlated stimuli were more discriminable when f_0 was high but less discriminable when it was low.

Separate repeated-measures ANOVAs were run on the d' values from Experiments 1a, b, in which Task and high versus low variant of the other spectral property, F_1 or f_0, were within-subjects variables. Task ranged across the two values listed at the bottom of the panels in Fig. 4: negatively (Negative) and positively (Positive) correlated f_0 or F_1 and Closure Duration values.

In the analysis of the f_0 x closure duration experiment (1a, Fig. 4a), the main effect of Task was not significant [$F<1$] nor was the main effect of F_1 value [$F(1,10) = 1.908$, $p>0.10$]. The two variables also did not interact significantly with one another [$F<1$].

In the $F_1 \times$ Closure Duration experiment (1b, Fig. 4b), the main effect of Task was again not significant [$F<1$], but the main effect of f_0 and its interaction with Task both were [f_0: $F(1,10) = 12.408$, $p = 0.006$; $f_0 \times$ Task: $F(1,10) = 5.808$, $p = 0.037$] because the negatively correlated stimuli were more discriminable than the positively correlated ones when f_0 was low, but the positively correlated stimuli were more discriminable than the negatively correlated ones when f_0 was high.

3.1.2. Discussion

The results of Experiment 1a show that closure duration does not integrate perceptually with f_0 even though it reliably covaries with it in stops contrasting for [voice]. Neither positively nor negatively correlated stimuli were more discriminable. The value of F_1 also did not alter the failure of these two dimensions to affect one another perceptually. This failure confirms the negative prediction of the auditory hypothesis: The covariation of closure duration with f_0 does not enhance the contrast because closure duration does not contribute to the low-frequency property. If Closure Duration and F_1 also fail to integrate perceptually, then the fact that f_0 differences are more discriminable in *MP* or *PM* than *MM* or *PP* combinations of F_1 and Closure Duration values does not, furthermore, disconfirm the auditory hypothesis.

The results of Experiment 1b also reveal no consistent pattern of integration of Closure Duration with F_1 but rather one that depends on f_0. The difference in discriminability between positively and negatively correlated stimuli reversed between the two f_0 levels. Discriminability thus does not correspond consistently to

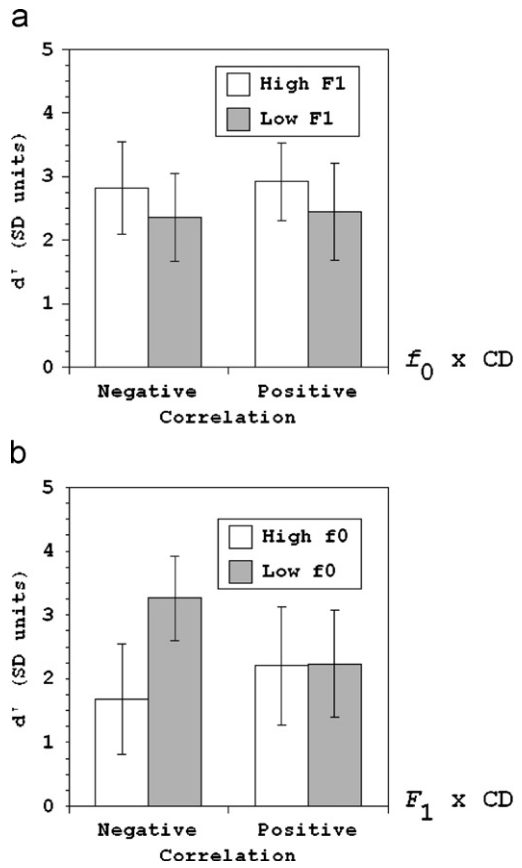

Fig. 4. Mean d' values by classification across listeners, with 95% confidence intervals, (a) Experiment 1a, f_0 × Closure Duration: white bars = high F_1, gray bars = low F_1. Negative = negatively correlated classification, Positive = positively correlated classification. (b) Experiment 1b, F_1 × Closure Duration: White bars = high f_0, gray bars = low f_0. Negative = negatively correlated classification, Positive = positively correlated classification.

the patterns observed in stops, and F_1 and Closure Duration do not show a consistent pattern of integration. A question that remains unanswered is why the perceptual interaction between F_1 and Closure Duration depends on f_0.

3.2. Experiments 2a, b: Integrality of F_1, f_0, and voicing continuation

These experiments compare the extent to which the putative contributors to the low-frequency property integrate in synthetic vowel-stop-vowel sequences and non-speech analogues of them constructed by removing all but the first formant. Voicing continuation (VC) into the closure was varied in both experiments, which differed in whether F_1 or f_0 at the edges of the vowels flanking the closure was varied.

3.2.1. Experiment 2a: F_1 × Voicing Continuation

Both the speech and non-speech versions of this experiment were conducted in Texas. Eighteen listeners participated in the speech version of the F_1 × Voicing Continuation (F_1 × VC) experiment, and 10 different listeners in the non-speech version.

3.2.1.1. Stimuli, speech. The speech stimuli in this experiment consisted of two 205 ms long, three formant vowels separated by an 80 ms stop closure. Fig. 5a shows how the two relevant dimensions of the stimuli were varied. Voicing continued for either 10 ms (left column) or 50 ms (right column) into the stop closure from the preceding vowel. At vowel offset and onset, F_1 fell either 600 Hz to 150 Hz (top row) or 350 Hz to 400 Hz

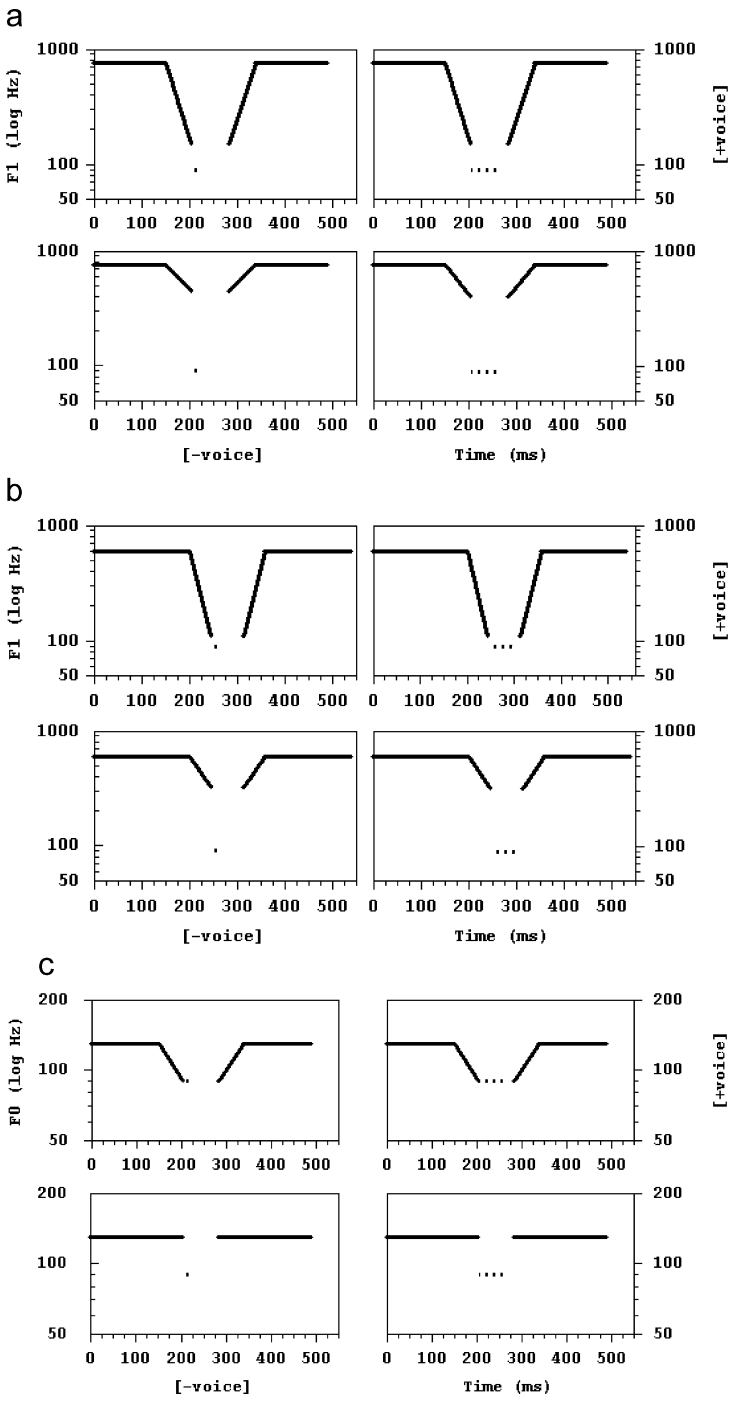

Fig. 5. 2×2 stimulus arrays showing the trajectories of the synthesis parameters for the F_1 and $f_0 \times$ Voicing Continuation experiments (Experiments 2a, b). (a) Experiment 2a, $F_1 \times$ Voicing Continuation, speech condition: the heavy lines represent the F_1 trajectory, and the dashes the continuation of voicing into the stop closure. Left column = short voicing continuation of 10 ms, right column = long voicing continuation of 50 ms, top row = low F_1 offset–onset frequency of 150 Hz (600 Hz fall from steady state), bottom row = high F_1 offset–onset frequency of 400 Hz (350 Hz fall), (b) Experiment 2a, $F_1 \times$ Voicing Continuation, non-speech condition: left and right same as (a), top = low F_1 offset–onset frequency of 110 Hz (490 Hz fall), bottom = high F_1 offset–onset frequency of 320 Hz (280 Hz fall), (c) Experiment 2b, $f_0 \times$ Voicing Continuation, speech and non-speech conditions: left and right same as (a), top = low f_0 offset–onset frequency of 90 Hz (40 Hz fall), bottom = high f_0 offset–onset frequency of 130 Hz (0 Hz fall).

(bottom row) from its steady-state frequency of 750 Hz. Both F_1 transitions lasted 55 ms. The steady-state frequencies of F_2 and F_3 were 1150 and 2400 Hz, and at vowel offset and onset they fell to 800 and 1750 Hz, respectively, conveying a bilabial place of articulation. These transitions lasted 40 ms. During the vowels, f_0 was held constant at 100 Hz, but it was reduced to 90 Hz during the interval of closure voicing. Closure voicing was otherwise synthesized by setting F_1 to 150 Hz and zeroing out the amplitudes of the higher formants.

3.2.1.2. Non-speech. The non-speech stimuli consisted of a gap between two single-formant vowel analogues, again synthesized through the parallel branch of the Klatt synthesizer (Fig. 5b). The amplitudes of all formants but F_1 were zeroed out and F_1's amplitude was constant throughout the stimuli. As in the speech stimuli, voicing continued either 10 or 50 ms into the gap. F_1 trajectories were slightly different from the speech stimuli: at vowel onset and offset, F_1 fell either 490 Hz to 110 Hz (top row) or 280 Hz to 320 Hz (bottom row) from its steady-state value of 600 Hz. The F_1 transition lasted 45 ms. These stimuli thus differ slightly from the speech stimuli, where F_1's steady-state value was 750 Hz, its offset–onset values were 150 and 400 Hz, and its transitions lasted 55 ms. As in the speech stimuli, f_0 was again fixed at 100 Hz during the vowel analogues and at 90 Hz during the gap voicing. F_1 was again set to 150 Hz during the gap. The gap's duration was fixed at 70, 10 ms shorter than that used in the speech stimuli, the duration of the preceding vowel analogue was 245 ms, and that of the following vowel analogue was 225 ms, 40 and 20 ms longer, respectively, than the vowels in the speech stimuli.[5]

The stimuli in which the duration of Voicing Continuation (VC) and the size of the F_1 fall are positively correlated, i.e, {*MM*: 10 ms VC, F_1 fall 350 Hz (speech) or 280 Hz (non-speech)} versus {*PP*: 50 ms VC, F_1 fall 600 Hz (speech) or 490 Hz (non-speech)}, differ most in the hypothesized low-frequency property, and have the *MM* combination that characterizes /p/ and the *PP* combination that characterizes /b/, respectively.

3.2.1.3. Stimulus presentation. Each of the six classifications began with 32 randomized training trials, followed by 80 randomized test trials, both with feedback following the listener's response. The order of the six classifications was counterbalanced across sessions.

3.2.1.4. Results. Fig. 6a displays the d' values, averaged across listeners in the speech and non-speech versions of this experiment (with 95% CIs). Overall, the listeners in the non-speech version discriminated the stimuli better than those in the speech version; in both versions, listeners discriminated stimuli in which F_1 and Voicing Continuation were positively correlated better than those in which these properties were negatively correlated.

Individual d' values were submitted to a repeated-measures ANOVA in which speech versus non-speech was a between-subjects variable and Task was a within-subjects variable. Task ranged across the two values listed at the bottom of Fig. 6a: negatively and positively correlated F_1 and Voicing Continuation values.

The main effect of Task was significant [$F(1,26) = 12.666$, $p = 0.001$], as was the effect of speech versus non-speech [$F(1, 26) = 14.528$, $p = 0.001$], but the two variables did not interact significantly [$F < 1$]. The absence of any interaction shows that the advantage of positively over negatively correlated stimuli was no greater for the speech than the non-speech stimuli.

Discussion of these results is postponed until after the results of Experiment 2b are presented.

3.2.2. Experiment 2b: $f_0 \times$ Voicing Continuation

Nineteen listeners participated in the speech version of this experiment, 23 different listeners in the non-speech version. Both versions were run in Texas.

[5]Parameter values differed in small ways between speech and non-speech stimuli in this and some other experiments because pilot work with each stimulus set showed that listeners were not equally sensitive to single-dimension differences in both kinds of stimuli. The results reported below show that these adjustments were not entirely successful, as listeners were generally more accurate in discriminating all non-speech stimulus pairs than the corresponding speech pairs. This result may, however, arise in part from the fact that the speech stimuli which listeners had to discriminate were within-category, which is known to be difficult. In any case, these differences cannot affect the hypotheses under test because they were intended only to make listeners equally accurate on the orthogonal single-dimension classification tasks.

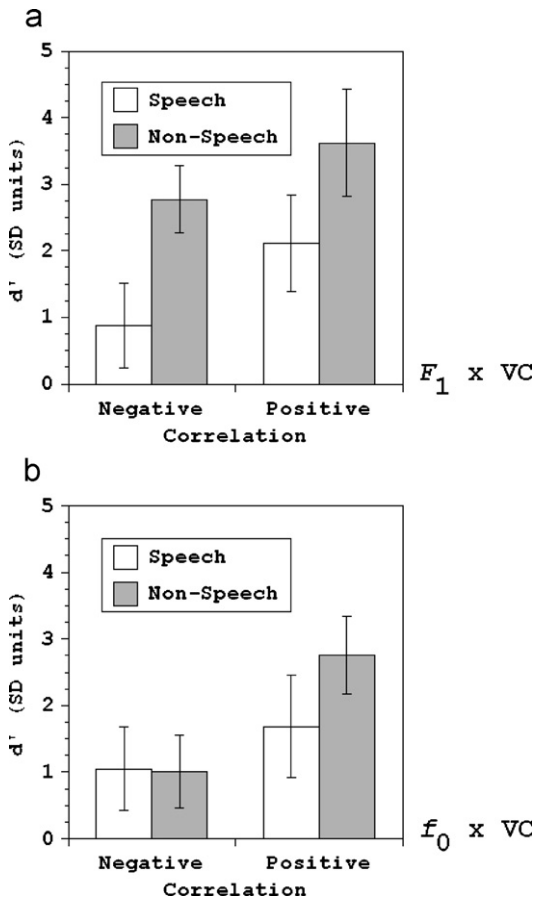

Fig. 6. Mean d' values by classification across listeners, with 95% confidence intervals, (a) Experiment 2a, $F_1 \times$ Voicing continuation: White bars = speech condition, gray bars = non-speech condition. Negative = negatively correlated classification, Positive = positively correlated classification, (b) Experiment 2b, $f_0 \times$ Voicing Continuation: same as (a).

3.2.2.1. Stimuli: speech. The speech stimuli again consisted of two 205 ms long, three formant vowels separated by an 80 ms stop closure. Fig. 5c displays the values of the parameters manipulated in producing the 2×2 array. As in the $F_1 \times$ Voicing Continuation experiment, voicing continued either 10 or 50 ms into the stop closure (left and right columns, respectively). The value of f_0 either remained constant at 130 Hz through both vowels (bottom row) or fell at vowel offset–onset to 90 Hz (top row). The duration of the falling f_0 transition was 50 ms. Closure voicing had a 90 Hz f_0 in all four stimuli. F_1 fell to 150 Hz at vowel offset–onset from its steady-state value of 600 Hz in all four stimuli. The values of F_2 and F_3 were the same as in the $F_1 \times$ Voicing Continuation experiment. The F_1 transitions lasted 35 ms, and the F_2 and F_3 transitions both lasted 40 ms.

3.2.2.2. Non-speech. Except that they had no F_2 or F_3, the non-speech stimuli were in all other ways identical to the speech stimuli in this experiment.

3.2.2.3. Stimulus presentation. Each of the six classifications began with 32 randomized training trials, followed by 80 randomized test trials, in all of which feedback followed the listener's response. The order of the six classifications was counterbalanced across sessions.

3.2.2.4. Results. Mean d' values across listeners (with 95% CIs) are shown for the speech and non-speech versions of this experiment in Fig. 6b. In both versions of the experiment, listeners discriminated positively correlated stimuli better than negatively correlated ones.

The d' values for individual listeners were again submitted to a repeated-measures ANOVA with speech versus non-speech as a between-subjects variable and Task a within-subjects variable (Task had the same two values it did in Experiment 2a). The main effect of Task was significant [$F(1,40) = 28.696$, $p < .001$], but not the main effect of speech versus non-speech [$F(1,40) = 1.689$, $p > .10$]. The speech versus non-speech contrast nonetheless interacted significantly with Task [$F(1,40) = 6.151$, $p = 0.017$], because the advantage of the positively over negatively correlated stimuli was greater for the non-speech than the speech stimuli.[6]

3.2.3. Discussion

Experiments 2a, b show that when F_1 or f_0 covary with Voicing Continuation so as to produce more or less low-frequency energy near and in the stop closure, stimuli are more discriminable than when they vary in the opposite way. Specifically, synthetic stimuli that mimic natural ones in combining low F_1 or f_0 offset–onset frequency with long continuation of voicing into the stop closure were more discriminable from stimuli combining high F_1 or f_0 offset–onset frequency with brief continuation of voicing than stimuli combining F_1 or f_0 with Voicing Continuation values in the opposite ways. We suggested above that this enhancement arises because these different acoustic properties actually integrate into a single IPP, called the low-frequency property, whose values are maximally different in such positively correlated stimuli and minimally different in negatively correlated ones.

Underlying this specific hypothesis is a more general one, namely, that auditorily similar properties integrate. This more general hypothesis predicts that the same enhancement should be observed in non-speech analogues in which these properties are positively correlated. This prediction was confirmed for the pairings of F_1 or f_0 with Voicing Continuation.

3.3. Experiment 3: Separability of $f_0 \times F_1$

This third set of experiments tested the last pairing of properties that the auditory hypothesis predicts would contribute to the low-frequency property, f_0 and F_1, which have each just been shown to integrate perceptually with the third, Voicing Continuation into the stop closure. Three different experiments were run: Experiment 3a compared listeners' responses to speech and non-speech stimuli, as in Experiments 2a, b, and Experiments 3b, c tested the robustness of the unexpected results obtained in Experiment 3a.

3.3.1. Experiment 3a: $f_0 \times F_1$: speech versus non-speech analogues

Fifteen listeners each were run in the speech and non-speech versions of this experiment; both versions were run in Texas.

3.3.1.1. Stimuli: speech. The speech stimuli in this experiment once more consisted of two 205 ms vowels separated by an 80 ms stop closure. Fig. 7a shows how the stimuli varied in f_0 and F_1; they combine the manipulations of these two properties in the F_1 and $f_0 \times$ Voicing Continuation experiments, while holding Voicing Continuation constant at 0 ms. The value of f_0 either remained constant at 130 Hz or fell to 90 Hz at vowel offset–onset (left versus right columns), and F_1 either fell 550 Hz to 200 Hz or 350 Hz to 400 Hz from its steady-state value of 750 Hz (top versus bottom rows). Both f_0 and F_1 transitions lasted 55 ms. F_2 and F_3 followed the same trajectories as in the speech conditions of the F_1 and $f_0 \times$ Voicing Continuation experiments (2a, b). All formant amplitudes were zeroed out during the closure.

3.3.1.2. Non-speech. The non-speech stimuli were identical to the speech stimuli, except that they lacked F_2 and F_3.

[6]The results for the negatively correlated stimuli of Experiment 2b are atypical in that discriminability was no greater in the non-speech condition than in the speech condition. Because the inclusion of higher formants in the speech condition may serve as a kind of perceptual noise, we had expected higher levels of discriminability for the non-speech stimuli (all else equal). This expectation was generally confirmed in the present study, and we have no explanation for the atypical pattern of Experiment 2b.

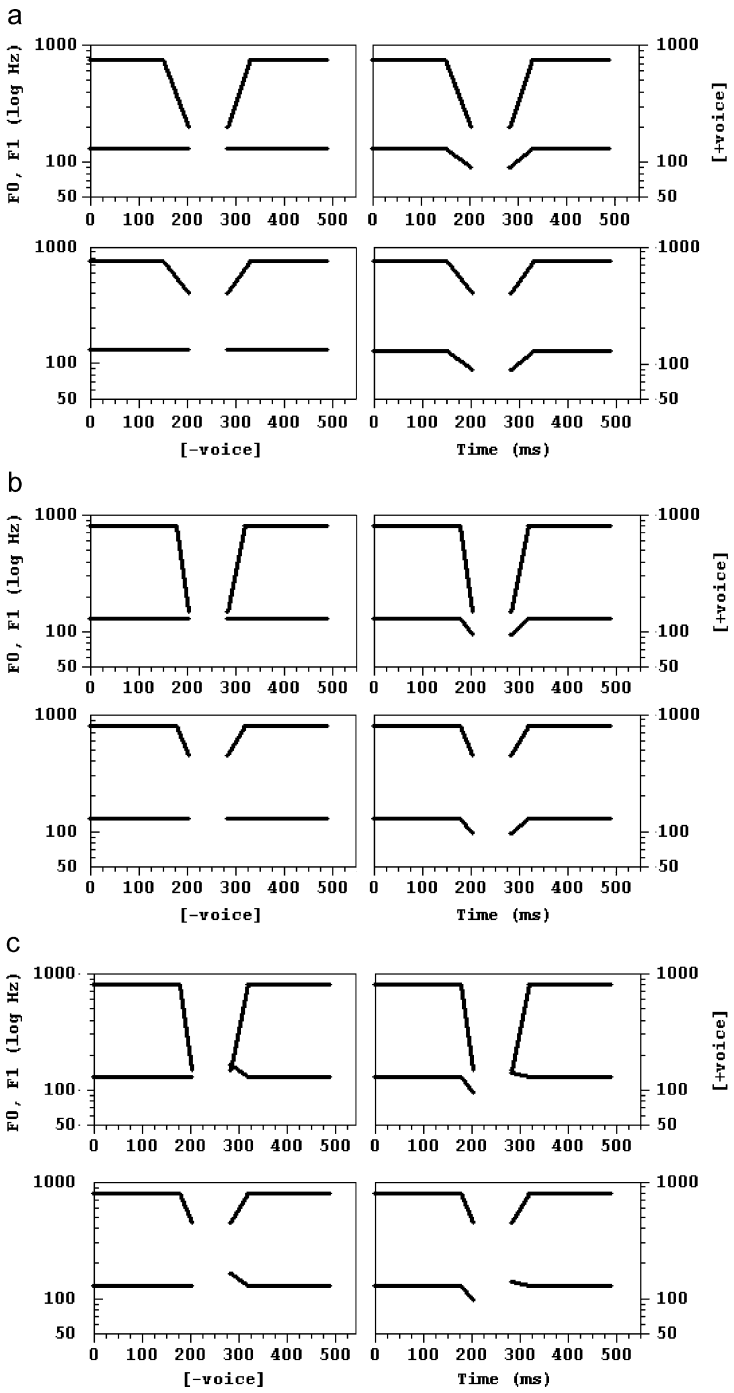

Fig. 7. 2×2 stimulus arrays showing the trajectories of the synthesis parameters in the $f_0 \times F_1$ experiments (Experiments 3a–c). Arrangement of stimuli corresponds to that in Fig. 2, in that positively correlated stimuli display the observed combinations of parameter values, and negatively correlated stimuli do not. (a) First $f_0 \times F_1$ experiment, speech and non-speech conditions: left = high f_0; offset–onset frequency of 130 Hz (0 Hz fall), right = low f_0 offset–onset frequency of 90 Hz (40 Hz fall), top = low F_1 offset–onset frequency of 200 Hz (550 Hz fall), bottom = high F_1 offset–onset frequency of 400 Hz (350 Hz fall), (b) second $f_0 \times F_1$ experiment, symmetric condition: left = high f_0 offset–onset frequency of 130 Hz (0 Hz fall), right = low f_0 offset–onset frequency of 95 Hz (35 Hz) fall, top = low F_1 offset-onset frequency of 150 Hz (650 Hz fall), bottom = high F_1 offset–onset frequency of 450 Hz (350 Hz fall), (c) asymmetric condition: left = high f_0 offset and onset frequencies of 130 and 165 Hz, respectively, right = low f_0 offset and onset frequencies of 95 and 140 Hz, respectively.

3.3.1.3. Stimulus presentation. Each of the six classifications began with 32 randomized training trials, followed by 80 randomized test trials, with feedback following the listener's response. The order of the six classifications was counterbalanced across sessions.

3.3.1.4. Results. Fig. 8a shows mean d' values across listeners in the speech and non-speech versions of this experiment (with 95% CIs). In both versions, listeners discriminated negatively correlated stimuli better than positively correlated ones, and they discriminated the non-speech stimuli better than speech stimuli for all differences.

The d' values of individual listeners were submitted to a repeated-measures ANOVA in which speech versus non-speech was a between-subjects variable and Task a within-subjects variable. The main effect of Task was significant [$F(1,28) = 5.748$, $p = 0.023$], as was the main effect of speech versus non-speech [$F(1,28) = 10.391$, $p = 0.003$], but these two variables did not interact significantly [$F < 1$]. Thus, this experiment produces a pattern of results opposite those predicted by our hypothesis, in that the combinations of f_0 and F_1 values which are predicted to produce minimally distinct values of the low-frequency property are significantly more discriminable than those predicted to have maximally distinct values for this IPP. Furthermore, this reversal was obtained for non-speech analogues as well as the original speech stimuli.

3.3.2. Experiments 3b, c: $f_0 \times F_1$, symmetric versus asymmetric f_0 trajectories

We explored the perceptual interaction between f_0 and F_1 in vowel-stop-vowel stimuli in two further experiments. The first of these additional experiments replicates the speech version of Experiment 3a. Its

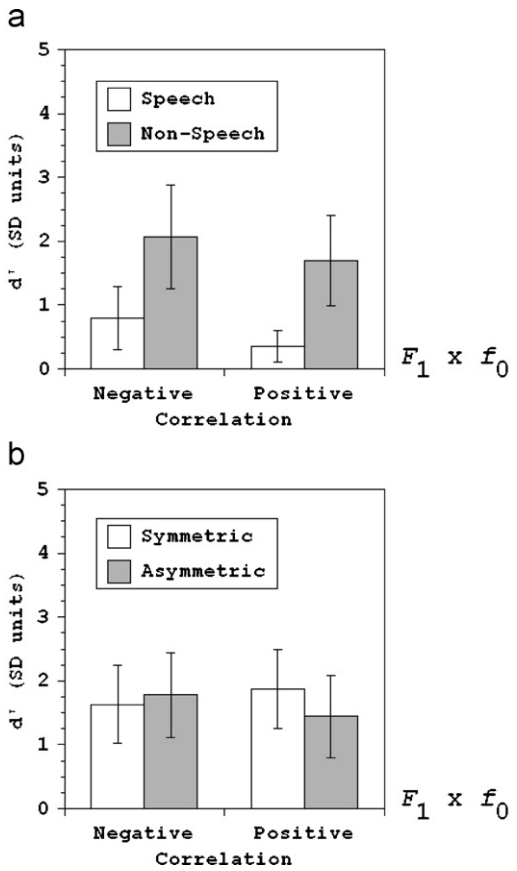

Fig. 8. Mean d' values by classification across listeners, with 95% confidence intervals, for $f_0 \times F_1$ experiments. (a) Experiment 3a: white bars = speech condition, gray bars = non-speech condition. Negative = negatively correlated classification, Positive = positively correlated classification, (b) Experiments 3b, c: white bars = symmetric condition, gray bars = asymmetric condition. Otherwise, same as (a).

purpose is to test the robustness of the reversal obtained there. The second experiment manipulates the f_0 trajectory of the vowel following the stop closure to test Silverman's (1986) claim that a lower f_0 onset frequency rather than a rising f_0 trajectory increases the likelihood of a [+voice] response to an intervocalic stop.

Silverman's claim was tested by manipulating the symmetry of the f_0 contours into and out of the stop closure. In Experiment 3b, like Experiment 3a, the f_0 trajectory into the vowel following the stop closure was the mirror image of that preceding the closure. We call this the "symmetric" f_0 trajectory. In Experiment 3c, on the other hand, the following vowel began with a falling f_0 trajectory, starting at either a high or low frequency relative to its steady-state value, and the preceding vowel ended either with the same level or falling f_0 trajectory as in the symmetric conditions. We call this the "asymmetric" f_0 trajectory. If Silverman is correct, then the asymmetric trajectory should produce the same pattern of results as the symmetric trajectory.

Twenty-three listeners were run with the symmetric stimuli, Experiment 3b, and 20 listeners with the asymmetric ones, Experiment 3c. Both conditions were run in Massachusetts.

3.3.2.1. Stimuli. The stimuli in these experiments were yet again two 205 ms five formant vowels separated by an 80 ms stop closure. Figs. 7b and 7c show how the stimuli varied. At vowel offset–onset in both Experiment 3b and 3c, F_1 either fell 650 Hz to 150 Hz or 350 Hz to 450 Hz from its 800 Hz steady-state value (top versus bottom rows). In Experiment 3b, the symmetric condition, f_0 was either constant at 130 Hz through both vowels or fell to 95 Hz at vowel offset–onset (Fig. 7b, left versus right columns). In Experiment 3c, the asymmetric condition, the f_0 trajectories at the end of the preceding vowel were the same as in the symmetric condition, but at the beginning of the following vowel, f_0 fell 35 Hz from 165 Hz or 10 Hz from 140 Hz to its 130 Hz steady-state value (Fig. 7c, left versus right columns). Both F_1 and f_0 transitions lasted 35 ms in both preceding and following vowels in both the symmetric and asymmetric stimuli. As in Experiment 3a, voicing ended with the offset of the preceding vowel and did not begin again until the onset of the following vowel. F_2 and F_3 followed the same trajectories as in earlier speech conditions. Unlike the speech stimuli used in the Experiments 2a, b, and 3a but like those used in Experiments 1a, b, these stimuli also had fourth and fifth formants, whose values were 3300 and 3850 Hz, respectively, throughout the stimuli, as Experiments 1a, b. No non-speech version of either experiment was run.

3.3.2.2. Stimulus presentation. Each of the six classifications began with 32 randomized training trials, followed by 96 randomized test trials, with feedback following the listener's response. The order of the six classifications was counterbalanced across sessions.

3.3.2.3. Results. Fig. 8b shows mean d' values across listeners in the symmetric and asymmetric conditions (with 95% CIs). In the symmetric condition (Experiment 3b), listeners discriminated the positively correlated stimuli slightly better than the negatively correlated stimuli. This finding is opposite that obtained with equally symmetrical stimuli in Experiment 3a. In the asymmetric condition (Experiment 3c), it was the negatively correlated stimuli that were slightly better discriminated.

Individual listeners d' values were submitted to a repeated-measures ANOVA in which symmetric versus asymmetric f_0 contours was a between-subjects variable and Task a within-subjects variable. The main effect of Task was not significant [$F<1$]; no significant difference was obtained between the symmetric and asymmetric conditions [$F<1$]; nor was there any interaction with Task [$F(1,41) = 1.291$, $p>.10$].

3.3.3. Discussion

The results of Experiments 3a–c do not confirm the predictions of our hypothesis. In Experiment 3a, in fact, we found that the negatively rather than the positively correlated stimuli were easier to discriminate for both speech and non-speech stimuli. Experiments 3b, c did not replicate this result, but indicated instead that f_0 and F_1 are probably separable. It is possible that adding F_4 and F_5 to the stimuli in Experiments 3b, c is responsible for this failure to replicate Experiment 3a, but it is difficult to see how the presence of these higher formants could have influenced percepts of properties much lower in the spectrum. This disconfirmation of our hypothesis requires us to modify the characterization of the low-frequency property (see Section 4). Finally, the failure to find any significant difference between the symmetric and asymmetric conditions in

Experiments 3b, c confirms Silverman's hypothesis that it is the level of f_0 rather than the direction of f_0 change that affects [voice] percepts.

4. General discussion

4.1. Summary of results

In Experiment 1a, where f_0 and closure duration were manipulated, listeners discriminated positively correlated stimuli no better than negatively correlated ones, regardless of F_1's value, while in Experiment 1b, where F_1 replaced f_0, the pair of stimuli that was easier to discriminate depended on f_0. These results show that closure duration does not integrate perceptually with either F_1 or f_0 at the edges of flanking vowels. The auditory hypothesis in fact predicted that two properties which are not auditorily similar, such as closure duration and F_1 or f_0 at flanking vowel edges, would not integrate perceptually.

The results of Experiments 2a, b were quite different: when the duration of voicing continuation into the closure replaced closure duration as the property manipulated together with F_1 or f_0, the positively correlated stimuli became consistently easier to discriminate than the negatively correlated ones. This was as true for non-speech analogues as it was for speech. These results confirm the specific positive prediction of the auditory hypothesis that these properties will integrate perceptually into the low-frequency property.

Finally, the results of Experiments 3a–c, in which F_1 and f_0 were manipulated, were inconsistent. In Experiment 3a, the negatively correlated stimuli were unexpectedly easier to discriminate than the positively correlated ones, for both non-speech analogues and speech. However, Experiment 3b failed to replicate the portion of Experiment 3a that used speech stimuli: negatively correlated stimuli were discriminated no better than the positively correlated ones. This failure to replicate suggests that these two properties do not integrate perceptually with one another after all.

In both Experiment 3a and 3b, the f_0 contours going into and out of the stop closure were mirror images of one another—f_0 was either high level into and out of the closure or fell into the closure and rose out of it. Experiment 3c tested whether f_0 must change symmetrically like this if it is not to integrate with F_1. It used f_0 contours that fell from a higher versus lower onset frequency in the vowel following the stop closure—f_0 was still level or falling at the end of the preceding vowel. If Silverman (1986) is correct in his claim that what matters is the f_0 level at the edges of flanking vowels and not its direction of change, then listeners should respond in much the same way to these asymmetric stimuli as they did to the symmetric ones used in Experiment 3b. This proved to be the case: as in Experiment 3b, the negatively correlated stimuli were not discriminated any more easily than the positively correlated ones in Experiment 3c. This also counts as a second failure to replicate the results of Experiment 3a and suggests that F_1 and f_0 at the edges of vowels flanking stop closures do not integrate with one another. If the low-frequency property were determined by simply how much low-frequency energy is present in and near the stop closure, then these two properties should have integrated in these stimuli. The fact they did not indicates that we need to redefine the low-frequency property. We do so in the next section.

4.2. Low-frequency spectral continuity not low-frequency energy

Two of the three pairs of acoustic properties integrated perceptually in the predicted way: both F_1 and f_0 integrated with Voicing Continuation. But repeated attempts to show that F_1 and f_0 also integrated with one another failed, and probably show that these two properties are instead separable.

Combining a low F_1 or f_0 at vowel edge with a long continuation of voicing into the adjacent closure produces a stimulus that is perceptually very different from one combining a high F_1 or f_0 at vowel edge with brief voicing continuation. The classification tasks showed that pairs of stimuli in which the amount by which F_1 or f_0 has fallen at the vowel edge correlates positively with the duration of Voicing Continuation are much easier to discriminate than stimuli in which these properties are instead negatively correlated. However, the discrimination of stimuli in which F_1 and f_0 are positively correlated at vowel edge is not easier than the discrimination of the corresponding negatively correlated stimuli.

What these results show is that we must redefine the low-frequency property (cf. Kingston & Diehl, 1994, 1995). That is, it is not the *amount* of low-frequency energy in the vicinity of the stop that is perceptually important but instead the *continuation* of low-frequency energy from the vowel into the consonant. Low-frequency energy continues well into the stop closure in intervocalic [+voice] stops but is abruptly cut off in [−voice] stops in this context. Low or falling F_1 or f_0 at the edges of vowels flanking stop closures enhances the percept that low-frequency energy continues across the edge into the closure.

This emphasis on low-frequency continuity brings our account back into accord with Stevens and Blumstein's (1981) original description of the "low-frequency property": "the voiced feature can be identified by testing for the presence of low-frequency spectral energy or periodicity over a time interval of 20–30 ms in the vicinity of the acoustic discontinuity that precedes or follows the consonantal constriction interval" (p. 29). They, too, described this perceptual property as an "integrated acoustic property" (p. 31).

Establishing (or rediscovering) that the IPP produced by the integration of F_1 or f_0 with voicing continuation is low-frequency continuity also explains another disconfirmed prediction. We had argued elsewhere (Kingston & Diehl, 1994, 1995) that F_1 and f_0 are both relatively low at vowel onset in syllable-initial stops with shorter but still lagging voice onset times (VOT) and that both F_1 and f_0 are relatively high with longer VOTs because that covariation increased the amount of low-frequency energy in the stop's vicinity. This hypothesis predicts that lowering the spectral center of gravity in the vowel would increase [+voice] responses for a given lagging VOT, even though there is no low-frequency energy whatsoever in the consonant interval.

In separate conditions in a categorization experiment, Molis and Diehl (1996) tested this prediction by either increasing the intensity of F_1 itself or by increasing the rate at which energy fell off with increasing frequency in the source spectrum. Neither manipulation shifted [voice] judgments relative to VOT. Both failures are expected if F_1 at vowel onset does not integrate with lagging VOT through their joint contribution to the height or concentration of energy at the bottom of spectrum. (A similar failure is now predicted for analogous increases in the intensity of the first harmonic.)

And when we add these failures to the failure of F_1 and f_0 to integrate with one another at the edges of vowels flanking intervocalic stops, we see that it is not the amount of low-frequency energy that matters perceptually.

Another problem with the failure of F_1 and f_0 to integrate is more apparent than real. That is, how can F_1 and f_0 not integrate with one another if each integrates with Voicing Continuation into the same IPP? The answer can be found in the redefinition of the low-frequency property as continuity of low-frequency energy between vowel and consonant intervals rather than the amount of low-frequency energy present near the stop. Low or falling F_1 or f_0 at vowel edges can each readily enhance the percept created by voicing continuation that low-frequency energy continues between the vowel and consonant intervals, without either influencing the other's enhancement of this percept. Indeed, without voicing actually continuing into the closure, neither F_1 nor f_0 at the vowel edge could create any spectral continuity at all between these intervals. If the low-frequency property instead corresponded to how much low-frequency energy is present near the stop, then F_1 and f_0 would instead be expected to integrate because they can both raise or lower its amount.

For our purposes, spectral continuity needs only to be greater between vowel and consonant when either F_1 or f_0 is low or falling at the vowels' edges. In such a case, energy at the vowels' edges is more continuous at low frequencies with that found in a neighboring stop closure into which voicing has continued than when F_1 or f_0 is high or rising. That is, perceptually important differences are ordinal not absolute. Similar proposals that the recognition of place of articulation depends on whether energy is continuous between consonant and vowel within some frequency range have been made by Kewley-Port, Pisoni, and Studdert-Kennedy (1983) and Lahiri, Gewirth, and Blumstein (1984).

4.3. Trading relations without integration

Nonetheless, when listeners categorize stops as [±voice], there are two ways in which F_1 and f_0 trade with the timing of voice offset/onset that do not depend on whether low-frequency energy continues from the vowel well into the consonant. That is, categorization of stops with respect to voice timing varies with F_1 or f_0 at

vowel edge in ways that cannot be attributed to the integration of these properties into the low-frequency property, even redefined as continuity of low-frequency energy between vowel and consonant.

4.3.1. F_1 and f_0 without Voicing Continuation

Both F_1 and f_0 trade with short-to-long lag VOT values in syllable-initial stops. Syllable-initial [+voice] stops in English and some other languages are typically produced without any voicing during the closure; voicing instead begins no earlier than shortly after the closure's release (Caisse, 1982; Docherty, 1989; Lisker & Abramson, 1964). These [+voice] stops are thus pronounced without any low-frequency energy continuing across the consonant–vowel border. Despite the absence of any such low-frequency continuity, a low F_1 or f_0 at voice onset consistently increases [+voice] responses for a given VOT (Benkí, 2001; Diehl & Molis, 1995; Haggard, Ambler, & Callow, 1970; Haggard, Summerfield, & Roberts, 1981; Kluender, 1991;[7] Lisker, 1975; f_0: Massaro & Cohen, 1976; F_1: Stevens & Klatt, 1974; Summerfield & Haggard, 1977; Whalen, Abramson, Lisker, & Mody, 1990).[8] These robust results are surprising in light of our explanation of F_1 and f_0's failure to integrate perceptually because voicing simply does not continue into the stop closure in the typical short lag allophone of syllable-initial [+voice] stops, so there is no voicing continuation for F_1 or f_0 to integrate with.

The best percept of low-frequency continuity is likely to arise in syllable-final and intervocalic [+voice] stops where voicing more often continues far into the stop closure, and F_1 and f_0 are low at the border between those two segments. When stops are instead syllable-initial in English, voicing typically does not begin in the stop closure but at the soonest right after the stop's release, in the common unaspirated or short lag allophone of syllable-initial [+voice] stops. As a result, low-frequency energy does not continue from the stop into the following vowel. F_1 or f_0 at vowel edge may nonetheless still trade with lagging VOT because listeners have observed that [+voice] stops have low F_1 and f_0 at vowel edges in the syllable-final and intervocalic contexts where voicing does continue into the closure and these spectral properties can integrate with it. These observations may lead them to treat low F_1 and f_0 at vowel edge as [+voice] correlates even in the syllable-initial contexts in which voicing does not continue into the stop. If this scenario is correct, these trading relations are a product of learned association and not integration.

4.3.2. Steady-state f_1 and f_0 and vowel and closure duration

Steady-state as well as offset F_1 and f_0 values trade with other acoustic correlates of the [voice] contrast in syllable-final and intervocalic stops. In categorizing syllable-final and intervocalic stops for [voice], listeners respond [+voice] more often when the preceding F_1 or f_0 offset value is low and also when the entire preceding vowel's F_1 or f_0 steady-state is low (cf. Fischer & Ohde,1990; f_0: Castleman & Diehl, 1996; F_1: Summers, 1988). These studies manipulated vowel or closure duration and not how long voicing continued into the stop.

The fixed classification experiments reported above (Experiments 1a, b) showed that neither F_1 nor f_0 at vowel edge integrates with following closure duration. We had in fact predicted that neither spectral property would integrate with this temporal property because combining low F_1 or f_0 with a short closure duration or high F_1 or f_0 with a long closure duration does not increase or decrease the amount of low-frequency energy near the consonant. These combinations also do not affect the continuity of low-frequency energy across the border between vowel and consonant. Nor can F_1 or f_0 at vowel edge covary with vowel duration to affect the amount or continuity of low-frequency energy. Even more surprising is the trading between steady-state F_1 or f_0 values and closure or vowel duration, as those values cannot affect the perceived continuity of low-frequency energy into the stop closure.

[7]José Benkí (p.c.) points out that Kluender (1991) collected responses from a non-human species, Japanese quail, as well as human listeners, and that the non-human listeners's responses cannot be determined by learned associations between VOT and F_1, even if the human listeners can. We still cannot tell from the categorization data reported by Kluender whether the two dimensions integrated for the non-human listeners.

[8]The fact that low-frequency energy does not continue across the consonant-vowel edge in [+voice] stops when they are pronounced with short lag VOT values may explain why relatively few languages implement the [voice] contrast with short versus long lag VOTs as compared to the more common voicing lead versus short lag (Keating, Linker, & Huffman, 1983). However, the more common realization itself reduces the phonetic differences between stops contrasting for [voice], because F_1 onset frequency differs little or not at all following voicing lead versus short lag syllable-initial stops. F_0 onset frequency still differs reliably between these stops, and thus contributes to the low-frequency percept (Kingston & Diehl, 1994).

Although it is unsurprising that neither F_1 nor f_0 integrate with closure duration, because no combination of these properties affects low-frequency (dis)continuity at the vowel–consonant border, the F_1 value at vowel edge nonetheless determines whether closure duration integrates perceptually with voicing continuation into the closure.

The likelihood that a listener will identify a postvocalic stop as [+voice] is known to be affected by the preceding vowel's duration (Denes, 1955; Raphael, 1972), the stop's own closure duration (Lisker, 1957), and by the ratio of these two durations to one another (Kohler, 1979; Port & Dalby, 1982), with the consonant:vowel (C:V) duration ratio being small for [+voice] stops and large for [−voice] ones. Two acoustic properties that contribute to the percept of the low-frequency property, F_1 at vowel edge and closure voicing, not only interact perceptually with this ratio, but do so in a way that depends on their own integration. Lisker (1978) showed that listeners respond [+voice] more often to a given closure duration if some voicing is present in the closure. More than 50 or 60 ms of voicing produces [+voice] responses even to stops with very long closure durations. Voicing probably shortens the perceived duration of the closure, either by filling it partly with low-frequency periodic energy or by weakening the low-frequency discontinuity between it and the flanking vowels. This shortening in turn further reduces the C:V duration ratio.

However, voicing does not make the closure sound shorter unless the voicing is spectrally continuous with the flanking vowels. Parker et al. (1986) were the first to demonstrate that the shortened percept depended on spectral continuity in a study of how judgments of the duration of a gap between two square waves are affected by low-frequency periodic energy in the gap. Their square wave-gap-square wave stimuli were non-speech analogues of an [aba-apa] series, in which the square waves mimicked the vowels and the gap mimicked the stop closure. The gap's duration varied, and the gap either began with low-frequency periodic energy—henceforth, pulsing—or had no pulsing. When the square waves' f_0 was either constant into and out of the gap or when it rose into the gap and fell out of it, pulsing in the gap did not affect listeners' judgments of the gap's duration. However, when the square waves' f_0 fell into the gap and rose out of it, pulsing produced more "short" responses for a given gap duration. Parker et al. argued that pulsing could shorten the gap in this condition because a falling–rising f_0 made the pulsing spectrally continuous with the flanking square waves, which permitted listeners to incorporate the pulsing into the square wave-gap-square wave percept.

This result did not allow us to identify which of two sources of spectral continuity permit voicing to shorten perceived closure duration in speech signals, because both F_1 and f_0 fall into and rise out of [+voice] stop closures. Kingston et al. (1990, published in Kingston & Diehl, 1995) used single-formant vowel analogues in which F_1 and f_0 contours could be independently varied, and found that a falling–rising F_1 but not a falling–rising f_0 contour permitted pulsing to shorten perceived gap durations significantly. This result showed that spectral continuity is created by progressive, continuous low-pass filtering of the signal from the vowel into the consonant analogue and not by the frequency match between the periodic sources in the vowel and consonant analogues.[9]

It remains surprising that Kingston et al. (1990) did not find that pulsing traded with gap duration when the f_0 contour was falling–rising, given that f_0 at vowel edge integrates perceptually with voicing continuation. The observed integration predicts the same increase in "short" responses when pulsing is present and f_0 is falling–rising as when F_1 is. This dissociation has two unexplained features. First, it differs from the others discussed here in that the two acoustic properties influence each other's perception in fixed classification but not categorization tasks; in the other cases, they do so in categorization but not fixed classification. Second, it is a judgment of whether a gap (or closure) is short or long that is affected more by pulsing when F_1 is falling–rising but not when f_0 is. Why should approximate spectral continuity with the first formant permit pulsing to be incorporated into the same percept as the flanking vowel analogues, but actual first harmonic (f_0) continuity alone not do so? Further research is required to explain both these features.

[9]It is worth asking whether spectral continuity is determined by (the analogue of) resonance but not source characteristics because that effect most closely resembles the progressive low-pass filtering of the signal by the vocal tract constriction in the original speech. In other words, does this perceptual interaction arise because listeners attribute the array of acoustic properties they hear to a sound-producing device whose physics is the same as the vocal tract's? According to Fowler (1990, 1991), similarities between listeners' responses to non-speech analogues and the original speech can arise only in this way. Space does not permit a response here, but the reader may wish to consult Diehl, Walsh, and Kluender's (1991) response to Fowler, as well their original paper (Diehl & Walsh, 1989). Fowler (1996) and Diehl et al. (2004) present further arguments.

4.3.3. Bias or sensitivity?

In the Introduction to this paper, we showed how a trading relation could arise from a shift in response bias alone for perceptually separable dimensions, as well as from actual integration of those dimensions. The only way to determine the source of a trading relation is to compare sensitivity to differences between pairs of stimuli in which the values of these dimensions are positively versus negatively correlated, as in the fixed classification tasks reported in this paper. It may turn out that the trading relations just discussed between F_1 and f_0, on the one hand, and VOT and vowel and closure duration, on the other, do not arise from the integration of these dimensions at all. That is, they may arise from shifts in response bias that are independent of changes in sensitivity brought about by integration of acoustic properties. Until sensitivity to correlated differences between these pairs of dimensions is measured, it is impossible to tell whether these trading relations represent only decisional integrality or perceptual integrality, too (see Macmillan et al., 1999, for further discussion).

4.4. Associative and gestural hypotheses

How do the present results square with either the associative or the gestural hypotheses, which were briefly described in the Introduction? Recall that the speech stimuli in each of our 2×2 arrays consisted of acoustic variants within the same phonological category. It is not obvious what perceptual effects the associative hypothesis would predict for such stimulus arrays. Normal speech experience might yield a negative correlation (in the sense previously used) between the two perceptual dimensions of an array if acoustic properties typically trade off within a category (e.g., [+voice] stops). Alternatively, normal variation in speech clarity between reduced forms and hyperarticulated forms (Lindblom, 1990) might produce a positive correlation between the same two dimensions. These opposing patterns of within-category acoustic covariation would, in turn, be expected to yield a perceptual advantage for the negatively correlated condition and the positively correlated condition, respectively, assuming that perceptual coherence derives from associative learning. The important point is that whichever direction of effect is predicted by the associative hypothesis, that prediction should hold for the speech conditions in all three sets of experiments reported here. The failure to observe such parallel results thus appears to disconfirm the associative hypothesis.

The present results supported the prediction of the auditory hypothesis that the pattern of discrimination performance for the speech stimuli should be in the same direction as that for the analogous non-speech stimuli. It is not clear how either the associative or the gestural hypotheses can explain these speech/non-speech parallels, at least without invoking additional assumptions.

4.5. Concluding summary

An IPP is the product of integration, so it may have practically the same value for stimuli that differ in their values for the individual acoustic properties. This many-to-one mapping of acoustic to intermediate perceptual properties goes a long way toward explaining how measurably different stimuli would be categorized as the same, i.e, how listeners can assign the same distinctive feature value to the physically different realizations that occur in different contexts. These intermediate perceptual properties may thereby link the concrete variety of the different allophones of phonemes to their phonological representations. These phonological representations may appear to be quite abstract but are in fact as concrete as the intermediate perceptual properties that convey them to listeners.

Acknowledgments

We gratefully acknowledge the support to the first and third authors through Grant R29 DC01708-03, 04, from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, and through Grant DBS92-12043 from NSF, and to the second and fourth authors from Grant R01 DC00427-12, 13, 14, 15 also from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We also thank José Benkí, Terry Nearey, and several anonymous reviewers for their help in improving the quality and clarity of the argument.

Appendix A

See Table A1.

Table A1
Average d' values (95% confidence intervals) obtained in the single-dimension tasks in each experiment

Experiment	Dimension of classification, value of other dimension			
	f_0, short CD	f_0, long CD	CD, low f_0	CD, high f_0
1a: High F_1	0.764 (0.727)	0.482 (0.402)	3.118 (0.784)	1.639 (0.779)
1a: Low F_1	0.750 (0.440)	1.129 (0.789)	2.144 (0.749)	1.664 (0.669)
	F_1, short CD	F_1, long CD	CD, low F_1	CD, high F_1
1b: High f_0	0.719 (0.750)	0.951 (0.704)	1.884 (0.579)	0.135 (0.197)
1b: Low f_0	2.163 (0.935)	0.618 (0.322)	1.975 (0.776)	3.192 (0.573)
	F_1, short VC	F_1, long VC	VC, low F_1	VC, high F_1
2a: Speech	0.993 (0.593)	1.882 (0.750)	0.568 (0.417)	0.411 (0.302)
2a: Non-speech	3.602 (1.104)	3.142 (0.759)	0.488 (0.552)	0.836 (0.583)
	f_0, short VC	f_0, long VC	VC, low f_0	VC, high f_0
2b: Speech	1.259 (0.553)	0.807 (0.473)	0.545 (0.505)	0.635 (0.317)
2b: Non-speech	0.925 (0.414)	1.373 (0.591)	1.668 (0.675)	0.866 (0.395)
	F_1, low f_0	F_1, high f_0	f_0, low F_1	f_0, high F_1
3a: Speech	1.026 (0.526)	0.769 (0.508)	0.327 (0.298)	0.231 (0.396)
3a: Non-speech	1.661 (0.676)	1.741 (0.682)	0.845 (0.497)	1.119 (0.627)
3b	1.369 (0.581)	1.075 (0.578)	0.561 (0.416)	0.557 (0.476)
3c	0.603 (0.349)	1.253 (0.559)	1.09 (0.623)	0.745 (0.481)

The first term in each column heading identifies the dimension along which the two stimuli differ, the second the value for the other dimension.

References

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 598–612.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.

Benkí, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, *29*, 1–22.

Best, C. T., Morrongiello, B. A., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception and Psychophysics*, *29*, 191–211.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, *66*, 1001–1017.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra from stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, *67*, 648–662.

Caisse, M. (1982). *Cross-linguistic differences in fundamental frequency induced by voiceless aspirated stops*. M.A. thesis, University of California, Berkeley.

Castleman, W. A., & Diehl, R. L. (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics*, *24*, 383–398.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, *27*, 761–764.

Diehl, R. L., & Kingston, J. (1991). Phonetic covariation as auditory enhancement: The case of the [+voice]/[−voice] distinction. In O. Engstrand, & C. Kylander (Eds.), *Current phonetic research paradigms: Implications for speech motor control, PERILUS XIV* (pp. 139–143). Stockholm: Stockholm University.

Diehl, R. L., Kingston, J., & Castleman, W. A. (1995). On the internal perceptual structure of phonological features: The [voice] distinction. *Journal of the Acoustical Society of America*, *97*, 3333 (abstract).

Diehl, R. L., & Molis, M. R. (1995). Effect of fundamental frequency on medial [voice] judgments. *Phonetica*, *52*, 188–195.

Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America*, *85*, 2154–2164.

Diehl, R. L., Walsh, M. A., & Kluender, K. R. (1991). On the interpretability of speech/nonspeech comparisons: A reply to Fowler. *Journal of the Acoustical Society of America*, *89*, 2905–2909.

Docherty, G. J. (1989). *An experimental study of the timing of voicing in English obstruents*. Ph.D. dissertation, University of Edinburgh.

Fischer, R. M., & Ohde, R. N. (1990). Spectral and duration properties of front vowels as cues to final stop-consonant voicing. *Journal of the Acoustical Society of America*, *88*, 1250–1259.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception and Psychophysics*, *27*, 343–350.

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

Fowler, C. A. (1990). Sound-producing sources as the objects of perception: Rate normalization and nonspeech perception. *Journal of the Acoustical Society of America*, *88*, 1236–1249.

Fowler, C. A. (1991). Auditory perception is not special: We see the world, we feel the world, we hear the world. *Journal of the Acoustical Society of America*, *89*, 2910–2915.

Fowler, C. A. (1992). Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics*, *20*, 143–165.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, *99*, 1730–1741.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum Associates.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America*, *47*, 613–617.

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading of cues in the voiced–voiceless distinction. *Journal of Phonetics*, *9*, 49–62.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *Journal of the Acoustical Society of America*, *109*, 764–774.

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, *73*, 1779–1793.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*, 419–454.

Kingston, J., & Diehl, R. L. (1995). Intermediate properties in the perception of distinctive feature values. In A. Arvaniti, & B. Connell (Eds.), *Papers in laboratory phonology IV* (pp. 7–27). Cambridge, UK: Cambridge UP.

Kingston, J., Diehl, R. L., Kluender, K. R., & Parker, E. M. (1990). Resonance vs. source characteristics in perceiving spectral continuity between vowels and consonants. *Journal of the Acoustical Society of America*, *88*, S55 (abstract).

Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America*, *97*, 1261–1285.

Kingston, J., Macmillan, N. A., Walsh Dickey, L., Thorburn, R., & Bartels, C. (1997). Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America*, *101*, 1696–1709.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among male and female talkers. *Journal of the Acoustical Society of America*, *87*, 820–856.

Kluender, K. R. (1991). Effects of first formant onset properties on VOT judgments can be explained by auditory processes not specific to humans. *Journal of the Acoustical Society of America*, *90*, 83–96.

Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 173–217). San Diego: Academic Press.

Kohler, K. J. (1979). Dimensions in the perception of fortis and lenis plosives. *Phonetica*, *36*, 332–343.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *Journal of the Acoustical Society of America*, *76*, 391–404.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Dordrecht: Kluwer.

Lisker, L. (1957). Closure duration and the intervocalic voiced–voiceless distinction in English. *Language*, *33*, 42–49.

Lisker, L. (1975). Is it VOT or a first formant transition detector? *Journal of the Acoustical Society of America*, *57*, 1547–1551.

Lisker, L. (1978). On buzzing the English [b]. Haskins Laboratories. *Status Report on Speech Research*, *55/56*, 181–188.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/versus/p/ in trochees. *Language and Speech*, *19*, 3–11.

Lisker, L., & Abramson, A. S. (1964). A cross-linguistic study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Macmillan, N. A., & Kingston, J. (1995). Integrality, correspondence, and configurality represent different degrees of perceptual interaction, not different types. In C.-A. Possamaï (Ed.), *Fechner Day 1995* (pp. 243–248). Cassis: International Society for Psychophysics.

Macmillan, N. A., Kingston, J., Thorburn, R., Walsh Dickey, L., & Bartels, C. (1999). Integrality of nasalization and F_1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision-rule interactions. *Journal of the Acoustical Society of America*, *106*, 2913–2932.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W., & Cohen, M. M. (1976). The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *Journal of the Acoustical Society of America*, *60*, 704–717.

Molis, M. R., & Diehl, R. L. (1996). First formant spectral properties and the initial stop-consonant [voice] judgments. *Journal of the Acoustical Society of America*, *99*, 2591 (abstract).

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, *19*, 347–373.

Nearey, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, *101*, 3241–3254.

Parker, E. M., Diehl, R. L., & Kluender, K. R. (1986). Trading relations in speech and nonspeech. *Perception and Psychophysics*, *39*, 129–142.

Port, R., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics*, *32*, 141–152.

Raphael, L. F. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, *51*, 1296–1303.

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

Silverman, K. E. A. (1986). F_0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica*, *43*, 76–91.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64*, 1358–1368.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas, & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1–38). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced–voiceless distinction for stops. *Journal of the Acoustical Society of America*, *55*, 653–659.

Summerfield, A. Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, *62*, 435–448.

Summers, W. V. (1988). F_1 structure provides information for final-consonant voicing. *Journal of the Acoustical Society of America*, *84*, 485–493.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica*, *47*, 36–49.