An articulatory examination of word-final flapping at phrase edges and interiors

Teruhiko Fukaya
Dani Byrd
Department of Linguistics, University of Southern California
fukaya@usc.edu, dbyrd@usc.edu

Formulations of flapping as a symbolic phonological rule suggest clear articulatory differences between flaps and stops, and often offer no overt explanation for why phrase boundaries should block the alternation. The present study explores the articulatory foundation of the distinction between flaps and non-flaps in word-final position. We examine kinematic and acoustic data for these articulations in phrase-final and phrase-medial positions and in falling- and level-stress contours. It is shown that a discrepancy exists between acoustic and articulatory durational patterning – while acoustic durations of flaps are shorter than those of non-flaps overall, their articulatory durations are not uniformly so. It is important to consider multiple potential articulatory sources – both spatial and temporal – for the acoustic shortness that characterizes flaps, including spatial reduction, temporal articulatory shortening, and changes in intergestural coordination. The kinematic data indicate that different sources of word-final flap shortness exist for different speakers and different prosodic conditions, suggesting that gradient variability in the spatiotemporal patterning of tongue-tip constrictions yields acoustic shortening in word-final flaps.

1 Introduction

Traditional phonological accounts of the occurrence of flaps in American English have generally been formulated as a symbolic phonological rule, sometimes variable (e.g. Zue & Laferriere 1979), in which an underlying alveolar stop consonant is replaced by or changed into a flap consonant. Such accounts in terms of alternation rules suggest clear articulatory differences between flaps and stops. Under a traditional interpretation, such a rule states that the underlying consonant /t/ is replaced in a word's realization by a distinct consonant [ɾ], suggesting a featural change and hence a shift in articulation. A typical example of the traditional flapping rule is as follows:

t => ɾ/[−consonantal] __(#)$V_{[-\text{stress}]}$,
where # may not be a phrase boundary (adapted from Hayes 1995 : 12)

It should also be noted that the two conditions on the environment (i.e. the stresslessness of the following vowel and the absence of a phrase boundary) are stipulations, and no explanation of these contextual requirements is provided by the rule itself, though see Kiparsky (1979: 437f.) for a foot-based account in which phrase boundary effects are more comfortably

Journal of the International Phonetic Association (2005) 35/1
doi:10.1017/S0025100305001891

accommodated. Rather than being concerned with the phonological characterization of flap distribution (on this see e.g. Kahn 1976, Kiparsky 1979, Selkirk 1982, Gussenhoven 1986, Banner-Inouye 1995, and Steriade 2000), this paper focuses on a phonetic understanding of the phenomenon, which, based on the phonological characterizations, is expected to critically involve prosodic context.

A foundational study characterizing the acoustics of flaps is that of Zue & Laferriere (1979), which observes flaps to be acoustically very short (10–40 ms); other studies with relevant acoustic data on flaps include Umeda (1977), Byrd (1993), and Stathopoulos & Weismer (1983). Other work has focused on the variable nature of flapping; for example, Patterson & Connine (2001) document the sensitivity of flapping to word frequency and morphological complexity.

Some studies have examined articulatory data to evaluate the phonetics of flapping. Stone & Hamlet (1982) examine acoustics and tongue and jaw movements during the production of /d/s and flaps in nonsense phrases consisting of repetitions of the syllable 'da' in an alternating stress/unstress pattern. They identify four different acoustic patterns for /d/s in the unstressed syllables: (i) a voiceless or partially voiced /d/; (ii) a voiced /d/ with a normally long closure duration; (iii) a short /d/, which was voiced or voiceless; and (iv) a flaplike allophone that 'appeared as a momentary decrease in the intensity of the preceding and following vowels and during which there was occasionally a small burst' (Stone & Hamlet 198: 404). Tongue gestures as well as consonant jaw height and subsequent opening jaw acceleration are found to be related to these acoustic patterns. 'Long /d/s [patterns (i) and (ii)] are produced with more close jaw positions than were short /d/s and flaps' (Stone & Hamlet 1982: 414), and opening jaw acceleration into the following vowel was larger with long /d/s than with short /d/s or flaps. Area of tongue–palate contact for the consonant was 'greatest during pattern 1, lesser during patterns 2 and 3, and least during pattern 4' (Stone & Hamlet 1982: 414). In addition, the patterns of jaw movement that they observe lead Stone and Hamlet to the supposition that [d] and [ɾ] are not actually two distinct sounds that vary allophonically in English, but rather a continuum of sounds roughly represented by the four acoustic patterns that they discuss (Stone & Hamlet 1982: 409).

In an attempt to answer the question of whether a general timing mechanism underlies the American English Flapping Rule, Turk (1992) examines word-medial intervocalic stops (/p, t, k, b, d, g/) in three stress environments: between an unstressed and a stressed vowel, between a stressed and an unstressed vowel, and between unstressed vowels. She finds that 'for all stops except /g/, a phonetic consequence of occurring in flapping environments is reduced length: total stop durations word-medially in all places of articulation tend to be longer in pre-stress position' (Turk 1992: 128). This general temporal reduction that Turk (1992) finds may have more dramatic or distinct acoustic consequences for alveolar stops, but the underlying motivation for the phenomenon is a prosodic one that does not pick out a single place of articulation for a symbolic alternation. This view is consistent with Browman & Goldstein's (1992) observation that spatial reduction occurs, for all sorts of articulations, in the prosodic context in which flapping obtains. They specifically note the tendency for reduction in glottal opening gestures in this context (thereby also helping explain the loss of voicelessness when a [ɾ] surfaces in a /t/ context). Browman & Goldstein comment 'the [temporal and spatial] reduction process would always reduce the oral gesture in this environment, but the contour that is perceived as a flap would simply be one of the possible output consequences, depending on the appropriate set of gestures' (p. 170).

In a recent kinematic study, de Jong (1998) evaluates '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' (p. 305). He investigates word-final alveolar stops (/t/ and /d/) in three prosodic environments: post-nuclear (unaccented), pre-nuclear and accented, and nuclear accented with a following unaccented vowel. He examines acoustic data including occlusion durations, voice onset times, and proportion of voicing in the closure, as well as spatial articulatory data including the positions of the jaw and the tongue. De Jong concludes 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' and points out that '[oral] kinematic measures generally do not exhibit quantization according to flap and [d] categories' (de Jong 1998: 309). He goes on to suggest that 'a gradient change in articulatory behaviour is giving rise to somewhat quantized acoustic results, which in turn give rise to consistent transcriptions' (de Jong 1998: 309).

As models of such a hypothesis, de Jong (1998) considers two possibilities: (i) that flaps are the result of the reduction of jaw movement (a lenition model) and (ii) that flaps are due to increased overlap between the tongue-tip gesture and the (following) vowel (a co-production model). He rejects the former because no consistent differences are found in jaw position between flaps and stops (de Jong 1998). He pursues the latter, pointing out that lower and more posterior tongue positioning observed in flaps is consistent with an articulatory co-production account in which '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' (de Jong 1998: 302), but also describes potential oversimplifications of such an account (p. 309).

The kinematic articulatory study of word-final flaps that we present below investigates the distinction between voiceless and voiced (flap) realizations of underlying word-final /t/ in the articulatory domain, entertaining the possibility that gradient changes in articulatory dimensions in alveolar consonants, which are specific to certain prosodic environments, give rise to quantal differences in percept (namely, yielding the percept of flap consonants), as suggested by de Jong (1998). We are also able to add a consideration of articulatory velocity profiles not available in de Jong's (1998) study.

Word-final /t/s occurring phrase-internally and at phrase boundaries in two stress contour patterns (falling and level) are examined. We report effects of phrase boundary and stress conditions on acoustic duration and kinematic properties of tongue-tip movement. Pursuing Turk's (1992) characterization of flaps as short stops, we consider distinct articulatory mechanisms (see figure 1) as possible sources for this acoustic shortening, enumerated as follows:

1. Articulatory spatial reduction of a constriction may shorten acoustic duration if the articulator spends less time in contact with the passive articulatory surface due to less extreme positioning in the mouth. (See figure 1a [less extreme final positioning] and figure 1b [less extreme initial positioning].)
2. Decreases in articulatory duration may shorten the acoustic constriction duration. (See figure 1c.)
3. Articulatory overlap among gestures may shorten acoustic duration due to truncation or undershoot of one gesture as another gesture encroaches on it. (See figure 1d.)

Crucially, changes in both the spatial and/or temporal articulatory domains can affect acoustic durational patterning. Importantly, these are not mutually exclusive.

We discuss and evaluate these possibilities in the experiment that follows, using kinematic signatures for these possible changes. The particular kind of kinematic evidence that would support each of these possibilities will be outlined in section 3.4.

2 Method

The present study examines word-final /t/ in four fully-crossed experimental conditions – phrase-internally versus phrase-finally and in a falling-stress contour versus a level-stress contour. Formulations of the flapping rule generally specify that for flapping to occur the alveolar consonant must not occur before a phrase boundary and must occur preceding an unstressed syllable. However, for word-final consonants this stress restriction is known to be

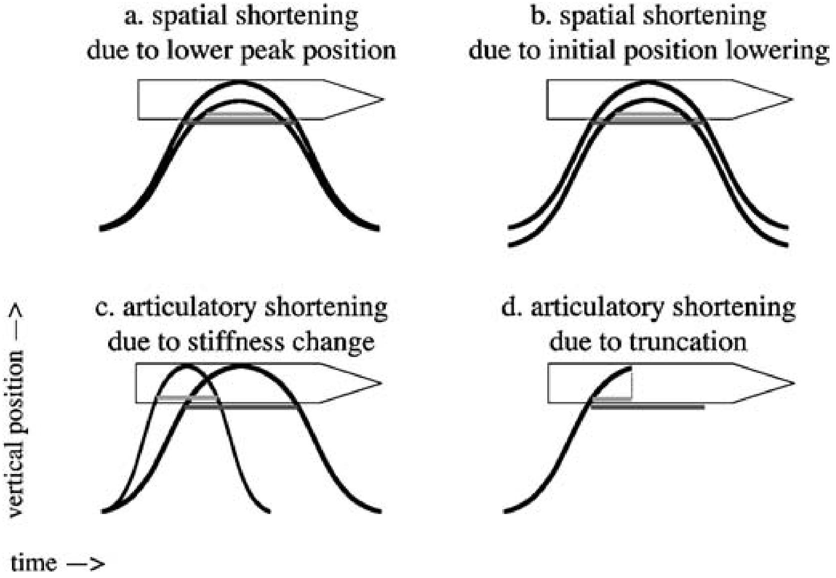

Figure 1 A schematic showing four articulatory modifications that could give rise to short acoustic constriction intervals. The box indicates compression of the active articulator against the palate; i.e. the first contact of the articulator with the palate is at the bottom of the box and further upward movement would compress the active articulator against the passive (palate) continuing an interval of linguapalatal contact. The dark grey horizontal bold line indicates the unmodified linguapalatal contact interval, and the light grey line indicates the contact interval resulting with the modified gesture.

flexible, and for this reason the two stress conditions are investigated here. (Note that we do not use a terminological distinction here between taps and flaps.)

Three speakers were recorded both acoustically (20 kHz sampling rate) and using the EMMA (Perkell et al. 1992) magnetometer system for movement tracking of the tongue tip (other articulators were also recorded but are not analyzed here). The movement data was sampled at 500 Hz, head-corrected, rotated to the occlusal-plane, and low-pass filtered at 25 Hz.

Table 1 Read stimuli (note that no underlining appeared in the stimuli as presented for the subjects). (Some of the sentences are based on examples in Cooper & Paccia-Cooper (1980: 151).)

Boundary condition	Stress condition	Sentence
No Boundary	Falling stress	Pat said that hot asparagus is the tastiest dish.
No Boundary	Falling stress	If you like to knit a lot, the store downtown has yarn on sale.
No Boundary	Falling stress	Ian said that late applications should be sent to the Dean's office.
No Boundary	Level stress	Pat said that hot apples are the tastiest dessert.
No Boundary	Level stress	Ian said that late inquiries should be sent to the Dean's office.
No Boundary	Level stress	If you can knit aptly, sweaters have a lot of appeal.
Boundary	Falling stress	When it's served hot, asparagus is the tastiest dish.
Boundary	Falling stress	If you like to knit, a lot of stores downtown have yarn on sale.
Boundary	Falling stress	Even if they're late, applications should be sent to the Dean's office.
Boundary	Level stress	When they're served hot, apples are the tastiest dessert.
Boundary	Level stress	Even if they're late, inquiries should be sent to the Dean's office.
Boundary	Level stress	If you like to knit, Alpine sweaters have beautiful yarn.

The target sentences are given in table 1. (In order to allow for examination of a set, albeit small, of lexical items for both phrasal conditions and for both stress conditions, some variability in the neighboring segments was unavoidable. However, the oral consonant preceding the target constriction formation was always a tongue-tip consonant, thereby creating a degree of consistency in the behavior of the articulator and constriction studied.) Stimuli were blocked by boundary condition (to simplify the reading task and ensure that the same prosodic boundary type was used in multiple repetitions). Ten repetitions were recorded, randomized within blocks separately for each subject. This yielded a total of 360 tokens for analysis. Regarding the level-stress sentences (which could potentially have marked falling-stress readings), listening confirmed that all speakers did stress the post-target syllable at least equally with the pre-target syllable.

The data analysis included both articulatory and acoustic measurements, though articulatory information is of primary interest in this study. The observed acoustic variables include the following (see Zue & Laferriere 1979 for the general approach):

1. acoustic duration – time between end of voicing (or if none, amplitude drop) to release burst (or if none, increase in amplitude)
2. narrow transcription – flap or /t/

presence or absence of:

3. voicing
4. pre-laryngealization (creaky voice/irregular glottal pulses preceding the flap closure)
5. post-laryngealization (creaky voice/irregular glottal pulses following the flap closure)

A USC-modified version of the Matlab-based MAVIS environment (Tiede et al. 1999) was used to measure algorithmically the tongue-tip sensor movement trajectory in the horizontal (x) and vertical (y) dimensions. The articulatory variables measured were:

6. constriction gesture extremum position (x & y)
    - at the moment of y-velocity zero-crossing at maximum constriction
7. constriction gesture displacement (x & y)
    - defined as the x or y difference between tongue-tip receiver positions at the zero-crossings at constriction onset and extremum
8. constriction (Euclidean) distance traveled
    - defined as the straight-line distance between tongue-tip receiver positions at the zero-crossings at constriction onset and extremum
9. constriction formation duration (i.e. articulatory duration)
    - defined by y-velocity zero-crossings at constriction onset and extremum
10. constriction gesture peak velocity (y)
11. acceleration duration (y)
    - duration from the time of zero-crossing at constriction onset to the time of constriction peak velocity
      (Note that this reflects the stiffness parameter of a gesture within a Task Dynamics model of speech production (Saltzman & Munhall 1989), i.e. the parameter shaping the internal temporal properties of an articulatory gesture.)
12. proportional time-to-peak-velocity (y)
    - the ratio of acceleration duration to total constriction formation duration
      (Note that the *ratio* of acceleration to total duration [i.e. proportional time-to-peak velocity] is chosen to differentiate truncation due to overlap from a stiffness change in the tongue-tip gesture because a gestural stiffness change could affect both acceleration and deceleration durations comparably.)

Figure 2 provides two sample tokens showing the onset point, peak velocity point, and extremum point for a flap realization (top) and a non-flap realization (bottom) (Speaker J). Note that while de Jong's 1998 study includes many important spatial variables, this study adds to his work by examining various TEMPORAL kinematic properties and tongue-tip y-velocity

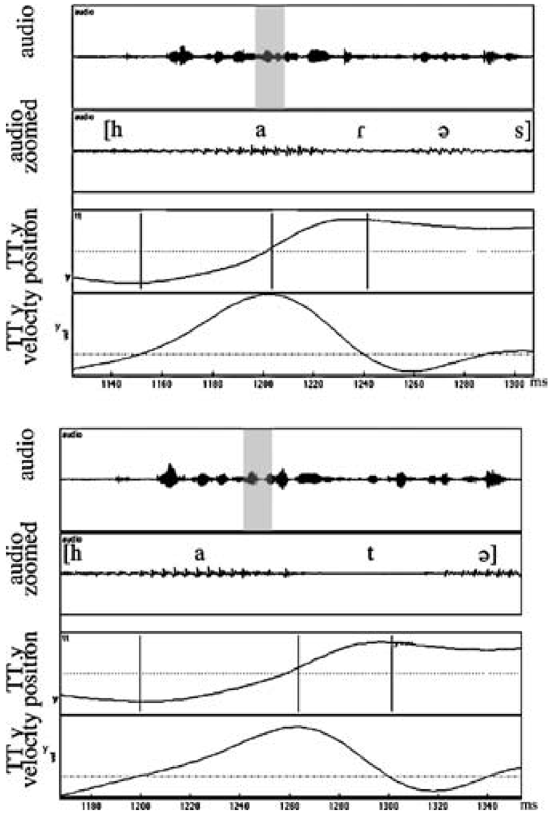

Figure 2 Two sample tokens showing the onset point, peak velocity point, and extremum point for a flap realization (top) and a non-flap realization (bottom). They are aligned at the onset-time-point, and a transcription is shown for the zoomed interval (indicated by the shaded box in the top panels). (Speaker K: 'Pat said that hot asparagus is the tastiest dish.')

profiles, as indicated by the last four measures above. Statistical analyses of the dependent variables were conducted separately for each subject using ANOVA,though a dependent measure was not always available for all tokens.

3 Results

3.1 Transcription and overall flap patterning

One author (TF), with phonetics training, transcribed the realization of the target /t/s, wearing headphones to listen to the target phrase as many times as necessary. The fairly straightforward goal of this procedure was to separate tokens into flap and non-flap realizations. As such, a brief summary of the transcription results is in order. Of the 358 transcribed tokens, 110 were recorded as flaps. Generally, flaps were transcribed in tokens having voiced constrictions; only 3 voiceless tokens were transcribed as flaps and only 12 voiced tokens were transcribed as non-flaps. (Riehl (2003) finds that for four of her six speakers, voicing and duration correlate equally well with flap identification in her forced-choice task and, for a fifth, voicing correlates only slightly more poorly.)

Table 2 A summary of the percentage frequency of word-final flaps (voiced realizations) in the experimental conditions. Percents are rounded to the nearest integer.

Speaker	Phrase-internal		Phrase boundary (stress conditions pooled)
	Falling stress	Level stress	
K 19 flaps, 101 [t]s	50% (n = 15/30)	10% (n = 3/30)	0% (n = 1/60)
J 40 flaps, 78 [t]s	90% (n = 27/30)	41% (n = 12/29)	0% (n = 1/59)
H 60 flaps, 60 [t]s	100% (n = 30/30)	100% (n = 30/30)	0% (n = 0/60)

Because of this consistency and the fact that we consider no underlying /d/s, the remaining statistical analyses contrast voiced versus voiceless realizations of the word-final /t/. That is, we take voicing as indicating a [ɾ] realization. We adopt this criterion for two reasons. First, the transcription of flaps was reliably and extremely highly correlated with voicing, and secondly, we feel that, given this consistency, a waveform inspection for the presence of voicing provides a more objective coding of the data than transcription (though in this case it aligns almost completely with the transcription judgments).

What proves to be of particular interest is the individual subject patterns. For Speaker K, almost all realizations of the word-final /t/ were [t]. However, the one exception to this was Speaker K's phrase-internal, falling-stress condition in which her realizations were variable, about 50% voiced [ɾ]. For Speaker J, voiceless [t] was realized uniformly at phrase-boundaries, with one exception: If the word-final consonant was phrase-internal, 41% were realized as voiced [ɾ] in the level-stress condition and 90% as voiced [ɾ] in the falling-stress condition. Finally, for Speaker H, all phrase-boundary realizations were voiceless [t] and all phrase-internal realizations were voiced [ɾ], regardless of stress pattern. These results are summarized in table 2. Overall, these patterns agree with flapping descriptions requiring no boundary. But the results are variable across subjects for the phrase-internal contexts. Subjects were either wholly or variably prone to flap in the falling-stress contour. Behavior in the level-stress contour generally indicated a lower propensity to flap than in the falling-stress contour, but there was between-subject variability.

The tendency to laryngealize varied idiosyncratically. Speaker H rarely laryngealized her flaps and consistently laryngealized her non-flaps. Speaker J also rarely laryngealized flaps and often laryngealized non-flaps, particularly at release. Speaker K occasionally laryngealized her flaps and often laryngealized her non-flaps.

3.2 Overall results for acoustic and articulatory duration

This coding of the data into voiced flaps and voiceless non-flaps yields acoustic results indicating, as expected, that flaps were short in acoustic duration – mean duration: [ɾ] 20 ms (7 ms SD) vs. [t] 43 ms (17 ms SD) – though there were also some short, voiceless closures. With the phrase boundary condition included, the voiceless tokens had a mean of 77 ms (47 ms SD). The mean flap duration is slightly shorter than that reported by Zue & Laferriere (1979); however, they state that '[a]s a phonetically defined group, flaps vary in duration from 10 to 40 ms' (p. 1044). Overall, the results indicate that the flaps were uniformly acoustically shorter than the non-flaps. (Given the somewhat short duration of the word-final voiceless non-flaps in the phrase-internal condition, it is conceivable that some category other than [t] might plausibly be identified with this production; however, the outstanding phonetic question of what the articulatory differences between the voiced [ɾ] and the longer voiceless realization are remains unchanged.)

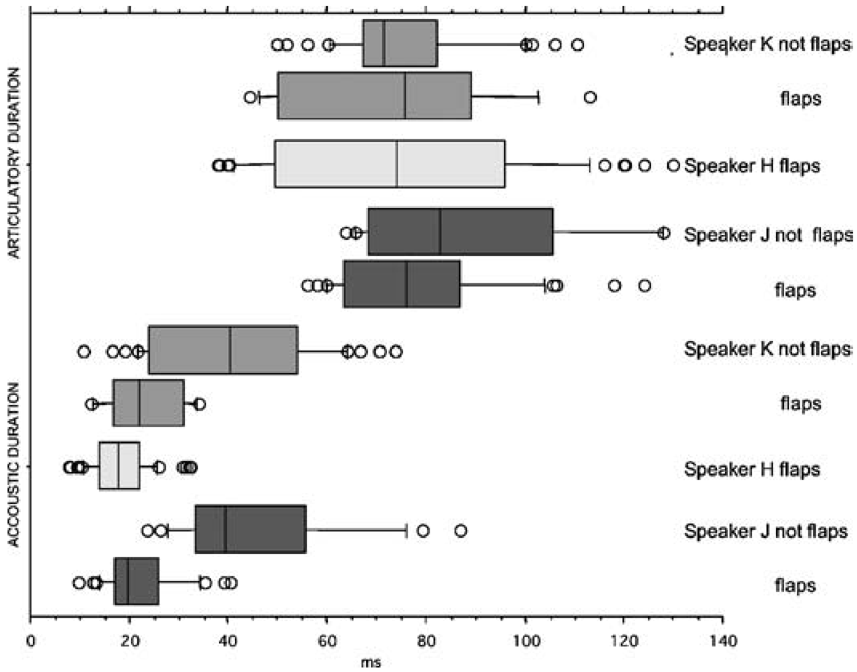

Figure 3 Articulatory and acoustic durations for flap and non-flap realizations in the phrase-internal condition. Box plots show the 10th, 25th, median, 75th, and 90th percentiles. Note the substantial overlap for articulatory durations as compared to the distinct distributions for the acoustic durations.

Though the acoustic durations of the flaps are, as expected, shorter than non-flaps, this does not correspond directly to the closing movement durations, which are NOT uniformly shorter for the flapped (voiced) tokens. This is indicated in figure 3 for the phrase-internal condition. While the mean articulatory duration of the flaps is significantly shorter than the non-flaps, there is a substantial amount of overlap in the distributions – a degree of overlap not seen in the acoustic domain.

The potential disjunct between acoustic and articulatory durational patternings and the complexity of individual differences are further emphasized when the no-boundary condition is examined for the two subjects who had both flap and non-flap realizations phrase-internally – Speakers J and K. A two-factor ANOVA testing for differences in articulatory and acoustic durations as a function of stress and flap/non-flap (i.e. voicing presence) finds for Speaker J main effects of flapping, but not of stress, on both duration values ($F(1,49) = 23.861$, $p < .0001$; $F(1,53) = 7.502$, $p = .0084$). However, there is an interaction effect on the articulatory (but not acoustic) duration ($F(1,53) = 14.396$, $p = .0004$) such that while the flaps are much shorter than non-flaps in the falling-stress condition (72 ms vs. 112 ms), this articulatory duration difference does not obtain in the level-stress condition (in which flaps are very slightly longer (7 ms)). However, for Speaker K no main effects or interactions exist, though there is a large, marginal effect for voicing ($F(1,38) = 3.456$, $p = .07$) that indicates the expected acoustic shortness of the flaps (23 ms vs. 41 ms) (presumably this fails to reach significance because there are only six flaps whose acoustic durations were measurable for this speaker). For Speaker K, differences in ARTICULATORY DURATION among flaps and non-flaps do not exist (means for flaps and non-flaps in both falling- and level-stress conditions are within 3 ms of one another). In summary, while durational differences between flaps and non-flaps exist in the acoustic domain, this is not paralleled by clear and consistent differences in articulatory durations.

3.3 Overall effects of phrase and stress conditions on duration and kinematics

The next step in the analysis is to determine what the effects of the prosodic conditions are on the acoustic duration and on the articulatory kinematics. In this section, we present the overall effects of phrase and stress condition so that general patterns associated with these prosodic factors can be understood independently of any categorical/transcriptional evaluation of the resulting consonant. In two-factor ANOVAs with the factors of phrase condition and stress condition (split for subjects), main effects of both factors were observed for acoustic duration and the measured articulatory dependent variables; all effects with $p < .05$ are reported – first for phrase condition and then for stress condition.

A comparison of the phrase-internal versus phrase-boundary conditions indicates that the phrase-boundary condition had longer acoustic and articulatory duration for all subjects (acoustic duration: Speaker J $F(1, 108) = 34.074$, $p < .0001$, Speaker H $F(1, 106) = 336.324$, $p < .0001$, Speaker K $F(1, 96) = 59.089$, $p < .0001$; articulatory duration: Speaker J $F(1, 112) = 34.074$, $p < .0001$, Speaker H $F(1, 116) = 55.305$, $p < .0001$, Speaker K $F(1, 103) = 45.03$, $p < .0001$). This difference also obtained for the acceleration phase of the constriction formation (Speaker J $F(1, 112) = 33.221$, $p < .0001$, Speaker H $F(1, 116) = 48.082$, $p < .0001$, Speaker K $F(1, 103) = 31.235$, $p < .0001$), and likewise for the proportional time-to-peak velocity. There were no crossover interactions, though two speakers (H and K) had an interaction effect indicating that the acoustic duration increase for the boundary was greater in the level-stress than in the falling-stress condition. The stiffness change (indicated by the acceleration duration) and consequent acoustic lengthening are typical of phrase-final articulations. Importantly, these longer durations are clearly incompatible with the short acoustic intervals that characterize flaps (see Parker & Walsh 1982), providing a quantitative foundation for why flaps uniformly do not appear in this context.

The phrase-boundary condition was also characterized by higher y-extremum values and larger Euclidean constriction displacements (y-extremum: Speaker J $F(1, 112) = 15.613$, $p < .0001$, Speaker H $F(1, 116) = 4.935$, $p = .0282$), Speaker K n.s.; Euclidean displacement: Speaker J $F(1, 112) = 6.969$, $p = .0095$, Speaker H $F(1, 116) = 5.734$, $p = .0182$, Speaker K $F(1, 103) = 4.415$, $p = .0381$). Other spatial effects of boundary, same in kind, include bigger y and x displacements for all speakers. Finally, there were no effects of boundary on peak velocity.

A comparison of the falling-stress versus level-stress contours, i.e. a main effect of stress condition, indicates that the falling-stress condition had shorter acoustic durations (Speaker J $F(1, 108) = 8.327$, $p = .0047$, Speaker H $F(1, 106) = 9.258$, $p = .003$, Speaker K $F(1, 96) = 59.089$, $p < .0001$). However, there was no difference in articulatory duration or acceleration duration EXCEPT that Speaker J had a shorter articulatory acceleration phase ($F(1, 112) = 6.089$, $p = .0151$) and total articulatory duration with falling-stress ($F(1, 112) = 6.636$, $p = .0113$) than level-stress. Only Speaker K shows an effect on peak velocity – higher peak velocity with falling stress, and only Speaker J shows an effect on proportional time-to-peak velocity. In the spatial domain, the falling-stress condition for Speakers J and H had a more retracted x-extremum (Speaker J $F(1, 112) = 8.723$, $p = .0038$, Speaker H $F(1, 116) = 18.872$, $p < .0001$) and a bigger x-displacement (Speaker H $F(1, 116) = 16.063$, $p < .0001$). Speaker K had a higher y-extremum ($F(1, 103) = 4.83$, $p = .0302$) in the falling-stress condition. There are no crossover interactions.

In summary, the data indicate that a phrase boundary gave rise to a stiffness lowering and consequent lengthening of the final tongue-tip gesture and to larger tongue-tip movements. Additionally, the falling-stress condition had shorter acoustic durations but varied in kinematic effects across speakers.

3.4 Sources of shortening

An objective of the analysis was to evaluate potential sources of the shortness that is known to characterize flaps in the acoustic domain and for which we sometimes see evidence in the articulatory domain (see figure 1 above). In the 'Introduction' section, we sketched out three possible sources – changes in the following domains: spatial lowering (either in peak or initial position), articulatory shortening due to a stiffness change, and/or articulatory truncation, which could have both spatial and temporal signatures. We emphasized that these may co-occur and mutually contribute to shortening. Each of these possibilities has particular kinematic observables. The spatial change shown in figure 1a giving rise to shorter acoustic constriction intervals is an overall smaller vertical displacement or (because the palate is sloped rather than perpendicular to the tongue-tip movement) a more retracted movement. This type of change would (by itself) yield no difference in articulatory duration (interval from movement onset to peak) but would exhibit a lower or more retracted extremum position AND a y (or x) displacement change. The possibility in figure 1b for acoustic shortening is the situation in which the extremum position changes due not to a displacement change but to a lower initial position of the articulator (likely due to a more open mouth at movement onset). This type of change would (by itself) exhibit no articulatory duration or displacement change. It is worth noting that the situation in figure 1b, in the absence of other modifications (e.g. in overlap) is, in general, incompatible with a standard Task Dynamics account (Saltzman & Munhall 1989) of articulatory control. Next, the situation in figure 1c shows articulatory shortening as the source of the acoustic shortening. This articulatory shortening is caused by a stiffness change (i.e. a change in the gesture's internal temporal properties) that would be indicated by a shorter acceleration duration. Finally, the last source of acoustic shortening shown – truncation (figure 1d) – could exhibit minimal to large differences in articulatory duration, extremum position, and/or displacement, depending on the degree of truncation. However, it is more helpful to consider that while truncation (alone) would yield no change in acceleration duration, it WOULD be indicated by a change in the proportional time-to-peak-velocity. These kinematic observables will be compared below for the flap and non-flap realizations in the tokens without phrase boundaries to evaluate possible sources of acoustic shortening.

The two data subsets in which variable realizations of the word-final /t/ are found WITH STRESS AND PHRASING HELD CONSTANT (see table 2 above) are (i) Speaker J phrase-internal level-stress and (ii) Speaker K phrase-internal falling-stress. These two datasets are used below to evaluate whether the voiced [ɾ] realizations of word-final /t/ differ substantially in their articulation from its voiceless [t] realizations. This allows us to investigate how the flaps and non-flaps differ without confounds of phrasing or stress patterning.

A third potential dataset cannot be statistically evaluated due to the small number of non-flaps in one cell – Speaker J's falling-stress realization, which included three non-flap tokens and 27 flap tokens. However, since these two groupings had a large difference in acoustic and articulatory duration, we will make one qualitative observation regarding an articulatory measurement (acceleration duration) that differs grossly between these flaps and non-flaps, since this could potentially be evaluated more thoroughly in follow-up studies.

3.4.1 The spatial account

Recall that two potential spatial changes could result in gross shortening of the acoustic constriction interval: a lower or more retracted extremum position(s) as a result of smaller movement displacement or a lower initial position with no overall displacement difference. ANOVA indicates that for Speaker J (phrase-internal level-stress), the y-displacement and the Euclidean distance measure were larger for flaps than for non-flaps ($p = .0005$); however, the y-extremum position was not significantly different. This suggests that it is the initial position in y that differs for the two, and indeed an ANOVA on onset y-position confirms that the tongue starts lower (6 mm) in the mouth for the flap tokens ($F(1,27) = 11.562$, $p = .0021$). For Speaker K (phrase-internal falling-stress), the flap tokens have lower y-extrema ($F(1,19) = 7.313$, $p = .0141$) (and a tendency for extrema to be more retracted [$F(1,19) = 3.78, p = .0668$ n.s.]). Displacement in the x component is also much smaller for the flaps ($F(1,19) = 16.673, p = .0006$), a difference of about 2.5 mm.

3.4.2 The temporal articulatory shortening account

There is the possibility that flaps are short stops because they differ in their internal articulatory temporal characteristics, i.e. that they are fast. In the absence of any amplitude changes, inherently shorter gestures would be indicated by higher peak velocities. In this case, peak velocities should be higher for [ɾ]s, thereby yielding faster, shorter movements. CRUCIALLY HOWEVER, since concomitant spatial differences might also exist, we take flap and non-flap acceleration duration (i.e. time-to-peak velocity) to be the best indicator of gestural stiffness.

The flaps do not uniformly exhibit faster velocity than the non-flaps. The ANOVA results indicate no effect on peak velocity or acceleration duration for Speaker K. Speaker J (level-stress) has large peak velocity differences such that the flaps (voiced [ɾ]) have higher peak velocities than the non-flaps (voiceless [t]) ($F(1, 27) = 26.038$, $p < .0001$). HOWEVER, no difference in the acceleration duration is found, suggesting that the velocity differences were due to the lower onset gestural position observed for these tokens, which consequently required a larger movement and higher peak velocity, and not due to a difference in gestural stiffness. (Note that there is no difference in the displacement/peak velocity ratio for flaps versus non-flaps either in this dataset.) As a note of further interest, for Speaker J's falling-stress tokens (recall n = 3 non-flaps and n = 22 flaps), we qualitatively observe a large difference in acceleration duration to accompany the shorter articulatory duration of the flaps; they have a 50% shorter mean acceleration duration than the non-flaps.

3.4.3 The truncation account

The truncation account suggests that flaps result from overlap with the following vowel, which could consequently cause truncation of the alveolar articulation. Falling-stress contexts have been reported to have increased gestural overlap (see, for example, de Jong, Beckman & Edwards 1993, de Jong 1998). A kinematic measure that helps identify this type of truncation is the proportional time-to-peak-velocity (i.e. the ratio of acceleration duration to total gesture duration). An increase in this value would indicate truncation because total duration would decrease disproportionately to the acceleration phase. The overlap account for flaps makes this prediction. Single-factor ANOVA testing for a difference in proportional time-to-peak-velocity for flaps versus non-flaps in the two relevant data subsets – Speaker J level-stress and Speaker K falling-stress – shows no significant differences.

3.5 Summary of results

An attempt to discern SYSTEMATIC AND UNIFORM kinematic differences between the stop and flap realizations is not particularly successful. Speakers and stress conditions appear to have different forces at work yielding the short acoustic closure intervals of the flaps. For Speaker K, the shortening seems most in accordance with a spatial reduction account since her flaps exhibit lower and more retracted extremum positions, paralleling results in de Jong (1998),but with no distinction from non-flaps in the articulatory or acceleration durations of the movement. This is the scenario sketched in figure 1a. For Speaker J in the level-stress condition, flaps have higher peak velocities, but this seems to be at least partially a consequence of their lower initial position in the mouth rather than an intrinsic change in gestural stiffness or reduction of the spatial end-position of the movement. Significantly, ARTICULATORY DURATIONS of the flaps versus non-flaps in these data subsets are NOT significantly different – as described above, differences in acoustic intervals are not mirrored directly in articulation. Thus, Speaker J's flaps in the level-stress condition appear to be distinguished, like Speaker K's, in the spatial domain, but by a more open posture of the mouth at the onset of the gesture and little change in the gestural specification, other than a consequent higher peak velocity. Finally, though by no means conclusive, for Speaker J in the falling-stress condition, we can speculate that the far shorter articulatory and acoustic durations may come from an articulatory foundation consistent with a stiffness change, as indicated by large differences in the acceleration duration measure. This is in line with figure 1c.

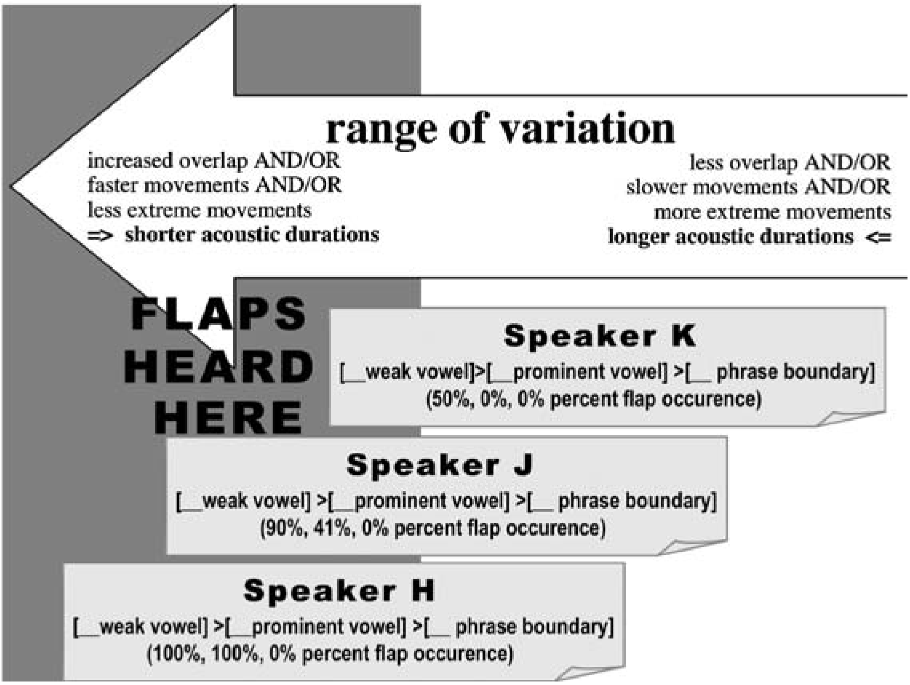

Figure 4 Schematic showing consistency in propensity for a flap to be heard as a function of articulatory patterning.

Overall, across subjects and stress conditions, we see evidence for changes consistent with the types shown in figure 1a and perhaps 1c. While a lower initial posture of the tongue in the mouth did occur for flaps for one speaker in one condition (Speaker J, level-stress), a difference in y-extremum position did not reach significance (flaps lower by 1.3 mm, $F = 2$, $p = .17$ n.s.); thus, this is not entirely consistent with the mechanism shown in figure 1b. Thus, overall, the kinematic data indicate that different sources of flap shortness exist for different speakers and for different prosodic conditions, with some propensity for spatial differences, either at movement initiation or ending, to differentiate flaps from non-flaps.

4 Discussion

We can see from figure 4 that the general ordering of the propensity for flaps to occur is shared among subjects, but because of individual differences (perhaps, for example, in speaking rate), the situations in which sufficient forces for shortening (or forces for sufficient shortening) exist differ among the subjects – Speaker H is most prone to produce short voiced stops for word-final /t/, Speaker K least, and Speaker J intermediately prone to do so.

The occurrence of word-final voiced flaps is highly frequent for some subjects, rare or variable for others. Further, if flaps resulted from the symbolic substitution of one consonant ([ɾ]) for another ([t]), we could expect those two consonants to be qualitatively and systematically different – for example, in their constriction location, constriction degree, or temporal properties. Such systematic qualitative differences were not found in this data, except in the sense that the flaps are shorter occlusions. Subjects and prosodic conditions clearly differed in the articulatory basis for these shorter occlusions. Gradient variability, in both the spatial and temporal domains, of tongue-tip constrictions can yield these acoustically short consonants. While this durational variation is salient (at least to linguists) and transcribed, the further conclusion, common to most introductory textbooks, that it is the result of symbolic substitution of one consonant for another (more rapid) consonant, may be overstated, at least for the word-boundary case.

A point of speculation may be relevant regarding word-internal flaps. Word-internal flaps are generally limited to occurring before stressless syllables (e.g. a flap can't occur in 'Utah' due to the presence of secondary stress on the second syllable). Recall that for the subject who had an overall stress effect, it was the falling-stress environment in which significant ARTICULATORY shortening was observed, suggesting that the forces creating short consonant intervals could be especially strong in this circumstance. Further, alveolars appear particularly susceptible to these effects (Browman & Goldstein 1992, Turk 1992), perhaps due to the biomechanical properties of the small tongue tip. It may be that the articulatory patterning that yields flaps gradiently in word-final contexts is phonologized for word-internal contexts.

5 Conclusion

The results of this articulatory study suggest that flaps do not differ categorically or uniformly (i.e. for all subjects) from non-flaps in the articulatory domain, despite their systematic properties – especially shortness – in the acoustic domain. These findings are consistent with de Jong's (1998) finding that 'kinematic measures generally do not exhibit quantization according to flap and [d] categories' (p. 309). In agreement with Turk (1992), certain linguistic environments seem likely to yield articulatory patterning conducive to very short acoustic consonant intervals. In the case of tongue-tip stops, these are recorded as flaps (Turk 1992). This implies that gradient variability in the spatiotemporal patterning of tongue-tip constrictions can yield these acoustically short consonants. Such variability could include speaker-specific combinations of articulatory posture and temporal organization within and between gestures.

Acknowledgements

This work was supported by NIH grant DC03172 and the USC Zumberge Fund. The authors thank an anonymous reviewer, Ken de Jong, Peter Ladefoged, Sungbok Lee, Vince Gracco, Elliot Saltzman, and Haskins Laboratories.

References

BANNER-INOUYE, S. (1995). *Trills, Taps and Stops in Contrast and Variation*. Ph.D. dissertation, UCLA.

BROWMAN, C. P. & GOLDSTEIN, L. (1992). Articulatory phonology: an overview. *Phonetica* **49**, 155–180.

BYRD, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics* **83**, 97–116.

COOPER, W. E. & PACCIA-COOPER, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press.

DE JONG, K. (1998). Stress-related variation in the articulation of coda alveolar stops: flapping revisted. *Journal of Phonetics* **26**, 283–310.

DE JONG, K., BECKMAN, M. E. & EDWARDS, J. R. (1993). The interplay between prosodic structure and hyperarticulation. *Journal of the Acoustical Society of America* **97**, 491–504.

GUSSENHOVEN, C. (1986). English plosive allophones and ambisyllabicity. *Gramma* **10.2**, 119–141. Instituut Nederlands, Univerity of Nijmegen.

HAYES, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

KAHN, D. (1976). *Syllable-based Generalizations in English Phonology*. Ph.D. dissertation, Indiana University, Bloomington.

KIPARSKY, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry* **10**, 421–441.

PARKER, F. & WALSH, T. (1982). Blocking alveolar flapping: a linguistic analysis. *Journal of Phonetics* **10**, 301–314.

PATTERSON, D. & CONNINE, C. M. (2001). Variant frequency in flap production. *Phonetica* **58**, 254–275.

PERKELL, J. S., COHEN, M., SVIRSKY, M., MATTHIES, M., GARABIETA, I. & JACKSON, M. (1992). Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movement. *Journal of the Acoustical Society of America* **92**, 3078–3096.

RIEHL, A. (2003) American English flapping: evidence against paradigm uniformity with phonetic features. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2753–2756.

SALTZMAN, E. L. & MUNHALL, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* **1**, 333–382.

SELKIRK, E. (1982). The syllable. In Van der Hulst, H. & Smith, N. (eds.), *The Structure of Phonological Representations* (*Part 2*), 337–383. Dordrecht: Foris Publications.

STATHOPOULOS, E. & WEISMER, G. (1983). Closure duration of stop consonants. *Journal of Phonetics* **11**, 395–400.

STERIADE, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. In Broe, M. B. & Pierrehumbert, J. B. (eds.), *Papers in Laboratory Phonology* V: *Acquisition and the Lexicon*, 313–334. Cambridge: Cambridge University Press.

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.

TIEDE, M. K., VATIKIOTIS-BATESON, E., HOOLE, P. & YEHIA, H. (1999). Magnetometer data acquisition and analysis software for speech production research. *ATR Technical Report TR-H 1999*. Kyoto, Japan: ATR Human Information Processing Labs.

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.

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

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