Acoustic correlates of voice quality settings: variation at the level of individual speaker

Alice Paver and Kirsty McDougall, University of Cambridge, UK

Voice quality (VQ), also described as 'the characteristic auditory colouring of an individual speaker's voice' [1], refers to the articulatory settings adopted by a given speaker. Perceptual assessment schemes such as the VPA protocol [1, 2] remain widely used, but their subjectivity has motivated interest in acoustic measures often assumed to be more objective. Although studies have reported acoustic correlates of laryngeal settings (e.g. CPP, HNR, spectral slope measures), these correlates are typically calculated across speakers, so apparent VQ effects may partly reflect inter-speaker differences [3]. Furthermore, few studies have investigated variation at the individual level, i.e. acoustic correlates of VQ when one speaker adopts different VQ settings in different recordings. By using stimuli with a range of VQ guises produced by the same speakers, this study isolates VQ-related acoustic change at the group level from broader speaker-specific variation.

Recordings of four British English male phoneticians producing the rainbow passage in multiple VQ guises were taken from the PASR dataset [4]. Six settings (laryngeal and supralaryngeal) were examined: modal (MOD), breathy (BRT), fronted tongue body (FTB), nasal (NAS), denasal (DEN) and lowered larynx (LLX). Acoustic measures previously linked to VQ (including CPP, H1*–H2*, HNR05–35, F1–F4) were extracted from vocalic segments using VoiceSauce in MATLAB. A linear mixed-effects model was fitted with VQ setting, acoustic measure and speaker as a three-way interaction, and segment as a random intercept. Pairwise comparisons focused on contrasts between modal and non-modal VQ settings.

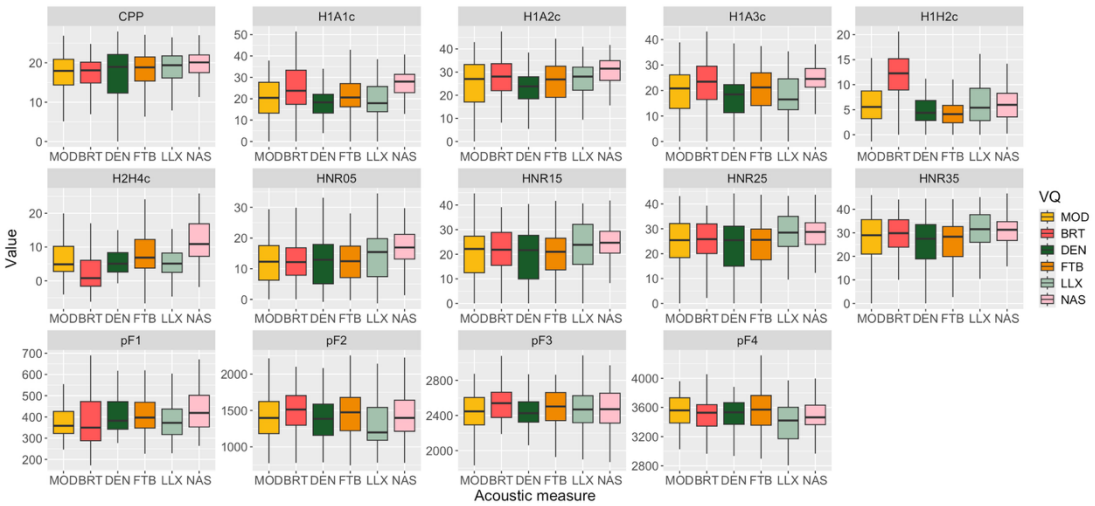

Figure 1. Boxplots of values for a selection of acoustic measures for each VQ setting. Each box represents values for all four speakers.

Results reveal a complex picture. Group-level trends, shown in Figure 1, showed significantly higher H1*–H2* for BRT relative to MOD, consistent with prior research. Additionally, novel findings include that LLX VQ elicited significantly lower F2 and F4 values, while NAS was associated with higher H1-H2* and H2-H4* than MOD. Follow-up analyses demonstrated substantial individual variation in both the direction and magnitude of VQ-related acoustic change. No single measure behaved consistently across all speakers. This paper will explore the individual variation in further detail. The findings have implications for phonetic understanding of the relationship between perceptual and acoustic analyses of VQ, which has applications to the fields of forensic phonetics, speech pathology and sociophonetics.

[1] Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge University Press.
[2] Mackenzie Beck, J. (2005). Perceptual analysis of voice quality: The place of Vocal Profile Analysis. In W. J. Hardcastle & J. M. Beck (Eds), *A Figure of Speech*. Routledge.
[3] Nechanský, T., Houzar, A., Bořil, T., & Skarnitzl, R. (2024). Controlled voice quality modifications: Acoustic, perceptual and ASR analysis. *IJSLL*, *31*(1), 49–76.
[4] Hughes, V., Foulkes, P., Harrison, P., Wormald, J., Xu, C., van der Vloed, D., & Kelly, F. (2022). Person-specific ASR [Data set]. http://pasr.york.ac.uk/