Phonological environment and the social evaluation of American English sibilants

JACOB B. PHILLIPS
Phonology Laboratory, University of Chicago

1 Introduction

Many speakers of English exhibit /s/-retraction, the process by which /s/ is produced or perceived approaching /ʃ/ in the context of /u/, especially in /stu/ clusters. Thus for a retractor, the word street is pronounced approaching shtreet. This has been robustly observed across different dialect groups of American English, including New York/Long Island (Kraljic 2008), Philadelphia (Gylfadottir 2015), Raleigh, North Carolina (Wilbanks 2017), Columbus, Ohio (Durian 2007), and southern Louisiana (Rutter 2011). It is so widespread that it has been referred to as a ‘general American innovation’ (Shapiro 1995), although it has additionally been observed outside the U.S., including Newfoundland, Canada (Clarke 2004), New Zealand (Lawrence 2000), England (Bass 2009, Glain 2014) and Scotland (Stuart-Smith et al. 2018). Unlike typical sound changes, there does not appear to be a gender effect in /s/-retraction production (Gylfadottir 2015), although women were found to lead the change in Raleigh (Wilbanks 2017).

Furthermore, while retraction is observed in /stu/ clusters across these different dialects of English, the same retraction is not observed in /spu/ or /sku/ clusters, either in frequency or rate of retraction, such that scream is rarely pronounced approaching shcream and spree is rarely pronounced approaching shpree (Shapiro 1995, Baker et al. 2011, Gylfadottir 2015). This asymmetric distribution poses a challenge to many accounts of the sound change, such as those that rely on overlapping gestures between the /s/ and the /ʃ/, predicting retraction in /spu/ clusters as the intervening /p/ has no associated lingual gesture. The observed asymmetry has received some attention examining the phonetic nature of the dis-

*This work would not have been possible without the contributions and invaluable research assistance of Hillel Steinmetz. I would like to thank Alan C. L. Yu, Laura Staum Casasanto and audiences at NWAV 46 for their advice, feedback and encouragement. I would also like to thank Alex Kramer for technical support and Giovanna Hooton and Paige Resnick for their help running participants in the laboratory. Funding in part provided by the University of Chicago Graduate Aid in Linguistics fund and the Sapir Scholarship program at the University of Chicago.
tribution, for example proposing that the affrication of /t/ is the driving force of retraction (Lawrence 2000), and similarly in phonological accounts, proposing that the underspecification of a place feature of /t/ allows for the assimilation across /t/ but not /{p, k}/ (Shapiro 1995). Yet little work has sought to examine the socioindexical nature of this variable and whether that index varies by phonological context.

While most of the studies examining /s/-retraction have focused primarily on production data, a small but growing body of work has looked at the perception and social evaluation of the phenomenon. For example, listeners have been shown to not shift their /sl/-/ʃ/ category boundaries after exposure to retraction in /stʃ/ clusters, like district, but do shift those boundaries if retraction is observed in unexpected environments, like colosseum (Kraljic et al. 2008), suggesting an attention to ongoing sound changing and/or compensation for coarticulation. Recent research in Austin, Texas found that /s/-retraction is below the level of speakers’ consciousness, with most speakers unaware of the phenomenon when questioned in interviews (Hinrichs et al. 2015). When prompted to give social meaning to the phenomenon, speakers provided diverse responses, often suggesting that it may be more Southern/hick, non-native, or the result of speech impediment. However, no studies to date have examined listener’s implicit social evaluation of /s/-retraction, either in /stʃ/ or /s{p, k}ʃ/ clusters.

In contrast, much work in sociolinguistics has rigorously examined the social evaluation of predominantly prevocalic sibilant variation in English. This indexical relationship has been found on various dimensions, with a fronted /s/ typically perceived as more middle class, feminine (Stuart-Smith 2007) and gay (Levon 2014). Similarly, a retracted /s/ has been demonstrated to be perceived as more rural/country (Campbell-Kibler 2011, Podseva & Van Hofwegen 2014), masculine (Zimman 2013) and lesbian (Podseva & Van Hofwegen 2014). Additionally the socioindexical meaning of /s/ variation has been found to vary by social context, indexing sexuality and gender in some contexts but not others (Pharao 2014). However, as these studies of sibilant variation generally focus on prevocalic /s/, they are largely agnostic to any role that the phonological environment may play in determining the socioindexical meaning of /s/.

The present study consists of a social evaluation ratings task that seeks to examine the socioindexical meaning of /s/ in /s{p, t, k}ʃ/ clusters in order to better understand the
complicated nature of the sound change. Specifically, this study asks if /s/ in /st\(\hat{u}\)/ clusters contributes different socioindexical meaning than /s/ in /s\{p, k\}\(\hat{u}\)/ clusters that can provide a better understanding of the asymmetrical phonological distribution as well as the lack of an expected gender effect in many communities.

2 Methods & materials

Stimuli: Eight model talkers (4 male, mean age 20) were recruited to record the auditory stimuli. All model talkers were native speakers of American English (four California, four Illinois) and received either payment or credit for an introductory linguistics course. The model talkers were seated in an isolated double-walled sound booth and were recorded on a Zoom H4n recorder with a Shure SM10A head-mounted microphone as they read a series of stop- and sibilant-initial lexical words off the computer screen in the carrier phrase: *Please say string again. String string string.* The wordlist for the sibilant-initial stimuli is provided in Table 1. The eight model talkers showed varying degrees of retraction in /st\(\hat{u}\)/ clusters, with one male and one female exhibiting forms categorically perceived by the researchers to be /\(\hat{S}\)/ rather than /s/, i.e. the retractors.

Table 1: Wordlist for stimuli creation

<table>
<thead>
<tr>
<th></th>
<th>/s</th>
<th>/sC</th>
<th>/sC(\hat{a})</th>
<th>/(\hat{S})/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/</td>
<td>sit</td>
<td>spit</td>
<td>spritz</td>
<td>shit</td>
</tr>
<tr>
<td>/t/</td>
<td>sing</td>
<td>sting</td>
<td>string</td>
<td>shin</td>
</tr>
<tr>
<td>/k/</td>
<td>sip</td>
<td>skip</td>
<td>script</td>
<td>ship</td>
</tr>
</tbody>
</table>

To create the /s\{p t k\}\(\hat{u}\)/ stimuli for the ratings task, the sibilant onsets from the corresponding minimal pairs were mixed digitally mixed using a Praat script (Darwin 2005) at four scaling ratios: 0%-/s/:100%-/\(\hat{S}\)/, 60%-/s/:40%-/\(\hat{S}\)/, 80%-/s/:20%-/\(\hat{S}\)/, and 100%-/s/:0%-/\(\hat{S}\)/. The scaling ratios were selected to include the extreme endpoints (0% and 100% /s/) and more naturalistic retraction values (60% and 80% /s/). Thus, for the word *script* the initial /s/ was removed and cross-spliced with the an onset mixed from *sip* and *ship*. Separate onset continua were then created for *spritz* and *string*. Stop-initial stimuli were included as fillers but were not manipulated. All files were normalized for intensity.
Participants: 342 (208 males, mean age 30) geographically diverse (113 urban, 179 suburban, 50 rural) native speakers of American English were recruited on Mechanical Turk and were paid for their participation in this study. Gender, age, and geographic environment were self-reported by the participants. No participants reported speech or hearing disorders/abnormalities. 41 additional participants were recruited but excluded from the analysis due to technical difficulties or non-attentive responses.

Procedure: In order to identify the attributes to include in the social evaluation ratings task, an initial open-ended panel study was conducted following on Amazon Mechanical Turk with 15 (10 male) native speakers of American English. Panel participants heard the unmanipulated sibilant-initial stimuli presented in Table 1 for each model talker and were asked to supply any information about the speaker’s identity, characteristics or traits in an open-ended question. For the male retractor, panel participants were split in describing his /st/ tokens as more masculine, athletic and straight than his other /s/ tokens or whether those same tokens were more gay, intellectual, and pretentious. The female retractor did not elicit strong changes in panel participants’ responses in her /st/ clusters, which were typically described as younger. Following the panel study seven attributes were selected including attractiveness, masculinity, friendliness, formality, shyness and sexuality. Additionally, geographic regions and environment types were included.

Table 2: Sample survey matrix
Please provide your impression of the speaker on the scales below:

| Attractive | ◦ ◦ ◦ ◦ ◦ ◦ ◦ ◦ Unattractive |
| Masculine  | ◦ ◦ ◦ ◦ ◦ ◦ ◦ ◦ Not at all masculine |
| Unfriendly | ◦ ◦ ◦ ◦ ◦ ◦ ◦ ◦ Friendly |
| Casual     | ◦ ◦ ◦ ◦ ◦ ◦ ◦ ◦ Formal |
| Shy        | ◦ ◦ ◦ ◦ ◦ ◦ ◦ ◦ Outgoing |
| Heterosexual | ◦ ◦ ◦ ◦ ◦ ◦ ◦ ◦ Homosexual |

Where does it sound like the speaker might be from? (select all that apply):

☐ The South   ☐ The Midwest   ☐ The Northeast
☐ West coast ☐ East coast ☐ The Southwest
☐ The City ☐ The Country ☐ The Suburbs
In the social evaluation ratings task, participants were assigned to one of the four retraction conditions (0%, 60%, 80% or 100% /s/) for one of the three clusters (/sp$/, /st$/ or /sk$/$), in order to allow for the different clusters to be rated and judged independently. Participants heard a total of 24 audio clips, with 8 targets and 16 fillers. Participants completed the task online and were instructed to use headphones. Each audio clip played once automatically and participants could optionally replay the clip as many times as they chose. After listening to each clip, participants were asked to rate the talker on each of the selected attributes presented on seven-point scales. The order and polarity of the attributes were randomized across participants, but held constant between trials for the same participant. Participants were additionally asked to select radio buttons corresponding to geographic categories. The order of the radio buttons was constant across participants and trials. A sample survey matrix is provided in Table 2.

Analysis: Linear mixed effects regression models were fitted separately for the different attributes (z-scored by participant), using the \texttt{lmer()} function of the \texttt{lme4} package in \texttt{R}, but only results for SEXUALITY and MASCULINITY will be presented here as those responses were most strongly conditioned by /s/ variation. Separate models were run for the male and female model talkers to allow for gender-specific differences in the evaluation of the different attributes. The models were constructed using a nested approach, incrementally adding variables as either fixed or random effects. Variables that improved model likelihood in any of the models were included in the models for the other attributes or gender. The fixed effects included in the analysis were WORD (SPRITZ, STRING, SCRIPT; treatment coded with STRING as base), retraction CONDITION (0, 60, 80, 100; scaled) and SPEAKER (F1, F2, F3, F4, M1, M2, M3, M4; treatment coded with F1 and M1 as base). Two-way interactions were included between the fixed effects. Random intercepts were included for TRIAL (scaled) and LISTENER. The state of origin of the model talkers as well as self-reported demographic categories of the listener (e.g. age, gender, geographic information) did not significantly improve model likelihood. To further examine the role of individual variability captured by SPEAKER, separate linear regressions were run on each model talker with WORD and CONDITION as independent variables.
Figure 1: Predicted sexuality (z-scored, y-axis) by model talker (x-axis), condition (color) and word (panel). A positive predicted sexuality value indicates a more HOMOSEXUAL response, while a negative value indicated a more HETEROSEXUAL response.

### 3 Results

**Sexuality:** For male model talkers, there was no main effect of WORD, CONDITION or the interaction of the two variables at the community level. There was a significant effect of SPEAKER, with model talker M3 more likely to be rated by listeners as more heterosexual ($\beta = -0.53, p < 0.001$) and model talker M4 as more homosexual ($\beta = 0.26, p < 0.05$).

For female model talkers, like male model talkers, there was no significant effect of WORD or CONDITION. There was a significant effect of SPEAKER, with model talkers F2 and F4 more likely to be evaluated as more homosexual (for F2: $\beta = 0.63, p < 0.05$; for F4: $\beta = 0.98, p < 0.01$). Figure 1 provides the predicted SEXUALITY ratings for each of the model talkers in the endpoint conditions (0 and 100% /s/), with listeners exhibiting significant influences of CONDITION or WORD in the evaluation of model talkers F4 and M4, both more likely to be rated as more homosexual, but not for the other model talkers.

Model talker M4, a male from Illinois, was significantly more likely to receive more homosexual ratings than the other males in the study. Additionally, M4 alone shows any
effect of WORD and CONDITION in the individual linear regression. For M4, there is a
significant interaction of WORD and CONDITION with listeners significantly more likely
to give a more heterosexual response in increased retraction conditions in the /sp/ and /sk/
clusters than in the /st/ clusters (for /p/: $\beta = -0.39, p < 0.001$; for /k/: $\beta = -0.36, p < 0.01$). Figure 2 illustrates listener’s predicted sexuality responses for M4 in all conditions.

Model talker F4, a female from Illinois, was also significantly more likely receive more
homosexual ratings than other female model talkers in the study. Like M4, F4 was the lone
female talker to show significant effects of WORD and CONDITION in the individual linear
regression. The observed main effect of CONDITION suggests that F4 was more likely to
receive to more heterosexual responses in increased retraction conditions ($\beta = -0.10, p < 0.05$). The additional interaction of WORD and CONDITION counteracts the main effect of
CONDITION, suggesting that she received more homosexual ratings in retracted conditions
for /sp/ clusters compared to /st/ clusters ($\beta = 0.32, p < 0.05$). Although it was trending
in the same direction for /sk/ clusters, it was not significant ($\beta = 0.20, p = 0.13$). Figure
3 illustrates listener’s predicted sexuality responses for F4 in all conditions.
Masculinity: For male model talkers, there was an observed main effect of Speaker, with model talkers M2 and M4 rated as less masculine (for M2: $\beta = -0.26, p < 0.05$; for M4: $\beta = -0.44, p < 0.001$). There was additionally a main effect of Word with /sk\r/ clusters being rated by listeners as less masculine than /st\r/ clusters ($\beta = -0.26, p < 0.05$). The interaction of Word and Condition suggests that listeners are significantly less likely to give more masculine responses to /sk\r/ clusters in increased retraction conditions than /st\r/ clusters ($\beta = -0.13, p < 0.05$). However, these effects were counteracted by the interaction of Speaker and Word, with speaker M3 more likely to receive more masculine responses in /sk\r/ words ($\beta = 0.33, p < 0.05$).

For female talkers, there was again a significant effect of Speaker, with model talkers F2 and F4 more likely to be evaluated as more masculine (for F2: $\beta = 0.21, p < 0.05$; for F4: $\beta = 0.50, p < 0.001$). Like for the male talkers, there was also a significant interaction of Word and Condition, suggesting that listeners are significantly more likely to give more masculine responses to /sk\r/ clusters in increased retraction conditions than /st\r/ clusters, in contrast to the effect for male talkers ($\beta = 0.13, p < 0.05$).

Model talker M1, a male from California, was consistently rated as above average for
masculinity compared to the other model talkers. M1 exhibited a main effect of WORD, with listeners significantly more likely to rate his /sku/ clusters as less masculine than his /stu/ clusters ($\beta = -0.27, p < 0.01$). This effect was further amplified by the interaction of WORD and CONDITION, as listeners were more likely to rate his /sku/ clusters as less masculine in increased retraction conditions than his /stu/ clusters ($\beta = -0.21, p < 0.05$).

Again, model talker F4, significantly more likely to be rated as more masculine than the other female model talkers, stood out in the individual linear regressions as the only female model talker to exhibit significant effects of WORD or CONDITION. The main effect of WORD suggests that listeners perceived /sku/ clusters produced by F4 to be more masculine than her /stu/ clusters ($\beta = 0.32, p < 0.01$). This effect is reinforced by the interaction of WORD and CONDITION, as F4’s /sku/ clusters are significantly more likely to be rated as more masculine in increased retraction conditions as compared to her /stu/ clusters ($\beta = 0.21, p < 0.05$).
4 Discussion

The goals of this study were to examine the ways in which phonological environment of /s/, that is the different clusters /s{p, t, k}o/, contribute to the social meaning of /s/-retraction, and conversely, the ways in which social evaluation of /s/ in these clusters can shed light on the asymmetric distribution of /s/-retraction. The results of the ratings task suggest that phonological environment plays a nuanced role in how listeners evaluate the sexuality and masculinity of different speakers.

For the evaluation of perceived sexuality, not only do individual participants vary significantly in how they rate the model talkers, but also individual speakers vary consistently in how they are rated. For speakers who are less likely to be perceived as homosexual, the phonological environment and the degree of retraction of /s/ do not further condition the evaluation as more or less homosexual. However, for speakers more likely to be perceived as gay or lesbian, the interaction of the environment and the degree of retraction significantly influence listeners’ perceptions of the speaker’s sexuality. For male speakers (specifically M4), they are more likely to be perceived as more straight in /sp\o/ and /sk\o/ clusters in increased retraction conditions. This suggests that a retracted /s/ in these clusters has the expected socioindexical meaning that a retracted /s/ carries in prevocalic environment, that is, more straight (for men). However, in /st\o/ clusters, increased retraction does not contribute to perceived straightness. The same pattern, albeit in reverse, is observed in female speakers more likely to be perceived as lesbian (F4): Increased retraction in /sp\o/ clusters contributes to a stronger perceived homosexuality than in /st\o/ clusters. Taken together, these results suggest that a retracted /s/ specifically in /st\o/ clusters does not index straightness (for men) or gayness (for women).

For the evaluation of perceived masculinity, the results for female talkers align with the perceived sexuality results. That is, a retracted /s/ in /sk\o/ clusters is perceived as more masculine than a retracted /s/ in /st\o/ clusters. This effect emerged individually as well for F4, a woman perceived as more lesbian and more masculine than her peers. However, the results for male speakers illustrate the reverse pattern, with a retracted /s/ in /sk\o/ clusters being evaluated as less masculine than a retracted /s/ in /st\o/ clusters. This pattern also emerged individually for M1, a male perceived as more masculine than his peers. Taken
together these results can be interpreted in a few different ways. Firstly, it may be the case that a retracted /s/ in /skɑ/ clusters indexes gender atypicality. Or secondly, it may be that the dampening of socioindexical value of /stɑ/ is of more consequence in speakers who are perceived to be more atypical on the relevant attribute, that is, (perceived) gay speakers for SEXUALITY and women for MASCULINITY. This hypothesis could be tested as a future direction of this work by having listeners only rate male or female talkers.

The weaker indexical role of a retracted /s/ in /stɑ/ clusters may be a result of listeners’ experience with the sound change. Many models of sound change propose that a change emerges when listeners stop compensating for extreme coarticulation and instead encode a new speech target (Ohala 1993, Harrington et al. 2008). Thus for /s/-retraction, the sound change occurs specifically because listeners stop compensating for coarticulation in /stɑ/ clusters. However, the results of this study suggest the reverse pattern – that speakers are compensating more for coarticulation in /stɑ/ clusters than in /s{p, k}ɑ/ clusters, possibly due to their experience with retraction in /stɑ/ clusters. Thus, as speakers are compensating for the retraction in /stɑ/ clusters, /s/ variability is less available to listeners as a sociolinguistics variable in that environment. Conversely, due to their rareness, a retracted /s/ in /spɑ/ or /skɑ/ clusters are not as strongly compensated for and thus can be assigned stronger indexical weight. It may be that this pattern is emerging because most of the listeners are not phonologized retractors and thus may still be expected to perceptually compensate for retraction, and as the sound change progresses, younger generations will compensate less and /stɑ/ variation may take on a stronger socioindexical role.

Furthermore, the prevalence of /s/-retraction in female speakers, whether they are reported to lead the sound change (Wilbanks 2017) or participate at equal levels (Gylfadottir 2014), may also contribute to the weakening of associations of masculinity and sexuality with a retracted /s/ in /stɑ/ clusters, but not in /spɑ/ or /skɑ/ clusters. Future examinations of production data may shed light on whether a gender effect is in fact observed in /s{p, k}ɑ/ clusters, with men exhibited greater numbers of retracted forms, albeit not to the degree of /stɑ/.
5 Conclusion

The results of this study suggest that a retracted /s/ in /spô/ and /skô/ clusters indexes a more masculine and straight (for men) or lesbian persona, as expected from previous research on a retracted /s/ in prevocalic environments. However, in contrast, a retracted /s/ in /stu/ clusters has weaker indexical associations with these attributes, especially in individuals perceived to be less typical for a given attribute. This suggests that /s/-retraction in /stu/ clusters is not strongly indexing masculinity, straightness (for men), or lesbianness. This study also found a tremendous degree of individual variability both with respect to how listeners evaluated the attributes and how the talkers were evaluated. Future studies that include more talkers with more robust social information about those talkers may shed more light on the socioindexical role of /s/-retraction.

REFERENCES


CLARKE, SANDRA. 2004. A note on several unusual fricative pronunciations on the southwest coast of Newfoundland. Regional Language Studies... Newfoundland 18. 15–17.


HARRINGTON, JONATHAN, FELICITAS KLEBER, and ULRICH REUBOLD. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern


