[LabPhon] Editor Decision - "Assessing mid-saggital tongue contours in polar coordinates using generalised additive (mixed) models"

From: **Kip Wilson** | admin@labphon.org
Wednesday 31 Jul, 20:55

To: **Mr Stefano Coretta** | stefano.coretta@gmail.com
Cc: **Stefano Coretta** | stefano.coretta@manchester.ac.uk

Dear Stefano Coretta,

The manuscript #202 entitled "Assessing mid-saggital (sic!) tongue contours in polar coordinates using generalised additive (mixed) models", which you submitted to the special collection “Techniques and Methods for Investigating Speech Articulation” of the Journal of the Association for Laboratory Phonology, has been refereed.

The comments of the reviewers are pasted in below. Reviewer B suggested that, after making some improvements, the work should be accepted. Reviewer A, however, argued that the work should be rejected because of its limited scientific contribution and some irremediable methodological shortcomings.

We believe that the submitted article addresses topics of relevance to laboratory phonology. However, we also largely share Reviewer A’s view that the paper is not innovative, be it a tutorial of the specified generalised additive mixed model applied to UTI data, or a critical review thereof. In addition, we note that the well-known problems of cross-subject comparison of UTI data make the treatment of mixed non-linear effects (a characteristic feature of these models) useless.

Because of these reports, we have decided to reject the manuscript. We are sorry to bring you this disappointing news. We hope that you can benefit from Reviewer A’s comments. Moreover, we hope that you will consider Laboratory Phonology as a possible outlet for your future work again.
Once again, thank you for submitting your manuscript to the special collection “Techniques and Methods for Investigating Speech Articulation” of the Journal of the Association for Laboratory Phonology.

Best wishes,
Lorenzo Spreafico (Editor for the special collection)
Alessandro Vietti (Editor for the special collection)
Mirjam Ernestus (General Editor)

Summary of action: Reject

Reviewer A:

I've read the paper "Assessing mid-saggital tongue contours in polar coordinates using generalised

additive (mixed) models" with great interest. After reading it, however, I felt rather disappointed. Particularly, I don't really see the scientific or methodological contribution of the paper. Unfortunately, I also don't see how it could easily be improved to make a substantial contribution (but I'll let the editor be the judge of that).

My main issues are threefold.

First: I have trouble to identify what is new. It's not the use of polar coordinates in analyzing UTI data: asa-scitation-org.proxy-ub.rug.nl/doi/10.1121/1.4919346.
It's also not the use of GAMs in analyzing UTI data: www.martijnwieling.nl/files/Noiray2019.pdf (though the authors do not analyze tongue contours). I therefore think the only novelty is that it analyzes tongue contours with GAMs.

Second: while this still seems potentially interesting, the author only uses it to analyze data for one subject (per analysis). The power of GAMs are that they allow for identifying non-linear patterns while *importantly* also taking into account individual and item-specific variation. To me it seems we would want to analyze data across individuals to reach any kind of general(izable) conclusion. But if I'm reading the manuscript correctly, tongue contours are not comparable between subjects due to different positioning of the probe and different anatomy (lines 146-149). So the analysis seems essentially useless to me, as there won't be any general conclusions possible. In addition, nothing is mentioned about probe movement during speaking. I would argue that the measurements across different pronunciations of one subject are not even comparable in this dataset. The reason for this is that the orientation of the probe compared to the jaw changes due to head and jaw movement. Even though the headset constrains movement of probe relative to head/jaw, there is still movement possible (otherwise the jaw could not move with respect to the headset) and given that the relative orientation of the probe to the head/jaw is not known, it is unclear how problematic this is. Even a small rotation will have a large effect on the differences inferred (since essentially different parts of the tongue are compared, but the incorrect assumption is that these are the same parts of the tongue, but that the tongue - rather than the probe - has moved). Of course, there is a bite plate recording, but this doesn't help, as there's only one recording per subject (and with the jaw closed). Consequently, to be able to analyze tongue contours (rather than to extract some landmarks), it's *absolutely essential* to know at all times the relative orientation of the probe compared to the head/jaw during speaking. Only in that case is it possible to take the intersection of the visible part of the tongue at different time points, and to truly investigate differences due to tongue position differences across conditions. It is clear that this is not the case for the data studied, and illustrating the GAM approach on this dataset suggests that this is an appropriate analysis, whereas this - unfortunately - is not the case. Advanced statistics can, unfortunately, not help solve esential problems in the data collection.

Third: the paper is supposed to be a short tutorial, but I don't see what it adds over much more extensive tutorials from Martón Soskuthy and Martijn Wieling (note that the correct reference for the latter tutorial is: www.sciencedirect.com/science/article/pii/S0095447017301377)? The present paper does offer a useful conversion function for visualization, but converting radial coordinates to cartesian coordinates is not hard and it doesn't change anything to the approach illustrated in the two tutorials. Due to the present paper's brevity (which for a tutorial is really not appropriate, I think) it also sometimes misses the mark. For example, it's really not appropriate to compare it01_gam_0 to it01_gam (there are several intermediate models which should be tested, so the

model comparison is really useless here and no informative conclusion can be reached on the basis of this model comparison). Essentially, even though there are multiple items (repeated multiple times) in the dataset, the author ignores this in his analysis. This also means that all results are overconfident, since all observations are treated as independent, which is not the case. I find this strange, given that the author does mention mixed-effects analysis quite frequently... Also I wonder whether the data is sorted appropriately to detect autocorrelation? Also Fig. 6a and b appear to be identical... I don't really understand the goal of section 3 is. To me it seems this Section could be left out entirely, as it's not really discussing the method and only discusses some individual differences. I fail to see why this is interesting (and really: why would it ever be OK to exclude 5 out of 12 people?). Finally, the paper mentions that this is an improvement over SS Anovas, but why it then doesn't compare the two approaches is unclear to me.

In sum: I don't see what the contribution to the field is of this paper over providing a few convenient functions for plotting polar coordinates (which warrants an R-package, but not a paper). I don't see a way to improve this to a state which would warrant publishing, without needing to collect new data where the position of the probe relative to the head/jaw movement is registered and where this information before analysis is used to correct the data. As it stands now, inferred tongue differences between voiced and voiceless pronunciations might simply be due to the probe shifting slightly (or an inappropriate analysis, unfortunately, as it ignores item variability). Even if this all were to be corrected, a tutorial paper should explain essential parts of the approach, rather then mention them only in passing.

Reviewer B:

The paper addresses an important question in the research of tongue ultrasound analysis: how can we specify multiple predictors and random factors in ultrasound data? This is an important contribution since more and more, we are using ultrasound data in studies that require these specifications, especially on sociolinguistic studies. This study therefore successfully demonstrates how GAMs can be implemented in ultrasound data analysis of tongue contours. The paper, therefore, addresses an important topic that is relevant for laboratory phonology.

In terms of originality, though the paper demonstrates the use of GAMs in ultrasound analysis, it does not create the statistical infrastructure but rather adapts it to tongue contours. However, this does not take any value from this, as also shown in Davidson (2006).

In terms of the methodology, it is methodologically sound, following GAMs implementation similar to Sóskuthy (2017). It shows a clear step-by-step description of the implementation on real data from two languages. In terms of the statistical model, it would be recommended to be clearer about the predictors and especially, the random factors, i.e. which they are and how they are specified in the formula, as well as interpreting results.

In relation to the R code, more explanation is necessary on what the input and output are. There are sections of code which are on grey (assuming it is the input) and another piece of code preceded by # (assuming it is the output of the code). This should be clarified to the reader.

In relation to the graphs, they need more working to be understood with more clarity. Sometimes, it requires a lot from the reader to make up for information or go back and forward to correctly interpret them. These are some general recommendations:
• Make more explanatory x and y labels, not just 'X' as the x label.
• Add measurement units to the x and y labels
• Normalise tick labels: they are smaller in figure 5.
• Check the Figure description in Figure 5. In the description is says vowels /a/ and /u/ but in the top-left labels of figures both say /u/.

Other general comments

In-text citations to R and the rticulate package are missing

More orthodox spelling in contractions: do not instead of don't and cannot instead of can't.

The reference to C2 and C3 appears in figures 2 and 3 before they are explained in line 111.
