Acoustic voice quality analysis of 1000+ speakers from across the UK

Ben Gittelson[1], Yang Li[2], Adrian Leemann[3]

[1]Amazon Alexa; [2]Phonetics Laboratory, TAL, University of Cambridge; [3]Department of Linguistics and English Language, University of Lancaster

[1]bwg2109@columbia.edu, [2]yl492@cam.ac.uk, [3]a.leemann@lancaster.ac.uk

The quality of a speaker's voice (VQ) is determined by laryngeal and supralaryngeal features (Ladefoged 1971). The acoustic correlates of these features have remained largely unexplored for speakers of British English. The aim of this paper is to present preliminary analyses of VQ measures that were applied on a corpus of 1000+ speakers from across the UK.

The data was crowdsourced through the *English Dialects App* (anonymous). In the app, users recorded ten sentences from *The Boy Who Cried Wolf* and reported demographic information. Three words were analyzed from 1,700 speakers across the UK using WebMAUS (Kisler et al. 2016): phrase-final "while" and phrase-initial "however" and "one." Using VoiceSauce (Shue 2011), we analyzed H1*-H2*, H1*-A1*, Cepstral Peak Prominence (CPP), Harmonic to Noise Ratios (HNR, at different frequencies), Subharmonic to Harmonic Ratios (SHR), and f0 (using STRAIGHT, cf. Kawahara 2001). For each measure, we extracted data on ten equidistant time points from each word, providing information about changes of VQ measures through time.

Regression models were fitted with the acoustic measures as response variables and various metadata categories such as such as age, gender, educational background, ethnicity, commuting distance and mobility as explanatory variables. This reveals a number of significant and sizeable effects, of which we note the more robust and consistent ones as follows:

While there are hints of lower f0 for black subjects than white ones, blacks have significantly lower CPP (e.g. by 2.76 dB than white counterparts in 'one'), suggesting lower periodicity; other VQ measures did not return consistent results. As expected, males have lower f0 (by around 85Hz in "one" and "however," 51Hz in phrase-final "while"), lower H1*-H2* (1.2 ~ 2.2 dB), lower HNR15 (1.7 ~ 2.2), with highly inconsistent CPP. CPP and f0 also decline significantly as participant age increases. There is no conclusive evidence for systematic geographic differences in VQ measures. There is, however, a tendency for higher CPPs in more rural areas, and lower CPPs in urban regions (tested on a subset of 20-30 year-old females). Further work will centre on geographically weighted regression to detect finer-grained dialectal variation in VQ (Stuart-Smith 1999). We will also discuss the correlations among various VQ measures and what they imply about differences in phonatory settings among different groups.

References

H. Kawahara, Jo Estill and O. Fujimura: Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT, MAVEBA 2001, Sept.13-15, Firentze Italy, 2001.

Kisler, T., Schiel, F. and Sloetjes, H. (2012): Signal processing via web services: the use case

Keating, P., Garellek, M. and Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. Proceedings of ICPhS 2015, Glasgow, Scotland.

WebMAUS, Proceedings Digital Humanities 2012, Hamburg, Germany, 30-34.

Ladefoged, P. (1971) Preliminaries to linguistic phonetics. Chicago: University of Chicago.

Anonyomous (in revision). The English Dialects App: the creation of a crowdsourced dialect corpus.

Shue, Y-L., Keating, P., Vicenik, C., and Yu, K. (2011). VoiceSauce: A Program for Voice Analysis. Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong.

Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. Urban voices: Accent studies in the British Isles, 203-222.