Abstract Proceedings, Annual Conference of the International Association for Forensic Phonetics & Acoustics, 3-5 August 2009, University of Cambridge, UK

Linear-scaling effects of co-articulation in the vowel space

Frantz Clermont
J. P. French Associates, Forensic Speech and Acoustics Laboratory, York, UK
akustikfonetiks@yahoo.com.au

The work reported here is part of a larger project whose aim is to gain a principled understanding of the acoustic-phonetic variability resulting from co-articulation. In this quest we target systematic modifications of the vowel-formant space, which are studied under the hypothesis that *the relative positioning of vowels for a given speaker is invariant and unaffected by consonantal context or location within a syllable*. The hypothesis brings into focus a "vowel-axis" configuration, whose implementation hinges on a new concept introduced by Broad and Clermont (*Speech Communication, 37: 175-195, 2002*) ─ the *vowel-formant ensemble* (VFE). Figure 1 depicts the VFE as a vertical array comprising the set of formants (e.g., F2), which are realised for a set of vowels at a given time slice and in a given context.

The characterisation of context effects is then approached by finding how VFEs for different time slices and contexts are scaled in relation to one another. Under the hypothesis stated above, all of a speaker's VFEs are similar to and thus linearly-scaled copies of one another. Figure 2 portrays the linear-scaling relation as a geometric similarity, such that any vertical slice is subdivided in the same proportion. Each VFE has its own scale relative to the mean VFE. The smaller the scale, the more compressed are the inter-vowel spacings (e.g., towards consonantal boundaries); the larger the scale, the more dilated are the spacings (e.g., near syllable centres).

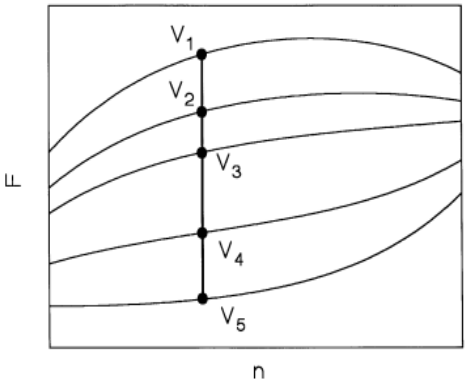

Figure 1: Hypothetical family of F2 transitions for a set of vowels in some Consonant-Vowel-Consonant (CVC) context. The abscissa represents time slices. The vertical line marks a VFE.

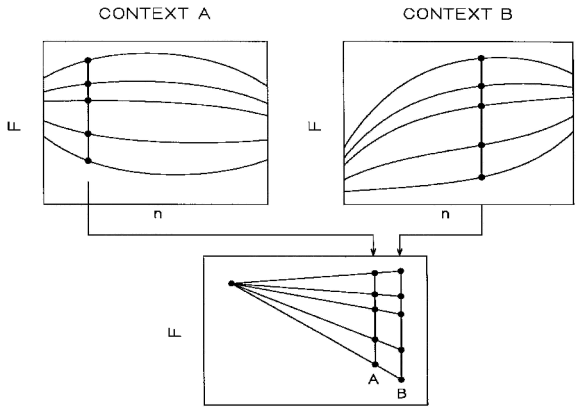

Figure 2: Top panels show families of hypothetical F2 transitions for a set of vowels in 2 hypothetical contexts A and B. From each context a VFE is selected and transferred in the bottom panel. Ensemble similarity is portrayed by the common intersection of all the lines connecting identical vowels in the 2 VFEs.

The linear-scaling of VFEs requires simple operations such as taking averages and fitting lines, which we demonstrate with formant data measured from 7 vowels in 7 CVd contexts (C = /h, b, d, g, p, t, k). The CVd syllables were recorded 5 times at one sitting by 4 adult-male, native speakers of Australian English. The application of VFE-scaling to our data gives results that do exhibit the similarity phenomenon and show its effectiveness for (1) tracking the effects of any single context; (2) comparing the effects of different contexts; and (3) normalising for context.