The influence of pitch and speaker sex on the identification of creaky voice by female listeners

Hannah White, Andy Gibson, Joshua Penney, Anita Szakay, Felicity Cox

Department of Linguistics, Macquarie University

hannah.white2@hdr.mq.edu.au; andy.gibson, joshua.penney, anita.szakay, felicity.cox@mq.edu.au

Abstract

Past work has raised questions about how creaky voice quality is identified in different voices, particularly whether greater pitch differences between modal and creaky voice facilitates creak identification or whether social expectations bias identification in certain voices. While the role of pitch and speaker sex in creak identification has been investigated, results have been equivocal. In this study, we used highly controlled stimuli to examine the extent to which pitch and/or speaker sex affect the identification of creak. A study of 130 Australian English-speaking female listeners found that pitch and speaker sex interacted in listeners' perception of creaky voice.

Index Terms: creaky voice, perception, speaker sex, pitch, Australian English

1. Introduction

Creaky voice is a non-modal voice quality, typically perceived to be rough, pulse-like and low in pitch [1, 2, 3, 4]. Acoustic analyses have shown there are many different phonetic realisations of creaky voice [2]; however perception research suggests that low pitch is a salient cue to the perception of all realisations despite their phonetic differences [5], and that the presence of creak can lead listeners to assign lower pitch ratings to these utterances compared to fully modal utterances [6].

In media reports and popular culture, creak is generally associated with young women's speech and is attached predominantly to negative connotations [7, 8, 9]. These negative associations are likely linked to the prominent use of creak by celebrities such as Kim Kardashian, Britney Spears and Paris Hilton [10]. Some linguistic perception research has corroborated the negative sentiments seen in media headlines. Such studies have found that voices with creak are rated as less competent, less educated, less trustworthy, less hirable and lower in solidarity than modal voices [11, 12]. These findings were especially strong in descriptions of female speakers compared to males [11, 12]. However, other studies have found more positive meanings associated with creak. Yuasa [13] found that male and female listeners rated American female speakers with creak as sounding casual, educated and genuine, and Gobl and Ní Chasaide [14] found that Irish listeners rated speakers with creak as sounding relaxed, intimate and unafraid.

In production research, many studies have suggested that creak is more prevalent in female speech, especially in American English [13, 15, 16, 17, 18]. However, Dallaston and Docherty [19] found that the majority of research in this area has been conducted in the United States, and that studies disproportionately focus on young women; therefore, more work is needed in order to empirically confirm whether creak is more prevalent in female speech than male speech. Studies from the United Kingdom have found creak to be more prevalent in the speech of men [20, 21, 22], as did a recent study of Australian English (AusE) [23], and a 2016 study of American English suggests no difference in creak prevalence between females and males [24]. Although the present study focuses solely on female listeners, future analyses will also include male listeners.

We know that creak is a feature used by both males and females in speech [16, 20, 24]; however, the general association of creaky voice with women in the wider population stands [7, 8, 9, 10]. This raises the question of whether listeners are biased to hear creak in female speech compared to male speech. Davidson [25] addressed this question, by investigating whether the identification of creak (i.e. whether a listener decides creak is present or not) is influenced more by listener expectations or by acoustic properties of the voice. Two male and two female speakers were used in her study. Within each sex one speaker had relatively high and one had relatively low average f0 [25]. The high male voice and low female voice were matched for average pitch as much as possible using natural speech. It was hypothesised that high identification of creak in female speech compared to male speech, regardless of pitch, could be explained by a bias for identifying creak in females due to listener expectations. However, it was also hypothesised that the difference in f0 between modal voice (typically higher for females) and creak (typically low for all speakers) would make creak more salient in female voices. Under this hypothesis, it was proposed that creak would be most identifiable in the high female voice, equally next identifiable in the low female and high male voices (as they were matched for f0) and least identifiable in the low male voice. Two experiments were run in which listeners were asked whether there was creak in the stimuli they were played. The second experiment used low female and high male voices that were more closely matched than in the first version. The results showed that listeners consistently false alarmed in the low male modal condition, i.e. they identified creak when it was not present. However, conflicting results across experiment versions meant it was not possible to attribute findings to any proposed hypothesis. In addition, due to the use of natural speech as stimuli, the high male and low female conditions could not be exactly matched for f0, making it impossible to unpack the influence of pitch and speaker sex on creak identification.

1.1. Research question and expectations

Our motivation is to disentangle whether pitch or speaker sex has greater influence on creak identification by using stimuli manipulated to ensure maximum control over speaker f0 and creaky voice. An experiment was designed to provide listeners with a highly controlled set of two-word noun phrases that varied according to speaker sex (female vs male), voice quality

(fully modal vs modal plus creaky component), and f0. Listeners were asked if they identified creak when presented with two types of stimuli: creaky, which contained creak, and modal, which did not contain creak (see Section 2.3). Source recordings from a male and female speaker were manipulated for f0 and presence of creak at the end of the phrase (i.e. creaky component). This resulted in four pitch conditions: low male, mid male, mid female, and high female. Importantly, mid male and mid female conditions had identical f0 contours, and differed only in source speaker sex. The process for creating the stimuli is described in Section 2.1.

We proposed three possible influences on identification of creaky voice (the first two from [25]):

1. Pitch: the difference in pitch between a speaker's modal voice and creaky voice has the greatest influence on accurate creak identification.
2. Speaker sex: the difference in the source speaker sex has the greatest influence on accurate creak identification.
3. Pitch and sex: pitch differences between modal voice and creaky voice and source speaker sex interact in creak identification.

In Table 1, we present our predictions for each scenario.

Influence	Creak condition	Modal condition
Pitch	hi-f > mid-f = mid-m > lo-m	hi-f > mid-f = mid-m > lo-m
Sex	hi-f = mid-f > mid-m = lo-m	lo-m = mid-m > mid-f = hi-f
Pitch & sex	hi-f > mid-f > mid-m > lo-m	hi-f > mid-m > mid-f > lo-m

Table 1: *Table of expectations of listener accuracy in creak identification for each proposed influence. hi-f = high female condition, mid-f = mid female condition, mid-m = mid male condition and lo-m = low male condition.*

If the pitch of the stimuli has the greatest influence on creak identification, we would expect creak to be more noticeable in the creak condition when the modal f0 is highest due to a clear difference in f0 between the modal and creak components in the stimuli. This would lead us to expect the highest creak identification accuracy in the high female pitch condition, followed by the mid female and mid male conditions (which have identical f0s), and lowest accuracy in the low male condition. We would expect the same patterning of results in the modal condition because the absence of creak would be most noticeable when modal f0 is highest. In the modal condition there is no substantial drop in f0 to trigger a creak response.

If speaker sex has the greatest influence on creak identification, we would expect creak identification accuracy patterns to be consistent within source speaker sexes in both creaky and modal conditions but different between the source sex conditions. In the creaky condition, we would expect listeners to be most accurate for the female voices and least accurate for the male voices regardless of f0 because of a bias towards hearing creak in female voices. In the modal condition, we propose that listeners would have lowest accuracy in not identifying creak in the female voice conditions due to this bias leading them to false alarm (i.e. identify creak for the female source stimuli when it is not present).

If pitch and speaker sex are interacting in creak identification, we propose that pitch would be a stronger influence on creak decisions in the most extreme pitch conditions (low male and high female). This means we would expect accuracy to be highest in identifying creak when it's present and not identifying creak when it's not present in the high female condition and lowest in the low male condition. In the mid conditions when f0 is the same regardless of speaker sex, we propose that speaker sex will mediate creak decisions. We would expect listeners to identify creak more in the mid female voice than the mid male voice in both the creaky and modal conditions due to the bias to identify creak in female voices. This would result in higher accuracy to creaky tokens and lower accuracy to modal tokens in the mid female condition compared to the mid male condition.

2. Methods

2.1. Stimuli

The stimuli were produced by one female (22 y.o.) and one male (28 y.o.), siblings, who are both native speakers of AusE. Each speaker recorded 30 two-word phrases. The phrases were chosen with reference to a list of adjective-noun pair bigrams extracted from ONZE [26] along with counts of how frequently each occurred in the corpus. Twenty frequent bigrams, including *huge pain*, *free time*, *large farm* and *main store*, were selected. A further 10 were created by pairing high frequency adjectives and high frequency nouns from the ONZE list, e.g. *warm tea* and *brown shoe*. All adjectives and nouns were monosyllabic and all contained long vowels and voiced codas in order to optimise pitch information available to the listeners throughout the duration of the stimuli.

Stimuli materials were recorded using a Sennheiser 416 microphone and Universal Audio Apollo Quad interface with a preamp at 44.1 kHz sampling rate into Logic Pro on a MacBook Pro computer. Speakers were asked to produce each phrase in modal voice with neutral intonation. In order to ensure the degree of pitch manipulation was similar across conditions, speakers were instructed to aim for target f0s. The female speaker was asked to produce the phrases at a target f0 of 175 Hz (mean across stimuli = 169 Hz, range = 163–177 Hz) and the male speaker was asked to aim for an f0 of 125 Hz (mean across stimuli = 122, range = 115–125 Hz). The male speaker is musically trained and coached the female speaker in producing the stimuli. Target f0s were approximated with reference to the the musical notes F3 for the female and B2 for the male. These are close approximates to 175 Hz and 125 Hz respectively. These targets were determined so the starting points of the f0 manipulations would be proportionate to the source f0 for each condition (discussed below).

F0 was manipulated in Praat [27] to create both modal and creaky pairs of each adjective-noun phrase (with creak in the final 40% of the noun in the creak condition). While we investigated various methods for creating creaky voice in our stimuli, we determined that manipulating f0 produced the most natural sounding creak. Firstly, two different pitch conditions were created per source voice (low male, mid male; mid female and high female). All the different pitch condition f0 manipulations were calculated proportionately. In lower pitch conditions (mid female and low male), f0 at the start of the adjective (first word in each phrase) was manipulated to be 85.6% of the source target. In the higher pitch conditions (high female and mid male), f0 at the start of the adjective was manipulated to be 120% of the source target. As a result, start f0 for the high female condition

was 210 Hz (120% of 175 Hz) and the start f0 for the low male condition was 107 Hz (85.6% of 125 Hz). These f0 values were chosen to reflect the mean f0s of 18–29 year old AusE speaking females and males found by [28]. F0 start for the mid male and mid female was 150 Hz.

All phrases were manipulated to have a gradual declination. In all conditions, the f0 at the end of the adjective/start of the noun rhyme was 93.3% of the f0 at the start of the adjective. In the modal condition, nouns (the second word in each phrase) were manipulated to have a gradual declination following from the adjective. The f0 at the end of the noun was 80% of the f0 at the end of the adjective/start of the noun rhyme. Mean f0s in the modal condition were 97 Hz in the low male condition, 135 Hz in the mid male and female conditions and 190 Hz in the high female condition. In the creaky condition, nouns were manipulated to have identical creaky voice starting 40% through the rhyme as follows: 70 Hz at 40%, 60 Hz at 50%, 50 Hz at 65%, 70 Hz at 75%, 60 Hz at 90% and 50 Hz at 100% of the rhyme duration. Manipulating f0 in this way created the percept of prototypical creaky voice with low and irregular f0 [2].

These manipulations resulted in 240 stimuli (30 phrases x 8 conditions: 2 x voice quality conditions - modal vs creak x 4 pitch conditions - low male, mid male, mid female and high female). Table 2 shows the f0 manipulations for each of the conditions. It shows that in the mid conditions, f0 values were identical for male and female source recordings.

Condition	Adjective		Noun	
	Start	Rhyme end	Rhyme start	Rhyme end
Modal lo-m	107 Hz	100 Hz	100 Hz	80 Hz
Modal mid-m	150 Hz	140 Hz	140 Hz	112 Hz
Modal mid-f	150 Hz	140 Hz	140 Hz	112 Hz
Modal hi-f	210 Hz	196 Hz	196 Hz	157 Hz
Creaky lo-m	107 Hz	100 Hz	100 Hz	creak
Creaky mid-m	150 Hz	140 Hz	140 Hz	creak
Creaky mid-f	150 Hz	140 Hz	140 Hz	creak
Creaky hi-f	210 Hz	196 Hz	196 Hz	creak

Table 2: *F0 manipulations of adjective and noun pairs by condition.*

2.2. Participants

Listeners were 130 AusE-speaking females who reported completing all of their schooling in Australia and no history of hearing loss. The mean listener age was 21 (range: 17–55) and they were all undergraduate students of linguistics or psychology. Listeners were compensated with either course credit or a $20 supermarket voucher for their time.

2.3. Procedure

The experiment was built in PsychoPy [29] and run online via Pavlovia.org. Participants were instructed to be seated in a quiet environment wearing headphones. Prior to starting the task, participants were provided with examples of creaky voice in male and female AusE voices, different to those used in the experiment. Examples were natural creak but followed the same structure as the stimuli (adjective-noun bigrams with creak on the noun). Additionally, participants heard eight practice items (one in each condition), in different male and female AusE voices. No feedback was provided during the practice. Participants then proceeded to the main experiment. For each trial participants were presented with a fixation cross for 500 ms, followed by the orthography of the phrase. After another 500 ms, the sound file played. Participants were asked to respond with the M and Z keys on their keyboard for whether they heard creak in the phrase or not, as quickly and as accurately as possible. Each participant was presented with all 240 stimuli across eight pseudo-randomised blocks of 30. Ten second breaks were provided between blocks.

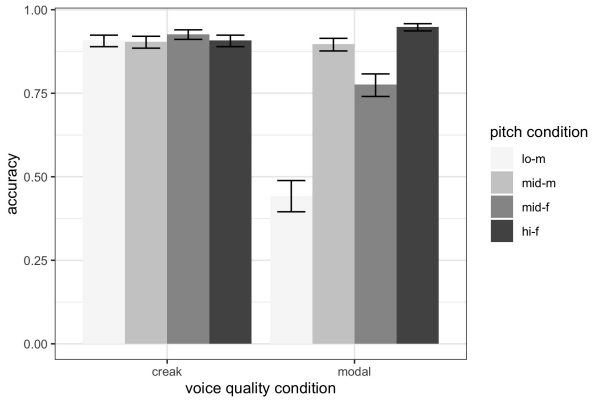

Figure 1: *Predicted accuracy results of generalised linear mixed effects model of each pitch condition (creaky voice condition and modal voice quality condition). Higher accuracy in the creak condition means more "creak" responses while higher accuracy in the modal condition means more "no creak" responses.*

3. Results

We limited our analysis to responses given after the offset of the adjective as it was the noun that contained the variable of interest (i.e. presence or absence of creak). We excluded responses that were quicker than 210 ms from the offset of the adjective or if they exceeded two standard deviations either side of the individual participant's mean response time. This resulted in the exclusion of 6.4% of the data leaving us with 29,213 data points for analysis.

A generalised linear mixed effects model was run using the lme4 package [30] in R [31] to investigate how listeners' accuracy at identifying creak presence or absence was affected by pitch and speaker sex in the different voice quality conditions (i.e. creaky and modal). Accuracy was included as the dependent variable and the independent variables were an interaction between four-level factor pitch condition (low male, mid male, mid female or high female) and two-level factor voice quality condition (creaky or modal), and scaled trial order with listener as a random intercept. We found a significant effect of trial order with listeners becoming more accurate throughout the duration of the experiment. The interaction between pitch condition and voice quality condition was also significant; model predictions are presented in Figure 1.

We conducted *post hoc* pairwise comparisons using the emmeans package [32] in R to further explore the interaction between pitch and voice quality. Results showed that in the creak condition, listeners were significantly more accurate at identifying creak presence in the mid female condition than in any other pitch conditions (low male: $p<0.05$; mid male: $p<0.01$; high

female: $p<0.05$). There were no other significant differences in the creak condition. In the modal condition, all pitch conditions were significantly different from each other (all $p<0.001$). Listeners were most accurate at identifying creak absence in the high female condition, followed by the mid male condition, the mid female condition and finally the low male condition. That is, they identified creak when it was not present most often in the low male stimuli and least often in the high female stimuli.

4. Discussion

As in previous literature [25, 33], our results show that listeners frequently false alarmed in the low male condition. In other words, they were more likely to identify creak in the low male condition when it was not present than for other conditions. However, we also see that the mid female condition also triggered false alarms when compared to the mid male and high female condition. Interestingly, listeners were significantly more accurate in identifying lack of creaky voice in the mid male condition despite the f0 manipulations being identical between the mid male and mid female conditions. These results point to listeners using both pitch and speaker sex as cues to identifying creaky voice: low pitch is the dominant cue triggering creak responses in the extreme pitch conditions regardless of speaker sex (i.e. in the modal condition the low male triggered inaccurate responses, but the high female triggered accurate responses). However, when the f0 is the same but speaker sex is different, the female source voice triggers creak responses.

Responses were at near ceiling accuracy in the creak condition. This tells us that listeners in this study were very good at identifying creak when it was present. Accuracy was significantly higher in the mid female creak condition compared with the other pitch conditions. We did not predict in any scenario that mid female stimuli would be rated more accurately than high female stimuli. We suggest that this may be due to a boost in the likelihood of identifying creak when pitch is low given a listener's expectations for sex.

Listeners may be making their judgments at least partially on the basis of the speaker's modal voice pitch (in this case, the adjective of the pair). This is highlighted in a comment left by one of the participants when asked if they noticed anything about the voices in the study:

> "The tone of the first word kind of gave away whether the second word would have creaky voice or not."

This is despite the adjective f0s being identical between the creaky and modal conditions. We suggest that, overall, when listeners hear a voice that is 'low for a female', or 'low for a male', they are more inclined to respond that they hear creak. In the creak condition, this would result in higher accuracy for the voices that sound low in the context of the speaker's sex (mid female and low male). In the modal condition, this boost would lead to false alarms for these speakers.

Even if listeners' judgments are affected by the pitch of the first word, it is clear that they are also sensitive to the creak manipulation. When listeners hear the creaky versions of the relatively high voices (high female and mid male), the difference in pitch between the initial higher modal voice and low creak leads to high rates of correct positive identifications of creak. In the modal condition, the absence of creak for these voices that sound relatively high given their sex is more obvious due to the lack of any dramatic drop in pitch. In future work, we will look at the reaction times to assess these possibilities. We may see faster responses where listeners' judgment relies on the pitch of the adjective.

If there is an overall boost in creak responses to the relatively low voices, we would expect to see greater accuracy to the low male voice in the creak condition. This may be absent in the present accuracy results due to the near ceiling effect. An analysis of reaction times may provide further clarity.

An alternative explanation for this over-identification of creak in the low male and mid female conditions could be based in experiences of hearing creak more often in low-pitched male and low-pitched female voices. A preliminary analysis of speaker creak prevalence by mean modal voice f0 suggests that creak may be more prevalent in male and female voices with lower mean modal f0s. The design of the present perception study does not enable us to determine whether the results in the modal condition are related to the association of creak with low pitch or the association of low-pitched voices using more creaky voice. Future research is needed to explore this from both a production and perception perspective.

It is possible that listener age impacts creak identification. Although the age range of listeners in the present study was quite large, the vast majority of listeners were less than 25 years old. The effects of listener age on creak identification would be an interesting area for future work.

5. Conclusions

This study set out to test how pitch and speaker sex influence female AusE-speaking females' identification of creaky voice. Results suggest that both pitch and speaker sex have an interrelated influence on creak identification by these listeners: for both female and male speakers, creak was over-identified in the lower-pitched condition. While previous work has not found solid evidence for sex and pitch influences [25], through the use of highly controlled manipulated stimuli, we have been able to tease apart these cues.

Findings from this study have implications for studies on the prevalence of creaky voice in speech. Many of these studies rely on manual annotation for identifying creak [19], which, as well as being a time-consuming and labour-intensive process, may lead to the over-identification of creak in lower-pitched male and female voices. There is a lot of recent research being conducted in the area of automatic creak detection methods, which may mitigate the biases of manual creak identification [19, 34, 35, 36, 37, 38].

In our future work we plan to investigate how male listener creaky voice identification compares to that of females. We also plan to explore whether reaction time data can shed further light on the role that speaker sex and pitch play in listeners' identification of creak.

6. Acknowledgements

This research was supported by a Macquarie University Research Excellence Scholarship to the first author and by Australian Research Council Grant DP190102164 and Australian Research Council Future Fellowship Grant FT180100462 to the fifth author. Thanks to members of the MQ phonetics lab, NZILBB lab and VUW Department of Linguistics for helpful discussions on this work.

7. References

[1] M. Garellek, "The phonetics of voice," in *The Routledge Handbook of Phonetics*, W. Katz and P. Assmann, Eds. Routledge,

2019.

[2] P. Keating, M. Garellek, and J. Kreiman, "Acoustic properties of different kinds of creaky voice," in *Proceedings of the 18th International Congress of Phonetic Sciences*. the University of Glasgow, 2015, Conference Proceedings, pp. 821.1–5.

[3] J. Peña, L. Davidson, and S. Orosco, "The independence of phrasal creak and segmental glottalization in American English," *JASA Express Letters*, vol. 1, p. 075205, 2021.

[4] L. Redi and S. Shattuck-Hufnagel, "Variation in the realization of glottalization in normal speakers," *Journal of Phonetics*, vol. 29, no. 4, pp. 407–429, 2001.

[5] L. Davidson, "Perceptual coherence of creaky voice qualities," in *Proceedings of the 19th International Congress of Phonetic Sciences*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds. Australasian Speech Science and Technology Association Inc, 2019, pp. 196.1–5.

[6] ——, "Contributions of modal and creaky voice to the perception of habitual pitch," *Language*, vol. 96, no. 1, pp. e22–e37, 2020.

[7] B. Ralston, "Vocal fry is not okay, eh? - the listener," *The New Zealand Listener*, 2012. [Online]. Available: https://www.noted.co.nz/life/life-in-nz/vocal-fry-is-not-okay-eh/

[8] M. M. Weber. (2017) Top five most annoying vocal habits. [Online]. Available: https://www.voiceempowerment.com/voice-empowerment-blog/2017/5/1/ten-most-annoying-vocal-habits-or-5

[9] N. Wolf, "Young women, give up the vocal fry and reclaim your strong female voice," *The Guardian*, 2015. [Online]. Available: https://www.theguardian.com/commentisfree/2015/jul/24/vocal-fry-strong-female-voice

[10] A. Croffey, "Vocal fry: Women changing voices to sound "creaky" like Kim Kardashian," *The Sydney Morning Herald*, 2016. [Online]. Available: https://www.smh.com.au/lifestyle/vocal-fry-women-changing-voices-to-sound-creaky-like-kim-kardashian-20160428-gogwvo.html

[11] R. C. Anderson, C. A. Klofstad, W. J. Mayew, and M. Venkatachalam, "Vocal fry may undermine the success of young women in the labor market," *PLoS ONE*, vol. 9, no. 5, p. e97506, 2014.

[12] J. Pittam, "Listeners' evaluations of voice quality in Australian English speakers," *Language and Speech*, vol. 30, no. 2, pp. 99–113, 1987.

[13] I. P. Yuasa, "Creaky Voice: A New Feminine Voice Quality for Young Urban-Oriented Upwardly Mobile American Women?" *American Speech*, vol. 85, no. 3, pp. 315–337, 2010.

[14] C. Gobl and A. Ní Chasaide, "The role of voice quality in communicating emotion, mood and attitude," *Speech Communication*, vol. 40, pp. 189–212, 2003.

[15] N. B. Abdelli-Beruh, L. Wolk, and D. Slavin, "Prevalence of vocal fry in young adult male American English speakers," *Journal of Voice*, vol. 28, no. 2, pp. 185–190, 2014.

[16] S. Melvin and C. G. Clopper, "Gender variation in creaky voice and fundamental frequency," in *Proceedings of the 18th International Congress of Phonetic Sciences*, T. S. C. for ICPhS 2015, Ed. the University of Glasgow, 2015, pp. 1–5 (Paper number 320). [Online]. Available: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0320.pdf

[17] R. J. Podesva, "Gender and the social meaning of non-modal phonation types," *Annual Meeting of the Berkeley Linguistics Society*, vol. 37, no. 1, pp. 427–448, 2013.

[18] L. Wolk, N. B. Abdelli-Beruh, and D. Slavin, "Habitual use of vocal fry in young adult female speakers," *Journal of Voice*, vol. 26, no. 3, pp. e111–e116, 2012.

[19] K. Dallaston and G. Docherty, "The quantitative prevalence of creaky voice (vocal fry) in varieties of English: A systematic review of the literature," *PloS ONE*, vol. 15, no. 3, p. e0229960, 2020.

[20] B. Gittelson, A. Leemann, and F. Tomaschek, "Using Crowd-Sourced Speech Data to Study Socially Constrained Variation in Nonmodal Phonation," *Frontiers in Artificial Intelligence*, vol. 3, p. 565682, 2021.

[21] C. Henton and A. Bladon, "Creak as a sociophonetic marker," in *Language, speech, and mind: Studies in honour of Victoria A. Fromkin*, L. M. Hyman and C. N. Li, Eds. Routledge, 1988, pp. 3–29.

[22] J. Stuart-Smith, "Glasgow: Accent and voice quality," in *Urban voices: Accent studies in the British Isles*, P. Foulkes and G. Docherty, Eds. Arnold, 1999, pp. 203–222.

[23] D. Loakes and A. Gregory, "Voice quality in Australian English," *JASA Express Letters*, vol. 2, no. 8, p. 085201, 2022.

[24] N. Abdelli-Beruh, T. Drugman, and R. H. Red Owl, "Occurence frequencies of acoustic patterns of vocal fry in American English speakers," *Journal of Voice*, vol. 30, no. 6, pp. 759.e711–759.e720, 2016.

[25] L. Davidson, "The effects of pitch, gender, and prosodic context on the identification of creaky voice," *Phonetica*, vol. 76, no. 4, pp. 235–262, 2019.

[26] E. Gordon, M. Maclagan, and J. Hay, "The ONZE Corpus," in *Creating and digitizing language corpora*, J. C. Beal, K. P. Corrigan, and H. L. Moisl, Eds. Palgrave Macmillan, 2007, vol. 2, ch. 4, pp. 82–104.

[27] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," 2018, [Computer program]. [Online]. Available: http://www.praat.org/

[28] Y. Leung, J. Oates, V. Papp, and S.-P. Chan, "Speaking fundamental frequencies of adult speakers of Australian English and effects of sex, age, and geographical location," *Journal of Voice*, vol. 36, no. 3, pp. 434.e1–434.e15, 2022.

[29] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv, "Psychopy2: Experiments in behavior made easy," *Behavior Research Methods*, vol. 51, pp. 195–203, 2019.

[30] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[31] R Core Team, "R: A language and environment for statistical computing," 2021. [Online]. Available: https://www.R-project.org/

[32] R. Lenth, "emmeans: Estimated marginal means, aka least-squares means," 2018. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[33] A. Li, W. Lai, and J. Kuang, "How do listeners identify creak? the effects of pitch range, prosodic position and creak locality in Mandarin," in *Proceedings of Speech Prosody 2022*, 2022, pp. 480–484.

[34] K. Dallaston and G. Docherty, "Estimating the prevalence of creaky voice: A fundamental frequency-based approach," in *Proceedings of the 19th International Congress of Phonetic Sciences*, S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Eds. Australasian Speech Science and Technology Association Inc, 2019, pp. 581.1–5.

[35] T. Drugman, J. Kane, and C. Gobl, "Data-driven detection and analysis of the patterns of creaky voice," *Computer Speech and Language*, vol. 28, no. 5, pp. 1233–1253, 2014.

[36] O. Murton, S. Shattuck-Hufnagel, J.-Y. Choi, and D. D. Mehta, "Identifying a creak probability threshold for an irregular pitch period detection algorithm," *The Journal of the Acoustical Society of America*, vol. 145, no. 5, pp. EL379–EL385, 2019.

[37] J. Villegas, K. Markov, J. Perkins, and S. J. Lee, "Prediction of creaky speech by recurrent neural networks using psychoacoustic roughness," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 355–366, 2020.

[38] H. White, J. Penney, A. Gibson, A. Szakay, and F. Cox, "Evaluating automatic creaky voice detection methods," *The Journal of the Acoustical Society of America*, vol. 152, no. 3, pp. 1476–1486, 2022.