An effect of flaps on the fourth formant in English

Natasha Warner
University of Arizona
nwarner@email.arizona.edu

Benjamin V. Tucker
University of Alberta
bvtucker@ualberta.ca

Very few segments of the world's languages have been shown to have a systematic effect on the fourth formant (F4). We investigate a large drop in F4 which sometimes occurs in conjunction with the flap in American English. The goal of the present work is to document this phenomenon, and to determine what phonological environments coincide with this large drop in F4. We measure data from six speakers producing words with medial flaps in various environments, such as *party*, *turtle*, *bottle*, *credit*, *harder*. We find that the combination of flap with a rhotic and to a lesser extend a syllabic /l̩/ leads to a larger drop in F4 than other flap combinations like a following /i/. Together with previous perceptual data, the findings support the conclusion that this feature of F4 results from transitions among articulations.

1 Introduction

When measuring vowel formants or reading spectrograms, linguists tend to focus on the first, second, and third formants (F1–F3), usually to the exclusion of higher formants such as F4. This is with good reason: it is well known that the first three formants provide the primary acoustic information for perception of vowels and sonorant consonants, and that higher formants such as F4 and F5 may provide information about the identity of the speaker, but they rarely contain segmental perceptual cues (Zhou et al. 2008). Only a few segments have been shown to have any systematic effect on F4 at all. In this paper, we examine very large shifts in F4, up to approximately 1000 Hz, which sometimes occur with a flap consonant in American English, especially in words like *quarter* [kʰɔɹɾɚ] and *turtle* [tʰɚɾl̩]. We test various factors about the environment of the flap to determine when such F4 patterns are prominent, and we examine how consistently such a drop is present.

Zhou et al. (2008) present a very detailed analysis of American English /ɹ/ as produced by two speakers, one who uses a retroflex and one who uses a bunched articulation of /ɹ/. They find that the retroflex speaker has a lowered F4 during /ɹ/ relative to surrounding vowels, while the bunched speaker does not. They model this effect through finite element method area functions of the vocal tract derived from MRI data. They show that the F4 is a resonance of the back cavity of the vocal tract, the cavity behind the palatal constriction, for both retroflex and bunched /ɹ/, but the two means of making /ɹ/ differ in whether it is a half-wavelength or a quarter-wavelength resonance. Johnson (2011) finds that this difference in F4 is sufficient to affect compensation for coarticulation in perception of the following segment.

Journal of the International Phonetic Association (2017) 47/1
doi:10.1017/S0025100316000219

First published online 7 July 2016

Espy-Wilson (2004) discusses the wide variability among speakers in how they articulate American English /ɹ/. She shows spectrograms of three speakers' productions of a nonsense word /waɹav/, representing varied articulations of the /ɹ/. She reports that the F4 during the /ɹ/ is approximately 400–520 Hz lower than during the surrounding vowels for the two speakers who show lowered F4 during /ɹ/. Derrick & Gick (2011) present ultrasound data on how speakers of North American English realize flapped or tapped /t d/ with /ɚ/ vs. a non-rhotic vowel both before and after it. They find categorical phonetic variability among four distinct tap/flap gestures. They find that certain articulatory patterns are more common across speakers when the tap/flap is only preceded by /ɚ/ or only followed by it, but they find extensive speaker-specific variability in articulations when the consonant is both preceded and followed by /ɚ/ (in the word *murder*). (Although Derrick & Gick distinguish taps from flaps, we will simply use 'flap' as a cover term in this paper for productions that could be taps or flaps, as our acoustic data does not distinguish the two.)

Some other segments have been found to affect F4 in various languages. Avelino & Kim (2003) show lower F4 with more retracted place of articulation of stops in the Pima language. The segments with retracted place have sometimes been called retroflex in past descriptive work on the language, but Avelino & Kim conclude that these segments do not actually involve retroflexion. Hamann (2003) also discusses F3 and F4 lowering in connection with retroflex consonants cross-linguistically. Lowering of F4 was found in Toda (Shalev, Ladefoged & Bhaskararao 1993), O'odham (Dart 1991) and Hindi (Stevens & Blumstein 1975).

In earlier work on spontaneous vs. careful speech, we investigated acoustic properties of American English flaps and intervocalic stops, focusing on amplitude, duration, and voicing rather than formant frequencies (Warner & Tucker 2011). While hand-labeling the data, we often noticed a very large drop in F4, timed to the flap consonant. Figure 1 shows several examples. This drop in F4 was not consistently present, although it did sometimes occur even if the consonant (expected to be a flap in careful speech) was extremely reduced, visible only as a slight dip in amplitude that could perhaps be called an approximant (Figures 1B, 1C). In the current work, we document this pattern of a drop in the fourth formant frequency, and investigate what factors in the surrounding phonological environment make it more or less likely to occur. We also compare the pattern in production to past data on listeners' ability or inability to use the drop in F4 as a perceptual cue (Warner, Fountain & Tucker 2009). The production and perception results together raise the question of whether this very large acoustic change timed to a particular segment is an artifact caused by overlap of articulations for neighboring segments, and if so, whether it can still serve as a useful perceptual cue.

2 Method

2.1 Materials

For this project, we measured data from a subset of the recordings used in Warner & Tucker (2011), which was a large-scale study investigating many aspects of the phonetics of intervocalic stops and flaps. The target consonants for that project were the stop phonemes /p t k b d g/ in intervocalic position with the following vowel unstressed, e.g. *choppy*, *automatic*, *abacus*, *already*. Target stop phonemes in each of six segmental environments were used (before a full vowel, e.g. *cheeky*, before syllabic /l̩/, e.g. *apple*, before /ə/, e.g. *academy*, before /ɚ/, e.g. *better*, after /ɚ/ or vowel–/ɹ/, e.g. *turkey*, and across a word boundary before /ə/, e.g. *would a*).[1] Warner & Tucker (2011) also tested stops both between two unstressed

[1] We do not intend to make a claim about the correct phonemic analysis of the /ɚ/ vowel as /ɚ/ vs. /əɹ/ in English, or about the underlying status of syllabic /l̩/, etc. We use phonemic slash marks rather than phonetic brackets to denote categories of the language as opposed to specific phonetic realizations, although analyses of the underlying form may differ.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

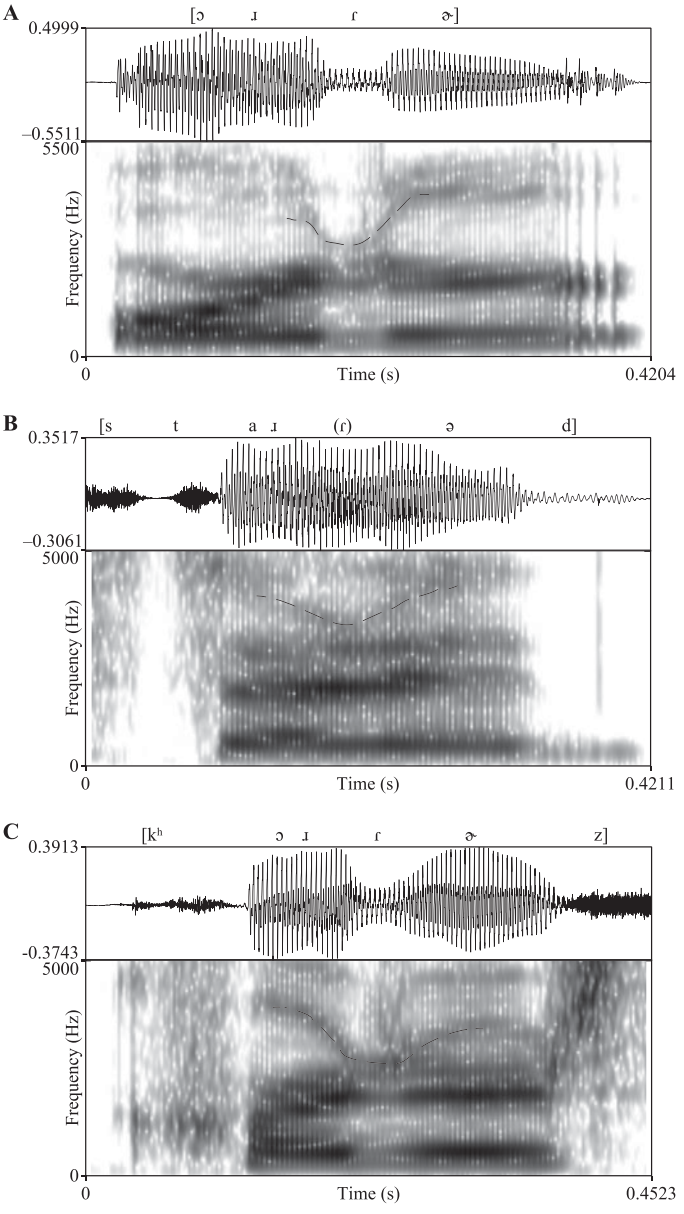

Figure 1 Waveforms and spectrograms showing examples of F4 drop, with dashed line superimposed on F4. (1A) An item from the current data (*order*, speaker C) with a drop in F4 at the time of the flap. (1B) Spontaneous conversation (*started*, speaker B) with extremely reduced consonant but F4 drop present. (1C) Read connected speech (*quarters*, speaker B) with F4 drop and a somewhat reduced consonant. (Compare Figure 4 for individual speakers.)

syllables (e.g. *vinegar*) vs. stops after a stressed vowel (e.g. *beggar*) where possible within the lexicon. Furthermore, that project included three speech styles: free conversation (where any stop phonemes speakers happened to produce in the target phonological environments were analyzed), story reading, and word-list reading. This resulted in a dataset of approximately 700 stop phoneme tokens per speaker, across many factors.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

For the current project, we are only investigating flaps, so we used only the words with /t d/ as the target consonant. We used only the word-list reading speech style from the Warner & Tucker (2011) materials. Because we are interested in documenting a little-known acoustic phenomenon, we chose to use careful speech and avoid the many additional sources of variability in spontaneous conversation. Of the six segmental environments in the earlier work, we used all except the word boundary condition (e.g. *what a*), since those had to be elicited with surrounding phrasal context, and thus the task differed somewhat from the rest of the conditions. We omitted the inter-unstressed environment from the earlier materials (e.g. *ability*), and used only the post-stress environment (e.g. *beauty*), because the earlier work found no strong effects of this stress manipulation, and because there are many gaps in the English lexicon for the inter-unstressed pattern. This left us with 100 items, with the target consonant phoneme /t d/ and its phonological environment (defined as having a following /i ə ɚ l̩/ or preceding /ɹ ɚ/) as conditions. In the 2011 study, we did not attempt to test both preceding and following environment separately, and this caused the mixture of preceding and following environments.

However, initial observations for the current work suggested that presence of an /ɹ/ or /ɚ/ in the environment, either before or after the target flap consonant, might be an important factor. Therefore, we recoded the items by presence vs. absence of a preceding /ɹ/ or /ɚ/ and by following vowel (using the same following vowel categories, /i ə ɚ l̩/), so that preceding and following environment can be examined separately. (We will refer to syllabic /l̩/ as a vowel for convenience, since it is acoustically similar to a vowel and all other conditions have a vowel in this position.) As in our earlier work, the only following vowel we used other than /ə ɚ l̩/ is /i/ because this is the only vowel that is common enough in this environment to supply enough items. Because the items were not originally chosen to test for effects of both preceding and following context, and because of constraints of what words are available in the lexicon, we did not have equal numbers of items in all conditions as defined for the current study (Table 1). The condition with preceding /ɹ ɚ/ and with following /l̩/ had only three words, *hurdle*, *turtle*, *fertilizer* (thus only one with /ɹdl̩/). Furthermore, the combination of preceding /ɹ ɚ/ and a following /ə/ never occurred among the items, with either /t/ or /d/.[2] The lack of information about this condition is unfortunate, but unless there is an interaction that affects specifically the combination of a preceding /ɹ ɚ/ and a following /ə/, the data from the other conditions, especially those with preceding /ɹ ɚ/, will provide most information about F4 in such environments.

One might further wonder whether the quality of the preceding vowel (among non-rhotic cases, e.g. /i/ vs. /a/ vs. /o/, etc.) influences F4 during the preceding vowel or during the target consonant. However, studying all combinations of preceding and following vowels and presence/absence of preceding /ɹ ɚ/ would require a very large number of items. We leave this as a topic for future research, because dividing the current item set into smaller categories based on previous vowel quality, even for those with no preceding /ɹ ɚ/, leaves too few items for reliable analysis.

2.2 Speakers

For the current work, we measured data from six speakers, a subset of those used in Warner & Tucker (2011). All six were native speakers of a rhotic variety of American English (two male

[2] These combinations are also relatively rare in the English lexicon, so even if we had originally designed the experiment around these conditions, we might not have been able to find enough items that are familiar to most speakers. A search of the NewDic dictionary (http://dingo.sbs.arizona.edu/~hammond/lsasummer03/newdic.txt) finds only four words with preceding /ɹ/ or /ɚ/, /d/, and with following /ə/ or [ɨ], most of which do not seem common enough to be readily pronounceable for most speakers (*cordovan*, *purdah*, *sordid*, *verdigris*).

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

Table 1 Items measured, arranged by the conditions that are analyzed in the current study. For preceding environment, /ɚ/ is considered as a type of preceding /ɹ/ (consonantal vs. vocalic status of the preceding [ɹ] quality is not at issue).

Following V	/ɹt/	non-ɹ /t/	/ɹd/	non-ɹ /d/
Full vowel /i/	party	beauty	sturdy	already
	hearty	catty		body
	thirty	mighty		caddy
	forty	eighty		daddy
		treaty		study
		pretty		buddy
		city		lady
		pity		steady
		committee		heady
		duty		needy
Syllabic /l̩/	turtle	battle	hurdle	cradle
	fertilizer[a]	bottle		model
		cattle		saddle
		vital		idle
		title		muddle
		fatal		puddle
		settle		huddle
		little		medal
		total		needle
				middle
/ə/	–	automatic	–	academy
		satisfaction		modify
		status		haddock
		united		credit
		data		edible
		competitive		incredible
		citizen		medicine
		legitimate		freedom
		notable		ridicule
		quota		audible
/ɚ/	charter	better	harder	cedar
	garter snake	water	larder	moderate
	barter	daughter	murder	ladder
	mortar	matter	orderly	spider
	quarter	writer	border	powder
		butter	order	trader
		letter	recorder	federal
		sweater	disorder	confederate
		motor		consider
		computer		odor

[a]Examination of the spectrograms and auditory judgment verify that all speakers produced this item with syllabic /l̩/, not with an /əl/ sequence, which could have contained light /l/ rather than dark /l̩/, thus potentially influencing the F4 differently.

and four female) with no substantial exposure to other languages before puberty. The speakers were students at the University of Arizona, and they received extra credit in an introductory course for participation.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

2.3 Procedures

The procedures are briefly summarized in this paper and described in greater detail in Warner & Tucker (2011). The current data came from the word-list reading portion of that experiment, and were randomized together with the /p b k g/ target items and the inter-unstressed items that are not included in the current work. Thus, speakers were not reading a long series of words with medial flaps, and the additional conditions serve as filler items for the current purpose. Each speaker read the list once, since we focused on including more items rather than more repetitions in order to maintain independence of data. Thus there were 600 tokens in total. Speakers were recorded using an Alesis CD recorder at 44.1 kHz. They were seated in a sound-protected booth, and used a high quality head-mounted microphone.

2.4 Measurements

We measured F4 at its maximum in the preceding and following vowel surrounding the target flap consonant, and at its minimum during the consonant if F4 remained clearly visible in the spectrogram during the consonant, or at the offset of F4 just before the consonant and onset of F4 after the consonant. Figure 2 shows examples. All measurements were made using Praat (Boersma & Weenink 2015).

Measuring F4 can be challenging because it often has very low amplitude, making it difficult to identify which frequency range is F4 at all. Formant trackers are frequently not reliable for higher formants, but pure hand-measurement is also not very accurate. Therefore, for each word, we first turned off formant tracking in order to avoid being misled by it, and identified the F4 contour across the duration of the word visually. Then we turned on the formant tracker in Praat, and used the Query formant frequency function to obtain a more accurate measurement of the formant that we had already visually identified as the F4. In order to avoid situations where the automatic formant tracker fails to pick up what can clearly be visually identified as F4, we set the number of formants to calculate to one greater than the number we expected there to actually be within the frequency range, normally using a setting of 6 formants within the first 5500 Hz. This usually resulted in the formant tracker finding an extra peak at some lower frequency where there was clearly no actual formant, but also locating a formant track relatively accurately along the visible F4. (Because this setting in any case locates more formants than are actually present, it worked well for both male and female speakers.) We then used the query formant frequency function to obtain the frequency of the track on the actual F4 (as identified visually), regardless of whether that was the fourth or the fifth (or occasionally the sixth) automatically located track. If the formant tracker did not place a formant on the region identified visually as F4, we adjusted the measurement time point slightly in order to measure at a time point with accurate formant tracking; on rare occasions, we adjusted the number of formants for the formant tracker to find instead.

Many words have a long steady plateau in the F4 throughout most of the preceding and following vowel, and sometimes throughout the word. The formant track identified by the software often shows some error variability, with the tracked formant jumping slightly from one time point to the next in a way that is inconsistent with stable F4 tracking. Rather than attempting to locate the time point in the preceding or following vowel with the mathematical maximum value for the formant track, if the formant track for the F4 had a long plateau in the vowel except for error variability, with no clear peak, we positioned the time point for measuring F4 of the vowel at the maximum amplitude of the vowel. This simply serves to locate a replicable time point if there is no peak F4 because it is too steady. For a few productions, the F4 was very unclear visually. However, it was possible to establish which faint elevation of energy was the F4 based on portions of the word where the F4 was clearer.

For productions where F4 did not continue throughout the consonant we measured two values, at offset and onset of the consonant, rather than a single value during the consonant.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

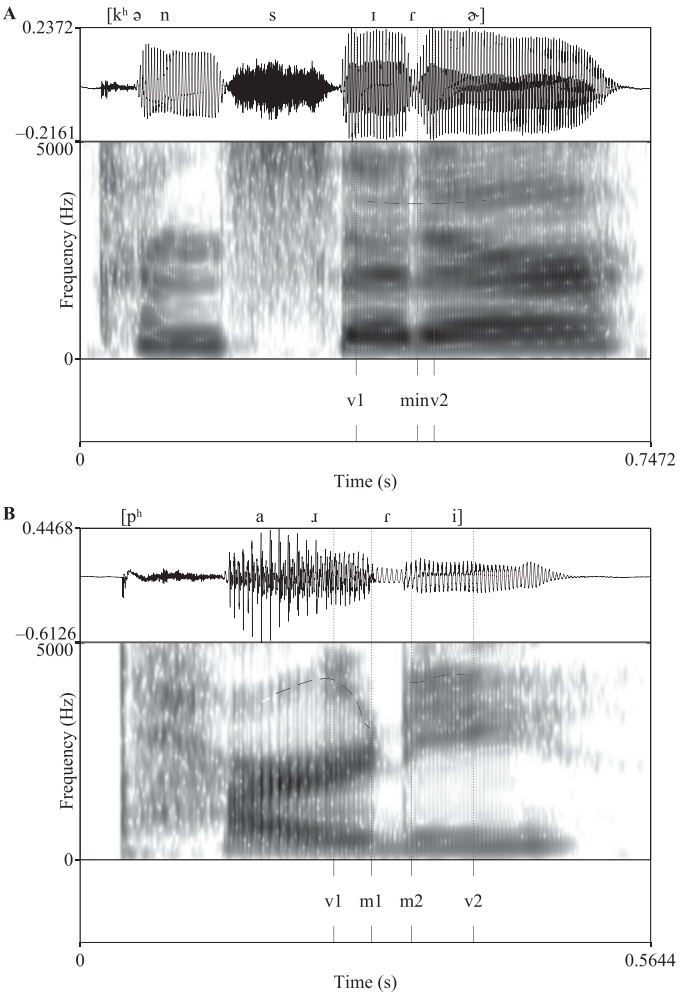

Figure 2 Examples of measurement points (v1 and v2 for the vowels, min or m1 and m2 for the consonant). (2A) *Consider* (Speaker B), with F4 continuing throughout the consonant. (2B) *Party* (Speaker C), with F4 ceasing during the consonant. The dashed line shows the approximate location of F4 near the consonant.

We averaged the two values in order to obtain a single value reflecting F4 at the boundaries of the consonant. This facilitated comparison with the productions that did have F4 throughout the consonant. We then subtracted the F4 value for the consonant from the average of the F4 measurements for the preceding and following vowel. This resulted in a measure of change (drop) in F4 during the flap consonant. A larger positive value reflects a larger drop, a negative value reflects a rise in F4 during the consonant. Since we measured at the maximum F4 during the vowels and minimum during the consonant, and there is some variability in the F4 track, this measure shows at least a small drop for most tokens, even those with rather steady F4 throughout. Thus, a measured F4 drop of 100 Hz for a given token usually does not reflect any appreciable change in F4 (see Figure 2A, with a measured F4 'drop' of 104 Hz and a rather steady F4). However, a measured F4 drop of 1000 Hz reflects a very large change in the F4 (see Figure 1A, with a measured F4 drop of 1338 Hz). Since we use F4 drop as a continuous dependent variable, we do not attempt to define a criterion of what counts as presence vs. absence of a drop.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

2.5 Statistical analysis

The statistical approach we choose for these data is by-subjects ANOVAs, using two separate designs to handle the non-fully-crossed design, as described below. The data are not ideal for statistical analysis. Some conditions have greater variability than others, violating the assumption of homogeneity of variance. Some conditions appear to have a slightly bimodal distribution, falling into tokens with a large F4 drop vs. tokens without an F4 drop (although see discussion below suggesting that the data is not actually bimodal). Because we wished to examine the effects of both preceding and following context together, there are few items in some conditions. These issues cause problems for either ANOVA or Linear Mixed Effects modeling. We choose to use ANOVAs to facilitate clear explanation of main effects across all levels of other factors (recall the explanation of main vs. simple effects in LME analysis in Clopper 2013), and because the focus of the work is on entirely categorical independent variables, which have less need of a regression-based method. In order to compensate for the problems in the distribution of the data, we choose to set the significance criterion at $\alpha = .01$ rather than .05. Although this is an imprecise method of compensating, it should greatly reduce the chance of a Type I error.

Because the data included no words with preceding /ɹ ɚ/ and with following /ə/, we analyzed the data using two separate ANOVAs. The first was a three-factor design with presence of preceding /ɹ ɚ/ (present, absent), consonant phoneme (/t d/), and following vowel (/i ɚ l̩/) as the factors. This excluded the conditions with following /ə/. The second ANOVA was a two-factor design with consonant phoneme (/t d/) and following vowel (/i ɚ l̩ ə/) as factors, using only the data without preceding /ɹ ɚ/. All ANOVAs were conducted on data averaged over items, so that there is one value for each speaker for each condition, consisting of the average of that speaker's production of all words (items) in that condition (effectively by-subjects ANOVAs). All factors in both designs were within subjects.

3 Results

3.1 F4 drop by phonological environment and across speakers

Figure 3 shows the average F4 drop data. In the three-factor design, the main effect of preceding /ɹ ɚ/ was significant ($F(1,5) = 18.25$, $p < .01$), as was the main effect of following vowel ($F(2,10) = 8.14$, $p < .01$). No other main effects or interactions were significant (phoneme: $F(1,5) = 1.41$, phoneme × preceding /ɹ ɚ/: $F < 1$, phoneme × following vowel: $F(2,10) = 1.84$, preceding /ɹ ɚ/ × following vowel: $F(2,10) = 3.29$; three-way interaction: $F(2,10) = 1.22$, all ps $> .05$). In the two-factor analysis, we found a significant main effect of the following vowel ($F(3,15) = 24.56$, $p < .001$), but no main effect of consonant phoneme ($F(1,5) = 4.86$, $p > .05$) and no interaction ($F < 1$). Figure 3 also suggests no consistent direction of effect for phoneme /t/ vs. /d/. Therefore, we collapsed the /t d/ conditions for all further analyses.

In order to determine which following vowels caused a larger drop in F4, we performed pairwise comparisons among the following vowel conditions. Because one condition with preceding /ɹ ɚ/ had only three items per speaker, and the variability in this condition was large, we chose to perform pairwise comparisons only among the conditions without preceding /ɹ ɚ/. Since the purpose of this work is to explore where the unexpected drop in F4 might occur, and we do not have a pre-existing hypothesis for what causes it, we performed all possible pairwise comparisons. We did not additionally use a correction for familywise error because we are using the conservative criterion of $\alpha = .01$, and further correcting that for familywise error would lead to extremely low power. The pairwise comparisons show that flaps before /i/ had a smaller drop in F4 than before any of the other following vowels (/i/ vs. /ɚ/: $F(1,5) = 55.65$, $p < .005$; /i/ vs. /l̩/: $F(1,5) = 23.15$, $p < .01$; /i/ vs. /ə/: $F(1,5) = 95.52$, $p < .001$). No

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

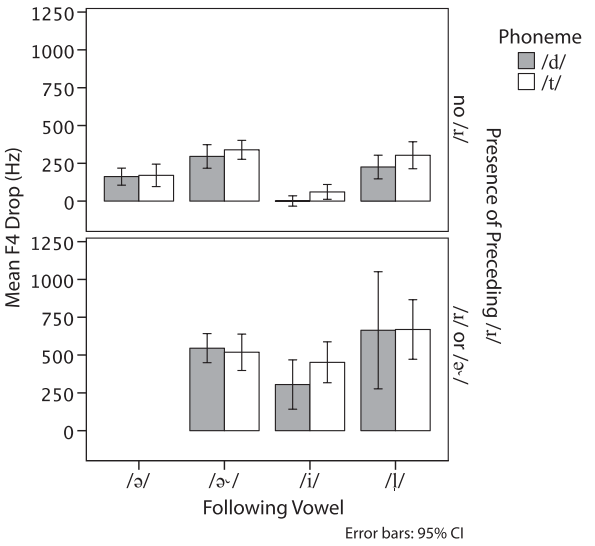

Figure 3 Average decrease in F4 from the average of the two surrounding vowels to the minimum during the flap consonant (if F4 is visible during it) or to the average at onset and offset of the flap consonant (if F4 ceases during the consonant), in Hertz.

other pairwise comparisons reached significance at the .01 level.[3] Thus, flaps with a following /i/ vowel have less drop in F4 than other flaps, but we cannot be sure of any further effects of the following vowel. Furthermore, the main effect of preceding /ɹ ɚ/ in the three factor analysis also shows that flaps with a preceding /ɹ ɚ/ have greater F4 drop than those without.

Considering the variability among speakers in use of retroflex vs. bunched /ɹ/ in American English, and the fact that Zhou et al. (2008) find an influence of retroflex but not bunched /ɹ/ on F4, one might wonder whether only speakers who use retroflex /ɹ/ have the drop in F4 that we point out, even though it seems to be timed to the flap and not the /ɹ ɚ/. Whether related to the /ɹ/ articulation or not, a speaker-specific articulation could be producing the drop in F4. Figure 4 shows the data for each speaker. Although some speakers show larger F4 drops than others, the phenomenon does not seem to be a property of a few speakers. Rather, all speakers show relatively large drops in F4 in at least some conditions, and the speaker with the least F4 drop is not consistent across conditions. It is possible that speakers produce both retroflex and bunched /ɹ/ variably depending on environment (Guenther et al. 1999), but Mielke, Baker & Archangeli (2016) find almost exclusive use of bunched /ɹ/ in speakers at the same university producing words with /ɹ/ in similar environments, so the current speakers are probably producing the F4 drop in spite of using bunched /ɹ/, not because of using the retroflex variant. Regardless of type of /ɹ/, it is at least clear that the large F4 drop is not limited to a few speakers.

[3] At the uncorrected .05 level, the pairwise comparisons of following /ə/ vs. /ɚ/ ($F(1,5) = 15.15$, $p = .011$) and /ɨ/ vs. /ɚ/ ($F(1,5) = 6.85$, $p < .05$) also reach significance, with following /ɚ/ having somewhat larger F4 drop than the other following vowels. If one were to apply the Bonferroni correction for familywise error in addition to using $\alpha = .01$, the required significance level would be .00167, which would mean that most tests would have extremely little power. For example, the pairwise comparison of /i/ to /ɨ/, which is significant at $\alpha = .01$ but not with additional Bonferroni correction, has an observed power to detect the actual observed effect size of .97 using $\alpha = .05$, a still relatively high .74 using $\alpha = .01$, but only .33 using $\alpha = .00167$ (with Bonferroni correction). Even with this stringent correction, the comparisons of /i/ to /ə/ and /i/ to /ɚ/ would be significant, supporting the claim that flaps with following /i/ have less F4 drop than with other following vowels.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

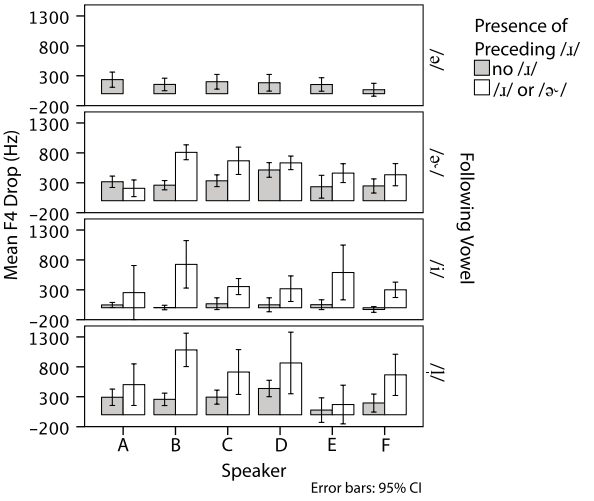

Figure 4 Average F4 drop (Hz) for individual speakers.

Speakers A and E are the two male speakers, whose formants might all be lower in the frequency range, and may show smaller changes as measured in raw Hertz. Examination of the maximum F4 during the preceding vowel shows that speaker A has the lowest F4 of all speakers, but speaker E has preceding vowel F4 in some conditions within the same range as the female speakers. While both male speakers have smaller F4 drops than the female speakers in some conditions, this is not consistent across conditions. Thus, the variability among speakers in size of F4 drop is not purely a matter of sex and therefore formant frequency range, although that may play some role.

3.2 F4 during neighboring vowels

Although the drop in F4 that we find is timed to the flap, not to any rhotic sound in the environment, it is possible that the F4 is influenced by a preceding or following /ɹ/ or /ə˞/ when one is present, as well. We analyzed the average F4 at the measurement points for the two surrounding vowels in order to determine the influence of the surrounding rhotic. Thus, this is an average for each token of the maximum F4 in the preceding and following vowel (or vowel + /ɹ/) context. These results appear in Figure 5. We analyzed these data using the same two ANOVA designs as for the F4 drop data (but collapsed over /t d/). The design excluding the following /ə/ context showed significantly lower vowel maximum F4 for conditions with preceding /ɹ ə˞/ ($F(1,5) = 41.19$, $p < .005$) and a significant main effect of following vowel among /i ə˞ l̩/ ($F(2,10) = 18.90$, $p < .001$), but no interaction ($F(2,10) = 2.55$, $p > .10$). The analysis using only conditions without preceding /ɹ ə˞/ also showed a significant effect of following vowel ($F(3,15) = 10.54$, $p < .005$). Examination of Figure 5 suggests the effect of following vowel primarily reflects lowered vocalic F4 for a following /ə˞/. Thus, a rhotic vowel or consonant preceding or following the flap lowers the F4 even at the time points that were measured as the maximum F4 of the surrounding vowels. This suggests that the values for F4 drop shown in Figure 3 (and Figure 6 below) may actually understate the change in F4 at the flap consonant, since F4 may already be lowered somewhat during the vocalic context in the conditions with the largest F4 drop at the flap. However, Figure 5 shows that the amount of depression of F4 during the vocalic context relative to non-rhotic vowels is on the order of 200 Hz, much smaller than the sudden drop in F4 at the flap.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

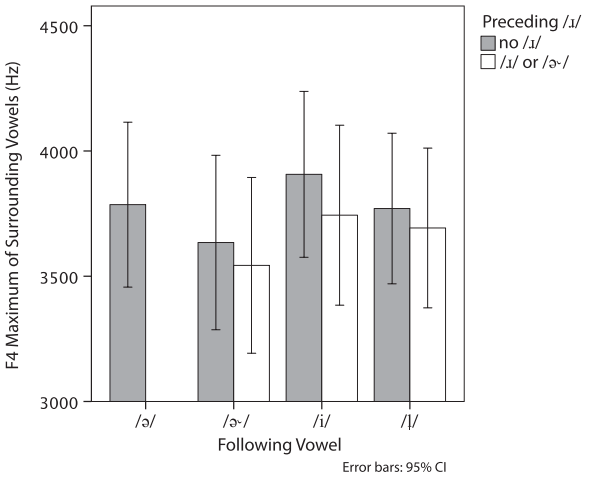

Figure 5 F4 (Hz) at maxima during preceding and following vowels (averaged over the preceding and following vowel for each token).

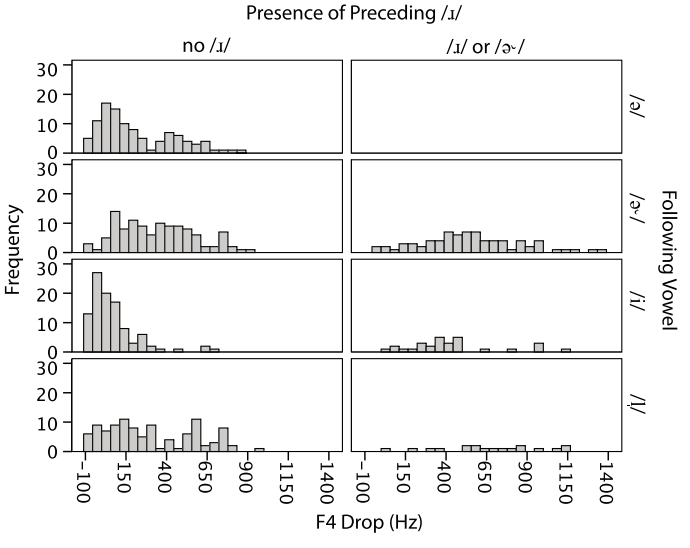

Figure 6 Histogram of F4 drop (Hz) data, by presence of preceding /ɹ ɚ/ and following vowel.

3.3 Is the F4 drop categorically present vs. absent?

Some conditions appear to have a bimodal distribution of F4 drop (Figure 6), which would indicate that some tokens have a drop in F4, and others do not. This could reflect a categorical difference in articulation. However, the 'dip test' (Hartigan & Hartigan 1985) showed no significantly bimodal distribution in any of the conditions (all ps $>$ .05). The condition without preceding /ɹ ɚ/ and with following syllabic /l/ was the only one that approached significance ($p = .084$). Thus, any tendency toward categorically different articulations is not clear enough to produce a significantly bimodal distribution.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

3.4 Asymmetric changes in F4 around the flap

In Figure 6, many tokens have an F4 drop of approximately 300–500 Hz. A symmetrical F4 drop of 300–500 Hz from the preceding vowel to the consonant, with an equal rise into the following vowel, would be a relatively small F4 drop among the spectrograms we have examined. What we often find, however, is that the F4 is asymmetric, with a large, clear drop/rise during the VC transition but rather steady F4 during the CV transition, or vice versa. Figure 2B above shows an example of this pattern. The calculated F4 drop is an average over the VC and CV transitions, which gives a smaller value. For example, the token of *party* in 2B was measured as having an F4 drop of only 355 Hz despite the large decrease in F4 during the VC transition.

In order to investigate the possibility of asymmetrical changes in F4 around the flap consonant, we calculated the F4 drop for the VC and the CV transitions separately. For the VC transition, we subtracted the F4 value at the minimum for the consonant (or onset of the consonant if F4 did not continue through it) from the F4 value measured in the preceding vowel. For the CV transition, we subtracted the value for the consonant from the value for the following vowel. We then subtracted the F4 change value for the VC transition from that for the CV transition. Thus, a negative value for this measure indicates a larger fall going into the flap consonant and a smaller rise in F4 from the consonant to the following vowel. The average value for this measure of F4 asymmetry was –54 Hz, and the median value was –30 Hz. That is, flaps tend to have slightly more F4 fall going into the flap than F4 rise coming out of it, but there is no large asymmetry in a consistent direction. Out of all tokens, 10% had a negative value below –474 Hz (much larger VC transition), and 10% had a positive value greater than 320 Hz (much larger CV transition). Such values represent a considerable discontinuity in the F4 pattern, since F4 typically varies little with changes among most speech segments.

4 Discussion

4.1 Degree and timing of the F4 drop

The results indicate that flap consonants are often accompanied by a large drop in F4, with F4 sometimes even dropping by 1000 Hz or more from the preceding vowel to the flap, and rising again after it. The drop in F4 is larger after a rhotic sound, whether the rhotic is consonantal or vocalic (/ɹ ɚ/), and it may be slightly larger before /ɚ/ as well, but this result is unclear. The drop in F4 is smaller before the vowel /i/ than before a following /ə/, /ɚ/, or syllabic /l̩/.

The results also show that a drop in F4 coinciding with flap consonants is relatively general across speakers, at least as indicated by those in this study: all six speakers studied here show this pattern to some extent, and none have a consistent lack of this pattern in F4 across conditions. Since Zhou et al. (2008) find that F4 is somewhat lower in retroflex productions of /ɹ/ than in bunched productions, one might suppose that speakers who habitually use retroflex /ɹ/ could have a larger drop in F4 than other speakers. The reverse might be true, though: if their F4 is lower during a neighboring /ɹ/ or /ɚ/, they might have less drop in F4 at the flap, because the F4 is lowered somewhat in the neighboring segment. However, Guenther et al. (1999) also showed that not all individual speakers use a consistent articulation (as also in Delattre & Freeman 1968 and earlier references cited therein). Rather, a given speaker's use of retroflex vs. bunched /ɹ/ can vary gradiently by phonological environment. In our recordings, we frequently see a stable F4 at the same frequency during an /ɹ/ as during other segments of the word, but in some tokens the F4 is somewhat lower during an /ɹ/ or particularly an /ɚ/ than elsewhere. Thus, some of the speakers may be making retroflex rhotics some of the time. However, the lowering of F4 during /ɹ ɚ/ is much smaller than the large, sudden drop of F4 with flap consonants, as shown in Figures 1 and 2B. We do not see a sharp drop in F4 timed to /ɹ/ the way Zhou et al. (2008) show in one token of /waɹav/ by their Speaker 4 (their

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

Figure 12), and it appears that only one of their three retroflex speakers showed this extreme effect of /ɹ/ alone, without a neighboring flap as in the materials we test.

The examples here show that the drop in F4 is timed to the flap consonant, not to a surrounding vowel or preceding /ɹ ɚ/. The change in F4 is rather sudden, and the data on asymmetry show that it usually happens on both sides of the consonant, although in some tokens it clearly happens during only the CV or only the VC transition (as in Figure 2B).

4.2 Explanations for the F4 drop

What causes this extreme shift (sometimes over 1000 Hz) in the frequency of the fourth formant, a formant which is known to have little relevance to speech perception? It is timed to the flap consonant, not to surrounding segments, and it sometimes occurs during only the transition into or out of the flap but not both. This suggests that the drop in F4 is caused by a specific articulatory configuration that the tongue sometimes passes through in the course of going between other articulations and the brief constriction of a flap consonant. Zhou et al. (2008: 4467) point out that 'higher formants are particularly responsive to smaller cavities in the vocal tract (e.g., piriform sinuses, sublingual spaces, laryngeal cavity)'. Since the F4 drop is somewhat larger before a following syllabic /l̩/, which may be prone to having such small cavities, the drop in F4 could be related to the articulation that occurs as the flap closure is laterally released into the /l̩/.

Furthermore, Zhou et al. (2008) conclude that for both retroflex and bunched /ɹ/, the fourth and fifth formants are resonances of the back cavity of the vocal tract, behind the palatal constriction. It seems likely that during the trajectory from the preceding vowel or /ɹ/ to the flap, and from the flap to the next vowel, the tongue sometimes alters the shape of the back cavity in a way that rapidly alters the wavelength of the standing wave associated with F4. It is not possible to determine through acoustic data exactly what change in articulation causes this, and neither is it possible to collect the type of sustained articulation MRI data that Zhou et al. (2008) use for a flap articulation, which by definition cannot be sustained. However, the fact that larger drops in F4 occur after a rhotic vowel or consonant suggests that the shape of the back cavity behind the palatal constriction for /ɹ ɚ/ may be involved. The fact that any substantial drop in F4 at all is rare before the vowel /i/, especially in the absence of a preceding /ɹ ɚ/, also leads to this conclusion, since /i/ also involves a close constriction near the palate, but a very different one than for /ɹ ɚ/. Stevens (1998: 279) points out that the center of gravity of the combined F3 and F4 peak is at its maximum for /i/. (Hamann (2002) discusses the incompatibility of the /i/ articulation with retroflexion, and this may be a related point.) Derrick & Gick's (2011) finding that there are several directions the tongue can move on the way into and out of a tap or flap closure, and that there is speaker variability as well as conditioning by surrounding rhotic sounds, also supports the conclusion that the variable F4 drop occurs because of the changes in resonating cavities during the transition between the surrounding sounds and the flap.

This explanation for the F4 drop need not be specific to the quick movement of the tongue during flaps. It could apply to non-flapped /t d/ as well. However, Stevens (1998: 358) predicts no change in F4 during the transition from /a/ to an alveolar stop closure, although this does not address vowel–rhotic–alveolar sequences. As an initial check of this possibility, we examined the recording of Speaker B holding a conversation (from Warner & Tucker 2011) for tokens of /t d/ not in flapping environment, with /ɹ/ or /ɚ/ nearby. Speaker B produces large average F4 drop in the current data, especially with a neighboring /ɹ/ or /ɚ/, and the conversation recording provides a source of words with the appropriate phonological environment, which was not a target of the wordlist. We identified three alveolar stop tokens (not flapped) with some sign of an asymmetrical drop in F4 in only one of the VC or CV transition (in *weird like*, *another day*, and *Purdue*). However, we also found several tokens with similar environments and no drop (*Carolyn's dorm*, *concert*, and *after taking*). We found

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

no tokens with a large F4 drop in a clearly visible F4. It may be that the longer closure period of a non-flapped /t d̪/ also allows the F4 drop to occur, but makes it more likely that it will occur only during the VC or the CV transition, but not both. During the longer closure, the oral cavity behind the closure may change shape through coarticulation with the next sound. Without the intervocalic environment interrupted only briefly by the flap, the F4 drop may also be less noticeable in the spectrogram.

As mentioned above, F4 and F5 are considered to be for the most part irrelevant for perception of speech sounds, although they may be used in perceiving whose voice one is hearing (Zhou et al. 2008). (F4 does influence listeners' percept of vowel quality when it is very close to a high F3, as in /i/, but this is because of the raising of F3, rather than because of any segment-specific property of F4; see Stevens 1998: 240, 279, 289.) Thus, listeners may have no reason to attend to F4 for purposes of perceiving most speech sounds, and one would then expect them to be rather insensitive to changes in F4 during the string of speech sounds, just as Japanese listeners are insensitive to the differences in F3 that form the major perceptual cue to American English /ɹ/ (e.g. Strange & Dittman 1984). The typically low amplitude and high bandwidth of higher formants could also discourage use of them as perceptual cues (Stevens 1998: 258ff.). However, a change of 1000 Hz in a formant frequency might still be noticeable to listeners. In a past perception study (Warner et al. 2009), we resynthesized tokens of *quarter* and *core* to manipulate the size of the F4 drop, and presented these stimuli to listeners for identification as either *quarter* or *core*. Results showed that listeners were only able to use the presence of an F4 drop as a perceptual cue to the presence of a flap in the stimuli made from a production where other perceptual cues were maximally ambiguous (only in the continuum created from a token of *quarter* produced with minimal intensity dip at the consonant). Even in that case, the perceptual effect of the F4 drop was very small, although it did reach significance.

Since a large F4 drop can occur even when the flap consonant is so reduced that it shows little change in intensity relative to surrounding vowels, and there might be few other perceptual cues to the flap's presence, we expected that the F4 drop might serve as a perceptual cue. However, the production results in the current work suggest another reason besides listeners' general lack of use of higher formants for why the F4 drop is such a weak perceptual cue to the presence of a flap (Warner et al. 2009): the F4 drop is only produced variably. In particular, very large F4 drops of more than 900 Hz (averaged over the VC and CV transitions) occur in only 3.4% of our data, and those of more than 1000 Hz occur in only 1.5% of the data. (To take asymmetric F4 drops into consideration, 4.6% of VC transitions have a drop of more than 1000 Hz, and 2.0% of CV transitions do.) Tokens with the flap too reduced to be perceptible based on other cues but with a very large F4 drop are probably rare.

Thus, it seems that this phenomenon is quite striking visually on a spectrogram, but that it is of little use to listeners for segmental perception in most tokens. Speakers do not produce it consistently as a typical feature of the flap segment. Instead, we conclude that the rapid drop in F4 timed to flaps occurs because of the combination of articulatory positions the tongue goes through on the way from the preceding sound to the flap and then to the following vowel. In particular, the combination of flap with a rhotic articulation and to some extent with a syllabic /l̩/ probably briefly creates a vocal tract configuration with a lower F4 than most other articulations, while the combination of a flap with a following vowel /i/ makes this vocal tract configuration unlikely. We conclude that the variably occurring large drop in F4 timed to flaps is the result of a combination of articulations of neighboring sounds with the flap consonant that occurs as a result of the transition between articulations.

Acknowledgements

We would like to thank Mary Dungan-Freiman & Karen Morian for their work on this project. We would also like to thank the anonymous reviewers for their helpful feedback.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press

References

Avelino, Heriberto & Sahyang Kim. 2003. Variability and constancy in the articulation and acoustics of Pima Coronals. *Annual Meeting of the Berkeley Linguistics Society*, vol. 29 (BLS 29), 43–53.

Boersma, Paul & David Weenink. 2015. Praat: Doing phonetics by computer [computer program], Version 6.0. http://www.praat.org/ (28 October 2015).

Clopper, Cynthia G. 2013. Modeling multi-level factors using linear mixed effects. *Proceedings of Meetings on Acoustics* (Acoustical Society of America 19(1)), 060028. http://scitation.aip.org/content/asa/journal/poma/19/1/10.1121/1.4799729.

Dart, Sarah. 1991. Articulatory and acoustic properties of apical and laminal articulations. *UCLA Working Papers in Phonetics* 79, 1–155.

Delattre, Pierre & Donald C. Freeman. 1968. A dialect study of American r's by X-ray motion picture. *Linguistics* 6(44), 29–68.

Derrick, Donald & Bryan Gick. 2011. Individual variation in English flaps and taps: A case of categorical phonetics. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 56(3), 307–319.

Espy-Wilson, Carol. 2004. Articulatory strategies, speech acoustics and variability. *Proceedings of From Sound to Sense*, June 2004, MIT, B62–B76.

Guenther, Frank H., Carol Y. Espy-Wilson, Suzanne E. Boyce, Melanie L. Matthies, Majid Zandipour & Joseph S. Perkell. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America* 105(5), 2854–2865.

Hamann, Silke. 2002. Retroflexion and retraction revised. *Papers on phonetics and phonology: The articulation, acoustics, and perception of consonants* (ZAS Papers in Linguistics 28), 13–26.

Hamann, Silke. 2003. *The phonetics and phonology of retroflexes*. Ph.D. dissertation, University of Utrecht.

Hartigan, J. A. & P. M. Hartigan. 1985. The dip test of unimodality. *The Annals of Statistics* 13, 70–84.

Johnson, Keith. 2011. Retroflex versus bunched [r] in compensation for coarticulation. *UC Berkeley Phonology Lab Annual Report*.

Mielke, Jeff, Adam Baker & Diana Archangeli, 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92(1), 101–140.

Shalev, Michael, Peter Ladefoged & Peri Bhaskararao. 1993. Phonetics of Toda. *UCLA Working Papers in Phonetics* 84, 89–123.

Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, Kenneth [N.] & Sheila Blumstein. 1975. Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics* 3, 215–233.

Strange, Winifred & Sybilla Dittmann. 1984. Effects of discrimination training on the perception of /rl/ by Japanese adults learning English. *Perception & Psychophysics* 36(2), 131–145.

Warner, Natasha, Amy Fountain & Benjamin V. Tucker. 2009. Cues to perception of reduced flaps. *Journal of the Acoustical Society of America* 125, 3317–3327.

Warner, Natasha & Benjamin V. Tucker. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America* 130, 1606–1617.

Zhou, Xinhui, Carol Y. Espy-Wilson, Suzanne Boyce, Mark Tiede, Christy Holland & Ann Choe. 2008. A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/. *The Journal of the Acoustical Society of America* 123(6), 4466–4481.

https://doi.org/10.1017/S0025100316000219 Published online by Cambridge University Press