From SS-ANOVA to GAMMs: Accounting for within and between-subject variation using generalized additive mixed models on ultrasound tongue contours

Jalal Al-Tamimi (Newcastle University); Matthias Heyne (University of Pittsburgh); Donald Derrick (University of Canterbury)

Background: Speakers vary in terms of the shape and morphology of their vocal tracts impacting on the size and curvature of their tongue (Lammert, Proctor & Narayanan, 2013 a & b). Often this leads to issues in generalizing results from Ultrasound Tongue Imaging (UTI) across speakers. These issues are particularly marked when using SS-ANOVA – the standard in UTI (Davidson, 2006). By using Generalized Additive Mixed models (GAMMs; Wood, 2017), accounting for within and between-subject differences is possible. GAMMs are gaining popularity in the community with a few studies applying it to UTI (Al-Tamimi, 2018; Heyne, Derrick & Al-Tamimi, 2019). In this paper, we first demonstrate how SS-ANOVAs lead to a generalization issue, propose a solution based on a transformation technique (cf. Heyne, 2016), and compare performance with GAMMs using both transformed and raw data. This will inform the field on the best ways in dealing with the full tongue contours as obtained from UTI. **Method:** 20 speakers (10 New-Zealand English, NZE; 10 Tongan), with a mean age of 40.3 (SD = 18) for NZE and 27.2 (SD = 8.3) for Tongan, were recorded playing a brass instrument (Trombone; cf. Heyne, 2016). UTI were captured using a Healthcare Logiq e (version 11) ultrasound machine with an 8C-RS wide-band microconvex array 4.0–10.0 MHz transducer. Speakers produced a list of real words in both NZE and Tongan (803 and 1,154 tokens respectively), followed by the five notes (Bb2, F3, Bb3, D4 and F4) at four intensities (*piano*, *mezzopiano*, *mezzoforte*, and *forte*). The UTI data were manually traced using GetContours for MatLab with a total of 7,428 tongue contours of sustained notes obtained at a third of note duration (3,715 for NZE, 3,713 for Tongan). Data were first scaled by converting them into polar coordinates, with translation using the estimated transducer location (cf. Heyne & Derrick, 2015). Then, the speaker with the smallest vocal tract was used as a reference to rotate and scale tongue contours in both angular and radial coordinates to align the position of the average contour's highest points during the production of the vowel/i:/ (FLEECE in NZE). Splines were then estimated using either SS-ANOVAs or GAMMs on raw and transformed data. SS-ANOVAs were fit on all tokens matching the desired condition, while for GAMMs, the distance values (Rho; dependent variable) were modelled as function of the angular values (Theta; time series), adjusted by the interaction between the language, note and its intensity. Our random effect was speaker-adjusted by the interaction between note and intensity allowing for speaker-specific normalization of both distance and angle. **Results:** Figure 1 shows an example of average tongue contours in NZE and Tongan for the note Bb2 at *forte*, with confidence intervals (CI) and regions of significant differences highlighted, using both raw and transformed data. SS-ANOVA plots (Fig. 1, a and b) display comparatively large intervals of significant difference with scaling considerably affecting tongue contour shapes. GAMMs plots (Fig. 1, c and d) show similar tongue contours, with smaller intervals of significant difference and consistency in regions of differences across analyses, both of which indicate major differences at the back and the front of the tongue. Overall, for the raw data (Fig. 1, d), the regions of significance are slightly variable and lead to smaller significant regions detected. **Conclusion:** SS-ANOVAs, the standard in UTI analyses, suffer from a few shortcomings when it comes to generalizing the results across speakers. GAMMs provide a direct way of accounting for both within and between-subject variation. Figures obtained with GAMMs using raw data lead to the same conclusions as when using transformed data, although CI are different, with slightly less precise estimates. Our recommendation is to use a transformation technique that is adapted for each dataset and rely on GAMMs to model tongue contours with the appropriate random effects structure. When using GAMMS, raw data will lead to the same conclusions, albeit with slightly more variation and smaller regions of significant difference.

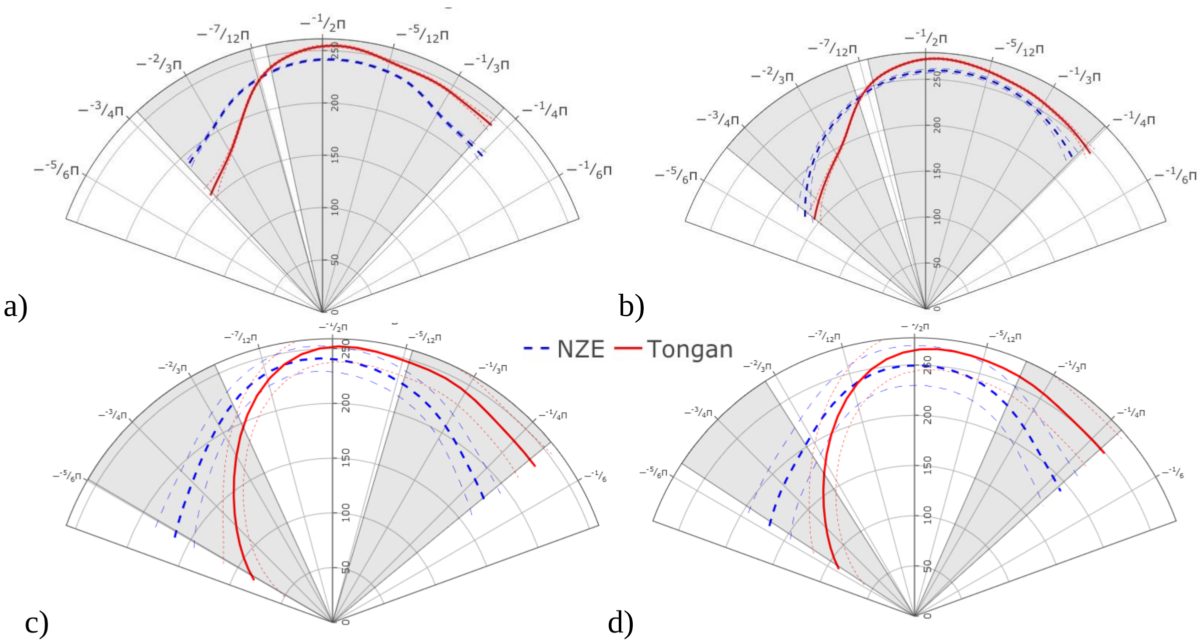

Figure 1: UTI using SS-ANOVA (top row) and GAMMs (bottom row) with (a and c) and without (b and d) transformation of the tongue contours for Bb2 *forte* in NZE and Tongan. Dashed lines indicate CI, while regions of significant difference are shown in gray.

References:

Al-Tamimi, J. (2018). A generalised additive modelling approach to ultrasound tongue surface: Quantifying retraction in Levantine Arabic back consonants. In Paper presented at “New Developments in Speech Sensing and Imaging - LabPhon16 Satellite event. Lisbon, Portugal. June 23
Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. JASA, 120(1), 407-415
Heyne, M. (2016). The influence of first language on playing brass instruments: an ultrasound study of Tongan and New Zealand trombonists. PhD thesis.
Heyne, M. and Derrick, D. (2015). Using a radial ultrasound probe’s virtual origin to compute midsagittal smoothing splines in polar coordinates. JASA, 138(6): EL509–EL514.
Heyne, M., Derrick, D., and Al-Tamimi, J. (2019). Native language influence on brass instrument performance: An application of generalized additive mixed models (GAMMs) to midsagittal ultrasound images of the tongue. Frontiers in Psychology, 10:2597.
Lammert, A., Proctor, M., and Narayanan, S. (2013a). Interspeaker variability in hard palate morphology and vowel production. JSLHR, 56: 1924–1933.
Lammert, A., Proctor, M., and Narayanan, S. (2013b). Morphological variation in the adult hard palate and posterior pharyngeal wall. JSLHR, 56: 521–530.
Wood, S. N. (2017). Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC, 2 edition.