# PERCEPTUAL INFLUENCES OF SOCIAL AND LINGUISTIC PRIMING ARE BIDIRECTIONAL

Dominique A. Bouavichith, Ian Calloway, Justin T. Craft, Tamarae Hildebrandt, Stephen J. Tobin, Patrice Speeter Beddor

University of Michigan
dombouav@umich.edu, iccallow@umich.edu, juscraft@umich.edu, tamhil@umich.edu, sjtobin@umich.edu, beddor@umich.edu

## ABSTRACT

During speech perception, listeners utilize and integrate both linguistic and social information. Previous research has demonstrated that social information can affect linguistic decisions, but little research has examined whether these effects are bidirectional. We probe at bidirectionality in an eye-tracking study in which, in each trial, participants first see a visual prime for one category (gender, lexical); participants then hear an auditory stimulus drawn from female-to-male and *shack*-to-*sack* continua and look to images of the non-primed category. Results replicate earlier findings that visual gender information can shift listeners' /s-ʃ/ boundary when paired with a non-prototypically gendered voice. The novel finding is that, in turn, visually primed linguistic information can shift listeners' judgments of speaker gender when paired with an ambiguous sibilant. Taken together, the results provide evidence of a bidirectional link between social and linguistic categorization in speech perception.

**Keywords**: speech perception, priming, social information, eye-tracking

## 1. INTRODUCTION

Listening is, in the context of conversational interactions, a social activity. The acoustic speech signal produced by a speaker conveys linguistic information that is interwoven with social information. These linguistic and social cues are not independent: they are not, for the most part, isolable in the acoustic input, and they interact in determining listeners' decisions about what a speaker is saying. For example, in early work, Ladefoged & Broadbent showed that listeners' vowel judgments are influenced by person-specific information about a speaker's overall vowel space [9]. Subsequent work has shown that socio-indexical information about a speaker and their speech variety not only biases listeners' linguistic choices, but also speeds the time course of those decisions [2, 21]. Thus, unsurprisingly, in their communicative exchanges with speakers, listeners are making informed use of socially structured acoustic variation. Even *expectations* about a speaker's social background (e.g., age, dialect, native language) serve to guide listeners' linguistic choices [4, 12, 16].

This study investigates the social category of gender, whose influences on phonetic perception—especially fricative perception—are well established. Sibilants differ in their spectral peak location, with /s/ having a higher-frequency peak than more posterior /ʃ/ [8]. Because spectral peak frequencies are, in general, higher for fricatives produced by women than men, there is, as Munson et al. point out, potential ambiguity across female and male speakers' productions: sibilants with peak frequencies in a targeted range could correspond to a male speaker's /s/ and a female speaker's /ʃ/ [14]. This difference is partially but not exclusively attributable to anatomical differences; sociophonetic factors also contribute to it [3]. Perceptual studies exploiting this ambiguity have found that, when American English listeners respond to ambiguous stimuli along a /ʃ/-/s/ continuum, they report hearing more /ʃ/ when the remainder of the utterance (e.g., (*sh)ack* or *(s)ack*) is produced by a woman than by a man [5, 11, 13]. Speaker gender also influences sibilant perception when the auditory context remains constant but listeners receive visual or other information about speaker gender [6, 14, 15, 18, 19].

Our primary goals in this study are to replicate previous findings that speaker gender informs listeners' linguistic decisions, and to address a further question: are the perceptual consequences of interacting social and linguistic information bidirectional? That phonetic judgments of ambiguous sibilants depend on expected speaker gender demonstrates that phonetic categories are socially malleable. However, are social categories similarly malleable, such that social judgments of gender-neutral voices depend on the expected phonetic category? Research on social categorization [7] and, for example, Sumner et al.'s proposal of dual encoding of social and linguistic information [20] lead us to predict bidirectional influences. In this study, we test this prediction using a visual world paradigm in which listeners are primed with either (i) social information as they make judgments about what word was produced or (ii) linguistic information as they make judgments about who produced a word.

## 2. METHODS

### 2.1. Stimuli

Auditory stimuli were 60 words drawn from a six-step linguistic *shack-sack* continuum and two five-step gender continua, one created from a female speaker's productions and the other from a male speaker's productions. All stimuli were created by splicing one of the six synthesized sibilants onto manipulated naturalistic (female or male) recordings of [æk].

The synthetic /ʃ-s/ continuum was created with the Klatt Synthesizer functionality in Praat [1]. Steps were generated with even spacing of third and fourth frication formant parameters. The sibilant continuum was created using the same parameters Munson [15] used, ranging between the values of his second and eighth continuum steps. Centers of gravity ranged from 3.2 kHz (more /ʃ/-like) to 7 kHz (more /s/-like).

The original versions of the naturalistic [æk] stimuli were extracted from "Say sack again" productions of one female and one male native speaker of American English. F0 and formant spacing (acoustic features associated with perceived gender) were modified in Praat to create the two five-step continua. Within each continuum, the mean F0 over the duration of the vowel was spaced evenly across consecutive steps, and the formant shift factor varied linearly. For the male speaker's continuum, mean F0 ranged from 135 Hz (unmodified) to 210 Hz, and the formant spacing factor ranged from 1.0 (unmodified) to 1.2 (greater spacing). For the female speaker's continuum, mean F0 ranged from 190 Hz (unmodified) to 90 Hz, and the formant spacing factor ranged from 1.0 (unmodified) to 0.83 (less spacing). Each sibilant was concatenated with each [æk] token to produce the 60 unique stimuli.

The visual stimuli consisted of black-and-white line drawings, corresponding to *shack* and *sack*, and grey-scale-converted photographs of female and male faces, rated as highly gender-prototypical, from the Chicago Face Database [10] (see Table 1).

### 2.2. Participants and Procedure

Twenty-five native English-speaking undergraduate students participated in the study. Auditory stimuli were presented over headphones and participants' eye movements were recorded using a remote monocular eye-tracker (EyeLink 1000 Plus, SR Research).

Each eye-tracking trial lasted approximately 10 seconds and asked either "What do you hear?" or "Who do you hear?". In *What* trials, participants saw a gender prime (female or male face) and subsequently looked to test images of *shack* and *sack* in response to an auditory stimulus. In *Who* trials, participants saw a lexical prime (*shack* or *sack*) and test images were female and male faces.

**Table 1:** Distribution of visual primes for sibilant and gender steps. Ambiguous stimuli are shaded. Shaded columns: 2 gender primes. Shaded rows: 2 lexical primes.


| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Female | ʃ | F | ʃ | F | ʃ/s | F | ʃ/s | F | s | F | s | F |
| | ʃ | F M | ʃ | F M | ʃ/s | F M | ʃ/s | F M | s | F M | s | F M |
| | ʃ | F M | ʃ | F M | ʃ/s | F M | ʃ/s | F M | s | F M | s | F M |
| | ʃ | M | ʃ | M | ʃ/s | M | ʃ/s | M | s | M | s | M |
| Male | ʃ | M | ʃ | M | ʃ/s | M | ʃ/s | M | s | M | s | M |

/ʃ/ ⟵ ⟶ /s/

Within each trial: (1) A prime image appeared in the center of the screen and was accompanied by the audio "You'll hear from" or "You'll hear". (2) After 1500 ms, two test images appeared on either side of the (then scaled-down) prime, accompanied by the audio instruction "Look at each {drawing/image}". (3) Participants were instructed to look at the prime image, with the audio "Look at the middle". (4) Participants heard, "{What/Who} do you hear?". (5) After 1000 ms, the prime image disappeared and the target audio trial (one of 60 stimuli) played.

To test for the effect of prime, stimuli ambiguous for gender and/or lexical item were presented (in different, randomized trials) with both primes; that is, stimuli ambiguous for *shack/sack* were presented with both gender primes and stimuli ambiguous for gender were presented with both lexical primes. Ambiguity was determined by pilot testing. Shading in Table 1 shows ambiguous stimuli receiving more than one prime. These stimulus/prime combinations were presented both in the originally female and in the originally male voices. Two repetitions of each trial resulted in 328 eye-tracking trials. Test image positions were counterbalanced across participants.

## 3. HYPOTHESES

We hypothesize that, consistent with previous work [6, 14, 15, 18, 19], listeners' perception of sibilants will be influenced by visual information about speaker gender. Thus, in *What* trials, listeners should fixate more on the *shack* image when the visual prime is a female face than when it is a male face.

For *Who* trials, we hypothesize that listeners will again be influenced by the visual prime—in this case, priming for lexical category. If the influences of social and linguistic primes are bidirectional then, just as a female prime should elicit more fixations on the *shack* image, so should a *shack* prime elicit more fixations on the female photo.

We also predict that, independent of the lexical prime, participants will fixate more on the female photo when an ambiguous stimulus sounds more like *shack* than like *sack*. This is the bidirectional corollary of an /ʃ/ percept in response to an ambiguous sibilant when the speaker turns out to be female [13].

## 4. RESULTS

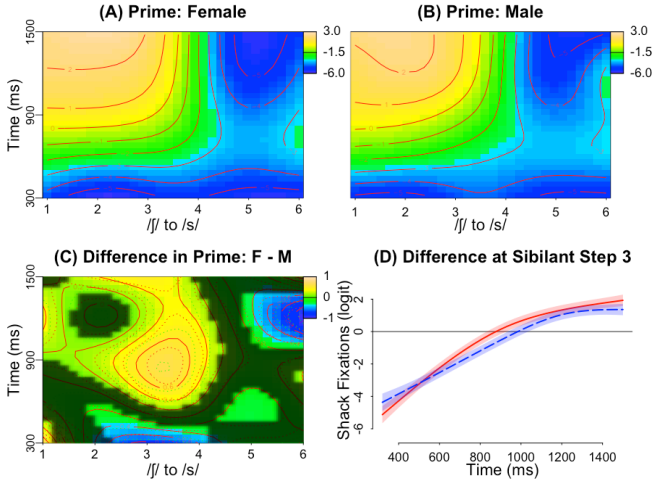
In the statistical analyses, fixations on the *shack* image (linguistic model) or on the female photo (social models) were submitted to Generalized Additive Mixed Models (GAMMs) using the *mgcv* [23] and *itsadug* [22] packages in R [17]. Only trials within areas of ambiguity, which received two primes (see Table 1), were included in the statistical models. Thus, the linguistic model predicts how gender-ambiguous voices at the centermost points of the perceived gender continuum affect sibilant categorization across an entire continuum from /ʃ/ to /s/. Likewise, the social models predict how ambiguous sibilant information at the /ʃ-s/ boundary affects gender categorization across all steps of the perceived gender continuum. Models were fit using the *bam* function with default smoothing. For GAMMs, visualization plays an essential role in significance testing and interpretation.

### 4.1. Linguistic models ("What do you hear?")

To test the first hypothesis that gender primes influence sibilant judgments, participants' proportion fixations on *shack* were modeled using a logistic GAMM. The effects of prime, speaker (original male/female), sibilant step, gender step, their interactions, and a random intercept for participant were included in the model. Of primary interest among the significant effects and interactions in the model output is that gender prime (male, female) significantly interacted with sibilant step in predicting proportion fixations on *shack*. Specifically, as expected, the model predicted a greater proportion of looks to *shack* when primed with a female face within ambiguous /ʃ-s/ steps along the sibilant continuum.

To visualize the effect of prime, we utilize a difference plot of the model output, which subtracts the effect of trials with a male prime from the effect of trials with a female prime. However, by way of background, we first *separately* show the model output for each prime condition. Figures 1A and 1B give the model-derived logit proportion fixations on *shack* when primed with a female (1A) and male (1B) face. Proportion fixations are represented by the color (or grey-scale) at each point across /ʃ/ to /s/ (x-axis) and time (y-axis); positive values, shown as warmer colors (or lighter grey-scale), indicate higher fixation proportions to *shack*. As would be expected, in both plots, the likelihood to fixate on either *shack* or *sack* increases across the time course of the trials (note stronger contrasts towards the top of the plots). Also as expected, the likelihood of fixating on *shack* is greatest for sibilant steps 1-3.

**Figure 1:** Model-derived fixations on *shack*. **A&B:** Proportion fixations across sibilant step (x-axis) and time (y-axis: 300 ms = 25 ms after sibilant offset) for female (**A**) and male (**B**) primes; positive values (warm colors / lighter grey-scale) = more looks to *shack*. **C:** Difference in proportion fixations (A minus B); unshadowed positive values = significantly more looks to *shack* in A relative to B (95% confidence interval). **D:** Effect of female (solid red) and male (dashed blue) prime on /ʃ-s/ step 3 over time.

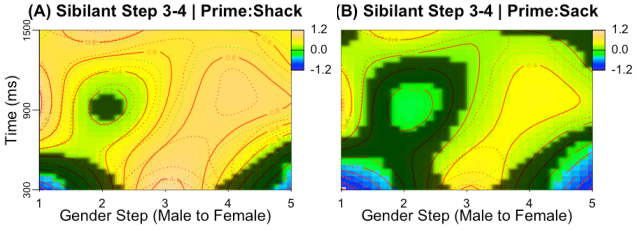



The *differences* between the effects of gender primes (1A minus 1B) are visualized in Figure 1C, where positive values (warmer/lighter) represent significantly (no shadow) more fixations on *shack* in 1A (female prime) relative to 1B (male prime). This increase holds across sibilant steps 2, 3, and 4. The greatest prime effect occurs in the most linguistically ambiguous region of the continuum, which we capture in Figure 1D as a "slice" taken at sibilant step 3. Figure 1D shows the difference, over time (x-axis), in logit proportion fixations on *shack* (y-axis) between female (solid red line) and male (dashed blue) primes at step 3 of the /ʃ-s/ continuum. In this more traditional representation of eye gaze patterns, we again see the expected effect of more looks to *shack* with the female prime.

### 4.2. Social models ("Who do you hear?")

To address the second hypothesis, that in the *Who* trials participants' fixations on the female or male photo will be influenced by a linguistic prime, we modeled participants' proportion fixations on the female face again using a logistic GAMM. The model included main effects of prime, speaker, sibilant step, gender step, their interactions, and a random intercept for participant; we provide the difference plots for this analysis.

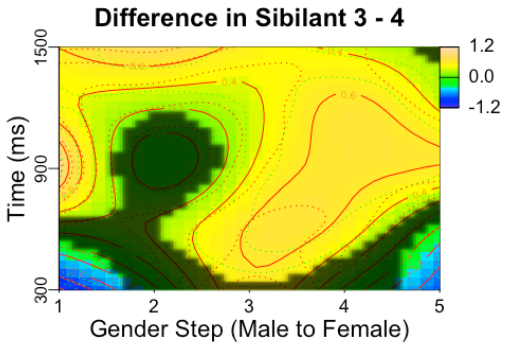
**Figure 2:** Model-derived fixations on female face. **A&B:** Difference in proportion fixations for sibilant step 3 (more /ʃ/-like) minus step 4 for *shack* (**A**) and *sack* (**B**) primes. Positive values = significantly more looks to *female* face in sibilant step 3 relative to 4.




Recall that the ambiguous /ʃ-s/ stimuli are sibilant steps 3 and 4 (Table 1). Figure 2A shows the difference in proportion fixations on the female face for sibilant step 3 (more /ʃ/-like) minus step 4 (more /s/-like) as a function of gender step and time, when primed with the *shack* image. Figure 2B gives the same information, but for prime *sack*. Both panels show modeled predictions, where positive values indicate significantly more fixations on the female face in step 3 relative to step 4. The significantly positive region is larger (both across gender steps and time) in 2A than 2B indicating that, as predicted, the *shack* prime elicits more looks to the female face than the *sack* prime. Within this model at sibilant step 3 (not shown), significant differences due to the effect of prime were found beginning around 700 ms and continued through the end of the trial. (Auditory information for gender is not available until vowel onset at 275 ms, so reliable target fixations could begin, at the earliest, at 475 ms assuming a 200 ms eye-movement programming delay.)

To address the third hypothesis that, *independent of priming condition*, participants will fixate more on the female photo when the stimulus sounds more like *shack* than *sack*, we ran an additional logistic GAMM with the same structure as before, but excluding prime. The output of this model as represented in Figure 3 shows, across the linguistic primes, an increase in likelihood to fixate on the female face (positive values) when more /ʃ/-sounding sibilants are produced by a more gender-ambiguous speaker (gender step 3). (As in Figures 1C and 2, only those areas where differences are significant are unshadowed.) This effect is strengthened as the voice becomes more prototypically female (steps 4 and 5), which we interpret as the integration of congruent gender cues in the rime of an auditory stimulus.

**Figure 3:** Model-derived fixations on female face, independent of prime.




## 5. DISCUSSION

This study provides evidence of a reciprocal influence of gender and sibilant priming on speech perception. Eye gaze patterns show that participants categorized ambiguous sibilants as /ʃ/ more often than /s/ when primed with a female face—a result that aligns with prior work showing that visual gender cues can modulate sibilant perception when the voice has reduced gender prototypicality.

Eye gaze patterns also indicate that participants categorized gender-ambiguous stimuli as being produced by a female speaker more often when primed with an image of *shack* than with an image of *sack*. Thus, visual information about sibilant category can modulate gender perception when a sibilant has reduced linguistic prototypicality.

More generally, our findings provide evidence that linguistic categories are socially malleable, and social categories (such as gender) are linguistically malleable: ambiguous linguistic content can be disambiguated through meaningful social information about a speaker *and* ambiguous social information for a speaker can be disambiguated through a meaningful linguistic percept.

Participants' responses also showed gradient influences of auditory sibilant information on gender judgments independent of priming condition: stimuli were more often categorized as female the closer the sibilant was to the /ʃ/ end of the sibilant continuum. (It remains to be shown whether the same pattern would have emerged in an experiment with no visual primes.) Gender differences in sibilant peak frequency presumably at least partly motivate the interrelated nature of these variables in perception. The results of this study further suggest that this spectral information may be encoded in such a manner as to aid both lexical access and social processing. Such an outcome offers support for perceptual frameworks [20] where the processing of linguistic and social information shows some degree of interactivity.

## 6. ACKNOWLEDGMENTS

This material is based on work supported by NSF Grant BCS-1348150 to Patrice Beddor and Andries Coetzee; any opinions, findings, and conclusions are the authors' and do not necessarily reflect the views of the NSF. We also thank audiences at the University of Michigan for helpful comments.

## 7. REFERENCES


[1] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341-345.

[2] Dahan, D., Drucker, S.J., Scarborough, R.A. 2008. Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 108, 710-718.

[3] Fuchs, S., Toda, M. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors? In: Fuchs, S., Toda, M., Zygis, M. (eds), *Turbulent Sounds: An Interdisciplinary Guide*. Berlin: De Gruyter Mouton, 281-302.

[4] Hay, J., Nolan, A., Drager, K. 2006. From *fush* to *feesh*: Exemplar priming in speech perception. *The Linguistic Review* 4: 351-379.

[5] Johnson, K. 1991. Differential effects of speaker and vowel variability on fricative perception. *Language and Speech* 34, 265-279.

[6] Johnson, K., Strand, E.A., D'Imperio, M. 1999. Auditory—visual integration of talker gender in vowel perception. *J. Phonetics* 27, 359-384.

[7] Johnson, K.L., Lick, D.J., Carpinella, C.M. 2015. Emergent research in social vision: An integrated approach to the determinants and consequences of social categorization. *Social & Personality Psychology Compass* 9(1), 15-30.

[8] Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108, 1252-1263.

[9] Ladefoged, P. Broadbent, D.E. 1957. Information conveyed by vowels. *J. Acoust. Soc. Am.* 29, 98-104.

[10] Ma, D.S., Correll, J., Wittenbrink, B. 2015. The Chicago Face Database: A free stimulus set of faces and norming data. *Behavior Research Methods* 47, 1122-1135.

[11] Mann, V.A., Repp, B.H. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception and Psychophysics* 28(3), 213-228.

[12] McGowan, K. 2015. Social expectation improves speech perception in noise. *Language and Speech* 58(4), 502-521.

[13] Munson, B., McDonald, E.C., DeBoe, N.L., White, A.R. 2006. The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *J. Phonetics* 34, 202-240.

[14] Munson, B., Ryherd, K., Kemper, S. 2017. Implicit and explicit gender priming in English lingual sibilant fricative perception. *Linguistics* 55(5), 1073-1107.

[15] Munson, B. 2011. The influence of actual and imputed talker gender on fricative perception, revisited. *J. Acoust. Soc. Am.* 130, 2631–2634.

[16] Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *J. Lang. and Social Psych.* 18, 62-85.

[17] R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

[18] Strand, E.A. 1999. Uncovering the role of gender stereotypes in speech perception. *J. Lang. and Social Psych.* 18(1), 86-99.

[19] Strand, E.A., Johnson, K. 1996. Gradient and visual speaker normalization in the perception of fricatives. In: Gibbon, D. (ed), *Natural Language Processing and Speech Technology: Results of the 3rd KOVENS Conference.* Berlin: Mouton de Gruyter, 14-26.

[20] Sumner, M., Seung, K., King, E., McGowan, K. 2014. The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psych.* 4, 1015.

[21] Trude, A.M., Brown-Schmidt, S. 2012. Talker-specific perceptual adaptation during online speech perception. *Lang. and Cognitive Processes* 27:7-8, 979-1001.

[22] van Rij, J., Wieling, M., Baayen, R., van Rijn, H. 2017. itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.3.

[23] Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R* (2nd edition). Chapman and Hall/CRC.