

Statistical models in speech production

Phil Hoole and Christian Geng

Statistical models have played a pervasive role in the analysis of articulatory data and modelling of speech production. There are probably two main motivations for such approaches. The first one is simple: Data reduction. Articulatory data can be derived, for example, from multiple fleshpoints (EMA, microbeam, Optotrak), from distances along gridlines applied to the vocal tract in 2 or 3 dimensions (cineradiography, MRI), or from nodes of a deformable mesh applied to the face (video). All these approaches can result in large numbers of raw coordinates. Apprehension of significant patterns in the data can be greatly facilitated if a small number of components can be derived that capture a large proportion of the variance. The basic properties of Principal Component Analysis (PCA) will be used to illustrate this. This provides a useful framework for the main part of the presentation, which is concerned with the second, and more far-reaching, motivation for statistical models; this is to use a data-driven approach to uncover the number and nature of the underlying dimensions of articulatory control. Despite a multiplicity of muscles at the periphery it is a common assumption that speakers organize their behaviour around a small number of independent functional degrees of freedom (and plausible control strategies for an articulatory model invariably take the same approach). It is a problem common to many fields of enquiry that the underlying behavioural dimensions cannot be observed directly. For reasons to be discussed, simply throwing a large amount of articulatory data at a PCA may not give the most revealing results. We will present a series of case studies from two main traditions that go beyond simple PCA: (1) guided factor analysis, introduced by Maeda. This allows the modeller to take basic knowledge about the articulatory apparatus into account, e.g. tongue is affected by jaw movement; lips and larynx may be correlated in rounded vowels, but are clearly anatomically independent. This approach has mostly been used for comprehensive models of individual speakers. (2) The PARAFAC approach of Harshman. This is an inherently multispeaker approach using speaker as a third mode to constrain the factor solution in interesting ways. Both these approaches have now been applied to a wide variety of data; advantages, limitations and practical considerations will be discussed. Finally, a brief outlook towards further models of potential interest for phonetics will be given.