Korean Journal of English Language and Linguistics, Vol 22, October 2022, pp. 1133-1158
DOI: 10.15738/kjell.22..202210.1133

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

A Mismatch in Completeness between Acoustic and Perceptual Neutralization in English Flapping

Gwanhi Yun (Daegu University)

Received: August 25, 2022
Revised: October 20, 2022
Accepted: October 30, 2022

Yun, Gwanhi
Professor, Department of English Language and Literature, Daegu University
201 Daegudae-ro, Gyeongsan Gyeongbuk, 38453, Korea
Tel: 053) 850-6025
Email: ghyun@daegu.ac.kr

ABSTRACT

Yun, Gwanhi. 2022. A mismatch in completeness between acoustic and perceptual neutralization in English flapping. *Korean Journal of English Language and Linguistics* 22, 1133-1158.

This study attempts to reveal the acoustical characteristics of flapped /t/s and /d/s as well as phonetic correlates of word-final /t,d/ contrast and to examine whether English native listeners distinguish a flapped /t/ and /d/ by using the durations of pre-flap vowels. For these purposes, production and perception experiments were administered for English native speakers. First, we found that word final devoicing does not occur in /t,d/ contrast and significant differences lie in many acoustic correlates, including durations of preceding vowels, stop closure durations, voicing duration and F0 of the preceding vowels. Second, the result showed the evidence that English flapping is incomplete neutralization, exhibiting that many acoustic properties differ between /t/ flaps and /d/ flaps in duration of pre-flap vowels, flap duration, voicing duration and VOT. Furthermore, the perception task yielded high perceptibility of word final /t,d/ contrast due to the availability of many acoustic cues. Next, it was shown that English listeners have difficulty in deciding whether a flap is an underlying /t/ or /d/. This suggests that complete neutralization engenders imperceptibility of /t/ flaps and /d/ flaps. Finally, our identification test revealed that the manipulation of the duration of the pre-flap vowels does not function as a perceptual cue for word medial /t/-/d/ contrast embedded in a flapping environment.

KEYWORDS

flapping, word final devoicing, (in)complete neutralization, acoustic correlates of voicing, (im)perceptiblity, length of pre-flap vowels

1. Introduction

Over past decades numerous studies have found that phonological rules are gradient, not categorical, including assimilation, neutralization, deletion, etc., with respect to some phonetic correlates of some phonological feature. Against this background, this study is mainly concerned with English flapping, a typical example of neutralization rules. One purpose of the current study is to address the question of whether flapping in American English is complete (i.e., categorical) or incomplete with respect to main acoustic correlates of flaps (i.e., gradient). Another goal is to explore whether the acoustic cue, i.e., the length of the preceding vowel is a primary perceptual cue to distinguish a flapped /t/ and a flapped /d/.

Flapping occurs in many phonological environments to different degrees primarily in American English (Yavaş 2011). Word medial /t,d/ are realized as flaps at 99% when preceded by a stressed vowel and followed by an unstressed vowel (e.g., *fatty, city, butter, ladder,* etc.) whereas word final /t,d/ become a flap in the same environment only at 19% (e.g., *at all, eat up*, etc.). In addition, flapping takes place between two unstressed vowels at 33% (e.g., *nationality*, *authority*, etc.). Furthermore, alveolars /t,d/ undergo flapping after the colored /r/ or before the following syllabic liquid, though less often (e.g., *porter, border, little, cattle*, etc.). In addition to the role of phonological environments, a multitude of conditioning factors have been found to affect the likelihood of flapping, including lexical frequency, morphological status, the following vowel, gender, etc. For example, there have been inconsistent findings about the role of gender in the rates of flapping. Herd et al. (2010) showed that flapping of word medial /t,d/ in nonce words occurred for females more frequently than males (89% vs. 86%) unlike the findings of previous studies (Byrd 1994, Sharf 1960). They also found that word medial /d/ undergoes flapping significantly more frequently than word medial /t/ (99% vs. 76%). Additionally, higher frequency words tend to be flapped more often than lower frequency words (Patterson and Connine 2001) whereas word frequency does not exert a crucial influence on the likelihood of flapping in Herd et al.'s (2010) study (88% vs. 89%). Also monomorphic words were shown to be more likely to be flapped than bimorphemic words (e.g., *city* >> *seating*, Patterson and Connine 2001).

Flap sounds are usually referred to as "allophones of /t/ and /d/ formed by a rapid movement of the tongue-tip making contact with the alveolar ridge, followed by immediate release" (Derrick et al. 2013, Kahn 1980, Turk 1992: 103). It has been suggested that the distinctive phonetic properites of flaps are voicing and shorter duration, compared to their counterpart stops /t,d/ (de Jong 1998, Zue and Laferriere 1979). Zue and Laferriere (1979) showed that the duration of flaps is comparatively shorter after stress, i.e., on average 26ms, ranging from 10ms to 40ms and flaps between two unstressed vowels are 40ms longer than those of the stress-induced short duration. Their study also showed that no release burst emerges in flaps. Charles-Luce (1997) also found that the indicator of voicing contrast in /t/ and /d/ is not closure duration but vowel duration. Following these studies, de Jong (1998) asserts that three sounds [t], [d] and [ɾ] lie in the duration continuum at phonetic level. Lavoie (2000) showed that the durations of /t/ flaps and /d/ flaps significantly differ, 34ms and 37ms, respectively and pointed out that flaps and stops differ in amplitude and sonority. Sung (2003) showed that English flaps and Korean flaps exhibit similar acoustic properties in the dimensions of closure duration (19ms, vs. 20ms), and percentages of voicing occurrences (82% vs. 95%). Based on these characteristics which play a crucial role in flaps being distinctive from stops /t,d/, many researchers suggested a defining feature for flaps, e.g., [+sonorant] (Kahn 1980), [+release] (Selkirk 1982), and [+extra short closure] (Steriade 2000).

This paper is organized as follows. In section 1, three theoretical issues associated with English flapping are discussed, and section 2 explicates the production experiment of /t/- and /d/-flapped words by English native speakers. Perception experiment is described in section 3 and section 4 discusses the result and concludes this paper.

1.1 Variation in Flapping

Variability in the pronunciation of words has been an important topic in the phonology-phonetics interface over past decades. This issue is associated with whether phonological rules are categorical or gradient. Joos (1942) and Port (1976) assert that phonological rules are by nature categorical or complete but incomplete flapping in American English is due to orthography or hyper-articulation as an effort to preserve the underlying voicing contrast (e.g., *latter* vs. *ladder*). In contrast, recent studies have shown that flapping is gradient or incomplete neutralization (Braver 2011, Herd et al. 2010, Patterson and Connine 2001, Simonet et al. 2008). For example, underlying phonological contrast is maintained as is evidenced by /t/ and /d/ traces at phonetic level. Specifically, vowels turned out to be longer before a flapped /d/ than a flapped /t/. This finding is interpreted to suggest that variation in flapping is an effort to maintain underlying phonological contrast. On top of that, lexical frequency is another conditioning factor to result in the extent of variation. To be specific, more variations in flapping were found in low-frequency words than in high-frequency words (e.g., LoF “wetting” [wɛtɪŋ]~[wɛɾɪŋ] vs. HiF “wedding” [wɛɾɪŋ]). Furthermore, morphological complexity also contributes to phonological variation with regard to flapping. For instance, complex words (e.g., “waiting” [wetɪŋ]~[weɾɪŋ]) showed more variation than monomorphemic words (e.g., “water” [wɔɾər]). Finally, phonological environment also exerts an influnece on the likelihood of variation of flapping. Herd et al. (2010) showed that word-medial /d/ is flapped more frequently than word-medial /t/ (99% vs. 76%, *leader* [lidər]~[liɾər] vs. liter [lɪtər]).

In addition to the variation in flapping at phonemic level, a multitude of previous studies have shown its variation at phonetic level, especially reflected in a variety of acoustic properties of flapped /t,d/. For one thing, the average pre-flap vowel duration showed variation from 6ms to 16ms, depending on the individual studies (Fox and Terbeek 1977, Herd et al. 2010, Yun 2015, Zue and Laferriere 1979). Also significant differences were found in pre-flap vowel duration between /t/ and /d/ flaps (Lavoice 2000, Turk 1992, Warner et al. 2009). These variations suggest that English flapping is not complete neutralization, but gradient, preserving the underlying contrast between /t/ and /d/. For flap duration, the closure duration of /t/ and /d/ flaps showed inconsistent characteristics. Some studies showed that /d/ flaps were longer than /t/ flaps (37ms vs. 34ms in Lavoie (2000), 30ms vs. 29ms in Herd et al. (2020) whereas others found the opposite patterns (18ms vs. 22ms in Turk (1992)). On the other hand, no significant differences were found between these two flaps (Zue and Laferriere 1979). It seems that a wider range of acoustic variations emerges in closure duration of flaps than pre-flap vowel duration. Ratios of the duration of the two vowels flanking flapped /t/ and /d were also found to be significantly different in many previous studies (Patterson and Connine 2001, Zue and Laferriere 1979). Intensity between a vowel and a flap was different between /t/ flaps and /d/ flaps variants, e.g., 10dB (Warner et al. 2009). Taken together, these acoustic properties seem to contribute to the mode of incomplete neutralization for English flapping, leading to phonetic variation at different weights.

English flapping which is applied to underlying /t/ and /d/ is apparently a typical example of phonological and phonetic variation. Given that previous research focused on the investigation of one or two acoustic cues for flapping, the present study attempts to provide a more comprehensive picture of acoustic cues involving English flapping. We investigate the presence of the multiple cues, including the duration of pre-flap vowel, closure duration of flaps, voicing duration of closure, F0 of pre-flap vowel, etc. centering on the realization of /t/ flaps and /d/ flaps (e.g., *writer* vs. *rider*).

1.2 Rule Ordering Pattern with Flapping

Rule ordering pattern is another issue associated with English flapping. English flapping has traditionally been known to be a typical example of opaque rule interaction in generative phonology. Fox and Terbeek (1977) proposed an opaque rule interaction whereby the rule of vowel lengthening induced by voiced obstruent precedes flapping. This rule ordering is based on the observation that vowels are longer before /d/ flaps than before /t/ flaps (e.g., *rider* >> *writer*). If flapping is applied prior to vowel lengthening, the incomplete neutralization with the respect to different pre-flap vowel duration cannot be given a suitable phonological account. This ordering indicates that vowel lengthening rule is not transparent in the surface form of flapped /d/.

Another type of opaque rule interaction is offered to give an account for Canadian variant of flapping (Chambers 2006, Joos 1942). For some speakers, especially in Vancouver and Ottawa in Canada, /a/ is raised to /ʌ/ before voiceless stop /t/ in /t/-flapped words (e.g., *typewriter* [...ʌjɾ..] vs. *rider* [...ajɾ..]). The asymmetry in the occurrence of raising between /t/-flapped and /d/-flapped words can be encouched within opaque type of rule ordering. To be specific, raising precedes flapping in Canadian English. Accordingly, raising rule is opaque in the surface.

Longer vowels before flapped /d/ over flapped /t/ and the presence of raised vowel before flapped /t/ over flapped /d/ are indicative of the trace of underlying contrast and of incomplete application of flapping. The phonological issue accompanied by flapping is beyond the scope of the current study. Rather, we focus on the phonetic variation regarding the realization of English flapping.

1.3 Phonetic or Phonological Uniformity with Flapping

It has been controversial whether uniformity is preserved at either phonological or phonetic dimension. Steriade (2000) argued for phonetic uniformity with the example of realization of stop sounds. For example, aspiration of the stop is maintained in morphologically related words (e.g., *mili[t^h]ary ≈ mili[t^h]aristic*) and furthermore, the flap allophone of /t,d/ is constant among the base forms and their derived words (e.g., *capi[ɾ]al ≈ capi[ɾ]alistic*). Tracking down the allophonic traces of the morphological family words, it is proposed that non-contrastive phonetic features, including [aspiration] and [extra short closure] characteristics of flap in the base form are uniformly observed in their corresponding inflected or derived forms.

On the other hand, phonological uniformity has been put forward in many studies (McCarthy 2001, Riehl 2003). According to this position, phonetic correlates of flaps such as closure duration, VOT, voicing duration, etc. are not uniform across derived words. Additional evidence comes from the finding that the syllabicity of /n/ fluctuates across inflected or derived words (e.g., *lighten≈lightens≈lightenend≈lightning*).

Although these opposing proposals demand further investigation into flapping, the current study is mainly concerned with the status of (in)complete application of flapping by looking into a variety of phonetic correlates of English flapping.

1.4 The Present Study

Assuming that English flapping invokes intricate theoretical issues covered in many previous studies, we attempt to achieve three research goals regarding the true merger of the underlying /t,d/ contrast. First, we investigate whether English flapping is completely neutralized or gradiently realized with respect to phonetic correlates of flapping. It is examined whether the differences in the phonetic correlates of the underyling voicing contrast in word-final /t,d/ (e.g., *wait* vs. *wade*) converge in the word-medial position, i.e., a flapping environment (e.g., *waiting* vs. *wading*). Unlike the tradition of generative phonology, a number of recent studies have shown that German final /t,d/ do not merge into fully voiceless stops with longer preceding vowels before /d/ than before /t/ and gradient voicing during /d/ (Roettger et al. 2014). Furthermore, while previous research mostly focuses on single or double acoustic cues, the current study examines a multitude of acoustic cues such as duration of pre-flap vowel, stop closure duration of flaps, voicing duration, perecentage of stop burst, and F0 of pre-flap vowel to reveal a more comprehensive status of English flapping. Although some argue that incomplete neutralization stems from the hyperarticulation of laboratory speech (Warner et al. 2006), a great majority of work reveals that word-medial flapping occurs above 80%. Accordingly, it is worthwhile examining a wider variety of phonetic corrlates centering on flappying to shed light on the debate of (in)complete flapping.

Second, we aim to provide phonological implications as to the interaction between flapping, vowel lengthening and polysyllabic shortening by measuring the pre-flap vowel length. Our potential findings are predicted to affect phonological approach to the interaction between flapping and lengthening of vowels preceding flaps in a different manner. First, if vowels before /t/ flaps are not significantly different from those before /d/ flaps, it might be indicative of complete neutralization and accordingly, it is proposed that flapping applies before vowel lengthening in rule-based phonology. In contrast, if we obtain the opposite finding, it leads to the implication that English flapping is incomplete neutralization and vowel lengthening precedes flapping, resulting in opaque rule interaction as suggested by Fox and Terbeek (1977). Additionally, we measured lengths of the vowel before word-final /t,d/ and the vowel preceding /t/ and /d/ flaps and compared them to see if polysyllabic shortening is applied in flapped words. If significant differences are found in the duration of pre-flap vowels between word-medial /t/ and /d/ words despite the application of polysyllabic shortening (e.g., *raiding* vs. *rating*), they might be another supportive evidence for incomplete flapping.

Last, we attempt to show whether the length of vowel preceding flaps is a perceptual cue sufficient to recover the underlying /t,d/ words. As described previously, a great bulk of studies have shown that the underlying contrast between /t/ and /d/ is preserved in one or two acoustic dimensions (Lavoie 2000, Patterson and Connine 2001, Turk 1992 among others). However, they did not demostrate that such subtle acoutic differences are the perceptual cues sufficient to distinguish /t/-flapped from /d/-flapped words. If delicate acoustic differences do not function as perceptual cues, a naturally arising question follows, i.e., why do speakers apply flapping incompletely though it does not benefit listeners' identification of flapped words. To bridge this research gap and answer the question, the present study examined whether pre-flap vowel length is a potential perceptual cue to discriminate /t/-flapped and /d/-flapped words.

2. Experiment 1: Production of English Flapping

2.1 Participants

Fifteen native speakers of American English participated in the production experiment. They were graduate students or office workers at Cornell University in the US. Seven were male and eight female. Their ages ranged from 18 to 56 (mean: 25, SD: 9.3). The participants received monetary compensation. They were from many parts of the US, including West Coast, Middle, East Coast, Southern US, etc. They had no speech hearing or utterance problems.

2.2 Materials

Two sets of English minimal pairs were randomly selected to see if (in)complete flapping occurs by comparing the acoustic properties of word-final /t,d/ with those of word-medial /t,d/ placed in a flapping environment. One group of tokens constitute twenty four minimal pairs where twelve words end with voiced alveolar /t/ and the other twelve words end with voiced alveolar /d/ (e.g., *beat* vs. *bead*). Each pair contains the identical vowel preceding the word-final /t,d/. Twelve vowels are followed by word-final /t,d/, such as /i,ɪ,ɛ,æ,u,ʊ,o,ʌ,ɑ,aɪ,aʊ/. All these forty eight words are monosyllabic.

To elicit flapping, another group of tokens was selected. They consist of derivational or inflectional words derived from the first group of words. To ensure a flapping environment, the twelve minimal pairs were selected for words formed by attaching affixes *(-ing, -ist, -y, -er*) to /t,d/-final words (e.g., *beading* vs. *beating*, *podder* vs. *potter*). Since these words are derived from the first group of words, they also contain the identical twelve vowels preceding the word-medial /t,d/. Furthermore, since they are attached with monosyllabic suffixes, they are bisyllabic English words with stress on the first syllable. A total of forty eight tokens were employed to obtain flapping data.

In total, 96 stimuli were used for the production materials (See Appendix for a full list of materials.). These tokens were embedded in a carrier sentence "Please say _____ ". 15 American English speakers were asked to read each carrier phrase as fast and naturally as they could to elicit the surface forms where flapping is likely to occur.

2.3 Procedures

The production experiment was carried out in a sound-attenuated booth in phonetics lab at Cornell University in the US. The 15 subjects were seated in front of the computer attached with a headset. They were instructed to read a list of stimuli embedded in the carrier sentence to a microphone fixed within the headset. The list of stimuli was written down in English orthography on several sheets of paper, and it contained three repetitions of the same tokens. The order of the stimuli was randomized; accordingly, the participants were asked to read it at a normal speaking rate.

After being recorded, all of the participants' audio data were analyzed in Praat (Boersma and Weenink 2020). Five acoustic parameters which are potential phonetic correlates of word-final voicing contrast in stops were measured to see if coda voicing is complete or incomplete neutralization; (i) duration of the preceding vowel, (ii) stop closure duration, (iii) voicing duration, (iv) percentage of stop burst, (v) F0 of the preceding vowel. These acoustic properties were compared between /t/-final and /d/-final words to see if there is any significant difference (e.g., *write* vs. *ride*).

In addition, four acoustic parameters were also measured and analyzed to examine whether flapping is complete or incomplete neutralization; (i) F0 of pre-flap vowel, (ii) VOT of /t/ flaps and /d/ flaps, (iii) ratio of the duration of pre-flap vowel to the length of post-flap vowel, (iv) percentage of aspiration. To see if there are significant differences in these dimensions between /t/-flapped and /d/-flapped words, these were compared. Since the identical stimuli were produced across subjects, the data were subjected to a one-way repeated-measures ANOVA for the statistical analyses. A total of 1,440 tokens were produced by 15 speakers (96 stimuli x 15 subjects) and the acoustic properties of word-final /t/ and /d/ were measured across these token words.

2.4 Results

2.4.1 Production of word-final /t,d/

Table 1 shows mean values of the acoustic parameters such as duration of preceding vowel (ms.), word-final /t,d/ stop closure duration (ms.), voicing duration (ms.), percentage of stop burst (%), and F0 of preceding vowel (Hz). As presented, mean duration of preceding vowel is 146 ms (SD = 38) for word-final /t/ and 227ms (SD = 55) for word-final /d/. Their difference reached significance ($F[1, 28] = 76.1$, $p < .0001$). This finding is consistent with those of many previous studies. For stop closure duration, its mean is longer for word-final /t/

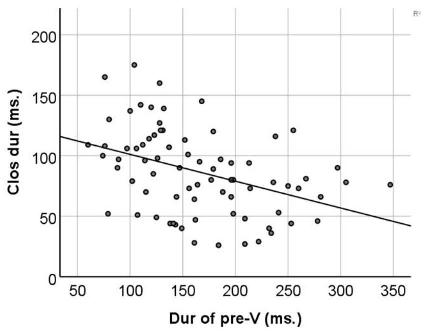
condition than word-final /d/ condition (100ms vs. 54ms, $F[1, 27] = 3217$, $p < .0001$). Furthermore, voicing duration is longer for word-final /d/ than word-final /t/ (42ms vs. 3ms, $F[1, 26] = 43.7$, $p < .0001$). An ANOVA was carried out with word-final consonant type (/t,d/) as within-subjects variable for the percentage of stop burst. Results showed that there is a significant effect (65% vs. 96%, $F[1, 28] = 9.28$, $p < .001$). Also the effect of word-final /t,d/ voicing on F0 of the preceding vowel approached significance (/t/:/d/ = 208Hz:167Hz, $F[1, 28] = 7.47$, $p = .01$). Additionally, the rates of stop burst were also significantly different. To be specific, 96% of word-final /d/ tokens were produced with burst whereas /t/-final words exhibited burst release with 65% ($F[1, 28] = 9.28$, $p < .01$). To confirm ANOVA results, linear mixed-effect models were run with estimates of four fixed effects and the results are presented in Table 1. As clearly illustrated, all four acoustic parameters were significantly different between final-/t/ and final-/d/ conditions and these findings are in line with ANOVA results.

Table 1. Results of Linear Mixed-effect Models for Four Acoustic Parameters for Word-final /t/ and /d/ Conditions

	Mean						
	/t/-final	/d/-final	Estimate	Std.Error	df	t-ratio	p-value
(Intercept)			227.1472	2.50	716	90.6	.000
Dur. of pre-V (ms.)	146	227	-81.0326	3.54	716	-22.8	.000
(Intercept)			54.8228	1.73	596	31.62	.000
Stop Clo Dur (ms.)	100	54	45.8263	2.69	596	17.02	.000
(Intercept)			42.4775	1.05	617	40.34	.000
Voicing Dur (ms.)	3	42	-39.4471	1.61	617	-24.4	.000
(Intercept)			167.7250	4.76	710	35.16	.000
F0 of pre-V (Hz)	208	167	41.1386	6.78	710	6.06	.000

The results for these four acoustic parameters, i.e., phonetic correlates of word-final voicing contrast are consistent with those of previous studies. English speakers produce word-final /t/ in the significantly differential manner from word-final /d/, thus preserving the underlying voicing contrast. Namely, word-final voicing contrast for /t, d/ is not lost but instead maintained with the acoustic traces of the underlying /t,d/. The five acoustic properties were statistically different between word-final /t/ and /d/ conditions.

In order to see whether these acoustic properties have any correlation with one another, Pearson correlation analyses were conducted. Analysis showed that there is a strong negative correlation between closure duration of final /t,d/ and duration of the preceding vowel ($r = -.141$, $p < .001$) as illustrated in Figure 1. It indicates that the longer the duration of preceding vowel, the shorter the duration of stop closure and voicing contrast between /t/ and /d/ is manifest in these two acoustic dimensions. Furthermore, it was revealed that a positive correlation between voicing duration and duration of preceding vowel also reached significance as illustrated in Figure 2. This implies that the longer the duration of preceding vowel is, the longer the voicing duration is. Analysis showed that there is a negative correlation between voicing duration and stop closure duration ($r = -.324$, $p < .01$). Taking these results together, we might conjecture that variants like word-final /d/ have shorter closure duration, are followed by longer vowel, and longer voicing duration than the variants of word-final /t/.



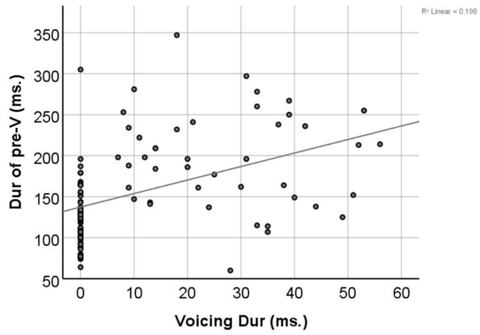
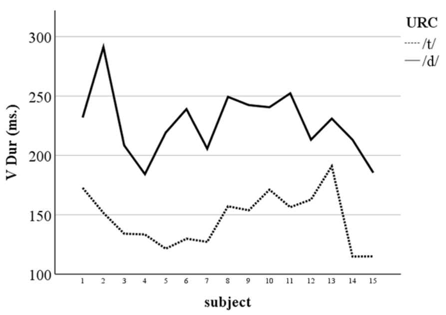
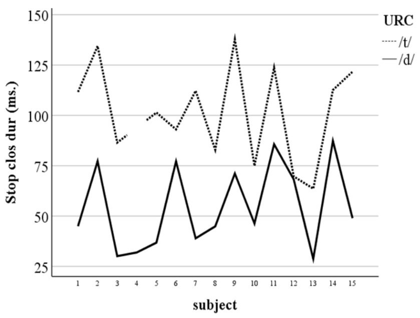


Fig 1 (left). Strong Negative Correlation between Duration of Preceding Vowel and Stop Closure Duration
Fig 2 (right). Strong Positive Correlation between Duration of Preceding Vowel and Voicing Duration

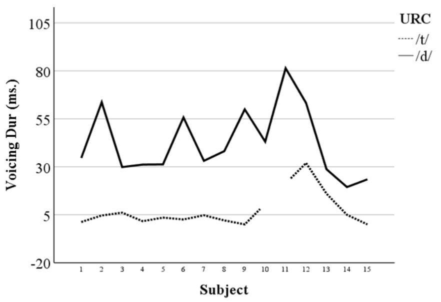
We also examined whether the siginficant differences between final-/t/ and final-/d/ are consistently observed among speakers. As seen in Fig. 3(a), vowels are longer before word-final /d/ than /t/ uniformly across all speakers. Stop closure duration also shows the consistent pattern of longer duration for final /t/ over /d/ condition in a uniform manner across speakers (Fig. 3(b)). Voicing duration patterns with the duration of preceding vowel without major speaker variation (Fig. 3(c)). Overall, words ending with /d/ show lower F0 of the preceding vowel than those with /t/ though minor speaker-dependent variability is observed as depicted in Fig. 3(d).



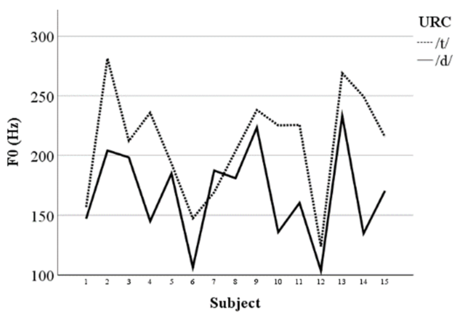
(a) Duration of preceding vowel



(b) stop closure duration



(c) voicing duration



(d) F0 of preceding vowel

Figure 3. (a) Duration of the Preceding Vowel by Subjects, (b) Stop Closure Duration by Subjects, (c) Voicing Duration by Subjects, and (d) F0 of the Preceding Vowel by Subjects. (Solid lines refer to word-final /d/ condition while dotted lines to word-final /t/ condition along all four dimensions.)

To sum up, word-final /t,d/ contrast in English is preserved for most acoustic correlates for voicing for all speakers who participated in this production study. Acoustic results suggest that word-final devoicing does not occur in American English unlike German final devoicing.

2.4.2 Production of /t/-flaps and /d/-flaps

Since one of the main purposes of the present study is to draw a more comprehensive picture of nature of English flapping, we investigated whether /t/ flaps and /d/ flaps show differences along a multitude of acoustic dimensions. Before looking into the detailed acoustic properties, all the tokens were aurally judged by a phonetically-trained researcher to see if the underlying word-medial /t,d/ are realized as flaps /ɾ/. In total, 94% of all tokens were judged as flaps. 89% of underlying /t/ stimuli were rated as flaps (SD: 0.3) whereas all /d/-tokens were considered as flaps at 100% (SD: 0.05). This difference in the judgement rates reached significance (F[1, 686] = 40.9, p = .000). This indicates that word medial-/d/s are more likely to be realized as flaps than word medial-/t/s in English. Our finding is consistent with Herd et al.'s (2010) finding that the former were flapped at 99% whereas the latter at 76% though the degrees are slightly different. Figure 4 illustrates that flapped /t/- and /d/-tokens overlap in duration; however, most of the tokens show below 50ms, which has been known as the boundary between alveolar stops and flaps. Accordingly, this high percentage (above 90%) of flapped tokens below 50ms validates the auditory judgment of flaps.

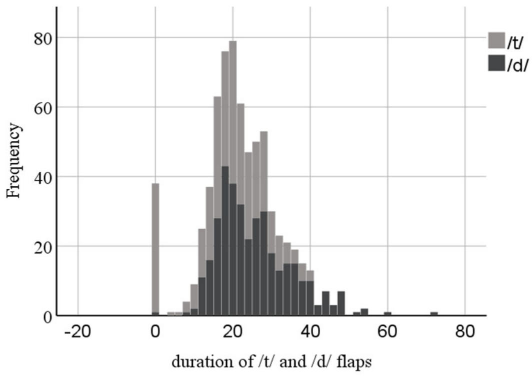


Figure 4. Distribution of Flapped /t/ and /d/ Tokens for 15 Speakers

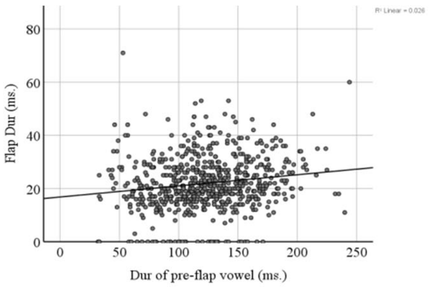
Table 2 presents six acoustic parameters of two types of flaps, including the duration of pre-flap vowel, flap closure duration, voicing duration, flap duration, VOT, duration of post-flap vowel, and F0 of pre-flap vowel. To determine whether word-medial /t,d/ are completely neutralized into flaps, these acoustic measures were subject to linear mixed-effect model with underlying word-medial /t,d/ contrast as fixed effect. First, vowels preceding /d/ flaps are 15ms longer than those preceding /t/ flaps like previous studies (Bravers 2014, Herd et al. 2010) (e.g., *wading* >> *waiting*). Second, duration of voicing portion during the flap closure is 8ms longer for /d/ flaps than for /t/ flaps. Third, the duration of /d/-flapped consonants is 7ms longer than that of /t/-flapped consonants. Fourth, VOT is longer for /t/ flaps than for /d/ flaps (25ms vs. 15ms). Additionally, vowels following /d/ flaps are significantly longer than those following /t/ flaps (e.g., *seeding* >> *seating*). Interestingly, both vowels flanking flaps are longer for /d/ flaps than /t/ flaps. The ratio of the duration of pre-flap vowel to the duration of post-flap vowel is reported as another cue to make a distinction between /t/ flaps and /d/ flaps (Herd et al. 2010). Our finding replicates Herd et al. (2010) by showing that the ratio is greater for /d/-flapped words than for /t/-flapped words (e.g., 1.07 vs. 1.01). F0s of the vowels preceding /t/ flaps, however, are not different from those preceding /d/ flaps.

Table 2. Results of Linear Mixed-effect Models for Six Acoustic Parameters for /t/-flaps and /d/-flaps

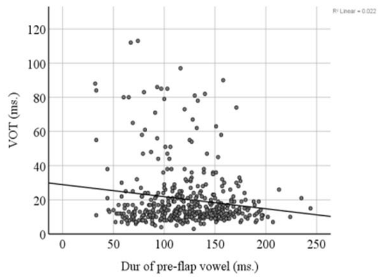
	Mean						
	/t/-flaps	/d/-flaps	Estimate	Std.Error	df	t-ratio	p-value
(Intercept)			130.6256	1.99	688	65.5	.000
Dur. of pre-flap V (ms.)	115	130	-15.3485	2.87	688	-5.34	.000**
(Intercept)			24.9525	.50	688	49.74	.000
Voicing Dur (ms.)	18	24	-5.9886	.72	688	-8.28	.000**
(Intercept)			25.0530	.49	688	50.15	.000
Flap dur (ms)	18	25	-6.3813	.72	688	-.8.86	.000**
(Intercept)			15.4107	1.14	419	13.41	.000
VOT (ms)	25	15	10.4014	1.67	419	6.19	.000**
(Intercept)			121.5167	1.68	688	71.68	.000
Dur of post-flap V	113	121	-8.5017	2.44	688	-3.47	.000**
(Intercept)	161	161	161.48	3.03	680	53.13	.000
F0 of pre-V (Hz)			-.0864	4.37	680	-.02	.984

These acousting findings indicate that the underlying word-final /t,d/ leave its traces in the word-medial flaps on the surface along a majority of acoustic correlates associated with voicing. They suggest that flapping in American English is characterized as an incomplete and gradient rule.

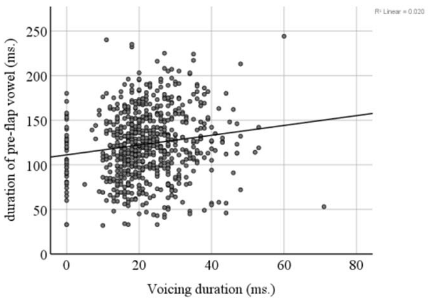
We examined whether there are correlations among these acoustic properties. Pearson correlation analysis showed that there is a positive correlation between flap duration and duration of pre-flap vowels (r = .161, p < .001) as illustrated in Figure 5 (a). This shows that the longer the duration of pre-flap vowels is, the longer flap duration becomes. Furthermore, it was shown that duration of pre-flap vowel has a negative correlation with VOT of flaps (r = -.149, p = .002). This finding suggests that the longer the duration of pre-flap vowel is, the shorter the VOT of flaps is. It makes sense that flapped-/d/ tokens tend to have longer duration of preceding vowel and shorter VOTs than flapped-/t/ tokens. Figure 5(c) shows a positive correlation between the duration of voicing and the duration of pre-flap vowel. It exhibits that the longer vowel tends to have longer voicing portion (r = .142, p = .000). Analysis showed that duration of pre-flap vowel has a positive correlation with duration of post-flap vowel as illustrated in Figure 5(d) (r = .258, p = .000). These results regarding significant correlations among five acoustic measures for intervocalic flaps indicate that they contribute to the maintenance of the underlying word-final /t,d/ contrast.



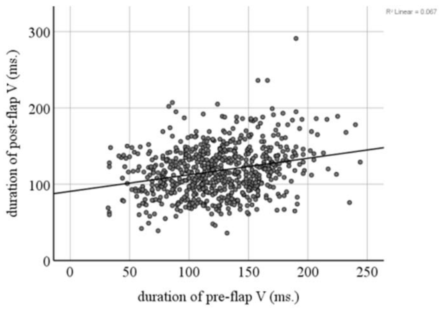
(a)



(b)



(c)



(d)

Figure 5 a. (top left). Positive Correlation between Duration of Pre-flap Vowel and Flap Duration
5 b. (top right). Negative Correlation between Duration of Pre-flap Vowel and VOT
5 c. (bottom left). Positive Correlation between Duration of Pre-flap Vowel and Voicing Duration
5 d. (bottom right). Positive Correlation between Duration of Pre-flap Vowel and Post-flap Vowel

Finally, we examined whether speaker variation exists for all the acoustic dimensions assoicated with /t/ flaps and /d/ flaps. As can be plainly seen in Figure 6, speaker variability emerges in all these acoustic measures. For example, some speakers reveal a larger mean differrence than other speakers between /t/ flaps and /d/ flaps in the duration of pre-flap vowels, voicing duration, flap duration, VOT and the duration of post-flap vowel.

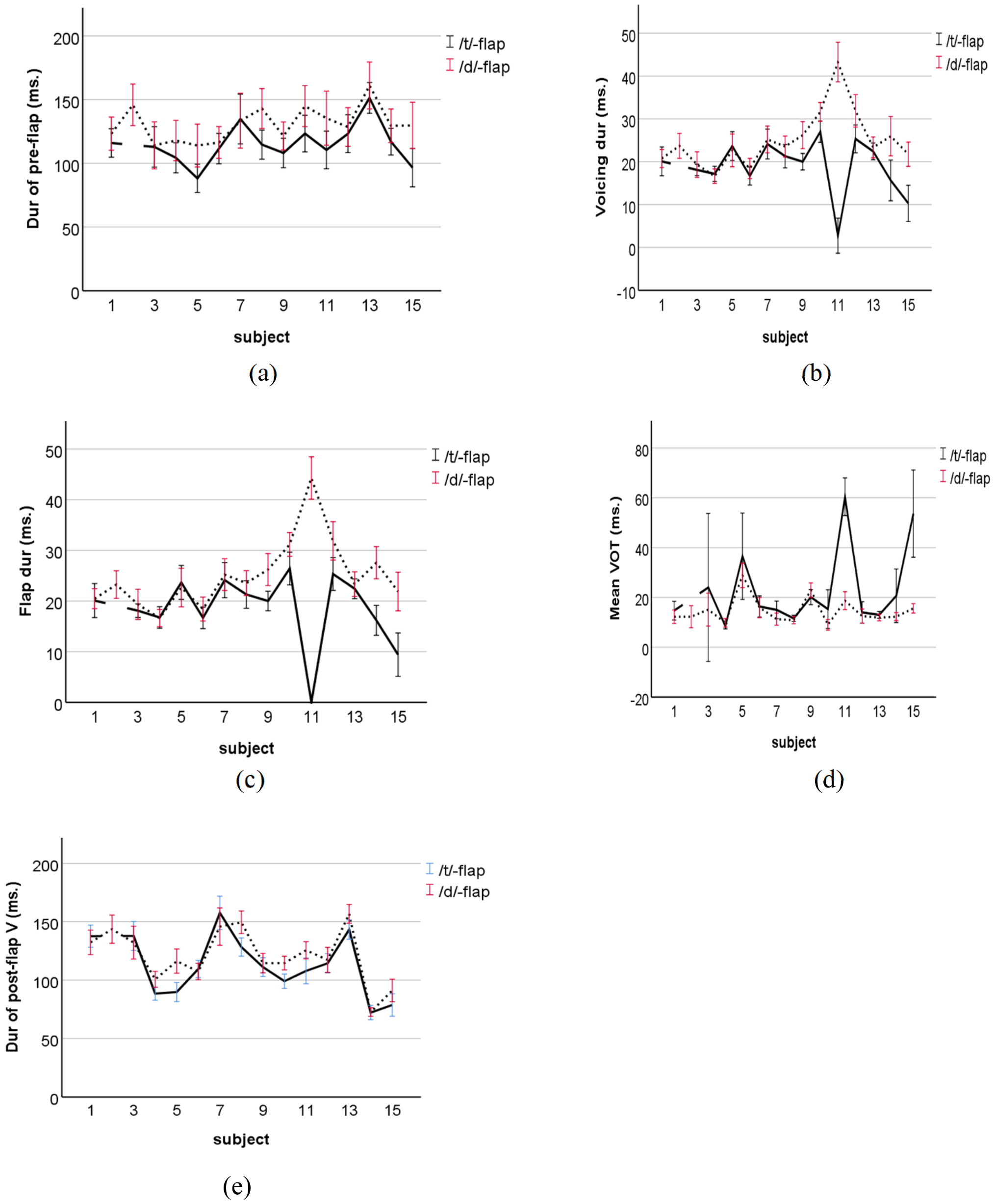


Figure 6. **(a) Duration of the Preceding Vowel by Subjects, (b) Voicing Duration by Subjects, (c) Flap Duration by Subjects, and (d) VOT by Subjects, (e) F0 of the Preceding Vowel by Subjects** (Solid lines refer to /t/-flap while dotted lines to /d/-flap condition along all five dimensions.)

To sum up, we investigated the acoustic correlates of word-final /t,d/ voicing contrast and explored whether such differences are maintained in word-medial /t/ flaps and /d/ flaps in order to determine whether English flapping is complete neutralization. Table 3 compares the acoustic properties of word-final /t,d/ with those of word-medial /t,d/ flaps. First, voicing contrast in word-final /t,d/ is evident in four distinct acoustic properties. Second, /t/ flap and /d/ flap differ in many acoustic measures such as duration of pre-flap vowels, flap durations, and voicing durations. These results can be interpreted to suggest that flapping in American English is mostly an incomplete neutraliation. Third, some acoustic property such as F0 of preceding vowel is completely neutralized in a flapping environment. Although other three acoustic dimensions preserve significant contrast between /t/ flap and /d/ flap, their differences in bimorphemic disyllables are smaller than those in monomorphemic words. For example, the durational gap between word-final /t/ and /d/ is 81ms whereas the difference in the duration of pre-flap is 15ms between /t/ flaps and /d/ flaps. In addition to the decrease of the gap between two types of flaps, the durations of the preceding vowels in a flapping environment are shorter, compared to those in word-final context. These reductions in the magnitude seem to stem from the application of polysyllabic shortening in English.

Table 3. Mean Values of Acoustic Parameters for Word-final /t,d/ and Word-medial /t,d/

	Word-final		Sig.	Word-medial		Sig.
	/t/	/d/	(diff)	/t/ flaps	/d/ flaps	(diff)
Dur. of pre-V (ms.)	146	227	✓ (81)	115	130	✓ (15)
Stop Clo Dur (ms.)	100	54	✓ (46)	18	25	✓ (7)
Voicing Dur (ms.)	3	42	✓ (39)	18	24	✓ (6)
F0 of pre-V (Hz)	208	167	✓ (41)	161	161	× (0)

3. Experiment 2: Perception of Word-final and Word-medial, Flapped /t,d/

3.1 Participants

The English-speaking participants were 24 university graduate students or staff enrolled at Cornell University in the US at the time of the experimentation. They were all native speakers of American English and used a variety of dialects, including West Coast, East Coast, Southern, Midwest, etc. Of these speakers, 10 also participated in the production experiment in the present study. All subjects were paid compensation for their participation in the perception experiment. Out of these participants, 15 were female and 9 were male, and none of them had speech or hearing impairment. Their mean age was 23 years old, ranging from 18 to 56 (SD: 7.9 years).

3.2 Materials

We used three groups of stimuli for perceptual materials in three blocks of experiments. First, to see whether word-final /t,d/ contrast is perceptible, 24 minimal pairs with final /t,d/ were selected as the stimuli for perception. These 48 words are identical to those mobilized in the production experiment administered in this study (See the Appendix). Twelve words end with voiceless alveolar /t/ whereas the other half of the pairs end with its voiced counterpart /d/. All of the minimal pairs contain the identical twelve vowels before the word-final /t,d/. These tokens were recorded by a native speaker of English and used as the listening stimuli in the perception experiment, i.e., AB Identification test.

In addition, to test whether the flapped words are completely neutralized perceptually, 24 minimal pairs were used for AB Identification test. They contain word-medial /t,d/ in a flapping environment where they are preceded by a stressed vowel and followed by an unstressed vowel. They are all disyllabic real words and identical to those employed in the production experiment (See the full list in the Appendix). As described previously, these pairs of words are morphologically derived words formed by attaching the suffixes like "~ing, ~ist, ~est, ~er, ~y". These /t/-flapped and /d/-flapped words balance in word length, and the suffixes and belong on average to high-frequency word groups.

Table 4. Word Pairs Used in Perceptual Experiment

	Word freq.	No of syllables	suffixes
/t/ flapped words	13,954	2	-ing, -est, -y, -er
/d/ flapped words	3,274	2	-ing, -ist, -y, -er

Table 5 shows the mean values of the acoustic properties of 48 /t/- and /d/-flapped words. The 24 pairs of words recorded by an English native speaker did not significantly differ in the five acoustic measures. These listening stimuli pairs seem to be completely neutralized and accordingly appropriate for AB Identification test. The duration of pre-flap vowel, flap duration, voicing duration, VOT and F0 of pre-flap vowel were controlled and balanced.

Table 5. The Acoustic Properties of Word Pairs Used in the Perceptual Experiment

	Dur. of prec.V (ms.)	Flap duration (ms.)	Voicing duration (ms.)	VOT (ms.)	F0 of prec.V (Hz)
/t/ flapped words	115	20	20	14	156
/d/ flapped words	123	20	20	12	163
	$p>.05$	$p>.05$	$p>.05$	$p>.05$	$p>.05$

The last group of listening stimuli for our perceptual study were manipulated to see whether the duration of the pre-flap vowel plays a crucial perceptual cue to identify medial-/t/ or medial-/d/. Five words in a flapping context were selected and recorded to create five continua of 10 steps by manipulating the length of the preceding vowel before word-medial /d/: {*beading, rider, kidding, padding, wading*}. Based on the recording of each of these words, five series of continua were synthesized. Aside from the length of the preceding vowel, other acoustic properties including formant frequency values, intensity, or aspiration remained intact. For example, the duration of (120ms) of the original vowel, /i/ in "beading" increased step-wise by lengthening the vowel by 10ms. Accordingly, a continua with 10 tokens were created for each word as exemplified in Table 6.

Table 6. Parameter Settings of the Vowels for the Resynthesized Materials: {*beading, rider, kidding, padding, wading*}

	Word	V-Dur	Word	V-Dur	Word	V-Dur	Word	V-Dur	Word	V-Dur
Step 1	beading	120	rider	175	kidding	70	padding	140	wading	195
Step size	↓	+10	↓	+10	↓	+10	↓	+10	↓	+10
Step 10	beading	210	rider	265	kidding	160	padding	230	wading	285

To sum up, the recordings of five stimuli served as the basis for generating five series of synthesized continua. These continua were created by lengthening the duration of the vowel preceding word-medial /d/. Accordingly, a total of 50 synthesized stimuli were employed as perception materials in the identification test and presented to 24 English native speakers. Thus, 1,200 responses were collected for analyses (50 tokens × 24 subjects).

3.3 Procedures

The perception experiments proceeded in three blocks consecutively. All participants took part in three identification tests in three blocks. In block I, Identification test was conducted to see whether English native speakers perceive word-final /t,d/ distinction. All participants were seated in front of a computer screen at sound-attenuated booth in a phonetics lab at Cornell University in the US. They were asked to listen to an English word ending with /t,d/ and to judge what word they heard. Two choices appeared on the computer screen where ① was a /t/-final word and ② /d/-final words (e.g., ① beat ② bead). These options were presented on the screen simultanenously with each listening stimulus. A total of 2304 listening tokens were presented one by one (24 minimal pairs × 2 repetitions × 24 subjects). Participants were asked to press the button 1 or 2 on the keyboard as quickly and accurately as possible. All the procedures of the perception task, including the randomization of the listening stimuli and the running of the procedure, including a pause among the trials was carried out with E-Prime 2.0 Professional.

After finishing the perceptual task in block I, participants proceeded to block II. Since another goal of the task is to examine whether English native listeners distinguish /t/ flaps from /d/ flaps, words that contained medial /t,d/ in a flapping environment were presented to them (e.g., "bedding" vs. "betting"). A total of 2304 listening tokens were presented (24 minimal pairs × 2 repetitions × 24 subjects). In the similar fashion to block I, all participants were instructed to respond to each listening token and press 1 or 2 once they made a judgment. They were forced to choose one between ① word-medial /t/ and ② word-medial /d/ (e.g., ① boating ② boding).

In block III, the identical subjects were instructed to listen to each of manipulated token differing in the vowel length and to decide whether it is word-medial /t/ or /d/ word embedded in a flapping context. They were asked to press 1 on the keyboard if what they heard was /t/-flapped word (e.g.,① boating) or to press 2 if it was /d/-flapped word (e.g., ② boding). Once they pressed the button, next token was presented aurally and two choices appeared on the screen simultaneously. A total of 1,200 tokens were presented (5 word continua × 10 steps × 24 subjects).

For analyses, the accuracy of the identification of the underlying word-final /t,d/ word was automatically obtained through E-prime and subjected to repeated-measures ANOVA as the dependent factor with underlying voicing as the independent factor in block I. In block II, we collected the accuracy of the identification of the underlying word-medial /t,d/ responses and analysis was carried out to see if there is any significant difference in the perception of two types of responses. In block III, the rates of responses for voiceless words (e.g., *beating, writer, kitting*, etc.) were calculated to examine how the tokens were identified and perceived as the underlying /t/ or /d/ words.

3.4 Results

3.4.1 Perception of word-final /t,d/ contrast

First, word-final /t/ tokens were identified as /t/-final words at 95% of accuracy and word-final /d/ stimuli were perceived as words ending with /d/ at 97% of accuracy. Accuracy did not differ according to the voicing of word-final alveolar ($F[1, 46] = 1.38$, $p > .05$). However, analysis of reponse times showed that word-final /t/ tokens led to faster RTs than their /d/ counterparts (1.6 sec vs. 1.7 sec, $F[1, 46] = 2.26$, $p < .05$). These findings suggest that English native listeners successfully perceive underlying voicing contrast /t,d/ which is reflected with remarkablly distinct acoustic parameters on the surface pronunciation. It seems that acoustically incomplete neutralization leads to complete perceptual distinction. Since there are sufficient acoustic cues to make a distinction between an underlying /t/ and /d/, it is evident that English native speakers have no difficulty identifying the underlying contrast by relying on those acoustic cues.

Furthermore, the uniform tendency of high accuracy of perception of word-final /t,d/ contrast is observed across speakers as illustrated in Figure 7. Mean accuracy was consistently higher for the perception of word-final /d/ than for that of word-final /t/ words with a greater majority of speakers. RTs were faster for the processing of word-final /t/ tokens than for word-final /d/ condition across speakers, showing little interspeaker variation as clearly delinenated in Figure 8. In a nutshell, it might be the case that although word-final /t/ tokens are responded to faster than word-final /d/ tokens for their perception, English speakers show the ceiling effect of accuracy for the identification of word-final /t,d/ contrast.

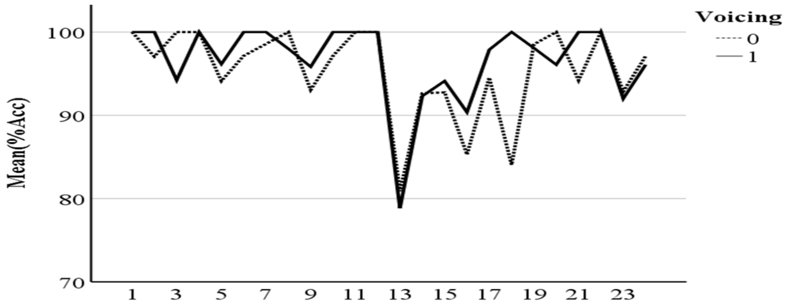


Figure 7. Accuracy by Subjects and Word-final /t,d/ Voicing (solid lines refer to percent accuracy of word-final /d/ tokens while dotted lines to that of word-final /t/ tokens)

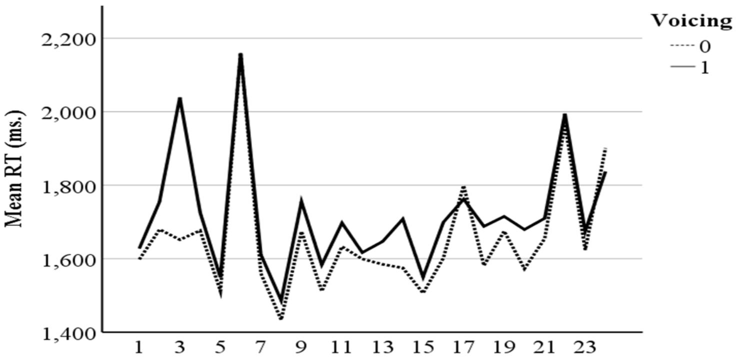


Figure 8. RTs by Subjects and Word-final /t,d/ Voicing (solid lines refer to RTs for word-final /d/ tokens while dotted lines to that of word-final /t/ tokens)

3.4.2 Perception of word-medial /t,d/ contrast

A two-way (2 × 24) repeated-measures ANOVA (Underlying voicing contrast × Word pairs) was conducted on the percent of correct responses. One goal was to see whether /t/-flapped words and /d/-flapped words are identified with different accuracy on the recovery of the underlying /t/ or /d/. Another purpose was to examine whether listeners recover the underlying /t,d/ in a differential manner according to individual word pair.

First, a main effect of underlying voicing contrast was observed. Analysis exhibited that significant differences were found on the accuracy of identification between /t/- and /d/-flapped words (F[1, 46] = 27.72, $p < .0001$). /d/-flapped words (e.g., "wading") were perceived more correctly at 63% than words containing an underlying /t/ (e.g., "waiting") at 48%. Near-chance rates of perception for /t/-flapped words indicate that they suffer from difficulty due to complete neutralization on the perception side. On the other hand, comparatively higher percentage of identification of the underlying /d/-medial words suggests that listeners seem to be susceptible to the recovery of voiced ones based on the surface voiced flaps. Both types collpased, overall correct percentages fell near chance at 55.5%, implying that on the perception side, /t,d/ flapping seems to have reached complete neutralization. RT data also show that there were no significant differences when listeners judged whether a flap is an underlying /t/ or /d/ medially (/t/-flapped words:/d/-flapped words=2033 ms:1975 ms, F[1, 46] = .17, $p > .05$).

In addition, we looked into the degrees of perceptual neutralization across individual /t,d/-medial word stimuli. Table 7 shows mean % correct for individual word pairs. Analysis showed that a main effect of word pair also reached significance. As is illustrated, percent of correct responses was significantly different by word pairs (F[23, 48] = 7.03 $p < .0001$). A majority of word pairs containing /t/ flaps and /d/ flaps were perceived near chance level below 55% except for five pairs (e.g., "hooting-hooding, wooting-wooding, writer-rider, rating-raiding, waiting-wading"). This finding suggests that English native listeners cannot reliably distinguish flapped /t/ from flapped /d/ although they realize the incomplete flapping on the production side as reported previously in the present study.

Table 7. Mean % Correct of Individual Word Pairs

Word pairs	Mean (% correct)	Std Dev	Std Err	Word pairs	Mean (% correct)	Std Dev	Std Err
hooting-hooding	91.7	27.8	2.84	louty-loudy	53.1	50.2	5.12
wooting-wooding	83.3	37.5	3.82	clouty-cloudy	53.1	50.2	5.12
writer-rider	75	43.5	4.44	sighting-siding	52.1	50.2	5.13
rating-raiding	66.7	47.4	4.84	retting-redding	52.1	50.2	5.13
waiting-wading	62.5	48.7	4.97	kitting-kidding	51	50.3	5.13
beating-beading	56.3	49.9	5.09	rutty-ruddy	50	50.3	5.13
betting-bedding	56.3	49.9	5.09	seating-seeding	49	50.3	5.13
butting-budding	55.2	50	5.1	cotter-codder	49	50.3	5.13
bitting-bidding	54.2	50.1	5.11	patting-padding	47.9	50.2	5.13
boating-boding	54.2	50.1	5.11	fattest-faddist	44.6	50	5.1
coating-coding	54.2	50.1	5.11	potter-podder	41.7	49.6	5.06
mooty-moody	53.1	50.2	5.12	putting-pudding	28.1	45.2	4.61

Furthermore, an interaction between underlying contrast and word pair was also found (F[23, 48] = 21.52, $p < .0001$). While some medial /t/ words were perceived more correctly than their counterpart /d/ words (e.g., "waiting-wading, rating-raiding"), other pairs showed the opposite pattern (e.g., "hooting-hooding, mooty-moody, wooting-wooding").

Word frequency did not have a consistent effect on the perception of the underlying /t,d/. To be specific, word pairs of lower frequency such as "hooting-hooding, wooting-wooding" benefited comparatively higher rates of perception. However, another group of higher frequency pairs such as "writer-rider, rating-raiding, waiting-wading" were hurt in the perception. Furthermore, another higher word frequency group such as "putting-pudding" was not perceived correctly with regard to the restoration of the underlying /t,d/ contrast.

Overall, the perception experiment shows that listeners suffer difficulty differentiating flapped /t/ and /d/ contrast although native English speakers make a distinction between /t/ flaps and /d/ flaps acoustically. This mismatch seems to arise due to the robustly delicate differences in the acoustic contrast in the realization of the underlying /t/ and /d/, i.e., subtle incomplete neutralization.

3.4.3 Perception of pre-flap vowel-length continuum

The factors of vowel length and words collapsed, the overall percentages of /t/-word responses were quite low at 35% as summarized in Table 8. This /d/ bias in native English listeners' responses (65%) might be due to the fact that the listening stimuli were synthesized from the medial /d/ words and they preserved /d/-related acoustic properties except for the variation of pre-flap vowel length. The rates of responses for /t/ words varied according to each word continuum (F[4, 795] = 11.14, p = .000).

Table 8. Mean % Responses for Word-medial /t/ Words by Individual Word Series

Stimuli	No.	Mean%	Std Dev(%rsps)	Std Err(%rsps)
beading~beating	160	44.4	49.8	3.94
rider~writer	160	30	46	3.63
kidding~kitting	160	19.4	39.6	3.13
padding~patting	160	32.5	47	3.71
wading~waiting	160	50.6	50.2	3.96

Figure 9 displays the identification results for the five synthesized word continua based on the manipulations of the length of pre-flap vowels. Each data point represents the mean percentages of responses for /t/-words such as "beating, writer, kitting, patting, waiting" as the pre-flap vowels lengthen by 10ms in a 10-step series. As clearly illustrated in Figure 10, three series of continua, e.g., {beading, rider, kidding} induced an inverse relation between the length of pre-flap vowels and /t/ judgment. The longer the pre-flap vowels were, the less the stimuli were perceived as /t/ flaps. For "beading" series, the length of pre-flap vowels affected the rates of responses for /t/ flapped words (F[9, 135] = 2.72, p < .01). The durations of pre-flap vowels being 120ms or 130ms (step 1 or 1), the stimuli were perceived as "beating" whereas the stimuli of longer pre-flap vowels tend to be judged rather as "beading". "rider" series of stimuli did not yield a significant effect of pre-flap vowel length on the perception of /t/ words (F[9, 135] = 1.86, p > .05). Regardless of whether pre-flap vowels varied along 10 steps, the stimuli were dominantly judged as "rider" rather than "writer". Although "kidding" series of stimuli were extremely likely to be judged as "kitting" at 30%, the degrees of length of vowel before the flap influenced the perception (F[9, 135] = 2.11, p < .05). That is, the longer the pre-flap vowels, the greater likelihood of perception of "kidding".

However, two series of listening continua, e.g., "padding, wading" did not yield the effect of the pre-flap vowel length on the perception of /t/ or /d/ words. Only the step 1 stimulus whose pre-flap vowel was 140ms was judged as "patting" at 62% whereas all other steps of stimuli invoked /t/-word perception below 45% (F[9, 135] = 1.7, p > .05). Likewise, an inverse relation between the durations of pre-flap and /t/ judgment did not emerge for "wading" series of synthesized stimuli (F[9, 135] = 0.85, p > .05).

These findings indicate that English native listeners' availability of the durations of the pre-flap vowels varies across the individual flapped words. Furthermore, the inverse relation between the length of pre-flap vowels and the /t/ judgment did not vary as to whether the vowel is a monophthong or diphthong. This result is not consistent with the suggestion made by Malécot and Lloyd (1968) that listeners may use the durations of monophthongs as cues for the judgment of word medial /t,d/, not the differences of pre-flap diphthongs.

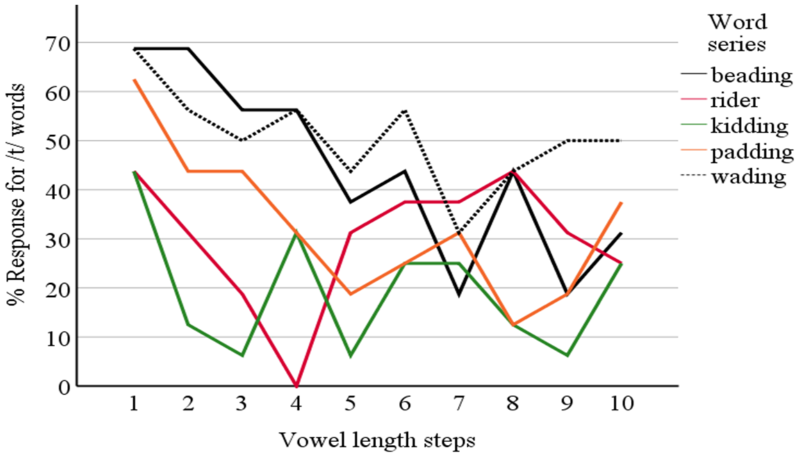


Figure 9. Percentages of /t/ Flapped Word Responses by the Length of Pre-flap Wowels for Five Word Continua Stimuli (i.e., beading, rider, kidding, padding, wading)

Taking together the subjects and the stimuli words, the perceptual identification of /t/ flap words seems to be affected by the manipulation of pre-flap vowels as seen in Figure 10. Specifically, the stimuli with longer pre-flap vowels were judged more as /d/ flapped words rather than /t/ flapped counterparts although the overall pattern reveals /d/-bias in listeners' responses ($F[9, 775] = 3.36$, $p < .001$).

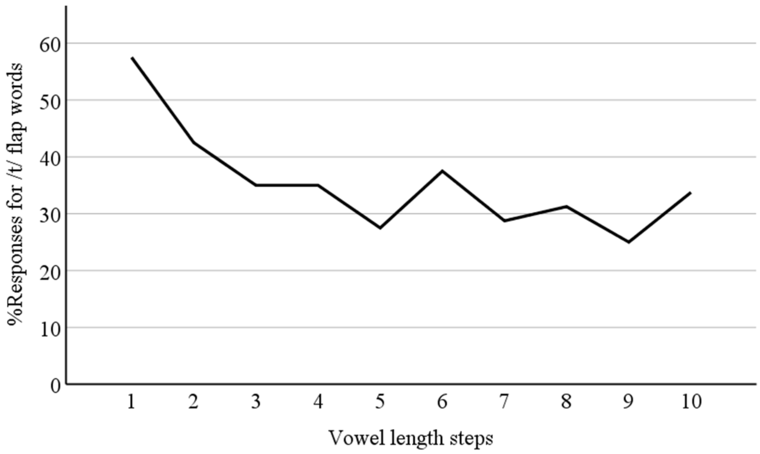


Figure 10. Percentages of /t/ Flapped Word Responses by the Length of Pre-flap Vowels

To sum up, the durations of the pre-flap vowels seem to function as the perceptual cues to distinguish /t/-flaps from /d/-flaps to a limited degree, especially depending on the individual pair words. This inverse relation between the length of pre-flap vowels and word medial /t/ or /d/ judgment is in line with the findings of production experiments, i.e., the vowel length being an indicator of the underlying voicing contrast (Charles-

Luce 1997, Herd et al. 2010). However, overall low percentages above seem to be indicative of lack of reliability of the pre-flap vowel length as a major perceptual cue to the decision.

4. General Discussion and Conclusion

In the present study, we administered production and perception experiments to see whether word medial /t/ and /d/ in a flapping environment are completely neutralized on the acoustical side and whether English native listeners can reliably distinguish a flapped /t/ from a flapped /d/. In this section, we first discuss evidence for English flapping as incomplete neutralization coupled with its acoustic properties and offers a phonological account for the incomplete flapping (Section 4.1). Next, we address complete flapping on the perception side based on the perception experiment, concluding this study (Section 4.2).

4.1 Acoustically Incomplete Flapping

In the acoustic experiment, we first found that the underlying /t,d/ contrast in word final position is substantially manifested in a wide range of acoustic dimensions. For example, the durations of preceding vowels were longer before word-final /d/ than before word-final /t/; voicing duration during stop closure was longer for /d/ than for /t/; closure duration was longer for /t/ than for /d/; F0s of preceding vowels were significantly higher for /t/ than for /d/. These marked contrasts confirm word-final voicing distinctions in English obstruents whose patterns are in opposition to word-final devoicing in German or Polish.

In addition, it was found that the underlying /t,d/ voicing contrast word-medially, i.e., in a flapping context, is still not completely neutralized unlike the traditional definition of categorical flapping expects. As clearly shown previously in Table 3, despite the shrinkage of the magnitude of the differences in voicing contrasts, the acoustic traces of the underlying opposition are plainly evident even in a flapping environment. Specifically, the vowels preceding /d/-flaps are 15ms longer than the pre-flap /t/ vowels (130ms vs. 115), replicating the finding of Patterson and Connine (2001); flap durations of /d/ flaps are longer than those of /t/ flaps (25ms vs. 18ms) although no significant difference was found in Zue and Laferriere (1979); voicing durations of /d/ flaps are longer than those of /t/ flaps (24ms vs. 18ms); VOTs of /t/-flaps are longer than those of /d/ flaps (25ms vs. 15ms); however, no difference was found in F0 of the vowels preceding /t,d/ flaps. These findings are consistent with those of previous production studies (Fox and Terbeek 1977, Herd et al. 2010, Patterson and Connine 2001, Sharf 1962, Stathopoulos and Weisner 1983, Turk 1992, Zue and Laferrier 1979). These acoustic correlates of voicing between word-final and word-medial position being considered, it makes sense to conclude that the application mode of flapping is dominantly incomplete at phonetic-fine details although flapping is prevalent in many English dialects (Eddington and Elzinga 2008).

The incomplete flapping found in this study provides some phonological implications in rule-based phonology or in constraint-based approach. First, the finding that the vowels preceding /d/ flaps are longer than those before /t/ flaps leads to the postulation that voicing induced vowel lengthening rule precedes flapping from the perspective of rule interaction as suggested in Fox and Terbeek (1977). This rule ordering gives an account for the opacity of vowel lengthening evident in “seeding >> seating”; (1) vowel lengthening (e.g., in “seeding”), (2) affixation (e.g., “seeding, seating”), and (3) flapping (e.g., “see[ɾ]ing, sea[ɾ]ing”). On the contrary, if flapping is applied before vowel lengthening in “seeding, seating”, the durational contrast in pre-flap vowels loses ways to be explained.

However, a closer scrutiny of the durational differences of pre-/d/ and /t/ flaps reveals that pre-flap vowels in disyllabic words are markedly shorter than the identical vowels in their corresponding monosyllabic words as described in Table 9. The differences in both the conditions reached significance (123ms vs. 187ms, $F[1, 1393] = 554.7$, $p < .0001$). In particular, the vowels preceding /d/ flaps in disyllabic words become mostly half shorter than the same vowels before word final /d/ in monosyllabic words. Phonological tense vowels seem to have undergone a dramatic reduction and be shorter than their corresponding lax vowels. Seemingly phonological tense vowels have become similar to lax vowels at phonetic-fine level. This might be due to the application of polysyllabic shortening in English whereby the stressed vowel is progressively shorter as the word becomes longer with the increased number of syllables (e.g., *spéed* >> *spéedy* >> *spéedily*, Port 1981, White and Turk 2010).

Table 9. Mean Vowel Durations (ms.) in Monosyllabic and Disyllabic Words

Monosyllabic words	Dur. (ms.)	Disyllabic words	Dur. (ms.)
bead:beat	228:151	beading:beating	132:116
bed:bet	203:139	bedding:betting	119:112
bid:bit	190:127	bidding:bitting	98:77
bode:boat	287:173	boding:boating	162:145
bud:but	198:150	budding:butting	121:110

Based on the finding regarding the shortened durations of pre-flap vowels in word medial position, we suggest that vowel lengthening is phonologically transparent in disyllabic words unlike the position of the rule-based approach, but phonetically opaque as is evident in the incomplete neutralization of flapping. (See a phonological OT-based account for English flapping and its phonological variation in Lee (2022)). Contrast preservation at phonemic level is lost but still holds true at phonetic-fine details in a flapping environment. By capturing the results of incomplete flapping at phonetic-fine level found in our study and incorporating lexically oriented constraint (e.g., Preserve(lexical contrast) into formal optimality theory, we propose a formal account for incomplete flapping and its accompanying phonetic contrast. Relevant constraints are explicated in (1). Flapping is triggered by a markedness constraint (1a) which induces the neutralization of /t,d/ contrast. (1b) is a type of an umbrella constraint which is responsible for the shortening of stressed vowels in longer words. Preserve(LexCont) is a lexically driven constraint which demands that lexical contrast should be preserved at phonemic or phonetic level. This is a constraint that plays a key role in maintaining the contrast in the underlying voicing of obstruents.

(1) Relevant constraints

a. *V{t/d}V: Intervocalic {t,d} are not allowed.

b. Polysyllabic Shortening (P-Short): Stressed vowels get shorter in longer words.

c. Preserve(LexCont): Preserve lexical contrasts.

d. *ɾ: a flap sound is not allowed.

Coupled with these constraints are the constraint ranking delineated below to give an account for incomplete flapping with respect to the durations of the pre-flap vowels. Candidate (2a) pair words preserve the underlying voicing contrast and thus pre-/d/ vowel is realized as longer than pre-/t/ vowel. These pairs are exempt from flapping and thus lose out fatally, violating the flapping-triggering markedness constraint. Both the /t,d/ words in candidate (2b) have undergone complete neutralization, and as a result have flaps along with the same length of the pre-/t,d/ flap vowels. However, the durations of the preceding vowels still are the same as those of the underlying tense vowels as polysyllabic shortening has not been applied. Consequently, they fatally fail to satisfy a higher ranked phonological contraint, P-Short. Candidate words in (2c) seem to be totally same as those in (2b) at phonemic level, but they are not in that both the pre-flap vowels are shortened as a result of the application of polysyllabic shortening. These words, however, do not survive the fierce competition since they fail to preserve the lexical contrast by a way of maintaining the voicing-induced vowel length difference. The candidate words in (2d) are selected as the final winner by satsifying the relevant phonological, lexical and phonetic constraints. These words take advantage of flapping (i.e., /t,d/ flaps), shortening of the pre-flaps due to the increased number of syllables, and incomplete neutralization by keeping the trace of underlying voicing contrast in pre-flap vowels.

(2) /seeding/–/seating/ => $[\text{siɾɪŋ}]_d$ – $[\text{siɾɪŋ}]_t$

"seeding-seating"	*V{t/d}V	P-Short	Pres(LContrast)	*ɾ
a. $[\text{sidɪŋ}]_d$ >> $[\text{sitɪŋ}]_t$	*!*	**		
b. $[\text{siɾɪŋ}]_d$ = $[\text{siɾɪŋ}]_t$		*!*	*	**
c. $[\text{sɪɾɪŋ}]_d$ = $[\text{sɪɾɪŋ}]_t$			*!	**
☞ d. $[\text{sɪɾɪŋ}]_d$ >> $[\text{sɪɾɪŋ}]_t$				**

(>> represents the longer duration of vowels preceding the medial /t,d/ and = refers to the same vowel duration)

The optimality theoretic grammar proposed in this study makes contribution to the architecture of the interface of phonology, phonetics and lexicon by incorporating lexically-conditioned phonetic detailed constraint, i.e., Preserve(LexCont) and P-Short into the evaluation ranking system. Furthermore, this grammar also is significant in that it captures the experimental result of incomplete neutralization of flapping evident in phonetic-fine detail level.

As previously mentioned, it has been controversial whether uniformity across derived words arises at phonological level or phonetic level (McCarthy 2001, Riehl 2003, Steriade 2000). In this study, we examined whether the voicing contrast in word-final /t,d/ in English enjoys the status of contrast in words placed in a flapping environment created by attaching suffixes like "-ing, -y, -er, -est, -ist". It was found that phonetic-details are uniformly observed across these derived words in a differential mode, depending on individual phonetic correlats. Note that a specific phonetic property of an allophone or a phoneme in the base form was not maintained in its derived form, but the contrast of phonetic correlates of voicing was preserved between the base and its derived form. For example, the durational contrast found in the vowels preceding word-final /t,d/ (e.g., "s<u>ee</u>d-s<u>ea</u>t") were maintained in or transferred to the vowel length contrast in the pre-flaps of their corresponding derived words (e.g., "s<u>ee</u>[ɾ]ing-s<u>ea</u>[ɾ]ing"). The contrast in stop closure duration is also maintained in a flap context. Furthermore, the contrast in voicing duration as well is uniform in the derived words as summarized in Table 3. These findings seem to support the idea of phonetic uniformity suggested by Steriade (2000) in that many phonetic correlates of voicing in morphologically simplex words are preserved in their derived words as well. (However, note that some researchers claim that English flaps are not derived from underlying /t,d/ but stored in the mental lexicon (Connine 2004, Jongman 2004, Jongman et al. 1992)). .

4.2 Imperceptible Complete Flapping

First, this study attempted to see whether word-final /t,d/ contrast is perceptible by English listeners. It was found that English native listeners had no difficulty in judging whether the word-final consonant is an underlying /t/ or /d/. This robust perceptibility of the underlying word-final /t,d/ above the accuracy of 96% is attributable conceivably to high availability of a variety of phonetic correlates of voicing contrast, including durations of preceding vowels, stop closure durations, voicing durations, F0 of preceding vowels, etc. It also confirms that English does not have word-final devoicing like German or Polish, and thus allows more marked contrast in word final coda position.

Next, we examined whether English listeners can distinguish a flapped /t/ from a flapped /d/ in disyllabic words. We found that English listeners did not made successful judgment about whether a flap is an underlying /t/ or /d/ although /d/-identification was slightly more accurate than /t/-identification. This finding seems to be inconsistent with that of Mitterer and Ernestus (2006) that listeners recover the reduced /t/. The overall accuracy was near chance-level (55.5%) as reported in previous section. This imperceptibility may lead to the speculation that complete neutralization has occurred in the listening stimuli used in the experiment. As described in Table 5, it is evident that a majority of acoustic properties do not differ between /t/ flap words and /d/ flap words. This lack of availability of perceptual cues to word-medial /t,d/ contrast may be of little help to restore an underlying /t/ or /d/. Previous studies have shown that any single acoustic cue such as the duration of pre-flaps alone does not suffice to make a decision between an underlying /t/ and /d/ although a multitude of cues are significantly different on the production side (Herd et al. 2010). Accordingly, it might be conjectured that even English native speakers suffer perceptual difficulty.

Finally, an experiment was conducted to determine whether the pre-flap vowel length affects the judgment on whether an flap is a /t/ variant or /d/ variant. The result showed that overall, it is not a major influencing factor in deciding an underlying phoneme. Although English native listeners made differential perception of /t,d/ based on the vowel length, the rates of responses were mostly near or below chance level for a range of manipulated vowel lengths. It is the case that English listeners cannot reliably distinguish a flapped /t/ from a flapped /d/.

In future studies, other acoustic cues such as flap durations, VOT and F0 of pre-flap vowels could be manipulated to determine whether these are used as reliable perceptual cues to make a distinction between a flapped /t and /d/. Additionally, this attempt also be made for L2 English speakers to see whether the delicate differences between these two are perceptible or imperceptible (Burrows 2014). Still it is not clear why some acoustic properties are significantly different enough to distinguish an underlying /t/ and /d/ on the production while their differences are imperceptible on the perception side. This asymmetry is worthwhile to examine further to draw a comprehensive picture of their relation from the functional perspective of communication.

References

Boersma, P. and D. Weenink. 2020. Praat: *Doing phonetics by Computer. Version 6.1.*

Braver, A. 2011. Incomplete neutralization in American English flapping: A production study. *Proceedings of the 34th Annual Penn Linguistics Colloquium* 17, 30-40

Braver, A. 2013. *Degrees of Incompleteness in Neutralization: Paradigm Uniformity in a Phonetics with Weighted Constraints*. Doctoral dissertation, Rutgers University.

Burrows, L. E. 2014. *The Learning of Phonetic Flap by English Language Learners: Data from Kindergarten* *and Fourth Grade*. Doctoral dissertation, University of Memphis.

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

Chambers, J. K. 2006. Canadian raising retrospect and prospect. *Canadian Journal of Linguistcs* 51, 105-118.

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

Connine, C. M. 2004. It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review* 11, 1084-1089.

Derrick, D., I. Stavness and B. Gick. 2015. Three speech sounds, one motor action: Evidence for speech-motor disparity from English flap production. *The Journal of the Acoustical Society of America* 137(3), 1493-1502.

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

Eddington, D. and D. Elzinga. 2008. The phonetic context of American English flapping: Quantitative evidence. *Language and Speech* 51(3), 245-266.

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

Herd, W., A. Jongman and J. Sereno. 2010. An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics* 38, 504-516.

Jongman, A. 2004. Phonological and phonetic representations: The case of neutralization. *Proceedings of the 2003 Texas Linguistics Society Conference*, 9-16.

Jongman, A., J. Sereno, M. Raaijmakers and A. Lahiri. 1992. The phonological representation of [voice] in speech perception. *Language and Speech* 35, 137-152.

Joos, M. 1942. A phonological dilemma in Canadian English. *Language* 18(2), 141-144.

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

Lavoie, L. M. 2000. *Phonological Patterns and Phonetic Manifestations of Consonant Weakening*. Doctoral dissertation, Cornell University.

Lee, M. K. 2022. A foot-based approach to English /t/-lenition. *Studies in Phonetics, Phonology and Morphology* 28(2), 259-280.

Malécot, A. and P. Lloyd. 1968. The /t/:/d/ distinction in American alveolar flaps. *Lingua* 19, 264-272.

McCarthy, J. 2001. *A Thematic Guide to Optimality Theory*. Research Surveys in Linguistics.

Mitterer, H. and M. Ernestus. 2006. Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34, 73-103.

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

Port, R. 1976. Influence of tempo on the closure interval cue to the voicing and place of intervocalic stops. *The Journal of the Acoustical Society of America* 59, 1208.

Port, R. 1981. Linguistic timing factors in combination. *The Journal of the Acoustical Society of America* 69, 262-274.

Riehl, A. K. 2003. American English flapping: Perceptual and acoustic evidence against paradigm uniformity with phonetic features. *Working Papers of the Cornell Phonetics Laboratory* 15, 271-337.

Roettger, T., B. Winter, S. Grawunder, J. Kirby and M. Grice. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43, 11-25.

Selkirk, E. 1982. The syllable. In H. van der Hulst and N. Smith, eds., *The Structure of Phonological*

Representations Part II, 337-383. Dordrecht: Foris.

Sharf, D. J. 1960. Distinctiveness of 'voiced T' words. *Amercian Speech* 35, 105-109.

Simonet, M., M. Rohena-Madrazo and M. Paz. 2008. Preliminary evidence of incomplete neutralization of coda liquids in Puerto Rican Spanish. In L. Colantoni and J. Steele, eds., *Laboratory Approaches to Spanish Phonology III,* 72-86. Somerville, MA: Cascadilla.

Steriade, D. 2000. Paradigm uniformity and the phonetics-phonology boundary. In M. Broe and J. Pierrehumbert, eds., *Papers in Laboratory Phonology* 83, 213-238.

Stathopoulos, E. T. and G. Weisner. 1983. Closure duration of stop consonants. *Journal of Phonetics* 11, 395-400.

Sung, E.-K. 2003. Acoustics and perception of flaps in American English and Korean. *Language Research* 39(4), 933-958.

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.

Warner, N., A. Fountain and B. V. Tucker. 2009. Cues to the perception of reduced flaps. *The Journal of the Acoustical Society of America* 125, 3317-3327.

Warner, N., A. Jongman, J. Sereno and R. Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32, 251-276.

Warner, N., E. Good, Al. Jongman and J. Sereno. 2006. Orthography vs. morphological incomplete neutralization effects. *Journal of Phonetics* 34, 285-293.

White, L. and A. Turk. 2010. English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics* 38, 459-471.

Yavaş, M. 2011. *Applied English Phonology*. Wiley-Blackwell.

Yun, G. 2015. Complete flapping through polysyllabic shortening in English. Paper presented at 2015 Annual Meeting on Phonology. UBC, Canada.

Zue, V. W. and M. Laferriere. 1979. Acoustic study of medial /t, d/ in American English. *The Journal of the Acoustical Society of America* 66, 1038-1050.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Appendix

Materials for Production Experiment (96 words)

	/d/		/t/	
Vowel type	**[d]**	**[ɾ]**	**[t]**	**[ɾ]**
/i/	bead	beading	beat	beating
	seed	seeding	seat	seating
/ɪ/	kid	kidding	kit	kitting
	bid	bidding	bit	bitting
/e/	wade	wading	wait	waiting
	raid	raiding	rate	rating
/ɛ/	bed	bedding	bet	betting
	red	redding	ret	retting
/æ/	pad	padding	pat	patting
	fad	faddist	fat	fattest
/u/	hood	hooding	hoot	hooting
	mood	moody	moot	mooty
/ʊ/	wood	wooding	woot	wooting
	pud	pudding	put	putting
/o/	bode	boding	boat	boating
	code	coding	coat	coating
/ʌ/	bud	Budding	but	butting
	rudd	ruddy	rut	rutty
/ɑ/	pod	podder	pot	potter
	cod	codder	cot	cotter
/aɪ/	ride	rider	write	writer
	side	siding	sight	sighting
/aʊ/	loud	loudy	lout	louty
	cloud	cloudy	clout	clouty