Examining within-session intra-speaker variation using computational models and phonetically-derived representations of speakers.

Lois Fairclough[1]

[1] *Department of Linguistics and English Language, Lancaster University, Lancaster, UK*
`l.fairclough@lancaster.ac.uk`

```
Work in Progress Poster

Eligible for the 'Best Student Paper Award'
```

One key cause of incorrect speaker identification is intra-speaker variation (Zhang, 2006). With speaker individuality as the overarching topic, and with a particular focus on intra-speaker variation, this project aims to explore the extent of variation within and between the voices of 20 male speakers aged 18-24, from two UK dialects. The project compares Manchester and Newcastle English speakers, using the ESRC funded Northern Englishes Data, collected by Haddican and Foulkes (2017). Here, there is potential to explore speaker accent variation alongside intra-speaker variation in one dialect.

Three 30-second spontaneous speech samples have been extracted for each speaker at three different time points from within the same recorded spoken interaction. This allows us to maintain our focus on intra-speaker variation by observing how robust different speaker modelling methods are to a simple form of within-speaker variation. Although a lot of forensic speech science research is concerned with intra-speaker variation across non-contemporaneous speech samples (see Rose, 2015), we propose that it is still of great value to also inspect intra-speaker variation within the same spontaneous speech event, similarly to that of Enziger & Morrison (2012) and Ross et al (2019), here differentiating with the use of cluster analyses.

To further investigate this form of intra-speaker variation, the present project compares computational models of speakers' speech samples with phonetically-derived representations of the same speakers' samples. They are compared by subjecting the different speaker models and representations to cluster analyses which expose the level of similarity between speakers. We compare vowel formants and MFCCs in relation to how these speakers cluster at these different time points within the same spoken interactions. We also compare Gaussian Mixture Models (GMMs) and raw speech measurements in a similar way. We are interested in whether or not speakers more-or-less "cluster" in the same ways across these different time points. We are also interested in how the different ways of modelling speakers affect the clustering patterns. We expect the findings to not only shed light on the robustness of the different speaker modelling methods across time, but we also expect to learn more about intra-speaker variation in spontaneous speech and across dialects.

References

Enzinger, E., & Morrison, G. S. (2012). The importance of using between-session test data in evaluating the performance of forensic-voice-comparison systems. In *Proceedings of the 14th Australasian International Conference on Speech Science and Technology* (pp. 137-140).

Haddican, William and Foulkes, Paul (2017). *A comparative study of language change in Northern Englishes.* [Data Collection]. Colchester, Essex: Economic and Social Research Council.

Rhodes, R. W. (2012). *Assessing the strength of non-contemporaneous forensic speech evidence* (Doctoral dissertation, University of York).

Ross, S., Earnshaw, K., & Gold, E. (2019, August). A Cautionary Tale for Phonetic Analysis: The Variability of Speech Between and Within Recording Sessions. In *19th International Congress of the Phonetic Sciences* (pp. 3090-3094). Australasian Speech Science and Technology Association Inc..

Rose, P. (2015, April). Forensic voice comparison with monophthongal formant trajectories-a likelihood ratio-based discrimination of "schwa" vowel acoustics in a close social group of young Australian females. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4819-4823). IEEE.