A linear-scaling method for normalising vowels in various consonantal contexts

Frantz Clermont
J.P. French Associates, Forensic Speech & Acoustics Laboratory, York, United Kingdom
akustikfonetiks@yahoo.com.au

It is common practice in forensic-comparison tasks to select formant-frequency measurements from the most stationary parts of vowel segments. There is, however, well-documented evidence from the acoustic-phonetics literature indicating that the putative targets of vowels are rarely attained because of co-articulation with preceding and following sounds, and that no particular time slice (or frame) of a vowel segment in consonantal context can be assumed to have complete immunity. It is therefore desirable to try and account (or normalise) for context effects with a view towards a more robust characterisation of speaker differences. Here we investigate a new approach to context normalisation.

The approach involves a shift in perspective from the traditional frame-to-frame evolution of a consonant's family of vowel-formant transitions towards a "vowel-axis" representation where frame-to-frame connections are removed and replaced by vertical lines connecting the formants for the ensemble of different vowels within the same frame. This new perspective helps pose the question of how vowel-to-vowel spacings within an ensemble vary from one frame to the next and from one consonant's family of formant transitions to another. Our hypothesis for a given speaker and a given formant is that the *relative spacings within an ensemble of vowels are invariant and unaffected by consonantal context or by time location within a syllable* (Broad, D J. & F. Clermont, "Linear scaling of vowel-formant ensembles (VFEs) in consonantal contexts", Speech Communication*,* 37, 175-195, 2002).

Under this hypothesis, all of a speaker's vowel-formant ensembles are geometrically *similar* within and between contexts, and therefore tied to one another by *linear scaling* of the speaker's average ensemble. Inverting the linear-scaling factors yields a normalised vowel space, in which all instances of a vowel's formant are standardised with respect to that vowel's average formant. Using a dataset of formant transitions from 7 vowels in 7 CVd contexts (C = /h, b, d, g, p, t, k/) recorded 5 times at 1 sitting by 4 adult-male, native speakers of Australian English, we give evidence in support of the similarity principle outlined above, and demonstrate the effectiveness of linear scaling as a method for constructing a speaker's vowel space in which context effects are substantially reduced and speaker-dependent patterns are thus enhanced. The method is also computationally inexpensive as it requires only simple operations such as taking averages and fitting straight lines.