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

Jalal Al-Tamimi[1] jalal.al-tamimi@newcastle.ac.uk, Matthias Heyne[2], & Donald Derrick[3]

[1]Newcastle University, UK; [2]University Pittsburgh, USA; [3]University of Canterbury, NZ

INTRODUCTION

- Speakers vary in terms of shape & morphology of vocal tract (cf. Heyne 2016)
  - Shape of hard palate & alveolar ridge impacts articulatory strategies (e.g., Dediu et al. 2017, 2019; Dediu & Moisik 2019; Lammert et al. 2013a&b), although acoustic consequences seem reduced due to individual adaption, e.g., for high front vowel /i/
  - Ratio of palatal & pharyngeal volumes influences vowel production (Fuchs et al. 2008; Lammert et al. 2013b)
  - Overall size of vocal tract -> more acoustical variability in females (Diehl et al. 1996; Simpson & Ericsdotter 2007; Weirich & Simpson 2014; Whiteside 2001) vs more articulatory variability in males (Simpson 2001 & 2002)

➢ Major impact on Ultrasound Tongue Imaging (UTI) due to lack of easily identifiable anatomical landmarks
  - Various normalization techniques have been proposed to try to account for these factors using...
  - Curvature of selected tongue shapes (Dawson et al. 2016, Ménard et al. 2012; Stolar & Gick 2013; Zharkova 2013a&b)
  - Relative articulatory height & fronting of a certain vowel tongue shape (Lawson & Mills 2014; Lawson et al. 2015; Noiray et al. 2014)

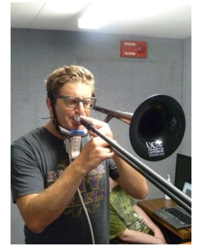
- Here, we hypothesize that an advanced statistical technique, generalized additive (mixed) models (GAMMs) can
  1) Account for *within* & *between-subject* variation
  2) Provide more accurate tongue contour estimates than smoothing spline analysis of variance (SSANOVA)

STATISTICAL TECHNIQUES

- SSANOVA (Gu 2002)
  - Standard technique for UTI data following Davidson (2006)
  - Not routinely used to model *within* & *between-speaker* variation
  - Data need to be expressed in polar coordinates to avoid errors most pronounced at edges (Heyne & Derrick 2015; Mielke 2015)
- GAMMs (Wood 2017)
  - Model non-linearity in time series (Hastie & Tibshirani 1986; Wood 2006 & 2017) by fitting smooths to curves (cf. polynomial regression)
  - Increasingly used in phonetics research to account for variation across time, e.g., formant trajectories (Sóskuthy 2017), positional changes of a single EMA sensor (Wieling 2018), or vocal tract constrictions in real-time MRI (Carignan et al. 2020)
  - Previously applied to UTI with fixed (Heyne et al. 2019) & variable timepoints (Al-Tamimi 2018)
  - Allow modelling of *within* & *between-subject* variation using random effects or factor smooths, fixed effect interactions, & non-linear interactions between time-series & predictor and/or factor smooths
  - Similar to mixed effects models with random intercepts & slopes but interactions can be non-linear within spline smooths & random effects (Al-Tamimi 2018; Heyne et al. 2019; Tamminga et al. 2016)

DATASET

- Participants: 10 Tongan & 10 New Zealand English (NZE) speaking trombone players (1 female each)
  - Mean age 40.3 (SD=18) for NZE vs 27.2 (SD=8.3) for Tongan players (for details see Heyne 2016)
  - Tongan is a typical Pacific language with small phoneme inventory (Garellek & White 2015)
- Recorded musical passages & word list reading in respective native languages using a GE Healthcare Logiq e (version 11) ultrasound machine with wide-band microconvex array transducer (4.0-10.0 MHz)
  - Transducer held in place using non-metallic jaw brace (Derrick et al. 2015) -> see photo on right
  - US screen images & audio recorded using shotgun microphone saved on external laptop at 58-60Hz

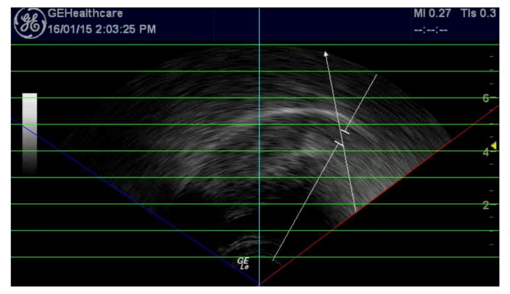


- US frames manually traced at vowel midpoint, 1/3 of note duration using `GetContours` (Tiede & Whalen 2015) in `MATLAB` (version 2015a)
  - 7,834 NZE & 4,422 Tongan vowel tokens, notes shown in Table below

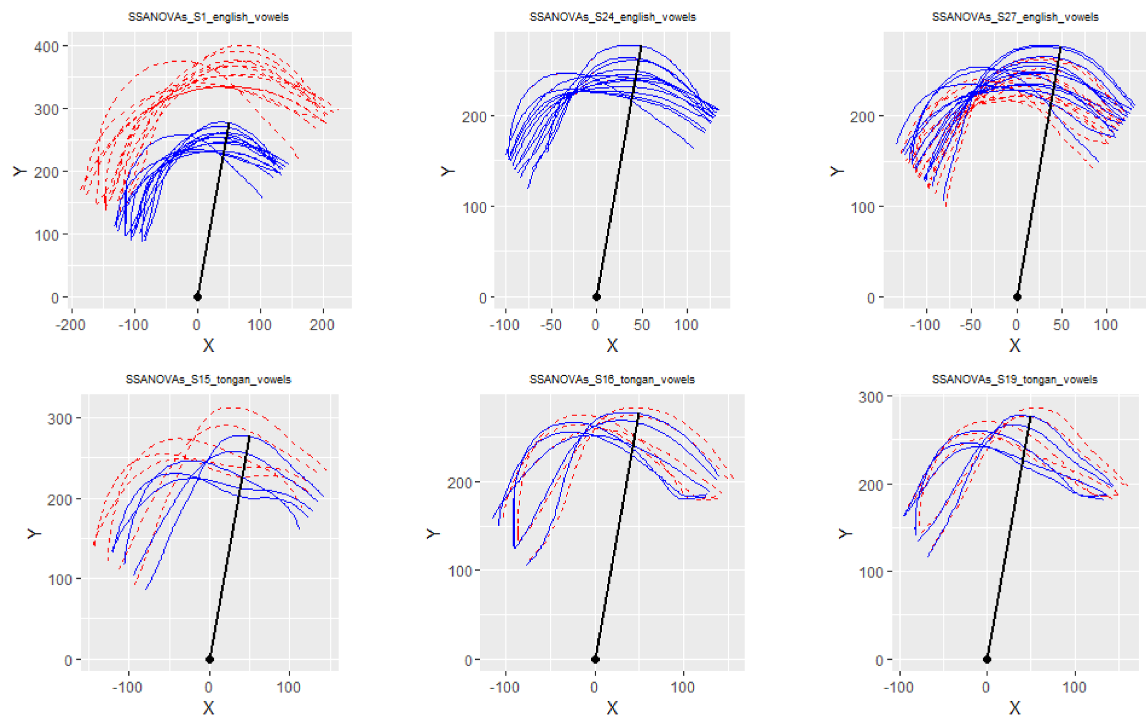
Note	Tongan				NZE			
Intensity	piano	mezzo-piano	mezzo-forte	forte	piano	mezzo-piano	mezzo-forte	forte
Bb2 (233Hz)	7	2	579	32	23	31	574	59
F3 (349Hz)	26	55	1,169	79	52	63	1,089	99
Bb3 (466Hz)	42	17	1,042	62	55	37	986	72
D4 (597Hz)	25	6	385	32	25	13	368	24
F4 (698Hz)	6	1	129	11	6	0	126	13

NORMALIZATION PROCEDURE

- Normalization method from Heyne 2016
  - Estimate 'virtual' origin of ultrasound probe on a by-subject basis (see Heyne & Derrick 2015) -> see example image on right



  - Transform points on tongue contours from Cartesian coordinates to polar coordinates using virtual origin = [0,0]
  - Estimate highest point of average tongue contour for vowel /i/ due to anatomical & stability cross-linguistically -> here we also used /o/ & /a/ for comparison
  - Rotate angle values (Theta) & scale radial length (Rho) to match values observed for speaker with smallest oral cavity space -> see middle plot in top row below; other plots show raw (dashed red lines) vs normalized (blue solid lines) vowel SSANOVA average curves

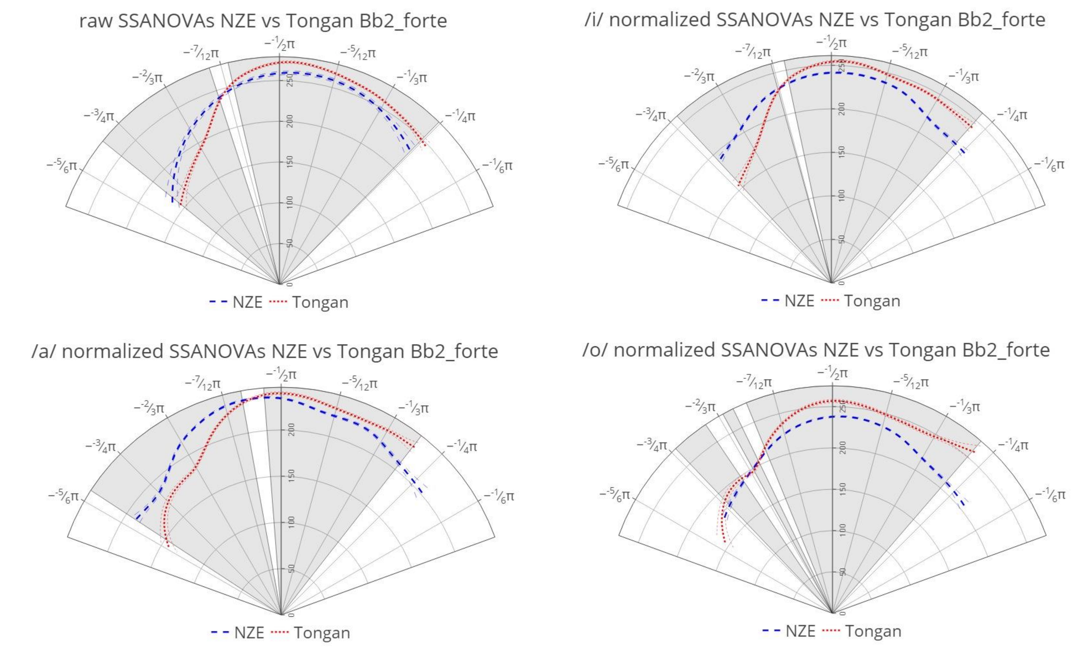


STATISTICAL DESIGN

- All raw data were imported to & normalized (if applicable) in `R` (version 3.6.2)
- SSANOVAs fit on all datapoints grouped by note & intensity (but without specifying any interactions) using `gss` package (Gu 2013)
  - Average splines visualized using `plotly` package (Sievert 2017) with regions of significant difference calculated where 1.96*SE confidence intervals did not overlap
- Auto-regressive GAMMs fit using `mgcv` package (Wood 2017)
  - Fixed factors: Interaction of Language, Note, & Intensity (*between-splines*) -> bam(rho ~ langNoteInt.ord + …, data=dfNotes, discrete=TRUE)
  - Smooths: Angle values (~time series) adjusted by interaction between Language, Note, & Intensity -> + s(theta, bs="cr", k=10)
  - Factor smooth interactions: Angle value by speaker (*between-subject*) adjusted by interaction between Language, Note, & Intensity (*within-subject*) -> + s(theta, subject, bs="fs", k=10, m=1, by=langNoteInt.ord)
  - Outcome variable: Tongue distance from virtual origin (normalized & raw Rho values)
  - Optimal models chosen by comparison & using R^2
  - Predicted smooths visualized using `plotly` with regions of significance obtained using `plot_smooth()` & `plot_diff()` from `itsadug` package (van Rij 2015)
- Comparison of SSANOVA & GAMM curve estimates by examination of size of intervals of significant difference for /i, a, o/-normalized & raw datasets

RESULTS

- Plots below show intervals of significant difference for SSANOVA average tongue contours for note Bb2 at *forte* intensity



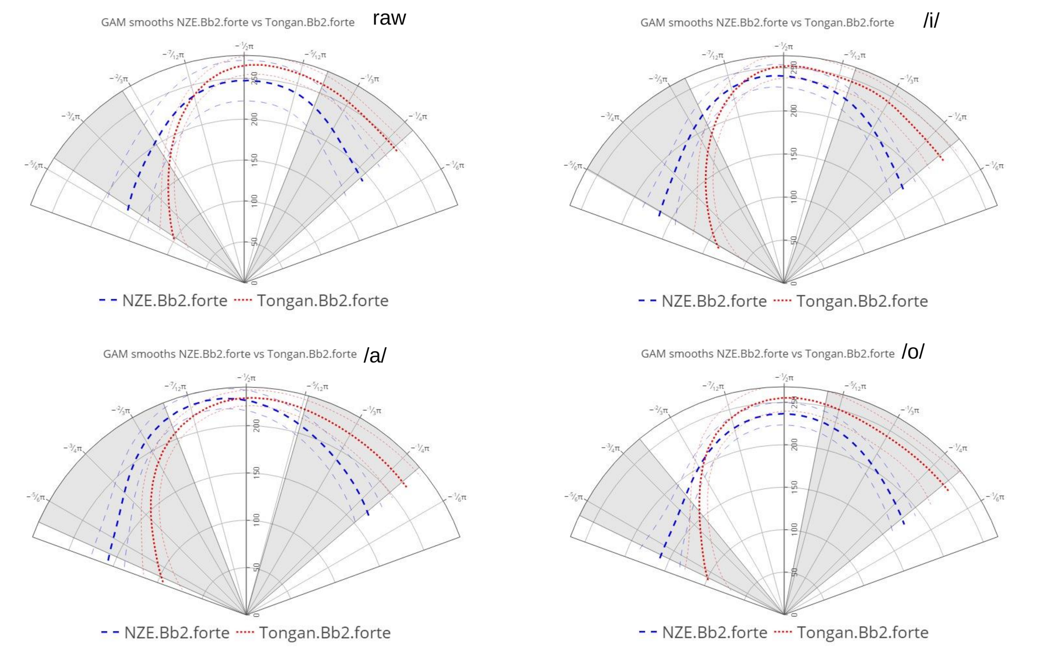
- SSANOVAs estimated on raw & normalized data show similarities in regions of significant difference (shaded areas; 95% CI) at front (right edge) & back of tongue
- Over-confidence in regions of significance -> likely inflated Type-I error due to omitted random effects
- All normalized datasets perform similarly to raw data, although contours look least 'noisy' for /i/ normalization
  - Contours fit on /a, o/-normalized data show additional inflections at edges, especially at back of tongue

FUNDING & ACKNOWLEDGEMENTS

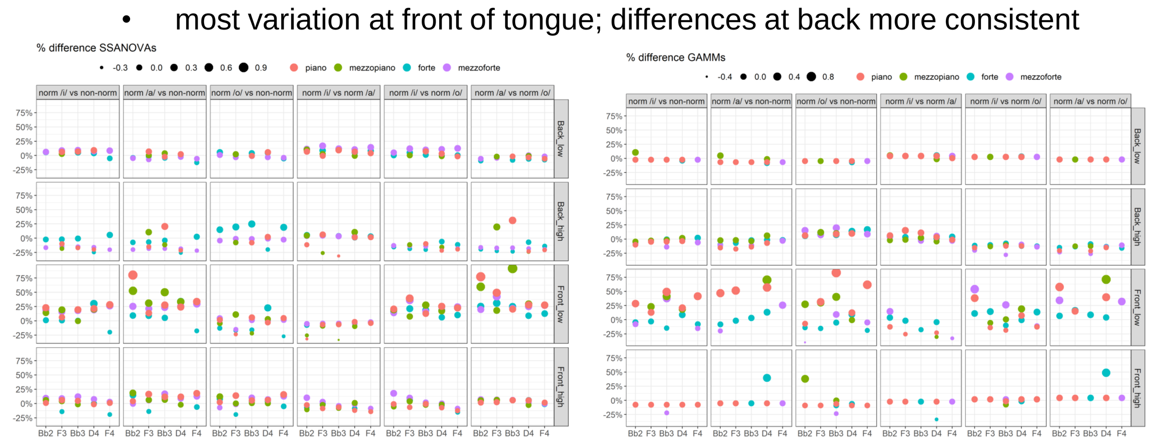
MH's research was supported by a Doctoral Scholarship from the University of Canterbury & funding from the NIH (research grant R01DC002852, PI Frank Guenther). JA-T prepared parts of this work while on sabbatical funded by the Leverhulme International Academic Fellowship (IAF-2018-016).
We appreciate the time volunteered by our research participants in Tonga & NZ, specifically the Royal Tonga Police Band, and are grateful for the manifold contributions of the late Romain Fiasson.

RESULTS (continued)

- Plots below show intervals of significant difference for GAMM smooths for note Bb2 at *forte* intensity



- GAMM smooths display smaller regions of significant difference, likely due to inclusion of random effects
- Similar results for raw & /i, o, a/ normalized datasets
- Percent difference plots below for SSANOVAs (left) & GAMMs (right) show similarities in regions of significance
  - most variation at front of tongue; differences at back more consistent



DISCUSSION & RECOMMENDATIONS

- Our results show differences in regions of significant difference for both raw & normalized datasets
  - Raw data findings more conservative
  - GAMMs outperform SSANOVA due to ease in modelling random effects and accounting for *within* & *between-speaker* differences
- Normalization using /i/ less variable than with /o/ or /a/ -> more variation observed in % difference plots for /o, a/-normalized data compared to raw &/or /i/
- GAMMs provide powerful framework to model *within* & *between-subject* differences allowing emergence of generalizations with confidence
- Recommendation ➔ Use normalization technique adapted to dataset, preferably with /i/-normalization or bite-plate
  - Raw data modelled with GAMMs show similarities, albeit with slightly more variation & smaller regions of significant difference.

REFERENCES