The contribution of source and filter to speaker characterisation

Vincent Hughes[1], Amanda Cardoso[2], Paul Foulkes[1], Peter French[1,3], Amelia Gully[1], Philip Harrison[1,3]
[1]*Department of Language and Linguistic Science, University of York, UK*
[2]*Department of Linguistics, University of British Columbia, Canada*
[3]*J P French Associates, York, UK*
vincent.hughes@york.ac.uk

Source-filter theory (Fant, 1960) has long been the dominant model of speech production in phonetics and speech science research. A key assumption of the theory is the independence of source and filter, such that different vowels can be produced with the same fundamental frequency (F0), and the same vowel can be produced with different F0s. However, a growing body of research has revealed non-linear interactions between source and filter. Air pressures within the resonating vocal tract have been shown to affect transglottal airflow and the vibration of the vocal folds, especially where the F0 or other high intensity harmonics are close to a formant frequency (Titze, 2008). The effects of non-linear source-filter coupling can be seen in Roberts (2012: 135). Other interdependencies between source and filter have also been found. Studies have shown covariance between F0 and formants, attributed to physiological factors (speakers with bigger larynges tend to have bigger vocal tracts, and therefore lower F0 and formants). Further, there are articulatory factors which lead to covariance in source and filter output. For instance, Gordon and Ladefoged (2001) report larynx raising in creaky voice and larynx lowering in breathy voice, both of which affect F1. This study examines the extent of source-filter interdependence in the context of speaker characterisation, with a view to establishing whether source and filter features capture complementary information that can be used to improve the accuracy of forensic voice comparison and speaker recognition.

Data from DyViS (Nolan et al., 2009) Tasks 1 and 2 for 90 speakers were used. A range of acoustic measures of source (F0 and laryngeal voice quality [VQ]) and filter (formants and MFCCs) were extracted from the vocalic portion of the hesitation marker *um*. High-quality recordings of both tasks and controlled segmental materials were used in order to remove the detrimental effects of channel, ensuring reliable acoustic measures (particularly for VQ measures). The current findings focus on source-filter independence, rather than absolute speaker discrimination performance.

Firstly, the raw data were examined. Figure 1 shows normalised between-speaker Kullback-Leibler divergences based on all of the source measures and all of the filter measures. No correlation was found here, such that two speakers could be very similar in terms of source measures, but very different in terms of filter measures, and vice versa. Secondly, the complementarity of source and filter measures was examined using speaker discrimination. Likelihood-ratio (LR) based testing was conducted using source and filter measures separately, as well as a fusion of the two. The filter measures consistently outperformed the source measures. However, when combining the source measures with the filter measures, improvements of as much as 64% in terms of the log LR cost (C_{llr}; Brümmer and du Preez, 2006) were found, compared with the filter measures alone. Taken together the results reveal considerable source-filter independence in terms of speaker characterisation. However, some individuals are more susceptible to improvements in performance than others. In this paper we explore this issue of individual variation within the context of overall system performance.

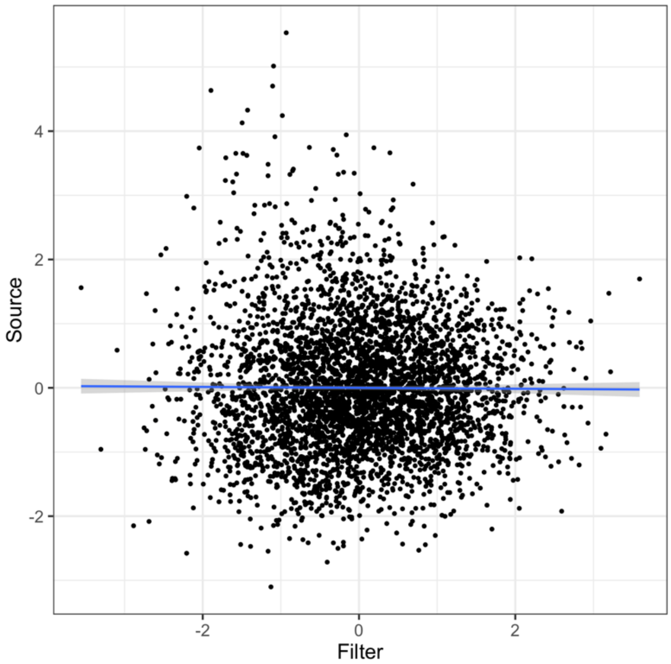

Figure 1. Scatterplot of normalised between-speaker distances based on filter measures and source measures fitted with a linear trend line ($R^2 < 0.001$)

References

Brümmer, N. and J. du Preez. (2006). Application-independent evaluation of speaker detection. *Computer Speech and Language*, **20(2-3)**, 230–275.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Gordon, M. and P. Ladefoged. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, **29(4)**, 383–406.

Nolan, F., K. McDougall, G. de Jong and 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.

Roberts, L. S. (2012) *A forensic phonetic study of the vocal responses of individuals in distress.* PhD Thesis, University of York, UK.

Titze, I. R. (2008). Nonlinear source-filter coupling in phonation: theory. *Journal of the Acoustical Society of America*, **123(5)**, 2733–2749.