

Signs of Partner Proximity in Formant Measurements: A First Look at the Relevance of Entrainment on Voice Identification Analysis

Ellen Grand and Georgina Brown

Linguistics and English Language, Lancaster University, UK
{e.grand|g.brown5}@lancaster.ac.uk

Members of Organised Criminal Groups (OCGs) are regularly persons of interest to security and law enforcement agencies, and so may be subjected to identification and investigation efforts. Voice analysis is one such identification method and is increasingly recognised by security agencies as a vital piece of intelligence.

This work examines one form of intra-speaker variation that is anticipated to be particularly relevant to recordings obtained in investigations of OCGs. Specifically, this work inspects how the relationships between speakers affect their communicative strategies, and how, in turn, these effects could go on to influence voice identification analyses. In order to address this, this study connects existing scientific literature on accommodation and entrainment to navigate their consequences on widely used voice identification methods. *Partner proximity* (Levitan & Hirschberg, 2011:3081) has been identified as a suitable framework through which to measure these effects. Partner proximity tracks the similarity between two speakers while they converse with one another, while comparing the degree of that similarity to when those speakers converse with others.

Building on the work of Earnshaw (2020), phase one of the current thesis examines signs of partner proximity in the formant frequency measurements taken from the FACE, TRAP and DRESS vowels in 15 paired tasks (30 speakers total) from the West Yorkshire Regional English Database (WYRED; Gold, Ross & Earnshaw, 2018). The vowels were examined collectively to ensure there were a suitable number of vowels to run statistical analysis on, and additionally, to look at a representation of the vowel space. Each task consisted of approximately 20 minutes of speech from which the first 10 clear realisations of each relevant vowel token were extracted; from these, formant frequency measurements were taken at the midpoint of the vowel. Euclidean distances and t-tests have been combined to infer the proximity between speakers and assess whether entrainment is taking effect in the speech of those speakers. The way in which this is approached within this methodology is to determine the similarity between speakers across different speaking sessions, and then assess the similarity between speakers when they speak to one another.

In general, the entrainment literature makes use of acoustic features that are frequent, and some experiments have incorporated the time sequencing of the acoustic features (Levitan & Hirschberg, 2011; Borrie, Lubold & Pon-Barry 2015; Wynn & Borrie, 2022). As the aims of this thesis are influenced by forensic applications, the features used in this study (formant frequencies of a small selection of vowels) align with those used to carry out forensic voice comparison analyses. Therefore, the current thesis adapts some of the methodology seen in entrainment studies to bring these features into focus whilst dealing with real life conundrums such as some loss of temporal detail.

The key findings from this first phase showed indications of some phonetic movement between speakers when conversing with one another. When considering the results seen in Earnshaw (2020), the current work yielded results that align with those findings. In addition, F0 measurements have been processed to produce Euclidean distances and t-test scores to contribute towards validation of the methodology as well as replicate previous works in entrainment and in forensic scenarios (Levitan & Hirschberg, 2011; Kruyt, 2023).

This initial study lays the groundwork to extend measures of partner proximity to other voice identification features and methods - specifically to speaker embeddings used within automatic speaker recognition.

References

- Earnshaw, K. (2020). *A forensic phonetic investigation of regional variation and accommodation in West Yorkshire* (Doctoral thesis, University of Huddersfield).
- Kruyt, J., de Jong, D., D'Ausilio, A., & Beňuš, Š. (2023). Measuring prosodic entrainment in conversation: A review and comparison of different methods. *Journal of Speech, Language, and Hearing Research*, 66(11), 4280–4314.
- Borrie, S. A., Lubold, N., & Pon-Barry, H. (2015). Disordered speech disrupts conversational entrainment: A study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. *Frontiers in Psychology*, 6, 1187
- Levitan, R., & Hirschberg, J. (2011, August). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech 2011* (pp. 3081–3084).
- Gold, E., Ross, S., & Earnshaw, K. (2018, September). The 'West Yorkshire Regional English Database': Investigations into the generalizability of reference populations for forensic speaker comparison casework. In *Proceedings of Interspeech 2018: Speech Research for Emerging Markets in Multilingual Societies* (pp. 2748–2752).