RESEARCH

Sound change and coarticulatory variability involving English /ɹ/

Bridget J. Smith[1], Jeff Mielke[1], Lyra Magloughlin[2] and Eric Wilbanks[3]

[1] North Carolina State University, Raleigh, NC, US

[2] University of Ottawa, CA

[3] University of California at Berkeley, US

Corresponding author: Bridget J. Smith (bridgetjoannsmith@gmail.com)

English /ɹ/ is known to exhibit covert variability, with tongue postures ranging from bunched to retroflex, as well as various degrees of lip protrusion and compression. Because of its articulatory variability, /ɹ/ is often a focal point for investigating the role of individual variation in change. In the studies reported here, we examine the coarticulatory effects of alveolar obstruents with /ɹ/, presenting data from a collection of sociolinguistic interviews involving 162 English speakers from Raleigh, North Carolina, and a pilot corpus of ultrasound and lip video from 29 additional talkers. These studies reveal a mixture of assimilatory and coarticulatory patterns. For the sound changes in progress (/tɹ/ and /dɹ/ affrication, and /stɹ/ retraction), we find increases over apparent time, but no effect of covert variability in our laboratory data, consisting mostly of younger talkers. When a sound change has already become phonologized to a new phonemic target with a correspondingly different articulatory target, the original variability is obscured. In comparison, post-lexical coarticulation of word-final /s z/ before a word-initial /ɹ/ more closely resembles /s z/ in tongue posture, with an effect of anticipatory lip-rounding that introduces a low-mid frequency spectral peak during the sibilant interval, and greater reduction in the frequency of this peak for talkers who transition more rapidly to the /ɹ/. In order to uncover the role of covert variability in a sound change, we must look to sounds that exhibit synchronically stable articulatory variability.

Keywords: sound change; covert; variation; coarticulation; rhotic; ultrasound

1 Introduction

This paper explores sound changes in American English that appear to be related to consonant coarticulation with /ɹ/, either synchronically or historically. English /ɹ/ has received special attention in work on sound change because it is implicated in a wide range of phonetic and phonological changes involving consonants and vowels, and because English /ɹ/ is a prototypical example of a sound that exhibits covert articulatory variation, as shown by Delattre & Freeman (1968) and others. Different coarticulatory patterns resulting from covert variability in English /ɹ/ are a likely starting point for the actuation and incrementation of sound change.

Baker et al. (2011) and De Decker & Nycz (2012) have suggested that articulatory variation is more likely to lead to sound change if it is covert, meaning that individuals in a community can exhibit different articulatory versions of the same sound patterns, for purely phonetic reasons, without noticeable acoustic results. Mielke et al. (2010) hypothesized that /s/ retraction in /stɹ/ clusters could differ according to which type of /ɹ/ was triggering it, but found that they could not directly assess the coarticulatory effects of bunched and retroflex /ɹ/ because all of their speakers bunched in the relevant

context. Baker et al. (2011) addressed this problem by focusing on the talkers without extreme retraction and identifying a difference among bunchers, namely degree of similarity between /s/ and bunched /ɹ/ tongue shape, which could account for the degree of coarticulation in speakers without categorical /s/ retraction. This paper represents an additional step in the search for overt signs of coarticulation across covertly variable tongue and lip postures, using data from a series of inter-related investigations into sound change and coarticulatory variability involving English /ɹ/,[1] and explores the mechanisms that may be at the root of these types of sound change.

The goal of this project is to study the articulatory variation that is believed to lead to sound change. To this end, we examine the distributions of several variables in a community of speakers, and we also examine the articulation of variants in the laboratory. These variables include known phonological patterns (/stɹ/ retraction and /tɹ dɹ/ affrication) and sequences of sounds that are expected to show similar phonetic effects (sequences of /s z/ and /ɹ/ both within words and across word boundaries). If we are taking the claims about the role of covert articulatory variation in the actuation of sound change seriously, we should attempt to find a community of speakers who are standing on the precipice of sound change: speakers who exhibit overt variation that can be attributed to covert differences in their speech production, but who do not yet have a shared phonological pattern that overrides the original phonetic motivation. To begin searching for this, in this paper, we are focusing our attention on a known covert articulatory variable (/ɹ/) and some familiar sound patterns involving /ɹ/. A major challenge of this work is to find patterns that vary within the community but that have not been phonologized already, so that the coarticulatory effects of different /ɹ/ variants can potentially be observed. Finding patterns like this is not trivial. Of all the variables under consideration, word-final /s z/ before word-initial /ɹ/ will turn out to be the best example of a phonetic effect of /ɹ/, while the more well-known /stɹ tɹ dɹ/ patterns turn out to already be phonologized in the community under investigation. We will conclude this paper with a discussion of future directions in the search for covert articulatory variation and its consequences.

1.1 Sound changes involving /ɹ/

The most well-studied /ɹ/-conditioned sound change is the retraction of /s/ in /stɹ/ clusters (e.g., *street* pronounced as [ʃtɹit]). It is unclear what the primary initiating factor of /str/ retraction was, but it is either directly or indirectly triggered by the following /ɹ/, and appears as a pronunciation variant in many English-speaking communities, such as Pennsylvania (Labov 1984; Gylfadottir 2015), Georgia (Phillips 2001), Louisiana (Rutter 2011), North Carolina (Piergallini 2011; Wilbanks 2017), Oklahoma (Rutter 2014), Newfoundland (Clarke 2008), the UK (Altendorf 2003; Bass 2009; Glain 2014), New Zealand (Lawrence 2000; Bauer & Warren 2008; Gordon & Maclagan 2008), and Australia (Stevens & Harrington 2016; Stevens et al. 2019). Two different mechanisms underlying /stɹ/ retraction have been postulated. Lawrence (2000) suggests that the /s/ assimilates to an affricated /t/, while most other researchers have argued that long-distance assimilation to the /ɹ/ is responsible (e.g., Shapiro 1995; Baker et al. 2011; Stevens & Harrington 2016).[2] We argue,

---

[1] Data from Wilbanks (2017) and Magloughlin (2018) are included alongside data from the current study, as all are part of a larger-scale project being conducted under the project "Phonological implications of covert articulatory variation" (NSF grant #BCS 1451475).

[2] Perhaps related to this is the phonotactic gap in English for /sɹ/. /ʃɹ/ initially evolved out of /skr/ clusters in West Germanic, in which /r/ may have been an apical tap or trill; however, later sporadic changes in English from phonotactically dispreferred /sɹ/ to the more prevalent cluster [ʃɹ] have also been documented when reduction forces these two sounds together, as in *grocery*, or in loanwords, such as *Sri Lanka*. We do not claim that covert variability is the only method by which a sound change may occur.

based on the evidence in this paper, that both mechanisms of actuation are possible, and may be idiolectally determined.[3]

A phenomenon that appears to be similar to /stɹ/ retraction, though less well studied, is the affrication of /t/ and /d/ in /tɹ/ and /dɹ/ clusters. Talkers with what we refer to as /tɹ/ and /dɹ/ affrication produce the stop portion of the cluster as a postalveolar affricate, such that *tree* and *dream* may sound like [tʃɹi] and [dʒɹim]. The phenomenon has been observed in American English (Read 1971), British English (Jones 1956; Treiman 1987), Australian English (Cox & Palethorpe 2007), and New Zealand English (Hay 2008; Maclagan 2010) (see Magloughlin 2018 for the first in-depth experimental study). The initial stage of /tɹ/ and /dɹ/ affrication was likely a change in the relative timing of gestures. Stops are released more slowly before approximants than before vowels, with longer voice onset time (VOT) and noisier release intervals (Klatt 1975). When a stop is released into a narrow constriction, the increased turbulence ("emergent affricates" in Ohala & Solé 2010) can be interpreted as intentional affrication, which can lead to change (Hall et al. 2006; Ohala & Solé 2010). The co-production of an /ɹ/ with a /t d/ release should color the resulting turbulent noise with the resonances particular to the intermediate gestures. We argue that this coarticulatory pattern should differ based on tongue posture, and that the degree of emergent affrication and its auditory consequences may vary as a function of the compatibility of articulatory gestures, as suggested in Baker et al. (2011). A large body of research has shown that sounds involving tongue tip gestures (e.g., /t d s l n/) exhibit more co-production with following tongue body gestures (e.g., /k g/) than vice-versa (Hardcastle & Roach 1979; Barry 1991; 1992; Byrd 1994; 1996; Byrd & Tan 1996; Chitoran et al. 2002; Kochetov & Goldstein 2005; Chitoran & Goldstein 2006; Kühnert & Hoole 2006; Kühnert et al. 2006). As such, we might expect that, for example, tip-up /t/ might be co-produced with a bunched /ɹ/ relatively easily, or even more easily with a tip-up /ɹ/, with only the addition of a pharyngeal gesture, but that covert variability in alveolar gestures, followed by an /ɹ/, which is also covertly variable, could reduce the co-producibility or enhance it based on the compatibility of the different gestures involved. However, once this change has become phonologized to a postalveolar affricate, we also expect to see a closure position that is more retracted than for prevocalic /t d/, regardless of /ɹ/ tongue shape.

The overt patterns we investigate (/stɹ/ retraction and /tɹ/ and /dɹ/ affrication) appear to have advanced beyond what we might expect to see if they were synchronic phonetic effects that could be the first step in sound change actuation. Therefore, in addition to these previously observed and apparently widespread patterns triggered by /ɹ/, we examine several sequences where we expect coarticulation between /ɹ/ and an alveolar sibilant to occur, giving us the opportunity to test this hypothesis. These include word-internal /ɹ/-sibilant sequences found in words such as *horse* (/ɹs/) and *bars* (/ɹz/), and all four combinations of /s z/ with /ɹ/ across a word boundary: /s#ɹ/ as in *this rock*, /z#ɹ/ as in *his rock*, /ɹ#s/ as in *our sink*, and /ɹ#z/ as in *our zinc*. Following Zsiga (1995), we expect to find more gradient patterns of coarticulation across word boundaries than in word-internal environments, where coarticulation may have become phonologized. We interpret the use of the term "phonologized" to mean that talkers are aiming for an acoustic and articulatory target that is distinct from the older (or underlying) form. We also expect phonetic patterns to exhibit dynamic changes relative to the conditioning environment,

---

[3] It is interesting to note that retraction in /stɹ/ is more extreme and frequent than retraction in /spɹ/ (Baker et al. 2011), even though /spɹ/ has no lingual consonant intervening between the /s/ and the /ɹ/. It is possible that the labial consonant blocks coarticulatory lip rounding, and this is potentially an interesting question for future research.

while we expect phonologized changes to show more temporal stability. For comparison, we also examine prevocalic /t d s z ɹ/ which are not in environments that involve consonant coarticulation.[4]

Previous research such as Baker et al. (2011) leads us to expect patterns like /stɹ/ retraction and /tɹ/ and /dɹ/ affrication to be instances of phonological assimilation for most talkers, and thus be less sensitive to the phonetic details of the local triggering /ɹ/ than during the initial stages of these changes. In other words, these talkers have performed the reclassification of /s/ before /tɹ/ into the /ʃ/ category, as modeled by Stevens et al. (2019). While /ɹ/ may continue to exert some influence on the preceding sound, it is not the same process as that which conditioned the initial change, especially if the target sound is now phonologically distinct. However, we expect the other sequences, particularly the ones that occur across word boundaries, to exhibit more straightforward effects of coarticulation, potentially similar to the coarticulation that occurred in the development of the more well-established patterns of retraction and affrication, and subject to individual variability in covert articulatory strategies.

We take heed of Bermúdez-Otero (2015), who cautions that a sound change, once phonologized, may generalize to less and less specific environments as the sound change is cyclically incremented, and Fruehwald (2012; 2016), who asserts that phonologization occurs at the earliest stages even in phonetically gradual sound change. But we are inspired by the potential revelations offered by somewhat hyperarticulated lab speech, which may shed light on whether these changes are in a stage of phonetically gradual sound change, or are categorical variants, or are in different stages for different talkers, which can be visualized, as suggested by e.g., Fruehwald (2012), Strycharczuk (2012), and Bermúdez-Otero (2015), as a bimodal distribution of variants, with one pattern for those who have categorically adopted a change, and another for those who are still in a stage of phonetic gradualness.

Examining coarticulation in sequences that do not appear to be involved in sound change is critical to understanding how sound change develops in the first place. Even isolated patterns that affect a small number of lexical items can offer important information about how coarticulation and conventionalization interact. While English /ɹ/ tongue shape is a well-known source of covert articulatory variation and is one of the factors we have focused on, many other possible sources of covert variation hold promise for detecting potential patterns of sound change actuation and implementation, as listeners and talkers try to map variable acoustic patterns onto different combinations of gestures and phonological categories.

1.2 Covert articulatory variation

Stevens (1972) demonstrated how small gestural changes may have a large acoustic impact in certain regions of the vocal tract, but relatively larger changes may have much smaller effects in other regions, with areas of large acoustic change corresponding to distinctive features, and defining sound categories within a language. Some of the larger changes in vocal tract configurations that have smaller acoustic consequences involve trading relations between regions of articulation, so that the loss or reduction of one type of gesture may be made up for by the addition of another. Some of these configurations involve very different articulation types that result in a similar acoustic effect. Covert articulatory vari-

---

[4] None of the instances of /ɹ/ include syllabic /ɹ/ (i.e., /ɚ/ or /ɝ/). As one might expect from the structure of English, all instances of /ɹz/ include morpheme boundaries (3rd person singular, possessive, and plural markers, and contractions), and are word-final, while only some of the instances of /ɹs/ include morpheme boundaries or are word-final. These are not analyzed separately. The full set of experimental stimuli are listed in the appendix.

ation occurs when talkers use different articulatory strategies to produce sounds that are not easily distinguished acoustically or auditorily. This may involve the use of completely different tongue postures, as in the case of /ɹ/ (e.g., Delattre & Freeman 1968), which may include any combination of pharyngeal, palatal, and labial constriction locations, or may result from trading relations between articulators to, e.g., maintain a lowered F2 for /u/ by trading between lip-rounding and tongue posture (e.g., Perkell et al. 1993). In the case of /ɹ/, talkers use bunched, retroflex, and other types of tongue shapes to produce /ɹ/ sounds, whose many-to-one articulatory-to-acoustic mapping is effectively indistinguishable by most listeners. A large-scale study on /ɹ/ in North American English showed that about half of all talkers produce only bunched /ɹ/ and half produce retroflex /ɹ/ in at least some phonological contexts (Mielke et al. 2016).[5] Other studies have shown a sampling distribution suggestive of a similar overall distribution (e.g., Uldall 1958; Espy-Wilson & Boyce 1994; Boyce & Espy-Wilson 1997; Alwan et al. 1997; Westbury et al. 1998; Guenther et al. 1999; Tiede et al. 2004). The acoustic difference between these /ɹ/ types is believed to be limited to the spacing of F4 and F5 (Zhou et al. 2008), making it difficult or impossible to identify the gestures underlying an isolated /ɹ/ just by listening – at the very least, the difference does not appear to be salient to most listeners (Twist et al. 2007). The choice of /ɹ/ variant (or variants) used by a given talker appears to be entirely idiolectal in North American English, with twins even being observed to use different tongue shapes (Magloughlin 2016). Research on Scottish English has shown that even though /ɹ/ tongue gestures may be covertly variable, they may become socially distributed, with the outcome that one variant may be the impetus for a sound change such as merger with schwar (Lawson et al. 2013), or derhotacization (Lawson et al. 2011), which leads to overt differences in /ɹ/ production. It is yet unclear how a covertly variable sound may become socially distributed, but Dediu & Moisik (2019) suggest that palate morphology may play a role in adoption of a particular variant, and there is still much that we do not understand about the perception of covert variability.

Variability in lip-rounding for /ɹ/ has not been studied as extensively as the lingual gestures in /ɹ/ or lip-rounding of /u/. However, we extend the characterization in Ladefoged & Maddieson (1996: 93) of rounding in vowels to describe three types of gestures that may be involved in consonant rounding. The first type of gesture involves *protrusion* of the lips, which draws the edge of the lips into a circular and extended shape, primarily using the orbicularis oris muscle, which can be applied to create a wide aperture with protruded lips (*outrounded*), as is the case for English postalveolars such as /ʃ/. The second type of gesture involves a narrow aperture with somewhat less protruded lips (*inrounded*), as is the case for prototypical /u/. The other gesture involved in rounding is lip *compression*, or vertical movement by either the upper or lower lip, or both. A number of researchers, such as Harrington et al. (2011) and Perkell et al. (1993), have demonstrated that some talkers use trading relations between lip-rounding and tongue posture to achieve a lowered F2 in /u/. Studies investigating /u/ have described a range of variability across talkers. For example, Daniloff & Moll (1968) reported that two of their three subjects used unrounded vowels at times, and Perkell & Matthies (1992) excluded a number of potential subjects because they failed to show any lip protrusion. Noiray et al. (2011: 340) summarize: "(i) protrusion is an inconsistent parameter for tracking anticipatory rounding gestures across individuals, more specifically in English; (ii) labial constriction (between-lip area) is a more reliable correlate." There is less literature about variability in lip-rounding in /ɹ/, but Lindau (1985) reports that three of her six American English speakers exhibited

---

[5] A study of 62 New Zealand English speakers by Heyne et al. (2018) found similarly that about 40% produce only tip-down (i.e., bunched) /ɹ/.

no rounding in /ɹ/. Alwan et al. (1997), while not specifically commenting on it, show tracings that suggest protrusion and compression for one talker, protrusion without compression (outrounding) for another, and compression without rounding for the remaining two. In other dialects of English, /ɹ/ may become labiodental as lip-rounding is taken to an extreme (Foulkes & Docherty 2000). Figure 1 shows two different tongue shapes for /ɹ/ (reported by Delattre & Freeman 1968), and lip images from six speakers producing word-initial /ɹ/ (from the production study reported in this paper).

Coronal stops and fricatives also display variability in both lip-rounding and tongue posture. Ladefoged & Maddieson (1996) state that dental and alveolar sounds, across and within languages, can be produced with either apical (tongue tip) or laminal (tongue blade) gestures, as well as with variability in place of articulation. For /t d/, Dart (1998) found that 67–70% of /t d/ tokens in a sample of 20 American English speakers were produced with an apical gesture, and that in a larger sample that included pilot data, 29% of /t/ productions were laminal. For /s/, the same study found that 52.5% of recorded tokens from 20 American English speakers were produced with laminal articulations and 42.5% with apical articulations. Bladon & Nolan (1977) found that seven out of eight British speakers used laminal /s/, and that half of the speakers' /t d/ productions were laminal. Ladefoged & Johnson (2014) suggest that the primary difference between /ʃ/ and /ʂ/ is whether the constriction is made with the tongue tip or blade, i.e., that /ʃ/ is necessarily laminal, but English does not have this phonemic distinction. Keating (1991) characterizes English postalveolars as being primarily laminal with an apical component, which can become the dominant gesture for some speakers: the constriction is usually fairly long, stretching somewhere between the palate and alveolar ridge, but those with a shorter constriction area are more likely to have a more dominant apical gesture. Ladefoged & Maddieson (1996: 148) point out "the secondary articulation of lip rounding is a feature of /ʃ/ in some languages, such as English and French, but it is not found in many other languages such as Russian." Within-language variability in /ʃ/ has not, to our knowledge been extensively discussed. Less articulatory investigation of lip-rounding in /s/ has been reported, but it has been found to vary in several studies that investigated rounding of other segments. Perkell & Matthies (1992) reported that control utterances containing /s/
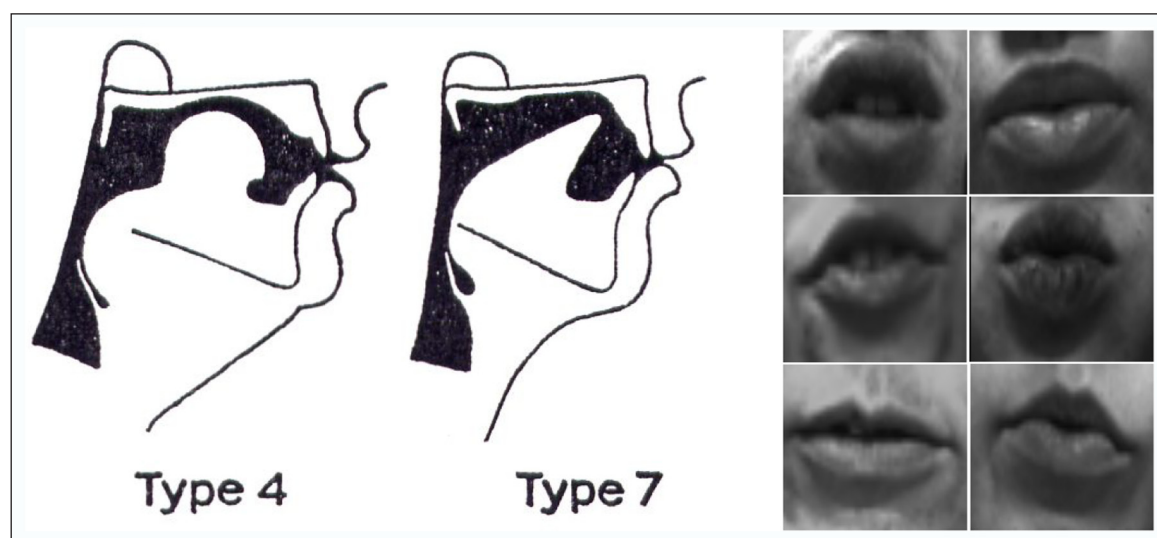

Figure 1: Left two panels: Bunched and retroflex /ɹ/ tongue shapes from (Delattre & Freeman 1968); Right panel: examples of lip rounding variability in word-initial /ɹ/, including open rounding (non-compressed out-rounding) (top left), in-rounding (middle left), compression only (bottom left), and varying degrees of rounding with compression (right panel).

exhibited more protrusion than those not containing /s/. Similarly, Bell-Berti & Harris (1981) reported that three out of six subjects were omitted because they unexpectedly showed activation of the orbicularis oris for /s/ in /isi/. These anecdotal findings suggest that even though /s/ may not be specified for rounding in the community-level grammar, it may be idiolectally rounded.

Lip-rounding is known to lower the frequency of all formants, as it lengthens the overall vocal tract, but either F2 or F3 is particularly affected because lip-rounding especially increases the length of the front cavity. Because /ɹ/ is characterized by the proximity or overlap of F2 and F3, lip-rounding can be viewed as an enhancement strategy. In some talkers, it may be used to compensate for lesser degrees of pharyngeal or palatal constrictions (i.e., *trading relations*, as discussed in Guenther et al. 1999, or as related to /u/, in Harrington et al. 2011). In fricatives and affricates, lip-rounding lowers the center of gravity of the spectrum and the spectral peak of the frication noise. Thus, lip-rounding has the potential to obscure the location of the constriction that is inferred by the listener.

Additional covert factors to consider include variability in individual vocal tract morphology and phonological settings. Several researchers (e.g., Mooshammer et al. 2004; Brunner et al. 2005; 2009; Rudy & Yunusova 2013) have found that vowels and coronal consonants may be produced with greater gestural variability by talkers who have a higher domed palate shape than those who have a flatter palate. While the variably produced sound may not be noticeably different in acoustic terms, variability in coronal gestures could affect coarticulatory patterns, which may be noticeable. Ellis & Hardcastle (2002) showed that talkers employ a variety of strategies when coarticulating two very different sounds (in their case, /n/ and /k/), including lack of assimilation, partial assimilation (gestural overlap or co-production), and complete assimilation. It seems likely that, in addition to gestural compatibility, individual phonological settings may also play a role in degree of assimilation between adjacent sounds. Other researchers have also reported variability in coarticulatory timing and degree of overlap (e.g., Byrd 1994; Byrd & Tan 1996), and velocity and magnitude of coarticulated gestures (e.g., Kuehn & Moll 1976; Ostry & Munhall 1985).

An important implication of covert articulatory variation, as suggested by Baker et al. (2011), is that the coarticulatory effects of different gestures should differ from each other, and produce between-talker differences in coarticulation strategies, which can result in different acoustic consequences with no overt cause or explanation. Covert articulatory differences cannot be spread through a speech community, but the acoustic consequences of variable coarticulatory patterns could be a locus for change. Bermúdez-Otero (2019) claims that individual level variability accounts for only "small-scale effects superimposed on larger macroscopic patterns that turn out to be driven by infra- and supra-individual mechanisms." We agree that the advance of a sound change in progress may in part be driven by infra- and supra-individual mechanisms, but argue that the actuation of a sound change, that is, the divergence of synchronic phonetic variability from diachronically stable variability, is not "small-scale", and must be brought about by individual variability. Bermúdez-Otero's claim, that an account of sound change beginning with covert variability must result in the eventual expansion of sound change to all regions in which the language is spoken, is a straw man. The advance of sound change necessarily depends on whether an innovation is adopted. As Bermúdez-Otero postulates, "learners reject an innovation if it is isolated or randomly scattered, but they adopt it and actively increment it if it accidentally displays an inverse correlation with age." Covert variability will be randomly distributed, thus the phonetic effects of covert variability on a particular sound pattern will also be randomly distributed. But a random distribution does not imply an equal distribution. A random distribution predicts that there may be larger clusters in

some locations at some times. If the different acoustic coarticulatory patterns that result from covert articulatory differences cluster in a younger age group in a particular place at a particular time, to a greater degree than the preceding generation, it could spark a sound change. Dediu & Moisik (2019) also argue: "when factors such as the frequency of the biased speakers in the community, their positions in the communicative network or the topology of the network itself change, sound change might rapidly follow as a self-reinforcing network-level phenomenon, akin to a phase transition."

Furthermore, since talkers with different covert variants may have different degrees of phonetic motivation for a particular sound pattern, covert articulatory differences provide a way to distinguish between the direct influence of articulatory factors (which apply to individuals) and emerging community-level speech norms. Following Baker et al. (2011) and De Decker & Nycz (2012), we predict that articulatory variation is more likely to lead to sound change if it is covert. However, we think it is probably premature to predict a typology of covertly-motivated sound changes, because our knowledge of covert variables is quite limited (e.g., there are well-studied cases like tongue shapes for /ɹ/ and lingual-labial trading relations for /u/), and a comprehensive search for previously unknown covert variables has (to our knowledge) not been conducted. Similarly, we think it is premature to conclude that covert articulatory variation was not involved in a particular sound change (c.f., Bermúdez-Otero 2019), because we simply do not know enough about covert articulatory variation. The goal of this paper is to explore the consequences of one familiar instance of covert articulatory variation that is known from the literature.

There have been some recent explorations of covert articulatory differences at the source, by examining biomechanical and anatomical variation. Harandi et al. (2017) use biomechanical modeling to investigate inter-speaker differences in muscle activation to achieve the same speech goals, and Dediu & Moisik (2019) describe the genetic basis for covert differences between groups of speakers, and note that most communities are heterogeneous with respect to covert articulatory differences that may be relevant to the phonetic motivation for a change or to the interpretation of innovative variants. The phonological relevance of covert articulatory differences is due to the fact that any sound pattern that is sensitive to the range of hyper- to hypo-articulated speech (see, e.g., Lindblom 1990) will be sensitive to differences in the articulation of its trigger.

Ohala (2011) has argued that the most obvious active role taken by speaker-listeners is to *maintain* accepted norms, and studies of variation and change have repeatedly revealed striking uniformity even in changes in progress and stable variation, and in the internal factors governing changes (Guy 1980; Labov 2009; Forrest 2015). Ohala (1989: 175) emphasizes the importance of "hidden" variation, i.e., that "speakers exhibit variations in their pronunciation which they and listeners usually do not recognize as variation." By definition, covert articulatory variables are linguistically relevant facts that cannot be set by the community. If the cause of coarticulation is covert, perceptual compensation for it is a more challenging task for the listener (Beddor 2009), and interaction between individuals with different degrees of coarticulation (and especially different gestures underlying that coarticulation) is particularly likely to result in the exaggeration of coarticulation that is necessary for articulatorily-driven sound change.

Baker et al. (2011) examined the articulation of /stɹ/ clusters as produced by American English speakers with and without apparent retraction patterns. The non-retracting speakers were used as a proxy for speakers who predated the beginning of /stɹ/ retraction, for whom the effect of /ɹ/ might have been truly coarticulatory. To observe covert articulatory variation and its potential phonological consequences in a community context, we have scaled up Baker et al.'s investigation in two ways: by investigating articulation and acoustics of several /ɹ/-conditioned phonological patterns in contexts with greater /ɹ/

variation (including ones where more categorical inter-speaker differences are expected), and by examining all of the variables from our laboratory study in a matched corpus study (a 162-talker sample of the Raleigh Corpus; Dodsworth & Kohn 2012) that allows us to compare the distribution of these variables in spontaneous speech, and distinguish changes in progress from stable variation.

2 Corpus study

To investigate the role of overt (auditory and acoustic) variation in phonological patterns related to /ɹ/, we examined these sound patterns in a 162-talker sample of the Raleigh Corpus (Dodsworth & Kohn 2012; Dodsworth 2014) – a collection of sociolinguistic interviews conducted with English speakers born and raised in Raleigh, North Carolina, which is an urban center in the southeastern United States. Speakers included in the analysis consisted of 89 women and 73 men, born between 1919 and 1996. To look for change over apparent time in the production of the target variables, we used a combination of auditory coding, automatic classification, and acoustic measurements.

2.1 Data sources

The Raleigh corpus interviews were recorded as described in Dodsworth & Kohn (2012) and Dodsworth (2017), using a lapel-mounted microphone, often in a speaker's home. None of the talkers from this subset reported any speech or hearing disorders, or displayed signs of any oral abnormality or neurological disease, as far as we are aware. All recordings were digitized at a sampling rate of 44.1 kHz with 16 bits per sample, and were force-aligned using the Penn Phonetics Lab Forced Aligner (Yuan & Liberman 2008). Some hand-correction made sure that speech intervals were aligned with the correct sections of the interview, but no hand-correction of segment boundaries was performed on the portions of the corpus that we used. Additional methods are described in the appropriate sections of this paper. We also collected laboratory speech from 29 speakers, who were mostly from North Carolina, referred to here as the May corpus. The laboratory data collection is described in more detail in Section 3, but we introduce the acoustic recordings here in order to present them alongside the Raleigh corpus data. The participants in the laboratory study read target words and phrases within a carrier phrase "give me a __ again". Token counts for each variable in the Raleigh corpus and the laboratory study described in Section 3 are listed in the supplementary files.

2.2 Auditory coding of overt variation
2.2.1 Auditory coding methods

In order to look for change over time, two of the authors and a third linguist listened to and rated sound clips containing each of the variables from the Raleigh Corpus and from the May corpus. A detailed description of coding methods and inter-rater statistics are listed in the Supplementary files, but here we note that /tɹ/ and /dɹ/ tokens were coded as "less affricated" (more [t d]-like) or "more affricated" (more [tʃ dʒ]-like), and tokens containing /s/ or /z/ were coded as "less retracted" (more [s z]-like) or "more retracted" (more [ʃ ʒ]-like). Missegmented or otherwise flawed tokens were removed from this and other analyses. "Less retracted" and "less affricated" judgments were recorded as 0, and "more retracted" and "more affricated" judgments were recorded as 1. The 1s and 0s were averaged across listeners for graphical purposes. The graphs of listener judgments represent the mean judgments across all tokens for each talker. Thus, values near 1 indicate agreement on the presence of affrication for /t/ and /d/ or retraction for tokens involving /s/, for most tokens for that talker. Values near 0 indicate rater agreement on the percept of lack of retraction or affrication for most tokens for that talker. The judgments for the

May corpus participants are added to the graphs with black outlines, for reference, but their values are not included in the loess curves shown in the graphs, or in the data used for statistical analysis.

A logistic mixed effects regression model was applied to raw (unaveraged) scores (0, 1) for each variable from the conversational tokens in the Raleigh Corpus, with centered year of birth and gender as fixed effects and talker, word, and rater as random effects: `lmer(score~birthyear*gender+(1|talker)+(1|word)+(1|rater)`. For /tɹ/ and /dɹ/ affrication and /stɹ/ retraction, additional methods were used, detailed in following sections.

2.2.2 Auditory coding results and discussion

The retraction judgments for word-initial and word-medial /stɹ/ are displayed by talker birth year and gender in Figure 2. The loess curves in the top panel suggest that initial /stɹ/ was more retracted for older males, but that in the later cohort, change in retraction follows a similar upward trajectory for both men and women. The figures in this paper display the relationship between birth year and phonetic measures using loess curves, while the regression models explore the related hypothesis that there is a linear relationship between birth year and the phonetic measures. In other words, what is illustrated in the figures is not identical to the questions addressed by the regression models reported in the text. In most cases, the figures do suggest a linear relationship, and when they do not, we will address this.

Averaged out over birth year, however, the /stɹ/ retraction model shows only a higher intercept for men ($z = 2.69$, $p < 0.005$), and a marginal overall decrease over birth year for men only ($z = –1.92$, $p = 0.55$). This decrease is probably attributable to a broader general sibilant retraction pattern that is observed in males born in the first half of the 20th century (Wilbanks 2017), as will be discussed in a later section. The lab data from the May corpus, not included in the statistics or loess curve, but overlaid on the graph as symbols with black outlines, shows an essentially bimodal distribution, with a cluster of 10 talkers born after 1990 who have very high rates of retraction. This is suggestive of complete phonologization of /stɹ/ → [ʃtɹ] for these talkers (or, in other terminology, "stabilization" of the sound change). Notably, the younger Raleigh corpus talkers also display a split, in which 8 talkers born between 1978–1990 are essentially responsible for the increase shown in the loess curve, while the other talkers received less than 50% retracted judgments.

Word-medial /stɹ/ retraction is shown in the bottom panel of Figure 2. There is again a dip followed by an increase in apparent time of retraction of word-initial /stɹ/, for both men and women. This does not show up as a significant change in slope in a linear model. However, modeling time with two orthogonal polynomials, motivated by the parabolic shape of the curves, yields a significant difference in the Akaike information criterion (AIC) between models ($X^2(2) = 10.4$, $p < 0.01$), and both genders show a strong positive effect of the 2nd polynomial ($z = 3.62$, $p < 0.001$), and neither slope nor intercept differ between men and women. Again, this initial decrease is attributable to the reduction of overall /s/ retraction over time, which interferes with our ability to visualize the trend of /ɹ/-motivated /stɹ/ retraction. The increase in the youngest generation of Raleigh speakers, as was the case with initial /stɹ/, appears due to a handful of very retracted talkers, whose bimodal distribution is even more apparent in this graph. The May corpus, again, not included in the statistical analysis or loess curve, but overlaid on the graph with black outlines, also shows a cluster of 10 talkers born after 1990 who have higher rates of retraction, suggesting a phonological leap in retraction is continuing for some of the talkers in this youngest generation.
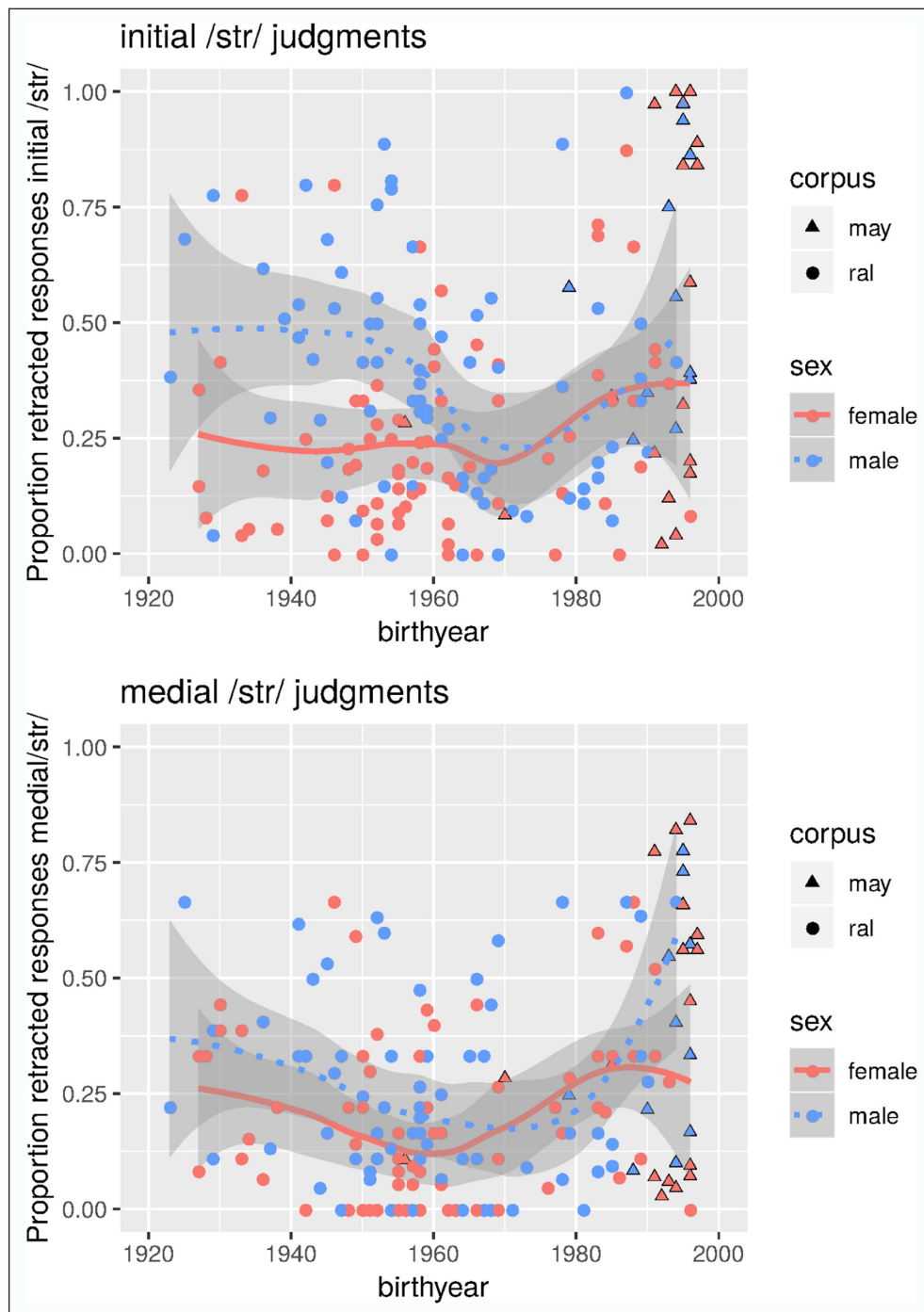
Figure 2: Retraction judgments for /stɹ/ in word-initial (top) and word-medial (bottom) contexts in Raleigh corpus and lab data.

The top panel of Figure 3 shows affrication of /tɹ/ increasing by birth year ($z = 2.6$, $p < 0.05$). The slope for males and females is not significantly different ($p = 0.74$), but the higher intercept for male talkers ($z = 2.95$, $p < 0.005$) is potentially due to the listeners' expectations of greater affrication for females, given that women tend to be leaders in sound change. It is also possible that the raters were paying attention to the center of gravity (COG), which is higher for females, and possibly less likely to be judged as [tʃ], especially when heard between randomized tokens containing male /tɹ/, which would have a lower COG. Although not included in the statistical analysis or model curves, it is worth noting that for both /tɹ/ and /dɹ/, the lab-produced tokens generally cluster in the
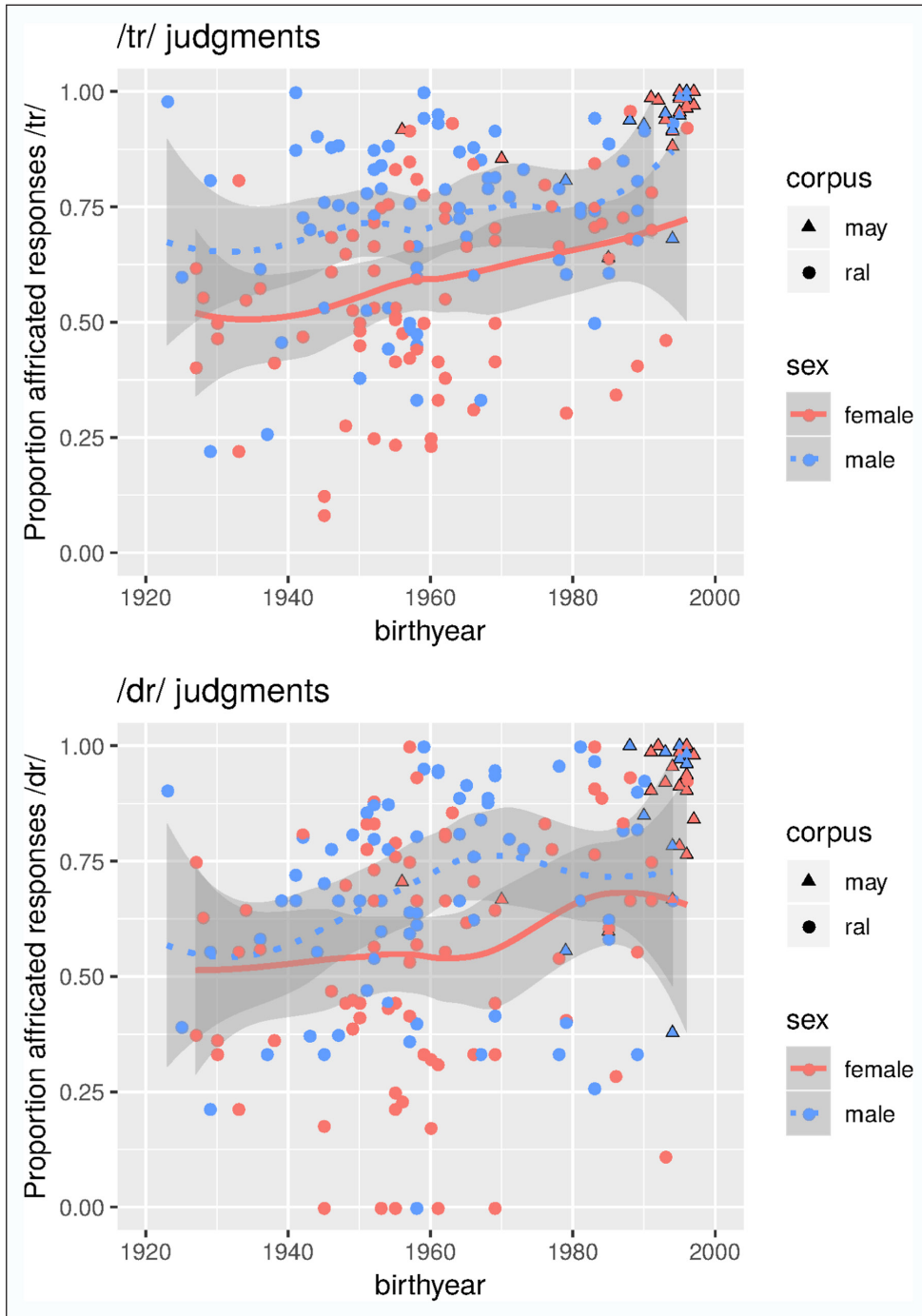
Figure 3: Affrication judgments for /tɹ/ (top) and /dɹ/ (bottom).

upper right corner. The median birth year of the lab talkers was 1994, with only five talkers born before 1990 (who had correspondingly lower affrication scores). This suggests that the youngest generation has completely phonologized/stabilized affrication of /tɹ/, with consistently affricated targets produced in more conservative lab speech. The trajectory of this shift for the older talkers looks to be fairly linear, with an approximately even distribution of variation in each synchronic time slice.

The bottom panel of Figure 3 also shows affrication of /dɹ/ increasing by birth year ($z = 2.02$, $p < 0.05$). The slope for males and females is not significantly different ($p = 0.7$), but again, the higher intercept for male talkers ($z = 2.2$, $p < 0.05$) is potentially due to listener expectation differences based on gender or COG. The trajectory

appears to be somewhat linear, with broader variability during the middle years of the change, as men abruptly increased their use of an affricated variant. It may be that at that point, some talkers started incrementing more quickly or adopting a phonologically affricated variant, while other talkers only continued the slower phonetic shift (or adopted a retrograde variant). Again, the younger May corpus talkers (not included in the loess curve), are primarily clustered in the upper right hand corner, most with completely phonologized/stabilized variants.

The results for all of the non-phonologized contexts are pooled into two graphs in Figure 4. The top panel includes the four cases of /ɹ/ followed by a sibilant (word-internal /ɹs/ and /ɹz/, along with /ɹ#s/ and /ɹ#z/), and the bottom panel includes the two cases of a sibilant followed by /ɹ/ (/s#ɹ/ and /z#ɹ/). The graphs have been collapsed in this
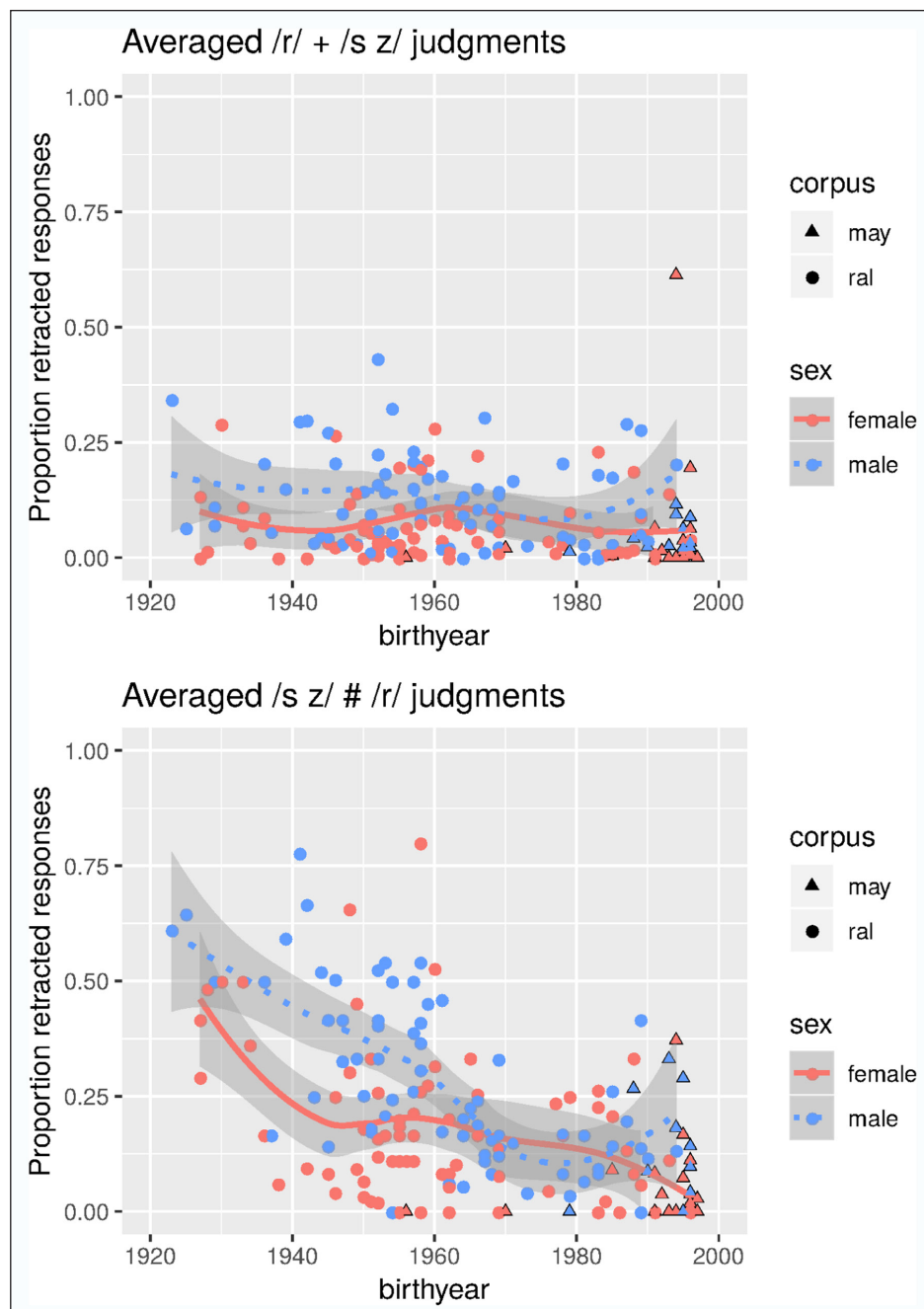
Figure 4: Retraction judgments for all /ɹ/ followed by /s z/ (top) and /s z/ followed by /ɹ/ (bottom).

way because the pattern is very similar across variables, and the token counts for some sequences (/ɹ#z/ in particular) are small.

Word-internal /ɹs/ shows little or no retraction for most talkers, and does not show a significant change in retraction over time for men or women, but does reflect a higher baseline of retraction for men, seen in their intercept ($z = 2.12, p < 0.05$). /ɹ#s/ also does not show significant retraction over time for men or women, but reflects a higher baseline of retraction for men ($z = 3.1, p < 0.01$). Word-internal /ɹz/ shows even less retraction for most talkers, and does not show any significant change in retraction over time for men or women, nor any difference in baseline retraction by gender. /ɹ#z/ has only three tokens in the entire Raleigh corpus, thus it was not subjected to statistical analysis, but we can speculate that any degree of retraction shown by any talker is likely to be idiosyncratic.

/s z/ before /ɹ/ is more comparable to /stɹ/, /tɹ/, and /dɹ/, in that the alveolar sound preceding the /ɹ/ might reflect some anticipatory coarticulation similar to that which may have conditioned the changes in /stɹ/, /tɹ/, and /dɹ/, being related to the tauto-morphemic patterns. Figure 4 (bottom) shows averaged retraction scores for /s z/ before /ɹ/, which reflects the pattern of general reduction in retraction of sibilants over time in Raleigh, with a negative linear slope for women (/s#ɹ/: $z = 3.85, p < 0.001$; /z#ɹ/: $z = 3.17, p < 0.005$), and a steeper reduction for men (/s#ɹ/: $z = 1.77, p = 0.079$; /z#ɹ/: $z = 23.1, p < 0.005$), who also have a higher baseline of retraction reflected in their intercept (/s#ɹ/: $z = 3.2, p < 0.005$; /z#ɹ/: $z = 3.4, p < 0.005$).

Figure 5 summarizes the auditory coding for all of the variables for the Raleigh corpus. The speakers are divided into the three generations defined by Dodsworth & Kohn (2012) based on in-migration patterns. The variables that show signs of increase over time are /tɹ/ affrication, /dɹ/ affrication, and word-medial /stɹ/ retraction. Word-initial /stɹ/'s recent increase is within generation 3 (born 1967–1996). The variables that show signs of decrease in retraction are initial /stɹ/ (within the older two generations), word-final /s/, and /z/ before /ɹ/. As far as we know, this decrease of retraction in apparent time is not limited to the /ɹ/ context, but a following /ɹ/ may be associated with more retraction in the older speakers who have it.

Figure 6 shows comparisons of /tɹ/ affrication, /dɹ/ affrication, and word-initial /stɹ/ retraction for Raleigh corpus speakers. The top panel shows that /tɹ/ and /dɹ/ affrication are correlated (adjusted $R^2 = 0.4364$ based on a simple linear regression of speaker
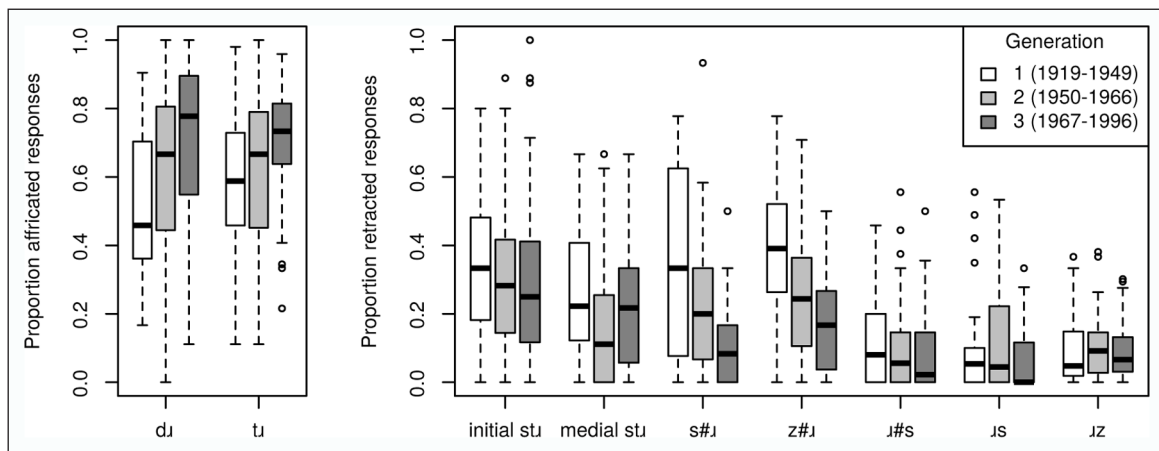

Figure 5: Summary of affrication judgments (left) and retraction judgments (right) by speaker generation. Values near 1 indicate agreement on the presence of affrication or retraction both within talker and across rater, while values near 0 indicate agreement of absence of affrication or retraction both within talker and across rater.
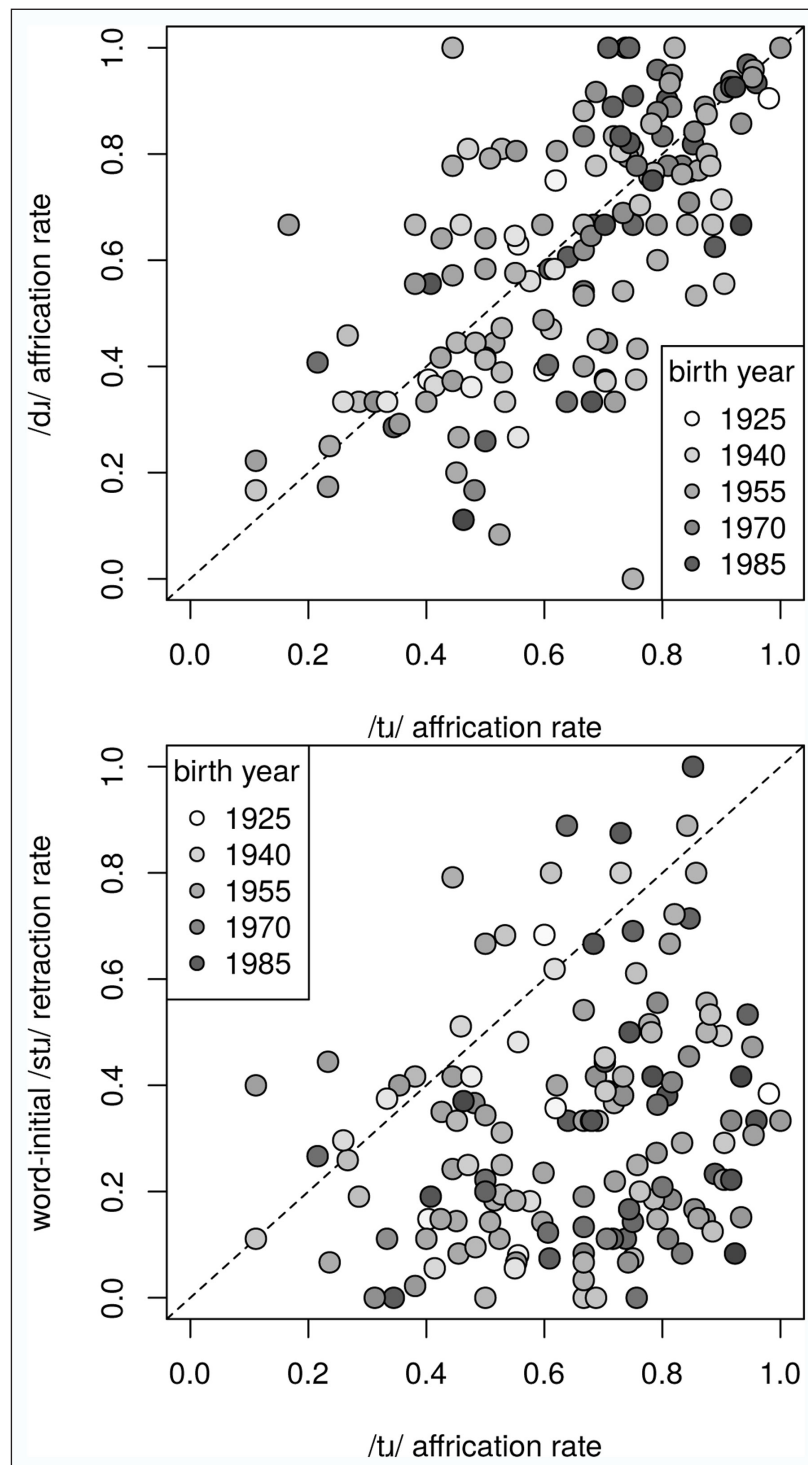
Figure 6: Comparison of variables by talker in the Raleigh corpus. Top: perceived affrication in /tɹ/ and /dɹ/ clusters; Bottom: perceived affrication in /tɹ/ and retraction in /stɹ/.

means), with more of the younger talkers clustering in the upper right quadrant, indicative of affrication of both clusters. The bottom panel shows that /stɹ/ retraction is less closely related to /tɹ/ affrication (adjusted $R^2 = 0.0431$). That is, a talker who affricates /tɹ/ is not particularly more likely also to retract /stɹ/, as seen by tokens in the lower right quadrant. However, the absence of any tokens in the upper left corner indicates that none of our participants retracted /stɹ/ without also affricating /tɹ/. This does not rule out the mechanism of /s/ retraction being triggered by /tɹ/ affrication. If we had seen

talkers populating the entire possible space, especially the upper left quadrant, we could have safely eliminated that hypothesis.

In summary, auditory coding of Raleigh corpus speech shows that affrication of /tɹ/ and /dɹ/ has been increasing in Raleigh over time. Retraction of /stɹ/ appears to be on the rise among the youngest generation in word-medial position. /s z/ in other /ɹ/-adjacent contexts do not appear to be becoming more retracted. However, this question is complicated by the fact that older talkers, particularly males, have more retracted /s/ in general. Generally retracted /s/ is known to be a stylistic marker of "Southernness" or a "country" way of speaking (Campbell-Kibler 2011; Podesva & Hofwegen 2014), though its usage in the Southern states has not been extensively examined. It appears that younger Raleigh speakers are using this variant less than older speakers, just as there is reduced usage of Southern-associated vowel variants in the younger generations (Dodsworth & Kohn 2012). The data presented in this section were based entirely on auditory coding by three listeners who differed considerably in their threshold for perceiving retraction and affrication of the speech tokens. Even with this variability, a pattern is apparent for /tɹ/, /dɹ/, and /stɹ/. In the next three subsections, we report non-perceptual approaches that may be able to corroborate or contradict the findings based on auditory coding.

2.3 Acoustic analysis of /stɹ/ retraction

In this section, we report on an acoustic approach to measuring /stɹ/ retraction in the Raleigh corpus (similar to that found in Wilbanks 2017), which is able to take into account the fact that a key reference point for /stɹ/ retraction: /s/, in many other contexts, is not stable over time in Raleigh.[6] Additionally, with this new analysis, we are able to see just how much the /ɹ/ contributes to /stɹ/ retraction above and beyond whatever retraction may occur in /st/ clusters.

2.3.1 Acoustic analysis methods

To calculate an acoustic measure of the change in /stɹ/ relative to /st/ across the Raleigh corpus subset, the center of gravity (COG) was measured from a 30 ms interval at the midpoint of the fricative interval in word-initial and medial singleton /s/, /st/, /ʃ/, and /stɹ/, and then a ratio was calculated of how far the COG of each /stɹ/ was from each talker's mean COG for /ʃ/, relative to each talker's average distance between /st/ and /ʃ/, as in Equation 1. Factor levels were contrast coded so that prevocalic /st/ for women is the baseline against which the other sounds are measured.

$$(1) \qquad \text{retraction ratio} = 1 - \frac{\text{mean(SH)} - \text{STR}_i}{\text{mean(SH)} - \text{mean(ST)}}$$

This method, following Baker et al. (2011), is described in more detail in Wilbanks (2017).[7] It was important to normalize the COG for /stɹ/ relative to some baseline COG measures because some talkers, in particular older men, have a lower COG for /s/ in many environments. Since we included this measurement for medial as well as initial tokens, and included baselines for singleton /s/, /st/, and /ʃ/, in addition to /stɹ/, the model is a bit more complex than the other analyses: `lmer(retraction_ratio~birthyear*phone_type*position*gender+preceding_manner+log(duration)+(1|talker)+(log(duration)|word))`.

---

[6] See Piergallini (2011) for an earlier study of /stɹ/ retraction in the Raleigh corpus.

[7] Tokens with duration under 40 ms or COG below 2,000 Hz or above 10,000 Hz were omitted as a basic quality control measure. Tokens preceding or following a known palatalizing environment were also omitted. Remaining tokens had COG measured in a 30 ms rectangular window centered on the midpoint, hopefully resolving any missegmentation issues caused by loose boundary placement from forced alignment.

2.3.2 Acoustic analysis results

Figure 7 presents COG measurements in raw Hz for all four consonant sequences of interest, separated by gender (based on data reported in Wilbanks 2017). This figure helps to visualize the degree of general Southern /s/-retraction over time, explains some of the results of the transformation, and highlights the necessity of using a retraction ratio to normalize /stɹ/ relative to /st/ and /ʃ/. One can see by raw Hz values that men and women both had /s/ and /st/ productions that were closer to their /ʃ/ productions in the oldest cohort, but that this retraction is especially pronounced for men, for whom these sounds almost overlap. Women (and men to some degree) from this cohort also have slightly more retraction in /st/ than in /s/ in other (non-palatalizing) environments (consistent with Baker et al. 2011). Even without accounting for the distance between /st/ and /ʃ/, there is a clear reduction in COG for women's /stɹ/ productions, especially word-medially. Men's /stɹ/ appears to be closely linked to the COG of their /s/ and /st/, which is increasing in apparent time, peaking and/or plateauing for the youngest cohort, which, as we saw, interfered with visualizing the effects of the /ɹ/ on the /s/ independently from the effects of more general /s/-retraction. The degree of difference between /s/ and /ʃ/ is increasing especially for men, suggesting reversal of an earlier sound change.
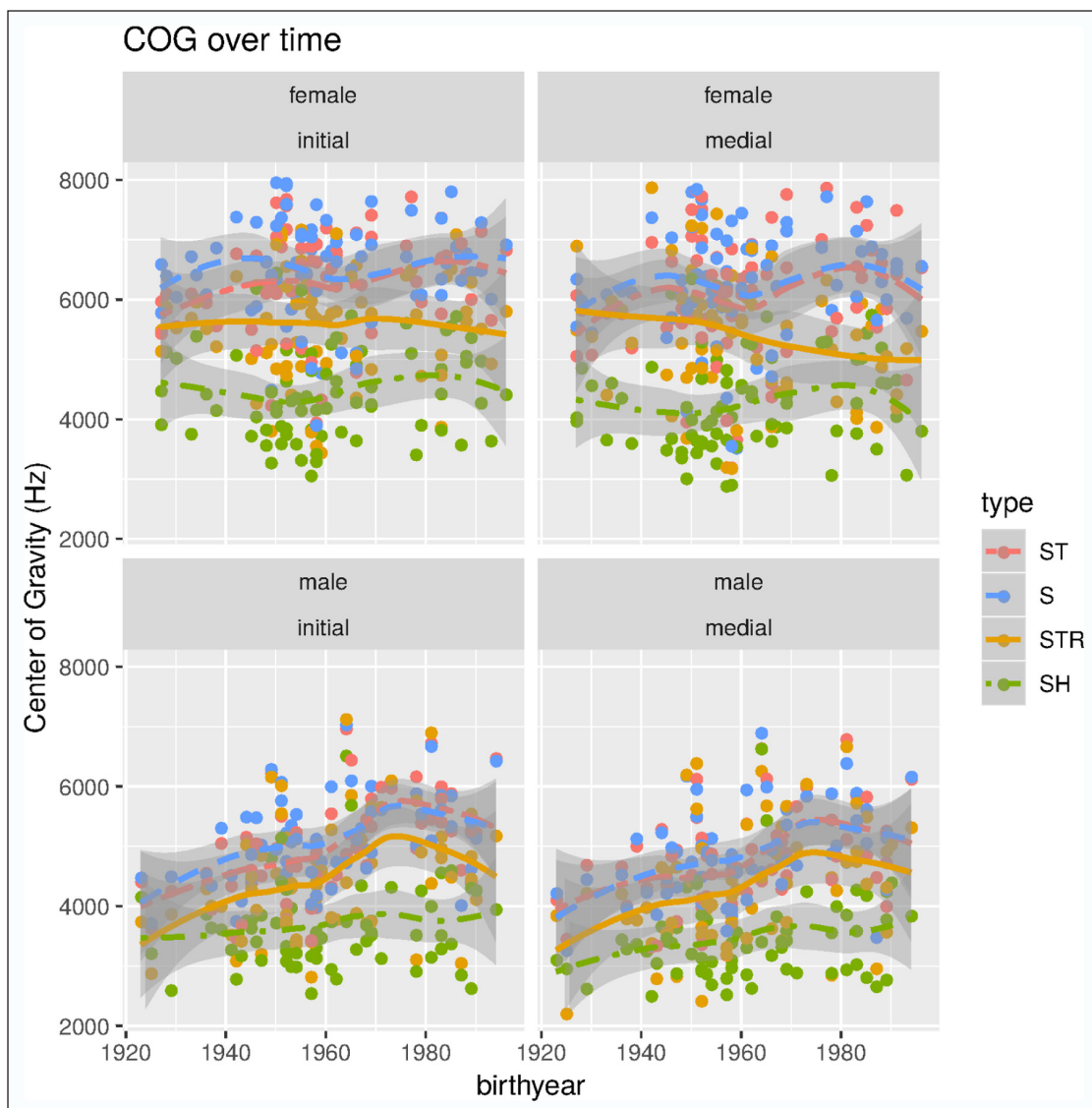

Figure 7: /s/ retraction in Raleigh: word-initial contexts (left) and word-medial contexts (right).

The inverted ratio of the difference between /ʃ/ and /stɹ/ relative to the difference between /ʃ/-/st/ is shown in Figure 8. The ratio is inverted so that more retraction is reflected by an increase in the retraction ratio, so that "up" equals "more". A higher ratio in the graph thus indicates a greater degree of similarity between /stɹ/ and /ʃ/. The top and bottom panels show word-initial and word-medial contexts (respectively) by birth year. Note that the degree of scatter is much greater for /stɹ/ than any of the other contexts, showing a great deal of phonetic (and perhaps phonological) variability in /stɹ/.

Full model outputs are in the supplementary files, and the many levels of complexity have been distilled into the most important points, so not all model estimates will be given here. The most apparent change is that /stɹ/ is becoming more retracted by birth year for women, word-initially (Est = 0.39, $p < 0.001$), and significantly more so medially (Est = 0.144, $p < 0.005$). But, as the graph illustrates, men's /stɹ/ productions appear relatively stable over time, requiring a negative adjustment by birth year relative to the women's increase over time, both initially (Est = −0.117, $p < 0.005$) and medially (Est = −0.152, $p < 0.05$). Longer duration segments have less retraction on average (Est = −0.38, $p < 0.001$). Unsurprisingly, both /stɹ/ and /ʃ/ are more retracted than /st/ for both men and women, both word-initially and medially, but /s/ is less retracted than /st/ for women in both contexts (initial: Est = −0.166, $p < 0.001$; medial is *n.s.* different), while the distance between /s/ and /st/ is smaller for men than for women in both
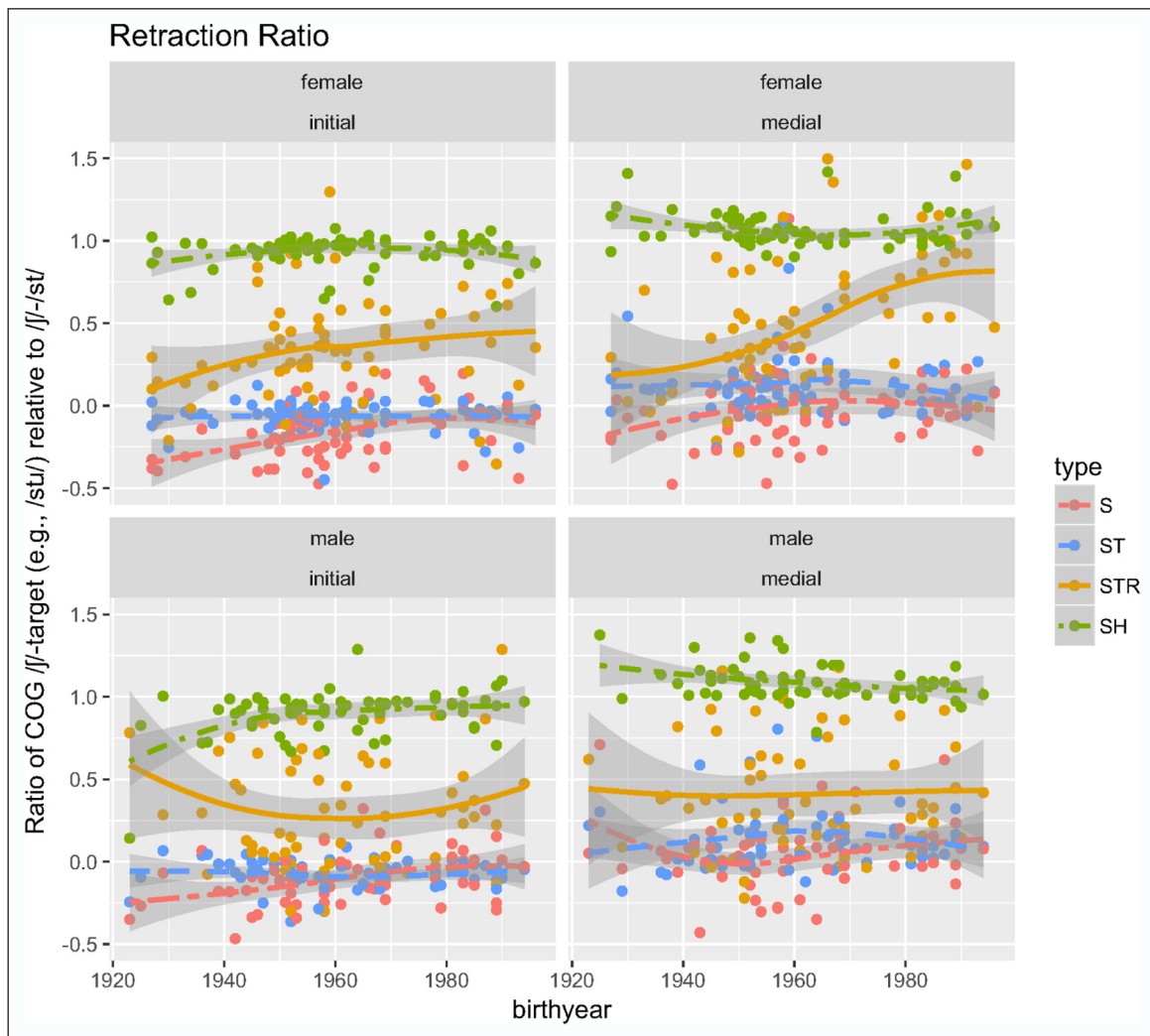

Figure 8: /s/ retraction in Raleigh: word-initial contexts (left) and word-medial contexts (right).

contexts (initial: Est = 0.08, $p < 0.001$; medial: Est = 0.06, $p = 0.055$). The retraction ratio of prevocalic /s/ for women is also increasing relative to /st/ because, as we saw in Figure 7, /st/ is becoming less retracted over time (initial: Est = 0.094, $p < 0.001$; medial: *n.s.* difference). All medial segments have a lower COG than initial segments, for men and women, but men's medial /ʃ/ shows a reduction in retraction ratio by birth year relative to the flatter trajectory for /st/ (Est = –0.099, $p < 0.01$).

In summary, /stɹ/ retraction seems to behave somewhat differently from /tɹ/ and /dɹ/ affrication. While the latter seems to have already been a change-in-progress in the older talkers, with a steady climb through the last generation, /stɹ/ has quite variably spanned the extremes between /s/ and /ʃ/ throughout this time sample, and is only now beginning to increase abruptly, for female talkers, starting with the youngest generation, as with many changes in Raleigh, NC. Phonetic (and perhaps phonological) variability is evident for /stɹ/ throughout the corpus, so this late increase is a noticeably different pattern. The spectral measures presented in this section provide confirmation that much of the /stɹ/ retraction that we noticed in older talkers was a result of their more general retraction of /s/, which is greatly reduced over apparent time, as with many other Southern features, and that the changes seen in the younger generation are reflective of a trend towards /stɹ/ retraction, which may have become phonologized for some of the youngest talkers, as we will see in the articulatory study, but which is also evidenced by several (especially female) talkers who have productions of /s/ that are as much as, and sometimes more retracted than their /ʃ/. The abruptness of this shift suggests rapid phonologization of a postalveolar sibilant target, but the fact that medial and shorter duration segments experience more retraction hint at the phonetic conditioning of the initial change.[8]

2.4 *Acoustic analysis of post-lexical /s z/ retraction*

The acoustic methods applied to /stɹ/ can also be usefully applied to word-final /s z/. Recall /s#ɹ/ and /z#ɹ/ were coded as retracted more often for older speakers. These auditory coding results do not indicate whether retraction is caused by /ɹ/ or if it is a general context-free feature of older speakers' /s z/. We have applied the same acoustic methods as in the previous subsection in order to see whether the older speakers' /s#ɹ/ and /z#ɹ/ are acoustically more retracted than their /s/ and /z/ in other contexts. Figure 9 shows talker-averaged ratios of their center of gravity (COG) of /s#ɹ/ and /z#ɹ/ relative to their general retraction of prevocalic /s z/, calculated by relative COG of /s#ɹ/ and /z#ɹ/ divided by COG of prevocalic /s z/. The majority of talkers have a lower COG preceding /ɹ/ (values less than 1), but a linear regression model of these talker averages suggests that there is no significant change over time, only hinting that men's /s#ɹ/ and /z#ɹ/ are getting a tiny bit closer to their prevocalic /s z/ ($p = 0.13$), which does not account for the large reduction in retraction seen in the older generations. This is evidence that the patterns where we saw a reduction in retraction before /ɹ/ are linked to a more global reduction in retraction for /s z/, and do not reflect a change in /ɹ/-conditioned retraction over time.

2.5 *Automatic classification of stop affrication*

Compared to /s z/ retraction, /tɹ/ and /dɹ/ affrication is harder to address with acoustic methods, because it involves a more complex set of phonetic differences, and because the differences are distributed across two phonetic segments (practically-speaking, two textgrid intervals).[9] Instead of measuring acoustic differences directly, an automatic classification

---

[8] We cannot rule out the possibility that the abrupt shift could potentially be due to dialect contact.
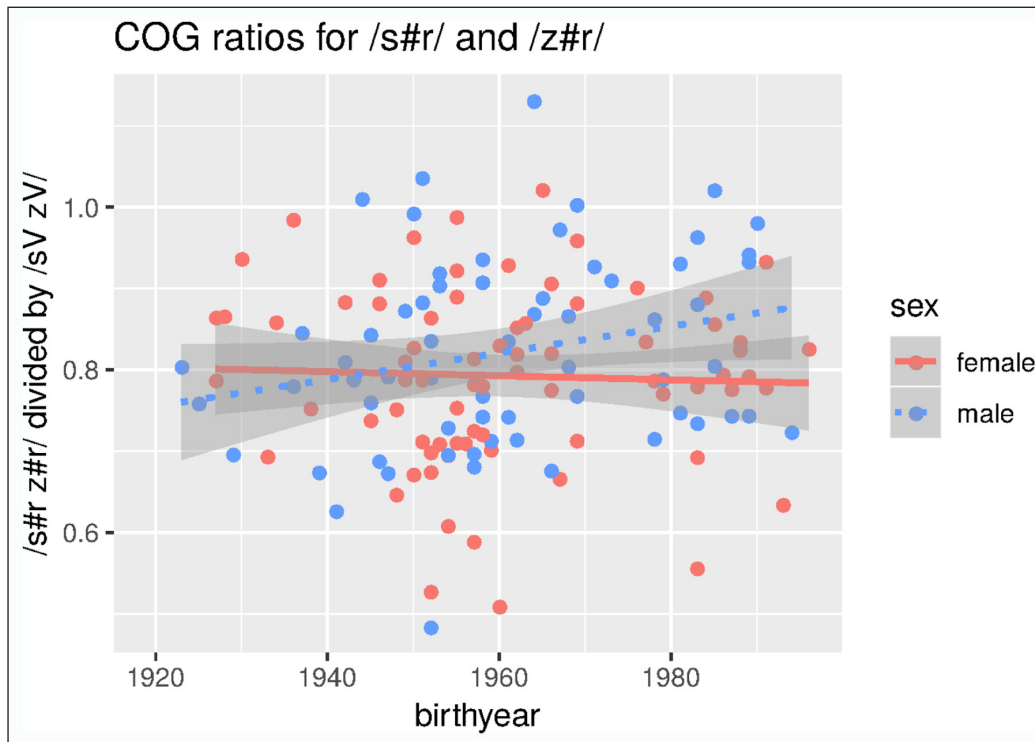[9] But see Hatcher (2017) for an acoustic investigation of /dɹ/ affrication in the Raleigh corpus.

Figure 9: Average retraction of /s#ɹ/ and /z#ɹ/ relative to general retraction of prevocalic /s z/, as shown by relative center of gravity (COG) of /s#ɹ/ and /z#ɹ/ relative to (divided by) COG of prevocalic /s z/. If anything, men show less retraction over time for post-lexical /s z/ before /ɹ/ relative to their prevocalic /s z/.

approach was applied (first presented in Magloughlin 2018). The addition of the forced aligner judgments is a source of converging evidence that one or more aspects of the signal can be more interpretable as an affricate than a stop. While the aligner and each of the three human judges may have picked up on different cues in making their judgments, the fact that affrication judgments increased by birth year in both studies strongly suggests /tɹ/ and /dɹ/ clusters have become more affricated over time, whichever cues that may entail.

2.5.1 Automatic classification methods

Automatic classification with forced alignment is a technique that uses likelihood values from forced alignment procedures to assign probability scores to target segments in corpus speech. This method uses the Penn Phonetics Lab Forced Aligner (P2FA) for English (Yuan & Liberman 2008), which maps word- and phone-level transcriptions to audio files, inserting phone boundaries based on transcriptions from a pronouncing dictionary. For each force-aligned segment, P2FA assigns a likelihood score based on information from its acoustic models. For this project (described in more detail in Magloughlin 2018), automatic classification involved extracting all non-function words with word-initial /t d tʃ dʒ/ + vowel or /ɹ/ sequences from the Raleigh corpus and force-aligning twice with P2FA (16 kHz acoustic models) – once using a version of the pronouncing dictionary that contained only stops for the target variants, and once using a *modified* version of the pronouncing dictionary that contained only affricates. It was assumed that affricated versions of /t/ and /d/ before /ɹ/ would closely resemble [tʃ] and [dʒ]. This technique, adapted from Yuan & Liberman (2009), was used in order to calculate an A(ffrication)-Score for every /t d tʃ dʒ/ token before a vowel or an /ɹ/. As shown in Equation 2, this was achieved by subtracting the probability (log likelihood score) associated with each segment's classification as a stop (T_1) from the probability of its classification as an affricate (T_2).

(2)
$$A(t) = \log p(t \mid T_2) - \log p(t \mid T_1)$$

In order to exclude outliers resulting from possible segmentation errors and sources, only A-Scores with values between –100 and +100 were included in the analysis (96% of tokens). Positive values indicate greater affrication. An additional step calculated A-Ratios, normalizing these scores by talker, relative to their prevocalic /t d/ and /tʃ dʒ/ productions (parallel to the COG ratio measure used for /s/ retraction in the preceding sections, following Baker et al. 2011 and Wilbanks 2017). A-Ratios show how close a particular talker's /t/ and /d/ before /ɹ/ are to their prevocalic /tʃ/ and /dʒ/. As illustrated in Equation 3, A-Ratios for /tɹ/ were calculated by subtracting their mean A-Score for prevocalic /t/ from their A-Score for each /tɹ/ token, and dividing that value by the difference between their mean A-Scores for prevocalic /tʃ/ and /t/.

(3)
$$\text{A-Ratio} = \frac{\text{observed A-Score} - \text{talker mean T A-Score}}{\text{talker mean CH A-Score} - \text{talker mean T A-Score}}$$

This method was also used to calculate ratios for the voiced tokens. A ratio closer to 0 indicates that the token is more stop-like, while a ratio closer to 1 is more affricate-like. A linear mixed effects model was applied to the A-Ratios for /tɹ/ and /dɹ/, with centered year of birth and gender as fixed effects and talker and word as random effects: `lmer(score~birthyear*gender+(1|talker)+(1|word))`. Degrees of freedom were obtained using a Satterthwaite approximation in order to provide approximate *p*-values. Full model outputs are included in the supplementary files for reference.

2.5.2 Automatic classification results

Figures 10 and 11 show /tɹ/ and /dɹ/ affrication, both of which are increasing in apparent time for men and women. Notably, men show a later increase in both /tɹ/ and /dɹ/ affrication, making the jump to affrication almost entirely within the last generation. Figure 10 shows the raw A-scores for prevocalic /t d/, /tɹ dɹ/, and /tʃ dʒ/. We can see the affrication of /tɹ/ and /dɹ/ increasing in apparent time, but the other consonants also show some variability, which we should take into account.

A-Ratios (shown in Figure 11) take changes in /t d/ and /tʃ dʒ/ into account, expressing the affrication of the stop-/ɹ/ as a point along a scale from that talker's plain stop to a phonemic affricate. We see confirmation that both /tɹ/ and /dɹ/ are increasing in affrication over apparent time, relative to the other sounds (/tɹ/: $t(123) = 4.4, p < 0.001$; /dɹ/: $t(104) = 6.0, p < 0.001$). For both /tɹ/ and /dɹ/, women display affrication earlier than men, as reflected in the lower male intercept (/tɹ/: $t(104) = 3.9, p < 0.005$; /dɹ/: $t(102) = 2.7, p < 0.01$), and a less steep slope for /dɹ/ ($t(102) = 2.3, p < 0.05$).

In summary, using automatic classification to determine probability of affrication confirms the auditory coding judgments reported above, in that both /tɹ/ and /dɹ/ are increasingly affricated over time, but additionally removes the potential for listener bias and socio-cognitive filtering applicable to age-grading and gender effects of this sound change. This method illustrates that women are slightly ahead in affrication for both /tɹ/ and /dɹ/, which was not the pattern reflected in auditory coding.

2.6 Corpus study summary

The study of consonant variation in the context of /ɹ/ in Raleigh revealed that /tɹ/, /dɹ/, and /stɹ/ have been undergoing change since the middle of the 20th century. This suggests that younger speakers are producing a phonologically-determined different target for the /t d s/ in these sequences, and not simply a coarticulated /t d s/. A phonologically
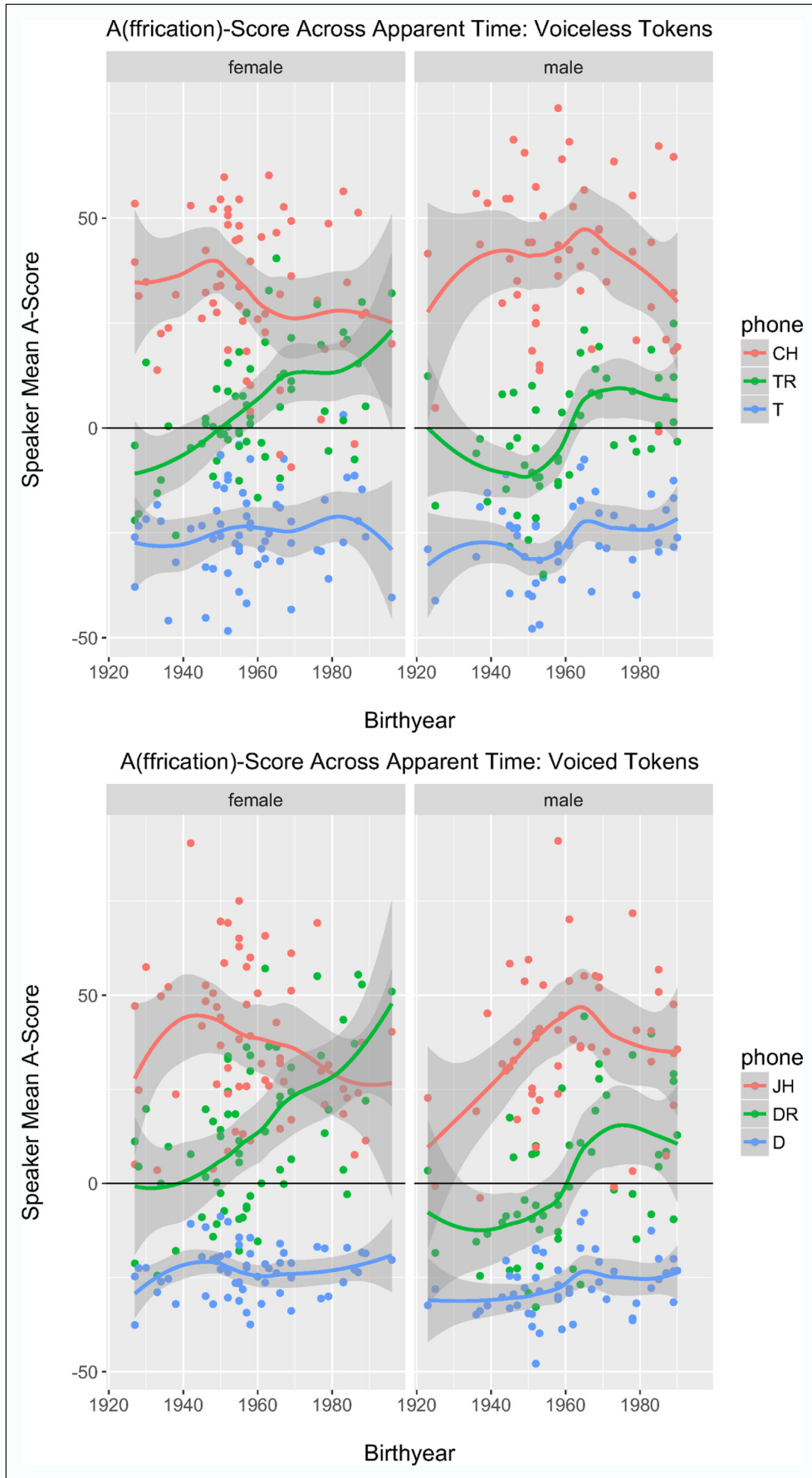
Figure 10: Mean A-Score of voiceless phones (top) and voiced phones (bottom) by sex and birth year (from Magloughlin 2018).
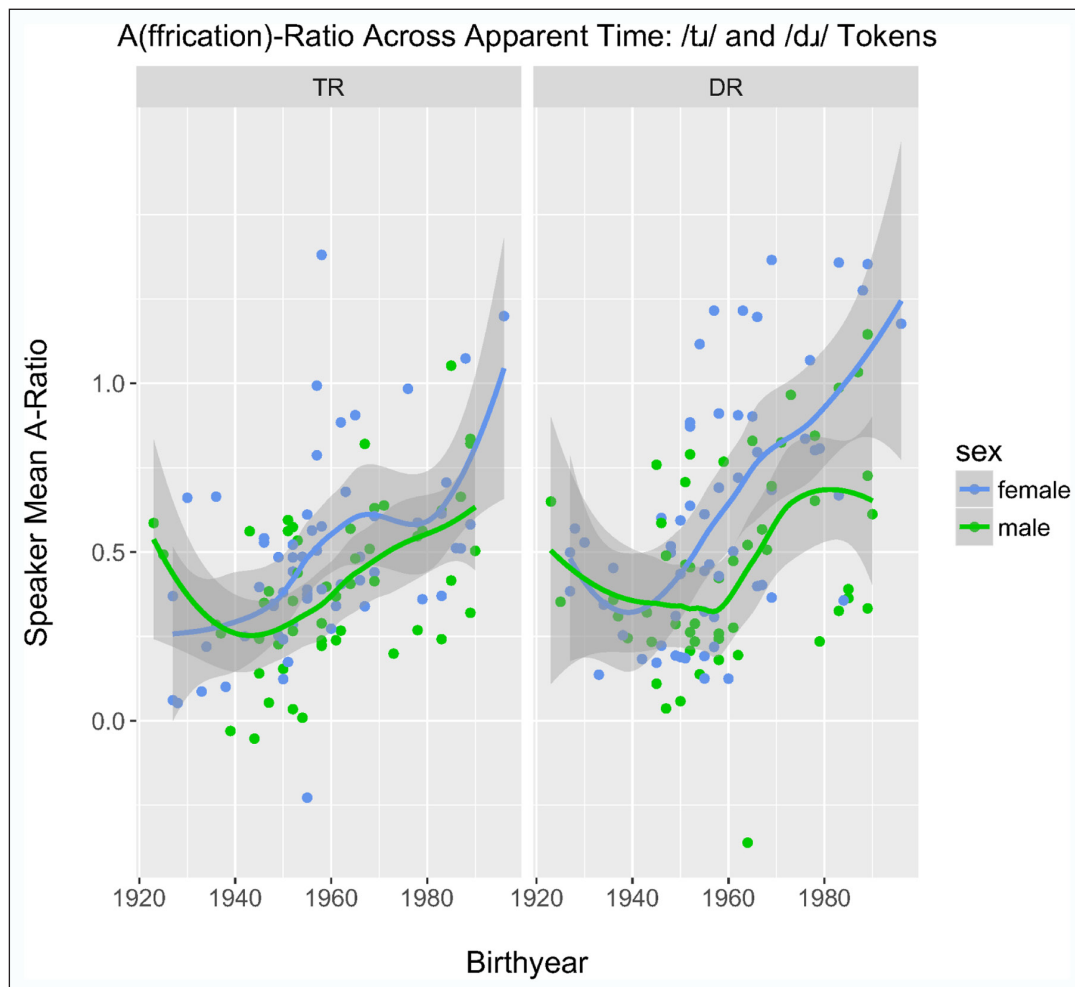
Figure 11: Mean A-Ratios for /tɹ/ (left) and /dɹ/ (right) by sex and birth year (from Magloughlin 2018).

affricated or retracted consonant is not expected to be sensitive to covert differences in /ɹ/ articulation in the same way as an unaffricated or unretracted target. In order to observe the role of covert articulatory variation in these patterns in Raleigh, we would need to study speakers who were born in the early-to-mid 20th century. The other sequences, all involving /ɹ/-adjacent sibilants, are not changing in apparent time, so it is more likely that covert articulatory differences could be observed in the realization of those sequences for a wide range of speakers, including the younger lab study participants described in the next section.

3 Laboratory study

The corpus study showed that /tɹ/ and /dɹ/ affrication and /stɹ/ retraction have been increasing in Raleigh, NC, particularly later in the 20th century, and that /s z/ retraction in other contexts decreased. No increase in /s z/ retraction was observed in contexts where they are directly adjacent to /ɹ/, either word-internally or across a word boundary. It is conceivable that the increases in affrication and retraction could be due to increasing coarticulation, or due to a shift toward new affricated and retracted targets, not directly due to coarticulation. The purpose of the laboratory study is to explore the articulation of stops and sibilants in the context of /ɹ/ for evidence that the patterns involving them are phonologically assimilatory (expected for /tɹ/ and /dɹ/ affrication and /stɹ/ retraction based on the corpus study) or coarticulatory (expected for the other variables), and

to examine the relationship, if any, between /ɹ/ tongue shape and retraction/affrication. Tongue shape is expected to play a basic role in the nature of coarticulatory effects but not learned phonological retraction or affrication patterns (such as where /t d s/ are realized as [tʃ dʒ ʃ] in the context of /ɹ/). But even if /tɹ/ and /dɹ/ affrication and /stɹ/ retraction are learned phonological patterns, we can expect to see evidence of coarticulation in the other variables (/s z/ before or after /ɹ/). To investigate these questions, we collected acoustic and articulatory data in the laboratory from speakers that overlap demographically with the Raleigh corpus speakers, in order to compare corresponding articulatory and acoustic data.[10]

3.1 Methods

3.1.1 Participants

29 native speakers of American English, recruited from a Raleigh, NC university community, participated in the production study, including 18 females and 11 males, ranging in age from 18 to 59 at the time of recording (median = 21). Participants reported no speech or hearing disorders, though subject 05 indicated "I had to learn how to say my Rs properly in elementary school" and subject 24 reported "a slight lisp sometimes". Neither were considered to be problematic for the purposes of analysis: subject 05 produced adult-like /ɹ/ across contexts, and we did not detect a lisp in subject 24's laboratory recorded speech. Participant details can be found in the supplementary files.

3.1.2 Stimuli

Stimuli consisted of familiar monosyllabic English words (and phonotactically probable nonwords) beginning with /ɹ tɹ t tʃ dɹ d dʒ/ as well as polysyllabic initial and medial /s ʃ stɹ/, post-lexical /s z/#/ɹ/, /ɹ/#/s z/, and word-internal /ɹs/ and /ɹz/ sequences. Other stimuli consisted of words and non-words that were relevant to other variables being studied in the same lab at the time of recording (a complete list of stimuli is included in the supplementary files). We should note that the post-lexical stimuli consisted of two not entirely related[11] content words (e.g., *kiss rock*, *bore zip*), so that many participants gave both words similar stress and some participants had a clear prosodic break between the two words, which greatly reduced coarticulation and interfered with our ability to discern differences between articulatory strategies based on relevant tongue postures. Tokens that exhibited a pause between the two words were removed from analysis, but the remaining variability in prosodic fluency is a source of additional noise that should be taken into account.

3.1.3 Procedure

The production study was conducted in a sound-attenuated booth. The reading task, which included experimental phrases in the form "give me a X again", typically took between 20 and 30 minutes to complete, following an initial set-up of approximately 10 minutes. Ultrasound, video, and acoustic data were collected simultaneously using a Terason t3000 ultrasound machine, running Ultraspeech 1.3 (Hueber et al. 2008) in direct-to-disk mode. A microconvex array transducer (8MC3 3–8MHz, 90-degree field of view) was used to image the mid-sagittal plane of the tongue. Ultrasound images were captured in 640 × 480 pixel bitmaps, generated at a rate of 60 fps. Video images were also captured in 640 × 480

---

[10] While the language backgrounds of the May corpus do not align exactly with the Raleigh corpus, it should be noted that the talkers are still a fairly homogeneous group, and thus we cannot extrapolate from our results to all varieties of English.

[11] While most word pairs were entirely unrelated, such as *bore* and *zip*, some word pairs may have had some tangential semantic relation, such as *kiss*, as in the rock band *KISS*, and *rock*, as in *rock and roll music*, which may not be salient for all participants.

pixel bitmaps at 60 fps, using an Imaging Source DFK 21BU04 1/4" closed-circuit TV camera recording in grayscale mode. Audio was recorded at a sampling rate of 44,100 Hz and a bit depth of 16 using an Audio-Technica AT803 lavalier microphone attached approximately one inch from the participant's mouth with an AT8418 instrument mounting clip. Audio was transmitted through a SoundDevices USBPre 2.0 preamplifier. Ultraspeech 1.3 (Hueber et al. 2008) was used to simultaneously capture the acoustic, video, and ultrasound data. Automatic file-naming during recording provided synchronous time stamp information, which resulted in the straightforward alignment of all three streams of data (acoustic, video, and ultrasound).

Participants were instructed to lean forward against a headrest. An Arduino-based device provided feedback to participants about their head orientation, based on a 9 degrees of freedom inertial measurement unit attached to a headband they wore. If their head orientation deviated by more than one degree from their initial orientation, they were not allowed to proceed through the word list until the deviation was corrected. They received a visual representation of head orientation to guide them back to their original position. Prior to recording, they were given a chance to familiarize themselves with this system. The occlusal plane was imaged using a tongue depressor, and the palate was imaged using a mouthful of water. Only the occlusal plane images were used in the present analysis, in order to orient all the images the same way.

3.1.4 Acoustic and perceptual analysis methods

For the corpus speech described above, we used relatively coarse measures: listener judgments, center of gravity and automatic classification, some of which we also applied to the laboratory speech. However, because we could directly tie articulatory events to acoustic data, we wanted to find more specific acoustic correlates of retraction and affrication (in addition to the perceptual classification described for the corpus and lab speech in the previous section). Acoustic analysis of /ɹ/ assimilation/coarticulation effects involved multitaper spectra of sibilant intervals, in order to identify spectral peaks which vary according to tongue and lip postures (see e.g., Koenig et al. 2013 for previous analyses of sibilants using multitaper spectra). For the analysis of spectral peaks, we used multitaper spectra, which were obtained using the spectRum package for R (Reidy 2013). Traditionally, spectra are estimated using a Discrete Fourier Transform (DFT), but multitaper spectra (MTS) exhibit reduced variance while maintaining greater temporal precision. This reduction in variability has very little effect on measuring spectral moments, such as COG; however, the location and amplitude of spectral peaks may be somewhat more accurately estimated with reduced variance (Reidy 2015).

3.1.5 Articulatory analysis methods

Articulatory data analysis had qualitative and quantitative components. We visually labeled each /ɹ/ tongue image as either retroflex or bunched, in order to classify talkers into those who produce retroflex /ɹ/ at least some of the time (subjects 06, 07, 11, 15, 16, 23, 27, and 30) and those who produce only bunched /ɹ/ (the other 21 participants).

In order to facilitate quantitative analysis of tongue position changes over time, we performed Eigentongues decomposition (i.e., principal component analysis (PCA) of pixel intensities in filtered and downsampled ultrasound images; Hueber et al. 2007). This analysis was performed as implemented by Carignan (2014) and described in Mielke, Carignan & Thomas (2017: 336). A similar PCA technique was used to analyze the video images, which were not filtered. For both ultrasound and video images, the first 50 principal components were retained for analysis. Previous work with similar data has shown that this includes a sufficient portion of the variance in the images (Mielke et al. 2017: 337).

The result of each PCA is a 50-dimensional vector representing each ultrasound and video frame. Following Hoole & Pouplier (2017) and Strycharczuk & Scobbie (2017), we performed Linear Discriminant Analysis (LDA) of the PCA output in order to quantify the articulatory similarity of coarticulated or assimilated consonants to word-initial /ɹ/, representing the local coarticulation source, and word-initial postalveolar consonants /ʃ tʃ dʒ/, which they may be more likely to resemble in a phonologized assimilation pattern. LDA models of pixel intensities in ultrasound and lip video images were generated for prevocalic /ɹ/ vs. all other phones and for /ʃ tʃ dʒ/ vs. all other phones, and were used to examine the similarity between target sounds and these reference sounds.

3.2 Results: Laboratory study

Here we examine the articulatory basis of the retraction and affrication patterns observed in the corpus study. In particular, we are interested in whether the various affected consonants are articulated more like the /ɹ/ that appears to have conditioned the change (at least historically) or like the /ʃ/, /ʒ/, /tʃ/, or /dʒ/ that they resemble at least superficially. This addresses the question of whether the articulatory gestures suggest phonological (or categorical) assimilation versus phonetic (or partially phonological) coarticulation. This comparison makes use of the fact that the tongue and lip positions used to produce /ɹ/ are similar but not identical to the way these postalveolar consonants are produced. LDAs, in this way, function as a quantitative measure of the similarities that might be found in tongue tracings, and they can capture differences that might be missed by, for example, simply hand-coding each token as tip-up or tip-down. To be fair, they can also miss important congruities that might be caught by a human eye, such as the substitution of an apical retroflex postalveolar fricative for a laminal postalveolar fricative, which may sound every bit as retracted, but appear to the LDA as more /ɹ/-like than /ʃ/-like. Thus, for our purposes, we will use the terminology "retracted" to refer to articulations that are more /ʃ/-like, with the understanding that /ɹ/-like articulations may also be retracted, but are not phonologically /ʃ/-like. This is a more conservative approach that may miss some articulatory retraction in terms of place of articulation being physically more retracted, or an entirely different gestural approach to retraction. Additionally, we must point out that, unlike palatographic studies, this method cannot quantify the degree to which the constriction is *physically retracted*, that is, located further back in the vocal tract. For the patterns that appear to be coarticulatory (resembling the local /ɹ/), we are interested in the contribution of tongue and lip-shape variability to different coarticulatory patterns. Since we have no expectation that post-lexical coarticulation will resemble phonological postalveolars, we will also look at the similarity between the /s z/ and /ɹ/ tongue and lip position, and their relative contribution to the observed acoustic features.

3.2.1 Articulatory results

Figure 12 shows articulatory trajectories for the /stɹ/ data. The *x*-axis is normalized time (according to acoustic duration), where the interval [0, 1] corresponds to the target segment (here /s/), the interval [–1, 0] is the preceding segment (here the vowel in the carrier phrase word *a*), and the interval [1, 2] is the following segment (here /t/). The time before –1 is scaled so that an interval of 1 matches the duration of the segment preceding or following the target. For this summary, a cubic smoothing spline was fit to each talker's data for each sequence of interest, using the default parameters of R's `smooth.spline` function. The top-left panel shows the LDA signal that is based on ultrasound frames segmented as /ɹ/ vs. ultrasound frames for all other speech segments, representing lingual similarity to /ɹ/. This shows that the /ɹ/ gesture peaks at the start of the /ɹ/ interval, and that it minimally affects the segmented /s/ interval. The bottom-left panel shows a similar
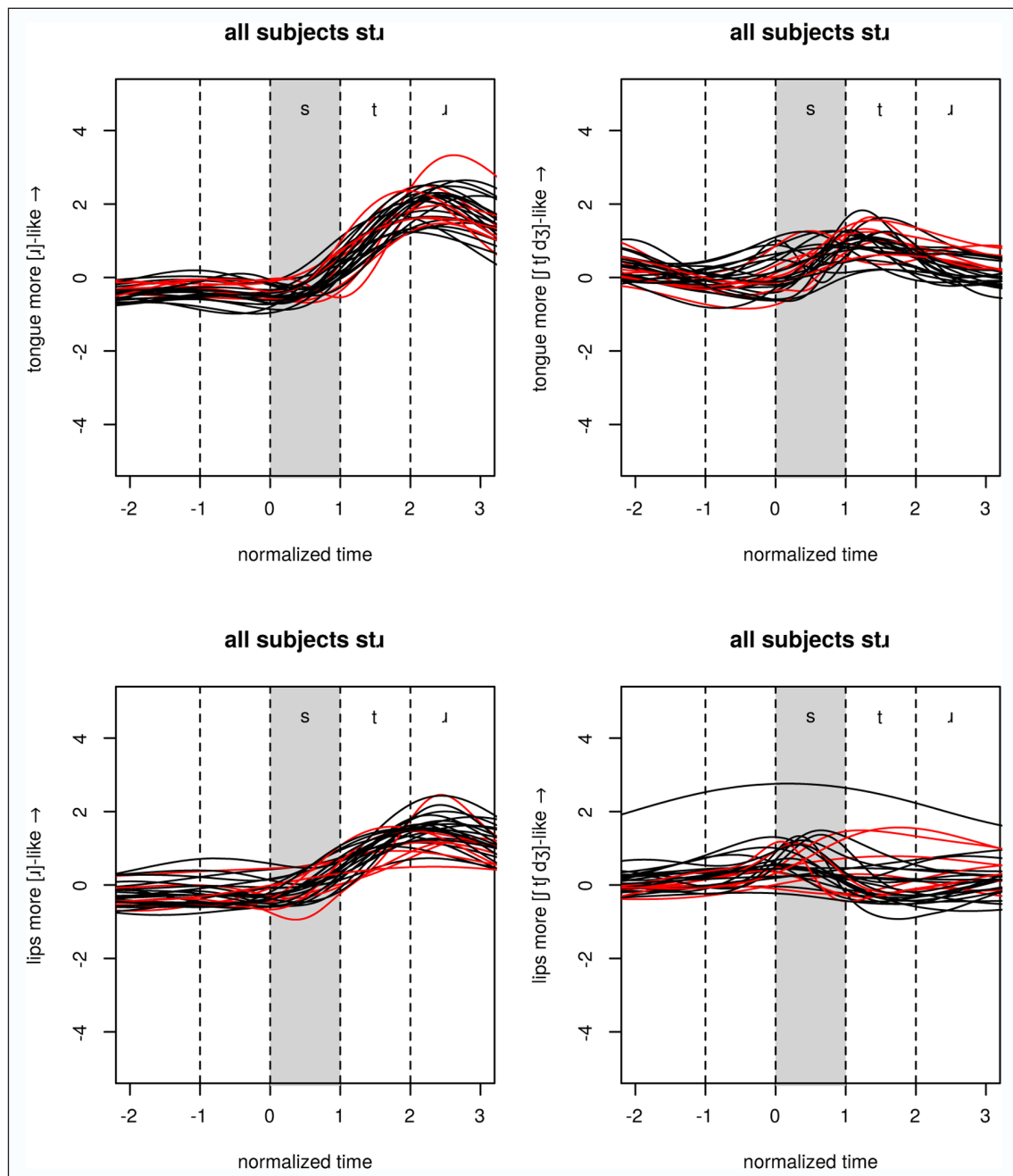
Figure 12: LDA signals (smoothed splines) for /stɹ/, showing timecourse of similarity to [ɹ] (left) and [ʃ tʃ dʒ] (right) for the tongue (top) and lips (bottom). The shaded interval from 0 to 1 is the /s/ segment interval, 1–2 is /t/, and 2–3 is /ɹ/. Data for talkers with retroflex /ɹ/ are shown in red.

pattern for lip video. The two panels on the right show the similarity of the tongue and lip images to postalveolar consonants /ʃ tʃ dʒ/. Several talkers show peaks during the /s/ interval, indicating that they are articulating the /s/ in a way that resembles a postalveolar consonant but does not particularly resemble an /ɹ/.

Figure 13 illustrates the differences between lip postures for /ʃ/, /ɹ/, and /tɹ/ for five talkers, using images created by averaging the downsampled lip video images over all the intervals for each phone. In general the lip constriction for /ʃ/ is more out-rounded and less compressed than /ɹ/, although there is variation across talkers, as suggested in the LDA. /tɹ/ is somewhat intermediate between /ɹ/ and /tʃ/, but more generally resembles /tʃ/ in phonologically affricating talkers.
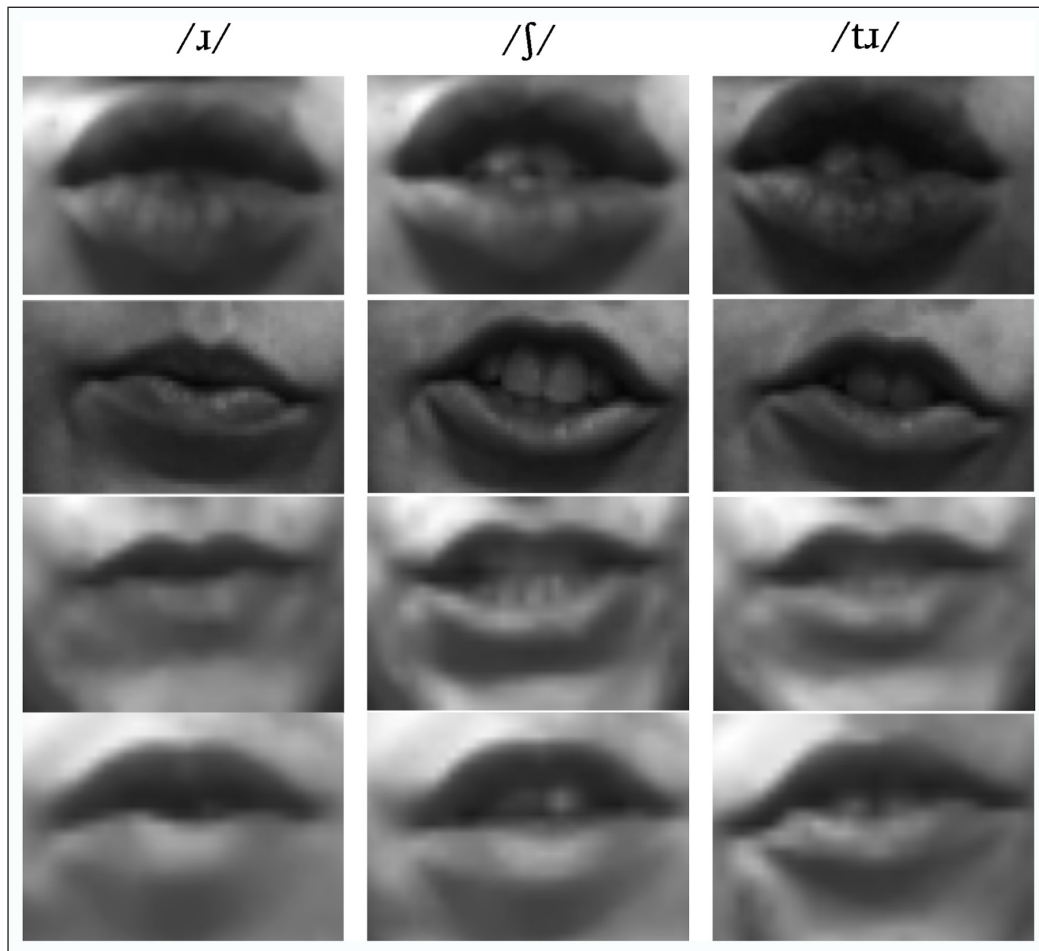
Figure 13: Examples of lip rounding evident in word-initial /ɹ/, /ʃ/, and /tɹ/. Talkers (from top to bottom) are 09, 03, 21, and 11.

The next set of figures provide summaries of the articulatory resemblance to /ɹ/ and postalveolar consonants for each of the variables in question, based on the middle 50% of each segment interval of interest. Figure 14 summarizes the lab data for /stɹ/. The /s/ interval for each participant is represented by a filled circle, where the size of the circle indicates the participant's rate of /stɹ/ retraction, according to the perceptual coding. The position of the circle indicates the similarity of the images in the middle 50% of the /s/ interval to /ɹ/ (*x*-axis) and to postalveolar consonants (*y*-axis); higher values mean greater similarity to the sound(s) listed on that axis. The left panel is based on lingual ultrasound images and the right panel is based on lip video images. Finally, the color of the circle indicates whether the talker produced any retroflex /ɹ/ (red for retroflex). Solid lines connect the circle to the point representing the following /t/, and dashed lines lead to the point representing the following /ɹ/.

The fact that all lines lead to the right means that the /t/ and /ɹ/ are more [ɹ]-like than the /s/ in the /stɹ/ clusters. In general, only the lips ever start out in positive /ɹ/ territory (meaning some /s/ tokens are produced with [ɹ]-like lip rounding), but many speakers have /s/ with positive values on the *y*-axis, meaning that they are similar to those same speakers' postalveolar consonants. The most retracted-sounding talkers (>50% retracted judgments, represented by larger circles) have a tongue shape that resembles postalveolars (positive signal value for postalveolars on the *y*-axis), while those with fewer retracted judgments (<50% retracted judgments, represented by smaller circles) had signal values
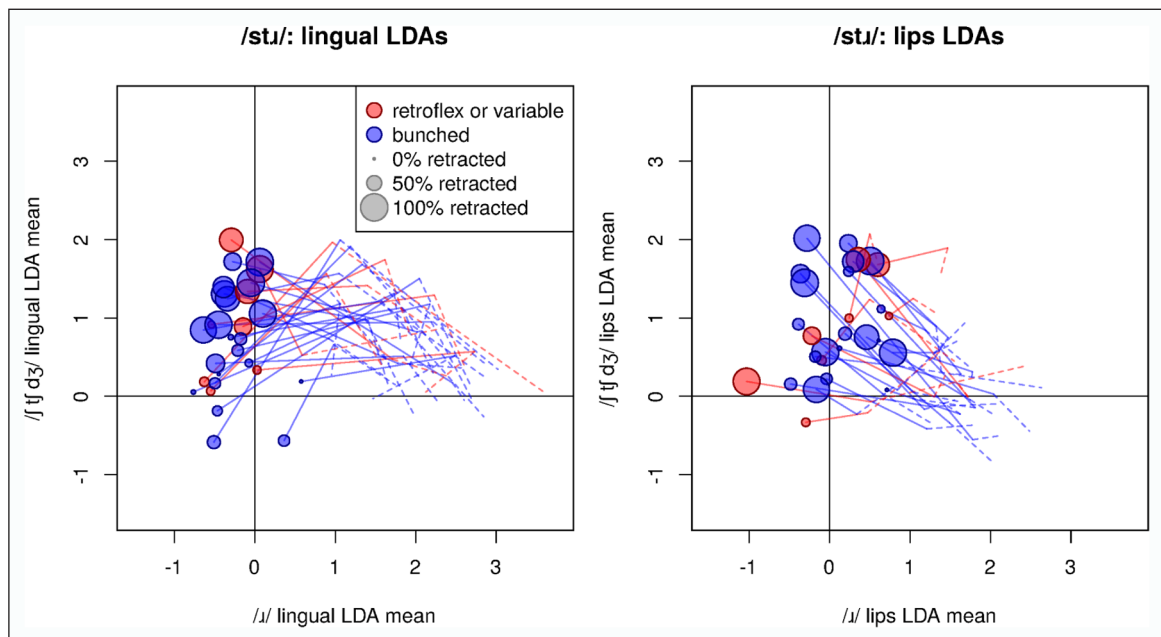
Figure 14: Similarity of /stɹ/ to /ɹ/ and postalveolar consonants. The circles are centered over the average LDA value of the /s/ interval, while the solid lines point to the average LDA value for the /t/ interval, and the dashed lines point to the average LDA of the /ɹ/ interval. Higher values mean greater similarity to the sound(s) listed on that axis. Retroflexers are in red and bunchers are in blue. The larger the circle, the more retracted judgments were given for that talker, while smaller circles represent less-retracted judgments.

near zero. The split point for more retracted versus less retracted talkers seems to occur near a signal value of 1. While this may not seem very high, coarticulation with /t/ and /ɹ/ will certainly alter the shape of the postalveolar so that it isn't identical to prevocalic postalveolars, which are also influenced by the tongue posture of the following vowel. Additionally, not all retractors will necessarily have a phonological [ʃ] as their intended target. Any postalveolar sound could be employed, since these are not phonemically distinct in English. While there is some similarity with the /ɹ/ shape (positive signal values on the *x*-axis), the retracted productions are not any more similar to /ɹ/ than the less retracted productions. That is, degree of /s/ retraction does not appear to be related to the degree of similarity with /ɹ/, but rather to how /ʃ/-like it is.

Even more interesting is the pattern for the /t/ interval in /stɹ/ clusters, which increases in its resemblance to postalveolars in all but seven of the most retracted talkers, for whom the *y*-axis signal value is already high during the /s/ interval. There is no indication of any difference between bunching and retroflexing talkers. This suggests that the /t/ in /stɹ/ clusters has become phonologically retracted for most talkers, and is not determined by the following /ɹ/ tongue shape. Thus, /s/ retraction, for some talkers, may be based on assimilatory coarticulation to the following /t/ (as suggested by Lawrence 2000). The /ɹ/ exerts an influence on the lingual gestures of the /t/ in /stɹ/ clusters such that the /t/ becomes both more /ɹ/-like and more /ʃ/-like. The /ʃ/-like component is anticipated during the production of the /s/. The LDAs do not tell us which regions of the tongue are more /ʃ/-like, whether it is the tip, blade, root, or all three, so we might imagine some variability across talkers with different gestures. The degree of /ʃ/-similarity of the /t/ does not seem to affect how retracted each talker's /s/ production is judged to sound, with some of the least-retracted sounding talkers having some of the highest postalveolar signals for the /t/ interval. So, if the /t/ is affecting the preceding /s/, it is only true for

some talkers. That is to say, retraction of the /t/ in /stɹ/ does not necessarily require that the /s/ will become retracted, though it may be a precondition for some.[12]

The peak of the lip-rounding gesture for /ɹ/ is temporally aligned with the onset of the /ɹ/, meaning that rounding occurs during a good portion of the /t/ interval, which is reflected in the signal value increase for /ɹ/ during the middle 50% of the /t/. If the /t/ interval has its own rounding specification, we might expect that to coincide with its onset as well, meaning the /s/ will contain some anticipatory rounding which will vary depending on the rounding specification for /t/. And here we see a mixture of patterns. A handful of talkers begin preparing for /ɹ/ rounding during the /s/, while other talkers have a different target for the onset of /t/, some of which are more similar to postalveolar lip-rounding. Of those who have the most postalveolar target for rounding during the /s/ interval, most have more retracted sounding productions. There are some other talkers whose rounding gesture doesn't resemble either /ʃ/ or /ɹ/, so we might guess that they have a different rounding specification for /t/ that does not resemble either, or that their anticipatory rounding does not extend as far.

Figure 15 summarizes the data on /tɹ/. This and the next three figures employ the same method as the /stɹ/ data in Figure 14 except that there are only two segments to show: the target segment and the adjacent /ɹ/. Circles for talkers whose lip images are shown in Figure 1 or whose tongue contours are shown below in Figure 17 are labeled. Nearly all of the participants were rated as having affricated /tɹ/ most of the time, and the left panel shows that nearly all of them produce the /t/ interval with a tongue posture that is similar to a postalveolar consonant. A few participants produce /t/ with a tongue posture
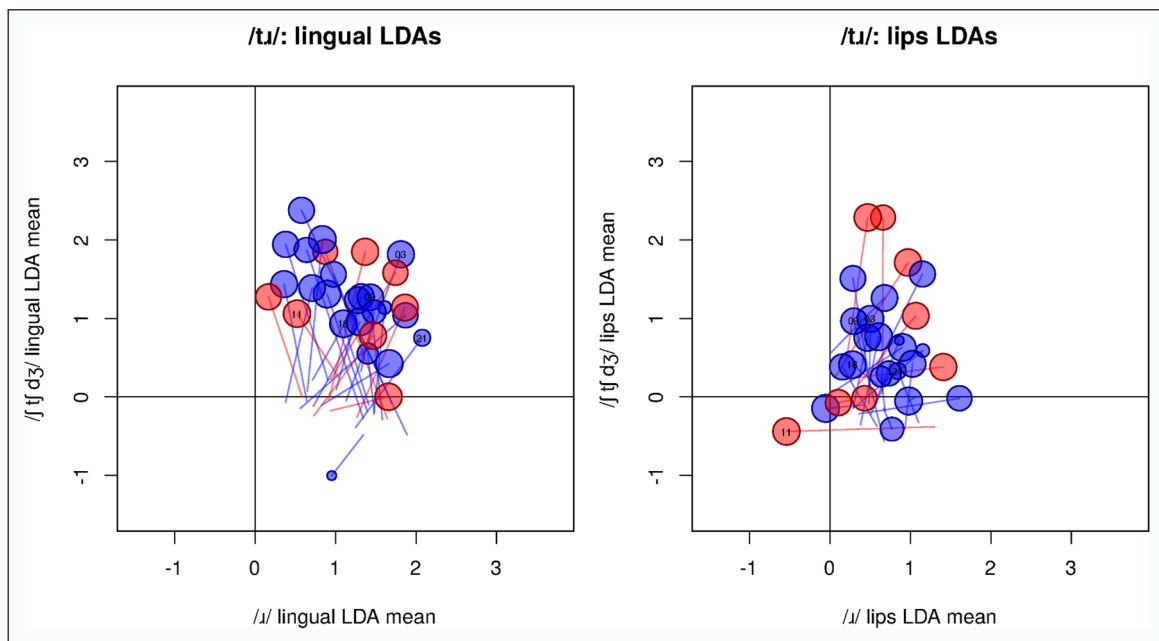

Figure 15: Similarity of /tɹ/ to /ɹ/ and postalveolar consonants. The circles are centered over the average LDA value of the middle 50% of the /t/ interval, while the lines point to the average LDA value for the middle 50% of the /ɹ/ interval. Higher values mean greater similarity to the sound(s) listed on that axis. Retroflexers are in red and bunchers are in blue. The larger the circle, the more affricated judgments were given for that talker, while smaller circles represent less-affricated judgments.

---

[12] We observed some productions in which the /t/ in /stɹ/ clusters was clearly affricated [tʃɹ], others where it appeared as a short-lag stop [dɹ], and others which were some kind of intermediate variant [dʒɹ]. We did not quantify this dimension of variability, but thought it might bear mentioning.

that is more like /ɹ/. There are no apparent differences between more- and less-affricated talkers, with the exception of one talker who was judged to have very little affrication, whose tongue shape is neither postalveolar nor /ɹ/-like. The other two talkers with low affrication scores show some degree of postalveolar tongue shape in line with some of the more affricated talkers, and a greater degree of similarity with /ɹ/. There are also no differences between bunchers and retroflexers that lend themselves to a simple explanation based on tongue shape. In general, the movement between the /t/ interval and the /ɹ/ interval (indicated by a solid line) is downward, meaning that the postalveolar consonant-like tongue shape is a property of the /t/ itself, and not the /ɹ/. The lip data for /tɹ/ (in the right panel) shows many talkers with postalveolar-like lip postures during the /t/ interval. The fact that nearly all of the circles in both panels of Figure 15 have positive values for the /ɹ/ LDA is consistent with these sounds being directly adjacent to and coarticulated with an /ɹ/. The fact that most have larger values for the /ʃ tʃ dʒ/ LDA than for /ɹ/ indicates that these stops have a postalveolar-like articulatory target.

Figure 16 summarizes the articulation of /dɹ/. Compared to /tɹ/, we observe that the /ʃ tʃ dʒ/ LDA signal values are generally higher, and the /ɹ/ LDA signal values are generally lower, which is consistent with a segmentation difference between /tɹ/ and /dɹ/: the /t/ interval includes the aspiration interval, when the tongue is free to assume an /ɹ/ shape, whereas the /d/ interval is dominated by the closure, and the middle 50% should be almost entirely the closure. As a result, it more strongly reflects the postalveolar tongue posture than the /tɹ/ intervals. Again, there are no apparent differences between more- and less-affricated talkers, with the exception of the one talker who was judged to have very little affrication, whose tongue shape is neither postalveolar nor /ɹ/-like. The other three non-affricators (who also happen to be bunchers) have postalveolar-like tongue postures, possibly suggesting that phonological retraction may have preceded affrication.
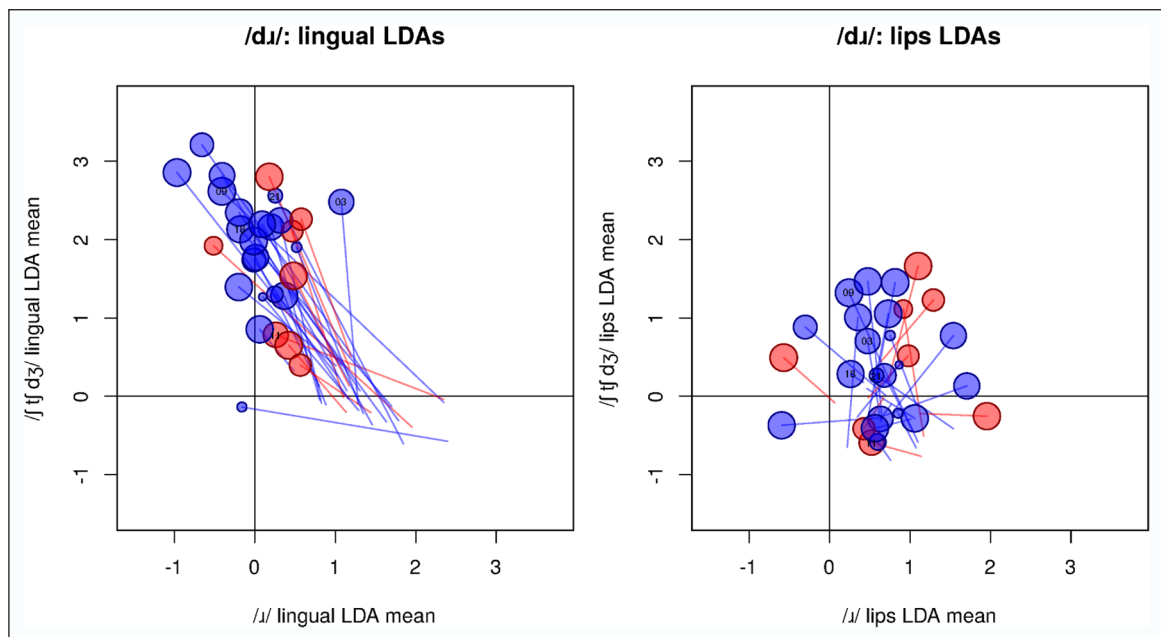

Figure 16: Similarity of /dɹ/ to /ɹ/ and postalveolar consonants. The circles are centered over the average of the middle 50% of the /d/ interval, while the lines move to the average LDA value for the /ɹ/ interval. Higher values mean greater similarity to the sound(s) listed on that axis. Retroflexers are in red and bunchers are in blue. The larger the circle, the more affricated judgments were given for that talker, while smaller circles represent less-affricated judgments.

To illustrate the articulatory similarity of alveolar stops before /ɹ/ to postalveolar affricates, which is shown abstractly by the linear discriminant analysis, Figure 17 shows midpoint tongue contour tracings from a typical participant from Magloughlin's (2018) study of a subset of 12 participants from the May corpus. The smoothing spline ANOVA comparisons show mean tongue contours for prevocalic /t d/, the clusters /tɹ dɹ/, and prevocalic phonological affricates /tʃ dʒ/ in thick lines; the thinner dotted lines represent the confidence intervals for these means. The tongue tip is to the right and the tongue root is in the bottom left. A pair of mean curves are significantly different where their confidence intervals do not overlap. Figure 17 shows that /t d/ before /ɹ/ have tongue body positions that are much more similar to /tʃ dʒ/ than to prevocalic /t d/ at both the midpoint of the closure interval (left) and the midpoint of the stochastic noise interval (right), even though /tʃ/ and /dʒ/ involve tongue root positions that are different from /t d/ with or without the influence of /ɹ/. In short, LDA captures the degree of similarity seen in these tracings, though it does not inform us in what specific regions the tongue shapes are similar or different.
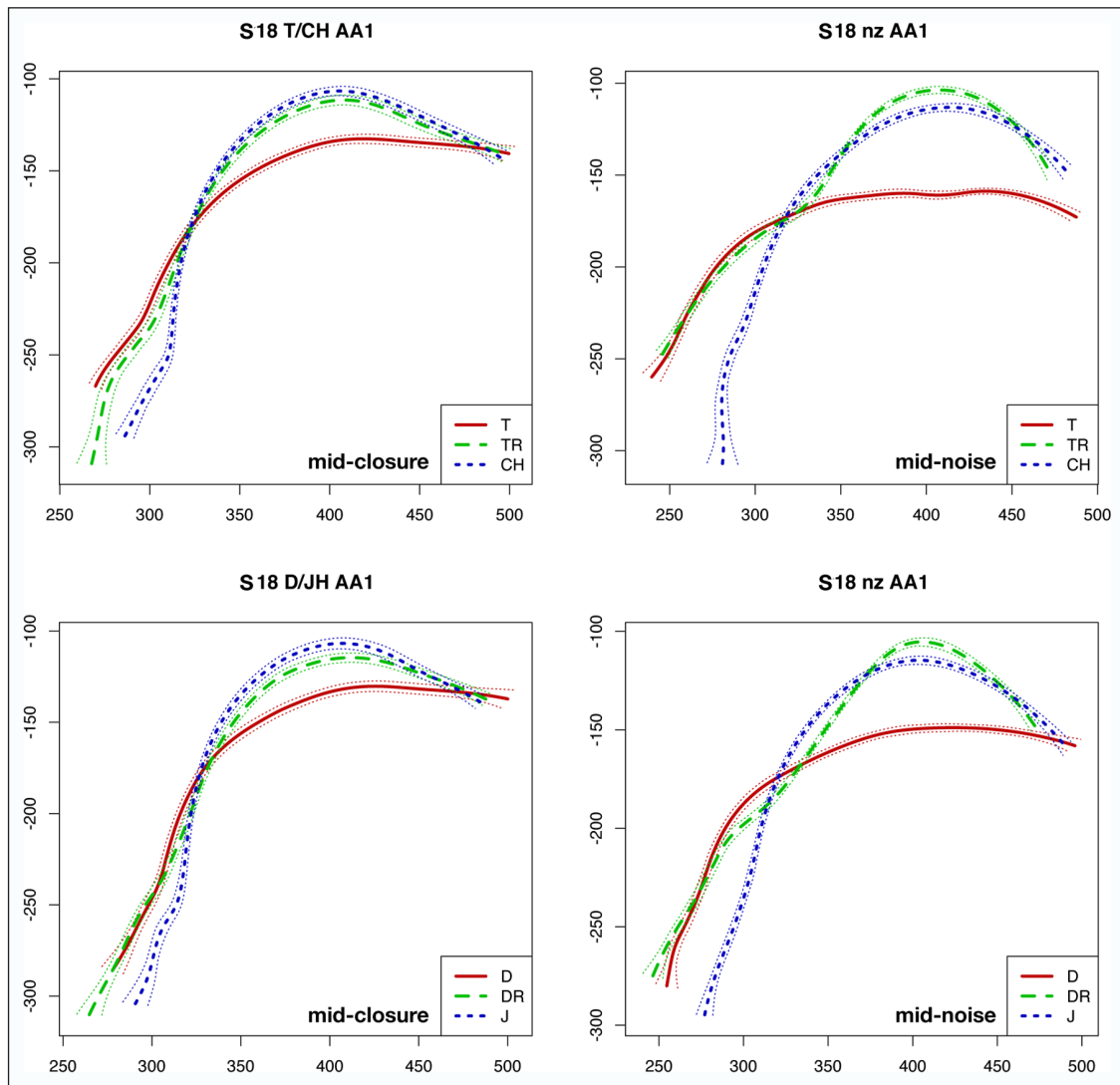

Figure 17: Sample SSANOVA comparisons (from Magloughlin 2018): Subject exhibits typical tongue contour patterns in /a/ vowel contexts at the tongue root and tongue tip/blade in voiceless (upper, /t tɹ tʃ/) and voiced (lower, /d dɹ dʒ/) contours, at the mid-point of the closure interval (left), and the mid-point of the stochastic noise interval (right).

Figures 14–16 have summarized the articulation of the three consonant clusters where affrication or retraction (/ʃ/-, /tʃ/-, or /dʒ/-likeness) seems to be phonological. This is evidenced by the high rate of perceptual classification as affricated or retracted, and the articulatory similarity to a postalveolar consonant. The remaining consonant sequences (/s#ɹ/, /z#ɹ/, /ɹ#s/, /ɹ#z/, and word-internal /ɹs/ and /ɹz/) had very low rates of perceived retraction in the lab speakers and in the youngest generation of speakers in the Raleigh corpus. We expect them to reveal more direct coarticulation between the sibilants and /ɹ/, notwithstanding the differences between within-word and post-lexical articulation.

Figure 18 shows /s/ and /ɹ/ across word boundaries. The top two panels show word-final /s/ before word-initial /ɹ/. In both panels, the circles representing the middle 50%
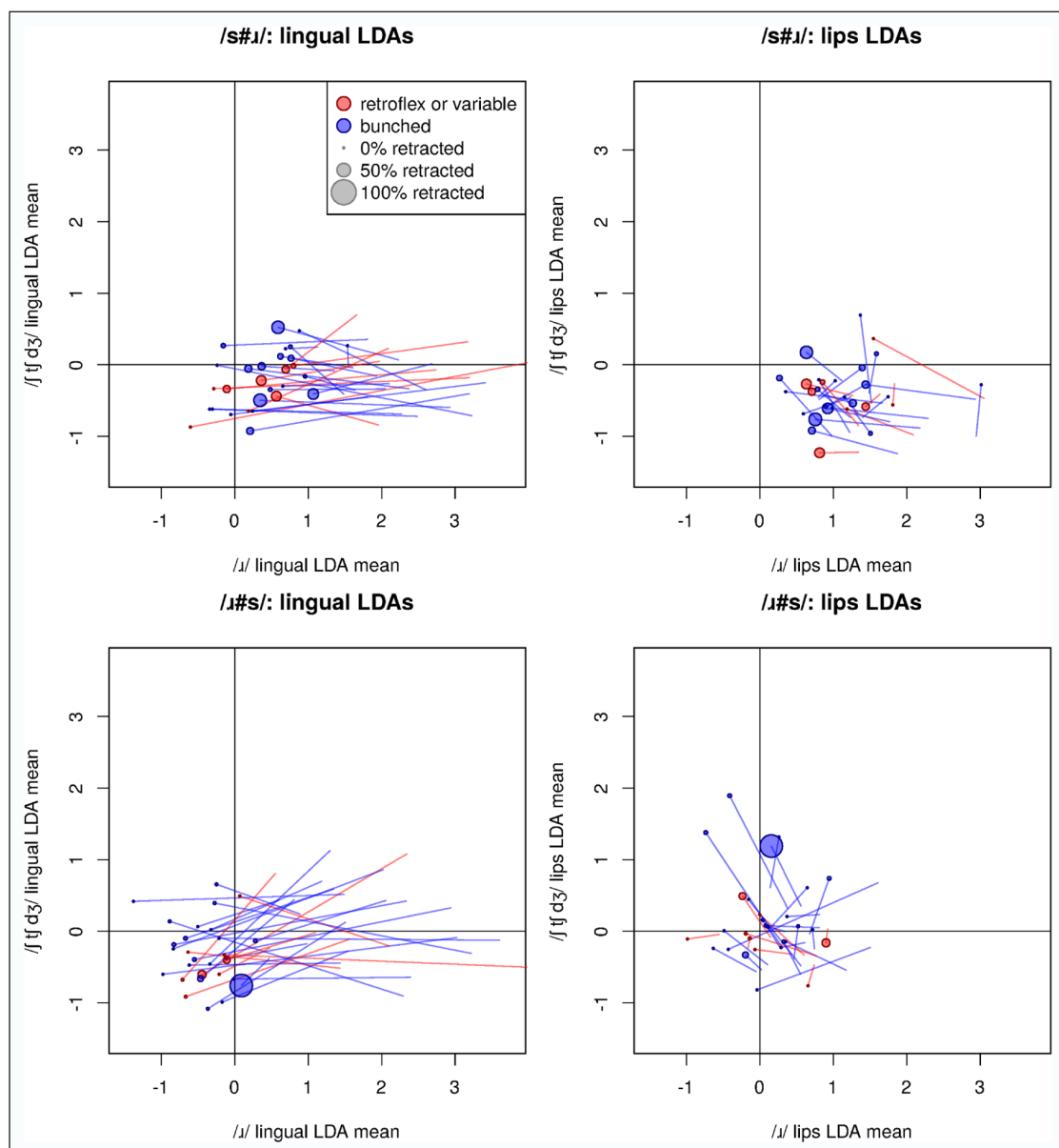

Figure 18: Similarity of /s#ɹ/ (top) and /ɹ#s/ (bottom) to /ɹ/ and postalveolar consonants. The circles are centered over the average LDA value of the middle 50% of the /s/ interval, while the solid lines point to the average LDA value for the middle 50% of the following /ɹ/ interval. Higher values mean greater similarity to the sound(s) listed on that axis. Retroflexers are in red and bunchers are in blue. The larger the circle, the more retracted judgments were given for that talker, while smaller circles represent less-retracted judgments.

of the /s/ interval are clustered just to the right of the origin, meaning they are somewhat similar to /ɹ/ and not similar to postalveolar consonants. In the lingual (left) panel, there is a bigger difference between the /s/ (circles) and the /ɹ/ (line endpoints) than there is in the labial (right) panel. The circles in the lip panel are further to the right, indicating that the /s/ interval has a lip posture that is similar to /ɹ/. None of the articulatory measures appear to be related to perceived retraction.

The bottom two panels in Figure 18 show the mirror-image sequence: word-initial /s/ after word-final /ɹ/. In this figure, the circles represent the second time interval (unlike all of the previous figures, where the /ɹ/ is after the affected consonant). The middle 50% of the word-initial /s/ interval shows no signs of coarticulation to /ɹ/. Importantly, the word-final /ɹ/ is generally not rounded, so the only obvious coarticulation source here is the tongue posture, but it does not show signs of influencing the middle part of the /s/.

Instead of showing figures for the voiced counterparts /z#ɹ/ and /ɹ#z/ and the word-internal /ɹs/ and /ɹz/, none of which show much coarticulation affecting the middle 50% of the sibilant interval, we turn to a summary figure which includes these sequences along with the ones we have seen so far. Figure 19 shows the mean LDA signal values for all of the variables, calculated separately for talkers who always bunched /ɹ/ (blue with solid circles) and talkers who retroflex at least sometimes (red with dashed circles).

The left panel summarizes the lingual LDAs. /stɹ/, /tɹ/, and /dɹ/ all show similarity to postalveolar consonants. /tɹ/ and the /t/ portion of /stɹ/ show the most similarity to /ɹ/, possibly because the release of the stop closure potentially occurs during the middle 50% of the interval segmented as /t/. All of the other variables are clustered on the opposite side of the origin, with the exception of /s#ɹ/ and /z#ɹ/, which show more lingual similarity to /ɹ/ than any of the sibilants that are preceded by /ɹ/. Any differences between bunching and retroflexing talkers are difficult to interpret, and do not seem consistent across categories, possibly due to segmentation differences.

The right panel summarizes the labial LDAs. /stɹ/, /tɹ/, and /dɹ/ again have the most similarity to the lip postures associated with postalveolar consonants. However, /dɹ/ has
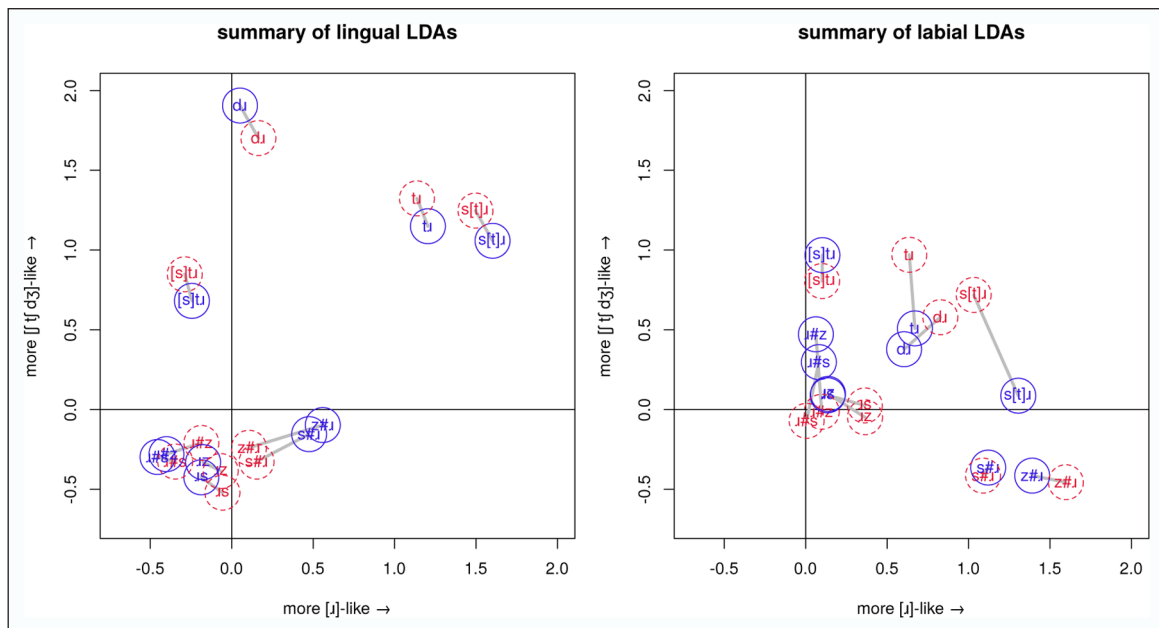

Figure 19: Summary of LDA signal means for all variables: circle center location shows mean LDA signal values for each variable, with talkers separated into groups who always bunched (blue solid lines) or who retroflexed at least sometimes (red dashed lines).

only as much labial similarity to postalveolars as bunching talkers' /ɹ#z/ or /ɹ#s/. Since we have no hypotheses about differences in coda /ɹ/ lip posture among talkers who never retroflex, this is hard to interpret. Similarly, we have no explanation for the apparent difference in lip posture for the /t/ interval of /stɹ/ or /tɹ/ between retroflexing and non-retroflexing talkers.

　/s#ɹ/ and /z#ɹ/ appear to be the best examples of coarticulatory effects in our data. Unlike /stɹ/, /tɹ/, and /dɹ/, the effects of neighboring /ɹ/ appear not to have been phonologized. /s#ɹ/ and /z#ɹ/ have more coarticulation than all the other sibilant-/ɹ/ sequences, comparable to what we see for the /t/ in /stɹ/, and there is no sign of similarity to postalveolar consonants, as we see for the apparently phonologized /stɹ/, /tɹ/, and /dɹ/. Furthermore, /s#ɹ/ and /z#ɹ/ both show a very similar, if small, difference between bunchers and retroflexers in lingual similarity to /ɹ/, something we don't observe for any of the other variables. This is consistent with the idea that /s z/ retraction that is triggered by a following /ɹ/ across a word boundary is the most phonetic of all the instances of retraction and affrication we have examined here. Another factor that makes it easier to observe covert articulatory differences in this context is that there is more covert variation here. The /ɹ/ in these sequences is always word-initial and prevocalic. In this context, retroflex /ɹ/ is frequent before nonhigh back vowels, and bunched /ɹ/ is frequent before high and front vowels (Ong & Stone 1998; Mielke et al. 2016). None of the other sequences allow this much variation in the triggering /ɹ/, because the other /ɹ/ instances are either in coronal consonant clusters or in coda position, both of which favor bunched /ɹ/. Since /s#ɹ/ and /z#ɹ/ are the most coarticulatory of all of the /ɹ/-conditioned patterns here, the next subsection explores /s#ɹ/ and /z#ɹ/ coarticulation in more detail.

3.2.2 Acoustic/articulatory results

Because word-final /s z/ followed by /ɹ/ are subject to coarticulatory processes, and are less strongly affected by phonologization of word-internal sound changes (c.f., Zsiga 1995; Fruehwald 2016), they give us the opportunity to understand what effect /ɹ/ has on preceding segments without having to worry as much about phonologically retracted segments. While we saw that /s/ and /z/ are more retracted before an /ɹ/ than a vowel, we wanted to isolate how much of this was attributable to tongue posture and how much to lip shape, so we extracted low- to mid-frequency spectral peaks to seek corroboration with articulatory measurements.

　Figure 20 shows multitaper spectra at the midpoint of /z/ before /ɹ/, with various tongue and lip configurations. We have observed a prominent mid-to-low-frequency spectral peak in coarticulated /s z/ (either in addition to a higher frequency peak in frication, or resulting in an overall lower peak), and its frequency drops during the course of the sibilant interval. Our data suggest that it is more often due to lip rounding than tongue retraction, as most talkers have some rounding involved, while fewer have co-produced /s z/ and /ɹ/. Higher degrees of rounding result in a narrow bandwidth mid-low-frequency peak for the front cavity, often in the range of F3. We have also observed considerable variation in the lip rounding associated with these /ɹ/ tokens (the shape as well as the timing), but this is yet to be quantified beyond individual talkers' LDAs. That is, there is no satisfactory method of comparing shapes between talkers in our data.

　For post-lexical /s z/ retraction, the linear discriminant analysis shows much earlier effects of rounding than of tongue posture changes for most talkers. The peak of the rounding gesture appears to be aligned to the onset of /ɹ/, causing varying degrees of gestural overlap with preceding sounds. While we don't expect rounding to contribute much to /s z/ retraction after /ɹ/ (because coda /ɹ/ is typically unrounded), we do think that this rounding gesture may have had some influence on the early stages of
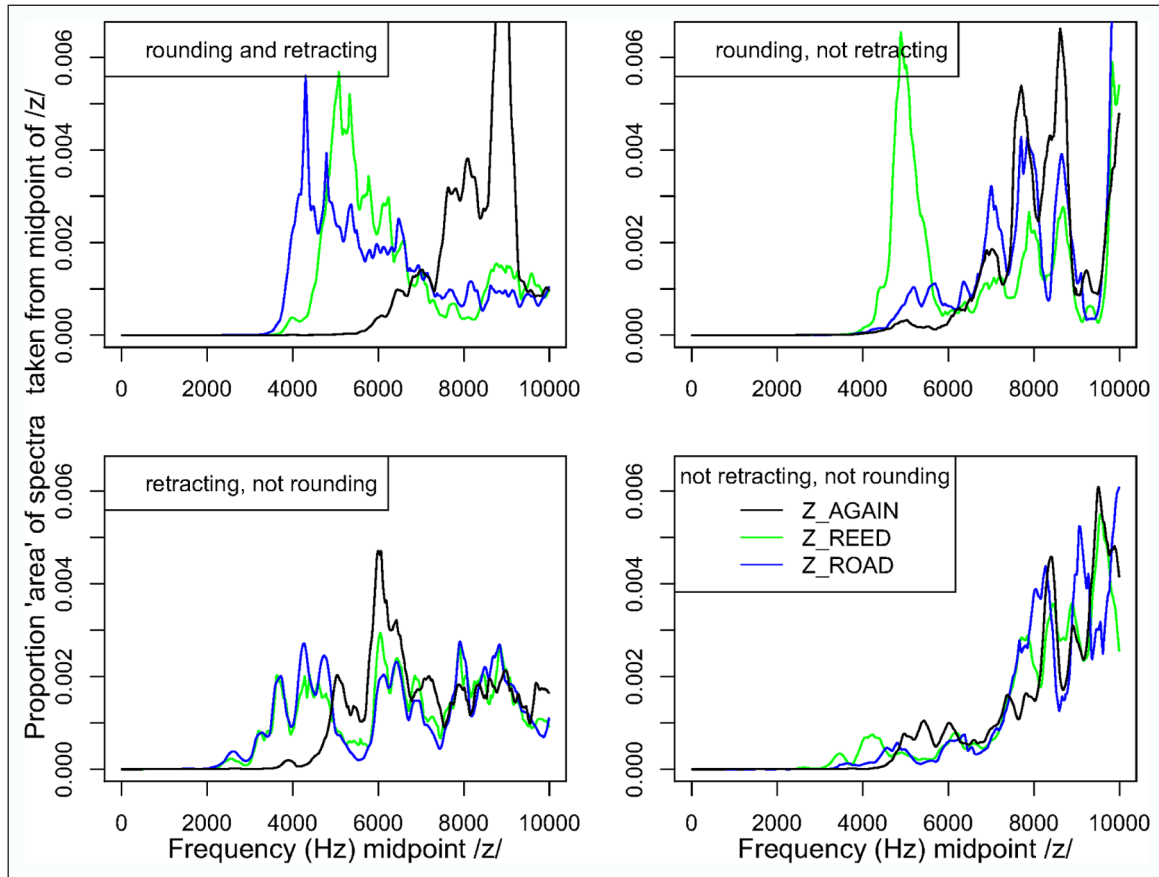
Figure 20: Multitaper spectra at midpoint of /z/ before /ɹ/, various tongue and lip configurations.

the retraction of /stɹ/ and the affrication of /t d/. A handful of talkers do have tongue postures with some similarity to /ʃ/ and show lower frequency peaks, suggesting robust coarticulation/anticipatory assimilation.

Figure 21 shows how both retraction and rounding cause a lower center of gravity in /s z/ before /ɹ/. Lip-rounding is known to spread across preceding sounds that are unspecified for rounding, and our articulatory data suggest that, for most talkers, the peak of the rounding gesture is aimed for the onset of the /ɹ/, which means that the /s z/ is at least partially rounded during most of its articulation. The rounding increases the size of the front cavity and may create a Helmholtz resonator, causing a large drop in the frequency of the spectral peak of the sibilance, and a corresponding reduction in the COG. When the tongue gestures for /ɹ/ begin early, an even greater decrease in spectral frequency is observed.

Combining the LDA signals for alveolar sibilants versus /ɹ/ with the spectral peak measurements, we examined the midpoint (40–60%) of the frication of /s/ and /z/ before /ɹ/, normalized using *z*-scored peak frequencies by talker to adjust for baseline talker differences. We can see that there is a significant decrease in the spectral peak concomitant with both tongue (ultrasound) and lip (video) signals in Figure 22. Notably, for the lip video LDA, we see that the peak frequency remains relatively consistent for LDA signal values below 0, after which the peak begins dropping quite a bit as the lip signal (rounding) increases. This suggests that there is no spectral peak frequency decrease without lip rounding. However, male /z#ɹ/ appears to be the exception, with a steady reduction as the signal increases, even with rounding < 0, and it is unclear why. Two challenges that we have yet to overcome include poor spectral peak tracking for
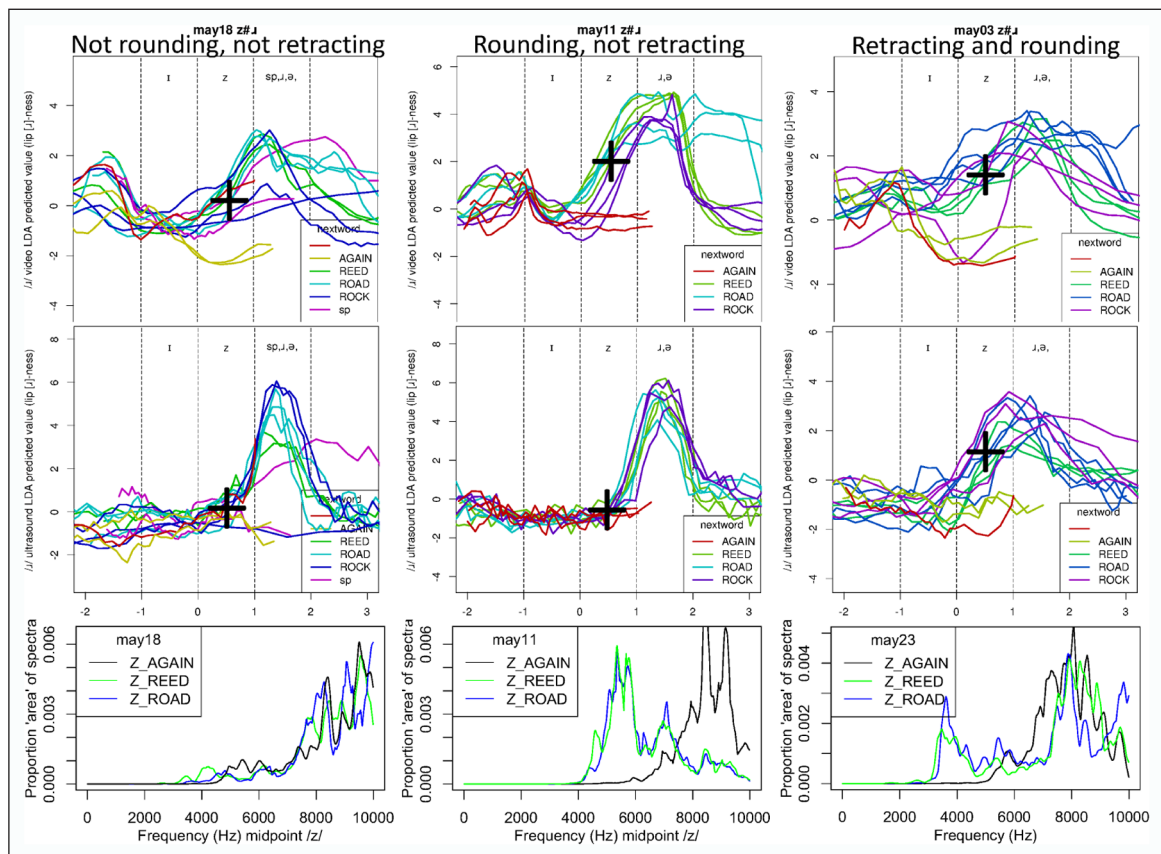
Figure 21: Both retraction and rounding cause a lower center of gravity in /s z/ before /ɹ/. Rounding additionally reduces the bandwidth of the lower frequency peak. First row = lip video, second row = tongue ultrasound, third row = multitaper spectra.

/z/, due to some (male) talkers having strong formants during frication, and difficulty with maintaining stable video over the timecourse of each session. Some participants would begin to slouch or sit up higher during recording, which may have been due to the flexibility of the headrest and not detected if it did not affect head angle. We have corrected our apparatus but have not yet been able to correct for this in the current data. We did, however, isolate 11 participants who had relatively stable video throughout the recording session, and their results are presented here. The following trends are generally consistent when including all talkers, but with much greater noise in the data. Mixed effects models were calculated separately for /s/ and /z/, with the token mean of the peak frequency in Hz for 40–60% of each token, as follows: `lmer(midpe akfreq~video+ultrasound+(1|talker)+(1|word))`. The inclusion of talker as a random effect deals with frequency differences related to vocal tract size, and more complicated models including a gender interaction did not have a significantly better fit. Models analyzing *z*-scored peaks showed essentially the same results, but are less useful in describing the acoustics.

For /s/, mixed effects modeling showed a decrease in peak frequency for the ultrasound (tongue) and video (lips) signal, with only the video showing a significant decrease (Est = 152, SE = 51.8, $t(59)$ = 2.9, $p < 0.001$), with about 152 Hz decrease estimated per LDA unit increase (from an intercept of 4185). For /z/, difficulty in peak tracking led to a slightly different outcome, with both ultrasound (Est = 150, SE = 49, $t(66)$ = 3.0, $p < 0.005$) and video (Est = 155, SE = 58, $t(66)$ = 2.7, $p < 0.01$) contributing about equally to the reduction in peak Hz frequency (Intercept = 4361 Hz).
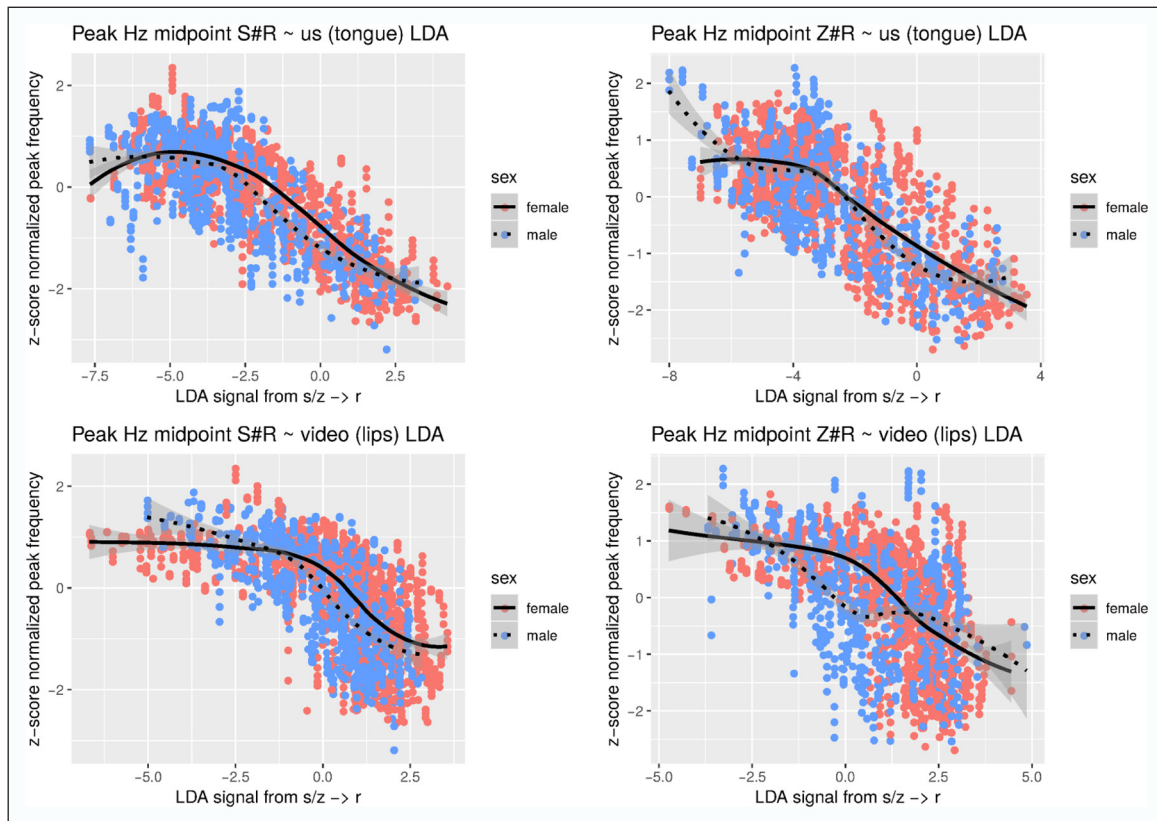
Figure 22: The more similar the /s z/ gesture is to the /ɹ/, the greater the decrease in the spectral peak during frication. Both retraction and rounding cause a lower frequency spectral peak in /s z/ before /ɹ/, but rounding does not influence the frequency of the peak at negative signal values, that is, without rounding, there is no decrease in the frequency of the peak. Values to the left of zero indicate more /s z/-like, while positive values to the right indicate more /ɹ/-like.

The coarticulation we observed for word-final /s z/ followed by /ɹ/ seems most likely to resemble the precursors of the coarticulation that may have led to the more clearly assimilatory patterns, but weaker due to the word boundaries. Both lip-rounding and tongue retraction can lead to lower peak frequencies (and consequently, COG). Lip-rounding increases the area of the front cavity and increases the length of the entire vocal tract, lowering the frequency of all resonances. If the aperture caused by lip constriction is small enough relative to the front cavity, it may create an extra low-frequency Helmholtz resonance. Lip-rounding additionally causes a decrease in bandwidth due to radiation damping (Fant 1970), so we might expect sharper peak frequencies. The tongue postures involved in producing an English /ɹ/ also increase the area of the front cavity, by creating a sublingual cavity (Alwan et al. 1997; Stevens 1999), which further reduces the peak sibilance, if sibilance can be maintained, which is not the case for all configurations.

Transitioning from an alveolar stop or sibilant to a rhotic is complicated by the postures used for each target, with some configurations allowing a rapid transition, and others that may make an excursion into "schwa space" (Gick & Wilson 2006). However, anticipatory rounding may spread across one or more preceding segments (e.g., Öhman 1966; Daniloff & Moll 1968; Bell-Berti & Harris 1981). We have found that there are varying degrees of roundness, both in protrusion and compression, and different timing of lip and tongue gestures. Thus, we hypothesize that reduction in spectral peaks and COG during the peak of frication, when not phonologically retracted, is attributable to lip-rounding, which may vary from talker to talker. The LDA analyses confirm that /ʃ/ rounding is different from

/ɹ/ rounding. Both types of sound exhibit a great deal of inter-speaker variability in shape and area of the labial constriction.

We had hypothesized that compatibility of tongue gestures between the alveolar and following rhotic might affect the percept of the coarticulated sounds, but as we noted earlier, there was a large amount of variability in the way the post-lexical stimuli were produced, with varying degrees of coarticulation related to prosodic factors besides the targeted covertly variable articulatory factors. In order to visualize the effects of reduced coarticulation, we plotted the peak Hz frequencies for /z#ɹ/ for the last 30 ms of the sibilance interval by the amount of time from the end of sibilance until the lowest F3 value, indicative of the peak of the /ɹ/ in Figure 23. Talkers with the highest peak frequencies in the last 30 ms of frication generally had a longer transition time into /ɹ/, while talkers with low peaks during the last 30 ms of frication were already beginning the transition to /ɹ/, regardless of bunching or retroflexion. While this does not shed light on the issue of particularities of tongue shape categories, it does provide a hint that faster transition times, aided by gestures that can be coordinated more quickly, do reduce the frequency in the spectral peak in a way that might change the percept of the resulting sound.

4 Discussion

In this paper, we have presented data from a series of inter-related investigations into sound change and coarticulatory variability involving English /ɹ/ in Raleigh, NC.[13] We reported our findings from multiple investigations of a corpus examining change over apparent time and multiple facets of a laboratory-based articulatory study examining consonantal coarticulation with /ɹ/.
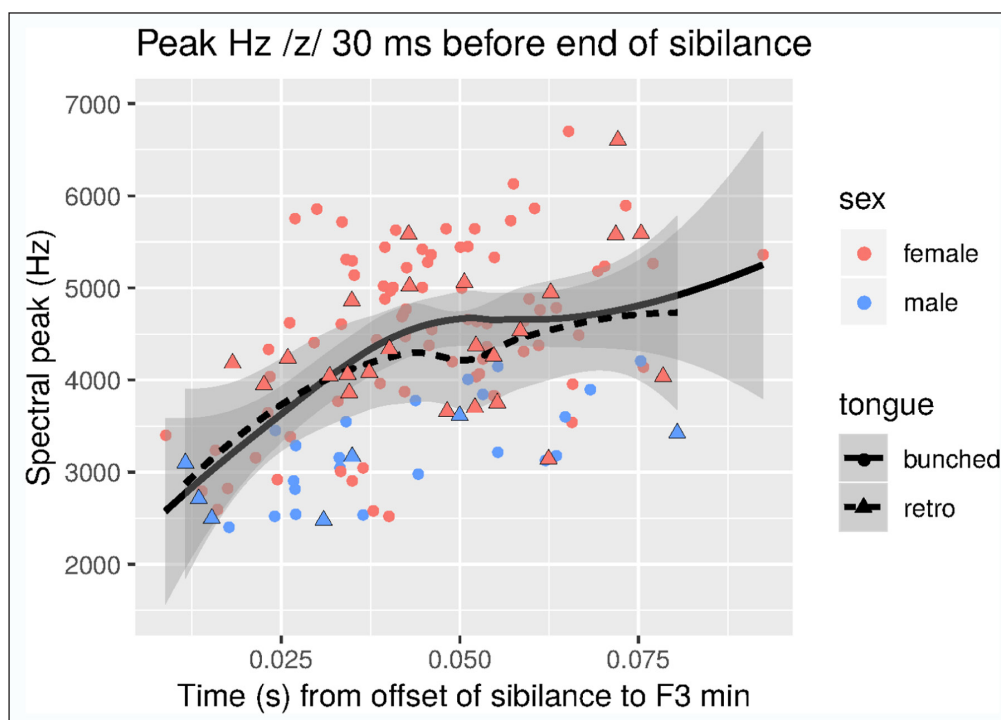

Figure 23: Peak Hz at 30 ms before the end of sibilance, plotted by difference in time between end of sibilance and peak /ɹ/ production (lowest F3 value). Less time between the offset of sibilance to the peak of the /ɹ/ produces lower spectral peaks in the stochastic noise.

---

[13] As described above, data from Wilbanks (2017) and Magloughlin (2018) have been included alongside data from the current study, as all are part of a larger-scale project.

The corpus studies revealed that /tɹ/ and /dɹ/ affrication are increasing over apparent time in Raleigh, NC. The trajectory of these related changes appears to be phonetically gradual, following unimodal distributions that increase linearly by birth year. The youngest generation appears to have phonologized these changes to postalveolar affricate targets, based on articulatory as well as acoustic measures, especially in more conservative lab-produced speech, in which all of the youngest talkers produce variants that are acoustically, perceptually, and articulatorily more similar to a postalveolar affricate than an alveolar stop or an /ɹ/. The corpus studies also suggested that /stɹ/ retraction in Raleigh, NC may be increasing for some talkers in some environments, but the distribution is much more scattered than we see for the phonetically gradual affrication of /t/ and /d/ before /ɹ/. The acoustic and perceptual data suggest a bimodal distribution for the youngest talkers, in which some have a phonologically postalveolar target, while others are still in a state of phonetic gradualness. Some amount of /s z/ retraction is also observed where these sounds meet /ɹ/ across word boundaries; although, there is less assimilatory overlap and it is not increasing over time. However, the retraction observed in these /s#ɹ/ and /z#ɹ/ contexts is suggestive of the type of variation that could have given rise to the other retraction (and possible affrication) patterns, but with less coarticulation, possibly due to different coarticulation strategies observed at word boundaries (as in Zsiga 1995; Ellis & Hardcastle 2002; Bermúdez-Otero 2016).

The articulatory studies showed the expected variation in /ɹ/ tongue postures, and additionally found that talkers produce onset /ɹ/ with generally compressed and sometimes in-rounded lips, while they usually produce /ʃ tʃ dʒ/ with generally open but protruded (outrounded) lips. The retracted /s/ in /stɹ/ and the affricated stops in /tɹ/ and /dɹ/ are all produced with tongue and lip gestures that are more similar to phonologically postalveolar consonants than to the /ɹ/ which triggers them – evidence that these patterns have been phonologized. On the other hand, the retracted /s z/ occurring word-finally before word-initial /ɹ/ involve tongue and lip gestures more similar to /ɹ/ than to postalveolar consonants – evidence that their acoustic retraction is more directly phonetically conditioned. The tongue and lip gestures involved with /ɹ/ articulation include pharyngeal, palatal, and labial constrictions. Co-production of alveolar sounds with any of these additional components can result in some degree of perceived retraction, vis-a-vis lower COG or spectral prominences.

The articulatory similarity we observed between /t/ and /d/ (before /ɹ/) and /tʃ/ and /dʒ/, suggests that, for the youngest talkers, these sounds are phonologically merged, though not in the sense of losing a distinction, since no minimal pairs are lost and no contrasts are neutralized. Given that allophones are not phonetically identical, and that coarticulation is required, there is no expectation that [tʃ] and [dʒ] before /ɹ/ will exhibit identical phonetic behavior to [tʃ] and [dʒ] in other (prevocalic) contexts, so the differences we observe between affricated variants of /t/ and /d/ before /r/, and phonological /tʃ/ and /dʒ/ (e.g., at the tongue root, see Figure 17) are consistent with a characterization of phonological merger (Magloughlin 2018), where, by merger, we mean strong similarity or identity of /t/ and /d/ (before /r/) with /tʃ/ and /dʒ/. The realization of the /s/ in /stɹ/ as [ʃ] in lab speech for several talkers suggests that /s/ in /stɹ/ clusters has phonologically merged with /ʃ/ for those talkers. No /ɹ/-loss has been noticed, but multiple talkers who produce a fully retracted [ʃ] in this environment also lack a [t] closure in the cluster, leading us to speculate that if this sound change were to advance throughout the community, it might be possible to see /stɹ/ merge with /ʃɹ/. But we limit our discussion to talkers in Raleigh, NC, and do not imagine that our speculation should apply to all communities for whom these changes are observed.

Examining relative timing of lip and tongue gestures showed that variability in lip-rounding in /s z/ before /ɹ/ accounted for much of the changes in the peak Hz of the midpoint of the sibilant interval. For individuals who exhibit less anticipatory rounding, or less rounding in general, a more retracted target might be posited to account for the spectral difference in others' speech. Regarding the actuation of sound change, this finding aligns with results found by Mann & Repp (1980), in which rounding contexts produced the expectation of a lower COG, while unrounded contexts produced the expectation for a higher COG, and thus the percept of a more retracted sound for the same stimulus. It is also in keeping with the findings of Yu (2013), in which individuals expected to have less sensitivity to local contextual effects are more likely to identify a rounded /s/ as /ʃ/. For /stɹ/, there is variability in the onset of the rounding gesture, as well as the degree of rounding, and whether it more closely resembles a postalveolar fricative, an /ɹ/, or some other configuration. Some talkers have rounding that extends over the course of multiple preceding segments, while others have a relatively late onset of rounding. Thus, we hypothesize that either coarticulation to /ɹ/-rounding or assimilation to a following post-alveolar variant of /t/, as we have seen, could have a similar triggering effect, which must necessarily depend on how these structures are represented individually. We imagine that there may be more than one route to /stɹ/ retraction, which is dependent on individual articulatory and phonological processes.

The literature on anticipatory rounding has shown talker- and language-specific variability in the expansion of the rounding gesture for /u/. For example, Daniloff & Moll (1968) showed extensive spreading of rounding across multiple segments, although their analysis is affected by the inclusion of /ɹ/ in some intervening sequences. Boyce (1990) showed that Turkish speakers generally allowed rounding to extend across multiple consonant segments, but English speakers had some consonants with their own rounding specifications. Many other studies (e.g., Bell-Berti & Harris 1981; Gelfer et al. 1989) have shown time-invariant rounding that is anchored to the onset of the rounded sound, and primarily affects the immediately preceding segment. Mann & Repp (1980) found a reduction in the perceptual influence of a rounded sound on a preceding fricative when a gap was inserted between the two segments. Listeners may also consider the spectral characteristics of the intervening stop, which, in our data showed some similarity to both /ɹ/ and /tʃ dʒ/ in tongue posture, but more similarity to /tʃ dʒ/ in the lip signal, not dissimilar to the pattern for /tɹ/. An explanation perhaps can be found in Gelfer et al. (1989), who observed that for some talkers, /st/ clusters had their own rounding gesture even in unrounded environments, so we must consider the possibility that the /t/ in these clusters also has its own rounding target, which may differ across talkers, and may spread to the preceding /s/.

While lip-rounding may explain early /stɹ/ retraction for some talkers, it does not explain how /t/ and /d/ before /ɹ/ became affricated and retracted. One thing that we can say regarding the relationship between /s#ɹ/ and /z#ɹ/ and the development of /tɹ/ and /dɹ/ affrication is that we noticed differences in the amount of time each talker took to get from /s z/ to /ɹ/, which led to different sounding coarticulation. The compatibility of lingual gestures between segments could influence this, but we were not able to directly address this in our data, due to differences in stress and speaking rate across talkers with different tongue gestures. However, we did see the spectral peak values decreasing more during sibilance for those talkers who transitioned more quickly into the following /ɹ/.

We hypothesize that /tɹ/ and /dɹ/ affrication may have resulted from some speakers being able to rapidly transition between the two sounds and create an intermediately coarticulated variant. While additional data would be required for us to isolate which configuration would most likely have led to this change, we did notice some intermediate

forms in /z#ɹ/ and /s#ɹ/ that would have been indicative of the early variants involved in /tɹ/ affrication. Some talkers had loss of frication and/or an excrescent schwa as they moved from the sibilant target to the /ɹ/. Some speakers had a transition in tongue posture during the sibilant, with the /ɹ/ tongue posture being reached by the end of the sibilant interval, with much lower-pitched sounding frication, and a couple of talkers even had full retraction during the midpoint of the sibilant. Still other talkers had labial frication between the sibilant and the /ɹ/ due to extreme lip-rounding. These variants resemble the variant pronunciations of /tw/ available in the early stages of /tw/ affrication, before the change has begun to accelerate, as reported in Columbus, OH by Smith (2013). We believe that /tɹ/ and /dɹ/ affrication may have emerged out of these sorts of variants, as the degree of coarticulatory overlap between /t d/ and /ɹ/ increased, and each new generation of talkers had to learn how best to replicate the resulting acoustic patterns.

5 Conclusion

Covert articulatory variability impacts the coarticulation strategies of variably produced sounds, which can have an auditory-acoustic effect. We hypothesized that these coarticulatory effects could actuate or increment a sound change by making coarticulation harder to compensate for. Our investigation of two related sound changes in progress that are triggered by /ɹ/ (/stɹ/ retraction and /tɹ/ and /dɹ/ affrication) indicate that these sequences involve approximation of postalveolar consonant tongue and lip gestures. Even those speakers who are not participating in /stɹ/ retraction still have a /t/ in that cluster that resembles the /t/ in affricated /tɹ/ (i.e., having positive LDA signals for both /ʃ/ and /ɹ/), possibly motivating this change; however, individual variability in tongue and lip gestures does not explain how it got to this point, and the resulting patterns suggest that the production of these clusters is already phonologically encoded for each talker. Thus we might expect to see different contributions of variable /ɹ/ gestures to these patterns driving the next set of changes, if any occur.

Post-lexical retraction (evidenced in lowered spectral peaks and COG) in /s#ɹ/ and /z#ɹ/ sequences shows the most difference between bunching and retroflexing speakers, probably because it is the context that allows the freest variation in /ɹ/ tongue shape, and also because, unlike the /stɹ/, /tɹ/, and /dɹ/ contexts, the acoustic results of this coarticulation have not been phonologized. But studying the lingual variation in /s#ɹ/ and /z#ɹ/ is complicated by the fact that most of the acoustic differences resulting from coarticulation are accounted for by lip rounding, which also varies across talkers. These facts may provide some clues about the preceding conditions for /tɹ/ and /dɹ/ affrication and /stɹ/ retraction, but additionally, we found that the reduction in frequency of spectral peaks was even stronger when the transition between the alveolar sibilant and the /ɹ/ unfolded more quickly. This pattern suggests that complementary and antagonistic tongue gestures could influence the timecourse of cluster coarticulation which could lead to different acoustic and phonological patterns.

In this study, we examined a range of sound patterns involving /ɹ/, in an effort to find some that involve direct coarticulation to particular /ɹ/ variants, which could motivate the actuation or incrementation of some sound changes involving /ɹ/. We found that the separate but related phenomena of /stɹ/ retraction and /tɹ/ and /dɹ/ affrication are each already phonological/stabilized patterns in Raleigh, such that younger speakers (including nearly all of our laboratory participants) produce sounds that are more similar to postalveolar consonants in these sequences. The other sequences we examined, all involving /s z/ immediately before or after /ɹ/, do not appear to involve phonologically retracted sibilants that resemble postalveolar consonants in their articulation. However, we only observed a large amount of coarticulatory variability in the sequences where the

sibilant precedes /ɹ/, which only occurs at word boundaries in our laboratory dataset. Complicating our investigation of /ɹ/ tongue shape effects, the most extreme effect of coarticulation that we found in this dataset is due primarily to lip rounding. We did find that the ability to co-produce alveolar sibilants with /ɹ/ lowered the spectral peaks in the resulting frication, but that because of prosodic limitations, we were unable to directly tie these to discrete tongue gestures such as tip-up or tip-down productions. We hope that future research can more directly tie this co-production to a model of covertly motivated sound changes. Future investigation of covert articulatory variation could focus on sequences where covert articulatory variation is expected, but associated phonological patterns have not been reported (i.e., not /stɹ/, /tɹ/, or /dɹ/); focus on communities where these sequences have not already undergone change; or focus on speakers who are old enough not to exhibit the retracted or affricated variants that may have developed in their communities.

Abbreviations

VOT = Voice Onset Time, AIC = Akaike Information Criterion, COG = Center of Gravity, P2FA = Penn Phonetics Lab Forced Aligner, DFT = Discrete Fourier Transform, MTS = Multitaper Spectra, PCA = Principal Component Analysis, LDA = Linear Discriminant Analysis

Additional File

The additional file for this article can be found as follows:

- **Supplementary file 1.** Appendix to 'Sound change and coarticulatory variability involving English /ɹ/'. DOI: https://doi.org/10.5334/gjgl.650.s1

Ethics and Consent

Research involving human subjects was ethically conducted in accordance with the Declaration of Helsinki and the U.S. Code of Federal Regulations, and was authorized by the Institutional Review Board at North Carolina State University (FWA00003429) under IRB protocol number 3074.

Competing Interests

The authors have no competing interests to declare.

References

Altendorf, Ulrike. 2003. *Estuary English: Levelling at the interface of RP and South-Eastern British English* 29. (Language in Performance 29). Tübingen, Germany: Gunter Narr Verlag.

Alwan, Abeer A., Shrikanth S. Narayanan & Katherine Haker. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I: The laterals. *Journal of the Acoustical Society of America* 101(2). 1064–1077. DOI: https://doi.org/10.1121/1.418030

Baker, Adam, Diana Archangeli & Jeff Mielke. 2011. Variability in English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23(3). 347–374. DOI: https://doi.org/10.1017/S0954394511000135

Barry, Martin. 1991. Temporal modelling of gestures in articulatory assimilation. In *Proceedings of the 12th International Congress of Phonetic Sciences* 4. 14–17. Université de Provence Aix-en-Provence.

Barry, Martin C. 1992. Palatalisation, assimilation and gestural weakening in connected speech. *Speech Communication* 11(4–5). 393–400. DOI: https://doi.org/10.1016/0167-6393(92)90044-8

Bass, Michael. 2009. Street or shtreet? investigating (str-) palatalisation in Colchester English. *Estro: Essex Student Research Online* 1(1). 10–21. https://www1.essex.ac.uk/journals/estro/documents/issue1/FullIssue.pdf#page=16.

Bauer, Laurie & Paul Warren. 2008. New Zealand English: Phonology. In Bernd Kortmann & Kate Burridge (eds.), *Varieties of English* 3. 39–63. Berlin and New York: Mouton de Gruyter.

Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85(4). 785–821. https://www.jstor.org/stable/40492954. DOI: https://doi.org/10.1353/lan.0.0165

Bell-Berti, Fredericka & Katherine S. Harris. 1981. A temporal model of speech production. *Phonetica* 38(1–3). 9–20. DOI: https://doi.org/10.1159/000260011

Bermúdez-Otero, Ricardo. 2015. Amphichronic explanation and the life cycle of phonological processes. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 374–399. Oxford: Oxford University Press.

Bermúdez-Otero, Ricardo. 2016. We do not need structuralist morphemes, but we do need constituent structure. In Daniel Siddiqi & Heidi Harley (eds.), *Morphological metatheory*, 385–428. DOI: https://doi.org/10.1075/la.229.13ber

Bermúdez-Otero, Ricardo. 2019. Individual differences and the explanation of sound change. *Glossa: A Journal of General Linguistics. Special Issue: Individuals, communities, and sound change.*

Bladon, R. Anthony W. & Francis Nolan. 1977. A video-fluorographic investigation of tip and blade alveolars in English. *Journal of Phonetics* 5. 185–193.

Boyce, Suzanne E. 1990. Coarticulatory organization for lip rounding in Turkish and English. *The Journal of the Acoustical Society of America* 88(6). 2584–2595. DOI: https://doi.org/10.1121/1.400349

Boyce, Suzanne E. & Carol Y. Espy-Wilson. 1997. Coarticulatory stability in American English /r/. *Journal of the Acoustical Society of America* 101(6). 3741–3753. DOI: https://doi.org/10.1121/1.418333

Brunner, Jana, Susanne Fuchs & Pascal Perrier. 2005. The influence of the palate shape on articulatory token-to-token variability. *ZAS Papers in Linguistics* 42. 43–67.

Brunner, Jana, Susanne Fuchs & Pascal Perrier. 2009. On the relationship between palate shape and articulatory behavior. *The Journal of the Acoustical Society of America* 125(6). 3936–3949. DOI: https://doi.org/10.1121/1.3125313

Byrd, Dani. 1994. Relations of sex and dialect to reduction. *Speech Communication* 15(1–2). 39–54. DOI: https://doi.org/10.1016/0167-6393(94)90039-6

Byrd, Dani. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24(2). 209–244. DOI: https://doi.org/10.1006/jpho.1996.0012

Byrd, Dani & Cheng Cheng Tan. 1996. Saying consonant clusters quickly. *Journal of Phonetics* 24(2). 263–282. DOI: https://doi.org/10.1006/jpho.1996.0014

Campbell-Kibler, Kathryn. 2011. Intersecting variables and perceived sexual orientation in men. *American Speech* 86(1). 52–68. DOI: https://doi.org/10.1215/00031283-1277510

Carignan, Christopher. 2014. *TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) software suite*. http://christophercarignan.github.io/TRAC-TUS.

Chitoran, Ioana & Louis Goldstein. 2006. Testing the phonological status of perceptual recoverability: Articulatory evidence from Georgian. In *Abstracts of the 10th Conference on Laboratory Phonology, Paris, June 29th–July 1st, 2006*, 69–70. http://web.archive.org/web/20061127051841/, http://aune.lpl.univ-aix.fr/~labphon10/abstracts/236.pdf.

Chitoran, Ioana, Louis Goldstein & Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in laboratory phonology* 7. 419–447.

Clarke, Sandra. 2008. Newfoundland English: Phonology. In Bernd Kortmann & Clive Upton (eds.), *Varieties of English* 1. 161–180. Berlin, New York: Mouton de Gruyter.

Cox, Felicity & Sallyanne Palethorpe. 2007. Australian English. *Journal of the International Phonetic Association* 37(3). 341–350. DOI: https://doi.org/10.1017/S0025100307003192

Daniloff, Raymond & Kenneth Moll. 1968. Coarticulation of lip rounding. *Journal of Speech, Language, and Hearing Research* 11(4). 707–721. DOI: https://doi.org/10.1044/jshr.1104.707

Dart, Sarah N. 1998. Comparing French and English coronal consonant articulation. *Journal of Phonetics* 26(1). 71–94. DOI: https://doi.org/10.1006/jpho.1997.0060

De Decker, Paul M. & Jennifer R. Nycz. 2012. Are tense [æ]s really tense? The mapping between articulation and acoustics. *Lingua* 122. 810–821. DOI: https://doi.org/10.1016/j.lingua.2012.01.003

Dediu, Dan & Scott R. Moisik. 2019. Pushes and pulls from below: Anatomical variation changes the relative probabilities of sound change pathways. *Glossa: A Journal of General Linguistics. Special Issue: Individuals, communities, and sound change*. DOI: https://doi.org/10.5334/gjgl.646

Delattre, Pierre & Donald C. Freeman. 1968. A dialect study of American r's by x-ray motion picture. *Linguistics* 6(44). 29–68. DOI: https://doi.org/10.1515/ling.1968.6.44.29

Dodsworth, Robin. 2014. *Network embeddedness and the retreat from Southern vowels in Raleigh* 20. 41–50. http://repository.upenn.edu/pwpl/vol20/iss2/6.

Dodsworth, Robin. 2017. Bipartite network structures and individual differences in sound change. Paper presented at the 4th International Workshop on Sound Change, University of Edinburgh, April 20–22, 2017. http://www.lel.ed.ac.uk/wsc/img/inv/Dodsworth.pdf.

Dodsworth, Robin & Mary Kohn. 2012. Urban rejection of the vernacular: The SVS undone. *Language Variation and Change* 24(2). 221–245. DOI: https://doi.org/10.1017/S0954394512000105

Ellis, Lucy & William J. Hardcastle. 2002. Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics* 30(3). 373–396. DOI: https://doi.org/10.1006/jpho.2001.0162

Espy-Wilson, Carol & Suzanne Boyce. 1994. Acoustic differences between "bunched" and "retroflex" variants of American English /r/. *Journal of the Acoustical Society of America* 95(5). 2823(A). DOI: https://doi.org/10.1121/1.409691

Fant, Gunnar. 1970. *Acoustic theory of speech production*. The Hague: Walter de Gruyter. DOI: https://doi.org/10.1515/9783110873429

Forrest, Jon. 2015. Community rules and speaker behavior: Individual adherence to group constraints on (ing). *Language Variation and Change* 27(3). 377–406. DOI: https://doi.org/10.1017/S0954394515000137

Foulkes, Paul & Gerard J Docherty. 2000. Another chapter in the story of /r/: 'Labiodental' variants in British English. *Journal of Sociolinguistics* 4(1). 30–59. DOI: https://doi.org/10.1111/1467-9481.00102

Fruehwald, Josef. 2016. The early influence of phonology on a phonetic change. *Language* 92(2). 376–410. DOI: https://doi.org/10.1353/lan.2016.0041

Fruehwald, Josef T. 2012. Redevelopment of a morphological class. *University of Pennsylvania Working Papers in Linguistics* 18(1). 10. https://repository.upenn.edu/pwpl/vol18/iss1/10/.

Gelfer, Carole E., Fredericka Bell-Berti & Katherine S. Harris. 1989. Determining the extent of coarticulation: Effects of experimental design. *Journal of the Acoustical Society of America* 86(6). 2443–2445. DOI: https://doi.org/10.1121/1.398452

Gick, Bryan & Ian Wilson. 2006. Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. In Louis Goldstein, Doug H. Whalen & Catherine T. Best (eds.), *Laboratory Phonology VIII: Varieties of phonological competence*, 635–659. Berlin: Mouton de Gruyter.

Glain, Olivier. 2014. Introducing contemporary palatalisation. In Theodora Lee (ed.), *Proceedings of the first Postgraduate and Academic Researchers in Linguistics at York (PARLAY 2013) conference*, 16–29. York: University of York. https://www.york.ac.uk/language/ypl/parlay/01/YPLPARLAY2013-02-Glain.pdf.

Gordon, Elizabeth & Margaret Maclagan. 2008. Regional and social differences in New Zealand phonology. In Bernd Kortmann & Kate Burridge (eds.), *Varieties of English* 3. 64–76. Berlin, New York: Mouton de Gruyter.

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. DOI: https://doi.org/10.1121/1.426900

Guy, Gregory R. 1980. Variation in the group and the individual: The case of final stop deletion. In William Labov (ed.), *Locating language in time and space* 1. 1–36. New York: Academic Press.

Gylfadottir, Duna. 2015. Shtreets of Philadelphia: An acoustic study of /str/-retraction in a naturalistic speech corpus. *University of Pennsylvania Working Papers in Linguistics* 21(2). 11. https://repository.upenn.edu/pwpl/vol21/iss2/11/.

Hall, Tracy Alan, Silke Hamann & Marzena Zygis. 2006. The phonetic motivation for phonological stop assibilation. *Journal of the International Phonetic Association* 36(1). 59–81. DOI: https://doi.org/10.1017/S0025100306002453

Harandi, Negar M., Jonghye Woo, Maureen Stone, Rafeef Abugharbieh & Sidney Fels. 2017. Variability in muscle activation of simple speech motions: A biomechanical modeling approach. *Journal of the Acoustical Society of America* 141(4). 2579–2590. DOI: https://doi.org/10.1121/1.4978420

Hardcastle, William J. & Peter J. Roach. 1979. An instrumental investigation of coarticulation in stop consonant sequences. In Harry F. Hollien & Patricia Hollien (eds.), *Current issues in the phonetic sciences*, vol. 9 (Current Issues in Linguistic Theory), 531–540. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.9.56har

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2011. The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern

British English. *Journal of the International Phonetic Association* 41(2). 137–156. DOI: https://doi.org/10.1017/S0025100310000265

Hatcher, Jessica. 2017. *Spectral measures of stop affrication: The case of /dɹ/ affrication in Raleigh.* Raleigh, NC: University of North Carolina MA capstone.

Hay, Jennifer. 2008. *New Zealand English.* Edinburgh University Press. DOI: https://doi.org/10.3366/edinburgh/9780748625291.001.0001

Heyne, Matthias, Xuan Wang, Donald Derrick & Kieran Dorreen. 2018. The articulation of /ɹ/ in New Zealand English. *Journal of the International Phonetic Association*, 1–23. DOI: https://doi.org/10.1017/S0025100318000324

Hoole, Philip & Marianne Pouplier. 2017. Öhman returns: New horizons in the collection and analysis of imaging data in speech production research. *Computer Speech & Language* 45. 253–277. DOI: https://doi.org/10.1016/j.csl.2017.03.002

Hueber, Thomas, Gérard Chollet, Bruce Denby & Maureen Stone. 2008. Acquisition of ultrasound, video and acoustic speech data for a silentspeech interface application. *Proceedings of the eighth International Seminar on Speech Production*, 365–369.

Hueber, Thomas, Guido Aversano, Gérard Chollet, Bruce Denby, Gérard Dreyfus, Yacine Oussar, Pierre Roussel & Maureen Stone. 2007. Eigentongue feature extraction for an ultrasound-based silent speech interface. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1245–1248. Honolulu, HI: Cascadilla Press. DOI: https://doi.org/10.1109/ICASSP.2007.366140

Jones, Daniel. 1956. *The pronunciation of English.* Cambridge: Cambridge University Press 4th edn.

Keating, Patricia A. 1991. Coronal places of articulation. In Carole Paradis & Jean-François Prunet (eds.), *The special status of coronals: Internal and external evidence* (Phonetics and Phonology 2), 29–48. San Diego, CA: Academic Press. DOI: https://doi.org/10.1016/B978-0-12-544966-3.50008-7

Klatt, Dennis H. 1975. Voice onset time, frication, and aspiration in wordinitial consonant clusters. *Journal of Speech, Language, and Hearing Research* 18(4). 686–706. DOI: https://doi.org/10.1044/jshr.1804.686

Kochetov, Alexei & Louis Goldstein. 2005. Position and place effects in Russian word-initial and word-medial stop clusters. *The Journal of the Acoustical Society of America* 117(4). 2571–2571. DOI: https://doi.org/10.1121/1.4788568

Koenig, Laura L., Christine H. Shadle, Jonathan L. Preston & Christine R. Mooshammer. 2013. Toward improved spectral measures of /s/: Results from adolescents. *Journal of Speech, Language, and Hearing Research* 56(4). 1175–1189. DOI: https://doi.org/10.1044/1092-4388(2012/12-0038)

Kuehn, David P. & Kenneth L. Moll. 1976. A cineradiographic study of VC and CV articulatory velocities. *Journal of Phonetics* 4(4). 303–320.

Kühnert, Barbara & Phil Hoole. 2006. Cohésion temporelle dans les groupes C1 /l/ initiaux en Français. In *XXVIe Journées d'Etude sur la Parole*, 545–548. Dinard, France: Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA); Association Francophone de la Communication Parlé (AFCP). https://hal.archives-ouvertes.fr/hal-00149245/.

Kühnert, Barbara, Phil Hoole & Christine Mooshammer. 2006. Gestural overlap and C-center in selected French consonant clusters. In *7th International Seminar on Speech Production (ISSP), Dec 2006,Ubatuba, Brazil*, 327–334. https://halshs.archives-ouvertes.fr/halshs-00185737/.

Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use: Readings in sociolinguistics*, 28–53. Englewood Cliffs, NJ: Prentice Hall.

Labov, William. 2009. Enigmas of uniformity. Talk presented at New Ways of Analyzing Variation (NWAV) 38, Ottawa, Canada. October 22–25, 2009.

Ladefoged, Peter & Ian Maddieson. 1996. *Sounds of the world's languages*. Oxford: Blackwell.

Ladefoged, Peter & Keith Johnson. 2014. *A course in phonetics*. Stamford, CT: Cengage Learning seventh edn.

Lawrence, Wayne. 2000. /str/→/ʃtr/: Assimilation at a distance? *American Speech* 75(1). 82–87. https://muse.jhu.edu/article/2713/. DOI: https://doi.org/10.1215/00031283-75-1-82

Lawson, Eleanor, James M. Scobbie & Jane Stuart-Smith. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics* 15(2). 256–268. DOI: https://doi.org/10.1111/j.1467-9841.2011.00464.x

Lawson, Eleanor, James M. Scobbie & Jane Stuart-Smith. 2013. Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics* 41(3–4). 198–210. DOI: https://doi.org/10.1016/j.wocn.2013.01.004

Lindau, Mona. 1985. The story of /r/. In Victoria A. Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 157–168. Orlando, FL: Academic Press.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H & H Theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-94-009-2037-8_16

Maclagan, Margaret. 2010. The English(es) of New Zealand. In Andy Kirkpatrick (ed.), *The Routledge handbook of world Englishes*, 152–164. London: Routledge.

Magloughlin, Lyra. 2016. Accounting for variability in North American English /ɹ/: Evidence from children's articulation. *Journal of Phonetics* 54. 51–67. DOI: https://doi.org/10.1016/j.wocn.2015.07.007

Magloughlin, Lyra. 2018. */tɹ/ and /dɹ/ in North American English: Phonologization of a coarticulatory effect*. Ottawa, Canada: University of Ottawa dissertation. DOI: https://doi.org/10.20381/ruor-21856

Mann, Virginia A. & Bruno H. Repp. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Attention, Perception, & Psychophysics* 28(3). 213–228. DOI: https://doi.org/10.3758/BF03204377

Mielke, Jeff, Adam Baker & Diana Archangeli. 2010. Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction. In Cécile Fougeron, Barbara Kühnert, Mariapaola D'Imperio & Nathalie Vallée (eds.), *Variation, detail, and representation* (LabPhon 10), 699–719. Berlin: Mouton de Gruyter.

Mielke, Jeff, Adam Baker & Diana Archangeli. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92(1). 101–140. DOI: https://doi.org/10.1353/lan.2016.0019

Mielke, Jeff, Christopher Carignan & Erik R. Thomas. 2017. The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: An ultrasound study. *Journal of the Acoustical Society of America* 142(1). 332–349. DOI: https://doi.org/10.1121/1.4991348

Mooshammer, Christine, Pascal Perrier, Christian Geng & Daniel Pape. 2004. An EMMA and EPG study on token-to-token variability. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel* 36. 47–63. http://www.haskins.yale.edu/staff/mooshammer/aipuk36_mooshammer.pdf.

Noiray, Aude, Marie-Agnès Cathiard, Lucie Ménard & Christian Abry. 2011. Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *The Journal of the Acoustical Society of America* 129(1). 340–349. DOI: https://doi.org/10.1121/1.3518452

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In Leiv E. Breivik & Ernst H. Jahr (eds.), *Language change: Contributions to the study of its causes* (Trends in Linguistics, Studies and Monographs 43), 173–198. Berlin: Mouton de Gruyter.

Ohala, John J. 2011. Accommodation to the aerodynamic voicing constraint and its phonological relevance. In *Proceedings of the International Congress of Phonetic Sciences XVII*, 64–67. https://www.internationalphoneticassociation.org/icphsproceedings/ICPhS2011/OnlineProceedings/SpecialSession/Session1/Ohala/Ohala.pdf.

Ohala, John J. & Maria-Josep Solé. 2010. Turbulence and phonology. In Susanne Fuchs, Martine Toda & Marzena Zygis (eds.), *Turbulent sounds: An interdisciplinary guide*, 37–97. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110226584.37

Öhman, Sven E.G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America* 39. 151–168. DOI: https://doi.org/10.1121/1.1909864

Ong, Darryl & Maureen Stone. 1998. Three-dimensional vocal tract shapes in /r/ and /l/: A study of MRI, ultrasound, electropalatography, and acoustics. *Phonoscope* 1(1). 1–13.

Ostry, David J. & Kevin G. Munhall. 1985. Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America* 77(2). 640–648. DOI: https://doi.org/10.1121/1.391882

Perkell, Joseph S. & Melanie L. Matthies. 1992. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within-and cross-subject variability. *Journal of the Acoustical Society of America* 91(5). 2911–2925. DOI: https://doi.org/10.1121/1.403778

Perkell, Joseph S., Melanie L. Matthies, Mario A. Svirsky & Michael I. Jordan. 1993. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study. *The Journal of the Acoustical Society of America* 93(5). 2948–2961. DOI: https://doi.org/10.1121/1.405814

Phillips, Betty S. 2001. Lexical diffusion, lexical frequency, and lexical analysis. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 123–136. DOI: https://doi.org/10.1075/tsl.45.07phi

Piergallini, Mario. 2011. */str/ retraction in North Carolina*. Raleigh, NC: North Carolina State University MA capstone.

Podesva, Robert J. & Janneke Van Hofwegen. 2014. How conservatism and normative gender constrain variation in inland California: The case of /s/. *University of Pennsylvania Working Papers in Linguistics* 20(2). 15. https://repository.upenn.edu/pwpl/vol20/iss2/15/.

Read, Charles. 1971. Pre-school children's knowledge of English phonology. *Harvard Educational Review* 41(1). 1–34. DOI: https://doi.org/10.17763/haer.41.1.91367v0h80051573

Reidy, Patrick F. 2013. spectRum. R package. https://github.com/patrickreidy/spectRum.

Reidy, Patrick F. 2015. A comparison of spectral estimation methods for the analysis of sibilant fricatives. *JASA Express Letters* 137(4). EL248–EL254. DOI: https://doi.org/10.1121/1.4915064

Rudy, Krista & Yana Yunusova. 2013. The effect of anatomic factors on tongue position variability during consonants. *Journal of Speech, Language, and Hearing Research* 56(1). 137–149. DOI: https://doi.org/10.1044/1092-4388(2012/11-0218)

Rutter, Ben. 2011. Acoustic analysis of a sound change in progress: The consonant cluster /stɹ/ in English. *Journal of the International Phonetic Association* 41(1). 27–40. DOI: https://doi.org/10.1017/S0025100310000307

Rutter, Ben. 2014. The acquisition of newly emerging sociophonetic variation: /str-/ in American English. *Journal of Child Language* 41(5). 1166–1178. DOI: https://doi.org/10.1017/S0305000913000342

Shapiro, Michael. 1995. A case of distant assimilation: /str/ → /ʃtr/. *American Speech* 70(1). 101–107. DOI: https://doi.org/10.2307/455876

Smith, Bridget J. 2013. *The interaction of speech perception and production in laboratory sound change*. Columbus, OH: The Ohio State University dissertation. http://rave.ohiolink.edu/etdc/view?acc_num=osu1374116504.

Stevens, Kenneth. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Edward E. David & Peter B. Denes (eds.), *Human communication: A unified view*, 51–66. New York: McGraw Hill.

Stevens, Kenneth N. 1999. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, Mary & Jonathan Harrington. 2016. The phonetic origins of /s/- retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58. 118–134. DOI: https://doi.org/10.1016/j.wocn.2016.08.003

Stevens, Mary, Jonathan Harrington & Florian Schiel. 2019. Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English. *Glossa: A Journal of General Linguistics. Special Issue: Individuals, communities, and sound change.* DOI: https://doi.org/10.5334/gjgl.620

Strycharczuk, Patrycja. 2012. *Phonetics-phonology interactions in presonorant voicing*. Manchester, UK: University of Manchester dissertation. https://personalpages.manchester.ac.uk/staff/patrycja.strycharczuk/Research_files/PhD_2_sided.pdf.

Strycharczuk, Patrycja & James Scobbie. 2017. Whence the fuzziness? Morphological effects in interacting sound changes in Southern British English. *Laboratory Phonology* 8(1). 7. DOI: https://doi.org/10.5334/labphon.24

Tiede, Mark K., Suzanne E. Boyce, Christy K. Holland & K. Ann Choe. 2004. A new taxonomy of American English /r/ using MRI and ultrasound. *The Journal of the Acoustical Society of America* 115(5). 2633–2634. DOI: https://doi.org/10.1121/1.4784878

Treiman, Rebecca. 1987. On the relationship between phonological awareness and literacy. *Current Psychology of Cognition* 7(5). 524–529.

Twist, Alina, Adam Baker, Jeff Mielke & Diana Archangeli. 2007. Are 'covert' /ɹ/ allophones really indistinguishable? *Penn Working Papers in Linguistics* 13(2). http://repository.upenn.edu/pwpl/vol13/iss2/16/.

Uldall, Elizabeth. 1958. American 'molar' R and 'flapped' T. *Revista do Laboratório de Fonética Experimental da Faculdade de Letras da Universidade de Coimbra* 4. 103–106.

Westbury, John R., Michiko Hashi & Mary J. Lindstrom. 1998. Differences among speakers in lingual articulation for American English /ɹ/. *Speech Communication* 26. 203–226. DOI: https://doi.org/10.1016/S0167-6393(98)00058-2

Wilbanks, Eric. 2017. Social and structural constraints on a phonetically-motivated change in progress: (str) retraction in Raleigh, NC. *University of Pennsylvania Working Papers in Linguistics* 23(1). 33. https://repository.upenn.edu/pwpl/vol23/iss1/33/.

Yu, Alan C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization*, 201–227. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199573745.003.0010

Yuan, Jiahong & Mark Liberman. 2008. Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics '08*. 5687–5690. http://languagelog.ldc.upenn.edu/myl/ldc/ICASSP_final.pdf.

Yuan, Jiahong & Mark Liberman. 2009. Investigating /l/ variation in English through forced alignment. In *Tenth Annual Conference of the International Speech Communication*

]u[

Association, 2215–2218. https://www.iscaspeech.org/archive/interspeech_2009/i09_2215.html.

Zhou, Xinhui, Carol Y. Espy-Wilson, Suzanne E. Boyce, Mark Tiede, Christy Holland & Ann Choe. 2008. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *Journal of the Acoustical Society of America* 123(6). 4466–4481. DOI: https://doi.org/10.1121/1.2902168

Zsiga, Elizabeth C. 1995. An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In Bruce Connell & Amalia Arvaniti (eds.), *Phonology and phonetic evidence: Papers in Laboratory Phonology iv*, 67–79.