DYNAMICS OF THE TONGUE CONTOUR IN THE PRODUCTION OF GUTTURAL CONSONANTS IN LEVANTINE ARABIC

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

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

ABSTRACT

This paper presents a dynamic account of the tongue contour imaged with Ultrasound Tongue Imaging (UTI) and quantified via Generalised Additive Mixed Models, during the articulation of guttural consonants (uvular, pharyngealised, pharyngeal) in Levantine Arabic. Gutturals are claimed to form a natural class and we aim to quantify the degree of (dis)similarity between the members of this class. UTI data were obtained from 8 participants (4 females), producing a variety of consonants (21) in the medial position of a disyllabic frame, with symmetric vowels /iː aː uː/. When compared to plain coronal consonants, and through a dynamic analysis of tongue contours throughout the VCV sequence, the three members of the guttural class show similarities in overall tongue changes towards the front, dorsum, back and root across the three vowel contexts, providing articulatory evidence for the legitimacy of gutturals as a natural class.

Keywords: Gutturals, Levantine Arabic, natural class, GAMMs, Ultrasound Tongue Imaging

1. INTRODUCTION

Guttural consonants are assumed to form a natural class due to specific phonological patterning, and/or to the use of a *common* oro-sensory zone in the pharynx [1, 2]. Its members were traditionally identified as pharyngeals /ħ ʕ/ (or epilaryngeal /ʜ ʢ ʡ/) and uvulars (/χ ʁ q/) [1]. Pharyngealised (/tˤ dˤ ðˤ sˤ/) [2], and/or Laryngeals (/ʔ h/) [3] are also considered to be part of this class, due to pharyngealised sharing a similar place of articulation to that of pharyngeals, albeit with a different degree of constriction [4, 5], or due to the increase in the frequency of the first formant in the vowels surrounding gutturals when compared with plain coronals [3]. However, phonetic research failed to find a common acoustic and/or articulatory correlate that unites all members of such class [6]. Indeed, gutturals in Arabic are variable in tongue configuration, degree of "retraction" and larynx rising/lowering [7, 8, 9, 4]. They favour regressive, rather than progressive feature spreading, which is strongest in pharyngealised consonants [10].

Following the "Laryngeal Articulator Model" (LAM) [11] "epilaryngeal" consonants are predicted to induce a maximal "retraction" of the tongue, with a back and down gesture and a raised larynx, "pharyngeal"/"pharyngealised" consonants to show partial "retraction" and "uvular" consonants to show nil "retraction". When considering the "pharyngeal"/"pharyngealised" pair, it is important to note that the former is assumed to have an upper-to-mid pharyngeal location (depending on the vowel, whereas the latter has a variable constriction location going from upper-to-low pharyngeal, depending on the language [11, 12].

Relying on the predictions from LAM, we propose that members of the guttural class (uvular, pharyngealised and pharyngeal) will share similarities in how they are realised, within consonants and across surrounding vowels. These similarities can be related to the constriction location, changes in specific sections of the tongue, e.g., tongue body, dorsum, root, or in the degree of tongue "retraction".

The aim of this study is then to assess whether there are any (dis)similarities between the members of the guttural natural class. We examine the full tongue contour (from tip to root) quantified via Ultrasound Tongue Imaging (UTI) and modelled via Generalised Additive Mixed-effects Models (GAMMs) on tongue shapes throughout the production of a VCV sequence. We quantify any changes not only at the tongue root, but also throughout the tongue contour, by using a 3D representation of the full tongue contour to guide our investigation.

2. METHOD

2.1. Data recordings

Ten Levantine Arabic Urban speakers (5 males, 5 females), aged 25-45 were recorded using synchronised UTI, ElectroGlottoGraphy (EGG), and audio recordings through a multichannel breakout [13]. The UTI data used a Mindray DP-6600, NTSC video output at 30fps (60fps de-interlaced), 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 EGG data used a 2-channel ElectroGlottoGraph, with larynx contact and height. Finally, the acoustic signal was recorded using a Roland Pro Microphone connected to a Roland Quad-Capture, sampled at 44.1 kHz with a 16 Bit quantisation in mono channel; the microphone was placed at approximately 15 cm distance from the speaker's mouth. Due to using the stabilisation headset, with the EGG electrodes and the UTI probe, the angle of view across all participants was identical.

2.2. Material

Speakers were asked to produce a list of real and nonce-words in the following frame /ˈʔVːˈCVː/ (V : = symmetric /iː aː uː/; C = all possible consonants in Levantine and other Arabic varieties = /b t d m n r f θ ð s z ʃ ʒ l w j k g x ɣ q tˤ dˤ ðˤ sˤ zˤ lˤ ħ ʕ ʔ h/), with three symmetric repetitions (with a maximum theoretical number of items of 2790 across all participants; 31 C * 3 V * 3 repetitions * 10 speakers). We then chose 21 consonants that were divided into 6 contexts, totalling 1940 items (21 C * 3 V * 3 repetitions * 10 speakers; with additional repetitions): **Plain** ⇒ /t d ð s z l/; **Velar** ⇒ /k g x ɣ/; **Uvular** ⇒ /q/; **Pharyngealised** ⇒ /tˤ dˤ ðˤ sˤ zˤ lˤ/; **Pharyngeal** ⇒ /ħ ʕ/; **Glottal** ⇒ /h ʔ/.

2.3. Data segmentation

Acoustic signals were first transliterated using the newly developed romanisation system for Arabic (ATR conventions) [14], then force-aligned using MAUS-universal language [15], with PraatAlign [16], adapted to Arabic. The boundaries obtained from the automatic forced alignement system were then manually corrected to prevent any errors, using criteria adapted from [5, 17]

2.4. UTI tracing

UTI data from 8 participants (4 males; 4 females) were analysed, using Articulate Assistant Advanced (AAA, version 2.18.04) [18]. First, the UTI video recordings were de-interlaced to 59.977 fps, to increase number of frames within each recording (frame length = 16 ms). The acoustic boundaries from the segmented speech were used in our subsequent analysis to guide landmarks selection. Because we wanted to dynamically track changes in tongue contours throughout the VCV sequence, we specified nine-time intervals (timeFrame) within a VCV sequence with a 25% interval shift, starting with the temporal location at 50% of the preceding vowel (V1) and ending at 50% of the following vowel (V2); the other seven intervals were equally spread across the remaining portion at 75% of V1; at 0%, 25%, 50%, 75%, and 100% of the medial consonant (C2), and at 25% of V2.

At each time interval, we traced the full tongue contour, in an unrotated view, from the visible portions on the UTI videos, using a 42 fanline shaped window, which was limited between the hyoid and the mandible bones. The tongue contours were first automatically traced using batch processing in AAA and were then all hand corrected. This resulted in a total of 13698 tongue splines. We extracted the 42 fanline coordinates in Polar coordinates with the Rho values representing tongue height in mm (r) and each fanline representing an angle value in Radians (φ). The first 4 and last 4 fanlines that were hidden by the hyoid and the mandible bones were excluded from subsequent analyses.

2.5. Statistical design

UTI data from 34 fanlines obtained throughout the VCV sequence from the nine intervals were considered as time series and were modeled via an Auto-Regressive Generalised Additive Mixed Model (AR-GAMM) using the package `mgcv` [19]. The total number of measurements submitted to our AR-GAMMs was 465732 datapoints (13698 splines * 34 fanlines). We used an ordered predictor for each of our predictors; Context (plain, velar, uvular, pharyngealised, pharyngeal, and glottal), Vowel (/iː aː uː/) and gender; order predictors reduce Type I error and increase power [20]. To account for inter-dependencies in the data (multiple speakers, items, contexts and two time series) and to allow for normalisation within and between speaker and gender, we used a maximal specification model that had the following specification:

- Outcome ⇒ Radius value in mm
- Fixed ⇒ Context by Vowel interaction by gender
- Smooths ⇒ Angle (34) and timeFrame (9) by the Context by Vowel interaction and by gender
- Tensor product interaction (ti) ⇒ between Angle and timeFrame by the Context by Vowel interaction and by gender
- Factor smooths interactions ⇒ Angle and timeFrame by speaker adjusted by Context by Vowel interaction
- Factor smooths interactions ⇒ Angle and timeFrame by word adjusted by gender

Our maximal model improved the fit when compared to a simpler one ($\chi^2_{(4)} = 8101.9$, $p<0.0001$); it accounted for 88.7% of the deviance explained in the data. We evaluated the structure of the model using the function `gam.check`. To

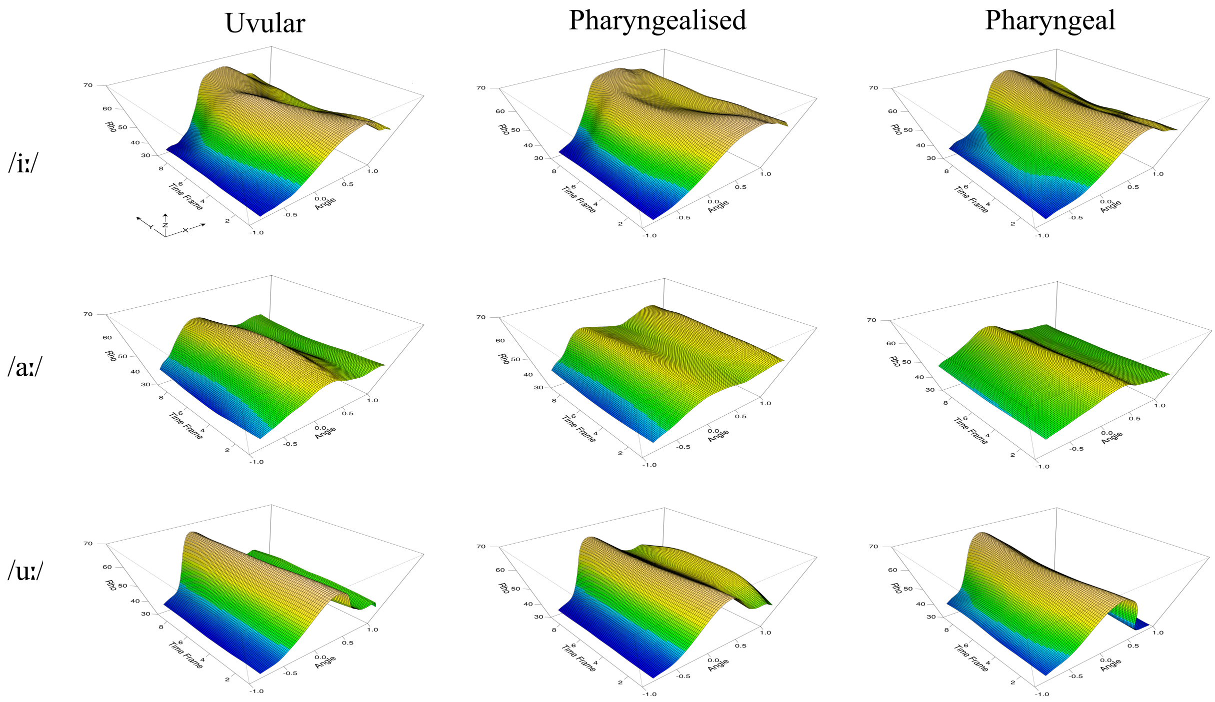

Figure 1: 3D surface contours for the three classes (Uvular Pharyngealised and Pharyngeal) in the vowels /iː aː uː/ (top, middle and bottom rows), according to the Angle (x-axis; oriented from bottom right to top right, tongue root to tip), timeFrame (y-axis; from bottom left to top left, V1 50% to V2 50%) and Rho (z-axis; bottom to top). Tongue height is indicated by both height within the image and colour continuum blue-green-orange.

quantify the overall tongue contour throughout the VCV sequence, we used two types of plots as 3D visualisations, with the angle values on the x-axis, the timeFrame on the y-axis and the tongue height (Rho) on the z-axis. The first type of plots is a 3D surface plot for a specific context by vowel interaction using `vis.gam` from `mgcv`. The second, is a differences plot between two tongue contours in a specific vowel environment (using `plot_diff2`) from `itsadug` [21]. We used predictions of our model for angle value ranging between -1 and 1, to estimate the constriction location in a similar fashion to that used in [22].

3. RESULTS

3.1. 3D surface plots

The results presented in Fig. 1 show the tongue surfaces for the interaction between specific contexts and vowels (uvular, pharyngealised and pharyngeal in columns; /iː aː uː/ in rows).

One interesting observation across the three contexts is the similarities in tongue surfaces with respect to the specific portion of the tongue that is impacted upon. For instance, starting with the three contexts in an /iː/ context (top row), we can easily observe tongue front and dorsum depression (angle = 0 to +0.84) and retraction towards the tongue back/dorsum area (angle = 0 to -0.84) and root. Uvular and Pharyngealised show similarities in tongue surfaces, with small differences related to dorsum root retraction and tongue height. In the three classes, most of the similarities are within the consonant itself spreading into the V2.

Within /aː/ (middle row), similar patterns are observed between the three contexts, with most of the changes observed towards the front, mid, back/dorsum and root of the tongue. Tongue body is lower than in /iː/, but shows minimal depression specifically in the pharyngealised context. Here again, most of the similarities are within the consonant spreading into V2.

Finally, within /uː/ (bottom row), most of the changes are observed within the uvular and the pharyngeal context, with less differences within the pharyngealised context. The three contexts show similarities in type of tongue changes from front to back, with differences in tongue height.

This first comparison showed how uvular, pharyngealised and pharyngeal contexts share similarities in how they impact on the tongue contour throughout the VCV sequence and within vowels. These changes are compatible with the “double bunching” for a “pharyngeal” place suggested by [11, 12], specifically for an /iː uː/ contexts, with changes within an /aː/ context located at the root of the tongue.

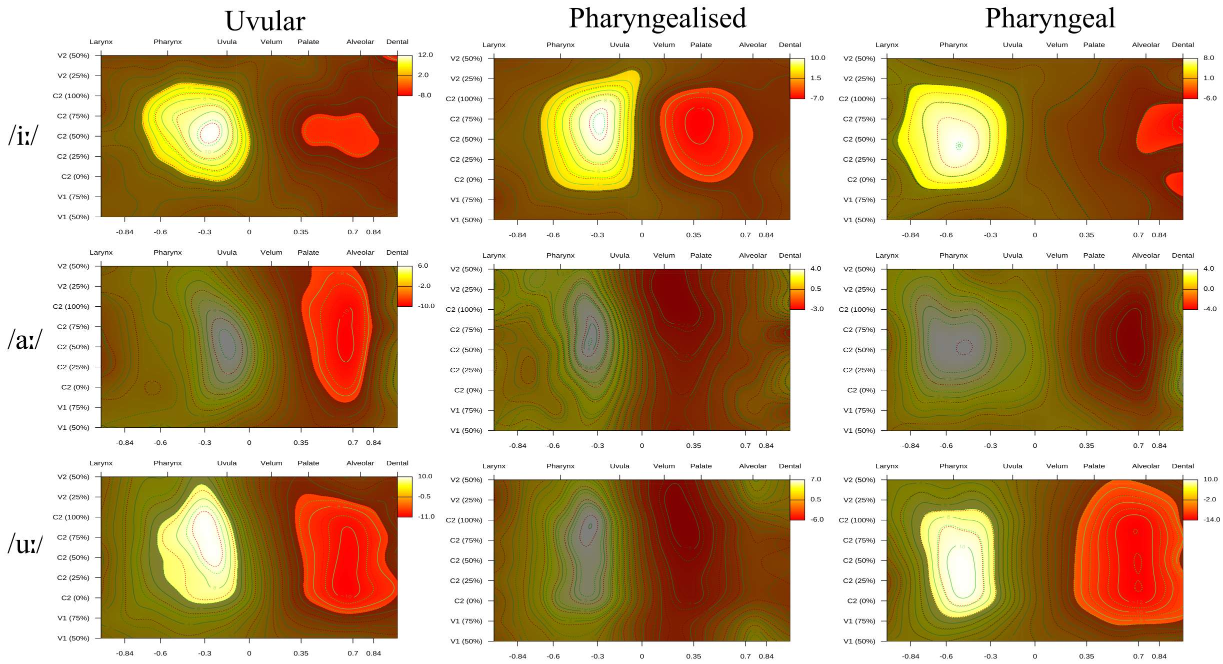

Figure 2: 3D difference plots, with statistically significant regions 95% CI highlighted. The plots show differences between each of the three classes (Uvular, Pharyngealised and Pharyngeal) and the plain class, in the vowels /iː aː uː/ (top, middle and bottom rows), according to the Angle (x-axis; tongue root to tip), timeFrame (y-axis; V1 50% to V2 50%), Rho (z-axis; tongue height difference indicated by lighter and darker colours, with lighter = increase in tongue height; darker = decrease), with estimated constriction location (secondary x-axis, top).

3.2. 3D difference plots

Given the similarities observed between the three contexts highlighted in Fig. 1, we wanted to formally assess whether there are any differences between each of the contexts and the plain coronal context. Results presented in Fig. 2 show the 3D difference plots, which compares the overall tongue contour differences across two contexts. It is interesting to note that across the three vowels, the results are similar and show overall that most of the differences between the three contexts and the plain coronal context are located at the front and the back tongue that goes into the root, specifically within pharyngeals. Most of the observed differences are located throughout the C2 that goes into V2 (up to the middle). Within V1, the only differences observed are around and after 75% of the vowel. The results suggest overall that the constriction location of "gutturals" is similar and is located within an upper to low pharyngeal that shows a gradient constriction across the contexts. Uvulars show the highest location, followed by pharyngealised and then pharyngeals.

4. DISCUSSION

This paper attempted to identify whether there are any (dis)similarities between uvular, pharyngealised and pharyngeal contexts. The results presented offer an empirical evidence compatible with legitimacy of the guttural natural class. Gutturals in Arabic show similarities in overall tongue shape and impacts on the front and back cavities. The uvular context shows partial "retraction" with a "raised" and "back" configuration following LAM [11]. The pharyngealised context has an intermediate tongue "retraction", with a back and mid-down gesture in /iː aː/, and a back and mid-up gesture in /uː/. The uvular and pharyngealised contexts are different in their constriction location with differences mostly in the degree of tongue rising. The pharyngeal context shows a near maximal "retraction" with a lowered tongue dorsum and tongue root changes, but with a fronted tongue position. The tongue dynamics via 3D whole Tongue contours and differences in comparison with a plain context, show *articulatory* differences throughout the C2 and in V2 (up to mid), with minimal differences within V1. This suggests that coarticulatory patterns are progressive in gutturals. The results of the estimated constriction location on the secondary x-axis suggest tongue root changes and retraction that are possibly indicative of larynx raising. Our results suggest gutturals to show possible double bunching in an /iː uː/ contexts, with tongue fronting towards the palatal region and tongue retraction towards the pharynx[11, 12].

5. ACKNOWLEDGMENT

This work is supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083), contributing to the IdEx Université Paris Cité - ANR-18-IDEX-0001; A British Academy/Leverhulme small research grant, UK (SG160181); A Leverhulme International Academic Fellowship, UK (IAF-2018-016), for the first author; and a personal post-doctoral 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).

6. REFERENCES

[1] J. McCarthy, "The phonetics and phonology of Semitic pharyngeals," in *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, P. Keating, Ed. Cambridge University Press, 1994, pp. 191–233.

[2] J. Sylak-Glassman, "An Emergent Approach to the Guttural Natural Class," *Proceedings of the Annual Meetings on Phonology*, pp. 1–12, 2014.

[3] B. Zawaydeh, "The phonetics and phonology of gutturals in Arabic," Ph.D. dissertation, Bloomington, IN: Indiana University, 1999.

[4] A. Laufer and T. Baer, "The emphatic and pharyngeal sounds in Hebrew and in Arabic," *Language and Speech*, vol. 31, pp. 181–205, 1988.

[5] J. Al-Tamimi, "Revisiting acoustic cues of pharyngealization in jordanian and moroccan arabic: Implications for feature specification," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 8, no. 1, pp. 1–40, 2017.

[6] M. Bin-Muqbil, "Phonetic and phonological aspects of Arabic emphatics and gutturals," Ph.D. dissertation, University of Wisconsin-Madison, 2006.

[7] C. Zeroual and G. Clements, "The feature [pharyngeal]," in *Features in Phonology and Phonetics: posthumous Writings by George N. Clements and Coauthors*, A. Rialland, R. Ridouane, and H. van der Hulst, Eds. Mouton de Gruyter: Berlin, 2015, pp. 247–276.

[8] F. Al-Tamimi and B. Heselwood, "Nasoendoscopic, videofluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic," in *Instrumental studies in Arabic phonetics*, B. Heselwood and Z. Hassan, Eds. John Benjamins, 2011, pp. 165–191.

[9] B. Heselwood and F. Al-Tamimi, "A study of the laryngeal and pharyngeal consonants in Jordanian Arabic using nasoendoscopy, videofluoroscopy and spectrography," in *Instrumental studies in Arabic phonetics*, B. Heselwood and Z. Hassan, Eds. John Benjamins, 2011, pp. 101–128.

[10] S. Hellmuth, *The Oxford handbook of Arabic linguistics*. Oxford University Press, 2013, ch. Phonology, pp. 45–70.

[11] J. Esling, S. Moisik, A. Benner, and L. Crevier-Buchman, *Voice Quality: The Laryngeal Articulator Model*. Cambridge University Press, 2019.

[12] S. Moisik, E. Czaykowska-Higgins, and J. Esling, "Phonological potentials and the lower vocal tract," *Journal of the International Phonetic Association*, pp. 1–35, apr 2019.

[13] A. Wrench and J. Scobbie, "High-speed cineloop ultrasound vs. video ultrasound tongue imaging: Comparison of front and back lingual gesture location and relative timing." in *Proceedings of the Eighth International Seminar on Speech Production (ISSP)*, 2008, pp. 57–60.

[14] J. Al-Tamimi, F. Schiel, G. Khattab, N. Sokhey, D. Amazouz, A. Dallak, and H. Moussa, "A Romanization System and WebMAUS Aligner for Arabic Varieties," in *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), © European Language Resources Association (ELRA), Licensed under CC-BY-NC-4.0*, Marseille, 20-25 June 2022, 2022, pp. 7269–7276.

[15] F. Schiel, "A statistical model for predicting pronunciation," in *Proc. of the International Conference on Phonetic Sciences, Glasgow, United Kingdom, Paper*, vol. 195, 2015.

[16] M. Lubbers and F. Torreira, "Praatalign: an interactive praat plug-in for performing phonetic forced alignment," https://github.com/dopefishh/praatalign, 2013-2016, version 2.0a.

[17] J. Al-Tamimi and G. Khattab, "Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops," *Journal of Phonetics*, vol. 71, pp. 306–325, nov 2018.

[18] A. Wrench, "Articulate assistant advanced user guide (version 2.17)," *Edinburgh: Articulate Instruments Ltd*, 2018.

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

[20] M. Sóskuthy, "Evaluating generalised additive mixed modelling strategies for dynamic speech analysis," *Journal of Phonetics*, vol. 84, 2021.

[21] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, "itsadug: Interpreting time series and autocorrelated data using gamms," 2017, r package version 2.3.

[22] C. Carignan, P. Hoole, E. Kunay, M. Pouplier, A. Joseph, D. Voit, J. Frahm, and J. Harrington, "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," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 11, no. 1, 2020.