Masculine toughness and the categorical perception of onset sibilant clusters

Jacob B. Phillips and Paige Resnick
Phonology Laboratory, University of Chicago, Chicago, Illinois 60637, USA
jbphillips@uchicago.edu, presnick@uchicago.edu

Abstract: Evidence from production studies demonstrate that /s/-retraction, a sound change in American English where /s/ approaches [ʃ], is progressing in apparent time and highly context-dependent, occurring overwhelmingly in /str/ clusters. This study reports findings from a phoneme categorization task, suggesting that listeners perceive sibilants in /str/ clusters less categorically than in /spr/ and /skr/ clusters and are becoming less categorical in apparent time for all clusters. Indicators of masculine toughness, including the model talker’s voice and face and the listener’s endorsement of masculine stereotypes, additionally contribute to a less categorical perception of sibilants in these environments.

© 2019 Acoustical Society of America
[BCS]
Date Received: April 23, 2019 Date Accepted: May 31, 2019

1. Introduction
An on-going sound change in many dialects of English, /s/-retraction is the process by which /s/ approaches [ʃ] in /str/ clusters (such that street may sound like ststreet) due to the coarticulatory influences of /r/, but less so in /spr/ and /skr/ clusters (scream rarely sounds like shcream). While a large body of work has examined the production of /s/-retraction (i.e., Baker et al., 2011; Gylfadottir, 2015; Shapiro, 1995; Wilbanks, 2017), less work has focused on its perception (cf. Kraljic et al., 2008; Stevens and Harrington, 2016). The present study reports findings from a phoneme categorization task consisting of nonce words in American English in which the onset sibilant in /sCr/ clusters—where C is taken to represent any voiceless stop /p, t, k/—was replaced with a step on a continuum from /s/ to /ʃ/, with more /ʃ/-like steps perceptually corresponding to an increased degree of retraction.

First, we examine the role of phonological environment, asking whether listeners exhibit different perceptual patterns by onset cluster in the same way that they exhibit different production patterns by onset cluster. We ask if their categorization strategies may vary between /str/ clusters on one hand and /skr/ and /spr/ clusters on the other, potentially accounting for the increased expectation for retraction in /str/ clusters. In numerous studies, listeners have been shown to compensate for coarticulation, shifting their contrast boundaries to recover the intended target (Mann and Repp, 1980). However, it is not immediately clear what role the contrast between /s/ and /ʃ/ plays in the perception of these clusters, as only /s/ is phonotactically licit preceding consonants other than /r/. Thus, it is possible that listeners may compensate for retraction in the traditional sense of a shifted response curve, or they may perceive sibilants less categorically in these environments. Regardless, the contrast between /s/ and /ʃ/ is neutralized in all three clusters, so if any context-dependent patterns of categorization are observed, they may be interpreted as a response to the increased expectation of coarticulation in /str/ clusters.

Second, we investigate apparent time trends in the categorization patterns of these clusters. Examinations of speech corpora have demonstrated that /s/-retraction is progressing in apparent time in /str/ clusters, with younger speakers producing more retracted variants than older speakers (Gylfadottir, 2015; Wilbanks, 2017). The present study asks if a parallel apparent time progression can be observed in perception. This builds off proposals that a sound change emerges when listeners do not compensate for extreme coarticulation, but rather encode a new speech target, which may persist in later productions even if the coarticulatory trigger becomes redundant or is lost altogether (e.g., Ohala, 1993). In the present study, we ask if listeners potentially perceive sibilant retraction as a cue for /r/ presence and thus perceive /ʃ/-like sibilants as /s/ in environments where coarticulation is anticipated.

a)Author to whom correspondence should be addressed.
Finally, we compare how different representations of masculinity influence the categorization of these clusters in order to better understand the potential socio-indexicality of retraction. Previous research has demonstrated that regardless of phonological content, a male’s fronted /s/ is evaluated as more gay and less masculine (Campbell-Kibler, 2011; Levon, 2014) while a backed, i.e., retracted, /s/ as more country and in some cases more masculine (Campbell-Kibler, 2011; Podesva and Van Hofwegen, 2014). While no research to our knowledge has examined the relationship between masculinity or sexuality and the production of /s/-retraction in /str/ clusters, Phillips (2018) found in a social evaluation task that retraction in /sCr/ clusters does not index a more masculine or straight identity for most talkers; however, a talker evaluated as more gay across the board was evaluated as more straight with a retracted /s/ in /spr/ and /skr/ clusters, but not in /str/ clusters. In categorization tasks like the present experiment, listeners have been shown to attune to these relevant social characteristics, accounting for that information with adjusted category boundaries (e.g., Strand, 1999). Furthermore, listeners’ attitudes about the relevant social characteristics, like their relative endorsement of gender stereotypes, contribute to their perceptual strategies and can strengthen the effect of these social factors (Campbell-Kibler, 2019). In the present study, we ask how listeners use community-defined indicators of masculinity, like the varied talker and co-present face, in the categorization of /sCr/ clusters and how these may differ from individually-defined factors like their relative endorsement of masculine stereotypes.

2. Methods and materials

2.1 Stimuli

Two white, male speakers from Iowa (age 19 and 21) recorded the auditory stimuli for this study, which consisted of nonce words with initial /SCr/ clusters (sprimbIe, shrimblIe, strimbIe, shrimbIe, scrimbIe, shcrimbIe), where the onset sibilant /S/ was any step on a continuum from /s/ to /S/. Nonce words were selected due to the phonotactic restrictions of English, neutralizing the contrast between /s/ and /ʃ/ preconsonantally. To create the continuum from /s/ to /ʃ/, the sibilant onsets from the prevocalic equivalents, simblIe and shimblIe, were extracted and modified to contain identical durations by the removal of random portions of the sibilant at zero crossings. Using a modified praat script originally created by Darwin (2005), the two sibilants were added at seven specified scaling ratios, provided, along with the centroid frequency measurements for each step, in Table 1. Each of the seven steps was cross-spliced onto the preconsonantal target word, creating a continuum from /s{p,t,k}rImbIal/ to /ʃ{p,t,k}rImbIal/. Both model talkers’ stimuli were paired with faces controlled for perceived attractiveness and normed as relatively more or less masculine than average from the Chicago Face Database (Ma et al., 2015).

2.2 Participants

200 (120 male, 161 white, 185 heterosexual) geographically diverse native speakers of American English participated online through Amazon Mechanical Turk and were paid $1 for an approximately five-minute study. The mean age of the participants was 34 (minimum 19, 1st quartile 28, median 33, 3rd quartile 40, maximum 60). All participants were using an internet connection located in the U.S. and reported being born in the U.S. and raised in a household using English as its primary language until age 12, which were collectively used as a metric for determining native American English status. An additional 38 individuals participated in this study but were excluded from analysis due to non-native status, non-attentive responses, or self-reported speech/hearing disorders.

Table 1. Scaling ratios and centroid frequency (CF) measurements for each onset stimuli.

<table>
<thead>
<tr>
<th>Step</th>
<th>Scaling ratio</th>
<th>CF: Talker 1</th>
<th>CF: Talker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95%/s/:5%/ʃ</td>
<td>7535 Hz</td>
<td>6649 Hz</td>
</tr>
<tr>
<td>2</td>
<td>80%/s/:20%/ʃ</td>
<td>7510 Hz</td>
<td>6621 Hz</td>
</tr>
<tr>
<td>3</td>
<td>65%/s/:35%/ʃ</td>
<td>7310 Hz</td>
<td>6302 Hz</td>
</tr>
<tr>
<td>4</td>
<td>50%/s/:50%/ʃ</td>
<td>6604 Hz</td>
<td>5978 Hz</td>
</tr>
<tr>
<td>5</td>
<td>35%/s/:65%/ʃ</td>
<td>5549 Hz</td>
<td>5088 Hz</td>
</tr>
<tr>
<td>6</td>
<td>20%/s/:80%/ʃ</td>
<td>4220 Hz</td>
<td>4562 Hz</td>
</tr>
<tr>
<td>7</td>
<td>5%/s/:95%/ʃ</td>
<td>3516 Hz</td>
<td>3683 Hz</td>
</tr>
</tbody>
</table>
2.3 Procedure

Participants heard 42 trials (7 steps × 3 consonant clusters × 2 model talkers = 42 trials), plus fillers interspersed to check for attentiveness. Participants were randomly assigned to a condition that included either faces normed as more masculine for both talkers, less masculine for both talkers, or one more and one less masculine face. Participants were instructed that the faces presented were pictures of the talkers. After each trial, participants responded with a key press corresponding to the orthographic representations presented on the screen, e.g., sprimble or shprimble, which were piloted for naturalness. Following the phoneme categorization task, participants evaluated the faces and voices used in the experiment on nine-point Likert scales for masculinity, toughness and attractiveness. Finally, participants completed a short survey, containing basic demographic questions as well as the Male Role Attitude Scale (MRAS, Pleck et al., 1993), which is a standardized psychological instrument designed to discern an individual’s relative endorsement of various stereotypes of masculinity. Specifically, we focus on toughness subscale, as defined by two questions: A guy will respect if he talks about his problems and A young man should be physically tough, even if he’s not big. The MRAS and the toughness subscale specifically have been used as explanatory variables in social evaluation (Levon, 2014) and phoneme categorization (Campbell-Kibler, 2019) of sibilants.

3. Results

Listeners’ responses (/s/ = 0 or /ʃ/ = 1) were modeled with logistic mixed effects regressions using the glmer() function in the lme4 package (Bates et al., 2015) in R (R Core Team, 2015). The fixed effects predictors included in the model were trial order (Order: 1–42, scaled), continuum step (Step: 1–7; scaled), onset cluster identity (Cluster: /Str/, /Spr/, /Skr/; Helmert-coded), participant’s age (Age: 19–60, log transformed to approximate normal distribution: 2.94–4.10; scaled), the participants toughness rating for the voice (VoiceToughness: 1–9, scaled), the participant’s toughness rating for the displayed face (FaceToughness: 1–9; scaled), and their own relative endorsement of male stereotypes of toughness from the MRAS survey (ToughnessEndorsement: 2–8; scaled). Other self-reported demographic information, including participant gender, sexuality, and location, did not improve model likelihood and were not included in the final model. A preliminary model was designed with all two- and three-way interactions between the fixed effects predictors and all interactions that did not reach a significance threshold of 0.05 were pruned from the final model. Additionally, a preliminary model included a maximally specified random effects structure, but failed to converge. The model reported is a result of progressive simplification of the random effects structure until convergence was reached, which includes random intercepts for participant with by-subject random slopes for Order and Cluster. The complete model in lme4 format is: Response ~ Order + Step * Age * ToughnessEndorsement + Step * Cluster * ToughnessEndorsement + Step * Cluster * FaceToughness + Step * Cluster * VoiceToughness + (1 + Trial | Subject) + (1 + Cluster | Subject).

The intercept of the model emerged as significant ($z = -7.09, p < 0.001$), suggesting that all else being equal, listeners are more likely to categorize the onset sibilant as /s/ than /ʃ/. This may be interpreted as a consequence of phonotactic restrictions of sibilants preconsonantly which neutralizes the phonological contrast between /s/ and /ʃ/ in these environments, allowing participants to answer /s/ at significantly higher rates than other phoneme categorization tasks. Similarly, this may be attributed to English orthography, with the /sCr/ sequences more common and acceptable than the /shCr/ sequences.

A main effect of Step was observed ($z = 21.27, p < 0.001$), which illustrates an increase in /ʃ/ responses as the scaling ratio of /ʃ/ increases, suggesting that listeners tend to perceive onset sibilants in consonant clusters categorically. The interaction of Step and Cluster is illustrated in Fig. 1(a). Visual inspection of the figure indicates that more /ʃ/ responses are predicted at higher, i.e., more /ʃ/-like, steps for /Spr/ and /Skr/ relative to /Str/ clusters, while no visually detectable differences in responses between the clusters can be observed at lower, i.e., more /s/-like, steps. This interaction emerged as significant in model, with fewer /ʃ/ responses predicted at higher steps in /Str/ clusters compared to /Spr/ or /Skr/ clusters ($z = 8.23, p < 0.001$). Additionally, fewer /ʃ/ responses were then predicted at higher steps in /Skr/ clusters compared to /Spr/ clusters ($z = 7.85, p < 0.001$). These findings suggest that listeners are the least categorical in their perception of /Str/ clusters and the most categorical in the perception of /Spr/ clusters.
No significant main effect of Age was observed, suggesting that all else being equal there are no shifts in the phonotactic bias against /ʃ/ responses in apparent time. However, as the interaction of Step and Age illustrates in Fig. 1(b), the youngest listeners appear to give fewer /ʃ/ responses than older listeners at higher, i.e., more /ʃ/-like, steps. The interaction of Step and Age in the model supports this observation ($z = 8.14, p < 0.001$), suggesting that younger listeners are less categorical in their perception of onset sibilant clusters than older listeners. The interaction between Cluster and Age did not reach the level of significance and was pruned from the final model, suggesting that younger listeners are not accounting for potential coarticulation at different rates than older listeners.

The role of masculine stereotypes in phoneme categorization was examined through three different variables: the model talker’s voice, the face presented, and the participant’s relative endorsement of masculine stereotypes of toughness. All faces normed by the Chicago Face Database (Ma et al., 2015) as more masculine were evaluated by our participants as tougher and more masculine than the faces normed as less masculine. One model talker was also consistently evaluated by our participants as tougher and more masculine than the other talker. However, while the FaceToughness and ToughnessEndorsement did not emerge as significant as main effects in the model, a main effect was observed for VoiceToughness ($z = 2.41, p < 0.05$), with higher toughness ratings of the voice generally predicting more /ʃ/ responses regardless of Step, i.e., regardless of how /ʃ/-like the stimulus is. Contrariwise, the interactions of each of these predictors of masculine stereotypes with Step, i.e., how /ʃ/-like the stimulus is, predict fewer /ʃ/ responses: Faces perceived as tougher (Step:FaceToughness: $z = -3.17, p < 0.01$), voices perceived as tougher (Step:VoiceToughness: $z = -1.97, p < 0.05$), and participants who more strongly endorse masculine stereotypes of toughness (Step:ToughnessEndorsement: $z = -3.92, p < 0.001$) all predict fewer /ʃ/ responses at higher, i.e., more /ʃ/-like, steps. This suggests that despite the across-the-board increase in /ʃ/ responses for tougher voices, all three indicators of masculine toughness serve to decrease the categoricity of sibilant perception in /SCr/ clusters, potentially speaking to the socio-indexicality of the onset sibilant, regardless of cluster.

Furthermore, ToughnessEndorsement also emerged as significant in its interaction with Cluster and Step ($z = -1.97, p < 0.05$), with fewer /ʃ/ responses predicted at higher steps for individuals who more strongly endorse masculine stereotypes of toughness in /Spr/ and /Skr/ clusters than /Str/ clusters. After accounting for the observation that listeners in general are less categorical in /Str/ clusters than /Spr/ and /Skr/ clusters and the observation that listeners who more strongly endorse masculine stereotypes are less categorical than their less stereotype-focused peers, the interaction of ToughnessEndorsement, Cluster and Step demonstrates that listeners who more strongly endorse masculine stereotypes exhibit an additional decrease in categoricity in /Spr/ and /Skr/ clusters relative to /Str/ clusters compared to listeners who less strongly endorse masculine stereotypes. This observation may suggest that the potential masculine socio-indexicality of pre-consonantal sibilants is more strongly linked to /Spr/ and /Skr/ than /Str/ clusters, as it is the listeners who may be attuning to performances of masculine toughness that are accounting for that masculine toughness in their categorization of onset sibilants, but specifically in /Spr/ and /Skr/ clusters where /ʃ/-retraction is least expected.
Using a heat map to more clearly illustrate the interaction between continuous
variables, Fig. 2 plots the model predictions for the interaction of ToughnessEndorsement,
Step, and Age. In the right panel, the youngest listeners who less strongly endorse tough-
ness stereotypes are more categorical than listeners with more strongly endorse such ster-
etotypes. In the center panel, there is no noticeable difference for median-aged listeners
depending on their relative endorsement of toughness stereotypes. And, in the right-
hand panel, older listeners are predicted to contrast with younger speakers, as older lis-
teners who more strongly endorse toughness stereotypes are predicted to exhibit more
categoricity than those who less strongly endorse toughness stereotypes. This observa-
tion is supported by the model, with the interaction of ToughnessEndorsement, Step,
and Age ($z = 5.20, p < 0.001$) predicting more /ʃ/ responses at higher steps for older par-
ticipants who more strongly endorse masculine stereotypes of toughness. This observa-
tion may suggest an apparent time change in the socio-indexicality of pre-consonantal
sibilants, with younger speakers who are more attuned to performances of masculinity
accounting for that in their perception.

4. Discussion

The first aim of this study was to determine whether listeners exhibit different categori-
ization strategies for onset sibilants depending on the identity of the following stop,
that is, are some clusters perceived more categorically than others? The findings of this
study suggest that listeners are less categorical in their perception of /Str/ clusters than
/Spr/ or /Skr/ clusters, continuing to perceive /s/ 50% of the time at even the final step
of the continuum in which the onset sibilant is 95% composed of a prevocalic /ʃ/ token.
In contrast, listeners perceive /ʃ/ around 75% of the time for /Spr/ and /Skr/
clusters at the final step, which, while notably higher than /Str/ clusters, is still much
lower than would be expected for a traditional prevocalic /s/-/ʃ/ categorization task.
This illustrates that despite clear differences in categorization strategies, for no cluster
are these onset sibilants perceived truly categorically: At no step are /s/ or /ʃ/ perceived
approaching 100%; rather in /Str/ clusters, onset sibilants are simply perceived addition-
ally less categorically than in /Spr/ and /Skr/ clusters. This lack of true categorical
perception is a result of the phonotactic restrictions of English that necessitated the
nonce word task to begin with: There is no contrast between /s/ and /ʃ/
preconsontanally and thus there is no true category boundary, allowing listeners to
perceive /s/ at all steps. However, the difference between the consonant clusters cannot
be explained by phonotactics alone, as the same lack of contrast exists for all clusters.
Rather, the decreased categoricity in /Str/ clusters may be a compensation for coarticu-
lation of sorts, in which listeners are less categorical in environments where coarticu-
lation is most expected. Crucially, compensation here looks different than may be
expected: Rather than the canonical shifted s-curve observed for other contrasts,
including prevocalic /s/-/ʃ/, perceptual compensation in /Skr/ clusters manifests itself as
decreased categoricity, and thus a general flattening of the response curve, especially at
higher, i.e., more /ʃ/-like, steps. This suggests that due to their experience with extreme
coarticulation in /Str/ clusters, listeners are more likely to accept [ʃ] as a production of
/s/ in /Str/ clusters compared to /Skr/ and /Spr/ clusters.

Second, we asked whether younger listeners, who presumably produce more
/s/-retraction than older listeners (Gylfadottir, 2015; Wilbanks, 2017), are less

Fig. 2. (Color online) Predicted /ʃ/ responses (hue) by step (x axis), relative toughness endorsement (y axis) and
age (panels). Age is binned (younger, middle, older) for data visualization, but analyzed as continuous. A lighter
tile indicates a stronger predicted /s/ response and a darker tile indicates a stronger predicted /ʃ/ response. A
more pronounced horizontal gradient indicates stronger categoricity.
categorical in their perception of these clusters than older listeners. The findings of this study confirm this hypothesis, finding that in general, younger participants are less categorical than older participants in their perception of the onset sibilant in all /SCr/ clusters. However, as no significant interaction was observed between age, cluster identity and continuum step, there is no evidence that younger listeners are more or less categorical in their perception of /Str/ clusters relative to /Spr/ or /Skr/ clusters. This suggests that while younger listeners are less categorical across the board than older listeners, the relative difference in categoricity between the clusters has remained constant in apparent time. That is, just as younger listeners are less categorical in their perception of /Str/ clusters than older listeners, so too are they less categorical in /Spr/ and /Skr/ clusters. This has implications for the trajectory of /s/-retraction if perception and production patterns progress in tandem, suggesting as younger speakers continue to produce more /s/-retraction in /str/ clusters, /skr/ and /spr/ clusters may continue to advance behind them as the new loci of the change. As the present study does not examine participants’ production in these clusters, this remains an avenue for future investigation.

Finally, we asked how different indicators of masculine stereotypes of toughness influence the categorization of onset sibilants in /SCr/ clusters in order to better understand the potential socio-indexicality of sibilant variation in these environments. The findings of this study suggest that indicators of toughness do play a role in phoneme categorization of these sibilants, with most, but not all, effects predicting more /s/ responses. This suggests that listeners, especially listeners who may be more attuned to subtle performances of masculinity, attribute a more /ʃ/-like onset sibilant to a performance of masculine toughness and account for that performance to recover the intended /s/ target. Furthermore, while listeners are on average less categorical in /Str/ clusters than /Spr/ and /Skr/ clusters due to their familiarity with coarticulation in those clusters, they are less categorical in /Spr/ and /Skr/ clusters than /Str/ clusters when accounting for potential performances of masculine toughness. This suggests that a retracted /s/ in /Str/ clusters less strongly indexes masculinity than a retracted /s/ in other environments, including other onset clusters. These results suggest that /s/-retraction as an ongoing sound change is not primarily about performing masculinity and toughness, which corroborates the social evaluation findings from Phillips (2018) in which a stereotypically gay talker was evaluated as more heterosexual with a retracted /s/ in /Skr/ and /Spr/ clusters, but not in /Str/ clusters. This is not to say, however, that /s/-retraction as a sound change is purely coarticulatory in nature and does not carry social meaning. Rather, more research is necessary to better understand precisely how listeners perceive variation and innovation in these environments and what meaning they assign such variation, including possible associations with broader categories, like class and ethnicity, or smaller traits, like outgoingness. In particular, we stress that the socio-indexical meaning of /s/ in these consonant clusters may not carry the associations expected from research on prevocalic sibilants.

Acknowledgments

Thank you to Alan C. L. Yu, Laura Staum Casasanto, and audiences at NWAV 47 and LSA 2019 for their comments and feedback.

References and links


