Working Papers of the Cornell Phonetics Laboratory
2003, vol. 15, pp. 271-337

American English Flapping: Perceptual and Acoustic Evidence Against Paradigm Uniformity with Phonetic Features*

Anastasia K. Riehl

This study investigates the claim that flapping patterns in American English are subject to phonetic paradigm uniformity constraints based on the phonetic feature [extra short closure], as proposed in Steriade (2000). The results of this study reveal that speakers do not maintain uniform paradigms with regard to flapping and that [extra short closure] is not an invariant acoustic cue for flap identification and therefore a questionable candidate for a phonetic uniformity constraint in the first place. American English flapping patterns therefore do not support a collapse of the phonetic and phonological components of grammar, as argued in Steriade (2000).

1. Introduction

This study investigates the claim that flapping patterns in American English are subject to phonetic paradigm uniformity constraints based on the phonetic feature [extra short closure], as proposed in Steriade (1996, 2000).[1] The claim is that a stop in a base form will be maintained in a corresponding inflected form, even if the phonological patterning predicts a flap in that environment. I investigate this claim by subjecting the preliminary study in Steriade (2000) to a larger, more controlled perceptual and acoustic study. The results of this investigation reveal that: a) flap/stop alternations cannot be explained by appealing to paradigm uniformity and, b) [extra short closure] is not an invariant cue to flap identity and therefore a questionable candidate for a phonetic uniformity condition in the first place. These results, therefore, do not lend support for a collapse of the phonetic and phonological components of the grammar, as argued in Steriade (2000).

The structure of this paper is as follows. Section 2 contains background information on phonological and phonetic accounts of flapping. Section 3 presents the

---

* An earlier version of this paper was presented as a poster at the 15th International Congress of Phonetic Sciences and published as Riehl (2003). I thank the attendees of this conference for their comments and suggestions. I would also like to extend my appreciation to Abby Cohn, Amanda Miller-Ockhuizen, and Draga Zec for valuable feedback on the various drafts of this work, also to Janet Pierrehumbert and Ayako Tsuchida for helpful discussions of the topic, and to Eric Evans for technical support.

[1] The 1996 reference refers to the version of the paper presented at the 5th Conference on Laboratory Phonology, while the 2000 reference refers to the version published in *Papers in Laboratory Phonology 5*. The two papers are essentially the same, though the earlier version contains a small amount of additional information. I refer to the 2000 paper throughout as it is the published version, though I will make reference on two occasions to the 1996 paper where it contains data not included in the later version.

concept of paradigm uniformity as applied to phonetic attributes and describes the flapping study in Steriade (2000). Section 4 provides information on the methodology and Section 5 the results. The discussion is in Section 6, and Section 7 has conclusions.

2. The phonology and phonetics of flapping

2.1 Phonological models of flapping

Flapping in American English has traditionally been understood as a phonological rule whereby intervocalic /t/ or /d/ becomes a flap before an unstressed vowel, as in the word *a**t**om*. Kahn (1980) describes the environment for flapping as one in which an alveolar stop becomes a flap following a [-consonantal] segment and preceding an unstressed syllable (whether a vowel or syllabic consonant). By [-consonantal] segment, he means to include a glide or /r/ which might follow a vowel, as in *par**t**y*. Some speakers may also flap in the environment following a nasal, but only when the preceding vowel nasalizes and the consonant is deleted, therefore preserving the aforementioned characterization. As Kahn (1980) points out, the requirement that the following vowel be unstressed is crucial to the environment, whereas the presence or absence of stress on the preceding vowel, sometimes said to play a role, is irrelevant, since flapping can occur following a stressed vowel, as in *u**tt**er* or an unstressed vowel, as in *obesi**t**y* (see also Hayes 1995). A purely descriptive formulation of the rule may look as follows:

/t/, /d/ → /ɾ/ / [-cons] _____ [+syllabic, -stress]

Figure 1: General description of flapping in American English

The rule also applies across word boundaries, as in the phrase *sit up*, and in such cases the [-stress] requirement may be relaxed. The present discussion concerns only the word-internal environment.

Important to the description of flapping is the observation that the rule applies optionally and has a greater tendency to occur in fast or casual speech. Whether or not the phenomenon is truly optional or is simply not yet fully understood is an issue to be addressed in the following subsections.

Various phonological accounts have attempted to identify more precisely the relevant prosodic environment for flapping. Kahn (1980) and Gussenhoven (1986) analyze flaps as ambisyllabic, Kiparsky (1979) as non-foot-initial, and Selkirk (1982) as syllable-final. For a summary of these perspectives, see Turk (1992). These different approaches do not bear directly on the issues presented in this paper, since the discussion will focus largely on variable flap realization in words with the same prosodic structure.

Phonological analyses of flapping also differ in the feature they assign to flaps. Kahn (1980) proposes [+sonorant] while Selkirk (1982) argues for [+release]. Steriade (2000) uses the feature [+extra short closure], citing similar length-based features proposed by Williamson (1977) and Banner-Inouye (1995). [extra short closure] will be discussed in more detail in Section 3.1.

2.2 Phonetic correlates of flapping

Several phonetic studies have investigated the acoustic or articulatory characteristics of flaps. In a comprehensive study of medial /t/ and /d/ in American English, Zue and Lafferiere (1979) examined alveolar stops in six different environments. Two of these environments produce segments commonly referred to as flaps. The first environment, which they call the "flapped" environment, describes a context where the stop follows a stressed syllable and precedes an unstressed syllable, such as in the word *fla**tt**er*. In this environment, flaps are characterized by brief tongue tip contact and immediate release, resulting in very short duration (and average of 26 ms within a range of 10-40 ms) and no release burst. Flaps in this context can have a "variety of acoustic realizations" (Zue and Lafferiere 1979:1043), including turbulent noise due to partial closures. The second environment, which Zue and Lafferiere refer to as the "unstressed" environment, describes a context where the stop falls between two unstressed syllables, such as in the word *complici**t**y*. In this environment, flaps tend to be longer (40 ms average). This study also produced some interesting findings regarding the "optionality" of flapping. In the "flapped" environment, the probability of occurrence of /t/ surfacing as a flap was .99. In the "unstressed" environment, however, the probability of occurrence of a flap was only .33, while the probability of occurrence of an aspirated stop was .66. Although these two environments are often conflated in the phonological characterization of flapping, these data suggest that there is a potentially important distinction between them. It is the second environment, between unstressed vowels, that will be the focus of this paper.

In addition to the flapping characteristics described above, DeJong (1998) found that the perception of flaps is most closely related to voicing during the closure and a lack of a release burst. Fujimura (1986) noted jaw weakening in flap articulations. Despite general findings that short duration, voicing during the closure and the lack of a strong burst correlate with flap identity, it is important to note that a number of studies have found the acoustic and articulatory properties of flaps versus stops to be somewhat less clear. For example, Stone and Hamlet (1982) and DeJong (1998) found that a number of the tokens in their studies could not be easily categorized as either stops or flaps, based upon articulatory and acoustic properties. Additionally, Zue and Lafferiere (1979:1048) note that a number of their /d/ tokens have some characteristics of both flaps and stops and that "to make a distinction between a long flap and a short unstressed /d/ would be highly subjective and likely to lead to misinterpretations."

2.3 Phonetic accounts of flapping

The fact that the phonological flapping rule is said to apply optionally, and the fact that several studies have produced tokens not easily categorized as either flaps or stops, have lead to speculation that flapping may be a gradient process of lenition, not a categorical one. This is a perspective taken by Umeda (1977), Stone and Hamlet (1982) and Fujimura (1986). Along this line of inquiry, Turk (1992) finds that all oral constrictions, not just alveolars, shorten in the flapping environment. In an articulatory and acoustic study, DeJong (1998) investigates the effect of focus and higher level prosodic position on alveolar stop production and concludes that while a prosodic by-product account can accommodate a great deal of the data, neither this model nor the categorical phonological rule can alone account for his findings. See DeJong (1998) for a more comprehensive review of the phonetic studies.

3. Paradigm uniformity

3.1 Paradigm uniformity and phonetic analogy

The concept of paradigm uniformity, sometimes referred to as paradigm regularity or analogy, has long been a part of the phonological literature (see e.g. Kiparsky 1982, as well as McCarthy 2001 and references therein for recent approaches in an Optimality Theoretic framework). A paradigm is a group of words that share a morpheme (e.g. *think*, *think-ing*, *un-think-able*). The term *uniform paradigm* refers to the observation that morphemes tends to be invariant in form across the members of a paradigm, despite difference in phonological context. Such uniformity is more likely to occur in productive paradigms, where the relationship between a base and inflected form is clear. Steriade (2000:313) states the condition of paradigm uniformity as follows:

> All surface realizations of μ, where μ is the morpheme shared by the members of paradigm x, must have identical values for property P.

Traditionally, uniform paradigms have been discussed only with respect to the surface phonology, for example the presence of a syllabic /n/ in all members of the paradigm {*lighten, lightens, lightened, lightening*}(McCarthy 2001) and therefore only to phonological features. Steriade (2000), however, extends the idea of paradigm uniformity to the concept of "phonetic analogy." Her claim is that uniform paradigms operate on non-contrastive phonetic attributes as well,[2] such that an allophone will surface in unexpected positions simply to satisfy a phonetic paradigm uniformity condition.

Steriade supports her case by offering examples of paradigm leveling in American English flapping, French schwa deletion and British English closure voicing. (The third case is only included in the 1996 version of the paper.) She concludes that non-contrastive phonetic features are subject to paradigm uniformity conditions in these cases, and that these attributes should therefore not be regarded as distinct from phonological features. Steriade ultimately takes this argument as evidence that the phonetic and phonological components of the grammar should be collapsed. Therefore, a

[2] Steriade refers to "phonetic features", but in this paper the term "phonetic attribute" will be used.

great deal rests on the empirical foundation of these three cases, and it is well worth revisiting them in careful detail. This paper focuses on the case of American English flapping. Barnes and Kavitskaya (2002) offer a reconsideration of the French schwa case.

3.2 Flapping and phonetic analogy: Steriade (2000)

Steriade bases her study of American English flapping on observations by Withgott (1983). Withgott noted that the words *militaristic* and *capitalistic*, although they bear the same stress pattern, differ in their /t/ allophones: the first /t/ in *mili**t**aristic* is a stop whereas the first /t/ in *capi**t**alistic* is a flap. Withgott attributes the difference to the base forms: *mili**t**ary* contains a stressed syllable following the /t/ and therefore a stop and *capi**t**al* contains a stressless syllable following the /t/ and therefore a flap. It is claimed that the inflected forms, which arguably do not have stress on the syllable following the /t/, simply retain the allophones from the base forms. With her study of flapping, Steriade aims to show that this phenomenon is part of a more general pattern of paradigm uniformity with regard to flapping in American English and that this uniformity is due to a non-categorical phonetic attribute, [extra short closure].[3]

Steriade argues for [extra short closure] by citing work by Zue and Lafferiere (1979) and Banner-Inouye (1995) who demonstrate that flaps tend to have shorter duration than stops. She then argues for the non-contrastiveness of this feature by drawing on Banner-Inouye's (1995) extensive study of flap-related contrasts and concluding that it is not necessary for any language to expand the phonological feature set in order to accommodate flaps specifically. Additionally, Steriade argues that the fact that a length contrast does not need to be specified at other places of articulation, despite the fact that all oral constrictions shorten in the flapping environment (Steriade cites Browman and Goldstein 1992, see also the aforementioned phonetic study by Turk 1992), is further evidence of the non-contrastiveness of [extra short closure]. Steriade therefore argues that this cue is at least one of the attributes that distinguishes a flap from

---

[3] Although Steriade (2000) refers to [extra short *closure*], it is clear from the data she cites that she intends this term to refer to total duration, meaning the duration of the closure plus VOT. I therefore assume that [extra short closure] refers to total duration.

a stop, and that such a cue, never contrastive in the phonology, would be a candidate for phonetic paradigm leveling effects. Steriade does acknowledge that other acoustic attributes are likely also involved in the characterization of a flap. However, it is necessary for the paradigm uniformity claim that the attribute she selects, [extra short closure], be a consistent indicator of flap identity; otherwise, there is no basis for the argument that the uniformity is based on a phonetic cue. I therefore assume that Steriade intends [extra short closure] to be the invariant flapping attribute, in the sense of Stevens, Keyser and Kawasaki (1986), and other acoustic attributes that work in conjunct with [extra short closure] are understood by Steriade to be redundant cues.

In order to test her hypothesis, Steriade asked 12 subjects to read two lists of words. The first list contained 5 target words: *voluntary*, *positive, negative, primitive* and *relative.* The expectation was that some speakers would place secondary stress on the syllable following the /t/ and thereby produce a stop and that other speakers would not place secondary stress on this syllable and thereby produce a flap. These words were randomized with five other words where all speakers were expected to flap, due to the /t/ following a stressed syllable and preceding a stressless syllable. These words, *fatal, fetish, totem, notary* and *rotary,* were included to ensure that speakers would not artificially produce stops in all tokens. On a second list, all ten of these words were inflected with the productive suffix *–istic,* a suffix which attaches to an adjective to produce another adjective meaning "having the qualities of X." The outcome includes several nonce forms, such as *primitivistic*, which the speakers claimed to be comfortable with.

Steriade assumes that in the inflected forms, there should be no secondary stress following the /t/. She bases this assumption on the lack of secondary stress following the /t/ in monomorphemic V'CVtV strings such as *meritocratic*. Therefore, she claims that the phonological rules alone would dictate that the speakers pronounce such forms with a flap.[4] However, Steriade predicts that if a speaker pronounces the base form with a stop

[4] The assumption in Steriade (2000) is that a phonological flapping rule should apply when the /t/ falls between two unstressed syllables. However, as mentioned in Section 2.2., the probability of occurrence of a flap was only .33 in this environment Zue and Lafferiere's (1979) study; therefore, this assumption may be faulty.

(due to a following syllable bearing secondary stress), then the speaker will also pronounce the inflected form with a stop, due to paradigm uniformity effects, even in the absence of secondary stress on the following syllable. Likewise, a person who articulates a flap in the base would be expected to pronounce a flap in the inflected form.

The results of Steriade's study appear to support her claims of phonetic paradigm leveling, although there is a small amount of variation in her data that is not accounted for. Eleven out of twelve speakers have identical allophones for every pair of base and inflected form. One speaker contains a single pair that is not uniform. The following figure summarizes the results reported in Steriade (2000). The figure is an adaptation of Table 17 in Steriade (1996), although the data it contains is the same as that reported in Steriade (2000).

Speaker →	1, 2, 3, 4, 5, 6, 7	8, 9	10, 11	12
*volun**t**ary*	t	t	t	t
*volun**t**aristic*	t	t	t	t
*posi**t**ive*	t	t	D	t
*posi**t**ivistic*	t	t	D	t
*primi**t**ive*	t	D	D	D
*primi**t**ivistic*	t	D	D	t
*rela**t**ive*	t	D	D	D
*rela**t**ivistic*	t	D	D	D
*nega**t**ive*	t	D	D	t
*nega**t**ivistic*	t	D	D	t

Figure 2: Results of study in Steriade (2000) (adapted from Table 17 in Steriade 1996), /t/ allophones for 12 speakers for the 5 target word pairs.[5] D=flap

Steriade (2000) derives these results through the relative ranking of two constraints. First, she argues that a constraint assuring paradigm uniformity in the case of categorical stress-PU(stress), can be broken down into a number of more specific constraints based upon the individual phonetic correlates of stress, such as duration, pitch accent and vowel

[5] The chart in Steriade (1996) also includes the pair *fetish/fetishistic* since one speaker produced stops in this pair, and the author therefore claims an additional 12 pairs provide evidence of a paradigm uniformity effect. This pair is not included here, however, since the /t/ in this and the other flapped pairs directly follows a stressed syllable (Zue and Lafferier's "flapped" environment) where all speakers are expected to flap, and the pair is therefore not a good test case for paradigm uniformity. The one speaker who produced stops in this pair likely did so do to the nature of the task, as discussed later in Section 6.1.1.

quality (Steriade 2000:321). The constraint active in the case of American English flapping is PU(stress, duration), defined as follows:

> PU (stress: duration)
> If two strings, Σ and Σ', stand in correspondence and if Σ is a stressed syllable, then Σ and Σ' are durationally equivalent.

This constraint assures that the duration of a stressed syllable in a base form will be maintained in a corresponding inflected form. This constraint is outranked by a second constraint, Reduction, which states that stressless vowels must be schwa. Since the duration of the vowel is restricted by the Reduction constraint, according to Steriade, the PU(stress, duration) constraint can be satisfied only by maintaining uniformity in the duration of the /t/. The result is therefore a set of word pairs where a stop in a base is maintained in the inflected form even in the absence of secondary stress. Likewise a flap in a base form results in a flap in the inflected form.

3.2.1 Issues about stress re: assumptions in Steriade (2000)

The present study challenges the claims about the existence of a paradigm uniformity effect in flapping and the presence of the [extra short closure] feature, to be discussed in the following subsection. Beyond these issues, however, there are certain problems with the assumptions regarding stress made in Steriade (2000) that should be kept in mind when considering that study's results.

First, Steriade proposes a series of PU(stress) constraints, as presented above, that break stress down into individual phonetic correlates. The assumption that stress, commonly understood to be a relational phonological property, can be broken down in such a manner, where PU(stress, duration) and PU(stress, pitch accent), for example, are entirely divorced from one another, and where stress is considered solely in the realm of phonetics, is entirely unmotivated. Furthermore, by basing her study on this assumption, that stress should be considered in the phonetics rather than the phonology, the results in Steriade (2000) can do nothing other than support her ultimate conclusion- that the paradigm uniformity effect is a phonetic one.

Second, although the analysis focuses on the duration of the /t/ allophones, the PU(stress, duration) constraint actually refers to the duration of the entire syllable. The length of the vowel, however, is meant to be controlled by Reduction, which dictates that all unstressed vowels are schwa. The assumption made in Steriade must therefore be that since schwas are relatively short in duration when compared to other vowels, any additional duration required by the constraint will be borne by the /t/. However, the analysis only works if schwas in unstressed syllables are of a consistently short length, an assumption unlikely to hold true. If a short stop followed by a relatively long schwa in a base could have the same duration as a long flap followed by a relatively short schwa in an inflected form, then the PU(stress, duration) constraint would be satisfied without uniformity of the /t/ phone.

Third, since the paradigm uniformity constraint assures correspondence between the duration of stressed syllables, and the manifestation of this correspondence is the form of /t/ allophone, the analysis crucially relies upon every stop in a base form being followed by secondary stress. If a stop in a base form were *not* followed by secondary stress, then the constraint would be unable to enforce uniformity in duration of the /t/, since the constraint enforces a correspondence between stressed syllables only. It is not clear that this assumption would always hold true.

Fourth, Steriade claims that a stop in a base form is an indication that the following syllable bears secondary stress; however, the author also claims that a stop in the inflected form is not an indicator of secondary stress but rather of paradigm uniformity. This argumentation is somewhat circular. If a stop is a diagnostic for stress in the base, why shouldit not be a diagnostic for stress in the inflected form as well? If it could instead be argued that the inflected forms with stops do in fact bear secondary stress, the argument that paradigm uniformity is based on a phonetic attribute would be superfluous. It is interesting to note, in relation to the preceding criticism, that inflected forms with stops *cannot* bear secondary stress while base forms with stops *must* bear secondary stress in order for the analysis to work.

The above issues will not be further explored in this paper; rather, the results of the present study challenge the empirical bases of the experiment in Steriade (2000) and therefore the conclusions they are argued to support. However, it is important to note that the theoretical underpinnings of the claims in Steriade (2000) are themselves subject to question.

3.3 Rationale for the present study

On the surface, the results of Steriade's study do appear to lend support for phonetic analogy. However, before these results can be viewed as conclusive evidence for paradigm uniformity effects at the phonetic level, two critical issues left unanswered by her study need to be addressed. First, do speakers maintain uniform paradigms with regard to flapping across multiple repetitions of a base and inflected form? Second, is [extra short closure] an invariant cue to flap identity?

First of all, do speakers maintain uniform paradigms with regard to flapping across multiple repetitions of a base and inflected form? In order to examine whether the /t/ allophone in a base is maintained in the inflected form, the base allophone must be determined. Such a comparison necessarily assumes that a speaker will always articulate the same allophone in the base. Someone who articulates *primitive* with a stop should therefore always articulate it with a stop. However, what if a speaker varies in his/her pronunciation of the base word, pronouncing [prɪmɪtɪv] on some occasions and [prɪmɪɾɪv] on others? It is not clear what paradigm uniformity would predict in such cases.

To address the issue of whether or not speakers consistently use the same allophone in a base or in an inflected form, speakers in the present study were asked to read the lists of words not once but twelve times. As will be seen in Section 4, the results reveal that in the vast majority of cases, speakers vary their choice of allophone across repetitions of a single form. Therefore, it is not clear that there is even any basis for evaluating paradigm uniformity.

The second issue concerns the choice of [extra short closure] as the invariant attribute that accurately distinguishes flaps from stops. Although Steriade cites work by Zue and Laferriere (1979) who found that flaps tend to have shorter durations than aspirated stops, she does not subject the tokens in her study to an acoustic analysis and therefore is unable to verify whether the tokens she perceives as flaps are actually shorter in duration than the tokens she perceives as stops, or whether other cues relate to flap/stop identification as well as or better than duration.

In order to address the issue of the relationship between perception and duration, each token in the present study was subject to impressionistic perceptual judgments by multiple listeners as well as to acoustic analyses of duration, VOT, and closure voicing. The results of this study reveal that while duration does correlate with perception fairly closely, it does not correlate perfectly, and for some speakers it correlates less well, or at least no better than, VOT or closure voicing. The results of this study therefore support a view of flapping whereby several cues work together, possibly in a trading relationship, and where no cue is necessarily invariant, as argued more generally by DeJong (1995).

4. Methodology

4.1 Recordings

Six subjects were recorded for this experiment, three males (Sp1, Sp2 and Sp3) and three females (Sp4, Sp5 and Sp6), all between the ages of 26 and 32, and all native speakers of American English from the northeastern United States. The subjects were asked to read lists of base forms and lists of inflected forms, as illustrated in Figure 3 and described below.

	Set A: Base Forms	**Set B: Inflected forms**
Target words	voluntary relative primitive positive negative	voluntaristic relativistic primitivistic positivistic negativistic
Flap words	fatal fetish totem notary rotary	fatalistic fetishistic totemistic notaristic rotaristic
Stop words	atomic attentive protective retentive fantastic	atomic attentive protective retentive fantastic

Figure 3: Composition of word lists

Half of the lists contained base forms (Set A in Figure 3)- the five target base words from Steriade (2000), the five flap words from Steriade (2000), and five words in which subjects were expected to use a stop allophone (due to the /t/ preceding a stressed syllable). The stop words did not appear on the wordlist in Steriade (2000) but were added to maintain a balance between flap and stop forms, so that the speakers did not get into a pattern of flapping the target words simply due to influence from the other flap forms on the list. The other half of the lists contained inflected forms (Set B in Figure 3 above)- the five target inflected words from Steriade (2000), the five inflected flap words from Steriade (2000), and the same five stop words as Set A. Both sets were randomized 12 times, resulting in 24 lists, and filler words were added to the beginning and end of each, in an effort to offset effects of list intonation. The two sets of lists were then intermingled such that a Set A list was followed by a Set B list, and so forth.

Before the recording, the subjects, who were all naive as to the purpose of the experiment, were given an opportunity to look over the lists. All of the subjects claimed to be comfortable with the nonce forms. Digital recordings were then made in the soundproof booth in the Cornell Phonetics Laboratory using an Electrovoice RE20 microphone. Each recording session took place in six short stages. A subject was given four lists at a time. After reading the four lists, the subject had an opportunity to take a break if desired. Most breaks were approximately two minutes long, with none longer than five minutes. The total time of each speaker's participation in the study was no more than 25 minutes.

4.2 Perceptual test

Four listeners completed a perceptual test to identify occurrences of stops and flaps in the data. The listeners were all graduate students at Cornell University with a background in phonetics and all native speakers of American English, one of these subjects being the author. The target tokens were divided into six sets, each set containing the repetitions of a single speaker. The sequence of tokens in each set was randomized, and a Perl script was created to run on each set, whereby the program would play each sound file twice and then, after prompting from the listener via pressing a key on the keyboard, the next token would be repeated twice, and so forth. The listeners heard the tokens through headphones, and on a sheet of paper they checked a box indicating whether they heard a stop or flap, in a forced choice experimental paradigm. After each set, the subjects were asked to take a break before beginning the next set (breaks ranged from ten minutes to 20 hours), with all sets completed within a twenty-four hour period. Before beginning each set, the subjects were given a tutorial which included a practice test of four to six tokens from the relevant speaker.

4.3 Acoustic analysis

The recordings were sampled at a rate of 22050 Hertz. The files were then labeled and analyzed in ESPS, Xwaves. The labels were assigned as follows. The beginning of the closure of the target /t/ was placed at the point where both the second formant (F2) and the third formant (F3) of the preceding vowel ended, or, in the cases where the formants remained throughout the closure, the label was placed at the point where the intensity of F2 and F3 decreased. The end of the closure was placed at the onset of a burst, or in cases where no burst was present, at the point where F2 and F3 of the following vowel began or increased in intensity. Fundamental Frequency (F0) was then calculated using the ESPS utility *getf0*, associated with Xwaves, at a step size of one millisecond.

A Perl script was used to place a VOT label at the first point (of at least five consecutive points, to avoid random spikes in voicing) where the probability of voicing was 1, starting at the closure label. The script calculated the duration of the closure, the VOT and the total duration (closure + VOT). The purpose of calculating total duration from a combination of closure duration and VOT, rather than from two hand-labeled points at the beginning and end of the segment, was to provide a more objective measure.

The following spectrograms and pitch tracks (which include probability of voicing) illustrate two different tokens with the hand-labeled "beginning closure" (bc) and "end closure" (ec) labels and the script-labeled "vot" label. The first set illustrates the word *relative* articulated by Sp3, an example of a canonical stop. The second set illustrates the word *negative* articulated by Sp6, an example of a canonical flap.

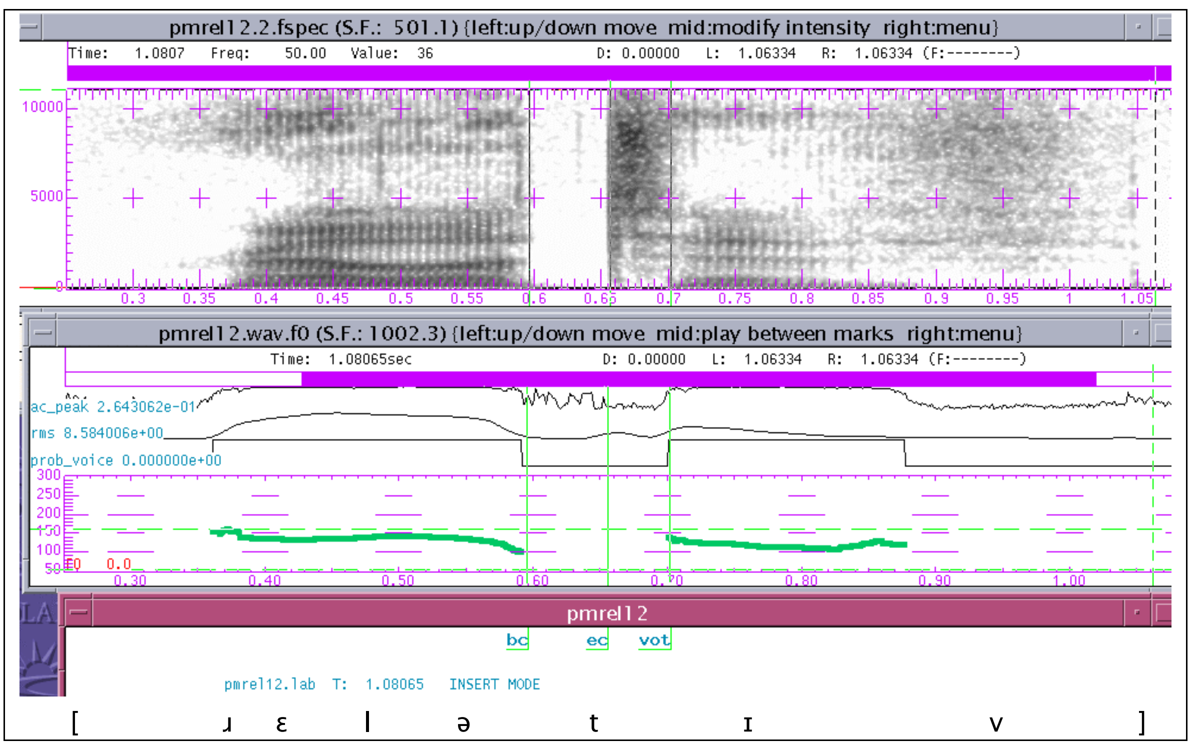

Figure 4: Spectrogram and pitch track of repetition 12 of *relative* by Sp3 illustrating a canonical stop.

In the above token of *relative* in Figure 4, the total duration is calculated as the duration of the closure ("bc" to "ec"), (58 milliseconds), plus the duration of VOT ("ec" to "vot"), (49 milliseconds). The total duration is therefore 107 milliseconds.

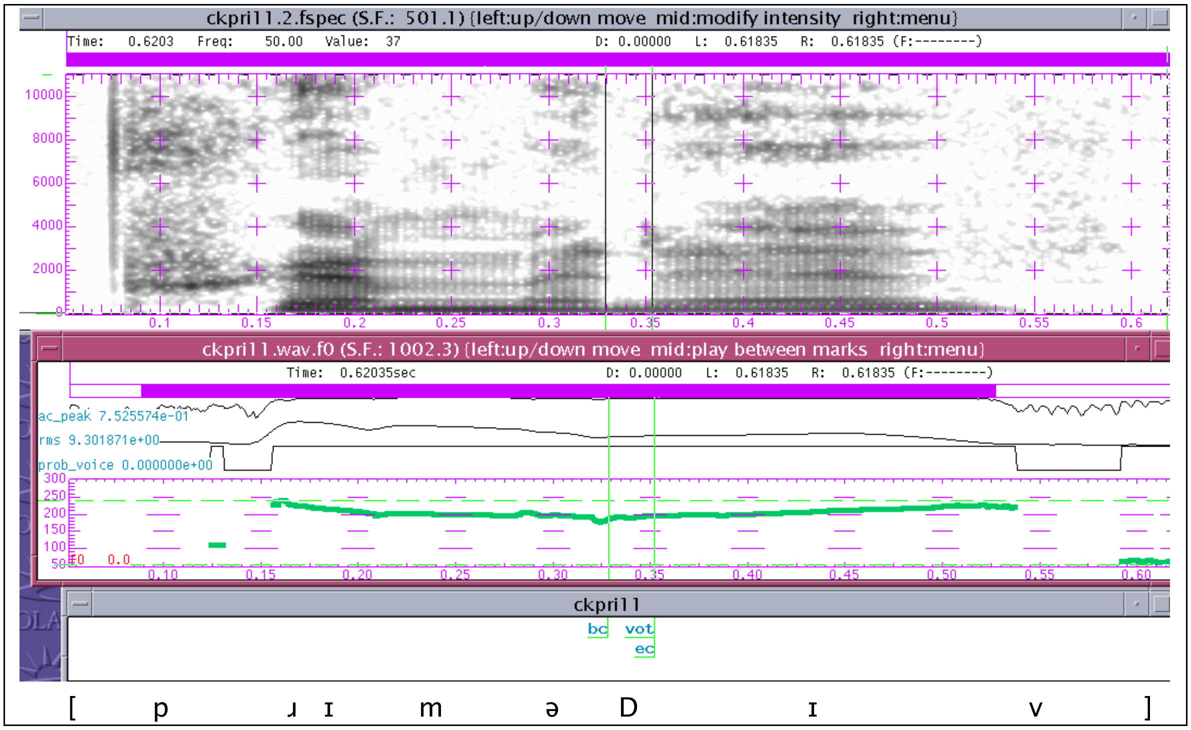

Figure 5: Spectrogram and pitch track of repetition 11 of *primitive* by Sp6 illustrating a canonical flap.

In Figure 5 above, the total duration of the /t/ allophone in *primitive* is calculated as the duration of the closure ("bc" to "ec"), (30 milliseconds), plus the duration of VOT ("ec" to "vot"), (0 milliseconds). The total duration is therefore 30 milliseconds.

As previously explained, the purpose of calculating the total duration from a combination of closure duration and VOT was to offer an objective measure. While the duration values calculated by the script seem appropriate for the vast majority of tokens, there is a problematic case involving some partially voiceless vowels for two speakers, Sp1 and Sp4. With most words, whether a token has a positive VOT, as in Figure 4 above, or a VOT of zero, as in Figure 5 above, voicing generally begins at the onset of formants. In fact, in forms such as Figure 5 above where no burst is present, it is the onset of formants that determines where the end closure (ec) label will be placed. In the case of the partially voiceless vowels, however, the VOT does not begin until mid-way through the vowel, and the script therefore reports a total duration that is likely in excess of the true duration of the segment. An example, a *positive* token by Sp1, is illustrated in Figure 6 below:

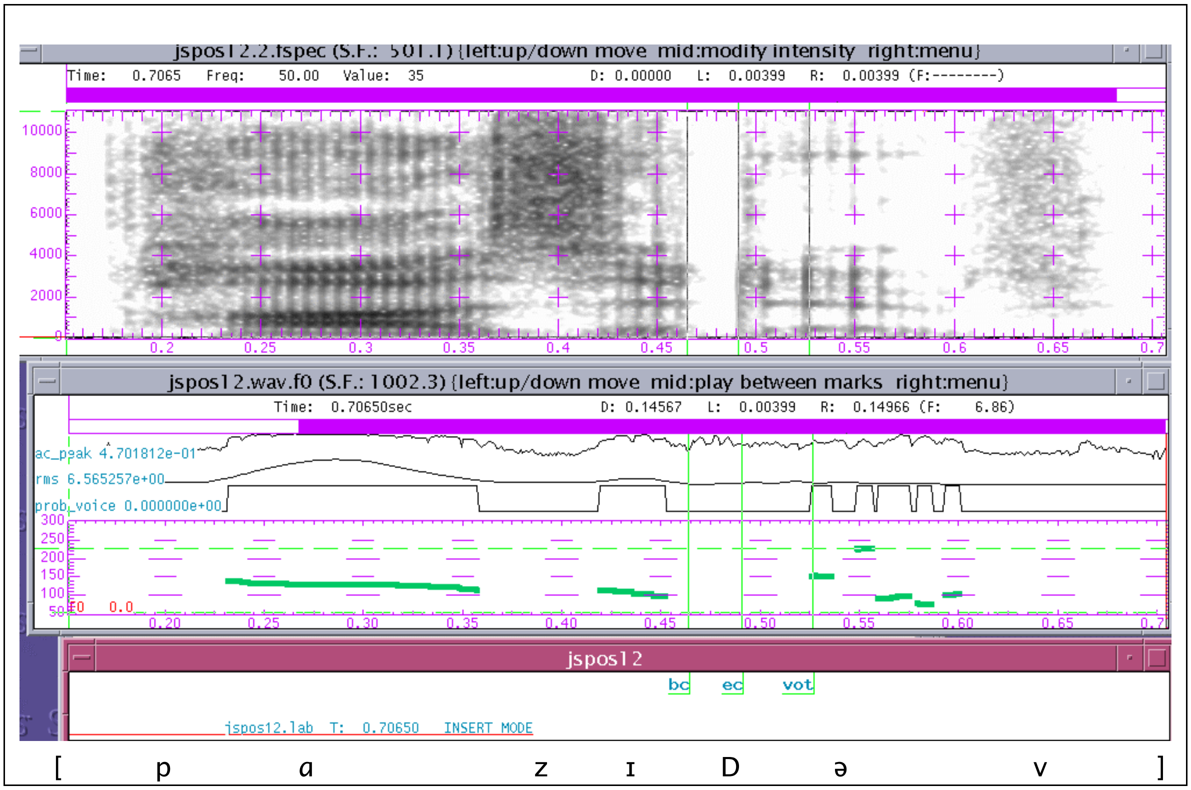

Figure 6: Spectrogram and pitch track of repetition 12 of *positive* by Sp1 illustrating a partially voiceless vowel following the /t/.

If total duration in this token is measured as closure duration plus VOT, than the result is a segment of 60 milliseconds. However, it seems clear from the spectrogram that the phone is not actually this long. If the total duration were instead calculated from the beginning of the closure to the onset of formants (reflected in the labeled figures above as “ec”), the duration would be only 21 milliseconds long. Such a large discrepancy in the duration of a single segment is troubling as it could affect the generalizations about durations of the /t/ allophones and have an impact on evaluating Steriade’s claim that [extra short closure] is the defining characteristic of a flap. In these cases of tokens with partially voiceless vowels, which constitute less than 4% of the target words, both duration figures (duration of “bc” to “vot” and duration of “bc” to “ec”) will be reported.

An additional Perl script was used to calculate the percent of voicing in the closure as well as the second half of the closure for each token. The results for the second half of the closure will be those reported in this paper. The reason for calculating the value in the second half of the closure is to more accurately capture the voicing of the closure as opposed to the voicing of the preceding vowel. The following spectrogram and pitch track of *negativistic* by Sp5 in Figure 7 illustrates a case where voicing from the preceding vowel carries over into the first half of the closure (if the end of the vowel is marked as the offset of formants as was done here):

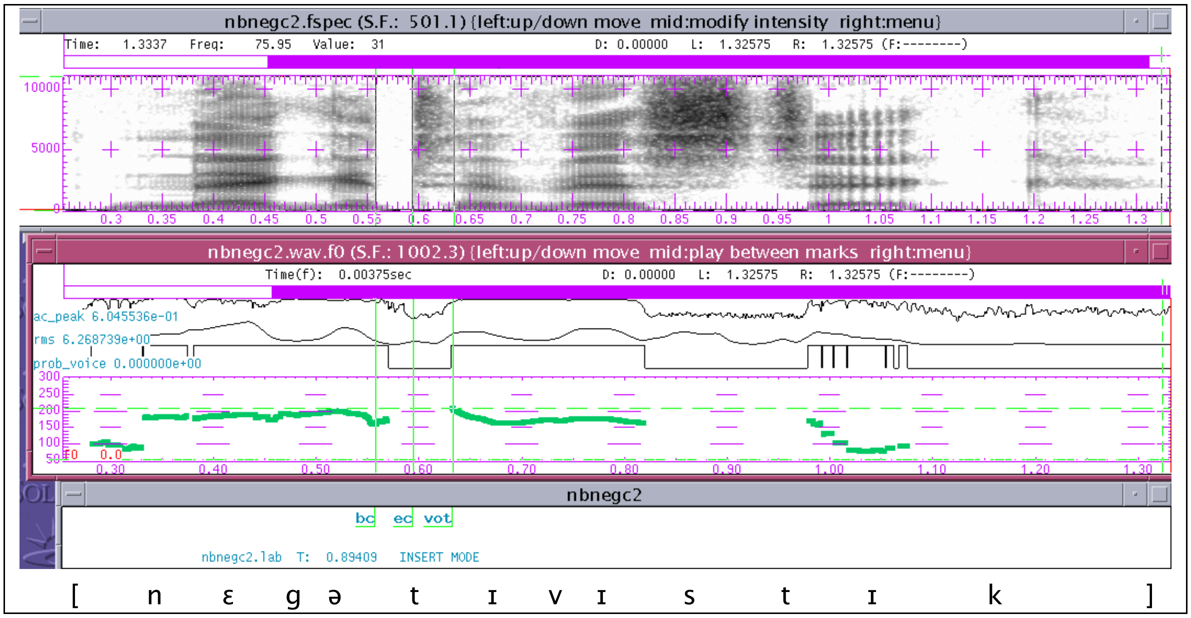

Figure 7: Spectrogram and pitch track of repetition two of *negativistic* by Sp5, showing voicing from the preceding vowel in the first half of the closure.

When closure voicing for the above token is calculated for the entire closure, 22% of the closure is voiced, whereas when it is calculated for the second half only, 0% of the interval is voiced.

In some cases, the closure duration of a speaker's /t/ is 0 ms due to an immediate burst, as in the following spectrogram of *positivistic* by Sp2 in Figure 8:

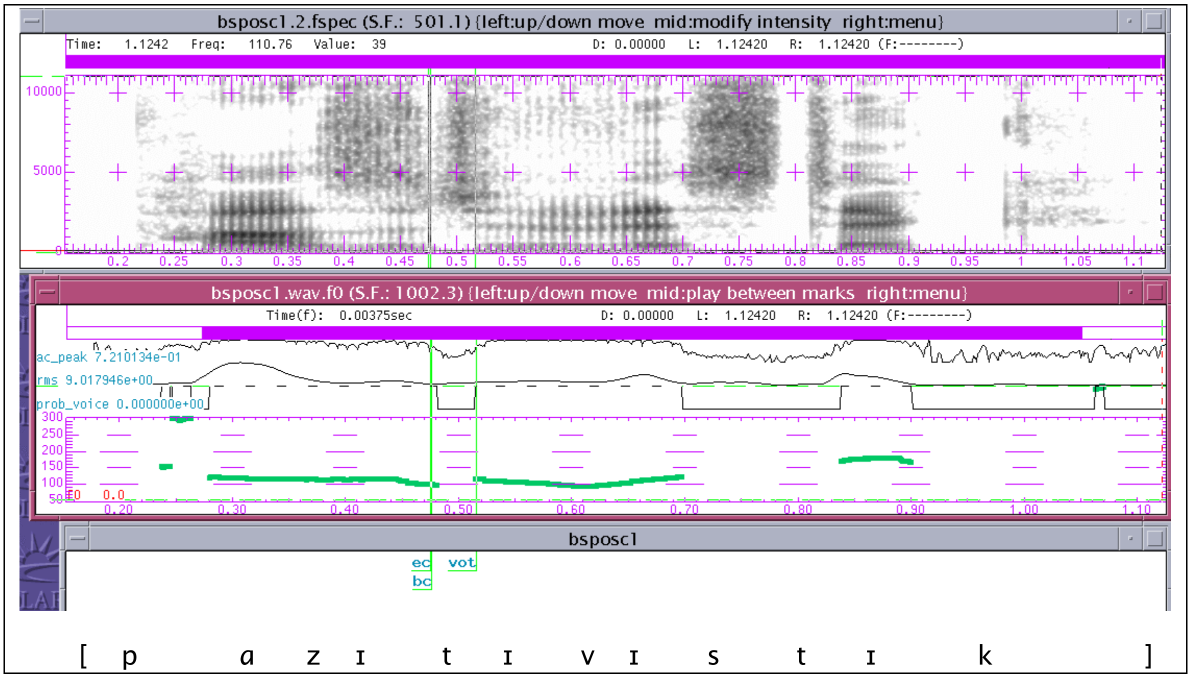

Figure 8: Spectrogram and pitch track of repetition one of *positivistic* by Sp2, showing a 0 ms closure.

In cases such as the above, closure voicing was not calculated, and the tokens are excluded from the results reported in Section 4.

The next section reports the results of these acoustic analyses.

5. Results

This section contains the results of the perception experiment as they relate to paradigm uniformity (5.1) followed by the results of the acoustic studies of duration (5.2), VOT (5.3) and closure voicing (5.4). In each case, the results for all of the target forms for each speaker are presented. The results for the canonical flap forms (*fatal, fetish*, etc.) and the canonical stop forms (*atomic*, *attentive*, etc.) are not included in this paper as they are peripheral to the present discussion. However, it is important to point out that the speakers performed as expected when reading these words: stops were perceived in 100% of the canonical stop forms and flaps were perceived in 98% of the canonical flap forms. (The 2% of the flap forms perceived as stops will be considered in Section 6.1.1.) Additionally, the pair *voluntary/voluntaristic* has been left out of this discussion, since all six of the speakers in this study were unable to articulate a flap in the environment following a /n/. (It is also worth noting that all 12 of Steriade's subjects pronounced these words with a stop as well.) With four target base forms and four target inflected forms repeated 12 times each by each of six subjects, less five missing tokens due to speaker error, the total number of tokens to be discussed in this study is 571.

Of these 571 tokens, the perception subjects were in total agreement on categorization of the /t/ allophone in 536 cases, leaving disagreement in only 35 tokens, or 6% of the total. Of these 35 tokens, three out of the four perception subjects agreed on the categorization in 30 of the cases, leaving a split categorization (two listeners identifying a segment as a flap and two identifying it as a stop) in only five cases, or less than 1% of the total cases. In the following subsections, a token will be considered a flap if at least three of the four listeners identified it as a flap, likewise with the stops. Tokens with a split categorization are identified as such.

5.1 Paradigm uniformity

In this section, the listeners' categorization of the /t/ allophones in the four target words are reported in Figures 10-13 by plotting the number of perceived flaps in the repetitions of a base form against the number in the corresponding inflected form for each speaker. The following schematic graph in Figure 9 illustrates how to read these graphs and how to identify potential cases of paradigm uniformity.

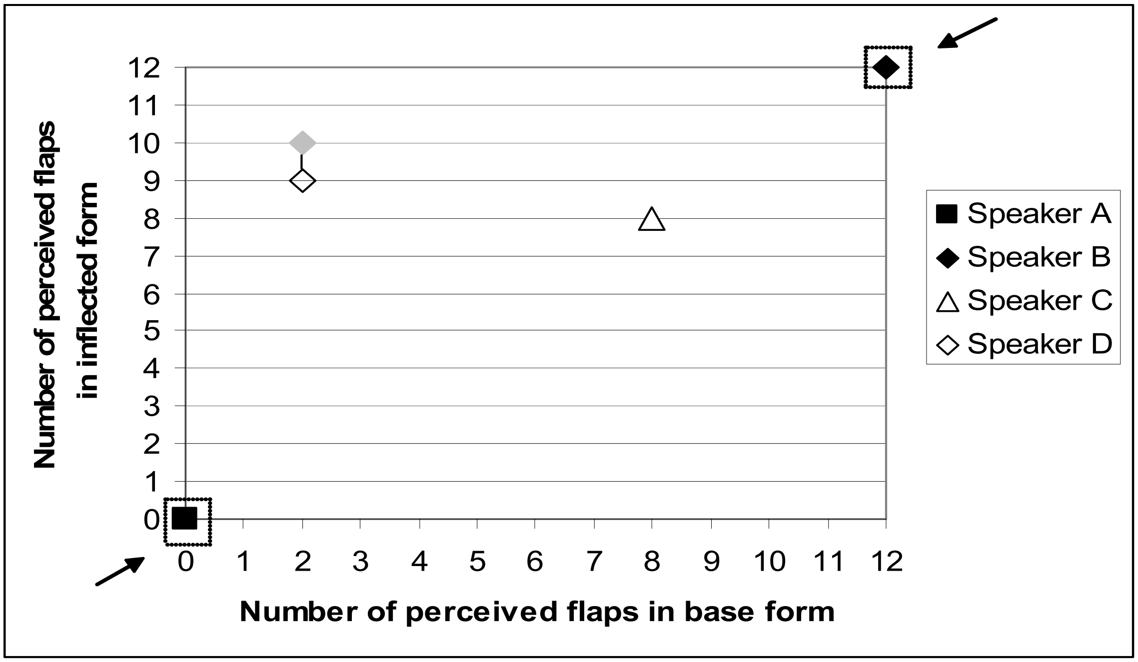

Figure 9: Schematic chart for a hypothetical word pair illustrating the format of Figures 10-13. Number of perceived flaps in a base form are plotted on the x-axis and number of perceived flaps in an inflected form are plotted on the y-axis. Boxes and arrows at points 0-0 and 12-12 are potential candidates for uniform paradigms (Speakers A and B); points anywhere else on the chart are not (Speakers C and D). The gray diamond attached to Speaker D's value indicates a missing token that may have resulted in an additional flap.

The number of perceived flaps in the base form of a word, out of 12 repetitions, is plotted on the x-axis while the number in the inflected form, out of 12 repetitions, is plotted on the y-axis. For example, Speaker A in Figure 9 above articulates 0 flaps in both the base and inflected forms; Speaker B articulates 12 flaps in both the base and inflected forms; Speaker C articulates 8 flaps in the base and 8 in the inflected form; and Speaker D articulates 2 flaps in the base and 9 in the inflected form. The shaded box attached to Speaker D's value point indicates that one of the speaker's inflected tokens is missing and, if articulated, may have resulted in an additional flap.

A potential uniform paradigm can be identified as a value point at 12-12 or 0-0, indicating that a speaker uttered either all flaps or no flaps (and therefore all stops) in both the base and inflected form. These areas of the graph are highlighted by a box and an arrow. This method of evaluation is based on the assumptions in Steriade (2000), where the author concludes that both consistent stop articulation or consistent flap articulation across a form indicates a uniform paradigm. I will argue in Section 6.1 that consistent flapping (a value at point 12-12) does not necessarily provide evidence for paradigm uniformity. However, at this point the data will be considered based upon the assumptions in Steriade (2000). A value that falls at any point on the graph other than 0-0 or 12-12 indicates a non-uniform paradigm. Therefore, the values for Speakers A and B are potential candidates for uniform paradigms, while the values for Speakers C and D are not. The five segments that received a split identification from listeners are considered of ambiguous status and are not included in the figures but are mentioned below each relevant figure.

The following Figures 10-13 display the results of the number of perceived flaps in each word pair for each speaker.

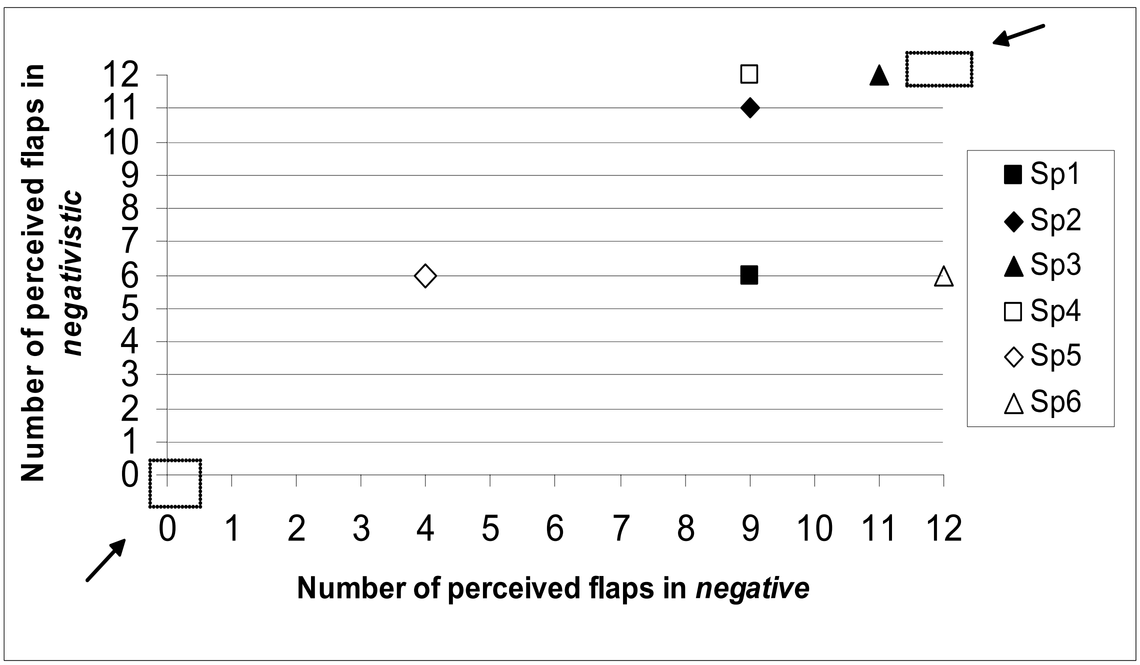

Figure 10: Number of perceived base flaps (x-axis) plotted against number of perceived inflected flaps (y-axis) in *negative/negativistic* for each speaker, out of 12 repetitions

As seen in Figure 10 above, no speaker maintains a uniform paradigm in the *negative/negativistic* pair. Sp3 comes close, however, with 11 flaps in the base and 12 in the inflected form. The other speakers articulate a mix of flaps and stops both within the repetitions of a single form as well as across the members of the paradigm. Sp1 articulates more flaps in the base than the inflected form, and all five of the other speakers articulate more flaps in the inflected form than in the base form. In general, with the exception of Sp5, all speakers have a greater tendency to produce flaps than stops. One of Sp4's base tokens and one of Sp5's inflected tokens received a split identification from the listeners. All other non-flaps were identified as stops.

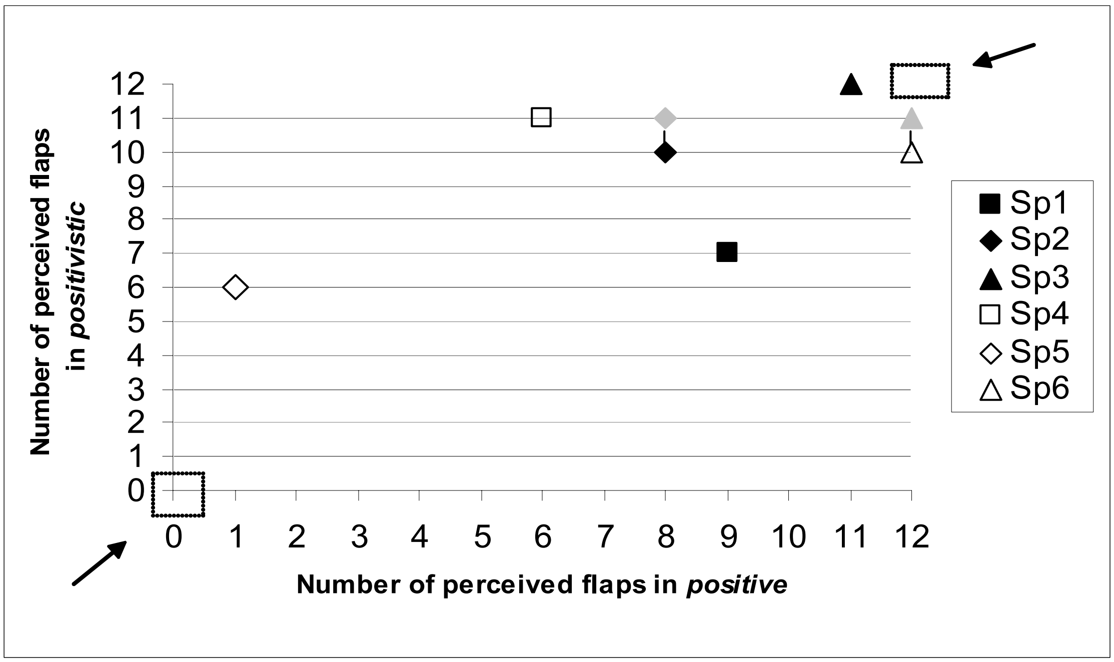

Figure 11: Number of perceived base flaps (x-axis) plotted against number of perceived inflected flaps (y-axis) in *positive/positivistic* for each speaker, out of 12 repetitions

As seen in Figure 11 above, no speaker maintains a uniform paradigm in the *positive/positivistic* pair, though two speakers come close, with Sp3 articulating 11 flaps in the base and 12 in the inflected form and Sp6 articulating 12 flaps in the base and 10 in the inflected form (with a possibility of 11 in the inflected form, had there been no missing token). All other speakers articulate a mix of flaps and stops both within the repetitions of a single form as well as across the member of the paradigm. In general, with the exception of Sp5, all speakers have a greater tendency to produce flaps than stops. One of Sp3's base tokens received a split identification from the listeners. All other non-flaps were identified as stops.

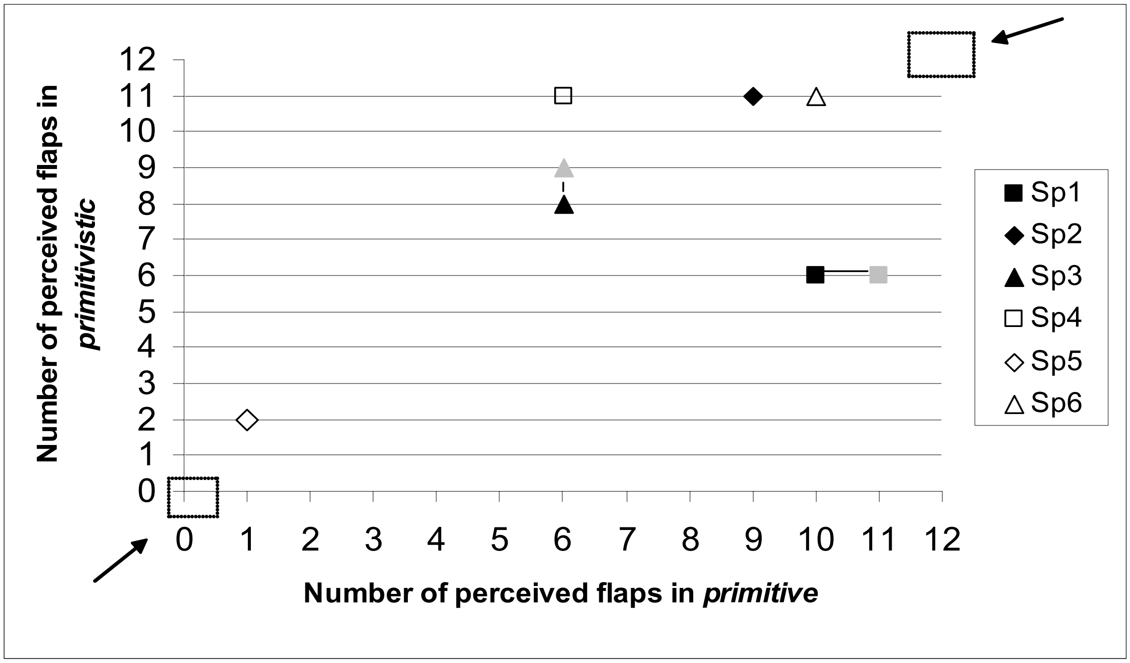

Figure 12: Number of perceived base flaps (x-axis) plotted against number of perceived inflected flaps (y-axis) in *primitive/primitivistic* for each speaker, out of 12 repetitions

As seen in Figure 12 above, no speaker maintains a uniform paradigm in the *primitive/primitivistic* pair. Sp6 comes close, however, with 10 flaps in the base and 12 in the inflected form. All other speakers articulate a mix of flaps and stops both within the repetitions of a single form as well as across the members of the paradigm. In general, with the exception of Sp5, all speakers have a greater tendency to produce flaps than stops. One of Sp3's base tokens received a split identification from listeners. All other non-flaps were identified as stops.

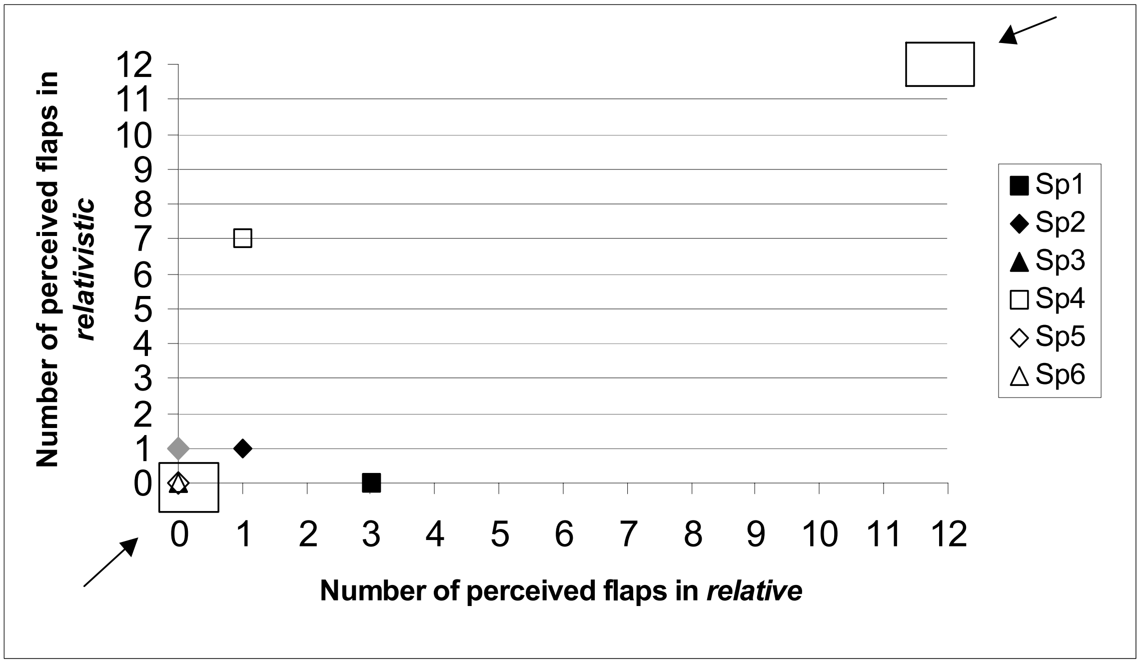

Figure 13: Number of perceived base flaps (x-axis) plotted against number of perceived inflected flaps (y-axis) in *relative/relativistic* for each speaker, out of 12 repetitions

As seen in Figure 13 above, three speakers, Sp3, Sp5 and Sp6, all maintain uniform paradigms in the *relative/relativistic* pair, articulating no flaps in either the base or inflected form (although as indicated by the shaded diamond, the repetitions of the inflected forms are out of 11, not 12, for Sp5). Sp2 also comes close to articulating a uniform paradigm, with one flap in both the base and inflected forms. In general, all speakers have a greater tendency to produce stops than flaps in this pair, although Sp4 articulates more flaps than stops in the base form. One of Sp1's base tokens received a split identification from the listeners. All other non-flap tokens were identified as stops.

5.1.1 Summary

The results of Figures 10-13 are summarized in Figure 14 below. Cases of uniform paradigms are indicated by a checkmark in the box identifying the appropriate word pair and speaker. Cases of near-uniform paradigms, where at least ten of twelve repetitions are consistent across a base-inflected pair for a speaker, are indicated with the number of base/inflected flaps in the appropriate box. Empty cells reveal instances of no observed paradigm uniformity.

Uniform paradigms are observed in only three cases, Sp1, Sp3 and Sp6's *relative/istic* pairs, where each speaker articulates all stops. In the other three word pairs, no speaker maintains a uniform paradigm, although two speakers produce near-uniform paradigms in two cases: Sp3 articulates 11 base and 12 inflected flaps in *negative/istic* and *positive/istic* while Sp6 articulates 10 base and 10 inflected flaps in *positive/istic* and 10 base and 11 inflected flaps in *primitive/istic*. Therefore, of the 24 possible pairs, uniform paradigms are observed in three cases, near-uniform paradigms are observed in four cases, and no uniform paradigms are observed in 17 cases. These results are summarized in Figure 14 below.

	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6
negative/istic			11/12			
positive/istic			11/12			10/10
primitive/istic						10/11
relative/istic	✓		✓			✓

Figure 14: Summary of results in Figures 10-13. ✓ = uniform paradigm. Numbers indicate cases of near-uniform paradigms, where the first value represents number of perceived flaps in the base and the second value represents number of perceived flaps in the inflected form. Empty boxes represent no case of paradigm uniformity observed.

5.2 Acoustic analyses

Measurements of duration, VOT and closure voicing were made for each /t/ phone, in accordance with the methodology described in Section 4.3. The results of these analyses are presented in the following subsections. For each acoustic cue, a figure containing the average value for perceived flap and stop tokens for each speaker is presented first. Then, each speaker's results are presented in a separate figure. In these latter figures, each /t/ allophone is grouped into a 10 millisecond category (for duration and VOT) or 10 percentage-point category (for voicing) in accordance with the perceived flap or stop categorization. In this format, it is possible not only to observe the range of values a speaker's perceived flaps and stops have in regards to the relevant attribute, but also to determine how well the perception of the allophone corresponds to the acoustic property. Each figure is based on 96 tokens (4 base forms + 4 inflected forms x 12 repetitions x 1 speaker), except in cases where one of the five missing tokens reduces that number.

Note that the scales in each figure differ. Adjusting the scales is necessary as speakers varied overall in the number of perceived flaps or stops articulated. The total number of flaps or stops that one speaker produced compared to the total number another speaker produced is not relevant to this study; rather, understanding the distribution of perceived flaps and stops as they relate to a given cue for a given speaker is what is important and therefore highlighted in the figures.

The structure of this section is as follows. Subsection 5.2.1 contains the results of the duration analyses; 5.2.2 contains the results of the VOT analyses, and 5.2.3 contains the results of the closure voicing analyses. Subsection 5.2.4 contains a summary of the three acoustic analyses.

5.2.1 Duration

As discussed in Section 4.3, total duration was calculated as closure duration plus VOT. Figure 15 below contains the average total durations, rounded to the nearest millisecond, for each speaker's perceived flap and stop tokens.

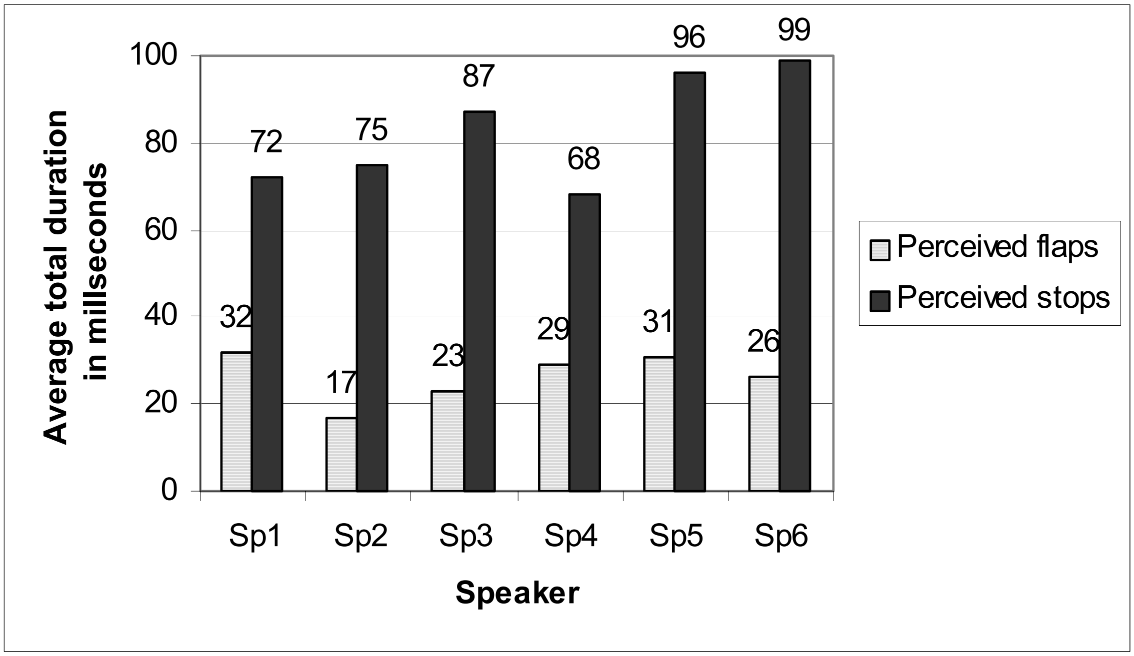

Figure 15: Average total durations of perceived flaps and stops for all speakers

As can be seen in Figure 15, the average duration of each speaker's perceived flaps is considerably shorter than that of each speaker's perceived stops. The smallest gap in duration between the allophones 39 milliseconds, as seen in the data for Sp4, whose average flap is 29 ms and average stop is 68 ms. The largest gap is 73 ms, as seen in the data for Sp6 whose average flap is 26 ms and average stop is 99 ms. This figure illustrates that each speaker does in fact exhibit a difference in average duration between perceived flaps and stops.

Figures 16 through 21 below display the total duration results for each speaker's /t/ allophones, in the format discussed at the beginning of the section. Split identification tokens are also included. In the case of Sp1 and Sp4, where additional total duration measurements were made, as discussed in Section 4.3 above, two figures are provided, the first where total duration is calculated as the distance between beginning of closure to VOT, as is the case in all of the other figures, and the second where total duration is calculated as the distance between beginning of closure and the onset of formants of the following vowel.

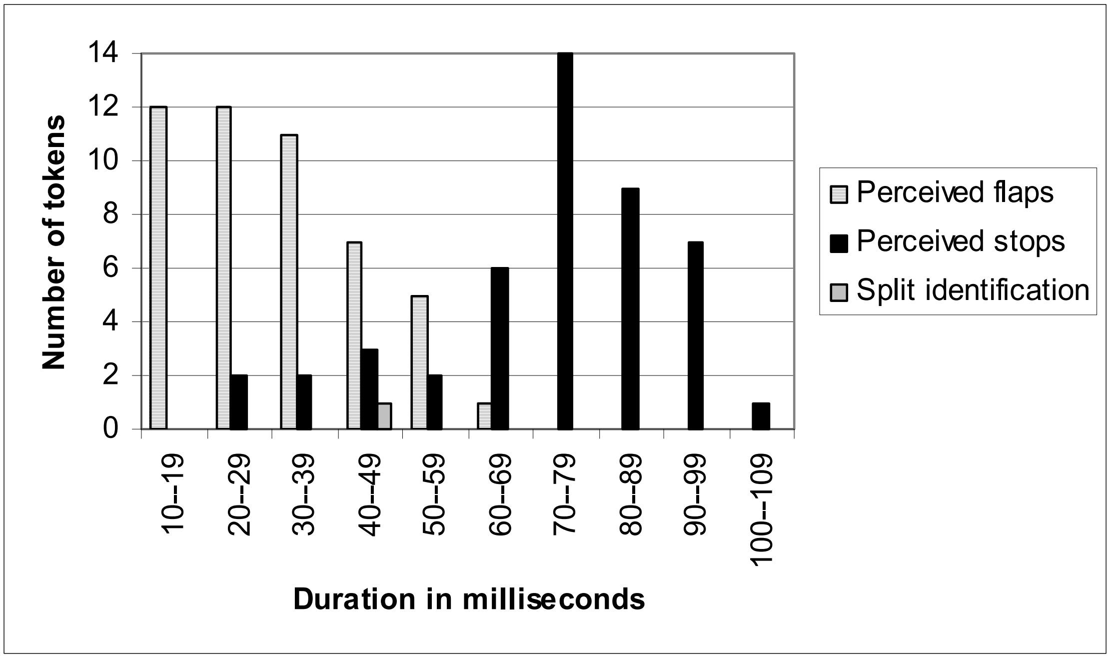

Figure 16a: Total durations of /t/ allophones for Sp1, calculated as closure duration + VOT

Figure 16a reports the total duration of Sp1's perceived flaps and stops when calculated as duration of closure + VOT. Sp1's 48 flap tokens have total durations ranging from 10 to 69 milliseconds, with the majority, 35 tokens, falling between 10 and 39 milliseconds. Sp1's 46 stops exhibit a wider range of durations, from 20 to 109 milliseconds, with a peak of 14 tokens in the 70 to 79 millisecond range. Although the two allophones fall into largely separate groups in terms of duration, there is a great deal of overlap in the 20 to 69 millisecond range. Additionally, Sp1 has one ambiguous token, which falls in the 40-49 millisecond range.

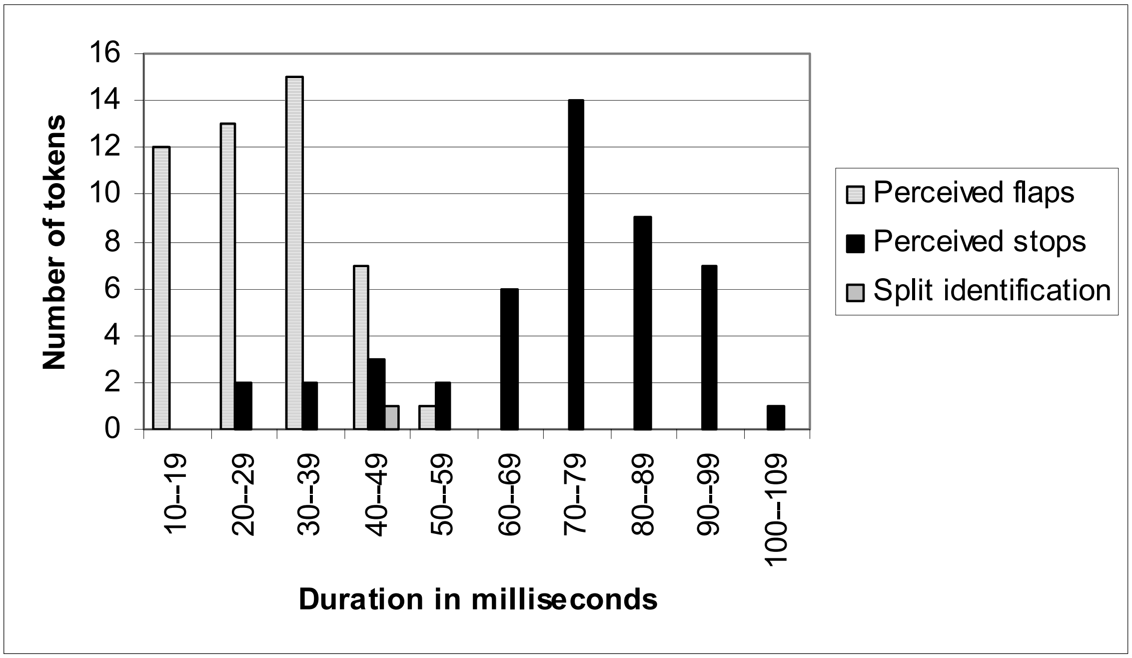

Figure 16b: Total durations of /t/ allophones for Sp1, calculated as distance from /t/ closure (offset of F2 and F3) to onset of F2 and F3 of the following vowel.

The revised Figure in 16b reports the total durations of Sp1's /t/ allophones as measured from the beginning of closure to the onset of F2 and F3 of the following vowel. These data in this figure differ from those in Figure 16a in that there is no longer any overlap between flaps and stops in the 60-69 millisecond range, and there is only one flap, as opposed to five, in the 50-59 millisecond range. Although these revised total duration values result in less overlap between the flaps and stops in Sp1's data, overlap does still exist in the 20 to 59 millisecond range.

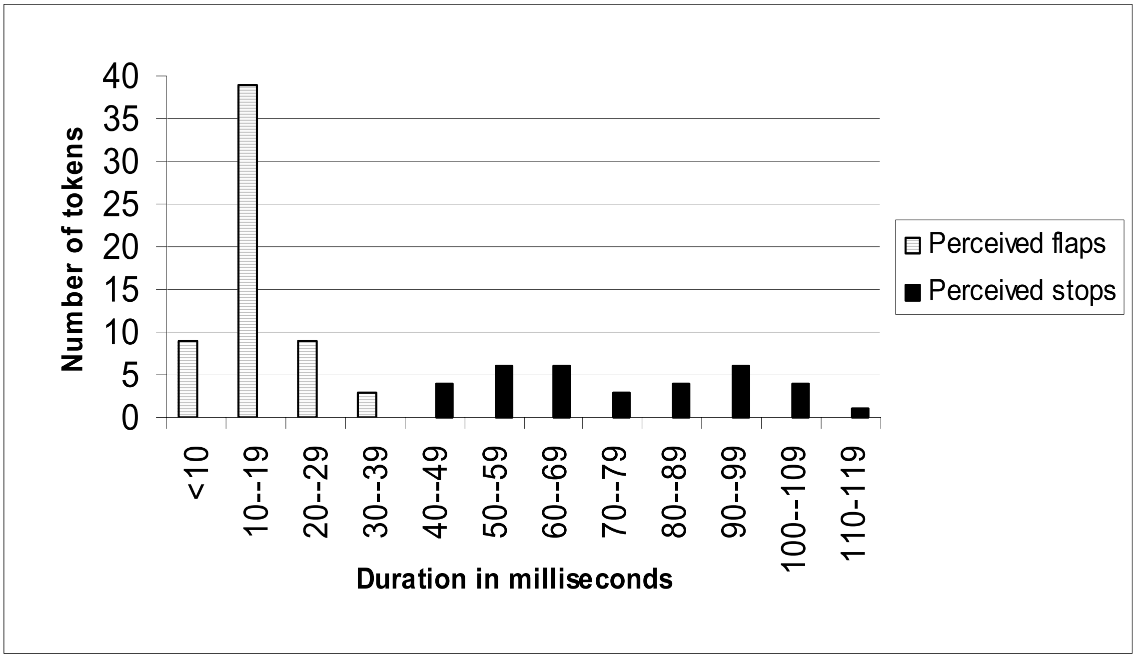

Figure 17: Total durations of /t/ allophones for Sp2 calculated as closure duration + VOT

As seen in Figure 17 above, Sp2's 60 perceived flap tokens range from 0 to 39 milliseconds, with the vast majority, 39 tokens, falling in the 10 to 19 millisecond range. Sp2's 34 perceived stop tokens show a fairly even distribution within the 40 to 119 millisecond range. The two allophones fall into distinct groups based upon duration in Sp2's speech, with all flap tokens at 39 milliseconds or less and all stop tokens at 40 milliseconds or greater.

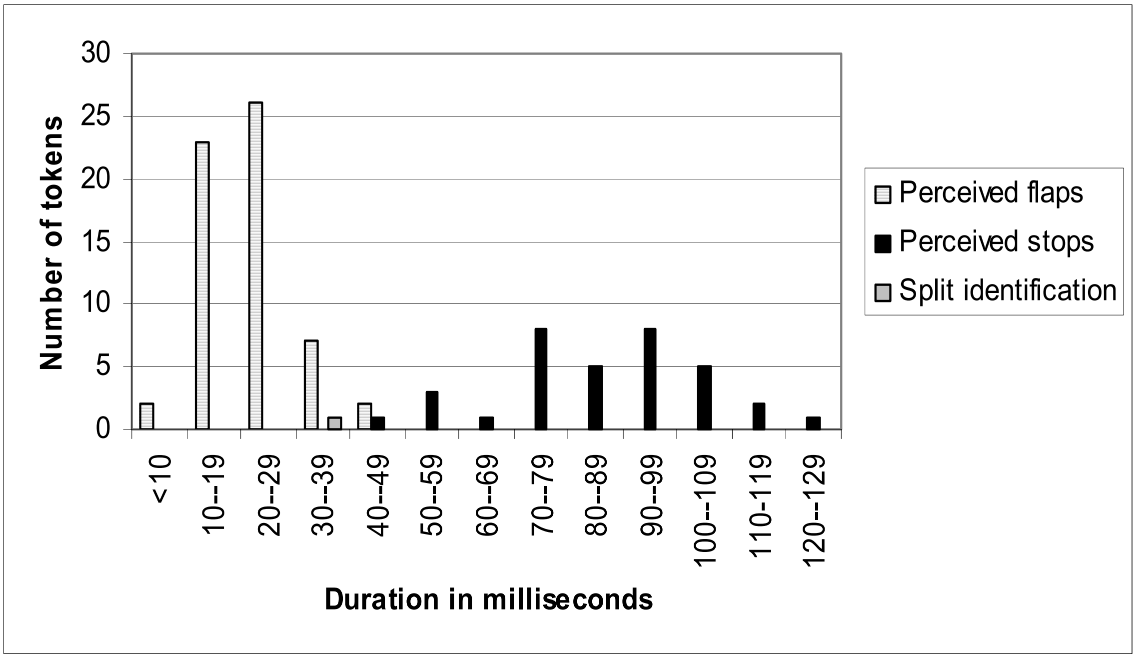

Figure 18: Total durations of /t/ allophones for Sp3 calculated as closure duration + VOT

As seen in Figure 18 above, Sp3's 60 perceived flap tokens range in duration from 0 to 49 milliseconds, with the majority, 49 tokens, falling between 10 and 29 milliseconds. The perceived stops display a wider distribution, falling between 40 and 129 milliseconds, with a concentration of 26 tokens falling between 70 and 109 milliseconds. Sp3's flaps and stops fall into two mostly distinct groups based upon duration, but with a small degree of overlap in the 40-49 millisecond range. Additionally, Sp3 has one ambiguous token, which falls in the 30-39 millisecond range.

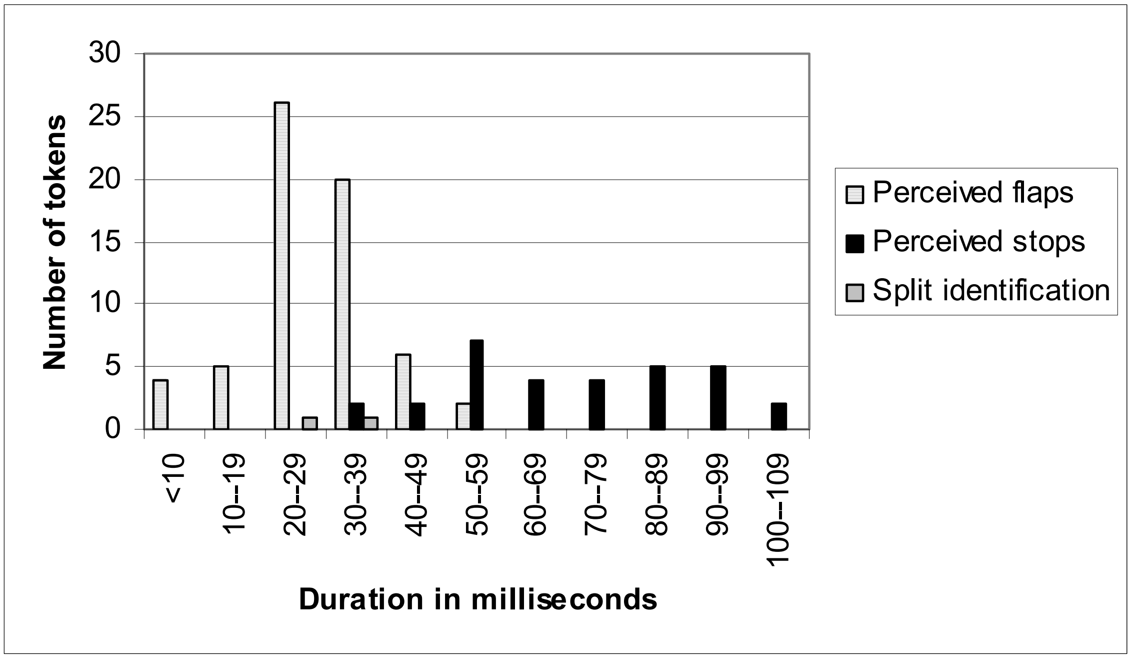

Figure 19a: Total durations of /t/ allophones for Sp4 calculated as closure duration + VOT

Figure 19a reports the total durations of Sp4's /t/ allophones calculated as duration of closure + VOT. According to this measure, the 63 perceived flap tokens fall between 0 and 59 milliseconds, with a vast majority, 46 tokens, between 20 and 39 milliseconds. Sp4's 31 perceived stops have a wider distribution, falling fairly evenly from 30 to 109 milliseconds. There is some degree of overlap between Sp4's flap and stop durations in the 30 to 59 millisecond range. Additionally, Sp4 has two ambiguous tokens, one in the 20-29 millisecond range and one in the 30-39 millisecond range.

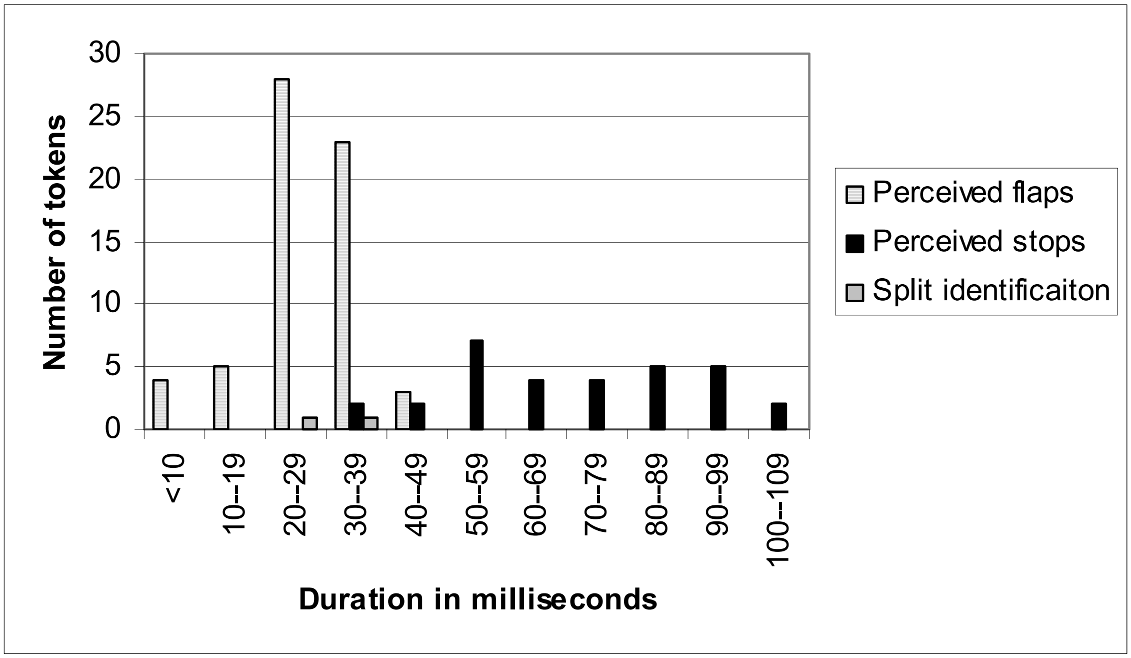

Figure 19b: Total durations of /t/ allophones for Sp4, calculated as distance from /t/ closure (offset of F2 and F3) to onset of F2 and F3 of the following vowel

Figure 19b reports the total durations of Sp4's /t/ allophones calculated as distance between beginning of closure and onset of F2 and F3 of the following vowel. The difference between these results and those in Figure 19a above is the lack of flaps in the 50-59 millisecond range, and the decrease of flaps (from 6 to 3) in the 40-49 millisecond range. The degree of overlap between flap and stop durations is less according to this measure of duration, although there is still some overlap in the 30 to 49 millisecond range.

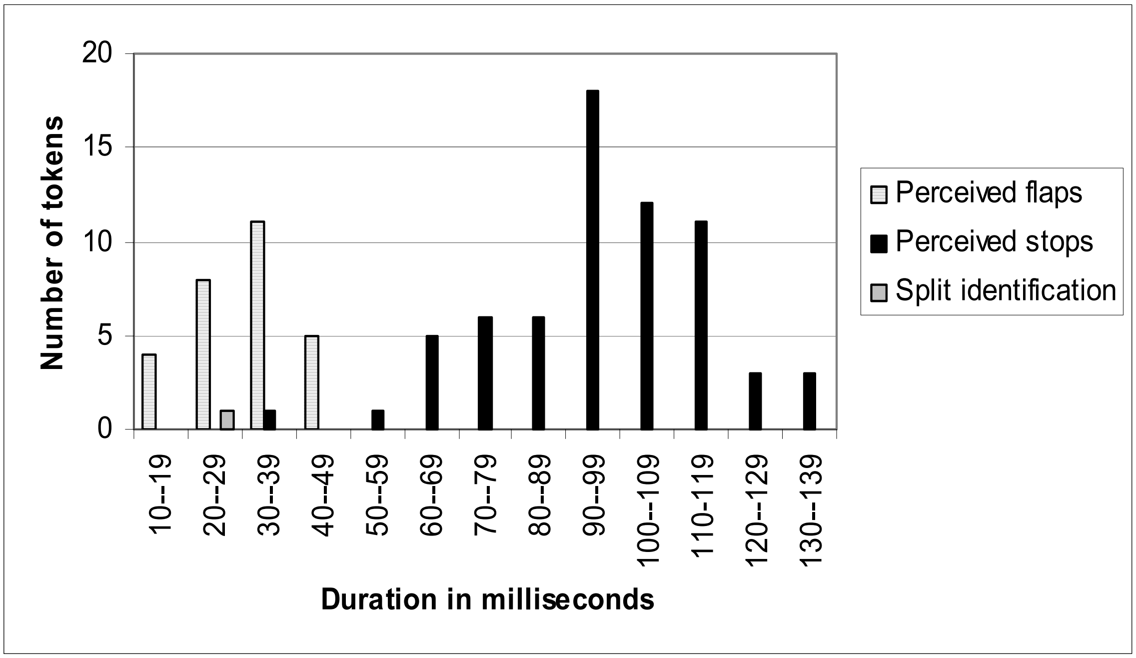

Figure 20: Total durations of /t/ allophones for Sp5 calculated as closure duration + VOT

As seen in Figure 20 above, Sp5's 28 perceived flap tokens fall between 10 and 49 milliseconds, with the majority, 11 tokens, in the 30 to 39 millisecond range. The 66 perceived stop tokens fall between 30 and 139 milliseconds, with the highest concentration, 18 tokens, in the 90-99 millisecond range. With the exception of one stop token in the 30-39 millisecond range, Sp5's flap and stop tokens fall into two separate groups based on duration, with flaps being 49 milliseconds or less and most stops being 50 milliseconds or more. Additionally, Sp5 has one ambiguous token, which falls in the 20-29 millisecond range.

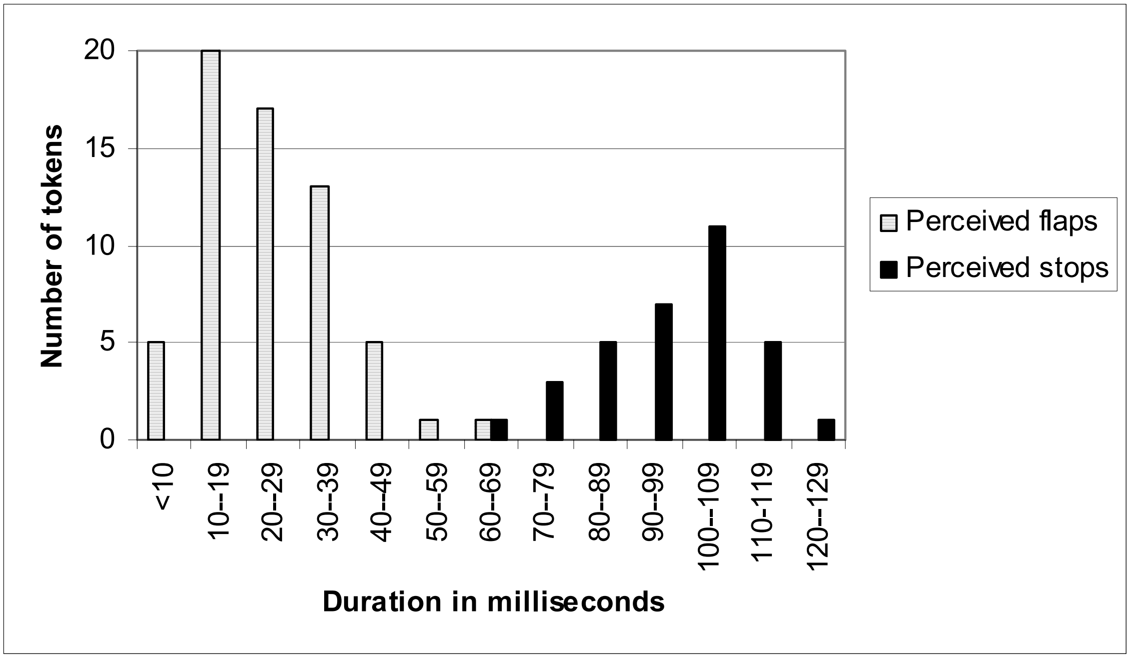

Figure 21: Total durations of /t/ allophones for Sp6 calculated as closure duration + VOT

As seen in Figure 21 above, Sp6's 62 perceived flap tokens fall between 0 and 69 milliseconds, with the majority, 40 tokens, between 10 and 39 milliseconds. The 33 perceived stops range from 60 to 129 milliseconds, with a peak of 11 tokens in the 100-109 range. With the exception of a small degree of overlap in the 60-69 millisecond range, Sp6's flap and stop tokens fall into two separate groups based upon duration, with flaps at 69 milliseconds or less and stops at 60 milliseconds or more.

5.2.2 VOT

As discussed in Section 4.3, VOT was calculated as the distance from the hand-labeled "end of closure" label to the first point (of five consecutive points) of voicing. Figure 22 below displays the average VOT of perceived flaps and stops for each speaker.

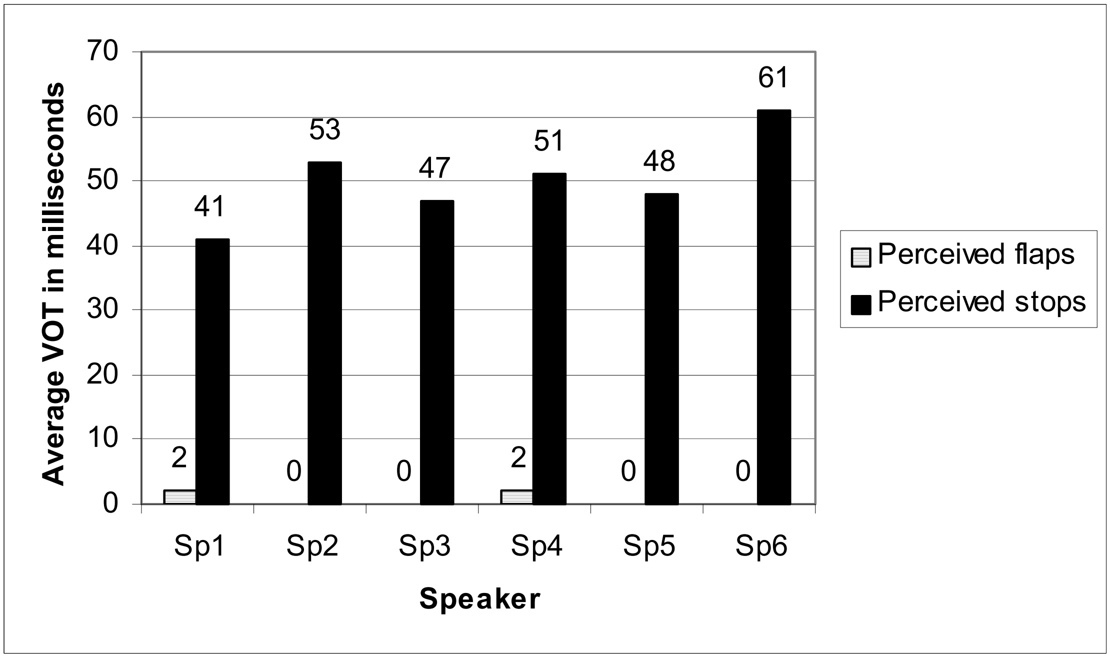

Figure 22: Average VOT values of perceived flaps and stops for all speakers

As seen in Figure 22, the average VOT of each speaker's perceived flaps is significantly shorter than that of each speaker's perceived stops. The smallest gap in VOT between the two allophones is 39 ms, as seen in the data for Sp1, whose perceived flaps have an average VOT value of 2 ms and perceived stops have an average VOT value of 41 ms. The largest gap is 61 ms, as seen in the data for Sp6 whose perceived flaps have an average VOT value of 0 ms and perceived stops have an average VOT value of 61 ms. This figure illustrates that each speaker does in fact exhibit a difference in average VOT value between perceived flaps and stops.

Figures 23 through 28 below report the VOT values of each speaker's /t/ allophones, in the format described at the beginning of the section.

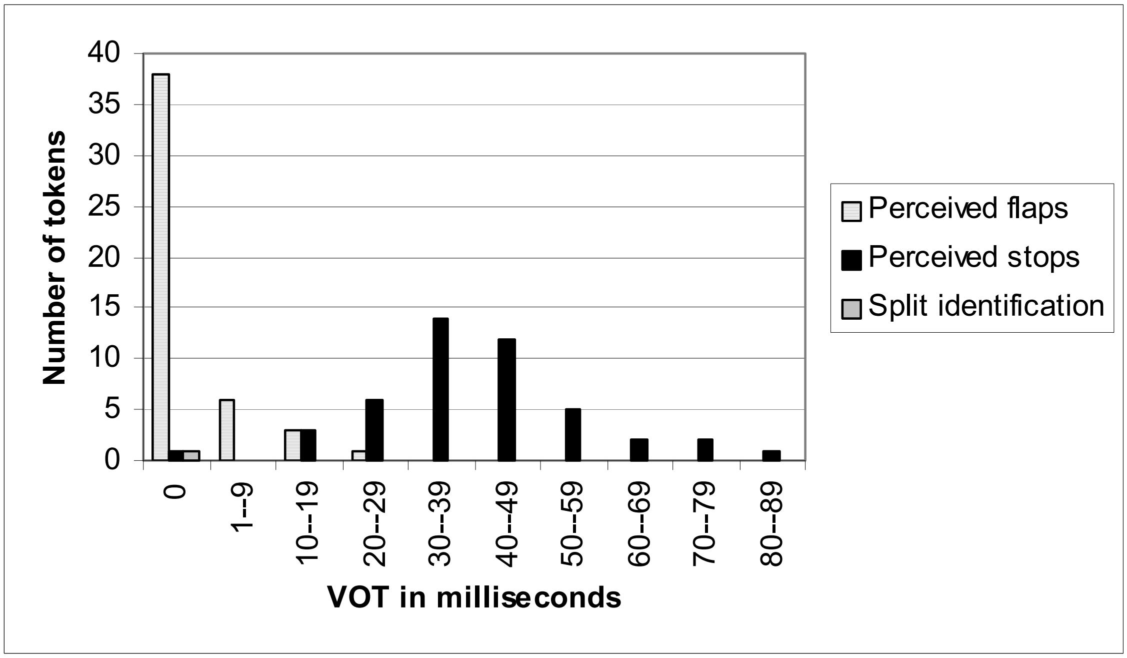

Figure 23: VOT values of Sp1 /t/ allophones

As seen in Figure 23 above, the vast majority of Sp1's perceived flaps, 38 tokens, have VOT values of 0, although an additional ten tokens fall between one and 29 milliseconds. The 46 perceived stop tokens have a wider range, with VOT values ranging from 0 to 89 milliseconds, with the majority, 26 tokens, falling between 30 and 49 milliseconds. Despite some degree of overlap with flap and stop tokens in the 0 to 29 millisecond range, the two groups are fairly distinct in Sp1's speech. Additionally, Sp1's one ambiguous token has a VOT value of 0.

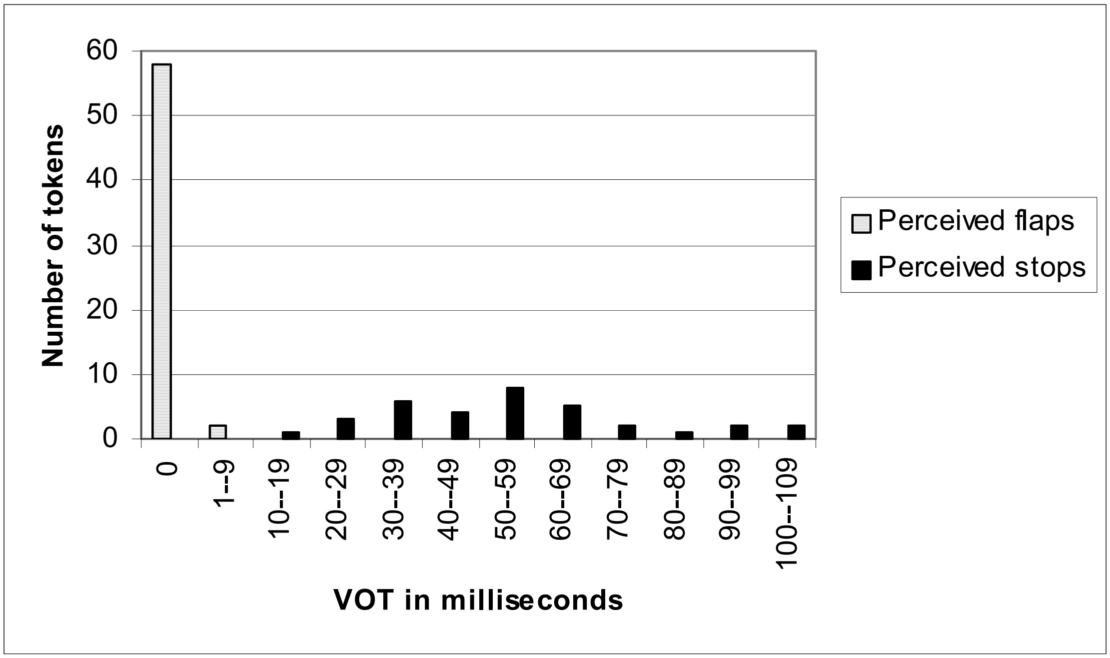

Figure 24: VOT values of Sp2 /t/ allophones

As seen in Figure 24 above, the vast majority of Sp2's perceived flaps, 58 tokens, have VOT values of 0, with two additional flap tokens falling between one and nine milliseconds. The 34 perceived stop tokens, on the other hand, have a much wider range of values, falling fairly evenly between ten and 109 milliseconds. Sp2's flap and stop tokens form two completely distinct groups with no overlap, with all flaps having VOT values of nine or less milliseconds and all stops having VOT values of ten or more milliseconds.

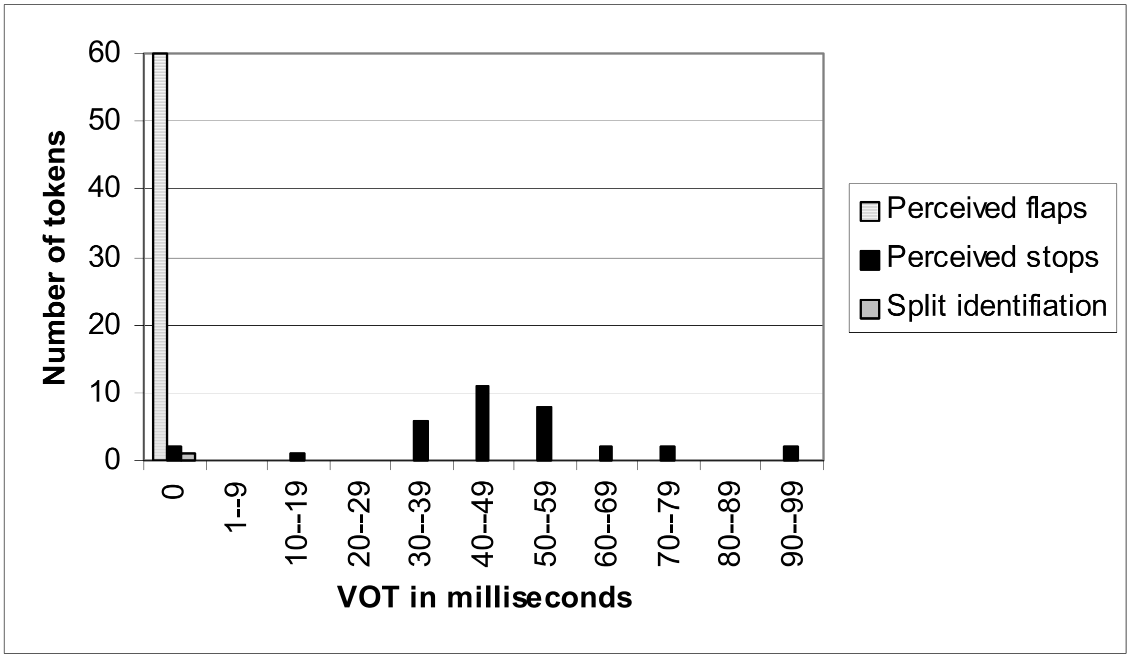

Figure 25: VOT values of Sp3 /t/ allophones

As seen in Figure 25 above, all of Sp3's 60 perceived flap tokens have VOT values of 0. The 34 perceived stops, on the other hand, have a much wider range, with values ranging from 0 to 99 milliseconds, with the majority of the stops, 25 tokens, falling between 30 and 59 milliseconds. Sp3's flap and stop tokens form two almost distinct groups, with an overlap of only two tokens in the 0 VOT category. Additionally, Sp3's one ambiguous token has a VOT value of 0.

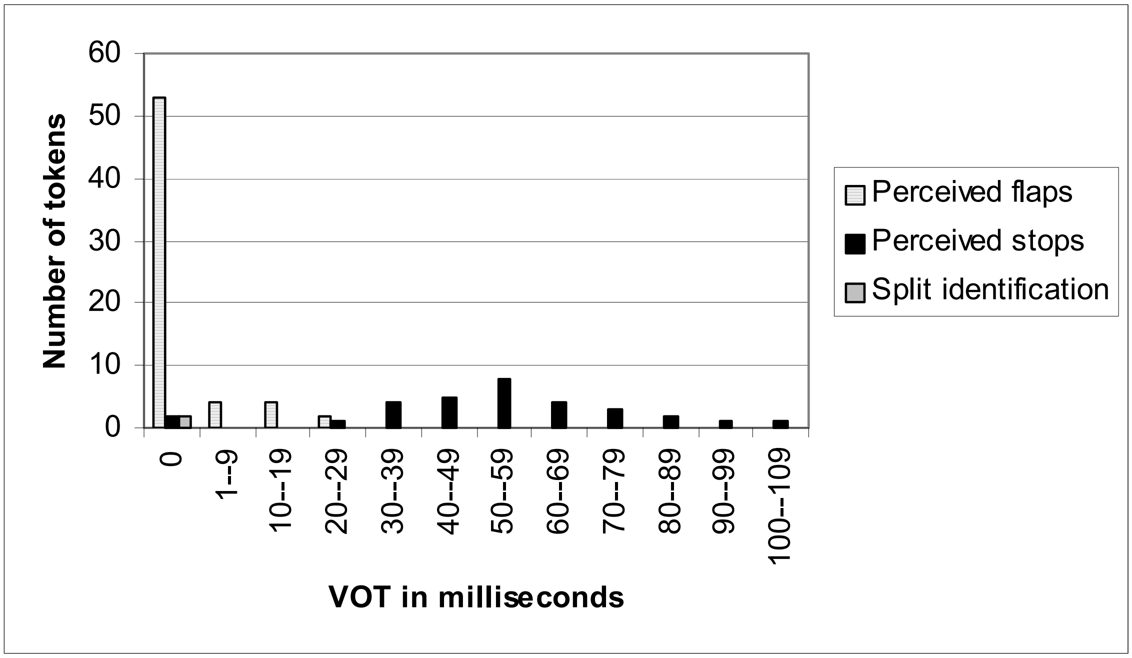

Figure 26: VOT values of Sp4 /t/ allophones

As seen in Figure 26 above, the vast majority of Sp4's perceived flaps, 53 tokens, have VOT values of 0, although ten tokens fall between the one and 29 millisecond range. The 31 perceived stops have a much wider range of VOT values, extending from 0 to 109 milliseconds, with a small peak in the 50-59 millisecond range. Despite some overlap, Sp4's flaps and stops fall into two fairly distinct categories, with flaps largely having values of 29 milliseconds or less and stops 30 milliseconds or more, with just three stops occurring in the flap region. Additionally, Sp4's two ambiguous tokens have VOT values of 0.

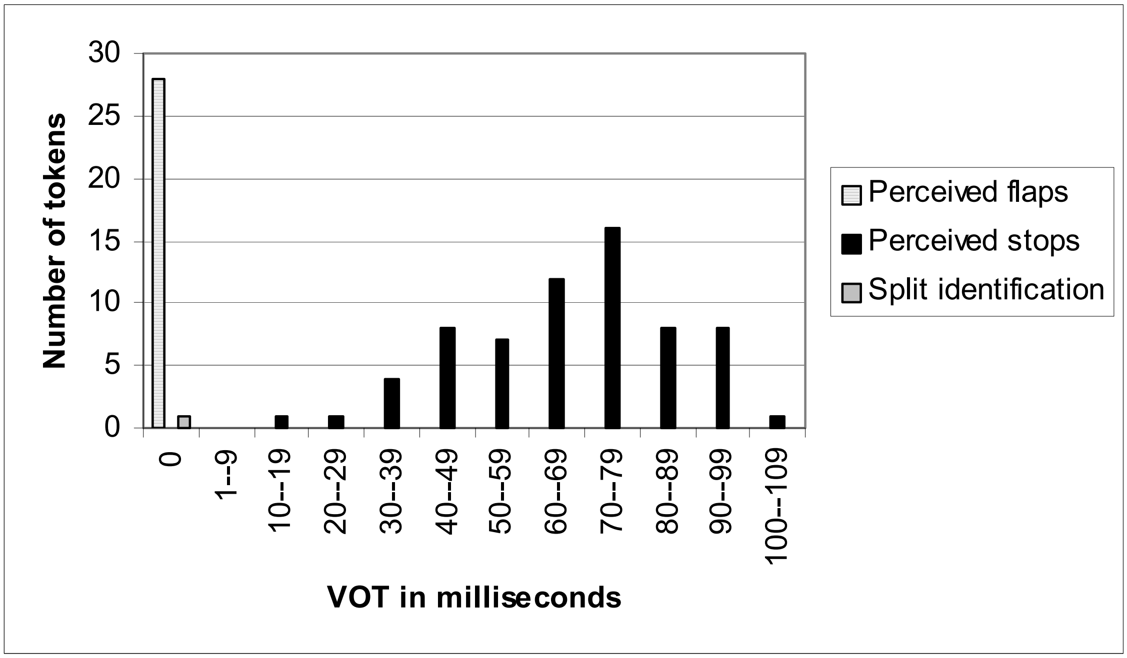

Figure 27: VOT values of Sp5 /t/ allophones

As seen in Figure 27 above, all of Sp5's 28 perceived flap tokens have VOT values of 0. The perceived stops, on the other hand, have a wide range of VOT values, from ten to 109 milliseconds, with a peak between 70 and 79 milliseconds. Sp5's flap and stop tokens therefore form two completely distinct groups with regard to VOT values, with all flaps having values of one millisecond or less and all stops having values of 10 milliseconds or more. Additionally, Sp5's one ambiguous token has a VOT value of 0.

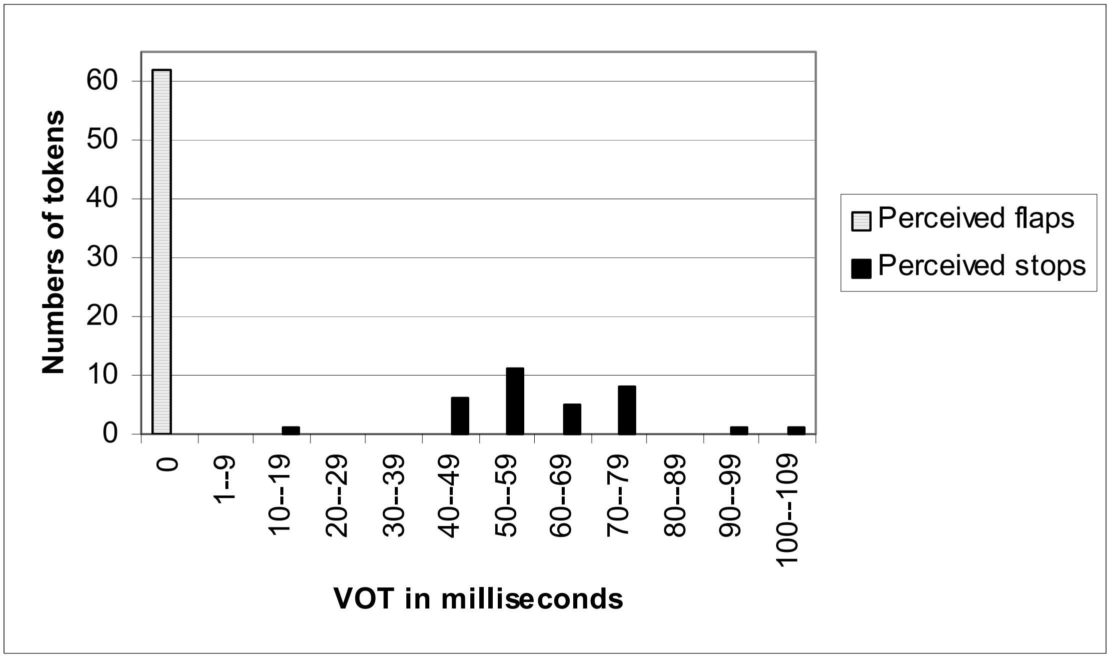

Figure 28: VOT values of Sp6 /t/ allophones

As seen in Figure 28 above, all of Sp6's 62 perceived flap tokens have VOT values of 0. The 33 perceived stop tokens, on the other hand, fall between a ten and 109 millisecond range, with the majority, 30 tokens, clustering between 40 and 79 milliseconds. Sp6's flap and stop tokens therefore form two completely distinct groups with regard to VOT, with all flaps having VOT values of 0 and all stops having values of ten milliseconds or more.

5.2.3 Closure voicing

As discussed in Section 4.3, percent of voicing was calculated over the second half of the closure for each speaker's /t/ phones. The average percentages for each speaker are reported in Figure 29 below.

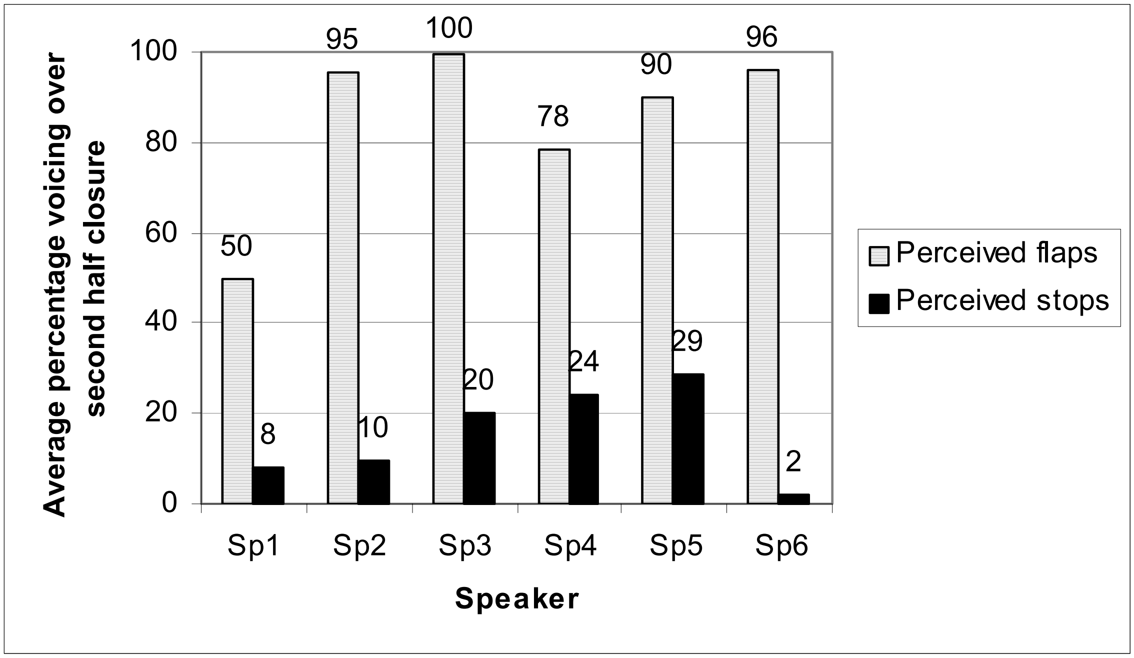

Figure 29: Average percentages of voicing over the second half of the closure of perceived flaps and stops for all speakers

As seen in Figure 29 above, the percent of voicing in the second half of the closure was much less for each speaker's perceived flaps than for each speaker's perceived stops. The smallest difference in percentage of voicing between the two allophones is 42%, as seen in the data for Sp1, where perceived flaps are an average of 50% voiced in the second half and perceived stops are an average of 8% voiced in the second half. The greatest difference is 94%, as seen in the data for Sp6, whose perceived flaps are an average of 96% voiced in the second half and perceived stops are an average of 2% voiced in the second half. This figure illustrates that each speaker does in fact exhibit a difference in average percentage of voicing over the second half of closure between perceived flaps and stops.

The following figures report the percentage of voicing in the second half of the closure for each speaker's /t/ allophones. As mentioned in Section 4.3, tokens with a closure duration of 0 ms have not been included in these calculations and are not reported in the figures. The number of such tokens excluded for each speaker is indicated in the summarization of each speaker's results.

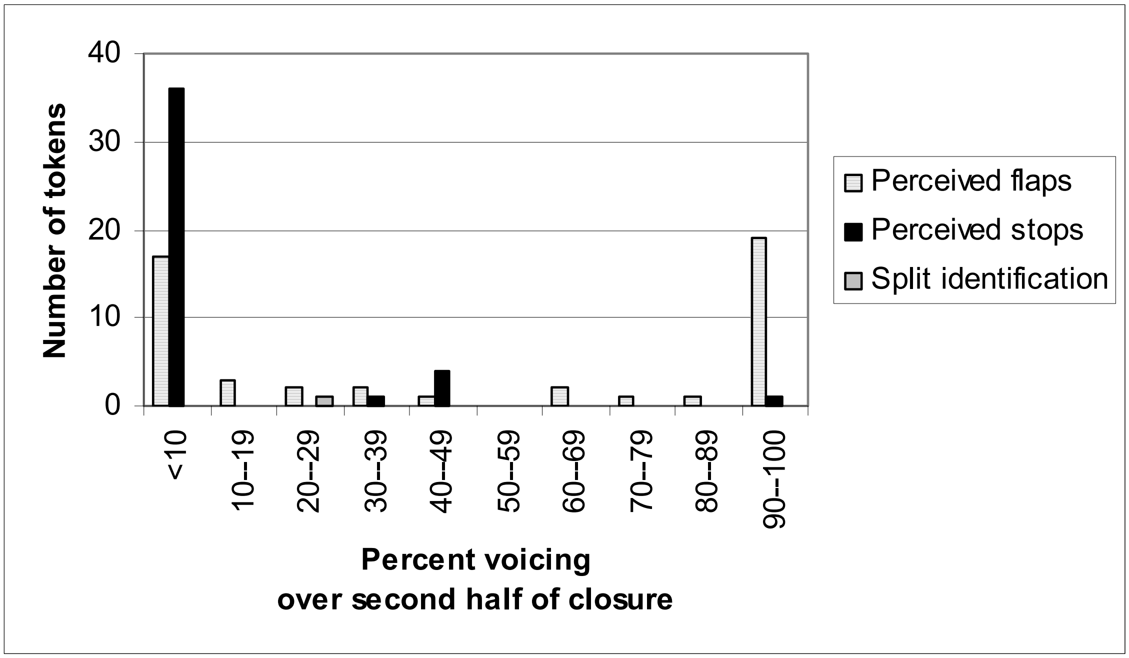

Figure 30: Percent of voicing over second half of /t/ closures for Sp1

As seen in Figure 30 above, the majority of Sp1's perceived flap tokens are equally split between 90-100% closure voicing (19 tokens), and less than 10% voicing (17 tokens), although there are flap tokens at most points in between as well, with a total of 12 tokens falling fairly evenly within the 10-89% range. The perceived stop tokens, on the other hand, are largely voiceless, with 36 having less than 10% voicing, although there are also four in the 30-49% range and one in the 90-100% range. In general, there is great overlap between the flap and stop tokens for this speaker. Additionally, Sp1's one ambiguous token falls in the 20-29% range. Four of Sp1's 95 tokens have been excluded from this chart due to 0 ms closures.

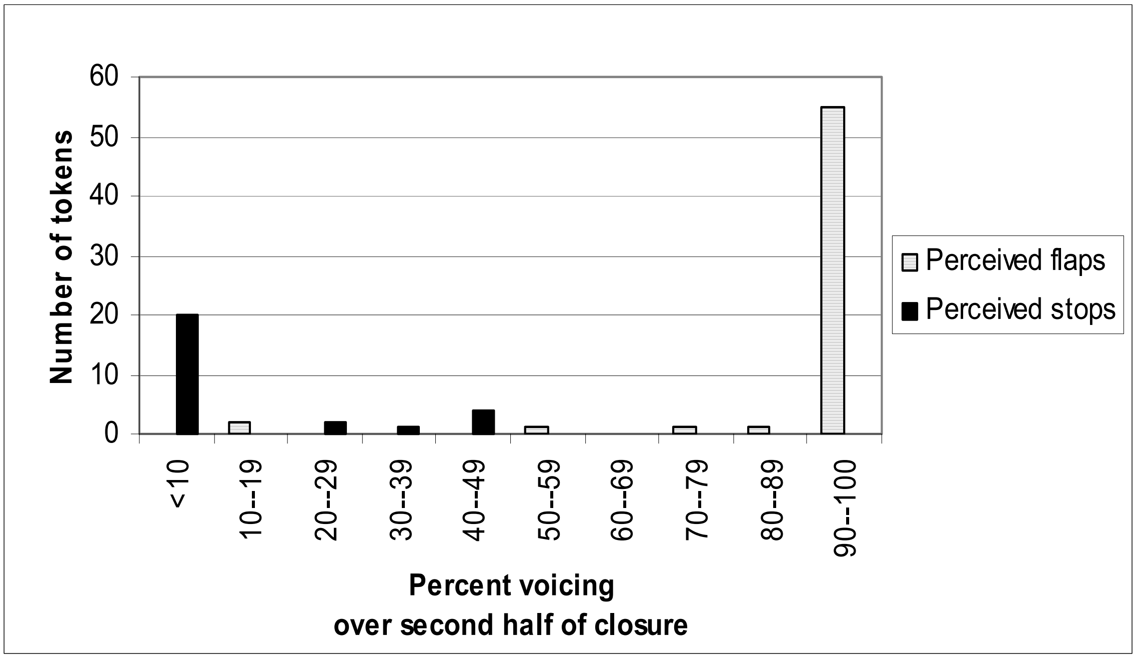

Figure 31: Percent of voicing over second half of /t/ closures for Sp2

As seen in Figure 31 above, the vast majority of Sp2's perceived flap tokens, a total of 55, are 90-100% voiced, although there are five tokens falling within the 10-89% range. The majority of perceived stops, 20, are less than 10% voiced, with seven tokens falling in the 20-49% range. Despite the spread of values for both the flaps and the stops, Sp2's tokens form two fairly distinct groups, with flaps generally falling at the 50% voiced point or above and stops falling at the 49% voiced point or below. Only two tokens, the flaps between 10-19% voiced, disrupt this even distribution. Six of Sp2's 94 tokens have been excluded from the figure due to 0 ms closures.

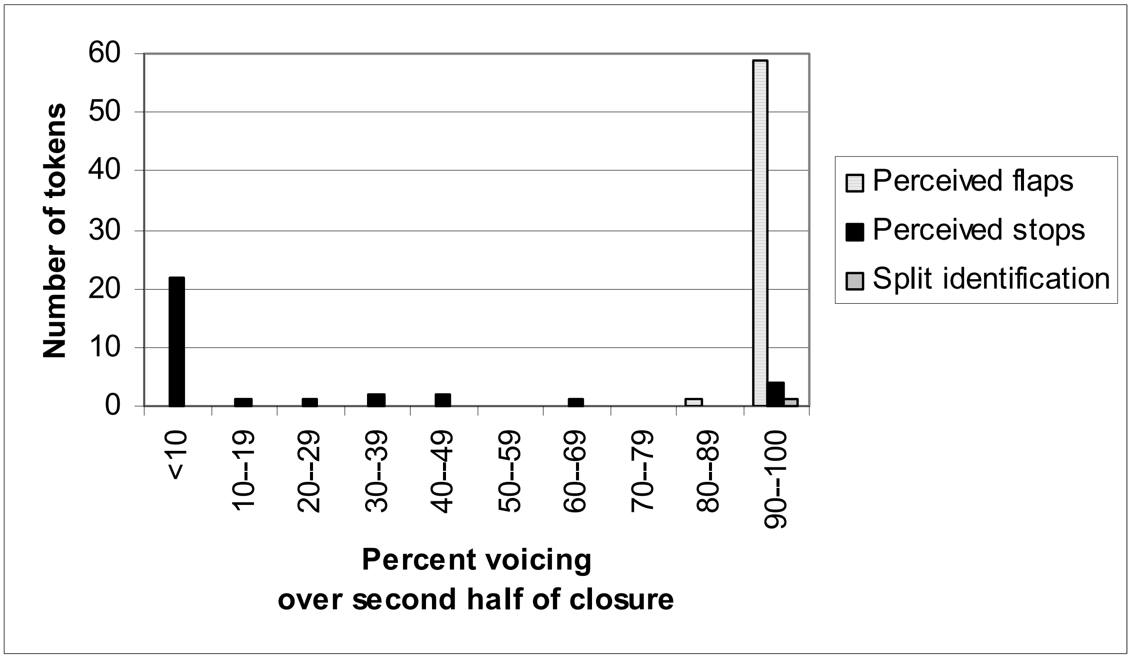

Figure 32: Percent of voicing over second half of /t/ closures for Sp3

As seen in Figure 32 above, the vast majority of Sp3's perceived flap tokens, 59, are 90-100% voiced, although there is one token in the 80-89% range. Although a majority of the perceived stop tokens, 22, fall below 10% voiced, they have a wider range of values than the flaps, with 11 tokens ranging from 10% to 100% voiced. Despite the wide range of closure voicing values for the stops, however, the flap and stop tokens in Sp3's speech form two fairly distinct groups, with flaps being 80% voiced or greater and stops being less than 80% voiced, with the exception of the four stop tokens within the 90-100% range. Additionally, Sp3's one ambiguous token falls in the 90-100% range. Two of Sp3's 95 tokens have been excluded from the figure due to 0 ms closures.

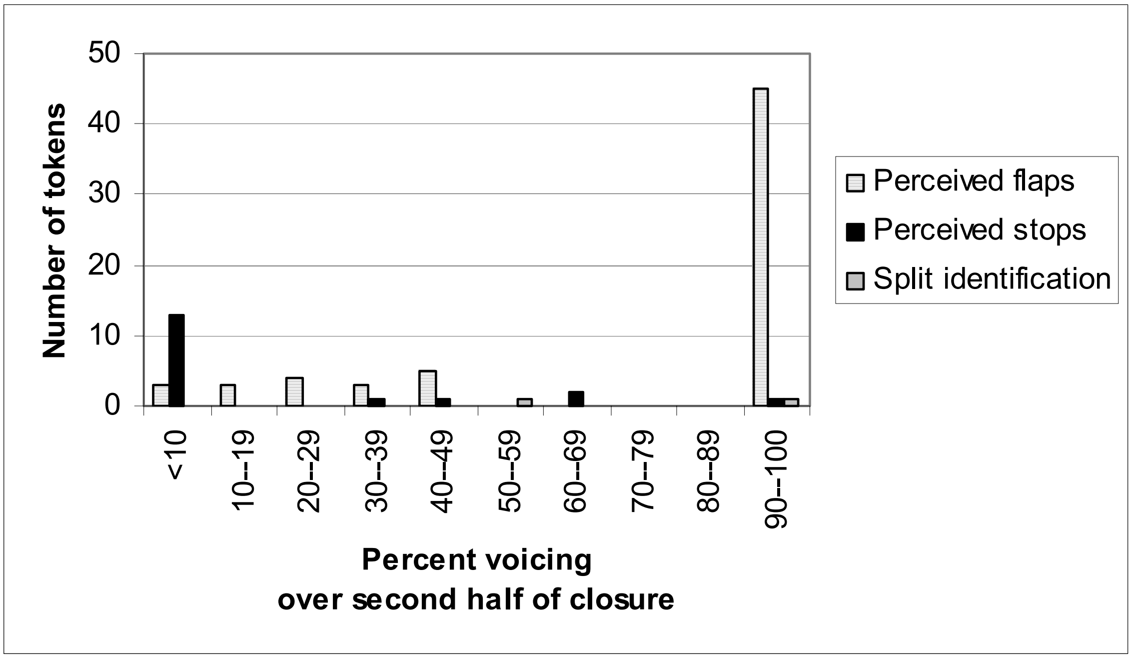

Figure 33: Percent of voicing over second half of /t/ closures for Sp4

As seen in Figure 33 above, the majority of Sp4's perceived flap tokens, 45, fall in the 90-100% voiced range. However, a large number of perceived flaps, 18 tokens, are from 0-49% voiced. Although the perceived stop tokens are largely less than 10% voiced (13 tokens), there are still five tokens that range from 30 to 100% voiced. In general, Sp4's flap and stop tokens exhibit a fairly high degree of overlap. Additionally, one of Sp4's ambiguous tokens falls into the 50-59% range and one falls into the 90-100% range. Thirteen of Sp4's 96 tokens have been excluded from the figure due to 0 ms closures.

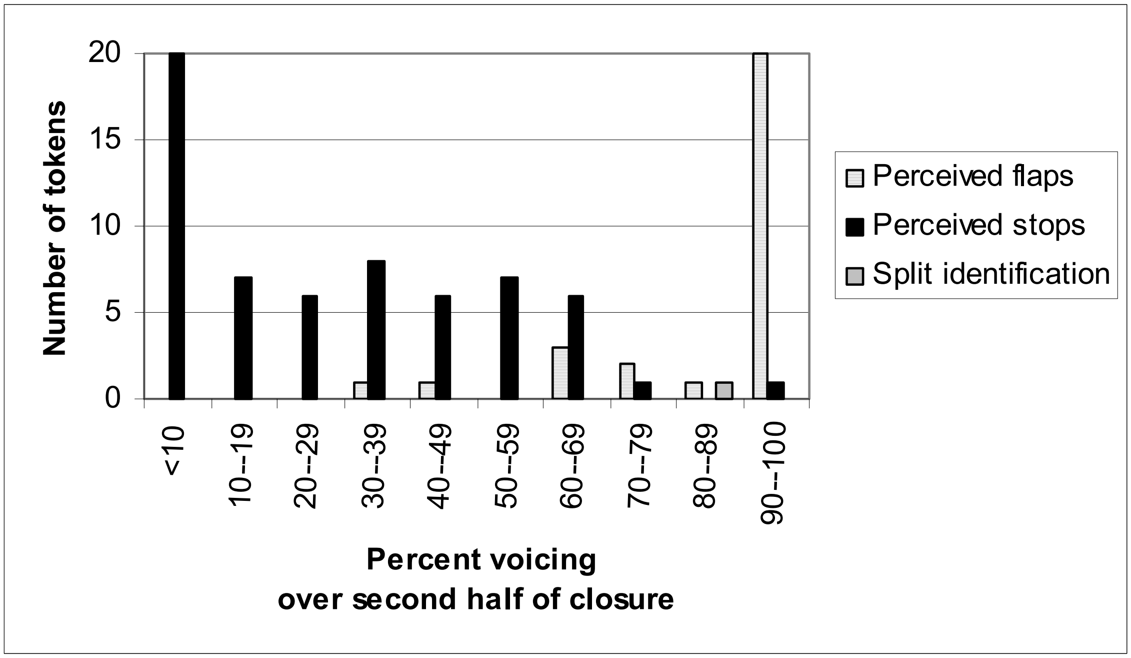

Figure 34: Percent of voicing over second half of /t/ closures for Sp5

As seen in Figure 34 above, the majority of Sp5's perceived flap tokens, 20, are 90-100% voiced, although there are eight remaining tokens that fall within the 30-89% voiced range. Regarding Sp5's perceived stops, even though the <10% category contains more tokens than any other (20 tokens), the majority of stops, 42 tokens, are spread throughout the range of categories, with most falling between 10% and 69% voiced, but with some extending to the 90-100% voiced range. In general, there is a great deal of overlap between Sp5's flap and stop tokens with regard to closure voicing. Four of Sp5's 95 tokens have been excluded due to 0 ms closures.

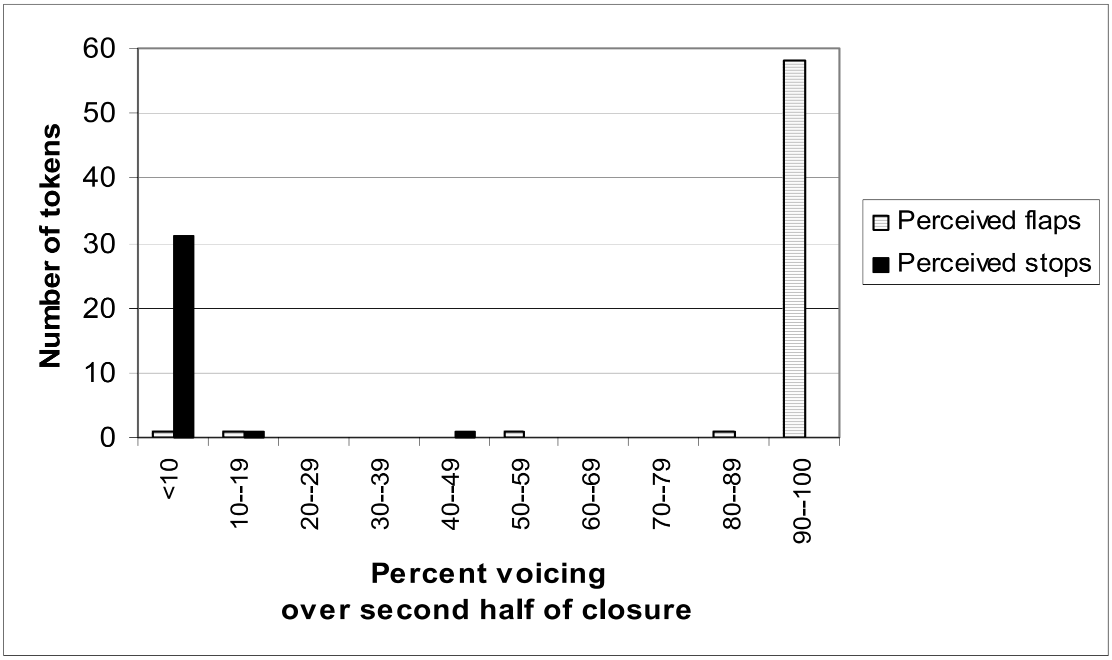

Figure 35: Percent of voicing over second half of /t/ closures for Sp6

As seen in Figure 35 above, the vast majority of Sp6's perceived flap tokens, 58, fall in the 90-100% range, with only four flaps falling into lower categories (0-59% voiced). The perceived stops, on the other hand, are largely less than 10% voiced (31 tokens), although two tokens fall into higher categories (10-49% voiced). In general, Sp6's flap and stop tokens fall into two neat groups, with flaps at or above 50% voiced and stops at 49% voiced or less. The only exceptions are the two flap tokens that fall within the 0-19% voiced range. Three of Sp6's 95 tokens have been excluded from the figure due to 0 ms closures.

5.2.4 Summary

Figure 36 below summarizes how well each of three acoustic measures presented above- duration, VOT and closure voicing, relate to each speaker's perceived flap and stop tokens. "A" represents a perfect relationship (no overlap of tokens), "B" represents a good relationship (an overlap of several tokens), and "C" represents a poor relationship (an overlap of a large number of tokens). These measures are not precise but are meant to characterize the general state of each speaker-cue relationship for the purposes of comparison.

	Sp1	Sp2	Sp3	Sp4	Sp5	Sp6
Duration	C	A	B	C/B[6]	B	B
VOT	C	A	B	B	A	A
Closure voicing	C	B	B	C	C	B

Figure 36: Relationship between the perception of each speaker's flap and stop allophones with duration, VOT and closure voicing. A= perfect relationship, B= good relationship, C= poor relationship.

As seen in Figure 36, all three acoustic cues have a poor relationship with the perception of stops vs. flaps for Sp1. For Sp2, both duration and VOT exhibit a perfect relationship with flap/stop perception, while closure voicing exhibits a good relationship. For Sp3, all three cues have a good, although not a perfect, relationship to flap/stop perception. For Sp4, VOT has a good relationship with flap/stop perception, closure voicing a poor relationship, and duration either a good or a poor relationship, depending upon the duration measure used. For Sp5, VOT has a perfect relationship with flap/stop perception, duration has a good relationship, and closure voicing has a poor relationship. For Sp6, VOT has a perfect relationship with flap/stop perception while both duration and closure voicing have a good relationship. Overall, VOT has the best relationship with flap/stop identification across all speakers while closure voicing has the worst relationship. Duration has a perfect relationship with flap/stop perception for only one speaker, Sp2, and it does not have the single best relationship with flap/stop perception for any speaker.

6. Discussion

6.1 Evidence against paradigm uniformity

The claim regarding paradigm uniformity and American English flapping in Steriade (2000) is that the allophone of /t/ a speaker articulates in the base form of a word will be maintained in the inflected form of a word, even if the phonology predicts a different outcome. More specifically, the expectation is that in the base form of a word, speakers will vary as to whether or not they place secondary stress on the syllable

[6] The two different ratings in Sp4's duration box relate to the two different duration measures applied to her data (see Section 4.3 above). The different measures both yielded a "C" rating for Sp1.

following /t/, and thus vary as to whether or not they produce a flap or a stop. In the inflected form of the word, however, the expectation is that no speaker will place secondary stress on the syllable following /t/. Although the phonology would therefore allow a flap in these words for all speakers, Steriade predicts that a flap will be articulated in the inflected form only if it was produced in the base form, due to paradigm uniformity constraints.

The paradigm uniformity claim in Steriade (2000) crucially depends on the assumption that a single speaker will systematically produce the base form of a word with one particular allophone. If a speaker were to vary his/her pronunciation of a base form, then there would be no standard on with which to judge the paradigm. The study in Steriade (2000) incorporates this assumption; the speakers were asked to read each word only once, and the allophone of /t/ a speaker used in the base form was assumed to be that person's standard allophone. Therefore, the expectation would be that if a speaker articulates a base and inflected pair multiple times, the same allophone of /t/ would surface each time. However, in the present study, where speakers were asked to read each word 12 times, only three of the 24 pairs were such that every repetition of base and inflected form contained the same /t/ allophone. Four other pairs, mentioned in Section 5.1.1, came close, where a speaker articulated the same /t/ allophone in at least ten of 12 repetitions of each of the base and inflected forms. Therefore, on a generous reading, only seven of the 24 pairs are possible candidates for a paradigm uniformity analysis.

However, even the four near-uniform cases, which are all cases of majority flaps, do not provide evidence for a paradigm uniformity condition; they simply do not offer evidence against it. In the case of an inflected form with no secondary stress following the /t/, Steriade (2000) states that she expects a flap to surface as the default, where paradigm uniformity is not at issue, as it does for example for the first /t/ in *meritocratic*. Therefore, by her own account, the appearance of a flap in both the base and inflected forms of a word is unremarkable: both occur as expected due to a lack of stress on the syllable following /t/. It is not necessary to say, in such cases, that the inflected form is retaining a characteristic of the base form. Therefore, the four near-uniform pairs in this study, and the 16 uniform pairs in Steriade (2000) that contain flaps, should not be taken as evidence for or against paradigm uniformity.

The other 17 pairs in this study (and, in fact, the four near-uniform pairs as well) all exhibit variation across the members of a paradigm and, in most cases, within repetitions of a single form. These pairs offer evidence against paradigm uniformity on two counts. First of all, the /t/ allophone in a base is not maintained in the inflected form, thus offering no evidence for the existence of uniform paradigms. Secondly, and perhaps more importantly, the same /t/ allophone is not even maintained throughout the repetitions of a single form. As speakers therefore clearly do not have a standard allophone in the base form, there is no basis for even discussing what the uniform paradigm would look like.

6.1.1 Variation and speech rate

If paradigm uniformity cannot explain the variation found in these data, and if an optional phonological rule cannot inform the issue, the question of what factors are responsible for when a /t/ surfaces as a flap vs. a stop remains perhaps one of more challenging issues regarding American English flapping patterns. Although this study does not directly address or answer this question, an examination of the results does reveal a striking observation: speakers tended to articulate stops during the early portions of the recording session and flaps during the later portions of the recording session.

In the case of four speakers (Sp1, Sp2, Sp3 and Sp4), the vast majority of stops occur during the first half of the repetitions. Excluding the *relative/relativistic* pair which contained largely stops for most speakers, the combined number of stops for these four speakers is 63. Of these 63 stops, 54 tokens (86%), occur in the first six repetitions, leaving only nine (14%), in the last six repetitions. The appearance of stops in the speech of the other two speakers, Sp5 and Sp6, is less predictable. However, it is worth noting that Sp5 pronounced the first repetition of every word as a stop.

An explanation for these statistics may come from initial observations of the flapping rule, that it occurs in informal and fast speech. It is likely that speakers felt more uncomfortable or had more of a desire to enunciate carefully in the beginning of the recording period. By the second half of the recordings, when the task and wordlist were familiar and the speakers felt comfortable, it is likely that they began speaking faster and more casually. Although formality and speech rate were not controlled in this study, a comparison of the recording times of each list for each speaker reveals that speech rate tended to accelerate as the repetitions proceeded for five of the six speakers. Figure 37 below displays the difference in recording time, from the beginning of the first target word on the list to the end of the last target word, between Lists 1 and 23 for each speaker. (List 23 is used in the comparison rather than List 24, since Lists 1 and 23 both contain base forms whereas List 24 contains inflected forms.)

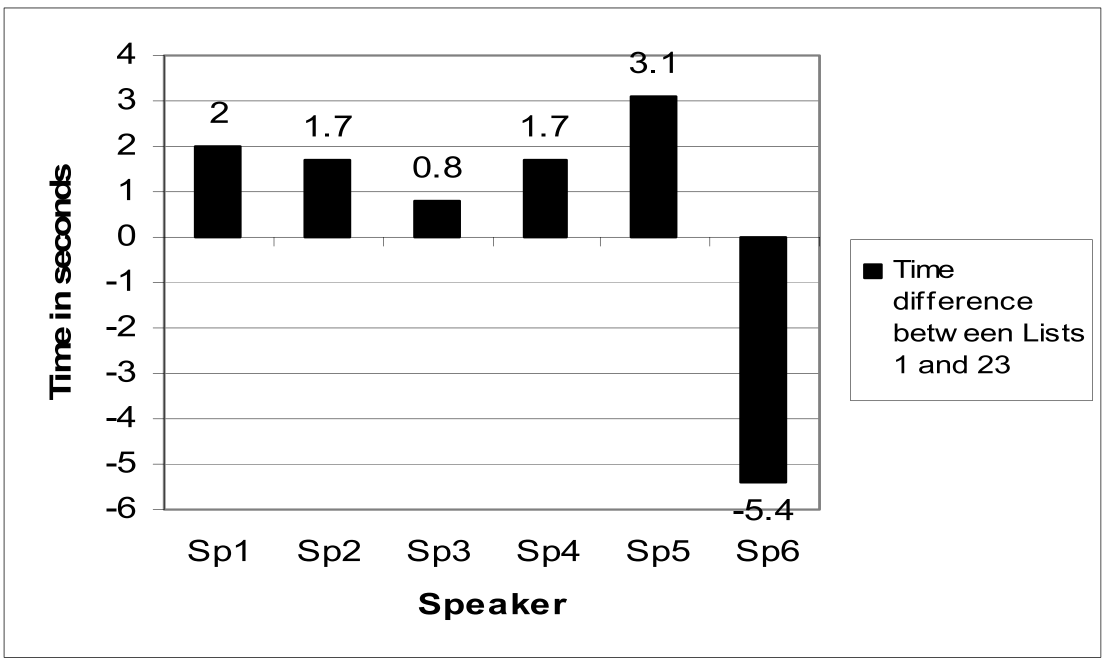

Figure 37: Difference, in seconds, between the total recording time of List 1 and List 23 for each speaker.

As can be seen in Figure 37 above, List 23 was read faster than List 1 for five of the six speakers. The smallest difference is .8 seconds, for Sp3, and the greatest difference is 3.1 seconds, for Sp5. (The typical duration of a word from these lists is 45-65 milliseconds.) While the significance of these data should not be exaggerated since other factors can affect the overall recording time of a list, the data do reflect a general trend of speakers accelerating speech rate as the recording task proceeds. Sp6 is the only speaker whose speech rate is faster for List 1 than for List 23. During the recording session, Sp6 asked, after her first session in the recording booth (reading Lists 1-4), if she was speaking too fast. I reiterated that Sp6 should simply read at whatever rate she considers her normal speaking rate. For the reminder of the lists, Sp6 read somewhat more slowly. Interestingly, Sp6 is also one of only two speakers whose stop tokens are not mainly confined to the first six repetitions. Instead, only four of her nine relevant stops fall in the first six repetitions. A further study that controls for speech rate and formality of speech would likely inform this issue.

These results are corroborated by data from the canonical flap forms mentioned in Section 4.1 above (*fatal, fetish*, etc.). Of the 720 total canonical flaps tokens (five base forms + five inflected forms x 6 speakers x 12 repetitions), I perceived 13 stops (the other perception subjects did not listen to this data). Of these 13 stops, all but two occurred in the first two repetitions of the words, meaning during the speakers' first short session in the recording booth. One of the speakers in Steriade (2000) also produced the first (and only) repetition of *fetish/istic* with stops.

The fact that speakers are much more likely to articulate a stop during the first half of the recordings and very likely to articulate a stop during the first couple of repetitions, may explain why Steriade (2000) found such a high number of stops in the pairs in her data (44 of 60 pairs) where the speakers were asked to repeat each word only once.

The full results of Steriade (2000), displayed in Figure 2 in Section 3.2, are reconsidered in the Appendix.

6.1.2 Appearance of paradigm uniformity in *relative/relativistic*

Of the 24 word pairs in the data, three are strong candidates for a paradigm uniformity analysis (Sp3, Sp5 and Sp6's repetitions of *relative/relativistic*). However, a couple of observations about these pairs suggest that it is not paradigm uniformity, but rather some other factor, that is responsible for the pattern. First of all, all three of these cases are from the *relative/relativistic* pair, suggesting that it is not paradigm uniformity, but rather some property of these words, that is causing the consistent appearance of a [t] phone. If paradigm uniformity were really at work, the pattern should be observed with other word pairs as well. Secondly, in each of these cases, the apparent uniformity arises because the speakers consistently articulate stops rather than flaps across the repetitions of the words. It is not clear, based on paradigm uniformity alone, why this should be the case; consistent articulation of flaps would also be in agreement with a paradigm uniformity analysis, according to Steriade (2000). This observation is even more striking when considering that all of the speakers articulated a majority of stops in this pair, when in the other three pairs, all speakers articulated a majority of flaps, with the exception of Sp5. These facts also contribute to the speculation that there is something about the *relative/relativistic* pair that encourages the production of stops rather than flaps.

One hypothesis for why the speakers in this study tend to produce stops rather than flaps in the *relative/relativistic* pair is that the Obligatory Contour Principle (OCP) (Leben 1973, Goldsmith 1976, McCarthy 1986) is playing a role in determining the /t/ allophone in these words. This pair is the only one where a sonorant consonant, /l/, precedes the /t/. As the flap is often described as a sonorant due to its acoustic properties (Ladefoged 1997), it is possible that the OCP constrains the sequence of two sonorants in these words for some speakers. (Note that the pair *military/militaristic*, first pointed out by Withgott (1983), as mentioned in Section 3.2, also contains an /l/ preceding the /t/). A further study testing this hypothesis is needed. Even if the OCP does not ultimately provide an explanation for the pattern observed in the *relative/relativistic* pair, it is likely that some other property of these words will. In any case, the argument that paradigm uniformity is responsible for the pattern in these word pairs is very weak.

6.2 Acoustic cues: duration, VOT and closure voicing

Since it is already clear that a paradigm uniformity analysis cannot be maintained for this data, the question of what invariant cue, if any, is responsible for distinguishing flaps from stops, is in some sense irrelevant. However, the question of whether it is correct to even assume that a single invariant acoustic cue could consistently separate the two allophones remains. Even though Steriade acknowledges that duration is likely just one of several cues relevant for distinguishing the two allophones from one another, it is crucial for the paradigm uniformity argument that [extra short closure] be an invariant indicator of flap identity, otherwise there is no basis for claiming that this non-categorical feature is subject to paradigm uniformity constraints. If it can be argued that [extra short closure] is not an invariant cue to flap identification, then this is a further argument against paradigm uniformity with phonetic features. The results of this study reveal that, in fact, [extra short closure] does not necessarily characterize the set of perceived flaps to the exclusion of perceived stops and is therefore not an invariant cue.

While the results of this study do reveal that duration is an important cue in the perception of a flap, they also reveal that duration is not an invariant cue. Only one speaker, Sp2, exhibits a complete separation in duration values between perceived stops and flaps, where all flaps are 39 ms or less and all stops are 40 ms or more. On the other hand, two speakers, Sp1 and Sp4, exhibit a great degree of overlap, while the other three speakers, Sp3, Sp5 and Sp6, exhibit a small degree of overlap. This means that for each of these speakers, there are one or more flaps with durations equal to or greater than those of stops. The following spectrograms and pitch tracks from Sp4 illustrate such a case. Although their durations are virtually identical, the first word, *primitive*, was perceived by all four listeners to contain a flap while the second word, *relativistic*, was perceived by all four listeners to contain a stop:

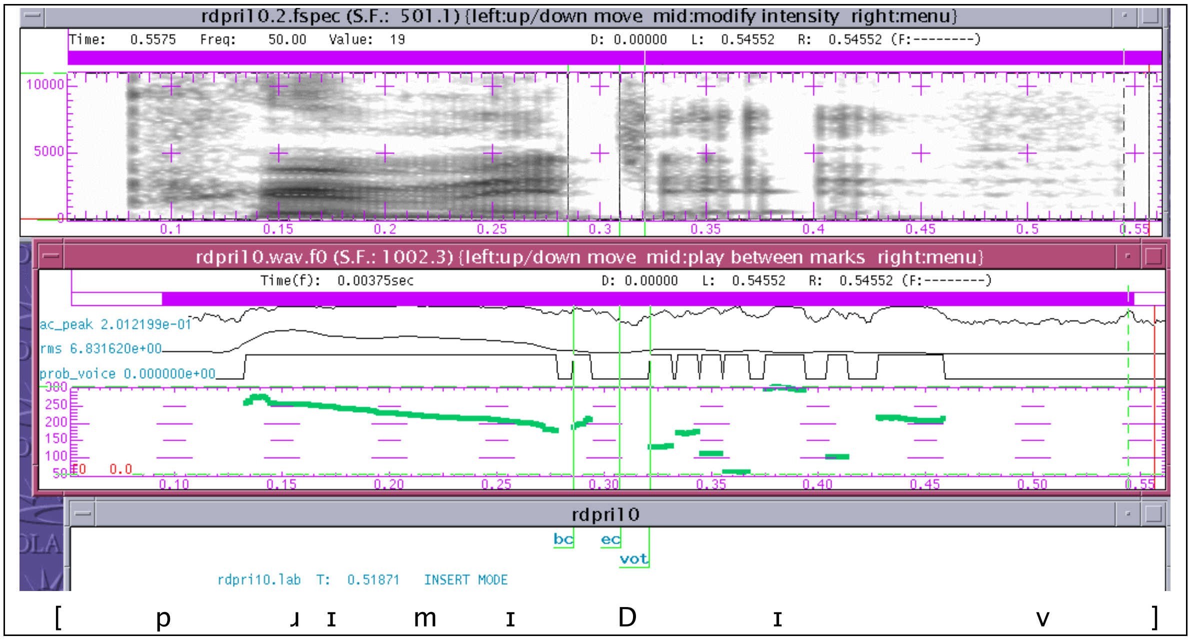

Figure 38: Spectrogram and pitch track of repetition ten of *primitive* by Sp4

When measuring duration of closure + VOT, the total duration of the /t/ in the token of *primitive* in Figure 38 above, perceived by all listeners to be a flap, is 36 milliseconds long. When measuring from beginning of closure to onset of formants, the total duration is 41 milliseconds long.

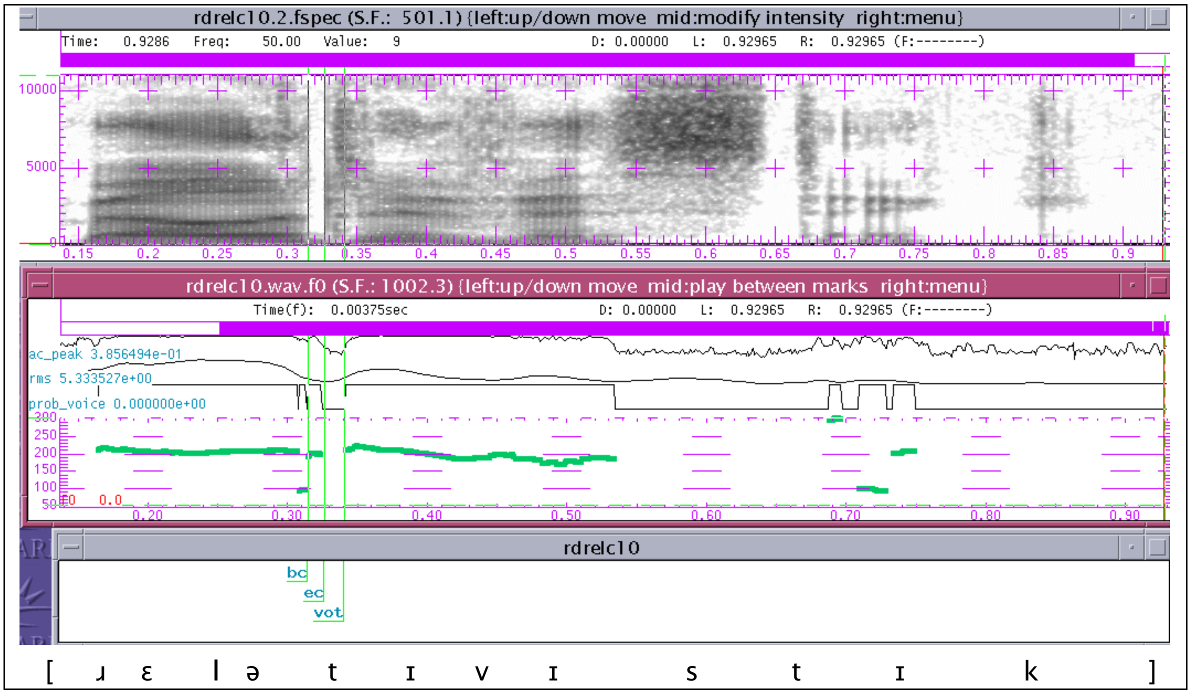

Figure 39: Spectrogram and pitch track of repetition 10 of *relativistic* by Sp4

When measuring duration of closure + VOT, the total duration of the /t/ in the token of *relativistic* in Figure 39 above, perceived by all listeners to be a stop, is 26 milliseconds long. When measuring from beginning of closure to onset of formants, the total duration is 37 milliseconds long. In short, the flap in *primitive* has a greater total duration than the stop in *relativistic,* (37 to 41 milliseconds for the flap, 26 to 37 milliseconds for the stop).

As the results in Section 5.2.1 indicate, and as the above examples illustrate, [extra short closure] does not necessarily capture the difference between flaps and stops, although it is clearly an important cue. Therefore, even if speakers had articulated uniform paradigms, it would not necessarily be correct to conclude that one invariant non-categorical feature, [extra short closure], was responsible for the uniformity.

The results of this study reveal that two other cues, VOT and closure voicing, are also relevant for distinguishing flaps from stops, and it is likely that additional cues not investigated here, such as intensity of aspiration, are also important. As seen in Section 5.2.4, VOT relates to perceived flaps and stops better than or as well as duration for all speakers, while closure voicing relates as well as duration for four speakers. Duration does not rank as the single best cue for any of the speakers in this study. Therefore, it is not accurate to claim that [extra short closure] is an invariant acoustic cue that all flaps necessarily have in common, and therefore it is not an appropriate candidate for evaluating a paradigm uniformity condition.

In fact, none of the three acoustic cues investigated in this study relate perfectly to flap/stop identification across the speakers. Rather, these cues, and perhaps others as well, seem to work in conjunction with one another to produce the end result- a sound that is perceived as either a flap or a stop. This concept of cue weighting, first proposed by Miller and Nicely (1955), may be able to explain some of the patterns found in this data.

As a possible example of cue weighting in these data, consider Sp6's borderline tokens. Figure 21 in section 5.2.1 reveals that this speaker has both a flap token and a stop token which fall in the 60-69 ms duration range. However, these two tokens are clearly distinguished in terms of closure voicing, whereby the flap is 100% voiced and the stop is 0% voiced, and also in terms of VOT, whereby the flap has a VOT of 0 ms and the stop has a VOT of 18 ms. As seen in Figure 35 in Section 5.2.3, Sp6 also has several tokens which overlap in terms of closure voicing. However, even though two of the flaps have closure voicing measures of less than 19%, causing them to pattern like the stops, their other two cues are strongly identifiable as flap cues: both tokens have VOT values of 0 ms and short durations of only 30 ms. Two other tokens in this figure have non-canonical measures, a stop in the 40-49% voiced range and a flap in the 50-59% voiced range. Again, the other cues allow these allophones to be easily identifiable: the token perceived as a flap has a VOT of 0 and a short duration of 33 ms while the token perceived as a stop has a VOT of 56 ms and a long duration of 98 ms. A further study that explicitly examines cue weighting and the perception of flaps, by controlling the various acoustic cues, would be informative.

6.2.1 Split identification tokens

Of the 571 total tokens in the data, five received split identification judgments from the listeners, meaning two identified a token as a stop and two as a flap. The following chart summarizes the characteristics of each of these five tokens with regard to the three acoustic cues examined. In addition, an impressionistic assessment of presence of a high frequency aspiration burst (based on visual examination of the spectrograms) is included.

	Total Duration	VOT	Closure voicing	Burst
Sp1- *relative*	42 ms	0 ms	29%	no
Sp3- *primitive*	38 ms	0 ms	100%	yes
Sp4- *negative*	28 ms	0 ms	57%	yes
Sp4- *positivistic*	33 ms	0 ms	100%	yes
Sp5- *negativistic*	29 ms	0 ms	80%	no

Figure 40: Summary of acoustic attributes of split identification tokens

All five split identification tokens have total durations and VOT values that fall within the expected flap range for each speaker. Additionally, the closure voicing measures for all but two of the tokens (Sp1's *relative* and Sp4's *negative*) fall clearly within the closure voicing measures of each speaker's flaps, with the two more ambiguous closure voicing tokens coming from speakers without a clear relationship between closure voicing and flap/stop identification. In short, it is not immediately clear from these measurements why at least the majority of these tokens would not be perceived by all listeners as flaps. It is possible that the presence of a high frequency aspiration burst in three of the tokens, a likely cue to stop identification not examined in this study, is responsible for the perceptual confusion. In the case of 30 other tokens that did nor receive unanimous perceptual judgments from listeners (3 of 4 agreed), 13 of the tokens have at least one attribute that is not canonical for the category in which they have been placed, while the other 17 seem to be perfect examples of their category.[7] Regardless of the cause of the conflicting judgments, these data corroborate the observation in 5.2.4 above, that a complex of cues, rather than a single invariant cue, is necessary to yield the perception of a flap.

The 35 tokens that do not receive unanimous perceptual judgments from the listeners raise an important question: is it productive to think of flaps and stops as two

[7] Of the 30 tokens that received 3-1 judgments from the listeners (only 3 of 4 agreed on the classification), there are several interesting imbalances in the data. Across the speakers, Sp4's data acquired the most such judgments- 11 of the 30, with the other 19 spread fairly evenly across the remaining speakers; across the word pairs, *positive/istic* acquired the most, with 19 such judgments, the other 11 being spread evenly across the other word pairs; across the listeners, two of the listeners cast the vast majority of the dissenting judgments- 10 for Listener 2 and 16 for Listener 3, while the other two listeners cast only two dissenting judgments each.

categorically distinct phones, or do canonical flaps and canonical stops simply represent opposite ends of the same spectrum? In other words, is flapping simply a gradient process of lenition? This is the perspective supported by the studies discussed in Section 3.3. Despite the presence of some ambiguous tokens in the present study, however, for the vast majority of tokens, 536 of 571, all four listeners agreed on the stop/flap categorization. Even so, a possibility suggested by DeJong (1998:309) based on his articulatory and acoustic study of flaps may apply: “a gradient change in articulatory behavior is giving rise to somewhat quantized acoustic results, which in turn give rise to consistent transcriptions.” The question of whether or not flapping is best viewed as a categorical phonological rule or a gradient process of lenition has yet to be settled. It remains, perhaps, the most intriguing question about the nature of flaps in American English.

6.2.2 Summary of acoustic analyses

The results from the acoustic analyses of total duration reveal that [extra short closure], though a fairly reliable indicator of flap/stop identification, does not necessarily describe the set of flaps to the exclusion of the set of stops and is worse or no better than VOT and/or closure voicing across all of the speakers. Therefore, [extra short closure] is not a good candidate for an invariant non-categorical flap attribute. In fact, none of the acoustic cues explored here can alone characterize the set of perceived flaps to the exclusion of the set of perceived stops. Rather, as the preliminary discussion of cue weighting suggests, the cues work in concert with one another to produce the perception of a flap or a stop. Therefore, if one were to claim that a pattern of flapping appears to exhibit paradigm uniformity effects, it would not be accurate to say that one non-categorical attribute is responsible for the observed effects, rather that a complex of cues work together to achieve the uniformity. Further, if a feature is needed to distinguish a flap from a stop, it would be most reasonable to choose an abstract feature that represents a group of phonetic attributes, such as a “cover feature” as suggested by Stevens’ et. al (1986), rather than a feature intended to represent a single phonetic attribute.

7. Conclusions

The results of this study clearly demonstrate that there is no evidence for paradigm uniformity with regard to flapping in American English. This evidence is twofold. First of all, when asked to repeat the target words multiple times, speakers in this study do not remain consistent in the articulation of a /t/ allophone either across the members of a paradigm or even within repetitions of a single form. Secondly, the results of the acoustic analyses reveal that no single acoustic cue is necessarily an invariant indicator of flap identity; rather, a complex of cues yield the perception of a flap or a stop. Therefore, even if paradigm uniformity were observed in a set of words, it would be inaccurate to claim that the uniformity is based on [extra short closure] or any other single non-categorical attribute. Since a paradigm uniformity analysis cannot be maintained for American English flapping, this phenomenon can also not be used as evidence for a collapse of the phonetic and phonological components of grammar, as argued in Steriade (2000).

References

Banner-Inouye, S. (1995) *Trills, Taps and Stops in Contrast and Variation.* PhD dissertation, UCLA.

Barnes, J. & D. Kavitskaya (2002) Phonetic analogy and schwa deletion in French. Paper presented at Berkley Linguistics Society 28, February 16, 2002.

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

De Jong, K. (1998) Stress-related variation in the articulation of coda alveolar stops: flapping revisited. *Journal of Phonetics* 26: 283-310

De Jong, K. (1995) On the status of redundant features: The case of backness and roundness in American English. In B. Connell & A. Arvaniti (eds.), *Papers in Laboratory Phonology 4* (pp. 68-86) Cambridge: Cambridge University Press.

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

Goldsmith, J. (1976) *Autosegmental phonology.* PhD dissertation, MIT

Gussenhoven, C. (1986) English plosive allophones and amibisyllabicity. *Gramma* 10: 119-141.

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

Kiparsky, P. (1982) Analogical change as a problem for linguistic theory. In *Explanation in Phonology*. *Publications in Language Sciences* 4: 217-236. Dordrecht-Holland: Fortis.

Kiparsky, P. (1979) Metrical structure assignment is cyclic. *Linguistic Inquiry* 10.3: 421-441.

Ladefoged, P. (1997) Linguistic phonetic descriptions. In W. Hardcastle and J. Laver (eds.) *The Handbook of Phonetic Sciences* (pp. 589-618) Oxford: Blackwell Publishers.

Leben, W. (1973) *Suprasegmental phonology*. PhD. dissertation, MIT

McArthy, J. (2001) Optimal paradigms. ROA-485

McCarthy, J. (1986) OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17: 207-263

Miller, G.A. & Nicely, P.E. (1955) Analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27: 338-53.

Riehl, A. (2003) American English Flapping: Evidence Against Paradigm Uniformity with Phonetic Features. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*. (pp. 2753-2756)

Selkirk, E. (1982) The syllable. In H. van der Hulst and N. Smith (eds.) *The Structure of Phonological Representations (Part 2)* (pp. 337-383) Dordrecht-Holland: Foris

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

Steriade, D. (1996) Paradigm Uniformity and the Phonetics-Phonology Boundary. Paper presented at the 5th Conference on Laboratory Phonology, Northwestern University, Evanston, Illinois

Stevens, K. N., S.J. Keyser & H. Kawasaki. (1986) Toward a phonetic and phonological theory of redundant features. In J.S. Perkell & D.H. Klatt (eds), *Invariance and Variability in Speech Processes* (pp. 426-449) Hillsdale, NJ: Lawrence Erlbaum

Stone, M. & S. Hamlet. (1982) Variation in jaw and tongue gestures observed during the production of unstressed /d/s and flaps. *Journal of Phonetics* 10: 819-415

Turk, A. (1992) The American English flapping rule and the effect of stress on stop consonant duration. *Cornell Working of the Phonetics Lab* 7: 103-133

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

Williamson, K. (1977) Multi-valued features for consonants. *Language* 53: 843-871

Withgott, M. (1983) *Segmental evidence for phonological constituents.* PhD dissertation, University of Texas, Austin.

Zue, V. W. & Laferriere (1979) Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America* 66.4: 1039-1050.

Appendix: A reconsideration of the data in Steriade (2000)

Figure 2 in Section 3.2 presents the results of the study in Steriade (2000) (adapted from her Table 17 in the 1996 version of the paper). Steriade argues that 59 of the 60 pairs in this table support a paradigm uniformity analysis. In light of the findings in the present study, however, it is possible to cast doubt on this argument for 56 of the pairs, as explained below in accordance with the different types of shading in the boxes.

Speaker →	1, 2, 3, 4, 5, 6, 7	8, 9	10, 11	12
voluntary	**t**	**t**	**t**	**t**
voluntaristic	**t**	**t**	**t**	**t**
positive	**t**	**t**	**D**	**t**
positivistic	**t**	**t**	**D**	**t**
primitive	**t**	**D**	**D**	**D**
primitivistic	**t**	**D**	**D**	**t**
relative	**t**	**D**	**D**	**D**
relativistic	**t**	**D**	**D**	**D**
negative	**t**	**D**	**D**	**t**
negativistic	**t**	**D**	**D**	**t**

Figure A (=Figure 2): Results of study in Steriade (2000), reconsidered.

<u>Dark gray box</u> (1 pair)- Non-uniform pair.

<u>Light gray boxes</u> (12 pairs)- These boxes reveal that all speakers produced only stops in the *voluntary/istic* pair. It is likely that the speakers were not able to produce a flap in the

environment following /n/. No speakers in the present study produced a flap in this pair. Further, Zue and Laffereie's (1979) study suggests that the likelihood of flapping in this environment is very low, as discussed in Section 2.2. In the environment following /n/ after a stressed vowel, the probability of occurrence of a flap is only .14. Although they did not examine /t/ following /n/ between unstressed vowels (as in the *voluntary/istic* pair), it is likely that the probability of occurrence would be much lower than .14, since they find that flapping after a stressed vowel is much more common (.99) than flapping between unstressed vowels (.33) It is therefore not surprising that none of the speakers in Steriade's (2000) study produced a flap in either member of this pair due to the environment itself.

<u>Horizontal-line boxes</u> (15 pairs)- These are pairs were both members of the paradigm contain flaps. As discussed in Section 6.1, although these pairs do not offer evidence against paradigm uniformity, they do not offer evidence for it, based upon the assumption in Steriade (2000) that speakers will flap in the inflected form as the phonological default.

<u>Diagonal-line boxes</u> (28 pairs, excluding *voluntary/istic*)- These boxes indicate that seven of the speakers produced only stops in all pairs. As discussed in Section 6.1.1, speakers in the present study had a strong tendency to produce stops during the first repetition of each word. This pattern is hypothesized to be related to speech rate and level of formality. It is therefore likely, or at least possible, that in the case of these seven speakers who produced *only* stops, they did so because of the nature of the task, not because of paradigm uniformity constraints.

<u>White boxes</u> (4 pairs)- These boxes indicate pairs containing stops by speakers who otherwise produced mainly flaps. These pairs provide the best argument for a paradigm uniformity condition in the data in Steriade (2000). The fact that these speakers articulate mainly flaps across the other pairs indicates that these speakers were likely not producing stops artificially due to the nature of the task; therefore, the fact that the speakers do

produce stops in both members of these pairs makes the pairs potential candidates for a paradigm uniformity analysis. However, since these pairs were not subject to multiple repetitions (found to reveal great variability in the present study), it is not clear how significant these results are.

To summarize, 56 of the pairs may well not result from a paradigm uniformity condition, while four arguably do. While the 56 disputed pairs are not inconsistent with such an analysis, they at least raise other possibilities that must be addressed before any conclusions can be drawn. Likewise, there may be other possible analyses of the remaining four pairs, which seem to offer the best evidence for a paradigm uniformity analysis.