Journal of Phonetics (1998) **26,** 283–310
Article ID: jp980077

Stress-related variation in the articulation of coda alveolar stops: flapping revisited

Kenneth de Jong

Department of Linguistics, 319 Memorial Hall, Indiana University, Bloomington, Indiana 47405, USA

Received 28 May 1997, revised 23 January 1998, accepted 9 July 1998

The present paper examines the plausibility of two models of flapping in American English: (1) a traditional model of flapping as a categorical switch from stop to flap production in a specified linguistic environment, and (2) a model of flapping in which flapping arises as a by-product of articulatory changes associated with the general implementation of prosodic structure. These models are tested against a corpus of X-ray microbeam records of English speakers producing utterances with word-final coronal consonants in the appropriate segmental context for flapping, but in varied prosodic locations. Tokens were submitted to perceptual, acoustic, and articulatory analyses. Results show that listeners consistently transcribe the presence of flaps according to acoustic differences in the presence of voicing during closure and a release burst. Transcriptions and lingual measurements, however, suggest that the difference between flaps and [d] is associated with gradient differences in lingual positioning. Some articulatory correlates of perceived flapping correspond to predictions of a model of increased co-production of vowels and consonants yielding lenited stops heard as flaps, but others do not. Problems raised by these results for both traditional and prosodic by-product models are discussed.

1. Introduction

This paper reports an investigation into the nature of coronal stop flapping in American English. Sifting through the literature on this phenomenon reveals a large number of issues and a range of explanations which run between two types of models which will be described below.

1.1. *American English flapping as a phonological rule*

The flapping rule is often cited as an example of a phonological alternation for the purposes of teaching basic phonological analysis (e.g., Wolfram & Johnson, 1982, Chap. 7). This rule specifies that alveolar stops are realized as flaps. It is often of particular interest in that it is potentially a neutralizing rule; both voiced and voiceless stops appear on the surface as voiced flaps.

0095–4470/98/030283 + 28 $30.00/0

The change is conditioned in two ways by the prosodic environment in which the stops appear. First, in most formulations of the rule, the stops must occur intervocalically, or as some claim (Kahn, 1980; see also Kenyon, 1940 and earlier editions, and Haugen, 1938), the alveolar segments, /r/, /d/, and /n/, can occur between the previous vowel and the flapped stop, such as in the word, *parting*. Second, in most formulations of the rule, the stops must occur between a stressed and an unstressed vowel. However, as is also pointed out by Kahn (1980), the critical part of the environment is the following unstressed vowel. The stops will also undergo flapping following unstressed syllables, such as in the word *obesity*. This prosodic condition on the flapping rule can be used to create a diagnostic for distinctions between entirely unstressed (marked in Vanderslice & Ladefoged, 1972, as reduced syllables) and secondarily stressed syllables (Hayes, 1995). For example, the second syllable of the word *vita* can be considered entirely unstressed since the medial stop typically undergoes flapping. By contrast, the word *veto* often (though not always) is produced with secondary stress on the final syllable, as is evidenced by an aspirated medial stop. It should also be noted that vowel reduction toward schwa often, though not always, occurs in these unstressed syllables as well; the second vowel in *vita* is reduced, while the second vowel in *veto* is not regardless of whether the /t/ is flapped or not.

In addition, the structural description of the rule can also span a word boundary, such as in the rather militant sentences used by Monnot and Freeman (1972), *he bit him* and *he beat him*. In this case, the phonological reduction of the third person object pronoun to a clitic can provide the necessary unstressed initial vowel to trigger flapping of the coda /t/ in *beat* and *bit*. This occurrence of flapping across a word boundary underscores the fact that stress pattern here is to be seen not entirely as a lexical property, but as a property of portions of produced utterances. It should also be noted that some descriptions of the flapping rule note that flapping can also occur before vowel-initial words where the vowel is not reduced (Kahn, 1980; Steriade, 1996). Thus, even though the literature generally treats flapping as a single phenomenon, the conditions for its application differ somewhat, depending on whether the structural description spans a word boundary or not.

In the 1930s and 1940s, there was considerable discussion as to the nature of the flapped stops. The first analyses of the phenomenon (Bloomfield, 1933; Twaddell, 1935) treated it simply as a voicing neutralization whereby [t]'s become voiced, and hence become the same as [d]'s. Others later on realized that this categorical neutralization was missing important aspects of the phenomenon. Trager (1942) described the segment produced by the flappping of /t/ as "voiced *t*, in opposition to a *d*". Other authors of this time period noted that there is some change in the super-glottal articulation of the stop — a change that was sometimes identified as producing a tap or flap instead of a stop (e.g., Kenyon, 1940, para. 163; Bloch, 1941; and Oswald, 1943). The occlusive articulation of the stop is converted, perhaps through some general reduction process, to a ballistic movement, such as those in the production of sonorant flaps and taps in languages such as Spanish and Korean. (The distinction between flap and tap suggested by Ladefoged and Maddieson, 1996, involves the degree to which tongue tip motion is tangential to the point of contact. Since the lenited stops examined here do exhibit tangential forward motion, and the most common term is flap, I will use the term flap for the English lenited stops. It is clear, however, that much more work on tongue motion must be done before we can be certain that there is a useful distinction to be made between flaps and taps.) Explicitly investigating the identification of English flapped [t] with tap in Spanish, Monnot and Freeman (1972) found subtle articulatory differences between them, but also found perceptual cross-identification of the two. Based on this perceptual identification, traditional generative treatments of the phenomenon have modeled flapping as the result of a rule which changes the segmental features of the stop to more resemble those of a tap in other languages.

1.2. *Flapping as a prosodic by-product*

An aspect of the flapping rule which was noted early on by Haugen (1938) which complicates the description of the phenomenon is the variability in the output of the flapping rule. The flapped consonants seem to vary in whether they should be called alveolar stops or alveolar flaps or taps. Haugen suggested that the flapped [t]'s actually fell along a continuum from a "fully articulated *d*" to a lightly articulated tap. Zue and Laferriere (1979) also found that there is a considerable variation in the durational characteristics of flapped *t*'s. Some of this variability could be described as due to categorical differences in the prosodic structure of the utterances. In lexically internal cases such as *veto*, there are differences which could be described as categorical differences in the existence of secondary stress on the second vowel. In cases of application across a word boundary, there is a similar categorical (or quasi-categorical) variation in the reduced, cliticized form and full forms of function words which might account for some of the variability. Since the rule within lexical forms seems to apply exclusively to stops before unstressed vowels, it seem likely that the absence of phrasal stress will also encourage flapping across a word boundary. In addition, prosodic boundaries intervening the target stop and the triggering vowel might keep the rule from applying. This is the approach taken, for example, by Nespor and Vogel (1986), who specify that the flapping rule is restricted to application internal to a prosodic utterance. Beyond these categorical approaches, the possibility of flapping being understood as an entirely gradient phonetic process of lenition has recently been suggested by Umeda (1977) and investigated in more detail by Stone and Hamlet (1982), and Turk (1992).

Stone and Hamlet (1982) studied reiterant productions of various sentences with syllables replaced by the CV /da/. They approached the resulting data from three perspectives, two articulatory and one acoustic, and found that the alveolar occlusions ranged from partially devoiced occlusions of longer duration, to archetypical flaps with no clear plosion characteristics. From an articulatory perspective, the occlusions varied in the dynamics of jaw movement such that the more *d*-like occlusions were associated with higher jaw positions and more forceful jaw opening movements. Electro-palatography showed, in addition, that the tokens varied in a correlated fashion in the amount and location of linguo-palatal contact. The more *d*-like tokens, which had higher jaw positions, showed more lingual contact in a more anterior position. More flap-like tokens, which had smaller jaw movements and no clear acoustic plosion characteristics, had smaller areas of linguo-palatal contact in a more posterior location. This study, then, suggested that, since there were no apparent categorical differences between the conditions with flaps and those with stops, and there was not a strong modality in the articulatory posture exhibited, flapping is gradient.

Fujimura (1986) similarly suggested that the difference in jaw activity in flaps and stops supports a unified explanation both for the prosodic conditions on the flapping rule and for the flapping itself. Macchi (1985) showed that unstressed syllables typically show less contrast in jaw position between the margins and nucleus of the syllable than do stressed syllables. Occlusive consonants have lower jaw positions, while (especially open) vowels are articulated with higher jaw positions. Thus, both of the conditions on the flapping rule will have the same effect, that of lowering the jaw in flapped alveolar stops. First, in the onset of an entirely unstressed syllable, the jaw will be unusually low, as Macchi (1985) showed. Second, the intervocalic environment is one in which the consonant is in close proximity to two vowels which both will demand a relatively low position of the jaw. Thus, other things being equal, the intervocalic position will be the one in which the jaw will be the lowest for the medial consonant. Flapping, then, could be a by-product of consonant-vowel coarticulation and the encoding of prosodic organization in the jaw movement profile.

De Jong, Beckman and Edwards (1993) describe another articulatory model which also might explain flapping. In this model, a major characteristic of unstressed syllables is a greater amount of overlap between neighboring gestures. Assuming the flapped stops are part of an unstressed syllable, along with the following unstressed vowel, they should exhibit greater coarticulatory overlap (co-production) with the neighboring vowels. Since neighboring vowels demand a lower jaw and (usually) a lower and more retracted tongue body, this increased overlap would tend to lower the jaw and lower and retract the tongue body, which in turn would weaken and retract the coronal closure. Turk (1992) also notes that the durational structure of flapped tokens roughly parallels the durational structure of tokens with labial stops in the same prosodic environment, suggesting that the prosodic organization of these utterances in general shortens the duration of the stops.

If a model of general processes associated with the expression of some aspect of prosodic structure can explain the appearance of flapped stops, one must consider the possibility that the flapping rule may just be an epiphenomenon created through inappropriately categorical analysis by linguists. Most formulations of the flapping rule are based on the observation that underlying alveolar consonants show up auditorily as a categorically different entity, flaps. This categorical difference could arise in a non-linearity in the mapping of articulatory behavior onto acoustic output, and thence, auditory impression. The observation of categorical differences in the consonants, then, could merely be the result of transcribers categorizing the acoustic results of gradient changes in articulation in terms of categorical differences found in other languages. The flapping rule thus might not be indicative of some aspect of the segmental phonology of English, but rather might be an indirect indicator of some aspect of the expression of the prosodic phonology of American English.

1.3. *Examining different aspects of flapping*

In order to effectively test these models, one must have access to the articulation, acoustics, and perception of flaps. The present paper presents analyses of a corpus of X-ray microbeam data which contains possible targets for the flapping rule. These analyses are designed to examine relationships between articulation, acoustics, perception, and prosodic structure, and thereby to determine: (1) whether flapping is consistently perceived across listeners, (2) how the occurrence of flapping is associated with prosodic structure, and (3) how flapping perception is associated with measurable aspects of articulation and acoustics. In the strongest version of a categorical rule account, flapping should be consistently perceived, may be variably related to prosodic structure (though phonological treatments are unclear on this point), and should be reflected in clear modal differences in articulation and acoustics. A prosodic by-product model requires that flapping be strongly associated with particular prosodic structures. Flapping may or may not be consistently perceived, and if consistently perceived, may or may not be associated with modal differences in acoustics or articulation. However, the prosodic accounts sketched above demand that there be measurable differences in articulation consistent with general prosodic effects on articulations found in other studies, and that these differences be associated with the perception of flapping.

The particular targets examined in this corpus are instances in which the flapping rule spans a word boundary. This prosodic condition is examined here, since it seems that the flapping rule is more likely to be variable in application, hence affording an opportunity to examine the conditions under which the flapping rule applies while controlling lexical content.

As in Stone and Hamlet (1982), different views of the data will be investigated and compared. The present corpus differs from that analyzed by Stone and Hamlet (1982) in three important ways. The articulatory view of the data is tongue flesh point trajectories in the sagittal plane. This view will allow for investigation of lingual positioning, especially as this articulation is coupled with the position of the tongue body and the jaw. Second, the tokens analyzed here are taken from the articulation of real lexical items, rather than reiterant speech. Third, the corpus studied here contains both lexically voiced and voiceless tokens; thus, the present study will examine both the flapping of /d/ and /t/.

There are three parts to the current study. First, transcription analysis is used to examine the relationship between the voiced/voiceless status of the consonant, the phrasal stress pattern, and the likelihood of being perceived as a flap. This analysis is used first to examine the connection of application of a flapping rule and the prosodic structure of the utterances in the present corpus, and second to give an indication as to the perceptual quality of the tokens which would form the basis of a categorical analysis of the phenomenon. The second part determines the oral articulatory and acoustic correlates of the transcription of flapping, in order to ascertain the degree to which perceptions reflect measurable differences in the oral articulation and/or the acoustics of the stops. The third part examines the correlates in more detail to get an indication of the likelihood that either the jaw-lowering or the simple coproduction account will account for the articulatory differences between stops and flaps in the corpus.

2. Methods

2.1. *The corpus*

The present paper reports analyses of part of a corpus of articulatory and time-aligned acoustic data gathered with the Wisconsin X-ray microbeam facility (Nadler, Abbs, & Fujimura, 1987). Microbeam systems use a narrowly-focused beam of X-rays to track the location of radio-opaque pellets attached to articulators of interest during speech. The present corpus contains records of the motion of seven pellets. Three 2.5 mm pellets were attached using a dental adhesive to the superior surface of the tongue at distances of 10, 30, and 50 mm from the tongue apex, as measured with the tongue extended. Two 3 mm pellets indexed jaw movement, one placed on a mandible molar and the other placed at the base of the mandible incisors. Two pellets also indexed the movement of the upper and lower lip. Of particular interest for the present are the tongue pellets and the incisor pellet. Tongue pellet locations were sampled at a rate of 100 Hz, and the jaw pellet was sampled at 50 Hz. The trajectory of each of the pellets was smoothed and oriented with respect to an estimate of the occlusal plane which was made prior to data acquisition. Also included in the corpus was a time-aligned acoustic record sampled at 10 kHz. For more information concerning these procedures, see de Jong (1995).

There were three subjects, two male (MB and SD) and one female (TG). All three were undergraduates at the University of Wisconsin and were paid for their participation in the experiment. They were all speakers of a Northern dialect of American English. None of the speakers exhibited appreciable diphthong centralization before voiceless obstruents, typical of many Canadian and some Northern American dialects.

The corpus consisted the two words, *toad* and *tote* (an American English term for a type of handbag), spoken in random order in a frame sentence, *I said*. "*Put the* ____ *on the table*." The complete corpus also contained other words which are not discussed here. The prosodic structure of the utterance was varied by having the speakers imagine different discourse settings in which they were to respond to someone having misheard a portion of the sentence. Their utterances, then, were to correct for the misheard portion. Speakers interpreted the instructions by placing nuclear (sentence) accent on the misheard item. Misheard items (and thus, nuclear accented items) were either *put*, *on*, or the target words. The corpus was submitted to an intonational analysis using the system of intonational transcription presented in Pierrehumbert (1980), and further developed in Pierrehumbert and Beckman (1988); further details of this analysis are presented in de Jong (1995). In addition to consistently placing nuclear accent on the appropriate item, speakers also consistently placed a pre-nuclear accent on the target word when it preceded the nuclear accented item. Thus, there were three prosodic environments for the target words: postnuclear, and hence unaccented; prenuclear and accented; and nuclear accented with a following unaccented *on*. The two male speakers had four repetitions of the two words in the three prosodic conditions for a total of 24 tokens each. The female speaker had six repetitions, for a total of 36 tokens. Due to various mechanical and software problems in porting and analyzing the data, two of SD's tokens were rendered unanalyzable.

2.2. *View* 1: *Transcription of the database*

The acoustic record of each token was extracted from the utterance using a digital wave-form editor. The initial splice was taken in the period of silence just before the onset of the initial [t] in the target word, and the final splice was taken toward the end of the [n] in the following *on* at a zero crossing. The spliced out tokens were then stored on disk in a fashion in which they could be played out with a simple software routine.

Four transcribers were asked to play the tokens from the disk and give relatively narrow IPA transcriptions of each token. They were encouraged to play the tokens as many times as necessary to be confident in their transcriptions. The first transcriber was the author, and the remaining three transcribers were advanced graduate students in the Phonetics Lab at UCLA. Each of them had a fair amount of fieldwork experience, and thus each was capable in the use of the IPA transcription system. The three transcribers were not made aware of the purposes of the study prior to transcription.

2.3. *View* 2: *Acoustic analyses*

The acoustic records were analyzed for differences between flapped and non-flapped stops which have been found in previous studies. Analyses were performed using waveforms and broad-band (300 Hz) spectrographic displays generated by the Kay CSL software on a PC. Specific variables used here were: (1) occlusion duration, measured as the duration of time between the damping of high-frequency energy indicative of coronal closure and the onset of high-frequency energy indicating release of the consonant; (2) voice onset time, measured as the latency of voicing from time of release; (3) voice offset time, measured as the latency of non-voicing from time of closure; (4) % voicing, measured as the proportion of time during the closure during which voicing was apparent; (5) V1-duration, measured as the latency of closure onset from onset of voicing for the previous vowel; (6) V2-duration, measured as the latency of nasal closure in *on* from the onset of modal voicing following the release of the target consonant; and (7) V/C ratio, V1-duration divided by the occlusion duration. In cases in which voicing proceeded throughout the closure, voice onset times were set to 0. However, it was often the case that with the presence of a clear burst, modal voicing was disturbed during the burst; thus, cases of voiced stops with strong bursts had short, positive voice onset times. In addition, as an index of the potential effect of differences in the point of articulation for flaps and stops, F2 values around the closure were obtained using a 12th order LPC analysis with a 20 ms window. Three additional measures were derived from these F2 traces; (8) F2 just before closure; (9) F2 in-transition, measured as the difference between the F2 in the temporal center of the previous vowel and that just before closure; and (10) F2 out-transition, measured as the difference between the F2 in the first voiced portion after the closure and the F2 in the temporal mid-point in the following vowel.

One expects flaps to: (1) have shorter occlusion durations, (2) have shorter voice onset times, and (3) have greater % voicing. With regard to the preceding vowel, one would expect a larger V/C ratio. With regard to F2 measures, one would expect higher F2 loci with more posterior closure locations, yielding more positive in-coming and more negative out-going transitions.

2.4. *View* 3: *Articulatory analyses*

Flesh point trajectories were analyzed for posture of tongue at the time of closure as well as the maximum raising position for the jaw. In order to align the various trajectories, the time of acoustic release of the alveolar consonant was taken as an alignment point. Pellet positions at that time, as well as at 20 ms increments before and after the alignment point, were extracted. In addition, jaw position measurements were taken in a dimension roughly corresponding to the major dimension of jaw movement. The amount of rotation of the jaw motion was determined by taking a random sampling of raised and lowered jaw positions and calculating the average angle from perpendicular to the occlusal plane; the average angle for TG was 3 degrees, and for SD was 13 degrees in an anterior direction. MB's jaw motion was essentially vertical with respect to the occlusal plane. For further discussion of these measurement procedures, see de Jong (1995).

From these raw measurements, several variables were extracted which are likely to index oral differences between flapped and non-flapped tokens, given the gradient models being examined here. The expectation from the gradient models mentioned above is that the flaps will: (1) have lower jaw positions at time of closure, here measured as the maximum position of the incisor pellet (lower IY); (2) have tongue body positions more antagonistic to the making of alveolar closures, here measured by the position of the dorsal pellet during closure (lower DX and DY); and (3) have more retracted tongue tip positions (as found by Stone & Hamlet, 1982), here measured as the raising and retraction of the tongue blade pellet (lower BX and greater BY). (Raising would be due to the fact that the hard palate slants upward behind the alveolar ridge.) In addition, one might expect two more results: (4) Since that the tongue surface is not a solid body, but can undergo a considerable amount of stretching and shrinking (there is nearly a 2-to-1 difference in the distance between the pellets during placement with the tongue extended and during speech with the tongue internal to the oral cavity), one might expect to find a greater amount of stretching of the tongue surface due to the antagonistic positioning of the tongue body for the neighboring vowels. This stretching is measured as the Euclidean distance between the three different tongue pellets (dorsal, mid, and blade) at time of closure (greater B-to-M, greater M-to-D). And (5), given the traditional description of flaps as ballistic movements of the tongue tip, one might also expect a greater amount of movement over the 40 ms spanning the time of closure, here measured as the sum of the distances traveled between two successive 20 ms intervals preceding and following the acoustic release of the consonant (greater delta-B).

In addition, for an analysis which examined the effect of the consonant on neighboring vowels, tongue pellet positions during the previous and following vowel were also examined. Time points of positioning were selected based on criteria reported in de Jong (1995). Positions in the previous vowel were chosen according to the position which most reflected the effects of a velar constriction gesture, usually the most retracted location for the dorsal pellet. Some tokens, however, did not exhibit any clear effect of a dorsovelar off-glide for the /o/; in these tokens, the position taken was the point of lowering and retraction in a dimension 45 degrees anterior of vertical. This criterion was chosen, since it selected a position which most closely corresponded to a speed minimum for the vowel. For the following vowel, the lowest point for the tongue-midsection pellet gave a consistent indicator of the opening for the following vowel.

3. Results

3.1. *Transcription*

Symbols the listeners used to transcribe the target alveolar consonants generally conformed to what one would expect of this corpus given the flapping rule. The target consonant was usually transcribed either as [t], as [d], or as [ɾ] by each transcriber, with a few exceptions. Transcriber 3 transcribed three tokens as having [n] in the coda of the target word, and transcriber 1 transcribed two tokens with [ð].

All four transcribers agreed in the transcription of the coronal stop in 65% of the cases. Another 20% of the cases had 3 of 4 transcribers in agreement, leaving 15% of the cases with no majority transcription. The disagreements were generally distributed across the transcribers, as is evident in comparing transcribers in a pair-wise fashion. The pairwise agreement between transcribers ranged from 57% (transcriber 1 vs. 3) to 69% (transcriber 2 vs. 4). For a single measure of inter-transcriber consistency, judgements of each token by each transcriber were compared with judgements of that same token by each other transcriber in a pair-wise fashion. There was 72% pair-wise

TABLE I. Confusion matrix for transcription of each token by pairs of transcribers

	[ɾ]	[d]	[t]	other
[ɾ]	168	62	3	16
[d]		73	23	3
[t]			143	0
other				1

agreement across the transcribers. If we consider just whether transcribers agreed on whether tokens should be considered flaps, the agreement rate rises to 81%. Finally, if we include the [ð] judgements as judgements that some sort of lenition has occurred, the agreement rate as to whether lenition has occurred rises to 85%. Thus, there is a fairly strong agreement among the transcribers concerning the occurrence of flapping, suggesting that the flap category is stable across listeners.

Though there is fairly strong evidence here that flapping is categorically present and consistently perceived across listeners, there is also some evidence in the transcriber disagreements that the present tokens represent a continuum. Table I gives a confusion matrix for the three categories of response collapsed across transcribers. As can be seen, a majority of the intertranscriber disagreements involved flaps and [d], not flaps and [t]. In general, there was more agreement on [t]; however, when there was a disagreement between transcribers, it was almost always between [t] and [d], not between [t] and a flap. This pattern suggests that the flap and the [t] reside some perceptual distance from one another, and that the [d] resides between the two.

To discuss the data more efficiently, each token will be categorized by the transcriptions in the following way. Tokens for which there was 100% agreement will be grouped into bins 1 (flaps), 3 ([d]'s), and 5 ([t]'s). Groups for which there was some disagreement will be grouped into bins 2 (flap or [d]), and 4 ([d] or [t]). The other three tokens (transcribed as [t] or flap) are included in a separate bin. Transcriptions with [n] and [ð] will be counted as flaps, which turn out to place each of these tokens in bin 2 (flap or [d]). The bin numbers 1–5 correspond to their location within an expected continuum of lenition.

Next, consider the relationship between perceived flapping and various categorical conditions on the tokens. Assuming a connection between degree of stress on a following vowel and the occurrence of flapping, one would expect the flapping rule to apply equally to /t/ and /d/ when the vowel in the following *on* is unstressed. In the current corpus, this occurred when nuclear accent was placed on either *put* or the target word. In these cases, *on* is unaccented and exhibits a strongly reduced vowel. (Evidence for this reduction comes from a comparison of tongue movement and position differences between accented and unaccented *on* in the current corpus and differences found in another corpus consisting of accented and unaccented content and nonsense words. Analyses, presented in de Jong, 1991, Chap. 3, show the differences for *on* are far greater and are consistent with a strong reduction toward schwa in the unstressed renditions of *on*.) When, however, nuclear accent is placed on *on*, one would expect the stop to be realized as either a [t] or a [d].

Table II summarizes the distribution of bin membership across accent, speaker, and lexical voicing. Expected cells in Table II to be populated with a flapping rule are boxed.

TABLE II. Numbers of tokens for each speaker and accent category in each transcription bin. Boxes indicate expected results given a standard flapping rule

	Postnuclear			Nuclear accented			Prenuclear accented		
	TG	MB	SD	TG	MB	SD	TG	MB	SD
TOAD									
1 (flap)	6		4	2		1	1		
2		3		4	2	3	1		1
3 ([d])					2		3	2	2
4							1	2	
5 ([t])									
other		1							
TOTE									
1 (flap)	3	2	3	1			1		
2	1	2				1			
3 ([d])									
4						1		3	
5 ([t])	2		1	5	4	1	5	1	4
other									
	EXPECTED FLAPS						EXPECTED STOPS		

Consider first the effect of accent on renditions of the word, *toad* (top half of Table II). While tokens in the unaccented (postnuclear) and prenuclear conditions behave more or less as expected, nuclear accented tokens did not. Postnuclear tokens by all of the speakers were always perceived as flapped by some of the transcribers. Speaker MB's tokens were sometimes perceived as having [d]. With the exception of one token by TG, the prenuclear tokens, as expected, were usually transcribed as [d]'s. MB's tokens tended to be perceived as a [t]. The nuclear tokens unexpectedly tended to be perceived as unflapped; this is particularly true of speaker MB.

The tokens of the word *tote* were more complicated, though similar (bottom of Table II). As expected, most prenuclear tokens were perceived as unflapped (with one clear exception by speaker TG), and most postnuclear tokens were perceived as flapped (with the same complication that MB's tokens tended to be heard as [d]). This flapping, however, was not as consistent, there being three unaccented tokens by two speakers that were transcribed as [t]. Nuclear tokens present a problem for the flapping rule even more clearly than did the tokens of *toad*. With the exception of one token by TG and one by SD, the nuclear tokens were clearly perceived as unflapped.

The flapping rule accurately describes about half of the data found here (39 of 82 tokens). If we liberally allow for transcribers to be off by one bin (i.e., insist only that one of the transcribers agrees with the rule), the proportion grows to 63 of 82. Thus, the transcriptions show two things. First, flapping across a word boundary does seem to be a perceptually categorical phenomenon, as evidenced by fairly high intertranscriber agreement. Second, the high number of tokens which do not conform to the rule show that the flapping rule is optional across a word boundary. However, this optionality is partly conditioned by prosodic location. In post-nuclear unaccented conditions, speakers usually flap, though not always. With nuclear accent on the target word, speakers usually do not flap.

3.2. *Acoustic and articulatory correlates of transcription categories*

In order to examine the relationship between transcriptions and acoustic and articulatory measurements, Table III shows mean values for each of the articulatory and acoustic measures for tokens which fell into each transcription bin. In order to help direct attention to aspects of the means which indicate differences between the transcription categories, the data were submitted to a series of two-way ANOVA's with subject and bin as independent variables and each measurement as a dependent variable. Because transitional bins (bins 2 and 4) do not represent categories but poor tokens of one of the other categories, only tokens for which there was transcriptional agreement (bins 1, 3, and 5) were included in the analysis. Analyses which produced F-values for the bin factor with p's < 0.0009 are in bold face in the table: (The p-value of 0.0009 was determined as the value which would give a family-wise p-value over all 54 analyses—18 variables by 3 word-conditions—of approximately 0.05. See Kirk, 1982, for a discussion of the family-wise P-value correction used here.) All analyses which produced F-values with p's < 0.0029 are boxed. (The p-value of 0.0029 is a value which would give a family-wise p-value over the 18 analyses for each word-condition of approximately 0.05.) Subject factors proved significant (at the 0.0009 level) for V1-duration for all the data, F2 and F2 transitions into the C for each word and for the data combined, as well as nearly all of the articulatory measures. None of the interactions were significant at the 0.0029 level in any of the analyses.

Acoustic correlates of the transcription categories (flap, [d], or [t]) were occlusion duration, voice onset time, proportion of voicing in closure (and the related voice offset time measure), the duration of the previous vowel, and F2 transition into the closure. Most of these differences are in the expected directions. Tukey post-hoc t-tests indicate that flaps have shorter occlusion durations, shorter voice onset times (indicating the general lack of a release burst), and a greater proportion of voicing in closure. However, examining vowel durations and F2-transition differences indicates different patterns. Here, flaps group with [d] in having longer vowel durations than [t], and flaps group with [t] in having smaller transitions than [d]. Among articulatory measures, the horizontal location of the tongue body, the vertical location of the blade, and amount of motion of the tongue-blade differed significantly between categories. However, among these, post-hoc tests indicate that only the tongue body effect is in the expected direction. The tongue body is more retracted during flaps than stops. The two tongue-blade measures grouped flap with [d] in having *lower* tongue blade positions and greater movement than [t]. Overall, differences between transcription categories are more robust for acoustic measures than for articulatory measures, and for phonemic /t/ than for phonemic /d/.

Interpreting these results with respect to flapping, however, is difficult because the ANOVA does not capture the relationship between the different transcription categories, nor does it address what is occurring in tokens for which subjects disagreed in transcription. To get a statistical measure which is more directed toward correlates of the

TABLE III. Average measurements for transcription categories

Speaker	TOAD Bin 1: Flap	TOAD Bin 2: Flap/Stop	TOAD Bin 3: [d]	TOAD Bin: 4: [d]/[t]	TOAD Bin 5: [t]	TOAD Average	TOTE Bin 1: Flap	TOTE Bin 2: Flap/Stop	TOTE Bin 3: [d]	TOTE Bin: 4: [d]/[t]	TOTE Bin 5: [t]	TOTE Average	ALL Bin 1: Flap	ALL Bin 2: Flap/Stop	ALL Bin 3: [d]	ALL Bin: 4: [d]/[t]	ALL Bin 5: [t]	ALL Average
Acoustic measures																		
TG Occlusion duration (sec)	0.019	0.300	0.039	0.018	**	0.026	0.024	0.031	**	**	0.065	0.050	**0.021**	**0.030**	**0.039**	**0.018**	**0.065**	0.038
Voice Onset Time (sec)	**0.003**	**0.007**	**0.049**	**0.067**	**	0.015	**0.009**	**0.015**	**	**	**0.057**	0.039	**0.005**	**0.010**	**0.049**	**0.067**	**0.057**	0.027
Voice Offset Time (sec)	0.019	0.025	0.025	0.018	**	0.022	**0.019**	**0.015**	**	**	**0.011**	0.013	**0.019**	**0.021**	**0.025**	**0.018**	**0.011**	0.018
% of Closure with Voicing	0.984	0.810	0.700	1.000	**	0.880	**0.830**	**0.568**	**	**	**0.209**	0.407	**0.933**	**0.729**	**0.700**	**1.000**	**0.209**	0.643
V1 duration (sec)	0.055	0.049	0.064	0.060	**	0.055	0.034	0.051	**	**	0.035	0.038	0.047	0.049	0.064	0.060	0.035	0.046
V2 duration (sec)	0.107	0.109	0.117	0.125	**	0.110	0.105	0.085	**	**	0.091	0.093	0.106	0.101	0.117	0.125	0.091	0.102
V1/Closure ratio	3.581	1.677	1.794	3.352	**	2.580	1.534	2.375	**	**	0.712	1.199	2.836	1.910	1.794	3.352	0.712	1.889
F2 at Occlusion (Hz)	1737	1706	1821	1799	**	1744	1700	1728	**	**	1559	1619	1725	1713	1821	1799	1559	1681
F2 transition into C (Hz)	404	363	545	528	**	421	310	396	**	**	255	291	**373**	**374**	**545**	**528**	**255**	356
F2 transition out of C (Hz)	− 78	− 36	− 99	− 16	**	− 64	0	− 33	**	**	− 43	− 32	− 52	− 35	− 99	− 16	− 43	− 48
MB Occlusion duration (sec)	**	0.036	0.034	0.029	**	0.034	0.033	0.036	**	0.033	0.040	0.036	**0.033**	**0.036**	**0.034**	**0.031**	**0.040**	0.035
Voice Onset Time (sec)	**	**0.055**	**0.135**	**0.077**	**	0.085	**0.000**	**0.000**	**	**0.028**	**0.116**	0.055	**0.000**	**0.041**	**0.135**	**0.048**	**0.116**	0.070
Voice Offset Time (sec)	**	0.028	0.020	0.021	**	0.024	**0.033**	**0.036**	**	**0.014**	**0.009**	0.018	**0.033**	**0.030**	**0.020**	**0.017**	**0.009**	0.021
% of Closure with Voicing	**	0.723	0.614	0.719	**	0.686	**0.999**	**0.991**	**	**0.486**	**0.245**	0.555	**0.999**	**0.790**	**0.614**	**0.579**	**0.245**	0.621
V1 duration (sec)	**	0.045	0.047	0.045	**	0.046	0.033	0.036	**	0.037	0.038	0.037	0.033	0.043	0.047	0.040	0.038	0.041
V2 duration (sec)	**	0.100	0.086	0.126	**	0.100	0.080	0.062	**	0.119	0.131	0.108	0.080	0.091	0.086	0.122	0.131	0.104
V1/Closure ratio	**	1.273	1.426	1.585	**	1.376	1.050	1.010	**	1.293	1.027	1.094	1.050	1.207	1.426	1.410	1.027	1.235
F2 at Occlusion (Hz)	**	1318	1379	1395	**	1351	1222	1255	**	1282	1261	1259	1222	1302	1379	1327	1261	1305
F2 transition into C (Hz)	**	207	351	248	**	262	91	141	**	204	129	143	**91**	**190**	**351**	**221**	**129**	202
F2 transition out of C (Hz)	**	−102	−17	−99	**	−73	− 74	− 91	**	−176	− 43	− 90	− 74	− 99	− 17	− 145	− 43	− 81
SD Occlusion duration (sec)	0.014	0.220	0.029	**	**	0.020	0.020	0.027	**	0.061	0.043	0.037	**0.016**	**0.023**	**0.029**	**0.061**	**0.043**	0.028
Voice Onset Time (sec)	**0.000**	**0.008**	**0.073**	**	**	0.016	**0.005**	**0.005**	**	**0.040**	**0.054**	0.035	**0.002**	**0.007**	**0.073**	**0.040**	**0.054**	0.025
Voice Offset Time (sec)	0.014	0.022	0.028	**	**	0.019	**0.020**	**0.027**	**	**0.013**	**0.014**	0.017	**0.016**	**0.023**	**0.028**	**0.013**	**0.014**	0.018
% of Closure with Voicing	1.000	0.999	0.967	**	**	0.994	**0.994**	**0.989**	**	**0.209**	**0.347**	0.569	**0.998**	**0.997**	**0.967**	**0.209**	**0.347**	0.781
V1 duration (sec)	0.042	0.046	0.045	**	**	0.044	0.021	0.040	**	0.009	0.031	0.027	0.034	0.045	0.045	0.009	0.031	0.036
V2 duration (sec)	0.088	0.111	0.128	**	**	0.104	0.074	0.106	**	0.094	1.101	0.093	0.083	0.110	0.128	0.094	0.101	0.099
V1/Closure ratio	3.476	2.153	1.524	**	**	2.640	1.133	1.467	**	0.149	0.776	0.879	2.597	2.016	1.524	0.149	0.776	1.760
F2 at Occlusion (Hz)	1340	1366	1362	**	**	1354	1276	1354	**	1287	1296	1295	1316	1364	1362	1287	1296	1324
F2 transition into C (Hz)	159	219	182	**	**	185	110	232	**	82	116	122	**140**	**222**	**182**	**82**	**116**	153
F2 transition out of C (Hz)	−116	−124	−199	**	**	−134	− 50	−116	**	−82	− 165	−122	− 91	− 123	− 199	− 82	− 165	− 128

TABLE III. Average measurements for transcription categories (cont'd)

Speaker	TOAD Bin 1: Flap	TOAD Bin 2: Flap/Stop	TOAD Bin 3: [d]	TOAD Bin: 4: [d]/[t]	TOAD Bin 5: [t]	TOAD Average	TOTE Bin 1: Flap	TOTE Bin 2: Flap/Stop	TOTE Bin 3: [d]	TOTE Bin: 4: [d]/[t]	TOTE Bin 5: [t]	TOTE Average	ALL Bin 1: Flap	ALL Bin 2: Flap/Stop	ALL Bin 3: [d]	ALL Bin: 4: [d]/[t]	ALL Bin 5: [t]	ALL Average
Articulatory measures																		
TG IY (mm)	34.6	34.7	34.3	34.7	**	34.6	34.7	34.6	**	**	35.0	34.9	34.6	34.7	34.3	34.7	35.0	34.8
DX (mm)	**9.4**	**9.8**	**10.4**	**9.9**	**	9.7	9.5	9.6	**	**	10.2	9.9	**9.5**	**9.7**	**10.4**	**9.9**	**10.2**	9.8
DY (mm)	42.1	42.7	41.9	41.7	**	42.2	42.4	42.4	**	**	43	42.7	42.2	42.6	41.9	41.7	43	42.5
BX (mm)	27	27	27.6	27.2	**	27.1	27	27.2	**	**	27.8	27.5	27	27.1	27.6	27.2	27.8	27.3
BY (mm)	41.8	41.8	41.9	42.1	**	41.9	41.8	41.7	**	**	43	42.5	**41.8**	**41.8**	**41.9**	**42.1**	**43**	42.2
Distance from M to B (mm)	9.6	9.6	9.3	9.5	**	9.5	9.5	9.6	**	**	9.7	9.6	9.6	9.6	9.3	9.5	9.7	9.6
Distance from M to D (mm)	8.1	8.6	7.9	7.9	**	8.2	8	8.1	**	**	8.2	8.1	8.1	8.4	7.9	7.9	8.2	8.2
Delta-B (mm)	3.3	2	4	3.5	**	3	2.9	2.1	**	**	2.5	2.5	**3.2**	**2**	**4**	**3.5**	**2.5**	2.8
MB IY (mm)	**	33.0	33.9	32.7	**	33.2	33.0	33.1	**	33.1	34.0	33.5	33.0	33.0	33.9	32.9	34.0	33.4
DX (mm)	**	**11.1**	**11**	**11.7**	**	11.2	10.6	11.1	**	11.4	11.3	11.2	**10.6**	**11.1**	**11**	**11.5**	**11.3**	11.2
DY (mm)	**	40.3	42.3	40.4	**	41	40.1	40.1	**	40.3	41.8	40.9	40.1	40.3	42.3	40.3	41.8	40.9
BX (mm)	**	24.9	25.1	25.7	**	25.1	24.6	25	**	25.4	25.6	25.3	24.6	24.9	25.1	25.5	25.6	25.2
BY (mm)	**	40.6	41.7	40.4	**	40.9	40.4	40.4	**	40.4	41.2	40.7	**40.4**	**40.6**	**41.7**	**40.4**	**41.2**	40.8
Distance from M to B (mm)	**	6.2	6.3	6.4	**	6.2	6.4	6.2	**	6.2	6.3	6.3	6.4	6.2	6.3	6.3	6.3	6.3
Distance from M to D (mm)	**	7.6	7.9	7.7	**	7.7	7.6	7.7	**	7.8	8.1	7.8	7.6	7.6	7.9	7.7	8.1	7.8
Delta-B (mm)	**	4.7	4.6	6.4	**	5	4.5	5.9	**	4.7	4.1	4.6	**4.5**	**5**	**4.6**	**5.4**	**4.1**	4.8
SD IY (mm)	31.6	31.9	31.6	**	**	31.7	31.9	32.0	**	32.8	31.7	31.9	31.7	31.9	31.6	32.8	31.7	31.8
DX (mm)	**9.8**	**11.2**	**10.6**	**	**	10.5	10.1	10.5	**	12	10.9	10.7	**9.9**	**11**	**10.6**	**12**	**10.9**	10.6
DY (mm)	38.6	39	39.5	**	**	38.9	38.6	39.7	**	41.2	39.5	39.4	38.6	39.1	39.5	41.2	39.5	39.2
BX (mm)	24.5	25.6	25.7	**	**	25.1	24.8	25.2	**	26.9	25.6	25.5	24.6	25.5	25.7	26.9	25.6	25.3
BY (mm)	39.9	39.9	40.2	**	**	39.9	39.8	40.8	**	41.5	40.8	40.6	**39.9**	**40.1**	**40.2**	**41.5**	**40.8**	40.3
Distance from M to B (mm)	7.6	7.5	7.8	**	**	7.6	7.7	7.3	**	7.1	7.6	7.5	7.6	7.5	7.8	7.1	7.6	7.6
Distance from M to D (mm)	7.1	7.1	7.3	**	**	7.1	7.2	7.4	**	7.8	7.2	7.3	7.1	7.1	7.3	7.8	7.2	7.2
Delta-B (mm)	7.4	5.9	6.4	**	**	6.7	6.8	7	**	4.5	3.9	5	**7.1**	**6.1**	**6.4**	**4.5**	**3.9**	5.8

perception of flapping in particular, measures were correlated with the number of times a token was transcribed as a flap. To factor out differences between the speakers, the number of flap transcriptions were regressed against effects coding variables indicating speaker first, and then the residual was regressed against the measurements. Table IV summarizes these analyses, indicating the partial correlation (*r*) values, together with the *t*-statistics and nominal significance levels for the associated regression coefficients. As with Table III, bold-face indicates effects significant at the 0.0009 level, while boxing indicates effects significant at the 0.0029 level.

The regression analysis indicates similar correlates to those indicated by the ANOVAs. Acoustically, flapped /d/'s correlate with shorter voice onset times and lower F2 loci. Flapped /t/'s correlate with shorter voice onset times, shorter closures, and the presence of voicing in the closure. A relatively weak effect of occlusion duration becomes apparent in a correlation of flapping with a higher V/C ratio when the data is pooled. Articulatory correlates are clear in the regression analyses, as well. Flapped /t/'s correlate with lower and more retracted blade pellet positions and increased movement of the blade pellet. None of the articulatory variables were significant predictors in the /d/-flapping analysis. Pooling the data reveals additional effects on dorsal x- and y-position and on blade x-position which appear as weak effects in both /t/ and /d/ analyses.

These results indicate that most of the types of correlates examined are related to the flapping phenomenon to some extent. Broad types examined here are aerodynamically influenced effects such as voicing and the presence of a burst, durations, formant loci, tongue body position, tongue blade position, and amount of movement of the tongue blade. However, different correlates occur, depending on whether the sound is lexical /t/ or /d/. For lexical /d/, blade x-position shows a trend toward retraction. The strongest correlate, however, is voice onset time. For lexical /t/, by contrast, there is a more robust difference in blade positioning and dynamics. The strongest correlate, like /d/, has to do with aerodynamically determined acoustic measures, in particular the presence of voicing in the closure and voice onset time. The acoustic differences are not surprising; the flapping rule indicates a change in voicing status for /t/, but not for /d/. The articulatory differences show that, in terms of tongue tip activity, [d]'s are very similar to flaps, while [t]'s differ from flaps in several respects.

Another result which appears in these analyses is that flap transcriptions correlate the strongest with aerodynamically related acoustic measurements. An obvious question which this raises is whether these acoustic factors are sufficient to explain the perception of flapping. If this were true, articulatory factors and F2 factors might still correlate with flap transcription because they happen to correlate with these stronger acoustic correlates. To examine this possibility in more detail, a stepwise regression analysis was performed to determine whether articulatory and formant factors had a demonstrably independent effect on transcriptions. To do this, flap transcriptions were regressed against the subject variables and voice onset time, voice offset time, and %voicing. The residual was then regressed against each of the variables boxed in Table III or Table IV. Table V shows partial *r*'s yielded by this analysis. Significance of the effects was determined using critical *r*-squared values for an alpha-level of 0.05 given in Wilkinson (1979). This analysis shows that one articulatory variable, DX, can be demonstrated to have an effect on the perception of flapping independent of aerodynamically influenced acoustic measures.

The overall conclusion suggested by this analysis, then, is that acoustic indices based on aerodynamically-related effects such as voicing and the presence of a burst are the

TABLE IV. Summary of regression of flap transcriptions against measurements

	TOAD			TOTE			ALL		
Acoustic variables	partial-*r*	*t* (40)	*p*	partial-*r*	*t* (40)	*p*	partial-*r*	*t* (40)	*p*
Occlusion duration (sec)	– 0.464	– 2.960	0.005	– 0.537	– 3.782	0.001	**– 0.522**	**– 0.578**	**< 0.0001**
Voice Onset Time (sec)	**– 0.572**	**– 4.196**	**< 0.0001**	**– 0.603**	**– 4.548**	**< 0.0001**	**– 0.601**	**– 6.614**	**< 0.0001**
Voice Offset Time (sec)	– 0.198	– 1.114	0.273	**0.555**	**4.017**	**0.0003**	0.220	2.156	0.034
% of Closure with Voicing	0.177	1.040	0.305	**0.786**	**7.814**	**< 0.0001**	**0.600**	**6.796**	**< 0.0001**
V1 duration (sec)	– 0.294	– 1.802	0.080	– 0.069	– 0.414	0.682	0.043	0.456	0.650
V2 duration (sec)	– 0.262	– 1.519	0.137	– 0.129	– 5.170	0.308	– 0.135	– 0.733	0.466
V1/Closure (Hz)	0.360	2.286	0.028	0.374	2.421	0.021	**0.436**	**4.232**	**< 0.0001**
F2 at Occlusion (Hz)	– 0.431	– 2.695	0.011	0.214	1.456	0.154	0.096	1.234	0.220
F2 transition into C (Hz)	– 0.502	– 6.646	0.002	0.053	0.295	0.770	– 0.063	– 0.317	0.752
F2 transition out of C (Hz)	0.085	0.362	0.719	0.274	1.626	0.113	0.113	0.891	0.375
Articulatory variables									
IY (mm)	– 0.120	– 0.769	0.447	– 0.199	– 1.243	0.222	– 0.198	– 1.945	0.059
DX (mm)	– 0.414	– 2.567	0.014	– 0.439	– 2.938	0.006	**– 0.432**	**– 3.977**	**0.0002**
DY (mm)	– 0.269	– 1.567	0.157	– 0.337	– 2.263	0.030	– 0.349	– 3.306	0.001
BX (mm)	– 0.454	– 2.850	0.007	– 0.420	– 2.773	0.009	**– 0.464**	**– 4.427**	**< 0.0001**
BY (mm)	– 0.214	– 1.207	0.235	**– 0.584**	**– 4.459**	**< 0.0001**	**– 0.487**	**– 4.967**	**< 0.0001**
Distance from M to B (mm)	0.177	0.825	0.415	– 0.049	– 0.390	0.699	0.008	0.021	0.983
Distance from M to D (mm)	– 0.126	– 0.055	0.987	– 0.255	– 1.469	0.150	– 0.244	– 0.685	0.496
Delta-B (mm)	0.061	0.382	0.705	**0.490**	**3.428**	**0.002**	0.296	2.771	0.007

TABLE V. Partial regression coefficients with aerodynamic variables factored out

Variable	TOAD	TOTE	ALL
DX	**− 0.440**		**− 0.392**
BX	− 0.372		
BY		− 0.216	0.017
Delta-B		0.144	− 0·001
Occlusion duration	− 0.376		− 0.282
V1 duration			− 0.171
F2-in-transition			− 0.210

clearest indicators of flapping. This result shows that, if a gradient change in articulatory organization (e.g., jaw positioning) is producing the perceived shift to the flap cateogory, it is due in large part to its combined effect with glottal activity in producing these acoustic attributes. However, these results do not indicate that the oral activity differences are not acoustically available, or that they have no effect on the perception of flapping. The independent contribution of the DX factor suggests that the perception of flapping is also tied to the oral articulation of the consonant.

3.3. *Articulatory correlates of flapping and prosodic models*

In this section, we turn to a more careful examination of particular measurements which have been shown to be correlates of flap transcriptions, as well as ones which gradient accounts suggest should be correlates. We begin by examining the two clearest correlates from the previous section, proportion of voicing in the closure and voice onset time. Fig. 1 plots the measures against one another.

This figure shows a pretty clear separation of tokens perceived as flaps (hollow symbols) from others for two of the three speakers. This is the type of distribution one might expect with an optionally applying categorical rule. TG's and SD's flapped tokens (upper and lower panels) all show very short voice onset times, indicative of absence of a strong burst. Tokens which appear to the upper right in each panel are fully voiced tokens which exhibited a strong burst which disrupted voicing at the release. Interestingly, several of TG's tokens, including one which was consistently transcribed as a flap, had very little voicing during closure. MB has less clear separation. To the extent that a generalization can be made about his tokens transcribed as flaps, it is when either there is no clear burst, or that there is voicing throughout closure.

Pursuing oral articulatory differences which might accompany these differences in voicing and strength of release, we turn to articulatory correlates of flapping. Fig. 2 plots average tongue pellet and jaw pellet positions for tokens for which transcriptions were consistent (in bins 1, 3, and 5). The tokens in 2 and 4 have been omitted for graphic clarity. First, flap jaw positions are lower only for MB, but not for TG or SD. This might be related to the results of de Jong (1995), which shows MB to be a more consistent "jaw user" in the production of coronal constrictions. All of the subjects show a generally lower tongue position in flaps, though TG shows a similar vertical positioning for flaps and [d]'s. Consistent across all of the subjects is a shifting of the overall tongue position to a more posterior position in flaps. The general shape of the (mid-sagittal) tongue

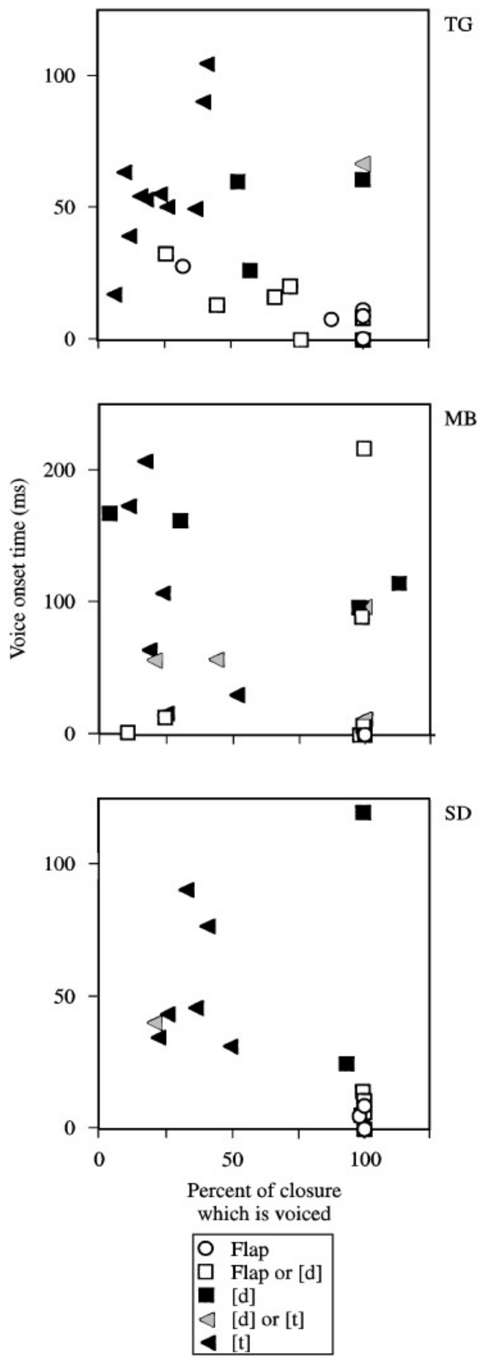

Figure 1. Voice onset time plotted against the percentage of closure exhibiting voicing. Tokens are coded by transcription category.

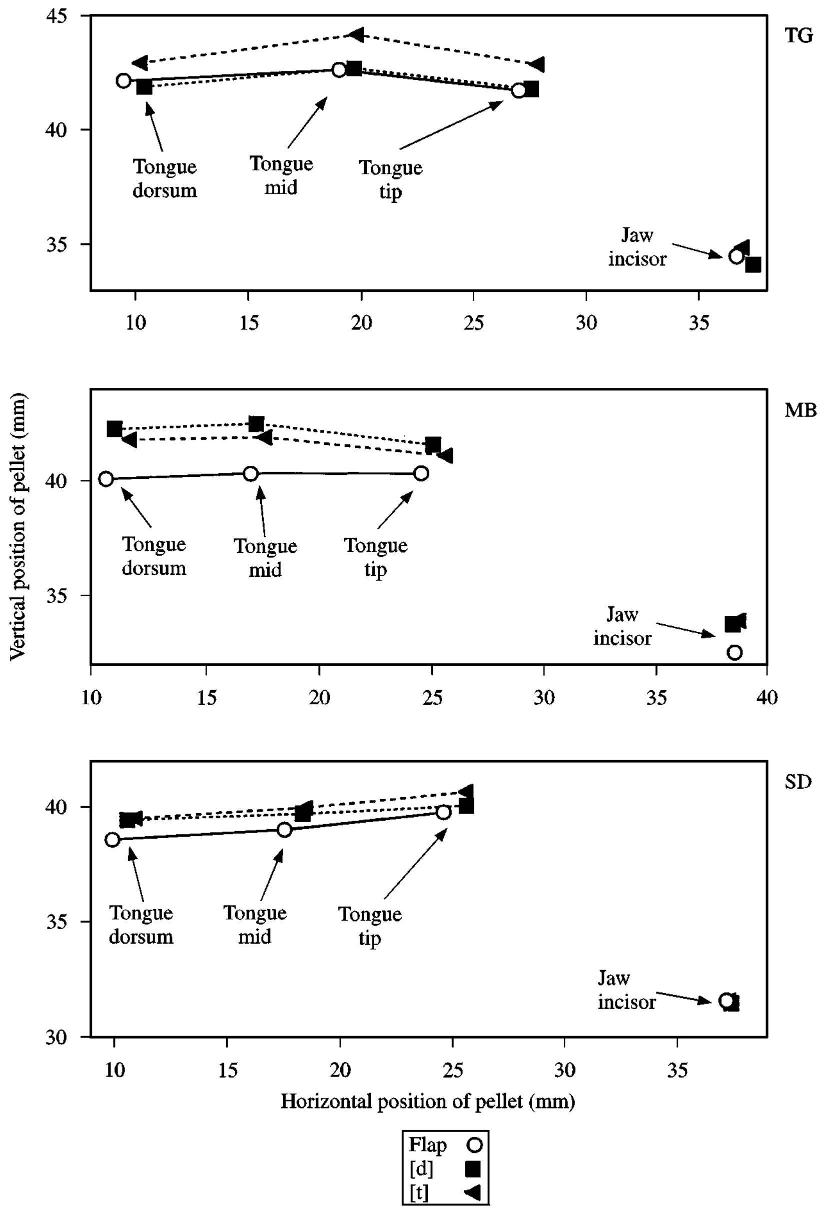

Figure 2. Mean pellet positions plotted in the sagittal plane. Positions are expressed in mm's from an origin located near the bend in the subjects' jaws.

surface is consistent across the different consonants within each subject, but not across subjects. TG and MB both exhibit a slightly convex shape, while SD exhibits a slight cupping of the surface.

To examine jaw position differences in more detail, Fig. 3 plots the position of the jaw in the target consonant. In this figure, it is plotted against the minimum jaw position in the vowel following the target consonant, in order to provide separation of each token, as

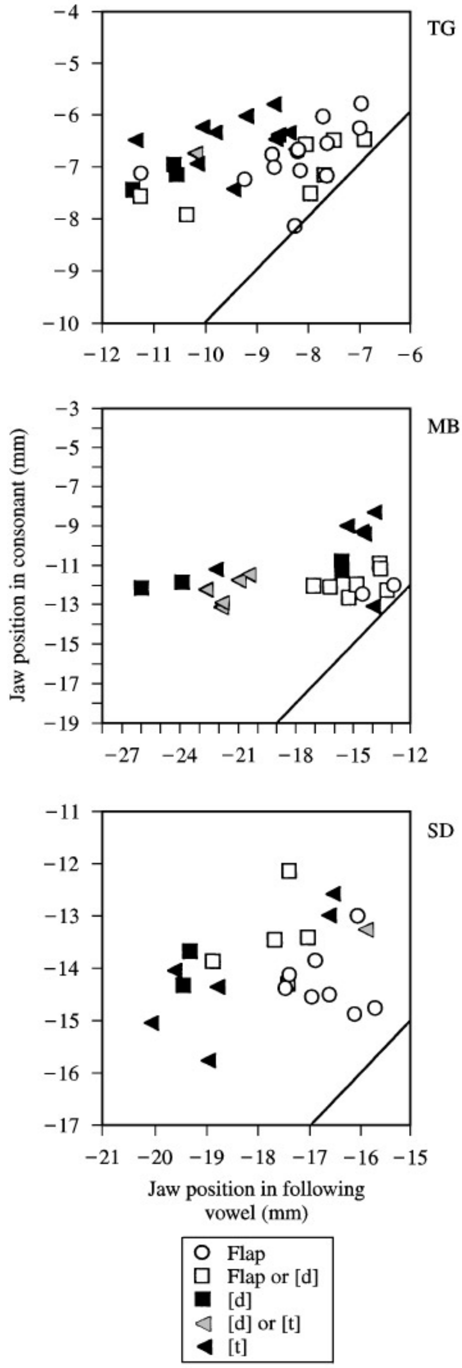

Figure 3. Jaw position in the target consonant plotted against its position in the following vowel for all three subjects. The diagonal line to the lower right of each panel indicates the x = y function, which indicates no opening movement from the consonant. Scales indicate mm's in a dimension rotated about an origin located near the bend in the subjects' jaws.

well as to give some notion of jaw movement out of the consonant. The diagonal line to the lower right indicates the x = y function; tokens on this line would indicate that the jaw had not moved in transition from the consonant to the following vowel. With some exceptions, there is a general separation between flap-like tokens (hollow symbols) and non-flaps in the horizontal dimension, indicating that the jaw is typically higher in vowels after a flap. This is expected if flapping happens before an unstressed and reduced vowel. There is also a pretty good separation in the diagonal dimension, indicating that jaw movement is smaller in flaps. What is not clear, however, is a separation of the tokens in a vertical dimension; hence the failure of the IY measurement to correlate with the perception of flapping. One notices a vertical separation between jaw position in flaps and [d]'s in MB's and SD's tokens; however, this separation is quite weak compared to the separation in the other dimensions. Thus, although jaw movement differences do correspond to the perception of flapping, the critical aspect of these movement differences for producing a difference in coronal closure, position during occlusion, does not differ consistently. This is the type of pattern expected if jaw movement and flapping are both connected with stress on the following syllable, but flapping is not caused by the jaw movement differences.

Fig. 4 plots the location of the blade pellet at the time of release for each of the speakers. What one notices immediately is that tokens which are unambiguously heard as [t] often exhibit higher or more anterior blade locations, suggesting a larger amount of contact between the tongue tip and alveolar ridge. Note that this difference in position separates [t] from flaps, while [d] positions are usually quite close to flap positions. Hence, blade pellet position failed to correlate with perceived flapping for /d/ tokens.

Fig. 5 plots the location of the dorsum pellet at the time of release for each of the speakers. Although the pattern is clearer with some speakers than others, a pattern of separation between flaps and non-flaps is apparent. Dorsal positions, as expected by a co-production account, are higher and/or more anterior (closer to the occlusion point) for [t]'s and [d]'s than for flaps. This holds true for both [d] and for [t].

These articulatory patterns do not support a simple jaw-movement reduction account of flapping, since jaw positions do not actually differ consistently between flapped and unflapped tokens at the time of closure. However, differences between tongue positions in flaps and stops are broadly consistent with a more general articulatory co-production account where blending of vocalic demands in the surrounding vowels and the demands of the consonant on tongue positioning produce a compromised lower and more posterior positioning corresponding to perceived flaps. This account is pursued further in the next section.

3.4. *Problems with a simple co-production account*

A closer examination of the articulatory data does not, however, consistently support a simple co-production account. A first problem involves the mutuality of the effect of the consonant and the neighboring vowels. With a simple co-production account where flapping is the result of blending the consonant and neighbouring vowel gestures, one would expect a mutual effect between the consonant and the neighbouring vowels. Examination of tongue body positions in the neighbouring vowels, however, does not confirm this expectation. Fig. 6 plots (to the left) dorsal positions for the vowel preceding the target consonant, and (to the right) dorsal positions in the vowel following the target consonant. There is a tendency for tongue body positions to be higher in vowels

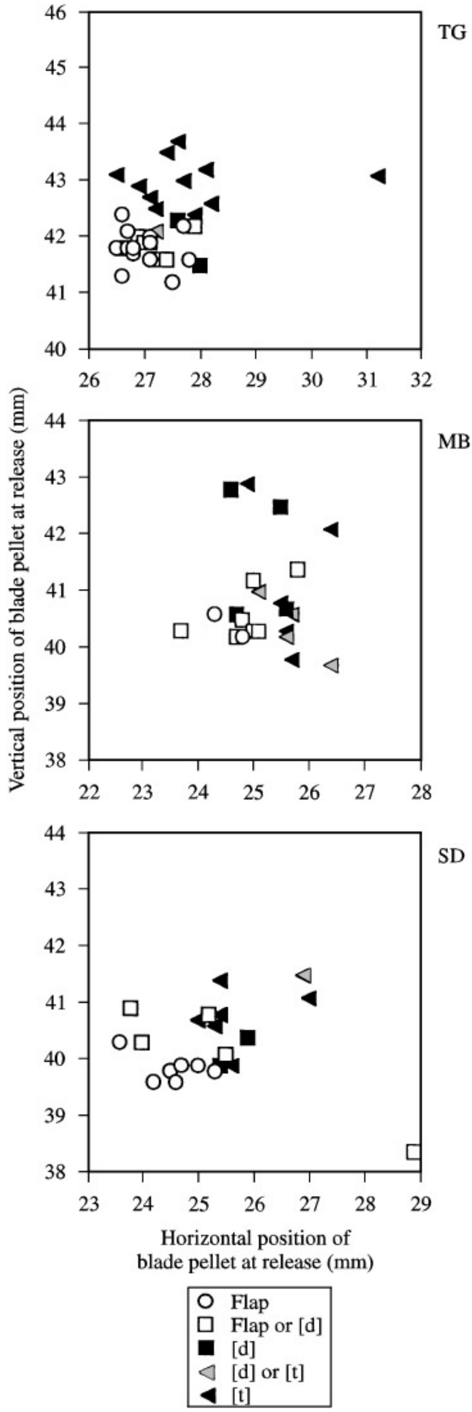

Figure 4. Position of a pellet placed 10 mm from the tongue apex at time of acoustic release plotted in the sagittal plane. X indicates distance in mm's parallel to the occlusal plane from the origin point. Y indicates distance in a perpendicular dimension.

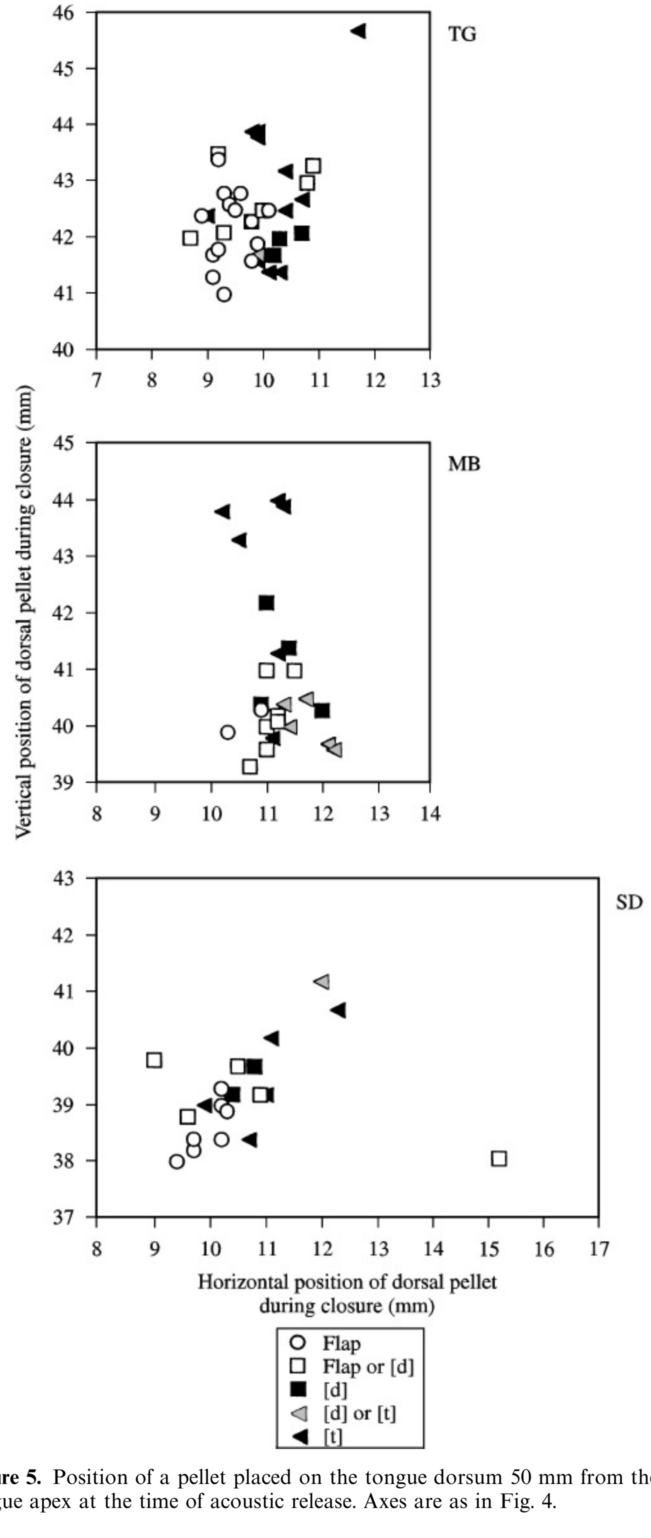

Figure 5. Position of a pellet placed on the tongue dorsum 50 mm from the tongue apex at the time of acoustic release. Axes are as in Fig. 4.

following flaps, as one would expect with a simple coproduction account. However, there is a similar tendency for tongue body positions to be lower in vowels preceding flaps. These tendencies were confirmed by an ANOVA with subject and a binary variable indicating two or more transcriptions as a flap (tokens with no fill in Fig. 6) as independent variables, and *x*- and *y*-positions for the previous and following vowels as dependent variables. Significance levels were set at a value such that the family-wise error for the block of analyses is equal to 0.05. This analysis revealed a large main effect of speaker, and a main effect of flapping on DY in the previous vowel [$F(1,76) = 7.00$, $p < 0.01$]. There was also a trend for the following vowel [$F(1,76) = 5.33$, $p < 0.03$]. The direction of the effects were opposite, DY for flaps being lower for the previous vowel and higher for the following vowel. Other effects were not significant. What is unexpected here is that the previous vowel actually shows less effect of the following consonant in cases in which the consonant is flapped.

Another detail of the data does not conform well to a simple co-production account. Fig. 7 illustrates tongue stretching effects by plotting the distance between the dorsal and mid pellet against the distance between the mid and blade pellet. A co-production account would predict, if anything, a greater amount of stretching in flaps, cases in which the demands of the vowels on the tongue body conflict more with those for the flapped consonant. Increasing amounts of stretching of the tongue would be indicated by tokens lying higher and further to the right. A cursory examination of this figure reveals generally a large amount of overlap between the flapped and unflapped tokens, thus yielding the poor correlations noted above. However, closer examination does reveal a general trend, apparent especially in MB and SD, toward more tongue stretching in stopped tokens. In addition, TG exhibits one stopped outlier in the same direction. The significance of this effect was tested by an ANOVA with speaker and flap transcription as independent variables, and distance between the pellets as the dependent variable. This analysis finds a strong main effect for subject and a weak main effect for flapping on the distance from blade to mid [$F(2,75) = 5.45$, $p < 0.025$]. Other effects were not significant. Note particularly that the lack of interaction between subject and flapping on blade-to-mid distance [$F(2,75) = 0.70$, $p > 0.45$] indicates that this main effect of flapping is not due to the single outlier for TG. Thus, there is a trend toward flapped tokens having more token-to-token similarity in tongue stretching, while stops tend to exhibit differences between tokens in pellet-to-pellet distance.

4. Discussion and conclusion

The current analyses were designed to determine correlates of perceived flapping, and by comparing and examining these correlates to ascertain the plausibility of a traditional account of flapping as due to a categorical rule and of an articulatory account of flapping as a by-product of more general prosodically governed processes.

Some aspects of the current data conform to the simple traditional categorical account. Transcribers were quite consistent in their transcriptions of whether flapping had occurred or not. Also, the strongest correlates of the transcription categories show a fairly clear separation of flaps from stops. The particular features implicated by this analysis are the presence of voicing in the closure and the lack of any disruption of modal voicing at release, indicating among other things the presence of a salient burst release. These acoustic attributes can reasonably be taken as representatives of featural changes

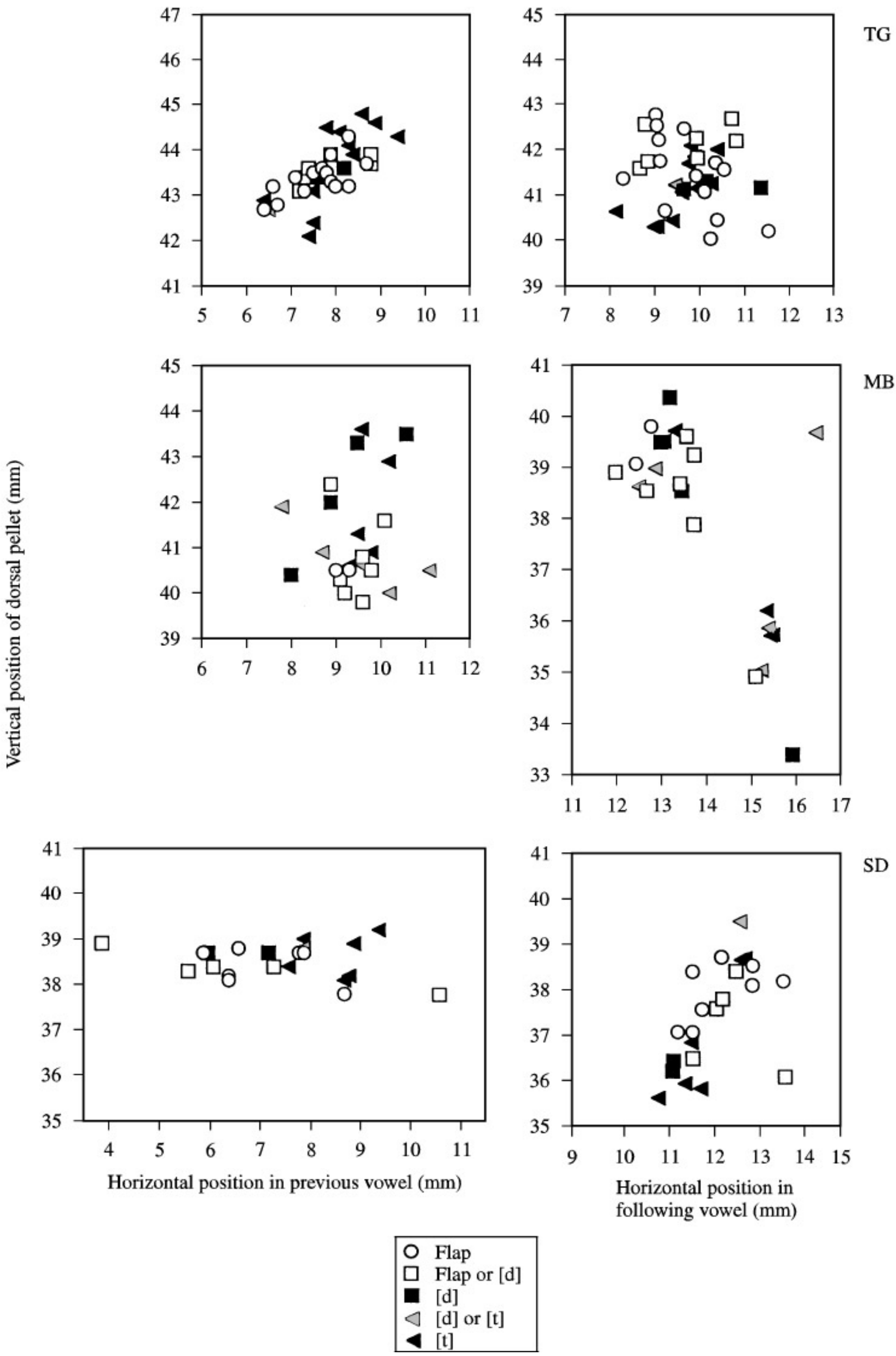

Figure 6. Position of dorsal pellet (placed 50 mm from tongue apex) at a time most indicative of vowel targets for vowels preceding (to the left) and following (to the right) of target consonants. Axes are as in Fig. 4.

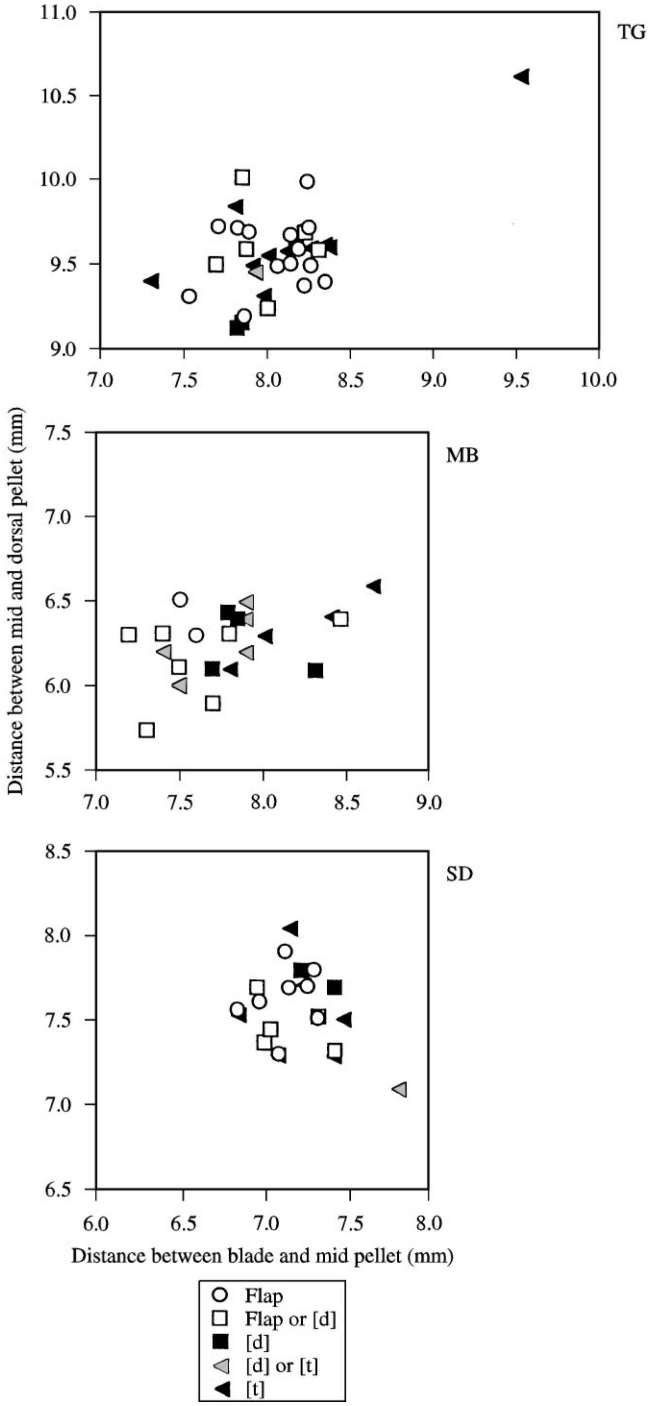

Figure 7. Distance between dorsal and mid tongue pellets plotted against the distance between mid and blade pellets. Greater tongue surface stretching is indicated by tokens appearing further up and to the right.

implicated in the phonological literature, [voiceless] → [voiced] and [obstruent] → [sonorant], respectively. (In the second case, works such as Steriade, 1993, point out that a major phonological difference between stops and sonorants is the existence of a release phase for the obstruents.)

There are other aspects of the current data which bear on a traditional account. First, the application of a flapping rule must be optional. Further, this optionality cannot be accounted for by including prosodic structure in the structural description of the rule; the current results show an inconsistent relationship between prosodic structure and the occurrence of flapping. The current speakers occasionally elected not to flap. Similarly, when the lexical item containing the target of the rule is nuclear accented, speakers usually, but not always, did not flap. It is possible that this optionality is due to the fact that the rule environment in the current corpus spanned a lexical boundary. If this is the cause for the speaker inconsistency, the rule would have to be formulated so as to be sensitive to lexical boundaries. This sort of reformulation, however, does not fit well with standard generative conceptualizations of the action of the rule, in which there is only one flapping rule, a post-lexical rule. Post-lexical rules are said to be sensitive to prosodic domains, but not to morphological structure. In addition, though this question cannot be addressed with the current corpus, it also seems quite likely that the flapping rule can also be suppressed word-internally in certain circumstances, such as talking to childern or in emphatic conditions. For example, Nespor and Vogel's (1986) treatment of flapping includes two caveats about the flapping rule not applying as expected; the first (p. 23) mentions the *t* in *water*, while the second (p. 225) discusses the same effect for the rule across a word boundary. Thus, the flapping rule must be couched within some sort of theoretical apparatus which allows it to relate probabilistically to the various conditions which trigger it.

Additionally, a closer examination of the data reveals a broad range of gradient aspects of flapping which are also not accounted for by simply citing a grammatical rule. First, there is one speaker in particular who produced a variety of tokens which were not well defined as flaps or stops by the acoustic criteria implicated by the other two speakers. Also, a second speaker produced voiceless flaps. Also, transcription inconsistencies reveal a fortis to lenis continuum ranging from [t] through [d] to flap. Second, standard generative conceptions of rule application also lead one to expect a well-defined articulatory expression of the featural differences in the rule output. Here, however, the separation of flaps from stops is much clearer in the acoustic measures than in the lingual measures. Thus, any model which leads one to expect a significant reorganization of lingual gestures corresponding to the occurrence of flapping is incorrect. Lingual positions and motions are indeed somewhat different between [t]'s and [d]'s; however, lingual motions for flaps and [d]'s are quite similar.

These observations suggest a model in which a gradient articulatory change produces acoustically quantized outputs which transcribers identify as a difference between flaps and stops. The two such explicit models which were examined here in some detail, however, at best broadly approximate the data. Jaw position does not differ consistently between flaps and stops. The general co-production model fares a bit better, since it predicts a difference in tongue-body positioning which does affect flap perception. Since this model relies on the antagonism between the coronal consonant and the neighbouring vowels in tongue-body positioning, it predicted the strong differences in tongue body positioning which were not necessarily accompanied by jaw position differences.

This model, however, seems to be an oversimplification in that it predicts mutual effects of vowels and consonants which do not occur consistently, and it does not explain why the lingual tissues show some decreased stretching in the flapped tokens. With respect to consonant-to-vowel effects, there is a strong directionality to the effects. Thus, if a co-production model is to account for flapping, it must insure co-production of the stop with the following vowel, not the preceding vowel. However, before pursuing such an account, it must be noted that the separation between flaps and stops is much clearer in tongue body positioning at the time of closure than it is in the middle of the following vowel. This pattern parallels the observation (such as by Hayes, 1995, and Turk, 1992) that flapping is not always accompanied by vowel reduction. Thus, one suspects that this connection between tongue body positioning in the following vowel and flap perception is, like the results for jaw positioning, due to parallel reduction effects on the consonant and on the vowel, rather than due to tongue body positioning in the vowel causing the reduction of the consonant to a flap.

With respect to tongue stretching, the co-production account claims that biomechanical constraints on the linkage between the tongue body and the tongue tip create stop lenition to flap. Since this is not a solid linkage, one might expect a certain amount of flexibility in the constraint which could be exploited by the articulatory sytem executing the dorsal and tip gestures simultaneously. Here, however, it appears that this potential flexibility is taken advantage of in cases where there is supposed to be *less* coarticulatory overlap, and hence less direct mechanical antagonism between the gestures. This observation suggests that, in the cases of flapped stops the speakers, to put it crudely, are not trying as hard to make a salient occlusion. This is the type of explanation pursued in more detail in de Jong (1995) for a broad range of stress-sensitive kinematic effects. Unstressed material, such as the flap and the vowel which follows it, receives less articulatory attention and thus exhibits more characteristics of what Lindblom (1990) calls hypo-articulate (or system oriented) speech.

The results of the analyses reported here point out significant problems with both the categorical rule account and with the prosodic by-product account. The conclusion, then, is that American English stop flapping across a word boundary can be described as a variable but quasi-categorical rule, so long as the objects of the rule's description are taken to be acoustic in nature. The results for oral kinematics are not very encouraging for a categorical rule description, in that kinematic measures generally do not exhibit quantization according to flap and [d] categories. This situation suggests that a gradient change in articulatory behaviour is giving rise to somewhat quantized acoustic results, which in turn give rise to consistent transcriptions. If this is right, from the speaker's perspective there is no reason to posit a rule which specifically demands the production of a flap before unstressed vowels. Rather, what is necessary is to understand the language's segmental and prosodic convention sufficiently to know when a salient consonant release is necessary, and when not. A final conclusion to this matter, however, will require both examination of glottal activity and aerodynamics in addition to the oral kinematics examined here, and examination of similar reduction processes word-internally and in other English dialects and other languages.

I thank Edward Flemming, Richard Wright, and Aaron Shryock for performing transcriptions of the data. I also thank Francis Nolan, Terry Nearey, Bruce Derwing, and two anonymous reviewers for helpful suggestions concerning the research reported here. Data acquisition was supported in part by the NSF under Grant No. IRI-8617852 to Mary Beckman and NIDCD under Grant No. T32DC00029 to the UCLA Phonetics Lab.

References

Bloch, B. (1941) Phonemic overlapping. *American Speech*, **16**, 278–284

Bloomfield, L. (1993) *Language*. (New York: Holt)

de Jong, K. (1991) *The Supralaryngeal Articulation of English Stress Accent*. Ph.D. dissertation, Ohio State University

de Jong, K. (1995) The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, **97**, 491–504

de Jong, K., Beckman, M. E. & Edwards, J. R. (1993) The interplay between prosodic structure and coarticulation. *Language and Speech* **36**, 197–212

Fujimura, O. (1986) Relative invariance of articulatory movements: an iceberg model. In *Invariance and Variability in Speech Processes*. (J. Perkell and D. Klatt, eds.), pp. 226–234. Hillsdale, N.J.: Lawrence Earlbaum Assoc

Haugen, E. (1938) Notes on "voiced t" in American English. *Dialect Notes*: **6** (16–17) 627–634

Hayes, B. (1995) *Metrical Stress Theory*. Chicago: University of Chicago Press

Kahn, D. (1980) *Syllable-based Generalizations in English Phonology*. New York: Garland

Kenyon, J. S. (1940) *American Pronuciation: a Textbook of Phonetics for Students of English*. Ann Arbor, Mich.: George Wahr

Kirk, R. E. (1982) *Experimental Design: Procedures for the Behavioral Sciences, 2nd ed*. Monterey, Calif.: Brook/Cole Pubs

Ladefoged, P. & Maddieson, I. (1996) *The Sounds of the World's Languages*. Oxford: Blackwell

Lindblom, B. E. F. (1990) Explaining phonetic variation: a sketch of the H&H Theory. In H.J. Hardcastle, & A. Marchal (eds.), *Speech Production and Speech Modeling, NATO ASI Series D: Behavioural and Social Sciences*, Vol. 55 (H.J. Hardcastle & A. Marchal, eds.), pp. 403–439. Dordrecht: Kluwer A.P.

Macchi, M. (1985) *Segmental and Supersegmental Features and Lip and Jaw Articulators*. Ph.D. dissertation, New York University

Monnot, M. & Freeman, M. (1972) A comparison of Spanish single-tap /r/ with American /t/ and /d/ in post-stress intervocalic position. In *Papers in Linguistics to the Memory of Pierre Delattre* (A. Valdman, ed.), pp. 409–416. The Hague: Mouton

Nadler, R., Abbs, J. H. & Fujimura O. (1987) Speech movement research using the new X-ray microbeam system. *Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn*, 1–7 Aug., 1987, vol. 1, pp. 221–224. Tallinn: Academy of Sciences of the Estonian Soviet Socialist Republic

Nespor, M. & Vogel, I. (1986) *Prosodic Phonology*. Dordrecht: Foris

Oswald, V. A. (1943) Voiced *t*—a misnomer. *American Speech* **18**, 18–25

Pierrehumbert, J. B. (1980) *The Phonology and Phonetics of English Intonation*. Bloomington, Ind.; Indiana University Linguistics Club

Pierrehumbert, J. B. & Beckman, M. E. (1988) *Japanese Tone Structure, Linguistic Inquiry, Monograph 15*. Cambridge, Mass.: MIT Press

Steriade, D. (1993) Closure, release and nasal contours. In *Nasality* (M. Huffman & R. Krakow, eds.), pp. 401–470. San Diego: Academic Press

Steriade, D. (1996) Paradigm uniformity and the phonetics-phonology boundary. Paper presented at the Vth Conference in Laboratory Phonology, Evanston, Illinous, July, 1996

Stone, M. & Hamlet, S. (1982) Variation in jaw and tongue gestures observed during the production of unstressed /d/s and flaps. *Journal of Phonetics*, **10**, 401–415

Trager, G. L. (1942) The phoneme "t": a study in theory and method. *American Speech*, **17,** 144–148

Turk, A. E. (1992) The American English Flapping Rule and the effect of stress on stop consonant durations. *Cornell Working Papers in Phonetics*, **7**, 103–133

Twaddell, W. F. (1935) *On Defining the Phoneme*, Language Monographs, No. 16. Baltimore: Waverly Press

Umeda, N. (1977) Consonant duration in American English. *Journal of the Acoustical Society of America* **61**, 846–858

Vanderslice, R. & Ladefoged, P. (1972) Binary suprasegmental features and transformational word-accentuation rules. *Language*, **48**, 819–838

Wilkinson, L. (1979) Tests of significance in stepwise regression. *Psychological Bulletin*, **86**, 168–174

Wolfram, W. & Johnson, R. (1982) *Phonological Analysis: Focus on American English*. Englewood Cliffs, N.J. Prentice Hall Regents

Zue, V. & Lafferriere, M. (1979) Acoustical study of medial /t/, /d/ in American English. *Journal of the Acoustical Society of America*, **66**, 1039–1050