Retraction of the whole tongue induced by pharyngealisation in Levantine Arabic: A between-subject account using Static and Dynamic PCA and GAMMs

Jalal Al-Tamimi[1], Pertti Palo[2]

[1]Université Paris Cité, CNRS, Laboratoire de Linguistique Formelle (LLF), Paris, France,
[2]Indiana University, Speech Production Laboratory, Bloomington, USA,
[1]jalal.al-tamimi@u-paris.fr; [2]pertti.palo@taurlin.org

Keywords: GAMMs, PCA, UTI, Static vs dynamic analysis, Pharyngealisation, Levantine Arabic

Introduction

Accounting for between and within-subject variability is a relatively easy task when one uses mixed effects regression [1, 2]. Modelling speaker (and items) as random effects allows for coefficients of the main effects to be adjusted to account for this type of variability [1, 2]. Using a maximal specification approach [1] allows for an accurate estimation of the within and between-subject variation, although this heavily depends on the structure of the data and the need to assess the model's fit to allow for a meaningful interpretation of the patterns observed. When dealing with Ultrasound Tongue Imaging (UTI) data, however, anatomical differences between speakers, for instance in tongue size, lead to important within and between subject variation that can hinder any cross-speaker generalisations, unless properly dealt with. The aim of our paper is to demonstrate how this can be achieved using two types of analyses, namely static and dynamic Principal Component Analysis (PCA), and static and dynamic Generalised Additive Mixed-effects Models (GAMMs). PCA emerges as a relatively easy, simple and robust approach to account for between-subject variability using UTI data [3] and has since been employed in various studies using UTI data on laterals [4], rhotic consonants [5] and acquisition of laterals/rhotics [6]. GAMMs on the other hand, while it has already been used on UTI data trying to account for within and between-subject variability [7, 8, 9], they are still not widely used by the community to account for between-subject and gender differences observed using UTI data.

In this study, we systematically compare the performance of the two approaches at quantifying within and between-subject and gender variability using UT data, by examining a particular contrast in Levantine Arabic: impact of the phonological secondary pharyngealisation vs plain coronal (plain henceforth) contrasts using a whole tongue approach. Previous research has shown that pharyngealisation has a general backing and retraction effect observed on the consonant itself and surrounding vowels [10]. The backing effect is related to the front-back dimension of the vocal tract, while, retraction is described as a general backing and lowering effect of the tongue dorsum and root. Using a whole tongue approach of the UTI data analysed via GAMMs, [8] showed that pharyngealisation led to a general retraction of the tongue dorsum and root, and a depression of the front part of the tongue. We re-examine the results of [8] and expand them by exploring how static and dynamic PCA perform in comparison to static and dynamic GAMMs on the same dataset. We demonstrate similarities between PCA and GAMMs and highlight issues surrounding their application and interpretation. We also put emphasis on strengths of GAMMs over PCA.

Method, corpus and data processing

Ten Levantine Arabic Urban speakers (5f, 5m), aged 25-45, were recorded using synchronised UTI, EGG, and audio recordings through a multichannel breakout [11]. The UTI data used a Mindray DP-6600, NTSC video output at 30fps, with a scan depth of 7.55cm, sampling Frequency of 5MHz, with an endocavity microconvex probe (10mm radius; 120°Field of View) with a metallic stabilisation headset (developed by Articulate Instruments). The acoustic signal was recorded using a Roland Pro Microphone connected to a Roland Quad-Capture, sampled at 44.1 kHz, 16 Bit quantisation in mono channel; the microphone was placed at ≈ 15 cm from the speaker's mouth. Due to using the stabilisation headset, the EGG electrodes and the UTI probe, the angle of view across all participants was identical.

Participants were instructed to produce a list of items in a /ˈʔVːˈCVː/ frame, with all possible consonants in LA in the three symmetric vowels /iː aː uː/; see more details of corpus, data collection and data analyses in [8]. UTI data from 8 speakers (4f, 4m) were automatically traced and manually corrected using AAA [12] at 9 timeframes within the VCV sequence (2 within each of the vowels; 5 within the consonant). The tongue contours were exported as 42 points in polar coordinates with tongue height in mm (r) and angle values in Radians (φ) in an unrotated view. Here, we only look at tongue contours obtained for all plain /t d ð s z l/ and pharyngealised /tˤ dˤ ðˤ sˤ zˤ lˤ/ contexts produced in the /iː aː uː/ vowel contexts and across all eight participants. For our data analyses, we used the centre 34 contour points, after removing the first/last four points that were hidden by the hyoid and mandibular bones.

Statistical approaches

We used four sets of modelling to account for both within and between-speaker and gender variation. Firstly, we started by using a dynamic PCA following [6] to account for the dynamic changes throughout the VCV sequence allowing us to identify the point of maximal tongue retraction and the averaged tongue contours to allow for an evaluation of the impact of pharyngealisation on the tongue contour. This was then followed by a static PCA following [5] on the point of maximal retraction. In comparison, we used the static and dynamic GAMMs adjusting for within and between-speaker and gender following [8] using a maximal specification approach [1]. For ease of comparison, for the PCA, we restricted our analysis to the VCV sequence with symmetric /iː/ sequences; for GAMMs, we accounted for the interaction between context and vowel by gender in our modelling, which provided a more streamlined modelling approach (for more details, see [8]).

For the PCA, we used the x and y polar coordinates (r and φ) for the 34 contour points that were z-scored per participant to adjust for speaker and gender specific variations (following [5, 6]). We then modelled these coordinates across the 9 intervals, using the function `princomp` from the `stats` package in `R` [13]. Once we obtained the variance explained for each PC with a loading of ≥ 5% [14], we plotted the data to evaluate changes related to both contexts and by proportional time. This allowed us to identify the point of maximal retraction and the averaged tongue contour. We then ran our static PCA on this point, and after obtaining the variance explained, we plotted the results to evaluate the differences related to each context. As a confirmation of the patterns observed and to assess the significance levels, each PC was submitted to a Linear Mixed-effects Modelling (LMM). For the dynamic PCA, we modelled each PC loading as a function of the interaction between the context and proportional time as fixed effects, with speaker and item as random intercepts and a by-speaker random slope for context; for the static PCA, context was modelled as a fixed effect, with speaker and item as random intercepts, without random slopes as they did not improve the models' fit.

The AR1-GAMMs model was fitted using `mgcv` [15] using the raw data. The model allowed for modelling tongue height (Rho) as a function of the interaction between context and vowel by gender as our fixed effects (ordered predictors), with two timeseries smooths for the 34 contour points and for the 9 timeframes, in addition to their interaction (using `ti`) adjusted by the fixed effects. Our random effects were modelled as factor smooths for each of speaker and item as a function of the two timeseries adjusted by the interaction between context and vowel (for speaker) and by gender (for item). Our optimal model accounted for 88.7% of the variance and improved the fit when compared to a simpler model ($\chi^2(4)$ = 8101.9, p<0.0001). For dynamic GAMMs, we explored the dynamic changes throughout the VCV sequence, first using 3D surface plots via the function `vis.gam` from `mgcv` [15] and then using the differences between the tongue contours in pharyngealised and plain contexts via the function `plot_diff2` from `itsadug` [16], with the secondary constriction location estimated following [17] based the model's predictions for angle values ranging between -1 and 1. For static GAMMs, we used the time of maximal retraction identified using the `plot_diff2` function, which was identical to that identified in the dynamic PCA located at timeframe 6 (=C2 at 75% = 62.5% proportional time). We quantified the tongue contour differences between the pharyngealised and the plain contexts using a custom-made plotting function adapted from [7] using the function `plot_diff` from `itsadug` [16].

Results of the PCA analyses

Table 1 presents the proportion of variance explained by the dynamic PCA (top row) and static PCA (bottom row). Starting with the dynamic PCA, the first four PCs explained 86.4% of the cumulative variance, with PC1 explaining 38.7% of the variance. Figure 1a shows the dynamic changes as a function of proportional time, with a clear lowering throughout the VCV sequence in the pharyngealised context, with the lowest value at the 62.5%, coinciding with the consonant's release. Figure 1b shows the averaged contour that shows pharyngealisation (dotted lines) impacting the whole tongue with overall retraction and tilting of the tongue (p<0.0001). Next, using timeframe 6 (=62.5% proportional time), the static PCA showed that the first four PCs explained 87.2% of the cumulative variance, with PC1 explaining 44.3% of the variance (Table 1, bottom row). The four plots presented in Figure 1c, d, e and f show the averaged tongue contours as a function of each PC. PC1 confirms the overall retraction and tilting of the whole tongue in the pharyngealised context (p<0.0001, Figure 1c dotted lines). PC2 showed a non-statistically significant raised tongue front, lowered tongue body and marginal retracted tongue dorsum (p=0.5, Figure 1d dashed lines); PC3 showed a lowered tongue mid and root (p<0.01, Figure 1e dotted lines) and PC4 showed a tendency for a raised tongue back and lowered root (p=0.09, Figure 1f dashed lines). The results obtained from the four PCs are correlated and confirm the general tongue retraction, tilting, backing and raising, with marginal tongue tip changes reported in [8, 10].

Table 1: Variance explained of PCA for dynamic (top) and static at 62.5% of proportional time (bottom)

PC	PC1	PC2	PC3	PC4	Total
Dynamic	38.7%	28.7%	12.3%	6.8%	86.4%
Static	44.3%	22.7%	12.7%	7.4%	87.2%

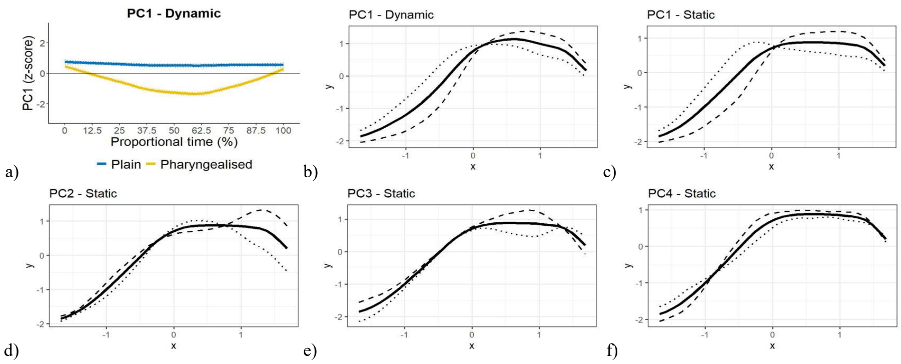

Figure 1: Results of the dynamic and static PCA in an /i:/ context: dynamic PCA with PC1 as a function of proportional time (a), PC1 loading as averaged across all timeframes (b), static PCA loading at timeframe 6(=62.5%), with averaged tongue contours on PC1 (c), PC2 (d), PC3 (e) and PC4 (f). Dashed lines = average contour +SD; dotted lines = average contour -SD

Results of the GAMMs analyses

Figure 2 shows the predictions from our GAMMs. First, the 3D surface plot across Angle, Timeframe and Rho shows the tongue, in the plain context (Figure 2a), at its highest position towards the hard palate (front tongue) with no major changes throughout the VCV sequence. The 3D surface plot for the pharyngealised context (Figure 2b) shows a clear tongue tip rising, tongue front and body depression, tongue back rising, and tongue dorsum and root retraction; changes observable throughout the VCV sequence with the strongest changes within the consonant (timeframes 3 to 7). Figure 2c shows the 3D difference plot between the pharyngealised and the plain contexts, which confirm a statistically significant tongue depression by a maximum of -6 mm (red) located between the alveolar ridge and the velum in addition to a statistically significant tongue back and dorsum retraction by a maximum of +10 mm (light colour) located between the uvula and the lower pharynx. These double depression and retraction are similar to a double-bunched production and are strongest at timeframe 6 (at C2 = 75%), which coincides with the consonant's release; the same position identified with the dynamic PCA above (at 62.5%). Figure 2d presents the 2D difference smooths between the pharyngealised and the plain contexts, which shows a statistically significant tongue tilting, with tongue front depression, tongue back and dorsum retraction, in addition to potential tongue root retraction in the pharyngealised context.

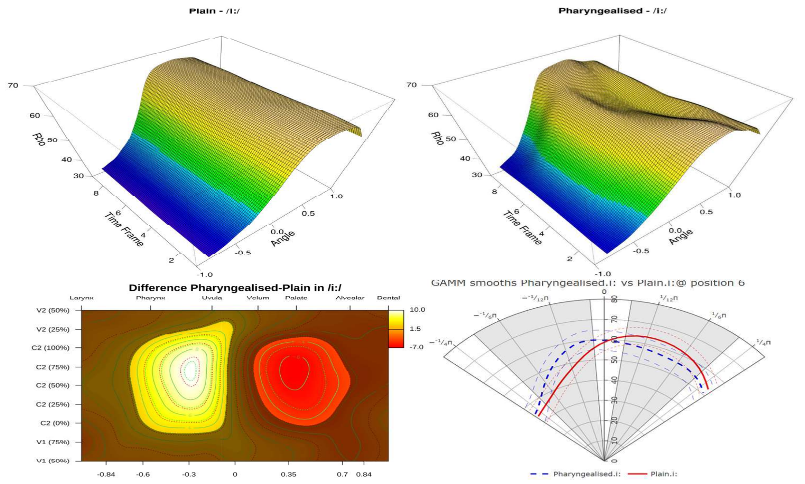

Figure 2: Results of the dynamic and static GAMMs in an /iː/ context: Dynamic GAMMs with 3D surface plot in the plain context (a; Angle on x-axis, tongue root to tip; timeframe on y-axis from V1 50% to V2 50%; height (Rho) on z-axis), 3D surface plot in the pharyngealised context (b), Dynamic 3D GAMMs difference between pharyngealised vs plain contexts (c; Angle on x-axis, tongue root to tip, timeframe on y-axis; V1 50% to V2 50%, height on z-axis; statistically significant tongue height difference indicated by lighter and red colours; lighter = increase in tongue height; red = decrease, with estimated constriction location on secondary x-axis at the top) and GAMMs difference smooths between pharyngealised vs plain contexts (d; polar coordinates with angle on x-axis and tongue height on y-axis; dashed/dotted lines indicate 95% CIs; shaded areas indicating significant difference).

Discussion and conclusions

This study confirmed prior results on the impact of pharyngealisation on the tongue shape reported in [8, 10]. Both PCA and GAMMs showed similar patterns. While PCA is quicker, it requires the use of more than one dimension to explain patterns in the data in addition to validating the results using LMMs. PC1 explained most of the variance in the data, but at a mere 38.7% for the dynamic PCA or 44.3% for the static and can sometimes be used as the sole PC to account for the changes observed in the plain vs pharyngealised contexts. This however means that 55 to 60% of the data is not accounted for. In addition, individual observations are used to construct the PCA, which in a sense mimics how mixed effects regressions work, albeit without any adjustments to fixed effects, random effects (especially items), nor random slopes (for either speaker or item) leading to potentially increased Type I error. Due to the fact that PCA's loadings are obtained for each observation, averaging and variance around the mean are used to account for between-speaker patterns, which can lead to over confidence in the patterns observed (e.g., our PC2 and PC4 results). GAMMs, on the other hand, offer fine-grained account of the data at the expense of heavier computation. They allow for a streamlined multidimensional modelling strategies that includes random effect's structure that is easier to implement than SSANOVAs [9] and provide the user with more interpretable outputs and powerful visualisations highlighting various patterns in the data. In conclusion, our results emphasise that GAMMs are relatively easy to apply on UTI data allowing for generalisations accounting for within and between-speaker and gender variation. It provides a unified framework to account for UTI data, without the need to apply any normalisation techniques as the coefficients are adjusted to account for between-speaker and gender differences [9].

Acknowledgments

This work is supported by: 1) a public grant overseen by the IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of the Labex Empirical Foundations of Linguistics - EFL; 2) a British Academy/Leverhulme small research grant, UK (SG160181; 2017-2019) and 3) a Leverhulme International Academic Fellowship, UK (IAF- 2018-016), for the first author; a personal postdoctoral grant from Eemil Aaltonen's foundation via The Finnish Foundations' Post Doc Pool, for the second author. We acknowledge usage of the High-Power Computing (HPC) facilities at Newcastle University, UK and at the CNRS/TGIR HUMA-NUM, IN2P3 and GENCI-IDRIS, France (Grant 2022-AD010613733). Many thanks to Danielle Turton and Ghada Khattab for supporting data acquisition.

References

[1] Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 68(3), 255–278.

[2] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59(4), 390–412.

[3] Johnson, K. (2008). *Quantitative methods in linguistics*. Blackwell Pub.

[4] Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Lab. Phon.: J. Assoc. Lab. Phon.* 8(1), 13.

[5] Nance, C., & Kirkham, S. (2022). Phonetic typology and articulatory constraints: The realization of secondary articulations in Scottish Gaelic rhotics. *Language*, 98(3), 419–460.

[6] Nagamine, T. (2023). Dynamic tongue movements in L1 Japanese and L2 English liquids. *Proc. 20th ICPhS*, Prague, 2442–2446.

[7] Heyne, M., Derrick, D., & 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. *Front. Psych.* 10(2597), 1–26.

[8] Al-Tamimi, J., & Palo, P. (2023). Dynamics of the tongue contour in the production of guttural consonants in Levantine Arabic. *Proc. 20th ICPhS*, Prague, 2095–2099.

[9] Al-Tamimi, J., Heyne, M., & Derrick, D. (2020). From SS-ANOVA to GAMMs: Accounting for within and between-subject variation using generalized additive mixed models on ultrasound tongue contours. *Proc. 12th ISSP*.

[10] Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Lab. Phon.: J. Assoc. Lab. Phon.* 8(1), 1–40.

[11] Wrench, A. A., & Scobbie, J. M. (2008). High-speed Cineloop Ultrasound vs. Video Ultrasound Tongue Imaging: Comparison of Front and Back Lingual Gesture Location and Relative Timing. *Proc. 8th ISSP*, 57–60.

[12] Wrench, A. A. (2018). Articulate Assistant Advanced User Guide (Version 2.17). *Edinburgh: Articulate Instruments Ltd*.

[13] R Core Team. (2022). *R: A language and environment for statistical computing* [Manual].

[14] Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

[15] Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC.

[16] van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*.

[17] Carignan, C., Hoole, P., Kunay, E., Pouplier, M., Joseph, A., Voit, D., Frahm, J., & Harrington, J. (2020). Analyzing speech in both time and space: Generalized additive mixed models can uncover systematic patterns of variation in vocal tract shape in real-time MRI. *Lab. Phon.: J. Assoc. Lab. Phon.* 11(1).