

On the relevance of locus equations for production and perception of stop consonants

LAWRENCE BRANCAZIO and CAROL A. FOWLER
*Haskins Laboratories, New Haven, Connecticut
and University of Connecticut, Storrs, Connecticut*

We examined the possible relevance of locus equations to human production and perception of stop consonants. The orderly output constraint (OOC) of Sussman, Fruchter, and Cable (1995) claims that humans have evolved to produce speech such that $F2$ at consonant release and $F2$ at vowel midpoint are linearly related for consonants so that developing perceptual systems can form representations in an $F2_{\text{ons}}$ -by- $F2_{\text{vowel}}$ space. The theory claims that this relationship described by locus equations can distinguish consonants, and that the linearity of locus equations is captured in neural representations and is thus perceptually relevant. We investigated these claims by testing how closely locus equations reflect the production and perception of stop consonants. In Experiment 1, we induced speakers to change their locus equation slope and intercept parameters systematically, but found that consonants remained distinctive in slope-by-intercept space. In Experiment 2, we presented stop-consonant syllables with their bursts removed to listeners, and compared their classification error matrices with the predictions of a model using locus equation prototypes and with those of an exemplar-based model that uses $F2_{\text{ons}}$ and $F2_{\text{vowel}}$, but not locus equations. Both models failed to account for a large proportion of the variance in listeners' responses; the locus equation model was no better in its predictions than the exemplar model. These findings are discussed in the context of the OOC.

It is well known among speech scientists that formant transitions provide listeners with important information for identifying the stop consonants /b/, /d/, and /g/ in syllable-initial position. However, it is a matter of debate how such transition information is used. In particular, the importance of these highly context-sensitive transitions to perceivers provides a challenge for theorists who claim that listeners use invariant information to identify stops (e.g., Fowler, 1986; Stevens & Blumstein, 1981). One early hypothesized invariant was a *locus* to which all second-formant transitions of a given consonant pointed back in time by extrapolation (Delattre, Liberman, & Cooper, 1955). Locus points were found for /b/ and /d/, but /g/ required two loci; and even the loci for /b/ and /d/ proved controversial. For example, Öhman (1966) found that the locus points were not invariant, and that they changed, depending on the identity of a preceding vowel. Later, Kewley-Port (1982) made extensive acoustic measurements of formant transitions for the stop consonants in various vocalic contexts and concluded against any invariant in the transition. Transitions, therefore, appear to be inherently context sensitive. As a result, theories assuming both the importance of invariance for perception and a role for formant transitions have taken different directions. One is to assume, as motor theorists have, that

invariance is found only at the level of production, and that specialized algorithms are needed to go from context-sensitive acoustic signals to invariant articulatory gestures (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). Another is to assume that perception is of invariant articulatory gestures, which are specified in the acoustic signal, and that either there is invariant information in the signal that has not yet been discovered, or the specification of the gesture in a signal differs in different contexts, but specifies nonetheless (Fowler, 1994). A third approach is to look for invariance not in the acoustic signals of individual utterances, but rather as a higher order property found when appropriate context-sensitive acoustic signals are related to one another (Sussman, 1989; Sussman, McCaffrey, & Matthews, 1991). This approach will be our main focus. A final solution is to abandon the notion of invariance altogether. In this approach, phonemes are considered to be abstract, acoustically (and optically) defined categories. Transitions are cues to be matched to templates, and any notion of invariance is unnecessary (see, e.g., Massaro, 1987).

Locus Equations

Sussman proposed a partial solution to the invariance problem that involves the derivation of *locus equations* based on second-formant transitions (e.g., Sussman, 1989; Sussman et al., 1991). To derive locus equations, second-formant ($F2$) values at the midpoints of a variety of vowels ($F2_{\text{vowel}}$) are plotted on the horizontal axis of a two-dimensional space, and $F2$ at the onset of voicing ($F2_{\text{ons}}$) following the release of a given consonant is plotted on the vertical axis (see also Klatt, 1987; Lindblom, 1963;

This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. We would like to thank Harvey Sussman, Björn Lindblom, Doug Whalen, and an anonymous reviewer for their helpful comments. Correspondence concerning this article may be sent to L. Brancazio, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511 (e-mail: lab93006@uconnvm.uconn.edu).

Nearey & Shammass, 1987). Regression line fits to these plots reveal a linear relationship between $F2_{\text{vowel}}$ and $F2_{\text{ons}}$. The regression lines for different consonants, or the locus equations that represent them, may be characterized by their slopes and intercepts. It has been found that the locus equation lines for different stop consonants occupy distinct regions in slope-by-intercept space (at least, if they are based on productions of speakers of a single language; Sussman, Hoemeke, & Ahmed, 1993).

Locus equations represent an invariant that is not found in a particular token utterance, but rather over sets of consonant–vowel (CV) syllables that share their initial consonant. Thus, locus equation slopes represent *relational* invariants (Sussman et al., 1991) rather than absolute invariants of the sort that Stevens and Blumstein (e.g., 1981) and others have sought (in which a single invariant cue for a consonant is found in all vocalic contexts).

The relational invariance approach has implications for construction of a locus-equation–based model of consonant perception. Because the systematicity captured by a locus equation is not explicitly contained within a single token, which is all that listeners hear at a given time, some sort of representational system is required so that a stimulus token may be related to other (nonpresent) tokens in order to uncover the relevant property. Sussman has proposed that the speech perception system has a neurally instantiated “2-D map of a bivariate acoustic space in which linear functions represent categories” (Sussman, Fruchter, Hilbert, & Sirosh, in press). In this space, $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ are the mapped acoustic parameters. In earlier papers, Sussman (1989; Sussman et al., 1991) proposed that within this array there are *combination-sensitive neurons* that respond to particular pairs of $F2$ at transition onset and $F2$ at vowel midpoint, extracted from the incoming speech signal. These combination-sensitive neurons specific to particular CV syllables are arranged in columns so that a set of neurons in a column all correspond to the same consonant. The $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ pairs represented in the array correspond to the region in $F2$ space captured by the locus equation for the particular consonant. The output of neurons in the array activates a higher level neuron representing the consonant; these higher level neurons may receive input from other systems as well, possibly ones processing burst and $F3$ information.

This proposed system was inspired by evidence of analogously operating mechanisms in the auditory systems of the bat, for echolocation (Olsen & Suga, 1991; Suga, O’Neill, Kujirai, & Manabe, 1983), and of the barn owl, for sound-source azimuth (Wagner, Takahashi, & Konishi, 1987). In both of these systems, combination-sensitive neurons have been isolated that respond to frequency pairs (in the bat) and frequency–phase-difference pairs (in the barn owl). The bat’s combination-sensitive neurons map the relative velocity of a moving object according to the Doppler shift of a frequency pair—the outgoing sonar pulse frequency and the returning pulse frequency. Relative velocity is a linear function of the Doppler shift; that

is, for a given relative object velocity, all pulse frequencies plotted against their echoes form a straight line. In the barn owl, frequency-specific neurons in the inferior colliculus have been found to respond to a particular phase between the sine waves of a common frequency at the two ears. These frequency–phase pairings correspond to interaural time difference, and thus, to location of the sound source in space (Wagner et al., 1987). Moreover, these neurons appear to be arranged by frequency on one axis and by phase difference on another axis, so that a third axis can be drawn that represents the emergent property of interaural time difference, thus forming the analogical basis for Sussman’s neural model of consonant perception, in which locus equation slope is an emergent property (Sussman, 1989; Sussman et al., in press).

In addition to developing his locus-equation–inspired model of speech perception, Sussman and colleagues (Sussman, Fruchter, & Cable, 1995; Sussman et al., in press) have offered a related account of why good regression fits to locus equations are found. They claim that there is a general preference of auditory systems for linearly related input parameters (as evidenced by the barn owl and bat systems). Perhaps in humans the organization of the speech perception system (which forms an auditory map of $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ space) is facilitated by such an orderly signal, and talkers have therefore evolved to coarticulate so as to produce the linear relation captured by locus equations. In other words, talkers are constrained to produce linearly ordered $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values in CVs to allow an efficient representational system to develop and function (Sussman et al., in press). This hypothesized constraint has been termed the *orderly output constraint* (OOC).

Purpose of the Present Research

The linear relation between $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ of CVs that share their initial consonant but differ in their vowels may be interesting both for what it may tell us about speech production and for what it may imply about speech perception. Accordingly, we focused in Experiment 1 on production issues, generating stimuli that we presented to listeners in Experiment 2, in which we focused on perception issues.

Regarding production, one question concerns the degree to which the slopes of the locus equations are, in fact, free to vary. Locus equation slopes are assumed to reflect the degree of coarticulatory overlap between a consonant and following vowel (Fowler, 1994; Krull, 1989; Sussman et al., 1991). An increase in $F2_{\text{vowel}}$ is associated with an increase in $F2_{\text{ons}}$ because of *anticipatory* coarticulation of the vowel during consonant production. At the moment of consonant release, the vocal tract’s configuration will already reflect movement toward the vowel’s target configuration; thus, the acoustic signal at release will differ according to the vowel target.

The slope differences associated with lines for different consonants may be understood in terms of consonants’ having different degrees of *coarticulation resistance* (see,

e.g., Bladon & Al-Bamerni, 1976; Recasens, 1985, 1991; cf. Fowler, 1994) to following vowels. In general, consonants resist coarticulatory overlap by a vowel to the extent that the vowel and consonant compete for use of the tongue. If a consonant is highly resistant to vowel influences (when production of the consonant gesture requires movement of articulators that are used in producing the vowel gesture), $F2_{\text{ons}}$ will vary little over vowel context (Recasens, 1985). Fowler (1994) has argued that locus equation slope is correlated with place of articulation (because consonants having different places of articulation differ in degree of coarticulation resistance) but is not a property of consonant place. This would suggest that any conditions that should increase or decrease a consonant's coarticulation resistance might change the slope of its locus equation. Research by Browman and Goldstein (1990) and Engstrand (1988), among others, shows that increases in speaking rate, or in the casualness of speech, are associated with increases in coarticulatory overlap and, by implication, decreases in coarticulation resistance. Compatibly, it has been found in a number of studies that manipulations of speech styles result in locus equation slope changes within a given place of articulation (Bakran & Mildner, 1995; Crowther, 1994; Duez, 1992; Krull, 1989). Experiment 1 was designed to explore the relationship between variability in slope and variability of speaking style. We also used it to explore the questions of whether slopes for different consonants remain distinctive under manipulations of speaking style, and whether different consonants always fall in distinct regions in slope-by-intercept space, as claimed by Sussman and colleagues (e.g., Sussman et al., in press; Sussman et al., 1991).

A second production issue concerns a finding from Chennoukh, Carré, and Lindblom (1995), who used the slopes and intercepts of locus equations produced by their production model as they varied coarticulatory overlap of a consonant with a variety of vowels. When locus equation slopes are plotted against their corresponding intercepts, the regression of intercept on slope within a given place of articulation reveals a good linear fit, and Chennoukh et al. have termed this regression equation a *second-order locus equation*, or SOLE. This linearity has also been observed by Crowther (1994) and Bakran and Mildner (1995) in natural productions. Linearity of SOLEs is of interest because we believe that it is not an outcome that the OOC would predict (as we will argue later), and it will lead us to ask whether an OOC account is required in order to explain first-order linearity. In Experiment 1, we sought to determine whether our speaking style manipulations would also result in good-fitting SOLEs.

As for perception, the major question we address here is whether evidence can be found to support the claim of Sussman and colleagues (e.g., Fruchter & Sussman, 1996; Sussman et al., in press; Sussman, et al., 1991) that the information captured by locus equations is relevant to perceivers. More specifically, we attempted in Experiment 2 to determine the degree to which the description

of second-formant transitions by the locus equation metric matches the information exploited by perceivers, by studying listener identification error patterns. Experiment 2 provides a comparison of the predictions of a speech perception model based on locus equations with another model that also uses $F2$ transition endpoints but not the assumption of a linear relation among syllables represented by these coordinates. We determined the relative extents to which each model accurately predicts the CVs in which listeners will correctly identify the consonant and those in which they will make errors. When, as predicted by each model, listeners did make errors, we asked whether the models correctly predicted the identity of the erroneous response. Generally, we attempted to determine whether a model in which listeners must be assumed to appreciate the historical tendency for the relation between $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ of consonant-invariant CVs to be linear better predicts listeners' data than models in which that assumption is not made.

EXPERIMENT 1

Many studies show that locus equation slopes are associated with degree of coarticulatory overlap between a particular consonant and following vowels. For example, Krull (1989) and Duez (1992) both found that slopes (for productions of a given place of articulation) were somewhat steeper in spontaneous speech than in list-read speech, and Bakran and Mildner (1995), Duez (1992), and Krull (1989) also found somewhat steeper slopes in unstressed syllables than in stressed syllables. Thus, in speaking styles in which more coarticulatory overlap is found (Browman & Goldstein, 1990; Engstrand, 1988), slopes are steeper. Further, Fowler (1994) demonstrated that consonants of different manner classes at the same place of articulation differed systematically in their locus equation slopes. Specifically, a fricative (a more resistant consonant; see Recasens, 1991) at a given place of articulation (such as /z/) had a shallower slope than did a stop at the same place (/d/).¹ Thus, with place of articulation held constant and coarticulation resistance changed, slope varied with the change in resistance. Likewise, two consonants that differ in place (/g/ and /v/) had indistinguishable slopes, perhaps because they were resistant to the same degree, but for different reasons: /g/ because it uses the tongue body and /v/ because it is a fricative. Therefore, Fowler argued that locus equations are not invariants for place of articulation; instead, the relation between place of articulation and slope is correlational.

In other studies, researchers have specifically manipulated coarticulatory overlap within a consonant category to observe changes in slope, as we did in the present study. Crowther (1994) attempted to increase coarticulatory overlap for the stop consonants by having a speaker produce them in a VCV context where $V1 = V2$ (so that the tongue position was closer to the vowel configuration when the consonant was produced), and to decrease overlap by having a speaker produce palatalized versions of

the consonants so that tongue position was more constrained. He found small increases in slope from the first manipulation, and large decreases in slope from the second. Using the Distinctive Region Model (Carré & Mraiyati, 1990), Chennoukh et al. (1995) synthesized VCVs with different degrees of coarticulation between V1 and V2 (thus varying the timing of the onset of V2 production with respect to the consonant closure). They found that when coarticulatory overlap was greatest, locus equation slopes were highest.

Our aim in Experiment 1 was to manipulate coarticulatory resistance in speakers' productions to cause increases and decreases in slope values for different consonants. We attempted to increase coarticulatory overlap (and thus slope) by encouraging speakers to be as far into the vowel gesture as possible at the release of the consonantal closure. We did this with the same technique as did Crowther (1994, with V1 = V2 in VCV productions), and we further encouraged coarticulatory overlap by having speakers produce these syllables at a fast rate. As a result, $F2_{\text{ons}}$ values should be closer to $F2_{\text{vowel}}$ values than in normal productions, and the slope should be steeper. We attempted to decrease coarticulatory overlap by instructing the speakers to precede the consonants with a central vowel (/Λ/), to attempt to maintain the central tongue configuration as much as possible up to the point of consonant release, and to speak at a slower rate (a necessity regardless, given the difficulty of the coarticulatory demands). The latter instructions were expected to result in less variability in $F2_{\text{ons}}$ values for a consonant across vowel contexts. That is, $F2_{\text{ons}}$ should vary less with changes in $F2_{\text{vowel}}$, and slopes should, therefore, be shallower.

Our intent was to determine whether we could induce large enough coarticulatory changes that the slope values for the different consonants would overlap. That is, we might find a high-resistance (low-overlap) /b/ slope equal to or lesser than a low-resistance (high-overlap) /d/ slope. Furthermore, we were interested in exploring whether regions in slope-by-intercept space would remain distinct for the different consonants under the manipulations. Although /b/ and /g/ are not consistently distinguished by slope alone, they are distinguished when both slope and intercept are considered (Sussman et al., 1991): They occupy distinct spaces in a slope-by-intercept plot. /d/ is distinguished from both /b/ and /g/ on the basis of slope alone, and thus is clearly distinguished by slope and intercept. Finally, our manipulations provided an opportunity to test whether the second-order locus equations reported by Chennoukh et al. (1995) for synthetic speech can also be found in natural productions.

Method

Subjects. One female and 2 male adult speakers of American English served as talkers in the experiment. All three were researchers at Haskins Laboratories and knew the purposes of the experiment. One (M1) was a trained phonetician, and the others (M2, F1) were the first and second authors of this paper.²

Materials and Procedure. The tokens produced by the speakers were of the form CV or VCV. The consonants used were /b/, /d/, and /g/, and the final vowels were /iʏ/, /ɪ/, /eʏ/, /æ/, /a/, /Λ/, /ɔ/, and /u/; these were the vowels used by Fowler (1994). Initial Vs were the central vowel /Λ/ in one condition and were the same as the postvocalic vowel in another condition. These were spoken at slow and fast rates, respectively. Speakers produced the tokens in three blocks of a single session. In the first block, they were presented with a list of CV syllables and were asked to produce at least five tokens of each one, spoken at a comfortable rate. In the next block, subjects were given a list of VCV disyllables with V1 = V2. They were instructed to speak these utterances at as fast a rate as possible without flapping the /d/s, and to produce five or more tokens of each one. In the final block, the subjects were given the same list of disyllables, except that the V1 was /Λ/ in all cases. They were instructed to produce at least five tokens of each of the utterances, and to speak them at a slow rate. They were explicitly instructed to attempt to maintain the central tongue configuration for the initial vowel into the closure position and thus to attempt to avoid anticipating the postconsonantal vowel until the consonant release. The utterances of M1 and M2 were recorded on audiotape in a sound-attenuating chamber; the utterances of F1 were recorded in a quiet, but not sound-attenuated, room.

The recorded tokens were filtered at 10.4 kHz and digitized at 20.8 kHz in SoundScope (GW Instruments) on a Macintosh Quadra 950 computer. Five tokens of each subject's productions of each stimulus were selected. Tokens were dropped if the consonant was flapped (so that there was no apparent burst or clear onset of voicing), or if it was spoken at a substantially lower amplitude than were the other tokens (because the stimuli were intended for use in a perceptual experiment). Otherwise, the first five tokens were selected. Acoustic measurements were made by the first author. They were taken on $F2$, following the procedure of Sussman et al. (1991), at vowel onset and vowel midpoint. The onset measure ($F2_{\text{ons}}$) was taken at the first discernible pitch pulse after the release of the consonant. The vowel midpoint measure was taken at a steady-state portion if one was present; at a midpoint of a diphthongal vowel; or at the maximum or minimum of a curve-shaped vowel pattern. Because the syllables were open, some of the diphthongs (notably /iʏ/ and /eʏ/) reached high frequencies at low amplitudes and lasted for fairly long durations. In such cases, the "midpoint" value was taken earlier than the actual temporal midpoint of the vowel, but still within the diphthongal glide. Spectral measures were taken from wide-band spectrographic displays (and narrow-band displays as well, when $F2$ and $F3$ or $F1$ and $F2$ appeared to be in close proximity) and from linear predictive coding (LPC) analysis (20 coefficients, 15-msec analysis window). Direct spectrographic measures took precedence over LPC measures when discrepancies occurred, particularly for the $F2_{\text{ons}}$ measurements.

Results

Effect of manipulations on slopes and intercepts.

The first question that we addressed was whether our attempts to manipulate coarticulatory overlap resulted in changes in slopes and intercepts for the three consonants. Based on the $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ measurements, we computed locus equations separately for the consonants /b/, /d/, and /g/ produced in the three conditions (fast with high overlap, normal, slow with low overlap; for simplicity, we will refer to these as *high*, *normal*, and *low*), spoken by the 3 subjects. The derivation of the locus equations consisted of regressing $F2_{\text{vowel}}$ on $F2_{\text{ons}}$, using the raw data points. Figures 1A–1C display the mean $F2_{\text{ons}}$ -by- $F2_{\text{vowel}}$ values of each CV spoken under the three con-

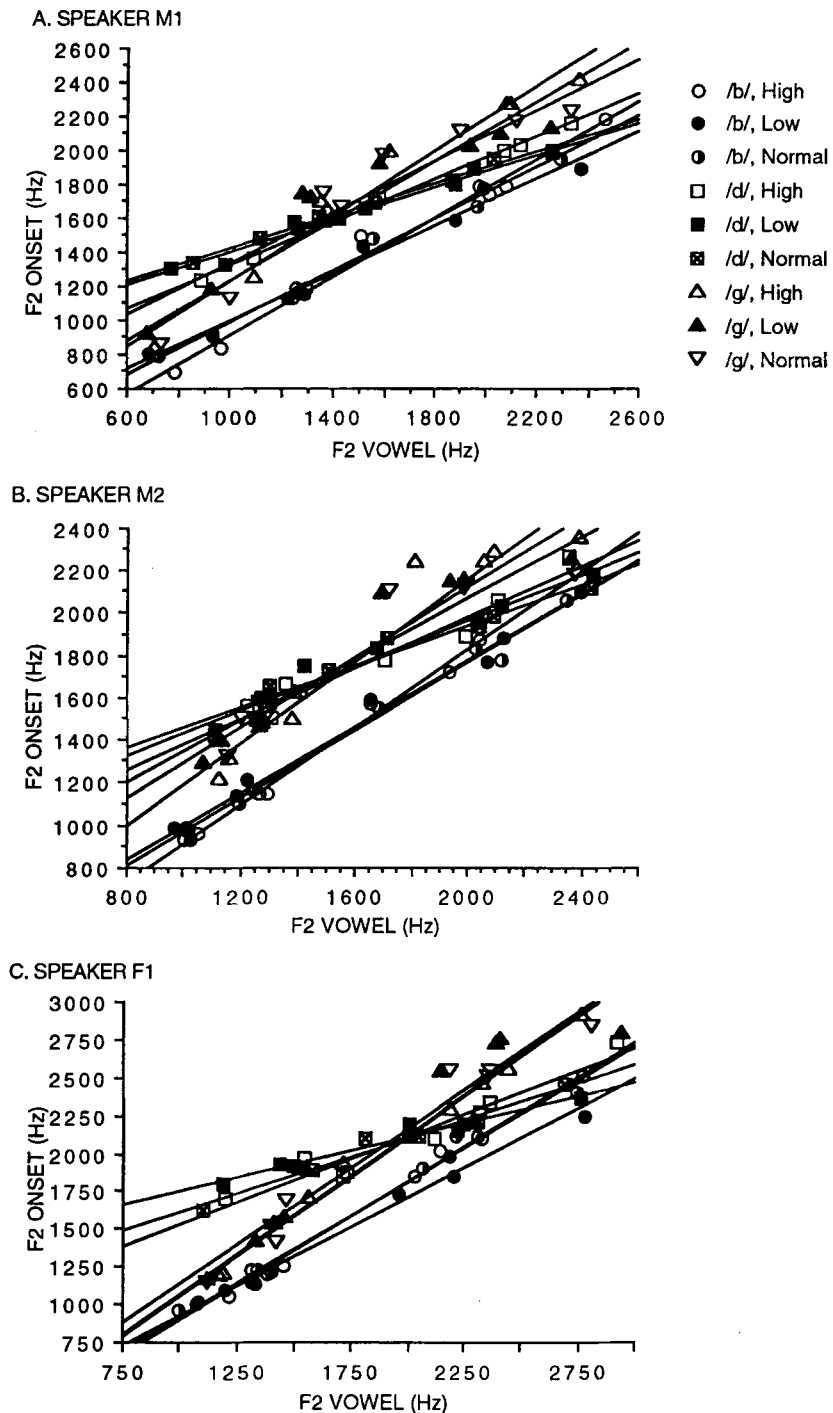


Figure 1. Individual-speaker locus equation plots for /b/, /d/, and /g/ in high-, normal-, and low-coarticulation-overlap conditions.

ditions, along with the locus equation lines for each speaker. The slopes and intercepts for the three consonants spoken under the different conditions by the different speakers (computed by token) are presented in Table 1, along with two measures of the fit of the regressions, R^2 and standard error of estimate.

Table 1 demonstrates that, overall, the high-overlap condition was effective in raising slopes, whereas the low-overlap condition was less successful at decreasing them. The high condition resulted in steeper slopes in all nine comparisons (although minimally so for speaker F1's /g/). The low condition resulted in shallower slopes

Table 1
Individual Speaker and Mean Regression Coefficients for
/b/, /d/, and /g/ Locus Equations Under
Different Overlap Conditions

Speaker	Consonant	Overlap	Slope	Intercept	R ²	SE of Estimates
M1	/b/	high	.860	53.199	.976	76.08
		normal	.762	223.330	.979	59.52
		low	.697	302.285	.978	57.17
	/d/	high	.634	695.753	.990	32.19
		normal	.474	950.039	.961	43.52
		low	.475	929.152	.955	50.68
	/g/	high	.956	277.693	.953	114.28
		normal	.875	361.862	.904	150.12
		low	.754	580.449	.886	144.20
M2	/b/	high	.920	-8.066	.985	55.23
		normal	.796	175.219	.986	48.49
		low	.786	208.005	.982	57.32
	/d/	high	.604	778.217	.932	67.15
		normal	.482	978.564	.908	68.60
		low	.532	901.782	.945	59.59
	/g/	high	.968	225.107	.917	133.55
		normal	.720	632.346	.852	132.60
		low	.828	467.844	.916	114.51
F1	/b/	high	.914	-11.910	.985	66.52
		normal	.887	37.803	.972	88.60
		low	.784	146.519	.970	81.04
	/d/	high	.587	938.561	.906	99.53
		normal	.490	1121.385	.936	63.89
		low	.360	1389.631	.893	62.30
	/g/	high	1.055	3.568	.983	78.83
		normal	1.049	26.650	.953	135.64
		low	1.024	116.665	.919	191.55
Means	/b/	high	.898	11.074	.982	65.94
		normal	.815	145.451	.979	64.87
		low	.756	218.936	.977	65.18
	/d/	high	.608	804.177	.943	66.29
		normal	.482	1016.663	.935	58.69
		low	.456	1073.522	.931	57.52
	/g/	high	.993	168.799	.951	108.89
		normal	.881	340.286	.903	139.45
		low	.869	388.319	.907	150.09

for two of the three consonants for Speaker M1, for only one consonant for M2, and for all three for F1 (although the effect was minimal for /g/). The slopes were submitted to a 3 (consonant) \times 3 (degree of overlap) repeated measures analysis of variance (ANOVA). There were significant main effects of both consonant identity [$F(2,4) = 23.65, p < .01$] and degree of overlap [$F(2,4) = 9.74, p < .05$] on slope, but no interaction [$F(4,8) < 1$]. Contrast tests on consonant (collapsing over degree of overlap) revealed that /b/ slopes were significantly greater than /d/ slopes [$F(1,4) = 25.60, p < .01$], as were /g/ slopes [$F(1,4) = 43.09, p < .01$]; however, /g/ slopes were not significantly greater than /b/ slopes [$F(1,4) = 2.26, p > .1$]. Contrast tests on degree of overlap revealed that high productions resulted in significantly steeper slopes than did both normal [$F(1,4) = 10.44, p < .05$] and low productions [$F(1,4) = 17.81, p < .05$]. Slopes for low productions were not significantly shallower than those for normal productions [$F(1,4) < 1$], despite a numerical trend in that direction (means for normal and low slopes were

.73 and .69, respectively). The slopes for /b/ and /g/ overlapped across the different conditions; compatibly, slopes for these consonants have been shown not to be reliably distinct in the literature, with some studies yielding steeper slopes for /b/ and some, steeper slopes for /g/ (e.g., Fowler, 1994; Lindblom, 1963; Nearey & Shammass, 1987; Sussman et al., 1991). However, the slopes for /d/ were distinct from all other slopes, with low /b/ falling the closest to any /d/ production (high /d/); their mean slopes were .697 and .634, respectively.

Effects of the coarticulatory overlap manipulations on intercept were closely tied to the effects on slope. As Table 1 shows, every instance of a change in slope was accompanied by a proportionate change in intercept in the opposite direction. A 3 (consonant) \times 3 (degree of overlap) repeated measures ANOVA was performed on the intercepts, revealing significant main effects of consonant [$F(2,4) = 16.07, p < .05$] and degree of overlap [$F(2,4) = 11.57, p < .05$], but no interaction [$F(4,8) < 1$]. Contrast tests revealed, for consonant, differences between /b/ and /d/ [$F(1,4) = 28.85, p < .01$] and /g/ and /d/ [$F(1,4) = 18.13, p < .05$], but no difference between /b/ and /g/ [$F(1,4) = 1.24, p > .1$]. The intercepts for /d/ were greater than all intercepts for /b/ and /g/. Despite the absence of a significant difference, intercepts for /g/ were generally greater than intercepts for /b/, with some overlap. Tests on degree of overlap revealed differences between high-overlap and normal productions [$F(1,4) = 11.87, p < .05$] and high-overlap and low-overlap productions [$F(1,4) = 21.44, p < .01$], with smaller intercepts for the high-overlap productions in both cases but no significant differences between normal and low-overlap productions [$F(1,4) = 1.41, p > .1$], although the mean intercept for the low-overlap productions was numerically greater than that for the normal productions.

To summarize, we found that our speech manipulations, as expected, did result in steeper slopes and smaller intercepts in the high-overlap condition, and we found a nonsignificant trend in the opposite direction for the low-overlap manipulation.³ The magnitudes of all of the slope and intercept changes were modest, and we did not find any overlap between /d/ slopes and intercepts and those of /b/ and /g/. Finally, the manipulations resulted in no reduction in linearity. Table 1 reveals that the R^2 s are no lower, and the standard error of estimates no higher, under the high and low manipulations than in the normal condition.

Examination of slope-by-intercept space and "second-order locus equations." Although neither slope nor intercept considered independently successfully distinguishes all three stop consonants from one another, when plotted against one another they have been shown to reveal nonoverlapping regions for each consonant (Sussman, 1994; Sussman, et al., 1991; Sussman & Shore, 1996). We examined whether our manipulations also resulted in distinct slope-by-intercept regions for the three consonants. Furthermore, we examined whether there were good-fitting second-order locus equations (Chen-

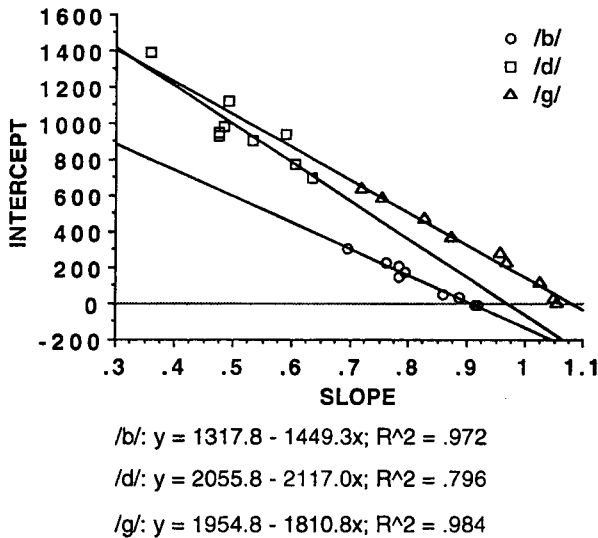


Figure 2. Plot of slope-by-intercept space for consonants under different overlap conditions spoken by all speakers. Regression lines are second-order locus equations.

noukh et al., 1995) for the three points (corresponding to the different coarticulation manipulations) in slope-by-intercept space for each consonant.

The locus equation coefficients of the 3 speakers' productions are presented in slope-by-intercept space in Figure 2. As the figure demonstrates, there is considerable variability in slopes and intercepts for each consonant; however, the consonants do fall within distinct regions. Thus, although slopes and intercepts considered individually did not reliably distinguish /b/ and /g/, together they do separate the consonants. When /b/ and /g/ have similar slopes, /g/ has a greater intercept; when they have similar intercepts, /g/ has a steeper slope.

The productions of each consonant appear to fall roughly on a line, consistent with the findings of Chenoukh et al. (1995). We computed SOLEs for each consonant by regressing the three intercept values for each consonant with the corresponding slope values from the different manipulations. The SOLEs and their coefficients are presented in Figure 2. The fits for /b/ and /g/ are quite good (R^2 for /b/ is .972; for /g/, .984); the fit for /d/ is more modest ($R^2 = .796$). When the regressions are performed separately for each speaker, the fits are improved (see Figures 3A–3C). Furthermore, interesting patterns emerge in these plots. The plots for Speakers M1 and M2 are nearly identical: The three lines are nearly parallel (more so for M1 than for M2), with the intercept for /g/ slightly greater than that for /d/, and the intercept for /b/ considerably smaller. This parallelism of second-order lines is not found for Speaker F1, however, but the regression lines still provide good fits to her productions.

There is a direct interpretation of the slopes and intercepts of these second-order equations. If the second-order locus equation lines each are formed by a pair of coordinates (slopes and intercepts of two first-order locus

equations), the slope of the new line is the $F2_{\text{vowel}}$ (multiplied by -1) of the point at which the two first-order locus equation lines intersect and the intercept is the $F2_{\text{ons}}$ of the same point.⁴ Therefore, the extent to which three or more locus equation lines fall directly on a line in slope-by-intercept space reflects the degree to which they all rotate about a single point; this may be observed in Figures 1A–1C.

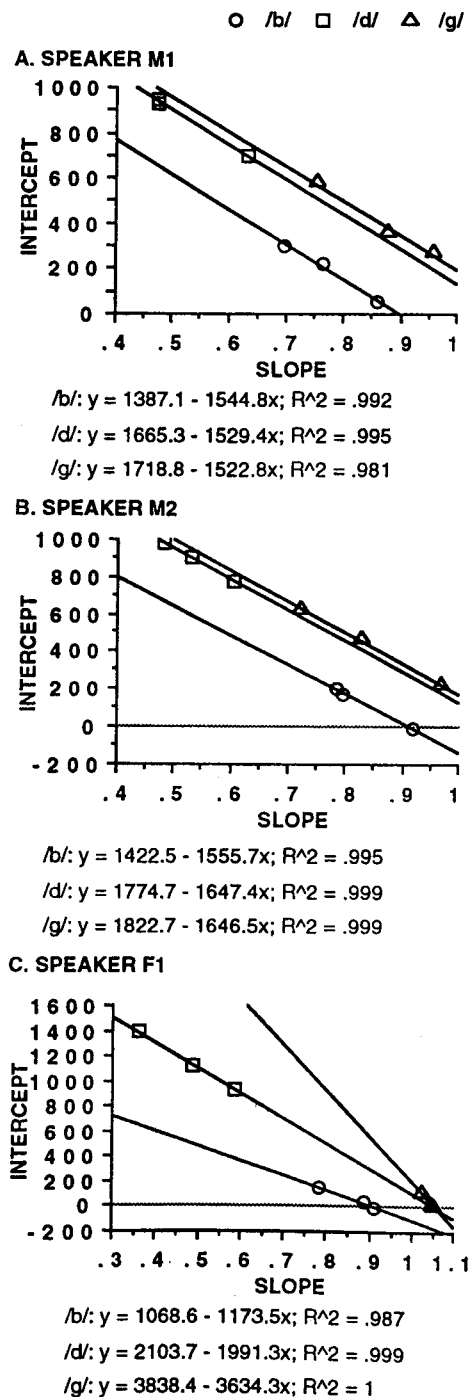


Figure 3. Individual-speaker second-order locus equation lines.

For Speakers M1 and M2, the $F2_{\text{vowel}}$ value at the point of rotation was essentially the same for all three consonants (hence the parallel second-order locus equation lines); this point of rotation was in the range of a central vowel (generally between the $F2_{\text{vowel}}$ values for /ʌ/ and /æ/). In contrast, Speaker F1's /b/ lines intersect at a back-vowel position. Her /g/ lines appear to intersect close to 3600 Hz; however, the slopes and intercepts of the three /g/ lines hardly vary from one another, so the rate and V1-context manipulations were not effective for them.

Locus equations for /g/ in front and back vowel contexts. Table 1 shows higher standard errors of estimate for /g/ than for /b/ and /d/, in every comparison but one. Indeed, /g/ is generally better fit by two regression lines than by one line. The better fit of separate lines for /g/ with front and with back vowels has been attributed to allophonic variation in /g/ productions, such that the tongue contact positions tend to be more palatal followed by front vowels and more velar followed by back vowels (Sussman et al., 1991). Thus, Sussman et al. (1991) suggest that it may be more appropriate to use the separate locus equations for the allophones (front, back) of /g/. Accordingly, we computed separate front-/g/ and back-/g/ locus equations for the present data set. The coefficients and measures of fit for each subject are presented in Table 2. Although there is a general trend toward steeper slopes and smaller intercepts in the high-overlap conditions and a trend in the opposite direction in the low-overlap conditions, these effects are not consistent across speakers and allophones (front, back) of /g/. Two 2 (al-

lophone) \times 3 (degree of overlap) repeated measures ANOVAs performed on the slopes and the intercepts revealed significant main effects of allophone on slopes [$F(1,2) = 54.87, p < .05$] and intercepts [$F(1,2) = 100.80, p < .01$], with shallower slopes and greater intercepts for the front-vowel allophones. However, effects of degree of overlap on slopes and intercepts and interactions between allophone and degree of overlap were nonsignificant.

In the previous section, we found that /b/, /d/, and /g/ do occupy distinct regions in slope-by-intercept space, and that different production styles of each consonant fall on lines in this space. Here we examine whether distinctive regions among different production styles are found when the /g/ allophones are considered separately. Figure 4 presents the slopes and intercepts of /b/, /d/, and the two allophones of /g/ spoken under the different conditions by all 3 speakers. The figure shows that the regions for the different categories overlap somewhat; for example, some front-/g/ points fall among /d/ points. However, most of this overlap is cross-speaker. In Figures 5A–5C, the slopes and intercepts for the 3 speakers are plotted separately, showing that there is no overlap of slope-by-intercept regions within the productions of a given speaker. Although there is no overlap among regions, a consonant produced under one manipulation may fall closer to productions of another consonantal category than to other members of its own category. For example, Speaker M1's fast front-/g/ and Speaker F1's normal front-/g/ fall closer to /d/ productions than to other front-/g/s. Thus, when /g/ is divided into two separate categories, nonoverlapping regions may be drawn around the consonants' coordinates in slope-by-intercept space, but these regions will be more diffuse and less distinctive than when /g/ is treated as a single class.

Table 2
Individual Speaker and Mean Regression Coefficients for /g/ Locus Equations Computed Separately for Front and Back Vowels, Under Different Overlap Conditions

Speaker	Vowel	Overlap	Slope	Intercept	R^2	SE of Estimates
M1	back	high	1.271	-84.118	.955	78.345
		normal	1.285	-109.732	.948	90.216
		low	1.332	-11.363	.971	63.883
	front	high	.572	1065.913	.946	38.608
		normal	.349	1417.392	.903	33.059
		low	.326	1404.582	.85	35.525
M2	back	high	1.224	-135.407	.891	56.327
		normal	1.353	-185.617	.473	86.982
		low	.842	409.843	.825	36.820
	front	high	.195	1875.182	.735	26.052
		normal	.107	1909.562	.392	33.194
		low	.245	1672.46	.894	21.488
F1	back	high	1.366	-417.209	.975	54.346
		normal	1.163	-150.963	.689	120.567
		low	1.106	-49.109	.834	72.347
	front	high	1.054	-10.738	.942	59.281
		normal	.535	1296.789	.714	83.019
		low	.255	2071.598	.413	93.656
Means	back	high	1.287	-212.245	.940	63.01
		normal	1.267	-148.771	.703	99.26
		low	1.093	116.457	.876	57.68
	front	high	.607	976.786	.874	41.31
		normal	.330	1541.248	.670	49.76
		low	.275	1716.213	.719	50.22

Discussion

In Experiment 1, we attempted to manipulate degree of coarticulatory overlap for each consonant, to increase and decrease the locus equation slopes and intercepts, with the intention of dissociating slope and intercept values from place of articulation distinctions. Our manipulations did cause the slopes and intercepts for each consonant to shift in the expected directions, supporting the view that slope is related to degree of coarticulatory resistance. However, the magnitudes of these shifts were modest. (The high manipulation caused somewhat smaller shifts in slope and intercept from the normal condition than those found by Crowther, 1994, who used a very similar manipulation.) /d/ remained distinct from /b/ and /g/ on the basis of slope alone, and all three consonants fell in distinct regions in slope-by-intercept space (although less clearly so when /g/ was divided into front- and back-context allophones). These findings and those of others (Bakran & Mildner, 1995; Crowther, 1994; Duez, 1992; Krull, 1989) show that although slope and intercept may well be *distinctive* with respect to place of articulation, because they are affected by factors independent of place, they are not *invariant*. Nonetheless, our

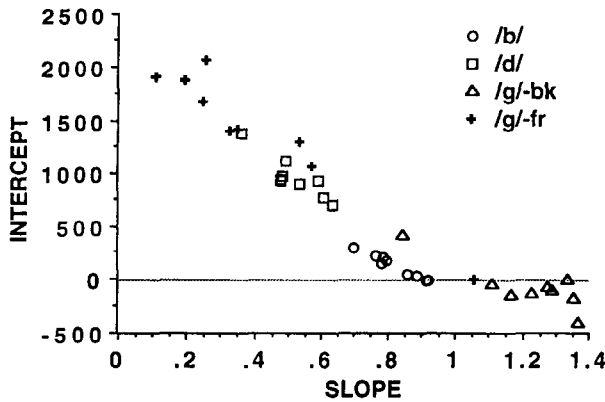


Figure 4. Plot of slope-by-intercept space for all speakers, with /g/ separated into front- and back-vowel allophones.

findings are consistent with the claims of Sussman and colleagues (e.g., Sussman et al., 1991; Sussman & Shore, 1996) that different consonants can be distinguished in slope-by-intercept space.

Our data raise the possibility that this distinctiveness in slope-by-intercept space is due to a higher order systematicity in which locus equations are embedded—namely, that captured by second-order locus equations. We found good linear fits of the slopes and intercepts from the different manipulations, as had Chennoukh et al. (1995) using synthetic speech. We additionally found interesting patterns in the SOLEs of 2 of our subjects: Their SOLEs were parallel to one another, indicating that the rotation points for their locus equation lines from the different manipulations (one rotation point per consonant) fell at approximately the same $F2_{\text{vowel}}$ positions. This pattern was not found by Chennoukh et al., however, and it may have been an artifact of our particular manipulations.

However, if the SOLE patterns are ultimately found to be robust phenomena, they will have implications for the OOC. As outlined in the introduction, the OOC claims that speakers produce CV syllables such that $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values are linearly arranged for each consonant, in order to facilitate the development of representational maps exploiting those acoustic parameters. Because these representations are for consonant classes, presumably independent of production variations (such as rate, stress, intervocalic context, etc.), it should follow that the linearity predicted by the OOC should be found even when all such variations in production are collapsed into a single consonant group. (In fact, our findings do not falsify such a prediction; because the slope and intercept changes due to our manipulations were modest in size, there is still strong linearity in locus equation fits computed across tokens from the different manipulations.) However, if the OOC-based linearity is indifferent to different production styles, then we believe there are no grounds for expecting strong linearity *within* a production style (as we found, although interpretation of our results in this regard is tempered by the limited magnitude

of the variability created by the manipulations). More significantly, we similarly see no basis in the theory for the finding that variations in speaking rate result in good-fitting SOLEs, indicating a systematic relationship between changes in slope and intercept due to these variations, again because the OOC does not make predic-

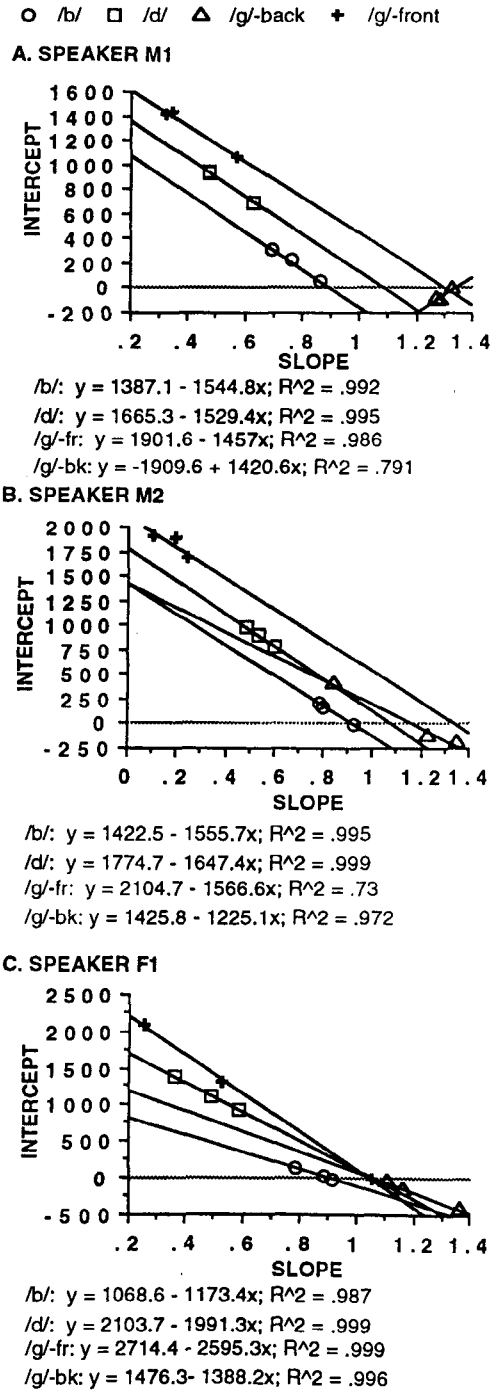


Figure 5. Individual-speaker second-order locus equation lines, with /g/ separated into front- and back-vowel allophones.

tion about linearity more specific than at the consonant-class level.

Sussman's basis for the OOC hypothesis stemmed from an appreciation of the strong and robust linearity between $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ (e.g., Sussman et al., 1995). He reasoned that, given the existence of specialized auditory systems which encode linearly related acoustic parameters, and given the established relevance of $F2$ transitions for consonant perception, it is therefore possible that the linearity captured by locus equations is a functional adaptation to support an auditory system that takes $F2$ transitions as inputs. The strength of this argument, however, depends on the uniqueness of the linearity of $F2$ locus equations; that is, plausibility of the OOC is weakened by the existence of other linear relationships that do not plausibly have a role in perception.

SOLEs may be considered one example of linearity that cannot serve a perceptual function and thus cannot reflect a perception-based production constraint. The linearity captured by SOLEs appears to be strong (in the limited data to date); yet it is not plausibly due to a production constraint serving a perceptual system that maps locus equation slopes against locus equation intercepts. Nor do first-order locus equations and SOLEs exhaust the linear relations between pairs of speech variables. A third example is found in first-formant locus equation plots. We computed locus equations for $F1$ transitions, using mean values from the appendix of Kewley-Port (1982). The regression coefficients, R^2 , and standard error of estimate values for $F1$ and $F2$ locus equations thus obtained are presented in Table 3. The three $F1$ locus equation lines have virtually the same slope, and very similar intercepts, so it is clear that these parameters are not distinctive for place of articulation. However, the linearity is strong; in fact, the standard error values for the $F1$ locus equations are actually *lower* than those for the $F2$ locus equations.⁵ Klatt (1987) also found good linear fits for $F1$ regressions. Thus, there are at least two instances of linearity without plausible perceptual relevance.

Thus far, we have argued that our results are consistent with the view that slopes and intercepts of locus equations, taken in tandem, are distinctive for place of articulation. However, we have challenged the plausibility of the OOC's explanation for the existence of locus equations—namely, that they are the result of a production constraint driven by the demands of the speech perception system. If the linearity is not a perception-driven

constraint, though, then what is its source? An alternative claim is simply that the linearity itself reflects a characteristic of coarticulation. An assumption of our work is that the slope of a locus equation line indexes the degree of *coarticulation resistance*, the extent to which a consonant resists coarticulatory overlap by a vowel. As shown by Recasens (e.g., 1985, 1991) and Farnetani (1990), resistance is greatest when the consonant and vowel compete for the same articulators, most importantly the tongue body.⁶ That the relation between $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ is close to linear means that a given change in $F2_{\text{vowel}}$ is associated with a magnitude of change in $F2_{\text{ons}}$ that is nearly the same everywhere along the function. From the perspective of our hypothesis, this implies that coarticulation resistance is generally the same between a consonant and the set of all vowels; it does not vary much across the set of vowels. This is not surprising, because vowels all use the tongue body, and so coarticulatory overlap by all of them should be resisted to approximately the same degree by a given consonant.

The coarticulation resistance account of the linearity of locus functions is considerably more economical than that of the OOC. First, it ascribes the linearity of locus functions to a theoretical construct that is already required to explain other data (e.g., Bladon & Al-Bamerni, 1976; Recasens, 1985). Second, it ascribes the linearity to behavior that is temporally local. That is, the degree of coarticulatory overlap of a consonant by a vowel is determined by a characteristic of a consonant—namely, the degree to which the vowel would compete with its achievement of its own gestural goals. In contrast, according to the OOC, speakers must maintain linearity by having their productions, distributed over time, conform to one another in $F2_{\text{ons}}$ -by- $F2_{\text{vowel}}$ space in order to meet the demands of an external constraint (the ostensible design of listeners' perceptual systems).

The coarticulation resistance account has a further advantage over that of the OOC. It can predict which consonants will be associated with steeper slopes and which with shallower ones. The OOC does not. Generally, consonants that use the tongue should be more resistant to coarticulatory overlap by vowels than those that do not. In addition, because the articulatory requirements for producing frication noise are more precise than those for producing stops, fricatives should have shallower slopes than those for homorganic stops. Both predictions were borne out in findings of Fowler (1994).

Sussman et al. (1995) provided two challenges to the coarticulation resistance account of linearity. The first is based on evidence that different vowels do coarticulate differentially with consonants. For example, data by Amerman (1970) show more anticipatory coarticulation by /a/, /ɪ/, and /æ/ than by /iʏ/ or /u/ in preceding consonants /p/, /b/, /t/, /d/, /s/, and /z/. The finding that coarticulation resistance is, to a degree, vowel specific would challenge the coarticulation resistance account of linearity if measured $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values fit their regression lines perfectly. However, they do not, and when re-

Table 3
Regression Coefficients for $F1$ and $F2$ Locus Equations,
Taking Mean Measurements Reported in Kewley-Port (1982)

Formant	Consonant	Slope	Intercept	R^2	SE of Estimates
$F1$	/b/	.272	241.7	.762	20.56
	/d/	.232	235.7	.943	8.56
	/g/	.243	221.9	.736	21.37
$F2$	/b/	.733	327.6	.983	44.39
	/d/	.299	1284.2	.959	23.17
	/g/	.803	567.8	.852	161.06

gressions are performed on subsets of the vowels, the slopes of the lines can change. A prediction based on Amerman's data is that slopes for the consonants /p/, /b/, /t/, /d/, /s/, and /z/ should be steeper computed over the vowels /a/, /ɪ/, and /æ/ than over vowels /iʏ/ and /u/. This prediction can be tested on the voiced consonants by using $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ measurements from Fowler (1994). The slopes for /b/, /d/, and /z/ (the subset of Amerman's consonants measured by Fowler) in the context of /iʏ/ and /u/ were .766, .578, and .540, respectively. In the context of /a/, /ɪ/, and /æ/, they were .863, .694, and .653. Likewise, in the data presented in Experiment 1, the slopes for /b/ and /d/ were shallower when computed for /iʏ/ and /u/ than for /a/, /ɪ/, and /æ/ for the normal and low conditions, although not for the high condition. In the normal condition, the slopes for /b/ and /d/ in the /iʏ/ and /u/ contexts were .795 and .529; in the /a/, /ɪ/, and /æ/ contexts, they were .922 and .545. In the low condition, the corresponding values were .715, .502, .820, and .530; however, in the high condition, they were .909, .689, .892, and .618. (However, Amerman's observations on the differing degrees of anticipatory coarticulation for the different vowels do not necessarily apply to the high-overlap speaking style. Our manipulation specifically encouraged anticipatory vowel coarticulation, so the manipulation may have preempted intrinsic vowel differences on this dimension.)

Sussman et al.'s (1995) second objection to the coarticulation resistance account derives from the nonlinearities in mapping from articulatory activity to the acoustic signal. Coarticulation resistance is a relation between articulatory gestures, but the linearity that locus equations describe is a relation between sets of acoustic variable pairs. Sussman et al. (1995) argue that the nonlinearities in the mappings between gestures and acoustic signals might cause larger departures from acoustic signal linearity than fits to locus equations exhibit; thus, a linear output in the acoustics requires a nonlinear relationship of coarticulatory overlap and vowel context in the articulatory domain. However, the locus equation regressions based on Amerman's data lend some support to the notion of a strong relationship between coarticulation resistance and acoustic linearity; deviations from linearity appear to be tied to deviations in degree of coarticulation resistance. Additionally, the Distinctive Region Model (Carré & Mrayati, 1990) produces linear locus equation functions when the experimenters fix the degree of vowel-to-consonant overlap (R. Carré, personal communication, July 1, 1996), further suggesting a link between acoustic linearity and uniform coarticulation resistance. Although we cannot yet assert that the construct of coarticulation resistance has sufficient explanatory power with respect to locus equations, the account is certainly falsifiable, by evidence demonstrating that uniform coarticulation resistance creates nonlinear acoustic relationships, and that nonuniform coarticulation resistance is required for acoustic linearity. However, such evidence has yet to be pre-

sented. Until the coarticulation resistance account is falsified, postulation of an OOC is extraneous.

A shortcoming of our coarticulation resistance account is that it does not provide an explanation for the finding of second-order linear fits (although we do not believe the hypothesis is inconsistent with SOLEs). We propose only that linearity is linked to uniform coarticulation resistance. Thus, we can explain why strong linearity is preserved separately for productions under each manipulation with the assumption that manipulations of coarticulation resistance have a uniform effect on the degree of coarticulatory overlap for each vowel. Our account does not explain why the separate lines for different speaking styles for a given consonant should all cross at a single "rotation point," resulting in good-fitting second-order locus equations. This is because the account makes predictions about locus equation slopes, but not about intercepts.

A last point to be raised from our results has implications for the perceptual study described in Experiment 2. It concerns whether /g/ is described better with one locus equation line or two. It has been argued from a (first-order) locus equation standpoint that two lines; one for front vowels and one for back vowels, more appropriately fit the $F2$ values for /g/ (Sussman et al., 1991; but see Fowler, 1994). However, examination of slope and intercept together suggests otherwise. First, the second-order regressions reveal strong regularity in /g/ productions when one line is computed for /g/ over all vowels, continuous with findings for /b/ and /d/ (specifically, the lines are parallel). Furthermore, in plots of slope against intercept, when a single line for /g/ is computed, /b/, /d/, and /g/ fall in clearly delineated regions. In contrast, when separate lines for /g/ in front and back vowel contexts are computed, the regions (while still nonoverlapping) are less clearly distinct, such that a back-vowel /g/ point may fall closer to /d/ points than to other back-vowel /g/ points. These observations both suggest that a single locus equation for /g/, in spite of its poor apparent fit to the $F2$ values, captures some regularity in production and should not be rejected in favor of two separate lines. However, a perceptual map of $F2$ space using a single-line fit for /g/ will do a poorer job of accurately representing /g/ tokens than would two /g/ lines; this apparent tradeoff was addressed in Experiment 2.

EXPERIMENT 2

So far, we have addressed the issues of whether locus equations describe relational invariants for place of articulation, and of the basis for the linear relation between the endpoints of $F2$ transitions. We found that locus equation slopes and intercepts vary as a function of degree of coarticulatory overlap, but that they remain distinctive for place of articulation under this variability. We argued that $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ coordinates have a strong linear relation because of coarticulation resistance and not because perceivers require it to be so. In this next section,

we examine the possible relevance of locus equations for perceivers in identifying stop consonants.

To review, Sussman has claimed that, because of an auditory-system preference for linearly related input parameters (documented in the auditory systems of the bat and the barn owl), human perceivers use the linear relationship between $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ to develop a representational system for speech perception. He has proposed that consonant perception uses “an auditory map similar to locus equation plots” (Sussman et al., in press). Furthermore, he has suggested specific neural mechanisms by which such a representational system might operate: combination-sensitive neurons respond to pairs of $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ frequencies that correspond to a particular consonant (Sussman, 1989; Sussman et al., 1991). In addition, neurons that fire for a given consonant (and thus respond to $F2$ values that fall close to the locus equation line) are arranged in columns in the neuronal structure. Sussman et al. (1991) argue that “in order for an entire stop consonantal equivalence class to be functionally tied together an acoustic/phonetically based constancy must exist. . . . For place of articulation, slope and y -intercept of the regression function can serve [as the emergent property that binds the members of a class]” (p. 1323). Thus, the fact that the input parameters to the neural system, $F2_{\text{ons}}$ and $F2_{\text{vowel}}$, are linearly related within a stop consonant class is a significant feature of the theory.

A number of predictions about consonant perception can be derived from a model of the sort that Sussman has proposed. First, a locus-equation-based model of perception makes the uncontroversial claim that $F2$ transitions have strong perceptual relevance. This is well grounded in the literature, dating back to the demonstrations of Liberman, Delattre, Cooper, and Gerstman (1954) that burstless two-formant stimuli can support stop place identification fairly well (although Liberman et al., 1954, and more recently, Ohde, 1988, have pointed out limitations of the informativeness of synthesized two-formant stimuli, particularly for velars).

In addition, a locus-equation-based model makes two other, more distinguishing assumptions. One is that the relevant information for consonant identification in $F2$ transitions can be sufficiently captured in two values, $F2$ at the onset of voicing and $F2$ at vowel midpoint (or transition offset; both have been used in locus equation measurement with similar outcomes; Nearey & Shammass, 1987), that the human perceptual system extracts in order to identify the consonant. In other words, the trajectory of an $F2$ transition is irrelevant as long as it has a particular onset and offset (Sussman et al., 1995).

The other distinguishing, and most crucial, assumption of a locus-equation-based theory of perception is that the linearity of CV tokens of a given consonant in $F2$ space is perceptually relevant. It is relevant because the neural representations are presumed to develop by capitalizing on the linear relationship, as outlined in the OOC, and further because the columnar-arrangement neural structure discussed by Sussman (1989) and Sussman et al.

(1991) uses the regression slopes and intercepts to bind neurons signaling members of a single consonant class. Sussman et al. (1991) suggest that the linear locus equation functions may be considered as “representing ‘CV prototypes’ conceptually similar to vowel category ‘prototypes’ (Grieser & Kuhl, 1989) in infant categorization” (p. 1321), and the authors illustrate these hypothesized prototypes as lines in $F2_{\text{ons}}$ -by- $F2_{\text{vowel}}$ space in their Figure 7. Fruchter and Sussman (1997; see also Fruchter, 1994, and Sussman et al., in press) claimed to support the perceptual relevance of locus equations with a study in which synthetic syllables with systematically manipulated $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values were presented to listeners for classification. They found that the regions in $F2$ space that elicited the greatest proportion of /b/, /d/, and /g/ responses tended to overlap with the regions where productions of the respective consonants fell (which were linear). However, these results do not provide strong evidence in favor of the view that linearity is perceptually relevant. Only three levels of $F3$ transition were used for each vowel, each corresponding to /b/, /d/, and /g/ transitions, respectively; the authors focused on responses to tokens with the “appropriate” $F3$ transitions for that response (i.e., “b” responses to tokens with /b/-like $F3$ transitions). Even under these favorable conditions, the correspondence of response patterns to locus equation patterns in productions was modest: Although there was overlap between the perception and production patterns, there was also considerable distinctiveness between the two. Indeed, Eek and Meister (1995) conducted a similar study, in which they manipulated $F2$ transitions in synthesized /pV/, /tV/, and /kV/ syllables, and also found partial overlap between production-based locus equation space and listener response patterns, but they interpreted their results as evidence against a role of locus equations in perception. However, neither of these studies provides a quantitative test of whether the linearity of locus equations has perceptual relevance.

A formal model that we believe would be consistent with the locus equation view would use locus equation lines to stand for the neural representations described in the theory. Each novel token would be matched against the lines in $F2$ space; whichever line the token falls on or closest to would “win.” Such a prototype model (see Sussman et al., 1991, Figure 7) appears to capture the intent of the research questions raised by Sussman et al. (1991). They asked: “Are tokens lying closest to the regression line better exemplars of the stop place category than tokens further off the line? Similarly, how far can $F2$ onset frequencies deviate around the locus equation function with tokens perceptually retaining their stop place affiliation?” (p. 1322). In our study, we quantified the locus equation model by computing the Euclidean distance in $F2$ space between tokens and locus equation lines for /b/, /d/, and /g/. According to the model, to the extent that a token is closer to one line than the others, it should be identified more often as the consonant corresponding to the line.

We will contrast this model with an alternative one in which only the first distinguishing assumption of the locus equation view (that $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ are sufficient to describe $F2$ transitions), and not the second (that locus equations are perceptually relevant) is made. We introduce the model solely to contrast it with the locus equation model. (That is, we do not propose it as a realistic model of speech perception.) Comparison of the two models isolates the assumption of perceptual relevance of linearity. In this second *exemplar* or *token-based* model, all of a listener's experience with stop consonants is stored, with each token syllable encoded in memory by its consonant identity and its $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values. When a new syllable is presented to a listener, the token is matched in parallel to each of the stored traces of consonants in memory. The outcome is a composite for each consonant type, with the contribution of each stored consonant (*memory trace*) to the composite weighted by its proximity in $F2$ space to the input token. (This approach is based loosely on the exemplar memory model of Hintzman, 1986, 1988.) If more memory traces coded as a particular consonant are closer to the new token than traces coded as other consonants, the token is identified as that consonant.

In the exemplar model, the information coded and used for identification is exactly that used by the locus equation model—namely, the onset and offset of $F2$ transitions. In this model, however, in contrast to the model based on locus equations, the linearity of the tokens in $F2_{\text{ons}}$ -by- $F2_{\text{vowel}}$ space does not factor into consonant identification. Rather, variability in production (i.e., deviations from linearity) plays a direct role for this model. If a token is presented that falls some distance from the locus equation line for its consonant, it may be correctly identified to the extent that the listener has encountered numerous other tokens that fall off the line more or less to the same extent. In fact, the model may correctly identify a token that is closer to the locus equation line for a different consonant than it is to its own line, if enough tokens of the same consonant have been presented that fall in the same region in $F2$ space. Thus, variability in production is exploited by the exemplar model. In contrast, our implementation of the locus equation model does not benefit from this variability. (The linear-based representation, even if it has a critical bandwidth about it, is still linear.)

We tested these models in Experiment 2 by comparing their classifications of CV tokens (based on their $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values) with the classifications of the same tokens by listeners. We addressed two questions. First, we asked how effective a model can be if it takes as its inputs $F2_{\text{ons}}$ and $F2_{\text{vowel}}$.⁷ To the extent that both models (linear prototype and exemplar) are effective or ineffective in predicting subjects' responses, we might find evidence concerning the appropriateness of the $F2_{\text{ons}}$ - $F2_{\text{vowel}}$ metric. Second, we asked whether there is any predictive advantage to the locus equation account's use of linear representations.

Method

Subjects. Thirty-two students from an introductory psychology class at the University of Connecticut served as subjects in the experiment. They received course credit for their participation. All reported normal hearing and were native speakers of English.

Stimulus materials. The speech stimuli presented to the subjects were the normal and high-overlap tokens of /b/, /d/, and /g/ in the contexts of /iʏ/, /ɪ/, /eʏ/, /æ/, /a/, /ʌ/, /ɔ/, and /u/ produced by speaker M2.⁸ The five tokens of each CV analyzed in Experiment 1 were used, resulting in a total of 240 stimuli. The high-overlap tokens, having been produced as VCV syllables, were excised from their V1 context in SoundScope, resulting in CV syllables.

Pilot work for this experiment (Brancazio & Mitra, 1994) revealed that subjects were close to ceiling in their identification of these syllables, with overall accuracy of 95.7%. This provided few errors to compare with the models' predictions. In the present study, we removed the bursts from the tokens. This served two functions. It lowered performance, and it removed a source of information, the burst, that is extraneous from the perspective of both models we are testing. Because we used natural stimuli, it was impossible to isolate another extraneous source of information, $F3$, without resorting to variable filters or to resynthesis of the natural speech, omitting $F3$.

The bursts were removed in SoundScope. In cases in which the burst was temporally independent of the onset of voicing (as visually determined in the acoustic waveform), a cut was made at a zero-crossing just before the onset of voicing. However, when the burst itself was voiced, the cut was made at a zero-crossing in an early pitch period so that as much of the burst was removed as possible while preserving the onset frequency of the transition.

We found that there were considerable differences in amplitude among the tokens. Accordingly, we amplified some of the tokens digitally; for fear of distortion, however, tokens were not amplified more than 200%. Even with this modification, some differences in amplitude remained.

After burst removal and amplification, the stimuli were output to audiotape. Four random sequence files were created, with each stimulus included once in each sequence. The stimuli were recorded on tape in these sequences.

Procedure. Subjects, run in groups of one to four, heard all 240 tokens over headphones in one of the four sequences, preceded by 5 practice tokens (selected from the 240). Eight subjects heard one sequence, 10 heard a second sequence, and 7 heard each of the other two sequences. The subjects were given response sheets numbered 1 to 245 with the selections "B," "D," and "G" and the numbers 1–5 next to each one. They were instructed to identify the consonant in each syllable. We explained to them that the syllables had been "altered" on a computer, so that it would be hard to hear the consonant in some cases, but that they should attempt to identify each one nonetheless, guessing if necessary. They were instructed to use the number scale to indicate their confidence in their response, with 5 for high confidence and 1 for guessing; however, analyses of the confidence judgments are not reported in the results.

Derivation of predicted responses. Predicted classifications of the speech tokens by the models were computed using Euclidean distance metrics. A number of studies have taken this approach to use acoustic parameters to predict listener confusions on stop consonants, as summarized by Krull (1990). For example, Krull (1990) found considerable success in predicting listener confusions of Swedish voiced stops by using $F2$, $F3$, and $F4$ at voicing onset, supplemented by burst duration, as acoustic parameters. However, to our knowledge, no study has applied the acoustic parameters to be used here—namely, $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ —in computing Euclidean distances.

For each of the models, distances from each consonant were computed for each token. The database for these computations consisted of the $F2$ values of CV productions by the five male speakers presented in Fowler (1994), and of Speaker M1's normal-overlap

CV tokens. This database was intended as a rough approximation to the subjects' perceptual history. All six speakers were from the northeastern United States, as were most of the listeners.

For the locus equation model, all the tokens in the database (720 tokens in all) were used to compute locus equations for /b/, /d/, and /g/. The Euclidean distances between a token and the nearest point on each consonant's line were computed by finding a perpendicular to each line that connected to the token's coordinates. A second locus equation model was formed as well, in which locus equation lines were computed separately for /g/ in the context of front and back vowels.

For the exemplar model, distances were computed between the token and each token in the database. These distances were transformed, so that small distances would have a proportionately greater effect than longer distances. This was done in two ways. First, any distance greater than 999 Hz was recoded as 999, and then all distances were converted by the equation

$$d_{\text{conv}} = 1 - \frac{d_{\text{orig}}}{1000}, \quad (1)$$

where d_{orig} is the Euclidean distance (in hertz), if less than 999; if the distance is greater than or equal to 999, then $d_{\text{orig}} = 999$.

The d_{conv} between a token and all of the stored tokens in the database were summed by consonant, by the equation

$$D_{i,c} = \sum_j (d_{\text{conv},i,j,c})^s, \quad (2)$$

where i is a test token, c is a consonant category (/b/, /d/, or /g/), j_c is a database token of consonant c , s is a squashing parameter, and $D_{i,c}$ is the summed distance score for each test token and the set of stored tokens of each consonant category.

Thus, distances were converted in Equation 1 to proportional contributions to a composite distance score, whereby a zero Euclidean distance became a contribution of 1, and a distance of 999 Hz or greater (the largest possible distance, from a /gi/ token to a /bu/ token, would be approximately 2500 Hz) became a near-zero contribution. The purpose of dividing by 1,000, rather than 2,500, was to help augment the difference in contributions between very small and relatively small distances. This goal was further achieved by raising the percent distance score to an exponent (the squashing parameter s); both 5 and 10 were used. (The proportional distances in Equation 1 were subtracted from 1.0 so that the squashing parameter's effect would be greatest for small, rather than large, distances.)

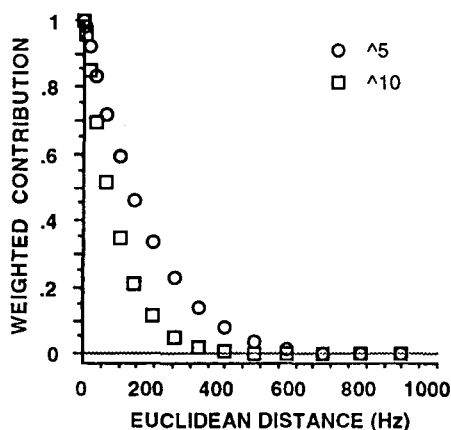


Figure 6. Plot of weighted contributions of stored token to an exemplar distance score, by Euclidean distance of stored token to test token.

This method is derived from that used by Hintzman (1986) in his exemplar memory model MINERVA.⁹ In Figure 6, the transformed proportional distance scores are plotted against Euclidean distance. Equation 2 resulted in three separate composite distance scores (composite matches to /b/, /d/, and /g/ tokens) for each test token. Because of the subtraction in Equation 1, a high composite score resulting from Equation 2 corresponds to a good degree of match for a test token. Therefore, each of the three distances for each token was scaled so that zero represented the largest $D_{i,c}$ score of all the test tokens across the three consonants, and 1.0 represented the smallest possible $D_{i,c}$ score.

In all four models (locus equation, locus equation with two /g/ lines, exemplar with exponent 5, and exemplar with exponent 10), each token is assigned three distance scores, one for each consonant. In the locus equation models, these are actual Euclidean distances, whereas in the exemplar models, it is a derived score based on Euclidean distances. In all of the models, however, the smallest score represents the best match. The predicted response (/b/, /d/, or /g/) by each model for each token is the consonant with the best match.

Results

Our main goal was to measure how well the models predicted subjects' response patterns. Accordingly, the majority of the analyses we will discuss were designed to address this question. First, however, we will briefly present data reflecting subjects' and models' response patterns independent of one another.

Subject accuracy. Overall, subjects correctly identified 66.2% of the tokens. Their mean percentage responses of /b/, /d/, and /g/ for each CV by rate are presented in Figure 7. Percent correct identification was submitted to 2 (coarticulatory overlap condition) \times 3 (consonant) \times 8 (vowel context) ANOVAs, by subjects (collapsed over five tokens per cell) and by items (collapsed over 32 subjects and five tokens per cell). All analyses are collapsed over the four test orders (the different orders did not affect accuracy systematically). In the analyses, the main effect of coarticulatory overlap was not significant by subjects or items [$F_1(1,31) = 1.88$, $p > .1$; $F_2(1,192) < 1$]. However, all other main effects and interactions were highly significant, at the $p < .0001$ level. The absence of a main effect of coarticulatory overlap is important (higher order interactions notwithstanding) because it indicates that differences in subject accuracy on tokens from the different conditions were not due to the shorter durations of the high-overlap, fast-rate syllables or to the fact that the high-overlap syllables were edited to remove the initial vowel. The significant higher order interactions indicate that performance varied for different CV syllables, spoken with different degrees of coarticulatory overlap; this is very much apparent in Figure 7.

Classification by models. We compared the performance of the four different models, two using linear (locus-equation-based) prototypes and two using the exemplar-matching algorithm. The locus equation models differ in that one uses one line for all /gV/ syllables, whereas the other has separate lines for /g/ in front and back vowel contexts. The model with two /g/ lines should

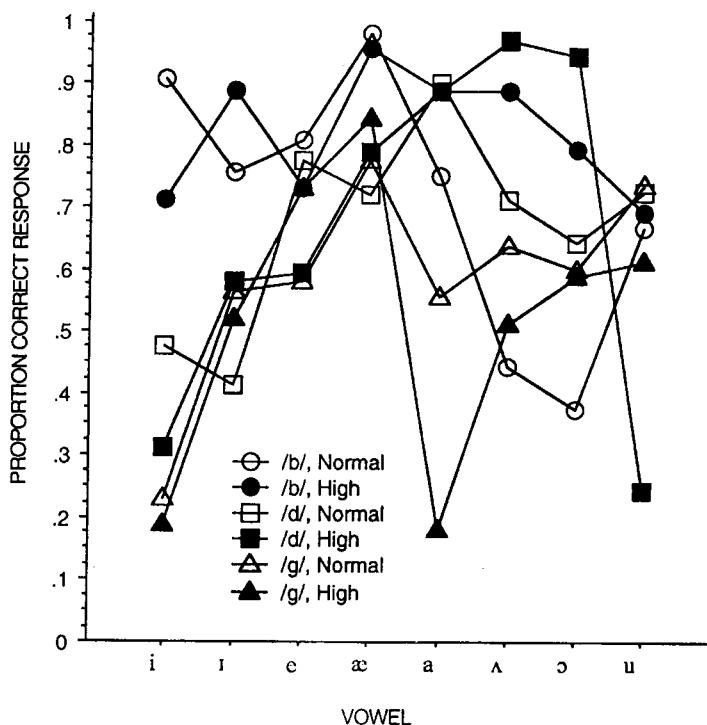


Figure 7. Subject accuracy in classification of tokens, by consonant, vowel, and coarticulatory overlap condition (normal vs. high).

classify tokens better, because it fits the $F2$ space for /g/ tokens more accurately. However, as discussed in Experiment 1, the appropriateness of considering /g/ as two separate allophones is debatable. The exemplar models differ from one another in the exponent used to increase the relative contribution of small distances. We expected that the model with the greater exponent (10) would classify tokens better than the one with the smaller exponent (5), because the greater exponent dampens to a greater extent the contributions of stored tokens at greater distances from the input token. These should be less informative for identifying the consonant.

The locus-equation lines against which the test tokens were matched are presented, along with the test tokens, in Figure 8. (The exemplar model contains too many data points to be plotted in usefully.) Figure 8A presents the locus equation lines for /b/, /d/, and /g/, and Figure 8B shows the lines for the /b/, /d/, and /g/-allophone lines. It is evident that the locus equation model will have trouble classifying back-vowel /g/ tokens, because the /g/ line does not fit their space effectively. In addition, the figure demonstrates that the space for /Ci v / (high $F2_{\text{vowel}}$) tokens is crowded, so that there should be a number of misclassifications of those tokens.

Overall, the models' accuracy in classifying the tokens was slightly better than the subjects' identification scores. The locus equation model (hereafter, LocEq) and the exemplar model with an exponent of 5 (hereafter, Exem-

plar 5) were equally accurate, correctly classifying 71.2% of the tokens (although their errors did not all occur on the same tokens). The locus equation model with /g/ allophones (LocEq+) and the Exemplar 10 model coincidentally also accurately classified equal numbers of tokens (77.1%).

The models' accuracy in classification is graphed in Figures 9A–9D, by consonant and vowel. Overall, the models' classification patterns bear some resemblance to that of the subjects, as can be seen through a visual comparison of Figures 7 and 9. Moreover, the models' response patterns were generally similar to one another: the models' percent agreement on classifications range from 76.7% (between the LocEq and Exemplar 5 models) to 90.4% (between the LocEq and LocEq+ models).

Models' prediction of subject responses. We examined the models' accuracy in predicting the pattern of subjects' classifications of the tokens in a number of different ways. First, we looked at the overall agreement in responses between the models and the subjects (without regard to the correctness of the classifications). Next, we considered the models' predictive strength with regard to subject accuracy—that is, how well they predict which items subjects will misidentify and what they will misidentify them as. Finally, we performed regression analyses on subjects' responses with the Euclidean distance scores computed in the models, to analyze the response patterns at a finer grain.

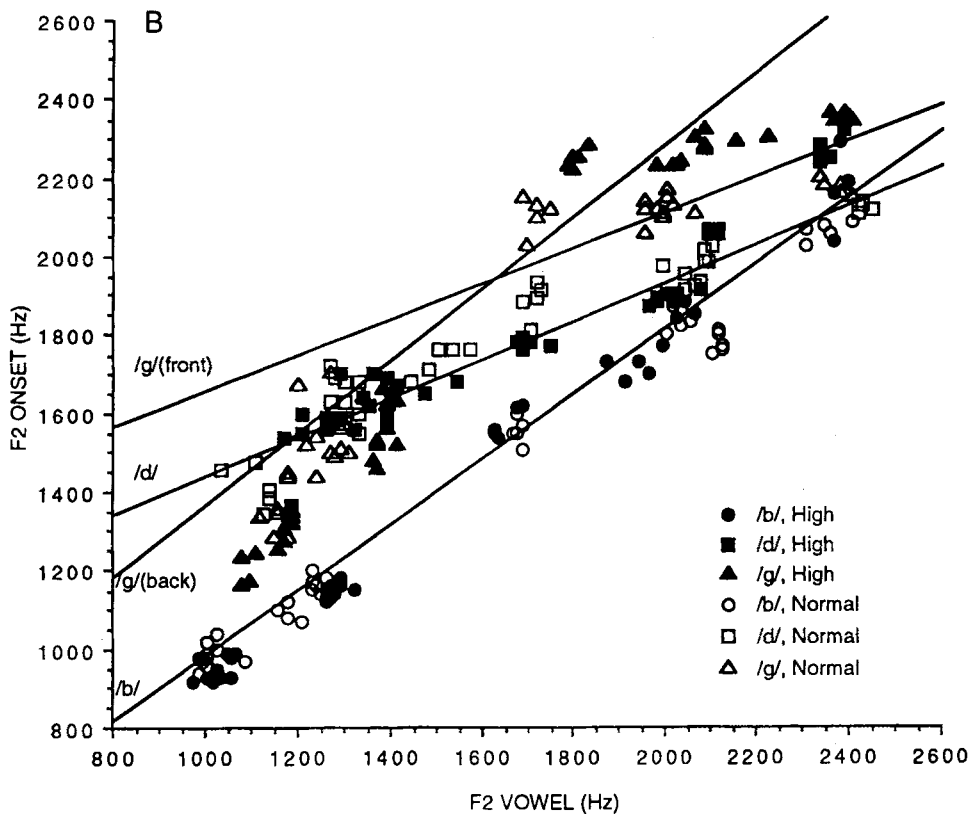
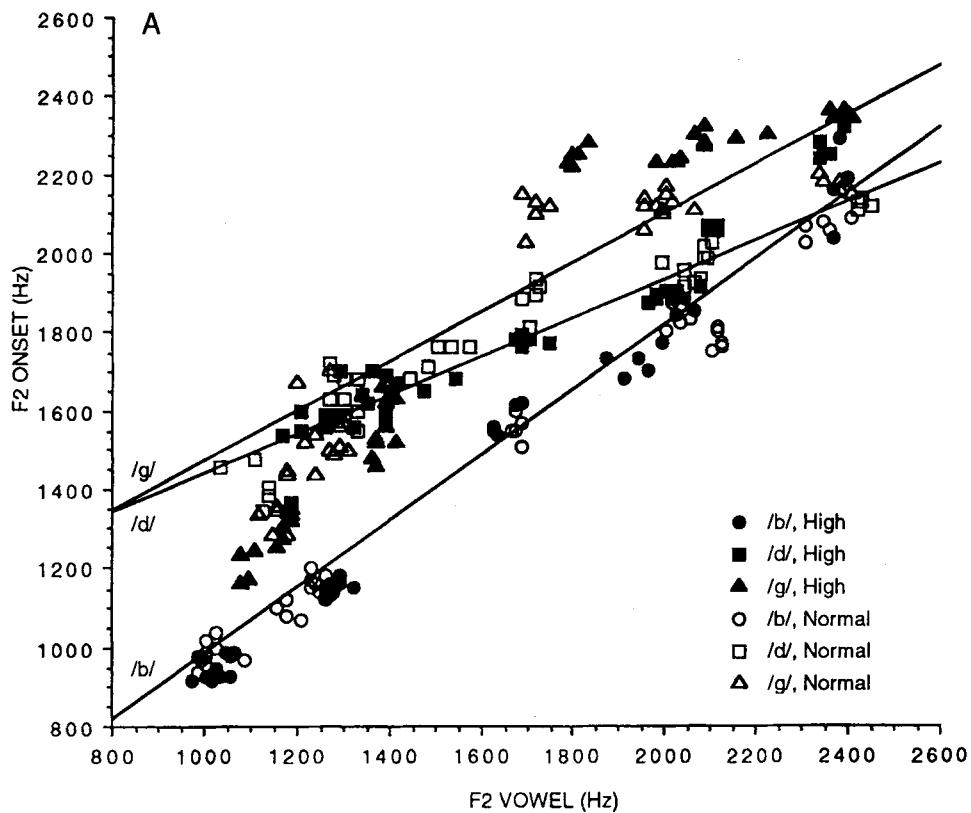


Figure 8. Locus equation templates computed from database of productions of 6 male speakers, plotted with tokens used in Experiment 2: (A) /b/, /d/, and /g/ lines; (B) /b/, /d/, and /g/-front and /g/-back lines.

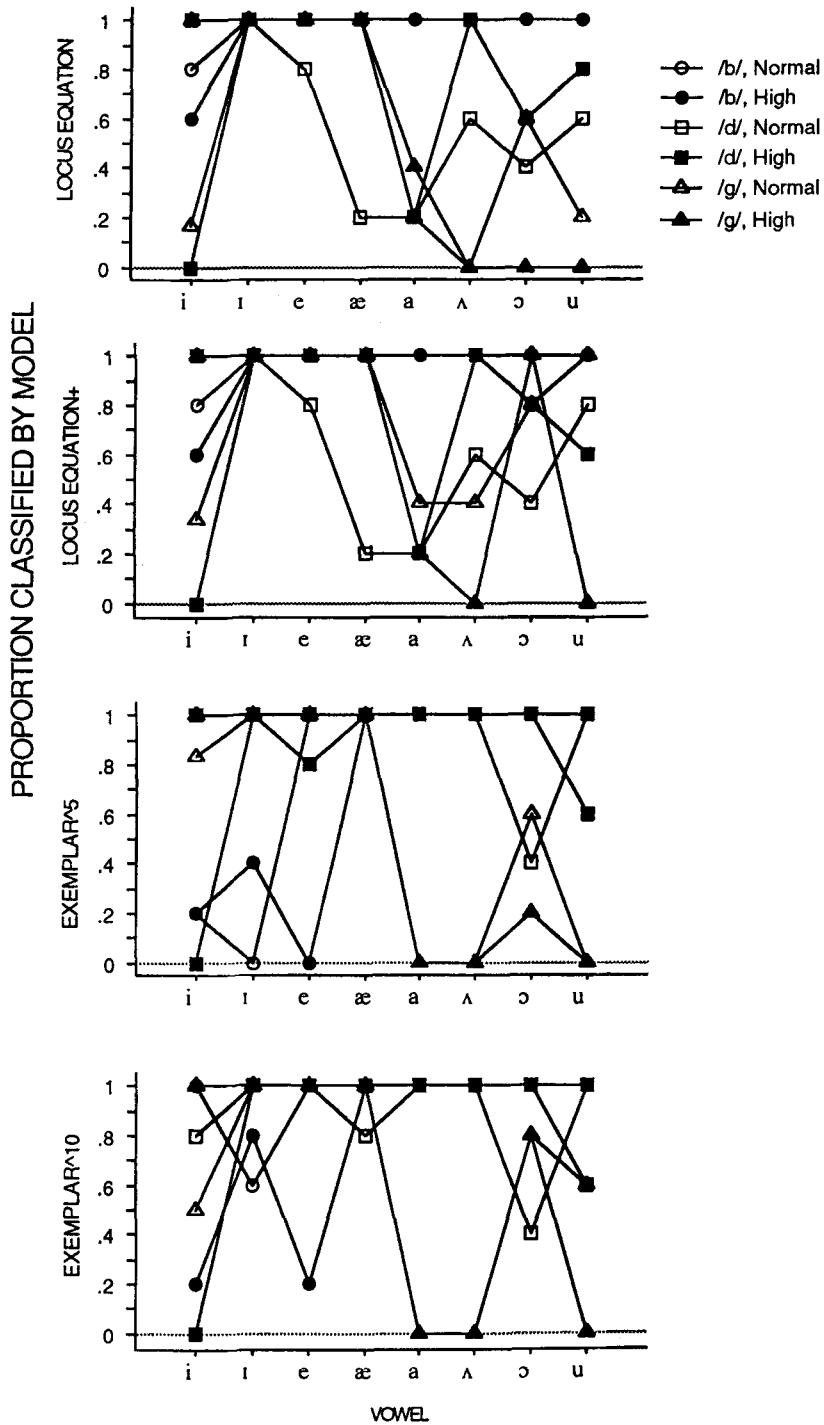


Figure 9. Models' accuracy in classification of tokens, by consonant, vowel, and coarticulatory overlap condition (normal vs. high).

Prediction of response patterns. In the first analysis, we looked at the percentage of subject responses that matched the response predicted by the models for each item. The identity of the test token did not enter into this analysis; it is indifferent to the accuracy of subject clas-

sifications. The results are presented in Table 4. The Exemplar 10 model performed the best, followed by the LocEq+, Exemplar 5, and LocEq models, respectively. The models all predicted between 55% and 62% of the subject responses; thus, the models' predictions were well

Table 4
Percent Agreement on Classifications
Between Subjects and Models

Model	Proportion of Agreement
LocEq	.555
LocEq+	.589
Exemplar 5	.576
Exemplar 10	.619

above chance (33%), but much of the variance in listener responses was not accounted for by any of the models.

The response prediction scores of the models were compared in repeated measures ANOVAs on percent response agreement, by subjects and by items. The main effect of model was significant by both subjects [$F(1,93) = 53.81, p < .0001$] and items [$F(2,717) = 4.79, p < .01$]. Post hoc comparisons of the models were also performed by subjects and by items; alpha was set at .01 to correct for error inflation. The subjects analysis found that all differences between models were significant at $p < .0001$, except for the comparison between the Exemplar 5 and LocEq models, which was marginally significant [$F(1,93) = 6.15, p = .015$]. The items analysis, on the other hand, revealed a significant difference between the Exemplar 10 model and the LocEq model [$F(1,717) = 13.69, p < .001$], and a marginal difference between the Exemplar 10 and Exemplar 5 models [$F(1,717) = 6.20, p = .013$], but no other significant differences. Thus, the more powerful of the exemplar models did better than the weaker LocEq model, and more marginally (significant only by subjects) than the stronger LocEq+ model, at predicting the pattern of subject responses. However, the differences between the models were small, and the total variance in subject responses predicted by all the models was modest.

Prediction of error patterns. In addition to considering how well the models predicted subjects' response patterns, we also examined the degree to which the models predicted which items subjects would misidentify and that to which they correctly predicted subjects' error patterns. We divided subject errors according to the responses predicted by the models. Table 5 presents the percentage of subject errors on items predicted by each model to be correctly identified, misidentified with the response given by the subject ("same error"), and misidentified with the response not given by the subject ("different error"). A 4 (model) \times 3 (prediction: predicted correct, same error, different error) repeated measures ANOVA was performed on the subject errors by subjects and items. There was a significant main effect of prediction in both analyses [$F(1,2,62) = 284.63, p < .0001$; $F(2,478) = 83.01, p < .0001$], as well as a significant interaction of model and prediction [$F(1,6,186) = 28.85, p < .0001$; $F(2,6,1434) = 2.99, p < .01$]. (There was no main effect of model, because the mean percent error, collapsed across prediction, was necessarily identical for all four models.) Post hoc analyses on prediction found that the percentage of errors predicted to be correct outnumbered both the

"same error" responses [$F(1,62) = 349.05, p < .0001$; $F(2,1,478) = 101.81, p < .0001$] and the "different error" responses [$F(1,62) = 492.53, p < .0001$; $F(2,1,478) = 143.76, p < .0001$]. The difference between "same error" and "different error" responses, with the former outnumbering the latter, was significant by subjects only [$F(1,62) = 12.32, p < .001$; $F(2, p > .05)$]. Thus, most subject errors were incorrectly predicted by the models to be correctly classified, but in cases in which a model did predict an error, it was (marginally) more often the case than not that the erroneous response was that predicted by the model (i.e., the proportion of "same error" was greater than the proportion of "different error"). These trends were compared by model in post hoc tests. For all four models, the comparisons of percentages of predicted correct to "same error" and to "different error" were highly significant ($p < .0001$) in both items and subjects analyses. Additionally, for all four models, the comparison between proportions of "same error" and "different error" were highly significant ($p < .0001$) in the subjects analysis. In the more conservative items analysis, the comparison for the Exemplar 5 [$F(2,1,1434) = 23.03, p < .0001$] and Exemplar 10 [$F(2,1,1434) = 22.35, p < .0001$] were again highly significant; however, the comparisons for the LocEq and LocEq+ models did not reach significance ($p > .05$ for both). Thus, error prediction appears to have been weaker in the locus equation models than in the exemplar models, although all of the models failed to predict the majority of subject errors.

Regression analyses. In the preceding analyses, the models' predictions were based on their categorical classifications of the tokens, and not on the magnitude of the distance scores (i.e., the distance to a locus equation line or composite distance to stored exemplars). Thus, the foregoing analyses did not distinguish, for example, two types of instances in which listeners selected /b/ as a response: In one case, the token /b/ was close to the /b/ line (in the linear model) or to a number of /b/ traces in memory (in the exemplar model) and far from the /d/ and /g/ lines/exemplars; in another case, /b/'s distance score was only moderately small and only slightly smaller than /d/ or /g/'s scores. Given that the models might capture information that influenced (but did not completely determine) subjects' responses, it was worthwhile to perform an analysis that could demonstrate a relationship between degree of competition between two response choices (based on distance scores) and subjects' response patterns.

Table 5
Percent Classifications by Models on
Tokens Misclassified by Subjects

Model	Proportion Predicted		
	Correct	Wrong (Same)	Wrong (Different)
LocEq	.660	.194	.146
LocEq+	.707	.167	.126
Exemplar 5	.647	.242	.111
Exemplar 10	.677	.226	.097

One way to do this was to conduct regression analyses of responses by distance scores. In one of the regression analyses that we conducted, each test token was represented three times: once by the percentage of subjects responding /b/, once by the percentage of /d/ responses, and once by the percentage of /g/ responses. Each percent response value (for each response choice) was paired with a distance score for the particular response (/b/, /d/, or /g/) for a given token. Thus, the correctness of a response (whether it was the consonant intended by the speaker) did not figure in the analysis. The difference between the distance score to the selected response and the smallest of the distance scores of the nonselected responses was used as the regressor. In this relative distance measure, if the selected response was the response predicted by a model, then the difference score was negative (because all other distances would be larger than that of the selected response); the distance score would always be positive if the selected response was not the predicted response. A negative coefficient indicated that subjects responded more often with consonants associated with smaller distances as compared with the distances of the alternative choices. In addition, we performed regressions using only the correct response's relative distance score for each token, along with subjects' rate of selection of the correct response.

The R^2 values for the all-response and correct-response regressions for each model are presented in Table 6. When all responses are considered, the models accounted for only a modest proportion of the variance in responses, between 36% and 46%. The Exemplar 10 model accounted for the greatest amount of variance, followed by the LocEq+ model. The LocEq model was the poorest predictor. When only the correct responses are included in the regression, the models fare very poorly: they accounted for less than 10% of the variance in subjects' rate of selection of the correct response. The same ordering of the models was found as in the all-response regression.

Plots of the all-response and correct-response regressions for the best-performing models, the LocEq+ and Exemplar 10, are presented in Figure 10. In the all-response regressions for both the locus equation and exemplar models, strongly negative and strongly positive distance scores overall tended to correspond to larger and smaller proportions of response selections, respectively. However, for responses with relative scores closer to zero (both positive and negative), proportions of subjects' response selections were widely distributed. The latter

finding is demonstrated by the correct-response regression plots in Figure 10, which show the very weak relation between relative distance and response selection for the correct responses, whose relative distance scores are mostly negative or positive but small. The poor fit in the correct-response regressions indicates that the apparent success of the models in the all-response regressions was due predominantly to their elimination of responses, but not to their selection of them. That is, responses with very high relative distance scores were rarely selected by subjects, but responses with strongly negative scores were not necessarily chosen in higher proportions than were those with weakly negative or weakly positive scores.

Statistical comparisons were made between the relative distance correlation coefficients of the four models, using a t test for comparing dependent r s (Cohen & Cohen, 1975, p. 53). In the all-response regressions, the Exemplar 10 model explained significantly more of the variance in the response patterns than did the Exemplar 5 [$t(717) = 10.45, p < .001$] and the LocEq model [$t(717) = 6.45, p < .01$], and the LocEq+ model explained more variance than did the LocEq model [$t(717) = 7.60, p < .01$] and the Exemplar 5 model [$t(717) = 2.93, p < .01$]; no other differences were significant. In the correct-response regressions, none of the differences between the models were significant.

Speaker-specificity and $F2$ models. Figure 8 reveals that the templates in the locus equation model often do not match the test tokens very well. Both the single-/g/ and the allophone-/g/ lines do not intersect the regions of the /g/ test tokens as well as the /b/ and /d/ lines do for the /b/ and /d/ tokens, respectively. Assuming that M2 is not an atypical speaker, this is merely a chance outcome of a small database. However, there are considerable speaker differences in slope and intercept values (cf. Sussman et al., 1991), and a realistic model must be able to overcome such differences. Although mechanisms have been suggested that normalize vowel spaces (e.g., Disner, 1980; Sussman, 1988, proposes a neurally based normalization model), none have been proposed that would convert $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values in such a way that locus equation slopes and intercepts would be normalized. However, if this could be accomplished, a speaker's own productions could be used to simulate normalized linear prototypes (or exemplars) to be used by the $F2$ -based perception models.

To simulate normalization, we provided the models with their best opportunity to explain listener identification patterns, by using M2's normal and fast productions (i.e., the test tokens) as the "memory" database rather than the productions of six different male speakers. The normal and fast productions were pooled to compute the locus equation coefficients. Distances were computed in the same way as before, except that in the exemplar models a jackknife technique was used so that a token was not matched against itself, but only against the other nine tokens of the syllable. Two other tokens, with the same vowel and rate as those of the test token but with the other two consonants, were also dropped in each exemplar dis-

Table 6
 R^2 Values for Regressions of Percent Subject Response Selections by Relative Distance Scores, With All Response Choices and Correct Responses Only

Model	Responses	
	All	Correct Only
LocEq	.367	.025
LocEq+	.436	.056
Exemplar 5	.388	.036
Exemplar 10	.459	.068

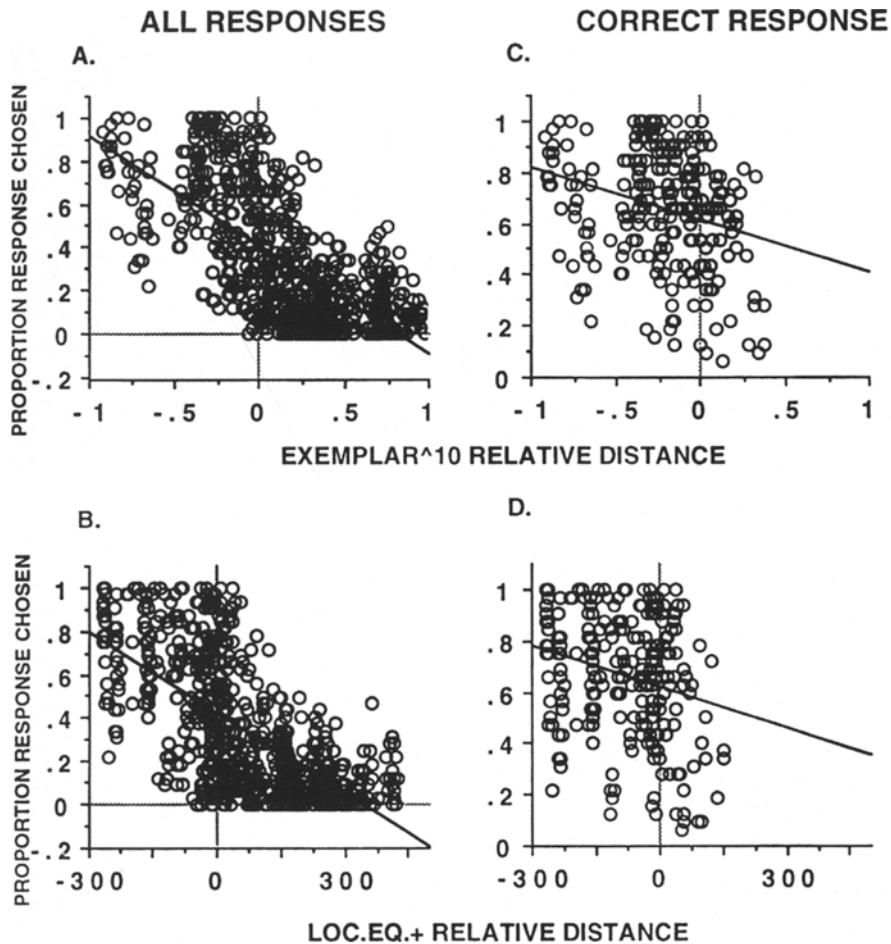


Figure 10. Regression plots of percentage of subjects choosing a particular response by relative distance scores produced by Exemplar 10 and LocEq+ models, with all response choices (A and B) and only correct responses (C and D) included. See text for description.

tance computation so that equal numbers of tokens of each consonant were used in each computation.

In classification analyses, the models all performed better than in the multiple-speaker database analyses, but they were ordered in their predictions of subject responses in the same way as in the multiple-speaker database, with the highest performance by the Exemplar 10 model (65% of the responses) and the lowest by the LocEq model (60.3% of the responses). In the relative distance regression analyses, however, the locus equation models improved their standings relative to the exemplar models (again, all models improved substantially from the earlier analysis). In the all-response regressions, the LocEq+ and Exemplar 10 models performed similarly (R^2 s of .507 and .503, respectively), and the LocEq model slightly outperformed the Exemplar 5 model (.477 and .448, respectively). In the correct-response regressions, the models still accounted for a minimal proportion of the variance in responses. The LocEq+ and LocEq models (R^2 s of .124 and .076, respectively) numerically outperformed the Exemplar 10 and Exemplar 5 models (R^2 s of .064 and

.041, respectively); however, the differences between the LocEq+ model and the Exemplar 10 model was not significant [$t(237) = 1.42, p > .05$], and the difference between the LocEq+ and the Exemplar 5 was marginal [$t(237) = 2.11, p < .05$]. In summary, all of the models performed better without the added variability of a multiple-speaker database, but comparisons among the models across the different analyses did not consistently demonstrate superior performance by either (locus equation or exemplar) approach.

Discussion

In Experiment 2, we tested the ability of models using locus equation prototypes or an exemplar database consisting of stored tokens' $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values to predict subjects' responses on burstless CV syllables. Overall, the variance in subject responses explained by the models (taken together) was significant, but modest (up to 65% of classifications predicted, and up to 51% of variance in regression analyses explained). Clearly, the metrics using $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ as parameters do capture

some systematicity in subjects' response patterns. However, they also fail to account for a large proportion of the patterning. The results indicate that the locus equations' relative success in classifying tokens (70%–80%, depending on the size of the database and whether separate /g/-allophone lines were used) should not be taken as evidence in favor of their perceptual relevance because the models' errors coincided only slightly with subjects' errors; most of the subjects' errors were on tokens that the models predicted to be correctly identified. The models' predictive strength (that is, their R^2 s being as high as they were in the regressions) appears to lie mostly in the fact that responses predicted to be extremely unlikely selections (those with large distance scores) were, in fact, rarely chosen. In contrast, some of the responses predicted to be more likely selections were chosen on a high proportion of decisions, but others were not. Thus, the models were weak in the sense that they had some success at eliminating some responses, but they were poor at predicting chosen responses. Our findings thus generally concur with those of the synthetic speech experiments by Fruchter and Sussman (1997) and Eek and Meister (1995), who found that the regions in $F2$ space corresponding to high-proportion selections of each consonant only partially overlapped with their respective locus equation lines. Thus, in response to the research questions posed by Sussman et al. (1991), we found that proximity to locus equation lines bears only a gross relationship to perceptual consonant classification.

The metrics using $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ are obviously insufficient to stand alone in a model; they apparently require supplementation by other sources of information, such as $F3$ and burst spectra. (For our experiment, however, burst information would not have helped; recall that the bursts had been removed from the stimuli.) Of course, it has not been claimed that $F2$ transitions alone are sufficient in a model; Sussman (e.g., Sussman et al., 1991) has argued that $F2$ transitions must be supplemented by other cues, although he has not made specific claims about the relative contributions of the various cues. However, given that the $F2$ transition is proposed as a primary cue, and one for which evolution shaped the auditory perceptual system, it might be expected to have more predictive strength than was found here. A different sort of account could be devised in which $F2$, $F3$, and burst information are not treated as independent cues (Dorman, Studdert-Kennedy, & Raphael, 1977), but rather are considered together; for example, a model could be based on prototypes defined in a three-dimensional space of $F2_{\text{ons}}$, $F2_{\text{vowel}}$, and $F3_{\text{ons}}$ (Lindblom, 1990; Sussman, 1991). We believe, however, that the OOC hypothesis cannot apply to such a model. According to the OOC, developing perceptual systems capitalize on linearly related input parameters when organizing representational maps; speakers produce a linear relation between $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ so that the speech perception system can map them. However, although it has been reliably demonstrated that $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ are consistently highly cor-

related, it is not the case that there is a linear relationship between $F2_{\text{ons}}$, $F2_{\text{vowel}}$, and $F3_{\text{ons}}$. (On the contrary, the three-dimensional scatterplots presented by Lindblom, 1990, and Sussman, 1991, reveal irregularly shaped regions for /b/, /d/, and /g/.) If a perceptual system organizes a representational map in the three-dimensional space of $F2_{\text{ons}}$, $F2_{\text{vowel}}$, and $F3_{\text{ons}}$, it is not exploiting a linear relationship between its input parameters. The point here is not that such a representational system is necessarily implausible, but rather that it is incompatible with the OOC. For the OOC to hold, $F3_{\text{ons}}$ would have to enter into a higher order representational system that would take as its inputs the output of the OOC-driven $F2$ representational subsystem as well as $F3_{\text{ons}}$ and possibly other cues such as burst information. (In fact, this type of model is discussed by Sussman et al., in press; see their Figure 17.)

The locus-equation-prototype models were compared with the exemplar models to determine whether listeners make specific use of the linearity of points in $F2_{\text{ons}}$ -by- $F2_{\text{vowel}}$ space. (We stress that the performance of the exemplar models should not be taken as a reflection on exemplar models as a class; this particular instantiation had a very restricted input of only two variables.) Our results clearly show no predictive advantage of incorporating locus equation lines into a model. In fact, the locus equation approach introduces a normalization problem that is more severe than in the exemplar models. It is noteworthy that the exemplar model outperformed the locus equation models when the database included productions of multiple speakers, but not when the database included only the productions of the test speaker. The lines computed over the productions of six other speakers differed from the test speaker's own locus equation lines, and thus his tokens did not fit as well to these lines; however, the greater range of $F2$ values for tokens of a given CV in the larger database apparently compensated for this somewhat in the exemplar models. For the locus equation template approach, if an individual speaker's lines deviate from population lines (as in the case of Speaker M2), then, for the model to be effective, a procedure must be available that transforms the speaker's lines to the norms. However, it is not clear that a sufficiently powerful normalization procedure could be devised. The differences in locus equation line coefficients among speakers are not simply a consequence of different vocal tract sizes; as demonstrated in Experiment 1 and elsewhere (e.g., Duez, 1992; Krull, 1989), locus equation slopes and intercepts are sensitive to a number of factors that influence degree of coarticulatory overlap, such as rate, stress, and casualness of speaking style. A normalization procedure would have to correct for all such factors, even though they may vary within speakers, even from word to word in continuous speech, as well as between speakers.

Across all of the analyses, we found the locus equation model with separate prototypes for the palatal and velar allophones of /g/ (for /g/ in the context of front and back

vowels, respectively) to be a consistently better predictor than the locus equation model with only one /g/ template. This is not at all surprising because the single-line fit to /g/ is relatively poor; many /gV/ tokens fall far off the regression line while falling close to one of the /g/-allophone lines. From a modeling standpoint, having separate /g/ templates is more effective. However, in Experiment 1, we found higher order systematicity in second-order locus equations for /g/, and we also found that the single-category /g/ was more clearly distinct from /b/ and /d/ in slope-by-intercept space than were the separate /g/-allophone categories; these findings can be taken as evidence of /g/'s coherence as a unitary category in production, along with the fact that perceivers hear all /g/s as members of one phonological category. We view this as an inconsistency in the locus equation model, since it requires two lines for /g/-allophones in spite of evidence that there is regularity captured across all productions of /g/.

We conclude our discussion of Experiment 2 by returning to the basic premise of the locus-equation-based model: that the linearity in $F2$ transitions captured by locus equations is perceptually relevant. Although our results are broadly consistent with the possibility that there are particular regions in $F2$ space for different consonants that are important for perception of consonants, they do not demonstrate any superiority of a locus equation model over a comparable model that does not use linear representations. Thus, our findings do not support the claim that locus equations have perceptual relevance.

GENERAL DISCUSSION

In Experiment 1, we demonstrated that slopes and intercepts for a consonant of a given place of articulation can change systematically under various speech manipulations, but that slope and intercept taken together remain distinctive for place, thus supporting Sussman's claim (e.g., Sussman et al., in press). However, we found systematic variability in $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values for a given CV under the different manipulations, which is captured by second-order locus equations, and we argued that such systematicity within a consonant class is not predicted by the OOC. We also argued that the linearity captured by locus equations is not unique, because other examples of linearity are found in the outputs of acoustic analyses of stop consonants (such as $F1$ locus equations and SOLEs) that do not serve a role in perception. Finally, we presented an account of the source of linearity in $F2$ values for stop consonants different from Sussman's OOC. We proposed that productions are roughly linear because coarticulation resistance is roughly equal for a given consonant in different vocalic contexts. In Experiment 2, we tested the perceptual relevance of locus equations by comparing subjects' classifications of CV tokens with their bursts removed to the predictions of a model using locus-equation-derived prototypes. We found that the locus equation model did

predict an above-chance proportion of subject responses, but that it left a large amount of the variance in responses unexplained. Furthermore, the model did no better than an alternative model that used the same $F2$ information but did not exploit the linearity of $F2$ values. Here, we will further discuss the findings of both experiments in reference to the OOC.

Sussman et al. (in press; see also Sussman et al., 1995) propose that, over the course of evolution, humans have exploited a predisposition of neural auditory systems (like those found in the barn owl and bat) to encode linear relationships in acoustic signals for the purpose of consonant perception. In this system, certain neurons respond to particular, linearly related, pairings of $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ values, extracted from a CV speech signal; when activated, they send information to a higher order system for the identification of a particular consonant. Sussman et al. (in press) argue that such a system has been able to develop because linearly related acoustic input parameters—namely, $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ —are available to be mapped. Thus, speakers have been subjected to an evolutionary constraint (the OOC) to produce CV syllables such that $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ are linearly related, in order to facilitate development of these representational maps.

There are some differences between the proposed consonant perception system and the bat and barn owl auditory systems motivating it that weaken the proposed analogy (cf. Pinker & Bloom, 1994). In particular, the higher order properties captured by the nonhuman systems (sound-source azimuth in the barn owl, object velocity in the bat) are continuous functions, not discrete categories like stop consonants. Secondly, these systems deal with *perfect* linear regressions; that is, by the laws of physics, a given Doppler shift for a given frequency infallibly signals a particular relative object velocity (at least in the case of the head-on approach); a particular phase shift for a particular frequency perfectly signals one interaural time difference, and thus one object location. Accordingly, they differ from the good, but far from perfect, fits of speech tokens to locus equation lines. In addition, locus equation lines, unlike Doppler-shift and phase-shift functions, are not parallel; a particular pairing of $F2$ values will not uniquely correspond to a single consonant category where the lines cross. Recall that locus equation metrics fail to classify speech tokens perfectly, achieving accuracy rates of only up to 80% (if speaker normalization is allowed). Finally, the multiple frequency-phase pairs for barn owls and the multiple frequency shifts for bats that constitute the linear relationships occur simultaneously in a stimulus, unlike in speech, where only one $F2_{\text{ons}}-F2_{\text{vowel}}$ pair occurs at a time. That is, there is linearity in the relation of acoustic variables that are contemporaneous in the stimulus input to bats and barn owls, but are historical in the stimulus input to humans.

Given the lack of parallelism between the proposed consonant system and the bat and barn owl systems,

there should be strong independent empirical support for the hypothesis of a neurally instantiated locus-equation-based representational system. For example, a finding that our perceptual data were better fit by the locus equation model than the exemplar model would have provided such support. However, we found that the exemplar model, which does not assume the presence of linear prototypes, predicts subject classifications at least as well as (and marginally better than) a locus-equation-based account. Furthermore, the results from Experiment 2 demonstrate that, if the perceptual system does have linear representations, their contribution to consonant perception is limited. The linear-prototype models at their most powerful (with separate /g/-allophone lines and a charitable simulation of speaker normalization) accounted for only about half of the variance in the response patterns by distance scores, and they predicted only about 60% of subjects' responses themselves. Furthermore, these results were for syllables with their bursts removed. If it is acknowledged that burst information contributes to consonant perception (Sussman et al., 1991), then the role of the transition information should be greater for perception of these stimuli when the bursts are removed. Thus, the perceptual relevance of locus equations for intact syllables would be even smaller than that observed for the burstless stimuli.

To summarize, we have argued that $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ differ fundamentally from the input parameters used by auditory systems documented to utilize linear functions. $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ are insufficient to distinguish stop place fully, and, compatibly, we found in Experiment 2 that if listeners use these values as acoustic cues in stop place categorization, the cues' contributions are moderate at best. Unlike the linear relationship of the frequencies of a pulse and its Doppler-shifted echo, which is sufficient for the bat to determine object velocity, $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ require supplementation from other acoustic parameters, such as burst spectra or $F3_{\text{ons}}$. We therefore question the plausibility of the OOC, on the grounds that the perceptual utility of $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ is not great enough to have merited the imposition of an evolutionary constraint on speech production. Furthermore, we believe that the findings of good fits to $F1$ locus equations and to second-order locus equations weaken the theoretical motivation for proposing such a constraint, because they provide examples of linearity that are unlikely to be due to perceiver-driven constraints.

Alternative Accounts

We have claimed that our results cast doubt on the plausibility of the OOC. We also concluded that the evidence for the perceptual relevance of the linearity captured by locus equations is weak; the locus-equation-prototype models' performance accounting for subjects' response patterns is modest, and it is no better than that of an alternative model without linearity built in. What sort of model, then, would be superior? Because we de-

signed our perceptual experiment with the particular aim of testing the locus equation account, our results do not offer specific suggestions on how to build a more successful model. However, given the general acceptance that $F2$ transition information is important for stop consonant identification, the model's relatively poor performance using this information is significant. It suggests that using $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ as relational cues that have a privileged processing stage (independent from the contributions of other sources of information) is a flawed approach. There are nonexclusive possible remedies. One is to take $F2$ information in the context of other sources of information, such as $F3$ and (although it is not relevant to the results of our experiment) the burst. Eek and Meister (1995) have suggested that perception is guided by the relation of the vowel's $F2$ to the strongest spectral peak in the burst. Earlier, Kuhn (1975, 1979) claimed that whichever formant corresponds to the front vocal tract cavity resonance will have greater perceptual relevance, and that the burst should be treated as continuous with the front cavity resonance. Dorman et al. (1977) added that the burst will be more closely tied to the major spectral peak of the vowel when transitions are shorter and, conversely, more independent of the major spectral peak when transitions are extensive, and that the relative contributions of the burst and transitions to perception vary according to these relationships. In another vein, Fant (1973), Lindblom (1990), Krull (1990), and Sussman (1991) have used static acoustic cues from $F2$ and $F3$ (and $F4$, for a Swedish contrast, in Krull, 1990), to form acoustic spaces corresponding to listeners' classifications. (As we discussed earlier, the latter set of approaches, which create multidimensional spaces utilizing a number of acoustic cues, is distinct from approaches in which the output of an $F2$ -processing system is integrated with the outputs of systems that process other cues, such as the model schematized by Sussman et al. [in press]. We argued that the former class of models is inconsistent with the premise of the OOC.)

Another remedy is to capture the information in $F2$ transitions in some other way than by two static values. It may well be the case that there is relevant information for consonant identification in the trajectory of the transitions that is lost in the $F2_{\text{ons}}-F2_{\text{vowel}}$ metric (see, however, Sussman et al., 1995). It should be borne in mind that the static $F2$ -parameter models (linear prototype and exemplar) did capture some variance in subject classifications and were well above chance in predicting subjects' responses. In other words, the models' bases for misclassifying a token corresponded to some extent to that of the subjects. This may suggest that the models use a correct source of information but under-sample it. Pols and Schouten (1978) found that burst removal impaired identification performance on voiceless stops more than on voiced stops, and that performance on the voiceless, but not voiced, stops was improved by having noise precede the onset. They concluded that the abrupt acoustic

onset brought on by burst removal masks the early pitch periods; furthermore, they offered the interpretation that the voiced stops are less affected because the transitions are longer and carry information over a greater temporal extent. This view runs counter to the assumption that only the frequency at the onset of voicing is crucial for place identification.

In opposition to the view that acoustic parameters are treated as cues to be processed in order to assign categories to the speech signal is the ecological, or direct-realist, theory of speech perception (Best, 1993, 1995; Fowler, 1986, 1994). In the direct-realist perspective, listeners perceive the place of articulation of a consonant because it is specified by the acoustic signal. That is, the theory rejects the popular assumption that phonemes are mental constructs whose representations are accessed by matching them to cues extracted from an acoustic signal. Instead, the view argues that phonemes are vocal tract events, and that these events are the objects of perception; information specifying the event is made available to the perceiver via the acoustics. The approach does not make reference to mental representations of phonemes, following the view that because the proximal stimulation is sufficient to support perception (a central assumption of the ecological approach), representations are not necessary (Shaw & Bransford, 1977; Shaw & Turvey, 1981).

On the surface, our results may appear to challenge direct realism, because they do show a moderate tendency for subjects to misidentify consonants when a token is better matched to a different consonant according to an acoustic “cue” ($F2$); such findings are generally consistent with a representational approach. In fact, if the acoustic information specifies place of articulation, there should not be any misidentifications in the direct-realist view (so long as the relevant information is picked up by the perceiver). However, the errors occurred on syllables with their bursts removed. From the realist stance, transitions with burst (i.e., an unaltered acoustic signal) specify place of articulation, but transitions alone do not necessarily. Transitions, artificially separated from the burst, may misleadingly index a different vocal tract event from the one that produced the signal. In other words, a different vocal tract event may produce an acoustic signal more like the (artificially burstless) transitions from the actual event than the event itself. Perceivers would thus be expected to misidentify consonants on the basis of this information.

Although locus equations are associated with representational approaches to perception, they are relevant to direct realism and, more broadly, to the class of theories that claim vocal tract gestures as the objects of speech perception (e.g., Liberman & Mattingly, 1985; Whalen, 1989), because of the connection between locus equations and coarticulation resistance. In gestural theories of speech perception (Fowler, 1986, 1996; Liberman & Mattingly, 1985), vocal tract gestures (which constitute phonetic segments) are units of production and of perception. For gestures to be perceived, the lis-

tener must parse the effects of one gesture in the acoustic signal from that of neighboring gestures, because the gestures have overlapping acoustic consequences due to coarticulation (see, e.g., Fowler & Brown, 1997). Thus, in a CV syllable, there is information in the burst and formant transitions that is parsed and attributed to the consonantal gesture, and information that is parsed and attributed to the vowel gesture (see, e.g., Fowler, 1984). Because gestures differ in their coarticulation resistance, they differ in their degree of coarticulatory overlap with their neighbors and therefore also in the degree to which their acoustic consequences are intertwined with the acoustic consequences of their neighbors (Recasens, 1985). Thus, a low-coarticulatory-resistance (high-overlap) gesture requires that more extensive acoustic consequences of the overlap be parsed in perception than does a high-coarticulatory-resistance one.

If locus equations are indices of coarticulation resistance, they also index (indirectly) the parsing necessary for a particular syllable. (This is essentially the point made by Sussman et al. [1995], when they wrote that locus equations reveal a “vowel-normalized $F2$ transition” [p. 3112] feature of stop consonants, in which “vowel-context-induced acoustic variability is ‘absorbed’” [p. 3123].) However, locus equations index coarticulation resistance by relating the extents of $F2$ transitions for a consonant in different vowel contexts. In other words, they are an *extrinsic* measure of coarticulation resistance—that is, a measure of the coarticulation resistance of a given token that requires reference to other tokens. However, the direct-realist parsing approach requires an *intrinsic* measure of coarticulation resistance—that is, a measure of the resistance of a token that can be determined on the basis of properties of that single token, without any external reference. This is necessary because in a realist theory, perception does not occur by comparing a presented token to stored information. An intrinsic index of coarticulation resistance must provide information about how much information should be parsed out for the consonant and the vowel, respectively, and allow for identification of the consonantal and vocalic gestures. This requires that the natural parsing lines in the acoustic signal must somehow make themselves available to the perceiver. Admittedly, we cannot suggest what this information might be, or how it is available to perceivers. Ideally for the account, however, there should be a higher order variable that captures this information, possibly taking the form of a function using dynamical information in the burst and transition. The crucial point is that there would not be a different function for each place of articulation, such that their outputs must be compared and the best fit selected, as in the locus equation approach.

Conclusions

To summarize, we induced changes in speaking style in productions of CV syllables that resulted in systematic changes in locus equation slopes and intercepts for /b/, /d/, and /g/ in Experiment 1. In spite of these slope and

intercept changes, the three consonants fell in distinct regions in slope-by-intercept space, although less clearly so when /g/ was divided into separate allophones for front- and back-vowel contexts. Thus, the findings of our production study generally support Sussman's claim (e.g., Sussman et al., in press) that slope and intercept are distinctive for place of articulation, at least for stop consonants. We also found that the locus equation slopes and intercepts of each consonant under the different manipulations could be fit to second-order locus-equation regressions; we argued that if these SOLEs hold up as robust phenomena, they may be problematic for the OOC theory. In our perception experiment, we found that a computational model of consonant perception based on a locus equation account proposed by Sussman et al. (1991) predicted subjects' classifications at above-chance levels, but that its overall predictive ability was modest at best. Furthermore, it performed no better at predicting responses than an exemplar-based model that did not capture the linearity of $F2_{\text{ons}}$ and $F2_{\text{vowel}}$. We concluded that these results argue against the postulation of perception models in which locus equation linearity is assumed to play an important role. Furthermore, we questioned the appropriateness of models in which $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ are treated as cues processed separately from other acoustic cues, and moreover those in which the $F2$ transition is reduced to these two static values. Finally, we paired our perception findings with other findings in the production domain (in particular, the fact that good-fitting locus equation functions can be found for $F1$ transitions) to argue against the plausibility of the OOC claim that the linearity in $F2_{\text{ons}}$ and $F2_{\text{vowel}}$ is a consequence of speakers' meeting the demands of perceivers. Instead, we proposed that Fowler's (1994) uniform coarticulation resistance hypothesis, in which perception plays no role, is better suited to explain locus equations.

REFERENCES

- AMERMAN, J. D. (1970). *A cinefluorographic investigation of the coarticulatory behavior of the apex and body lingual articulators*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.
- BAKRAN, J., & MILDNER, V. (1995). Effect of speech rate and coarticulation strategies on the locus equation determination. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIII International Congress of Phonetic Sciences* (Vol. 1, pp. 26-29). Stockholm: KTH and Stockholm University.
- BEST, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 289-304). Dordrecht: Kluwer.
- BEST, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Cross-language speech perception* (pp. 171-204). Baltimore: York Press.
- BLADON, R. A. W., & AL-BAMERNI, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics*, **4**, 137-150.
- BRANCAZIO, L., & MITRA, J. (1994). A reconsideration of locus equations as invariants for place of articulation of stop consonants [Abstract]. *Journal of the Acoustical Society of America*, **95**, 2923.
- BROWMAN, C. P., & GOLDSTEIN, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I* (pp. 341-376). Cambridge: Cambridge University Press.
- CARRÉ, R., & MRAYATI, M. (1990). Articulatory-acoustic-phonetic relations and modelling, regions, and modes. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 211-240). Dordrecht: Kluwer.
- CHENNOUKH, S., CARRÉ, R., & LINDBLOM, B. (1995). On the "locus equation" and its relation with the consonant place of articulation [Abstract]. *Journal of the Acoustical Society of America*, **97**, 3242.
- COHEN, J., & COHEN, P. (1975). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Wiley.
- CROWTHER, C. S. (1994). Modelling coarticulation and place of articulation using locus equations. *UCLA Working Papers in Phonetics*, **88**, 127-148.
- DELATTRE, P. C., LIBERMAN, A. M., & COOPER, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, **27**, 769-773.
- DISNER, S. F. (1980). Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America*, **67**, 253-261.
- DORMAN, M. F., STUDDERT-KENNEDY, M., & RAPHAEL, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, **22**, 109-122.
- DUEZ, D. (1992). Second-formant locus patterns: An investigation of spontaneous French speech. *Speech Communication*, **11**, 417-427.
- EEL, A., & MEISTER, E. (1995). The perception of stop consonants: Locus equations and spectral integration. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIII International Congress of Phonetic Sciences* (Vol. 1, pp. 18-21). Stockholm: KTH and Stockholm University.
- ENGSTRAND, O. (1988). Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *Journal of the Acoustical Society of America*, **83**, 1863-1875.
- FANT, G. (1973). Stops in CV-syllables. In G. Fant (Ed.), *Speech sounds and features* (pp. 110-139). Cambridge, MA: MIT Press.
- FARNETANI, E. (1990). V-C-V lingual coarticulation and its spatiotemporal domain. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 93-130). Dordrecht: Kluwer.
- FOWLER, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359-368.
- 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. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception & Psychophysics*, **55**, 597-610.
- FOWLER, C. A. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, **99**, 1730-1741.
- FOWLER, C. A., & BROWN, J. M. (1997). Intrinsic f_0 differences in spoken and sung vowels and their perception by listeners. *Perception & Psychophysics*, **59**, 729-738.
- FRUCHTER, D. (1994). Perceptual significance of locus equations [Abstract]. *Journal of the Acoustical Society of America*, **95**, 2977.
- FRUCHTER, D., & SUSSMAN, H. M. (1997). The perceptual relevance of locus equations. *Journal of the Acoustical Society of America*, **102**, 2997-3008.
- GRIESER, D., & KUHL, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, **25**, 577-588.
- HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.
- HINTZMAN, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.
- KEWLEY-PORT, D. (1982). Measurements of formant transitions in naturally produced stop consonant-vowel syllables. *Journal of the Acoustical Society of America*, **72**, 379-389.
- KLATT, D. H. (1987). Review of text-to-speech conversion for English. *Journal of the Acoustical Society of America*, **82**, 737-793.
- KRULL, D. (1989). Consonant-vowel coarticulation in spontaneous speech and reference words. *PERILUS*, **10**, 101-105.

- KRULL, D. (1990). Relating acoustic properties to perceptual responses: A study of Swedish voiced stops. *Journal of the Acoustical Society of America*, **88**, 2557-2570.
- KUHN, G. M. (1975). On the front cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America*, **58**, 428-433.
- KUHN, G. M. (1979). Stop consonant place perception with single-formant stimuli: Evidence for the role of the front-cavity resonance. *Journal of the Acoustical Society of America*, **65**, 774-788.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.
- LIBERMAN, A. M., DELATTRE, P. C., COOPER, F. S., & GERSTMAN, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, **68** (Whole No. 69), 1-13.
- LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.
- LINDBLOM, B. E. (1963). *On vowel reduction* (Report No. 29). Stockholm: Royal Institute of Technology, Speech Transmission Laboratory.
- LINDBLOM, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403-440). Dordrecht: Kluwer.
- MASSARO, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.
- NEAREY, T., & SHAMMASS, S. (1987). Formant transitions as partly distinctive invariant properties in the identification of voiced stops. *Canadian Acoustics*, **15**, 17-24.
- OHDE, R. N. (1988). Revisiting stop-consonant perception for two-formant stimuli. *Journal of the Acoustical Society of America*, **84**, 1551-1555.
- ÖHMAN, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, **39**, 151-168.
- OLSEN, J. F., & SUGA, N. (1991). Combination-sensitive neurons in the medial geniculate body of the mustached bat: Encoding of relative velocity information. *Journal of Neurophysiology*, **65**, 1254-1274.
- PINKER, S., & BLOOM, P. (1994). Authors' response: Humans did not evolve from bats. *Behavioral & Brain Sciences*, **17**, 183-184.
- POLS, L. C. W., & SCHOUTEN, M. E. H. (1978). Identification of deleted consonants. *Journal of the Acoustical Society of America*, **64**, 1333-1337.
- RECASENS, D. (1985). Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language & Speech*, **28**, 97-114.
- RECASENS, D. (1991). An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics*, **19**, 177-192.
- SHAW, R. E., & BRANSFORD, J. (1977). Introduction: Psychological approaches to the problem of knowledge. In R. E. Shaw & J. Bransford (Eds.), *Cognition and the symbolic processes* (pp. 1-39). Hillsdale, NJ: Erlbaum.
- SHAW, R. E., & TURVEY, M. T. (1981). Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 343-416). Hillsdale, NJ: Erlbaum.
- 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: Erlbaum.
- SUGA, N., O'NEILL, W. E., KUJIRAI, K., & MANABE, T. (1983). Specificity of combination sensitive neurons for processing of complex bisonar signals in auditory cortex of the mustached bat. *Neurophysiology*, **49**, 1573-1625.
- SUSSMAN, H. M. (1988). The neurogenesis of phonology. In H. A. Whitaker (Eds.), *Phonological processes and brain mechanisms* (pp. 1-23). New York: Springer-Verlag.
- SUSSMAN, H. M. (1989). Neural coding of relational invariance in speech perception: Human language analogs to the barn owl. *Psychological Review*, **96**, 631-642.
- SUSSMAN, H. M. (1991). The representation of stop consonants in three-dimensional acoustic space. *Phonetica*, **48**, 18-31.
- SUSSMAN, H. M. (1994). The phonological reality of locus equations across manner class distinctions: Preliminary observations. *Phonetica*, **51**, 119-131.
- SUSSMAN, H. M., FRUCHTER, D., & CABLE, A. (1995). Locus equations derived from compensatory articulation. *Journal of the Acoustical Society of America*, **97**, 3112-3124.
- SUSSMAN, H. M., FRUCHTER, D., HILBERT, J., & SIROSH, J. (in press). The orderly output constraint: A functional role for highly correlated, linearly related components in the speech signal. *Behavioral & Brain Sciences*.
- SUSSMAN, H. M., HOEMEKE, K. A., & AHMED, F. S. (1993). A cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. *Journal of the Acoustical Society of America*, **94**, 1256-1268.
- SUSSMAN, H. M., MCCAFFREY, H. A., & MATTHEWS, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, **90**, 1309-1325.
- SUSSMAN, H. M. & SHORE, J. (1996). Locus equations as phonetic descriptors of consonantal place of articulation. *Perception & Psychophysics*, **58**, 936-946.
- WAGNER, H., TAKAHASHI, T., & KONISHI, M. (1987). Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *Journal of Neuroscience*, **7**, 3105-3116.
- WHALEN, D. H. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Perception & Psychophysics*, **46**, 284-292.

NOTES

1. Sussman and Shore (1996), in contrast, reported that /d/ and /z/ did not differ significantly in slope or intercept. However, 14 of the 22 subjects reported by Sussman and Shore had a steeper slope for /d/ than for /z/, the direction of difference found by Fowler (1994), suggesting that the lack of significance is due to a lack of power. Sussman and Shore further found that a discriminant function analysis successfully distinguished /d/, /z/, and other coronals, taken as a group, from labials and velars. However, they did not test whether a discriminant function analysis would be able to distinguish between different consonants within the coronal class. In fact, when plotted in slope-by-intercept space, most of the /d/ and /z/ productions reported by Sussman and Shore fall in distinct regions, indicating that locus equation coefficients do distinguish consonants within a place class.

2. We saw no possible harm in using ourselves as subjects. Our aim was to vary coarticulation resistance and therefore locus equation slope as much as possible. The only way we as subjects could bias results in the intended direction was to have followed our own instructions especially well.

3. In order to address a reviewer's concern that the slope and intercept changes could have been due to changes in $F2_{\text{vowel}}$ as well as in $F2_{\text{ons}}$, we recomputed the slopes and intercepts for each speaker in two ways. First, we took the $F2_{\text{vowel}}$ values from the normal productions and regressed the $F2_{\text{ons}}$ values from the high and low productions on them, and second, we took the $F2_{\text{ons}}$ values from the normal condition and regressed them on the $F2_{\text{vowel}}$ values from the high and low conditions. To the extent that the changes in slope and intercept in the high and low conditions from the normal condition are independent from changes in $F2_{\text{vowel}}$, the recomputed locus equations using the normal-production $F2_{\text{vowel}}$ values should be similar to the true locus equations for the high and low conditions. Furthermore, to the extent that the slope and intercept changes are related to changes in $F2_{\text{ons}}$, the recomputed locus equations using the normal-production $F2_{\text{ons}}$ values should be more similar to the locus equations for the normal condition. In fact, both outcomes were obtained: in the first analysis, the mean slopes and intercepts in the high and low conditions (averaged across consonants) were .828 and .718, and 361.9 and 502.2 (the original means were .823, .693, 328.0, and 560.3); in the second analysis, the means were .710, .696, 510.9, and 569.2. Thus, there was little change in the first analysis, and substantial change in the second (for the high condition); moreover, in the second analysis, the recomputed means were very similar to those of the normal condition (.726 and 500.8). This indicates that the

observed slope and intercept changes due to the manipulations were, in fact, primarily a consequence of changes in $F2_{\text{ons}}$.

4. There is a straightforward algebraic proof of this. The intersection of two lines $y_1 = m_1x_1 + b_1$ and $y_2 = m_2x_2 + b_2$ is

$$\left(\frac{b_2 - b_1}{m_1 - m_2}, \frac{m_2 b_1 - m_1 b_2}{m_2 - m_1} \right).$$

The slope of the second-order line (formed by plotting the first-order slopes of these points on the x -axis and the first-order intercepts on the y -axis) is

$$\frac{b_2 - b_1}{m_2 - m_1},$$

and thus equals (-1) multiplied by the x -coordinate of the intersection of the two lines. The intercept of the second-order line can be determined by inserting the coordinates (x_1, y_1) into the general equation

$$y = \left(\frac{y_2 - y_1}{x_2 - x_1} \right) * x + b,$$

to equal

$$\frac{x_2 y_1 - x_1 y_2}{x_2 - x_1}.$$

Substitution of first-order slopes for x s and intercepts for y s reveals this to be identical to the y -coordinate of the intersection point.

5. Standard error of estimate (mean squared variance off the regression line) is a better basis for comparison than R^2 , because there is more variance in $F2_{\text{ons}}$ than there is in $F1_{\text{ons}}$ (covering a range of about 1400 Hz rather than about 500 Hz). A given amount of deviation off of a line for a shallow slope results in a lower R^2 than does the same amount of deviation for a steeper slope.

6. An exception is /g/, which uses the tongue body, but shows large changes in $F2_{\text{ons}}$ with changes in $F2_{\text{vowel}}$. /g/ may not resist coarticulation in a language such as English, because a shift in its place of articulation due to coarticulation with a vowel does not make it confusable with any other stop in the language. Although /d/ uses the tongue tip, not the tongue body, Recasens (1985) argues that the physical linkage of the tongue tip to the tongue body can cause the achievement of a tongue-tip gesture to constrain movement of the tongue body as well.

7. Sussman (e.g., Sussman et al., 1991) does acknowledge that locus equations cannot fully support consonant perception and suggests that the outputs of the $F2$ -processing system interact with outputs of other cue-processing systems (for $F3$ and burst). Thus, the model would not be expected to predict subjects' responses perfectly. However, because locus equations are the backbone of the model, and presumably the predominant cue, they should be expected to predict a large proportion of responses.

8. Only the normal and high-overlap tokens were used to limit the number of tokens subjects listened to, and because the low-overlap condition was not effective in changing the locus equation slopes. We chose tokens produced by Subject M2, who showed the greatest overall change in slopes from normal to high productions for the three consonants and thus offers the greatest degree of variability in token space.

9. Hintzman uses 3 as the exponent. We used higher power exponents, however, to reduce further the contributions of all but the closest matches. This should improve the performance of the model. However, recall that the exemplar model is still constrained to use the same information that the linear-template model uses. Our aim is to see whether a model can perform as well as these without any use of prototypes.

(Manuscript received May 13, 1996;
revision accepted for publication January 22, 1997.)