CREAKY VOICE IN THREE SCOTS VARIETIES: USING F0-BASED IDENTIFICATION TO CONSIDER SOCIAL AND LINGUISTIC FACTORS

Joe Pearce

Glasgow University Laboratory of Phonetics, University of Glasgow
j.pearce.1@research.gla.ac.uk

ABSTRACT

While creak in American English is linked to female speakers [1], research on Scots accents using Vocal Profile Analysis (VPA) [2] has found male voices to be creakier than female voices [3, 4]. However, VPA cannot account for within-speaker variation in creak [4]. Here, I consider the potential use of the f0-based method of quantifying creak [5, 6] for examining how social factors (age, gender, area) and linguistic factors affect use of creak in spontaneous speech in Glasgow, Lothian and Insular Scots (n=95).

Creak use was higher among Insular Scots speakers, but showed no difference by age or gender. It was also constrained by linguistic factors (phrase-final position, potential glottal stops, vowel onsets). This work demonstrates how f0-based identification of creak allows consideration of the role of social and linguistic factors; the potential future applications of this method are also discussed.

Keywords: Voice quality, creaky voice, sociophonetics, Scottish English, gender

1. INTRODUCTION

Creaky voice is a mode of phonation characterised by a variety of acoustic characteristics [7], notably low fundamental frequency (f0). This property of some types of creaky voice has been exploited in the f0-based method of automatically identifying creak [5, 6] which codes for creaky voice when f0 drops below a speaker-specific threshold. Here, I consider the potential applications of this method in identifying creaky voice in Scots and Scottish English accents, where Vocal Profile Analysis (VPA) has connected the use of creak with high socio-economic status in male speakers in Edinburgh [8, 9], male speakers in Glasgow [3], and male adolescents in Dumfries, Inverness and Aberdeen [4]. Understanding more about the function of creak requires the ability to consider cases where creak occurs intermittently in more detail than is possible in VPA [4], either due to linguistic factors, or because speakers are using creak to express non-permanent stances, like disengaged affect [10] or toughness[11].

Creak in English is favoured by the presence of glottal stops, which can either be realised with full closure and favour creak in surrounding segments, or be realised without a full closure and manifest as creak in surrounding segments [12, 75] [13]. Glottal stops are common in Scots: In conversational speech, [14] reports a 76% glottalisation rate for /t/ in Glaswegian and [15] reports an 85% glottalisation rate in Edinburgh. Glottals have also been reported in Orkney and Shetland [16, 17]. Creak is also favoured by phrase-final position [18] (see [19] for a review), and phrase-initial vowel onsets [20]. Scottish accents are therefore ideal for exploring the application of this method, allowing us to consider a case where creak is used for both linguistic and social meaning.

Here, I consider creak in data taken from the Scots Syntax Atlas (SCOSYA) [21] in a large sample of younger (approx. 18-25) and older (approx. 65+) male and female speakers from three Scots dialect areas: Glasgow and Lothian, where creak use has been investigated previously, and Insular Scots spoken in Orkney and Shetland, where no previous research has investigated the use of creak. I consider how creak use varies by age, gender and Scots variety, controlling for linguistic factors that favour intermittent use of creak.

2. METHODS

2.1. Corpus

95 speakers from SCOSYA were selected: 19 speakers of Insular Scots (Orkney and Shetland), 29 speakers from Lothian (Edinburgh and surrounding areas), 47 speakers from Glasgow (incl. Greater Glasgow). Speakers were chosen on the basis of an auditory assessment of the level of background noise and recording quality, with noisy and lower quality recordings excluded to improve the results of the forced alignment and subsequent F0 tracking. The varying number of speakers between areas reflect differences in the total available number of each

area, as well as differences in usable recordings. 180 seconds of speech were extracted from each speaker. Time-aligned transcriptions provided by SCOSYA were converted to TextGrids using [22], then force aligned using the Montreal Forced Aligner v.1.1 [23, 24].

2.2. F0-based automatic identification of creak

I followed [6] in using an f0-based method of identifying creak automatically. Files were processed at 16 kHz using MacReaper [6], a drag-and-drop implementation of REAPER [25] for MacOS, to obtain Glottal Closure Instants (GCIs). As [6] notes, REAPER appears to be reliable at detecting GCIs at low and irregular f0, making it well-suited to analysis of creak. Where creak manifested as multiply-pulsed or aperiodic, I found that REAPER tended not to identify all GCIs, resulting in identification of lower f0 and thus classification of creak; however, this is only a passing observation and future work could investigate whether this occurs at scale.

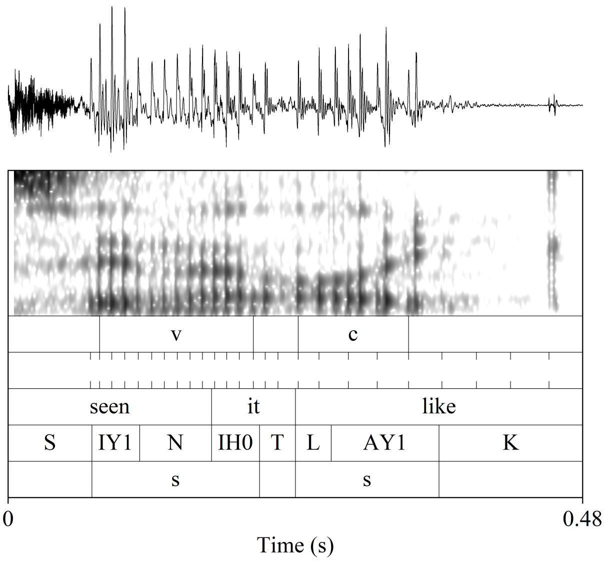

Figure 1: Example of creak identification for Scott, a younger male Glasgow speaker. Tiers show (1) creak identification, (2) GCI detection, (3) words, (4) phones, and (5) initial sonorant stretches

The automated procedure considered only GCIs that occurred within stretches of sonorants. [26] found that this improves the performance of f0-based identification of creak. 52,684 stretches were originally identified, which included 1,188,908 GCIs. Selections were hand-checked for local background noise, constructed dialogue, and major forced alignment errors, which were excluded. After exlusions, 30,792 stretches remained, containing 613,151 GCIs.

Following [6], local f0 was calculated for each GCI by taking the inverse of the time between each GCI within a voiced stretch. Antimodes were then detected using an automated procedure [6] in [27], which `modes`[28](v.0.7.0) to identify the f0 mode of non-creaky speech, the f0 mode of creaky speech, and the antimode between them. Example output is shown in Fig 2. Antimodes were inspected visually for verification. Three speakers were excluded at this stage on the basis that reliable antimodes could not be identified (1 YF Glasgow, 1 OF Glasgow, 1 OF Lothian). 92 speakers remained for analysis.

Speech from the original stretches was then separated into new creaky and non-creaky stretches based on whether local f0 fell below an individual speaker's antimode: In this way, a stretch could either be coded entirely as creaky or non-creaky, or separated into smaller stretches to allow for shorter cycles of creak. An example is shown in Fig. 1: Original stretches are shown on the bottom tier, GCIs on the second tier, and classification as voice or creak on the first tier. I quantified the percentage of creak used by each speaker and group by dividing the total duration of creaky stretches by the total duration of all sonorant stretches. Separation into creaky and non-creaky stretches created 31,698 stretches, of which 26,922 (85%) were not creaky and 4,776 (15%) were creaky. This was equivalent to a total amount of 3,759s (63 mins) of relevant sonorant stretches, of which 340s (9%) was creaky.

2.3. Statistical analysis

The presence of creak was analysed using a binomial generalized linear mixed-effects model in `lme4`[29] (v1.1-28)in R v4.1.2[27]. Effects were added sequentially starting from a minimal model and model fit compared using log-likelihood ratio tests with `anova()`. Fixed effects (reference levels in bold) in the final model were Duration of stretch (log-transformed, scaled), Speech Rate (calculated locally in syllables per second, log-transformed and scaled), potential Glottalisation (**None**, pre-glottalisation context, glottalable context), Phrase-final Position (**Non-final** or Final), Vowel (**Medial vowel**, none, initial vowel, both medial and initial), Contains /l,w,j/ (**No /l,w,j/** or contains /l,w,j/), and Area (**Glasgow**, Lothian, Insular). The final model also included uncorrelated random slopes for Participant by Duration and Speech Rate, and Words by Duration and Speech Rate, where 'Words' was the words that made up the stretch. Fixed-effects that were tested but which did not improve the fit of the model were as follows: Contains Nasal, Contains Rhotic, Gender (male or female) and Age (older or younger).

Linguistic factors were estimated from the forced aligner output, and so are best understood as proxies for linguistic factors that favour creak. Phrase-Final Position was estimated based on whether the stretch was followed by a pause. Phrase-initial Vowels are coded within the factor Vowel by identifying phrase-initial stretches that began with a vowel. Glottalable contexts are cases adjacent to a /t/, which itself occurs in a context that could be produced as a glottal stop [30, p. 183-184]. Pre-glottalisation contexts were identified as stretches followed by a /p/, /t/, /k/ or /tʃ/ that occurs in a context that could favour pre-glottalisation [30, p. 183-184]. This aimed to account for the fact that /p/ and /k/ can also be glottalised in Glasgow [31].

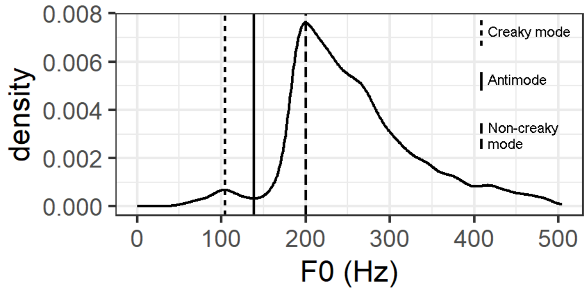

Figure 2: An antimode analysis of the f0 distribution for Alice, a younger female Glasgow speaker

3. RESULTS

3.1. Linguistic factors

Increased voiced stretch Duration and Speech Rate both reduced the log odds of creak (Duration $\beta = -1.121$, $SE(\beta) = 0.054$, $p<0.001$; Speech Rate $\beta = -0.458$, $SE(\beta) = 0.0333$, $p < 0.001$).

By contrast, creak was favoured by Phrase Final position (log(odds) = 0.475, $SE(\beta) = 0.045$, p $<$ 0.001), and a stretch occurring in a Glottalable Context (log(odds) = 0.454, $SE(\beta) = 0.057$, p $<$ 0.001). A stretch being followed by a potential Preglottalisation context did not favour creak. Furthermore, containing an Initial Vowel (log(odds) $= 0.660$, $SE(\beta) = 0.085$, p $<$ 0.001) and containing no vowel also favoured creak (log(odds) = 1.445, $SE(\beta) = 0.088$, p $<$ 0.001), while containing both a Medial and Initial vowel did not favour creak. Containing /l,w,j/ also favoured creak (log(odds) = 0.177, $SE(\beta) = 0.052$, p $<$ 0.005).

3.2. Social factors

The overall percentage of creak used by each Area, Gender and Age group is shown in Fig. 3. Insular Scots voice quality stands out as particularly creaky: Being an Insular speaker increased the log(odds) of a stretch being creaky by 0.610 ($SE(\beta) = 0.197$, p $<$ 0.01). However, no significant difference was found by Gender or Age.

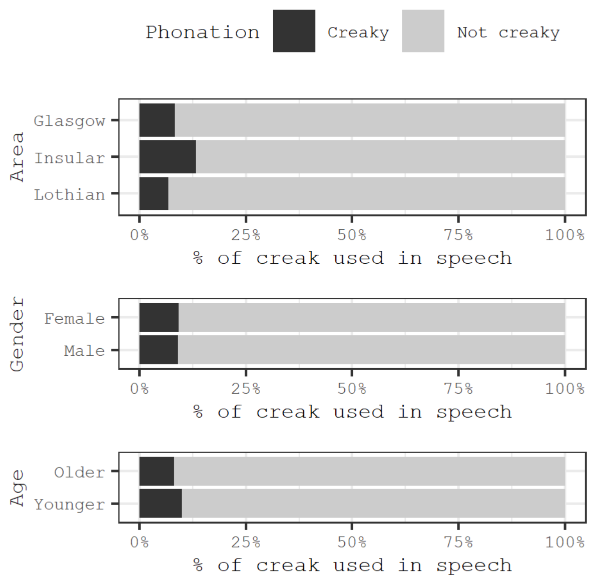

Figure 3: Percentage of creak used across all voiced speech by Area, Gender and Age

4. DISCUSSION

Previous work [5, 6, 32] has shown the potential for f0-based separation of creaky and non-creaky speech on the basis of f0 antimode. Here, this method was used to separate creaky and non-creaky speech, and model how both social and linguistic factors favour the presence of creak in Scots.

4.1. Linguistic variation

Longer stretches were more likely to be identified as non-creaky, showing a limitation of the present analysis: As creak is separated from non-creak, new stretches are created, and these tend to be short if creak is present, as creak is rarely present throughout an entire stretch. This means that it is difficult to make a distinction between speakers who use creak often, but for short periods of time (e.g. for glottals), from those whose creak use may contribute more substantially to their overall voice quality. Future research might consider using a different unit of analysis: for example, separating speech into 10 ms frames following [26]. Determining what unit of analysis to work with may require considering whether there is a threshold for the presence of creak under which creak does not contribute to the perception of a creaky voice and rather only contributes to the perception of a linguistic phenomenon (e.g. a glottal stop).

There is no surprise here in that vowel onsets, phrase-final position, and potential glottal stops favour creak. However, no previous work appears to have considered the effect of linguistic factors on f0-based estimates of creak. The success of estimating these from the forced aligner rather than hand-coding for each linguistic effect shows the usefulness of f0-based estimation of creak as a coarse-grained method of estimating creak use in large amounts of data, even in spontaneous speech. This makes it possible to distinguish between linguistic uses of creak and potential other reasons, making it useful for sociophonetic research.

4.2. Social factors

Creak was more prevalent in Insular Scots than in Glasgow Scots, while no significant difference was found between Glasgow and Lothian speakers. While [33] find that the use of local Shetland lexical, phonological and morphosyntactic features is lower among younger speakers, they note that the situation is highly complex with different speakers using local forms at different rates. One possible interpretation of the high use of creak in Orkney and Shetland is that creak may be a characteristically local voice quality that younger speakers may be maintaining, while their use of other certain local linguistic forms decreases. Future research could consider differences between Orkney and Shetland, which while grouped together here on the basis of shared historical and phonological characteristics, do form separate varieties, as well and what the function of creak is in Insular Scots.

However, outside of Area, no social factors improved the fit of the model. Unlike [3, 4], I found similar rates of creak between male (9.0%) and female speakers (9.1%). However, as [26] notes, f0-based identification of creak performs better on female speech than male speech; the lack of a higher rate of creak for male speakers could reflect poorer identification of creak in male speakers. I also found similar rates of creak between older (8.1%) and younger (9.9%) speakers. This suggests that in Glasgow and Lothian, creak may serve a more linguistic than social function. However, factors not considered here may be important to consider in future research: Considering the relationship between ethnicity and creak use in other varieties [32, 34] and high creak and high socio-economic status in Edinburgh [8], these social factors may be important in understanding the indexical meanings of creak in Scots accents. Additionally, the present analysis did not consider any of the potential conversational or affective functions of creak, which have been suggested to underlie gender differences in creak found in other varieties (e.g.[10]).

5. CONCLUSION

Work using only a single voice quality measure is often difficult to interpret because measures of Harmonics-to-Noise Ratio and spectral tilt cannot be interpreted in absolute terms in isolation [35], an issue that becomes more apparent when working with corpus data. F0-based categorisation of creak shows promise in this regard: As many forms of creaky voice are characterised by F0 that is low within in a speaker's F0 range, or tracked as such due to F0 irregularity, this method could be used for coarse-grained separation of creaky and non-creaky voice. This could then form the basis of subsequent acoustic analysis of both creaky and non-creaky voice, in turn allowing a more complete view of the types of creak used and where creak fits in a speaker's wider acoustic voice profile.

F0-based estimation of creak has allowed for consideration of how creak use patterns according to both social and linguistic factors, but future research is needed to determine the meaning of creak in Scots beyond relationships with macro-social categories. This present study demonstrates the potential use of the method in sociophonetic research, such as analysing large amounts of spontaneous speech where linguistic context may differ between speakers. It also shows the promise of the method more generally: while this analysis is restricted to macro-level social categories and previously known linguistic constraints, future research could code data for factors such as stance, affect, turn-taking, or topic to gain a closer insight into the indexical, affective and pragmatic meanings of creak in a particular variety or context.

6. REFERENCES

[1] I. Yuasa, "Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile american women?" *American Speech*, vol. 85, pp. 315–337, 10 2010.

[2] J. Laver, S. Wirz, J. Mackenzie, and S. Hiller, "A perceptual protocol for the analysis of vocal

profiles," *Edinburgh University Department of Linguistics Work in Progress*, vol. 14, pp. 139–155, 1981.
[3] J. Stuart-Smith, "Voice quality in Glaswegian," in *Proceedings of the 14th ICPhS*, 1999, pp. 2553–56.
[4] J. Beck and F. Schaeffler, "Voice quality variation in Scottish adolescents: Gender versus geography," in *Proceedings of the 18th ICPhS*, 2015.
[5] K. Dorreen, "Fundamental frequency distributions of bilingual speakers in forensic speaker comparison," Master's thesis, University of Canterbury, 2017.
[6] K. Dallaston and G. Docherty, "Estimating the prevalence of creaky voice: a fundamental frequency-based approach," in *Proceedings of the 19th ICPhS*, M. T. Sasha Calhoun Paola Escudero and P. Warren, Eds. Canberra, Australia: ASSTA, 2019, pp. 532–536.
[7] P. Keating, M. Garellek, and J. Kreiman, "Acoustic properties of different kinds of creaky voice," in *Proceedings of the 18th ICPhS*, 2015.
[8] J. Esling, "Voice quality in edinburgh: A sociolinguistic and phonetic study," Ph.D. dissertation, University of Edinburgh, 1978.
[9] ——, "The identification of features of voice quality in social groups," *JIPA*, vol. 8, pp. 18–23, 1978.
[10] R. Podesva, "The affective roots of gender patterns in the use of creaky voice," in *ETAP 4*, 2018.
[11] N. Mendoza-Denton, *Homegirls: language and cultural practice among Latina youth gangs*. Malden, MA: Blackwell Pub, 2008.
[12] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. Blackwell, 1996.
[13] G. Docherty and P. Foulkes, "Sociophonetic variation in 'glottals' in Newcastle English," in *Proceedings of 14th ICPhS, San Francisco*, J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey, Eds., 1999, pp. 1037–1040.
[14] J. Stuart-Smith, "Glottals past and present: A study of T-glottalling in Glaswegian," *Leeds Studies in English*, vol. 30, pp. 181–204, 1999.
[15] E. Schleef, "Glottal replacement of /t/ in two british capitals: Effects of word frequency and morphological compositionality," *Language Variation and Change*, vol. 25, no. 2, pp. 201–223, 2013.
[16] P. Sundkvist, ""Standard English" as spoken in Shetland's capital," *World Englishes*, vol. 30, no. 2, pp. 166–181, 2011.
[17] H. Schmitt, "Orkney English phonology: observations from interview data," *Scottish Language*, vol. 34, pp. 58+, 2015.
[18] 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., 1988, pp. 3–29.
[19] L. Davidson, "The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 12, no. 3, pp. 1–18, 2021.
[20] L. Dilley, S. Shattuck-Hufnagel, and M. Ostendorf, "Prosodic constraints on glottalization of vowel-initial syllables in American English," *JASA*, vol. 95, no. 5, pp. 2978–79, 1994.
[21] J. Smith, D. Adger, B. Aitken, C. Heycock, E. Jamieson, and G. Thoms, "The Scots Syntax Atlas," https://scotssyntaxatlas.ac.uk, 2019.
[22] R. Fromont, "trs2grid.jar," 2022. [Online]. Available: https://sourceforge.net/projects/labbcat/
[23] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: trainable text-speech alignment using Kaldi," in *Proceedings of the 18th Conference of the ISCA*, 2017.
[24] ——, "Montreal Forced Aligner [computer program]. v.1.1.0," 2017. [Online]. Available: http://montrealcorpustools.github.io/Montreal-Forced-Aligner/
[25] D. Talkin, "REAPER: Robust Epoch And Pitch EstimatoR," 2015. [Online]. Available: https://github.com/google/REAPER
[26] H. White, J. Penney, A. Gibson, A. Szakay, and F. Cox, "Evaluating automatic creaky voice detection methods," *JASA*, vol. 152, no. 3, pp. 1476–86, 2022.
[27] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021.
[28] S. Deevi and D. Strategies, *modes: Find the Modes and Assess the Modality of Complex and Mixture Distributions, Especially with Big Datasets*, 2016, r package v 0.7.0.
[29] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
[30] A. Cruttenden and A. C. Gimson, *Gimson's pronunciation of English*, 8th ed. London: Routledge, 2014.
[31] O. McCarthy and J. Stuart-Smith, "Ejectives in Scottish English: a social perspective," *JIPA*, vol. 43, pp. 273—98, 2013.
[32] A. Szakay and E. Torgersen, "A re-analysis of f0 in ethnic varieties of London English using REAPER," in *Proceedings of the 19th ICPhS*, M. T. Sasha Calhoun Paola Escudero and P. Warren, Eds. Canberra, Australia: ASSTA, 2019, pp. 1675–78.
[33] J. Smith and M. Durham, "A tipping point in dialect obsolescence? Change across the generations in Lerwick, Shetland," *Journal of Sociolinguistics*, vol. 15, no. 2, pp. 197–225.
[34] D. Loakes and A. Gregory, "Voice quality in australian english," *JASA Express Letters*, vol. 2, no. 8, p. 085201, 2022.
[35] M. Garellek, "The phonetics of voice," in *The Routledge Handbook of Phonetics*, 2019, pp. 75–106.