

Speaker-variance ratios in forensically-realistic vowel formant data: Normalising for consonantal context

Frantz Clermont

J. P. French Associates & University of York, Forensic Speech & Acoustics Laboratory, U.K.
akustikfonetiks@yahoo.com.au

In a previous study (Clermont, French, Harrison and Simpson, 2008) we showed that the upper formant-frequencies, F3 in particular, carry the bulk of the speaker information contained in central segments of short vowels (as in the words KIT, DRESS, TRAP, LOT) of Standard Southern British English (DyViS corpus: Nolan, McDougall, de Jong and Hudson, 2006). This finding is based on 25 young-male speakers' recordings of unscripted telephone conversation, in which all phonetic occurrences of a given vowel were marked centrally and their individual formants (F1, F2, F3, F4) were averaged to represent that vowel. The use of per-speaker averages of vowel formants facilitated our statistical analyses within and between speakers, but it effectively blurred any consonantal effects that might have persisted as far as the vowel centres.

Our 2008 experiments left pending, therefore, the question of whether co-articulatory effects on the selected vowel segments might have inflated the between-to-within speaker variance in favour of the upper formants. Simpson (2008) shed some light on this question by performing speaker-discrimination analyses for each place- and manner-of-articulation category of the vowels' initial and final consonantal contexts. The upper formants still fared relatively better than the lower ones, but the discrimination ability of all four formants improved as a result of phonological categorisation. Taking context into account is therefore beneficial, and suggests that taking advantage of systematic context effects might also be effective.

To capture systematic modifications of a speaker's vowel-formant space, we define a formant's values for a *set of vowels* at a given time location and in a given context as a *vowel-formant ensemble* (VFE), and adopt the hypothesis (Broad and Clermont, 2002) that *the relative spacing of the vowels within a VFE is mostly invariant with respect to context or location within a syllable*. This property of VFEs implies a *family* structure tied by *linear-scaling* relationships, which here are quantified from our 2008 data and then inverted to normalise with respect to place-of-articulation contexts. Using speaker-variance ratios we show that context normalisation enhances the discriminating power of F2 to a competitive level vis-à-vis F3. We argue that this result confirms the effectiveness of our approach, and we discuss its forensic implications.

References

- Broad, D.J. and F. Clermont (2002). Linear scaling of vowel-formant ensembles (VFEs) in consonantal contexts. *Speech Communication*, **37**: 175-195.
- Clermont, F., P. French, P. Harrison and S. Simpson (2008). Population data for English spoken in England: A modest first step. *Abstract Proc. Annual Conf. of the International Association for Forensic Phonetics and Acoustics (IAFPA)*, Lausanne, Switzerland, July 2008.
- Nolan, F., K. McDougall, G. de Jong and T. Hudson (2006). Introducing DyViS: A dynamic study of British English for forensic purposes. *Abstract Proc. Annual Conf. of the International Association for Forensic Phonetics and Acoustics (IAFPA)*, Gothenburg, Sweden, July 2006.
- Simpson, S. (2008). Testing the speaker discrimination ability of formant measurements in forensic speaker comparison cases. *Unpublished MSc Thesis*, University of York, UK.