Analysing tongue contours with multivariate Generalised Additive Models

Stefano Coretta[1], George Sakr[2]

[12]University of Edinburgh, UK,
[1]s.coretta@ed.ac.uk, [2]georges.sakr@ed.ac.uk

Keywords: ultrasound tongue imaging, generalized additive models, multivariate

Background. Ultrasound Tongue Imaging (UTI) has seen a lot of recent advancements, among which more portable equipment set ups and faster and more accurate tongue contour tracing techniques. In particular, the implementation of a DeepLabCut (DLC) model for markerless pose estimation (Mathis et al. 2019) of tongue imaging data by Wrench and Balch-Tomes (2022) has afforded us with new ways to quantify tongue articulator points position and movement. The DLC model tracks the position of 11 points on the midsagittal surface of the tongue thus moving away from the use of fixed fan lines, which only allow for a vectorised movement along each fan line corresponding to the point where the tongue surface crosses the fan line. The DLC tracked points move freely in 2D space, so that for each point we obtain X and Y coordinates. This way of tracking individual points on the tongue better captures their movement than using a fixed fan coordinate system.

Problem. However, the standard method of analysing tongue contours using Generalised Additive Models (GAMs) in which the X coordinate is replaced by the fan line number and the Y corresponds to the Y coordinate can no longer be used, since the method returns both X and Y coordinates for each point and no fan lines and since, especially towards the back of the tongue, the contour tends to "curl" on itself.

Solution. The solution is to use multivariate GAMs that model X and Y coordinates as a multivariate Gaussian distribution. This method allows us to model tongue contours that curl on themselves and to model points on the tongue surface that move freely on the 2D space, rather than just along a straight fan line.

Example. We will discuss the pros and cons of using multivariate GAMs by example with DLC tracked tongue data from a study on pharyngealised consonants in Lebanese Arabic and a study on tongue root displacement in voiced stops in Italian and Polish. The figure on the side shows the prediction plot generated by a multivariate GAM applied to the Lebanese Arabic data. Each speaker is plotted on one row and each column shows the 95% CIs of the predicted tongue contours for plain vs emphatic consonants followed by one of five vowels. The multivariate GAM allows us to observe that while some speakers produce retracted tongue root, others do not.

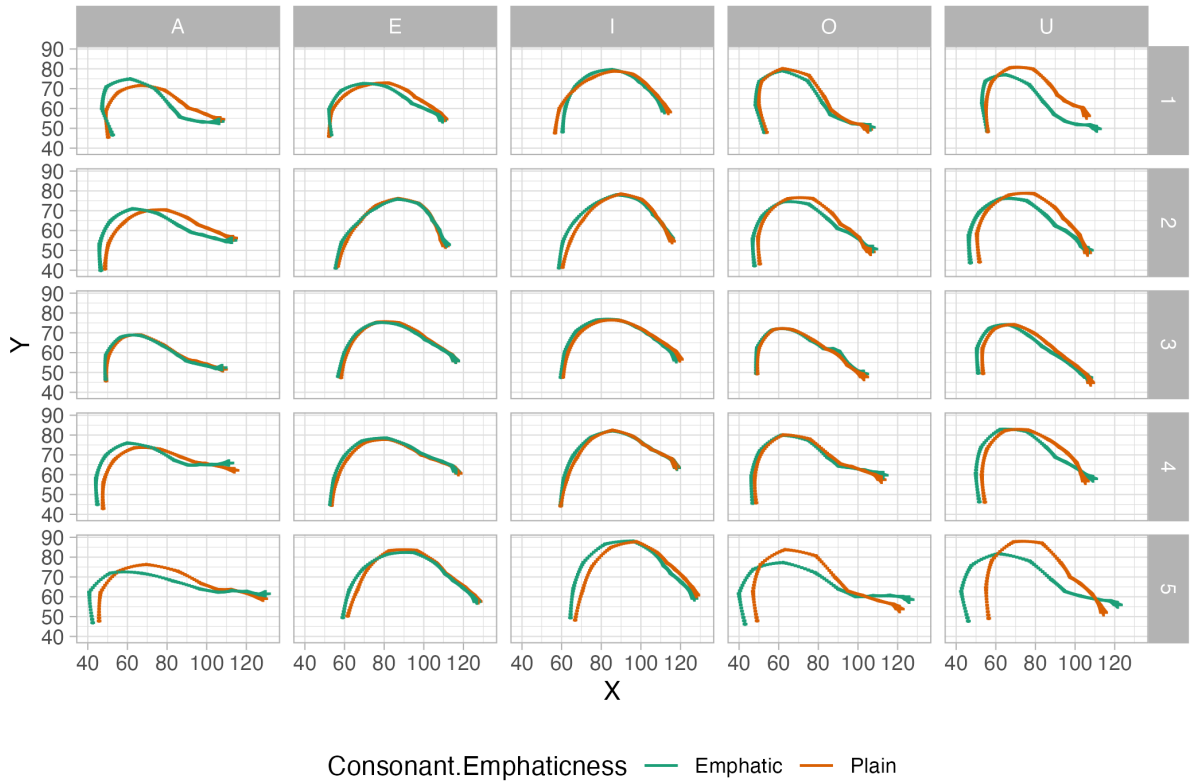

Figure 1 Predicted tongue contours from a multivariate GAM.