IDENTRODUCTION

Locus equations are linear regressions of the frequency of the second formant transition sampled at its onset (F2 onset) on the frequency of the second formant sampled in the middle of the following vowel (F2 vowel) for a single consonant coarticulated with a range of vowels. The F2 onset is plotted on the y axis and the F2 vowel on the x axis. Locus equation variables, F2 onset and F2 vowel, are extremely highly correlated in natural speech (Sussman et al., 1991, 1993, 1995). The phonetic origin(s) of this F2 transition endpoint correlation remain obscure, and theories concerning it are at this time speculative and controversial (Fowler, 1994; Sussman et al., 1995, 1997). What is not controversial is the fact of the robust, stable correlation between locus equation variables across speakers and languages (Swedish: Lindblom, 1963; Krull, 1988, 1989; Canadian English: Nearey and Shammass, 1987; American English: Sussman et al., 1991, 1995, 1997; Fowler, 1994; Thai, Cairene Arabic, Urdu: Sussman et al., 1993; Spanish: Celdran and Villalba, 1995; Estonian: Eek and Meister, 1995). As the above references indicate, the high correlation between the locus equation variables has now been replicated in at least six different laboratories, and we may regard it as a well-established fact, and probably a language universal.

From their discovery by Lindblom during a study of vowel reduction in Swedish (Lindblom, 1963) up to the present day, locus equations have been employed descriptively, particularly as indices of degree of consonant–vowel (CV) coarticulation (Krull, 1989). In a departure from this descriptive tradition, Nearey and Shammass (1987), Sussman (1989) and Sussman et al. (1991, 1995) have postulated that the highly correlated locus equation variables, in combination, could be cues for consonant place of articulation perception. The basis for this concept is that the particular regression of F2 onset on F2 vowel depends on the consonant. Each regression function can be interpreted as defining an equivalence class of CV transitions, and each equivalence class maps to a particular consonant. The classes intersect to some extent, so they are only “partially” distinctive as to consonant; in those regions of intersection some other, non-locus equation dimension (such as the burst, cf. Dorman et al., 1977) is assumed to carry the distinctive information. Attributing to locus equations a role in consonant place of articulation perception offers a new perspective on a major portion of the noninvariance dilemma as it is manifested in speech perception (Sussman et al., 1991).

Briefly, “noninvariance” means there is lacking a transparent map between a highly variable signal and a unit of the linguistic message. A particularly baffling, paradigmatic case of noninvariance has been the variability of F2 transitions associated with the stop consonants /b,d,g/ coarticulated in CV syllables with a range of vowels (Liberman et al., 1954, 1967; Delattre et al., 1955; Liberman and Mattingly, 1985). If a normalization routine were to be explicitly and mathematically derived for the variable F2 transitions within a place category, “correcting” for vowel-contexts, the function would do well to resemble a typical locus equation scatterplot. In the normal course of obstructed + vowel productions, the F2 onset–F2 vowel data self-organize into the equivalence classes mentioned above, which we have called “vowel-normalized F2 transitions” (Sussman et al., 1995). We use this term in a descriptive, not in any formal or implemental sense.

A locus equation-based perspective on place of articulation perception stresses the organization of allophones of a phonetic category into a linearly correlated distribution of F2 coordinates in acoustic space. What makes this linear distribution of F2 coordinates nontrivial is that locus equa-
tions might not be articulatorily inevitable. Recent pilot work using the articulatory synthesis model, APEX (Stark et al., 1996; Lindblom et al., 1997), which generates an articulation-to-formant mapping, has shown that the full set of biomechanically possible articulatory configurations (e.g., lips, mandible, shape and position of tongue body and tip), for an apical closure for a given vowel context, is very broad. Thus, speakers’ acoustic output ((F2 vowel, F2 onset)) can vary, in principle, over an extensive acoustic range, and linearity across the vowel space is by no means assured. Linear relationships among F2 transition endpoints appear to be “special” in the sense of being implicitly “selected” by speakers from a large range of articulatory possibilities. Phonetic- logically logical candidates to account for the linearity include (i) articulatory constraints operating to optimize speech production, and/or (ii) motor optimizations working in the service of perception. Part of the purpose of this study is to test the second possibility.

It should be noted that the locus equation theory of obstructed place of articulation perception across vowel contexts has been misinterpreted by some as implying that F2 onset in relation to F2 vowel is a specifier of consonant place of articulation rather than simply an important cue (Fowler, 1994). This is not the case. Rather, F2 onset in relation to F2 vowel is claimed to yield a partially distinctive cue (distinctive in many vowel contexts but not all) that typically functions as an important element of a polyphonic, redundant cue set. The locus equation-based hypothesis is not an invariance theory. It should not be so considered because it does not claim there is invariance in the signal. Rather, it claims there is modally a pattern of variation in the signal that can be readily mapped and usefully abstracted by a neural system. Rather than “invariance,” this notion of the relationship between waveform and phoneme might be termed “mapable variance” or “orderly variance” where “mapable” or “orderly” are to be defined strictly in terms of the feature-mapping properties of sensory neural systems (Sussman et al., 1997).

In the following sections, some arguments for including locus equations in a theory of obstructed place of articulation perception across vowel contexts will be reviewed. This review will also motivate the research to be reported below, the thrust of which is to compare the acoustic production space of locus equations to the perceptual space of listeners instructed to identify stop consonants in synthetic CV continua orthogonally varying F2 onset and F2 vowel, the latter over a full range of vowel contexts. This experiment is designed to test the relevance of locus equation acoustic variables to the perception of stop consonants.

A. Rationale for the perceptual relevance of locus equations

1. Robustness of the phenomenon

One idea that locus equation data clearly refute is the Motor Theory notion that acoustic variability precludes establishing phonetic categories based on the acoustic signal:

Putting together all the generalizations about the multiplicity and variety of acoustic cues, we should conclude that there is simply no way to define a phonetic category in purely acoustic terms. (Liberman and Mattingly, 1985, p. 12)

For if phonetic categories were acoustic patterns, and if, accordingly, phonetic perception was properly auditory, one should be able to describe quite straightforwardly the acoustic basis for the phonetic category and its associated percept. (ibid.)

When viewed in terms of the linear and tightly clustered F2 onset–F2 vowel relationship, F2 transition variability by stop place category is about as orderly as one can find anywhere in Nature. As noted previously, the consonant-dependent F2 onset–F2 vowel relationships are stable and robust across speakers, dialects, languages and even in the face of articulatory perturbations produced by bite blocks (Sussman et al., 1995). Discriminant analyses of speaker functions, using slope and y-intercept as predictor variables, have shown 100% correct classification of the functions into stop place categories [English: Sussman et al. (1991); Spanish: Celletan and Villalba (1995)]. Slope-y-intercept values may be somewhat modulated by speaker-induced changes in speaking style (Krull, 1989) or rate (Bakran and Mildner, 1995), but sufficient contrast among consonant place categories is maintained despite these modulations.

2. Perceptual studies of stop place of articulation cued by locus equation parameters

It has long been known that the F2 transition, as well as the noise burst, is an important cue for stop place (see Sussman et al., 1991 for a complete review of this literature). Well before the locus equation metric was first derived (Lindblom, 1963), Liberman et al. (1954) manipulated F2 onset frequency, from a point 480 Hz below to 720 Hz above, the steady-state F2 resonances of seven vowel contexts in a two-formant synthesis study of stop consonant place perception. They concluded that “the results of this experiment show that the direction and degree of second-formant transitions can serve as cues for the aurally perceived distinctions among the stop consonants” (p. 9). While their Figure 5 showed respectable separability of ‘‘b, d, g’’ regions of perceptual space across vowel contexts, there were problematic contexts where the F2 transition was not very effective in cueing alveolars (e.g., /d, t/ in places of h/ contexts).

In an effort to relate the results of this classic study to the locus equation parametrization of the F2 transition, Liberman et al.’s Figure 3 was replotted. We transformed their seven F2 vowel frequencies, their eleven F2 transition onsets (per vowel), and their identification values1 to (x, y, z) coordinates to be compared with locus equation acoustic data obtained from five male speakers taken from Sussman et al. (1991). Figure 1 shows the obtained results. The identification data displayed as a gray scale are referred to as an “identification surface.” In analogy to a spectrogram, high rates of consonant identification are indicated by darker areas. The acoustic data to be compared are overlaid (as white circles). While two-formant pattern playback steady-state
vowel target frequencies (essentially an $F_2$-prime vowel) cannot be expected to exactly match $F_2$ vowels from our speakers (Bladon and Lindblom, 1981), the correspondence between perception results and acoustic data are quite reasonable. The labial consonant identifications closely parallel the acoustic data. The “d” identifications also agree reasonably well with the superimposed locus equation data. The “g” results are noticeably weaker as areas of “g” perception predominate at $F_2$ onsets $>2250$ Hz, and little acoustic $F_2$ onset data falls in this region. Conversion of the Sussman et al. acoustic data to $F_2$ prime values would increase the degree of correspondence to the Liberman et al. perceptual data in the upper-right quadrant of these figures, i.e., for /g/ in front vowel contexts.

A more recent perceptual test of locus equation parameters, $F_2$ onset and $F_2$ vowel, as cues for stop consonant place perception was performed by Eek and Meister (1995). Using Estonian CVV syllables beginning with voiceless unaspirated stops /p,t,k/ followed by nine long vowels, tokens from one speaker were used as stimuli for identification by 13 listeners. The $F_2$ onsets “without noise components of the transitional part of the burst” were varied and presented to listeners for identification. In agreement with the Liberman et al. (1954) study, all rising $F_2$ transitions were heard as “p,” while falling transitions were heard as “t” in back vowel contexts and “k” in front vowel contexts. Missing from the identification responses were alveolar “t” in front vowel contexts and velar “k” in back vowel contexts. The authors concluded that the best perceptual performance (75%–100%) resulted when the strongest peak of the burst was combined with a measure of $F_2$ vowel-prime (rather than the $F_2$ onset–$F_2$ vowel parameters).

The research to be reported below was an attempt to replicate the results of Liberman et al. (1954), employing more realistic five-formant stimuli and an expanded vowel space. Also, advantage was taken of the presence of the $F_3$ by independently manipulating it (three levels of $F_3$ are employed—one level appropriate for [b], one for [d], and one for [g]). Correspondence of burstless CV stimulus identifications to acoustic locus equation data was evaluated to test the perceptual relevance of locus equation acoustic variables to stop place perception.

FIG. 1. Identification surfaces for “b,d,g” across seven vowel contexts. Data ($F_2$ vowel, $F_2$ onset, identification frequencies) transformed from Liberman et al. (1954) and displayed in locus equation coordinate space. Overlaid acoustic data from production (white circles) based on data from five male speakers taken from Sussman et al. (1991).
I. METHODS

A. Synthetic speech stimuli

The stimuli were synthesized using the Klatt synthesizer, KLSYN88 (Klatt, 1980), running on a VAXStation under VMS. The parameter files that controlled the syntheses were derived from files generated by Klattalk, a text-to-speech program (Klatt, 1987). The process of creating the experimental stimuli began with ten Klattalk files, one for each bV syllable [ba], [bi]. The ten monophthongal vowel contexts, in order of ascending F2 vowel, were /σαυειαι/. The stimuli were given a male voice quality. The [b]'s were made burstless by zeroing the track for the variable parameter “amplitude of burst.” Several variables were held constant across all tokens: (i) the duration of all tokens was set to 300 ms, (ii) the pitch track of all tokens was set equal to that of the [ba] token, (iii) the durations of the F2 and F3 transitions were set to 48 ms for all tokens, and (iv) the VOT of all tokens was set to 6 ms. The F2 and F3 tracks from the end of their transitions (48 ms post-onset) through the end of the vowel (300 ms post-onset) were adjusted to reflect the mean F2 and F3 vowel values reported by Sussman et al. (1991) for a sample of ten adult males (Tables I and II). The adjustment was carried out by setting the end of the F2 and F3 transitions to the appropriate mean target value and then raising or lowering all following points of the F2 or F3 track to splice cleanly with the end of the transition. For example, the frequency of the F2 of [bi] 48 ms post-onset was changed from the original Klattalk value of 1994 Hz to the Sussman male mean value of 2180 Hz. Then, all subsequent values in the F2 track were raised 186 Hz (2180 minus 1994).

The F2 and F3 onsets were varied orthogonally. Three levels of F3 transition, one appropriate for [b], one for [d], and one for [g], were combined with 11 levels of F2 onset per vowel (11 F2 onset values per vowel×10 vowels×3 levels of F3 = 330 total stimuli for the experiment). The F3 onset values were adjusted to reflect the Sussman male means (Table II). As in the Liberman et al. (1954) experiment, the F2 onset series spanned a range from 480 Hz below to 720 Hz above the appropriate F2 vowel frequency. Intervals in the F2 onset series followed a stepwise equidistant plan such that there were, including the endpoints, 11 onset values per vowel separated by equal intervals on a Bark scale. Straight-line transitions were interpolated between F2/F3 onsets and their respective targets using the “draw” command of KLSYN88. The F2 vowel×F2 onset space sampled in this experiment is shown in Fig. 2. Table I provides frequency values for the stimulus coordinates shown in Fig. 2.

The waveforms based on the completed. DOC files were

<table>
<thead>
<tr>
<th>Level</th>
<th>F2 onset (Y) values (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1675 1788 1842 1905 1935 2345 2392 2511 2625 2900</td>
</tr>
<tr>
<td>2</td>
<td>1507 1620 1674 1737 1767 2181 2228 2349 2465 2745</td>
</tr>
<tr>
<td>3</td>
<td>1355 1467 1522 1584 1614 2029 2077 2199 2316 2599</td>
</tr>
<tr>
<td>4</td>
<td>1216 1328 1383 1445 1474 1889 1937 2059 2177 2462</td>
</tr>
<tr>
<td>5</td>
<td>1089 1200 1256 1317 1346 1759 1807 1930 2047 2332</td>
</tr>
<tr>
<td>6</td>
<td>970 1082 1138 1198 1228 1639 1687 1809 1926 2211</td>
</tr>
<tr>
<td>7</td>
<td>859 972 1028 1088 1117 1527 1575 1696 1813 2096</td>
</tr>
<tr>
<td>8</td>
<td>754 868 925 985 1014 1423 1470 1591 1707 1989</td>
</tr>
<tr>
<td>9</td>
<td>653 769 827 888 917 1325 1372 1492 1608 1887</td>
</tr>
<tr>
<td>10</td>
<td>555 674 733 795 824 1233 1280 1399 1514 1791</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>F2 vowel (X) values (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>955 1068 1122 1185 1215 1625 1672 1791 1905 2180</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>F3 onset values (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>b-like</td>
<td>2381 2313 2403 2420 2270 2343 2365 2400 2357 2460</td>
</tr>
<tr>
<td>d-like</td>
<td>2418 2406 2458 2527 2359 2566 2559 2569 2557 2667</td>
</tr>
<tr>
<td>g-like</td>
<td>2170 2178 2195 2227 2172 2551 2617 2664 2632 2904</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>F3 vowel values (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>b-like</td>
<td>2471 2357 2518 2546 2284 2460 2483 2473 2518 2728</td>
</tr>
</tbody>
</table>
B. Consonant identifications

Stimuli were delivered to subjects via a reel-to-reel tape deck (Tandberg TD 20A) connected to an amplifier and headphones (Realistic stereo mixing console, Adcom GFA-535L amplifier, splitter box, beyerdynamic DT100 headphones) in a quiet room. The six subjects were all graduate students between the ages of 25 and 40 years old. All subjects reported normal hearing. All were native speakers of an American or Canadian English dialect. Subjects 1, 3, and 4 were male, while subjects 2, 5, and 6 were female. With the cooperation of the experimenter, subjects were asked to set their own level to be comfortable but moderately loud. An oscilloscope reading for a calibration tone played at this level setting was recorded, so the chosen level could be duplicated for subsequent sessions with a given subject. Subjects were trained in the task sufficiently (220 randomly selected stimuli) to become comfortable with a relatively rapid stimulus presentation rate, adopted to prevent subjects from becoming bored later with the task. No feedback was given during training except to initially assure subjects that by the end of training they would be adapted to the task rate. Initially, the trials seemed speeded to some subjects, but by the end of training all acknowledged the presentations were comfortably paced. In order to limit fatigue, a maximum of two tapes (720 stimuli) were tested on a given day. Thus, each subject listened to eight replications per stimulus over a minimum of four days. Although fewer subjects were used than in Liberman et al. (1954), each subject was run through more trials to discern whether there was significant between-subject variation. Subjects were asked to identify each stimulus as most similar to "b," "d," "g," "w," or "no consonant." The responses "w" and "no consonant" were not part of the response repertoire of Liberman et al. (1954), but were added to this replication design, after the first author had listened to the stimuli, in order not to force subjects to make sometimes nonsensical identifications. Subjects were warned to use the last alternative only if there was really no consonant, not if the consonant was simply ambiguous.

C. Analysis

Identification frequencies were tabulated and entered along with F2 onset and F2 vowel of the eliciting stimuli into graphics software (MATLAB version 4.2c.1) to produce the "identification surfaces." Each identification surface is based on 110 triplets of the form \( (F2\text{ vowel}, F2\text{ onset}, \text{identification frequency}) \)—the identification frequencies are added on as the \( z \) values. The software interpolates between sample points in a manner analogous to dropping a cargo net over a 3-D scatterplot of the identification data. The method of shading, known as Gouraud shading, is piecewise bilinear; the shading of each patch varies linearly and interpolates the corner values of the rhomboids (The MathWorks, Inc., 1992).

A chi square method used to statistically compare the identification results with the acoustic data is described in Sec. II C.

II. RESULTS

The results are first presented as identification surfaces along with overlaid locus equation acoustic data. It is with this type of display that the correspondence of perceptual responses to locus equation acoustic data can best be judged. A chi square analysis of the degree of correspondence between identification and acoustic data appears in a later section. There is also a brief presentation of the effects of \( F3 \).

A. Identification surfaces

Recall from Fig. 2 that there were 110 combinations of \( F2 \) vowel and \( F2 \) onset composing a set of stimuli for a given level of \( F3 \). The format of Fig. 2 is that of a normal locus equation acoustic space, so that each vertical line of values is an \( F2 \) onset series (11 levels) for a given vowel context, and the ten vowel contexts are arranged horizontally.
in order of ascending $F_2$ vowel frequency. Each of the points in Fig. 2 represents a stimulus in the perception experiment, so that there is identification data for each of them. If the identification data are graphed on a $z$ axis, we can create a surface by interpolating between adjacent triplets $<x,y,z>$. The overall effect is as if the individual $F_2$ onset series identification curves for a given consonant identification were slices of bread, each slice being a different vowel context, and the slices were assembled into a loaf, which we then view from above. These surfaces have the decided advantage of being able to simultaneously display all vowel contexts for a given consonant identification and level of $F_3$. For translation to a two-dimensional medium, the $z$ axis is coded via a gray scale—the higher the rate of identification, the darker the shading—similar to the intensity coding of a sound spectrogram. All figures show subject-pooled data (i.e., $z$ ranges from 0–48).

Although $F_3$ was a variable in this experiment, it had little effect, so only surfaces for appropriate levels of $F_3$ (b-like for “b,” d-like for “d,” or g-like for “g”) will be shown, as they are representative. (The effects of $F_3$ did show an interesting overall pattern, which will be described after the main result.) For the consonant identifications “b,” “d,” and “g,” corresponding token-level locus equation acoustic data from five male native speakers of English producing “beat, bit, bait,... , deet, dit, date,... , geet, git, gate,...,” five times each (data taken from Sussman et al., 1991) are overlaid (white circles). The overlays allow appraisal of the correspondences between the distribution of the acoustic data and features of the perception data.

There are clear peaks and valleys in the identification surfaces, a peak (darker region) representing a high rate of identification as a given consonant. Turning first to the “b” surface (Fig. 3), it can be summarized by saying that a high rate of “b” response is indicated when the $F_2$ onset is lower than the $F_2$ vowel. In the back vowel region ($F_2$ vowel ca. 1000 Hz), a high rate of “b” identification is achieved even when the $F_2$ onset is slightly above the $F_2$ vowel. The overall effect for “b” identifications is of a massif, or ridge, rather than of a single peak. The overlaid acoustic data correspond very well with this ridge of “b” perception.

The “d” surface is shown in Fig. 4. The “d” peak in the back vowel region ($F_2$ vowel <c. 1300 Hz) exactly complements the ridge for “b” i.e., it is clear that the “b” and “d” curves do not overlap in back vowel contexts. On the other hand, in the front vowel contexts ($F_2$ vowel c. 1600 Hz) the “d” peaks are lower than in the back vowel region and overlap the “b” ridge. Noting the attenuated intensity of the “d” peaks in the front vowel region, there is nevertheless a good correspondence between the “d” identification peaks and the overlaid [d] acoustic data.

Referring to the “g” surface (Fig. 5), in front vowel contexts the peak region for “g” complements the peaks for “b” and “d.” Actually, it appears there may be, near the joint boundaries, cases of “triple-point” front vowel context stimuli that are almost perfectly confusable among the three stops. The front vowel peak for “g” is attenuated above an $F_2$ vowel of c. 1750 Hz. The back vowel region is characterized by a relative lack of “g” responses, except for one vowel context (/a/, in the region of $F_2$ vowel at c. 1100 Hz). The correspondence between the “g” peak and the overlaid acoustic data is fair in the front vowel region. The back vowel correspondence is problematic, inasmuch as there is very little “g” identification but plenty of [g] acoustic data distributed in the region.

The patterns of “w” and “no consonant” responses were not sufficiently interesting to warrant a graphical presentation. In general the “w” surfaces were relatively featureless. This turned out to be a case of between-subject variation, with two subjects showing great activity against a background of four showing none. The few “w” responses were concentrated in the back vowel region at the lower limit of the $F_2$ onset series, where they overlap the “b” curve in subject-pooled data, or cross over the “b” curve in those subjects that gave a “w” response. The “no consonant” surfaces also showed little activity. Subjects were instructed to use this response sparingly. There seemed to be a gradient from the back vowel region to the front vowel region, with higher and more widespread activity with increasing $F_2$ vowel. As might be expected, “no consonant” activity at the lower $F_2$ vowel values (below ca. 1750 Hz) was concentrated among those stimuli having a relatively flat $F_2$ consonant-vowel transition. Above ca. 1750 Hz, the CV transition trajectory seemed to have less influence; it seemed

---

**FIG. 3.** Identification surface for “b” overlaid with acoustic data (white circles) from production of bVt syllables [taken from Sussman et al. (1991)].
especially difficult to evoke a stop identification with the vowel context /i/, even for a wide range of \( F_2 \) transitions. As with "w," there was considerable between-subject variation in the frequency and distribution of the "no consonant" responses, with two subjects particularly prone to use it compared to the other four.

**B. Effects of \( F_3 \)**

Table III presents a summary of the effects of level of \( F_3 \) on stop place perception. A certain degree of symmetry, of an antiparallel type, is evident in the table. Particularly, there was no effect of the \( F_3 \) condition on "b" identification in back vowel contexts and on "g" identification in front vowel contexts, while there were effects on "d" versus "g" identification in back vowel contexts and on "b" versus "d" identifications in front vowel contexts. In comparing these patterns of \( F_3 \) effects with locus equation acoustic data, a strong relationship emerges, namely, that there is a lack of \( F_3 \) effects in those regions in which there is a lack of overlap between the different stop places of articulation (back vowel [b] and front vowel [g], while there are tradeoff effects between the overlapping stops in the region of their overlap [d] and [g] overlap in back vowel space, [b] and [d] in front vowel space). These tradeoffs are in the natural directions, with g-like \( F_3 \) elevating "g" versus "d" identifications, and b-like \( F_3 \) elevating "b" versus "d" identifications.

**C. Quantitative comparison of acoustic and identification data**

In an attempt to quantify the degree of relationship, or in some sectors of locus equation space, the lack of direct relationship, between locus equation acoustic data and the identification frequencies, a chi square analysis was performed.

Expected frequencies were generated by a model in which identification frequencies directly reflect the distribution of acoustic data local to a stimulus in the identification experiment. Locality was defined in terms of nearest neighbors in locus equation acoustic space. For each \( F_2 \) vowel, \( F_2 \) onset point sampled in the identification experiment

<table>
<thead>
<tr>
<th>Effect on</th>
<th>For back V contexts</th>
<th>For front V contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;b&quot;</td>
<td>none</td>
<td>b-like elevates, g-like depresses</td>
</tr>
<tr>
<td>&quot;d&quot;</td>
<td>g-like depresses</td>
<td>b-like depresses</td>
</tr>
<tr>
<td>&quot;g&quot;</td>
<td>g-like elevates</td>
<td>none</td>
</tr>
</tbody>
</table>
(Fig. 2), the Euclidean distance between it and each of the entire set of acoustic data points overlaid onto the identification surfaces was computed. Next, the 48 nearest-neighbor acoustic data points were identified, and then profiled in terms of their initial consonant. So, for example, for the identification stimulus at the sixth level of $F_2$ onset in the vowel context series /e/, the 48 nearest neighbors from the acoustic data set included 9 bVt tokens, 25 dVt tokens, and 14 gVt tokens. This profile, 9 b’s, 25 d’s, and 14 g’s, was then adopted as an expected set of identification frequencies for that stimulus. There was no strong theoretical basis for adopting any particular neighborhood parameters, except to preserve some notion of locality. Other choices as to the nature of the neighborhoods (size, basis, weighting by rank or distance, etc.) could have been used.

For the purposes of the chi square analysis, some modification was made to the frequencies observed in the experiment. First, “b” and “w” responses were lumped, since the expected frequencies included no “w” category and “w” and “b” are both labials. Second, the observed frequencies were selected from the three levels of $F_3$ so that the “b” frequencies were taken from the experiment with a b-like $F_3$, the “d” frequencies from the experiment with a d-like $F_3$, and the “g” frequencies from the experiment with a g-like $F_3$. The different levels of $F_3$ could have been analyzed separately or simply lumped rather than selectively combined, but it was decided to make the format of the chi square analysis parallel to that of the identification surfaces presentation, in which $F_3$ was for the most part deempa-sized due to its relatively slight effects. Third, the “no consonant” responses of the identification experiment were simply ignored since there was no comparable category of expected responses and the relative frequencies of such responses in the experiment was small. Finally, the combined selected observed frequencies were rescaled to 48 to make the scale of the observed frequencies equal to that of the expected frequencies. The computations described above account for the observed frequencies in the chi square analysis not being whole numbers.

Given the expected frequencies and the observed frequencies modified as described, the chi square of each set of frequencies, for b, d, and g, for each point of locus equation space sampled in the identification experiment, was computed. A chi square using an expected frequency less than 2 was considered invalid (Guilford and Fruchter, 1978), so those cases were left out of the analysis. In consideration of this, the particular analysis protocol had to be varied by sample point. In areas of locus equation acoustic space in which only one consonant is represented, i.e., the nonoverlapped areas of [b] in back vowel space and [g] in front vowel space, the chi square had no degrees of freedom, and no interpretation was possible. However, in a substantial subset of these cases, the observed and expected frequencies are actually identical, and a statistical analysis seems superfluous. Thus, in the chi square results table (Table IV), “no df” means the analysis could not be applied because df=0, while “equal” means the analysis similarly could not be applied but also seems unnecessary due to the identity of observed and expected. For the cases in which two or three consonants were represented among the expected frequencies, i.e., there was some integration or overlap among the consonants in the acoustic neighborhood, a chi square with 1° or 2° of freedom, as appropriate, was computed and evaluated as not significant (ns), significant at the $p=0.05$ level (*), or significant at the $p=0.01$ level (**). Yate’s correction for continuity was applied in those cases in which it was relevant.

As can be seen from Table II (which has an arrangement identical to Table I and parallel to Fig. 2), there is generally close agreement of observed and expected frequencies in the regions of nonoverlapped acoustic data, [b] in back vowel space and [g] in front vowel space. This lends some credence to the nearest-neighbor method used to derive expected frequencies—the method yields good predictions unless there is overlap among the stops. A departure of observed from expected frequencies can be seen in virtually every region of overlap among acoustic tokens. This result is strongest in the back vowel region at higher $F_2$ onset frequencies, which is the region of overlap between [d] and [g], and is also strong but relatively weaker in the front vowel region at lower $F_2$ onset frequencies, which is the region of overlap between [b] and [d]. There is one anomalous set of results, for the /i/ context series at higher levels of $F_2$ onset, that can be partly attributed to the breakdown of the nearest-neighbor
method. For the part of the series at the upper rightmost corner of the sample space, the same set of neighbors was sampled repeatedly simply at longer distances as the top three sample points recede from the margins of the acoustic data.

III. DISCUSSION

The results of this perception study closely resemble those obtained with only two-formant synthesis in Liberman et al. (1954) (cf. Fig. 1). The current study should be viewed as a cross-validation of the results of that pioneering work. Recently, convergent results were also presented by Eek and Meister (1995), although they give a very different interpretation than will be offered here. Conflicting interpretations aside, we can be reassured by the fact that three different laboratories, each using independently constructed stimuli, have separately observed the same general pattern of context-dependent consonant identifications in perception tests of burstless stimuli orthogonally varying $F2$ vowel and $F2$ onset.

The notion that a map of locus equation space somewhere in the auditory system could contribute significantly to consonant place identification is supported by the good match between the locus equation acoustic data and corresponding peaks of the identification surfaces. The darkest areas of the identification surfaces, indicating unequivocal identification of particular stops, can be thought of as analogous to partial phonological homunculi (at least as can be envisioned in these limited acoustic dimensions), while the overlaid acoustic data could represent the inputs that organized the homunculi. The correspondence between the surfaces and the acoustic data is much better for the “b” and “d” surfaces than it is for the “g” surface. Especially problematic is the lack of correspondence between the “g” surface and [g] acoustic data in back vowel contexts. This finding is interpreted here to mean that the $F2$ onset–$F2$ vowel combination is a more important cue for /b/ and /d/ than it is for /g/, at least for the current type of stimulus.

Several characteristics of the current stimuli may have contributed to the lack of “g” identifications in back vowel contexts. First and foremost, the stimuli were burstless, while bursts for natural [g] are generally the most prominent among the voiced stops (Smits et al., 1996). Strong bursts being such a dependable correlate of [g], it is understandable that a natural-sounding burstless [g] is difficult to achieve. Even for the acoustic region yielding the most consistent “g” responses—front vowel contexts coupled with the higher $F2$ onset values—some subjects preferred to respond “no consonant” at a comparatively high rate. This general disadvantage for “g” is compounded in the back vowel region, where it competes with “d” in a region of [d]–[g] overlap in locus equation acoustic space (cf. the section to follow on the Dominance hierarchy hypothesis). Second, the stimuli were constructed with values of $F2$ onset and $F3$ onset measured at the first glottal pulse, following the standard protocol for locus equation measurements. For the g-like $F3$ condition stimuli this means that the “velar pinch,” the convergence of $F2$ and $F3$ onset typically found in [g] tokens, has been attenuated compared to natural tokens. Usually, the most intense overlap of the $F2$ and $F3$ onsets may be found during the [g] burst, and by the first glottal pulse the $F2$ and $F3$ transitions have already diverged somewhat, so that if samples are taken at the first glottal pulse the acute front end of the $F2$–$F3$ pinch will be removed. Presuming the pinch is an important cue for /g/, its attenuation in these stimuli probably contributes further to a suppression of “g” identifications. Finally, the values of voice onset time (VOT) and transition duration were held constant across all stimuli, and the specific values selected—a VOT of 6 ms and a transition duration of 48 ms—are briefer than those for natural [g] tokens. For example, a long VOT is frequently cited as a distinctive correlate of velars [mean [g] VOT = 25 ms (N = 30; Nossair and Zahorian (1991))], so that a short VOT might be another factor tending to suppress “g” identifications.

A. Dominance hierarchy hypothesis

In interpreting consonant identification data very similar to that of this study, Eek and Meister (1995) concluded that there was no clear relationship between token-level locus equation acoustic data and the identification patterns. Such a dismissive conclusion is premature. A dominance hierarchy hypothesis is here offered to help conceptualize the relationship between the token-level acoustic data and the identification patterns for the burstless stimuli used in this study. Figure 6 presents a schematic of the hypothesis. Each outline represents a particular stop consonant’s cloud of points in locus equation acoustic space. It is an abstract rendition of the combined scatterplots for [b,d,g] with the addition that the opacity of the “clouds” models the postulated perceptual dominance effect in regions of acoustic overlap. In combined scatterplots [d] and [b] data overlap in front vowel contexts while [d] and [g] data overlap in the back vowel region. Regarding the identification surfaces (Figs. 3–5), it can be seen that front vowel “d” and back vowel “g” are in a sense missing from the identification surfaces, similar to the results of Eek and Meister (1995) as well as the replotted data of Liberman et al. (1954). These identification patterns can be understood in terms of the proposed dominance hierarchy, namely, b>d and d>g. The idea is that a “b” identification will tend to prevail when tokens fall in the region of overlap between [b] and [d] in the front vowel region while, likewise, a “d” identification will tend to prevail when tokens fall in the region of overlap between [d] and [g] in the back vowel region. The stops [b] and [g] do not overlap, so their dominance relation is irrelevant.

Cross-linguistic support for the dominance hierarchy hypothesis can be obtained from the results of Eek and Meister (1995) for Estonian. A labial “p” was heard in 90%–100% of the cases in front vowel contexts, while “t-syllables” were heard as “p-syllables” in 70%–93% of the cases. In back vowel contexts the predominant consonant identification was alveolar “t” with no “k” responses reported (except before /a/ as also shown in our data). These Estonian data confirm the pattern of b>d in front vowel contexts and d>g in back vowel contexts for burstless stimuli.

The dominance hierarchy hypothesis framework can also be used to interpret some of the results found in Dorman et al. (1977), in which the vowel-dependent cue value of
bursts versus formant transitions for /b,d,g/ recognition was assayed. Referring, for example, to Dorman et al.’s Figure 4, it can be seen that the burst carries significant cue value for /d/ recognition in some front vowel contexts, but not back vowel contexts. Conversely, the burst is most valuable for /g/ recognition in back vowel contexts. Combining these observations with the patterns described by the dominance hierarchy hypothesis, one could simply conclude that the burst carries most weight in those situations in which the $F_2$ vocalic transitional cue is not really distinctive—“b” versus “d” in front vowel contexts and “d” versus “g” in back vowel contexts—cuing the alternative disfavored in the dominance hierarchy.

The dominance hierarchy hypothesis should be viewed as relative to stimulus properties, particularly the burstlessness of the current stimuli. The hypothesis describes the pattern of consonant identifications prevailing in nondistinctive $F_2$ onset–$F_2$ vowel situations when the decisive burst cue is a null. Apparently, in such cases, the stop place with the less prominent burst in natural speech prevails. Probably, if the burst is not null, it can override the default identifications described by the dominance hierarchy. Referring again to Fig. 6, a reversal of dominance relations could be envisioned as switching the position of the clouds, now the “d” cloud obscuring “b” rather than “b” obscuring “d,” for example. The dominance relation based on $F_2$ onset–$F_2$ vowel is $b>d$ in front vowel contexts, but appropriate burst information should be able to override this, favoring $d>b$, and likewise, the dominance relation based on $F_2$ onset–$F_2$ vowel is $d>g$ in back vowel contexts, but appropriate burst information should be able to override this, favoring $g>d$ (cf. Walley and Carrell, 1983). Varying VOT or further manipulating $F_3$ transition parameters might also produce dominance instabilities or reversals.

Actually, for the [b]–[d] overlap region of front vowel space, the reversal of dominance relations has already been observed. Although the subject-pooled data shows “b” dominating over “d,” at least two of the subjects clearly show the reverse relation, $d>b$. Despite this between-subject variability, the authors are encouraged to suppose that $b>d$ is the prevailing dominance pattern at the group level because group data from the current experiment, Liberman et al. (1954), and Eek and Meister (1995) all show a dominance of “b” over “d” perception in front vowel space.

The between-subject variation noted above is one indication that the dominance of “b” over “d” in front vowel space is not as robust as the dominance of “d” over “g” perception observed in back vowel space (the latter dominance relationship showed no between-subject variation). The comparative instability of the $b>d$ relation can also be discerned in the statistical comparisons to be discussed in the next section.

B. Chi square analysis

Significant differences between expected frequencies based on the local densities of acoustic tokens and the observed identification frequencies were found in the areas of overlap between stop consonants (Table IV). This result reinforces the conclusions previously discussed, namely that there is a dominance hierarchy in effect for the stimuli of these experiments, such that one consonant is perceived in preference to another one overlapping it in locus equation acoustic space. In contrast, regions with no overlap among stops show either exact or fairly close agreement between the expected and observed identification frequencies, although this was not always amenable to statistical test due to there sometimes being no degrees of freedom. Inspection of Table IV also indicates that the most consistent dominance effect (i.e., divergence of observed from expected frequencies) is in back vowel space in the region of [d] and [g] overlap (where $d>g$), while the dominance of $b>d$ perception in front vowel space is considerably weaker, indeed at several points absent (the nonsignificant values).

One unanticipated insight afforded by the chi square analysis concerns an apparent contrast in the mode of dominance in the back vowel region versus the front vowel region. The pattern in back vowel space at higher levels of $F_2$ onset is very commonly that “d” perceptions are more frequent than would be expected on the basis of the local density of [d] tokens, and there are as well a significant number of [d] tokens in the region, leading to a very strong dominance of “d” perception overall. A somewhat different pattern often occurs among the front vowels at lower levels of $F_2$ onset. Here again “d” perceptions are often more frequent than would be expected on the basis of the local density of [d] tokens, i.e., they carry relative perceptual weight, but there are comparatively few [d] tokens in the area and an overwhelming number of [b] tokens. Thus, the somewhat weak and unstable dominance of “b” perception over “d” perception in the front vowel region might be viewed as due in part to an opposing interaction between the greater perceptual weight of [d] tokens and the far greater density of [b] tokens, while the comparatively strong, stable dominance of
“d” perception over “g” perception in the back vowel region could be ascribed to the perceptual weight of [d] reinforced by its considerable density in that region.

C. Conclusion

In summary, there is ample evidence that F2 onset and F2 vowel, in combination, are significant cues for the perception of stop consonant place of articulation. These components of the speech signal are likely to be mapped together and extracted as a feature, which we have termed the vowel-normalized F2 transition, during speech perception. Of course, the F2 transition has long been considered an important cue for stop consonant place, but we are now considering a particular parametrization/coding of the F2 transition: in terms of its endpoints. The form of the postulated feature-extracting map could be a topographic representation of locus equation acoustic space (F2 vowel×F2 onset).

A coherent pattern of integration of the F2 transition with two other sets of cues, the burst and F3, can be noted. In the previous discussion of the pattern of tradeoffs between F2 transition cues and burst cues, as described by Dorman et al. (1977), it was concluded that the burst carries most weight in those situations in which the F2 vocalic transient cues are not really distinctive—“b” versus “d”, in front vowel contexts and “d” versus “g” in back vowel contexts—cueing the alternative disfavored in the postulated dominance hierarchy. The pattern of F3 effects in this experiment, as was summarized in Table III, is somewhat parallel to this view of the pattern of burst effects. As with the burst, F3 carries most weight in those situations in which the F2 vocalic transient cues are not really distinctive. There is a lack of F3 effects in those regions in which there is a lack of overlap between the different stop places of articulation (back vowel [b] and front vowel [g]), while there are tradeoff effects between the overlapping stops in the region of their overlap ([d] and [g] in back vowel space, [b] and [d] in front vowel space). These tradeoffs are in the natural directions, with g-like F3 elevating “g” versus “d” identifications, and b-like F3 elevating “b” versus “d” identifications. Thus, it seems that one key to understanding the patterns of both burst and F3 effects on stop place perception is to view them in relation to the pattern of locus equation data. The way in which burst perceptual weight and F3 perceptual weight both correlate with the overlap versus distinctiveness of locus equation acoustic data argues strongly for a locus equation interpretation of F2 transitional cues.

As a closing note it is interesting that computational studies exploring properties of time-delayed neural networks also provide indirect support for F2 onset–F2 vowel playing an important role in stop place perception. When neural networks are fed sampled consonant–vowel waveforms from natural speech and instructed to classify phonetic categories by place of articulation the most effective cues for this task were reported to be F2 onset and F2 vowel frequencies (Waibel et al., 1989).

1Values for the z axis (ID frequency) were derived by digitizing enlargements (×5.26) of the ID graphs from the left half of Fig. 3 in Liberman et al. (1954). The enlarged figures were divided into /b/ functions, /d/ functions, and /g/ functions and then separately scanned (120 dpi, 8 bit/pixel PICT files) and digitized. Estimated digitizing error was ±1.5, i.e., less than 2.5%.

2A complete chi square results table is available from the second author upon request.

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC2014, National Institute on Deafness and Other Communication Disorders, to the second author. Portions of this manuscript are based on the Ph.D. dissertation of the first author, presented to the University of Texas at Austin. We would also like to acknowledge the constructive editorial comments of Winifred Strange, Terry Nearey, and Diana Krull.


Downloaded 31 Aug 2013 to 128.119.168.112. Redistribution subject to ASA license or copyright; see http://asadl.org/terms


