

Disentangling acoustic and social biases in creaky voice perception: The effects of f0 and face gender on creakiness ratings

Jeanne Brown*, Department of Linguistics, McGill University, Montreal, Canada, jeanne.brown@mail.mcgill.ca

Meghan Clayards, Department of Linguistics, School of Communication Sciences and Disorders, McGill University, Montreal, Canada

*Corresponding author.

Creaky voice has historically been associated with men’s speech, supported by acoustic studies. Since around 2010, however, sociolinguistic work alongside public discourse has perpetuated greater creaky voice use by women, typically implementing impressionistic coding. This study investigates whether this recent shift can be attributed to perceptual social and acoustic biases related to (perceived) speaker gender and pitch (f0), respectively. Using a matched-guise paradigm, 40 Canadian English listeners rated the perceived creakiness of the same modal and creaky voices—altered to have ambiguously gendered formants and median f0s (115, 135, 155 Hz)—paired with female and male faces. Bayesian regression analyses revealed strong effects of voice quality and moderate effects of f0: Creaky and lower-f0 stimuli were rated as creakier. No overall effect of face gender was found. However, a weak interaction between face gender and f0 suggests a possible gender prototypicality bias: At lower f0s, female faces were rated as slightly creakier than male faces and at higher f0s, male faces were rated creakier than female faces. These findings show that neither acoustic nor gender-based biases alone can account for widespread reports of women-led creaky voice use. Several possible explanations for this discrepancy are discussed.



1. Introduction

Along with the development of more sophisticated phonetic measures and statistical modelling in recent years, the analysis of voice quality has seen rapid growth (Garellek, 2022). Creaky voice in particular has seen more systematic investigations from both phonetic as well as sociolinguistic perspectives, given its diverse uses in languages of the world, for phonological contrast, prosodic and segmental structures, pragmatic and discourse-related function, indexicality, as well as the construction of personae (Davidson, 2021). More recently, sociophoneticians have raised questions about the perceptual dimensions of creaky voice and how they interact with acoustic and articulatory features of voice (e.g., Kreiman et al., 2014) as well as social evaluations, stereotypes, and expectations of voice (e.g., Ligon et al., 2019; White et al., 2024). Despite this growing body of work, a persistent mismatch remains between how creaky voice is produced and how it is perceived—particularly in terms of gendered associations. While traditional qualitative research (e.g., Henton & Bladon, 1988) and acoustic work (e.g., Klatt & Klatt, 1990; Gittelsohn et al., 2021) both link creak primarily with male speakers, more recent impressionistic studies and public discourse have emphasized its use by women (Yuasa, 2010; Podesva, 2013; Quenqua, 2012; for example). This paper addresses this gap by directly examining the effects of perceived speaker gender and variation in speaker f_0 on perceptual creakiness ratings. Focusing on creaky voice perception, this work aims to clarify how the social climate surrounding linguistic behavior and quantifiable acoustic cues jointly shape the way listeners process and interpret speech.

2. Background

2.1 Creaky voice phonetics

Within phonetics, *voice* or *phonation* broadly refers to the acoustic signal produced by the articulatory system (i.e., the glottis and supralaryngeal tract), grounded in acoustic physics and physiology (Garellek, 2019; Kreiman & Sidtis, 2011). In contrast, *voice quality* or *phonatory quality* denotes the holistic percept of this acoustic signal (or of a person's voice), grounded in cognition (Johnson & Babel, 2023; Kreiman & Sidtis, 2011). *Voice qualities* or *phonation types* represent specific patterns of vibration that can be situated along a uni-dimensional continuum of vocal fold aperture (Ladefoged, 1971), more traditionally, or within a multi-dimensional space collectively defined by various acoustic cues (Keating et al., 2023a; Kreiman & Sidtis, 2011), stipulated in more contemporary work on voice.

Creaky voice is a non-modal voice quality (or phonation type) typically characterized acoustically by a low fundamental frequency (f_0), irregular vocal pulses, and decreased transglottal airflow (e.g., Gordon & Ladefoged, 2001; Keating et al., 2015; Wright et al., 2019). Articulatorily, creaky voice is produced by increasing adductive tension and decreasing longitudinal tension of the vocal folds, allowing the vocal folds to be compressed and thick (Keating et al., 2015; Wright

et al., 2019). This configuration permits little airflow through the glottis, resulting in some vibration, albeit slow and aperiodic. Perceptually, creaky voice is described as having a low, rough, croaking or crackling sound to the ear, like “a rapid series of taps” (Catford, 1964, p. 34) or popping corn (Henton & Bladon, 1988). Keating et al. (2015) describe multiple sub-types of creaky voice, each characterized by different acoustic cues. At present, there is ongoing debate about whether these types are perceptually distinct or form one broad perceptual category (see Davidson, 2019b vs. Garellek, 2015; Gerratt & Kreiman, 2001) as well as how individual acoustic cues contribute to creaky voice perception (Khan et al., 2015). Considering the limited empirical perception data, we employ the broader term “creaky voice” to refer to the holistic percept of creakiness, rather than any specific acoustic sub-type.

2.2 The production-perception mismatch

Phonation types can be contrastive in some of the world’s languages (see Keating et al., 2023a for a review), similar to f_0 contrasts in tonal languages. Creaky voice can also serve a prosodic or segmental function in a variety of languages (see Davidson, 2021), marking intonational phrase boundaries (e.g., Crowhurst, 2018; Dilley et al., 1996; Redi & Shattuck-Hufnagel, 2001) or enhancing segmental contrasts (e.g., Garellek, 2022; Pierrehumbert, 1995). This paper focuses instead on non-contrastive creaky voice, especially discussing its sociolinguistic uses.

Voice quality is known to have many sociolinguistic functions in a variety of languages, conveying pragmatic information and to construct individual personae (Callier, 2013; Laver, 1968; Pillot-Loiseau et al., 2019; Podesva, 2013; Sicoli, 2010; Yuasa, 2010, among others). Socially-constrained variation in creaky voice use has been extensively studied in English varieties (see Dallaston & Docherty, 2020; Davidson, 2021, for reviews), with increasing work on other languages published only recently (e.g., Burin, 2022; Duarte-Borquez et al., 2024; Johnson & Babel, 2023; Sebregts et al. 2023; Uusitalo et al., 2024). Previous acoustic studies of various English varieties—including both seminal and contemporary work on creaky voice—have shown that *men are creakier than women* across a variety of speech contexts ranging from spontaneous speech in naturalistic settings to read wordlist recordings in a laboratory (e.g., Brown & Sonderegger, 2025; Gittelsohn et al., 2021; Hanson & Chuang, 1999; Irons & Alexander, 2016; Iseli et al., 2007; Klatt & Klatt, 1990; Loakes & Gregory, 2022; Syrdal 1996). Older impressionistic analyses, i.e., those relying on audio-visual coding of creaky voice, have reached similar conclusions with respect to gender (again in varied speech contexts), suggesting that creak can be interpreted as a sign of masculinity, authority, and socio-economic status in British varieties especially (Abercrombie, 1967; Esling, 1978; Henton & Bladon, 1988; Laver, 1968; Stuart-Smith, 1999).

There is current impressionistic work—which still relies on audio-visual creaky voice identification methods—also sourced from various speech contexts, but examines American

English varieties almost exclusively. Contrary to the aforementioned acoustic and older impressionistic work on creaky voice, these studies show that *women are creakier than men* (e.g., Abdelli-Beruh et al., 2014; Melvin & Clopper, 2015; Podesva, 2013; Wolk et al., 2012; Yuasa, 2010), coinciding with the emergence of that same claim in mainstream American media in the late 2000s (Fessenden, 2011; Grim, 2015; Jaslow, 2011; Steinmetz, 2011; Quenqua, 2012, for example). Creaky voice is often referred to as vocal fry in popular discourse and media, and a sudden surge in use of this term around 2010 is observable in Google Ngrams (Google Books, 2025). Yuasa (2010) suggests that the increased use of creaky voice by young, upwardly-mobile American women is triggering a shift in popular perception towards a more educated, urban-oriented, nonaggressive and informal interpretation. Others explicitly asked listeners to provide qualitative judgements of modal and creaky voices, then compared listener ratings across voice qualities. Most studies exposed listeners to the same speakers producing both modal and creaky voice (in lab speech in Anderson et al., 2014 and Lee, 2016; in spontaneous speech from online sources in Stewart et al., 2024), while others only compared modal speakers and creaky speakers (recorded in-lab in Gallena & Pinto, 2021). Conversely, Ligon et al. (2019) did not make use of any audio stimuli, opting to train their listeners to identify and differentiate various voice qualities (modal and creaky voice included) and then ask them to associate the voice qualities with affective/emotive traits. Contrasting with Yuasa's (2010) interpretation of creaky voice use, these speaker perception studies reach the consensus that creaky voice leads to overt negative sentiments: Women exhibiting creaky voice were often perceived as less competent, attractive and hireable (note that Lee, 2016, finds comparable negative judgement of both men's and women's creaky voice). Social perception work on creak has placed less focus on attitudes in other English varieties; the studies that do exist find similar but less consistent negative perceptions (in New Zealand English, Calhoun & White, 2025 and Pittam, 1987; mixed findings in Irish English, Gobl & Ní Chasaide, 2003; in British English, Liu & Xu, 2011; and in Canadian English, Goodine & Johns, 2014). When exposed to two young women's voices with creak and two others without creak, Canadian English listeners judged women exhibiting creaky voice as more ditzy and lazier, as well as less assertive, responsible and hardworking than modally-voiced women (Goodine & Johns, 2014).

These methodological differences highlighted by the studies described above appear to impact conclusions drawn from investigations of creaky voice across gender: Acoustic analyses (almost exclusively) find more creak for men, and impressionistic analyses describe more creak for men prior to the 2000s, but more creak for women thereafter. While there is a remarkable lack of empirical evidence indicating a change in voice quality over time (see Brown & Sonderegger, 2025, for a review), there is compelling evidence for change in the social view of creaky voice. The production-perception mismatch displayed in creaky voice research demonstrates that perception does not always reflect acoustics or articulation. While acoustic studies rely on

objective measures, the overarching goal of this work is to model the *perceptual* instantiation of creaky voice. Creaky voice as a perceptual phenomenon has attracted both public and academic interest in recent years, and speakers and listeners show the ability to reliably attend to it as well as develop strong intuitions about it. Acoustic correlates alone therefore function as indirect measures of creaky voice.

Some work has made efforts to bridge the gap between articulation, acoustics, and perception, notably Kreiman et al.'s (2014) psychoacoustic model of voice (and Kreiman et al., 2021, thereafter) which identifies a set of articulatorily-grounded acoustic measures that all contribute individually to the perception of voice quality. While this model may provide a quantitative measure of voice quality that more accurately reflects its perception compared to other acoustic studies of voice, the discord between older impressionistic studies of creaky voice and more contemporary ones remains unresolved. If a change in voice quality production (articulation and acoustics) has not occurred over time, then there is no motivation for any changes to the psychoacoustic model of voice over time. The psychoacoustic model of voice in its current state does not allow for the integration of social factors in voice perception, failing to account for any variation or change (at the individual or group level), and is not designed to assess *creaky* voice perception specifically. One plausible explanation for the divergence between creaky voice production and varying perceptions relies on the assumption that listener expectations can heavily influence perception. Accordingly, Section 2.3 discusses the well-attested effect of perceived speaker characteristics on speech perception, and Section 2.4 explores how perceptual judgements of voice can be shaped by speaker pitch and gender.

2.3 Sociolinguistic perception: Speech perception as a function of speaker perception

Sociolinguistic perception is a vast topic that integrates perspectives from diverse fields of research, including social psychology, cognitive science, sociolinguistics and phonetics. It is now well-known that variation in the speech signal can affect the social evaluations of the speaker (as exemplified in the previous section with respect to creaky voice usage). The *matched-guise* technique was first introduced by Wallace Lambert and his colleagues, highly influential in that it operationalized a method to systematically analyze listeners' social evaluations in relation to linguistic features of speech. In Lambert et al.'s 1960 study, the same speaker's voice was recorded in French and in English, effectively producing two different guises which were presented to listeners without revealing that they originated from the same person. Listeners were then tasked with assigning subjective ratings of 14 character traits to the guises (Lambert et al., 1960). Since then, the paradigm has been applied to various linguistic features and personal characteristics, including, but not limited to, gender, race/ethnicity, region of origin, sexual orientation (see D'Onofrio, 2016; Drager 2010; Foulkes & Hay, 2015, for overviews).

This is generally thought to occur because listeners hold stereotypes towards social groups and attribute those stereotypes to individuals whom they perceive as belonging to that group on the basis of linguistic features alone (Johnson, 2000; Lippi-Green, 1997). The vast and varied body of work collectively shows the wide-reaching effect that linguistic variation has on social evaluations of speaker attributes.

Furthermore, the relationship between phonetic and social information in speech perception is bidirectional. Social information about the speaker, whether actual, expected, or perceived, can bias speech perception (distinctly from social perception). As a variation of the matched-guise design and following long-standing work arguing for the perceptual integration of visual and auditory information (e.g., Campanella & Belin, 2007; McGurk & MacDonald, 1976), Strand and Johnson (1996) spearheaded a method to examine the effect of perceived speaker attributes (social perception of the speaker) on speech perception/processing. In search of evidence for a visually-driven (or socially-driven) speaker normalization effect on sibilant perception, they paired ambiguously-gendered voices with prototypical male and female faces to isolate a *face gender effect*. They found that when Central Ohio English listeners (15 women and 9 men) were tasked with identifying whether they heard /s/ or /ʃ/ (2AFC) from a continuum of synthesized sibilants, they identified more /s/ when audio was presented with the male face than the female face, indicating a lower frequency threshold for /s/ perception (following general gender patterns of sibilant production). These results provide evidence for integration of the visual perception of gender with the acoustics in the speech signal in gradient sibilant categorization. Thus, priming listeners with social information about the speaker, encouraging a specific speaker perception and creating a certain expectation for that speaker's voice and language use, can affect their perception of various linguistic variables (generally believed to be linked to those relevant social groupings) in otherwise acoustically identical speech. In addition to perceived speaker gender (Alderton, 2020; Bouavichith et al. 2019; Jessee & Calder, 2025; Johnson et al., 1999; Lindvall-Östling et al. 2020; Strand & Johnson, 1996; Strand 1999; Yu, 2022), perceived speaker race/ethnicity (Babel & Russell, 2015; Kutlu et al., 2022; Staum Casasanto, 2010), age (Drager, 2011; Hay et al., 2006b, Koops et al., 2008; Walker & Hay, 2011), dialect (Hay & Drager, 2010; Hay et al., 2006a; Niedzielski, 1999), socio-economic class (Hay et al., 2006b), sexual orientation (Mack & Munson, 2012), as well as micro-sociological categories like personae (D'Onofrio, 2015; see D'Onofrio, 2020, for a review), have been shown to influence speech perception in diverse ways (see D'Onofrio, 2016; Drager 2010; Foulkes & Hay, 2015; Weatherholtz & Jaeger, 2016, for overviews).

Expectations and stereotypes about speakers have even been shown to trigger shifts in production in some cases, such as *expectation-driven convergence* (e.g., Vaughn & Kendall, 2019; Wade et al., 2023; and see Auer & Hinskens, 2005, for a review). In Wade et al. (2023),

non-Southern out-group speakers converged to more monophthongal /aɪ/ (a stereotypically Southern American English pronunciation) when listening to speech from a *labeled* Southern American English speaker, despite the speaker acoustically producing a Midland American English accent. On the other hand, some work does not yield clear social priming effects in the expected directions (e.g., Squires, 2013; Lawrence 2015, Juskan, 2016; Walker et al., 2019). Juskan (2016) situates their inconsistent priming effects within the psychological literature and suggests a number of conditions that need to be met in order for priming effects to emerge as expected: 1) The stimuli must be based on a highly salient linguistic variable/feature that listeners are explicitly aware of (i.e., a stereotype); 2) the linguistic variable must vary on a continuum rather than discretely; and 3) the prime and the acoustic stimuli must not be so mismatched that listeners will not accept that the prime and stimuli are combined. While these conditions may explain some of these failures to replicate priming effects on perception, the expectation that priming effects should be invariable or even similar across studies is controversial (see Juskan, 2016, for a discussion).

Given that this paper's main topic, the production-perception mismatch in creaky voice, relates directly to varying gender expectations, only studies examining perceived gender effects on speech perception will be discussed further. Strand and Johnson's (1996) study has inspired other (quasi-)replications in recent years (e.g., Bouavichith et al. 2019; Jessee & Calder, 2025; Munson et al., 2017), which attests to the robustness of the face gender effect in sibilant perception. Munson et al. (2017) used both explicit and implicit priming, resulting again in lower frequency /s/ perception for male faces and male-suggestive cues. Likewise, Jessee and Calder (2025) show that when listeners were told that speakers were transgender, they perceived a higher frequency /s/ for the feminine voice and a lower frequency /s/ for the masculine voice than listeners who were not given that gender information. The face gender effect is also not limited to sibilant perception. Johnson et al. (1999) apply their 1996 paradigm to the perception of the /ʊ/-/ʌ/ contrast and find that the expectation for men's formants to be lower frequency than women's formants is substantiated by a face gender effect: Participants tend to perceive the vowel category boundary at lower F1 frequencies when stimuli is paired with a male face and at higher F1 frequencies when paired with a female face. Alderton (2020) examines the perception of /u/-fronting (a.k.a., GOOSE-fronting) in Standard Southern British English, a sound change led by women but not yet carrying social salience. They find a significant interaction between listener gender and face gender, men identifying fronter /u/ vowels when primed with a woman's face, despite failing to find a significant effect of face gender alone (Alderton, 2020). In Yu (2022), listeners were presented with a gendered face and (ambiguously gendered) audio stimuli along a voicing continuum (crossing VOT and f0) and instructed to identify whether they heard a /b/ or /p/ (2AFC). Listeners who were exposed to a male face showed less reliance on

VOT compared to those exposed to a female face or those who were given no visual information, consistent with typically-male acoustic behavior (VOT differences are less distinct for men than women) (Yu, 2022). In summary, these studies illustrate the influence of socio-indexical and paralinguistic information on speech perception in that expectations about a speaker's gender or other personality traits can prime listeners to interpret the speech signal in ways that conform to the social expectation.

2.4 Systematic analyses of creaky voice perception

The apparent mismatch between acoustic and impressionistic methods of identifying creaky voice has been noted in previous perceptual work (namely, Davidson, 2019a and White et al., 2024). The tendency to impressionistically identify more creakiness in women's speech was first formally hypothesized by Davidson (2019a) to be due to two possible factors: acoustic pitch of the speech and perceived speaker gender. The *Acoustic hypothesis* (referred to as the pitch contrast scenario in White et al., 2024) stipulates that creaky voice is more perceptible in higher pitched modal voices because of a larger pitch differential between modal and creaky voice, compared to lower pitched modal voices which have a smaller pitch differential (Davidson, 2019a). There is uncontroversial evidence for an upper bound of approximately 80 Hz on the f_0 range of creaky voice (e.g., Blomgren et al., 1998; Hollien & Michel, 1968; Leung et al., 2022). Some studies report similar f_0 ranges of creaky voice for both men and women (respectively, 49 and 48 Hz on average in Blomgren et al., 1998; respectively reaching 46 Hz and 59 Hz minimums in Keating & Kuo, 2012) while others report somewhat lower f_0 ranges for men's creak than for women's creak (respectively, 55 Hz and 74 Hz on average in Brubaker et al., 2016; and 70 Hz and 88 Hz on average in Davidson, 2019a). Despite these differences, the hypothesis generally holds in assuming that creaky voice is always produced at a low f_0 , and therefore when habitually high-pitched voices (generally women's voices) lower dramatically to reach the f_0 threshold for creakiness, it is perceptually more salient than when habitually low-pitched voices (generally men's voices) lower moderately to a similar threshold. The *Bias hypothesis* (referred to as the gender bias scenario in White et al., 2024) presents the alternative that listeners are simply biased to assume that creak is more common in women's speech compared to men's, given social stereotypes (Davidson, 2019a). In reality, these two hypotheses are difficult to tease apart as they often co-occur: Men typically have lower modal pitch whereas women have higher modal pitch. White et al. (2024) describe a third hypothesis or scenario that combines the former two, suggesting that the speaker habitual pitch will have the strongest impact on the perception of creaky voice when voices have distinct pitches, and only when pitch is no longer informative will speaker gender affect perception. The predictions of these hypotheses are shown in **Table 1**.

	Creaky voice condition	Modal voice condition
Acoustic hypothesis (pitch contrast scenario)	hi-f > mid-f = mid-m > lo-m	hi-f > mid-f = mid-m > lo-m
Bias hypothesis (gender bias scenario)	hi-f = mid-f > mid-m = lo-m	lo-m = mid-m > mid-f = hi-f
Acoustic + Bias hypothesis (pitch contrast + gender bias scenario)	hi-f > mid-f > mid-m > lo-m	hi-f > mid-m > mid-f > lo-m

Table 1: Predictions of each hypothesis for listener accuracy in the identification of creaky voice (adapted from Davidson, 2019a, and White et al., 2024).

Assuming such an acoustic bias, the most relevant cue to the identification of creaky voice is a larger pitch differential between modal and creaky voice. That is, voices with a larger creaky-to-modal voice pitch differential should facilitate both the identification when creak is present and the non-identification when it is absent. Therefore, in both the creaky voice condition (i.e., identifying creak when it is present) and the modal voice condition (i.e., not identifying creak when it is absent), this hypothesis predicts more accurate creak identification in the highest-pitched voice (hi-f), comparably lower accuracy in the mid-pitched voices (mid-f and mid-m), and then lowest accuracy in the lowest-pitched voice (lo-m). On the other hand, if we assume a gender bias, then the most relevant cue to creaky voice identification is gender, specifically women’s voices eliciting more creaky voice responses in all contexts. Following this idea, creaky voice is predicted to be more accurately detected in women’s voices (hi-f and mid-f, regardless of pitch) when it is present, and also more inaccurately detected (over-detected) in women’s voices when it is absent. In the case of the combined acoustic and gender bias scenario, pressures from the creaky-to-modal voice pitch differential and gender interact: A larger pitch differential and women’s voices are predicted to lead to more creaky voice identification. White et al. (2024) posit that pitch will play a larger role than gender in the extreme pitch conditions, with a high-pitched voice (hi-f) facilitating accurate creak identification in the creaky condition and non-identification in the modal condition, whereas responses to a low-pitched voice (lo-m) will exhibit less accuracy in both the identification and non-identification of creaky voice. They predict that gender will only play a role in creak decisions in conditions where the voices are similar in pitch. In the creaky condition, they expect the women’s voice (mid-f) to elicit more accurate identification of creak than in the men’s voice (mid-m), while in the modal condition, they expect the women’s voice to elicit more inaccurate identification of creak than in the men’s.

To test these hypotheses, both Davidson (2019a) and White et al. (2024) conducted experiments in an attempt to disentangle the effects of pitch and gender on creaky voice perception. Crucially, stimuli for these experiments fit four speaker profiles: a high-pitched female, a mid-pitched female, a (quasi-)equally mid-pitched male, and a low-pitched male. Davidson (2019a) made

use of natural speech stimuli from high quality podcast recordings, extracting modal and creaky speech stimuli from eight voices (with average f_0 s of 226 Hz and 194 Hz for the hi-f voice, 164 Hz and 152 Hz for the mid-f voice, 132 Hz and 145 Hz for the mid-m voice, and 111 Hz and 90 Hz for the lo-m voice). Davidson (2019a) also included other conditions such as location/extent of creak (none, partial, or whole) and type of utterance (fragment or full sentence), but these will not be discussed extensively here. Conversely, White et al. synthesized their stimuli from male (28 y.o.) and female (22 y.o.) modal voices recorded in a lab, manipulating mean f_0 (190 Hz for the hi-f voice, 135 Hz for the mid-f and mid-m voices, and 97 Hz for the lo-m voice) and voice quality (inserting cycle-to-cycle f_0 irregularity) while leaving gendered formant ratios untouched. White et al. limited their stimuli to bigrams (adjective noun pairs), synthesizing creak into the rhyme of the second word in the creaky voice condition. Both studies ran a creak identification task, requiring binary decisions from participants. Despite recruiting speakers of different varieties of English, 54 Americans (for each of two versions of the experiment) in Davidson (2019a) and 258 Australians in White et al., the results of both studies shared some similarities.

In Davidson's (2019a) first version of the experiment, a weak tendency to accurately identify more creak in female voices (both hi-f and mid-f) than in males' was found. However, in the second version of the experiment (when the mid-pitched male and female voices were more closely matched), the gender difference in both creaky conditions disappears. In the modal condition, participants were more accurate at detecting the absence of creak in the hi-f voice, and less accurate for the lo-m voice, falsely detecting creak when there was none. Due to inconsistencies in the results of both versions of the experiment, Davidson (2019a) concludes that there is no robust evidence for either an acoustic bias or a gender bias. Aside from gender and pitch, across both of Davidson's (2019a) experiments, participants were worse at identifying creak in the partially creaky condition than the wholly creaky condition, suggesting that creak is less salient in a prosodic position where it is expected (utterance-finally).

White et al. (2024) find similarly mixed evidence for a gender or acoustic bias. In the creaky voice condition, creak identification is slightly less accurate for the lo-m voice (but the mid-m voice patterns with the female voices), either partially corroborating Davidson's female voice skewed gender bias effect in her first study (2019a) or supporting the acoustic bias towards more creak identification for women's higher pitched voices. Moreover, the strong false-alarms of creak in the lo-m voice in the modal condition from Davidson's second study (2019b) are also replicated, indicating an acoustic bias in the opposite direction as predicted (i.e., men's voices inducing more creaky voice percepts). These two findings show that for men's low-pitched voices, listeners are less likely to identify creak when it is present (creaky voice condition), while also being more likely to identify creak when it is not present (modal voice condition). Lower overall creaky voice identification accuracy for the lo-m voice suggests that listeners struggle

more to distinguish creaky voice from low modal pitch in men's voices compared to other voices (hi-f, mid-f, and mid-m). Interestingly, White et al. find additional evidence for the predicted gender effect in the modal voice condition, more creak inaccurately identified (false-alarmed creak) in the mid-f voice than the mid-m voice in this case. In view of the two false-alarmed creak findings, White et al. conclude that the combined pitch contrast and gender bias scenarios have explanatory power, but suggest an alternative underlying mechanism: a *pitch given gender bias*. When listeners hear modal voices that are low given the listeners' expectations for gender (i.e., lo-m and mid-f), they are more likely to identify creak when there is none.

Li et al. (2023) conduct a comparable study in Mandarin, a language which, importantly, has not been reported to carry any social associations between gender and creaky voice. Stimuli originated from declarative sentences produced with modal and creaky voice by a high-pitched female speaker recorded in a lab. F0 and formant manipulations were performed to create the low-pitched male stimuli. Conditions included pitch range/gender (hi-f vs. lo-m), creak extent (mono-syllabic creak vs. multi-syllabic creak), and prosodic position (final vs. non-final). Forty native Mandarin listeners were tasked to identify characters pronounced with creaky voice. Li et al. find that Mandarin listeners consistently identify more creak in the low-pitched male voice: low pitch facilitating creak identification but also increasing false-alarmed creak identification in modal speech as in Davidson (2019a) and White et al. (2024). Furthermore, they find that sentence-final position inhibited the identification of creak, confirming the same effect of prosodic position in Davidson (2019a).

Altogether, these results do lend some support to the combined pitch contrast and gender bias scenario (summarized in **Table 2**). Confirming at least a small effect of the predicted gender bias, Davidson (2019a) finds slightly more creak identified in women's voices than in men's voices, though only in one of two versions of her experiment, and White et al. (2024) find slightly more creak identified in women's modal voices compared to men's, when pitch is matched and comparing women's creaky voices (high and mid-pitched) to low-pitched men's creaky voices. Confirming a stronger acoustic effect related to pitch but in the unexpected direction, Davidson (2019a), White et al. and Li et al.'s (2023) studies all find increased creak identified in low-pitched men's (modal) voices. Considering the broader motivations for these perception studies, these results alone do not fully explain the overwhelming tendency to identify more creak in women's voices in impressionistic studies across the sociolinguistic body of literature. If there is a dominant acoustically-grounded inclination to find more creak in low-pitched men's voices, evidenced to largely eclipse effects of social stereotypes which lead to women's voices being perceived as creakier, then how do so many listeners (trained phoneticians and speech-language pathologists alike) consistently continue to identify more creak for women?

	Creaky voice condition	Bias	Modal voice condition	Bias
Davidson (2019a): exp. ver. 1	hi-f = mid-f > mid-m = lo-m	Gender	hi-f = mid-f = mid-m = lo-m	None
Davidson (2019a): exp. ver. 2	hi-f = mid-f = mid-m = lo-m	None	hi-f > mid-f = mid-m > lo-m	Acoustic
White et al. (2024)	hi-f = mid-f = mid-m > lo-m	None	hi-f > mid-m > mid-f > lo-m	Acoustic + Gender
Li et al. (2023)	lo-m > hi-f	None	hi-f > lo-m	Acoustic or Acoustic + Gender

Table 2: Actual results on listener accuracy in the identification of creaky voice from Davidson (2019a), White et al. (2024), and Li et al. (2023) alongside the hypotheses (see Table 1) supported.

2.5 Research questions

Public discourse from roughly the last decade often reports on extreme creaky voice usage by women, a pattern attested in recent sociolinguistic and sociophonetic research (e.g., Podesva, 2013; Yuasa, 2010), despite acoustic evidence consistently showing that men tend to produce more creaky voice (Gittelsohn et al., 2021; Klatt & Klatt, 1990; among many others). The goal of this project is to reconcile differences between the production and perception of creaky voice by uncovering the perceptual pathway by which this discrepancy arises. We test two hypotheses (originally from Davidson, 2019a): an acoustic pitch contrast bias in which creak is more perceptible in higher-pitched voices due to greater contrast with modal pitch, and a gender bias in which listeners expect women to be creakier. The primary research question addressed here investigates whether voice f_0 and visual face gender both independently show quantifiable effects on creaky voice perception in Canadian English. Using a matched-guise paradigm, we isolate the social effect from the acoustic effect to determine whether either or both contribute to a perceptual asymmetry.

The precise experimental design implemented to test these questions draws from the methods from Davidson (2019a), White et al. (2024), and Strand & Johnson (1996). To disambiguate effects of an acoustic bias from a gender bias on the perception of creaky voice, f_0 and gender need to be treated independently. F_0 values are therefore restricted to a f_0 range (and formant structure) ambiguous for gender, and clearly gendered faces are then paired with the ambiguously-gendered voices in a matched-guise paradigm. If a gender bias exists in creaky voice perception, then a priming effect of face gender should be observable even if voice quality

and pitch are held constant. Specifically, if the listeners' gender bias leads them to expect more creak in women's speech, then a woman's face should prime increased creakiness percepts, but if their gender bias presumes more creak in men's speech, then a man's face should prime increased creakiness percepts. Alternatively, if an acoustic bias exists, then f_0 values, independent of gender perception, should influence creaky voice perception. Predictions of an acoustic bias can also differ depending on the direction of the f_0 effect: If a larger pitch differential between modal and creaky voice is crucial in the perception of creak, then a higher f_0 value should result in more creaky percepts; otherwise, if low pitch is the most relevant cue to creaky voice, then a lower f_0 value should induce more creaky percepts. A combined acoustic and gender bias (like the pitch given gender bias in White et al., 2024) is also possible and would be substantiated by a face gender effect that differs as a function of f_0 .

This line of inquiry opens up new avenues for understanding of how voice perception is not merely a function of physical (acoustic) input, but also of socially structured expectation. In the broader landscape of sociophonetic research, this study also contributes to filling gaps in the existing literature. While there has been increasing interest in creaky voice, few studies have directly tested how social perception shapes creaky voice perception, especially in non-American varieties of English. Although prior work has shown that speech perception is influenced by social expectations, especially around gender, this research has largely focused on segmental features like sibilants and vowels. Thus far, there has been comparably less attention given to more complex and multi-dimensional features of speech, notably non-modal voice qualities like creak.

3. Method

First, a norming study (Section 3.1) was conducted to determine what voice settings would be appropriate for our ambiguously gendered stimuli. Clearly gendered faces were then paired randomly with the gender-ambiguous audio stimuli in a matched-guise paradigm (Section 3.2) to assess the effect of face gender priming on creak perception. Listener ratings of creaky and modal voices were elicited along a continuous scale from not creaky at all to extremely creaky. All stimuli from the norming study and creaky voice perception experiment (training, practice, and trials), original audio recordings, scripts, datasets, and saved models are available on the paper's OSF page (<https://osf.io/f45yh/>).

3.1 Norming perceived gender

To create a gender ambiguous voice, the Praat (Boersma & Weenink, 2025) "Change gender" function was used, which varies both formant ratio and f_0 . A norming study was conducted to choose stimuli that were rated as ambiguously gendered.

3.1.1 Participants

Eighteen English-dominant speakers, born and raised in (and currently living in) Canada, without any hearing difficulties nor cochlear implants (9 women, 9 men, 18 to 71 years of age, median age: 38) were recruited online through Prolific (Palan & Schitter, 2018).

3.1.2 Stimuli

A 27-year-old female native Canadian English speaker's voice was recorded using a MixPre-3 audio recorder and a Shure SM10A headset microphone. Sampling rate was 44.1 kHz and intensity was set to 70 dB for all recordings. Ten phonetically balanced utterances, Harvard Sentences (IEEE, 1969; list 1), were pronounced in a neutral modal voice and intonation was kept consistent across sentences, with a falling contour. The f_0 was manipulated to create three new median f_0 values: 115 Hz, 135 Hz, and 155 Hz, roughly matching the ambiguous gendered pitch values used in Davidson (2019a) and White et al. (2024). The formant shift ratio was manipulated to create five values ranging from most prototypically male-like (longer vocal tract) formants to most prototypically female-like (shorter vocal tract) formants: 0.8, 0.85, 0.9, 0.95, 1. Both variables were fully crossed for all 10 modal utterances, creating 150 trial stimuli. All experimental trials were randomized, and each participant was only shown 75 of the 150 total stimuli in order to keep the task short. Creaky voice utterances were not used in the norming study because the goal was to elicit gender judgements based solely on f_0 and formant values, and modal voice is best for maintaining a clear f_0 throughout the utterances.

3.1.3 Experimental design

The experiment was conducted online using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). The task was estimated to take roughly 15 minutes and participants were paid at a rate of £9.00/hour, amounting to £2.25 per participant. Participants began a headphone screener (Woods et al., 2017) and were given a maximum of two attempts to pass it.

Instructions for the gender rating task were presented on the first screen, followed by three practice trials. The practice trials included audio of one utterance at the mid-point and both ends of the manipulated continua (lowest f_0 and most male-like formants, highest f_0 and most female-like formants, and mid f_0 and ambiguous formants). In both the practice and experimental trials, participants were presented with a fixation cross alongside audio from the continua of manipulated utterances. As soon as the audio finished playing, the next screen prompted participants to rate the voice along a 5-point scale for gender prototypicality, with 5 indicating a very feminine voice and 1 indicating a very masculine voice. Participants then clicked to continue to the next trial.

3.1.4 Results

The results in **Figure 1** (right) show a clear trend towards more masculine-sounding ratings as median f_0 and formant shift ratio decreases, and more feminine-sounding ratings as median f_0

and formant shift ratio increases. It was determined (through visual inspection of **Figure 1** left) that a formant shift ratio of 0.9 provided the most ambiguously gendered responses, and all three f0 values were retained to examine the effect of pitch (within a generally ambiguous range) on creakiness ratings.

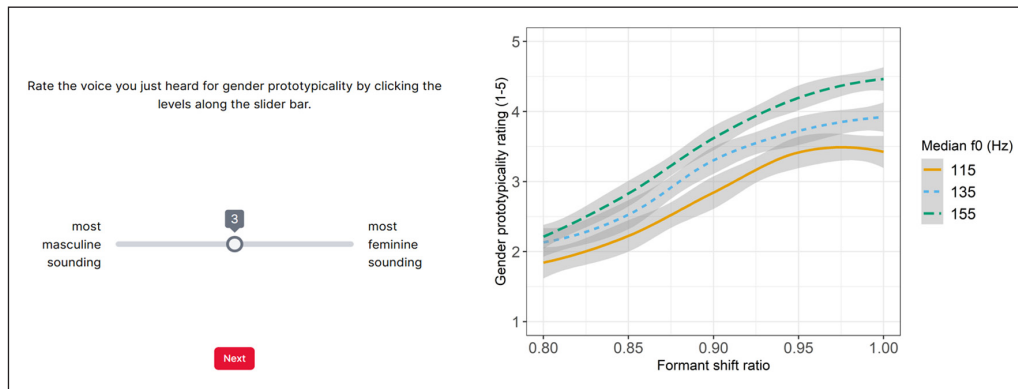


Figure 1: Gender prototypicality rating screen that was presented to participants (left) and aggregated gender prototypicality ratings (1 = most male-sounding, 5 = most female-sounding) plotted by formant shift ratio and median f0 (right).

While there was minor variation by utterance, no particular utterances elicited qualitatively different responses than the others, observable in Figure A1 (Appendix A) and from a Bayesian ordinal cumulative regression model (utterance: $\hat{\beta} = -0.36$, $\hat{\sigma} = 0.21$, CI = $[-0.79, 0.07]$; Table A2 in Appendix A). Individual variation in gender prototypicality ratings can be observed in Figure A3 (of Appendix A). While some participants skewed towards more feminine-sounding ratings across most of the manipulated continuum (see participant 44382292 and 44382167), no participants skewed towards more masculine-sounding ratings. This can be explained by the source of the original recordings, which was a female speaker, causing certain listeners to perceive a more feminine voice regardless of the manipulated f0 and formants.

3.2 Creaky voice perception experiment

The main perception experiment addressed the study's primary research question targeting how creaky voice perception might be influenced by an acoustic bias or a gender bias, i.e., whether manipulated f0 and presented face gender independently affect listeners' creakiness ratings.

3.2.1 Participants

Recruitment for the creaky voice perception experiment followed the same procedure and recruitment criteria as in the norming study. We had a target sample size of 40 and we excluded and replaced any participants who failed to complete the study ($n = 6$). We had planned to also exclude participants with evidence of poor understanding or attention to the task as evidenced by a pattern of random responses, however no participants met this exclusion criteria. Forty new

English dominant speakers, (14 women and 26 men, aged 18 to 71 years old, median age: 32) were included. None were involved in the norming study.

3.2.2 Stimuli

The same 27-year-old female native Canadian English speaker's voice was recorded using a MixPre3 audio recorder and a Shure SM10A headset microphone. Sixty new phonetically balanced sentences (i.e., not those used in the norming study), lists 3–8 of the Harvard Sentences (IEEE, 1969), were pronounced by the speaker. We chose to elicit somewhat natural productions of modal and creaky voice (like Davidson, 2019a) to avoid complications in simulating creaky voice acoustically. Thirty sentences were produced in a neutral modal voice, and 30 sentences were produced with roughly the first half of the sentence in modal voice and the second half in creaky voice. F0 tracks are often unreliable in creaky voice production due to their characteristic f0 irregularity, therefore f0 manipulations are often ineffective on creaky voices. As such, partially creaky sentences were preferred over fully creaky sentences so that the f0 manipulations could be perceived across all stimuli, obvious in the modal voice portions. Because the creaky utterances were produced naturally, the exact timing of creaky voice was variable. To provide clearer characterization of the original creaky utterances, we examined the range of variation in the proportion of creaky voice. Durations of creak were estimated using an audio-visual coding method to identify the onset of creak within the original creaky utterances. Approximate proportions of creak were then calculated (by dividing the duration of creak by the entire utterance duration) and are plotted in Figure A4 (Appendix A). The range of proportions of creak was 0.40 to 0.77 with a mean of 0.58 for all 30 creaky utterances. Original modal utterances did not contain any creak. Speech rate varied to some extent, ranging from 2.16 to 3.94 vowels per second with a mean of 3.15 vowels/second (see Figure A5 in Appendix A for speech rate plots by voice quality and median f0). Intonation was again kept consistent across sentences by the speaker, with a falling contour. The mean f0 across vowels within all the utterances in the original recordings was 187 Hz. Instead of using multiple speakers with varying mean modal pitches (as did Davidson, 2019a), we followed White et al. (2024) in manipulating f0 values to achieve more control. As described above, we used three levels of f0: 115 Hz, 135 Hz, and 155 Hz. The 30 unique modal and 30 unique creaky utterances were each split into three groups of 10, and each group was f0-shifted to one of the three values, creating 60 audio stimuli. Following results of the norming study, the formant shift ratio was set to 0.9 for all stimuli.

Impressionistically, the “change gender” function introduced some distortions that may lead to a percept of creakiness in some of the modal stimuli. To check whether the creaky stimuli were still objectively creakier and, importantly, that the different levels of f0 were equally creaky, we implemented an acoustic analysis on all formant and f0-shifted stimuli. The acoustic analyses of the 60 audio stimuli were conducted using PraatSauce (Kirby, 2018), a Praat script for spectral measures, and another Praat script for f0-related measures (Brown & Sonderegger, 2025). A

total of 533 vowels were analyzed, measured at 3 equidistant points, at 25, 50 and 75% of the vowel duration, from which vowel means were calculated. The acoustic correlates of creak examined in this analysis include the proportion of unreliable f_0 tracks (i.e., the proportion of vowels for which Praat could not track the f_0 consistently), in addition to more common measures within the creaky voice literature, specifically $H1^* - H2^*$, CPP, and $HNR < 500$ Hz (see method in Brown & Sonderegger, 2025, for a more in-depth discussion of these measures and how they relate to creaky voice). $H1^* - H2^*$ values that depended on unreliable (0 or NA) f_0 values were removed (n points = 495, 30.96% of points over all $H1^* - H2^*$ tracks and n vowels = 128) and extreme $HNR < 500$ Hz values were excluded (n vowels = 14). **Figure 2** illustrates empirical differences between vowels in modal utterances and those in creaky utterances across median f_0 values. Observational trends show higher proportions of unreliable f_0 tracks and lower $H1^* - H2^*$, CPP and $HNR < 500$ Hz values for creaky utterances compared to modal utterances, providing acoustic evidence for increased creakiness in the creaky stimuli. This voice quality difference in acoustic measures is mostly consistent across f_0 values (with the exception of $H1^* - H2^*$), indicating that the f_0 manipulation did not seem to directly affect creakiness in the stimuli.

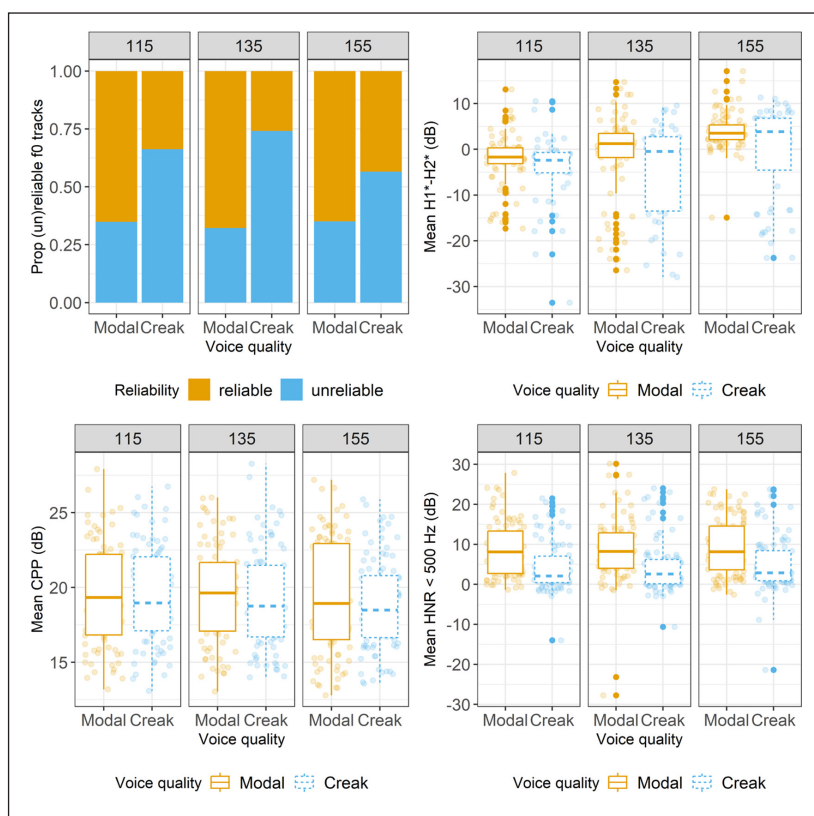


Figure 2: Selected acoustic correlates of creak (mean proportion of reliable/unreliable f_0 tracks, mean $H1^* - H2^*$ in dB, mean CPP in dB, and mean $HNR < 500$ Hz in dB) plotted by voice quality and median f_0 .

The 60 audio stimuli were paired with 60 faces from the London Set of Faces from the Faces Research Lab (DeBruine & Jones, 2017); 30 were presented with a unique male face and 30 with a unique female face. The faces selected from this database came from people under 30 years old, as these faces would match the original speakers' age more closely. The sampled faces also matched the general distribution of ethnicities across the larger database, including 20 white, 4 black, 3 East Asian and 3 West Asian faces (per gender). All participants were exposed to all f_0 levels, but in order to achieve randomization of the unique faces within gender and across utterances, participants branched off into two alternating groups. Odd-numbered participants (in terms of order of recruitment) were shown utterances ending in numbers 1, 2, 3, 4, or 5 with female faces and utterances ending in numbers 6, 7, 8, 9, or 0 with male faces, while even-numbered participants were shown the same utterances with flipped face gender assignments. As such, face gender variation within utterances is between-subjects. **Table 3** shows the distribution of stimuli across conditions. All stimuli (audio utterances paired with randomized faces within gender) were then randomized in the experimental trials.

Median f_0 / FSR	Face gender	Odd participants (n = 20)		Even participants (n = 20)	
		Modal	Creaky	Modal	Creaky
115 Hz / 0.9 FSR	F	Utt. 1–5	Utt. 31–35	Utt. 6–10	Utt. 36–40
	M	Utt. 6–10	Utt. 36–40	Utt. 1–5	Utt. 31–35
135 Hz / 0.9 FSR	F	Utt. 11–15	Utt. 41–45	Utt. 16–20	Utt. 46–50
	M	Utt. 16–20	Utt. 46–50	Utt. 11–15	Utt. 41–45
155 Hz / 0.9 FSR	F	Utt. 21–25	Utt. 51–55	Utt. 26–30	Utt. 56–60
	M	Utt. 26–30	Utt. 56–60	Utt. 21–25	Utt. 51–55

Table 3: Structure of stimuli across all conditions: median f_0 (115 Hz, 135 Hz, 155 Hz) and formant shift ratio (FSR, 0.9), voice quality (modal, creaky), face gender (F, M), and branching of participants. Utterance numbering has been modified from that in the larger list of Harvard Sentences to better convey that utterances are unique (correspondences to the original list and utterance numbering from the Harvard Sentences are provided in Table A6 of Appendix A).

3.2.3 Experimental design

The experiment followed the same protocol as the norming study, including the headphone screener, but instead of the gender prototypicality task, they performed the creakiness rating task. Instructions were presented on the first screen, followed by a training session, four practice trials, and 60 experimental trials.

The training phase aimed to orient participants to the acoustic and social cues to creaky voice, ultimately intending to reduce variability in participants' ratings that might arise from unfamiliarity. The training first described creaky voice to the participants in an accessible and simplified way, noting its rough, croaking/crackly sound to the ear, acknowledging its common name in public discourse, *vocal fry*, and contrasting it with modal voice. We avoided descriptions that made any mention of low pitch so as not to prime participants into associating creaky voice with low pitch a priori. Participants were presented with four examples of creaky voice. Next, participants were shown two faces that matched the broad popular perceptions of speakers associated with creaky voice use: a young female American celebrity (Kim Kardashian) and an old British actor (George Sanders, a.k.a. Shere Khan, from *The Jungle Book*). Text explained that these speakers are known for their very creaky voices. An image of their face, an audio excerpt of a fully creaky utterance, and the corresponding transcription for each was included. On the following screen, participants were exposed to two, more local, speakers (both public figures in Canada but less well-known than the previous American and British celebrities), a young man (Lenni-Kim Lalande) and a young woman (Francesca Farago), roughly the same age as the speaker who recorded the stimuli. On-screen text explained that speakers can manipulate their voice quality and make use of both modal and creaky voice within an utterance. An image of their faces, an audio excerpt of a partially creaky utterance, and its transcription (with the creak in bold font), were presented. Unlike other studies that only expose participants to training stimuli similar to the experimental materials, our training included (American and British) stereotypical examples of creaky voice. This approach provides participants with familiar anchors, allowing them to recognize and contextualize creak as a socially and acoustically meaningful feature.

The practice trials included stimuli from both ends of the manipulated continua (lowest f_0 and ambiguous formants with male face, highest f_0 and ambiguous formants with female face) in modal and creaky voice. In both the practice and experimental trials, participants were presented with the image of a person's face (either male or female) alongside audio from the continua of manipulated utterances (see **Figure 3** left). As soon as the audio finished playing, the next screen prompted participants to rate the voice along a visual analog scale (VAS) for creakiness (see **Figure 3**, right), with 100 indicating an extremely creaky voice and 0 indicating modal voice (no creak at all). We preferred to use a gradient measure of creaky voice perception (motivated in other acoustic perception work, Munson et al., 2010; Urberg-Carlson et al., 2008), requiring listeners to rate the level of creakiness rather than provide binary judgements (as in Davidson, 2019a and White et al., 2024). Participants then clicked to proceed to the next trial. Reaction time was measured from the moment the screen with the creakiness rating scale appeared to the moment the "Next" button was clicked. Participants were allowed to change their selection freely prior to clicking "Next."

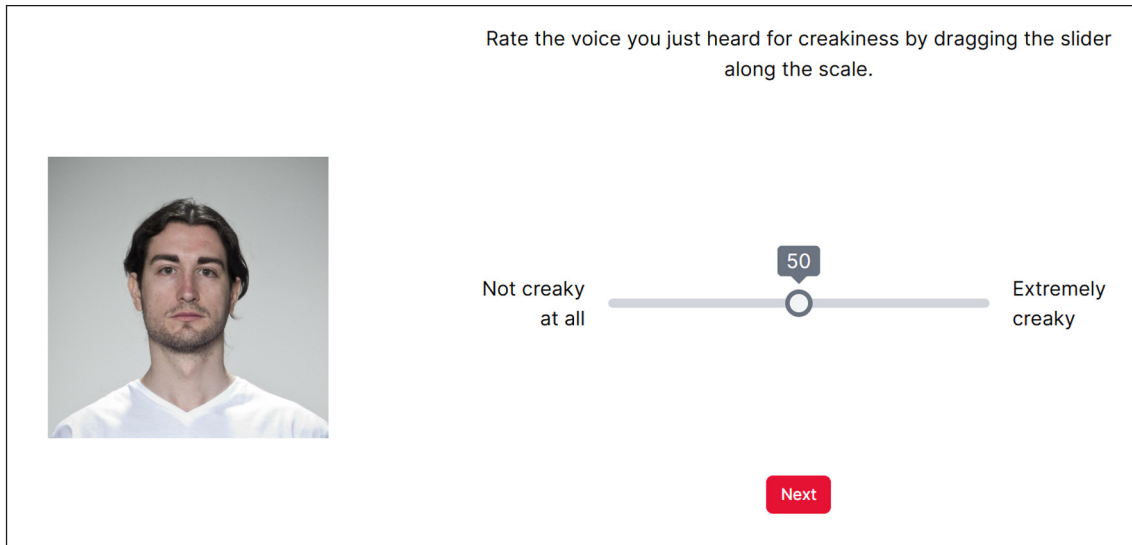


Figure 3: Screen with face and audio stimuli (left) and creakiness rating screen (right) presented to participants.

Following the experimental trials, a short debrief questionnaire was administered. Participants were asked three open-ended questions: whether they found the task difficult, what cues (both auditory and visual) they thought influenced their ratings, and if there was any additional information they would like to share.

3.2.4 Statistical analysis

Using the *brms* package (Bürkner, 2017) in R, a Bayesian zero-one-inflated beta regression model was fitted to the response data (creakiness ratings). A beta regression was chosen because the dependent variable is bounded (by the VAS) (Sonderegger & Sóskuthy, 2025) and the zero-one-inflated variant was implemented because the endpoints (0 and 100 on the original VAS or 0 and 1 in proportions) are included in the possible responses (Heiss, 2021; see also Zellou et al., 2024). The exact model formula can be found in Table A7 of Appendix A. Model structure was chosen based on theoretical importance of the predictors to the research questions. Fixed effects included face gender, voice quality and median f_0 , henceforth f_0 . The effect of face gender on creakiness ratings is of primary interest for the research questions in this paper, crucial to the investigation into how (perceived) gender might affect the perception of creaky voice. The individual effect of voice quality on creakiness ratings is included to confirm that participants are capable of distinguishing between modal and creaky voice. The effect of f_0 is included to test the two hypotheses: i) that a higher modal-to-creaky f_0 difference (higher f_0 values) leads to increased creakiness percepts (acoustic bias); and ii) that low pitch is an important cue to creaky voice, leading to more creakiness ratings for lower f_0 values. All two-way and three-

way interactions between these predictors were also included to assess whether a face gender effect on creakiness ratings differs by voice quality or by f_0 value, but also to confirm that voice quality contrasts (measured perceptually by creakiness ratings) are maintained across different ambiguous pitch ranges. Both two-level predictors, face gender and voice quality, were standardized (centered and divided by 2 standard deviations) and the only multi-level predictor, f_0 , was Helmert contrast coded (centered and orthogonal) so that each contrast corresponded to the difference between that level and the mean of the previous levels (Sonderegger, 2023). Maximal varying-effect¹ structure was implemented into the model. By-participant, by-utterance, and by-face varying intercepts were included, allowing participants, utterances and faces to vary in their baseline creakiness ratings. Varying correlated i) by-participant slopes for face gender, voice quality, f_0 , and their two-way and three-way interactions; ii) by-utterance slopes for face gender; and iii) by-face slopes for voice quality, f_0 , and their interaction, were also included, allowing the effects of the predictors on creakiness ratings to vary in direction and size by participant, utterance, and face. In addition, to account for more participant variability in the use of the 0–100 slider scale, we included by-participant varying intercepts on the ϕ , α and γ parameters. This allowed participants to differ (within the model) in the precision of responses (i.e., variance/dispersion; ϕ) and in endpoint usage (α and γ). The model was fitted with flat priors on all parameters and a weakly informative $LKJ(1.5)$ prior on the correlation terms.

A shifted log-normal regression model was also fitted to the reaction time data, using the same model structure as for the creakiness rating data (see Table A12 in Appendix A). A log-normal regression was chosen because the dependent variable is bounded and can only be positive (Sonderegger & Sóskuthy, 2025), and the shifted variant was chosen because it has been shown to be particularly well-suited to reaction time data, which has a shifted lower bound from 0 to at least 200 ms (Lindeløv, 2019). Because the reaction time data is not directly related to the main research questions, it will only be discussed briefly in the results that follow.

4. Results

4.1 Creakiness ratings

A total of 2,400 responses (40 participants x 60 trials) were collected. The distribution of creakiness ratings is plotted below in **Figure 4**. The distribution appears to be bimodal, suggesting that many tokens are rated as very creaky (50–100), and fewer are rated less creaky (0–50), with a limited number of tokens rated as modal (~0–10). When split by voice quality condition, it is observed that few creaky utterances are rated as somewhat modal, but that comparably more modal utterances are rated as somewhat creaky.

¹ Varying effects are the Bayesian equivalent to *random* effects in frequentist models (McElreath, 2018).

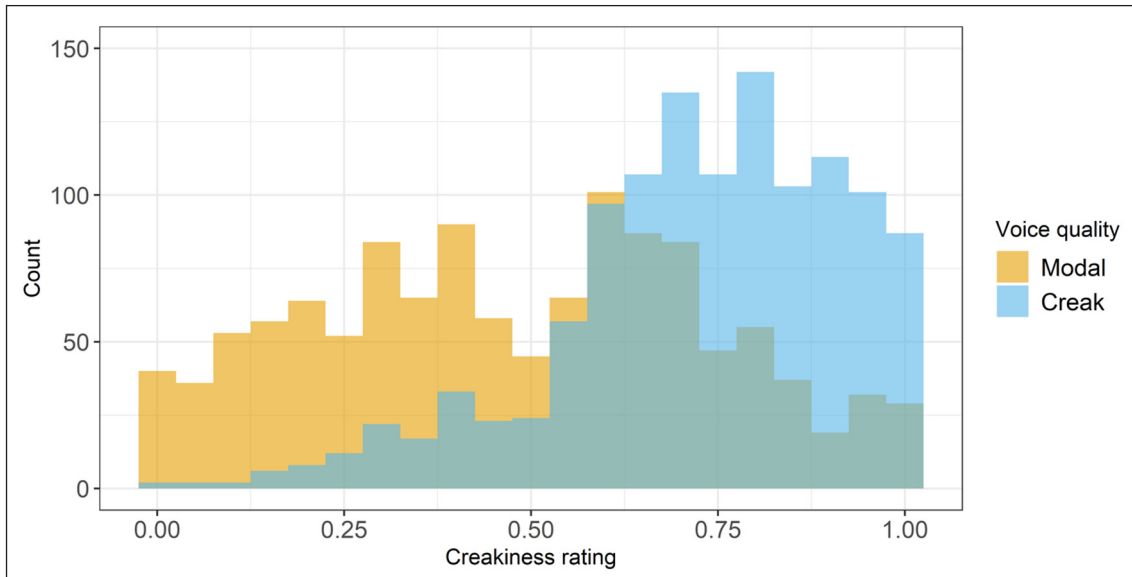


Figure 4: Distribution of creakiness ratings, colored by voice quality condition.

Group-level results are plotted in **Figure 5**, and the zero-and-one-inflated Bayesian regression model table for fixed effects is shown in **Table 4** (the full model summary is available as Table A7 of Appendix A). As expected, the main effect of voice quality on creakiness ratings is significant and clearly observed at all f_0 levels and face gender levels (compare x-axes in **Figure 5**). In the creaky voice condition (when the second half of the utterance contains creak), the voice is rated to be creakier than in the modal voice condition (when the entire utterance is modal) (creak – modal: $\hat{\beta} = 1.04$, $\hat{\sigma} = 0.12$, CI = [0.80, 1.30]; **Table 4**). Corroborating the distribution by voice quality in **Figure 5**, modal stimuli creakiness ratings center around approximately 0.5, meaning that participants judged most modal stimuli as still containing some level of creaky voice. On the other hand, few of the creaky stimuli were rated as modal (visible from the empirical creakiness ratings in **Figure 5**). While more subtle, the main effect of manipulated median f_0 is also significant. When f_0 values increase, creakiness ratings decrease (135Hz – 115Hz: $\hat{\beta} = -0.12$, $\hat{\sigma} = 0.05$, CI = [-0.22, -0.02]; 155Hz – (135Hz + 115Hz): $\hat{\beta} = -0.10$, $\hat{\sigma} = 0.03$, CI = [-0.16, -0.04]; **Table 4**). This is consistent with more creakiness perceived at lower f_0 values, which is borne out in **Figure 5** (compare facets). The face gender effect alone does not reach significance (F – M: $\hat{\beta} = 0.03$, $\hat{\sigma} = 0.03$, CI = [-0.04, 0.10]; **Table 4**), only significant in interaction with f_0 . When f_0 values increase to 155 Hz, the face gender effect reverses relative to the two lower f_0 values, 115 Hz and 135 Hz (facegender*(155Hz – (135Hz + 115Hz)): $\hat{\beta} = -0.05$, $\hat{\sigma} = 0.02$, CI = [-0.10, -0.01]; **Table 4**). However, when comparing estimated marginal means using the emmeans package (Lenth, 2023) in R, that is, calculating the predicted effect of face gender at different f_0 levels, there is insufficient evidence for any significant effect at any given level.

Averaged over voice quality levels, when the median f_0 is 115 Hz, 135 Hz, or 155 Hz, the face gender effect fails to reach significance, but trends are compatible with the aforementioned face gender effect reversal (respectively, F–M: $\hat{\beta} = 0.08$, CI = $[-0.04, 0.20]$; $\hat{\beta} = 0.08$, CI = $[-0.04, 0.20]$; $\hat{\beta} = -0.08$, CI = $[-0.19, 0.03]$). From visual observation of **Figure 5**, we can see that at both 115 Hz and at 135 Hz, female faces prime creakier ratings (the blue triangular points are slightly higher than the orange circular points) but that this trend reverses at 155 Hz, male faces priming creakier ratings instead (orange circular points are higher than blue triangular points), though these observations remain speculative in the absence of statistical confirmation. In any case, the data from this experiment shows that a face gender effect is limited at best.

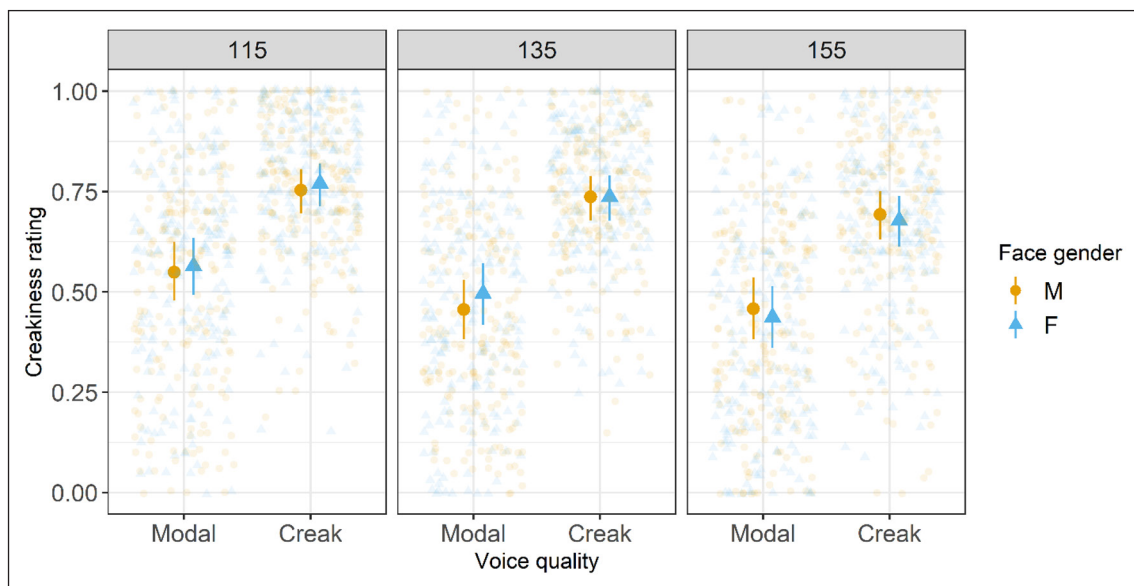


Figure 5: Predicted creakiness ratings (foreground points) with 95% credibility intervals and empirical creakiness ratings (background points) by face gender (color and shape), voice quality (x-axis) and f_0 (facets).

Empirical individual variation plots for creakiness ratings are available in Appendix A (Figure A8, A9 and A10) but will not be extensively discussed here. Figure A8 shows that participants do not differ in the direction of the effect of voice quality; utterances containing creaky voice are consistently rated as creakier than the modal utterances across all participants. In comparison to the effect of voice quality, there is much more individual variation in the effect of f_0 (see Figure A9), which suggests that this effect is less robust. Likewise, there is individual variation in the direction of the effect of face gender, at least for the participants who do show observable differences between gendered faces (see Figure A10). Overall, participants do vary in the range of ratings they attribute to the predictors (voice quality, f_0 and face gender). Some participants make use of the full VAS scale while others display more moderate responses, and some participants

exhibit large rating differences between experimental conditions whereas others only show slight differences. These differences in quantity of between-participant variation are confirmed by the by-participant varying intercept estimates in the model (sd ϕ intercept: $\hat{\beta} = 0.61$; sd α intercept: $\hat{\beta} = 1.73$; sd γ intercept: $\hat{\beta} = 1.81$; see Table A7 in Appendix A).

Fixed effects				
Coefficient	$\hat{\beta}$	$SE(\hat{\beta})$	95% CI	P_d
Intercept	0.48	0.11	[0.27, 0.69]	100
ϕ (phi) Intercept	2.10	0.10	[1.91, 2.30]	100
α (zoi) Intercept	-4.06	0.36	[-4.82, -3.42]	100
γ (coi) Intercept	0.75	0.53	[-0.19, 1.88]	93.97
FG(F - M)	0.03	0.03	[-0.04, 0.10]	79.40
VQ(cr - mo)	1.04	0.12	[0.80, 1.30]	100
f0_1(135 - 115)	-0.12	0.05	[-0.22, -0.02]	98.92
f0_2(155 - (135 + 115))	-0.10	0.03	[-0.16, -0.04]	99.98
FG(F - M):VQ(cr - mo)	-0.04	0.06	[-0.16, 0.08]	73.00
FG(F - M):f0_1(135 - 115)	0.00	0.05	[-0.09, 0.09]	52.18
FG(F - M):f0_2(155 - (135 + 115))	-0.05	0.02	[-0.10, -0.01]	98.98
VQ(cr - mo):f0_1(135 - 115)	0.10	0.10	[-0.09, 0.29]	84.00
VQ(cr - mo):f0_2(155 - (135 + 115))	-0.01	0.06	[-0.12, 0.10]	59.95
FG(F - M):VQ(cr - mo):f0_1(135 - 115)	-0.10	0.08	[-0.26, 0.06]	88.60
FG(F - M):VQ(cr - mo):f0_2(155 - (135 + 115))	0.03	0.04	[-0.06, 0.11]	74.40

Table 4: Zero-and-one-inflated Bayesian regression model table for fixed effects (FG = face gender, VQ = voice quality, f0 = new median f0) on creakiness ratings. Probabilities of direction (p_d) of credible effects are bolded.

4.2 Reaction time

For the reaction time data, long reaction times were excluded based on visual inspection of the distribution plot (see Figure A11 left panel in Appendix A), eliminating responses with reaction times over 15 seconds from the data ($n = 153$). The full summary of the shifted log-normal Bayesian regression model results is available in Table A12 of Appendix A. The main effect of voice quality on reaction time (in milliseconds) is significant. Overall, the modal voice condition leads to longer reaction times than the creaky condition (creak - modal: $\hat{\beta} = -0.06$, $\hat{\sigma} = 0.03$,

CI = $[-0.11, -0.01]$; Table A12), shown in **Figure 6** (negative slopes of lines). The main effects of f0 and of face gender were not found to be significant individually (see Table A12), but did reach significance thresholds in interaction, suggesting an opposite face gender effect at 135 Hz compared to 115 Hz (facegender*(135Hz – 115Hz): $\hat{\beta} = -0.06$, $\hat{\sigma} = 0.03$, CI = $[-0.12, -0.00]$; Table A12). While a comparison between estimated marginal means does not reach significance for any of the f0 levels, it appears from **Figure 6** that male faces lead to longer reaction times than female faces at 135 Hz (in both modal and creaky voices) and at 155 Hz (for creaky voices), but at 115 Hz, female faces lead to longer reaction times (specifically for modal voices). Because these reaction times are relatively long and may reflect more than just immediate processing, these results should be treated with caution. However, they do provide some evidence that incongruence between the face gender and the expected f0 for that gender lead to slower responses. This suggests that the faces and voices were integrated by participants.

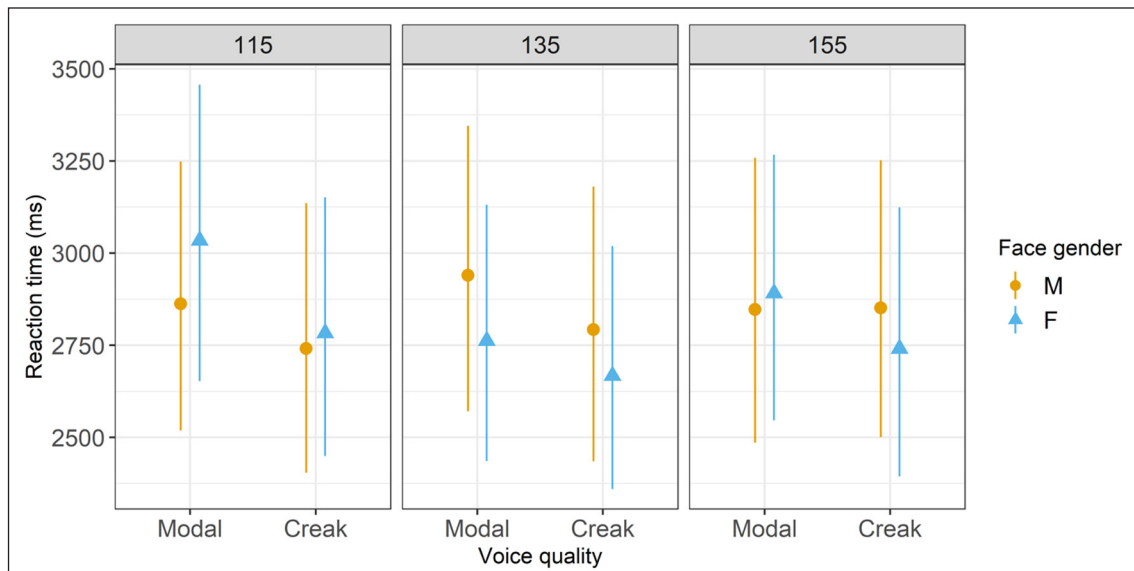


Figure 6: Predicted reaction times (in ms) with 95% credibility intervals by voice quality (x-axis), f0 (facets), and face gender (color and shape).

Empirical individual variation plots for reaction time are available in Appendix A (Figure A13, A14 and A15) but will not be discussed here.

5. Discussion

5.1 General discussion of results

The acoustic analysis of the stimuli showed more acoustic creakiness (higher proportions of unreliable f0 tracks, lower $H1^* - H2^*$, lower CPP and lower $HNR < 500$ Hz) for the (partially)

creaky utterances compared to the modal utterances. In the creakiness perception experiment, participants rated utterances containing creaky voice as creakier than those containing little to no creak (in modal voice). These results confirm that listeners can consistently recognize and distinguish creaky voice from modal voice in our experiment, validating that our f_0 and formant-altered modal and creaky voice stimuli induced perceptually distinct creakiness judgments.

Despite some individual variation across listeners, voices with lower median f_0 were rated as creakier than voices with higher median f_0 overall. This provides evidence that low pitch is closely related to perceptual creakiness, even within a relatively restricted range, aligning with previous perceptual studies (Davidson, 2019a; Li et al., 2023; White et al., 2024). Notably, our acoustic analyses of the stimuli did not clearly show increased creakiness for lower f_0 utterances. The proportions of unreliable f_0 tracks, CPP values and $\text{HNR} < 500$ Hz values were roughly similar for all f_0 levels, while only $\text{H1}^* - \text{H2}^*$ values decreased (indicative of more acoustic creak) alongside f_0 . This suggests that low f_0 is a perceptually useful cue to creaky voice, but that low f_0 does not necessarily lead to stronger acoustic cues to creaky voice. However, acoustic work has found covariation between creaky voice and low f_0 , suggesting that the relationship may be bidirectional (e.g., in American English in Davidson, 2020; in White Hmong in Garellek et al., 2013; in Mandarin in Kuang, 2017; and in Cantonese in Zhang & Kirby, 2020). Additionally, a recent study applying mediation analysis to creaky voice claims that a number of its acoustic correlates are at least partially mediated by f_0 (Brown, 2025). While the results of this study do show trends towards raised creakiness ratings for lower-pitched modal voices specifically (around 115 Hz), they are not statistically supported. Nevertheless, this trend is consistent with previous work showing increased false-alarm rates in lower low-pitched (roughly 100 Hz) modal voices (Davidson, 2019a; Li et al., 2023; White et al., 2024).

Face gender priming weakly influenced creakiness ratings in relation to f_0 values, despite not affecting creakiness ratings in isolation. Observational trends suggest that for ambiguously-gendered voices, a female face might prime increased creakiness perceptions for lower-pitched voices, but a male face might prime increased creakiness perceptions for higher-pitched voices. While not identical, these results share distinct similarities to those in White et al. (2024) in a call back to their speculation about a pitch given gender bias scenario. At matched mean f_0 values of 135 Hz, they found more creak identified in the woman's voice than the man's voice (and also more creak for a low-pitched man's modal voice), thus proposing that the underlying cause of this gender effect is low pitch, given the expected pitch range for gender. This hypothesis can account for our study's increased creakiness ratings for low-pitched voices (115 Hz and 135 Hz) primed by a female face, but fails to explain why a male face might prime listeners to perceive more creaky voice at a higher pitch (155 Hz). It is important to note, however, that this pattern of reversal is only observed in the empirical and model prediction plots and is not clearly statistically significant. If (and only if) we assume that this trend is reliable, we could argue that the pitch given gender bias may be less restricted than White et al. initially proposed, affecting

not only *low* pitch given gender but also *high* pitch given gender in a more general expectation-reliant hypothesis. Some motivations and implications of such a hypothesis are explored further in following sections (5.2 and 5.3).

As for reaction time results, slower reaction time was observed for modal voice utterances than for creaky voice utterances, possibly indicating that listeners experienced more difficulty in determining whether there was absence of creak. Visual trends (Figure 8) point to longer reaction times for unexpected pitch-gender combinations, that is for lower-pitched utterances (115 Hz) primed by a woman’s face and higher-pitched utterances (135 Hz and creaky 155 Hz) primed by a man’s face. This is consistent with the idea that linguistic perception is influenced by social expectations. Walker & Hay (2011) show that when word age (determined by the typical age of speakers who often use that word) and voice age are congruent, lexical access is facilitated, resulting in faster reaction times and higher accuracy than when they conflict. They take this as evidence that incongruency between linguistic and social information increases processing costs. Following the same reasoning, when speaker f0 and face gender create conflicting gender expectations, this could lead to increased processing demands and uncertainty, even though they are not being asked to make a gender judgement. These reaction time effects therefore give tentative support that our face gender manipulation created gendered social expectations for the stimuli. That said, surprisingly, this proposed effect reverses for modal voices at 155 Hz, with women’s faces showing slightly higher reaction times. However, given that our reaction times are not based on button-press responses and thus our data is very noisy, conclusions drawn from only reaction time data remain speculative.

Overall, this study shows that (Canadian English) listeners’ perception of creaky voice is affected by speaker f0 even when controlling for gender biasing formant ratios. However, creakiness is less convincingly impacted by independently providing a gender cue with a face. Situating these results with respect to the hypotheses or scenarios described by Davidson (2019a) and White et al. (2024), they provide additional evidence for an acoustic hypothesis (pitch contrast bias), but again in the opposite direction as predicted: *lower* pitch (and/or a smaller modal-to-creaky-voice pitch differential) increasing perceived creakiness, as well as weak evidence for a bias hypothesis (gender bias) in which the exact direction of said bias is inconclusive—but seems to differ according to pitch (see **Table 5**).

Creakiness ratings	Hypothesis tested	Hypothesis predictions
155 Hz < 135 Hz < 115 Hz	Acoustic	low-f0 < high-f0
M < F M < F F < M	Gender	M < F

Table 5: Summary of the listener creakiness ratings across median f0 values and face genders (for both voice qualities), with respect to the hypotheses and predictions tested.

As perception work on creaky voice has expanded, the results have revealed reliable trends. First, lower-pitched voices are perceived as creakier, which coincides with the acoustic literature that finds stronger creaky voice correlates for men (whose voices are typically lower-pitched) than for women (whose voices are typically higher-pitched). Second, briefly setting Li et al. (2023) aside because it does not vary f_0 within gender, there is limited evidence in support of stronger creak judgements for women. Collectively, the tendency to find more creak in actual/perceived women's voices compared to men's is inconsistent in Davidson (2019a), which only occurs in the first study, but not the subsequent replication; and in White et al. (2024), creak is unlikely to occur in the real world, given that it is only observable if male and female voices are matched at ambiguous pitch values. If anything, creak is weak in the current study and largely dependent on pitch as well. Thus, perception studies are *not* at odds with production studies of creaky voice. Rather, the perception of creaky voice in recent sociolinguistic and sociophonetic studies relying on impressionistic coding is at odds with i) (almost) all previous acoustic studies of creaky voice (e.g., Brown & Sonderegger, 2025; Gittelsohn, 2021; Klatt & Klatt, 1990); ii) older sociolinguistic studies of creaky voice (in the U.K.), also relying on impressionistic coding (e.g., Henton & Bladon, 1988; Stuart-Smith, 1999); and iii) to some degree, current systematic perceptual sociophonetic studies of creaky voice (e.g., Davidson, 2019a; White et al., 2024). Given this, the central question now shifts from explaining a production-perception mismatch in creaky voice to explaining how so many recent sociolinguistic studies converge on increased creak in women's voices through impressionistic coding. What makes the perception of creaky voice in recent impressionistic studies so different from both its acoustic realization and its perception in older impressionistic studies and controlled perception studies? This paper argues that neither a social gender bias nor an acoustic pitch-contrast bias fully accounts for the inflated perception of creak in women's voices. Identifying the source of conflict between diametrically opposed findings on gendered creaky voice use has proven to be complex and multifaceted, requiring further investigation. The rest of this discussion is dedicated to providing directions for future sociophonetic research on creaky voice.

5.2 Questioning the validity of measures

If incongruent pitch-gender pairings subvert listener expectations, then how does this lead to increased perception of creak? As mentioned above (Section 5.1), White et al. (2024) suggest that this mechanism is related to low pitch with respect to the expected pitch of that gender. Assuming that low pitch is an important cue to creaky voice, this proposal follows. What is not so clear is why, in the current study's results, a higher-pitched voice relative to expected pitch of a given face gender might lead to increased perceived creakiness. One possibility is that while our experiment (and potentially other creak perception experiments) was designed to evaluate perceived creakiness, listeners may not clearly be basing their ratings on creakiness

alone. Listeners could be integrating other perceptual judgements of gender non-conformity, unnaturalness, or salience, for example, into their ratings. As a result, increased “creakiness” could in theory reflect increased “weirdness” to listeners instead.

This paper highlights how the perceptual identification of creaky voice is subject to social, cognitive or acoustic biases. It is worth noting, however, that existing acoustic studies also suffer from methodologically heterogeneity. Altogether, they implement a varied range of acoustic correlates to quantify/identify creaky voice: Some rely on one or two measures, usually low f_0 or $H1^* - H2^*$ (Pépiot, 2014; Loakes & Gregory, 2020; Syrdal 1996; Szakay, 2012; Szakay & Torgersen, 2015); some argue for a minimum of two measures, $H1^* - H2^*$ and CPP (Garellek & Esposito, 2023; Seyfarth & Garellek, 2018); whereas others employ numerous measures (Brown & Sonderegger, 2025; Hanson & Chuang, 1999; Iseli et al., 2007; Kuang, 2017; Lortie et al., 2015), and even apply dimensionality-reduction methods (Johnson & Babel, 2023; Keating et al., 2023a); and some others use automatic creak detection tools which can be variably based on a single acoustic cue (Sebregts et al., 2023; Szakay & Torgersen, 2019) or multiple acoustic cues (Irons & Alexander, 2016; Kuang, 2017). This is a testament to the lack of consensus on the precise set of acoustic correlates to creak. The psychoacoustic model of voice (Kreiman et al., 2014; Kreiman et al., 2021) (introduced in Section 2.2) describes various acoustic cues as perceptually validated and individually necessary to the analysis of voice, and Keating et al.’s (2023a) cross-linguistic multi-dimensional analysis of voice argues for even more cues to be implemented. However, few perceptual studies thereafter (and even fewer perceptual studies of *creaky* voice) have made use of all of these reportedly key parameters, usually selecting a subset of these. As such, another possible explanation for the production-perception mismatch could postulate that previous studies do not use acoustic measures that reflect perceptual creakiness ratings. Perhaps dimensionality reduction methods could better integrate the various acoustic measures, including those from the psychoacoustic model of voice, and better represent the multi-dimensionality of the holistic percept of creaky voice. At the same time, some acoustic cues may be less perceptually relevant to creak than presumed in previous acoustic studies (e.g., Huang, 2019; Keating et al., 2023b; Khan et al., 2015). Without more in-depth empirical examination of this proposal, it cannot be certainly refuted. That said, given the abundance, breadth and varied use of cues, as well as consistency in results of previous acoustic literature of creak, it would be surprising if a slight modification to the method of analyzing acoustic measures would suddenly result in opposite findings (i.e., those that would support more creak acoustically for women).

Another alternative is that the use of creaky voice varies by gender in a way that is not captured by rates or averages used in previous studies and is more perceptually salient in women than in men. Investigations into creaky voice thus far have largely been restricted to descriptions of prevalence or average quantities of creak in the voice, often ignoring nuances in creaky voice usage. Burin (2022), for instance, notes more variability in women’s production of creaky voice

than in men's in American English, and increased pitch variability for women is reported to be widespread (e.g., Gisladdottir et al., 2023). It is therefore conceivable that there exist gender differences in creaky voice usage as a function of creak location, intensity, type, or variance, for example. Whether these differences in use map to more perceptual salience remains empirically understudied, however.

5.3 Effects of expectation, awareness, and salience

In traditional impressionistic studies of creaky voice (e.g., Henton & Bladon, 1988, Stuart-Smith, 1999), the social expectation was for men, older upper-class men in particular, to have creakier voices. This expectation is in alignment with phonetic work corroborating this pattern in voice acoustics (e.g., Klatt & Klatt, 1990) and voice articulation, indicating physiological differences in vocal fold length and thickness as a function of gender (e.g., Hollien, 1974). Prior to the early 21st century, social expectations and acoustic expectations of voice were compatible. As of roughly 2010, it has become apparent that social expectations of voice have shifted to women exhibiting creakier voices (at least among most North American, English-speaking communities), despite a clear lack of evidence for any change in articulation (e.g., Zhang, 2021) or acoustics (e.g., Brown & Sonderegger, 2025) over time. At present, it appears that social expectations of voice conflict with phonetic expectations of voice.

Labov (1972) proposes a classification of sociolinguistic variables (i.e., socially-stratified linguistic behavior) as a function of speaker/listener awareness. The concept of awareness is then often attributed to *salience* of the linguistic variable (but see Jaeger & Weatherholtz, 2016, for a more information theoretic model of salience as a function of surprisal, frequency, and perceived social informativeness). Labov describes three levels of awareness: 1) *indicators* are below the threshold of conscious awareness, never noticed by speakers/listeners; 2) *markers* are typically perceived at least subconsciously as listeners can manipulate the variable usage depending on context, suggesting implicit awareness of them; and 3) *stereotypes* are socially marked (often stigmatized) features characterized by explicit awareness, usually subject to meta-linguistic commentary. Considering this classification, creaky voice does not clearly fall under one single level of awareness. In American popular discourse, creaky voice (better known as vocal fry) is likely considered to be a stereotype associated with young women. Speaker and listeners are highly conscious of it and are often advised to avoid using it due to its controversial social evaluation. These strong opinions in the mainstream media have potential to influence general perception of creaky voice but are less likely to be an entirely faithful representation of widespread views of creak (see Coupland, 2014; Trudgill, 2014, for discussions of media and language). More broadly, it seems that creaky voice could be considered a marker, used by different social groups but often in a subconscious way, i.e., speakers are not fully aware that they are using it. As such, it is possible that creaky voice variation by gender is twofold, acting as a stereotype when women use it, but as a mere marker when men use it. In the current state

of affairs, these pressures are both at odds, pulling creaky voice perception in two different directions and obscuring direct gender effects (independent of f_0) in this study. In fact, explicit and implicit social evaluations might induce different perceptual behaviors and involve separate cognitive processing (see Evans, 2008, for a review). From another perspective, Juskan's first necessary condition for effective priming is sufficient social salience of the variable (2016), and conflicting social expectations might prevent the satisfaction of this condition.

Campbell-Kibler (2009) introduces a similar concept to this, sociolinguistic "bullet-proofing," showing that the social meaning of the variable (ING) (i.e., the use of -ing vs. -in) depends on listeners' perceptions of the speaker. While -ing is generally associated with higher intelligence and -in with lower intelligence, this pattern only emerged for working-class non-Southerners. Southerners were consistently rated as less intelligent regardless of their (ING) use, and middle-to-upper class (non-Southern) speakers were effectively bullet-proof to any effect of (ING) variation on perceived intelligence. This demonstrates that the social evaluation of linguistic variation can differ greatly as a function of listeners' expectations for/stereotypes about distinct social groups. It is possible that men's voices are bullet-proof to creaky voice use in the sense that creaky voice use is not perceived as socially significant for them and is therefore somehow less salient or perceptible to the listener.

Another possible explanation for the increase in reported creakiness in women's speech found in recent sociolinguistic studies is creaky voice in men may undergo perceptual normalization due to its frequency, an automatic process that is below the level of consciousness. Because creaky voice is acoustically common in men's voices and listeners are frequently exposed to instances of creak in men's voices, its presence may be filtered out or perceived as less salient (akin to exemplar-based categorization, see Drager & Kirtley, 2016, or Johnson, 1997). This process would be comparable to reduced salience of creak in prosodically expected positions, identified consistently less utterance-finally (Davidson, 2019a; Li et al., 2023). This could make creaky voice in men less noticeable and thus under-identified in impressionistic coding, artificially inflating the identification of creak in women's speech. However, this account does not explain why controlled perceptual experiments such as Davidson (2019a) do not similarly show increased creak judgments for women. If creaky voice in men was genuinely less salient, we might expect listeners in these perception studies to rate male voices as less creaky overall. Given that the opposite occurs, normalization effects alone are not sufficient to account for the gender asymmetries found in impressionistic sociolinguistic studies and controlled perceptual experiments.

5.4 Limitations of the current study

Participant demographic characteristics were not considered in this analysis, following Davidson's (2019a) finding that listener gender did not affect creaky voice identification and White et al.'s (2024) concurring decision not to account for listener effects. There is some evidence, however,

that listener characteristics can influence perception (e.g., Yu & Zellou, 2019, among others). Preliminary empirical plots by listener gender and age are provided in figures A16, A17 and A18 (Appendix A) and suggest a possible gender effect whereby female listeners are more likely to rate voices as creakier than male listeners, but no consistent effects of age. As such, the inclusion of listener demographic information in formal modelling of creaky voice ratings may allow for a more thorough understanding of creaky voice perception.

In addition, closer examination of the stimuli and questionnaire data from the creak experiment provided some useful information about how participants perceived the stimuli and felt about the experiment. Generally, participants did not find the task to be very difficult. A few participants noted that the voices did not match the faces or that they ignored the faces entirely, suggesting that they may have been less influenced by the face gender primes. This could explain the weak face gender effect observed in this study. Other participants remarked that the voices/audio sounded a bit unnatural, distorted or noisy, which could be concern for the ecological validity of the stimuli. The Praat “Change gender” function did seem to introduce some distortion into the voice signal (both the original recordings and manipulated stimuli can be accessed on the OSF page), possibly leading some listeners to identify creakiness even in the modally-voiced utterances as seen in **Figure 4** above. The modal stimuli also appeared to be produced at a slightly faster speech rate than the creaky stimuli (see Figure A5 in Appendix A), which could facilitate creak (Brown & Sonderegger, 2025), also potentially contributing to the somewhat creaky-rated modal stimuli. As a result, differences in creakiness ratings between modal and creaky stimuli may be less pronounced.

Moreover, in the questionnaire data, a few participants shared that they were familiar with creaky voice; one participant explicitly stated their dislike for it. However, a systematic assessment of listeners’ attitudes towards creak, or their stereotypes about it prior to participation in the experiment, was not included in this study. At present, it has not yet been empirically demonstrated that Canadian English listeners hold the same social biases as American English listeners, potentially explaining weaker face gender priming in the current study. There is some evidence (also from a matched-guise face priming design) that Canadian English listeners differ in their perception of speech and stereotypes compared to American English listeners (Kutlu et al., 2022). Nevertheless, it has been shown that Canadian listeners do hold similar negative judgements of women who use creaky voice (Goodine & Johns, 2014) as American listeners (e.g., Anderson et al., 2014). Narratives promoting criticism of creaky voice are likely to reach Canadian audiences due to the pervasive influence of American media complemented by coverage in mainstream Canadian outlets and internet blogs/op-eds (e.g., Chattopadhyay, 2015; Weber, 2017). Furthermore, the training phase itself may have affected participants’ responses. The presentation of two stereotyped and comparably extreme examples of creak could have inadvertently raised the perceptual threshold for what counts as creaky, possibly decreasing

sensitivity to local and/or more subtle instances of creak. On the other hand, although our description of creaky voice deliberately avoided reference to low f_0 /pitch, participants still tended to associate lower f_0 s with greater creakiness, which supports the reliability of this effect.

It is also possible that some of the ambiguous voices and their pairings with some of the faces could have led to transgender or non-binary gender voice percepts. Gender expectations and stereotypes related to transgender and non-binary creaky voice use remain severely understudied (Eckert & Podesva, 2021), and the few studies that do exist present opaque results. As a consequence, it is unclear how and to what extent these might impact listener creakiness ratings. Becker et al. (2022) find that creaky voice variation is only predicted by gender in interaction with speech style and hypothesize that larger style-shifts amongst non-binary AFAB individuals on testosterone and trans men (regardless of hormonal status) may be attributable to an avoidance of features ideologically associated with cis women's speech. This study's findings are indicative of the complex relationship between creaky voice use and projections of personal identity (with respect to gender in this case).

By the same token, it is possible that the (often negative) stereotypes surrounding young women's creaky voice are not specific to creaky voice, but rather to a particular persona. The combination of individual meaningful elements (e.g., linguistic features) to construct a more broad and complex meaningful entity (e.g., a persona) is referred to as *bricolage* within the domain of third-wave sociolinguistics (Eckert, 2008; originating from fields of sociology and anthropology, Hebdige, 1984; Lévi-Strauss, 1962). In Podesva's (2007) study of a young gay doctor's speech, he determined that this speaker made use of extensive falsetto voice quality (greater duration, f_0 range and maximum), exaggerated stop releases (longer and more intense bursts), and lexical choices (e.g., "dear") to form a so-called diva persona. Likewise, in some cases, creaky voice may be used concurrently with other sociolinguistic variables to construct specific personae. Generalization towards exact personae is not clear at present: Creaky voice has been proposed to be linked to the indexation of character or affective traits like aloofness, disengagement, and negativity (e.g., Podesva, 2018; Pratt, 2018), but also to upward-mobility, authority, and toughness (e.g., Dilley et al., 1996; Mendoza-Denton, 2011; Yuasa, 2010). For instance, creaky voice in combination with *uptalk* and the use of "like" as a discourse marker could convey a persona of a ditzy Millennial or Gen Z woman, whereas in combination with low modal pitch and slow speech rate it could convey the entirely different persona of a confident businesswoman. The distinct personae, instead of creaky voice use specifically, could also prompt different social and affective evaluations by listeners, influencing their perception of creaky voice in a variety of ways.

6. Conclusion

This study found evidence that speaker f_0 affects the perception of creaky voice, with low f_0 s increasing creakiness ratings even within a gender-neutral pitch range with gender-neutral

formant structure. While face gender primes did not independently influence creakiness ratings, subtle interactions between perceived face gender and f_0 suggest that socio-physiological expectations may shape listener judgements to some extent, especially when pitch and gender cues are incongruent. These effects were weak at best, however, failing to provide convincing evidence for a robust social gender bias in creaky voice perception. Crucially, our findings do not support the notion that increased creak perception in women's voices can be explained by a widespread gender bias or an acoustic bias alone. Instead, they point to a more complex interplay between acoustic cues and social expectations, one that may be contingent on methodological decisions, listener awareness and/or the salience of gendered voice norms. Alongside previous systematic production and perception studies of creaky voice, the present results demonstrating a marked lack of empirical evidence for greater perception of creakiness in women's voices cast further doubt on the pervasive narrative—perpetuated in popular discourse and recent influential sociolinguistic studies—that women are often creakier. From a methodological standpoint, these findings raise important questions about impressionistic coding practices in analyses of voice, especially when divorced from acoustic or perceptual validation.

Broadly, this paper explores how acoustic and social cues to speaker identity can influence listener voice perception. It highlights the limitations of phonetically-grounded but socially agnostic models of voice perception, advocating for more nuanced approaches that integrate acoustic and social information into cognition. By bringing together acoustic, perceptual, and social dimensions of speech, we can better understand how creaky voice manifests in production, molds listener judgements, and reveals broader patterns of language use and social meaning.

Data accessibility statement

All stimuli, code and data are provided on the paper's OSF page at <https://osf.io/f45yh/>, doi:10.17605/osf.io/f45yh.

Additional file

The additional file for this article can be found as follows:

- Appendix A. DOI: <https://doi.org/10.16995/labphon.24285.s1>

Ethics and consent

This research was reviewed and approved by the McGill Research Ethics Board 2 under REB# 419–0319. All participants provided informed consent to participate in the study, in accordance with ethical research guidelines.

Acknowledgements

We are grateful to Morgan Sonderegger for his expert insight, discussion, and comments on various versions of this work, to Eleanor Chodroff for her guidance on methodological decisions during the project's conception, and to Abby Walker and Charlotte Vaughn for the thoughtful conversations that helped sharpen our thinking. We also appreciate the contributions of colleagues at McGill in MCQLL and P* and audiences at [mot^h]2025 for their feedback and discussions. We thank two anonymous reviewers and associate editor, Yao Yao, whose engaged feedback and constructive suggestions substantially strengthened this manuscript.

Funding information

This work was supported by the Social Sciences and Humanities Research Council of Canada [435-2024-0996] grant to Meghan Clayards, as well as the Fonds de recherche du Québec – Société et culture [373124] grant, the Arts Graduate Research Enhancement and Travel Awards program at McGill, and the CRBLM Travel Award to Jeanne Brown.

Competing interests

Meghan Clayards is a member of the editorial board for *Laboratory Phonology*, which is on a voluntary basis. All other authors declare that they have no competing interests.

Author contributions

Jeanne Brown and Meghan Clayards jointly conceptualized and designed the study and interpreted the results. Jeanne Brown created the stimuli, conducted the experiment, led the data analysis,

and drafted the manuscript. Meghan Clayards supervised the project and provided funding. Both authors revised the manuscript, read, and approved the submitted version.

References

- Abdelli-Beruh, N. B., Wolk, L., & Slavin, D. (2014). Prevalence of vocal fry in young adult male American English speakers. *Journal of Voice*, *28*(2), 185–190. <https://doi.org/10.1016/j.jvoice.2013.08.011>
- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.
- Alderton, R. (2020). Speaker gender and salience in sociolinguistic speech perception: Goose-fronting in Standard Southern British English. *Journal of English Linguistics*, *48*(1), 72–96. <https://doi.org/10.1177/0075424219896400>
- Anderson, R. C., Klofstad, C. A., Mayew, W. J., & Venkatachalam, M. (2014). Vocal fry may undermine the success of young women in the labor market. *PLoS ONE*, *9*(5), e97506. <https://doi.org/10.1371/journal.pone.0097506>
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behaviour Research Methods*, *52*(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Auer, P., & Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect change: Convergence and divergence in European languages* (pp. 335–357), Cambridge University Press. <https://doi.org/10.1017/CBO9780511486623>
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *Journal of the Acoustical Society of America*, *137*(5), 2823–2833. <https://doi.org/10.1121/1.4919317>
- Becker, K., Khan, S. u. D., & Zimman, L. (2022). Beyond binary gender: Creaky voice, gender, and the variationist enterprise. *Language Variation and Change*, *34*(2), 215–238. <https://doi.org/10.1017/S0954394522000138>
- Blomgren, M., Chen, Y., Ng, M. L., & Gilbert, H. R. (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *Journal of the Acoustical Society of America*, *103*(5), Pt 1, 2649–2658. <https://doi.org/10.1121/1.422785>
- Boersma, P., & Weenink, D. (1992–2025): Praat: Doing phonetics by computer [Computer program]. Version 6.1.16, retrieved 12 August 2020. <https://www.praat.org>
- Bouavichith, D. A., Calloway, I., Craft, J. T., Hildebrandt, T., Tobin, S. J., & Beddor, P. S. (2019). Perceptual influences of social and linguistic priming are bidirectional. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 1039–1043. Australasian Speech Science and Technology Association and International Phonetic Association.
- Brown, J. (2025). Acoustic correlates in the production of creaky voice: Mediation by f0 [Oral presentation]. *72nd Annual CLA Conference*. Montreal, Canada. <https://cla-acl.ca/programmes/congres-de-2025-meeting.html>

- Brown, J., & Sonderegger, M. (2025). A sociophonetic study of creaky voice across language, gender and age in Canadian English-French bilinguals. *Journal of Phonetics*, *112*, 101431. <https://doi.org/10.1016/j.wocn.2025.101431>
- Brubaker, H., Whitfield, J. A. & Schoonmaker Rodgers, J. (2016). Fundamental frequency characteristics of modal and vocal fry registers. *Honors Projects*, *267*. Bowling Green State University. <https://scholarworks.bgsu.edu/honorsprojects/267>
- Burin, L. (2022). *Perception and accommodation among French learners of English: An acoustic and electroglottographic study of creaky voice*. [Doctoral dissertation, Université Paris Cité].
- Bürkner, P. (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>.
- Calhoun, S., & White, H. (2025). What makes iconic pitch associations “natural”: The effect of age on affective meanings of uptalk and creak. *Language and Speech*, 238309251314863. <https://doi.org/10.1177/00238309251314863>
- Callier, P. (2013) *Linguistic context and the social meaning of voice quality variation*. [Doctoral dissertation, Georgetown University].
- Campanella, S., & Belin, P. (2007) Integrating face and voice in person perception. *Trends in Cognitive Sciences*, *11*(12), 535–543. <https://doi.org/10.1016/j.tics.2007.10.001>
- Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change*, *21*(1), 135–156. <https://doi.org/10.1017/S0954394509000052>
- Catford, J. C. (1964). Phonation types: The classification of some laryngeal components of speech production. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, J. L. M. Trim (Eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961* (pp. 26–37). Longmans.
- Chattopadhyay, P. (2015, July 28). ‘Vocal Fry’ undermines empowered young women, says Naomi Wolf. *The Current* [Radio broadcast]. CBC Radio. <https://www.cbc.ca/radio/thecurrent/the-current-for-july-28-2015-1.3170502/vocal-fry-undermines-empowered-young-women-says-naomi-wolf-1.3170511>
- Crowhurst, M. J. (2018). The influence of varying vowel phonation and duration on rhythmic grouping biases among Spanish and English speakers. *Journal of Phonetics*, *66*, 82–99. <https://doi.org/10.1016/j.wocn.2017.09.001>
- Dallaston, K., & Docherty, G. (2020). The quantitative prevalence of creaky voice (vocal fry) in varieties of English: A systematic review of the literature. *PLoS ONE*, *15*(3). e0229960. <https://doi.org/10.1371/journal.pone.0229960>
- Davidson, L. (2019a). The effects of pitch, gender, and prosodic context on the identification of creaky voice. *Phonetica*, *76*(4), 235–262. <https://doi.org/10.1159/000490948>
- Davidson, L. (2019b). Perceptual coherence of creaky voice qualities. In S. Calhoun, P. Escudero, M. Tabain and P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 147–151, Australasian Speech Science and Technology Association and International Phonetic Association.
- Davidson, L. (2020). Contributions of modal and creaky voice to the perception of habitual pitch. *Language*, *96*(1), e22–e37. <https://doi.org/10.1353/lan.2020.0013>

- Davidson, L. (2021). The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *Wiley Interdisciplinary Reviews Cognitive Science*, 12(3), 1547. <https://doi.org/10.1002/wcs.1547>
- DeBruine, L., & Jones, B. (2017). Face Research Lab London Set [Dataset]. <https://doi.org/10.6084/m9.figshare.5047666.v5>
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24, 423–444. <https://doi.org/10.1006/jpho.1996.0023>
- D'Onofrio, A. (2015). Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics*, 19(2), 241–256. <https://doi.org/10.1111/josl.12115>
- D'Onofrio, A. (2016). *Social meaning in linguistic perception*. [Doctoral dissertation, Stanford University].
- D'Onofrio, A. (2020). Personae in sociolinguistic variation. *WIREs Cognition Science*, 11(6), e1543. <https://doi.org/10.1002/wcs.1543>
- Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass*, 4(7), 473–480. <https://doi.org/10.1111/j.1749-818X.2010.00210.x>
- Drager, K. (2011). Speaker age and vowel perception. *Language and Speech*, 54(1), 99–121. <https://doi.org/10.1177/0023830910388017>
- Drager, K., & Kirtley, J. (2016). Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 1–24). Cambridge University Press. <https://doi.org/10.1017/CBO9781139680448.003>
- Duarte-Borquez, C., Van Doren, M., & Garellek, M. (2024). Utterance-final voice quality in American English and Mexican Spanish bilinguals. *Languages*, 9, 70. <https://doi.org/10.3390/languages9030070>
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), 453–476. <https://doi.org/10.1111/j.1467-9841.2008.00374.x>
- Eckert, P., & Podesva, R. (2021). Non-binary approaches to gender and sexuality. In J. Angouri & J. Baxter (Eds.), *The Routledge handbook of language, gender, and sexuality* (pp. 25–36). Routledge. <https://doi.org/10.4324/9781315514857>
- Esling, J. (1978). The identification of features of voice quality in social groups. *Journal of the International Phonetic Association*, 8(1–2), 18–23. <https://doi.org/10.1017/S0025100300001699>
- Evans, J. S. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Fessenden, M. (2011, December 9). 'Vocal fry' creeping into U.S. speech. *Science*. <https://www.science.org/content/article/vocal-fry-creeping-us-speech#:~:text=Since%20the%201960s%2C%20vocal%20fry,or%20females%20varies%20among%20languages>
- Foulkes, P., & Hay, J. B. (2015). The emergence of sociophonetic structure. In B. MacWinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 292–313). John Wiley & Sons. <https://doi.org/10.1002/9781118346136.ch13>

- Gallena, S. K., & Pinto, J. A. (2021). How graduate students with vocal fry are perceived by speech-language pathologists. *Perspectives of the ASHA Special Interest Groups*, 6(6), 1554–1565. https://doi.org/10.1044/2021_PERSP-21-00083
- Garellek, M. (2015). Perception of glottalization and phrase-final creak. *Journal of the Acoustical Society of America*, 137(2), 822–831. <https://doi.org/10.1121/1.4906155>
- Garellek, M. (2019). The phonetics of voice. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 75–106). Routledge.
- Garellek, M. (2022). Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics*, 94, 1–22. <https://doi.org/10.1016/j.wocn.2022.101155>
- Garellek, M., & Esposito, C. M. (2023). Phonetics of White Hmong vowel and tonal contrasts. *Journal of the International Phonetic Association*, 53(1), 213–232. <https://doi.org/10.1017/S0025100321000104>
- Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *Journal of the Acoustical Society of America*, 133(2), 1078–1089. <https://doi.org/10.1121/1.4773259>
- Gisladottir, R., Helgason, A., Halldorsson, B., Helgason, H., Borsky, M., Chien, Y., & Stefansson, K. (2023). Sequence variants affecting voice pitch in humans. *Science Advances*, 9(23). <https://doi.org/10.1126/sciadv.abq2969eabq2969>
- Gittelsohn, B., Leemann, A., & Tomaschek, F. (2021). Using crowd-sourced speech data to study socially constrained variation in nonmodal phonation. *Frontiers in Artificial Intelligence*, 3, 565682. <https://doi.org/10.3389/frai.2020.565682>
- Gobl, C. & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, 40(1–2), 189–212. [https://doi.org/10.1016/S0167-6393\(02\)00082-1](https://doi.org/10.1016/S0167-6393(02)00082-1)
- Goodine, A., & Johns, A. (2014). “Would you like fries with thaaaat?” Investigating vocal fry in young female Canadian English speakers. *Strathy Student Working Papers on Canadian English 2014*, 1–15.
- Google Books. (2025, April 10). *Vocal fry* [Infographic]. Google Books Ngram Viewer. https://books.google.com/ngrams/graph?content=vocal+fry&year_start=1800&year_end=2022&corpus=en&smoothing=3
- Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, 29(4), 383–406. <https://doi.org/10.1006/jpho.2001.0147>
- Grim, R. (2015, March 31). My girlfriend went to a speech therapist to cure her vocal fry. *Vice*. <https://www.vice.com/en/article/i-took-my-girlfriend-to-a-speech-therapist-to-cure-her-annoying-vocal-fry-988/>
- Hanson, H. M., & Chuang, E. S. (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, 106(2), 1064–1077. <https://doi.org/10.1121/1.427116>
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4), 865–892. <https://doi.org/10.1515/LING.2010.027>

- Hay, J., Nolan, A., & Drager, K. (2006a). From fush to feesh: Exemplar priming in speech perception. *Linguistic Review*, 23(3), 351–379. <https://doi.org/10.1515/TLR.2006.014>
- Hay, J., Warren, P., Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34, 458–484. <https://doi.org/10.1016/j.wocn.2005.10.001>
- Hebdige, D. (1984) *Subculture: The meaning of style*. Methuen.
- Heiss, A. (2021). A guide to modeling proportions with Bayesian beta and zero-inflated beta regression models. <https://doi.org/10.59350/7p1a4-0tw75>
- Henton, C., & Bladon, A. (1988). Creak as a sociophonetic marker. In L. Hyman & C. Li (Eds.), *Language, speech and mind: Studies in honor of Victoria A. Fromkin* (pp. 3–29). Routledge. <https://doi.org/10.4324/9781003629610-2>
- Hollien, H. (1974). On vocal registers. *Journal of Phonetics*, 2(2), 125–143. [https://doi.org/10.1016/S0095-4470\(19\)31188-X](https://doi.org/10.1016/S0095-4470(19)31188-X)
- Hollien, H., & Michel, J. F. (1968). Vocal fry as a phonational register. *Journal of Speech and Hearing Research*, 11(3), 600–604. <https://doi.org/10.1044/jshr.1103.600>
- Huang, Y. (2019). The role of creaky voice attributes in Mandarin tonal perception. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 1465–1469. Australasian Speech Science and Technology Association and International Phonetic Association.
- Irons, S. T., & Alexander, J. E. (2016). Vocal fry in realistic speech: Acoustic characteristics and perceptions of vocal fry in spontaneously produced and read speech. *Journal of the Acoustical Society of America*, 140(4), 3397–3397. <https://doi.org/10.1121/1.4970891>
- Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *Journal of the Acoustical Society of America*, 121(4), 2283–2295. <https://doi.org/10.1121/1.2697522>
- Jaeger, T. F., & Weatherholtz, K. (2016). What the heck is salience? How predictive language processing contributes to sociolinguistic perception. *Frontiers in Psychology*, 7(1115), 1–5. <https://doi.org/10.3389/fpsyg.2016.01115>
- Jaslow, R. (2011, December 16). Are “creaking” pop stars changing how young women speak? *CBS News*. <https://www.cbsnews.com/news/are-creaking-pop-stars-changing-how-young-women-speak/>
- Jessee, E., & Calder, J. (2025). The cisgender listening subject in sociolinguistic perception: Transgender identity affects sibilant categorization in American English. *Journal of Sociolinguistics*, 0, 1–14. <https://doi.org/10.1111/josl.12702>
- Johnson, F. L. (2000). *Speaking culturally: Language diversity in the United States*. Sage. <https://doi.org/10.4135/9781452220406>
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–165). Academic Press.

- Johnson, K., & Babel, M. (2023). The structure of acoustic voice variation in bilingual speech. *Journal of the Acoustical Society of America*, 153(6), 3221–3238. <https://doi.org/10.1121/10.0019659>
- Johnson, K., Strand, E., & D’Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, 27(4), 359–384. <https://doi.org/10.1006/jpho.1999.0100>
- Juskan, M. (2016). *Production and perception of local variants in Liverpool English: Change, salience, exemplar priming*. [Doctoral dissertation, Albert Ludwig’s University].
- Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. University of Glasgow.
- Keating, P., Garellek, M., Kreiman, J., & Chai, Y. (2023b, May 8–12). Acoustic properties of subtypes of creaky voice [Poster presentation]. *The 184th Meeting of the Acoustical Society of America*. Chicago, IL, USA. <https://doi.org/10.1121/10.0018918>
- Keating, P., Kuang, J., Garellek, M., Esposito, C., & Khan, S. (2023a). A cross-language acoustic space for vocalic phonation distinctions, *Language*, 99(2), 351–389. <https://doi.org/10.1353/lan.2023.a900607>
- Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *Journal of the Acoustical Society of America*, 132(2), 1050–1060. <https://doi.org/10.1121/1.4730893>
- Khan, S., Becker, K., & Zimman, L. (2015). The acoustics of perceived creaky voice in American English. *Journal of the Acoustical Society of America*, 138(3), 1809–1809. <https://doi.org/10.1121/1.4933741>
- Kirby, J. (2018). Praatsauce: Praat-based tools for spectral analysis. <https://github.com/kirbyj/praatsauce>
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87(2), 820–857. <https://doi.org/10.1121/1.398894>
- Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics*, 14(2), Article 12.
- Kreiman, J., Gerratt, B., Garellek, M., Samlan, R., & Zhang, Z. (2014). Toward a unified theory of voice production and perception. *Loquens*, 1(1), e009. <https://doi.org/10.3989/loquens.2014.009>
- Kreiman, J., Lee, Y., Garellek, M., Samlan, R., & Gerratt, B. (2021). Validating a psychoacoustic model of voice quality. *Journal of the Acoustical Society of America*, 149, 457–465. <https://doi.org/10.1121/10.0003331>
- Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Wiley-Blackwell. <https://doi.org/10.1002/9781444395068>
- Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *Journal of the Acoustical Society of America*, 142(3), 1693–1706. <https://doi.org/10.1121/1.5003649>

- Kutlu, E., Tiv, M., Wulff, S., & Titone, D. (2022). Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cognitive Research: Principles and Implications*, 7, 7. <https://doi.org/10.1186/s41235-022-00354-0>
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. University of Chicago Press.
- Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, 60(1), 44–51. <https://doi.org/10.1037/h0044430>
- Laver, J. D. M. (1968). Voice quality and indexical information. *International Journal of Language & Communication Disorders*, 3(1), 43–54. <https://doi.org/10.3109/13682826809011440>
- Lawrence, D. (2015). Limited evidence for social priming in the perception of the BATH and STRUT vowels. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. University of Glasgow.
- Lee, K. E. (2016). *The perception of creaky voice: Does speaker gender affect our judgments?* [Master's thesis, University of Kentucky]. <https://doi.org/10.13023/ETD.2017.032>
- Lenth, R. (2023). emmeans: Estimated marginal means, aka least-squares means. R package version 1.8.8. <https://doi.org/10.32614/CRAN.package.emmeans>
- Leung, Y., Oates, J., Papp, V., & Chan, S.-P. (2022). Speaking Fundamental frequencies of adult speakers of Australian English and effects of sex, age, and geographical location. *Journal of Voice*, 36(3), 434.e1–434.e15. <https://doi.org/10.1016/j.jvoice.2020.06.014>
- Lévi-Strauss, C. (1962). *La pensée sauvage*. Plon.
- Li, A., Lai, W., & Kuang, J. (2023). Creaky voice identification in Mandarin: The effects of prosodic position, tone, pitch range and creak locality. *Journal of the Acoustical Society of America*, 154(1), 126–140. <https://doi.org/10.1121/10.0019941>
- Ligon, C., Rountrey, C., Rank, N. V., Hull, M., & Khidr, A. (2019). Perceived desirability of vocal fry among female speech communication disorders graduate students. *Journal of Voice*, 33(5), 805.e21–805.e35. <https://doi.org/10.1016/j.jvoice.2018.03.010>
- Lindeløv, J. K. (2019). Reaction time distributions: An interactive overview. Retrieved June 7, 2025 from <https://lindeloev.github.io/shiny-rt/>
- Lindvall-Östling, M., Deutschmann, M., & Steinvall, A. (2020). An exploratory study on linguistic gender stereotypes and their effects on perception. *Open Linguistics*, 6(1), 567–583. <https://doi.org/10.1515/opli-2020-0033>
- Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Routledge.
- Liu, X., & Xu, Y. (2011). What makes a female voice attractive? In W. Sum Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 1274–1277. City University of Hong Kong.
- Loakes, D., & Gregory, A. (2022). Voice quality in Australian English. *JASA Express Letter*, 2(8), 1–8. <https://doi.org/10.1121/10.0012994>

- Lortie, C. L., Thibeault, M., Guitton, M. J., & Tremblay, P. (2015). Effects of age on the amplitude, frequency and perceived quality of voice. *AGE*, 37(117). <https://doi.org/10.1007/s11357-015-9854-1>
- Mack, S., & Munson, B. (2012). The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the "gay lisp" stereotype. *Journal of Phonetics*, 40(1), 198–212. <https://doi.org/10.1016/j.wocn.2011.10.002>
- McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748. <https://doi.org/10.1038/264746a0>
- Melvin, S., & Clopper, C. G. (2015). Gender variation in creaky voice and fundamental frequency. In The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. University of Glasgow.
- Mendoza-Denton, N. (2011). The semiotic hitchhiker's guide to creaky voice: Circulation and gendered hardcore in a Chicana/o gang persona. *Journal of Linguistic Anthropology*, 21(2), 261–280. <https://doi.org/10.1111/j.1548-1395.2011.01110.x>
- Munson, B., Edwards, J., Schellinger, S., Beckman, M., & Meyer, M. (2010). Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana. *Clinical Linguistics & Phonetics*, 24(4–5), 245–260. <https://doi.org/10.3109/02699200903532524>
- Munson, B., Ryherd, K., & Kemper, S. (2017). Implicit and explicit gender priming in English lingual sibilant fricative perception. *Linguistics*, 55(5), 1073–1107. <https://doi.org/10.1515/ling-2017-0021>
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 62–85. <https://doi.org/10.1177/0261927X99018001005>
- Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. <https://doi.org/10.1016/j.jbef.2017.12.004>
- Pierrehumbert, J. (1995). Prosodic effects on glottal allophones. In O. Fujimura & M. Hirano (Eds.), *Vocal fold physiology 8: Voice quality control* (pp. 39–60). Singular Press.
- Pillot-Loiseau, C., Horgues, C., Scheuer, S., & Kamiyama, T. (2019). The evolution of creaky voice use in read speech by native-French and native-English speakers in tandem: A pilot study. *Anglophonia*, 27. <https://doi.org/10.4000/anglophonia.2005>
- Pittam, J. (1987). Listeners' evaluations of voice quality in Australian English speakers. *Language and Speech*, 30(2), 99–113. <https://doi.org/10.1177/002383098703000201>
- Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, 11(4), 478–504. <https://doi.org/10.1111/j.1467-9841.2007.00334.x>
- Podesva, R. J. (2013). Gender and the social meaning of non-modal phonation types. In C. Cathcart, I-H. Chen, G. Finley, S. Kang, C. S. Sandy, & E. Stickles (Eds.), *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 37(1), 427–448. <https://doi.org/10.3765/bls.v37i1.832>

- Podesva, R. J. (2018) The affective roots of gender patterns in the use of creaky voice [Invited presentation]. *Experimental and Theoretical Approaches to Prosody 4*. University of Massachusetts, Amherst. <https://www.youtube.com/watch?v=ZtqHpia7Iy8>
- Pratt, T. C. (2018). *Affective sociolinguistic style: An ethnography of embodied linguistic variation in an arts high school*. [Doctoral dissertation, Stanford University].
- Quenqua, D. (2012, February 27). They're, like, way ahead of the linguistic currrrve. *New York Times*. <https://www.nytimes.com/2012/02/28/science/young-women-often-trendsetters-in-vocal-patterns.html?searchResultPosition=1>
- Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29(4), 407–429. <https://doi.org/10.1006/jpho.2001.0145>
- Sebregts, K., Vriesendorp, H., Quené, H., & White, Y. (2023). Creaky voice in L2 English and L1 Dutch. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, 1841–1845. International Phonetic Association.
- Seyfarth, S. & Garellek, M. (2018). Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics*, 71, 425–450. <https://doi.org/10.1016/j.wocn.2018.09.001>
- Sicoli, M. (2010). Shifting voices with participant roles: Voice qualities and speech registers in Mesoamerica. *Language in Society*, 39(4), 521–553. <https://doi.org/10.1017/S0047404510000436>
- Sonderegger, M. (2023). *Regression modeling for linguistic data*. MIT Press.
- Sonderegger, M., & Sóskuthy, M. (2025). Advancements of phonetics in the 21st century: Quantitative data analysis. *Journal of Phonetics*, 111, 101415. <https://doi.org/10.1016/j.wocn.2025.101415>
- Squires, L. (2013). It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of Sociolinguistics*, 17(2), 200–237. <https://doi.org/10.1111/josl.12025>
- Staum Casasanto, L. (2010). What do listeners know about sociolinguistic variation? *University of Pennsylvania Working Papers in Linguistics*, 15(2), Article 6.
- Steinmetz, K. (2011, December 15). Get your creak on: Is “vocal fry” a female fad? *Time*. <https://healthland.time.com/2011/12/15/get-your-creak-on-is-vocal-fry-a-female-fad/>
- Stewart, C. F., Kling, I., & D'Agosto, A. (2024). Modal register, vocal fry, and uptalk: Identification and perceptual judgments of inexperienced listeners. *Journal of Voice*. Advance online publication. <https://doi.org/10.1016/j.jvoice.2024.02.028>
- Strand, E. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, 18(1), 86–100. <https://doi.org/10.1177/0261927X99018001006>
- Strand, E., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, 14–26. <https://doi.org/10.1515/9783110821895-003>
- Stuart-Smith, J. (1999). Voice quality in Glaswegian. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*, 2553–2556. Edward Arnold.

- Syrdal, A. K. (1996). Acoustic variability in spontaneous conversational speech of American English talkers. *Proceedings of the 4th International Conference on Spoken Language Processing*, 438–441. <https://doi.org/10.1109/ICSLP.1996.607148>
- Szakay, A. (2012). Voice quality as a marker of ethnicity in New Zealand: From acoustics to perception. *Journal of Sociolinguistics*, 16(3), 382–397. <https://doi.org/10.1111/j.1467-9841.2012.00537.x>
- Szakay, A., & Torgersen, E. (2015). An acoustic analysis of voice quality in London English: The effect of gender, ethnicity and f0. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–4. University of Glasgow.
- Szakay, A., & Torgersen, E. (2019). A re-analysis of f0 in ethnic varieties of London English using REAPER. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 1675–1678. Australasian Speech Science and Technology Association and International Phonetic Association.
- The Institute of Electrical and Electronics Engineers. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17(3), 225–246. <https://doi.org/10.1109/IEEESTD.1969.7405210>
- Urberg-Carlson, K., Munson, B., & Kaiser, E. (2008). Gradient measures of children’s speech production: Visual analogue scale and equal appearing interval scale measures of fricative goodness. *Journal of the Acoustical Society of America*, 125(4), 2529. <https://doi.org/10.1121/1.4783533>
- Uusitalo, T., Nyberg, L., Laukkanen, A., Waaramaa, T., & Rantala, L. (2024). Has the prevalence of creaky voice increased among Finnish university students from the 1990’s to the 2010’s? *Journal of Voice*, 38(3), 697–702. <https://doi.org/10.1016/j.jvoice.2021.12.006>
- Vaughn, C., & Kendall, T. (2019). Stylistically coherent variants: Cognitive representation of social meaning. *Revista de estudos da linguagem*, 27(4), 1787–1830. <https://doi.org/10.17851/2237-2083.0.0.1787-1830>
- Wade, L., Embick, D., & Tamminga, M. (2023). Dialect experience modulates cue reliance in sociolinguistic convergence. *Glossa Psycholinguistics*, 2(1), 1–30. <https://doi.org/10.5070/G6011187>
- Walker, A., & Hay, J. (2011). Congruence between “word age” and “voice age” facilitates lexical access. *Laboratory Phonology*, 2(1), 219–237. <https://doi.org/10.1515/LABPHON.2011.007>
- Walker, M., Szakay, A., & Cox, F. (2019). Can kiwis and koalas as cultural primes induce perceptual bias in Australian English-speaking listeners? *Laboratory Phonology*, 10(1), 1–29. <https://doi.org/10.5334/labphon.90>
- Weatherholtz, K., & Jaeger, T. (2016). Speech perception and generalization across talkers and accents. *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.95>
- Weber, M. M. (2017, May 10). Top five most annoying vocal habits. *Voice Empowerment*. <https://www.voiceempowerment.com/voice-empowerment-blog/2017/5/1/ten-most-annoying-vocal-habits-or-5>

- White, H., Penney, J., Gibson, A., Szakay, A., & Cox, F. (2024). Influence of pitch and speaker gender on perception of creaky voice. *Journal of Phonetics*, *102*, 1–15. <https://doi.org/10.1016/j.wocn.2023.101293>
- Wolk, L., Abdelli-Beruh, N., & Slavin, D. (2012). Habitual use of vocal fry in young adult female speakers. *Journal of Voice*, *26*(3), e111–e116. <https://doi.org/10.1016/j.jvoice.2011.04.007>
- Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception & Psychophysics*, *79*(7), 2064–2072. <https://doi.org/10.3758/s13414-017-1361-2>
- Wright, R., Mansfield, C., & Panfili, L. (2019). Voice quality types and uses in North American English. *Anglophonia*, *27*, 1952. <https://doi.org/10.4000/anglophonia.1952>
- Yu, A. C. L. (2022). Perceptual cue weighting is influenced by the listener's gender and subjective evaluations of the speaker: The case of English stop voicing. *Frontiers in Psychology*, *13*, 840291. <https://doi.org/10.3389/fpsyg.2022.840291>
- Yu, A. C. L., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics*, *5*(1), 131–150. <https://doi.org/10.1146/annurev-linguistics-011516-033815>
- Yuasa, I. (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>
- Zellou, G., Barreda, S., Lahrouchi, M., & Smiljanić, R. (2024). Learning a language with vowelless words. *Cognition*, *251*, 105909. <https://doi.org/10.1016/j.cognition.2024.105909>
- Zhang, Z. (2021). Contribution of laryngeal size to differences between male and female voice production. *Journal of the Acoustical Society of America*, *150*(6), 4511–4521. <https://doi.org/10.1121/10.0009033>
- Zhang, Z., & Kirby, J. (2020). The role of F_0 and phonation cues in Cantonese low tone perception. *Journal of the Acoustical Society of America*, *148*(1), EL40–EL45. <https://doi.org/10.1121/10.0001523>

