Assessing midsaggital tongue contours in polar coordinates using generalised additive (mixed) models

Stefano Coretta

1 Introduction

Since the publication of the seminal paper by Davidson (2006), statistical modelling of whole tongue contours has been dominated by the use of Smoothing Splines Analysis of Variance (SSANOVA), which undoubtedly brought a conspicuous advancement for understanding the articulation of speech sounds. Some of the limitations of modelling tongue contours with SSANOVA is that separate models are needed for different phonetic contexts even within a single speaker, and secondly SSANOVA as implemented in most studies does not include random effects (which have been shown to be extremely important XXX). The general difficulty felt with SSANOVA and tongue contours has favoured alternative methods like Principal Component Analysis.

On the other hand, developments in the statistical and programming world have seen the emergence of a highly flexible technique, Generalised Additive models (GAMs). GAMs have being increasingly adopted in linguistics as a means to model dynamic speech data. Indeed, Sóskuthy (2017) explicitly suggests the use of GAMs with tongue contours

This paper introduces an implementation of GAMs with tongue contours using polar coordinates. The use of polar GAMs is illustrated with ultrasound tongue imaging data comparing voiceless and voiced stops. The R package `rticulate` has been developed to facilitate the use of the model, and it is briefly introduced here.

1.1 Ultrasound tongue imaging

Ultrasound imaging is a non-invasive technique for obtaining an image of internal organs and tissue. 2D ultrasound imaging has been successfully used for imaging sections of the tongue surface (for a review and applications in field settings, see Gick 2002). To image the tongue, the transducer placed in contact with the sub-mental triangle (the area under the chin) aligned either with the mid-sagittal or the coronal plane. The ultrasonic waves propagate from the transducer in a radial fashion through the aperture of the mandible and get reflected when they hit the air above the tongue surface. This 'echo' is captured by the transducer and translated into an image like that in Figure 1.

1.2 Generalised Additive models

Generalised additive modelling is a more general form of non-parametric modelling that allows fitting non-linear as well as linear effects. Generalised additive models, or GAMs, are built with smoothing splines. Smoothing splines are also at the heart of SSANOVA. In

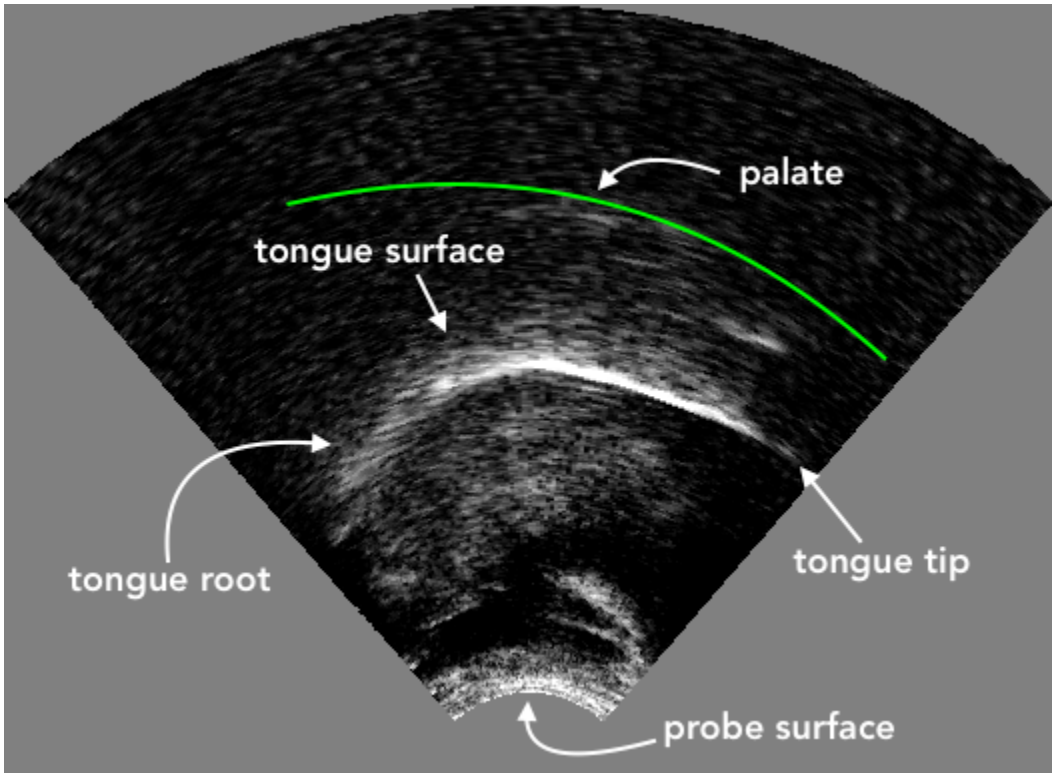

Figure 1: An ultrasound image of the tongue

GAMs, however, smoothing splines maximise the fit to the data while being constrained by a smoothing penalty estimated from the data itself which prevents overfitting. GAMs are powerful and flexible models that can deal with non-linear data efficiently. Moreover, random effects can also be implemented in GAMs, as generalised additive mixed models (or GAMMs).

Tongue contours as extracted from ultrasound imaging can be efficiently model using GAM(M)s. The random effects part of GAMMs also constitutes an improvement over traditional SSANOVA, which usually implements fixed effects only. Moreover, GAMMs can reduce autocorrelation, for example by allowing separate smooths to be fitted to the individual contours/trajectories, or by including a first-order autoregression model. For a technical introduction to GAM(M)s, see Zuur (2012) and Wood (2017).

1.3 Polar coordinates

Mielke (2015) and Heyne & Derrick (2015b), Heyne & Derrick (2015a) have shown the benefits of using polar coordinates of the tongue contours rather then cartesian coordinates. Polar coordinates are constituted by pairs of radius and angular values, which define a point relative to the origin of the coordinate system. The point is describes with a radius, which corresponds to the radial distance from the origin, and the angle from the reference radius. Tongue contours, due to their shape, tend to have increasing slope at the left and right edges, in certain cases tending to become almost completely vertical. The almost

verticality of the contours has the effect of increasing the variance of the fitted contours (and hence increased confidence intervals), and is some cases it even generates uninterpretable curves. When tongue contours are expressed with polar coordinates, on the other hand, the variance is reduced and the fitted contours generally reflect more closely the underlying tongue shape. Mielke has implemented a series of R (R Core Team, 2018) functions for fitting polar SSANOVAs to tongue contours in cartesian coordinates. Plotting is subsequently obtained by reconverting the coordinates to cartesian.

2 Polar GAM(M)s

Having shown the benefits of using GAMMs with dynamic data and polar coordinates, polar GAMMs for modelling tongue contours are introduced here. Polar GAMMs are GAMMs fitted to tongue contours using polar coordinates. As with SSANOVA, the coordinates of the fitted contours are converted to cartesian coordinates for plotting. A polar GAM is constructed as follows: the radius coordinates is the outcome variable, a smooth term over the angular coordinates is the predictor. The smooth term enables modelling the non-linearities of the contours. Predictors such as consonant or vowel type, or speech rate, can be specified in the model. The predicted polar coordinates that are returned by the model can then be converted to cartesian coordinates using the cartesian coordinate of the origin that defines the polar system. The polar origin is either known or estimated from the data, depending on the ultrasonic system used.

To illustrate the use of polar GAMs, an example will given in the following sections. The main properties of the R package `rticulate` will also be discussed in relation to the experiment. The function `polar_gam()` accepts cartesian coordinates, which are converted into polar using a user specified origin or the origin estimated from the data. The GAM is fitted on the polar coordinates and the predicted values are converted back to cartesian using the same origin for plotting.

2.1 Data collection and processing

Synchronised audio and ultrasound tongue imaging data have been recorded from 4 speakers of Italian. An Articulate Instruments Ltd™ set-up was used for this study (**??**). The ultrasonic data was collected through a TELEMED Echo Blaster 128 unit with a TELEMED C3.5/20/128Z-3 ultrasonic transducer (20mm radius, 2-4 MHz). A synchronisation unit (P-Stretch) was plugged into the Echo Blaster unit and used for automatic audio/ultrasound synchronisation. A FocusRight Scarlett Solo pre-amplifier and a Movo LV4-O2 Lavalier microphone were used for audio recording. The acquisition of the ultrasonic and audio signals was achieved with the software Articulate Assistant Advanced (AAA, v2.17.2) running on a Hawlett-Packard ProBook 6750b laptop with Microsoft Windows 7. Stabilisation of the ultrasonic transducer was ensured by using a headset produced by Articulate Instruments Ltd™ (2008).

Before the reading task, the participant's occlusal plane was obtained using a bite plate (Scobbie et al., 2011). The participants read nonce words embedded in the frame sentence *Dico* ____ *lentamente* 'I say ____ slowly'. The words follow the structure $C_1V_1C_2V_2$, where

C_1 = /p/, V_1 = /a, o, u/, C_2 = /t, d, k, g/, and $V_2 = V_1$. Each speaker repeated the stimuli six times.

Spline curves were fitted to the visible tongue contours using the AAA automatic tracking function. Manual correction was applied in those cases that showed clear tracking errors. The time of maximum tongue displacement within consonant closure was then calculated in AAA following the method in Strycharczuk & Scobbie (2015), described in what follows. A fan-like frame consisting of 42 equidistant radial lines was used as the coordinate system. The origin of the 42 fan-lines coincides with the centre of the ultrasonic probe, such that each fan-line is parallel to the direction of the ultrasonic signal. Tongue displacement was thus calculated as the displacement of the fitted splines along the fan-line vectors. The time of maximum tongue displacement was the time of greater displacement along the vector that showed the greatest standard deviation. The vector search area was restricted to the portion of the splines corresponding to the tongue tip for coronal consonants, and to the portion corresponding to the tongue dorsum for velar consonants.

The cartesian coordinates of the tongue contours were extracted from the ultrasonic data at the time of maximum tongue displacement (always within C2 closure). The contours were subsequently normalised within speaker by applying offsetting and rotation relative to the participant's occlusal plane (Scobbie et al., 2011). The dataset is thus constituted by x and y coordinates of the tongue contours that define respectively the horizontal and vertical axis. The horizontal plane is parallel to the speaker's occlusal plane.

2.2 Fitting a polar GAM

GAMMs can be fitted in R with the `gam()` function from package `mgcv` (Wood, 2011, 2017). `bam()` is a more efficient function when the datasets has several hundreds observations. The package `rticulate` has been developed as a wrapper of the `bam()` function to be used with tongue contours. The special function `polar_gam()` can fit any specified GAM model to tongue contours coordinates, using the same syntax of `mgcv`. The function accepts tongue contours either in cartesian or polar coordinates. In the first case, the coordinates can be transformed into polar before fitting. The function `plot_polar_smooths()`, used for plotting the estimated contours, converts the coordinates back into cartesian.

A GAM in R can be specified with a formula that uses the same syntax of `lme4`, a commonly used package for linear mixed-effects models. The `mgcv` package allows to specify smoothing spline terms with the function `s()`. This function takes the term along which a spline is created (for example, a time series, or x-coordinates). Within `s()` the user can choose, among other arguments, what kind of spline to use, and the grouping factor (namely, the factor with the levels to be compared). For a more in-depth introduction to GAM(M)s in R targeted to linguists, see Sóskuthy (2017) and Wieling (2017).

Due to differences in the placement of the probe and the speaker's anatomy, different portions of the tongue are likely to be imaged across speakers. For this reason, it is recommended to fit separate models for each participant, rather then aggregate all of the data in a single model.

As means of illustration, the following paragraphs will show how to fit a polar GAM with data from one of the 4 Italian speakers.

We can start from a simple model in which we test the effect of C2 place, vowel, and

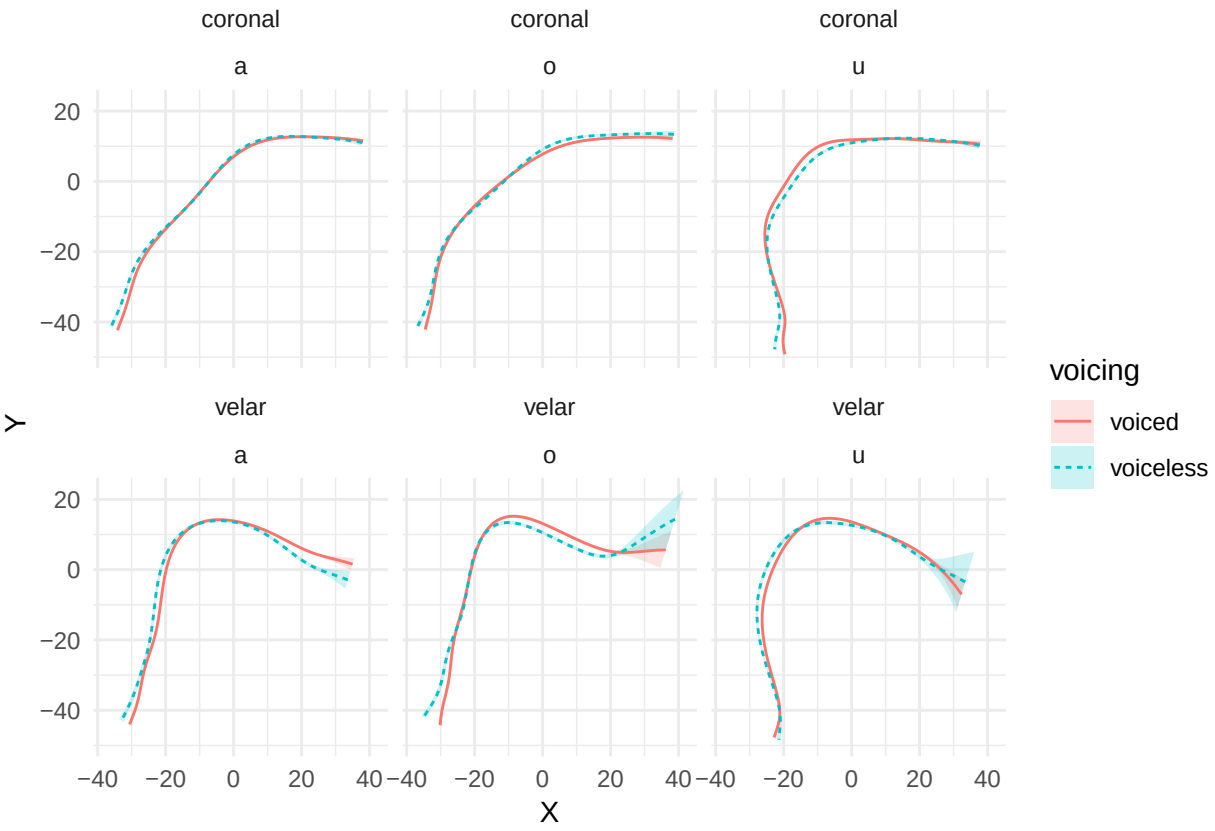

Figure 2: Estimated tongue contours of IT01 depending on C2 place, vowel and C2 voicing.

voicing on tongue contours. Modelling different contours for each combination of the three predictors can be achieved with the `by` argument in the difference smooth function and including the related parametric term. `vc_voicing` is an ordered factor which specifies for each contour the combination of C2 place, vowel, and voicing. The following code fits the specified model to the contour data of IT01. When running the code, the coordinates of the estimated origin used for the conversion to polar coordinates are returned. The author can optionally specify those manually, the contours are not in a fan-like system (like the data exported from AAA).

```
## The origin is x = 14.3900999664996, y = -65.2314226131983.
```

The function `plot_polar_smooths()` can be used to plot the estimated contours. The shaded areas around the estimated contours are 95% confidence intervals. Note that, differently from SSANOVA, statistical significance can't be assessed from the overlapping (or lack thereof) of the confidence intervals.

One way of assessing significance is to compare the ML score of the full model against one without the relevant predictor, using the function `compareML()` from the `itsadug` package. Both the parametric term and the difference smooth need to be removed in the null model.

```
## The origin is x = 14.3900999664996, y = -65.2314226131983.
```

```
## it01_gam_0: Y ~ s(X)
##
## it01_gam: Y ~ vc_voicing + s(X) + s(X, by = vc_voicing)
##
## Chi-square test of ML scores
## -----
##        Model     Score Edf Difference     Df  p.value Sig.
## 1 it01_gam_0 12395.227   3
## 2   it01_gam  7423.356  36   4971.871 33.000  < 2e-16  ***
##
## AIC difference: 10258.62, model it01_gam has lower AIC.
```

References

Articulate Instruments Ltd™. 2008. Ultrasound stabilisation headset users manual: Revision 1.4. Edinburgh, UK: Articulate Instruments Ltd.

Davidson, Lisa. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120(1). 407–415. doi:10.1121/1.2205133.

Gick, Bryan. 2002. The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association* 32(02). 113–121.

Heyne, Matthias & Donald Derrick. 2015a. Benefits of using polar coordinates for working with ultrasound midsagittal tongue contours. *The Journal of the Acoustical Society of America* 137(4). 2302–2302.

Heyne, Matthias & Donald Derrick. 2015b. Using a radial ultrasound probe's virtual origin to compute midsagittal smoothing splines in polar coordinates. *The Journal of the Acoustical Society of America* 138(6). EL509–EL514. doi:10.1121/1.4937168.

Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *The Journal of the Acoustical Society of America* 137(5). 2858–2869. doi:10.1121/1.4919346.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Scobbie, James M., Eleanor Lawson, Steve Cowen, Joanne Cleland & Alan A. Wrench. 2011. A common co-ordinate system for mid-sagittal articulatory measurement. In *QMU CASL Working Papers*, 1–4.

Sóskuthy, Márton. 2017. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. arXiv preprint arXiv:1703.05339.

Strycharczuk, Patrycja & James M. Scobbie. 2015. Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English. In *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5.

Wieling, Martijn. 2017. Generalized additive modeling to analyze dynamic phonetic data: a tutorial focusing on articulatory differences between l1 and l2 speakers of english. *The Mind Research Repository (beta)* (1). doi:10.1016/j.wocn.2018.03.002.

Wood, Simon. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73(1). 3–36.

Wood, Simon. 2017. *Generalized additive models: An introduction with r*. Chapman and Hall/CRC 2nd edn.

Zuur, Alain F. 2012. *A beginner's guide to generalized additive models with R*. Highland Statistics Limited: Newburgh.