

Social and acoustic factors in the perception of creak

Amy Hemmeter

North Carolina State University

October 24th, 2015

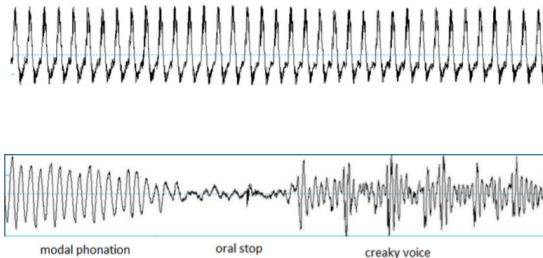
New Ways of Analyzing Variation 2015

Toronto, Ontario, Canada

What is creak?

- ▶ Creak is a phonation type in which the arytenoid cartilages are held tightly together, so that the vocal folds can vibrate only at the anterior end (Ladefoged 2010, p. 150)
- ▶ Creak is characterized by low frequency and damped glottal pulses (Epstein 2002).
- ▶ Creak generally occurs in the last 1-2 syllables of an utterance, cross-linguistically

- ▶ Compare the following [i] vowel which I produced with modal voicing on Praat in Figure 1 with an example of utterance-final creak from production data I collected in MI in 2013 in Figure 2. Both are 156.55 milliseconds in duration (Boersma and Weenink, 2013)



Who uses creak?

- ▶ The short answer is: everyone
- ▶ In older research, creak was associated with masculinity, and thought to be a robust marker of male speech (Henton and Bladon, 1988, p. 21).
- ▶ In the past five years, however, there have been a fairly large number of studies linking women to creak usage in the US (Yuasa, 2010; Podesva 2011; Wolk et al, 2011; Anderson and Nguyen, 2004; Mendoza-Denton, 2011).
- ▶ My research from Michigan found that men and women use creak equally often in sociolinguistic interviews, but that women use it more often in reading passages

Perception of creak: acoustic considerations

- ▶ McGlone (1967) found a negligible difference in F0 range between men and women when using creak, despite the fact that in modal vowels the F0 differs quite widely.
- ▶ Men generally have an F0 of 100 to 140 Hz (Hollien and Jackson, 1973; Krook, 1988), while women's F0 usually falls between 175 and 240 Hz (Stoicheff, 1981; Krook, 1988).
- ▶ The difference between modal and creak F0 is much larger for women than for men; listeners may attend to this difference when perceiving creak in women's voices, acoustically

Research questions

Is the association of creak with women due to acoustic or to social cues? Is it harder for listeners to hear creak from "male" speakers than from "female" speakers with gender-ambiguous F0?

Overview

- ▶ After a training session on recognizing creak, listeners were asked to listen to utterances and determine whether creak was used in those utterances
- ▶ Utterances were paired with photos of faces to prime listeners to expect the stimulus voice to be of a certain gender
- ▶ 20 participants between the ages of 18 and 24

| | Normative Female Voice | Gender Ambiguous Voice | Male normative voice |
|-------------|------------------------|------------------------|----------------------|
| Male face | | X | X |
| Female face | X | X | |

Preparation: Gender Ambiguous Stimuli

- ▶ I performed a perception experiment to ensure that the ambiguous voices were actually ambiguous - speakers can still be perceived as male at high pitches and female at low pitches; listeners are attending to other acoustic cues as well (Zimman 2013)
- ▶ Target stimuli came from two tall women (~6 feet tall each) to account for some of the acoustic consequences of sexual dimorphism in humans (namely, formant values and vocal tract length), whose pitch I altered via synthesis
 - ▶ Shorter men don't sound much more "feminine" than tall ones, possibly because it's more socially costly to be perceived as a feminine-sounding short man than to be perceived as a masculine-sounding tall woman

Preparation: Gender Ambiguous Stimuli

- ▶ In addition to the target stimuli from the two tall women, I had filler utterances from two normative men and two normative women
- ▶ Utterances were taken from real sociolinguistic interview data with speakers in Michigan
- ▶ All utterances were originally produced with creak in the final 1-2 syllables of the utterance

Synthesis: gender-ambiguous stimuli

- ▶ I lowered the speakers' pitches as far as they could go without the appearance of creak-like phonation
- ▶ I performed a perception experiment with just the audio of the pitch-altered stimuli to participants, along with the unaltered gender-normative stimuli
- ▶ Listeners believed that the speakers were male 27.5% of the time
- ▶ Linguists know from the McGurk Effect that (McGurk & McDonald 1976) that perception is multi-modal, and therefore a 27.5% male judgment rate is likely enough to sway listeners to expect a certain gender with the addition of gender-normative photos

Synthesis: gender-ambiguous stimuli

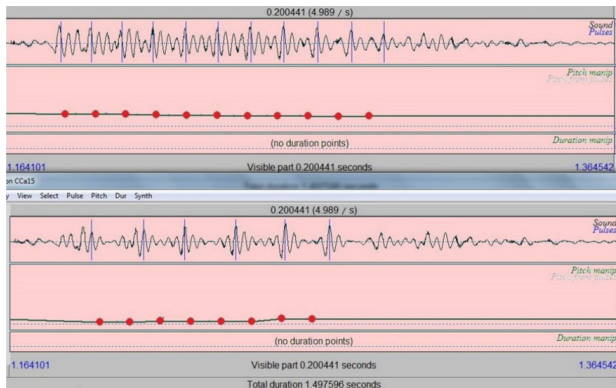
Pre (Tall Woman 1)

Post (Tall Woman 1)

Pre (Tall Woman 2)

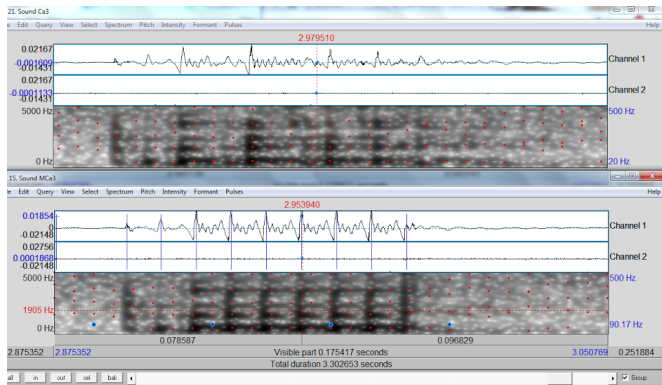
Post (Tall Woman 2)

Synthesis: creaky voice



Original modal recording
Synthesized creaky voice

Synthesis: modal voice



Original creaky recording
Synthesized modal voicing

Experiment design

- ▶ Experiment was designed using PsychoPy
- ▶ Two parts: a training portion, and the “real” experiment
- ▶ Participants were presented with an audio stimulus (either a gender normative female voice, a gender ambiguous voice, or a normative male voice) and a photo stimulus (either a male face or a female face) to prime for gender (the photos were only used in the experiment, not the training)
- ▶ Photo stimuli came from the Face Place database, from the Tarrlab at Brown University (now at Carnegie Mellon)
- ▶ Forced choice: could answer “creak” or “no creak”
- ▶ Data was collected on accuracy and reaction time

Training

- ▶ Creak was explained orally, first with an extended vocalization with creaky voice, then in a sentence without creak and a sentence with creak in utterance-final position
- ▶ Participants were presented with twelve unaltered utterances, all taken from readings of the Rainbow Passage by Michigan speakers in 2013
- ▶ Participants responded with either “creak” or “no creak,” and were presented with feedback, either “correct” or “incorrect”
- ▶ Participants got three chances to pass the training before they were told they could not participate in the rest of the experiment
- ▶ One-third of participants failed the training

Experiment



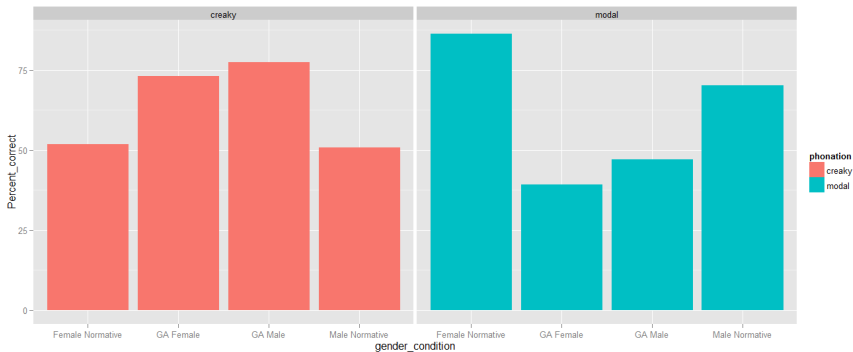
CREAK

NO CREAK

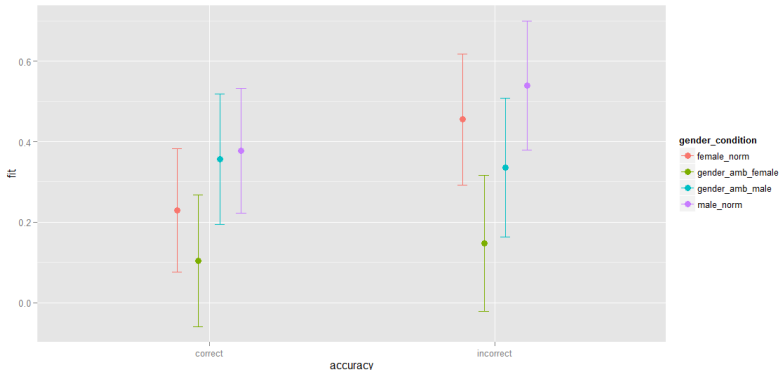
Statistical methods

- ▶ Reaction time: Mixed effects linear regression model
 - ▶ Independent variables: an interaction between gender condition and accuracy, plus phonation and random intercepts of participant
- ▶ Accuracy: Mixed effects logistic regression
 - ▶ Independent variables: an interaction between gender condition and phonation plus random intercepts of participant

Results: Accuracy



Results: Reaction Time



Accuracy

- ▶ From these results, it seems as though there are some effects from social expectations, but the big split seems to be between normative and non-normative voices
- ▶ Listeners seem to have a bias towards creak in the non-normative voices
- ▶ More accurate in the creaky conditions with non-normative voices
- ▶ Requires more sophisticated methodology in synthesizing “gender ambiguous” voices

Reaction Time

- ▶ Participants' reaction times were slower for normative stimuli overall, but especially when they were wrong
- ▶ Gender priming seems more important in reaction time than in accuracy
 - ▶ Listeners take slightly longer to respond with male faces in the gender ambiguous condition than with female faces, especially when they are correct
- ▶ Reaction time is fastest with the gender ambiguous female condition – perhaps because a lower pitched voice coming from a female face tends more often to be created in natural speech

Conclusion

- ▶ There are some social differences because listeners treat gender ambiguous voices with female photos differently than they do gender ambiguous voices with male photos
- ▶ Acoustic differences are quite small, and the split between normative and non-normative voices is bigger both in reaction time and accuracy; listeners are sensitive to altered voices
- ▶ This study raises methodological questions about the creation of gender-ambiguous stimuli – more needs to be taken into account than vocal tract size and pitch

References

- Epstein, M. A. (2002). Voice quality and prosody in English (Doctoral dissertation, University of California Los Angeles).
- Henton, C. G. (1986). Creak as a sociophonetic marker. *The Journal of the Acoustical Society of America*, 80(S1), S50-S50.
- McGlone, R. E. (1967). Air flow during vocal fry phonation. *Journal of Speech, Language and Hearing Research*, 10(2), 299.
- 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.

References

- Podesva, R. J. (2011). Gender and the social meaning of non-modal phonation types. In 37th annual meeting of the Berkeley Linguistics Society (pp. 12-13).
- Wolk, L., Abdelli-Beruh, N. B., & Slavin, D. (2012). Habitual Use of Vocal Fry in Young Adult Female Speakers. *Journal of Voice*, 26(3), e111-e116.
- Yuasa, Ikuko Patricia (2010). Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? *American Speech* 85(3):315-337.
- Zimman, Lal (2013, January). Perceived gender in context: gendered style among transgender speakers. Paper presented at the Linguistic Society of America Annual Meeting, Boston, MA.

Acknowledgements

Thanks to Robin Dodsworth, Erik Thomas and Jeff Mielke for advising me on this experiment. Thanks also to Cadwell Turnbull, Karissa Wojcik, Angela Tramontelli Stephanie Gardner, Laura Barnes and Kellam Barta for lending me their voices and finally thanks to Mike Fox and Jon Forrest for helping me with the statistics.