Removing the disguise: the matched guise technique and listener awareness

Abstract

Sociophonetic perception is often studied using versions of the matched guise technique. Linguists using this technique appear united in the methodological assumptions that participants believe the manipulation and that this belief influences perception below the level of introspective awareness. We report an audiovisual matched guise experiment with a novel ’unhidden’ instruction condition. The basic task is a replication of the Strand effect (Strand, 1999; Strand & Johnson, 1996). Participants in the ’unhidden’ condition were instructed that the man or woman in the photo did not represent the voice they were listening to. Participants in both guises exhibited the Strand effect to nearly numerically identical extents. This result suggests that participants need not believe a link exists between a voice and a purported social category for visually-cued social information to influence segmental perception. We explore the implications of this result for the MGT and for theories of social awareness and speech perception more broadly.

Introduction

A great deal of attention has been paid in the phonetics and sociophonetics literatures to the perception of the voiceless fricatives [ʃ] and [s] in English. To a first approximation, these fricatives differ in the distance between the point of lingual articulation and the teeth, which give them their characteristic sibilance (Fant, 1960; Shadle, 1991). English [s] has a short resonating chamber behind the teeth; it is typically produced by holding the tongue blade near enough to the alveolar ridge to cause turbulent airflow. English [ʃ] has a comparatively larger resonating chamber; it is typically produced with a more posterior, palato-alveolar tongue position and lip rounding both of which serve to reinforce this posteriority. But listeners do not perceive via first approximations. Indeed, these two fricatives have been exciting to researchers precisely because of the sensitivity listeners bring to their perception and how that perception interacts with both linguistic and social knowledge.

Coarticulatory and Social Information Influence [ʃ]-[s] perception

Listeners are sensitive to articulatory mismatches between the fricatives [ʃ]-[s] and neighboring sounds. Whalen (1984) conducted a series of experiments to investigate listeners' responses to articulatory mismatches in synthetic speech. Overall, the result of these investigations was that subcategorical phonetic mismatches slow phonetic judgments. In onset position, in isolation, or in coda position, misleading coarticulatory information inhibited reaction times. Listeners, Whalen cautions in the conclusion, are sensitive to articulatory patterns that are below the level of conscious awareness and not available to direct experimenter scrutiny. While listeners will readily fill-in missing or ambiguous information, the presence of actively *conflicting* articulatory information is inhibitory.

A commonly used methodology involves the creation of synthetic fricative continua. These continua have endpoints in prototypical examples of [ʃ] and [s] with some number of equal-sized acoustic steps generated, synthesized, or even mixed between these. Somewhere in the middle of such a continuum will be fricative-like noise that is ambiguous as to category membership: not clearly a [ʃ] and not clearly an [s]. May (1976) paired a continuum from [ʃ] (2.9 kHz) to [s] (4.4 kHz) with synthetic [æ] vowels to form CV pairs. May found that listeners perceived a higher proportion of the fricative continuum as [ʃ] when paired with vowel stimuli from a smaller vocal tract. The logic here is that smaller resonating chambers between the lingual articulation and teeth will have a higher mean frequency than larger resonating chambers. Listeners' use of apparent vocal tract size in perception reflect their knowledge of this variation (Munson, 2011).

Mann and Repp (1980) replicated this finding, extending it to natural productions of vowels spoken by a male or female-identified talker. Similar to May's results with simulated vocal tract size, Mann & Repp found a higher proportion of the fricative continuum was heard as [ʃ] when paired with the speech of the female talker. This early work, like others of the period (Ohala, 1984), theorized size as being a relatively deterministic feature of talker sexual dimorphism. One consequence of this view is that gender-related variation in the speech signal is considered mechanistic, universal, and following from purely physical laws. Vocal tract size is presumably not available for individual performance and so listener knowledge of this variation can be correspondingly simple. Vocal tract size may influence perception, but it does so implicitly, automatically, and below the level of introspective awareness.

Strand and Johnson (1996) conducted a pair of experiments investigating the influence of purported gender of a talker on the perception of the [ʃ]-[s] boundary. In their experiment 1, listeners heard a [ʃ]-[s] continuum paired with voices previously normed as prototypical female, non-prototypical female, non-prototypical male, and prototypical male voices. The result replicates Mann and Repp (1980) and extends it to show that the influence of a gendered voice correlates with the protypicality of that voice (exp1). They then extend this research to show that presenting listeners with prototypically-gendered videos of their purported talker can, again, shift perceptions of the [ʃ]-[s] such that listeners report hearing a higher proportion of the continuum as [ʃ] when watching a female talker and a higher proportion of [s] when watching a male talker. The AV condition of their experiment 2 is reminiscent of McGurk and MacDonald (1976) and is presented in that context. A striking feature of the McGurk Effect is its automaticity; participants can not choose to perceive the two components of a fused percept independently. It is unclear from Strand and Johnson (1996) and subsequent work whether the perceptual influence of visually-presented social information is implicit and automatic, like vocal tract size, the McGurk effect, etc., or whether the effect disappears when listeners are aware of the guise manipulation.

This is an incomplete sample of the literature on the perception of these fricatives. We hope, however, that the message is clear that even when arriving at a purely linguistic percept, listeners' judgments depend on a rich constellation of evidence and expectation. Vocal tract size, formant transitions, following vowel quality (Mann & Repp, 1980), and coarticulatory cues, along with the acoustic properties of the fricative itself, can all shape how listeners categorize that fricative. Rather than relying on a single, invariant, phonetic cue, listeners take the entire fricative and context into account Whalen (1991).

One imagines such exquisite sensitivity to the phonetic cues conveying linguistic category membership might restrict language users' freedom to communicate and perceive social information via the same phonetic signal. This would be the prediction of a phonetic theory in which linguistic information and social information battle for control of the air waves –where listeners must normalize away social variation to recover linguistic information. Instead, with these fricatives, at least, we can observe the opposite. The fricatives /ʃ/ and /s/ often carry social meaning (Mack & Munson, 2012; Podesva & Kajino, 2014) with /s/ being "perhaps the most iconic phonetic variable in the field" (Calder, 2018). The implication is that the social and linguistic meanings of particular phonetic cues are not in competition with one another.

Phonetics, Speech Perception, and the social-construction of gender

It has long seemed normal in speech research to imagine that gender is a simple, binary projection from biological sex onto social identity (Daniel, Lorenzi, da Costa Leite, & Lorenzi-Filho, 2007; Samoliński, Grzanka, & Gotlib, 2007). However, if these biological tendencies were deterministic we would expect to see differentiation begin at puberty. It does not. In fact, prior to the onset of puberty, girls' oral and nasal cavities tend to be larger than those of boys (Samoliński et al., 2007). If anything, we should expect lower formants and lower center and peak frequencies for girls, inverting the adult pattern. Instead what we observe is that listeners can differentiate the voices of children as young as 4 years of age using vowel formant frequencies (Perry, Ohde, & Ashmead, 2001). Schellinger, Munson, and Edwards (2017) report a pair of experiments in which participants heard words produced by children between the ages of 2 and 5, and provided continuous ratings identifying fricatives, vowels, and gender typicality. Children typically show gendered patterns in speech at age 4 and up despite vocal tract length being non-distinct for this cohort. It is critical to remember that formants and fricatives are the result of not purely vocal tract biology but also articulator coordination. Even without biologically-differentiated vocal tracts, people who identify as male or female can perform that identity through gestural style. Vowels, in both their linguistic and social aspects, are the acoustic consequence of gestural control.

Gender is more likely the product of, rather than an explanation for, linguistic variation Eckert and Podesva (2021). Just as with words, genders are arbitrary; both the category labels and their acoustic correlates are language specific (Johnson, 2005, 2006) and the constellation of meanings are socially-constructed in interaction (Eckert, 2008). The formant ratios that distinguish 'male' from 'female' in Norwegian are markedly different from the formant ratios that do this in Danish (Johnson, 2006); what it means to be 'male' versus 'female' is quite different in Thailand than in Japan (Alpert, 2014; Käng, 2013). Children don't perform adult-like vowel formant patterns because they were born tiny men and women, children perform adult-like vowel formant patterns because they identify as a gender and are participating in the sylistic bricolage (Zimman, 2017) available to communicate that gender to others. Humans are meaning-making agents, not deterministically resonating meat tubes and expert listeners of a language know this.

The existence of this knowledge questions awareness and control. In the earliest sociophonetic perception research it was still possible to imagine that the kind of knowledge listeners drew on to perceive gender was knowledge of primary biological traits. We now understand that, instead, the influence of gender-based expectations in speech perception like that investigated here is evidence of the influence of cultural knowledge on what are traditionally understood to be purely linguistic decisions (Boyd, Fruehwald, & Hall-Lew, 2021). Just as vowel height, lip rounding, and syllable affiliation influence the perception of fricative place, so too do socially-constructed gender categories.

Matched Guise

The Matched Guise technique (MGT) has been deployed in numerous configurations but, at its core, the technique pairs a single linguistic signal (identical recordings, an identical speaker, identical texts, etc.) with multiple purported social categories to elicit the influence of those cues on participants' linguistic judgments (Campbell-Kibler, 2005, 2007) or language attitudes (CHAN, 2021; Hadodo, this volume). In their foundational use of the technique, for example, Lambert, Hodgson, Gardner, and Fillenbaum (1960) found that bilingual Montrealer's voices evoked quite different social judgments in French vs English guises, providing evidence that listeners are able to perceive and connect social information in the voice to ideological framing of social types. In social speech perception research, cross-modal audio/visual matched guise studies are common in which visual information serves as a 'guise' for identical voice recordings; researchers sometime disregard the social information in voices Rubin (1992) and sometimes take the combination of voice and visual stimuli into account (Campbell-Kibler, 2016; Gnevsheva, 2017; McGowan, 2015). But uniting these linguistic researchers, and delineating them from colleagues in social psychology (for discussion, see Rosseel & Grondelaers, 2019), is the methodological assumption that the connection of voice to social type happens below the level of conscious awareness. Awareness here, though generally not explicitly acknowledged, appears to be construed narrowly as participants' ability to identify and comment on the existence of a guise manipulation. Researchers attempt to deceive participants about the intentional use of guise to elicit evidence of social evaluation in language attitudes, segmental speech perception, memory, etc.

It may be assumed that the matched guise technique works because listeners are unaware of the guise manipulation. Researchers go to great lengths to ensure this lack of awareness (e.g. Grondelaers & van Gent, 2019; Pharao & Kristiansen, 2019)). However, the majority of studies cannot speak to this lack of awareness during phonetic perception because the data provided by the participants is relatively late in processing and involves layers of potential introspection and evaluation that block access to the initial online percept for listeners and researchers alike. McGowan and Babel (2020) performed an audio/visual MGT with both a task designed to get at phonetic perception of individual segments and a sociolinguistic interview intended to investigate listeners' judgements about the purported speaker. Every participant was shown both guises and while segmental and social perceptions were aligned with the identity of the purported talker in the initial guise presentation, these perceptions diverged in the second guise – with phonetic perceptions remaining unchanged and social evaluations tracking the change of guise. Of particular relevance to the present study, despite the fact that the fricatives used in McGowan and Babel (2020) were not different across guises, participants often commented on how the fricatives participated in communicating the purported social identity. This work raises the likelihood of at least two levels of sociophonetic perception and suggests that further work is needed to understand the role of awareness, and the necessity of deception, for the "complex, multi-layered process" of perception (Babel, this volume).

This paper reports an audiovisual matched guise experiment with both standard 'hidden' and novel 'unhidden' instruction conditions. The basic task is a replication of Strand and Johnson (1996). Listeners are asked to identify an ambiguous word as *sack* or *shack* on a [ʃ]-[s]continuum given manipulated beliefs about the gender identity of the talker (Stecker & D'Onofrio, this volume; Tripp & Munson, 2022). As described above, numerous previous replications have found that listeners perceive more of the ambiguous continuum as [ʃ] when they believe the speaker identifies as a woman and more as [s] when they believe the speaker identifies as a man and that, furthermore, this effect is bi-directional, with fricative type influencing perception of gender for an ambiguous voice (Bouavichith et al., 2019). Unusually, participants in the present study's 'unhidden' Instructions condition were briefed, in the instructions, about the guise manipulation. They were instructed that the man or woman in the photo was not associated with the voice they were listening to. (Campbell-Kibler, 2021), using a similar manipulation, finds that listeners have some ability to disregard social information when making accentedness or attractiveness judgments but that influence of available social information, particularly from the voice, is difficult to disregard completely. In the present study, participants were asked to provide a *sack/shack* lexical decision either with, or without, explicit instructions to disregard the visual stimulus.

Method

Participants

120 participants (self-identified 59 female, 61 male; ages 20 to 75) were recruited to complete the online experiment online. These participants were recruited through prolific.co and had provided language history and demographic data as part of Prolific's general pre-screening questionnaire. Participation was restricted to a standard sample of desktop computer users located in the USA, aged 18-100, who spent most of their childhoods in the US, first language English, primary language English, with no known language or hearing difficulties. Additionally, due to an audio playback restriction imposed by Apple Computer, the Safari web browser could not be used. Participants were urged only to accept the task if they could do so in a quiet space, free from distractions and wearing headphones for the 6 to 10 minute duration of the experiment (average time 6:51). Headphone usage was not verified within the instrument. No participants' data were excluded from analysis. Participants were paid $3 for their time, pro-rated from a projected rate of $20/hour (actual rate: $26.29/hour). This same instrument was piloted in the Speech Perception lab of The Ohio State University and, while reaction times online were generally slower than in-person, results from the online administration were generally consistent with results collected

under laboratory conditions. Four participants were excluded for low accuracy rates (below 85%).

Stimulus Materials

Auditory Stimuli. The auditory stimuli used in this study are the same wav-format files used in (Bouavichith et al., 2019). The stimuli, which were generously shared with us, contain two parts, both of which are drawn from synthetic continua: a fricative onset and a VC rime. The fricative onsets comprise a synthetic six step /ʃ-s/ continuum. These steps were generated with the Klatt Synthesizer in Praat (Boersma, 2001) using parameters identical to Munson (2011) ranging between the values of Munson's second and eighth continuum steps (which were, in turn, based on the parameters used in Strand and Johnson (1996)). Centers of Gravity ranged from 3.2 kHz (/ʃ/-like) to 7 kHz (/s/-like).

For the VC rime, two additional continua were modified from natural productions of /[æk]/ spoken by one male-identifying and one female-identifying talker in the carrier phrase "Say sack again". Two five-step gender continua were created by evenly spacing mean F0 across consecutive steps such that the male /aek/ continuum increased F0 frequency and formant spacing from their unmodified values while the female talker's /aek/ continuum decreased both parameters from unmodified. Each synthesized fricative token was concatenated with each CV rime of /aek/ resulting in a total of 60 unique auditory stimuli. These manipulations are described in greater detail in Bouavichith et al's section 2.1 and summarized visually in Table 1. Unlike MGT studies that ask a talented, multi-dialectal talker to consciously change their speech style (e.g. Wright, 2023), the talkers for these stimuli were asked to produce stimuli representing the gender identity they habitually perform. As these talkers were advanced doctoral students in a linguistics PhD program, some of the elements of such an identity are likely available to conscious reflection, but many of these indexical features are likely implicit even for them.

	F0	F2:F3		/s/ → /ʃ/		
	135Hz	1.0		Sack	s/shack	Shack
				.	.	.
Male Voice				.	.	.
				.	.	.
	210Hz	1.2	↓	Sack	s/shack	Shack
	190Hz	1.0		Sack	s/shack	Shack
				.	.	.
Female Voice				.	.	.
				.	.	.
	90Hz	.83	↓	Sack	s/shack	Shack

Table 1
Bouavichith et al. (2019) auditory stimulus continua

Explicit Evaluations of Auditory Stimuli

Voices carry social information. To better understand how our auditory stimuli might influence participants' perception of the identities of the two talkers, we elicited explicit social ratings. Participants who completed the in-person pilot version of the experiment were asked to make judgements regarding the gender, gender prototypicality, and sexuality of a natural production of *sack* produced by each of the two talkers. Participants listened to the recording and then selected from a fixed set of responses; no free form responses were elicited.

Participants' judgments of the female voice indicate general agreement about the gender identity of the speaker. Most participants (87%) indicated the gender of the speaker to be female, 2 participants indicated trans female more specifically, and 3 were unsure or otherwise unable to determine the gender of the speaker. For the female voice, average prototypicality ratings (in which, for a given gender, 0 is least prototypical, and 5 is most prototypical) were 4.3/5 if the participant had indicated 'female', and 2.75/5 if the participant had indicated 'trans female'. Judgements of the voice's sexuality were more variable, with 54% indicating they were unsure, 40% indicating the speaker was most likely heterosexual, and 1 participant each indicating the speaker was most likely bisexual or another sexuality.

Participants' judgments of the male voice suggest similar agreement.76% of participants indicated the gender of the speaker to be male, while 1 participant indicated trans male, and 21% were unsure of the gender of the speaker. Average prototypicality ratings were lower for the male speaker but similarly consistent: 3.6/5 if the participant had indicated the voice belonged to a 'male' speaker, and 2/5 if they had indicated the person speaking was a 'trans male'. As with the female voice, judgements of the voice's sexuality were more variable. 65% indicated they were unsure, 14% indicated the speaker was most likely heterosexual, and 16% indicated homosexual and, again, 1 each indicating the speaker was most likely bisexual, or another sexuality not listed.

Visual Stimuli. The visual stimuli used in this study, again identical to the images used in (Bouavichith et al., 2019), are shown in Figure 1. These included two face images, used for the guise manipulation, which were retrieved from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2015), a resource containing high-resolution, normed images of faces indexed by gender and ethnicity. As in Bouavichith et al., CFD-WF-015-006-N was selected as the representation of the gender-protypical female talker and CFD-WM-029-023-N was selected as the representation of the gender-prototypical male talker. Both images were converted to greyscale at the command line using ImageMagick (LLC, n.d.).

Additionally, two gray-scale line drawings were used as visual representations of *shack* and *sack*. These images were used in place of orthographic targets both to maintain consistency with Bouavichith et al's design and to facilitate future eye tracking investigation of this phenomenon.

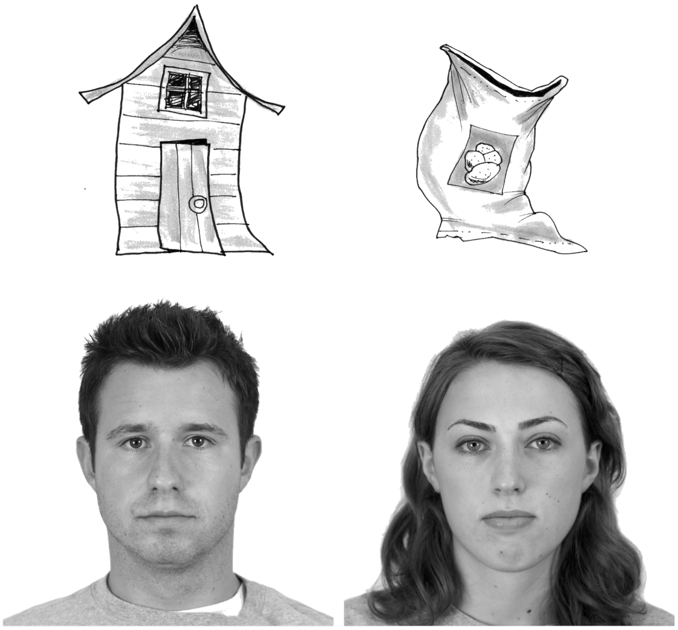

Figure 1. Visual Stimuli comprised *shack* and *sack* targets (top) and a gender-protypical 'male' and 'female' face (bottom)

Procedure

The experiment was created in OpenSesame v3.3 (Mathôt, Schreij, & Theeuwes, 2012) and exported for the web using OSWeb v1.4.14.0. Modifications to the experiment included translating portions of the python code into JavaScript and adding code to collect Prolific IDs and provide proof of completion to Prolific at the end of the experiment. This experiment was hosted on a JATOS (Lange, Kuhn, & Filevich, 2015) instance hosted on an Ohio State University Linguistics Department server.

Participants received a link to the experiment via Prolific and used their own computers, keyboards, and headphones to complete the experiment.

In a between-subjects design, participants were randomly assigned to one of two awareness conditions. These conditions differed only in the initial information provided as to the nature of the experiment. Participants in the *hidden* condition experienced a standard Matched Guise task. They were given no information about the task or the stimulus materials beyond the general instructions for completing the experiment: listen to the voice, press 'z' if you heard the word on the left, press 'm' for the word on the right. Participants in the *unhidden* condition also received this instruction and were given a partial debriefing regarding the task. They were informed that– while they would see faces onscreen while hearing words– the voices in a given trial were not produced by the person shown in the images, the images had been downloaded from a database of photographs created at the University of Chicago for experimental use, and that the auditory and visual stimuli were in no way related to each other. Participants were divided equally among these two conditions. Neither awareness condition was informed about the synthetic nature of the auditory stimuli.

Additionally, participants were assigned to one of two gender congruity conditions. Although the manipulated rimes sounded gender ambiguous to us, and had been rated as ambiguous by Bouavichith et al. (2019)'s pilot participants, the possibility remained that the voices, particularly at the end-points, might be perceived incongruously with the faces as in, for example, **?**[1].

In congruous trials, the faces and voices were paired such that participants were

[1] We are choosing the words 'congruous' and 'incongruous' intentionally to suggest faces and voices may pattern together in particular ways in listeners' experience and perception with no implied claim that voices may 'match' or 'mismatch' in some way that suggests either experimenters or participants have veridical access to an objective reality

only presented with auditory stimuli from the female talker's continuum alongside the female face and tokens from the male talker were only presented alongside the male face. In incongruous trials, by contrast, auditory stimuli from the female talker's continuum were only ever presented alongside the male face and tokens from the male talker's continuum were only ever presented alongside the female face. Half of participants were randomly assigned to each congruity condition. Each participant heard all 60 auditory stimuli; 30 paired with the male face and 30 paired with the female face.

In each trial, participants were shown one of the two faces for 1500ms. Following this initial presentation, the face remained onscreen and was flanked by the *shack* and *sack* images. Simultaneously, one of the auditory stimuli was played over the headphones. The trial ended when the participant pressed an appropriate key on their physical keyboard and their response and reaction time data were uploaded to the JATOS instance. In both congruous and incongruous conditions, all 60 unique trials (30 per face) were presented twice to each participant for a total of 120 trials.

Predicted Results

Face: male or female

Consistent with previous results, we expect to replicate the Strand effect; in general, we anticipate that more of the [ʃ]-[s] continuum will be heard as [ʃ] when participants are shown the female face and more to be heard as [s] when participants are shown the male face. However, these general predictions about the Face presentation when the congruence of auditory and visual components of the guise are taken as a whole.

Congruence: pairing of face and voice

To our knowledge, the influence of congruence has not been directly investigated for listeners' joint perception of gender and fricative place. Johnson, Strand, and D'Imperio (1999) tested AV integration of Male and Female faces with prototypical and non-prototypical gendered voices in a vowel quality perception task. They find what appears to be an incongruence effect with the prototypical male voice; listeners reported no difference in perceived vowel quality with this voice in either Face condition (Johnson et al., 1999, p. 376, Table 4). For this reason, we anticipate a replication of the Strand effect on fricative identification in our congruous trials (when Face and Voice do not conflict) but a failure to replicate for the incongruous trials (when Face and Voice provide conflicting social information). This difference may be stronger with the male voice, given both Johnson, Strand, and D'Imperio's finding but also King (2021).

We make a similar prediction for reaction times. Johnson et al. (1999) did not collect reaction time data, but McGowan (2011) reports longer reaction times for incongruous trials, albeit in a very different task, and **?** would seem to suggest that this should hold for listeners' identification of fricatives on a [ʃ]-[s] continuum. Specifically, we predict longer reaction times, in general, for the Incongruous conditions. Furthermore, when gender information is most clear, at gender continuum steps 1 and 2 for the Male talker and at gender steps 4 % 5 for the Female talker, and in conflict with the presented Face, listeners' response times should be slower.

Since strong phonetic correlates of gender, F0 and F3, have been manipulated over the course of the VC rime continua in our auditory stimuli, we anticipate that the effect of incongruous face and voice should be strongest for the natural end points of the continua where the difference is most salient and weaker as phonetically-cued gender information becomes more ambiguous. These stimuli have been independently normed for ambiguity (Bouavichith et al., 2019, p. 1040, Table 1) in the 2nd and 3rd levels of the rime continua. This means we anticipate an interaction between Face and Rime step but only in the incongruous trials and only at the extremes of the rime continuum.

Guise: Hidden or Unhidden

The primary goal of this experiment was to explore the role of listener awareness and control in the matched guise technique. The tremendous care researchers take to ensure that the guise manipulation is hidden from participants suggests a kind of imagined fragility. From this view: listeners who become aware of the guise manipulation will have introspective access to and deliberative control over the influence of visual social information on perception. If this is true, explaining the guise manipulation, in the unhidden condition, should have a strongly negative effect on the Strand effect. Alternatively, if the influence of social information is not available to introspection or deliberative control, we should see no change between the (traditional) hidden matched guise and the unhidden guise.

Additionally, we speculate that there may be a response time difference between the Hidden and Unhidden guises even if there is no apparent difference in percept between the conditions. It can certainly be the case that participants will arrive at the same behavioral responses via different cognitive processing paths, perhaps drawing on different levels of knowledge and awareness, and that these differences may be visible in response times between the Instruction conditions.

Results

Participants provided a total of 14,400 trials (120 trials from each of 120 online participants). It is not clear what it means to be 'accurate' when asked to perceive fricatives from a continuum so accuracy was calculated only for responses to the [ʃ] and [s] endpoints. Overall, participants were highly accurate (96.8%) but four participants were excluded from further analysis for accuracy below the pre-determined 85% threshold reducing the total number of trials to 13,920. Trials were coded as correct if the participant responded 'shack' to onset step 1 or 'sack' to onset step 6. The four excluded participants all scored 67.5% accuracy or lower.

An additional 50 trials were excluded due to response times that were either too fast or too slow. To reduce the effects of response time outliers on subsequent analyses, all response times shorter than 50 ms (N=0) and longer than 5000ms (N=50) were excluded. The 5000ms response time cutoff was used instead of imposing an in-experiment time limit on responses to a trial to ensure that participants were required to respond to each trial. After these exclusions were applied, the data from 13,870 trials were left for analysis (approximately 96.3% of the initial data set). The majority (96.8%) of the remaining response times were within a range between 200 and 2000ms. To increase normality of the distribution of response times across participants, the remaining response times were log-transformed.

[ʃ]-[s] **Percepts**

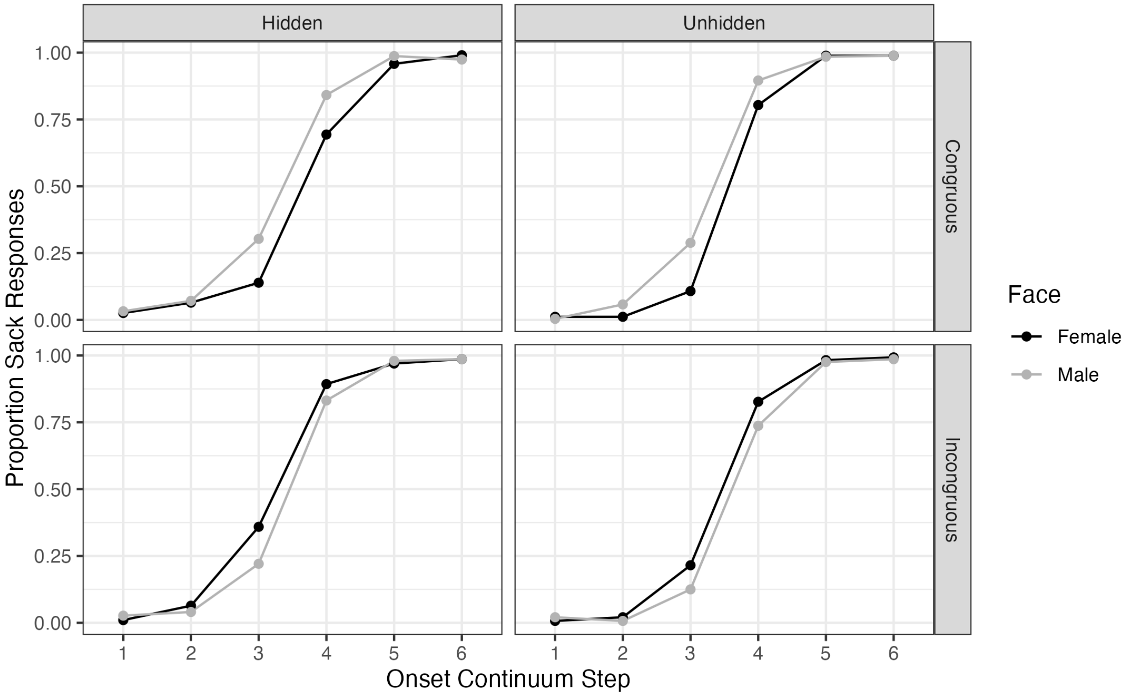

Figure 2. Proportion 'sack' responses plotted as a function of [ʃ]-[s] fricative (Onset) continuum steps and purported gender presented by the face.

Figure 2 presents listeners' percepts on this 2AFC task. The horizontal axis in each of these four plots is the fricative (syllable Onset) continuum step. Step 1 of the continuum is most [ʃ]-like, step 6 is the most [s]-like, steps 3 & 4 are the most ambiguous. Darker lines in Figure 2 present trials using the female Face; lighter lines present trials using the male Face. The Hidden and Unhidden instruction conditions are represented by the left and right columns of figures, respectively. The rows present the Congruous blocks where Face and Coda speaker voice shared a gender identity (top) and Incongruous trials where Face and Coda speaker voice mismatched in gender identity (bottom).

A successful replication of the Strand effect would mean that a higher proportion of the ambiguous stimuli would be heard as [s] when the purported gender suggested by the face is male than when the face is female. This pattern appears to hold in both the Hidden and Unhidden conditions, but only when gender identity of the talker who produced the CV rime stimuli was congruous with the gender presented in the visual portion of the guise. From Figure 2 it would appear that listeners' reported percepts more closely track the voice of the talker than the face in the picture when these sources of information are incongruous.

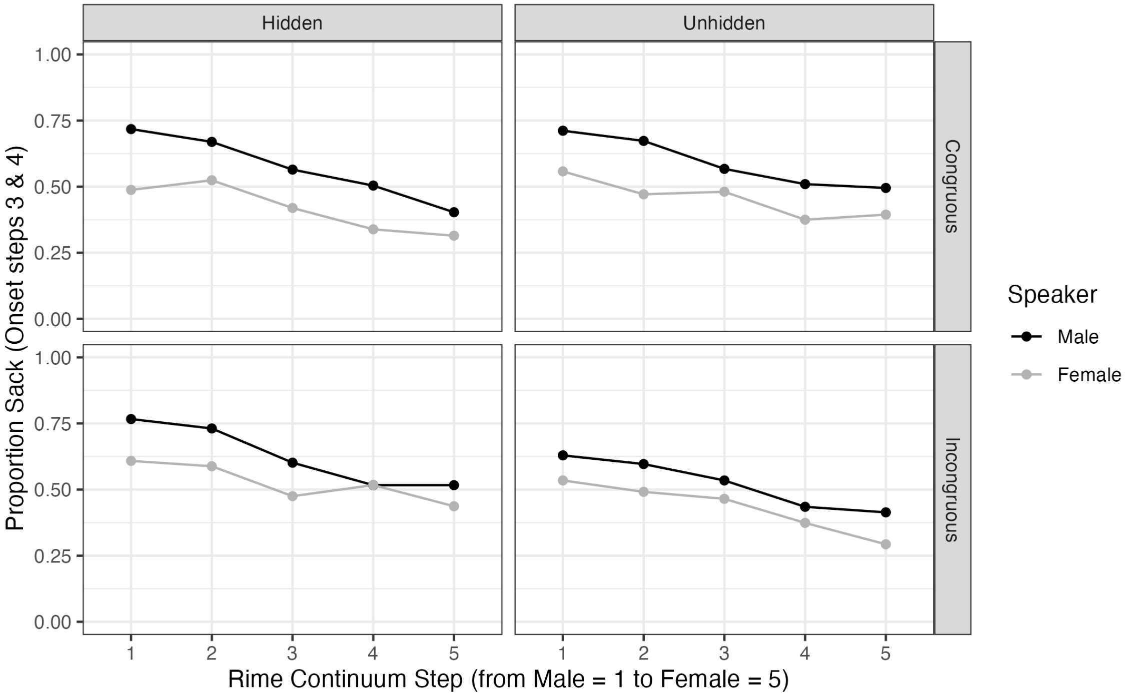

Figure 3. Proportion 'sack' responses on ambiguous fricative trials plotted as a function of CV rime continuum steps and gender identity of stimulus talker.

We predicted that, since strong phonetic correlates of gender have been manipulated over the course of the VC rime continua, the effect of incongruence should be strongest for the end points of the continua where the social information presented by the voice is, presumably, most salient and weaker as phonetically-cued gender information becomes more ambiguous. Figure 3 suggests that this prediction is at least partially borne out. Figure 3 plots proportion 'sack' responses to the ambiguous portion of the [ʃ]-[s] continuum (steps 3 & 4) as a function of rime continuum step. The meaning of line color has changed in this figure. Dark lines represent the male talker and lighter lines represent the female talker. Step 1 on this continuum includes the most natural token for the male talker and the most manipulated token for the female talker while step 5 includes the most natural token for the male talker and the most manipulated token for the female talker. As before, columns present the Hidden and Unhidden conditions while rows present the Congruous and Incongruous blocks.

In a 2AFC task with unbiased stimuli, chance is 50%. Responses at the .5 line in figure 3 suggest that the ambiguous fricatives remained ambiguous while responses that tend to be above this line reflect a tendency toward [s] percepts and responses that tend to be below this line reflect a tendency toward [ʃ]. Across all 4 conditions we observe a declination from highest-proportion [s] responses in step 1 of the F0 continua to lowest in step 5. When face and voice were congruous, virtually all male-voice (and male face) responses are above or at 50% 'sack' and virtually all female-voiced (and female face) responses are at or *below* 50% 'sack'. This is the same pattern that can be observed at Onset continuum steps 3 & 4 in figure 2. It is not clear from Figure 3 alone if there is any difference at all between the Congruous and Incongruous conditions. However, it is important to recall about the bottom row of this figure that male talker responses in the incongruous trials were presented with a female face while female talker trials were presented with a male face. Even a weakly-significant Strand effect would predict that the female talker, particularly on the more ambiguous continuum steps, should show more 'sack' responses consistent with having been shown a male face and no such effect is evident in this plot.

Indeed, a striking feature of figures 2 and 3 is how the apparent influence of gender information flips between congruous and incongruous conditions in the former but remains essentially constant in the latter. Taken together, these plots suggest that cues to gender in F0 is a stronger predictor of listeners' reported percept in this matched guise task than just the purported gender of the face.

Finally, the main objective of this experiment was to explore the role of listener awareness in the matched guise technique. Here again there may be differences between the congruous and incongruous conditions that will be better understood through quantitative analysis, but the overall trend is clear. If there is an effect of explaining to participants that the voice and face in the matched guise task are unrelated to each other, that effect is so weak as to be essentially invisible in these visual interrogations of the data. Categorical responses in the Hidden and Unhidden instruction conditions appear to be identical.

Logistic Regression and Quantitative Analysis

These qualitative assessments of listener responses can be examined further through quantitative analysis. Through model comparison we initially arrived at a logistic mixed model to predict percept with Congruity condition, instruction condition, Onset step, Face and Rime step with interactions for all but Rime step. This model was justified by model selection but given the notorious difficulty of interpreting a 4-way interaction and the preceding visual interrogation of the data, we opted to separate Congruence into a pair of 3-way models. Using `glmer()` lme4, we divided the data into congruous and incongruous subsets and fitted a pair of logistic mixed models (estimated using ML and BOBYQA optimizer) to predict percept with Instruction condition, Onset.step, Face and Rime step (`percept ~ Instruction * Onset.step * Face + Rime.step`) The models included random intercepts for subject. All categorical predictors were coded using contrast coding.

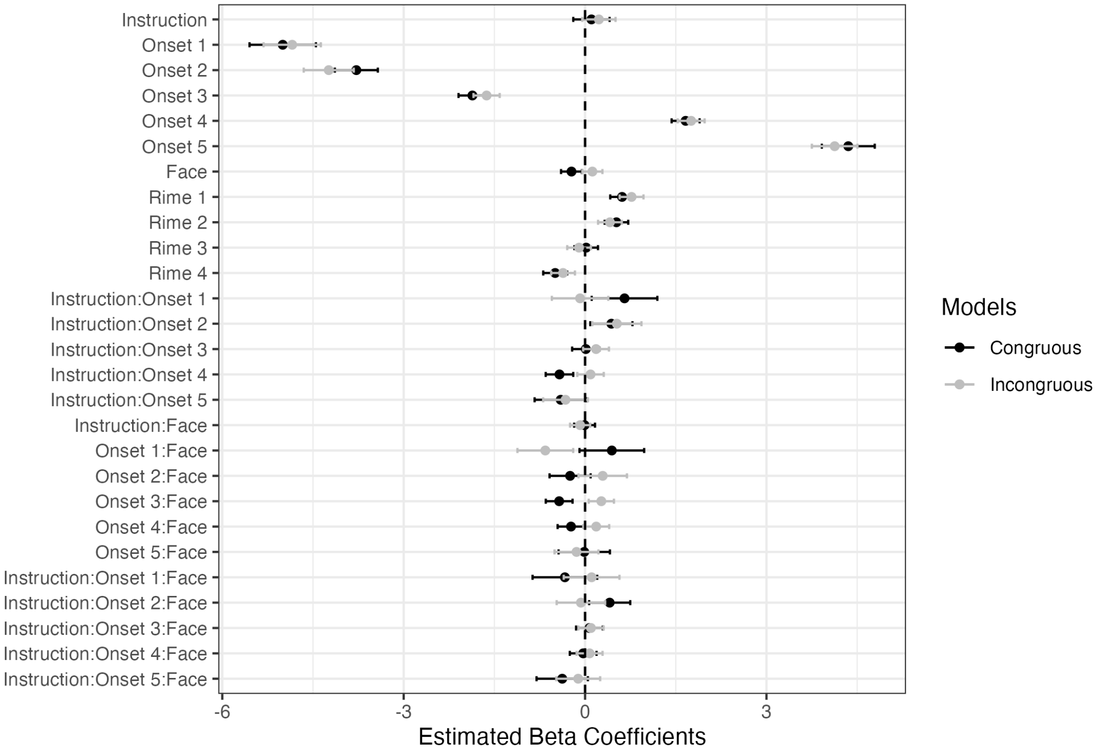

Figure 4. Estimated Beta coefficients for listener responses in the Congruous (black) and Incongruous (gray) logistic regression models plotted with 95% confidence intervals.

Beta coefficients for the two separate logistic mixed models are plotted together in Figure 4. Terms plotted to the left of the dashed zero line have a negative influence on 'sack' percepts in the model while terms plotted to the right have a positive influence. As a consistency check we can observe that the levels of the Onset continuum behave in precisely the expected ways and all levels are statistically significant predictors of percept in both models. Onset step 1 ([ʃ]) is negatively associated with 'sack' responses and significant in both the Congruous ($\beta = -5.00, SE = 0.28, p < 0.001$) and Incongruous ($\beta = -4.84, SE = 0.24, p < 0.001$) models. Onset step 5 ([s]) is positively associated with 'sack' responses and significant in both the Congruous ($\beta = 4.35, SE =, p < 0.001$) and Incongruous ($\beta = 4.12, SE = 0.19, p < 0.001$) models.

As visual inspection of the data suggests, this study includes a replication of the

Strand effect in the Congruous condition. There is a main effect of Face in the model ($\beta = -0.22, SE = 0.09, p < 0.05$). Face is negatively associated with 'sack' responses suggesting that, with these stimuli, at least, it is more appropriate to understand the effect of Face as an increase of 'shack' responses given the female Face. The inclusion of the interaction term for Onset and Face allows us to see that the effect of Face is greatest on the ambiguous Onset steps 3 ($\beta = -0.43, SE = 0.11, p < 0.001$) and, to a lesser extent, 4 ($\beta = -0.23, SE = 0.11, p < 0.05$).

However, the Strand effect observed in the Congruous condition is not attributable entirely to the main effect of Face. Rime F0 is also significant; Rime level 1, the male end of the continuum, is positively associated with 'sack' responses ($\beta = 0.61, SE = 0.10, p < 0.001$) as is Rime level 2 ($\beta = 0.52, SE = 0.10, p < 0.001$). Rime level 3, where the continuum is most gender ambiguous, is not statistically significant. Rime level 4, on the female end of the continuum, is negatively associated with 'sack' responses and significant ($\beta = -0.49, SE = 0.10, p < 0.001$).

Unsurprisingly, the Strand effect has not been replicated in the incongruous condition. As is visible in the bottom row of Figure 2, the effect of Face on 'sack' responses is not significant. The interaction of Onset and Face also behaves quite differently in the Incongruous model. Onset x Face is negatively associated with 'sack' responses at Onset step 1 ($\beta = -0.66, SE = 0.24, p < 0.001$) but positively associated with 'sack' responses and significant at Onset step 3 ($\beta = 0.27, SE = 0.11, p < 0.05$).

Interestingly, the significant effect of Rime observed in the Congruous model also holds, nearly identically, in the Incongruous model. Rime level 1, the male end of the continuum, is again positively associated with 'sack' responses ($\beta = 0.77, SE = 0.10, p < 0.001$) as is Rime level 2 ($\beta = 0.41, SE = 0.10, p < 0.001$). Rime level 3 is also not statistically significant in the Incongruous model. Rime level 4, on the female end of the continuum, is negatively associated with 'sack' responses and significant

$(\beta = -0.36, SE = 0.10, p < 0.001)$.

Finally, the quantitative analysis of the primary objective of this experiment, exploring the effect of unhiding the matched guise manipulation from participants, largely supports the qualitative analysis. As can be observed in Figure 4, there is no significant main effect of Instruction condition in either model. Still, a somewhat more nuanced picture emerges from the interactions of Instruction condition with Onset and the 3 way interaction of Instruction, Onset, and Face in the Congruous trials. The interaction of Instruction with Onset is significant, or nearly so, at every step of the fricative continuum other than the most significant. In the [ʃ]-like portion of the continuum, the interaction with face is positively associated with 'sack' responses at step 1 $(\beta = 0.65, SE = 0.28, p < 0.05)$ and 2 $(\beta = 0.44, SE = 0.18, p < 0.05)$. The interaction of guise with the most ambiguous onset step is not significant $(\beta = 0.011, SE = 0.12)$. The interaction of Instruction with Onset step 4, on the [s] end of the continuum is negatively associated with 'sack' responses and statistically significant $(\beta = -0.43, SE = 0.12, p < 0.001)$. Instruction x Onset step4 is also negatively associated with 'sack' responses but does not reach significance at the standard alpha level $(\beta = -0.40, SE = 0.22, p = 0.067)$. The 3-way interaction of Instruction x Onset x Face is positively associated with 'sack' responses at step 2 $(\beta = 0.41, SE = 0.17, p < 0.05)$ and weakly, but not significantly, negatively associated with 'sack' responses at step 5 $(\beta = -0.38, SE = 0.21, p = 0.080)$.

There is also no main effect of Instruction in the Incongruous trials. The 3-way interaction of Instruction x Onset x Face, while justified by model selection for inclusion in this model, also does not reach statistical significance. However the 2-way interaction of Instruction with Onset step is positively associated with 'sack' responses at Onset step 2 $(\beta = 0.53, SE = 0.21, p < 0.05)$ and approaches significance at step 3, where it is weakly positively associated $(\beta = 0.18, SE = 0.11, p = 0.095)$ and step

5 where it is weakly negatively associated ($\beta = -0.32, SE = 0.19, p = 0.086$).

Response Times

As with the logistic regression models, we again opted to separate Congruence into a pair of 3-way models for linear mixed model analysis of our log-transformed response time data. Using `lmer()` lme4, we reused the congruous and incongruous subsets created for the logistic regression models and We fitted a linear mixed model (estimated using REML and nloptwrap optimizer) to predict logRT with Guise, Onset, Face and Rime (`logRT ~ Instruction * Onset * Face + Rime`). The models included random intercepts for subject. All categorical predictors were coded using contrast coding. Beta coefficients for both models are plotted in Figure 5. Terms plotted to the left of the zero line are associated with a decrease in log response time while terms plotted to the right of the zero line are associated with an increase in log response time. Notably, the longest response times are associated with the most ambiguous steps of the [ʃ]-[s] onset continuum. Onset step 3 is positively associated with response time and significant in both the congruous ($\beta = 0.08, SE = 0.007, p < 0.001$) and incongruous ($\beta = 0.07, SE = 0.007, p < 0.001$) models. The same is true of step 4 in the congruous ($\beta = 0.07, SE = 0.007, p < 0.001$) and incongruous ($\beta = 0.07, SE = 0.007, p < 0.001$) models as well. On the other hand, steps 1, 2, and 5 are all negatively associated with response time and also significant in both models (see Figure 5).

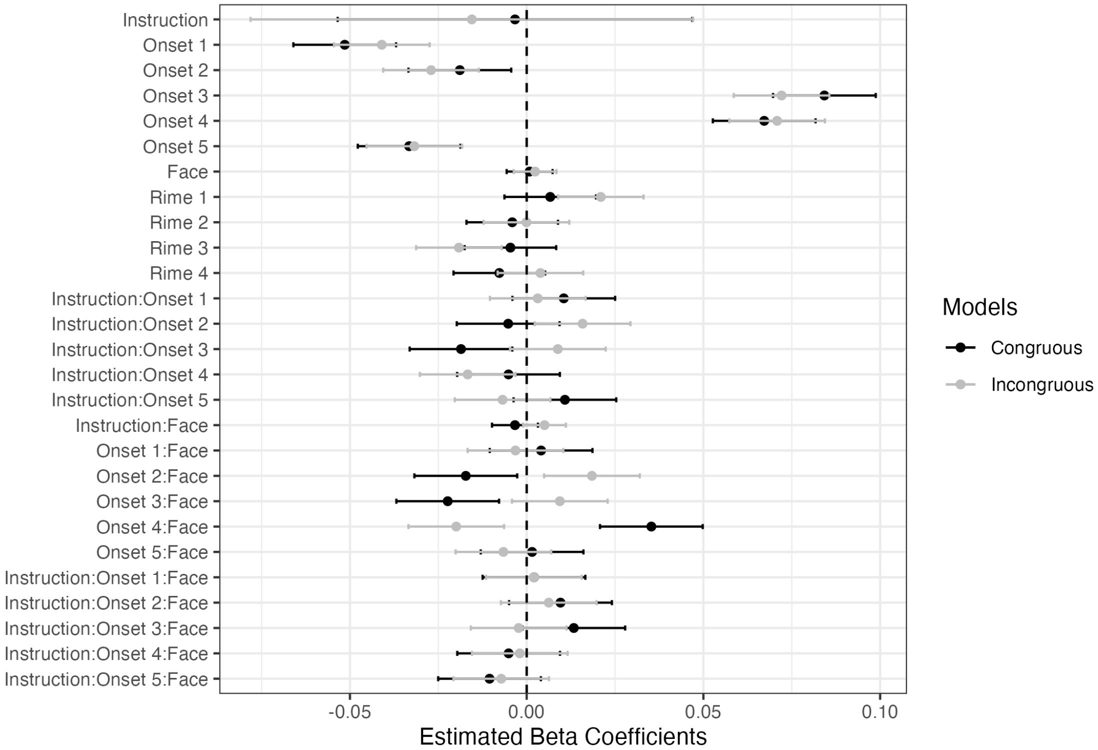

Figure 5. Estimated Beta coefficients for log-transformed response times in the Congruous (black) and Incongruous (gray) linear regression models plotted with 95% confidence intervals.

We predicted overall slower response times in the Incongruous than Congruous conditions and this prediction is not borne out by the data. Apart from generally higher variability in the incongruous conditions, there is no positive or negative trend in response times between the two Congruity models. For example, within the Incongruous model response times given the interaction of Onset step 3 * Face are longer ($\beta = -0.009, SE = 0.007, p = 0.17$), which would seem to support our prediction, but response times for Onset step 4 * Face are shorter ($\beta - 0.02, SE = 0.007, p < 0.01$), the opposite of what we predicted. The exact opposite pattern appears within the Congruous model where response times are shorter given Onset 3 * Face ($\beta = -0.02, SE = 0.007, p < 0.01$) but longer given Onset step 4 * Face ($\beta = 0.04, SE = 0.007, p < 0.001$). These crossing patterns can be seen in Figure 5.

Given the replication of the Strand effect in the Congruous, but not the Incongruous conditions described in the previous section, it may be notable that there is a significant main effect of Face in the Congruous model where it is negatively associated with response time ($\beta = 0.22, SE = 0.08, p < 0.05$) and not significant in the Incongruous model.

Discussion

The question that motivated this study was a desire to understand the role of listener awareness and control in the matched guise technique. We believe that the careful measures researchers generally employ to obscure the nature of the guise manipulation from participants is attributable to a long-held assumption in the sociolinguistics literature that social knowledge is high-level knowledge, available to introspective control, and that this differs from linguistic knowledge which is low-level knowledge, unavailable to control (Campbell-Kibler, 2016). The results of the present study are inconsistent with this imagined fragility of the influence of social knowledge. Revealing the nature of the guise manipulation did not significantly influence listener responses in either the congruous or incongruous conditions. Nor did this revelation have a significant influence on response times in either condition.

The finding that the Matched Guise effect holds for speech perception both when hidden from the participant and when unhidden is inconsistent with a model of processing in which social knowledge simply acts as a filter on linguistic knowledge. Social knowledge influences perception even when listeners are aware that it is, or may be, false. This result parallels previous results for accentedness and attractiveness judgments (Campbell-Kibler, 2021). A similar result may be present, for social information, in the within-participants guise manipulation of McGowan and Babel (2020). In that study, the authors use participants' metalinguistic commentaries to assess the extent to which the guise manipulations were or were not 'believed'. The results of the present study suggests that that belief may be irrelevant. The present result also gives additional context to studies demonstrating influence of social knowledge even when listeners have no reason to believe the guise manipulation (Hay & Drager, 2010; Hay, Nolan, & Drager, 2006; Niedzielski, 1999). It is unclear whether social knowledge will prove to be as resilient to awareness as the obligatory McGurk effect (McGurk & MacDonald, 1976) which persists even when participants actively identify that the face and voice in the experiment are mismatched (Green, Kuhl, Meltzoff, & Stevens, 1991), but the suggestion is that it will.

The gender identity of the talker who produced the VC Rime supplemented Face in the Congruous conditions to make the Strand effect even stronger; the mechanism may prove similar to the way lip-rounding accentuates the backness of back vowels. In the Incongruous conditions, though, listeners' perception of the [ʃ]-[s] continuum tracked the VC Rimes, rather than the purported gender of the Face. This pattern was strongest in the least-ambiguous portions of the Rime continuum and weakest in the most-ambiguous. In a sense, by separating trials by congruity of face and voice we have replicated Strand and Johnson (1996)'s exp1 and exp2 simultaneously. One wonders, looking back at their exp2, whether this classic result was *also* a congruous condition in which listeners had sufficient gender information from the voice to supplement the purported information from the Face. Even the non-prototypical voices used in that study did pattern, in exp1, in weakly gendered ways. This finding may provide some insight into recent failures to replicate the original Strand effect (Schellinger et al., 2017; Wilbanks, 2022).

The phonetic correlates of gender manipulated in the VC rimes for this study are F0 and formant ratios. However, these may not be the only cues listeners are drawing upon with their knowledge of US English. Surely, F0 and vowel formant ratios *can* *be* important to listeners, just as voice onset time and vocal fold vibration can be important cues to the voicing of /t/ and /d/. But as Lisker (1986) catalogs, there are 16 cues to this apparently simple feature in English, any of which might be sufficient to communicate voicing, but none of which is required. In this study we have used manipulated stimuli that obscure, over the course of two gender continua, the gender identity of the talker who produced the basis token for that continuum. At an explicit level, these continua *sound ambiguous* to the experimenters in much the way that **?**'s stimuli do not sound obviously mismatched. But our perception results suggest that listeners are still aware, albeit implicitly, of the gender identity we have attempted to obscure by altering the phonetic correlates of gender.

Conclusion

Decades of research since Strand and Johnson (1996)'s original finding have demonstrated that a visual cue can shift fricative perceptions when paired with an ambiguously-gendered voice (although cf Munson 2017 and Wilbanks 2022). Bouavichith et al. (2019) even demonstrated with eye-tracking that this effect is fast and bi-directional. One could come away from Strand & Johnson's exp1 and exp2 and subsequent replications with a theoretical model in which visually-cued social information and phonetically-cued social information exert equivalent influence on speech perception. Prototypically-gendered voices can shift perception of a [ʃ]-[s] continuum and prototypically-gendered visual information can as well. However, listeners' behavior in our Congruous and Incongruous conditions is inconsistent with such a model and suggests, instead, that when visually-cued and phonetically-cued social information are in congruence, they can enhance one another. When, on the other hand, these information sources conflict, it is the phonetically-cued social information that will dominate (Campbell-Kibler, 2021).

It is unlikely that fricatives are unique in this respect. For example, the incongruous results seen in this study are, perhaps, predicted by lack of Face effect for Johnson et al. (1999)'s vowel perception results in exp2 given a stereotypical face (particularly, in that study, for the male voice). As listeners, we do not have veridical access to the speech sounds intended by a talker. Instead, we must combine the speech signal with our phonological knowledge, lexical knowledge, social expectations, visual input, expectations of the social world (Babel, this volume) and other sensory information to arrive at a percept. The implication is that perception is more holistic than is dreamt of in our phonologies. Category boundaries, whether for speech sounds or social categories, are fuzzy and perception needs to be fast. We retain knowledge of, and use, detailed social and linguistic knowledge at both high and low levels of processing. Enumerating the phonetic correlates of gender may not be the wrong question, but it is certainly premature given the limitations of current theory to account for what listeners actually do. A better question is something like "what kinds of knowledge do listeners draw on during perception and when?"

(Barrett, Zimman, Davis, & Raclaw, 2014, p. 205) writes, "any assumption of essentialism will ultimately marginalize those individuals who do not fit the essentialist understandings of human behavior". It may not feel brutal or reductive to read May (1976)'s findings about large and small vocal tracts as if they refer to male and female vocal tracts, respectively, but it does necessarily imply that tall, long-necked women and short, squat-necked men need to find some other way of labeling themselves. The idea that male voices come from large bodies and female voices come from small bodies need not be literally true for the phonetic and perceptual correlates of size to become enregistered alongside other features in the creation of gendered personae (D'Onofrio 2020). Our prediction that incongruity in face and voice would slow listener judgments was not supported. It is tempting to interpret this as evidence that, unlike misleading coarticulatory information, listeners are aware of the diversity of gender expression, but this is not a question the current study can resolve.

What the current study can resolve is that listeners' social knowledge of speech is not delicate. The present result is equally inconsistent with a model that disregards social knowledge entirely and with any model of speech perception that presumes *all* social knowledge to be late, high-level, and available to introspective control. Part of what listeners know when they know a language includes the simultaneous patterning of 'linguistic' and 'social' information in a shared phonetic signal. Social knowledge is not a weakly-associated prime; Social knowledge and linguistic knowledge are deeply intertwined in speech perception and it is perverse to assume that the language subsystem underlying this ability would necessarily distinguish them.

References

Alpert, E. R. (2014). *Language, gender, and ideology in japanese professional matchmaking.* (Unpublished doctoral dissertation).

Babel, A. (this volume). A semiotic approach to awareness and control. *Journal of Sociolinguistics*, *42*(1), XXX.

Barrett, R., Zimman, L., Davis, J., & Raclaw, J. (2014). The emergence of the unmarked. *Queer excursions: Retheorizing binaries in language, gender, and sexuality*, 195–223.

Boersma, P. (2001). Praat. *a system for doing phonetics by computer. Glot International*, 341–345.

Bouavichith, D. A., Calloway, I. C., Craft, J. T., Hildebrandt, T., Tobin, S. J., & Beddor, P. S. (2019). Bidirectional effects of priming in speech perception: Social-to-lexical and lexical-to-social. *The Journal of the Acoustical Society of America*, *145*. doi: 10.1121/1.5101933

Boyd, Z., Fruehwald, J., & Hall-Lew, L. (2021). Crosslinguistic perceptions of /s/ among english, french, and german listeners. *Language Variation and Change*, *33*(2), 165–191.

doi: 10.1017/S0954394521000089

Calder, J. (2018). From "gay lisp" to "fierce queen": The sociophonetics of sexuality's most iconic variable. In K. Hall & R. Barrett (Eds.), *The oxford handbook of language and sexuality* (pp. 1–23).

Campbell-Kibler, K. (2005). *Listener perceptions of sociolinguistic variables: The case of (ing)* (Unpublished doctoral dissertation). Stanford University.

Campbell-Kibler, K. (2007). Accent,(ing), and the social logic of listener perceptions. *American speech*, *82*(1), 32–64.

Campbell-Kibler, K. (2016). Toward a cognitively realistic model of meaningful sociolinguistic variation. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 123–151).

Campbell-Kibler, K. (2021). Deliberative control in audiovisual sociolinguistic perception. *Journal of Sociolinguistics*, *25*(2), 253–271.

CHAN, K. L. R. (2021). Verbal guise test: Problems and solutions. *Academia Letters*.

Daniel, M. M., Lorenzi, M. C., da Costa Leite, C., & Lorenzi-Filho, G. (2007). Pharyngeal dimensions in healthy men and women. *Clinics*, *62*(1), 5–10.

Eckert, P. (2008). Variation and the indexical field 1. *Journal of sociolinguistics*, *12*(4), 453–476.

Eckert, P., & Podesva, R. J. (2021). Non-binary approaches to gender and sexuality. *The Routledge handbook of language, gender, and sexuality*, 25–36.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague, The Netherlands: Mouton.

Gnevsheva, K. (2017). Within-speaker variation in passing for a native speaker. *International Journal of Bilingualism*, *21*(2), 213–227.

Green, K., Kuhl, P., Meltzoff, A., & Stevens, E. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the mcgurk effect. *Attention, Perception, & Psychophysics*, *50*, 524-536. Retrieved from http://dx.doi.org/10.3758/BF03207536 (10.3758/BF03207536)

Grondelaers, S., & van Gent, P. (2019). How “deep” is dynamism? revisiting the evaluation of moroccan-flavored netherlandic dutch. *Linguistics Vanguard*, *5*(s1).

Hadodo, M. (this volume). Situating experience in social meaning: Ethnography, experiments and exemplars in the enregisterment of istanbul greek. *Journal of Sociolinguistics*, *42*(1), XXX.

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4), 865-892.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3), 351-379.

Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. Remez (Eds.), *The handbook of speech perception* (pp. 363–389).

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499.

Johnson, K., Strand, E. A., & D’Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of phonetics*, *27*(4), 359–384.

Käng, D. B. (2013). Conceptualizing thai genderscapes: transformation and continuity in the thai sex/gender system. In *Contemporary socio-cultural and political perspectives in thailand* (pp. 409–429). Springer.

King, E. T. (2021). *Speaker and group specificity in spoken word recognition* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The journal of abnormal and social psychology*, *60*(1), 44.

Lange, K., Kuhn, S., & Filevich, E. (2015, 06). "just another tool for online studies” (jatos): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), 1-14. doi: 10.1371/journal.pone.0130834

Lisker, L. (1986). “voicing” in english: A catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and speech*, *29*(1), 3–11.

LLC, I. S. (n.d.). *Imagemagick.* Retrieved from `https://imagemagick.org`

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*(4), 1122–1135. Retrieved from `https://doi.org/10.3758/s13428-014-0532-5` (Database can be accessed at https://chicagofaces.org/)

Mack, S., & Munson, B. (2012). The association between/s/quality and perceived sexual orientation of men’s voices: implicit and explicit measures. *Journal of Phonetics*, *40*(1), 198–212.

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

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314–324.

May, J. (1976). Vocal tract normalization for /s/ and /š/. *Haskins Laboratories Status Report on Speech Research*(SR-48), 67–73.

McGowan, K. B. (2011). *The role of socioindexical expectation in speech perception* (Unpublished doctoral dissertation). University of Michigan, Ann Arbor, MI.

McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, *58*(4), 502–521.

McGowan, K. B., & Babel, A. M. (2020). Perceiving isn’t believing: Divergence in levels of sociolinguistic awareness. *Language in Society*, *49*(2), 231–256.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Munson, B. (2011). The influence of actual and imputed talker gender on fricative perception, revisited (l). *The Journal of the Acoustical Society of America*, *130*(5), 2631–2634.

Niedzielski, N. (1999, March). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62-85.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of f_0 of voice. *Phonetica*, *41*(1), 1–16.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*(6), 2988–2998.

Pharao, N., & Kristiansen, T. (2019). Reflections on the relation between direct/indirect methods and explicit/implicit attitudes. *Linguistics Vanguard*, *5*(s1).

Podesva, R. J., & Kajino, S. (2014). Sociophonetics, gender, and sexuality. *The handbook of language, gender, and sexuality*, 103–122.

Rosseel, L., & Grondelaers, S. (2019). Implicitness and experimental methods in language variation research. *Linguistics Vanguard*, *5*(s1).

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative english-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–531.

Samoliński, B. K., Grzanka, A., & Gotlib, T. (2007). Changes in nasal cavity dimensions in children and adults by gender and age. *The Laryngoscope*, *117*(8), 1429–1433.

Schellinger, S. K., Munson, B., & Edwards, J. (2017). Gradient perception of children's productions of/s/and/θ/: A comparative study of rating methods. *Clinical Linguistics & Phonetics*, *31*(1), 80–103.

Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *Journal of phonetics*, *19*(3-4), 409–424.

Stecker, A., & D'Onofrio, A. (this volume). Recognizing uptalk: Memory and metalinguistic commentary for a sociolinguistic feature. *Journal of Sociolinguistics*, *42*(1), XXX.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of language and social psychology*, *18*(1), 86–100.

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In *Konvens* (pp. 14–26).

Tripp, A., & Munson, B. (2022). Perceiving gender while perceiving language: Integrating psycholinguistics and gender theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(2), e1583.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics*, *35*, 49-64.

Whalen, D. H. (1991). Perception of the english/s/–/ʃ/distinction relies on fricative noises and transitions, not on brief spectral slices. *The Journal of the Acoustical Society of America*, *90*(4), 1776–1785.

Wilbanks, E. (2022). *The integration of social and acoustic cues during speech perception* (Unpublished doctoral dissertation). University of California, Berkeley.

Wright, K. E. (2023). Housing policy and linguistic profiling: An audit study of three american dialects. *Language*.

Zimman, L. (2017). Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and/s. *Language in Society*, *46*(3), 339–370.