Indexical weight is distributed: Evidence from social evaluations of /s/

Grace Brown, Meghan Sumner & Robert J. Podesva*

Abstract. Though it is well established that listeners can infer social meanings from linguistic variation, less is known about how listeners use properties of the linguistic signal to form such indexical relationships from moment to moment. We examine this question using the well-documented association between /s/ center of gravity (CoG) and perceived masculinity. Using a matched guise paradigm, we ask how two properties of the phonetic signal, the F2 of the vowel following /s/ and speaker voice, modulate listeners' masculinity evaluations of sibilant acoustics. Furthermore, we investigate whether this modulation depends on the variability of the /s/ tokens listeners hear within the experimental context by comparing listeners who heard a single invariant /s/ token per categorical /s/ CoG condition (fronted, mid, backed) with those who heard /s/ tokens that varied with the phonological environment. Bayesian modeling confirmed that high-CoG (fronted) /s/ is typically perceived as less masculine than low-CoG (backed) /s/, replicating prior findings. Critically, masculinity judgments were primarily influenced by speaker voice, with /s/ CoG and F2 of the following vowel serving as secondary, modulating cues to speaker masculinity. These findings suggest that social evaluation reflects integration of the full available signal rather than extraction of a single variable.

Keywords. sociolinguistics; perception; phonetics; masculinity; /s/

1. Introduction. Language users possess structured knowledge about the linguistic patterns of groups of speakers and of individuals. This knowledge allows us, as perceivers of language, to infer social meaning from linguistic variation, mapping linguistic forms onto social categories, performances, and stances (Eckert 2008; Silverstein 2003; Foulkes & Docherty 2006). This link is exemplified by the well-documented association between sibilant acoustics and perceived gender characteristics: realizations of /s/ with a high center of gravity (CoG) are associated with female voices, femininity, and non-normative masculinity (Munson 2007; Campbell-Kibler 2011; Boyd et al. 2021). Yet, form-meaning mappings are not fixed; social meanings of /s/ can extend to concepts beyond gender (Stuart-Smith 2007; Munson 2007; Munson et al. 2006; Linville 1998) or may not map onto gender at all (e.g., Calder & King 2022), and gender can be performed and inferred through other linguistic variables, such as pitch characteristics (Leung et al. 2021; Levon 2014; Munson 2007; Terango 1966) or use of creaky voice (Yuasa 2010). Given this complexity, the inferring of a particular social meaning from a linguistic variant is necessarily shaped by the context in which it occurs (Eckert 2008; Silverstein 2003). This principle has been well-established in sociolinguistic theory, and empirical work on register (Pharao et al. 2014) and listener ideologies (Levon 2014) has shown that social and discursive context shapes which indexical relationships listeners activate. What remains less well understood is the mechanism by which phonetic signal properties guide social inference. The present study asks: how do elements of context that are always perceivable in the linguistic signal (i.e., speaker voice and phonological context) contribute to our understanding of social meaning mappings?

* Authors: Grace Brown, Stanford University (grcbrown@stanford.edu), Meghan Sumner, Stanford University (sumner@stanford.edu) & Robert J. Podesva, Stanford University (podesva@stanford.edu).

1.1. VARIATION IS CENTRAL TO SPEECH PERCEPTION. Listeners are highly sensitive to linguistic and acoustic variation both within and across speakers, and they use it to process upcoming words, craft detailed acoustic memories, and adapt to new speakers. Palmeri et al. (1993) demonstrated that listeners recognize previously heard words more quickly and accurately when repeated in the same voice than in a different voice, providing evidence that talker-specific acoustic detail is encoded in memory. Speaking rate and amplitude are similarly retained, and familiarity with a particular voice improves perceived intelligibility of speech in noise (Bradlow & Pisoni 1999). These talker-specificity effects are well accounted for by exemplar-based theories of speech perception. For instance, Goldinger (1996) demonstrated that episodic memory traces for spoken words retain acoustic details, showing that voice-specific priming effects persisted across study-test delays and were sensitive to fine-grained acoustic similarity between different-voice repetitions. Building on this, Goldinger (1998) proposed a fully episodic theory of lexical access in which the mental lexicon consists of accumulated perceptual traces rather than abstract representations, with each stored episode preserving the social and phonetic details of its instantiation. Johnson (1997) developed a parallel exemplar model specifically addressing talker variability, arguing that speech perception proceeds without abstract speaker normalization and that incoming speech is matched against stored exemplars that retain, for instance, coarticulatory and indexical detail alongside phonological category information. Together, these frameworks converge on the view that talker-specific properties are not filtered out during perception but are retained and actively used.

Listeners' sensitivity to variation extends to the coarticulatory structure of the signal. Early evidence for this comes from Mann & Repp (1980), who showed that listeners perceive more instances of /s/ than /ʃ/ when a fricative is followed by the rounded vowel /u/ than by /a/, interpreting this pattern as perceptual compensation for the known coarticulatory effect of rounded vowels on preceding fricatives. This finding established that listeners use the acoustic properties of a following vowel to inform their interpretation of an adjacent fricative. Beddor et al. (2013) extended this line of evidence to anticipatory coarticulation in real time, using a visual-world paradigm to show that listeners exploit anticipatory vowel nasalization to project upcoming phonological content before it is acoustically realized and that this sensitivity is gradient and individually variable. Finally, Bradlow & Bent (2008) demonstrated that listeners adapt rapidly to talker-specific phonetic patterns and that they can generalize that adaptation across talkers sharing an accent. They argued that this pattern indicates a perception system that is simultaneously attuned to individual voice properties and capable of abstracting across talkers. Together, this body of work establishes that listeners have the perceptual resources to use both speaker voice and coarticulatory F2 as phonetic context in social evaluation.

1.2. GENDERED PERCEPTIONS OF /S/. Within the indexical field framework (Eckert 2008), the high-CoG (or fronted) /s/ and low-CoG (or backed) /s/ variants are each associated with a constellation of potential social meanings, many of which are related to gender identity and performance. The indexical field of high-CoG /s/ links this variant to female voices, femininity, and for male speakers, non-normative masculinity (Munson et al. 2006; Munson 2007). Campbell-Kibler (2011) demonstrated that English listeners consistently activate this mapping between high-CoG /s/ and non-normative masculinity mapping when evaluating speakers, rating those who produce high-CoG /s/ as less masculine across multiple speakers. Boyd et al. (2021) extended this finding to show that the mapping is stable for English listeners across variation in speaker language.

Evidence from Danish illustrates that the indexical field for high-CoG /s/ contains multiple potential mappings and that which one is activated depends on linguistic context. Pharao et al. (2014) found that a high-CoG /s/ variant activated a non-normative masculinity mapping only when it appeared in a register already associated with that social meaning; in a register linked to streetwise heterosexual masculinity, the same variant activated no such evaluation. This finding is precisely what the indexical field framework predicts: the same form has multiple potential meanings, and sociolinguistic context selects among them. The question the present study pursues is whether phonetic context, specifically the F2 of the following vowel and the properties of the speaker's voice, operates as an analogous selection mechanism within the indexical field for /s/.

1.3. MATCHED GUISE. The matched guise technique is frequently implemented by isolating a single phonetic variable by cross-splicing tokens into otherwise identical frame utterances, licensing the attribution of rating differences to the manipulated feature alone (Lambert et al. 1960). The findings reviewed in Section 1.2 rest substantially on this logic, and it has substantially advanced our understanding of social meanings associated with /s/ and other linguistic variables. However, the method has an inherent structural limitation that bears directly on the theoretical question we pursue: by holding the phonetic context constant across guises, it cannot reveal how that context shapes which mapping listeners activate. The matched guise paradigm, by design, tests form-meaning mappings under conditions that eliminate the phonetic variation through which context-selective inference is assumed to normally operate.

Levon (2014) provides evidence that this limitation has empirical consequences. Using a modified matched guise paradigm designed to vary listener beliefs about speaker identity, Levon investigated how ideologies about gender, sexuality, and class in British English shaped listeners' social perceptions of mean pitch, /s/ spectral characteristics, and TH-fronting. He found that listeners' gender ideologies simultaneously enabled certain form-meaning associations and blocked others, such that the same phonetic variant (e.g., high pitch and high-CoG /s/) could be rendered socially meaningful or socially inert depending on the ideological beliefs of the listener. Together with (Pharao et al. 2014), this work shows that both social and discursive context modulate which social meanings are activated by a particular linguistic form. The present study extends this line of inquiry to phonetic context, asking how specific, recoverable properties of the acoustic signal, such as the F2 of the vowel following /s/ and speaker voice, function as the context through which listeners use /s/ to inform judgments of speakers' masculinity.

1.4. PHONETIC CONTEXT: SOURCES AND PREDICTIONS. We focus on two properties of the phonetic signal that are always present and therefore always inferable from the linguistic signal. The first is the F2 of the vowel immediately following /s/. Through coarticulation, /s/ tokens preceding front vowels (high F2) tend to have higher CoG than those preceding back vowels (low F2), such that CoG and F2 covary in a systematic, phonologically predictable way (Soli 1981). This covariation suggests that F2, as an acoustic reflection of vowel frontness, is a property of the acoustic signal that partially determines how front an /s/ is perceived to be. F2 constitutes a phonetic context that could bear on /s/ CoG-to-social-meaning mapping listeners activate: a high-CoG /s/ preceding a front vowel is phonetically expected, while the same CoG value preceding a back vowel is marked, which could invite social inference. Listeners who are sensitive to this coarticulatory relationship may therefore weight the same CoG value differently depending on the F2 of the following vowel.

The second property is speaker voice. Because no two voices are identical, and because lis-

teners encode talker-specific acoustic detail episodically (Palmeri et al. 1993; Bradlow & Pisoni 1999), the properties of a speaker's voice constitute a phonetic context that is always available when /s/ is evaluated. Furthermore, a speaker's voice carries its own information about gender and masculinity, which can either converge with or diverge from the gender information typically carried by a high or low CoG /s/.

1.5. RESEARCH QUESTIONS AND PREDICTIONS. The present study addresses two broad research questions. First, how do variable speaker voice and phonetic environment influence listeners' social evaluations of /s/? Second, how does the stability or variability of /s/ tokens within an experiment shape the degree to which these factors enter into social evaluation? We address these questions through three fine-grained analyses, each of which includes a secondary question about whether the effect of interest varies by token condition, which captures within-experiment /s/ variability. (Q1) On average, how do listeners evaluate the masculinity of fronted /s/, and does this differ between the single-token and multi-token conditions? We expect the default CoG-to-masculinity mapping to replicate regardless of token condition, establishing a baseline against which the modulating effects of phonetic context can be assessed. (Q2) How does speaker voice influence listeners' masculinity evaluations of fronted /s/, and does this vary by token condition? We expect speaker voice to modulate the default mapping in both conditions, with two possible, divergent realizations of this effect: attenuation or reversal of the CoG-to-masculinity mapping when voice and CoG provide divergent cues to masculinity or greater CoG effects on masculinity in voice contexts that aren't often perceived as highly masculine or non-masculine. (Q3) How do speaker voice and the F2 of the following vowel jointly influence listeners' masculinity evaluations of fronted /s/, and does this vary by token condition? We expect F2 to modulate the CoG-to-masculinity mapping in the multi-token condition, where the coarticulatory relationship between CoG and the following vowel is preserved, but not in the single-token condition, where it is disrupted. We further predict that the specific direction of the F2 effect to be contingent on the phonetic context provided by speaker voice.

2. Methods.

2.1. PARTICIPANTS. Data were analyzed from 539 participants, recruited online via Prolific. All participants were adults residing in the United States who reported speaking American English as a first language. Participants were compensated $2.75 for the task, which took approximately 11 minutes.

2.2. DESIGN AND STIMULI. The experiment used a 2 x 3 x 2 mixed design. Token condition (single vs multi) was manipulated between subjects. Within subjects, participants heard stimuli varying in /s/ condition (front, mid, back) and in speaker (Speaker A, Speaker B), with six trials per /s/ condition and speaker combination, alongside filler controls containing no /s/. The two speakers were selected to differ in voice characteristics relevant to perceived masculinity; specifically, Speaker A's sibilant-less productions were rated by listeners in a norming study as "typically masculine" and Speaker B's sibilant-less productions were rated as "non-masculine". This selection ensures that speaker voice constitutes a phonetic context that varies, independently of /s/ condition, along the same social dimension under investigation.

Stimuli were constructed by splicing /s/ tokens into 18 unique sentence frames that were normed as not eliciting strong associations with gendered performances. Each sentence contained a single, word-initial instance of /s/. In the single-token condition, a single /s/ token was selected

for each /s/ condition and spliced into each speaker's productions of each sentence frame. This held CoG constant within a given /s/ condition across the phonological environments in which it appeared, thereby decoupling /s/ acoustics from the coarticulatory relationship with the following vowel. In the multi-token condition, /s/ tokens were produced in the same sentence frame into which they were later spliced, preserving the potential covariation between CoG and the acoustic properties of the following vowel context. Tokens for the multi-token condition were selected as follows: all /s/ productions falling within the recorded /s/ CoG range shared by both talkers were identified, partitioned into thirds, and the token closest to the midpoint of each third was selected as the representative for the front, mid, and back /s/ conditions, respectively. The contrast between token conditions allows us to assess the role of phonetic stability versus variability in shaping social evaluations.

	Single-Token	Multi-Token
fronted /s/	8203 Hz	8033 Hz
mid /s/	6928 Hz	6502 Hz
backed /s/	5654 Hz	5605 Hz

Table 1. Table showing the mean center of gravity (Hz) for each /s/ condition and splicing condition

2.3. PROCEDURE. Participants completed the experiment through the web-based recruitment platform, Prolific.co. The experiment was coded using the JavaScript library jsPsych (de Leeuw et al. 2023). Participants completed 72 listening trials, presented in a random order. In a given trial, participants heard a single spoken sentence that played automatically. The auditory stim-ulus was presented alongside a prompt to rate the masculinity of the speaker of that sentence along a visual analog scale with endpoints "Extremely masculine" (scored as 1) and "Not at all masculine" (scored as 0). To ensure that participants heard the entire sentence, clicks along the visual analog scale would not be registered until after the audio clip played in its entirety. A par-ticipant's rating along the scale could not be changed within a trial, and the experiment would automatically advance to the next trial after registering a response along the scale. After complet-ing the experimental trials, participants completed a short survey about their gender beliefs and a demographic survey.

2.3.1. ANALYSIS. Masculinity rating was predicted from fixed effects of /s/ condition (backed, mid, fronted), token condition (single, multi), speaker (A, B), the raw F2 of the following vowel, and /s/ duration, with all five factors entered as a fully-crossed interaction. Acoustic controls for the proportion of creaky voice within the utterance, utterance-level mean F0, and utterance-level F0 range were included as additional fixed effects. All continuous predictors were mean-centered prior to model fitting. To accommodate the bounded nature of the rating scale, responses were modeled using a zero-one-inflated beta (ZOIB) distribution in R (Fang 2025; R CoreTeam 2024). The precision parameter (phi) and zero-one inflation parameters (zoi, coi) were each modeled as a function of the speaker. This captured the possibility that the two speakers differed, in addition to mean masculinity ratings, in the dispersion and boundary-response tendencies they elicited from listeners. Although speaker was included as a predictor of the precision and boundary-inflation

parameters, posterior estimates for these effects were negligible[†].

All R-hat values were below 1.01 and bulk and tail ESS exceeded 2400 for all parameters, indicating sufficient chain convergence and mixing (Bürkner 2017). Posterior predictive checks confirmed that the model adequately reproduced the observed distribution of ratings, including the proportions of responses at the endpoints of the masculinity rating scale.

For the main effect of /s/ condition and the speaker by /s/ condition interaction, we report observed means alongside posterior estimates. For the joint influence of speaker, F2, token condition, and /s/ condition, we report posterior-predicted values rather than observed means, as the four-way interaction involves a continuous predictor evaluated at the means of several acoustic controls.

3. Results. The results bear on three related questions: whether the expected CoG-to-masculinity mapping replicates in our data, how speaker voice modulates that mapping, and how the F2 of the following vowel interacts with both. We found that fronted /s/ is consistently perceived as less masculine than backed /s/, replicating prior findings across both token conditions. However, speaker voice emerges as a stronger cue to perceived masculinity than /s/ condition, and for one speaker, the /s/-guise effect disappears entirely in the single-token condition. Finally, the direction and magnitude of the F2 effect on /s/ evaluations diverges by speaker in the multi-token condition, suggesting that listeners integrate coarticulatory information into their social evaluations in speaker-contingent ways. Taken together, these findings indicate that listeners do not evaluate /s/ CoG in isolation: both speaker voice and phonetic environment shape how the CoG-to-masculinity mapping is expressed, and the degree to which they do so depends on whether listeners were exposed to stable or variable /s/ tokens.

3.1. GENERALIZED EFFECT OF /S/ COG ON PERCEIVED MASCULINITY. Consistent with prior work, listeners rated speakers as less masculine in trials with fronted /s/ than in trials with mid or backed /s/, and this held across both single-token and multi-token conditions (Figure 1). Effect sizes are reported as odds ratios from the ZOIB model, with observed mean differences on the original 0-1 rating scale provided as reference. The backed/fronted contrast was credible in both the single-token condition (OR $= 1.12$, 95% HPD: $[1.072, 1.117]$, obs. mean difference $= 0.022$) and the multi-token condition (OR $= 1.11$, 95% HPD: $[1.058, 1.16]$, obs. mean difference $= 0.010$). The mid/fronted contrast was likewise credible in both conditions, though smaller (single-token: OR $= 1.05$, 95% HPD: $[1.001, 1.10]$, obs. mean difference $= 0.012$; multi-token: OR $= 1.08$, 95% HPD: $[1.027, 1.113]$; obs. mean difference $= 0.011$. Although effect sizes were small, the finding that fronted /s/ trials received the lowest masculinity ratings was stable across token conditions. This constitutes a replication of Campbell-Kibler (2011) and Boyd et al. (2021) and further establishes a baseline against which the modulating effects of phonetic context can be assessed.

3.2. ROLE OF SPEAKER VOICE. Speaker voice emerged as a powerful cue to perceived masculinity, accounting for substantial variance in ratings above and beyond /s/ condition (Figure 2). Averaged across /s/ conditions, listeners were 7.49 times more likely to rate Speaker A as more masculine than Speaker B in the single-token condition (95% HPD: $[7.00, 8.03]$) and 6.02 times more likely in the multi-token condition (95% HPD: $[5.61, 6.45]$). These odds ratios are several times larger than those observed for the /s/ condition contrasts in either token condition (Sec-

[†] No posterior-predicted effect size diverged meaningfully from zero.

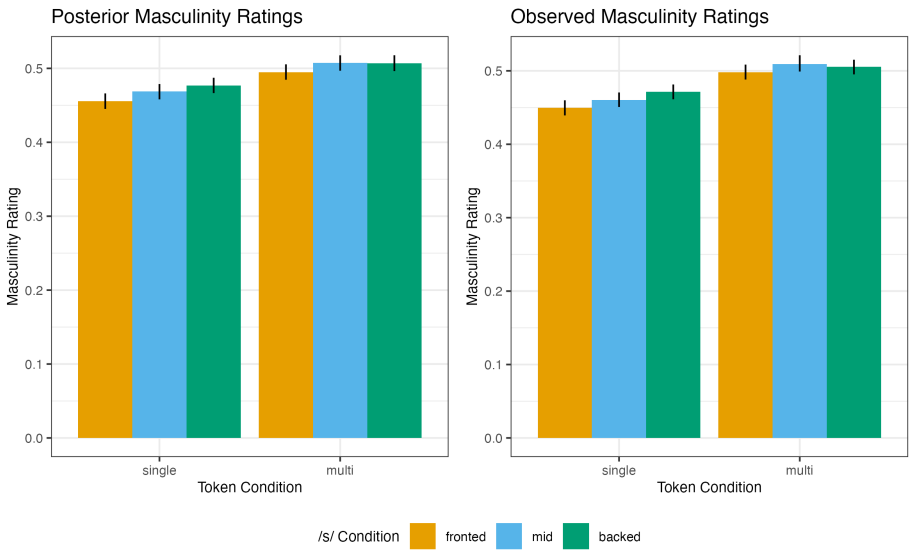

Figure 1. Observed and posterior-predicted effects of /s/ condition and token condition on masculinity ratings. 1 on the y-axis corresponds to the "Extremely masculine" end of the visual analog scale, and 0 corresponds to "Not at all masculine".

tion 3.1). The somewhat larger speaker separation in the single-token condition is consistent with the interpretation that when coarticulatory phonetic context is available, indexical weight is distributed more broadly across signal properties rather than concentrated in the voice alone. These results confirm that the properties of a speaker's voice constitute a phonetic context that dominates listeners' masculinity evaluations, with /s/ condition modulating that evaluation rather than driving it.

For Speaker A, whose voice is perceived as more masculine than Speaker B, listeners showed credible backed/fronted and mid/fronted contrasts in the single-token condition (backed/fronted: OR $= 1.18$, 95% HPD: $[1.112, 1.125]$, obs. mean difference $= 0.032$; mid/fronted: OR $= 1.10$, 95% HPD: $[1.035, 1.116]$, obs. mean difference $= 0.014$). For Speaker B, whose voice provides a phonetic context that already indexes non-masculine gender performance independently of /s/, the credible intervals for both contrasts in the single-token condition included 1 (backed/fronted: OR $= 1.107$, 95% HPD: $[0.994, 1.115]$, obs. mean difference $= 0.011$; mid/fronted: OR $= 1.00$, 95% HPD: $[0.932, 1.08]$, obs. mean difference $= 0.007$), indicating no reliable /s/ condition effect. This is consistent with listeners treating Speaker B's already non-masculine voice as sufficient phonetic context to anchor their masculinity evaluation, with the sibilant cue providing no additional indexical information beyond what the voice already specifies.

In the multi-token condition, credible /s/ condition effects were observed for both speakers, with the relative effect size reversing across speakers compared to the single-token condition. For Speaker A, the backed/fronted and mid/fronted contrasts were credible but smaller than in the single-token condition (backed/fronted: OR $= 1.09$, 95% HPD: $[1.025, 1.115]$, obs. mean difference $= 0.010$; mid/fronted: OR $= 1.07$, 95% HPD: $[1.007, 1.14]$, obs. mean difference $= 0.012$), suggesting that when there is greater within-speaker variability of /s/ CoG, some of the indexical work previously done by voice alone is redistributed across the available

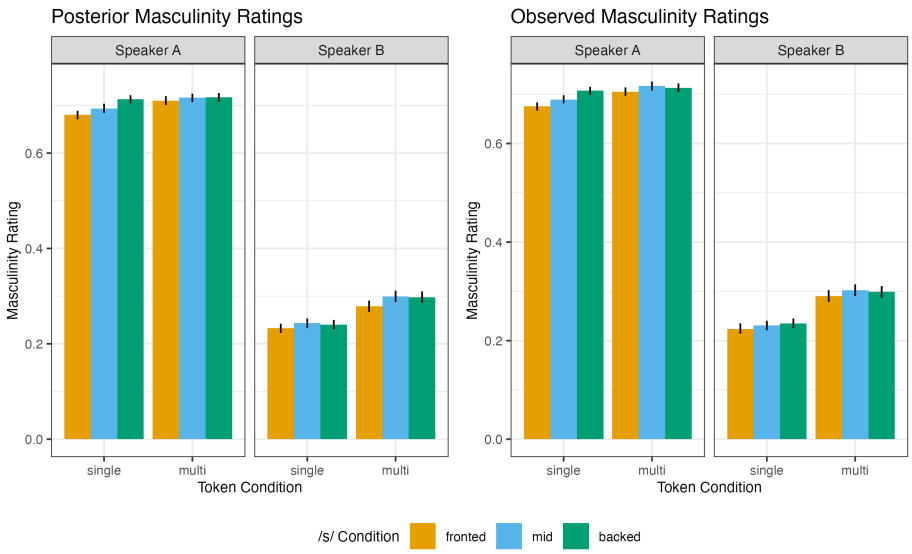

Figure 2. Observed and posterior-predicted effects of /s/ condition, token condition, and speaker on masculinity ratings. 1 on the y-axis corresponds to the "Extremely masculine" end of the visual analog scale, and 0 corresponds to "Not at all masculine".

signal properties. For Speaker B, the contrasts were credible and of comparable magnitude to Speaker A's (backed/fronted: OR $=$ 1.13, 95% HPD: $[1.049, 1.21]$, obs. mean difference $= 0.009$; mid/fronted: OR $= 1.08$, 95% HPD: $[1.001, 1.16]$, obs. mean difference $= 0.012$), indicating that the within-speaker /s/ variability present in the multi-token condition was sufficient to reinstate the /s/ condition effect.

3.3. JOINT INFLUENCE OF SPEAKER VOICE AND PHONETIC ENVIRONMENT. To examine the combined influence of speaker, F2 of the following vowel, token condition, and /s/ condition, we fit a model including the four-way interaction among these predictors (Figure 3). We report posterior-predicted slopes for the F2 x /s/ condition interaction, evaluated separately by speaker and token condition, at mean values of acoustic controls (proportion creak, F0 range, mean F0, /s/ duration).

In the single-token condition, where /s/ CoG was decoupled from the coarticulatory relationship with the following vowel, F2 did not modulate the /s/ condition effect for either speaker. This illustrates that when the coarticulatory relationship between CoG and F2 is disrupted, F2 is not recruited as phonetic context for social evaluation. In other words, listeners did not treat F2 as independently informative about /s/ once its coarticulatory basis was removed.

In the multi-token condition, where /s/ CoG naturally covaried with the phonological environment, F2 modulated the CoG-to-masculinity mapping in ways that diverged by speaker. For Speaker A, listeners became increasingly likely to rate fronted /s/ as less masculine as F2 rose, such that /s/ condition masculinity rating differences were strongest in high F2 contexts (95% HPD: $[-0.029, -0.002]$). For Speaker B, the effect ran in the opposite direction: listeners rated backed and mid /s/ as less masculine as F2 increased (backed: 95% HPD: $[-0.035, -0.016]$; mid: (95% HPD: $[-0.028, -0.008]$)), such that /s/ condition masculinity rating differences were strongest in low F2 contexts. These contrasting patterns indicate that listeners used the coarticu-

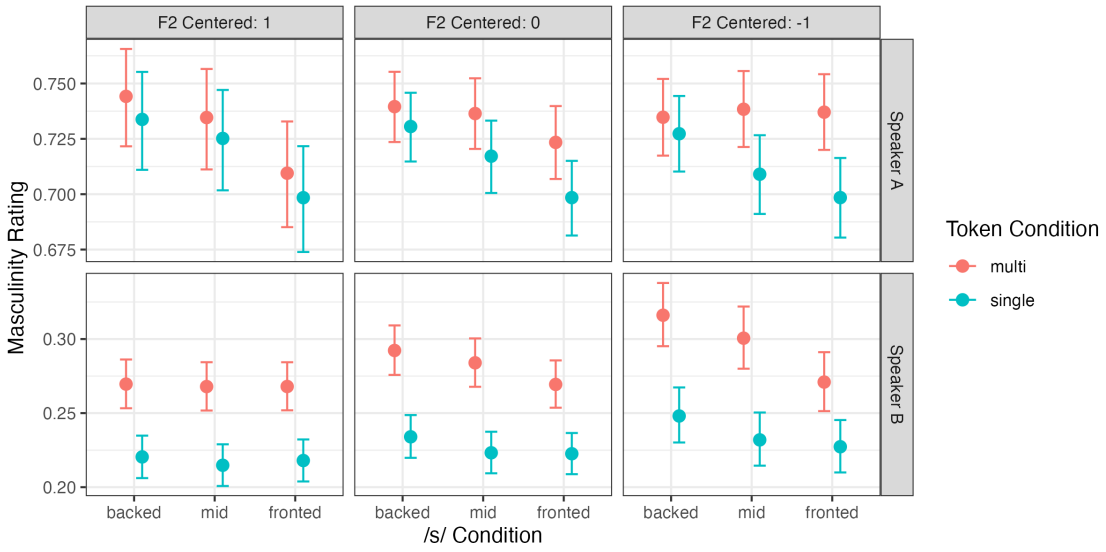

Figure 3. Posterior-predicted conditional effects of /s/ condition, token condition, and speaker at varying values of F2 of the vowel following /s/. F2 values are z-scored across both speakers, such that 0 corresponds to the mean F2, and +/- 1 corresponds to a single standard deviation above/below the mean F2.

latory relationship between /s/ CoG and F2 as phonetic context for social evaluation, but that the mapping this context selected differed depending on the social information provided by speaker voice.

4. Discussion. The central finding of this study is that the default relationship between /s/ CoG and perceived speaker masculinity is selectively activated depending on the phonetic context in which the variant occurs. Speaker voice and, when preserved through natural token variability, the F2 of the following vowel both shape the degree to which perceived masculinity is informed by /s/ condition. These findings are consistent with the context-dependence posited by the indexical field framework (Eckert 2008) and illustrate how properties of the phonetic signal itself can function as context through which social meaning mappings are activated or suppressed.

More broadly, the results support a view in which social evaluation is not a matter of extracting a single variable from the acoustic stream and mapping it onto meaning. To arrive at a masculinity judgment, listeners integrate their knowledge of coarticulation, the indexical properties of the voice, and /s/ quality jointly. The token condition manipulation reveals two distinct modes of this integration. When listeners are exposed to stable /s/ variants and variable speaker voices, as in the single-token condition, they utilize the variation available to them (i.e., /s/ condition and speaker voice) to make masculinity judgments. This results in relatively consistent /s/ condition rating behavior across phonological environments. When listeners are exposed to variable /s/ variants and variable speaker voices, as in the multi-token condition, they use the full range of variation available to them (i.e., phonetic detail in and around /s/ as well as speaker voice), thereby producing more variable /s/ condition rating behavior that differs across both speakers and phonetic environments. The shift from the single- to the multi-token condition is therefore not simply a matter of adding more information. Rather, it changes the structure of the evaluation process itself, expanding the set of signal properties listeners treat as relevant and producing a

richer, more context-differentiated pattern of social inference.

These patterns are consistent with an exemplar-based account of speech perception (Johnson 1997; Goldinger 1996, 1998). If listeners encode coarticulatory and talker-specific details of phonetic tokens in memory, those details are available to inform social inference. Listeners are known to be sensitive to the coarticulatory relationship between fricatives and adjacent vowels (Mann & Repp 1980), providing a perceptual basis for the F2 effects we observe. The coarticulatory relationship between /s/ CoG and F2 of the following vowel must be present in the signal for F2 to modulate social evaluation, and when this relationship is disrupted, as in the single-token condition, F2 carries no evaluative weight. Speaker voice effects are likewise consistent with the episodic encoding of talker-specific detail (Palmeri et al. 1993; Bradlow & Pisoni 1999).

Where Pharao et al. (2014) and Levon (2014) both demonstrate context-selective activation of indexical mappings at the level of register knowledge and listener ideologies, respectively, the present findings show that the phonetic signal itself plays an analogous role. Pharao et al. (2014) showed that register determines whether listeners link fronted /s/ with non-normative masculinity. Levon (2014) showed that listeners' ideologies about gender similarly mediate when fronted /s/ is linked with non-normative masculinity. This study shows that the recoverable properties of the acoustic signal operate as an additional mediating force, one that is, in principle, available on every utterance and to every listener regardless of register knowledge or gender ideologies. The Speaker B finding illustrates this most clearly. Speaker B was normed as "non-masculine"; when asked to provide judgments about his masculinity, listeners may already have sufficient cues from his voice, aside from /s/, to anchor their evaluation, rendering the /s/ cue redundant. When there is increased variability of /s/ in the multi-token condition, listeners differentiate /s/ conditions for both speakers, producing effects of similar magnitude for Speaker A and Speaker B. Phonetic context in the signal can therefore override voice-level indexical information when it is present and informative, a pattern that prior matched guise studies (e.g., Campbell-Kibler 2011; Boyd et al. 2021), by using "normative" speakers (with respect to social categories like gender) and invariant tokens, were not designed to detect.

5. Conclusion. How listeners form indexical links between linguistic form and social meaning depends not on any single variant, but on the integration of the full available signal. This study demonstrates that properties inherent to and recoverable from the phonetic signal itself, here, the F2 of the following vowel and the indexical properties of the speaker's voice, constitute one source of that context for drawing inferences about the relationship between /s/ and perceived speaker masculinity. Listeners do not extract /s/ quality from the acoustic stream and evaluate it in isolation; they integrate their knowledge of coarticulation, what they know about the speaker's voice, and /s/ quality jointly to arrive at a social judgment. The variability of the /s/ signal itself determines which of these sources of information listeners treat as relevant. Stable variants license a narrower integration over voice and categorical /s/ quality, while variable variants license a broader integration that includes the fine phonetic detail surrounding the target segment.

References

Beddor, Patrice Speeter, Kevin B. McGowan, Julie E. Boland, Andries W. Coetzee & Anthony Brasher. 2013. The time course of perception of coarticulation. *The Journal of the Acoustical Society of America* 133(4). 2350–2366. https://doi.org/10.1121/1.4794366.

Boyd, Zac, Josef Fruehwald & Lauren Hall-Lew. 2021. Crosslinguistic perceptions of /s/ among

English, French, and German listeners. *Language Variation and Change* 33(2). 165–191. https://doi.org/10.1017/S0954394521000089. Edition: 2021/07/08.

Bradlow, Ann R. & Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106(2). 707–729. https://doi.org/10.1016/j.cognition.2007.04.005.

Bradlow, Ann R. & David B. Pisoni. 1999. Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America* 106(4). 2074–2085. https://doi.org/10.1121/1.427952.

Bürkner, Paul-Christian. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80(1). 1 – 28. https://doi.org/10.18637/jss.v080.i01. https://www.jstatsoft.org/index.php/jss/article/view/v080i01.

Calder, J. & Sharese King. 2022. Whose gendered voices matter?: Race and gender in the articulation of /s/ in Bakersfield, California. *Journal of Sociolinguistics* 26(5). 604–623. https://doi.org/10.1111/josl.12584.

Campbell-Kibler, Kathryn. 2011. Intersecting Variables and Perceived Sexual Orientation in Men. *American Speech* 86(1). 52–68. https://doi.org/10.1215/00031283-1277510.

Eckert, Penelope. 2008. Variation and the Indexical Field. *Journal of Sociolinguistics* 12. 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x.

Fang, Liu. 2025. zoib: Bayesian Inference for Beta Regression and Zero-or-One Inflated Beta Regression. https://CRAN.R-project.org/package=zoib.

Foulkes, Paul & Gerard Docherty. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34(4). 409–438. https://doi.org/10.1016/j.wocn.2005.08.002.

Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(5). 1166–1183. https://doi.org/10.1037/0278-7393.22.5.1166.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251–279. https://doi.org/10.1037/0033-295X.105.2.251. Place: US.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullennix (eds.), *Talker Variability in Speech Processing*, 145–165. San Diego: Academic Press 1st edn.

Lambert, W., R. Hodgson, Robert Gardner & Samuel Fillenbaum. 1960. Evaluational Reactions to Spoken Languages. *Journal of abnormal and social psychology* 60. 44–51. https://doi.org/10.1037/h0044430.

de Leeuw, Joshua, Rebecca Gilbert & Böjrn Luchterhandt. 2023. jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments. *Journal of Open Source Software* 8. 5351. https://doi.org/10.21105/joss.05351.

Leung, Yeptain, Jennifer Oates, Siew-Pang Chan & Viktoria Papp. 2021. Associations Between Speaking Fundamental Frequency, Vowel Formant Frequencies, and Listener Perceptions of Speaker Gender and Vocal FemininityMasculinity. *Journal of Speech, Language, and Hearing Research* 64(7). 2600–2622. https://doi.org/10.1044/2021$_JSLHR-20-00747$.

Levon, Erez. 2014. Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society* 43(5). 539–566. https://doi.org/10.1017/S0047404514000554.

Linville, Sue Ellen. 1998. Acoustic Correlates of Perceived versus Actual Sexual Orientation in Mens Speech. *Folia Phoniatrica et Logopaedica* 50(1). 35–48. https://doi.org/10.1159/000021447.

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

Munson, Benjamin. 2007. The Acoustic Correlates of Perceived Masculinity, Perceived Femininity, and Perceived Sexual Orientation. *Language and Speech* 50(1). 125–142. https://doi.org/10.1177/00238309070500010601.

Munson, Benjamin, Sarah Jefferson & Elizabeth McDonald. 2006. The influence of perceived sexual orientation on fricative perception. *The Journal of the Acoustical Society of America* 119. 2427–37. https://doi.org/10.1121/1.2173521.

Palmeri, Thomas J., Stephen D. Goldinger & David B. Pisoni. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(2). 309–328. https://doi.org/10.1037/0278-7393.19.2.309.

Pharao, Nicolai, Marie Maegaard, Janus Spindler Møller & Tore Kristiansen. 2014. Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society* 43(1). 1–31. https://doi.org/10.1017/S0047404513000857. https://www.cambridge.org/core/journals/language-in-society/article/indexical-meanings-of-s-among-copenhagen-youth-social-perception-of-a-ph 76B4354FEDB55A819FADDF321620BC33.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Words and Beyond: Linguistic and Semiotic Studies of Sociocultural Order* 23(3). 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2.

Soli, Sigfrid D. 1981. Second formants in fricatives: Acoustic consequences of fricativevowel coarticulation. *The Journal of the Acoustical Society of America* 70(4). 976–984. https://doi.org/10.1121/1.387032.

Stuart-Smith, Jane. 2007. Empirical evidence for gendered speech production: /s/ in Glaswegian. In J Cole & J Hualde (eds.), *Laboratory Phonology 9* Phonology and phonetics, 65–86. New York, USA: Mouton de Gruyter.

R Core Team, R. 2024. R: A language and environment for statistical computing. https://www.R-project.org/.

Terango, Larry. 1966. Pitch and Duration Characteristics of the Oral Reading of Males on a Masculinity-Femininity Dimension. *Journal of Speech and Hearing Research* 9(4). 590–595. https://doi.org/10.1044/jshr.0904.590.

Yuasa, Ikuko Patricia. 2010. Creaky Voice: A New Feminine Voice Quality for Young Urban-Oriented Upwardly Mobile American Women? *American Speech* 85(3). 315–337. https://doi.org/10.1215/00031283-2010-018.