Speaker- and group-specific information in formant dynamics: a forensic perspective

Vincent Hughes & Paul Foulkes
Department of Language and Linguistic Science, University of York
vincent.hughes@york.ac.uk

There is growing interest in analysing speech as a time-varying signal, such as modelling the dynamics of vowel formant trajectories rather than traditional examinations at static midpoints. While the attention to dynamic properties is relatively new in phonetics and sociolinguistics, formant trajectories have been used widely for some time in forensic voice comparison (FVC) as an effective means of discriminating between speakers within homogeneous populations (i.e. within regional and social groups). The initial motivation for testing the speaker-specificity of formant dynamics was based on a broad assumption about the cognitive representation of segments (Nolan 1997, McDougall, 2004, 2006), described succinctly by Nolan (1997): (phonetic/phonological) *targets* are specified by a grammar and thus shared by all speakers who use that language. By contrast, the phonetic pathways between targets are not specified at the level of the language, but individually acquired and administered. It therefore makes sense to examine the pathways between targets (e.g. onsets and offsets of vowels), rather than the targets themselves (vowel midpoints), in order to identify speaker-specific phonetic patterns.

Such a view is commensurate with various models of phonology and of speech production/perception, but not all: usage-based models, for example, permit in principle any aspects of phonetic and phonological structure to be learned and represented at a cognitive level, without specifying that all speakers necessarily share solely an abstract representation. To test these assumptions we conducted a series of studies which examine the speaker- and group-specific information encoded in the dynamics of formant trajectories. Using the F1, F2, and F3 trajectories of PRICE fitted with polynomial equations extracted from sociolinguistic recordings from four regional varieties of British English we conducted forensic speaker-discrimination tests. Consistent with previous studies, error rates (i.e. misclassifications) were considerably lower when using dynamic information compared with static midpoint data. However, the polynomial coefficients were also able to predict a speaker's regional background above chance level. The same tests were also conducted using the formant trajectories of FACE extracted from the ONZE corpus (REF) for speakers from different class and age groups. Again, formant dynamics outperformed static midpoints in terms of speaker discrimination, but could also predict social group.

These results show that considerable between-speaker variation is present in formant dynamics, as Nolan et al. predict, but also that there is shared knowledge of fine-grained phonetic information about the dynamics of vowel production within regional and social groups. These data, therefore, offer some support for a model of learning and representation that includes more detail than just phonemic targets.

References

McDougall, K. (2004) Speaker-specific formant dynamics: an experiment on Australian English /aI/. International Journal of Speech, Language and the Law 11(1), 103-130.

McDougall, K. (2006) Dynamic features of speech and the characterisation of speakers: towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law* 13(1), 89-126.

Nolan, F. (1997) Speaker recognition and forensic phonetic. In W. J. Hardcastle and J. Laver (eds.) *The Handbook of Phonetic Sciences*. Oxford: Blackwell. pp. 744-767.