Voice Quality of Hesitations: Acoustic Measures and VPA ratings

Amanda Cardoso[1], Paul Foulkes[1], Peter French[1,2], Philip Harrison[1,2], Vince Hughes[1], Colleen Kavanagh[1,3], and Eugenia San Segundo[1]

[1]*Department of Language and Linguistic Science, University of York, UK*

{amanda.cardoso|paul.foulkes|peter.french|vincent.hughes|colleen.kavanagh|eugenia.sansegundo}@york.ac.uk

[2]*J P French Associates, York, UK*

philip.harrison@jpfrench.com

[3]*RCMP, Canada*

colleen.kavanagh@rcmp-grc.gc.ca

Voice quality (VQ) provides a great deal of speaker-specific information and is widely regarded as a useful tool to discriminate speakers by forensic speech scientists (Gold and French 2011, San Segundo et al. in press). However, the range of anatomical, motoric, and socio-cultural factors that affect a speaker's VQ makes it difficult to analyse (Kreiman et al. 1993). All specific vocal settings are intermittent and not permanent, in that they are manifested only in certain segments (e.g. phonation is only manifested in voiced segments). Given these complications, is it possible to reduce auditorily-assessed VQ to a set of acoustic measures?

The acoustic analysis of VQ metrics is complicated by the common practice to apply those metrics to sustained vowels, which are infrequent in natural speech and only partially representative of a speaker's vocal range. However, an exception to this is vocalic hesitations (*uh*s and *um*s), which occur frequently in speech, are often longer than lexical vowels, and often occur adjacent to unfilled pauses. The acoustics of hesitations also provide good speaker discriminatory information (Tschäpe et al. 2005, Hughes et al. 2016). Therefore, hesitations may help to answer whether and which acoustic correlates of VQ can be compared with perceived VQ settings.

San Segundo et al. (in press) produced auditory ratings of VQ settings for each speaker in the DyViS corpus (Nolan et al. 2009), using a modified version of the vocal profile analysis (VPA) scheme (Laver et al. 1981). Our data consists of 33 speakers who received the highest and lowest scalar degree ratings across the VQ dimensions and who, therefore, represent extreme speakers within this corpus. We focus here only on creaky, breathy, and neutral (i.e. VPA settings rated 0).

Reliable acoustic measures to distinguish breathy and creaky phonation have been widely debated (e.g. Gobl and Ní Chasaide 1992, Klug 2014, Keating et al. 2015, Klug et al. 2018). Breathy and creaky voice have lower f0 values than modal voice. Creaky voice is also claimed to have higher jitter, i.e. frequency variation between consecutive f0 cycles. Breathy voice has higher positive values of H1-H2 (amplitude difference between first and second harmonic) and H1-A3 (amplitude difference between first harmonic and third formant) than creaky voice. Acoustic measures of f0, jitter, H1-H2, and H1-A3 were taken from the sustained vowel portions for each of the 33 speakers' hesitations (mean/median=18, range=11–39) from DyViS Task 1.

We found that some acoustic measures do align with the speakers' VPA ratings (Table 1). Speakers who are perceived to have creaky phonation tend to have lower f0 than other speakers (Figure 1), and higher jitter (Figure 2). Conversely, H1-H2 and H1-A3 do not correspond well to VPA ratings, with little difference between speakers classified as breathy and creaky, and breathy values lower than those for creaky - contra previous studies (e.g. Ní Chasaide and Gobl 1997, Esposito 2010).

VQ of hesitations are comparable to overall VQ, but there are also VQ patterns specific to hesitations, which further supports the use of hesitation for speaker discrimination.

Table 1. Average acoustic measures of voice quality for speakers rated as creaky, breathy or neutral.

VPA VQ	*f0 (Hz; mean)*	*Jitter (mean)*	*H1-H2 (dB; mean)*	*H1-A3 (dB; mean)*
Creaky	99	0.92	3.98	9.64
Breathy	109	0.85	3.27	9.34
Neutral	109	0.82	1.69	6.10

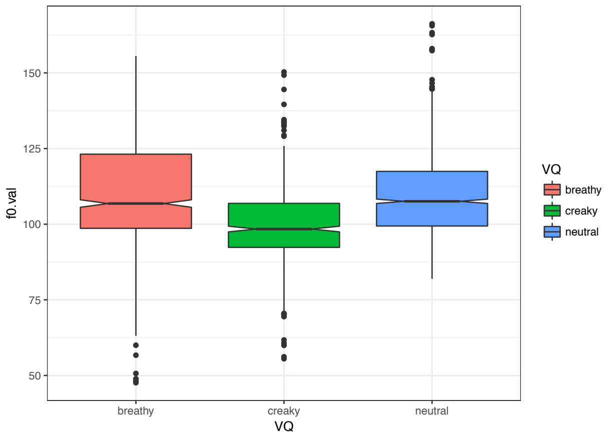

Figure 1. Distribution of f0 (Hz) values across speakers rated as creaky, breathy, or neutral.

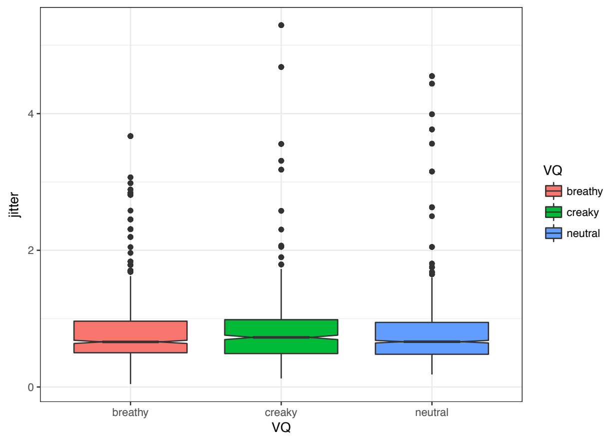

Figure 2. Distribution of jitter values across speakers rated as creaky, breathy, or neutral.

References

Esposito, C. M. (2010). Variation in contrastive phonation in Santa Ana del Valle Zapotec. *Journal of the International Phonetic Association*, 40(2), 181–198.

Gobl, C. and Ní Chasaide, A. N. (1992). Acoustic characteristics of voice quality. *Speech Communication*, 11(4), 481–490.

Gold, E. and French, P. (2011). International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law*, 18, 293–307.

Hughes, V., Foulkes, P. and Wood, S. (2016). Formant dynamics and durations of um improve the performance of automatic speaker recognition systems. *Proceedings of the 16th Australasian Conference on Speech Science and Technology (ASSTA)*.

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

Klug, K. (2014). *The Role of Voice Quality in Forensic Speech Science*. MSc Dissertation, University of York.

Klug, K., Kirchhübel, C., French, P. and Foulkes, P. (2018). Do the acoustics support the perception? The example of breathy voice, *IAFPA Conference*, Huddersfield, UK.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A. and Berke, G. S. (1993). Perceptual Evaluation of Voice Quality Review, Tutorial, and a Framework for Future Research. *Journal of Speech, Language, and Hearing Research*, 36(1), 21–40.

Laver, J., Wirz, S., Mackenzie-Beck, J. and Hiller, S. (1981). A perceptual protocol for the analysis of vocal profiles. In *Work in Progress*, Department of Linguistics (Edinburgh: University of Edinburgh). (Reprinted in Laver, J., 1991, The Gift of Speech Edinburgh University Press).

Ní Chasaide, A. and Gobl, C. (1997). Voice source variation. In Hardcastle, W.J. and Laver, J. (eds.) *The handbook of phonetic sciences.* Oxford: Blackwell, 427-461.

Nolan, F., McDougall, K., de Jong, G. and Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research, *IJSLL*, 16(1), 31–57.

San Segundo, E., Foulkes, P., French, P., Harrison, P., Hughes, V. and Kavanagh, C. (in press). The use of the Vocal Profile Analysis for speaker characterization: a methodological proposal. *Journal of the International Phonetic Association*.

Tschäpe, N., Trouvain, J., Bauer, D. and Jessen, M. (2005). Idiosyncratic patterns of filled pauses. *IAFPA Conference*, Marrakesh, Morocco.