Journal of Phonetics 38 (2010) 504–516

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

An acoustic and perceptual analysis of /t/ and /d/ flaps in American English

Wendy Herd *, Allard Jongman, Joan Sereno

Linguistics Department, University of Kansas, Lawrence, KS 66044, USA

ARTICLE INFO

Article history:
Received 11 December 2008
Received in revised form
4 June 2010
Accepted 25 June 2010

ABSTRACT

This paper presents an acoustic and perceptual study of alveolar flaps in American English. In the acoustic study, vowel duration differences in disyllabic tokens replicated previous findings in that vowels preceding /d/ were significantly longer than those preceding /t/. Flap frequency was also analyzed based on a method of distinguishing flapped from unflapped stops on a speaker-by-speaker basis. It was discovered that females flapped more often than males and that participants were more likely to flap when they were less aware of the contrast between /t/ and /d/. Contrary to past research, neither word frequency nor morphological complexity affected flap frequency in the present study.

In the perceptual study, four naturally produced word pairs were used to manipulate underlying representation (/t/ or /d/), vowel duration preceding the flap, and word frequency. Vowel duration alone did not predict the listeners' perception of flapped /t/ and /d/; word frequency, where high frequency words were identified correctly more often than low frequency words, and a d-bias, where /d/ flaps were identified correctly more often than /t/ flaps, did prove significant. Unlike previous research, this study uses nonarbitrary values to distinguish flapped from unflapped tokens and draws connections between the acoustic and perceptual results.

1. Introduction

In American English, when a /t/ or /d/ occurs between a stressed and unstressed syllable, it often surfaces as an alveolar flap [ɾ], creating an environment where /t/ and /d/ might be neutralized. According to the previous research of Lehman and Heffner (1940) and Chen (1970), who examined monosyllabic words contrasting voiced and voiceless consonants in final position, vowel duration is significantly longer preceding voiced segments than voiceless segments and consonant duration of voiceless segments is significantly longer than that of voiced segments. This led researchers to expect vowel duration and/or consonant duration differences to aid in the differentiation of /t/ and /d/ when flapped in disyllabic words, avoiding complete neutralization of the two segments. In the past forty years, several acoustic and perceptual studies have been carried out in order to investigate whether /t/ and /d/ are neutralized when flapped; however, these studies fall short because no clear distinctions have been drawn between the acoustic nature of flaps and the acoustic nature of flapped tokens listeners perceive as /t/ or /d/.

1.1. *Acoustic analysis of flaps*

Sharf (1962) who analyzed the production of monosyllabic words ending in /t/ and /d/, like 'cat' and 'cad', and disyllabic words with medial /t/ and /d/ in flapping environments, 'catty' and 'caddy', found that vowels were 91 ms longer on average before word-final /d/ than /t/ in monosyllabic words but that there was only a 9 ms difference between vowels preceding /d/ and those preceding /t/ in disyllabic flapped words. In contrast to Chen's (1970) findings, where the consonant durations of voiceless segments were longer than their voiced counterparts, Sharf (1962) found that the average closure duration of flapped /d/ (30 ms) was only 4 ms longer than that of flapped /t/ (26 ms). Similarly, Fox and Terbeek (1977) and Zue and Laferriere (1979) found that although /t/ and /d/ appear to neutralize when flapped, vowels preceding /d/ flaps were significantly longer than those preceding /t/ flaps, but the closure durations of /d/ flaps and /t/ flaps were not significantly different. In studies measuring how stress affects the duration of English stops, various researchers have reported that post-stress /t/ and /d/ occur with significantly shorter durations than their pre-stress counterparts, reporting durations similar to those found in the above-mentioned studies. Stathopoulos and Weismer (1983) found that when /t/ occurred in a flapping environment, it measured 41 ms on average, while /d/ measured 44 ms, and Lavoie (2000) reported /t/ measured 34 ms on average while /d/ measured 37 ms; however, Turk (1992) found shorter durations with /t/ measuring 22 ms and /d/ measuring 18 ms. More recently, Patterson and Connine (2001) searched the SWITCHBOARD corpus for minimal pairs like 'putting' and 'pudding' in which /t/ and /d/ are flapped. Like previous researchers, Patterson and Connine (2001) found that average duration of vowels preceding /d/ flaps was 16 ms longer than that of vowels preceding /t/ flaps, a statistically significant

* Corresponding author. Tel.: +1 785 864 3450; fax: +1 785 864 5724.
E-mail addresses: wenherd@ku.edu, zacwendy@sbcglobal.net (W. Herd).

0095-4470/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2010.06.003

difference. In addition to replicating the results of past research, they noted that flapping occurred significantly less often when the tokens were low frequency words and when the tokens were morphologically complex, i.e., contained two morphemes as in 'beating'.

While Sharf (1962), Fox and Terbeek (1977), Zue and Laferriere (1979), and Patterson and Connine (2001) document vowel length differences that acoustically distinguish an underlying /t/ flap in a word like 'writer' [ˈɹaɪɾɚ] from 'rider' [ˈɹaɪːɾɚ] with an underlying /d/, they report inconsistent consonant duration differences and make no attempt to look at a specific dialect of American English. The one speaker in the Sharf (1962) study was from Detroit, but Fox and Terbeek (1977) and Zue and Laferriere (1979) refer to their participants only as speakers of an American English dialect. In the Patterson and Connine (2001) study, the minimal pairs were drawn from the SWITCHBOARD corpus without restriction to one regional dialect, and with individual tokens within minimal pairs produced by different speakers, where vowel duration differences could be due in part to speaker variation. Additionally, these studies included no perceptual tests to evaluate how salient the observed vowel differences are to listeners.

1.2. Perceptual studies of flaps

One of the earliest perceptual studies was conducted by Sharf (1960), who played recordings of a male and a female speaker reading tokens with /t/ and /d/ in flapping environments to 12 native speakers of American English. Listening participants were instructed to check the word they thought they heard on a numbered list of target words with minimal pair counterparts, i.e., *catty* and *caddy*. Listeners were able to correctly perceive the female's productions 86.5% of the time and the male's productions 64% of the time; however, Sharf (1960) did not describe how it was determined if a speaker were producing an alveolar stop or a flap, but he noted that the female speaker "used a [t] sound" rather than a flap, which would account for the difference in accuracy between the perception of the female speaker's and the male speaker's productions (107). Malécot and Lloyd (1968) also administered a perception test in which they randomized minimal pairs with medial /t/ and /d/ produced by five native speakers of an Eastern American dialect. Fifty native English-speaking participants took the test which required them to circle the correct spelling of the word they heard, so participants would hear [ˈkɪɾi] and then circle either 'kitty' or 'kiddie', a similar format to that used by Sharf (1960). The listeners performed near chance level, averaging 56.6%, and researchers found that listeners had a general /d/ bias, probably due to the voiced quality of flaps. After noticing the /d/ bias in listeners' responses, Malécot and Lloyd (1968) visually examined the spectrograms of the minimal pairs, looking for a correlation between vowel length and percent of /d/ responses by listeners. Although no vowel durations were reported, they did report a strong positive correlation between the length of monophthong vowels and /d/ judgments; however, no such correlation was found for diphthongs.

These two early perceptual studies both fall short, Sharf (1960) in its use of only two speakers, one of whom did not produce /t/ as a flap, and both studies in their lack of definitive results as to listeners' abilities to distinguish /d/ and /t/ when flapped. While the findings of Malécot and Lloyd (1968) indicate that listeners may have used monophthong vowel length as a cue for their judgments, they did not confirm that the speakers made significant or consistent vowel length differences when producing words with /d/ and /t/ medially. The last group of studies discussed combines both acoustic analyses and perception tests to investigate whether /t/ and /d/ are neutralized when flapped.

1.3. Acoustic and perceptual studies of flaps

Fisher and Hirsh (1976) completed an early acoustic and perceptual study investigating whether /t/ and /d/ were neutralized when flapped. The acoustic portion of this study replicated the findings of Sharf (1962) and Zue and Laferriere (1979) with vowels preceding /d/ measuring significantly longer than those preceding /t/. They also found that flapped /t/ measured longer than flapped /d/, but not significantly so. The perceptual portion of the study was based on the transcriptions of six phonetically trained judges, who Fisher and Hirsh (1976) reported could reliably distinguish between flapped /d/ and flapped /t/. Unfortunately, the acoustic results of this study are based on only a small number of speakers (six), with the vowel length data based on only half of those participants, and the perceptual data are based on the judgments of phonetically trained judges, not telling us whether the difference between flapped /t/ and flapped /d/ is salient to naïve speakers of American English.

Charles-Luce (1997) conducted an acoustic and perceptual study that examined how the manipulation of listener-present/listener-absent and biased passage/neutral passage conditions affected the /t/ and /d/ contrast in flapping environments. Based on measurements of closure duration, first vowel duration, and overall word duration, Charles-Luce (1997) first found, like past researchers, that closure duration was never a significant indicator of voicing. Vowel duration differences, on the other hand, significantly predicted /t/ and /d/ flaps only in the biasing passages and not in the neutral context, indicating that speakers maintained the voicing difference when the target word was predictable. No significant vowel duration or word duration differences were found independent of the two conditions: biased/neutral passage and listener-present/listener-absent. The perceptual study yielded more surprising results in that listeners perceived /t/ and /d/ correctly 81% of the time, 80% for listener-absent and 82% for listener-present, a surprisingly high percentage correct when compared to the results of Sharf (1960) and Malécot and Lloyd (1968).

In similar studies of semantic predictability, Charles-Luce and Dressler (1999) and Charles-Luce, Dressler, and Ragonese (1999) returned to the question of how semantic predictability affected the production and perception of /t/ and /d/ when flapped. The studies found that children near the age of seven and older adults tended to use closure duration to indicate voicing, apparently hyperarticulating to the point that they were no longer producing /t/ and /d/ as flaps. During the perceptual portion of the studies, the hyperarticulated productions of the children near age seven and the older adults were unsurprisingly identified more accurately than those of the other children and the young adults. More interesting, the percentage of tokens perceived correctly ranged from 56% to 73% across all groups. The inclusion of hyperaticulated /t/ and /d/, where /t/ and /d/ are not flapped, could account for the different perceptual results reported in Sharf (1960), Malécot and Lloyd (1968), Charles-Luce (1997), and Charles-Luce et al. (1999).

1.4. Motivation for the present study

The acoustic and perceptual studies of flaps conducted over the past forty years leave questions that have yet to be answered. While acoustic studies have consistently shown that the closure duration of /t/ and /d/ when flapped is not an accurate cue to voicing, studies vary over the role of preceding vowel duration as a cue to voicing. Earlier studies documented a 7–16 ms difference in vowel duration, where vowels were longer preceding /d/ flaps, but the more recent Charles-Luce (1997) study found that neither vowel duration nor word duration were statistically significant independent cues. Furthermore, researchers disagree as to what constitutes a flap, using various arbitrary measures, like 50, 10–40 ms, or half the duration of an average [t] or [d]. Similarly, the perceptual studies have varied dramatically. While some studies, like Malécot and Lloyd (1968), have reported a percentage of tokens perceived correctly very near chance levels, others, like Charles-Luce (1997), have documented considerably higher levels of accuracy.

In addition to the question of whether /t/ and /d/ are acoustically or phonetically neutralized in flapping, the studies have not considered that regional dialect differences may also have contributed to the varying results. No one has set out to study the production and perception of vowel and closure duration differences in /t/ and /d/ flaps within one regional dialect. As a portion of their study, Patterson and Connine (2001) compared the frequency of flapping between the speakers of the Northern dialect and the North Midland dialect of American English. Although they and Byrd (1994) predicted that no variation in flapping would occur across dialects, both studies reported differences in flap frequency. Byrd (1994) found that American English speakers from the North and the Northeast and speakers from the North Midland produced no significant variation in flapping the word ‘water’, but speakers from the North and the Northeast were significantly less likely to flap across morpheme-boundaries in the phrase ‘suit in’ than their North Midland counterparts. Like Byrd (1994), Patterson and Connine (2001) reported that speakers of the Northern dialect of American English produced flaps less often, with 74.8% of tokens being flapped, than speakers of the North Midland dialect, who produced only 81.6% of flappable tokens as flaps. Since past research indicates that flaps can differ in frequency across dialect regions, it is also possible that the vowel durations preceding those flaps and the durations of the flaps themselves may differ as a function of dialect region.

This paper combines an acoustic and perceptual study in order to investigate whether /t/ and /d/ have neutralized in flapping environments for 20 speakers (10 male, 10 female) of a North Midland dialect of American English. The speakers produced both monosyllabic minimal pairs ending in /t/ and /d/ and disyllabic minimal pairs with medial /t/ and /d/ in a flapping environment. These stimuli were balanced in terms of lexical status, vowel quality, word frequency and morphological complexity, and they were read in three different contexts: pseudorandomized list, carrier sentence, and minimal pairs. The data provide a method of distinguishing flapped from unflapped stops on a speaker-by-speaker basis.

Based on previous research, average vowel duration differences before flapped /t/ and /d/ for this group of speakers should be near the range 7–16 ms. As in previous research, no difference in the duration of the flaps themselves based on whether they were underlying /t/ or /d/ is expected. It is also predicted that speakers will be less likely to distinguish flapped /t/ and /d/ through either vowel duration or consonant duration differences in the pseudorandomized list than in the other two contexts, because they will be the least aware of the contrast. Alternately, speakers are predicted to be most likely to distinguish flapped /t/ and /d/ through duration differences or hyperarticulation in the minimal pairs context, because they will be the most aware of the contrast. Finally, in spite of vowel length differences, it is anticipated that listeners who hear a word with a medial flap will perceive underlying /t/ and /d/ near chance levels, indicating that the contrast between /t/ and /d/ may be measurable when one uses acoustic instruments but that the contrast has neutralized in natural speech. If listening participants perform significantly above chance levels, it is predicted that those elevated results will be the effect of a d-bias in the listeners' responses due to the voiced nature of flaps and of word frequency, where listeners will choose the more frequent word when faced with a choice of a high frequency or low frequency candidate.

2. Experiment 1: acoustic study

2.1. Method

2.1.1. Participants

Thirty native American English speakers from the University of Kansas participated in the acoustic portion of this study. Of these speakers, five males and three females were excluded from analyses, because they reported living more than 50 miles outside of Kansas on the dialect survey, which was administered to all participants. Two other female speakers, the last two recorded, were also excluded prior to analyzing the data in order to balance gender within the participant groups. Analyses discussed in this study are therefore based on 20 speakers, 10 females and 10 males, who were born in or within 10 miles of Kansas and who had not lived outside of the area for more than one year. All participants were recorded reading three different lists of words and nonwords containing word-final and word-medial /t/ and /d/.

2.1.2. Stimuli

The stimulus lists contained 28 disyllabic minimal pairs contrasting /t/ and /d/ in flapping environments, half of these with diphthongs as the first vowel and half with monophthongs. Based on Sharf (1960) and Malécot and Lloyd (1968), diphthongs may not show the same vowel duration differences as monophthongs. Additionally, the lists included 26 monosyllabic minimal pairs contrasting /t/ and /d/ word-finally. The monosyllabic pairs were counterparts to the disyllabic minimal pairs, so the disyllabic pairs ‘tidal’–‘title’ and ‘tidy’–‘tighty’ had the corresponding pair ‘tide’–‘tight’. Because two of the monosyllabic pairs, ‘tide’–‘tight’ and ‘wide’–‘white’, matched up with two disyllabic pairs each, there were 28 disyllabic minimal pairs but only 26 monosyllabic pairs as shown in Table 1. When no lexical monosyllabic counterparts were available for disyllabic minimal pairs, for example in the minimal pair ‘pouter’–‘powder’, a near minimal pair was used, ‘pout’–‘cowed’. The monosyllabic minimal pairs allowed the researchers to measure an average /t/ and /d/ duration for each speaker, enabling them to better define what would be considered a flap for each individual speaker.

Additionally, 28 disyllabic words that do not exist in minimal word pairs were used along with 28 nonword counterparts, for example ‘shedding’ and ‘shetting’. Of the 28 disyllabic nonword minimal pairs, the medial consonant was /d/ in half of the lexical items and /t/ in the other half.

To sum up, the stimulus lists include 28 disyllabic word pairs (56) and 28 disyllabic word–nonword pairs (56), so 112 disyllabic words and nonwords were read in three contexts ($112 \times 3 = 336$) by 20 speakers for a total of 6720 disyllabic tokens. Additionally, 26 monosyllabic word pairs (52) were read in three contexts ($52 \times 3 = 156$) by 20 speakers for a total of 3120 monosyllabic tokens. A total of 9840 monosyllabic and disyllabic tokens were recorded. Analysis of the speakers' flap frequency, vowel duration, and flap duration of these word–nonword minimal pairs should shed light on how those measures interact with the underlying representation of the flap (/t/ or /d/) and the lexical status of the stimuli.

In addition to balancing monophthongs and diphthongs in the word list, vowel quality, word frequency, and morphological

Table 1
Lexical monosyllabic and disyllabic minimal pairs.

Vowel	Word final /d/	Word final /t/	Intervocalic /d/	Intervocalic /t/
aɪ	ride	write	rider	writer
	tide	tight	tidal	title
	wide	white	widest	whitest
	bide	bite	biding	biting
	side	sight	sided	sighted
	tide	tight	tidy	tighty
	wide	white	wide-ish	white-ish
eɪ	grade	grate	grader	grater
	raid	rate	raiding	rating
	fade	fate	faded	fated
aʊ	cowed	pout	powder	pouter
	cloud	clout	clouded	clouted
oʊ	bode	boat	boded	boated
	code	coat	coding	coating
i	Swede	sweet	Swedish	sweetish
	lead	cleat	leader	liter
ɪ	kid	kit	kiddie	kitty
	bid	bit	bidder	bitter
ɛ	bed	pet	pedal	petal
	wed	wet	wedding	wetting
æ	mad	mat	madder	matter
	pad	pat	padding	patting
ʌ	bud	but	budding	butting
	thud	shut	shudder	shutter
u	tude	toot	tudor	tutor
	rude	root	rudy	rooty
ɑ	odd	ought	odder	otter
	plod	plot	plodding	plotting

complexity were also considered. An attempt was made to include a variety of vowels and suffixes. Using the Francis-Kučera (1982) written word frequencies for the disyllabic lexical items, a paired samples *t*-test conducted comparing words with underlying /t/ ($\bar{x}=33.00$, s.d.=85.5) and /d/ ($\bar{x}=15.24$, s.d.=34.2) found that the two lists did not significantly differ in word frequency [$t(41)=-1.214$, $p=0.23$]. Likewise, when another paired samples *t*-test was performed comparing the number of morphemes in words containing /t/ ($\bar{x}=1.64$, s.d.=0.49) and /d/ ($\bar{x}=1.74$, s.d.=0.45), the number of morphemes in the two lists did not differ significantly [$t(41)=1.071$, $p=0.29$].

2.1.3. Procedure

Participants read the above-mentioned tokens in three different contexts. In the first context, the speakers read a list consisting of the 164 target words and nonwords pseudorandomized with 51 fillers. During context two, speakers read the same list with the words and nonwords inserted into the carrier sentence, 'Please say _______ again.' For the third context, speakers read a paired flap list, in which the 164 tokens were listed as minimal pairs and were pseudorandomized with 20 fillers. Half of the participants, five males and five females, read the paired flap list with /t/ tokens first and half, five males and five females, with /d/ tokens first. Participants were instructed to speak as naturally as possible. The lists were constructed in such a way that speakers would be unaware of the goal of the study in context one's word list and context two's carrier list, resulting in natural speech, but the speakers would be focused on the difference between minimal pairs in the paired flap list (context three), possibly resulting in more salient differences between /t/ and /d/.

Speakers were recorded in an anechoic chamber via an Electro-Voice RE-20 microphone at the University of Kansas. Nineteen speakers were recorded onto a flash card using a Marantz Portable Solid State Recorder (PMD 671) at a sampling rate of 22.05 kHz. One speaker was recorded on DAT tape and later digitized using PRAAT software at a sampling rate of 22.05 kHz with a low-pass filter setting of 10 kHz (Boersma & Weenink, 2005).

2.2. Measurements

Five acoustic measurements were taken for each disyllabic word or nonword using PRAAT software: word duration, first vowel duration, a ratio of first vowel duration to word duration, a ratio of first vowel to second vowel duration, and /t/ or /d/ duration (Boersma & Weenink, 2005). In addition, word final /t/ or /d/ duration was measured for all monosyllabic words. Fig. 1 illustrates where these measurements were made.

For tokens read in isolation or in pairs, word duration was measured from the onset to the offset of the visible waveform and spectrogram, and for tokens read in the carrier sentence, word length was measured between changes in the cyclic nature of the waveform and changes in the spectrogram at word boundaries. The first vowel and the second vowel for disyllabic words and nonwords were measured from the onset of the first formant to the offset of the second as seen in spectrograms. For disyllabic items, the /t/ or /d/ duration was measured from the offset of the first vowel to the onset of the second. The /t/ or /d/ duration of monosyllabic tokens was measured from the offset of the first vowel to the end of the visible waveform and the end of the stimuli in the spectrogram when items were read in isolation or in pairs. When read in the carrier sentence, the end of the consonant duration for monosyllabic tokens was measured as the change in the cyclic nature of the waveform and a change in the spectrogram at the word boundary.

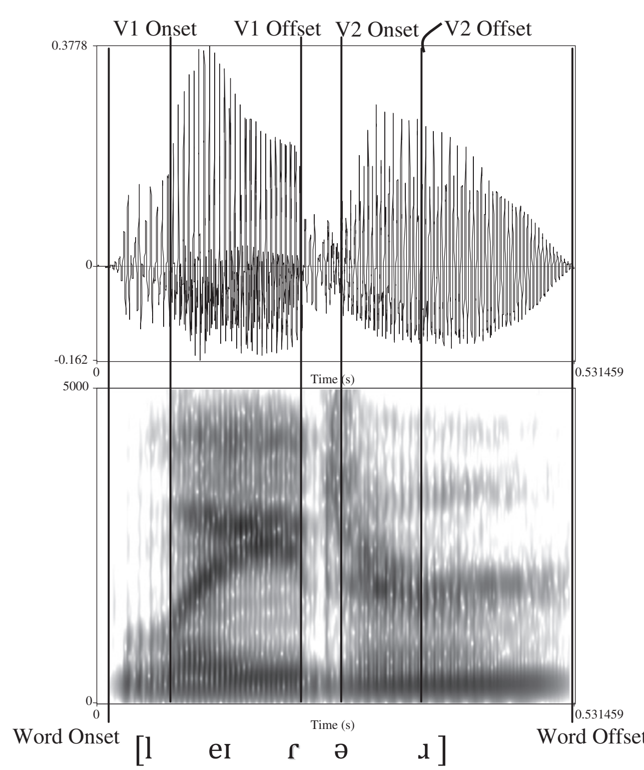

Fig. 1. Acoustic measurements in the word 'later'.

2.3. Results

2.3.1. Defining flaps

As mentioned earlier, previous studies used one value, usually a number between 30 and 50 ms, as the boundary between alveolar flaps and stops for all speakers, a method which fails to take speaker variability into account. Instead of using an arbitrary number to define flaps across all speakers, each speaker's productions were analyzed separately in this study. First, the target consonant durations of all of a single speaker's monosyllabic and disyllabic tokens (156 monosyllabic words and 336 disyllabic words and nonwords) were graphed, as shown in Fig. 2, resulting in a bimodal distribution. Using duration to distinguish flapped from unflapped tokens may raise concerns that /d/-tokens would be classified as flaps more often than /t/ due to the shorter duration of voiced consonants; however, Fig. 2 illustrates that /t/ and /d/-tokens classified as flaps overlap in duration. Next, based on a gap in the distribution between flapped and fully articulated stops, a value distinguishing the two was assigned to each speaker. As can be seen in Fig. 2, the distribution for speaker 3, a clear boundary falls between 52 and 62 ms, so all tokens below 52 ms were considered flapped and all above were considered unflapped for this speaker. For all speakers, there was at least one empty bin separating flapped from unflapped stops. Table 2 lists the value at which flaps were defined for each of the 20 speakers. The average flap cut-off value across all 20 speakers was 56.25 ms, with a range of 43–69 ms. It should be noted that the average flap duration for each speaker ranged from 24 to 41 ms, values very similar to those previously documented.

Fig. 3a–c illustrates the distribution of /t/ and /d/ tokens with the flap cut-off marked by a vertical line for Speakers 10, 15, and 19. The flap cut-off values for these speakers represent the low end (43 ms), mean (55 ms), and high end (69 ms) of the flap cut-off range.

Since all /t/ and /d/ monosyllables were measured from the end of the vowel to the end of the stimulus in contexts 1 and 3 and from the end of the vowel to the onset of the following word ('again') in context 2, the duration of unflapped /t/ varied by context as shown in Fig. 4a–c. While the duration varied, it is important to note that the gap between flapped /t/ and unflapped /t/ was visible in all contexts.

Flaps were defined individually using duration, because this has long been the dependent variable measured and used to

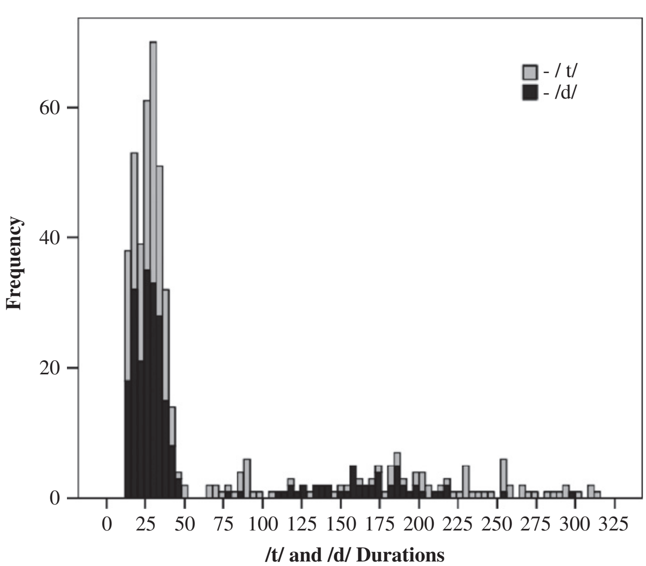

Fig. 2. Distribution of all monosyllabic and disyllabic /t/ and /d/ tokens for speaker 3.

Table 2
Value below which each speaker's /t/ and /d/ productions were considered flaps and average duration of each speaker's flaps.

Speaker number	Flap cut-off value (ms)	Average flap duration (ms)
1	57	32
2	61	35
3	52	37
4	61	40
5	61	33
6	58	27
7	53	27
8	43	24
9	57	30
10	43	25
11	60	35
12	60	35
13	53	31
14	60	29
15	69	41
16	47	27
17	57	27
18	59	35
19	55	28
20	59	32

distinguish between flaps and their longer unflapped counterparts. However, as Lavoie (2000) pointed out, since flaps and stops differ in sonority, amplitude measures of these segments should differ also. In her study, Lavoie (2000) found that RMS amplitude measures were higher for flaps than stops. Using a similar method, we measured the raw amplitude of the closure of /t/ and /d/ in five monosyllabic pairs, like 'mat' and 'mad', and in five disyllabic pairs, like 'matter' and 'madder' for each speaker. The tokens from context 2 (sentential context) were used, because this ensured that the word-final /t/ and /d/ segments would occur preceding a vowel, an environment similar to the word-medial segments. Of this subset of 400 tokens, 64% were defined as flapped, 28% were defined as unflapped, and 8% were excluded due to being unreleased or glottalized. The amplitude of the tokens defined as flaps (68 dB) measured significantly higher than that of unflapped tokens (64 dB) [$t(365)=3.796$, $p<0.0001$]; however, intensity measures of flapped and unflapped tokens overlapped. While we do not feel that amplitude alone can be used to distinguish between flapped and unflapped /t/ and /d/, these amplitude measures confirm that our method of defining flaps on a speaker-by-speaker basis differentiates between the less sonorous /t/ and /d/ stops and the more sonorous flaps.

The benefits of defining flaps individually in this manner include that the duration at which /t/ and /d/ are considered flapped is based on the speaker's productions, taking speaking rate and any other speaker variation into account. This process also reflects that alveolar flaps as a class are shorter in duration than alveolar stops, but it does not imply any other relationship between flaps and stops; whereas, using a value like half the duration of an alveolar stop or one standard deviation below the mean of an alveolar stop assumes a connection between the durations of flaps and stops. The flap cut-off value used in this study is preferable to other values used in previous studies because it is a nonarbitrary value based on a natural gap in each speaker's distribution of /t/ and /d/ tokens, allowing flaps to be defined on a speaker-by-speaker basis.

2.3.2. Disyllables

2.3.2.1. Flap frequency. Using the speaker-by-speaker values listed in Table 2, each intervocalic /t/ or /d/ utterance was defined as flapped or unflapped. Of the 6720 disyllabic tokens recorded,

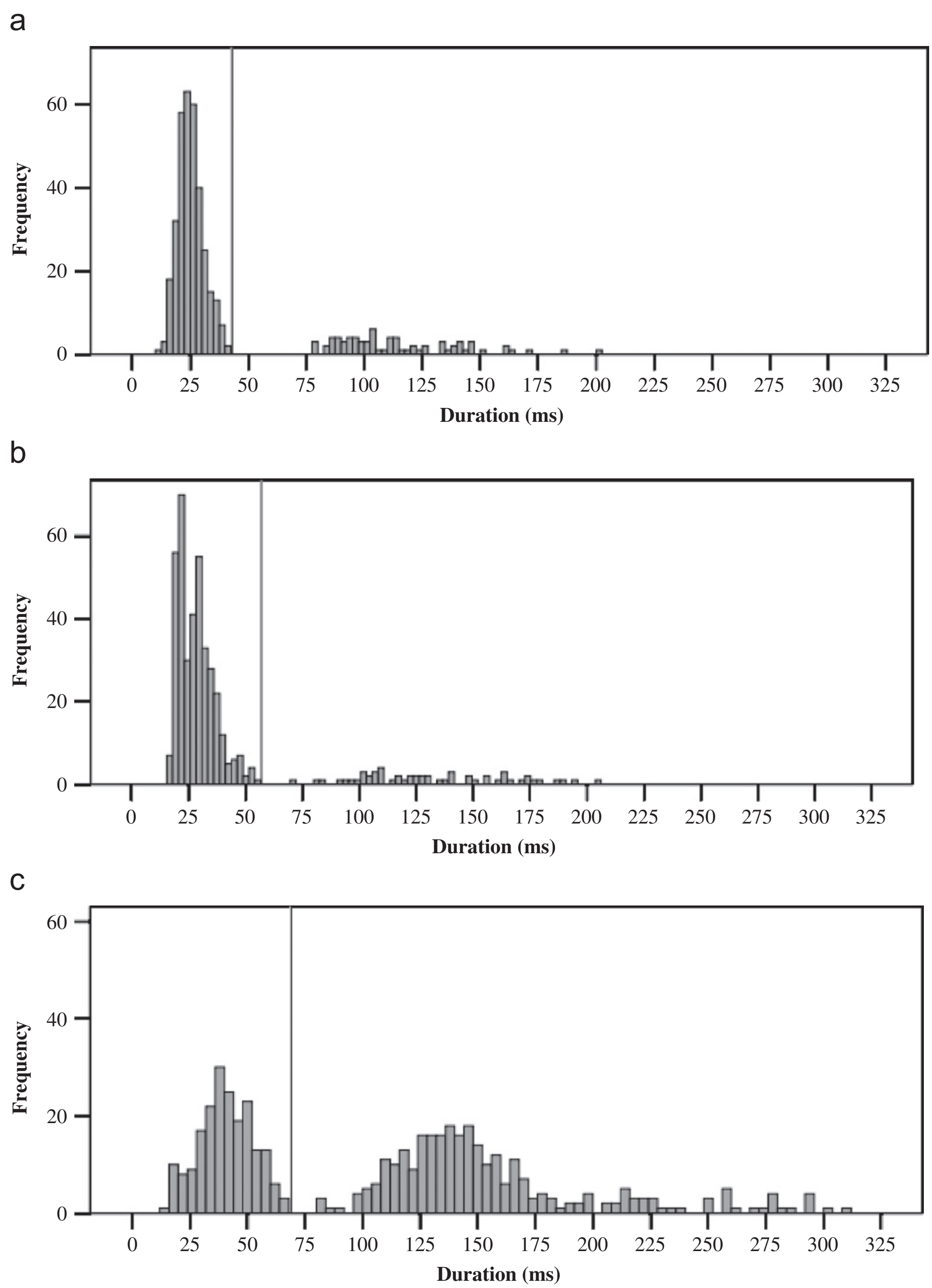

Fig. 3. Distributions of all /t/ and /d/ tokens for speaker 10 with a 43 ms cut-off value (a), for speaker 19 with a 55 ms cut-off value (b), and speaker 15 with a 69 ms cut-off value (c).

49 or 0.7% were excluded due to pronunciation errors or measurement difficulties. Of the remaining 6671 flappable disyllabic words and nonwords, 88% were pronounced as flaps. As predicted, word medial-/d/ was flapped more frequently at 99% than word medial-/t/ at 76%, and this difference proved significant in a paired samples *t*-test [$t(331)=-38.526$, $p<0.0001$]. These results may leave readers under the misconception that /t/ and /d/ surface as flaps more frequently than as fully articulated stops; however, that is only the case in the flapping environment defined above. In other contexts, such as when /t/ and /d/ occur word-finally in monosyllables, flapping is less frequent or even absent. Although the purpose of this study is not to analyze word-final /t/ and /d/ in monosyllables, it is interesting to note how word-final /t/ and /d/ were produced in this study. Of the 3120 monosyllables produced by speakers, 69% were realized as [t] or [d], 14.9% were produced as flaps, all of which occurred in the sentential context 2, 15.9% were unreleased or glottalized, and 0.2% were excluded due to pronunciation errors. The /t/ segments occurred as [t] 66%, as flaps 5.6%, and unreleased or glottalized 28.1% while the /d/ segments were produced as [d] 72%, as flaps 24.2%, and unreleased 3.5%. It should be noted here that, due to the large number of talkers and stimuli, statistical analyses from this point onward are based only on disyllables and on averages across speakers by gender.

A $2\times2\times3$ (Underlying Representation × Gender × Context) repeated measures Analysis of Variance was conducted on flap

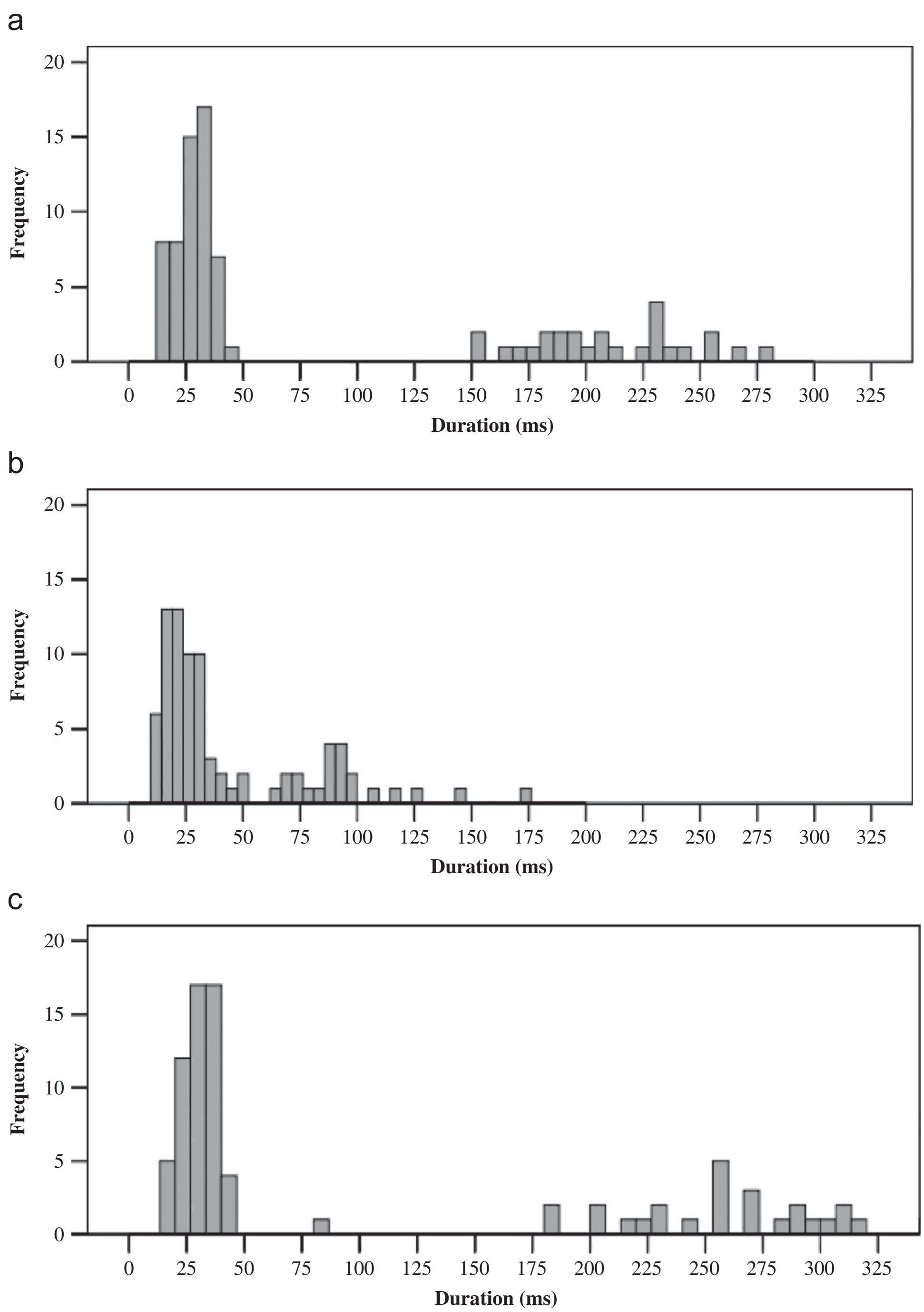

Fig. 4. Distribution of all /t/ tokens in (a) context 1 (pseudorandomized list), (b) context 2 (carrier sentence), and (c) context 3 (minimal pair list) as produced by speaker 3 in both monosyllables and disyllables.

frequency and main effects of Gender and Context were found. Females flapped more frequently with a mean flap frequency of 89% than males with 86% [$F(1, 326)=23.092$, $p<0.0001$]. As seen in Fig. 5, the difference was due to the percentage of /t/ tokens flapped since 99% of the /d/ tokens were flapped for both genders, contributing to an interaction between Underlying Representation and Gender [$F1, 326)=21.646$, $p<0.0001$]. It is important to point out that the interaction between Gender and Context was nonsignificant, so females flapped more frequently than males across all contexts [$F(2, 326)=1.070$, $p=0.344$]. Note also that these results contradict the findings of Sharf (1960), Zue and Laferriere (1979), and Byrd (1994) who concluded that male speakers produced flaps more frequently than females. The different results could be due to the smaller number of participants in the Sharf (1960) and Zue and Laferriere (1979) studies, two (one male and one female) and six (three males and three females) participants, respectively. In addition, the Byrd (1994) study included nasal flaps and across-word-boundary flaps in addition to intervocalic flaps, while the present study specifically investigates intervocalic oral flaps.

The main effect of Context [$F(2, 326)=10.124$, $p<0.0001$] is illustrated in Fig. 6, where it can be seen that /t/ is flapped most frequently in the pseudorandomized list context, that is the context read first and the context in which the participants were least aware of the /t/ and /d/ contrast in flappable minimal pairs like ‘writer’ and ‘rider’. A Bonferroni post hoc analysis determined

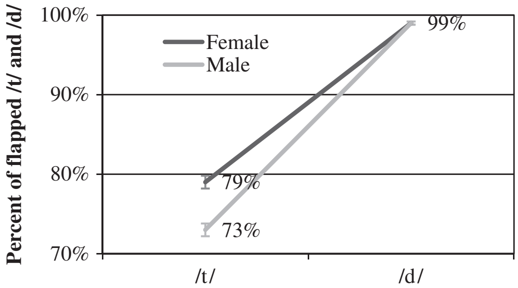

Fig. 5. Flap frequency of /t/ and /d/ by gender. Error bars reflect standard error.

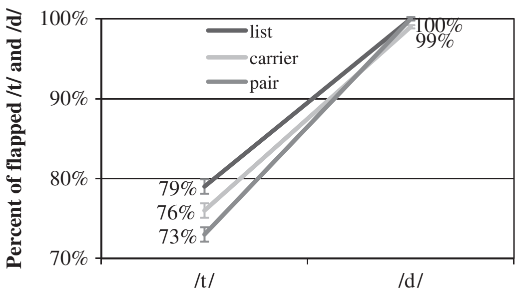

Fig. 6. Flap frequency of /t/ and /d/ by context. Error bars reflect standard error.

that /t/ was flapped more often in context 1 than in context 2 ($p=0.043$) or context 3 ($p=0.003$). Although all three contexts did not significantly differ from one another in overall flap frequency, it is clear that speakers flapped /t/ more frequently when they were less aware of the /t/ and /d/ contrast due to the pseudorandomized nature of the list than in either the second carrier sentence context, when they had already seen the words in a list and may have become aware of the contrast, or in the third paired context, when they focused on the contrast due to the opposition of minimal pairs. There was also an interaction between Underlying Representation and Context, because the flap frequency of the /t/ disyllables varied according to context, but the flap frequency of /d/ disyllables remained near 100% regardless of context [$F(2, 326)=10.080$, $p<0.0001$].

According to Patterson and Connine (2001), high frequency and monomorphemic words are more likely to be flapped than low frequency and bimorphemic words. In order to test whether word frequency affected flapping, the Francis-Kučera (1982) written word frequencies for the disyllabic words were used to divide the words into two groups, high frequency and low frequency, based on a mean cut-off of 25 per million. After determining that the mean frequency of all disyllabic lexical items was 25, words with frequencies lower than 25 were labeled low frequency words and words with frequencies at and above 25 were labeled high frequency. An independent samples *t*-test was conducted comparing the flap frequency of the high frequency words to that of the low frequency words. The high frequency disyllables were flapped 88%, and the low frequency disyllables were flapped 89%, a nonsignificant difference [$t(499)=0.290$, $p=0.772$]. In another independent samples *t*-test, the words were divided into two different groups, very low frequency words (0 words per million) and very high frequency words (100 words per million and above), to see if an effect of word frequency could be obtained. The very low frequency disyllables were flapped 84%, and the very high frequency disyllables were flapped 85%, another nonsignificant difference [$t(147)=0.301$, $p=0.764$]. To see if the number of morphemes in a lexical item affected flap frequency, an independent samples *t*-test was used to compare the flap frequency of monomorphemic lexical items to bimorphemic ones. Target stops were produced as flaps at a rate of 88% in monomorphemic words and 89% in bimorphemic words, a nonsignificant difference [$t(499)=0.170$, $p=0.865$]. Based on these results, it seems fair to conclude that neither word frequency nor morphological complexity of lexical items affected how often /t/ and /d/ were flapped in the present study. The different results could be due to the different collection methods used in Patterson and Connine (2001) and the present study. While Patterson and Connine (2001) used the SWITCHBOARD speech database, the tokens in this study were produced by speakers in an anechoic chamber. Additionally, Patterson and Connine (2001) only investigated the flap frequency of /t/, but the present study investigates the flap frequency of both /t/ and /d/. Finally, the tokens in this study were produced an equal number of times by the same subjects; however, the number of repetitions used in the Patterson and Connine (2001) study depended on the number naturally occurring in the SWITCHBOARD speech database, ranging from 1 occurrence to 674, and these lexical items were not produced by the same speaker.

Finally, using the 28 disyllabic words that do not exist in minimal pairs with their 28 nonword counterparts, a paired samples *t*-test was performed comparing the flap frequency of words to that of nonwords. The words were flapped 90% of the time while the nonwords were flapped 85%, a significant difference [$t(165)=2.737$, $p=0.007$]. Additionally, in a 2×2 (Gender × Lexical Status) repeated measures Analysis of Variance, in which Lexical Status represents whether the disyllables are words or nonwords, the interaction between Gender and Lexical Status proved nonsignificant [$F(1, 164)=0.001$, $p=0.884$]. This indicates that the difference in flap frequency between words and nonwords occurred across all speakers regardless of gender and that it was not driven by the higher flapping frequency of female speakers.

The above flap frequency results suggest that females, contrary to past research, flap disyllabic /t/ tokens more frequently than males. Flap frequency was also affected by context, so the more aware speakers are of a contrast between flappable /t/ and /d/, the less likely they are to flap /t/. And although word frequency and morphological complexity of lexical items do not affect how often they are flapped, lexical items are flapped more frequently than nonwords.

2.3.2.2. Measures of flap duration. Using only flapped disyllabic tokens, a paired samples *t*-test was carried out comparing the mean duration of /t/ flaps to that of /d/ flaps. Similar to prior studies, /d/ flaps measured longer in duration (30.4 ms) than /t/ flaps (29.5 ms), a difference in duration in the opposite direction of /t/ and /d/ unflapped stops, where the voiceless alveolar stops are longer than the voiced. However, unlike previous studies, the small difference between /t/ and /d/ flaps was significant [$t(335)=4.087$, $p<0.0001$], probably due to the large sample size. Like previous studies, it must be concluded that flap duration is not a cue to the underlying voicing of /t/ or /d/ flaps. While overall flap duration differences between /t/ and /d/ fall below 1 ms, speaker variation does exist as shown in Table 3. For example, speakers 14 (7 ms) and 18 (6 ms) exhibit a larger mean difference than the other speakers. A perceptual study could shed light on whether or not this larger difference is perceived by listeners as a cue for voicing.

2.3.2.3. Measures of vowel duration. Four measurements reflecting vowel duration were recorded for each disyllabic token: first vowel duration, first vowel duration to second vowel duration ratio, word length, and first vowel duration to word length ratio. All four measures will be reported during the paired samples *t*-tests, because

Table 3
Speaker variation in flapped consonant duration of disyllables.

Sp. #	/t/ dur. (ms)	s.d.	Min.	Max.	/d/ dur. (ms)	s.d.	Min.	Max.	Diff. (ms) (/t/ − /d/)
1	31	9.7	9	55	32	9.7	12	56	−1
2	33	9.3	12	57	36	9.4	12	56	−3
3	27	7.9	12	43	26	7.6	12	46	1
4	40	6.0	27	60	41	7.4	15	58	−1
5	33	10.5	10	59	33	10.7	10	57	0
6	27	7.5	14	56	26	7.2	12	48	1
7	26	6.3	17	49	26	7.1	14	42	0
8	24	5.3	11	41	24	5.8	14	41	0
9	29	9.8	14	54	29	8.2	15	53	0
10	25	5.2	12	40	25	5.5	14	42	0
11	35	9.0	10	57	36	9.5	11	57	−1
12	33	8.7	15	57	36	8.8	15	59	−3
13	28	8.8	10	46	31	8.6	12	51	−3
14	25	8.8	10	56	32	10.4	11	59	−7
15	42	16.1	16	66	39	10.9	18	59	3
16	27	7.8	13	45	27	7.1	14	45	0
17	27	7.5	14	51	28	8.0	15	55	−1
18	28	7.2	16	44	34	7.6	18	46	−6
19	27	7.8	18	54	28	7.2	18	52	−1
20	32	8.9	15	56	30	7.2	16	54	2

Table 4
Significant vowel length measures in /t/ and /d/ flapped tokens.

Measurement	/d/ flap disyllables	/t/ flap disyllables	Difference /d/ − /t/	Level of significance
V1:V2	1.76	1.68	0.08	[$t(335)=4.981$, $p<0.0001$]
Word length	508 ms	499 ms	9 ms	[$t(335)=6.051$, $p<0.0001$]
V1:Word	0.282	0.272	0.01	[$t(335)=4.000$, $p<0.0001$]

these measurements replicate the methods used in past studies; however, only the first vowel duration will be reported in the repeated measures Analysis of Variance as the results of the other three related variables produced nearly identical results. Paired samples *t*-tests were conducted on the four measurements mentioned above to see if first vowel duration varied depending on whether the disyllabic word contained a /t/ flap or a /d/ flap. As predicted, vowels preceding /d/ flaps (137 ms) were 6 ms longer on average than those preceding /t/ flaps (131 ms) [$t(335)=11.809$, $p<0.0001$]. Table 4 details the significance of the other vowel measurements. The results replicate past findings that first vowel to second vowel ratio, word length, and first vowel to word ratio are longer for tokens with /d/ flaps than with /t/ flaps.

A $3\times2\times2$ (Context × Gender × Diphthong) repeated measures Analysis of Variance was conducted on vowel duration. Contrary to predictions made based on Sharf 1960 and Malécot and Lloyd 1968, the interaction between vowel duration and Diphthong proved to be nonsignificant [$F(1,\ 156)=1.965$, $p=0.163$]. The average 5 ms vowel difference between monophthongs preceding /t/ (112 ms) and those preceding /d/ (117 ms) did not differ significantly from the 7 ms vowel difference between diphthongs preceding /t/ (143 ms) and those preceding /d/ (150 ms).

A further analysis was carried out in which monomorphemic disyllabic pairs, like 'petal' and 'pedal', were compared to bimorphemic disyllabic pairs, like 'wetting' and 'wedding'. The purpose of this analysis was to see if vowel duration differences preceding flapped /d/ and /t/ would be greater for bimorphemic pairs than for monomorphemic pairs since the bimorphemic pairs are derived from root words, in this case 'wet' and 'wed', which contrast in vowel duration preceding /t/ and /d/. A paired samples *t*-test found a trend, where the difference in vowel duration for bimorphemic pairs was greater than that of monomorphemic pairs [$F(1,\ 58)=3.214$, $p=0.072$]. The vowel duration preceding /d/ was 8 ms greater than that preceding /t/ for bimorphemic pairs, like 'wedding' and 'wetting'. In the case of monomorphemic pairs, like 'pedal' and 'petal', only a 3 ms difference was measured. These results are based on a small sample of stimuli, three monomorphemic pairs and five bimorphemic pairs, so a larger set of stimuli matched in morphemic structure could lead to clearly significant results.

Similar to the speaker variation in flapped consonant duration discussed above, speakers also varied in vowel duration as shown in Table 5. Whereas some speakers, like speakers 5 (−2 ms) and 17 (1 ms), produced a smaller vowel duration difference between disyllabic tokens containing a /d/ and /t/ flap than the 6 ms average, other speakers, like speakers 13 (11 ms), 15 (10 ms), and 18 (28 ms), produced a larger than average difference. These speaker variations could be used in future perceptual studies to ascertain whether listeners use these consonant and vowel duration differences to decide if a flapped alveolar stop is a /t/ or /d/ or whether they rely more heavily on word frequency or a general d-bias to make these judgments.

2.4. Discussion

Instead of using arbitrary numerical cut-offs, like 10–40 or 50 ms, to distinguish the consonant duration of stops and flaps, the cut-off values in this study were determined on a speaker-by-speaker basis by visually analyzing the /t/ and /d/ distribution of each speaker. Across all speakers, the average flap cut-off value was determined to be 56 ms with individual cut-off values ranging from 43 to 69 ms.

Using the flap frequency of speakers based on these values, it was found that females flapped /t/ tokens more frequently than males, in direct contradiction with the findings of Sharf (1960), Zue and Laferriere (1979), and Byrd (1994). As discussed briefly above, the findings of the present study are more reliable because

Table 5
Speaker variation in vowel duration preceding flaps in disyllables.

Sp. #	V dur. preceding /t/ (ms)	s.d.	Min.	Max.	V dur. preceding /d/ (ms)	s.d.	Min.	Max.	Diff. (ms) (/d/ – /t/)
1	137	37.4	30	211	144	39.8	44	232	7
2	131	39.8	47	212	136	42.9	35	216	5
3	117	33.8	45	198	123	37.7	49	228	6
4	152	42.8	56	248	156	44.3	53	269	4
5	133	41.7	50	278	131	42.7	47	265	−2
6	122	32.6	51	198	129	35.4	58	220	7
7	129	37.2	48	196	134	39.3	62	212	5
8	132	35.7	38	215	138	36.7	43	222	6
9	118	31.6	67	182	127	39.1	61	205	9
10	110	31.0	33	170	115	33.0	32	195	5
11	130	40.0	48	213	133	41.0	43	224	3
12	126	36.7	43	204	133	40.8	47	232	8
13	118	38.5	45	205	129	41.2	29	214	11
14	149	41.0	64	241	158	41.2	63	245	9
15	110	33.7	73	176	120	44.0	70	193	10
16	140	36.0	61	250	147	36.3	68	230	7
17	110	31.7	24	184	111	33.7	32	209	1
18	140	36.1	68	210	168	50.9	87	291	28
19	145	36.7	38	242	148	36.5	61	261	3
20	130	33.4	54	222	138	34.8	55	233	8

they are based on a larger number of speakers, 10 males and 10 females, a larger number of flappable tokens (112 disyllabic tokens) read in three contexts, and a more well-defined flapping environment (intervocalic post-stressed) than the previous research.

The context in which the flappable tokens were read also affected the frequency with which the tokens were pronounced as flaps. When speakers read words and nonwords in the first context, a pseudorandomized list within which the words were read in isolation and the speakers were unaware of the contrast being investigated, participants were more likely to flap intervocalic /t/ and /d/. In the second context, a pseudorandomized carrier sentence list during which the speakers may have become aware of the contrast since they were seeing the list for the second time, the participants flapped to a great degree, though the difference was not significant. Finally, in the third context, a minimal pairs list within which the words were set in contrastive /t/ and /d/ opposition, making the participants hyperaware of the contrast, flapping occurred the least and significantly differed from the list context. Therefore, the more aware the speakers were of the /t/ and /d/ contrast, the less likely they were to flap.

Based on the results reported above, it is unclear whether participants in this study contrast intervocalic flap voicing in disyllabic words by using consonant and vowel duration. While it is true that the /d/ flaps measured significantly longer than the /t/ flaps, the 0.9 ms difference could not be used by a listener as cue for voicing. Similarly, a vowel duration difference preceding /t/ flaps and /d/ flaps, where vowels measured 6 ms longer preceding /d/ flaps, was proven statistically significant. While this difference appeared consistently, pointing to an incomplete neutralization of /t/ and /d/ when flapped, and while the difference supports the findings of previous acoustic studies that were based on more varied speaker populations, it is also unlikely that listeners could use a 6 ms difference in preceding vowel duration to distinguish /t/ flaps from /d/ flaps. The comparison of vowel duration differences between monomorphemic and bimorphemic disyllables also produced interesting results, indicating that any vowel duration contrasts that exist preceding flapped /t/ and /d/ could be due to a relationship between monosyllable root words ('wet' and 'wed') and bimorphemic disyllabic flapped words ('wetting' and 'wedding'). Since a contrast is maintained through vowel duration in the root word, the contrast is likewise maintained in the corresponding bimorphemic word.

In order to evaluate whether /t/ and /d/ have been perceptually neutralized when flapped, a perceptual study is necessary. If the /t/ and /d/ flaps are contrastive due to vowel duration, we should expect to find a correlation between longer vowel duration and the correct perception of /d/ flaps. If /t/ and /d/ are neutralized perceptually, higher word frequency and a general d-bias, not acoustic measures, could be correlated with the correct perception of /t/ and /d/.

3. Experiment 2: perceptual study

3.1. Method

3.1.1. Participants

Thirty-four American English speakers from the University of Kansas participated in the perceptual portion of this study. Of these speakers, 24 were female and 10 were male, and none of them had lived outside the United States for more than one year. Seventeen of the female and nine of the male participants were born in Kansas and had not lived outside of the area for more than a year. The remaining participants were from other parts of the United States, including Massachusetts (1F), Illinois (3F), Virginia (1F), Minnesota (1F), Colorado (1M), and North Carolina (1F), but, like the participants from Kansas, those from outside of Kansas identified either 'caught' and 'lot', 'pen' and 'shin', or both as rhyming pairs. Based on their responses to the dialect questionnaire, the prevalence of flapping across the United States, and the similarity of flap frequency, vowel duration differences, and flap duration differences reported in the current acoustic study and in studies based on speakers from other areas in the United States, all of the above-mentioned participants were included in the following analyses.

3.1.2. Stimuli

In order to control word frequency and underlying representation, the four word pairs in Table 6 were chosen. There are four words that contain underlying /d/, and these are matched with words that only differ by containing underlying /t/. Of these pairs, two consist of high frequency /d/ words and low frequency /t/

Table 6
Word pairs used in perceptual study.

/d/	Word frequency	/t/	Word frequency
leader	187	liter	4
wedding	34	wetting	3
tidal	1	title	106
madder	0	matter	377

Table 7
The vowel duration difference (v. diff.) of word pairs used in the perceptual study.

Word pair	Mean condition		Enhanced condition		Opposite condition	
	v. diff. (ms)	*z*-score	v. diff.	*z*-score	v. diff.	*z*-score
leader–liter	8	0.233	34	3.256	−31	−4.302
wedding–wetting	7	0.116	20	1.628	−14	−2.326
tidal–title	9	0.349	30	2.791	−5	−1.279
madder–matter	9	0.349	22	1.860	−31	−4.302

words according to the Francis-Kučera (1982) written word frequencies. Additionally, the other two pairs consist of low frequency /d/ words and high frequency /t/ words, balancing both word frequency and underlying representation.

Next, using natural productions from the recordings made for the production experiment, the vowel duration difference between the vowels preceding /d/ and /t/ was controlled in order to create three different conditions as seen in Table 7. Speaker 11 was chosen, because she flapped all repetitions (pseudorandomized list, sentence context, and minimal pairs) of the above-mentioned pairs and the vowel durations preceding flapped /d/ and flapped /t/ could be mapped onto the three different vowel conditions. For the mean condition, speaker 11 produced 'leader' with a 124 ms vowel preceding /d/ and produced 'liter' with a 116 ms vowel preceding /t/ in the minimal pairs list context. This pair was used for the mean condition, because the vowel duration difference between the vowel preceding /d/ and the vowel preceding /t/ is 8 ms, close to the 6 ms mean difference observed in the production study. For the enhanced condition, the vowel preceding /t/ in 'liter' was 90 ms during the sentence context, so the sentence context repetition of 'liter' was paired with the minimal pairs repetition of 'leader', because the difference between the vowel preceding /d/ (124 ms) and the vowel preceding /t/ (90 ms) is 34 ms, over one standard deviation above the mean vowel duration difference reported in this study. Likewise, for the opposite condition, the vowel preceding /t/ in liter (116 ms) measured 31 ms longer than the vowel preceding /d/ (85 ms). In other words, participants hear the opposite cues because vowels are longer preceding /t/ than /d/ and the difference is over one standard deviation below the mean reported in this study.

As can be seen in Table 7, the duration difference between vowels preceding flapped /d/ and flapped /t/ is within one millisecond of the mean vowel duration difference reported in the present acoustic study (6 ms). The difference in the enhanced condition is at least one standard deviation above the mean, where there is a larger, or more enhanced, vowel duration difference than the average found in Experiment 1. The tokens in the opposite condition have vowel duration differences that are at least one standard deviation below the mean, where the vowel is actually longer preceding /t/ flaps than /d/ flaps, so the vowel duration difference goes in the opposite direction of what is expected. It is important to note that the tokens in the mean, the enhanced, and the opposite conditions have not been manipulated; the conditions are created by carefully matching the tokens produced by speaker 11 during three different repetitions.

If listeners use vowel duration as a cue to decide whether they hear an underlying /t/ or /d/ when presented with the auditory stimulus [ˈliɾɚ], they should perceive the tokens in the enhanced condition correctly most often, followed by the mean condition, and then the opposite condition. Since flaps are voiced segments, listeners may perceive them as voiced alveolar stops, leading to a higher percentage of 'd' responses and a general d-bias. Listeners may also depend on word frequency and choose a more frequent candidate like 'leader' and 'title' over less frequent items like 'liter' and 'tidal'.

3.1.3. *Procedure*

Participants were asked to complete a forced-choice identification task, which was designed using Paradigm software (Tagliaferri, 2008). Items from the word pairs listed above were presented one token at a time over headphones. Participants then saw two possible choices, 'leader' and 'liter', for example when they heard the stimulus [ˈliɾɚ], and chose a response by mouse-clicking on the word. Participants were required to center the mouse on the computer screen by clicking inside a circle before hearing the next stimulus. Possible orthographic representations were counterbalanced across listeners, so half of the participants saw 't' words on the left side of the screen and 'd' words on the right, and the other half saw 'd' words on the left and 't' words on the right. The stimuli were blocked by the vowel duration conditions (mean, enhanced, and opposite), with the order of the blocks also being counterbalanced across listeners. Within a block, each token was presented four times in random order.

3.2. *Results*

A $2\times2\times3$ (Underlying Representation × Word Frequency × Vowel Duration Difference) repeated measures Analysis of Variance was conducted on the percent of correct responses. The two levels of Underlying Representation were determined by whether the underlying flap in the word was /t/ or /d/, the two Word Frequency levels were high versus low, and the three Vowel Duration Difference conditions were mean, enhanced, and opposite.

First, no main effect of Vowel Duration Difference was found [$F(2, 66)=1.390$, $p=0.256$]. The percentage correct of the mean, enhanced, and opposite conditions fell near chance at 52%, 52%, and 48%, respectively. However, as predicted, main effects of both Underlying Representation [$F(1, 33)=24.105$, $p<0.0001$] and Word Frequency [$F(1, 33)=22.556$, $p<0.0001$] were found. Words containing an underlying /d/, like 'leader', were perceived correctly more often at 57% than words containing an underlying /t/, like 'liter, at 44%. It is also not surprising that the high frequency words, like 'leader' and 'title', were responded to more correctly at 59% than low frequency words, like 'liter' and 'tidal', at 42%.

In addition to these main effects, an interaction between Word Frequency and Underlying Representation was also found [$F(1, 33)=8.423$, $p=0.007$]. While high frequency /d/ and /t/ items are perceived correctly near the same levels, 62% and 55%, respectively, the change in correct responses for low frequency /t/ words (33%) as compared to high frequency /t/ words (55%) is greater than that for their /d/ word counterparts (51–62%), resulting in a significant interaction. Low frequency /t/ words appear to be at a double disadvantage, because they can neither be helped by word frequency effects nor by a d-bias. While there was no main effect of Vowel Duration Difference, there was an interaction between Vowel Duration Difference and Underlying Representation. Although the percentage of correct responses for /d/ words remains the same across all three conditions (mean—56%, enhanced—57%, and

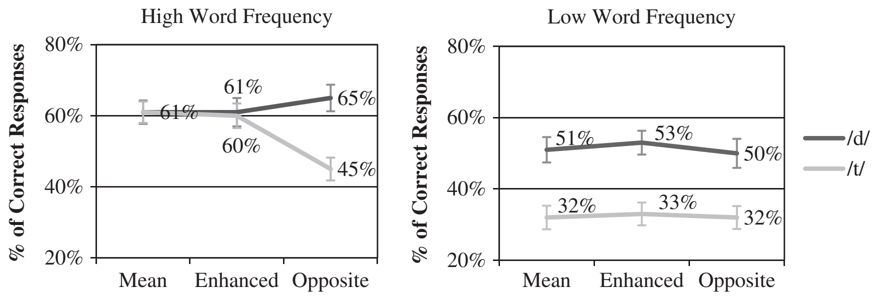

Fig. 7. Percent of correct responses by Word Frequency, Underlying Representation, and Vowel Duration Difference. Error bars reflect standard error.

opposite—58%), the responses to /t/ words decrease in the opposite condition (mean—47%, enhanced—47%, and opposite—39%). When listeners were presented with unexpected cues, shorter vowel duration preceding /d/ than /t/ flaps, the /t/ words were hurt by these opposite cues. However, the percentage correct for /d/ words, which benefited from a d-bias, did not decline.

Finally, a three-way interaction between Vowel Duration, Underlying Representation, and Vowel Duration Difference was also found. As seen in the right panel of Fig. 7, the percent of correct responses to /d/ words and /t/ words remained roughly the same across conditions, with no interactions occurring. Looking at the left panel of the figure, the interaction between Vowel Duration Difference and Underlying Representation explained above can be seen again. In addition to showing that /t/ words are hurt more by the opposite cues than /d/ words, this figure also shows that the main effect of Underlying Representation is more robust than the main effect of Word Frequency. If Word Frequency affected the correct perception of /t/ and /d/ flaps as much as Underlying Representation, there would also have been an interaction between Underlying Representation and Vowel Duration Difference for low frequency words, but as shown in the right panel of Fig. 7, this interaction does not occur.

3.3. Discussion

The careful manipulation of Underlying Representation, Word Frequency, and Vowel Duration Difference has allowed the present study to unravel the effects of these three variables on the perception of flaps. When listeners must judge whether a flap is an underlying /t/ or /d/, vowel duration alone does not affect their perception. Instead, listeners use word frequency and a general d-bias to make their decisions, with the d-bias more robustly affecting their choices. Since previous perceptual studies did not control word frequency and vowel duration, the large range of flap perception reported (56.6–86.5%) could have been caused by uncontrolled effects of word frequency and interactions between underlying representation, word frequency, and vowel duration differences as well as by the inclusion of unflapped tokens.

4. Conclusions

This study attempted to determine how often American English speakers flap intervocalic post-stressed /t/ and /d/, whether American English speakers differentiate /t/ and /d/ when flapped, and, if these sounds are acoustically distinguishable, whether American English speaking listeners can perceive the difference between flapped /t/ and flapped /d/. The speaking participants in Experiment 1 consistently flapped /d/ at ceiling levels (99%) but flapped /t/ significantly less frequently (76%). Additionally, all speakers were more likely to flap /t/ in the context where they were least aware of the /t/ and /d/ contrast, and female speakers flapped /t/ more frequently than male speakers. These flap frequency findings are based on a new method of distinguishing flapped from unflapped /t/ and /d/ using the distribution of each speaker's /t/ and /d/ consonant durations, a more objective method than that used in the past. After using this method to identify flapped /t/ and /d/ for all speakers, the vowel durations preceding the flaps and the durations of the flaps themselves were analyzed. With respect to producing differences between flapped /t/ and /d/, speakers produced vowels preceding flapped /d/ 6 ms longer on average than vowels preceding flapped /t/; however, flapped /t/ was less than a millisecond shorter than flapped /d/. In other words, speakers used vowel duration preceding /t/ and /d/ to acoustically distinguish the two sounds, and they produced a statistically significant flap duration difference between flapped /t/ and /d/, but the difference is too small to help perception.

While the vowel duration difference was measurable and statistically significant, the results of Experiment 2 show that listeners cannot reliably distinguish flapped /t/ from flapped /d/. Instead of vowel duration acting as a major cue to the voicing contrast between flapped /t/ and /d/ during the perceptual experiments, listeners relied more heavily upon a d-bias, where flapped /d/ was perceived correctly more often than flapped /t/, and upon word frequency, where high frequency words were perceived correctly more often than low frequency words.

In future studies, /t/ and /d/ durations could be measured in disyllables where they can be flapped, i.e., 'attic', and where they cannot be flapped, i.e., 'adapt'. Distributions graphed with these measurements would allow researchers to clearly distinguish flapped /t/ and /d/ from their unflapped counterparts in disyllabic contexts, instead of using both monosyllabic and disyllabic contexts. Additionally, a stimulus list containing medial /t/ and /d/ in monomorphemic pairs like 'petal–pedal' and bimorphemic pairs like 'wetting–wedding' could be analyzed to determine if the vowel duration difference between bimorphemic pairs is greater than that between monomorphemic pairs, a result that could have a phonological basis in optimality theory. The present study has set the groundwork to more reliably distinguish flapped from unflapped tokens and has found that while flapped /t/ and /d/ are acoustically distinguishable due to vowel duration differences, these small differences are not used as perceptual cues by native speakers of American English.

References

Boersma, P., & Weenink, D. (2005) *Praat: Doing phonetics by computer* (Version 4.3.12) [Computer Program]. Retrieved May 11, 2005, from 〈http://www.praat.org/〉.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*, 39–54.

Charles-Luce, J. (1997). Cognitive factors involved in preserving a phonemic contrast. *Language and Speech*, *40*, 229–248.

Charles-Luce, J., & Dressler, K. M. (1999). The effects of semantic predictability in non-pathological older adults' production of a phonemic contrast. *Clinical Linguists and Phonetics*, *13*, 199–217.

Charles-Luce, J., Dressler, K. M., & Ragonese, E. (1999). Effects of semantic predictability on children's preservation of a phonemic voice contrast. *Journal of Child Language*, *26*, 505–530.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*, 129–159.

Fisher, W. M., & Hirsh, I. J. (1976). Intervocalic flapping in English. *Chicago Linguistics Society*, *12*, 183–198.

Fox, R. A., & Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics*, *5*, 27–34.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin Company.

Lavoie, L. M. (2000). *Phonological patterns and phonetic manifestations of consonant weakening*. Ph.D. dissertation, Cornell University.

Lehman, W. P., & Heffner, R-M. S. (1940). Notes on the length of vowels (III). *American Speech*, *15*, 377–380.

Malécot, A., & Lloyd, P. (1968). The /t/:/d/ distinction in American alveolar flaps. *Lingua*, *19*, 264–272.

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica*, *58*, 254–275.

Sharf, D. J. (1960). Distinctiveness of 'voiced T' words. *American Speech*, *35*, 105–109.

Sharf, D. J. (1962). Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech*, *5*, 26–30.

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

Tagliaferri, B. (2008). *Paradigm: Perception research systems* [Computer Program]. Retrieved March 23, 2008, from 〈http://www.perceptionresearchsystems.com/〉.

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

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