Temporal dynamics of /s/-retraction in American English

Jacob B. Phillips*
Phonology Laboratory, University of Chicago, USA

*University of Chicago, Department of Linguistics, 1115 58th Street, Rosenwald Hall Room 224, Chicago, Illinois 60637, USA

Abstract
This study explores the nature of /s/-retraction, a sound change in American English by which /s/ approaches [ʃ] in the context of /r/. Through an examination of phonological, prosodic, and temporal factors this study asks if /s/-retraction is better categorized as an assimilatory or coarticulatory process. Results support a coarticulatory account, as retraction is observed to be highly dependent on the duration of the sibilant and the relative distance from the coarticulatory trigger. While no evidence for prosodic strengthening was observed across the board, its interaction with sibilant duration suggests that retraction is resisted phrase-initially. Additionally, this study provides strong empirical evidence for individual variability in the effects of these predictors, including some individuals for whom prosodic position is more meaningful in conditioning sibilant production than phonological environment.

Keywords: Coarticulation, /s/-retraction, sound change, prosodic strengthening, individual variability

1. Introduction
Coarticulation and assimilation are two distinct processes that yield similar results: one speech sound changes in some form to be more similar to a nearby sound. Assimilation is phonological process by which a speech sound categorically changes in some form of its representation to be more similar to neighboring sounds. In contrast, coarticulation is a phonetic process by which a speech sound exhibits variable, gradient changes in its realization to be more similar to neighboring sounds. Thus, at their simplest, phonology – and thus assimilation – is more categorical, and phonetics – and thus coarticulation – is more gradient (Keating, 1988, 1990; Pierrehumbert, 1990; Zsiga, 1993, 1997)

*Corresponding author
Email address: jbphillips@uchicago.edu (Jacob B. Phillips)
Both assimilation and coarticulation have been observed robustly cross-linguistically (Farnetani & Recasens, 2010), although coarticulatory patterns have been demonstrated to vary significantly from language to language (Beddor et al., 2002; Manuel, 1999; Ohman, 1966), suggesting it is not a purely mechanical process. In contrast to assimilation, in which the change is feature specification is observed equally across the entire sound, coarticulation is dynamic and highly local, with the degree of coarticulation increasing as it nears the coarticulatory trigger (Keating, 1985, 1988; Whalen, 1990; i.a.)

The present study focuses on /s/-retraction, a sound change in progress in American English that has been characterized as both assimilation and coarticulation in the literature. Observed throughout the continental United States, Canada and abroad, /s/-retraction is the process by which /s/ is produced approaching [ʃ] in the context of /r/, especially in /str/ clusters like street (Archangeli et al., 2011; Baker et al., 2011; Durian, 2007; Gylfadottir, 2015; Lawrence, 2000; Mielke et al., 2010; Rutter, 2011; Shapiro, 1995; Wilbanks, 2017). While there is remarkable variation both within and between speakers in the production of /s/ in all /sCr/ clusters, where ⟨C⟩ is taken to represent a voiceless stop /{p,t,k}/, /s/ is significantly more retracted in /str/ than in /spr/ (spree) or /skr/ (scream) clusters (Baker et al., 2011; Gylfadottir, 2015), to the extent that most existing research has focused exclusively on /s/-retraction in /str/ clusters.

This study seeks to better understand the underlying mechanism of the change through an examination of the temporal dynamics of /s/-retraction. Specifically, this study examines how various factors including sibilant duration, prosodic boundaries and gestural timing – in addition to the identity of the intervening stop – condition the observed retraction over the course of the sibilant in /sCr/ clusters. A dynamic trajectory, in which the degree of retraction increases over the course of the sibilant, and durational conditioning, in which retraction is resisted in sibilants with longer durations, would suggest a coarticulatory account of the sound change. In contrast, categorical and stable trajectories that are not strongly conditioned by duration or speech rate would suggest an assimilatory account of retraction.

1.1. Retraction

Throughout the short history of research on /s/-retraction, there has been much debate as to the causes of the observed retraction and how to best account for it, often focusing on /str/ clusters alone. In the first dedicated account of /s/-retraction, Shapiro (1995) proposes that /s/ exhibits long-distance place of articulation assimilation from /t/ across the intervening /t/. Based on his observation that retraction is not seen in /spr/ and /skr/ clusters, Shapiro explicitly prohibits retraction in those clusters by proposing that /p/ and /k/ as non-coronal segments block the assimilation of place features. Lawrence (2000) agrees that /s/-retraction is an assimilatory process observed only in /str/ (and /stʃ/) clusters, but counters that /s/-retraction is inherently a local process due to affrication on the /t/. This account necessitates two ordered processes,
where the intervening /t/ is first affricated to /tʃ/ as a result of its adjacency to /r/, following the robust sound change in English in which /t/ and /d/ affricate preceding /r/ as in truck /trʌk/ or drink /drɪŋk/ (Read, 1975; Smith, 2013). Subsequently, /s/ locally assimilates the place features of /tʃ/ to yield a final /ʃʃʃ/ cluster. The local assimilatory account also explains the observed asymmetry between /str/ and /s{p,k}r/ clusters, as the intervening /p/ and /k/ do not affricate or bear post-alveolar place features. More recent research by Archangeli, Baker, and Mielke (Archangeli et al., 2011; Baker et al., 2011; Mielke et al., 2010) challenge these cut-and-dry assimilatory accounts, proposing that /s/-retraction is fundamentally a coarticulatory process for most speakers. Baker et al. (2011) observed gradient, intermediate forms between prevocalic /s/ and /ʃ/ in all /sCr/ clusters, both acoustically and articulatorily, although /s/ in /str/ is significantly more /ʃ/-like than in /spr/ or /skr/ clusters. For some speakers, proposed to be phonologized retractors, /s/ in /str/ clusters is perceptually and acoustically more /ʃ/-like: however, it is rarely observed to be fully within the range of canonical /ʃ/ for any speaker (Mielke et al., 2010). Many additional studies have also observed gradient, intermediate forms in /str/ clusters (Durian, 2007; Gylfadottir, 2015; Wilbanks, 2017); however Rutter (2011) observed more categorical /s/ or /ʃ/ productions in /str/ environments.

While many investigations of /s/-retraction have focused on word-initial /str/ clusters, some studies have examined the role of the position of the cluster within the word. In each of these studies, an effect of word position has been replicated, with increased retraction observed in word-medial positions (Durian, 2007; Gylfadottir, 2015; Wilbanks, 2017). Additionally, apparent time trends for the interaction of birth year and word position have been observed, with younger speakers showing even more retraction word-medially than older generations (Gylfadottir, 2015; Wilbanks, 2017). With these findings in mind, Durian (2007) and Wilbanks (2017) suggest that word-medial positions may be the locus of the actuation of /s/-retraction. Parallel effects of word-position have also been observed for prevocalic instances of /s/ in which retraction is not expected (Phillips et al., 2018), with word-initial prevocalic /s/ produced with a higher centroid frequency than word-medial instances of /s/, suggesting that regardless of phonological environment, sibilants enhance phonological contrast, and thus resist retraction, word-initially. No work to our knowledge has examined the role of phrase positions or sibilant duration in conditioning /s/-retraction.

1.2. Timing, duration & strengthening

In order to identify the underlying process of /s/-retraction – whether it is better characterized as coarticulation or assimilation – the present study examines the temporal dynamics of /s/, focusing on the influence and interactions of three factors: timing relations, sibilant duration and prosodic position.

Within articulatory phonology (Browman & Goldstein, 1992, 1995) in which speech production is a series of dynamic gestures, coarticulation is the natural result of overlap in the temporal specification of neighboring gestures. Thus, coarticulation is inherently a process concerned with timing relations and over-
lapping gestures (Barry, 1985, 1991; Beddor, 2009; Cho, 1998; Delvaux et al., 2012; Iskarous & Kavitskaya, 1993, 1995, 2000, i.a). Consonant clusters present particular challenges to gestural planning as they require concurrent and potentially conflicting demands on an articulator. This results in an articulation in which the individual consonants are not discretely or canonically articulated (Byrd & Tan, 1996).

Rather, the articulations of consonants in clusters overlap substantially (Recasens et al., 1993) and can be significantly reduced (Barry, 1985; Kerswill, 1985; Nolan, 1992). In a study of English fricatives, Zsiga (1995) examined /s/-/ʃ/ sequences across word boundaries, such as in a phrase like press your point, observing that /s/ can approach /ʃ/ as a result of the gestural overlap across the word boundary. Despite resulting in a similar acoustic realization as historical assimilations of /s/ to /ʃ/ preceding /ʒ/, for example press/pressure, Zsiga utilized electroplagography to demonstrate that the coarticulation across word boundaries can be differentiated from phonological assimilation in its articulatory configuration. Expanding this investigation, Zsiga (2000) analyzed the acoustic realization of gestural overlap at word boundaries in both English and Russian, finding that English and Russian exhibit different patterns of gestural overlap, with English exhibiting a more gradient and variable production, suggesting coarticulation, and Russian exhibiting a more categorical and consistent production, suggesting assimilation.

In the absence of articulatory data, gestural timing information can be garnered from the acoustic signal. One means of doing so is to calculate the approximate alignment measurements, including left-edge, c-center and right-edge, from acoustic landmarks (Selkirk & Durvasala, 2013; Ruthan et al., 2018). The c-center (consonant center, Browman & Goldstein, 1988) is the temporal midpoint of an onset singleton or cluster relative to an anchor point, usually the offset of the vowel. Browman & Goldstein (1988) demonstrated that in English, words with simplex onsets, like lot, and words with complex onsets, like plot or splot use the same timing relations between the consonant onset and the vowel. Specifically, Browman & Goldstein demonstrate that the interval between the c-center of the simplex onset to the anchor point is comparable to the interval between the c-center of the complex onset and the anchor point. However, the same relationship is not observed for simplex and complex codas. Furthermore, variation in timing relations can be observed between languages, with some exhibiting a c-center effect (e.g. Goldstein et al., 2007 for Georgian; Marin, 2013 for Romanian) but other exhibiting a preference for other timing relations (e.g. Goldstein et al., 2007 for Tashlhiyt Berber; Pouplier & Beaus, 2011 for Slovak; Shaw et al., 2011 for Moroccan Arabic), and others alternating depending on the phonological environment (e.g. Brunner et al., 2014 for German; Hermes et al., 2013 for Italian).

Another factor tightly intertwined with coarticulation is duration and speech rate: Segments produced with longer durations have been shown to exhibit less coarticulation than segments in identical phonological environments produced with shorter durations (Fourakis, 1991; Fowler, 1981; Gay et al., 1974; Gendrot & Adda-Decker, 2005; Iskarous & Kavitskaya, 2010; Lindblom, 1963; Tsao et al., 2007).
In work on Swedish consonant-vowel coarticulation, Lindblom (1963) found that vowel reduction is a function of segment duration. Lindblom observed that vowels are less likely to reach their acoustic targets, or be produced with canonical on- or off-glide movements, as the duration of the vowel decreases. Lindblom proposes that vowel reduction is primarily an intrinsic result of human physiology and motor planning, by which targets simply cannot be reached in the amount of time provided by the speech rate. Whalen (1990) also found a duration effect, looking at vowel-to-vowel coarticulation in nonce words in English. Crucially, Whalen did not present participants with the entire nonce word until the utterance had began, which resulted in less anticipatory coarticulation in the unknown condition, suggesting that coarticulation is planned by the speaker. Gendrot & Adda-Decker (2005) examined a corpus of broadcast French and German, finding that vowel space for both languages shrinks concentrically as the duration of the vowels examined decreases, providing strong evidence for the relationship of duration and coarticulation outside of the laboratory. In contrast, Mok (2011) examined vowel-to-vowel coarticulation in Thai, a language with a phonological vowel length distinction, finding that there is no simple one-to-one relationship between duration and coarticulation. Additionally, Mok proposes that a speaker’s effort to maintain clarity can override effects of segment duration. Furthermore, Cho (2004) suggests that duration factors alone cannot account for the degree of coarticulation observed and related factors of prosodic strengthening must be taken into account.

Prosodic position has also robustly been demonstrated to influence speech production, yielding prosodic strengthening. Prosodic strengthening is the process by which sounds at prosodic landmarks, such as phrase boundaries and under prominence, are enhanced temporally or spatially, while the specific nature of the observed strengthening varies depending on the sound or landmark. Prosodic lengthening, has been robustly observed both domain-finally (Byrd, 2000; Byrd et al., 2006; Cho, 2002, 2006; Onaka, 2006, i.a.) and domain-initially (Byrd & Saltzman, 2003; Cho, 2002, 2006; Cho & Keating, 2001, 2009; Cho & Kim, 2014; Fougeron, 2001; Fougeron & Keating, 1997; Keating et al., 2003; Onaka, 2003, i.a.). However, due to the variation between sounds and boundary types, some sounds have been observed not to exhibit lengthening, or to even exhibit shortening at prosodic boundaries (Cho et al., 2017; Cho & Kim, 2014; Cho & McQueen, 2005). Cho & McQueen (2005) found that voice onset time (VOT) in Dutch voiceless stops shortens under prosodic prominence, exhibiting the reverse pattern as English voiceless stops. While both sounds may be described with the same [− voice] feature, the two differ with respect to the [+/− spread glottis] feature. Thus in Dutch, by shortening VOT, voiceless stops are enhancing the [− spread glottis] feature. Cho & Kim (2014) observed the same pattern for voiceless stops in English /sC/ clusters shortening at prosodic boundaries, where voiceless stops following /s/ contrast with prevocalic voiceless stops by having the phonetic feature [− spread glottis]. Of particular interest to this study is the other side of Cho & Kim (2014), which found that /s/ in /sC/ clusters did exhibit the expected prosodic lengthening at phrase boundaries.

In addition to temporal enhancement, segments exhibit spatial enhancement
as a means of maximizing phonological contrasts, a form of highly localized hy-
found that /i/ in English is raised acoustically and articulatorily adjacent to
boundaries, but fronted in accented positions, both dimensions that serve to
enhance the contrast between /i/ and other vowels in English. While voic-
ing contrasts have been robustly observed to be enhanced at boundaries (Cho
& Kim, 2014; Cho & McQueen, 2005; Clayards & Knowles, 2015; Cole et al.,
2007), less work has found evidence for place of articulation contrasts in con-
sonants. Looking at /n/ and /t/ in English, Cho & Keating (2009) found
prosodic strengthening on a variety of dimensions including peak contact and
nasal energy for /n/ and RMS energy and burst centroid frequency for /t/.
Keating et al. (1999) found evidence for increased linguopalatal contact in non-
sibilant obstruents, but crucially found less evidence for spatial strengthening
in the sibilant obstruents examined. This generalization was also replicated by
Clayards & Knowles (2015), who found stronger evidence for voicing contrast
enhancement than place enhancements at word boundaries for sibilants. These
findings contrast with the place enhancement effect of initial positions observed
by Phillips et al. (2018), demonstrating that word-initial instances of prevocalic
/s/ are characterized by a lower centroid frequency than word-medial prevocalic
/s/, effectively maximizing the phonological distance between /s/ and /ʃ/.

Furthermore, prosodic strengthening is manifested by coarticularatory resis-
tance and local hyperarticulation, inherently intertwined with the processes of
lengthening and contrast enhancement (Cho & Keating, 2009; Fougeron & Keat-
ing, 1997). Thus, adjacency to prosodic boundaries not only lengthens segments,
but it can also serve to resist coarticulation, effectively leading to local hyper-
articulation. Similarly, Cho (2004) found that vowels at boundaries of higher
prosodic positions (Prosodic Word vs. Intermediate Phrase vs. Intonational
Phrase) resist coarticulation with neighboring vowels. Cho suggests that the
coa rticularatory resistance is not simply a form of prosodic strengthening, but
additionally an effort on the speaker’s role to achieve clarity. This view of coar-
ticulation suggests that coarticulated speech is more ambiguous and increases
effort for the listener (e.g., Manuel, 1990; Lindblom, 1990), contrasting with
recent research that coarticulation is beneficial to the listener, providing more
information about upcoming sounds (Beddor et al., 2013; Pycha, 2016; Scarbor-

1.3. Hypotheses and predictions

The present study is designed to better understand the nature of /s/-retraction
through an examination of the ways in which the following consonants, adja-
cency to a prosodic boundary, sibilant duration and timing factors influence
the degree and trajectory of retraction observed. Specifically, this study asks
if retraction is best characterized as a coarticulatory or assimilatory process,
which can be elucidated through the examination of several concrete factors.
In this section, I outline the specific, testable predictions made by the different
accounts.
A model of retraction as a coarticulatory process first predicts intermediate forms between prevocalic /s/ and /f/ in all /s{p,t,k}r/ clusters, as observed by [Mielke et al. (2010)]; but with the intermediate forms in /str/ clusters more advanced, i.e. more /f/-like than in /s{p,k}r/ clusters. Additionally, a coarticulatory model of retraction predicts that retraction will be highly localized and dynamic, with the degree of retraction increasing throughout the duration of the sibilant as the distance from /r/ decreases. Furthermore, a coarticulatory model of retraction predicts that retraction should be resisted in segments with longer durations, whether that be a result of adjacency to a prosodic boundary or due to speech rate and timing, following [Lindblom (1963)]. If prosodic lengthening is observed, this would potentially decrease the degree of retraction phrase-initially compared to phrase-medially, parallel to the observed pattern for reduced retraction word-initially vs. word-medially [Durian (2007); Gylfadottir (2015); Wilbanks (2017)]. Conversely, adjacency to a prosodic boundary may also serve to shorten the duration of the intervening stop to enhance the [+spread glottis] feature, following [Cho & Kim (2014)], decreasing the temporal gap between /s/ and /r/ and potentially encouraging greater retraction in phrase-initial positions.

A model of retraction as primarily a phonological process proposes that /s/ in /str/ clusters assimilates the [+anterior] feature of /r/ for retractors but retains the [+anterior] feature for non-retractors. Such an account predicts that speakers produce forms that more categorically fall into the speakers /s/ and /f/ ranges, as observed by [Rutter (2011)]. An assimilatory account would also preclude retraction in /spr/ or /skr/ clusters, as the long distance assimilation is only possible across the under- or unspecified /t/ [Shapiro (1995)]. As a phonological process, an assimilatory account of retraction does not predict dynamic trajectories of retraction that vary as a function of the duration of the sibilant or the distance from /r/, as feature-spreading is categorical and would be expected to observed at similar degrees over the course of the sibilant [Manuel (1987)]. Furthermore, in an assimilatory account, prosodic strengthening may occur at phrase boundaries, enhancing place contrast between sibilants. This would predict that retractors would produce a lower centroid frequency phrase-initially to maximally distinguish /str/ from /s/, while non-retractors would produce a higher centroid frequency to distinguish /str/ from /f/.

## 2. Methods

### 2.1. Participants

Thirty participants were recruited from the University of Chicago community and received either payment or course credit for their participation. Twenty-one participants identified as male and nine as female. All participants were college-aged (mean age 19) native English speakers, raised across the continental United States, with higher concentrations in the Northeast and Midwest. Most participants reported being raised in suburban areas (n=17), with fewer participants from urban (n=8) or rural (n=5) areas. No participants reported
speech or hearing disorders/abnormalities. One additional participant took part in the study but was not included in the analysis as he was raised outside the U.S.

2.2. Stimuli

Five target sibilant environments were used for the study, with /s/ in pre-vocalic (sage) or preconsonantal environments (sprain, strain scrape) as well as prevocalic /ʃ/ (shade). Additional filler words included word-initial stops, nasals and approximants. A complete list of target and filler words is provided in Table 1.

<table>
<thead>
<tr>
<th>Targets</th>
<th>Fillers</th>
</tr>
</thead>
<tbody>
<tr>
<td>/s/</td>
<td>sage</td>
</tr>
<tr>
<td>/spr/</td>
<td>sprain</td>
</tr>
<tr>
<td>/str/</td>
<td>strain</td>
</tr>
<tr>
<td>/skr/</td>
<td>scrape</td>
</tr>
<tr>
<td>/ʃ/</td>
<td>shade</td>
</tr>
<tr>
<td></td>
<td>bait, brake, dame, drake, gaze, grape, jade, knave, lace, mace, rave</td>
</tr>
</tbody>
</table>

The target and filler words were placed in carrier phrases designed to contrast adjacency to an intonational phrase (IP) boundary. In the initial condition, the target word was the first word of a second sentence (e.g. I don’t know what he said to me. STRAIN or drake is maybe what he said). In the medial condition, the target word was preceded by a single word in the second phrase (e.g. I don’t know what he said. Maybe STRAIN or drake is what he said to me). It is worth noting that while IP-medial, the second carrier phrase was still intermediate phrase (IntP) initial, making the relative difference in the boundary between the two conditions small but controlled. Both carrier phrases were controlled for syllable count and the sounds preceding and following and the target words. Each target word and filler was also included in the second position of the coordinated structure (e.g. strain or drake and drake or strain). This secondary target was included for the naturalness of the carrier phrase but was not analyzed. Secondary targets were randomly paired with the primary targets and not all pairs were included for time considerations. The carrier phrases are reproduced in Table 1.

2.3. Procedure

Participants were seated in an isolated double-walled sound booth and were recorded on a Marantz Solid State PMD661 with a Shure SM10A head-mounted microphone. Participants were instructed to read the stimuli presented on the computer screen aloud at a normal speaking rate with no particular emphasis, stress, or focus. The stimuli were presented using PsychoPy [Peirce, 2008] in four blocks of 32. Each block contained each target word in both carrier phrases (2 × 16 = 32). The order of the stimuli within each block was randomized. The
Table 2: Design of stimulus sentences contrasting initial vs. medial sentence positions, with the target in capital letters, added here for clarity but presented to participants without emphasis.

<table>
<thead>
<tr>
<th>Initial</th>
<th>Medial</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I don’t know what he said to me.</em></td>
<td><em>I don’t know what he said.</em></td>
</tr>
<tr>
<td><em>STRAIN or drake is maybe what he said.</em></td>
<td><em>Maybe STRAIN or drake is what he said to me.</em></td>
</tr>
</tbody>
</table>

pace of the study was determined by the participants who advanced to the next trial by pressing the space key.

Recordings were manually checked for errors or disfluencies and subsequently forced aligned using FAVE (Rosenfelder et al., 2011) which determines phone-level boundaries using the HTK toolkit (Young, 1994) and the CMU American English Pronouncing Dictionary (Weide, 1998) to determine phonemic representations of words. Following forced alignment, phone-level boundaries for the target sibilants and intervening stops were manually corrected by the researcher.

2.4. Measurements

2.4.1. Spectral measurements

To characterize retraction, the centroid frequency (CF) and peak frequency (PF) of the sibilant were examined. Centroid frequency is the first of the four spectral moments (centroid frequency, standard deviation, skewness and kurtosis) and is frequently used to distinguish /s/ from /ʃ/ in English (Jongman et al., 2000; Shadle & Mair, 1996), and particularly in work examining English /s/-retraction (Baker et al., 2011; Gylfadottir, 2015; Mielke et al., 2010; Wilbanks, 2017). Peak frequency is another commonly used feature for distinguishing place of articulation in sibilants, but has been less often used in examinations of /s/-retraction (Rutter, 2011). For both measurements, /s/ is typically characterized by higher spectral energy than /ʃ/ due to the shorter oral cavity anterior to the constriction necessary for the production of /s/ relative to /ʃ/. Additionally, both spectral measurements have been shown to be highly variable depending on the speaker (Hughes & Halle, 1956) and their gender (Nittrouer, 1995; Stuart-Smith, 2007) and socio-economic class (Stuart-Smith, 2007). As /s/-retraction is a sound change in which /s/ approaches /ʃ/ in acoustic and articulatory realizations, a more retracted /s/ is expected to have a lower spectral energy than a less retracted /s/.

A custom Praat (Boersma & Weenink, 2001) script automatically extracted centroid frequency and peak frequency measurements for all intervals labeled as /s/ or /ʃ/ at eleven equidistant points (at 10% increments of the fricative’s duration from 0% to 100%) using 40 ms Hamming windows with preemphasis at 80 Hz and an examined frequency range from 500 to 12000 Hz. Measurement from the first and last 20% of the fricative were not included in the analysis.
The same script also measured the duration of the sibilant, the duration of the following stop, the word duration, and the labels of the following four segments.

While both spectral measurements were considered, ultimately, centroid frequency was selected as the measurement to best characterize /s/-retraction for this study. In the preliminary statistical analysis, there was no significant difference in model likelihood between the models using centroid frequency or peak frequency as the dependent variable. Due to the equal likelihood of the two models, a quadratic discriminant analysis (QDA) was performed to determine which measurement was a better predictor of the contrast. Using a subset of the data collected in this study, QDA classifiers were fit to the Gaussian distributions of prevocalic /s/ and /ʃ/, first using centroid frequency as the cue of contrast. The remainder of the data was then run through through the classifier to determine which distribution (/s/ or /ʃ/) a given measurement most likely belongs to. A second set of classifiers were fit using peak frequency. The QDA classifiers using centroid frequency outperformed those using peak frequency, exhibiting greater accuracy in discriminating prevocalic /s/ and /ʃ/, leading to the selection of centroid frequency as the measurement of analysis in this study.

Prior to modeling, CF values (in Hz) were normalized using z-scoring for all participants to allow for comparison of values between individuals and genders. To obtain the z-scored CF value, the mean and standard deviation for each participant’s /s/ was calculated across conditions (both prosodic and phonological) and timepoints. The formula for z-score normalization is provided in (1).

\[
CF_{z-scored} = \frac{\text{CF of segment} - \text{speaker mean CF of /s/}}{\text{speaker standard deviation of /s/}} \quad (1)
\]

2.4.2. Duration measurements

The duration of the sibilant was measured to examine whether prosodic lengthening occurs in IP-initial vs. medial positions in /sCr/ clusters. It was also included in the spectral analysis of retraction as an explanatory variable in order to examine the potential role of duration in conditioning spectral variation, as segments with longer duration have been shown to exhibit less coarticulation (Lindblom 1963). Sibilant duration was taken from the onset and offset of aperiodic frication as evident in the spectrogram.

The duration of the stop was also measured to examine whether prosodic shortening, due to the [− spread glottis] feature of stops following /s/, on the intervening stop can be observed in /sCr/ clusters in phrase initial position, following (Cho & Kim 2014) for /sC/ clusters. Stop duration was taken from the offset of frication to the onset of formant structures as evident in the spectrogram.

2.4.3. Gestural timing measurements

The c-center of each onset was calculated to examine the role that gestural timing may have on sibilant production, as sibilants nearer to the onset nucleus are thus closer to /r/ and expected to exhibit higher degrees of gestural overlap and thus more coarticulation. In the absence of articulatory data, the acoustic
signal can be used to calculate left-edge, c-center, and right-edge, following Selkirk & Durvasala (2013) and Ruthan et al. (2018). The distance from a given timepoint to the c-center was then calculated and included as an explanatory variable in the spectral analysis to examine the role of speech timing on observed coarticulation.

Prior to including c-center as an explanatory variable, the relativized standard deviation (RSD) was compared for all three temporal alignment measurements to confirm the c-center calculated acoustically best characterizes the coordination of English consonant onsets in this study. Using the manually-corrected forced-aligned textgrid, the duration from each measurement to an anchor point, the end of the vowel, was automatically calculated using a custom Praat script for all /r/, /Cr/ and /sCr/ clusters. Left-edge was calculated as the duration from the onset of the first consonant to the anchor and right-edge as the duration from the release of the /r/ to the anchor. C-center was calculated as the duration from the midpoint, in the case of simplex /r/ onsets, or mean of the midpoints of the onset consonants, in the case of /Cr/ and /sCr/ clusters, to the anchor point. RSD (SD/mean) was calculated separately for each place of articulation (/p/: rave∼brake∼sprain; /t/: rave∼drain∼strain; /p/: rave∼grape∼scrape). In all three places of articulation, c-center exhibited the lowest RSD, suggesting that it is the most consistent measurement of temporal alignment.

2.5. Statistical analysis

Data were analyzed with linear mixed-effects regressions which allow for the inclusion of both fixed and random effects. Models were created using the lmer() function from the lme4 (Bates et al., 2015) package in R (R Core Team, 2015). All p-values were obtained using normal approximation, following Mirman (2014).

3. Results

3.1. Duration analysis

The model for sibilant duration (ms) included fixed effects for trial order (Order, 1–128), the prosodic position of the target word (Position, IP-Initial vs. IP-Medial), and the identity of the following consonant (FollSegment, ∅ for prevocalic, and /p/,/t/,/k/). The inclusion of self-reported gender of the speaker and other demographic information, including participant age, ethnicity, sexuality or geographic origin, did not improve model likelihood. Order was scaled and centered at 0; Position was sum-coded; FollSegment was treatment-coded with ∅ as base. The model included by-subject random intercepts to account for interspeaker variability and by-subject random slopes for Position*FollSegment to allow for individual variation in the effect of this interaction. No by-item random intercepts or slopes were included, as there was only one unique word for each of the target environments (prevocalic (sage), /spr/ (sprain), /str/ (strain), and /skr/ (scrape)).
For the model on stop duration, only words with /sCr/ clusters were included in the analysis as there was no measurement to analysis in words with prevocalic /s/. The model for stop duration (ms) was identical model presented for sibilant duration with the distinction that the following segment (FOLLSEGMENT) was not a relevant measure, so stop identity (STOPIDENT) was selected instead. STOPIDENT was treatment coded with /p/ as base.

However, in neither the sibilant duration model nor the stop duration model was duration significantly modulated by phrase position. These findings suggest that the relatively small difference between intonational phrase-initial and intonational phrase-medial/intermediate phrase-initial is not great enough to manifest itself in prosodic lengthening or, in the case of the post-/s/ stop, possible prosodic shortening. In the following section, we turn to a spectral analysis of /s/ to examine possible prosodic strengthening and possible effects of duration, regardless of the phrase position, on retraction.

Table 3: Significant main effect and interaction summary statistics from the lmer model on CF. Model: CF ~ Order + (TimePoint + FOLLSEGMENT + Position + SDuration + CCDistance)^3 + (1 + Order + (Position * FOLLSEGMENT) | Speaker).

<table>
<thead>
<tr>
<th></th>
<th>Est.</th>
<th>SE</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.62</td>
<td>0.06</td>
<td>10.13</td>
<td>***</td>
</tr>
<tr>
<td>TimePoint</td>
<td>0.36</td>
<td>0.03</td>
<td>14.08</td>
<td>***</td>
</tr>
<tr>
<td>FOLLSEGMENT- /p/</td>
<td>-0.41</td>
<td>0.08</td>
<td>-5.41</td>
<td>***</td>
</tr>
<tr>
<td>FOLLSEGMENT- /t/</td>
<td>-1.12</td>
<td>0.10</td>
<td>-11.20</td>
<td>***</td>
</tr>
<tr>
<td>FOLLSEGMENT- /k/</td>
<td>-0.71</td>
<td>0.08</td>
<td>-8.55</td>
<td>***</td>
</tr>
<tr>
<td>TimePoint: FOLL- /p/</td>
<td>-0.20</td>
<td>0.03</td>
<td>-5.79</td>
<td>***</td>
</tr>
<tr>
<td>TimePoint: FOLL- /t/</td>
<td>-0.23</td>
<td>0.03</td>
<td>-6.74</td>
<td>***</td>
</tr>
<tr>
<td>TimePoint: FOLL- /k/</td>
<td>-0.24</td>
<td>0.03</td>
<td>-7.19</td>
<td>***</td>
</tr>
<tr>
<td>TimePoint: Pos-medial</td>
<td>0.39</td>
<td>0.03</td>
<td>11.47</td>
<td>***</td>
</tr>
<tr>
<td>SDuration: FOLL- /p/</td>
<td>0.13</td>
<td>0.04</td>
<td>3.04</td>
<td>**</td>
</tr>
<tr>
<td>SDuration: FOLL- /t/</td>
<td>0.18</td>
<td>0.05</td>
<td>3.75</td>
<td>***</td>
</tr>
<tr>
<td>SDuration: FOLL- /k/</td>
<td>0.12</td>
<td>0.05</td>
<td>2.62</td>
<td>**</td>
</tr>
<tr>
<td>Pos-medial: SDuration</td>
<td>-0.17</td>
<td>0.07</td>
<td>2.54</td>
<td>*</td>
</tr>
<tr>
<td>Pos-medial: SDuration</td>
<td>-0.20</td>
<td>0.06</td>
<td>-3.18</td>
<td>**</td>
</tr>
<tr>
<td>Pos-medial: SDuration</td>
<td>-0.13</td>
<td>0.07</td>
<td>-1.92</td>
<td>*</td>
</tr>
<tr>
<td>SDuration: CCDistance</td>
<td>0.08</td>
<td>0.02</td>
<td>2.95</td>
<td>**</td>
</tr>
<tr>
<td>SDuration: CCDistance</td>
<td>0.07</td>
<td>0.03</td>
<td>2.41</td>
<td>*</td>
</tr>
<tr>
<td>TimePoint: Pos-medial</td>
<td>-0.05</td>
<td>0.02</td>
<td>-2.82</td>
<td>**</td>
</tr>
<tr>
<td>TimePoint: Pos-medial</td>
<td>-0.09</td>
<td>0.02</td>
<td>-5.55</td>
<td>***</td>
</tr>
</tbody>
</table>

*p < 0.05; **p < 0.01; ***p < 0.001

3.2. Spectral analysis

The model for centroid frequency (CF, z-scored) of /s/ included fixed effects for trial order (ORDER, 1–128), the prosodic position of the target word
(POSITION, IP-Initial vs. IP-Medial), the timepoint over which the measurement was extracted (TIMEPOINT, from 3 (18.18%) to 9 (81.81%) of the total sibilant duration), and the identity of the following consonant (FOLLSSEGMENT, ∅ for prevocalic, and /p/, /t/, /k/). Additionally, the duration of sibilant (SDURATION, in ms) and the distance from the measurement timepoint to the c-center of the onset (CCDISTANCE, in ms) were included in the model. Like with the duration models, the inclusion of self-reported gender of the speaker and other demographic information, including participant age, ethnicity, sexuality or geographic origin, did not improve model likelihood.

All continuous variables were scaled and centered at 0, including ORDER, TIMEPOINT, SDURATION and CCDISTANCE. POSITION was sum-coded; FOLLSSEGMENT was treatment-coded with ∅ as base. The model included by-subject random intercepts to account for interspeaker variability and by-subject random slopes for POSITION*FOLLSSEGMENT to allow for individual variation in the effect of this interaction. No by-item random intercepts or slopes were included, as there was only one unique word for each of the target environments (prevocalic (sage), /spr/ (sprain), /str/ (strain), and /skr/ (scrape)).

Table 3 provides a summary of significant main effects and interactions in the centroid frequency analysis. Only significant predictors, defined as having a p-value less than 0.05, are reported.

The first step of the analysis is to examine the coarticulatory effects of the consonant cluster, determining if the results confirm the reported asymmetric distribution of /s/-retraction, in which higher degrees of retraction, i.e. lower centroid frequency values, are observed in /str/ clusters than /spr/ or /skr/ clusters. Figure 1 plots the z-scored centroid frequency for each target environment, including prevocalic /s/ and /S/ across all participants. Visual inspection of the data suggests that lowered centroid frequency values occur in all /sCr/ clusters compared to prevocalic /s/, with the lowest values in /str/ clusters. The results of the model support these observations, with a significant lowering effect of the following consonant in all /sCr/ clusters (for /p/: t = 5.41, p < 0.001; for /t/: t = −11.20, p < 0.001; for /k/: t = −8.55, p < 0.001). These findings provide evidence for the coarticulatory influence of /r/ in all clusters, with the highest degree of coarticulatory found in /str/ clusters.

As reported in Table 3, there is a main raising effect of timepoint across all clusters, suggesting that the centroid frequency of /s/ rises throughout the sibilant duration, making it more distinctive from /S/ over the course of its production (t = 14.08, p < 0.001). However, of particular interest to this study is how /sCr/ clusters differ from prevocalic /s/ in their temporal dynamics. Figure 2 plots the predicted centroid frequency for each target environment at seven timepoints over the course of its production. In Figure 2 the relative rising over the course of the sibilant can be observed in all /s/ environments, but /sCr/ clusters appear to taper off or even lower over the final third of their examined production in contrast to prevocalic /s/ which continues its rise. These observations are confirmed in the results of the model, with all three clusters showing a lowering effect of the interaction of timepoint and following consonant cluster (for /p/: t = −5.79, p < 0.001; for /t/: t = −6.74, p < 0.001;
for /k/: $t = -7.19, p < 0.001$). These findings suggest that the lowering effect of the consonant clusters increases throughout the duration of the sibilant further separating preconsonantal /s/ from prevocalic /s/. This provides evidence for the role of locality in retraction, suggesting that as the temporal distance from /r/ decreases, the degree of coarticulation observed increases.

An additional aim of the experiment was to manipulate adjacency to a prosodic boundary to examine if retraction is resisted in phrase-initial positions in the same way it is in word-initial positions (Gylfadottir 2015; Wilbanks 2017). Firstly, as previously mentioned, IP-initial sibilants did not exhibit significantly longer durations, contra Cho & Kim (2014). If segments were
to lengthen word-initially, then the hypothesis predicts that they would resist coarticulation and exhibit higher centroid frequency. Regardless of lengthening, segments would also be expected to maximize contrast at prosodic boundaries, further contributing to higher centroid frequency values. However, no main effect of prosodic position was observed, either in lengthening or strengthening (see Section 4 for a discussion as to why). Despite the lack of main effect of phrase position, Figure 3 suggests that initial and medial positions appear to be distinguished over the course of the sibilant production. However, Figure 3 suggests centroid frequency raising, i.e. less retraction, in phrase-medial positions relative to phrase-initial positions rather than the predicted inverse. Furthermore, Figure 3 shows that phrase-medial /s/ production is more dynamic and parabolic than phrase-initial trajectories and with more variance in the predicted values. This account is confirmed by the model ($t = 11.47, p < 0.001$), suggesting that the main centroid frequency raising effect of timepoint is further enhanced in phrase-medial positions.

Although duration was not successfully manipulated between phrase-initial and medial instances of /s/, the duration of the sibilant was nonetheless a critical factor in conditioning the observed retraction. Figure 4 plots the predicted centroid frequency values for the target environments for sibilants of differing durations. While the predicted mean of prevocalic /s/ and /ʃ/ are shown to be characterized by a relatively flat slope, suggesting little influence of segment duration, the /sCr/ clusters all exhibit increased centroid frequency values as the duration of the segment increases. This is confirmed by the model, suggesting that in /sCr/ clusters, longer durations encourage higher centroid frequency value, i.e. less retraction (for /p/: $t = 3.04, p < 0.01$; for /t/: $t = 3.75, p < 0.001$; for /k/: $t = 2.62, p < 0.01$). These results are in keeping with predictions for the role of duration, finding that increased duration effectively decreases the
Figure 4: CF values for each phonological environment as a function of sibilant duration. All CF values are z-scored. Values for /ʃ/ are shown for comparison but were not included in the model.

cocartulatory power of the adjacent consonants, here the stop-rhotic clusters.

Duration further interacts with phrase position and following segment in conditioning centroid frequency of /s/. In Figure 5, the phrase-initial (left panel) instances of /sCr/ exhibit the same effect of duration as the interaction of duration and target environment pictured in Figure 4 with visibly more positive slopes; that is, sibilants in /sCr/ clusters with a longer duration exhibit higher centroid frequency values than segments with a shorter duration. However, the phrase-medial (right panel) instances of /sCr/ do not appear to exhibit as strong an effect, suggesting that the coarticulatory resistance of increased duration is stronger in initial than medial positions. This observation is supported by the model, with the interaction of sibilant duration, phrase position and following segment having a dampening effect on centroid frequency in medial positions (for /p/: t = -2.54, p < 0.05; for /t/: t = -3.18, p < 0.01; for /k/: t = -1.92, p < 0.05). These findings suggest that the coarticulatory resistance of increased segment duration is weaker phrase-medially than phrase-initially. Furthermore, while the interaction of phrase position and following consonant cluster was not significant in the pooled model, the inclusion of by-subject random slopes for that interaction, which significantly improved model likelihood, hints to further individual differences in how phrase position affects /s/ production in consonant clusters. This observation is explored further in Section 4.2.

The duration of the sibilant further interacts with the distance from the c-center and the identity of the following consonant. In /spr/ and /skr/ clusters, when the duration of the sibilant is longer and the examined timepoint is further from the c-center for the onset cluster, suggesting that there is less overlap between the clusters, the sibilant is more likely to resist coarticulation. This is manifested in a raising of the centroid frequency of /s/ (for /p/: t = 0.08, p < 0.01; for /k/: t = 0.07, p < 0.05). However, this resistance to coarticulation was
not observed in /str/ clusters ($t = 0.01, n.s.$), suggesting that timing relations are more critical in conditioning coarticulation in environments in which weaker coarticulation is expected.

Finally, sibilant duration and distance from the c-center both further interact with phrase position and timepoint across all the different phonological environments. Lower centroid frequency values were observed phrase-medially at later timepoints when the duration is sibilant is longer than for the same timepoint and phrase position when the duration of the sibilant is shorter. This is manifested in a lowering interaction effect of sibilant duration, phrase position and timepoint ($t = -2.82, p < 0.01$). Similarly, lower centroid frequency values were observed phrase medially at later timepoints when the distance from the c-center is greater. This is manifested in a lowering interaction effect of c-center distance, phrase position and timepoint ($t = -5.55, p < 0.001$).

4. Discussion

While the highest degree of retraction was observed in /str/ clusters, this study also provides clear empirical evidence for retraction in /spr/ and /skr/ clusters as well, with /s/ in all clusters exhibiting lower centroid frequency values compared to prevocalic environments. Furthermore, in no cluster could the sibilant be described as canonically /ʃ/: instead, gradient forms intermediate between prevocalic /s/ and /ʃ/ were observed, in line with (Baker et al., 2011) for all /sCr/ clusters and numerous findings of intermediate forms in /str/ clusters (Labov, 2001; Durian, 2007; Mielke et al., 2010; Gylfadottir, 2015), in contrast to the categorical findings of Rutter (2011). These intermediate forms begin to challenge an account of retraction as assimilation, as phonological reanalysis would predict more canonical instances of /ʃ/ in environments conditioning
retraction. Further support for a coarticulatory account is proposed in Section 4.1 through a discussion of the temporal dynamics of sibilant production in /sCr/ clusters. And finally, special attention is given to individual variation in /s/ production in Section 4.2.

4.1. Temporal dynamics of retraction

The results of this study provide strong evidence for a coarticulatory model of /s/-retraction in which the extent and degree of retraction are dependent on temporal factors, manifested in the effects of four distinct predictors: the duration of the sibilant, the timepoint over the course of the sibilant production at which measurement was taken, the distance from that timepoint to the c-center, and the prosodic position of the sibilant. Although these factors are distinct on the surface, they all fundamentally speak to the relative strength of the coarticulatory trigger – /r/.

As predicted, sibilants in all /sCr/ clusters with longer durations exhibited higher centroid frequencies, effectively decreasing retraction, with the strongest effect in /str/ clusters. This effect of duration strongly supports a coarticulatory account of retraction, as less coarticulation is expected in segments with longer durations. Similarly, as predicted, the later the timepoint – and thus the closer to /r/ – the lower the centroid frequency for all /sCr/ clusters. This sheds light on the highly local nature of retraction, despite acting across a intervening stop, further supporting a coarticulatory account of retraction.

In much the same way, the distance from the timepoint to the c-center of the onset characterizes not just the distance from /r/, but also the relative timing of the /s/ and the /r/. While distance from the c-center did not appear to significantly condition the observed retraction on its own, it emerged as a significant predictor in its interaction with the duration of the sibilant, with increased retraction observed further from the c-center for /spr/ and /skr/ clusters. Surprisingly, the same effect was not observed for /str/ clusters, suggesting that the relative timing of the consonants plays a greater role in conditioning coarticulation in environments where coarticulation is less prevalent, like in /spr/ and /skr/ clusters. This may effect not have emerged in /str/ because coarticulation is expected regardless of the relative timing of the consonants or because the possibility that affrication on the intervening stop may have masked any such effect.

While a coarticulatory account of retraction predicted increased retraction in phrase-medial positions, both as a result of prosodic lengthening and strengthening in which hyperarticulation is observed at boundaries following (Fougeron & Keating, 1997), there was no interaction of prosodic position and cluster identity. Furthermore, sibilant duration was not successfully conditioned by adjacency to a prosodic boundary, with no difference observed in the length of sibilants between the IP-initial and medial positions. Taken together, these findings may suggest that the contrast between the prosodic positions was not great enough to manifest evidence for prosodic strengthening, since the positions analyzed as IP-medial can also be described as IntP-initial, and thus still
adjacent to a relatively structurally high prosodic boundary despite being comparatively lower than IP-initial positions. Nonetheless, prosodic strengthening can be observed in the interaction of prosodic position with duration and target clusters, with the resistance to retraction, as a result of increased sibilant duration, weakened in phase-medial positions. In other words, more retraction is expected phrase-medially regardless of the duration of the sibilant, while retraction in initial positions is more dependent on sibilant duration. Furthermore, the lack of a general effect of prosodic strengthening of retraction may be due to the incredible amount of individual variation observed in this interaction, examined in depth in Section 4.2.

Taken together, these findings illustrate that /s/-retraction is a gradient, dynamic process that is heavily conditioned by factors that influence the relative distance from the sibilant to the rhotic, suggesting that it is best characterized as a coarticulatory rather than assimilatory process. Furthermore, if this account is embraced, this study provides further evidence that coarticulation is a planned, phonologically-constrained, language- and speaker-specific process [Mannel 1990] and cannot be viewed through a strictly biomechanical lens.

4.2. Individual Variation

The community level findings of gradient intermediate forms, visualized in Figure 1, support a coarticulatory model of retraction, which is further evidenced through an examination of the individual patterns different speakers exhibit in their production of /s/ in these environments. Figure 6 plots each individual’s mean retraction ratio, rather than z-scored CF, for each /sCr/ cluster. The retraction ratio [Mielke et al. 2010], provided in 2, outputs a number between 1 and 0, where 0 represents that a given value is identical to that individual’s mean prevocalic /s/ and a 1 represents that a given value is identical to that individual’s mean prevocalic /ʃ/.

\[
\text{Retraction Ratio} = \frac{\text{CF of segment} - \text{speaker mean CF of } /s/}{\text{speaker mean CF of } /ʃ/ - \text{speaker mean CF of } /s/}
\]  

As Figure 6 illustrates, most individuals exhibit a retraction ratio between 0 and 0.5 for all clusters, suggesting intermediate forms closer to /s/ than /ʃ/. For four individuals, the observed retraction ratio for /str/ is greater than 0.75. Additionally, this study replicates the sizable gap between 0.6 and 0.8 observed by [Mielke et al. 2010], which correlates with their perceptual categorization of speakers as retractors or non-retractors (in their definition, a retractor is an individual whose /str/ cluster would be perceived as /ʃtr/ by a trained listener). The replication of this gap may be indicative of a naturally occurring distinction between retraction as a result of coarticulation and phonologized sound change. It is also noteworthy that very few participants in the study would be categorized as retractors, especially given the relative youth of the participants recruited.

The results of this study point to a tremendous degree of interspeaker variation that can contribute to better understanding the nature of the actuation and propagation of the sound change. Many models of sound change propose that a
sound change begins when a speaker produces extreme coarticulation which is then misinterpreted by the listener as a new speech target (Ohala, 1993, i.a.). Baker et al. (2011) examine /s/-retraction as a key to understanding sound change actuation, as it is naturally biased toward interspeaker variability because English /t/ is articulated with a stable degree of variation due to the lack of a perceptual distinction for listeners between the bunched and retroflex varieties. The present study expands upon the findings of Baker et al., examining individual variation not just in phonological environment but also in prosodic position.

Despite the lack of main effect for phrase position, the interaction of prosodic position and phonological environment shows a high degree of inter-speaker variation. The results of the model presented in Section 3 capture community-level observations, while accounting for individual variation by the inclusion of by-subject random slopes and intercepts. In Figure 7, the random slopes for phrase position, phonological environment and their interaction are plotted for each individual, providing a visual representation of the reality of individual variation. Participants are sorted by their conditional mode for by-subject random intercept for the centroid frequency of /s/, displayed in the top left panel. The remaining two panels in the top row display the random slopes for trial order or phrase position, with some individuals showing significant deviations from the mean for phrase position in both directions. The middle three panels display the random slopes for following segment (/p/, /t/ and /k/), showing the individual variation with respect to the effect of the following consonant. Of particular note are the random slopes for /t/ (FOLLSEGMENT3), highlighting the increased inter-speaker variation in the context of /t/ compared against the other consonant clusters. The visualization of the first six panels suggest that less variation is observed in the effect of prosodic position than phonological context and almost no visible variation in the effect of trial order. The remaining three panels on the bottom row illustrate the by-subject random slopes for the interaction of prosodic position and following segment, which improved model likelihood despite not reaching significance in the group model. Again, like the variation observed in the context of /t/, these panels demonstrate the individual
Figure 7: Caterpillar plot for the mixed effects model, with by-subject random intercepts and by-subject random slopes for phrase position (\textit{Position}: 2 = phrase-medial), phonological environment (\textit{FollSegment}: 2 = /p/; 3 = /t/; 4 = /k/), and their interaction. Participants are ordered by their conditional mode for by-subject random intercepts.

variability that this interaction has on the degree of retraction and captures the deviations that the group-level analysis cannot. The caterpillar plot illustrates the nature and magnitude of this variation, with some subject showing a mild effect, suggesting smaller deviations from the group norm, and a few individuals showing a robust effect in contrast to the group norm.

To better understand the nature of individual variation beyond the random slopes in Figure 7, separate linear regressions were run on each speaker’s produc-
Figure 8: CF for /s/ and /ʃ/ in different phonological environments and phrase positions for three individuals, illustrating the by-subject variation in the interaction of phrase position and following segment. The left and center panels illustrate individuals with no significant effect of the interaction. The right panel shows an individual with a significant effect of the interaction of position, timepoint and following segment in all three clusters.

As running additional linear regressions on each speaker decreases the power of each model and increases the likelihood of type I errors, these models were only interpreted to supplement the findings of the pooled model. The linear regressions were identical in form to the pooled model, but by nature excluded all random effects. The results of three participants exhibiting distinctive patterns are visualized in Figure 8. Speaker 9701 (left) exhibits no effect significant effect of position, target, timepoint or their interaction. Speaker 9604 (center) exhibits a clear effect of phonological environment in /str/ clusters but not in /spr/ or /skr/ clusters, but shows no effect of timepoint, position or its interaction with phonological environment. Speaker 9603 (right) show a significant interaction of timepoint, phonological environment and phrase position.

While Speaker 9603 would not be described as an across-the-board retractor,
as her /str/ production phrase-medially overlaps with her prevocalic /s/ production, yielding a mean retraction ratio of 0.08, her /str/ production phrase-initially is not only significantly different from her prevocalic /s/ and her phrase-medial /str/ values but is closer to her mean prevocalic /ʃ/, yielding a mean retraction ratio value of 0.67. Furthermore, lowered centroid frequency values are observed phrase-initially in /spr/ and /skr/ as well, to the extent that some of those tokens were perceived by the researcher to be [ʃ]. Only prevocalic /s/ and /ʃ/ do not vary between phrase-initial and medial positions. In this way, her /sCr/ production seems to be completely modulated by prosodic position rather than the place of the following consonant, with canonical /s/ articulations phrase-medially and intermediate articulations between /s/ and /ʃ/ phrase-initially.

Despite the group level effect that inhibits retraction phrase-initially in sibilants with longer durations, a significant degree of individual variation is shown in the interactions of phonological environment and phrase position in both directions, indicating increased retraction initially for some speakers and medially for others. And it stands to reason that its speakers like 9603 who exhibit increased retraction phrase-initially, despite the trend in the opposite direction and findings of increased retraction word-medially rather than word-initially, may be the actuators of sound change. Specifically, these extreme coarticulatory values of /s/ are observed phrase-initially, precisely where they are most prominent and ripe for target reanalysis. The extreme coarticulatory values exhibited by speakers like 9603 in /spr/ and /skr/ clusters suggest that /s/-retraction as a sound change may progress beyond /str/ clusters.

5. Conclusion

The present study demonstrates that /s/-retraction is a dynamic process, changing in response to the strength of coarticulatory triggers. As the relative distance between the examined timepoint and the /ʃ/ decreases, the degree of retraction can be seen to increase. Furthermore, retraction is observed not just in /str/ clusters, but also to a lesser extent in /spr/ and /skr/ clusters, with those environments exhibiting the same temporal dynamics albeit at an overall baseline of less retraction. These temporal factors that condition retraction provide evidence to support a coarticulatory, rather than assimilatory, account of /s/-retraction in contrast to the earliest accounts of the phenomenon.

This study also demonstrates the observed retraction is dependent not just on the phonological context of /s/, but on also on its prosodic position. While no main effect of prosodic position was observed, adjacency to an intonational phrase interacted subtly with the other predictors to suggest that retraction is resisted in phrase-initial positions. This finding, parallel to previous findings for resistance to retraction word-initially, suggests that prosodic strengthening in /sCr/ clusters manifests itself as hyperarticulation toward the canonical prevocalic /s/, enhancing the phonological contrast between /s/ and /ʃ/. Furthermore, a tremendous degree of individual variation was observed between
the thirty participants recruited, including participants who not only exhibited increased retraction phrase-initially, but also participants for whom the prosodic position was more meaningful in conditioning the sibilant realization than the phonological environment. Building on proposals that sound change actuation results from the misinterpretation of individual variation and extreme coarticulation, the individual patterns of increased retraction phrase-initially serve not just as the ideal loci for target misinterpretation for /str/ clusters due to their prosodic prominence, but also for potential environments for sound actuation in /spr/ or /skr/ clusters, building off the observed prerequisite coarticulatory variation in the clusters by other participants.

Acknowledgements

I would like to thank Alan C. L. Yu for his guidance, feedback, and encouragement. I am also grateful for the support of Laura Staum Casasanto, Josh Falk, Katie Franich and Susan Lin. Finally, I would also like to thank audiences at LabPhon 16, MidPhon 21, NWAV 45 for their feedback on different stages of this work.

References


