

Investigating articulatory variability across recording sessions using a functional clustering technique

Lorenzo Spreafico¹, Alessandro Vietti², Alessia Pini³, Simone Vantini⁴

¹*Università degli Studi di Bergamo*, ²*Libera Università di Bolzano*, ³*Università Cattolica del Sacro Cuore*, ⁴*Politecnico di Milano*

The ultrasound tongue imaging technique [1] is used to study the spatial and temporal properties of lingual articulation. Over the years, there has been increasing standardization in the statistical analysis of tongue profiles from ultrasonographic data [2, 3, 4]. Despite this, the current methods do not yet offer effective solutions for comparing data across recording sessions or from different speakers. In this paper, we show how a non-parametric clustering technique can accommodate for possible misalignment of functional data and therefore how it can be used to investigate intraspeaker articulatory variability across recording sessions. This technique allows the simultaneous co-registration and clustering of curves produced in different recording sessions via a functional k-medoid alignment [5, 6], which represents an improvement over the analysis based on k-mean alignment presented in [7] as it provides cluster centroids which are more faithful to the elicited data.

In order to test the potential of the technique, we ran a pilot experiment. First, we devised a list of reduplicated C₁V₁.C₁V₁ Italian words and non-words in which vowels were restricted to /i, e, ε, a, o, u/ and consonants to /p, t, k, s, ʃ/ (e.g.: ['pa:pa, 'te:te, 'ʃo:ʃo]). Then we collected ultrasound data using a SonixTablet machine controlled from a PC running Articulate Assistant Advanced software [8]. Data was recorded at about 100fps with 69 beam-formed echo pulses evenly spread over a 134-degree field of view. Data presented here are from the recordings of a trained adult male phonetician who read the randomized list of stimuli twenty times while fitted with the UltraFitty headset [9] to stabilize the ultrasound probe. After ten repetitions of the stimuli list, the headset was repositioned to simulate two distinct recording sessions.

Collected data were used to validate the clustering technique and test whether it was able to detect that the tongue shape was different or similar for:

- (i) the articulation of the same vowel in the same phonetic conditions but across different recording sessions, e.g. variants of [a] by the same speaker in recording session₁ and recording session₂;
- (ii) the articulation of the same vowel under different conditions within a recording session, e.g. positional variants of [a] in [pa], [ta], [ka], [sa], [ʃa].

To this aim, for each pair of vowels in the dataset we ran a 2-, 3-, 4- and 5-clusters partition. We adopted the L2 distance between curves to compute a dissimilarity index of curves-medoids within each cluster, and affinities - namely shift and dilation - as the group of admissible warping functions. We used the index to evaluate the alignment/clustering of the curves to their medoid templates. Results confirm the quality of the clustering technique for comparing data across recording sessions. For example, Figure 1 illustrates how the tongue profiles for [a] and [e] that were recorded in distinct sessions are correctly reported to two homogeneous clusters. However, Figure 2 firstly shows that by increasing the number of clusters, the quality of classification does not markedly improve. This evidence suggests that fine-graded articulatory variability that is present in the original data is not captured, namely positional variables are not identified. Secondly, the reduction in the dissimilarity between the case without and with alignment supports the importance of managing the possible curve misalignment before performing any statistical analysis, either while collecting data - possibly by resorting to a bite plane trace that provides a consistent reference to allow images to be rotated and translated [10] -, or while pre-processing data for the statistical analysis.

During the presentation, we will discuss how this can depend both on the clustering technique and on the accuracy limits of the ultrasound tongue imaging technique, and we will give information regarding the optimum settings for obtaining the best results.

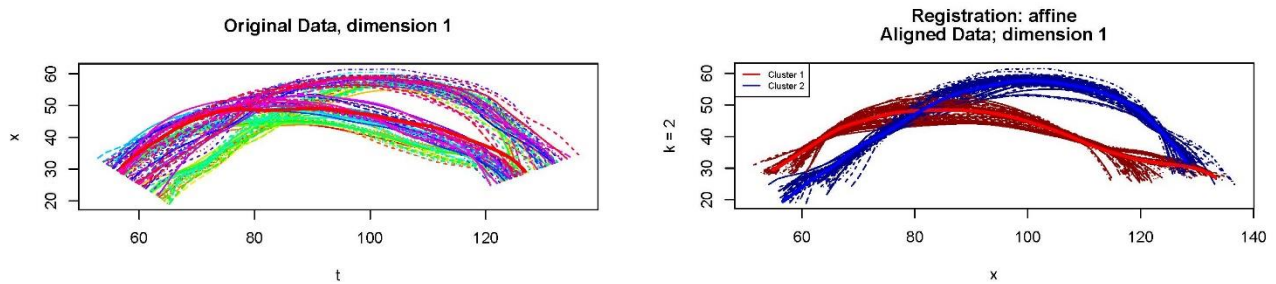


Figure 1. *Original curves (left) and aligned and clustered curves via 2-medoid alignment with affine warping (right). The colors in the left panel refer to the different positional variants of [a] and [e].*

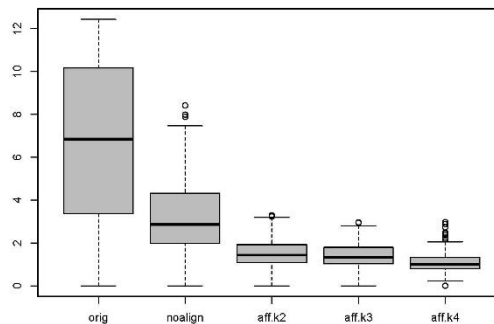


Figure 2. *Boxplot dissimilarity indexes for original, non-aligned, 2-, 3-, and 4-clusters partition for [a] and [e].*

- [1] Stone, M. A guide to analysing tongue motion from ultrasound images. 2005. *Clinical linguistics & phonetics* 19, 455-501.
- [2] Davidson, L. Comparing tongue shapes from ultrasound imaging using smoothing/spline analysis of variance. 2006. *The Journal of the Acoustical Society of America*, 120, 407-415.
- [3] Dawson, K. M., Tiede, M. K., & Whalen, D. H. Methods for quantifying tongue shape and complexity using ultrasound imaging. 2016. *Clinical linguistics & phonetics*, 30, 328-344.
- [4] Zharkova, N. Using ultrasound to quantify tongue shape and movement characteristics. 2013. *The Cleft Palate-Craniofacial Journal*, 50, 76-81.
- [5] Sangalli, L. M., Secchi, P., Vantini, S., Vitelli, V. k-mean Alignment for Curve Clustering. 2010. *Computational Statistics and Data Analysis*, 54, 1219-1233.
- [6] Sangalli, L. M., Secchi, P., Vantini, S., Vitelli, V. Functional clustering and alignment methods with applications. 2010. *Communications in Applied and Industrial Mathematics*, 1, 205-224.
- [7] Vietti, A. Pini, A, Spreafico, L., Vantini, S. Towards cross-subject comparison of tongue profiles. A Functional Data Analysis approach. 2017. *Oral presentation at Ultrafest VIII*, Potsdam 4-6 October 2017.
- [8] Articulate Instruments. 2011. *Articulate Assistant Advanced user guide*. Version 2.17.
- [9] Spreafico, L., Pucher, M., Matosova, A. UltraFit: A Speaker-friendly Headset for Ultrasound Recordings in Speech Science. *Proceedings of Interspeech 2018*, 1517-1520.

[10] Scobbie, J., Lawson, E., Cowen, S., Cleland, J., Wrench, A. A common coordinate system for mid-sagittal articulatory measurement. 2010. *Working Paper WP-20*. QMU Edinburgh.