Humanities and Social Sciences Communications

Article in Press

https://doi.org/10.1057/s41599-026-07108-z

Social perception of creaky voice in Mandarin Chinese: everyone's gender matters

Received: 25 July 2025

Accepted: 16 March 2026

Cite this article as: Yao, Y., Li, M., Chang, C.B. Social perception of creaky voice in Mandarin Chinese: everyone's gender matters. *Humanit Soc Sci Commun* (2026). https://doi.org/10.1057/s41599-026-07108-z

Yao Yao, Meixian Li & Charles B. Chang

Social perception of creaky voice in Mandarin Chinese: Everyone's gender matters

Abstract

Recent years have witnessed increasing interest in the social meanings of non-modal voice qualities, but most existing studies focus on English, especially in the North American context. This paper reports a perceptual study of the social meanings of creaky voice in Mandarin Chinese for mainland Chinese listeners. The study used a large set of resynthesized stimuli including multiple talkers and pairs of utterances differing only in voice quality (creaky vs. modal). Sixty Mandarin listeners completed a social perception experiment in which they collectively evaluated 38 talkers (presented in creaky or modal voice quality) on four socio-demographic dimensions (age, gender, sexuality, education) and 19 traits related to personality (e.g., confident, genuine, pretentious) and communicative style (e.g., engaging). Results of a factor analysis and mixed-effects models indicated multiple effects of creaky voice on the perception of talker age, gender, and warmth; further, these effects interacted with both talker gender and listener gender, in ways that often differed from previously documented patterns for North American English. These findings shed light on the multifaceted indexicality of creaky voice in Mandarin and contribute to mounting evidence of crosslinguistic and crosscultural variation in the social meanings of non-modal voice qualities.

Keywords: gender, sociolinguistics, speech perception, social perception, creaky voice, voice quality, Mandarin Chinese, crosscultural variation

Introduction

In recent years, there has been increasing interest in the socio-indexical functions of creaky voice as a non-modal voice quality. Creaky voice differs from modal voice (i.e., the neutral mode of vocal fold vibration that all languages share) both physiologically and acoustically. Prototypical creaky voice is characterized by low-amplitude and irregular vocal fold vibration (Henton and Bladon, 1988) and an acoustic profile comprising low fundamental frequency (f_0), shorter glottal pulses (i.e., low open quotient), and a shallow spectral tilt (Klatt and Klatt, 1990). Other types of creaky voice, such as vocal fry, doubly- or multiply-pulsed voice, and aperiodic voice, have slightly different acoustic features but share the auditory quality of creakiness (Batliner et al., 1994; Gerratt and Kreiman, 2001; Hedelin and Huber, 1990; Keating et al., 2015; Klatt and Klatt, 1990; Redi and Shattuck-Hufnagel, 2001). In this paper, we focus on non-pathological, non-contrastive creaky voice (henceforth, "creak"), which is often observed to occur utterance-finally and to cooccur with low f_0 in natural speech (Ladefoged, 1982; Monsen and Engebretson, 1977).

The most widely documented social effects on creak production and perception are related to gender, although there is some evidence that the use of creak also varies by age, education, social class, and ethnicity (Esling, 1978; Gittelson et al., 2021; Loakes and Gregory, 2022; Podesva, 2011; Szakay, 2012; Trudgill, 1974). Earlier reports, mainly on British English dialects, described creak as predominantly a feature of men's speech (Esling, 1978; Henton and Bladon, 1988; Stuart-Smith, 1999), a pattern evident in more recent research as well (Gittelson et al., 2021; Brown and Sonderegger, 2025). The association of creak with men is understandable, given that creak often cooccurs with low f_0 and men's f_0 is typically much lower than women's. Crucially, this association leads to creak being linked to masculinity as well as other stereotypically masculine traits (e.g., authority, dominance, status), therefore contrasting with breathiness, which is often associated with femininity in Western cultures (Laver, 1968). Studies on other varieties of English reported converging findings, such as American and Australian English listeners perceiving creak more favorably in men's than women's speech (Pittam, 1987) and creak being vital to the construction of a "hardcore Chicano gangster" persona characterized by toughness in the U.S. (Mendoza-Denton, 2011).

Apart from gender, creak may also index low energy, social withdrawal, disengagement, and/or boredom (Laver, 1980), possibly due to its naturally low amplitude and frequent occurrence in phrase-final position with weakened prosodic prominence. Young speakers of North American English have been found to associate creak with "coolness", an affective style encompassing low vocal effort, chilled manners, emotional disengagement, and less institutionally-oriented stances (Greer, 2015; Greer and Winters, 2015; Pratt, 2018; Zimman, 2017). Similarly, creak can signal low arousal, lack of interest in the current topic, detachment from the current discourse, or interactional misalignment with the interlocutor (Gobl and Ní Chasaide, 2003; Grivičić and Nilep, 2004; Lee, 2015). In line with "low energy", creak has also been found to be a preferred voice quality for describing discomfort, pain, and misery crosslinguistically, including in Bengali, English, and Tzetzal (Mayan) (Brown and Levinson, 1978; Torres et al., 2020; Wilce, 1997).

Despite the iconic association between creak and masculinity, some studies have argued that the use of creak is rising among women, such that creak is becoming (or has already become) a female speech feature in English-speaking populations, especially in the U.S. (Yuasa, 2010). For example, two-thirds of young American women produced creak in read speech in one study (Wolk et al., 2012), while every American woman in a different sample aged 18–50 produced at least one creaky episode in spontaneous speech (Oliveira et al., 2016). The spread of creak among women is likely assisted by women's tendency to imitate or reinforce creak when conversing with a creaky interlocutor (Borrie and Delfino, 2017; White et al., 2023). Along these lines, several studies observed American women to be creakier than their male counterparts in different regions of the U.S., across different ethnic groups, and across a wide age range (Yuasa, 2010; Podesva, 2011; Melvin and Clopper, 2015). Women were more than twice as likely as men to produce creak in spontaneous speech (Yuasa, 2010; Podesva, 2011), whereas the gender gap was narrower in read speech (Melvin and Clopper, 2015). More recent research has presented conflicting findings, however. No gender difference was found in young Americans' overall creakiness in read speech in one study (Abdelli-Beruh et al., 2016), while a cross-dialect literature review found no conclusive evidence that American women are becoming creakier over time, pointing out limitations such as low speaker diversity (predominantly young urban college-educated females) and shallow time depth (no data prior to 2010) (Dallaston and Docherty, 2020).

Putting aside the debate regarding the relative frequency of creak in women's speech, one might wonder: why would women adopt the creaky speech style, given that neither of the core social meanings of creak—masculinity and coolness—is particularly feminine? One possible explanation is that women use creak to distance themselves from conventional stereotypes of women and instead approximate a socially dominant group, one that is masculine, professional, and authoritative. The trend for American women to speak with a low-pitched voice, especially in professional contexts, was noted several decades ago (Yuasa, 2010) and is reflected in the common use of creaky and low-pitched voice by female newscasters and journalists (Dilley et al., 1996; Nissen et al., 2020; Yau, 2020).

Nevertheless, even though American women's creak could be perceived as professional-sounding, it is not consistently well-received by listeners, especially those from different backgrounds. Although demographically similar listeners (i.e., college students at U.S. universities) perceived creaky young women positively as "professional, urban, educated, and upwardly mobile" (Yuasa, 2010), other studies with larger and more diverse listener samples found that, compared to modal counterparts, creaky young women were rated lower in intelligence, trustworthiness, competence, level of education, and employability (Anderson et al., 2014; Foulks, 2020; Kunnemann, 2017; Lee, 2016). In contrast, men's use of creak either improved listeners' impressions (e.g., by increasing perceived attractiveness and masculinity) (Greer and Winters, 2015; Lee, 2016) or resulted in a less negative impact compared to creak's effect on the perception of female talkers (Lee, 2016). To some extent,

negative perception of women's creak may reflect a listener's gender attitudes and ideologies. In general, male listeners may prefer a female voice signaling a small body size (i.e., high pitch, wide formant dispersion, breathy voice) (Xu et al., 2013; Ohala, 1983) whereas women's production of low-pitched creak goes against such a preference. Furthermore, because creak is linked with masculinity and authority, listeners may reject a woman's creak if they perceive her to be using an "accent" to which she is not entitled (Chao, 2017; Chao and Bursten, 2021).

Examining the existing literature on the social perception of creak, we note two critical limitations: (1) low linguistic and cultural diversity, and (2) the entanglement of voice quality with pitch. First, previous work is overwhelmingly based on users of English, mostly monolingual speakers of American and British varieties, although the use and functions of non-modal voice qualities are known to vary across language varieties (Esposito and Khan, 2020; Kreiman et al., 2010; Pennock-Speck, 2005; Tian et al., 2019), between monolingual and bilingual speakers (Cantor-Cutiva et al., 2023; Gibson and Summers, 2021; Kim, 2017), and within the language repertoire of bilingual individuals (Agathe and Claire, 2013; Brown and Sonderegger, 2024; Gibson and Summers, 2021; Zhu et al., 2022), with some exceptions (Sebregts et al., 2023, 2024; Brown and Sonderegger, 2025). Relatedly, the majority of previous research was situated in Western cultural contexts, which is problematic given the profound crosscultural differences observed in the construction of women's voice and affective speech (Holliday et al., 2023; van Bezooijen, 1995; Starr, 2015; Yuasa, 2010). Second, the recurrent confounding of creak with low pitch—found throughout the literature, especially when natural speech stimuli were used—means we actually have little information about the social meanings of creak specifically. As mentioned earlier, creak has a natural tendency to occur with a low-pitched voice, and these qualities are intertwined in perception as well, with low pitch promoting the perception of creakiness and vice versa (Davidson, 2018, 2020; Kuang and Liberman, 2015, 2016, 2018). However, creak and low pitch are not inseparable: they have independent articulatory mechanisms, can each be linguistically contrastive on their own (Gordon and Ladefoged, 2001), and may contribute to social perception separately (Parker and Borrie, 2017). Focusing on English, Parker and Borrie (2017) reported that creak had positive effects on likability and intelligibility for high-pitched female speakers but unfavorable effects for low-pitched female speakers, although these findings were based on a small sample of speakers ($N = 8$), limiting their generalizability. Some previous research used (re)synthesized stimuli for better stimulus control, but usually included only a few talker voices and/or utterances (Gobl et al., 2002; Gobl and Ní Chasaide, 2003; Xu and Lee, 2018).

The current study addressed both limitations by examining the social meanings of creak in a less-studied linguistic and sociocultural context—Mandarin Chinese as used by speakers from mainland China—using a sizable set of resynthesized stimuli that were carefully controlled to differ only in terms of voice quality (creaky vs. modal). Although a full-scale typological investigation of the linguistic and social functions of creak is beyond the scope of the current study, Mandarin is well-positioned to provide insight into possible crosslinguistic variation in the social perception of creak because of a central feature of Chinese languages: lexical tones. Tones create new word meanings with systematic prosodic variations, primarily in pitch pattern but secondarily in other features such as voice quality. In Standard Mandarin, there are four main tones (Tones 1–4), and low-register tones (e.g., Tone 3) are associated with creak in both perception and production (Belotel-Grenié and Grenié, 2004; Huang, 2024a; Kuang, 2017, 2018). This association is also found in other Chinese varieties, such as Tianjin Mandarin and Cantonese (Davison, 1991; Yu and Lam, 2014). Crucially, the association of creak with tone means that creak serves a different linguistic function in Mandarin as compared to English. At the same time, factors such as (change in) a cue's use may be associated with its social meanings across languages (see, e.g., de Leeuw and Chang, 2024; Nodari et al., 2019; Sankoff, 2004). Given that creak is used differently in Mandarin

than English, could creak in Mandarin therefore show a different pattern of social perception from English as well?

Relatively little is known about the social perception of creak by Mandarin listeners. In one study presenting different versions of a short utterance synthesized with different pitch ranges and voice qualities, Mandarin listeners rated creaky voices as less attractive than breathy and modal voices (Xu and Lee, 2018). In another study with naturally produced stimuli from one female and one male talker, Mandarin listeners perceived the talkers as less good-looking and more gender-neutral when in their creaky guise than in their modal guise and assigned a lower value to an object when described by a creaky talker than by a modal talker (Li and Mok, 2023). An overall listener bias against creak was also found in a similar study with one natural and one resynthesized voice (Li et al., 2023). None of these studies reported any effects of talker gender, in line with previous work on creak in Taiwan Mandarin speakers (Kuang, 2018).

Building upon previous research, the current study sought to examine the social implications of Mandarin creak related to the talker's demographic background, personality, and communicative style. More specifically, we focused on whether and how these social meanings would (1) interact with the talker's gender and/or the listener's gender, and (2) diverge from previous findings on English. Notably, the current study included a large set of talker voices with multiple utterances per voice and strictly controlled f_0 and other acoustic aspects of the stimuli, allowing social perceptions of talkers' creaky guise to be interpreted more straightforwardly as due to the presence of creak specifically and as generalizable across different voices.

Methods

Participants

Three groups of participants were recruited: talkers, raters, and listeners. All participants were self-identified native Mandarin speakers between ages 18 and 35 years, born and raised in mainland China with no history of speech or language disorders. Additionally, talkers and listeners self-identified as either (cisgender) female or male. The talkers ($N = 40$; 20F, 20M; $M_{age} = 24.9$, $SD = 3.4$) recorded the raw speech materials that were used as the basis for the resynthesized perceptual stimuli. The raters ($N = 22$; 11F, 11M; age 18–29), including a trained speech-language pathologist (female; age 24) as the expert rater, provided goodness judgments on the resynthesis output for selecting the stimuli used in the social perception experiment. Lastly, the listeners ($N = 60$; 33F, 27M; $M_{age} = 24.0$, $SD = 3.0$) completed the social perception experiment.

Since Mandarin speakers in this age group have typically learned English as a foreign language, we administered an online Language History Questionnaire (LHQ; version 3) (Li et al., 2014) to the listeners before the social perception experiment. Data from the LHQ indicated that almost all listeners ($N = 59$) reported knowing English as a foreign language, with an average age of acquisition of 8.8 ($SD = 4.7$). Their self-reported English proficiency level (averaged across ratings of listening, speaking, reading, and writing abilities on a seven-point scale) was 3.8 ($SD = 1.1$). Five listeners reported prior residence in an English-speaking country: the United Kingdom (6–14 months) or the U.S. (6–30 months). Three other listeners reported prior residence (no more than 18 months) in other countries where they might be exposed to foreign languages (France, Singapore, Spain, Sweden). These periods of overseas residence were typically for study exchanges and internships starting no earlier than secondary school.

Apart from gender, female and male listeners were similar on most socio-demographic dimensions, but we observed two between-gender differences pertaining to sexuality. First, although relatively few listeners reported being gay or bisexual overall ($N = 8$; 7F, 1M), female listeners

were much more likely to identify as such than male listeners were. In all, 51 listeners (26F, 25M) self-reported as straight, five as bisexual (4F, 1M), and three as gay (3F, 0M); one male did not report his sexuality. Second, considerably more female listeners than male listeners reported knowing at least some gay individuals (females: 9.1% "know many", 81.8% "know some", 9.1% "know none"; males: 3.7% "know many", 66.7% "know some", 29.6% "know none").

Stimuli

The stimuli in the social perception experiment were constructed in a five-step process. First, natural read speech was recorded by the talkers. Each talker read 120 emotion-neutral Mandarin sentences, which covered all the segments and tones in Mandarin and averaged ten characters (syllables) in length (see *Data availability* for a link to the Open Science Framework (OSF) repository providing a full list of the sentences and other study materials). Due to COVID restrictions, the recordings were made online using the talker's own equipment at a quiet location of the talker's choice. Second, the recordings were segmented into consecutive voiced portions (e.g., vowels, sonorant consonants) and voiceless portions (e.g., voiceless consonants).

Third, a resynthesis procedure was applied to each utterance to create two versions: creaky and modal. Crucially, the procedure resynthesized the voiced portions with a Klatt synthesizer (which tracked and recreated the f_0, formants, and intensity profiles of the original audio) and then concatenated the resynthesized voiced portions with the adjacent (voiceless) portions with smoothed boundaries. Creakiness was created by adding double pulsing points in the resynthesis script, resulting in doubly-pulsed creak throughout the voiced portions. The procedures for resynthesizing creaky and modal utterances were otherwise identical. Thus, the two versions of each utterance had identical f_0 contours, formants, and intensity profiles, but one was consistently creaky from beginning to end while the other was consistently modal. Figure 1 shows example waveforms of the creaky and modal versions of the same vowel.

Fourth, the resynthesized utterances were screened for comprehensibility (i.e., listeners' subjective impression of understandability) and naturalness. After conducting an initial round of quality control by excluding obviously flawed tokens (e.g., unintelligible, substantially distorted, presence of background noise), we then passed the remaining tokens ($N = 3,768$) to raters to evaluate comprehensibility and naturalness separately on a five-point scale. Each token was evaluated by at least two raters including the expert rater. Based on the raters' evaluations, we excluded two talkers (1F, 1M), whose resynthesis quality was exceptionally low due to their fast speech rates, and selected the best 6–11 utterance pairs (creaky vs. modal) for the remaining talkers. Thus, the final set of stimuli consisted of 38 talkers (19F, 19M) and 328 utterance pairs, which showed overall high comprehensibility ($M = 4.2$, $SD = 0.5$) and intermediate to high naturalness ($M = 3.2$, $SD = 0.5$). Modal utterances were slightly more comprehensible than their creaky counterparts (modal: $M = 4.2$, $SD = 0.5$, creaky: $M = 4.1$, $SD = 0.5$; paired $t(327) = 3.190, p = 0.002$); however, the two sets of utterances did not differ significantly in naturalness (modal: $M = 3.2$, $SD = 0.6$, creaky: $M = 3.2$, $SD = 0.5$; paired $t(327) = 0.609, p = 0.543$).

Finally, utterance collections were created by concatenating the tokens selected for each talker into one sound file per voice quality. Thus, each talker was represented by a collection of creaky utterances and, separately, by a matching collection of modal utterances. These collections, averaging 8.6 utterances in length ($SD = 1.2$), were used as the stimuli in the social perception experiment.

To confirm that the creaky stimuli were indeed acoustically creakier than the modal stimuli, we conducted acoustic analyses of their spectral tilt (specifically, the corrected amplitude difference between the first and second harmonics of the spectrum, $H_1{}^*{-}H_2{}^*$) and harmonic-to-noise ratio (HNR) in the 0–3,500 Hz spectrum in VoiceSauce (Shue et al., 2011). The results of these analyses

Table 1: Distribution of talkers across high- and low-f_0 groups, by gender, with data about their f_0 characteristics.

Gender	f_0 group	N	Mean f_0 (Hz)	SD	Range
F	high	9	235.3	9.2	(222.5, 250.3)
F	low	10	202.5	9.5	(184.0, 215.4)
M	high	12	126.3	4.5	(121.0, 135.1)
M	low	7	106.2	8.2	(91.0, 117.2)

showed that, compared to the corresponding modal stimuli, the creaky stimuli had significantly lower values for both $H_1{}^*{-}H_2{}^*$ (creaky: $M = -7.2$ dB, $SD = 6.5$, modal: $M = 4.5$ dB, $SD = 3.4$; paired $t(327) = 45.335, p < 0.001$) and HNR (creaky: $M = 29.2$ dB, $SD = 4.0$, modal: $M = 43.6$ dB, $SD = 5.2$; paired $t(327) = 84.871, p < 0.001$). These differences were consistent with previously documented acoustic differences between modal and creaky speech (Keating et al., 2015; Klatt and Klatt, 1990).

Additionally, to check whether the stimuli were representative of natural talker variation in f_0, we measured each talker's average f_0 across their selected stimuli and, further, classified the talkers as high- or low-f_0 talkers using by-gender means as the dividing lines. The results of the f_0 analyses suggested that the talkers had normal f_0 levels for their respective gender (female: $M = 218.0$ Hz, $SD = 19.1$, male: $M = 118.9$ Hz, $SD = 11.6$). Moreover, they were roughly evenly distributed across the different f_0 groups (see Table 1), meaning that the stimuli were not obviously biased toward high- or low-f_0 talkers.

Procedure

The study procedure for listener participants consisted of an initial language background questionnaire (described above) and three other tasks, in the following order: (1) a reading task, (2) a social perception experiment, and (3) a questionnaire about gender and sexuality. The goal of the reading task was to collect data on the listeners' baseline f_0 level and voice quality in their own production. Listeners were tested in person in Hong Kong and completed all tasks in one session while seated before a desktop computer and wearing a headset. The whole session took about one hour and was conducted entirely in Mandarin. The items used in these tasks are viewable on OSF (see *Data availability*).

In the reading task, listeners provided samples of their own speech by reading aloud five short Mandarin sentences. They were instructed to read the sentences at a comfortable and natural volume and pace.

In the social perception experiment, listeners were told that they would hear a series of voices created by speech engineers and provide impressionistic evaluations of the imaginary talkers. More specifically, they were told that the engineered voices would be used in virtual talk shows in the future, and that their impressionistic evaluations would therefore help in matching the voices with virtual characters later. Given that the resynthesized speech stimuli were intermediate to high on perceived naturalness (see *Stimuli*), these instructions provided a plausible context in which some speech samples could sound not fully natural. To prevent the experiment from being too long, during the experiment, each listener was presented with the stimuli of a subset of 14 talkers (balanced across talker genders and voice qualities), blocked by talker. Each listener heard only one voice quality (creaky or modal) of a given talker, and each talker was evaluated by at least 10 listeners per voice quality. For each talker, listeners first made forced-choice classifications of four demographic properties: age ("16 and under", "17–20", "21–25", "26–30", "31–35", "36–40"),

education level ("less than high school", "high school", "college", "postgraduate"), gender ("male", "female"), and sexuality ("gay/homosexual", "straight/heterosexual", "decline to answer"). For each categorical classification of talker gender and sexuality, listeners also rated their confidence on a five-point Likert scale (5 = most confident). Listeners rated the typicality of the talker's voice for the perceived gender on a five-point Likert scale as well (5 = most typical). Finally, listeners rated the talker on 19 impressionistic traits related to personality and communicative style (e.g., confident, genuine, pretentious, engaging), each on a five-point Likert scale (5 = highest degree of the given trait). These rating items were based on previous studies of the social meanings of creak. Additional rating items not analyzed in this study concerned the talker's height and body size and how the talker's voice made the listener feel. Throughout the rating process, listeners could choose to hear the auditory stimuli repeatedly.

The post-experiment questionnaire included 11 items about the listener's attitudes concerning gender equality and homosexuality, their own sexual orientation, and their social network. The gender attitude items consisted of seven statements oriented toward gender equality (e.g., "Both men and women can have a gentle temperament"; $N = 2$) or conventional gender stereotypes (i.e., gender inequality, such as "Men manage external affairs, while women manage domestic affairs"; $N = 5$). For each item, listeners were asked to indicate how much they (dis)agreed with the statement on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). The sexuality attitude items consisted of two statements oriented toward either tolerance ("I believe that both homosexuality and heterosexuality should be accepted by society") or intolerance of homosexuality ("Personally, I don't particularly like homosexuality"), which listeners could indicate (dis)agreement with on the same Likert scale. Finally, listeners could indicate their own sexual orientation from among four choices ("straight/heterosexual", "gay/homosexual", "bisexual", "decline to answer") and the approximate number of gay individuals they knew from among three choices ("none", "some", "many"). The latter four items pertaining to sexuality were all optional.

Statistical analysis

To quantify listeners' social attitudes, we calculated two attitudinal scores per listener, one concerning gender and the other concerning sexuality, on the basis of the questionnaire data. Two listeners did not respond to the sexuality attitude items and thus did not receive sexuality attitude scores. The gender attitude score was averaged across the listener's responses to all gender attitude items, with responses to equality-oriented statements coded according to the original scale and responses to inequality-oriented statements coded in reverse (i.e., 5 = strongly disagree, 1 = strongly agree). The sexuality attitude score was averaged between the two sexuality attitude items, with responses on the intolerance-oriented statement coded in reverse as above. In this manner, higher gender attitude scores and higher sexuality attitude scores reflected an overall more positive attitude toward gender equality and toward homosexuality, respectively.

The perceptual dataset consisted of talker evaluations gathered across 840 trials (60 listeners × 14 talker guises per listener), which were analyzed statistically in a series of mixed-effects models. The models were built to examine the effects of voice quality, talker gender, and listener gender on talker evaluations while controlling for random fluctuations by talker and by listener. Both talker gender and listener gender were sum-coded (TalkerGender: F = 1, M = −1; ListenerGender: F = 1, M = −1), while voice quality was treatment-coded with modal as the reference level (VQuality: modal/MOD = 0, creaky/CRK = 1). Thus, a simple effect of voice quality indicates an overall difference in the outcome variable between modal and creaky conditions when talker gender and listener gender are controlled at average levels. Classifications of talker demographics were coded numerically and modeled separately. Age and education classifications were coded ordinally as 1–6

and 1–4, respectively (starting with 1 = "16 and under" / "less than high school"), while gender and sexuality classifications, after removing null responses, were coded as binary (0 = "female", 1 = "male"; 0 = "straight", 1 = "gay"). Ratings of the 19 impressionistic traits first underwent a factor analysis to reduce dimensionality, and then the principal factors were modeled separately.

All statistical analyses were conducted in R (R Development Core Team, 2023). The models were initially built with maximal fixed- and random-effect structures using the *lmerTest* package (version 3.1.3) (Kuznetsova et al., 2017) and then underwent backward elimination to remove obviously non-significant fixed predictors ($p > 0.1$) and ensure model convergence. For brevity, we only cite critical coefficients from the final models in the paper; however, the full report of model-building is available on OSF (see *Data availability*), along with full model outputs and datasets. To interpret significant interactions, post-hoc pairwise comparisons were conducted with the *emmeans* package (Lenth et al., 2025) with Bonferroni adjustment.

Results

Listeners' speech and social attitudes

Based on their speech samples, listeners were not unusual in their own f_0 level or creakiness. Overall, f_0 values for the listeners were very similar to those of the talkers matching their gender (female listeners: $M = 214.8$ Hz, $SD = 16.3$, female talkers: $M = 218.0$ Hz, $SD = 19.1$; male listeners: $M = 125.4$ Hz, $SD = 14.1$, male talkers: $M = 118.9$ Hz, $SD = 11.6$). No listeners produced audible creak. Thus, our sample of Mandarin listeners was unlikely to be influenced in their perceptual judgments by idiosyncratic aspects of their own speech.

Listeners' attitudes concerning gender and sexuality differed significantly between women and men (see Figure 2), but not between self-reported heterosexual and non-heterosexual participants. Female listeners showed significantly higher gender attitude scores, reflecting a greater orientation toward gender equality, than male listeners (females: $M = 4.1$, $SD = 0.7$, males: $M = 3.2$, $SD = 0.6$; Welch-corrected two-tailed $t(58.0) = 5.481, p < 0.001$). Female listeners also showed significantly higher sexuality attitude scores, reflecting greater acceptance of homosexuality, than male listeners (females: $M = 4.3$, $SD = 0.6$, males: $M = 2.8$, $SD = 1.1$; Welch-corrected two-tailed $t(35.5) = 6.243, p < 0.001$). However, no significant difference was found between the heterosexual and non-heterosexual listeners, in either gender attitude scores (heterosexual: $M = 4.0$, non-heterosexual: $M = 3.6$; Welch-corrected two-tailed $t(9.2) = 1.034, p = 0.328$) or sexuality attitude scores (heterosexual: $M = 4.1$, non-heterosexual: $M = 3.5$; Welch-corrected two-tailed $t(12.7) = 1.667, p = 0.120$), although these results are limited by the small number of non-heterosexual listeners ($N = 9$).

Perception of talker demographics

Listeners were overall highly accurate in identifying a talker's gender ($M_{acc} = 96.5\%$). Given the high accuracy rate, unsurprisingly, the mixed-effects model of the likelihood of accurate gender identification showed no significant effects of voice quality, talker gender, or listener gender (see *Data availability* for a link to full model outputs). Gender identification errors, albeit rare ($N = 29$, out of 840 responses), were most likely to occur with female listeners hearing female talkers' creaky guise ($N = 14$). Furthermore, they were distributed among more than one-third of listeners ($N = 21$) and almost one-fourth of talkers ($N = 9$), suggesting that the errors were not caused by one or two individuals with idiosyncratic speaking or listening styles.

More nuanced patterns of talker gender perception appeared in ratings of gender identification

confidence and of voice gender typicality (see Figure 3 and Figure 4). The two ratings were positively correlated ($r(838) = 0.68, p < 0.001$), and both ratings were significantly higher when the gender identification was correct than incorrect (confidence: Welch-corrected two-tailed $t(28.7) = 6.850, p < 0.001$; typicality: Welch-corrected two-tailed $t(29.1) = 5.990, p < 0.001$). Mixed-effects models revealed a significant three-way interaction among voice quality, talker gender, and listener gender for both ratings (confidence: $\beta = -0.088, p = 0.039$; typicality: $\beta = -0.191, p < 0.001$). Post-hoc pairwise comparisons of estimated marginal means showed that female listeners gave significantly higher confidence and typicality ratings for female modal than female creaky talkers (confidence: $est. = 0.261, p = 0.028$; typicality: $est. = 0.399, p = 0.003$), but significantly lower confidence and typicality ratings for female creaky than male creaky talkers (confidence: $est. = -0.478, p < 0.001$; typicality: $est. = -0.569, p = 0.002$). In addition, female listeners gave female creaky talkers significantly lower typicality ratings than male listeners did ($est. = -0.608, p = 0.002$). Taken together, these results corroborated the error patterns in suggesting that female listeners hearing female creaky talkers were the most prone to perceptual confusion about talker gender. Since talker gender was one of the critical independent variables, trials where the talker's gender was not correctly identified were removed from subsequent analysis, leaving 811 trials in the dataset.

In terms of sexual orientation, talkers were perceived much more often as straight (67% of the time; 540 of 811 responses) than as gay (12% of the time). The substantial number of null sexual orientation responses ($N = 170$) came mainly from six listeners who mostly declined to identify sexuality. Excluding these null responses, the mixed-effects model of sexuality identification found a significant effect of listener gender ($\beta = 0.621, p = 0.021$) and significant interactions of talker gender and listener gender ($\beta = -0.510, p = 0.029$) and of voice quality, talker gender, and listener gender ($\beta = 0.682, p = 0.046$). Post-hoc pairwise comparisons showed that the effects were mainly driven by female listeners being more likely to judge male modal talkers as gay than male listeners were ($est. = 2.262, p = 0.013$). The mixed-effects model of sexuality identification confidence showed a significant effect of identification response ($\beta = -0.626, p < 0.001$) and a significant interaction between identification response and talker gender ($\beta = -0.432, p < 0.001$). Compared to identifying female talkers as gay, listeners were more confident about identifying female talkers as straight ($est. = 1.058, p < 0.001$) and male talkers as gay ($est. = 1.029, p < 0.001$). Overall, these results suggested that, compared to male listeners, female listeners more readily perceived male talkers as gay; however, both female and male listeners were less confident about identifying female talkers as gay.

As for age and education level, listeners' estimations were overall quite close to the talkers' actual age and education level ($M_{age} = 24.9$ yr; median education level: college), probably because they came from the same demographic group. In almost half of age estimation responses ($N = 391$), talkers were perceived to be 21–25, and the vast majority of responses ($N = 711$) fell in the range of 17–30. Correspondingly, talkers were most often judged to be college-educated ($N = 405$). Estimations of age and education level were positively correlated (Spearman's rank correlation $\rho(811) = 0.65, p < 0.001$). Despite the lack of wide variation in age estimation, the mixed-effects model of estimated age range (coded ordinally as 1–6, starting with 1 = "16 and under") showed a marginal interaction of voice quality and talker gender ($\beta = -0.133, p = 0.072$). Post-hoc comparisons found that male creaky talkers sounded older than both female creaky talkers ($est. = 0.428, p = 0.027$) and female modal talkers ($est. = 0.399, p = 0.022$). Male modal talkers tended to be perceived as younger than male creaky talkers but older than female modal talkers; however, neither trend was significant (male modal vs. male creaky: $est. = -0.237, p = 0.149$, male modal vs. female modal: $est. = 0.162, p = 1.000$). In fact, the talkers' actual ages significantly differed between genders, female talkers being about two years younger on average (females: $M_{age} = 23.9$, $SD = 2.4$, males: $M_{age} = 26.1$, $SD = 3.9$; Welch-corrected two-tailed

$t(30.1) = -2.073, p = 0.047$). Together with the model results, these findings suggested that the small age gap between genders was more reliably perceived when male talkers were creaky. The model of estimated education level showed no significant effects of voice quality, talker gender, or listener gender.

Perception of talker personality and communicative style

Listeners' ratings of 19 impressionistic traits ($N = 15,409$; 811 trials $\times$ 19 traits) were entered into a latent factor analysis to reduce the dimensionality of the data. Based on recommendations from the parallel analysis and scree test (Sakaluk and Short, 2017), the factor analysis yielded three latent factors, which we renamed, respectively, COMPETENCE, LIKABILITY, and WARMTH. Table 2 shows the loading of each rated trait on each factor; a higher value indicates a larger portion of variance explained by that trait. As shown in Table 2, COMPETENCE was positively loaded by traits associated with competence (e.g., diligent, professional) and negatively loaded by traits associated with incompetence (e.g., hesitant, lazy). On the other hand, WARMTH was positively loaded by traits typically associated with warmth and solidarity (e.g., genuine, friendly) and negatively loaded by traits associated with hostility or distance (e.g., aggressive, pretentious). Thus, we interpret higher scores on COMPETENCE and WARMTH as indicating greater perceived competence and warmth, respectively. The factor LIKABILITY was positively loaded by traits related to both competence and warmth, and is thus interpreted as a general indicator of talker likability. Three scores of talker impression, one per factor, were calculated for each perceptual trial and used in subsequent statistical analyses. The three factors were all centered around zero (COMPETENCE range: $(-3.309, 3.692)$; LIKABILITY range: $(-3.027, 3.245)$; WARMTH range: $(-3.933, 3.227)$) and collectively explained 44% of the variance in the impressionistic ratings.

Mixed-effects models found effects of voice quality on some, but not all, factor scores. For COMPETENCE, there was a significant effect of talker gender ($\beta = 0.228, p = 0.002$) and a significant interaction between talker gender and listener gender ($\beta = 0.087, p = 0.049$), but no effect of voice quality (see Figure 5). Post-hoc comparisons showed that female listeners gave significantly higher COMPETENCE scores to female than male talkers ($est. = 0.630, p = 0.001$), but male listeners did not ($est. = 0.281, p = 0.597$). For LIKABILITY, there were no significant effects of voice quality, talker gender, or listener gender (see Figure 6). As for WARMTH, there was a marginal effect of talker gender ($\beta = -0.162, p = 0.077$) as well as significant two-way interactions between voice quality and talker gender ($\beta = 0.267, p = 0.006$) and between talker gender and listener gender ($\beta = 0.167, p = 0.002$) and a significant three-way interaction between voice quality, talker gender, and listener gender ($\beta = -0.206, p = 0.006$) (see Figure 7). Post-hoc comparisons showed significant differences related to voice quality and talker gender only for male listeners. To further examine the interactions, we refitted the model of WARMTH scores on trials from male listeners only, and again found an effect of talker gender ($\beta = -0.309, p = 0.002$) and an interaction of voice quality and talker gender ($\beta = 0.458, p < 0.001$). Post-hoc comparisons showed that creak significantly lowered male listeners' WARMTH scores for male talkers ($est. = -0.576, p = 0.009$); in contrast, it increased their WARMTH scores for female talkers (creaky: $M = 0.055$, 95% CI [−0.272, 0.381], modal: $M = -0.286$ [−0.613, 0.041]), although this latter trend was not significant ($est. = 0.340, p = 0.303$).

Post-hoc analyses exploring pitch effects

Although the social perception experiment isolated the effect of voice quality by acoustically controlling f_0 in the stimulus materials, we further disentangled voice quality effects from possible pitch effects through a series of post-hoc statistical analyses. In particular, we built augmented

Table 2: Factor loadings of rated traits on three factors: COMPETENCE, LIKABILITY, and WARMTH. Shading indicates no loading.

Trait	COMPETENCE	LIKABILITY	WARMTH
confident	0.59	0.36	
diligent	0.60	0.27	0.21
formal	0.67		
professional	0.68	0.21	
intelligent	0.57	0.38	
authoritative	0.60		
lazy	−0.50		
charismatic	0.35	0.58	
approachable		0.80	0.36
friendly		0.82	0.35
convincing	0.42	0.51	
engaging	0.33	0.62	
aggressive	0.33		−0.62
genuine	0.24		0.59
gentle			0.61
pretentious			−0.42
hesitant	−0.44		
fashionable	0.25	0.40	
casual	−0.44		

mixed-effects models (of perceived age, gender identification confidence, voice gender typicality, and WARMTH) containing an additional fixed effect for talkers' overall pitch level (i.e., f_0 group; see Table 1), which was allowed to interact fully with the other fixed effects. In the augmented model of perceived age, f_0 group had no significant effect and did not interact with other fixed effects. In the augmented models of gender identification confidence and voice gender typicality, however, f_0 group participated in a significant four-way interaction with voice quality, talker gender, and listener gender (confidence: $\beta = -0.196, p = 0.012$; typicality: $\beta = -0.325, p = 0.003$). Post-hoc pairwise comparisons showed that creak significantly affected female listeners' perception of low-f_0 females, leading to lower gender identification confidence ($est. = -0.388, p = 0.001$) and lower voice gender typicality ($est. = -0.527, p = 0.019$); however, it did not have this effect on female listeners' perception of high-f_0 females or either male f_0 group, nor on male listeners' perception of any f_0 group. In the augmented model of WARMTH, f_0 group participated in a marginal three-way interaction with voice quality and talker gender ($\beta = -0.299, p = 0.056$). Post-hoc pairwise comparisons showed that the effect of creak mainly targeted male listeners hearing high-f_0 male talkers, leading to lower WARMTH scores for creaky than modal talkers ($est. = -0.769, p = 0.028$).

In short, effects of creak on social perception showed a complex interaction with natural f_0 variation across talkers. On the one hand, for gender-rated variables, the effect of creak targeted female listeners' perception of female talkers with below-average f_0, echoing Parker and Borrie's (2017) findings. On the other hand, for a personality-related variable (WARMTH), it targeted male listeners' perception of male talkers with above-average f_0.

Discussion

In this study, we examined possible gender effects on the social perception of Mandarin creak, finding effects of both talker gender and listener gender. Given that these findings contrast with previous research on Mandarin (Xu and Lee, 2018; Li et al., 2023; Li and Mok, 2023),which reported no gender effects, and previous research on English (Pittam, 1987; Greer and Winters, 2015; Chao and Bursten, 2021; Anderson et al., 2014; Foulks, 2020; Lee, 2016), what underlies the different patterns of social perception of Mandarin creak in the current study? We believe there are three factors at play in the current results: (1) perceptual relationships of creak with pitch, (2) gendered perception of creak in Mandarin, and (3) crosslinguistic variation in the social perception of voice quality.

First, creak tends to cooccur with low pitch in natural speech (Kuang, 2017, 2018), and the two properties are also correlated in perception (Davidson, 2018; Kuang and Liberman, 2018; Davidson, 2020; Huang, 2024a; Yang, 2011; Yu and Lam, 2014), leading to their confluence in most research on creak perception. In one of the few studies that systematically disentangled voice quality and pitch, vocal fry affected the perceived intelligence and likability of young American women unfavorably when pitch was low, but favorably when pitch was high (Parker and Borrie, 2017). Our results suggest that, acoustically controlling for pitch, doubly-pulsed creak can significantly affect Mandarin listeners' perception of talker gender (less female-sounding), age (older), and warmth (less warm). The directions of the effects are generally compatible with the potential effects of creak lowering perceived pitch; crucially, however, the effects were manifested via interactions with talker gender and listener gender (which we discuss further below). Recall, for example, that our post-hoc analyses showed effects of creak targeting the gender identification of low-f_0 female talkers, for female listeners primarily. This pattern could be due to low-pitched female talkers being perceived with even lower pitch when creaky and, hence, becoming gender-ambiguous to listeners, contrasting with previous work on English (Davidson, 2020). Alternatively, creak may not have affected the perceived pitch, but instead the perceived gender directly; in this case, too, the effect of creak would be most obvious for low-pitched females because this is where it would lead to gender ambiguity. The absence of this effect for low-f_0 male talkers may be related to a weak perceptual distinction between creaky and modal voices in low-pitched male speech, where listeners may "hallucinate" creak even if the talker is not creaky (Davidson, 2018). Thus, at least some of our findings may be due to perceptual relationships between creak and pitch.

Second, whereas previous studies reported no gender effects on the perception and production of Mandarin creak (Kuang, 2018; Li and Lai, 2023), the current results reveal several gendered patterns. Importantly, the gender effects involved not only talkers' gender, but also listeners' gender, which was rarely examined in previous research. The two most noteworthy effects were in (1) female listeners' higher confusion in gender identification for creaky than modal female talkers, and (2) male listeners' lower warmth scores for creaky than modal male talkers, contrasting with their tendency toward higher warmth scores for creaky than modal female talkers. Regarding the first gender effect, although the directionality and locus of this effect on female talkers are both expected (as discussed above), the manifestation of the effect only among female listeners is unexpected in light of the literature, which reported no effect of listener gender on perception of talker gender (Brown et al., 2021; Hancock and Pool, 2017). We hypothesize that this pattern reflects differential use of creaky voice cues in women's speech by female and male Mandarin listeners: for female listeners, women's creak negatively impacts their identifiability as women, possibly via the lowering of perceived pitch, while male listeners may perceive women's creak as a distinct feminine speech style—a hypothesis we return to below. Regarding the second gender effect, male listeners' perception of lower warmth in creaky male talkers may be related to the effect of creak on perceived age. Recall that creak

tended to make male talkers sound older. Because male talkers were, coincidentally, marginally older than the male listeners (talkers: $M = 26.1$, $SD = 3.9$, listeners: $M = 23.8$, $SD = 3.9$; Welch-corrected two-tailed $t(39.0) = 1.980, p = 0.055$), creak may have made male talkers sound considerably more mature—and, thereby, less friendly or approachable—to younger male listeners. As for male listeners' tendency to perceive higher warmth in creaky than modal female talkers, in light of previous research suggesting that creak can be used to portray female sexuality in Chinese (Callier, 2010), it is possible that creak increases the sex appeal of female talkers for straight men; we hypothesize that such a "sexy speech" register could provide a mechanism through which creak increased the perceived warmth of female talkers for our predominantly straight male listener sample. However, this account needs to be validated in future research, as the above trend was not significant and the questionnaire item most closely related to sex appeal ("Does the speaker sound charismatic/attractive?") was not a direct contributor to warmth scores.

Finally, we return to the major caveat to our present knowledge of how voice quality is socially perceived: the limited linguistic and cultural diversity represented in the literature. Our results indicate that creak can be associated with social meanings independent of its linguistic function, as shown in the case of Mandarin. If we compare the current results from Mandarin listeners with previous results from American English listeners, both similarities and differences can be identified. On the one hand, both speech communities exhibit some gender-specific patterns in social perception of creak; on the other hand, the social meanings of creak are substantially different between the two languages. In American English, creak tends to be associated with masculinity and coolness, and young creaky women may be socially penalized. Interestingly, some of our findings suggest that creak may be associated with masculinity in Mandarin as well. For example, creak in Mandarin makes a young woman sound more gender-ambiguous and a young man sound older (and, therefore, more like a prototypical man). Crucially, however, this association does not negatively impact social perception of young creaky women for Mandarin listeners of either gender. On the contrary, at least in terms of warmth, young male Mandarin listeners disfavored creak in fellow young men while tending to favor creak in young women. The fact that both English and Mandarin listeners seem to be influenced by the inherent association of creak and masculinity, yet exhibit significantly different gendered perceptions of creak, attests to the value of crosslinguistic and crosscultural investigations for advancing scientific understanding of sociolinguistic perception.

We are, however, cautious to point out a few limitations of the current study, which argue for further research on the social perception of voice quality. First, gender was given a simplified, binary treatment, both in recruitment and in study design. We targeted only female and male genders in recruiting talkers and listeners and considered primarily these two genders in designing the study materials. Furthermore, gender identity was conflated with biological sex throughout the study. For example, in the social perception experiment, listeners were asked not to identify talkers' gender identity per se, but rather their sex. Second, the implementation of creak was in terms of one specific type: amplitude-modulated, doubly-pulsed creak. Although our resynthesis process could generate this type of creak over whole utterances, it was not able to replicate natural variety in the realization of creaky voice, including types with distinct acoustic patterns (e.g., vocal fry), positional specificity (e.g., phrase-final creak), and tonal or register specificity (e.g., low tone). In regard to creak in tonal languages, period doubling may also lead to a "low tone" bias, although the effect is allegedly weaker in amplitude-modulated doubling (the type of doubling used in our creaky stimuli) than in frequency-modulated doubling (Huang, 2023, 2024b). Finally, although our resynthesized stimuli were highly comprehensible and reasonably natural, they did not reach ceiling levels of comprehensibility and naturalness. Because of this, we constructed a cover story for listeners (see *Procedure*) to explain why some stimuli might not sound fully natural. Despite the plausibility of this story in the age of virtual assistants and synthetic voices, it is possible that results could shift if

listeners were, instead, asked to evaluate real talkers uttering maximally natural-sounding stimuli.

In closing, we envision several directions for future research. For one, there is ample room to expand the current line of research to consider a wider spectrum of gender identities, including more gender-minority groups. In addition, the current paradigm could be extended to other non-modal voice qualities (e.g., breathiness) for a more comprehensive picture of the social meanings of voice qualities in Mandarin. Given that the current study asked a limited number of questions about listeners' gender- and sexuality-related ideologies, future studies could add much more nuance to the relationship we observed between attitudes and social perception. Lastly, the presence of crosslinguistic and crosscultural variation in the social meanings of voice qualities points toward many potential research avenues concerning bilingual populations and the development (and change) of social perception in the context of bilingual language acquisition. Do bilingual listeners show similar or distinct social perceptual patterns regarding voice qualities across their different languages? How, and when, do bilinguals acquire novel social perceptual patterns in a second (or third) language? Addressing such questions will contribute to a broader and more inclusive understanding of how voice qualities are cognitively represented and socially perceived.

Data Availability

Datasets and supplementary files related to the current study are available in the OSF repositories https://osf.io/srp7u and https://osf.io/cejgf. The repositories include: (1) datasets and statistical analysis reports (including R scripts and full model outputs) that allow readers to replicate the analyses reported in the paper; (2) the list of questions (with annotations) in the social perception experiment; (3) the list of sentences used to elicit natural speech recordings; and (4) a description of the resynthesis process using a Klatt synthesizer.

References

Abdelli-Beruh NB, Drugman T, Red Owl RH (2016) Occurrence frequencies of acoustic patterns of vocal fry in American English speakers. Journal of Voice 30(6):759.e11–759.e20

Agathe B, Claire P (2013) The influence of language and speech task upon creaky voice use among six young American women learning French. In: Bimbot F, Cerisara C, Fougeron C, et al (eds) Proceedings of the 14th Annual Conference of the International Speech Communication Association (INTERSPEECH 2013). International Speech Communication Association, Lyon, France, p 2395–2399

Anderson RC, Klofstad CA, Mayew WJ, et al (2014) Vocal fry may undermine the success of young women in the labor market. PLoS ONE 9(5):e97506

Batliner A, Burger S, Johne B, et al (1994) MÜSLI: A classification scheme for laryngealizations. Report 45, Ludwig-Maximilians-Universität München, Friedrich-Alexander-Universität Erlangen-Nürnberg

Belotel-Grenié A, Grenié M (2004) The creaky voice phonation and the organisation of Chinese discourse. In: Proceedings of the 1st International Symposium on Tonal Aspects of Languages (TAL 2004). International Speech Communication Association, Beijing, China, p 5–8

Borrie SA, Delfino CR (2017) Conversational entrainment of vocal fry in young adult female American English speakers. Journal of Voice 31(4):513.e25–513.e32

Brown J, Sonderegger M (2024) Creaky voice variation across language, gender and age in Canadian English–French bilingual speech. In: Cho T, Kim S, Holliday J, et al (eds) Proceedings of the 19th Conference on Laboratory Phonology. Hanyang Institute for Phonetics and Cognitive Sciences of Language, Seoul, South Korea, p 309–310

Brown J, Sonderegger M (2025) A sociophonetic study of creaky voice across language, gender and age in Canadian English-French bilinguals. Journal of Phonetics

Brown KM, Dahl KL, Cler GJ, et al (2021) Listener age and gender diversity: Effects on voice-based perception of gender. Journal of Voice 35(5):739–745

Brown P, Levinson SC (1978) Universals in language usage: Politeness phenomena. In: Goody EN (ed) Questions and Politeness: Strategies in Social Interaction. Cambridge University Press, Cambridge, UK, p 56–311

Callier P (2010) Voice quality, rhythm and valorized femininities, poster presented at Sociolinguistics Symposium 18, Southampton, UK

Cantor-Cutiva LC, Bottalico P, Webster J, et al (2023) The effect of bilingualism on production and perception of vocal fry. Journal of Voice 37(6):970.e1–970.e10

Chao M (2017) An echoic account of vocal fry. Master's thesis, San Francisco State University, San Francisco, CA

Chao M, Bursten JRS (2021) Girl talk: Understanding negative reactions to female vocal fry. Hypatia 36(1):42–59

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

Davidson L (2018) The effects of pitch, gender, and prosodic context on the identification of creaky voice. Phonetica 76(4):235–262

Davidson L (2020) Contributions of modal and creaky voice to the perception of habitual pitch. Language 96(1):e22–e37

Davison DS (1991) An acoustic study of so-called creaky voice in Tianjin Mandarin. UCLA Working Papers in Phonetics 78:50–57

de Leeuw E, Chang CB (2024) Phonetic and phonological L1 attrition and drift in bilingual speech. In: Amengual M (ed) The Cambridge Handbook of Bilingual Phonetics and Phonology. Cambridge University Press, Cambridge, UK, p 721—745

Dilley L, Shattuck-Hufnagel S, Ostendorf M (1996) Glottalization of word-initial vowels as a function of prosodic structure. Journal of Phonetics 24(4):423–444

Esling J (1978) The identification of features of voice quality in social groups. Journal of the International Phonetic Association 8(1–2):18–23

Esposito CM, Khan SD (2020) The cross-linguistic patterns of phonation types. Language and Linguistics Compass 14(12):e12392

Foulks N (2020) Listeners' attitudes towards young women with glottal fry. Master's thesis, East Tennessee State University, Johnson City, TN

Gerratt BR, Kreiman J (2001) Toward a taxonomy of nonmodal phonation. Journal of Phonetics 29(4):365–381

Gibson TA, Summers C (2021) A perceptual study of cross-linguistic influence on vocal fry use in women exposed to two languages. International Journal of Bilingual Education and Bilingualism 24(3):373–385

Gittelson 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

Gobl C, Ní Chasaide A (2003) The role of voice quality in communicating emotion, mood and attitude. Speech Communication 40(1–2):189–212

Gobl C, Bennett E, Ní Chasaide A (2002) Expressive synthesis: How crucial is voice quality. In: Proceedings of 2002 IEEE Workshop on Speech Synthesis. Institute of Electrical and Electronics Engineers, Piscataway, NJ, p 91–94

Gordon M, Ladefoged P (2001) Phonation types: A cross-linguistic overview. Journal of Phonetics 29(4):383–406

Greer SDF (2015) The perception of coolness: Voice quality and its social uses and interpretations. Master's thesis, University of Calgary, Calgary, Canada

Greer SDF, Winters SJ (2015) The perception of coolness: Differences in evaluating voice quality in male and female speakers. In: The Scottish Consortium for ICPhS 2015 (ed) Proceedings of the 18^{th} International Congress of Phonetic Sciences. University of Glasgow, Glasgow, UK, p paper number 0883

Grivičić T, Nilep C (2004) When phonation matters: The use and function of *yeah* and creaky voice. Colorado Research in Linguistics 17(1):1–11

Hancock AB, Pool SF (2017) Influence of listener characteristics on perceptions of sex and gender. Journal of Language and Social Psychology 36(5):599–610

Hedelin P, Huber D (1990) Pitch period determination of aperiodic speech signals. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '90), vol 1. Institute of Electrical and Electronics Engineers, Piscataway, NJ, p 361–364

Henton C, Bladon A (1988) Creak as a sociophonetic marker. In: Hyman LM, Li CN (eds) Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin. Routledge, London, UK, p 3–29

Holliday J, Walker A, Jung M, et al (2023) Bringing indexical orders to non-arbitrary meaning: The case of pitch and politeness in English and Korean. Laboratory Phonology 14(1):1–24

Huang Y (2023) Phonetics of period doubling. PhD thesis, University of California, San Diego, San Diego, CA

Huang Y (2024a) The effects of vocal fry and period doubling on the perceived naturalness of Mandarin tones. In: Cho T, Kim S, Holliday J, et al (eds) Proceedings of the 19th Conference on Laboratory Phonology. Hanyang Institute for Phonetics and Cognitive Sciences of Language, Seoul, South Korea, p 431–432

Huang Y (2024b) Perception and imitation of period-doubled phonation: Pitch and voice quality. Journal of the Acoustical Society of America 156(2):1391–1412

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. University of Glasgow, Glasgow, UK, p paper number 0821

Kim JY (2017) Voice quality transfer in the production of Spanish heritage speakers and English L2 learners of Spanish. In: Perpiñán S, Heap D, Moreno-Villamar I, et al (eds) Romance Languages and Linguistic Theory 11: Selected Papers from the 44th Linguistic Symposium on Romance Languages (LSRL), London, Ontario. John Benjamins Publishing, Amsterdam, The Netherlands, p 191–207

Klatt DH, Klatt LC (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. Journal of Acoustical Society of America 87(2):820–857

Kreiman J, Gerratt BR, Khan SD (2010) Effects of native language on perception of voice quality. Journal of Phonetics 38(4):588–593

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

Kuang J (2018) The influence of tonal categories and prosodic boundaries on the creakiness in Mandarin. Journal of the Acoustical Society of America 143(6):EL509–EL515

Kuang J, Liberman M (2015) Influence of spectral cues on the perception of pitch height. In: The Scottish Consortium for ICPhS 2015 (ed) Proceedings of the 18th International Congress of Phonetic Sciences. University of Glasgow, Glasgow, UK, p paper number 0435

Kuang J, Liberman M (2016) The effect of vocal fry on pitch perception. In: Proceedings of the 2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Institute of Electrical and Electronics Engineers, Piscataway, NJ, p 5260–5264

Kuang J, Liberman M (2018) Integrating voice quality cues in the pitch perception of speech and non-speech utterances. Frontiers in Psychology 9:2147

Kunnemann DB (2017) Perceptions of the effects of vocal fry on aspirational careers in prospective job markets, undergraduate honors thesis, University of Arkansas, Fayetteville, AR

Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software 82(13):1–26

Ladefoged P (1982) The linguistic use of different phonation types. UCLA Working Papers in Phonetics 54:28–39

Laver J (1980) The Phonetic Description of Voice Quality. Cambridge University Press, Cambridge, UK

Laver JDM (1968) Voice quality and indexical information. British Journal of Disorders of Communication 3(1):43–54

Lee KE (2016) The perception of creaky voice: Does speaker gender affect our judgments? Master's thesis, University of Kentucky, Lexington, KY

Lee S (2015) Creaky voice as a phonational device marking parenthetical segments in talk. Journal of Sociolinguistics 19(3):275–302

Lenth RV, Banfai B, Bolker B, et al (2025) emmeans: Estimated marginal means, aka least-squares means [R package], version 1.11.0-003. Available online: https://cran.r-project.org/package=emmeans

Li A, Lai W (2023) How do listeners evaluate creak: A matched-guise study in Mandarin Chinese, paper presented at the 97th Annual Meeting of the Linguistic Society of America, Denver, CO

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

Li P, Zhang F, Tsai E, et al (2014) Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. Bilingualism: Language and Cognition 17(3):673–680

Li Q, Mok P (2023) A perception study on voice quality and stance in Mandarin Chinese. In: Skarnitzl R, Volín J (eds) Proceedings of the 20th International Congress of Phonetic Sciences. Guarant International, Prague, Czechia, p 1786–1790

Loakes D, Gregory A (2022) Voice quality in Australian English. JASA Express Letters 2(8):085201

Melvin S, Clopper CG (2015) Gender variation in creaky voice and fundamental frequency. In: The Scottish Consortium for ICPhS 2015 (ed) Proceedings of the 18^{th} International Congress of Phonetic Sciences. University of Glasgow, Glasgow, UK, paper number 0320

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

Monsen RB, Engebretson AM (1977) Study of variations in the male and female glottal wave. Journal of the Acoustical Society of America 62(4):981–993

Nissen S, Randle QB, Johnson JL, et al (2020) Prosodic elements for content delivery in broadcast journalism: A quantitative study of vocal pitch. Electronic News 14(2):63–77

Nodari R, Celata C, Nagy N (2019) Socio-indexical phonetic features in the heritage language context: Voiceless stop aspiration in the Calabrian community in Toronto. Journal of Phonetics 73:91-–112

Ohala JJ (1983) Cross-language use of pitch: An ethological view. Phonetica 40(1):1–18

Oliveira G, Davidson A, Holczer R, et al (2016) A comparison of the use of glottal fry in the spontaneous speech of young and middle-aged American women. Journal of Voice 30(6):684–687

Parker MA, Borrie SA (2017) Judgments of intelligence and likability of young adult female speakers of American English: The influence of vocal fry and the surrounding acoustic-prosodic context. Journal of Voice 32(5):538–545

Pennock-Speck B (2005) Voice quality and the construction of female voice styles. In: Asociación Española de Estudios Anglo-Norteamericanos, de Leonardo JJCG (eds) Actas XXVIII Congreso Internacional AEDEAN. University of Valencia, Valencia, Spain, p 407–414

Pittam J (1987) Listeners' evaluations of voice quality in Australian English speakers. Language and Speech 30(2):99–113

Podesva RJ (2011) Gender and the social meaning of non-modal phonation types. In: Cathcart C, Chen IH, Finley G, et al (eds) Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society. Berkeley Linguistics Society, Berkeley, CA, p 427–448

Pratt TC (2018) Affective sociolinguistic style: An ethnography of embodied linguistic variation in an arts high school. PhD thesis, Stanford University, Stanford, CA

R Development Core Team (2023) R: A language and environment for statistical computing, version 4.4.0. http://www.r-project.org

Redi L, Shattuck-Hufnagel S (2001) Variation in the realization of glottalization in normal speakers. Journal of Phonetics 29(4):407–429

Sakaluk JK, Short SD (2017) A methodological review of exploratory factor analysis in sexuality research: Used practices, best practices, and data analysis resources. Journal of Sex Research 54(1):1–9

Sankoff G (2004) Adolescents, young adults and the critical period: Two case studies from "Seven Up". In: Fought C (ed) Sociolinguistic Variation: Critical Reflections. Oxford University Press, Oxford, UK, p 121—139

Sebregts K, Vriesendorp H, Quené H, et al (2023) Creaky voice in L2 English and L1 Dutch. In: Skarnitzl R, Volín J (eds) Proceedings of the 20th International Congress of Phonetic Sciences. Guarant International, Prague, Czechia, p 1841–1845

Sebregts K, Vriesendorp H, Quené H, et al (2024) Long-term phonetic convergence vs. speaker-specificity: Creaky voice in L2 English. In: Cho T, Kim S, Holliday J, et al (eds) Proceedings of the 19th Conference on Laboratory Phonology. Hanyang Institute for Phonetics and Cognitive Sciences of Language, Seoul, South Korea, p 387–388

Shue YL, Keating P, Vicenik C, Yu K (2015) VoiceSauce: A program for voice analysis. In: Lee WS, Zee E (eds) Proceedings of the 17^{th} International Congress of Phonetic Sciences. City University of Hong Kong, Hong Kong, p 1846–1849

Starr RL (2015) Sweet voice: The role of voice quality in a Japanese feminine style. Language in Society 44(1):1–34

Stuart-Smith J (1999) Glasgow: Accent and voice quality. In: Foulkes P, Docherty GJ (eds) Urban Voices: Accent Studies in the British Isles. Arnold, London, UK, p 201–222

Szakay A (2012) Voice quality as a marker of ethnicity in New Zealand: From acoustics to perception. Journal of Sociolinguistics 16(3):382–397

Tian J, Zhou Y, Kuang J (2019) Cross-linguistic variation in the phonetic realization of "breathier voice". In: Calhoun S, Escudero P, Tabain M, et al (eds) Proceedings of the 19^{th} International Congress of Phonetic Sciences. Australasian Speech Science and Technology Association Inc., Canberra, Australia, p 1450–1454

Torres PJ, Henry SG, Ramanathan V (2020) Let's talk about pain and opioids: Low pitch and creak in medical consultations. Discourse Studies 22(2):174–204

Trudgill P (1974) The Social Differentiation of English in Norwich. Cambridge University Press, Cambridge, UK

van Bezooijen R (1995) Sociocultural aspects of pitch difference between Japanese and Dutch women. Language and Speech 38(3):253–265

White H, Penney J, Gibson A, et al (2023) Convergence of creaky voice use in Australian English. In: Skarnitzl R, Volín J (eds) Proceedings of the 20^{th} International Congress of Phonetic Sciences. Guarant International, Prague, Czechia, p 1791–1795

Wilce JM (1997) Discourse, power, and the diagnosis of weakness: Encountering practitioners in Bangladesh. Medical Anthropology Quarterly 11(3):352–374

Wolk L, Abdelli-Beruh NB, Slavin D (2012) Habitual use of vocal fry in young adult female speakers. Journal of Voice 26(3):e111–e116

Xu A, Lee A (2018) Perception of vocal attractiveness by Mandarin native listeners. In: Klessa K, Bachan J, Wagner A, et al (eds) Proceedings of the 9th International Conference on Speech Prosody. International Speech Communication Association, Poznań, Poland, p 344–348

Xu Y, Lee A, Wu WL, et al (2013) Human vocal attractiveness as signaled by body size projection. PLoS ONE 8(4):e62397

Yang RX (2011) The phonation factor in the categorical perception of Mandarin tones. In: Lee WS, Zee E (eds) Proceedings of the 17^{th} International Congress of Phonetic Sciences. City University of Hong Kong, Hong Kong, SAR, p 2204–2207

Yau E (2020) Podcaster prosody: Creaky voice and Sarah Koenig's journalistic persona. Lifespans and Styles 6(2):2–10

Yu KM, Lam HW (2014) The role of creaky voice in Cantonese tonal perception. Journal of the Acoustical Society of America 136(3):1320–1333

Yuasa IP (2010) Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? American Speech 85(3):315–337

Zhu S, Chong S, Chen Y, et al (2022) Effect of language on voice quality: An acoustic study of bilingual speakers of Mandarin Chinese and English. Folia Phoniatrica et Logopaedica 74(6):421–430

Zimman L (2017) Affective stance and voice quality in a pervasively creaky speaker: Stance objects as a tool for investigating indexical meaning, paper presented at New Ways of Analyzing Variation (NWAV) 45, Vancouver, Canada

Author contributions

Y.Y. and C.B.C. conceived the experiment; Y.Y. and C.B.C. acquired funding; Y.Y. and C.B.C. obtained ethics approval; Y.Y. and M.L. created the stimulus materials; Y.Y. and M.L. conducted the experiment; Y.Y. and C.B.C. curated the data; Y.Y. and M.L. analyzed the results; and Y.Y. and C.B.C. wrote the original draft of the manuscript. All authors contributed to designing the experiment and to manuscript review and editing.

Competing interests

The authors declare no competing interests.

Ethical approval

The study was approved by the Institutional Review Board of the Hong Kong Polytechnic University on Apr 6, 2022 (ethics approval number HSEAR20220405004). The approval covered all work reported in this article. All study protocols were performed in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments.

Informed consent

Written informed consent was obtained from all participants, who received a nominal amount of monetary compensation for their time and effort. Consent was obtained between July and December 2023, by the researcher from the participant directly, and covered participation, data use, and data publication. All participants were fully informed of possible risks of the study as well as of the anonymity and intended use of their data; they were also debriefed on the aims of the research.

Figure legends

Figure 1: **Waveforms of sample stimuli.** The waveforms show resynthesized [a] produced by female talker F017. (a): creaky version. (b): modal version.

Figure 2: **Scores for gender attitudes and sexuality attitudes, by listener gender.** A: gender attitudes. B: sexuality attitudes. The maximum score for each is 5. Each horizontal line indicates the median; each box, the interquartile range; and whiskers, values within 1.5 times the interquartile range.

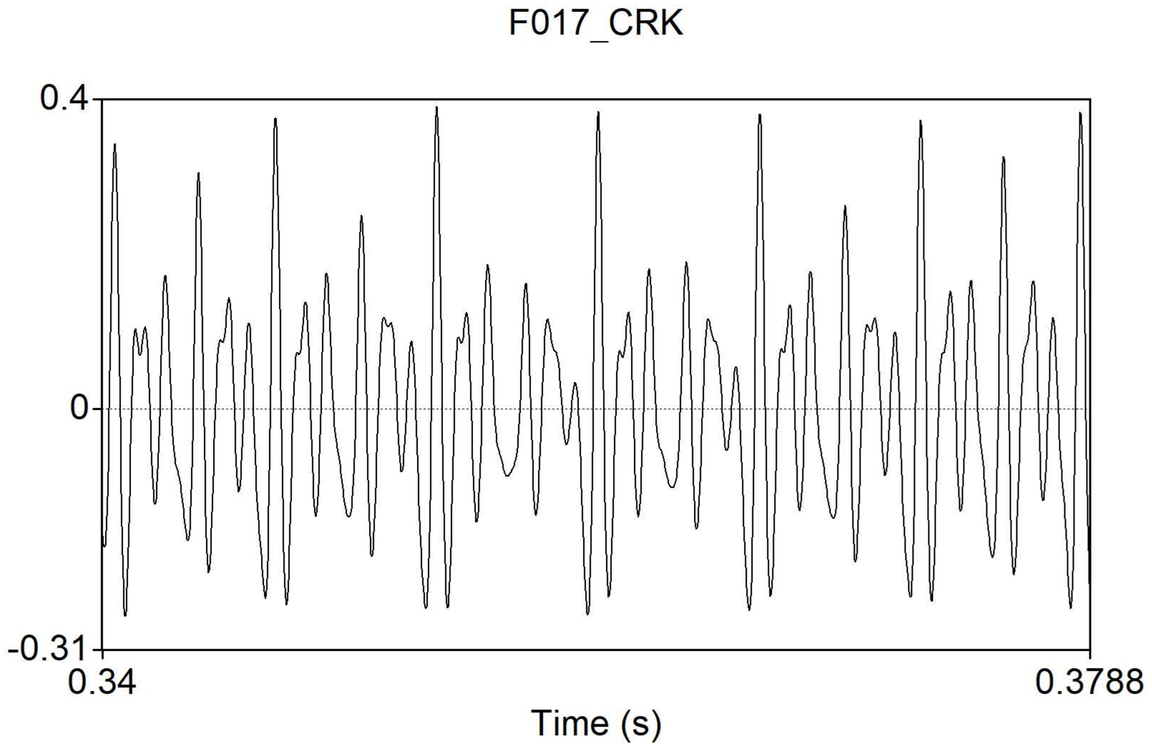
Figure 3: **Confidence ratings for gender identification judgments, by listener gender, talker gender, and voice quality.** "Most confident" corresponds to a rating of 5. Error bars indicate standard deviations.

Figure 4: **Ratings of talkers' voice gender typicality, by listener gender, talker gender, and voice quality.** "Most typical" corresponds to a rating of 5. Error bars indicate standard deviations.

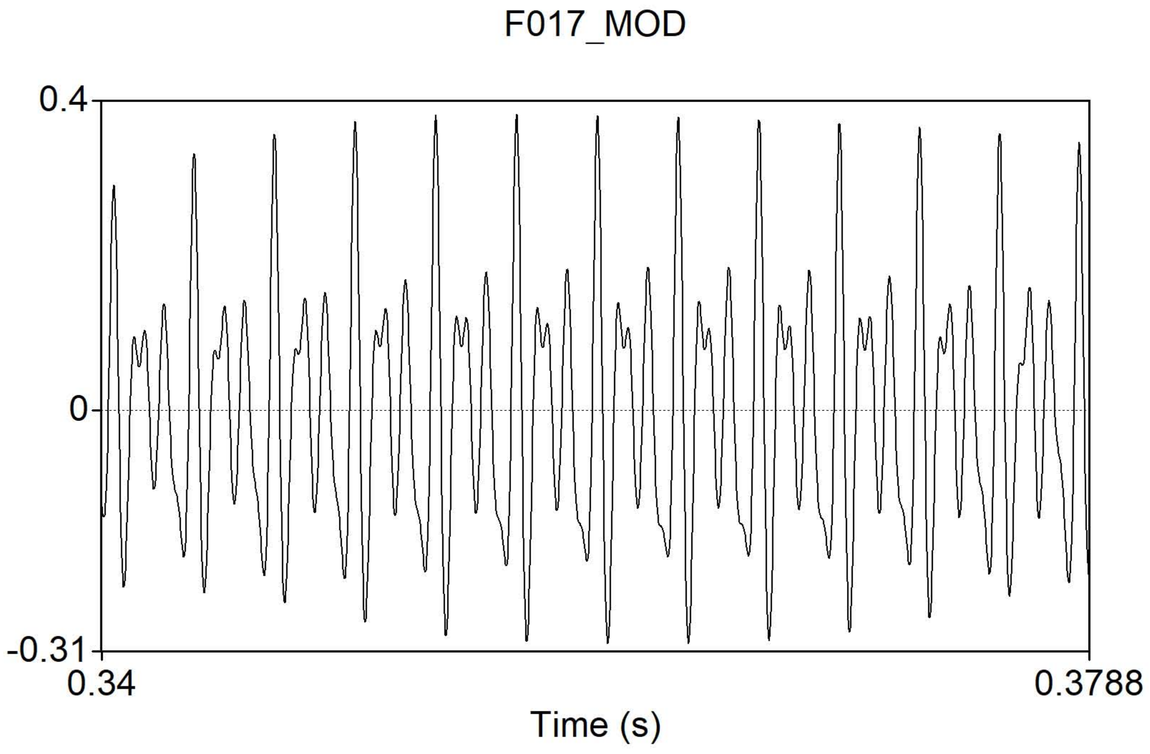
Figure 5: **Mean COMPETENCE scores, by listener gender, talker gender, and voice quality.** Error bars indicate standard deviations.

Figure 6: **Mean LIKABILITY scores, by listener gender, talker gender, and voice quality.** Error bars indicate standard deviations.

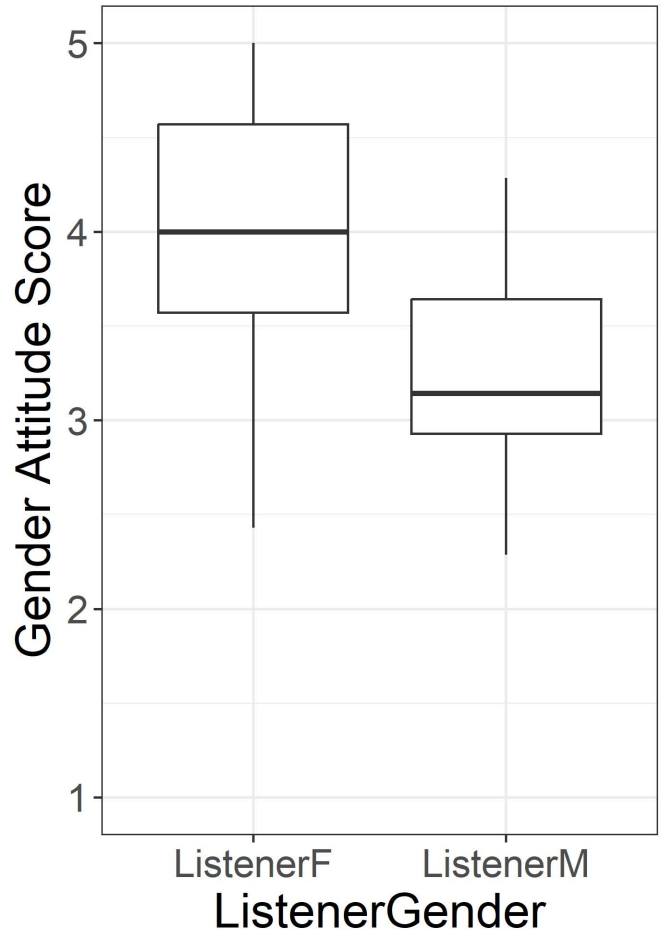
Figure 7: **Mean WARMTH scores, by listener gender, talker gender, and voice quality.** Error bars indicate standard deviations.

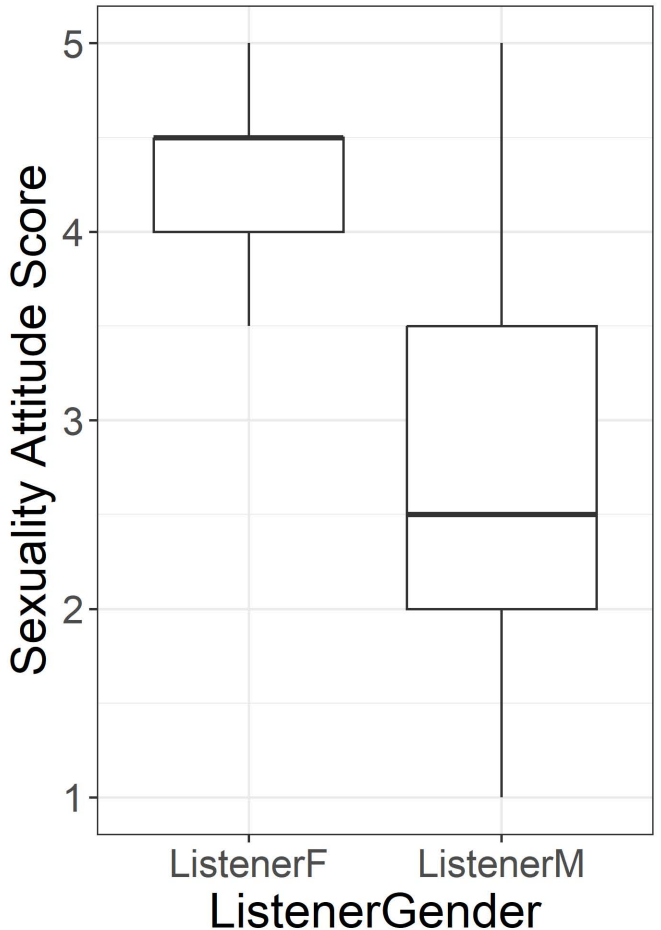

F017_CRK
0.4
0
-0.31
0.34
0.3788
Time (s)

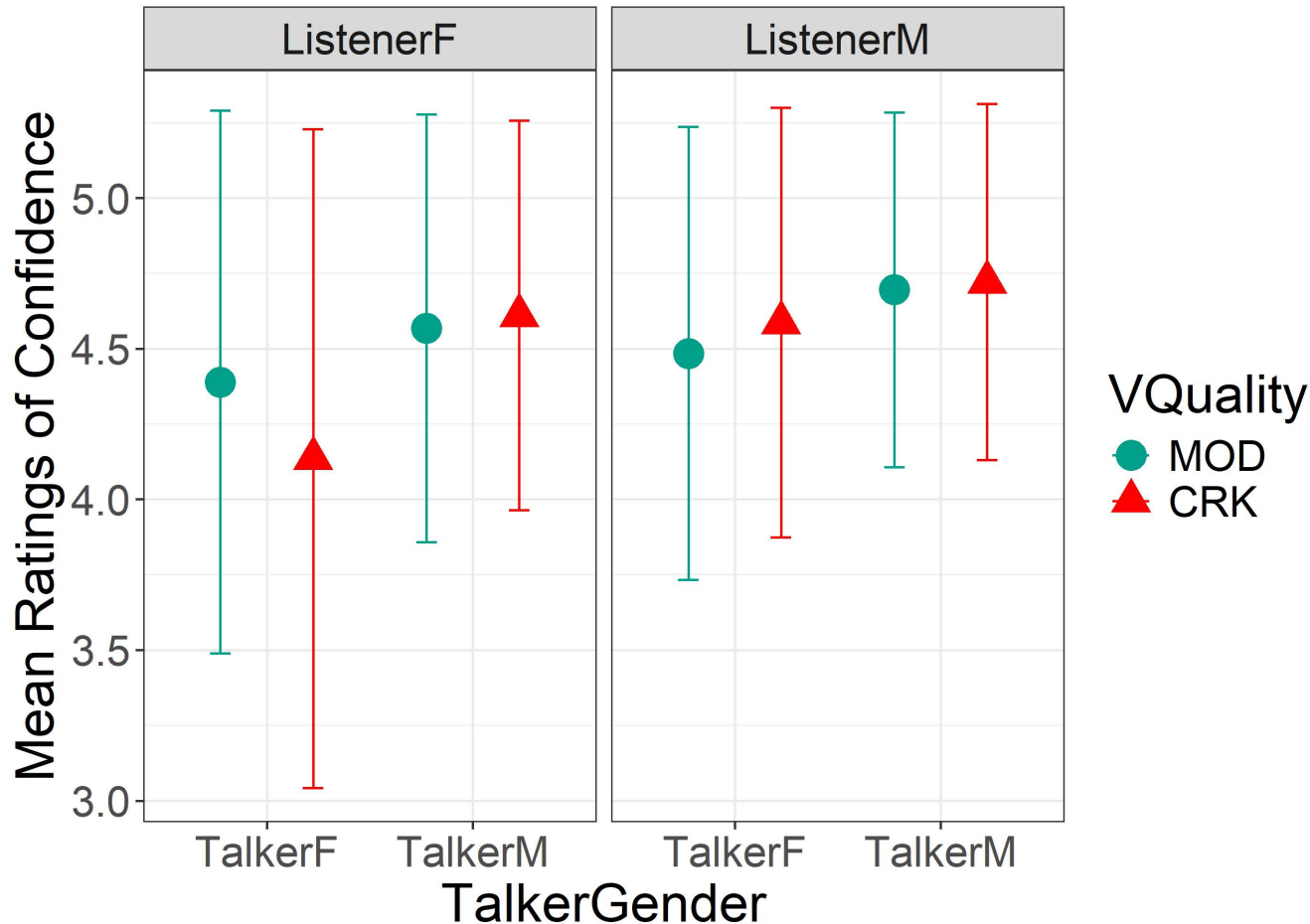
(a)

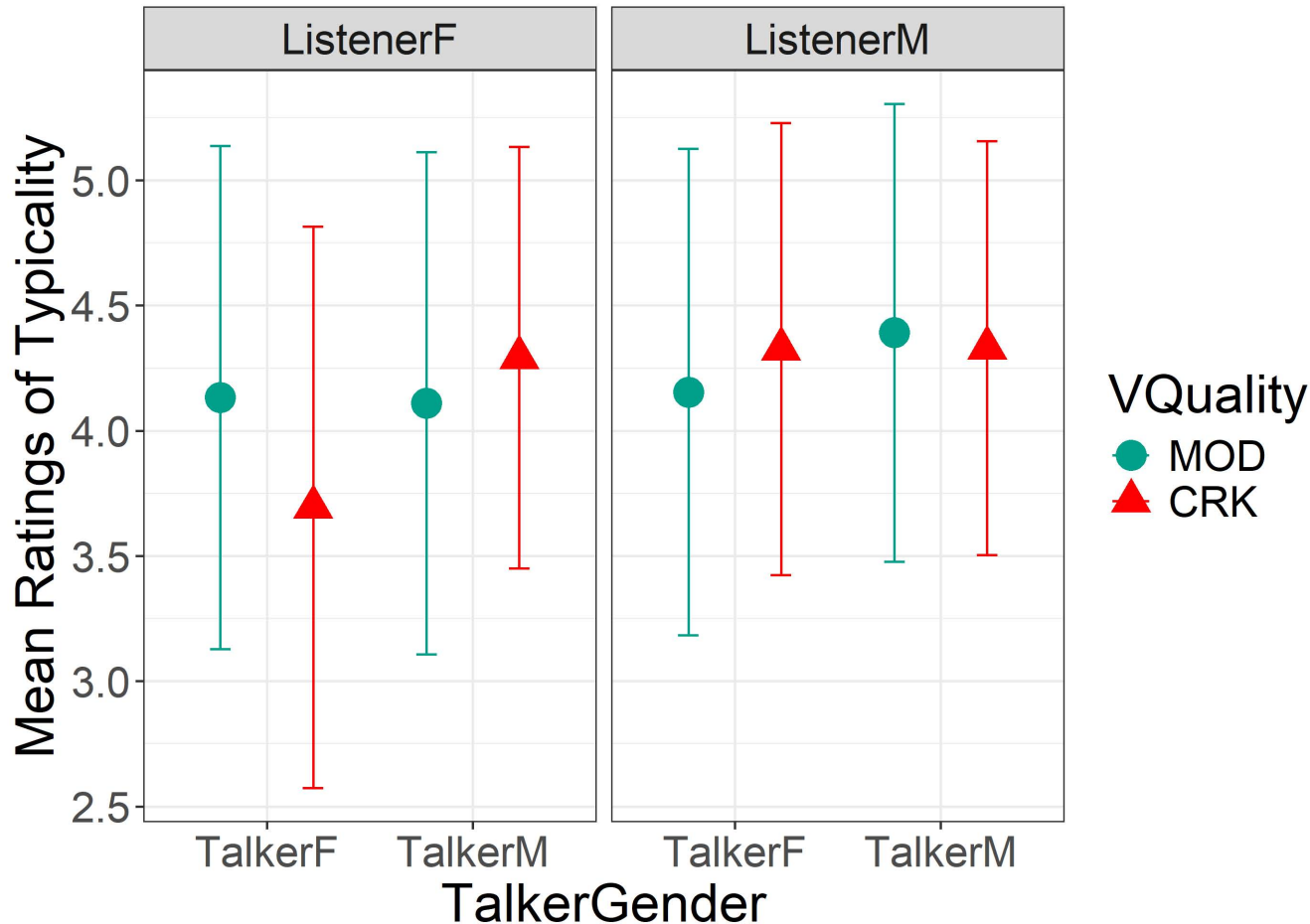

F017_MOD
0.4
0
-0.31
0.34
0.3788
Time (s)

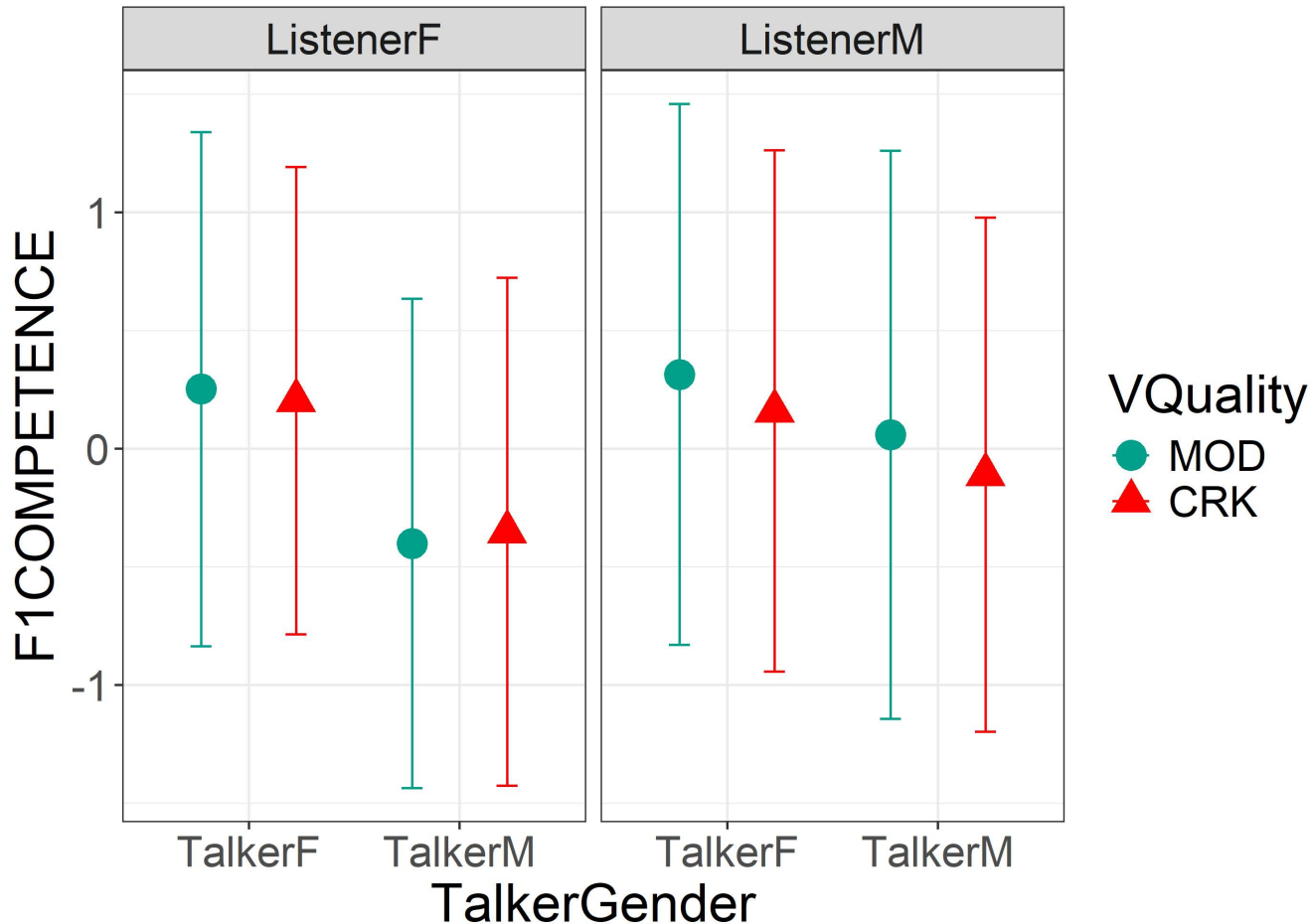
(b)

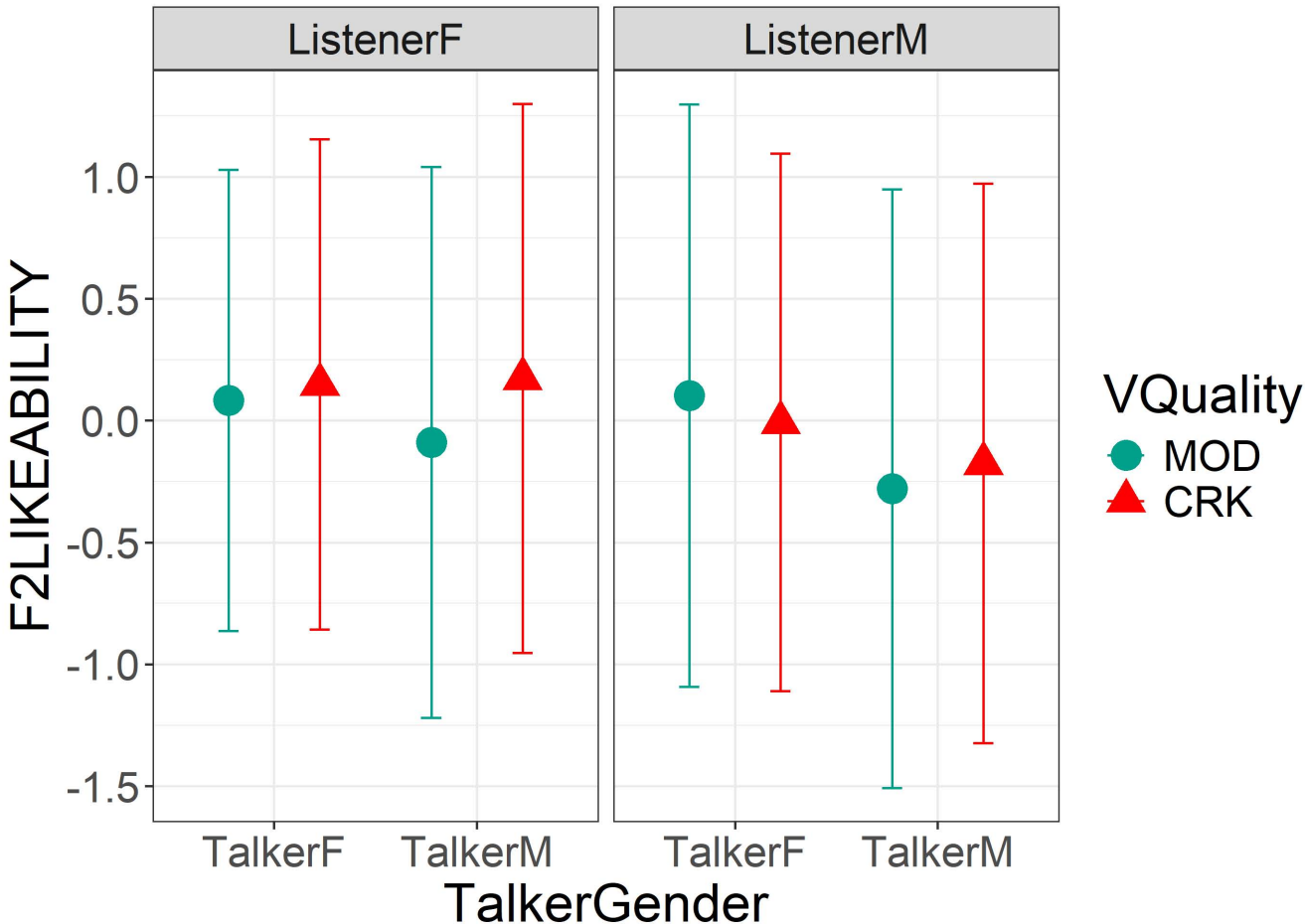

Gender Attitude Score
5
4
3
2
1
ListenerF
ListenerM
ListenerGender

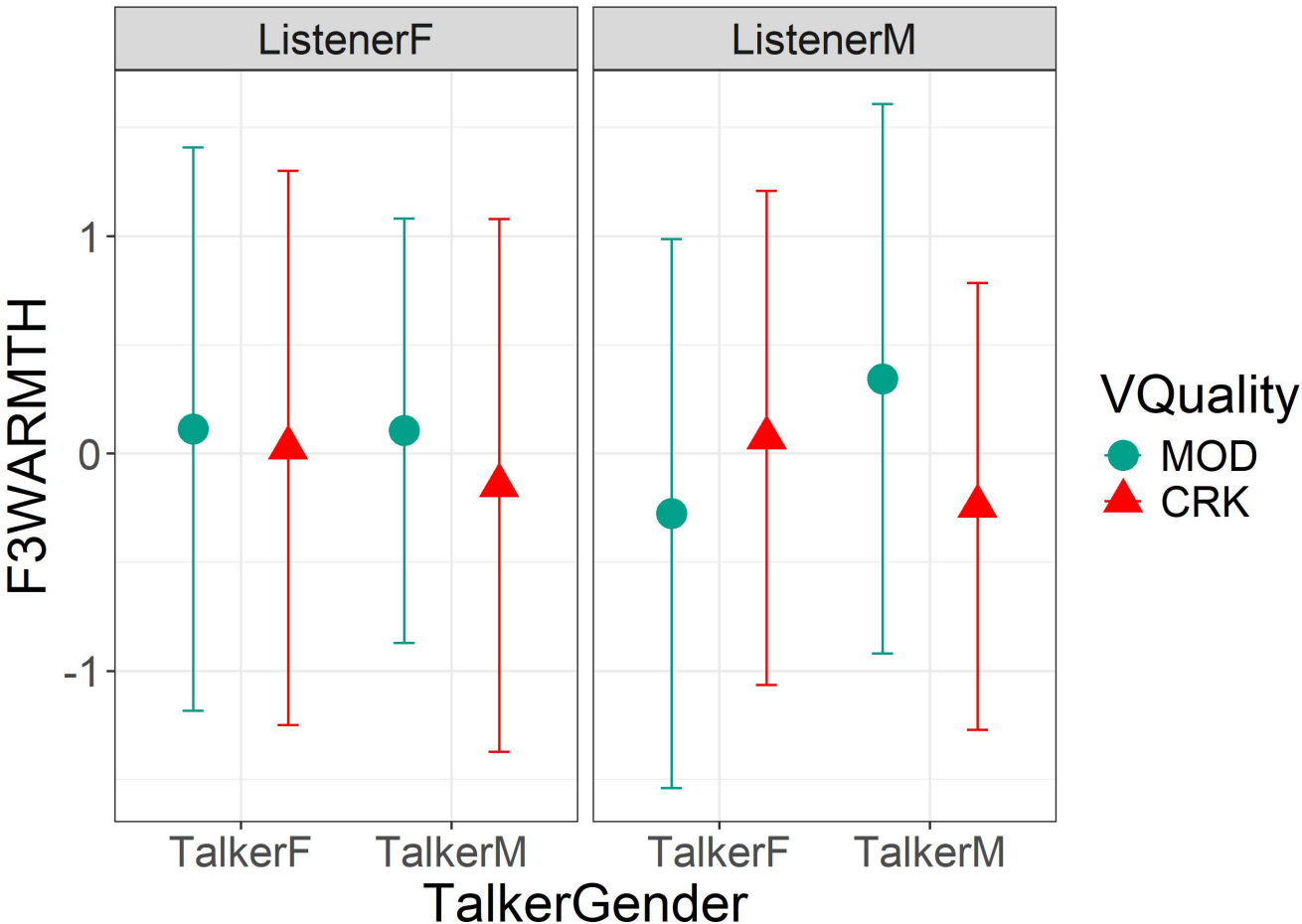

Sexuality Attitude Score
5
4
3
2
1
ListenerF
ListenerM
ListenerGender


ListenerF
ListenerM
Mean Ratings of Confidence
5.0
4.5
4.0
3.5
3.0
TalkerF
TalkerM
TalkerF
TalkerM
TalkerGender
VQuality
MOD
CRK


ListenerF
ListenerM
Mean Ratings of Typicality
5.0
4.5
4.0
3.5
3.0
2.5
TalkerF
TalkerM
TalkerF
TalkerM
TalkerGender
VQuality
MOD
CRK


ListenerF
ListenerM
F1COMPETENCE
1
0
-1
TalkerF
TalkerM
TalkerF
TalkerM
TalkerGender
VQuality
MOD
CRK


ListenerF
ListenerM
F2LIKEABILITY
1.0
0.5
0.0
-0.5
-1.0
-1.5
TalkerF
TalkerM
TalkerF
TalkerM
TalkerGender
VQuality
MOD
CRK


ListenerF
ListenerM
F3WARMTH
1
0
-1
TalkerF
TalkerM
TalkerF
TalkerM
TalkerGender
VQuality
MOD
CRK