Journal of Phonetics (2001) **29,** 1–22
doi.10.006/jpho.2000.0128
Available online at http://www.idealibrary.com on IDEAL®

Place of articulation and first formant transition pattern both affect perception of voicing in English

José R. Benkí

Department of Linguistics, University of Michigan, 1076 Frieze Bldg., Ann Arbor, MI 48109-1285, U.S.A.

Received 19th August 1999, and accepted 10th October 2000

Voice onset time (VOT) and first-formant (F_1) transition characteristics are both important phonetic cues for voicing categorization of pretonic stops in English. These cues also covary with place of articulation in production, such that bilabials are produced with shorter VOT values, and velars are produced with longer VOT values. This study investigated the effects of place of articulation and F_1 transition pattern on voicing categorization of VOT continua. Listeners categorized for voicing and place of articulation the stop consonant in VCV and CV stimuli for which the variables F_1 transition pattern, place of articulation (bilabial, alveolar, and velar), and VOT were fully crossed. A logistic regression analysis shows that bilabial and alveolar stimuli were more likely than velar stimuli to be classified as voiceless, largely independent of F_1 transition pattern. Increasing F_1 onset frequency and shortening transition duration also made voiceless judgments more likely. The magnitude of the F_1 transition effects was considerably larger than the place of articulation effect. The results are inconsistent with models of consonant classification in which acoustic–phonetic cues for place of articulation are not involved in the perception of the voicing contrast. However, the observed perceptual interaction between place of articulation and voicing may be consistent with either a feature- or segment-based model of consonant classification.

1. Introduction and background

Speech segments are often analyzed and described using the phonological categories of distinctive features, such as voicing, place of articulation, manner, etc. (Jakobson, Fant & Halle, 1951; Chomsky & Halle, 1968). The existence of individual features as entities independent from each other is supported by early work showing that the acoustic or phonetic cues that determine the perception of one feature, such as voicing, are different than the cues for another feature, such as place of articulation. Voice onset time (VOT), the time from release burst to onset of laryngeal vibration, is known to determine

E-mail: benki@umich.edu

0095–4470/01/010001 + 22 $35.00/0

whether a particular stop consonant is perceived as voiced or voiceless regardless of the place of articulation (Liberman, Delattre & Cooper, 1958; Lisker & Abramson, 1970). Similarly, the transitions of the higher formants along with release bursts determine the perceived place of articulation of a stop consonant regardless of its status as voiced or voiceless (Cooper, Liberman, Delattre & Gerstman, 1952). Analyses of phoneme identification in noise experiments by Miller & Nicely (1955) are consistent with phonological features being perceived as independent units, with limited redundancy among different features. Nevertheless, interactions between place of articulation and voicing in both production and perception have been documented.

The present study investigates perceptual interactions between place of articulation and voicing features, in an effort to clarify their status as separate categories. The acoustic–phonetic and perceptual studies that support independence of features remain to be reconciled with the long-standing invariance problem in phonetics, that a single phonological representation can have different phonetic, articulatory, and acoustic realizations in different situations or contexts. An important problem for phonetics, if distinctive features are to be maintained as phonological units, is to describe the nature of such variation and to show that such variation is constrained in a manner consistent with distinctive features as independent categories. Alternatively, it may turn out that features are not basic phonological units.

1.1. *Production and perception of the [voice] contrast*

Many languages exhibit the voicing ([voice])[1] distinction with two series of stops, voiced ([+voice]) and voiceless ([−voice]), which differ primarily in the relative timing between the consonantal release and the onset of laryngeal vibration, known as voice onset time (VOT) (Lisker & Abramson, 1964). While the exact implementation of VOT for the [voice] contrast differs among languages, in general the [−voice] category is signaled by a longer VOT than the [+voice] category. For English, [+voice] stops in utterance-initial position are released simultaneously with the onset of voicing, for a VOT of approximately zero, while [−voice] stops are aspirated with a substantial lag, at least 30 ms, following the release and onset of voicing.

While English [+voice] stops are transcribed phonemically as /b d g/ and [−voice] stops as /p t k/, the same stops are transcribed phonetically in utterance-initial position (and possibly in pretonic position as well) as [p t k] and [p^h t^h k^h], respectively. English speakers rarely prevoice utterance-initial stops (Caisse, 1982; Docherty, 1992). I will use the standard *phonemic* transcription of the English stops (/b d g/ *vs.* /p t k/) in the rest of this paper.

As Lisker (1986) notes, the timing relation between oral release and onset of laryngeal vibration produces a number of acoustic cues, primarily manifested in differences in formant transitions. In [+voice] stops, which typically have VOT values less than or equal to 0 ms, voicing begins simultaneously with the release. As a result, acoustic energy from vocal fold vibration excites the first formant (F_1) during the entire consonant–vowel (CV) transition. Low-frequency periodic acoustic energy from laryngeal vibration

[1] To avoid confusion between the voicing feature and laryngeal vibration, the phonological feature [voice] is indicated with square brackets. The voiced ([+voice]) and voiceless ([−voice]) values of this feature are also indicated with square brackets. Other uses of the term *voice* without square brackets refer to laryngeal vibration. Square brackets are also used below to indicate a fine IPA transcription.

makes F_1 audible during its rise from the consonantal release to the vowel steady-state frequency.

The long VOTs for [−voice] stops result in first-formant (F_1) transitions that differ radically from [+voice] transitions. The frequency of F_1 at voicing onset frequency is much higher for [−voice] stops than for [+voice] stops. Since voicing onset occurs much later than the release occurs in [−voice] stops, F_1 is not excited until very late in the CV transition, at which time the vocal tract is close to the vowel steady-state configuration. This delay in the excitation of F_1 is also known as F_1 cutback (Liberman *et al*., 1958).

Early work in speech perception has verified the role of VOT in the perception of the [voice] contrast in utterance-initial position for speakers of English and other languages (Liberman *et al*., 1958; Lisker & Abramson, 1970). In perceptual studies of the [voice] contrast, the role of VOT is quantified by a number known as the VOT boundary value. Listeners are more likely to classify stops with VOT values below the boundary value as [+voice] (e.g., /b d g/). Stops with VOT values above the VOT boundary are more likely to be classified as [−voice] (e.g., /p t k/). A number of studies show that, in addition to VOT, the effects of F_1 transition and frequency at voicing onset are important in the perception of [voice] (Stevens & Klatt, 1974; Lisker, Liberman, Erickson, Dechovitz & Mandler, 1977; Summerfield & Haggard, 1977; Kluender, 1991; Pind 1999). These studies show that F_1 transition properties and VOT have a trading relation (see Repp (1982) for a review of trading relations) for the [voice] contrast. The relationship between VOT and F_1 in [voice] classification is such that an increase (decrease) in VOT can be offset by lengthening (shortening) the F_1 transition and lowering (raising) the frequency of F_1 at voicing onset. Thus, VOT and F_1 transition manipulations both have equivalent effects on [voice] perception.[2] The effect of F_1 transition manipulations on [voice] classification can be quantified in terms of the resulting shift in the VOT boundary for [voice] classification.

1.2. *Interaction between [voice] and place of articulation*

Both VOT and F_1 transition characteristics can be viewed as the phonetic or acoustic correlates of the [voice] contrast, on the basis of both acoustic and perceptual studies. However, there is some evidence that the phonologically orthogonal contrasts of place of articulation and [voice] interact with each other, both acoustically and perceptually. While the realization of the [voice] contrast is similar across different places of articulation, a limited covariation between place of articulation and VOT has been observed in a variety of languages, with bilabial stops having shorter VOT values than alveolar stops, which in turn have shorter VOT values than velar stops (Fischer-Jørgensen, 1954; Peterson & Lehiste, 1960; Lisker & Abramson, 1964). This covariation has been verified recently in studies of variation in VOT in a variety of contexts within a single language (Cooper, 1991*a*, *b*; Docherty, 1992; Jessen, 1998), as well as in the cross-linguistic study by Cho & Ladefoged (1999). Aerodynamic, physiological, and gestural timing proposals have been put forth to explain the observed pattern of covariation between place and

[2] Stevens & Blumstein (1981) argue for an alternative point of view, in which the [voice] contrast is conveyed by the integrated acoustic property of low-frequency periodic energy in the vicinity of the release, encompassing not only F_1 cutback cues but also f_0 differences and intensity of aspiration differences. While this view introduces another level of representation, it is consistent with the view that VOT and F_1 transition properties trade with each other perceptually in [voice] categorization.

VOT. The aerodynamic explanations suggest that the smaller air cavity behind more posterior constrictions and/or the concomitant larger cavity in front of the constriction delay initiation of vocal fold vibration (Hardcastle, 1973; Maddieson, 1997; Stevens, 1998). On the physiological side, the greater mass and contact area involved in a velar constriction may result in releases that are slower than releases from anterior constrictions (Hardcastle, 1973; Summerfield, 1974; Diehl & Kluender, 1987). A third proposal is that vocal fold opening durations are constant across place of articulation, with the result that aspiration intervals are increased to account for shorter closure intervals for velar stops (Zue, 1976; Weismer, 1980; Maddieson, 1997; cf. Lisker & Abramson, 1964; Umeda, 1977).

Interaction between the perception of [voice] and the perception of place of articulation has also been investigated. Some studies have found place of articulation to affect [voice] classification (Kuhl & Miller, 1975, 1978; Miller, 1977), paralleling covariation between VOT and place of articulation in production. In these studies, listeners classified bilabial stops as [−voice] more often than velar stops, consistent with the observed covariation in production. Alveolar stops were classified intermediately between bilabials and alveolars. Another set of studies (Lisker, 1975; Kluender, 1991) have failed to confirm any interaction between the perception of these two phonological categories. In the remainder of this section I discuss the findings and shortcomings of these studies, and the implications for the present investigation of whether the perception of place of articulation and [voice] interact with each other.

In a series of experiments aimed at showing the cross-species generality of sensitivity to VOT, Kuhl & Miller (1975, 1978) obtained chinchilla and human identification functions for [voice] judgments on a VOT continuum for different places of articulation. Kuhl and Miller found a strong similarity between the human and nonhuman categorization results, including an apparent perceptual interaction between [voice] and place of articulation, with shorter boundary value for bilabials (27 ms) than alveolars (35 ms), and a shorter boundary value for alveolars than velars (42 ms). However, Kuhl and Miller note that their results are ambiguous as to whether place of articulation actually affects [voice] categorization because the stimuli confound F_1 transition with place of articulation. Kuhl and Miller (as well as many other studies) used the formant trajectories specified in Lisker & Abramson (1970). Each place of articulation in these stimuli has a different F_1 transition, such that the bilabial continuum has a sharper F_1 transition than the alveolar continuum, which in turn has a sharper transition than the velar continuum.

Using stimuli with identical F_1 transitions to avoid the confound in the Lisker & Abramson (1970) stimuli, Miller (1977) obtained distinct boundary values for each place of articulation. Her results confirm a role for place of articulation in [voice] categorization, with a bilabial boundary value of 25 ms, an alveolar boundary value of 28 ms, and a velar boundary of 29 ms. Since the continua for each place of articulation had identical F_1 transitions, the different boundary values for different places of articulation provide support for perceptual interaction between [voice] and place.

In another cross-species study, Kluender (1991) presents data that show that humans and Japanese quail are very similar in [voice] categorization of synthetic stimuli varying in VOT, F_1 transition pattern, and place of articulation. While [voice] categorization for both species was largely predictable from VOT and F_1 transition pattern, consistent with a general auditory explanation for the effect of F_1 on [voice] categorization, place of articulation had no significant effect on [voice] categorization. The frequency of F_1 at

voicing onset strongly affected [voice] categorization as predicted, and explained most of the variation in VOT boundary values, while F_1 transition duration accounted for less of that variation, consistent with Summerfield & Haggard (1977). The lack of effect of place of articulation is inconsistent with the place of articulation hypothesis and the findings for F_1 transition characteristics strongly support the F_1 transition hypothesis.

As Kluender notes, the results reported in Kuhl & Miller (1975, 1978) do not address the issue of perceptual interaction between [voice] and place of articulation because the Lisker & Abramson (1970) stimuli do not control for F_1 transition across place of articulation. Among these three studies, then, the only inconsistency is the positive and negative findings of an effect of place of articulation on [voice] categorization by Miller (1977) and Kluender (1991), respectively. One possible reason for the negative finding in Kluender (1991) is that the stimuli in that study lacked bursts, while the stimuli used in Miller (1977) contained bursts, providing more acoustic cues to place of articulation. Additionally, given the small effect size of place on [voice] categorization as found by Miller (1977), other sources of variability may have prevented a positive result from reaching significance in Kluender (1991). The implications for the present study are (1) both place of articulation and F_1 transition should be controlled as separate factors, (2) cues to place of articulation should be as robust as possible, and (3) enough data should be collected to ensure sufficient statistical power.

The experiment reported below investigates whether place of articulation affects the perception of [voice] by measuring how VOT continua are labeled when F_1 transition and place of articulation are manipulated in a full factorial design. Higher [−voice] categorization rates, or lower VOT boundary values, for bilabial and/or alveolar stops than for velar stops, independent of F_1 transition characteristics, would support a role for place of articulation affecting the perception of [voice].

The F_1 transition manipulations are expected to affect [voice] categorization rates as has been reported by the several studies cited above. The inclusion of F_1 transition as a factor serves two purposes: (1) to provide a replication of the experiments reported by Kluender (1991), in which F_1 transition pattern was a factor along with place of articulation and VOT, and (2), to provide a comparison of effect size for any potential place of articulation effect.

2. Method

To test the effects of place of articulation and F_1 transition on [voice] classification, forced-choice consonant classification data of CV and VCV stimuli from 12 subjects were collected using a fully-crossed design with the following factors: VOT (0–65 ms in 5 ms steps), place of articulation (bilabial, alveolar, and velar), F_1 transition (four F_1 transitions varying in onset frequency and duration), context (initial, or CV and intervocalic, or VCV), and pairing (anterior and posterior). The pairing manipulation was necessary because the response boxes only permitted four different responses per trial, so that continua from only two different places of articulation could be presented in a single block.

The context manipulation (initial or intervocalic) provides for two degrees of place of articulation information. In the initial condition with CV stimuli, all of the place of articulation cues reside in the burst and higher formant transitions. In the intervocalic condition with VCV stimuli, additional cues are present in the VC transition.

2.1. *Stimuli*

The stimuli were 24 VOT continua synthesized with the Sensyn implementation of the cascade branch of the Klatt synthesizer (Klatt & Klatt, 1990) at 12-bit resolution and 10 kHz sampling rate. The fully-crossed design of the place of articulation, F_1 transition pattern, and context factors (three places × four F_1 transition patterns × two contexts) yielded 24 continua. All stimuli were matched for peak amplitude. Stimuli with the stop in initial context (CV stimuli) were 400 ms long, and the stimuli with the stop in intervocalic context (VCV stimuli) were 600 ms long.

The F_1 transitions varied in onset frequency. The F_1 onset frequency and transition duration were selected to be comparable to those of the stimuli in Kluender (1991). Following consonant release, bandwidth of F_1 (B1) was constant at 130 Hz. Both onset and duration manipulations were factorially combined for four combinations: high/short (450 Hz onset, 30 ms duration), high/long (450 Hz onset, 60 ms duration), low/short (180 Hz onset, 30 ms duration), and low/long (180 Hz onset, 60 ms duration).

Variation in VOT, from 0 to 65 ms in 5 ms steps, was controlled by manipulating the AF, AV and AH synthesis parameters. All stimuli began with a 10 ms release burst using AF. Voice onset was initiated by ramping up AV from 20 to 60 dB over 5 ms. Between the release burst and voice onset, the AH noise source at 36 dB excited the formants F_2, F_3, F_4, and F_5. For all stimuli, f_0 started at 100 Hz, rose linearly to 130 Hz during the middle of the vowel, then fell linearly to 70 Hz for the end of the vowel. The waveform of each stimulus was examined to confirm the actual VOT. For measurement purposes, measured VOT was defined as the interval between the release burst and the zero-crossing that began the first pulse to reach approximately 5% of the steady-state vowel amplitude. The 5% criterion was adopted because the slight overlap between the AH and AV sources at voice onset made determination of the exact instant of voice onset difficult, while contributing to more natural sounding stimuli. This measurement method resulted in measured VOT values that corresponded closely but were slightly higher than the synthesizer VOT. The average difference was 12.1 ms, with $\sigma = 1.71$ ms. The changing f_0 at voice onset probably also contributed to the slight discrepancy between synthesizer and measured VOT. There was no covariation in this discrepancy with place of articulation or F_1 onset type.

Place of articulation cues for the CV stimuli consisted of F_2 and F_3 transitions and bursts appropriate for bilabial, alveolar, and velar stops. The formants above F_3 were constant, with $F_4 = 3250$ Hz, $F_5 = 3700$ Hz, and $F_6 = 4990$ Hz. Release bursts were synthesized by using the frication source (AF) to briefly excite the formants above F_1 for the first 10 ms of the stimulus. For the bilabials, the AF source was set to 45 dB at release, 35 dB at 5 ms after release, and turned off afterwards. The alveolar and velar stimuli had AF set to 49 dB at release, 40 dB at 5 ms after release, and turned off afterwards. Measurements of the rms value from a 20 ms window centered on the release indicated that the formant patterns interacted with the AF settings. The alveolar burst (32.6 dB SPL) was stronger than the velar (18.9 dB SPL) and bilabial (17.0 dB SPL) bursts.

The VCV stimuli were constructed by preposing a 140 ms neutral vowel and a 60 ms silent gap to the CV stimuli. The end of the neutral vowel contained formant transitions appropriate for the bilabial, alveolar, and velar places of articulation matched to the formant transitions of the rest of the stimulus. The neutral vowel was synthesized with the AV parameter 10 dB lower than for the main vowel so that the consonant was perceived as pretonic (preceding a stressed vowel). The F_1 offset frequency of the schwa at

consonant closure, 180 Hz, and length of the 60 ms silent gap that followed were selected so that the VOT crossover would occur roughly in the middle of the VOT continuum.

2.2. *Subjects*

Twelve adults were recruited from the University of Massachusetts community to participate in the study and were paid \$6/hr for their time. All listeners were phonetically naive native speakers of English, and none reported any hearing problems.

2.3. *Procedure*

Four subjects at a time were assigned to separate PC-controlled response stations in a quiet room. Stimuli were presented binaurally over TDH-39 headphones, and the listeners adjusted the volume to a comfortable level. Classification data were collected via the response boxes over six 2-hr sessions, no more than one session per day. Each session consisted of twelve 7-min blocks with frequent breaks.

Because the response stations had only four buttons, trials were blocked by context (initial or intervocalic) and partially by place of articulation. Three different pairings of place of articulation were used. Each block contained tokens from a single context, two different places of articulation, all four F_1 transition patterns, and all VOT values. For example, the bilabial/alveolar pairing had labels "bah/pah/dah/tah" for the initial context (CV stimuli) condition and "uh-bah/uh-pah/uh-dah/uh-tah" for the intervocalic context (VCV stimuli) condition.

The 112 stimuli (2 places of articulation $\times$ 4 F_1 transitions $\times$ 14 VOTs $=$ 112 stimuli) in each block were presented approximately every 3 s in random order without replacement. While there are many more possible block orders than groups of subjects, making complete counterbalancing impossible, each group performed the blocks in a different order, subject to the constraint that consecutive blocks always differed in both pairing and context. A minimum of 20 judgments per stimulus per subject were collected. Some subjects were able to participate in extra sessions, so that up to 30 judgments were collected for certain stimuli for those subjects. However, the results from each subject were weighted equally in the population-averaged statistical model below.

2.4. *Results*

The results are presented first by plots of [voice] classification as a function of synthesizer VOT with place of articulation (Fig. 1) and F_1 transition pattern (Fig. 2) as parameters. Second, statistical significance is assessed with a logistic regression analysis (Tables I and II). Finally, the relation between [voice] categorization and the main variables of VOT, F_1 transition pattern, and place of articulation is quantified using the results of the logistic regression analysis (Fig. 3).

In the broadest terms, the effects of both place of articulation and F_1 transition manipulations are significant and show limited interaction. Bilabials and alveolars were categorized more often as [$-$voice] than velars. The high/short F_1 transition gave rise to more [$-$voice] judgments than the low/long transition, with the other two F_1 transition patterns patterning in between. The factors of context and pairing and their interactions are either insignificant or have effect magnitudes that are much smaller than those of place of articulation and F_1 transition.

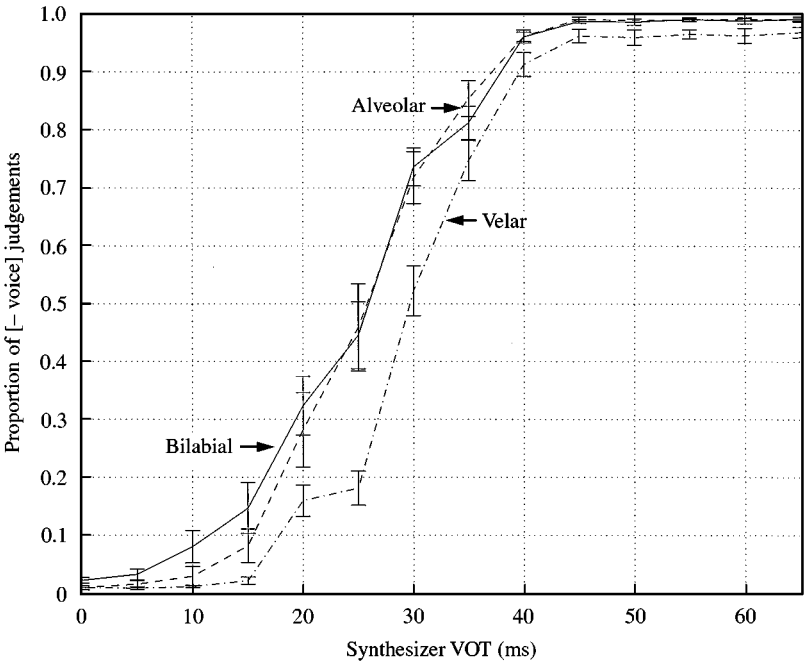

Figure 1. Classification of [voice] by place of articulation, pooled across F_1 transition pattern and context condition. The mean proportions across subjects of /p/, / t/, and /k/ responses for each value of synthesizer VOT are plotted with standard error bars.

Any misclassifications by place, such as a "bah" response for a stimulus with velar formant transitions, were excluded from consideration. Approximately, half of the total place misclassifications were due to one subject. Most of this subject's misclassifications, constituting 14% of the subject's total responses, were in the beginning sessions. The subject otherwise patterned like the other subjects. The misclassifications constitute 2.4% of the total responses, indicating that place of articulation information was robust.

2.4.1. *Identification functions*

Identification functions for [voice] in Figs 1 and 2 show VOT boundary shifts arising from both place of articulation and F_1 transition variation, averaged across subjects. Collapsing across all factors besides place of articulation, the functions in Fig. 1 present the results with place of articulation as a parameter (three lines). The bilabial and alveolar functions show a shorter VOT boundary value than the velar function, indicating that bilabials and alveolars more likely than velars to be classified as [−voice]. Fig. 2 presents the results with F_1 transition pattern as a parameter (four lines) and collapsing across all other factors. The high/short F_1 transition gave rise to the most [−voice] judgments, while the low/long F_1 transition correlated with the least [−voice] judgments. As expected, the F_1 transitions with conflicting effects of duration and onset frequency produced intermediate proportions of [−voice] judgments, with the low/short transition producing slightly more [−voice] judgments than the high/long transition.

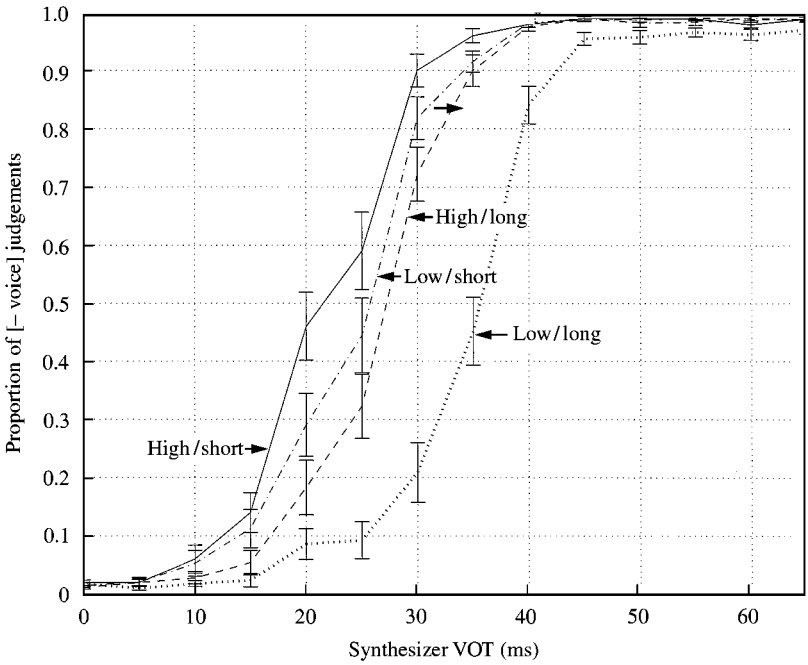

Figure 2. Classification of [voice] by F_1 transition pattern, pooled across place of articulation and context condition. The mean proportions across subjects of [−voice] responses (/p t k/) for each value of synthesizer VOT are plotted with standard error bars, parameterized by F_1 transition pattern (low/long, high/long, low/short, and high/short).

2.4.2. *Logistic regression analyses*

While the results of the binary data of the present study could be evaluated for statistical significance using methods more traditional in speech perception, such as an ANOVA on the boundary values, a logistic regression technique was used to evaluate the results. Logistic regression (LR) models of [voice] classification were generated using the maximum-likelihood method, which generates coefficient estimates that make the observed pattern of data most likely (for a thorough introduction to logistic regression, see Hosmer & Lemeshow, 1989). The LR framework models the probability of a [−voice] judgment as

$$p([-voice]) = \frac{1}{1 + e^{-\beta}} \tag{1}$$

where β is the linear combination in Equation (2) of predictor variables $X_1, \ldots, X_n$:

$$\beta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{2}$$

The coefficients $\beta_1, \beta_2, \cdots, \beta_n$ are estimated and assessed for statistical significance. Note that the predictor variables can be either categorical or continuous. The sign on a coefficient β_n indicates the direction of the effect, so that a positive β_n coefficient

TABLE I. Pooled logistic regression model of [−voice] classification. The independent variables are VOT, F_1, PLACE, CONTEXT, PAIRING, $F_1 \times$ PLACE, PLACE × CONTEXT, and PLACE × PAIRING

Variable	β	95% CI	G	df	Sig
Constant	− 5.8385	(− 5.9304, − 5.7466)			
VOT	0.2160	(0.2127, 0.2193)	73911.218	1	0.0000
F_1			5291.009	3	0.0000
Low/long	− 1.6712	(− 1.7237, − 1.6187)			
High/long	0.0709	(0.0246, 0.1172)			
Low/short	0.5294	(0.4826, 0.5762)			
High/short	1.0709	(1.0220, 1.1198)			
PLACE			1601.607	2	0.0000
Bilabial	0.4700	(0.4316, 0.5084)			
Alveolar	0.3067	(0.2687, 0.3447)			
Velar	− 0.7767	(− 0.8163, − 0.7371)			
CONTEXT			31.437	1	0.0000
Initial	− 0.0766	(− 0.1034, − 0.0497)			
Intervocalic	0.0766	(0.0497, 0.1034)			
PLACE × F_1			163.993	6	0.0000
Bilabial × low/long	− 0.1696	(− 0.2351, − 0.1041)			
Bilabial × high/long	− 0.2616	(− 0.3273, − 0.1959)			
Bilabial × low/short	0.2333	(0.1679, 0.2988)			
Alveolar × low/long	0.1863	(0.1210, 0.2516)			
Alveolar × low/short	0.1462	(0.0809, 0.2115)			
Alveolar × high/long	− 0.0689	(− 0.1342, − 0.0036)			
PLACE × CONTEXT			n/a	2	0.6576
PAIRING			12.374	1	0.0004
Anterior	− 0.0480	(− 0.0747, − 0.0213)			
Posterior	0.0480	(0.0213, 0.0746)			
PLACE × PAIRING			21.911	2	0.0000
Anterior × bilabial	0.0827	(0.0449, 0.1205)			
Anterior × alveolar	− 0.0723	(− 0.1099, − 0.0347)			
	Log-likelihood = − 17623.048				

indicates a positive correlation between the variable and the event, while a negative coefficient indicates a negative correlation.

In principle, both a boundary value analysis and a LR analysis of the same data will lead to the same conclusions. Furthermore, with Equations (1) and (2), LR coefficients can be easily translated into boundary values such as those that might be otherwise obtained by a probit analysis or interpolation of an identification function. However, there are a number of advantages of a LR analysis in comparisons of the effects of the different variables within the same study, as well as in other experiments. Exponentiating each LR coefficient results in an odds ratio, which represents how much more likely a [−voice] response is when the given factor is present in the stimulus. The odds ratios

TABLE II. Population-averaged logistic regression model of [−voice] classification. The independent variables are VOT, F_1, PLACE, CONTEXT, PAIRING, $F_1 \times$ PLACE, PAIRING, and PLACE × PAIRING

Variable	β	95% CI	t (df = 11)	Sig	α_F/k	$\alpha_F/7k$
Constant ($k = 1$)	− 7.2534	(− 8.3457, − 6.1610)	− 14.6150	**0.0000**	0.0500	0.0071
VOT ($k = 1$)	0.2703	(0.2340, 0.3066)	16.4028	**0.0000**	0.0500	0.0071
PLACE ($k = 3$)						
Bilabial	0.5558	(0.1042, 1.0074)	3.4708	**0.0052**	0.0167	0.0024
Alveolar	0.3686	(− 0.0581, 0.7952)	2.4363	0.0330		
Velar	− 0.9244	(− 1.2437, − 0.6051)	− 8.1641	**0.0000**		
F_1 ($k = 4$)						
Low/long	− 2.1073	(− 2.6926, − 1.5219)	− 10.7315	**0.0000**	0.0125	0.0018
High/long	0.0920	(− 0.0568, 0.2408)	1.8422	0.0925		
Low/short	0.6569	(0.4399, 0.8738)	9.0241	**0.0000**		
High/short	1.3584	(0.9695, 1.7474)	10.4119	**0.0000**		
$F_1 \times$ PLACE ($k = 6$)						
Bilabial × low/long	− 0.2295	(− 0.5335, 0.0745)	− 2.4215	0.0339	0.0083	0.0012
Bilabial × high/long	− 0.3153	(− 0.4220, − 0.2086)	− 9.4782	**0.0000**		
Bilabial × low/short	0.2965	(0.1087, 0.4842)	5.0648	**0.0004**		
Alveolar × low/long	0.2256	(0.0041, 0.4470)	3.2679	**0.0075**		
Alveolar × high/long	0.1821	(0.0526, 0.3116)	4.5099	**0.0009**		
Alveolar × low/short	− 0.0759	(− 0.1763, 0.0244)	− 2.4267	0.0336		
CONTEXT ($k = 2$)						
Initial	− 0.0928	(− 0.3851, 0.1995)	− 0.8230	0.4280	0.0250	0.0036
Intervocalic	0.0928	(− 0.1995, 0.3851)	0.8230	0.4280		
PAIRING,						
PAIRING × PLACE ($k = 3$)						
Anterior	− 0.0669	(− 0.1354, 0.0016)	− 2.7558	0.0187	0.0167	0.0024
Anterior × bilabial	0.1164	(0.0376, 0.1953)	4.1630	**0.0016**		
Anterior × alveolar	− 0.0956	(− 0.1930, 0.0018)	− 2.7673	0.0183		

can be used to compare experimental factors such as place of articulation and F_1 transition pattern with each other directly rather than through a third factor such as VOT. These odds ratios can be compared with those from other studies as well. Finally, LR analysis directly reports separate model parameters representing the boundary value, or "null effect" for all variables (the constant β_0), and the effects on categorization for each variable (the β_n). For continuous variables, such as VOT, the slope of the identification function can also be computed, although this was not carried out in the present study.

Two logistic regression models are presented, a pooled LR model computed by pooling all the subjects together, and a population-averaged LR model computed by averaging coefficient estimates across individual subject LR models. The pooled LR model serves as a rough evaluation of which variables are significant. However, a pooled analysis assumes that the subjects are homogenous, so any major conclusions should not be made without confirming subject homogeneity.

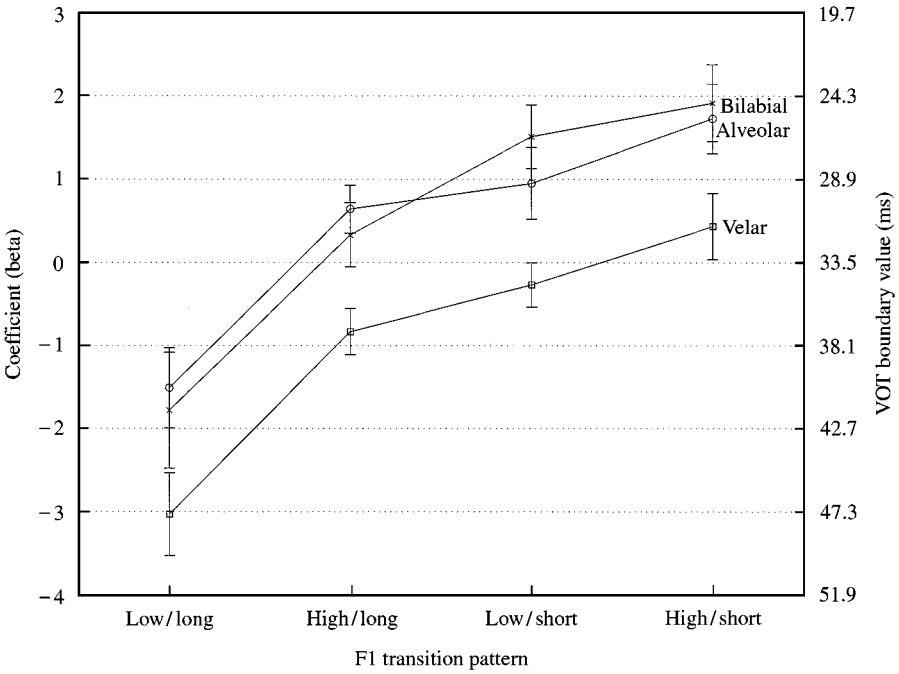

Figure 3. Population-averaged LR model coefficient means with 95% confidence intervals (not Bonferroni-adjusted) for F_1 transition pattern and PLACE. Each point represents the mean across subjects of the sum of the PLACE, F_1, and $F_1 \times$ PLACE coefficients for that particular F_1 and PLACE combination. The lines connecting points for each place of articulation are solely intended to highlight any interactions with F_1 transition pattern. The scale on the right side of the plot indicates the equivalent VOT boundary value.

An alternative to using a pooled model as the primary analysis tool is using the pooled analysis as a first step toward identifying the significant variables, but basing conclusions on a two-phase analysis following the approach described by Lorch & Myers (1990). This approach is used here. A LR model for each subject is estimated using the set of variables identified as significant or near-significant by the pooled analysis. The population-averaged model consists of the coefficients averaged across subjects. Each coefficient is evaluated for statistical significance using *t*-tests. The results of the *t*-tests serve to confirm or reject subject homogeneity.

Table I contains the coefficient estimates of the pooled LR model. The dependent variable (DV) is a [−voice] judgment, with the independent variables (IV) of VOT (continuous), PLACE (bilabial, alveolar, or velar), F_1 (low/long, high/long, low/short, high/short) CONTEXT (initial or intervocalic), PAIRING (anterior or posterior). The pooled LR model also contains the interaction terms $F_1 \times$ PLACE, PLACE $\times$ CONTEXT, and PLACE $\times$ PAIRING. The coefficient estimate for each variable is given by β, which represents the change in the log odds associated with a one-unit change for continuous independent variables such as VOT (measured in ms). Categorical variables, such as PLACE, are deviation coded, meaning that the β for each level of a variable (for example, PLACE has levels bilabial, alveolar, and velar) represents the effect on the log odds of that particular category relative to the geometric mean of all the categories for

that variable. The confidence limits on each estimate β_i are calculated by $\beta_i \pm z_{1-\alpha/2} \times \text{S.E.}(\beta_i)$ for $\alpha = 0.05$.

Note that under the deviation coding scheme, the values of the final levels of PLACE (velar) and F_1 (high/short) are linear combinations of the other coefficients. To compute the confidence intervals of the coefficients that are linear combinations of the linearly independent coefficients, the sample variance $\hat{\sigma}_L^2$ is needed. The sample variance of a linear combination is given by

$$\hat{\sigma}_L^2 = \sum_j \hat{\sigma}_j^2 + \sum \sum_{j \neq j'} r_{jj'}, \hat{\sigma}_j \hat{\sigma}_{j'}$$

where $\hat{\sigma}_j^2$ is the variance of the jth variable and $r_{ij'}, \hat{\sigma}_j \hat{\sigma}_{j'}$ is the covariance of the jth and j'th variables (Hosmer & Lemeshow, 1989, pp. 52–53).

In addition to the corresponding coefficient estimates and confidence intervals, each variable is also presented with the G statistic, which represents the change in deviance when the variable is removed from the model (Hosmer & Lemeshow, 1989, p. 14), and the significance level of G for the corresponding degrees of freedom. The G statistic is chi-square distributed with degrees of freedom as noted. The overall goodness-of-fit of the model is quantified by the log-likelihood value.

The pooled LR model shows that the three main variables, VOT, F_1, and PLACE, are each significant ($p < 0.00005$ for each variable). Other variables, CONTEXT ($p < 0.00005$), PAIRING ($p < 0.0004$), $F_1 \times$ PLACE ($p < 0.00005$), and PLACE $\times$ PAIRING ($p < 0.00005$) are also significant but explain quite modest amounts of variation, as reflected by the coefficient magnitudes and goodness-of-fit values being an order of magnitude smaller than the three main variables, with the exception of $F_1 \times$ PLACE. The PLACE $\times$ CONTEXT interaction variable does not have a significant G ($p < 0.6576$), and therefore is excluded from the population-averaged LR model.

The PAIRING variable and the interaction variable PLACE $\times$ PAIRING are included to check for any effects of particular place of articulation combinations. The anterior value for the dichotomous PAIRING variable represents those trials in which the *other* place of articulation in the block was anterior. For example, for trials with bilabial stimuli, the anterior trials were blocked with alveolars while the posterior trials were blocked with velars.

Coherence across subjects is tested in the population-averaged LR model in Table II. The population-averaged LR model contains the same independent variables as the pooled LR model except for the non-significant variables. First, a LR model for each listener was computed using all of the significant variables in Table I. Second, the coefficients were evaluated for significance using two-tailed single-group t-tests at the 0.05 significance level. Because several comparisons are involved, the Bonferroni method was used to set the familywise Type I error rate at $\alpha_F = 0.05$, and to expand the confidence intervals. The coefficients were divided into seven families of k comparisons each. The Bonferroni expansion of confidence intervals for each estimate β_i, were calculated using the formula $\beta_i \pm t \times \text{S.E.}(\beta_i)$, where the critical value for t was selected using df $= 11$, $\alpha = 0.05$, and the appropriate value of k for the family (see Myers & Well, 1991, Table D.8, p. 629). The unadjusted t statistic for each test is reported. In the fifth column, significance is declared (indicated by bold typeface) for a particular test if the significance level falls below that of α_F/k. The Bonferroni significance level for each family is indicated in column 6. A stricter criterion, $\alpha_F/7k$, is shown in the final column, which was computed

by dividing the Bonferroni significance level for each family by 7, the total number of families.

As in the case of the pooled LR model, the values of the final levels of PLACE (velar) and F_1 (high/short) in Table II are linear combinations of the other coefficients. These values were computed for each subject, and formed the sample for computing means and standard errors (used to calculate the confidence level), just as with the linearly independent levels of PLACE and F_1. Equivalently, the means and standard errors could have been computed from the population-averaged means, variances, and covariances of the linearly independent levels of PLACE and F_1.

According to the t-tests and confidence intervals, the effects of VOT, F_1 transition pattern, place of articulation, and one coefficient of the pairing variable are significant at both the Bonferroni-adjusted 0.05 level as well as the stricter $\alpha_F/7k$ level. The effect of F_1 is also highly reliable in that no overlap is shown for the confidence intervals of the different coefficient estimates. Reliable differences between the low/long and low/short as well as the high/long and high/short indicate that transition duration is an important contributor to [voice] categorization. Similarly, reliable differences between the low/long and high/long as well as the high/short and low/short indicate that F_1 frequency at voicing onset is also an important contributor to [voice] categorization. For PLACE, the confidence intervals for bilabials and alveolars overlap with each other while neither overlap with the estimate for velars, indicating that bilabials and alveolars together pattern distinctly from velars. The interaction term between PLACE and F_1 is significant. Therefore, the coefficient values are interpreted together combining the effects of both variables in Section 2.4.3.

The effect of CONTEXT is not reliable across subjects. The confidence intervals do not overlap for the remaining variable, PLACE × PAIRING, indicating a small but reliable effect across subjects for the pairing manipulation. Bilabials are slightly more likely to be categorized as [− voice] when paired with alveolars than with velars. Approaching but not quite reaching significance is the effect of alveolars being slightly more likely to be categorized as [− voice] when paired with velars than with bilabials. The size of these effects are small, and are discussed below in terms of odds ratios. No obvious explanation is available for the interaction between pairing and place of articulation.

2.4.3. *Interpretation of F_1 and PLACE coefficients*

As mentioned above, the effects of F_1 transition, PLACE, and F_1 × PLACE coefficients should be interpreted together, since the interaction variable F_1 × PLACE is significant. Fig. 3 shows the sum of the mean logistic coefficient estimates for each combination of F_1 transition and PLACE. For example, the coefficient sum for the bilabial low/long combination is the sum of the bilabial coefficient, the low/long coefficient, and the bilabial × low/long coefficient. These sums were computed for each subject and then averaged across subjects for Fig. 3. A critical value of $t = 2.201$ was selected for computing the 95% confidence intervals (two-tailed, df = 11), which are not Bonferroni-adjusted. The scale on the right side of Fig. 3 shows the equivalent VOT boundary values.

Bilabials with the high/short F_1 transition are more likely to be judged [− voice] than any other F_1–PLACE combination, and have the highest coefficient sum at slightly greater than 2.0. On the other extreme, velars with the low/long F_1 transition are least

likely to lead to [−voice] judgments, having the lowest coefficient sum at slightly less than −3.0.

The main effect of PLACE is shown by all the velar coefficient sums being lower than the corresponding bilabial and alveolar coefficients for all F_1 transitions. Bilabial and alveolar coefficient sums seem to overlap, however, suggesting no real difference between these two places of articulation. Indeed, the confidence intervals of the means of the bilabial and alveolar coefficients (without taking into account the interaction terms) overlap, as shown in Table II.

Likewise, the main effect of F_1 transition is manifested by the high/short coefficient sums being higher than the corresponding low/short coefficients, which are in turn higher than the corresponding high/long coefficient sums, which are higher than the low/long coefficient sums, for all three places of articulation. Both the F_1 onset frequency manipulation (high *vs.* short) and the F_1 transition duration (low *vs.* long) show substantial effects, in that all four F_1 transition patterns result in different classification rates for a given place of articulation.

The nonparallelism of the lines in Fig. 3 is diagnostic of a statistically significant interaction between place and F_1 transition, also indicated by some of the $F_1 \times$ PLACE terms in Table II reaching significance. The significance of $F_1 \times$ PLACE appears to result from bilabials with long F_1 transitions being slightly more likely than alveolars to be classified as [−voice] alveolars, with the situation reversed for the short F_1 transitions.

The magnitudes of the effects of VOT, F_1 transition pattern, place of articulation, and pairing can be directly compared with each other using odds ratios, computed by exponentiating the difference between appropriate coefficients from Table II. The odds ratio ψ represents how much more likely or unlikely a [−voice] judgment is for stimuli with a particular factor, for an ambiguous setting of all of the other variables. Selected odds ratios are displayed in Table III.

The VOT odds ratio of $\psi(5\text{ ms VOT}) = 3.8629$ means that stimuli with additional 5 ms of VOT make stimuli 4 times more likely to be perceived as [−voice] within the boundary region. Doubling the exponential results in squaring the odds ratio, so $\psi(10\text{ ms VOT}) = 14.9220$. For the categorical variables of F_1 transition pattern and place of articulation, the odds ratios in Table III represent the difference in likelihood for the two different settings involved. For example, ψ(high/short, low/long) = 31.9982 means that stimuli with the high/short F_1 transition are about 32 times more likely than stimuli with the low/long transition to be perceived as [−voice]. The magnitude of the F_1 transition effects found in the present study are substantially larger than the magnitude of the effect of place of articulation.

TABLE III. Selected odds ratios with 95% Bonferroni-adjusted confidence intervals using the population-averaged coefficient estimates from Table II

Variable	Odds ratio $\psi = e^{\beta}$	95% CI
ψ (5 ms VOT)	3.8629	3.2222, 4.6309
ψ (10 ms VOT)	14.9220	10.3829, 21.4454
ψ (high/short, low/long)	31.9982	12.2237, 83.7623
ψ (bilabial, velar)	4.3940	2.2811, 8.4638
ψ (anterior × bilabial)	1.1234	1.0383, 1.2156

The final odds ratio, the only pairing variable to reach significance, is slightly above 1.0, indicating a small but reliable (the CI does not include 1.0) effect of pairing. It is not a true odds ratio, since two levels of pairing are not being compared, but represents an effect relative to the average odds for the variable in question. The effect on likelihood, however, is much smaller than that of any of the other variables in the experiment.

2.4.4. *Summary of results*

The logistic regression analysis of [−voice] judgments confirms the roles of place of articulation and F_1 transition in [voice] classification. Logistic regression coefficients for bilabials and alveolars are significantly higher than for velars, indicating that [−voice] judgments are more likely for bilabials and alveolars than for velars, all other things being equal. Somewhat independently, F_1 transitions also affect [voice] classification, with probability of [−voice] judgments and coefficient size being highest for the high/short transition, followed by the low/short, high/long, and low/long transitions in that order. Both transition duration and frequency of F_1 at voicing onset contribute to [voice] categorization. Some interaction exists between place and F_1 transition, much of which is due to the effect of alveolars on [−voice] judgments being intermediate between bilabials and velars for the high/short F_1 transition. The effects of F_1 transition were much greater than those of place of articulation.

There were mixed results for the context and pairing factors. The context manipulation, intervocalic or initial position for the consonant, was not significant, either as a main effect or in interaction with place. The lack of a significant effect for the context-by-place interaction indicates that the additional place cues present in the intervocalic condition had no effect on [voice] classification. A slight though significant interaction between place and pairing occurred, such that [−voice] judgments were slightly more likely for bilabials when paired with alveolars than with velars, and [−voice] judgments were slightly more likely for alveolars when paired with velars than with bilabials. No obvious explanation exists for the pairing effect.

3. Discussion

3.1. *Comparison with previous studies*

The main results, that the place of articulation of a pretonic consonant and the F_1 transition into the following vowel both have a significant effect on [voice] classification, is partly consistent with Miller (1977) for place of articulation effects, and consistent with Kluender (1991) for F_1 transition effects.

With regard to the effect of F_1 transition pattern on [voice] categorization, the results of the present study are largely consistent with the findings of Kluender (1991). Higher F_1 onset frequencies and shorter transition durations (the present study indicates that the effect of transition duration is as strong as the onset frequency, while Kluender (1991) reports only a minor role for transition duration) increase the probability of a [−voice] percept. As for the basis of this effect, Kluender (1991) points out two sets of relevant findings: (1) that humans and nonhumans show similar effects of F_1 transition pattern on [voice] categorization of VOT continua, and (2) that nonsimultaneity judgments of two-component sinewave nonspeech analogues of VOT stimuli (Pisoni, 1977) show effects of lower component transition duration and onset frequency that are analogous

to effects of F_1 transition pattern on [voice] categorization (Hillenbrand, 1984; Parker, 1988). Kluender (1991) argues that the most parsimonious explanation for these converging results is that general auditory mechanisms, not any specialized speech adaptations, are responsible for effects of F_1 transition on [voice] categorization. Kluender & Lotto (1994) further argue that it is unlikely that the effects of F_1 transition pattern result from learning by either humans or nonhumans.

The place of articulation effect, that bilabials and alveolars are more likely than velars to be classified as [−voice], is consistent with Miller (1977), except that in the present study, alveolars pattern with bilabials instead of intermediately between velars and bilabials. The difference in likelihood between bilabials and velars, shown in Table III, is slightly greater than that resulting from about 5 ms VOT, which is close to the corresponding difference of 4.8 ms in Miller (1977). Furthermore, F_1 transition pattern and place of articulation show little interaction. In contrast with Miller (1977) and the present study, Kluender (1991) reports a null effect of place of articulation on [voice] categorization. I speculate that the lack of an effect of place of articulation could be due to decreased attention by the participants to the place of articulation contrast, possibly because of the lack of release bursts in the stimuli, and that the trials were blocked by place of articulation. In the present study, subjects categorized the stimuli—which had both release bursts and formant transitions—by [voice] *and* place of articulation.

Different release bursts across place of articulation do not seem likely as a simple explanation for the effect of place of articulation on [voice] categorization. Lisker (1975) notes that burst amplitude is a cue for [−voice] judgments in stops. While the velar stimuli in the present study had a higher AF setting than the bilabial stimuli, rms measurements indicate that the bilabial and velar burst amplitudes were comparable, while the alveolar burst amplitude was stronger. However, the alveolars patterned with the bilabials with respect to [voice] categorization. While burst amplitude has an effect on [voice] categorization, there is also an effect of place of articulation that does not appear to be explained by differences in bursts across place of articulation.

3.2. *Implications for phonological feature theory*

The dependence of [voice] categorization on place of articulation suggests the possibility that the basic phonological categories are not features but segments. If the decision rule for [voice] categorization depends upon the output of other phonological categorization process, then the perceptual processes investigated here might be better characterized as segment categorization rather than distinctive feature value categorization. Investigating how place of articulation influences [voice] categorization should help answer the question of whether features or segments are basic categories for speech perception.

There are a number of logically distinct manners in which the place of articulation features could influence [voice] categorization. A possible explanation for the present results is that [voice] categorization is dependent on the phonetic cues generally thought to determine the perception of place of articulation. In other words, the phonetic cues of release bursts and higher formant transitions are shared by the [voice] and place of articulation decisions. The idea that the perception of one particular event or category (such as the feature [voice]) follows from a variety of diverse stimulus attributes has a basis in the theory of perception developed by Egon Brunswik (for an overview of

Brunswik's work, see Postman & Tolman, 1959). In Brunswik's theory, various stimulus attributes—no single one of which is necessary or sufficient—with their own weights are used by the perceptual process in question.

A number of researchers have looked specifically at speech perception with the idea that multiple categories might be signaled by a number of shared stimulus attributes. The studies include Mermelstein's (1978) investigation of VC sequences in which the vowel and consonant are signaled by shared acoustic cues, the extension and reanalysis reported by Whalen (1989), the work by Massaro and colleagues on multimodal speech perception (Massaro, 1987; also see specific studies cited below), and work by Kingston and Diehl (1994) on cue-sharing between the [voice] contrast and other contrasts through a variety of mechanisms. Nearey's (1990, 1992, 1997) logistic regression models of VC and CV categorization provide a useful framework for modeling how orthogonal phonological contrasts are conveyed by potentially overlapping sets of acoustic–phonetic cues. His framework can be applied to the case of place of articulation and [voice] categorization to determine whether the features themselves or segments determined by particular combinations of place of articulation and [voice] values are basic units. Below, I discuss the perception of [voice] and place of articulation features using terminology adapted from Nearey's work on the perception of segments.

The results of the present study seem to rule out a *primary-cue feature model*, in which [voice] categorization is dependent solely on acoustic cues arising from VOT. The results are compatible, however, with a *secondary-cue feature model*, in which [voice] categorization is dependent both on acoustic cues arising from VOT and on the acoustic cues for place of articulation, such that bilabial and alveolar formant transitions condition more [−voice] percepts than velar acoustic cues. Both models of perception are termed feature models because they are both compatible with feature values as real categories. Note that under the secondary-cue feature model, an effect on the perception of place of articulation by VOT and other acoustic–phonetic cues traditionally associated exclusively with [voice] could be possible.

Another possible explanation for interaction between [voice] and place of articulation categorization is the presence of a *segment bias* in either the primary- or secondary-cue feature model favoring /p t g/ over /b d k/. One interpretation of the present results is that bilabials and alveolars have a shorter VOT boundary (hence /p t/ over /b d/) than velars (hence /g/ over /k/). Results from Whalen's (1989) extension and reanalysis of Mermelstein's (1978) data, and subsequent reanalysis in Nearey (1990), suggest constrained dependence between orthogonal phonological classifications. Nearey argues that mere cue-sharing between phonological categories is an insufficient explanation, but one possibility is a bias that shifts the boundaries between certain categories to favor some of the response categories at the expense of the others. For the consonant categorization data presented here, a segment bias could make /p t g/ responses more probable than /b d k/.

A third possibility is a *segment model*, in which the [voice] contrast, mediated by VOT and F_1 transition characteristics, is perceived in a different way for velars than for alveolars and bilabials. This outcome would mean that features such as [voice] and place of articulation are at best epiphenomena, and that the smallest recombinable phonological categories are segments.

Nearey (1997) speculates that the diphone bias in VC classification might represent a number of functions, including phonotactic constraints implemented as contextual features in the Massaro & Cohen (1983) fuzzy logical models, lexical preferences for

certain words or nonsense syllables, or optimal decision rules when the center of the decision space borders only a subset of the response categories. Since the stop consonants considered in the present study are all actual English phones in legal phonological contexts, it seems unlikely that a segment bias would represent phonotactic or inventory constraints. However, lexical biases or decision rules to explain center effects are possibilities.

Further investigation of the issues of how overlapping cues are perceived and the primacy of either features or whole segments as phonological categories will require experiments in which acoustic cues to both [voice] and place of articulation are varied systematically. A few previous studies have addressed the issue in this manner. Sawusch & Pisoni (1974) report classification data of CV stimuli drawn from a VOT continuum, a F_2–F_3 onset frequency (bilabial to alveolar) continuum, and a combined continuum ranging from /ba/ to /ta/. Their analyses suggest that [voice] and place of articulation categorization interact, but do not clarify the nature of the interaction. A series of experiments reported in Oden & Massaro (1978, 1980) in which listeners categorized elements of a stimulus array for [voice] and place of articulation, offer support for cue-sharing effects. The arrays consisted of synthetic speech CV tokens for which the cues to [voice] (VOT, and in one experiment, aspiration amplitude) were crossed with place of articulation (F_2 and F_3 transitions) over a range of levels appropriate for /bæ/, /pæ/, /tæ/, or /dæ/ responses. A complex pattern of cue-sharing between [voice] and place of articulation categories was found. High VOT values (and in the third experiment, higher aspiration amplitudes), appropriate for [−voice] stops, made bilabial judgments more likely than alveolar judgments. Additionally, in one of the experiments, high F_2 and F_3 onset frequencies, appropriate for alveolars, made [−voice] judgments more likely than [+voice] judgments. However, the direction of the latter cue-sharing effect is inconsistent with Miller (1977), as Oden and Massaro acknowledge, and the present study. Recall that Miller (1977) reports that alveolar stimuli lead to more [+voice] judgments than bilabials do, and that in the present study, velars are more likely than alveolars or bilabials to be classified as [+voice].

While Oden & Massaro (1980) assume that syllables, not phonemes or features, are basic categories (their footnote 1), their fuzzy logical model analyses of results pooled across subjects support three explanations for their cue-sharing results equally well: (1) a secondary cue effect explanation analogous to the secondary-cue feature based model discussed above, with [voice] and place of articulation categorization dependent on the same acoustic cues; (2) low-level psychoacoustic interactions in which VOT could influence the perceived value of F_2 and F_3 onset, or F_2 and F_3 onset could influence the perceived value of VOT; and (3) the need for particular acoustic cue weights for particular segments, as implemented by what they call phoneme prototype modifiers. While the first two explanations for cue-sharing, secondary cue effects and psychoacoustic interactions, are logically distinct, they make very similar predictions, and it seems difficult to imagine an experiment that would distinguish the two explanations. Importantly, both hypotheses are compatible with a secondary-cue feature-based model of speech perception. The third explanation seems compatible with a segment or syllable-based model, but definitely not a feature-based model, since the phoneme prototype modifiers relate particular acoustic cues to particular response categories or segments. More stimulus array data, as well as analyses of the reliability of particular effects across subjects, such as the population-averaged LR model presented in this paper, are needed to resolve the conflict between segment- and feature-based models.

In a recent critique of Nearey (1997), Kluender & Lotto (1999) point out that Nearey's empirical approach to modeling speech perception as pattern recognition may provide an adequate description of classification but fail to provide an adequate explanation. They argue instead that explanations for speech perception are likely to be found in general auditory processes, among other areas, and that a successful logistic model of labeling behavior, which merely provides an accurate mathematical description, could obscure the underlying reasons for such behavior.

Such criticisms apply as well to the approach proposed here for uncovering the connections between place of articulation and [voice] categorization. The research program advocated by Kluender & Lotto (1999) of seeking explanations for phonetic phenomena *outside* of phonetics is an epistemologically rigorous model of investigation (see Lindblom (1980) and Ohala (1990) for further discussion of this approach to phonetics). However, the current state of the art in speech perception falls short of a complete description of the acoustic cues for phonological contrasts. The present study and the stimulus array experiments proposed here address that gap, which must be filled along the way to a more profound understanding of speech perception.

4. Conclusion

The logistic regression analysis of the [voice] categorization data presented here are two-fold. First, the role of F_1 transition characteristics in [voice] classification is confirmed, with the effects being consistent with previous studies such as Kluender (1991). High F_1 onset frequencies and short F_1 transition durations increase the probability of [−voice] percepts in pretonic consonants. This finding is entirely compatible with the view that the effect of F_1 transition on [voice] classification arises from general auditory mechanisms.

Second, the manipulation of place of articulation and F_1 transition pattern as independent factors permitted teasing apart the similar effects of those cues. Pretonic consonants with F_2 and F_3 transitions appropriate for bilabials and alveolars are classified as [−voice] more often than analogous stimuli with velar transitions, largely consistent with results reported by Miller (1977). This effect of place of articulation obtains independent of the well-known F_1 transition effects on [voice] classification, which were replicated here. Higher F_1 onset frequencies and steeper transitions made [−voice] judgments more likely, and for the range of manipulation in the present study, the effects of F_1 transition were larger than those of place of articulation.

The interdependence between place of articulation and [voice] classification could be compatible with the notion of features as psychological categories, depending on how the two decisions interact. If the interaction is that of a secondary cue dependence, in which [voice] and place of articulation depend on the same acoustic cues, or a segment bias, in which certain combinations of [voice] and place of articulation features are favored over other combinations regardless of acoustic cues, then a feature-based model of classification is viable. However, if it turns out that different segments require different weightings of acoustic cues, then the interaction reported here is consistent with segments, and not features, being basic categories in speech perception. A stimulus array classification study and analysis using the framework used by Nearey (1990, 1992, 1997) could help decide between the different explanations for the interdependence between place of articulation and [voice] classification.

This work was supported by Grant No. R-29-DC01708-2 from the National Institute of Deafness and Communicative Disorders, National Institutes of Health to John Kingston and was carried out in partial fulfillment of requirements for the PhD degree at the University of Massachusetts at Amherst. I am grateful for a predoctoral fellowship at the Department of Linguistics, Ohio State University while I worked on much of this project. Preliminary versions of this work were presented at the third joint meeting of the Acoustical Societies of America and Japan, Honolulu, Hawaii and colloquia at the Max-Planck-Institute for Psycholinguistics and University of Michigan. For important comments and advice, I thank dissertation committee members Lyn Frazier, Neil Macmillan, John McCarthy, as well as Pam Beddor, Keith Johnson, Cecilia Kirk, Keith Kluender, Terry Nearey, and audiences at MPI and the University of Michigan. Reviews and comments from Peter Bailey, Jorgen Pind, Terry Nearey, and Doug Whalen helped to improve this paper enormously. All errors are my own. The support and guidance of John Kingston are especially appreciated.

References

Caisse, M. (1982) *Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops*. MA thesis, University of California.

Cho, T. & Ladefoged, P. (1999) Variation and universals in VOT: evidence from 18 languages, *Journal of Phonetics*, **27,** 207–229.

Chomsky, N. & Halle, M. (1968) *The sound pattern of English*. New York: Harper & Row.

Cooper, A. (1991*a*) *An articulatory account of aspiration in English*. PhD dissertation, Yale University.

Cooper, A. (1991*b*) Laryngeal and oral gestures in English /p, t, k/. In *Proceedings of the international congress of phonetic sciences* 12 (Aix), **2,** 50–53.

Cooper, F. S., Liberman, A. M., Delattre, P. & Gerstman, L. (1952) Some experiments on the perception of speech sounds, *Journal of the Acoustical Society of America*, **24,** 597–606.

Diehl, R. L. & Kluender, K. R. (1987) On the categorization of speech sounds. In *Categorical Perception* (S. Harnad, editor), pp. 226–253. Cambridge: Cambridge.

Docherty, G. J. (1992) *The Timing of Voicing in English obstruents*. Berlin: Foris Publications.

Fischer-Jørgensen, E. (1968) Voicing, tenseness, and aspiration in stop consonants, with special reference to French and Danish. In *Annual report of the Institute of Phonetics*, **3,** 63–114. Copenhagen: University of Copenhagen.

Hardcastle, W. J. (1973) Some observations on the Tense-Lax distinction in initial stops in Korean, *Journal of Phonetics*, **1,** 263–271.

Hillenbrand, J. (1984) Perception of sine-wave analogs of voice-onset time stimuli, *Journal of the Acoustical Society of America*, **75,** 231–240.

Hosmer, D. W. & Lemeshow, S. (1989) *Applied logistic regression*. New York: John Wiley and Sons.

Jakobson, R., Fant, C. G. M. & Halle, M. (1951) *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge: MIT Press.

Jessen, M. (1998) *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: John Benjamins.

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

Klatt, D. H. & Klatt, L. C. (1990) Analysis, synthesis, and perception of voice quality variations among male and female talkers, *Journal of the Acoustical Society of America*, **87,** 820–857.

Kluender, K. R. (1991) Effects of first formant onset properties on voicing judgments result from processes not specific to humans, *Journal of the Acoustical Society of America*, **90,** 83–96.

Kluender, K. R. & Lotto, A. J. (1994) Effects of first formant onset frequency on voice judgments result from auditory processes not specific to speech, *Journal of the Acoustical Society of America*, **95,** 1044–1052.

Kluender, K. R. & Lotto, A. J. (1999) Virtues and perils of an empiricist approach to speech perception, *Journal of the Acoustical Society of America*, **105,** 503–511.

Kuhl, P. K. & Miller, J. D. (1975) Speech perception by the chinchilla: voiced–voiceless distinction in alveolar plosive consonants, *Science*, **190,** 69–72.

Kuhl, P. K. & Miller, J. D. (1978) Speech perception by the chinchilla: identification functions for synthetic VOT stimuli, *Journal of the Acoustical Society of America*, **63,** 905–917.

Liberman, A. M., Delattre, P. C. & Cooper, F. S. (1958) Some cues for the distinction between voiced and voiceless stops in initial position, *Language and Speech*, **1,** 153–167.

Lindblom, B. (1980) The goal of phonetics, its unification and application, *Phonetica*, **37,** 7–26.

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

Lisker, L. (1986) "Voicing" in English: a catalog of acoustic features signaling /b/ versus /p/ in trochees, *Language and Speech*, **29,** 3–11.

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

Lisker, L. & Abramson, A. S. (1970) The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th international conference of phonetic sciences* (B. Halá, M. Romportl & P. Janota, editors), pp. 563–567. Prague: Academia.
Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D. & Mandler, R. (1977) On pushing the voice onset time (VOT) boundary about, *Language and Speech*, **20,** 209–216.
Lorch, R. F. & Myers, J. L. (1990) Regression analyses of repeated measures data in cognitive research, *Journal of Experimental Psychology*: *Learning*, *Memory*, *and Cognition*, **16,** 149–157.
Maddieson, I. (1997) Phonetic universals. In *The handbook of phonetic sciences* (J. Laver & W. J. Hardcastle, editors), pp. 619–639. Oxford: Blackwells.
Massaro, D. W. (1987) *Speech perception by ear and eye*: *a paradigm for psychological inquiry*. Hillsdale: Erlbaum.
Massaro D. W. & Cohen, G. C. (1983) Phonological context in speech perception, *Perception and Psychophysics*, **34,** 338–348.
Mermelstein, P. (1978) On the relationship between vowel and consonant identification when cued by the same acoustic information, *Perception and Psychophysics*, **23,** 331–335.
Miller, J. L. (1977) Nonindependence of feature processing in initial consonants, *Journal of Speech and Hearing Research*, **20,** 519–528.
Miller, G. A. & Nicely, P. E. (1955) An analysis of perceptual confusions among some English consonants, *Journal of the Acoustical Society of America*, **27,** 338–352.
Myers, J. L. & Well, A. D. (1991) *Research design and statistical analysis*. New York: HarperCollins.
Nearey, T. M. (1990) The segment as a unit of speech perception, *Journal of Phonetics*, **18,** 347–373.
Nearey, T. M. (1992) Context effects in a double-weak theory of speech perception, *Language and Speech*, **35,** 153–172.
Nearey, T. M. (1997) Speech perception as pattern recognition, *Journal of the Acoustical Society of America*, **101,** 3241–3254.
Oden, G. C. & Massaro, D. W. (1978) Integration of featural information in speech perception, *Psychological Review*, **85,** 172–191.
Oden, G. C. & Massaro, D. W. (1980) Evaluation and integration of acoustic features in speech perception, *Journal of the Acoustical Society of America*, **67,** 996–1013.
Ohala J. J. (1990) There is no interface between phonology and phonetics: a personal view, *Journal of Phonetics*, **18,** 153–171.
Parker, E. M. (1988) Auditory constraints on the perception of stop voicing: the influence of lower-tone simultaneity, *Journal of the Acoustical Society of America*, **83,** 1597–1607.
Peterson, G. E. & Lehiste, I. (1960) Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, **32,** 693–703.
Pind, J. (1999) The role of *F*1 in the perception of voice onset time and voice offset time, *Journal of the Acoustical Society of America*, **106,** 434–437.
Pisoni, D. B. (1977) Identification and discrimination of the relative onset time of two component tones: implication for voicing perception in stops, *Journal of the Acoustical Society of America*, **61,** 1352–1361.
Postman, L. and Tolman E. C. (1959) Brunswik's probabilistic functionalism. In *Psychology*: *a study of a science*. *Vol. 1*. *Sensory*, *perceptual*, *and physiological formulations* (S. Koch, editor), pp. 502–564. New York: McGraw-Hill.
Repp, B. (1982) Phonetic trading relations and context effects: new evidence for a phonetic mode of perception, *Psychological Bulletin*, **92,** 81–110.
Sawusch J. R. & Pisoni, D. B. (1974) On the identification of place and voicing features in synthetic stop consonants, *Journal of Phonetics*, **2,** 181–194.
Stevens, K. N. (1998) *Acoustic Phonetics*. Cambridge: MIT Press.
Stevens, K. N. & Blumstein, S. E. (1981) The search for invariant acoustic correlates of phonetic features. In *Perspectives on the study of speech* (P. D. Eimas & J. L. Miller, editors), pp. 1–38. Hillsdale: Erlbaum.
Stevens, K. N. & Klatt, D. H. (1974) Role of formant transitions in the voiced–voiceless distinction of stops, *Journal of the Acoustical Society of America*, **55,** 653–659.
Summerfield, A. Q. & Haggard, M. P. (1977) On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants, *Journal of the Acoustical Society of America*, **62,** 435–448.
Summerfield, Q. A. (1974) Processing of cues and contexts in the perception of voicing contrasts. In *Preprints of the 1974 Stockholm speech communication seminar*, Vol. 3 (G. Fant, editor), pp. 77–86. Upsala: Almqvist and Wiksell.
Umeda, N. (1977) Consonant duration in English, *Journal of the Acoustical Society of America*, **61,** 846–858.
Whalen, D. (1989) Vowel and consonant judgments are not independent when cued by the same information, *Perception and Psychophysics*, **46,** 284–292.
Weismer, G. (1980) Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations, *Journal of Phonetics*, **8,** 427–438.
Zue, V. (1976) *Acoustic characteristics of stop consonants*. Indiana University Linguistics Club.