PERCEPTUAL COHERENCE OF CREAKY VOICE QUALITIES

Lisa Davidson

New York University
lisa.davidson@nyu.edu

ABSTRACT

This study examines the acoustic properties of two types of creaky voice in English, and investigates whether listeners identify them equally as implementations of creaky voice. The two types are prototypical creaky voice (low, semi-regular F0 and damping between glottal pulses), and multiply-pulsed creaky voice (glottal openings that alternate in higher and lower amplitude). Analyses of F0, HNR, H1-H2 and subharmonic-to-harmonic ratio (SNR) indicate that modal voice differs from both prototypical and multiply pulsed creaky voice for all cues except SNR. However, none of these cues consistently distinguish the two types of creak. The perceptual study shows that listeners identify creak equally often for both types relative to modal voice. Listeners may be primarily sensitive to the cues that consistently encompass both creaky types, leading them to be treated as a perceptually coherent group.

Keywords: phonation, creaky voice, voice quality, speech perception

1. INTRODUCTION

Within the linguistic phonetic literature on voice quality, a number of studies have observed that the quality often referred to as “creaky voice”, “creak” or “laryngealization” is actually a collection of phonation properties that seem to pattern together. One way these different types of creaky voice are observed to act is as a prosodic element (e.g. marking phrase endings), in languages like English, German, Finnish, Swedish or Chinese [1-7].

Although different authors do not always use the same labels for types of creaky voice, there is general agreement about what the types are. Nearly all of these studies include the most canonical form of creaky voice (called ‘prototypical creaky voice’ by [8]), with a low F0, strong damping between glottal pulses, and semi-regular periods [9, 10]. Another type is ‘aperiodic’, characterized by glottal pulses that vary considerably in both frequency and duration such that almost no periodicity is evident [8, 11]. A third type is ‘multiply-pulsed’ creak, sometimes called diplophonia, which is characterized by glottal openings that alternate in higher and lower amplitude, with the entire duration between higher amplitudes being semi-periodic [8, 12, 13]. ‘Non-constricted creak’ has also been observed, in which F0 is low and irregular, but these properties are accompanied by glottal spreading and higher airflow, not constriction [4].

Keating et al. [8] summarize the acoustic properties that are expected for each type based on the literature and their own measurements. Two of their predictions are examined in the stimuli used for the current study: prototypical and multiply pulsed creaky voice. Table 1 is based on their expectations for the acoustic properties that will be present and/or distinguish between these two types of creak. The measures they consider include F0, harmonics-to-noise ratio (HNR, higher values indicate greater noise and lower periodicity in the signal), H1-H2 (the difference in the amplitude of the first and second harmonics: a measure of glottal constriction, with lower numbers indicating greater constriction), and subharmonic-to-harmonic ratio [14] (SHR values should be higher for multiply pulsed creak, which should have more subharmonics).

Table 1: Acoustic properties expected to characterize types of creak. No checkmark means this property is variable or unknown.

Acoustic correlate	low F0	irreg F0, high HNR	low H1-H2	high SHR
prototypical	✓	✓	✓	
multiply pulsed		✓	✓	✓

While studies have reported that several types of phonation fall under the umbrella of “creaky voice”, fewer studies examine whether listeners consider all of these types equal representations of the category of (non-pathological) creaky voice, or whether some types are more representative of this voice quality (but see [15]). One study that presented voices as disordered found that listeners could distinguish between the qualities of multiply pulsed, amplitude modulated, and noisy/aperiodic voices, using a similarity scale from 1 to 7 [12]. This suggests that listeners can attend to differences in acoustic properties that have been attributed to creaky voice. However, it is unclear whether listeners similarly distinguish acoustic information about creak when it is linguistic and not disordered. While it is predicted that listeners will consider both prototypical creaky and multiply pulsed voice as creaky compared to modal voice, it is possible that one type will be treated as a more archetypal by listeners and will therefore be more frequently labelled as an instance of creaky voice. Alternatively, listeners may treat both types as a coherent perceptual grouping of creaky voice.

This study has 3 goals: (1) to acoustically determine whether Keating et al's. [8] acoustic predictions distinguish the creak types in this study as shown in Table 1, (2) to determine whether listeners accurately identify the presence of creak in either fully or partially creaky utterances, and (3) to investigate whether listeners are equally likely to identify prototypical creaky voice and multiply pulsed voice as exemplars of creak.

2. METHODS

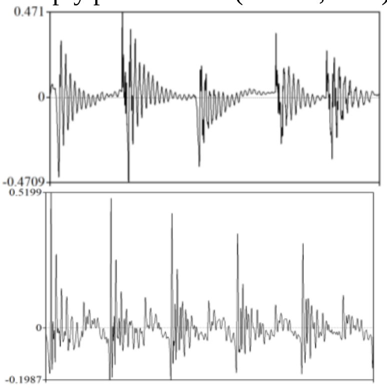
2.1. Stimuli

The talkers for the stimuli were four female podcast hosts who record in professional settings, two with a high average modal pitch (Hi-1, 2: ~195Hz) and two with a lower average (Lo 1, 2: ~150Hz) (see Figure 2). The modal F0 differences were included to examine whether potential acoustic differences distinguishing type of creak could vary depending on the height of a speaker's modal F0.

Stimuli were 3-4 word phrases taken from the ends of sentences (mean dur = 909ms; e.g., 'many months ago', 'kids one day'). There were 3 categories for voice quality: modal, prototypical (proto) creak, and multiply pulsed (mult) creak. The two creak types also interacted with creak duration: whole utterance, or partial creak, in which creak was produced on the second 50% of the utterance (e.g. "my older f̰r̰ḭḛn̰d̰s̰ a̰r̰ḛ"). Partial creak is included because it may be that participants are at ceiling in the whole creak conditions, so potential differences between mult and proto creak might be more likely to emerge in the partial creak condition. Listeners heard 5 utterance types: fully modal, fully proto creaky, fully mult creaky, partially proto creaky and partially mult creaky. There were 3 tokens for each of the creak types and 6 for the modal type. All phrases had final falling F0 and neutral semantic content.

The creak types were identified by visual inspection of waveforms and spectrograms in Praat [16]. Prototypical and multiply pulsed creak were chosen because they were the most common two types found in the podcasters' speech and are visually distinct. Proto creak is characterized by a low and irregular F0, with a long closed phase. Mult creak contains two sets of glottal openings which alternate regularly in amplitude and length (see Fig 1). The creak type stayed relatively consistent throughout the short utterances.

Figure 1: Example of prototypical creak (top, Lo-1) and multiply pulsed creak (bottom, Hi-2)



2.2. Participants and procedure

Participants were recruited on Amazon Mechanical Turk (N=54, 29M, 25F). In training, listeners were provided with text descriptions of creaky voice and with 2 audio examples each of partial prototypical (labelled as creaky) and modal phonation. In the test, they were told to determine (yes/no) whether audio phrases contained creaky voice anywhere in the file.

3. ACOUSTIC ANALYSIS OF STIMULI

To determine whether the acoustic properties of the modal, proto and mult stimuli conform to the hypotheses in [8], VoiceSauce [17] was used to measure the properties discussed in their paper: F0, H1*-H2* ('*' indicates a correction for formant values), HNR<3500Hz, and SHR. The STRAIGHT algorithm was used for F0. Because there were different phonemes in each file and many consonants are not appropriate for voice quality measurements, measurements were taken over one vowel in the modal and creaky portions of each file in order to maintain as much consistency as possible.

The results of the measurements are illustrated in Figure 2. For each acoustic measure, a linear mixed effects model of speaker*quality (modal, proto, mult) with token as a random intercept was carried out in R, and for almost every measure, all of the main effects and interactions were significant at p<.01 (with the exception of the interactions between mult and each low pitched speaker for SHR, and main effects of speaker for H1*-H2*, which are not significant.) These results were further investigated by using a linear model to compare voice quality within the individual speakers.

Results for F0 show significant differences for all voice quality types for all 4 speakers. In all cases, modal F0 is significantly higher than F0 for both creaky types (p < .001). Within the creak types, F0 is higher for mult than for proto creak for Hi-1 (β=28.86, z=12.0, p<.001) but lower for Lo-2 (β= -33.19, z=-17.5, p<.001). There were no significant differences for Hi-2 or Lo-1.

Results for HNR < 3500Hz show that HNR for the modal quality is significantly higher than for either of the two creaky qualities (p < .001). For Hi-1, HNR is significantly higher for mult than for proto (β=1.99, z=3.88, p<.001), but the opposite is true for the other 3 speakers (Hi-2: β=-17.42, z=-24.25, p<.001, Lo-1: β=-8.96, z=-15.9, p<.001, Lo-2: β=-4.31, z=-6.52, p<.001).

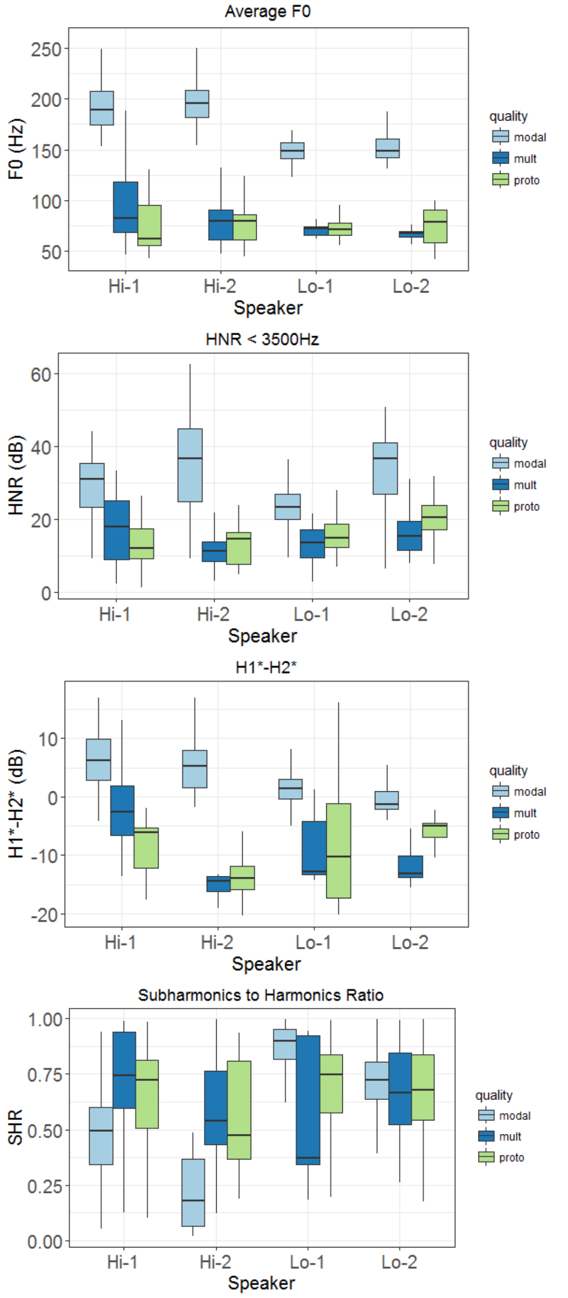
For H1*-H2*, again the modal quality is significantly higher than for either of the creaky qualities for all speakers (p < .001). For Hi-1, proto is significantly lower than mult (β=-4.9, z=-8.86, p<.001), while for Hi-2 and Lo-2, proto is significantly higher than mult (Hi-2: β=8.92, z=-14.07, p<.001, Lo-2: β=13.07, z=13.70, p<.001). There is no significant difference between proto and mult for Lo-1 (β=.35, z=.76).

For SHR, the results are more complicated. For Hi-1, the only significant difference is that modal is lower than mult (β=-.37, z=-2.29, p=.05). For Hi-2, modal is lower than both mult (β=-.35, z=-3.87, p<.001) and proto (β=-.34, z=-3.46, p<.002). For both low-pitched speakers, modal is higher (contra predictions) than proto (Lo-1: β=.32, z=3.78, p<.001, Lo-2: β=.11, z=4.23, p<.001). There is no significant difference between proto and mult.

The results for the acoustic measures confirm some of the predictions for the comparison of modal and creaky voice: both prototypical creak and multiply pulsed creak are lower in F0, HNR < 3500Hz, and H1-H2. For F0, Keating et al. reported that the relationship between F0 and multiply pulsed creaky voice was unknown; this study demonstrates that F0 is lower for both types of creak than for modal voice. SHR does not always conform to the prediction that it should be higher for multiply pulsed creak than modal voice. It is true for both high-pitched speakers, but not the low-pitched speakers. This reason for this difference between speakers with higher vs. lower modal F0 is not immediately clear and is an area for future investigation.

In the stimuli, some differences between prototypical and multiply voiced creak were observed, but they do not seem to be systematic across speakers. For each acoustic variable, some speakers show significantly higher values, while others show lower values. This suggests that at least the cues found in these stimuli do not necessarily provide consistent information for listeners to distinguish between different types of creak across all speakers. However, there are reliable cues to distinguish between creaky and modal phonation. In the following perception study, listeners' ability to distinguish between these two creak types, and between creaky vs. modal voice, is examined.

Figure 2: Acoustic properties of the modal, multiply pulsed, and prototypical creak stimuli. From top to bottom: Average F0, HNR<3500Hz, H1*-H2*, SHR.



4. PERCEPTION STUDY

In the perception study, results are reported for how accurate listeners were in indicating that creak was present (for mult and proto creak types, both whole or partial creak), or not present (for modal tokens). The creak types are divided by whole and partial creak, since as hypothesized (and as shown below), listeners could be nearly at ceiling on utterances that were creaky throughout their whole duration.

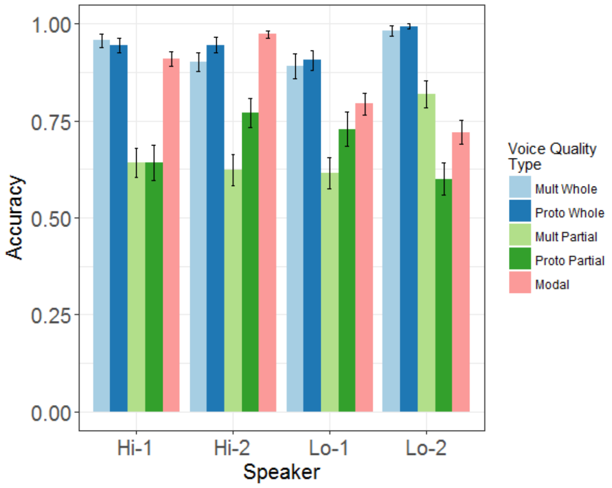
Results are shown in Figure 3. Accuracy (proportion of correct responses) was analysed with a mixed effects logistic regression in R with speaker and type + amount of voice quality (modal, whole proto, partial proto, whole mult, partial mult) as factors (amount of creak was not treated as a separate variable because the modal tokens were not broken down into whole and partial.) These factors were sum coded and stimulus and participant were included as random intercepts.

Results show that overall, listeners are significantly less accurate on Lo-1 (β=-.62, z=-3.35, p<.001) and more accurate on Hi-2 (β=.43, z=2.13, p=.03). Participants were significantly more accurate on both whole creak types and less accurate on both partial creak types (all *p*<.001) (there was no significant result for modal).

Tukey post-hoc tests using *multcomp* [18] provide further relevant comparisons for individual speakers. All speakers are less accurate on both types of partial creak than both types of whole creak (all *p*<.01). For both higher pitched speakers and for Lo-1, there is no difference in accuracy on modal vs. either whole creak condition, but for Lo-2, listeners were less accurate on modal than on either type of whole creak (*p*<.01). Compared to accuracy on modal tokens for higher pitched speakers, the findings suggest that listeners are sometimes false alarming on the modal tokens of the speakers with lower average pitch.

In the comparison between partial proto and mult, where potential differences between the creak types are most likely to emerge, results show that there are only significant differences between proto and mult for two speakers, but the differences are in the opposite direction. For Lo-2, listeners are more likely to identify multiply pulsed creak as creaky than prototypical creak (β=-1.1, z=-4.2, p<.001). For Hi-2, listeners accurately identify prototypical creaky more often (β=.79, z=2.98, p=.02)

Figure 3: Accuracy for modal and both whole and partial multiply pulsed and prototypically creaky utterances. Accuracy for creaky tokens = participants responded yes; accuracy for modal tokens = participants responded no.



5. DISCUSSION

These results indicate that there are consistent acoustic cues that distinguish modal from creaky voice, such as F0, HNR<3500Hz, and H1-H2 [10, 19], but these same cues do not reliably distinguish between prototypical and multiply pulsed creaky voice for all speakers. Moreover, higher vs. lower modal F0 of the speaker does not seem to interact consistently with any acoustic cues to distinguish between proto and mult creaky voice. The absence of a consistently higher SHR for mult as compared to either modal or proto is unexpected based on Keating et al's [8] predictions; further study is necessary to understand the acoustic cues that best characterize multiply pulsed creak.

The perception study outcome is consistent with the acoustic analysis of the stimuli since listeners did not consistently distinguish between mult and proto creak. As predicted, the whole creak condition showed a ceiling effect, as listeners showed 90%+ accuracy on both proto and mult creak. The significantly lower accuracy on partial creak indicates that it is more difficult for listeners to identify creak when it only occurs following modal voice during the last 50% of the utterance. Yet even here, there are only significant differences between proto and mult creak for two speakers, and they are in opposite directions. Since none of the acoustic cues explored here are unique to only these speakers, listeners may be keying into different potential cues that are unexamined here to distinguish partial proto from mult for these two speakers.

These perception results are consistent with the assumption in the literature that despite characteristic differences in how these types of creak present in spectrograms and waveforms, at least prototypical and multiply pulsed creak behave mostly as a coherent category that listeners are equally likely to identify as creaky voice, perhaps because listeners are most sensitive to low F0 and signal decay to identify creaky voice [20]. Future research should also include aperiodic or non-constricted creak to determine whether they also pattern like prototypical and multiply pulsed creak in perception.

6. REFERENCES

[1] Henton, C., and Bladon, A., 1987. Creak as a sociophonetic marker. *Language, speech and mind: studies in honor of Victoria A. Fromkin*, L. Hyman and C. Li, eds., pp. 3-29, London: Routledge.

[2] Kreiman, J., 1982. Perception of sentence and paragraph boundaries in natural conversation. *J. Phon.*, 10, 163-175.

[3] Ogden, R., 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *J. Int. Phon. Assoc.*, 31, 139-152.

[4] Slifka, J., 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice*, 20: 2, 171-186.

[5] Garellek, M., and Seyfarth, S., 2016. Acoustic differences between English /t/ glottalization and phrasal creak. *Proceedings of Interspeech 2016*, 1054-1058.

[6] Belotel-Grenie, A., and Grenie, M., 2004. The creaky voice phonation and the organisation of Chinese discourse. Proceedings of the International Symposium on Tonal Aspects of Languages, Beijing, China.

[7] Surana, K., and Slifka, J., 2006. Is irregular phonation a reliable cue towards the segmentation of continuous speech in American English? *Proceedings of Speech Prosody 2006*, paper 177.

[8] Keating, P., Garellek, M., and Kreiman, J., 2015. Acoustic properties of different kinds of creaky voice. ICPhS 2015, Glasgow, Scotland.

[9] Laver, J., *The Phonetic Description of Voice Quality*, Cambridge: Cambridge University Press, 1980.

[10] Garellek, M., to appear. The phonetics of voice. *Handbook of Phonetics*, W. Katz and P. Assmann, eds., New York: Routledge.

[11] Batliner, A., Burger, S., Johne, B. *et al.*, 1993. MÜSLI: A classification scheme for laryngealizations. *Proceedings of the ESCA workshop on prosody*, 176-179.

[12] Gerratt, B., and Kreiman, J., 2001. Toward a taxonomy of nonmodal phonation. *J. Phon.*, 29, 365-381.

[13] Redi, L., and Shattuck-Hufnagel, S., 2001. Variation in the realization of glottalization in normal speakers. *J. Phon.*, 29: 4, 407-429.

[14] Sun, X., 2002. Pitch determination and voice quality analysis using Subharmonic-to-Harmonic Ratio. *Proceedings of IEEE ICASSP*, 333-336.

[15] Ishi, C., Sakakibara, K.-I., Ishiguro, H. *et al.*, 2008. A method for automatic detection of vocal fry. *IEEE Transactions on Audio, Speech and Language Processing*, 16: 1, 47-56.

[16] Boersma, P., and Weenink, D., *Praat: Doing phonetics by computer [Computer program]*, version 6.0.23, 2016.

[17] Shue, Y.-L., Keating, P., Vicenik, C. *et al.*, 2011. VoiceSauce: A program for voice analysis. *Proceedings of the XVII International Congress of Phonetic Sciences*, pp. 1846-1849, Hong Kong: International Phonetic Association.

[18] Hothorn, T., Bretz, F., and Westfall, P., 2008. Simultaneous inference in general parametric models. *Biometrical Journal*, 50: 3, 346-363.

[19] Kreiman, J., Gerratt, B. R., Garellek, M. *et al.*, 2014. Toward a unified theory of voice production and perception. *Loquens*, 1: 1.

[20] Hollien, H., and Wendahl, R., 1968. Perceptual study of vocal fry. *J. Acoust. Soc. Am.*, 43, 506-509.