

# Listeners' perception of voice similarity in Standard Southern British English versus York English

*Kirsty McDougall*

*Department of Theoretical and Applied Linguistics, University of Cambridge, UK*

kem37@cam.ac.uk

This paper reports part of a research programme which explores 'perceived voice similarity' (PVS), the notion that within a group of speakers of the same sex and age, listeners will perceive certain speakers as sounding more similar to each other than others. Findings from this research focussing on various aspects of Standard Southern British English (SSBE) and on a voice parade case in East Anglia were presented at IAFPA 2010 and 2011 (McDougall 2011, Nolan *et al.* 2010, 2011). The present study considers the extent to which the acoustic correlates of PVS are consistent across accents, and reports results from a study of York English (YE) in comparison with the SSBE findings.

For the SSBE experiment, 15 male speakers, aged 18-25 years, were selected from the *DyViS* database (Nolan *et al.* 2009). For the YE experiment, 15 male speakers of the same age were selected from the newly developed *YorViS* database of YE<sup>1</sup> which contains recordings of the same format as *DyViS*. For both experiments, spontaneous speech from a telephone call task was used to create the stimuli: two utterances per speaker, each approximately 3 seconds in duration. Within each experiment, each speaker was matched with all other speakers and with himself to form 120 pairings. 20 listeners (speakers of British English; a different group for each experiment) rated the (dis)similarity of the paired voice samples on a nine-point scale from 'very similar' to 'very different'. Multidimensional scaling (MDS) was applied to the ratings to derive five perceptual dimensions for each accent whose correlation with long-term fundamental frequency, articulation rate, and long-term formant analysis of F1, F2, F3 and F4 was tested using Spearman's formula.<sup>2</sup>

Long-term fundamental frequency plays an important role in PVS in voice similarity, yielding significant correlations with perceptual dimensions in both accents. It correlates significantly with the first perceptual dimension for SSBE ( $r = 0.804$ ), indicating that it is of key importance for this accent. In YE, long-term  $f_0$  correlates significantly with dimension 3 ( $r = 0.689$ ), while the upper formants, F3 and F4, appear to show greater levels of importance by correlating significantly with respectively dimensions 1 ( $r = 0.536$ ) and 2 ( $r = 0.514$ ). F3 does not correlate with any perceptual dimension for the SSBE responses (at the time of writing, F4 information is not available for SSBE). The lower formants appear to behave in a similar fashion across the two accents, with F2 ranking higher (significant correlation with dimension 2 in SSBE ( $r = 0.514$ ) and with dimension 3 in YE ( $r = 0.557$ )) than F1 (significant correlation with dimension 4 in both accents: SSBE  $r = 0.675$ , YE  $r = 0.718$ ). Articulation rate did not achieve a significant correlation with any of the MDS dimensions in either accent, possibly due to the short duration of the stimuli.

Implications of the findings for voice parade construction, in particular with respect to the choice of foil voices, will be discussed.

## References

K. McDougall (2011) 'Acoustic correlates of perceived voice similarity: a comparison of two

accents of English.’ Paper presented at the International Association for Forensic Phonetics and Acoustics Annual Conference, Vienna, 24-28 July 2011.

F. Nolan, J.P. French, K. McDougall, L. Stevens and T. Hudson (2011) ‘The role of voice quality ‘settings’ in perceived voice similarity.’ Paper presented at the International Association for Forensic Phonetics and Acoustics Annual Conference, Vienna, 24-28 July 2011.

F. Nolan, K. McDougall, G. de Jong & T. Hudson (2009) The *DyViS* Database: Style-Controlled Recordings of 100 Homogeneous Speakers for Forensic Phonetic Research. *International Journal of Speech, Language and the Law*, **16.1**, 31–57.

F. Nolan, K. McDougall and T. Hudson (2010) ‘Perceived voice similarity and acoustic measures.’ Paper presented at the International Association for Forensic Phonetics and Acoustics Annual Conference, Trier, 18-21 July 2010.

<sup>1</sup> British Academy Small Grant No. SG112892 to K. McDougall ‘A Cross-Accent Investigation of the Phonetic Correlates of Perceived Voice Similarity’.

<sup>2</sup> Thanks are due to Erica Gold, Philip Harrison and Christin Kirchhübel for their contributions to these phonetic analyses.