Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation

CAROL A. FOWLER
Haskins Laboratories, New Haven, Connecticut

This experiment explored the information for place of articulation provided by *locus equations*—equations for a line relating the second formant ($F2$) of a vowel at midpoint to $F2$ of the formant at consonant-vowel (CV) syllable onset. Locus equations cue place indirectly by quantifying directly the degree of coarticulatory overlap (*coarticulation resistance*) between consonant and vowel. Coarticulation resistance is correlated with place. The experiment tested predictions that when coarticulation resistance varies due to properties of the consonant other than place of articulation (in particular, due to manner of articulation), locus equations would not accurately reflect consonantal place of articulation. These predictions were confirmed. In addition, discriminant analyses, using locus equation variables as classifiers, were generally unsuccessful in classifying a set of consonants representing six different places of articulation. I conclude that locus equations are unlikely to provide useful place information to listeners.

In several publications (e.g., Sussman, 1989, 1991; Sussman, Hoemeke, & Ahmed, 1993; Sussman, McCaffrey, & Matthews, 1991), Sussman and colleagues have explored properties of "locus equations" as "relational invariants" signaling stop-consonant place of articulation. Locus equations, first described by Lindblom (1963), are equations for a line relating the second formant ($F2$) of a variety of vowels at their nearest approximations to their steady states (henceforth $F2_V$) to $F2$s of a given consonant at the acoustically defined syllable onset of a consonant-vowel (CV) sequence (henceforth $F2_O$ for $F2$ at syllable onset). Lindblom found good fits of lines to the data from one Swedish speaker when he plotted $F2_V$ against $F2_O$ pairs for a given consonant, /b/, /d/, or /g/, produced in the context of a variety of vowels. Slopes of the lines were distinct for each consonant; they were steepest for /g/ and flattest for /d/. Sussman and colleagues (see also Nearey & Shamass, 1987) have followed up this research using speakers of English. Sussman et al. (1991) found slopes varying between .73 and .97 for twenty speakers producing /b/, slopes between .27 and .50 for speakers producing /d/, and slopes between .54 and .97 for /g/, in the context of 10 different vowels. In group analyses, values of the slope for /b/, /d/, and /g/ were .91, .54, and .79, respectively. As compared with Lindblom's findings, the values for /b/ and /g/ are out of order. The difference

could be either a language difference or an individual difference; two of Sussman et al.'s 20 speakers also had higher slopes for /g/ than for /b/. In any case, in statistical analyses, slopes were distinct for all three consonants. Further, in discriminant analyses, slopes classified the three consonants by place with near-perfect accuracy, and slopes and intercepts classified them by place with perfect accuracy. Success was moderate when $F2_V$ and $F2_O$ were used as classifiers.

Sussman and colleagues (e.g., 1989; Sussman et al., 1991) have proposed locus equations as relational invariants that listeners might use to recover stop-consonantal place of articulation information from acoustic speech signals. Of course, the locus equations themselves, and even the acoustic information needed to construct them, are not given in an acoustic CV signal that a listener might hear. When a talker produces a CV, a single x, y pair (i.e., a single $F2_V$, $F2_O$ pair) is made available, not a line or an equation for a line. Sussman (1989) proposed a model, devised by analogy to findings on the neural system for sound localization in the barn owl (e.g., Wagner, Takahashi, & Konishi, 1987), in which the variety of $F2_V$, $F2_O$ pairs that a listener might receive when a given consonant-initial syllable is produced constitute a neural column or slab that effectively constitutes the locus line. Any $F2_V$, $F2_O$ pairing that a listener might recover will fall closest to one locus line and, in this way, place is uniquely determined. The discriminant analysis performed by Sussman et al. (1991) that comes closest to testing the feasibility of this account (by using $F2_V$ and $F2_O$ as classifying variables, rather than using slope and intercept) classified the consonants by place with a success rate of about 77%.

The purpose of the present paper is to explore the locus equations further in regard to the information they

might provide for place of articulation in general. This exploration extends the application of the locus equation beyond that intended by Sussman and colleagues. In most references to locus equations, Sussman and colleagues restrict the invariant's scope to one of signaling place of articulation for stop consonants only.[1] However, three lines of evidence, to be reviewed briefly below, convince me that place of articulation is not subsumed under manner features for the language user. Given that a viable candidate invariant for stop place of articulation must explain the structure of featural space as language users know it, and especially as they perceive it, this evidence suggests that any invariant for stop place must be an invariant for place of articulation in general.

Featural Space in Linguistic Competence

In linguistic theories of phonology until the mid 1970s, the features of a phonological segment were listed in a column. This notation ascribed no internal structure to features within a column. However, linguistic evidence has shown that features do have relational structure. With the advent of "autosegmental" phonologies (Goldsmith, 1976), investigators began to explore this structure and to seek ways of capturing it in representations that were different from that of the feature column. Based on a cross-linguistic investigation of featural dependence and independence in phonological processes (that is, an investigation of phonological processes or rules in which sets of features frequently, rarely, or never participate jointly), Clements (1985) proposed a hierarchical structure of features that bears remarkable similarity to the structure of the vocal tract itself. In the structure, features that never or rarely participate jointly in phonological processes separate early on in the structure and those that commonly participate separate late. For present purposes, the important observation is that, based entirely on evidence of featural dependence and independence, Clements represented place features *as a class* splitting from manner features *as a class* at a juncture that he called a "supralaryngeal tier."[2] Thus, evidence from the distribution of features across the entries in language users' lexicons indicates that, for language users, the places of articulation of stop consonants are not subsumed under the stop-manner features.

Normal language users have just two ways of acquiring this featural structure—by producing speech and by perceiving it. Deaf speakers and, complementarily, hearing persons who cannot speak each have just one of those two ways. Existing evidence suggests that information for a general place feature is available either way.

Language Production

In speech production, of course, consonants that are identified as having the same place feature have the same constriction location. That they share place in speech planning is suggested by evidence from spontaneous errors of speech production. Van den Broecke and Goldstein (1980) write, for example, "When we examine which fea-

tures are most often involved in a switch . . . place is by far most affected across all data sets" (p. 62). Place is involved, not place within distinct manner classes. Further, to my knowledge, no model of language production (e.g., Dell, 1986; Shattuck-Hufnagel, 1979) has found it useful in accounting for error patterns to propose stop-place features that are distinct from, say, fricative-place features. Thus, the place features that, via investigation of linguistic phonologies, have been ascribed to speakers' competence or knowledge are the same as those that speakers plan and utter.

Listening

The evidence regarding perception is compatible. If stop places of articulation were distinct from places of articulation for other manner classes of consonant—that is, if place features were subsumed under manner features— then mishearings in which manner is misheard while place is preserved should be rare. But they are not, apparently. For example, Bond and Garnes (1980) report errors in which manner and place mishearings both occur as single-feature mishearings, and they do not report either error type as rare.

Also relevant for present purposes, it is known in the perception literature that the same formant transitions that are used along with noise frication as information for /s/ are heard as /t/ when a stop-appropriate interval of silence is interposed between the /s/ frication and the vowel (Best, Morrongiello, & Robson, 1981). Accordingly, not only are the place features for alveolar fricatives and alveolar stops the same according to the three sources of evidence just reviewed, their perception is based in part on the same acoustic information in the signal—the very information that, in voiced stops, contributes with $F2v$ to the locus equation.

In short, in linguistic competence, place features cross manner features; they are not nested under manner. Language users can acquire that competence either by producing or by perceiving speech. In either domain, place information exhibits the same crossed, rather than nested, character in relation to manner.

Given the foregoing evidence, it is of interest to ask whether locus equations can serve as invariants for the psychologically real feature, "place of articulation," and not just for the stop-consonant places to which they have been applied. Before addressing that question, however, I consider the general concept of "invariant" itself as it is used, variously, in the literature. The aim here is to pin down the sense in which the locus equation is an invariant. I will suggest that it is an invariant in a very weak sense of the term. In particular, in other usages, invariants are "specifiers"; that is, they uniquely determine a property. However, as Sussman and colleagues understand, locus equations are not specifiers; rather, they are, at most, cues. They provide information for stop place, but not sufficient information to specify stop place. Next, I explore reasons why excellent fits are obtained when lines are fit to $F2v$, $F2o$ pairs and why /b/, /d/, and /g/

have distinctive slopes. This exploration leads to an experiment that extends work on the locus equation and further tests its adequacy as a source of information for place of articulation.

Invariants as Specifiers or Cues

Probably the best-known work on invariants for phonetic features in speech is the research by Stevens and Blumstein (1981). They have sought unique patterns in acoustic speech signals that are invariably present when their associated phonetic feature is present in any talker's utterance. Because a given pattern is invariably present when its feature is produced, and because the pattern is unique to the feature, it can specify the feature—that is, determine it uniquely. In the view of Stevens and Blumstein, specification is likely to mean that the perceptual system "responds in a distinctive way when a particular sound has these properties so that the process of decoding the sound into a representation in terms of distinctive features can be a fairly direct one" (p. 2). That is, invariants excite feature detectors (see also Stevens, 1989), and the relevant detected features map directly onto distinctive features of the language.

Stevens and Blumstein's (1981) original proposal for invariants for places of articulation—unique shapes of the spectrum near consonant release—has been shown not to be adequate to specify place (Lahiri, Gewirth, & Blumstein, 1984) and not to be as salient to listeners as other information for place in the signal (Blumstein, Isaacs, & Mertus, 1982; Walley & Carrell, 1983). However, their search for invariants as specifiers continues (e.g., Lahiri et al., 1984).

Stevens and Blumstein's (1981) use of the term "invariant" has some similarities to Gibson's more general usage (1979; Reed & Jones, 1982) that describes information in light, air, and other informational media that permit perceivers to know their world and to act effectively in it. For Gibson, as they are for Stevens and Blumstein, invariants are unique patterns in informational media that are invariably present when the property or event in the environment about which they provide information is present. Further, because the pattern is invariably present in the presence of its corresponding property in the environment, and because it is unique to it, it specifies the property.

Gibson's (1979) concept of invariant incorporates an explanation of why we should expect to find invariants as specifiers in informational media: Invariants are invariably present in the presence of their corresponding properties or events, because they are lawfully caused by the properties or events, not merely associated with them. Properties and events in the environment of the perceiver are causal sources of patterning in light and air, and these caused patterns serve as specifiers. Because distinct properties or events are likely to structure the light and air distinctively, caused structure in light and air can provide sufficient information for their causes; that is, they can specify them.

One reason why Gibson's (1979) general theory of perception has seemed unlikely to serve as a useful basis on which to develop a theory of speech perception (see Fowler, 1986 and the following commentary) is that the phonological or phonetic primitives that a listener is presumed to recover—phonetic features, for example—are generally not considered to cause, in an unmediated way, the structuring of the acoustic speech signal. That is, features are believed to be in the mind of the talker, and not to be transparently reflected in the vocal-tract actions that do causally structure the signal (e.g., Hammarberg, 1976; Pierrehumbert, 1990; Repp, 1981). Both in the field of linguistics, however (e.g., Browman & Goldstein, 1990), and in psychology (Fowler, Rubin, Remez, & Turvey, 1980; Saltzman & Kelso, 1987), theories are under development in which phonological primitives are presumed to be gestures of the vocal tract. This kind of development allows for the possibility that the phonological primitives of speakers' utterances do immediately structure the acoustic signal and hence may be specified by acoustic invariants in Gibson's sense of the term.

In the context of the foregoing characterization of invariants, in what sense are locus equations invariants? The answer is that they are invariants in a very limited sense. They are not in the acoustic signal when, say, CVs are produced. That is, as noted, a CV only provides a pair of points; locus equations are not in the signal at all. Rather, they reflect relationships that are, it is claimed, in the head of a perceiver. The second sense in which locus equations are not invariants is that they do not specify place of articulation. They provide information for place, but they do not provide sufficient information. That is to say, locus equations are, at most, cues for place ("at most" because experiments on their use by listeners have not yet been reported). Sussman and colleagues are aware that locus equations are not specifiers; they (e.g., Sussman, 1991) are actively exploring additional information (e.g., $F3$ at syllable onset) that will improve their classification of /b/-, /d/-, and /g/-initial utterances by place of articulation using information in the acoustic signal that is available to a listener and information in the head that has been acquired. Accordingly, the force of the foregoing is to suggest not that these researchers are claiming more for locus equations than the equations can provide, but rather, that the term "invariant" has been used in the literature, both in speech perception and in perception generally, with a considerably stronger meaning than Sussman and colleagues intend. The only sense in which the locus equation is an invariant is that, in the judgment of Sussman et al. (e.g., 1991), the slope does not vary much across speakers for a given consonant, and values for the consonants differ significantly.

Why Do Locus Equations Provide
Any Information for Place?

It is interesting and informative to consider why $F2$ at consonant release ($F2_O$) is a positive function of $F2$ at vowel midpoint ($F2_V$) and why the slopes for /b/, /d/,

and /g/ differ in magnitude. The functions have a positive slope, because talkers coarticulate—that is, they overlap the production of serially ordered consonants and vowels. Accordingly, if a vowel has a high $F2$, $F2$ will also be relatively high at the acoustic onset of the syllable, because vowel production began before consonant release, and vowel production affects the acoustic signal at release. If a vowel has a low $F2$, $F2$ will be low at acoustic-syllable onset for the same reason. Therefore, $F2_V$, $F2_O$ points tend to fall on a line with positive slope. The magnitude of the locus equation slope reflects the extent of consonant-vowel overlap (see also Duez, 1992; Krull, 1989; Sussman et al., 1993). Phonetic segments differ in the extent to which they resist overlap by neighboring segments, an effect called "coarticulation resistance" by Bladon and Al-Bamerni (1976). In research on Catalan Spanish, Recasens (e.g., Recasens, 1984) has shown, for example, that four consonants characterized by different degrees of contact between the tongue dorsum and the palate—contact decreases progressively in the series dorsopalatal /j/, alveopalatal /ɲ/ and /y/, and alveolar /n/—exhibited less coarticulatory influence from preceding and following vowels the greater the degree of tongue dorsum contact by the consonant. More generally and intuitively, consonants that use the same main articulator as a neighboring vowel do not permit the vowel to pull the tongue very far away from the consonant's characteristic locus of constriction.

In findings by Sussman et al. (1991), /b/ had the steepest slope. Production of /b/ does not involve the tongue at all; accordingly, the vowel has considerable freedom to coarticulate with /b/. The consonant /d/ had the shallowest slope; it is produced by the tongue blade, forward of the dorsum. The rather steep slope for /g/ is somewhat surprising; however, there is considerable data showing that velar stops do not resist coarticulatory overlap very much—and, indeed, that they do not strongly resist being pulled from their place of articulation—by vowels (e.g., Öhman, 1967; Perkell, 1969). The reason for this may be that in Swedish and English, the languages studied, there are no similar consonants with places close to /g/ and /k/. Accordingly, talkers can permit considerable /g/- and /k/-vowel coarticulation, because listeners will not be confused by, say, a /g/ that has been shifted in its place by a front or back vowel. They will not be confused because there is no consonant with the place of articulation of fronted or backed /g/.[3] (However, see the analysis of front and back /g/s below for a qualification of this explanation.)

A conclusion based on these considerations is that the slope of the locus equation reflects coarticulation resistance more or less directly. It provides information for place indirectly, only because variation in place of articulation is an important source of variation in coarticulation resistance. In turn, this suggests that a reason that locus equations can only serve as cues, and not as specifiers, is that variables other than place of articulation affect coarticulation resistance. That consideration led to the design of the experiment next to be described.

Whether a source of information for place is a cue or a specifier, it should signal place of articulation as talkers produce it (e.g., van den Broecke & Goldstein, 1980) and as listeners perceive it (e.g., Greenberg & Jenkins, 1964; Singh, Woods, & Becker, 1972). That is, /d/, /t/, /z/, and /s/ should be associated with the same locus equation because they have, and are perceived to have, the same place of articulation; however, /g/ and /v/, for example, should be associated with different locus equations because they have, and are perceived to have, very different places of articulation. But these expected outcomes will hold only if differences between consonants other than place—in particular, voicing and manner differences between consonants—do not affect coarticulation resistance and hence do not affect the slope of the locus equation lines. However, it is quite likely that stop and fricative manners of articulation will be associated with different coarticulation resistances (cf. Recasens, 1989). The reason is that the articulatory requirements for producing fricatives are considerably more delicate than they are for producing stops. To produce a stop, all that is required is that a complete constriction be made in the vocal tract. To produce /d/ or /t/, for example, the tongue blade must stop the flow of air from the lungs by making a constriction on the alveolar ridge of the palate. The constriction can be made with more force than necessary with no obvious acoustic consequences. However, producing /s/ requires that talkers create a narrow channel for airflow between the raised tongue blade and the alveolar ridge of the palate. If the constriction is too narrow, the resulting consonant will be a stop, not a fricative. If the constriction is too wide, airflow will not become turbulent, and the result will be a vowel.

This reasoning led me to predict, specifically, that fricatives will have shallower slopes than stops at the same place of articulation. More vaguely, it also seemed possible that a stop and a fricative having different places of articulation might have the same slope, the one because its coarticulation resistance derived from its place of articulation and the other because about the same resistance was associated with its particular manner of articulation.

METHOD

Subjects

The subjects were twelve students at Dartmouth College. Equipment failure led to poor recordings from two subjects; that left data from five males and five females. The subjects participated for course credit, and all were native speakers of American English.

Materials and Procedure

The materials consisted of word and nonword CVt sequences. The first consonant of each monosyllable was one of /b/, /v/, /ð/, /d/, /z/, /ž/, and /g/. Each initial consonant was followed by each of the eight vowels /iy/, /I/, /ey/, /æ/, /ʌ/, /a/, /ɔ/, and /uw/. Letter sequences representing each CVt monosyllable were printed in large type on two sheets of paper. The subjects were instructed to read each monosyllable once. If the experimenter indicated that his or her pronunciation was the intended one, then the subject went on to produce five more tokens of the syllable. If the pronuncia-

tion was not as intended by the experimenter, the subject was given the expected pronunciation, which he or she repeated once and then produced five more times, or until there were five useable tokens of each utterance, mispronunciations and stuttered productions being excluded. Utterances were recorded on audiotape in a sound-attenuating chamber.

Acoustic measurements were taken from five tokens of each CVt produced by each talker, using procedures similar to those described by Sussman et al. (1991). Syllables were filtered at 10 kHz, digitized at 20 kHz, and input to MacSpeechLab II (GW Instruments) on a Macintosh computer. Measures of $F2$ were taken at vowel onset ($F2_O$) and at vowel "midpoint" ($F2_V$). Vowel onset was identified as the onset of voicing in the second formant of the vowel following consonant release. Following Sussman et al. (1991), the midpoint was the vowel steady state, if any; for U- or inverted U-shaped formant patterns, it was the value at the trough or peak. If there was a monotonic change in frequency, the formant value at the temporal midpoint of the vowel was selected. Other measures that were taken were $F3$ at syllable onset, the acoustically defined duration of the vowel, and the temporal point (msec from $F2_O$) at which vowel midpoint was measured. However, only $F2$ measures will be discussed here. Spectral measures were taken initially from wide-band spectrographic displays. Duplicate measurements were taken from linear predictive coding analyses of the syllables (using 25 coefficients; analysis window, 25 msec). If these were discrepant by more than 50 Hz, fast Fourier transforms were also examined, and the median of the three measures was taken as the value to be used in data analysis.

$F2$ onsets were easy to identify when syllables began with a stop. For /z/ and /ž/, vowel onset was identified as the first pitch pulse after offset of visible high-frequency frication in the spectrographic display. In /v/- and /ð/-initial syllables, occasionally the waveform revealed frication in pitch pulses following consonant release. In those cases, visible evidence of frication ending in vocalic pitch pulses was relied upon.

RESULTS AND DISCUSSION

I begin by examining the conditions that overlap with those examined by Sussman et al. (1991)—that is, syllables beginning with /b/, /d/, and /g/. Following that, all conditions are examined. (Mean $F2_O$s and $F2_V$s and their standard errors are presented in the appendix.)

Stop Consonants

Figure 1(a–c) presents the data on /b/, /d/, and /g/ in a manner similar to figures in Sussman et al. (1991), in that they are averaged across talkers. The eight points plotted in each figure represent $F2_V$, $F2_O$ pairs for each of the eight vowels averaged across talkers. In these regressions, slopes for /b/, /d/, and /g/ are, respectively, .8 ($R^2 = .97$), .48 ($R^2 = .95$), and .71 ($R^2 = .93$). These compare with values of .91, .54, and .79 in the analogous group data of Sussman et al. (from their Figure 2). These findings replicate those of Sussman et al. in their ordinal patterning, although the lines have flatter slopes overall than in the earlier study.

The good fits of the points in Figure 1(a–c) may give a misleading picture of the data from the perspective of a listener, however. In the figures, points are averaged over talkers and tokens. But, of course, listeners do not hear averaged data. Figure 2 presents the data token by token, with lines for /b/, /d/, and /g/ presented on the
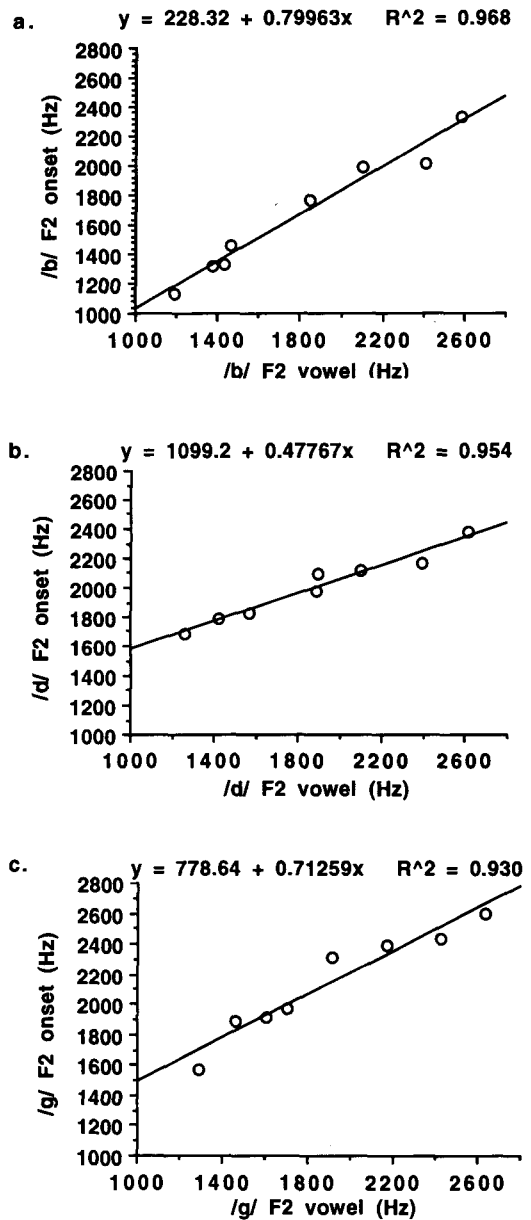

Figure 1. Group scattergrams and regression lines for the consonants /b/, /d/, and /g/ averaged over subjects.

same axes. Here, it is apparent that there is considerable spread about each line and there is the chance that points may not always fall closest to their own regression line.

One way to look at the distinctiveness of the slopes and intercepts of the three consonants is to fit regression lines to the data of individual subjects and to test the statistical significance of slope and intercept differences between consonant pairs. (Notice that this procedure can give estimates of the slope, intercept, and fit that differ somewhat from those yielded by procedures that yielded data in Figures 1 and 2.) Average slopes across subjects are .8, .47, and .68 for /b/, /d/, and /g/, respectively. In this
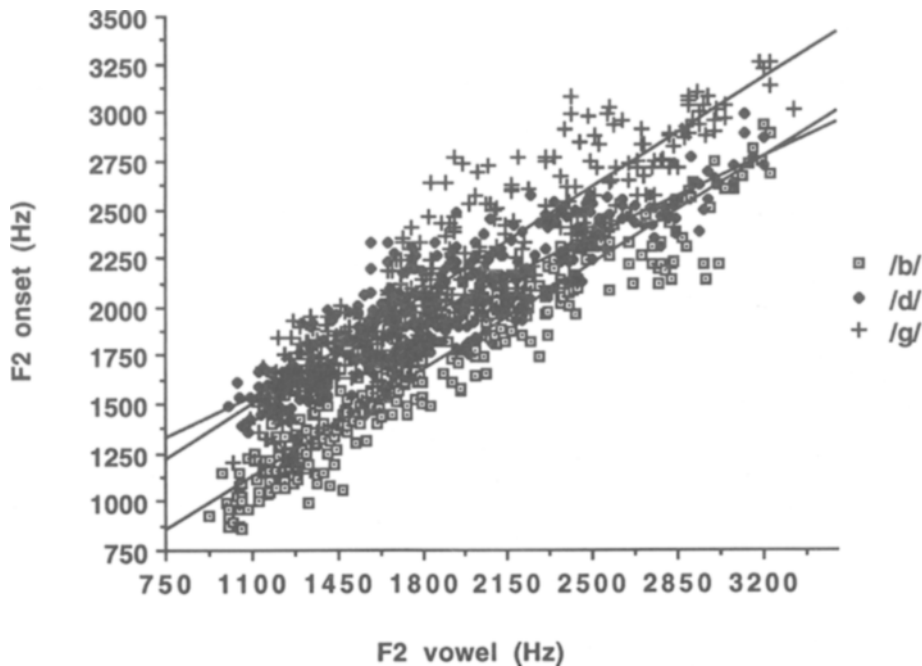
Figure 2. Scattergrams and regression lines for /b/, /d/, and /g/ unaveraged.

analysis, slopes for /b/ and /g/ differ only marginally [$t(9) = 1.98$, $p = .08$, two tailed]; differences between the other pairs are significant [/b/-/d/: $t(9) = 13.02$, $p < .0001$; /d/-/g/: $t(9) = 6.01$, $p = .0002$]. Slopes and intercepts are highly correlated ($r = -.90$ across the 70 pairs of values; that is, 7 consonants by 10 subjects). Accordingly, the intercept cannot add much predictive information beyond information already provided by the slope. However, in this case, it does distinguish /b/ from /g/ [$t(9) = 5.76$, $p = .003$].

For the most part, the present data pattern like those of Sussman et al. (1991). The major exception so far is that /b/ and /g/ differ only marginally in slope. Another finding from the earlier study that is replicated here is that the slope of the locus equation for /g/ is considerably flatter in the context of front as compared with back vowels. (This can be seen in Figure 1c, where the upper four points represent the four front vowels.) There is a tendency for a similar pattern in syllables with initial /b/ and /d/, but it is considerably weaker. Figure 3 plots values of the slope for the three consonants in the context of front as compared with back vowels. In an analysis of variance (ANOVA) with factors of consonant (/b/, /d/, /g/) and vowel context (front, back), both main effects and the interaction are significant [consonant: $F(2,18) = 5.67$, $p = .01$; vowel context: $F(1,9) = 11.76$, $p = .0075$; interaction: $F(2,18) = 8.92$, $p = .002$]. The significant interaction reflects the fact, evident in Figure 3, that, whereas all consonants have lower slopes in the context of front vowels than they do in the context of back vowels, the effect is especially marked for /g/. Possibly,

/g/ is more resistant to coarticulation by front vowels than it is to that by back vowels, because front vowels can pull its place close to that of more front consonants. However, there are no consonants in English with places of articulation farther back than /g/; accordingly, it may be pulled back without causing confusion.

A striking aspect of the outcome of this analysis is that in the context of front vowels, /g/ has numerically the lowest of the three slopes, whereas in the context of back vowels, it has the highest. Further, /g/'s slope in the con-
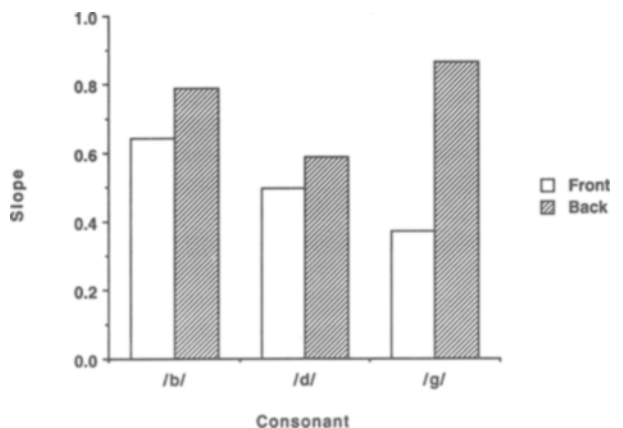

Figure 3. Slopes of regression lines for /b/, /d/, and /g/ produced in the context of front and back vowels.

text of back vowels is not distinct statistically from /b/'s slope in the same context [$t(9) = .45$, $p = .66$].

A question arises as to whether, given the marked slope differences for /g/ in different contexts, listeners should be thought to use separate locus equations for /g/ in the context of front and back vowels. Several considerations suggest not. One is that both in the present data and in that of Sussman et al. (1991), slopes for /g/ in the context of back vowels are not distinguished from those for /b/. Second, the separate regression lines do not appear to improve the fit of data points to locus equation lines. Across subjects, R^2 averages .74 and .71 for front and back vowel contexts, respectively. This value may not be comparable to the value computed over the whole set of /g/-initial syllables, because the latter analysis is based on twice as many data points as the former. However, elimination of half the data (i.e., data from half of the subjects) from the analysis that includes all eight vowels yields an R^2 of .88. A more important consideration suggesting that only one locus equation should be computed for /g/ is that there is no evidence that listeners hear distinct consonants; indeed, there is evidence that they do not. The consonants are transcribed the same way phonetically, and they are spelled the same way orthographically. Further, there is evidence that listeners factor fronting and backing coarticulatory effects in their perception of /g/ (Mann, 1980, 1986). Accordingly, it appears most reasonable to suppose that listeners construct one locus equation (if any) for /g/. In the following analyses, front and back vowel contexts are not considered further.

The next step in the analysis was to perform discriminant analyses to determine whether the variables, $F2_V$ and $F2_O$, can be used successfully to classify consonants by place of articulation. Sussman et al. (1991) report discriminant analyses performed separately on male and female talkers on grounds that "the discriminant function analysis should not be expected to derive a normalization factor to relate gender groups" (p. 1319). Whether or not the procedure is realistic from the point of view of perception, success in classifying syllables by place of articulation is considerably greater if data from males and females are analyzed separately. The procedure of Sussman et al. was followed here.

In the discriminant analysis, $F2_V$, $F2_O$ pairs were used to classify consonants as /b/, /d/, or /g/. A jackknife procedure was used; that is, data from four of the five subjects in each group were used to establish an algorithm for classifying the consonants, and then the algorithm was applied to data from the fifth subject. The procedure was repeated for each subject. Table 1 shows the data for males and for females. On average, 70% of classifications were correct for males, compared with 62.5% for females.

The classifications are not as successful as those of the data from male and female subjects in the analyses of Sussman et al. (1991) (78% and 76%, respectively). Most likely, the reason for this is that in the present analyses, data from only four subjects could be used to generate

Table 1
Results of Discriminant Analyses Using $F2_V$ and $F2_O$
to Classify Consonants /b/, /d/, and /g/ into Place Categories

Consonant	Percent Correct	Number of Cases Classified in Each Cell		
		/b/	/d/	/g/
Male Subjects				
/b/	88.5	177	12	11
/d/	55	31	110	59
/g/	66.5	21	46	133
Mean	70.0			
Female Subjects				
/b/	83	166	34	0
/d/	44.5	36	89	75
/g/	60.0	18	62	120
Mean	62.5			

Note—Values in bold represent correct classifications by place.

classification algorithms in the discriminant analyses, whereas in the study by Sussman et al., data from nine subjects was used. In any case, results are qualitatively quite similar except that, in the present data, relatively more /d/s were classified as /g/ than in the analyses of Sussman et al. The results of both sets of analyses make it clear, however, that $F2_V$ and $F2_O$ pairs cannot specify place; at most, the values can serve jointly as cues.

In the theory proposed by Sussman (1989), locus lines reflect a relationship that is encoded in the brain, and listeners associate $F2_V$ and $F2_O$ pairs from spoken syllables with the appropriate neural slab, represented by a locus line, to extract place information from the signal. Further analyses of the present data attempted to test this idea directly. In these analyses, data from four subjects within a gender group were used to estimate the locus equations for /b/, /d/, and /g/. These were assumed to compose the relationship between $F2_V$ and $F2_O$ in memory. Then $F2_V$, $F2_O$ pairs from the remaining subject were input and the line they fell closest to was computed. The place of articulation associated with that line was considered the model's place identification, and accuracy was assessed. This was done separately for each subject in the experiment. Distance from a line was computed using both euclidean and city block distances; however, there was almost no difference between the two in terms of classification success.

Results from the euclidean-distance version of the model follow. For males, 70.2% of points fell closest to the line representing the place of articulation of the spoken syllable (81%, 60%, and 70% for /b/, /d/, and /g/, respectively); for females, 66.3% of classifications were accurate (83%, 43%, and 70% for /b/, /d/, and /g/, respectively). Thus, results were not very different from those of the discriminant analyses.

All Seven Consonants

Locus equation slope values, intercepts, and R^2s (computed separately for each subject and averaged) are provided in Table 2 for the set of seven consonants. Individ-

ual linear fits to the data of all 10 subjects ranged from a low of $R^2 = .76$ for the infrequent $/\check{z}/$ to $R^2 = .93$ for $/b/$. In ANOVAs on the slope values, the effect of consonant was highly significant $[F(6,54) = 50.27, p = .0001]$. In post hoc tests, most slope differences were significant. I had predicted that $/z/$ and $/d/$, having the same place of articulation but different coarticulation resistances, would have different slopes. This occurred. The slope for $/z/$, .42, was significantly shallower than that for $/d/$, .47 $[t(9) = 3.09, p = .01]$. This probably reflects the extra resistance to coarticulation by a fricative as compared with that of a stop. I also speculated, without having any particular cases in mind, that a stop and a fricative, having different places of articulation but the same resistance to coarticulation, might have the same slope values. This may have occurred as well. The consonants $/g/$ and $/v/$ had statistically equivalent slopes $[t(9) = .8, p = .45$, with slope values .68 and .73, respectively]. Similarly, $/\delta/$ and $/d/$ had indistinguishable slopes $[t(9) = 1.29, p = .23$, with values .50 and .47, respectively]. The intercepts in these problematic cases pattern as they should, however. $/g/$ is distinct from $/v/$ $[t(9) = 4.02, p = .003]$ and $/d/$ is distinct from $/\delta/$ $[t(9) = 6.38, p = .001]$, but it does not differ from $/z/$ $[t(9) = 1.50, p = .17]$.

Discriminant analyses using the jackknife procedure were used to classify the seven consonants by place separately for male and female subjects. Table 3 presents the results of those analyses. They were generally unsuccessful. Since the classification was by place, classifications of $/z/$ as $/d/$ and vice versa were counted as correct. As the table shows, classifications were correct for males about 40% of the time (chance is 18.4%—that is, 1/7 for rows other than $/d/$ and $/z/$; 2/7 for $/d/$ and $/z/$). For females, the value is 33.3%. Most of the time, the place category receiving the modal number of classifications was the correct category. However, in both analyses, many more incorrect than correct classifications were made. Analyses restricted just to the set of fricatives were only a little more successful: for males, performance averaged 56%, and for females, 48% (chance is 25%).

Table 2
Slope Values Sorted by Magnitude, Intercepts, and Fits, with Significance Levels of Interesting Comparisons Marked

Consonant	Slope		Intercept		R^2
$/b/$	.80		244.3		.93
		$p = .001$		$p = .005$	
$/v/$	.73		336.9		.92
		$p = .42$		$p = .003$	
$/g/$	.68		815.5		.84
$/\delta/$	.50		902.8		.87
		$p = .23$		$p = .001$	
$/d/$	.47		1120.4		.85
		$p = .01$		$p = .17$	
$/z/$	.42		1078.1		.84
		$p = .03$			
$/\check{z}/$	.34		1407.8		.76

Table 3
Outcome of the Discriminant Analyses on the Seven Consonants of the Experiment

Consonant	Percent Correct	Number of Cases Classified in Each Cell						
		$/b/$	$/v/$	$/\delta/$	$/d/$	$/z/$	$/\check{z}/$	$/g/$
			Male Subjects					
$/b/$	10.5	**21**	107	23	26	20	1	2
$/v/$	55.5	25	**111**	32	11	18	0	3
$/\delta/$	30.0	28	32	**60**	22	53	2	3
$/d/$	57.0	3	0	13	**72**	42	46	24
$/z/$	52.0	17	23	35	39	**65**	16	5
$/\check{z}/$	15.0	1	0	4	62	19	**30**	84
$/g/$	59.5	1	0	1	34	4	41	**119**
Mean	39.9							
			Female Subjects					
$/b/$	12.0	**24**	108	22	14	32	0	0
$/v/$	50.0	42	**100**	30	7	21	0	0
$/\delta/$	23.0	31	28	**46**	19	42	24	10
$/d/$	40.5	5	2	7	**37**	44	46	59
$/z/$	32.5	33	9	59	23	**42**	24	10
$/\check{z}/$	16.5	7	1	10	39	41	**33**	69
$/g/$	59.0	0	13	11	23	14	21	**118**
Mean	33.3							

Note—Values in bold represent correct classifications by place.

It is clear from the table that although the discriminant function is given the values $F2_V$ and $F2_O$ as the bases on which to make its classifications, it is not constructing an algorithm that has much in common with the locus equations. The consonants $/b/$ and $/v/$ had reliably distinct slopes and intercepts, but $/b/$ is classified as $/v/$ more often than it is classified as $/b/$. The consonants $/v/$ and $/g/$, which had the same slope, are rarely confused in the discriminant analysis. Finally, $/\delta/$ had the same slope as $/d/$, but it is confused with $/d/$ less than it is with $/z/$.

Again, distance models were tried in which $F2_V$, $F2_O$ pairs, based on data from four of the five subjects in a gender group, were used to estimate locus lines. Data from the fifth subject were used to test the extent to which $F2_V$, $F2_O$ pairs from a new subject would fall closest to the appropriate line. As before, this procedure was repeated 10 times, once for each subject. With euclidean distances, the success rate was 44% for males and 36% for females. Success rates were slightly lower using city block distances.

Figure 4 reveals the difficulties encountered by these classification schemes. There is considerable overlap in the points for different consonants. It is frequently the case (56%–64% of the time, according to the last analysis reported) that a given $F2_V$, $F2_O$ pair falls closest to the wrong locus line. Figure 5 provides scattergrams individually for each consonant.

GENERAL DISCUSSION

Locus Equations

Sussman's (1989) proposal that locus equations may serve a role in perception of stop place of articulation in speech perception rests, in part, on an analogy between perception of place in humans and sound localization in
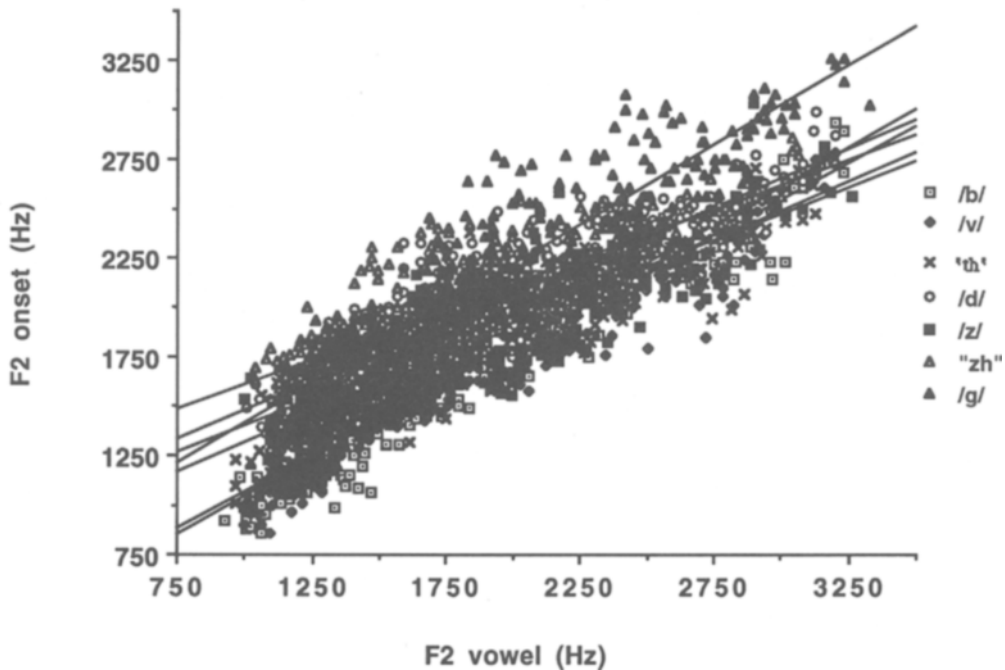
Figure 4. Scattergrams and regression lines for the set of seven consonants unaveraged.

the barn owl. Research on sound localization in the barn owl reveals that information about the difference in phase between given frequency components of a signal received at the two ears, itself ambiguous as to location, combines in the barn owl's brain with information about the component frequencies themselves having those phase differences to jointly specify location in space. Sussman's rather clever proposal by analogy was that $F2_O$, ambiguous as to place, may combine in the human brain with $F2_V$ to signal place of articulation. The present results, as well as those of Sussman et al. (1991), suggest an important difference between use of phase differences and frequency in the barn owl and use of $F2_V$ and $F2_O$ in humans. In particular, whereas phase difference × frequency pairs *specify* location in space, $F2_V$ × $F2_O$ pairs do not specify place of articulation or even stop place of articulation. The present findings extend earlier ones, moreover, in showing, in two ways, that the locus equations provide poor information for place. First, in the data averaged over subjects, the patterning of significant differences in slope and intercept is not according to shared and different places of articulation; most seriously, /z/ and /d/, which share place of articulation, have distinct locus equations. Second, in the data considered as the listener must deal with them—token by token—there is sufficient spread of points around each line and overlap in points for different consonants, that analyses designed to classify consonants by place fare poorly. A reason for the poor fits is that locus equations do not reflect place of articulation

immediately, but, perhaps among other things, represent a variable correlated with place—coarticulation resistance.

Nothing in the foregoing discussion forces the conclusion that $F2_V$, $F2_O$ pairs serve no role in perception of place. The discriminant and distance analyses both sort the seven consonants into place categories with better-than-chance accuracy. To date, no direct tests of the perceptual use of $F2_V$, $F2_O$ pairs have been reported. However, there is evidence that place can be perceived without the pairs being available. For example, Blumstein and Stevens (1980) synthesized the initial portions of CV stimuli—truncated syllables ranging in duration from 10 to 46 msec—and synthesized the syllables to test their hypothesis that place of articulation for stop consonants is specified in a patterning of spectral energy in the close vicinity of stop release. When stimuli included the stop burst, identification of the consonant as /b/, /d/, or /g/ (in the context of /i/, /a/, and /u/) was well above chance, even for the 10 msec stimuli. The importance of the finding in the present context is that there was no $F2_V$ in these truncated syllables. Accordingly, $F2_V$ is not needed (and therefore neither are the relations captured by locus equations), at least to classify consonants as /b/, /d/, or /g/.

The foregoing is not to suggest that locus equations themselves are uninteresting. In fact, they appear to provide an interesting metric for studying the extent of coarticulatory overlap, which may vary not only with coarticulation resistance due to place and manner of articulation, but also, for example, with stress (Krull, 1989) and speak-
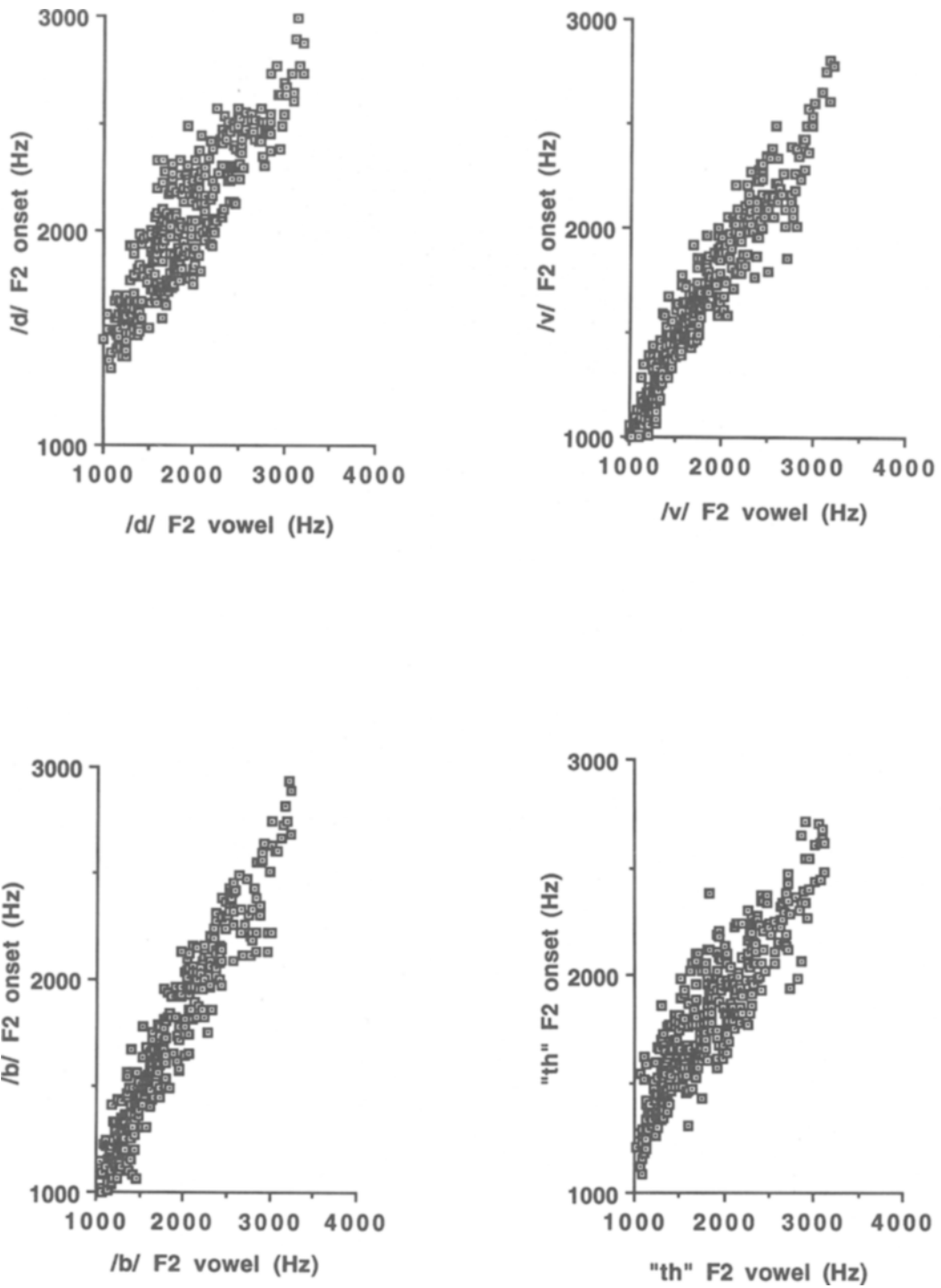
Figure 5. (Above and facing page) Scattergrams for the seven consonants presented separately.
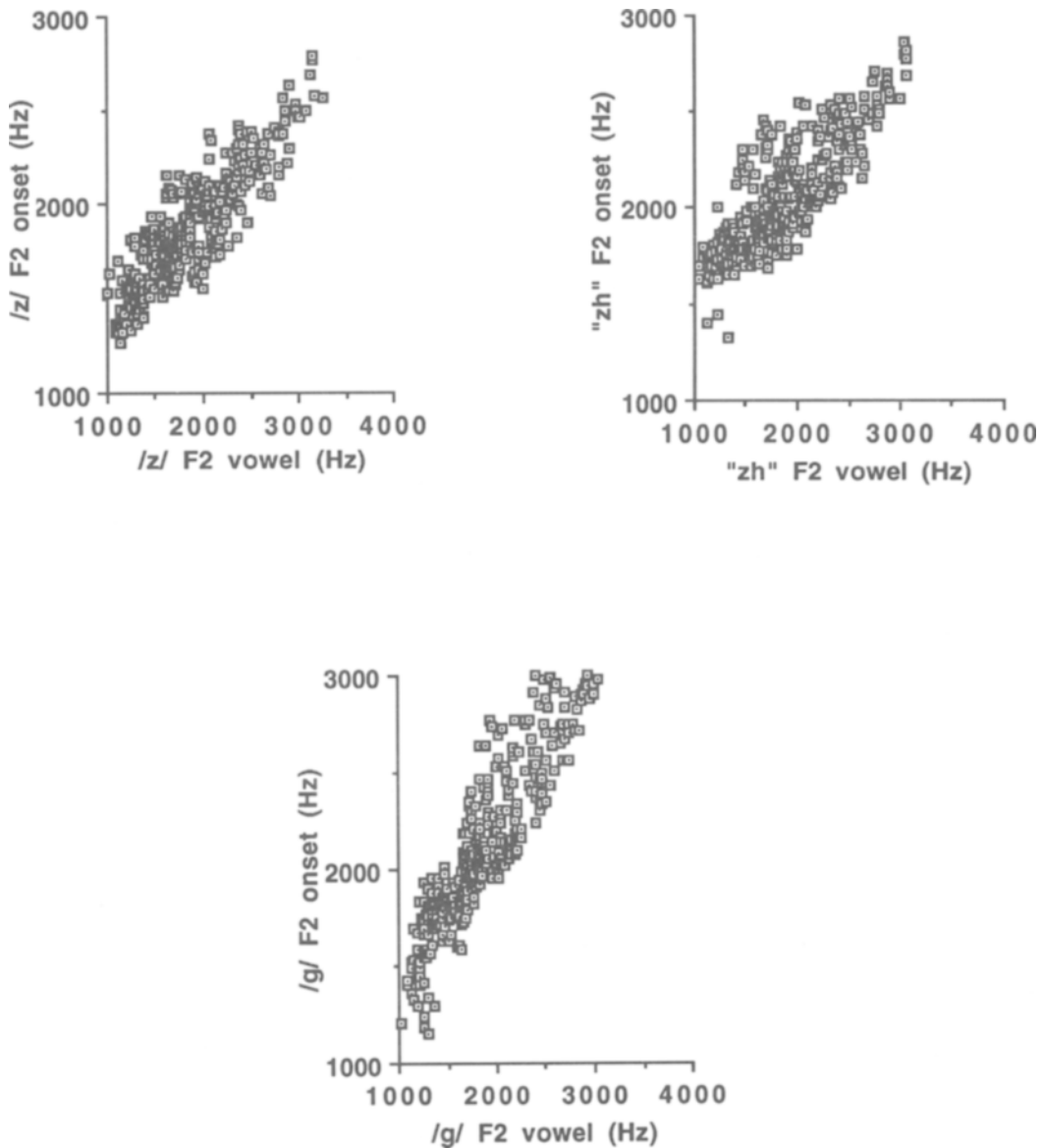
Figure 5 (continued).

ing style (Duez, 1992). It is only to suggest that the equations do not provide sufficient information for place of articulation. The information for place must classify consonants by place as listeners do, and listeners have to be shown to detect and use the information.

What's Left?

Four general points of view regarding place perception remain, two holding that invariants in the signal specify place, and the other two—one new to the literature here—holding that there are no invariants.

1. Invariants that specify phonetic features. One view ascribed in the introduction to Stevens and Blumstein (1981; see also Stevens, 1989) holds that invariant

information in the acoustic speech signal specific to each phonetic feature will eventually be found. Features themselves are components of a linguistic message, not necessarily of the articulatory actions that produce the signal. Therefore, invariants in the signal will map onto representations of the features in the head of the listener. Although the original idea that place of articulation is specified by spectral cross-sections has not proven viable, the search continues. Kewley-Port (1981) found that information conveyed over time by "running spectra" (successive spectral sections in a CV syllable) were more successful than Stevens and Blumstein's static spectra in classifying the stop consonants by place of articulation. Blumstein and colleagues (e.g., Lahiri et al., 1984) have

also sought information conveyed over time following stop release. To date, wholly adequate specifiers for place have not been found.

2. Invariants for phonetic gestures. In this second view, developed by extrapolation from Gibson's (1979) theory of direct perception (e.g., Fowler, 1986; Fowler & Rosenblum, 1991), there are invariants for phonetic properties of an utterance. In order for there to be invariants, in the theory, the phonetic properties must be the immediate physical causes of the invariants that specify them. That is, the phonetic properties must be intrinsic to the articulatory actions that implement spoken utterances.

In the theory of articulatory phonology of Browman and Goldstein (1986, 1992), "phonetic gestures" of the vocal tract replace phonetic features as primitive components of a linguistic message. Given that point of view, it is possible (but it is not a claim of articulatory phonology) that each phonetic primitive causes a patterning in the acoustic signal that is unique to it and that, therefore, specifies it.

Opposed to that possibility, there remains the problem that coarticulation changes the particular way that consonants and vowels are produced and makes the acoustic speech signal context sensitive and lacking invariants for phonetic gestures. There may be a solution to this problem as well. Phonetic gestures appear to be achieved by synergies of the vocal tract (e.g., Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984)—that is, organized relations among articulators that permit invariant achievement of a gesture's macroscopic goal in context-sensitive (equifinal) ways. For example, a bilabial constriction for /b/, /p/, or /m/ is achieved by an organization among the jaw and the two lips that permits closure at the lips to be achieved with differential contributions of the three articulators. If the jaw is relatively low, even unexpectedly low due to an unanticipated external perturbation to the jaw (Kelso et al., 1984), the lips contribute relatively more to closure, and closure is achieved. If the jaw is relatively closed, the lips contribute less to closure. If, in unperturbed speech, a low vowel, such as /a/, coarticulates with /b/, causing the jaw to lower during closing for /b/, the lips will contribute more to closure than if a high vowel coarticulates with /b/ with an associated higher position of the jaw. Synergies may be seen as implementing selective degrees of resistance against coarticulation—high resistance to any coarticulating gestures that would prevent achievement of its macroscopic goal and lower resistance to gestures that permit equifinal goal achievement.

Because gestural goals are achieved in the vocal tract in this view, something invariant occurs in articulation when a gesture with a given place is produced—namely, constriction at the required place (which, in the case of /g/, for example, may be a region). Given this, there is the possibility that the invariance in articulation causes invariance in the acoustic signal. The acoustic invariants are, as yet, undiscovered, perhaps just because, to date, there is little research seeking invariant acoustic consequences for invariant gestural causes.

3. There are no invariants and no specifiers in the acoustic speech signal. Most theories of speech perception fall into this category. In these theories—that include views as otherwise diverse as Massaro's fuzzy logical model of perception (Massaro, 1987) and Liberman and Mattingly's motor theory (Liberman & Mattingly, 1985)—the signal generally does not specify phonetic properties, and it does not include invariants. For Massaro, syllable prototypes in memory are associated with the constellation of acoustic and optical cues for the syllable that can occur when the syllable is produced. However, generally only a subset of the cues is available to a perceiver. To choose one syllable prototype over others as the syllable type produced by the talker, it is necessary only that the cues from the signal be most consistent with those of just that one syllable prototype.

In the motor theory, the signal cannot specify phonetic properties of the intended message, because coarticulation causes "encoding" (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) of the information for successive consonants and vowels in the acoustic speech signal. (According to Liberman & Mattingly, 1985, p. 26, "no higher order invariants have thus far been proposed, and we doubt that any will be forthcoming.") In the theory, an innate vocal-tract synthesizer includes knowledge of acoustic consequences of phonetic gestures, and it uses that knowledge to recover phonetic gestures from the signal.

In both theories, because the signal does not specify the phonetic message, something inside the speaker—syllable prototypes or a knowledgeable vocal-tract synthesizer—has to augment the information in the signal.

4. Specification without invariance. Heretofore, invariants have been discussed as if they are the only specifiers possible. Here, I raise the possibility that a speech signal may specify its phonetic message without the specification being achieved by invariants.

This point of view, as I will elaborate it here, shares most of the assumptions of (2) above: (a) primitives of a linguistic utterance are gestures implemented by synergies of the vocal tract; (b) synergies ensure that coarse-grained gestural goals are invariantly achieved, albeit in context-sensitive ways, as described above; and (c) the consequent acoustic signal specifies its gestural source. The view diverges from (2) in its answer to the question of whether the way the gestures are specified is by means of invariants in the acoustic speech signal.

In one sense, there cannot be specification without any invariance. If three acoustic signals all specify "bilabial closure," then there is invariant *information* (namely, the information "bilabial closure occurred") in the three signals. The question is whether the information, invariantly present, is instantiated in an invariant property of the acoustic signal. An example from another domain, one in which invariant instantiation *is* found, may make this clear. When an object approaches an observer (or vice versa) there is information in the optic array for "time to contact"— that is, the time at which the object will

contact the observer, or vice versa (Lee, 1976). Not only is the information always there, but, in addition, the structure in the light that instantiates the information can be given the same description—namely, the ratio of an optical angle for an approaching object (the size of the object's optical contours in the reflected light) and its rate of expansion in the optic array (A/Å).

The question is whether specification is always achieved by information that can be given an invariant description across instantiations. I do not know the answer, but will raise the possibility that it need not be.[4] As Kuhn (1975) has pointed out, following Fant (1960), $F2$ and $F3$ of vowels sometimes are affiliated with either the front or the back vocal-tract cavity (that is, the cavity in front of or behind the constriction point for the vowel). In /i/, associated with a forward constriction point, $F3$—the higher resonance—is a resonance of the (short) front cavity and $F2$ is a back-cavity resonance. In /u/, with a back constriction point, the cavity affiliations of $F2$ and $F3$ are the reverse of those for /i/. Acoustically, there is information in the relative intensities of the formants to signal their respective affiliations: When $F2$ is the front-cavity resonance, it is considerably more intense than $F3$; when $F3$ is the front-cavity resonance, the intensities are more nearly equal. There is, in addition, some evidence that listeners use the relative amplitude of $F2$ and $F3$, in identification of /i/ versus /y/, for example (Aaltonen, 1985).

As an example, a consonant constriction made after a vocalic one creates its own front and back cavities that will differ in their lengths from those of the preceding vowel. Imagine a consonantal constriction that shortens the front cavity and lengthens the back cavity. Changes in the lengths of the cavities should be reflected in the formant transitions from vowel to consonant, and the transitions in relation to the recoverable constriction point for the vowel may specify the consonantal place of articulation. However, whether $F2$ and $F3$ each either rises or falls will depend on which formant is the front- and which the back-cavity resonance and which cavity shortens or lengthens. If place of articulation is specified in the signal more or less in the manner outlined, then apparently it is not specified by an invariant analogous to (A/Å). I do not see that such an outcome would matter for the perceiver. The essential thing (at least in a theory of perception as direct) is for phonetic properties to be specified.

REFERENCES

AALTONEN, O. (1985). The effect of relative amplitudes of $F2$ and $F3$ on the categorization of synthetic vowels. *Journal of Phonetics*, **13**, 1-9.

BEST, C. T., MORRONGIELLO, B., & ROBSON, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, **29**, 191-211.

BLADON, A., & AL-BAMERNI, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics*, **4**, 137-150.

BLUMSTEIN, S., ISAACS, E., & MERTUS, J. (1982). The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, **72**, 43-50.

BLUMSTEIN, S., & STEVENS, K. (1980). Perceptual invariance and onset spectra for stop consonants in different vocalic contexts. *Journal of the Acoustical Society of America*, **67**, 648-662.

BOND, Z., & GARNES, S. (1980). Misperceptions of fluent speech. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 115-121). Hillsdale, NJ: Erlbaum.

BROWMAN, C., & GOLDSTEIN, L. (1986). Towards an articulatory phonology. *Phonology*, **3**, 219-252.

BROWMAN, C., & GOLDSTEIN, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, **18**, 299-320.

BROWMAN, C., & GOLDSTEIN, L. (1992). Articulatory phonology: An overview. *Phonetica*, **49**, 155-180.

CLEMENTS, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, **2**, 225-252.

DELL, G. (1986). A spreading-activation theory of retrieval in speech production. *Psychological Review*, **93**, 283-321.

DUEZ, D. (1992). Second-formant locus patterns: An investigation of spontaneous French speech. *Speech Communication*, **11**, 417-427.

FANT, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

FOWLER, C. A. (1986). An event approach to the study of speech perception from a direct realist perspective. *Journal of Phonetics*, **14**, 1-28.

FOWLER, C. A., & ROSENBLUM, L. D. (1991). The perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 33-59). Hillsdale, NJ: Erlbaum.

FOWLER, C. A., RUBIN, P., REMEZ, R., & TURVEY, M. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Eds.), *Language production I: Speech and talk* (pp. 373-420). London: Academic Press.

GIBSON, J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

GOLDSMITH, J. (1976). *Autosegmental phonology*. Bloomington: Indiana University Linguistics Club.

GREENBERG, J., & JENKINS, J. (1964). Studies in the psychological correlates of the sound system of American English. *Word*, **20**, 157-177.

HAMMARBERG, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics*, **4**, 113-133.

KELSO, J. A. S., TULLER, B., VATIKIOTIS-BATESON, E., & FOWLER, C. A. (1984). Functionally-specific articulatory cooperation following jaw perturbation during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 812-832.

KEWLEY-PORT, D. (1981). *Representation of spectral change as cues to place of articulation in stop consonants*. Unpublished doctoral dissertation, City University of New York.

KRULL, D. (1989). Consonant-vowel coarticulation in spontaneous speech and in reference words. *PERILUS*, **10**, 101-105.

KUHN, G. (1975). On the front-cavity resonance and its possible role in speech perception. *Journal of the Acoustical Society of America*, **58**, 428-433.

LAHIRI, A., GEWIRTH, L., & BLUMSTEIN, S. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants. *Journal of the Acoustical Society of America*, **76**, 391-404.

LEE, D. (1976). A theory of visual control of braking based on information about time to collision. *Perception*, **5**, 437-459.

LIBERMAN, A., COOPER, F., SHANKWEILER, D., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.

LIBERMAN, A., & MATTINGLY, I. (1985). The motor theory revised. *Cognitive Psychology*, **21**, 1-36.

LINDBLOM, B. (1963). *On vowel reduction* (Report No. 29). The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden.

MANN, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, **28**, 407-412.

MANN, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r." *Cognition*, **24**, 169-196.

MASSARO, D. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

NEAREY, T., & SHAMASS, S. (1987). Formant transitions as partly distinctive invariant properties in the identification of voiced stops. *Canadian Acoustics,* **15,** 17-24.

ÖHMAN, S. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America,* **41,** 310-320.

PERKELL, J. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study.* Cambridge, MA: MIT Press.

PIERREHUMBERT, J. (1990). Phonological and phonetic representation. *Journal of Phonetics,* **18,** 375-394.

RECASENS, D. (1984). V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustical study. *Journal of Phonetics,* **12,** 61-73.

RECASENS, D. (1989). Long range coarticulatory effects for tongue dorsum contact in VCVCV sequences. *Speech Communication,* **8,** 293-307.

REED, E., & JONES, R. (EDS.). (1982). *Reasons for realism: Selected essays of James J. Gibson.* Hillsdale, NJ: Erlbaum.

REPP, B. (1981). On levels of description in speech research. *Journal of the Acoustical Society of America,* **69,** 1462-1464.

SALTZMAN, E., & KELSO, J. A. S. (1987). Skilled action: A task-dynamic approach. *Psychological Review,* **94,** 84-106.

SHATTUCK-HUFNAGEL, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ: Erlbaum.

SINGH, S., WOODS, D., & BECKER, G. (1972). Perceptual structure of 22 prevocalic English consonants. *Journal of the Acoustical Society of America,* **52,** 1698-1713.

STEVENS, K. (1989). On the quantal nature of speech. *Journal of Phonetics,* **17,** 3-45.

STEVENS, K., & BLUMSTEIN, S. (1981). The search for invariant correlates of phonetic features. In P. Eimas & J. Miller (Eds.), *Perspectives of the study of speech* (pp. 1-38). Hillsdale, NJ: Erlbaum.

SUSSMAN, H. (1989). Neural coding of relational invariance in speech: Human language analogs to the barn owl. *Psychological Review,* **96,** 631-642.

SUSSMAN, H. (1991). The representation of stop consonants in three-dimensional acoustic space. *Phonetica,* **48,** 18-31.

SUSSMAN, H., HOEMEKE, K., & AHMED, F. (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., McCAFFREY, H., & MATTHEWS, S. (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.

VAN DEN BROECKE, M., & GOLDSTEIN, L. (1980). Consonant features in speech errors. In V. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 47-65). New York: Academic Press.

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.

WALLEY, A., & CARRELL, T. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation. *Journal of the Acoustical Society of America,* **73,** 1011-1022.

NOTES

1. This is not everywhere the case. In their 1991 study, for example, Sussman et al. stated: "The overall purpose of the research was to present a refocused conceptualization of a traditional, but elusive candidate for place of articulation . . . The departure point was to seek invariant acoustic patterns related to place of articulation" (p. 1309). However, Sussman made clear in his review of this manuscript that the intended application of the locus equations is for stop consonants only, perhaps even just "b-d-g."

2. One piece of evidence that Clements reports for this representation is that in English, consonants that share place but differ in manner (in particular, /t/, /d/, /n/, and /l/) all assimilate in place to certain following consonants. For example, they become interdental preceding /θ/ (as in "eighth," "width," "tenth," and "health").

3. In his review of this manuscript, Stevens pointed out that across the $F2_v$, $F2_o$ points for /g/, there is a shift in the cavity affiliation of $F2$. It is the front-cavity resonance in the context of back vowels, and so it may be affected by vowel rounding; it is the back-cavity resonance for front vowels.

4. I must beg the reader's indulgence here. I have already considered the most popular view that there is no specification and no invariance. In this last case, I want to assume that there *is* specification and to ask whether specification requires invariance. Specifiers for place have not been found, and my lack of expertise in the acoustics of speech precludes very educated guessing. I must ask the reader to suspend the accurate assessment that my examples here lack detail and do not particularly inspire confidence that specifiers for place will be found. The question is: If such specifiers were to be found, would they be invariants?

APPENDIX
Values of $F2$ at Onset and $F2$ at Midpoint Averaged Over the 10 Subjects

Consonant	$F2_o$	$F2_v$	$F2_o$	$F2_v$	$F2_o$	$F2_v$	$F2_o$	$F2_v$
		Central or Back Vowels						
	/ʌ/		/a/		/ɔ/		/uw/	
/b/	1328	1435	1319	1373	1131	1187	1461	1461
	23.7	25.7	23.9	22.5	25.0	23.8	26.7	31.3
/v/	1327	1402	1312	1354	1164	1193	1504	1530
	22.7	23.4	26.1	22.1	23.8	20.3	25.4	28.7
/ð/	1620	1495	1563	1362	1476	1220	1861	1806
	28.3	23.2	32.6	23.4	37.6	23.2	34.6	30.0
/d/	1825	1572	1788	1423	1683	1262	2096	1896
	31.4	25.4	32.3	24.3	33.9	24.3	38.8	34.6
/z/	1691	1554	1664	1413	1609	1272	1936	1842
	25.7	27.7	30.1	25.5	33.3	24.5	31.6	30.3
/ž/	1939	1606	1909	1455	1849	1298	2070	1887
	28.6	28.6	32.8	22.7	35.4	24.9	33.8	30.1
/g/	1916	1608	1885	1468	1573	1289	1966	1708
	29.3	25.0	24.0	21.0	39.2	25.9	41.7	30.1
		Front Vowels						
	/iy/		/I/		/ey/		/æ/	
/b/	2334	2588	1994	2104	2012	2406	1763	1847
	42.5	55.0	35.0	43.6	35.1	48.7	30.1	33.5
/v/	2218	2556	1953	2079	1939	2355	1735	1828
	41.2	51.0	33.2	39.7	31.0	42.2	28.3	32.6
/ð/	2208	2555	1935	2032	1960	2354	1809	1839
	40.9	51.9	36.3	39.2	36.6	45.7	31.9	33.0
/d/	2376	2613	2119	2101	2164	2394	1980	1888
	41.6	51.0	41.0	40.4	38.4	46.2	36.5	33.2
/z/	2193	2550	1923	1997	1954	2304	1832	1843
	43.4	54.2	36.2	40.7	34.8	43.1	29.7	33.6
/ž/	2318	2506	2089	2035	2139	2286	2032	1849
	40.1	48.3	36.3	38.3	34.4	41.6	32.1	33.1
/g/	2596	2630	2385	2168	2436	2421	2308	1912
	53.9	52.0	47.3	38.9	51.6	47.8	48.2	33.4

Note—Values below means are standard errors.