Extraction of Persian coronal stops from ultrasound images using linear discriminant analysis

Reza Falahati
Scuola Normale Superiore di Pisa

Vahid Abolghasemi
University of Shahrood

Introduction

Ultrasound is an appealing technology which could be used for imaging the vocal tract. Similar to other techniques, it also has some limitations. Tracing the tongue contours in ultrasound images is a very time consuming task. Thirty minutes of tongue imaging at 60 fps will result in 108,000 images. Several different approaches to this problem have been proposed (Angul & Kambhamettu 2003; Baker 2005; Li et al. 2006; Fasel and Berry 2010; Tang et al. 2012; Hueber 2013; Pouplier & Hoole, 2013; Sung et al. 2013), with promising results. This study uses a new application called TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) developed by Carignan (2014) for extracting time-varying articulatory signals in large-scale image sets and compares the results with Falahati (2013) who has manually traced the tongue contours. The research question followed here is whether the automatic tracing can capture the articulatory differences between the simplified and unsimplified consonant clusters in Persian. The consonant clusters under study are composed of coronal stops [t d] followed and preceded by non-coronal consonants (i.e., $V_1C_1C_2\#C_3V_2$) where target coronal stops (i.e., C_2) could be optionally simplified.

Methodology

In order to choose the ultrasound images for processing, TextGrid was used to mark the target coronal consonants as well as the preceding and following consonants (i.e., C_1 & C_3) and also the two vowels adjacent to the three consonants in the middle. After choosing the images of interest, a feature reduction/extraction technique was applied. The open source software suite TRACTUS implemented in MATLAB was used for such end. The first step was to specify the border of ultrasound fan within the images followed by filtering. The goal at this stage was to strike a balance between the tongue contours and image noise (see Figure 1 top left). Choosing the region of interest (i.e., ROI) was the next step in the process. At this time the area of image showing the range of tongue contour movement was created (see Figure 1 top right). The final stage in using TRACTUS was to generate PC scores. This was the result of applying principle component analysis (PCA, Hueber et al. 2007) to the processed data. PC scores represent "the degree to which the imaged vocal tract matches a limited set of articulatory configurations which are identified by the PCA model" (Carignan & Mielke, p. 4). The combinations of PCs result in heatmaps illustrating the means (see Figure 1 bottom).

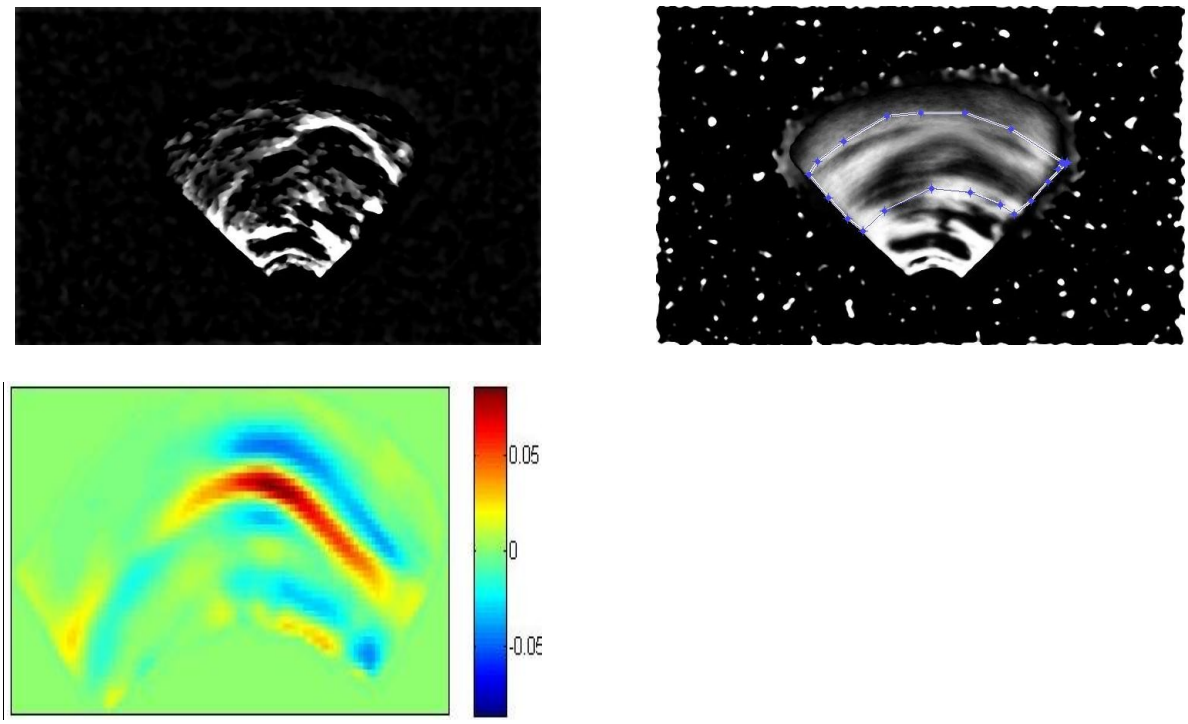

Figure 1: Top left: filtered image; Top right: polygonal ROI; Bottom: Heatmap for PC1.

TRACTUS tool is helpful up to the point of creating the PC scores from the ultrasound data. Once the PC scores were created, they were transformed via liner discriminant analysis (i.e., LDA) to create signals as inputs to an LDA model with classes for the simplified and unsimplified coronal stops [t d] as well as the remaining sounds. The articulatory signals generated for individual tokens in our study are analogous to tracing one specific point of the tongue over temporal dimension to generate gestural scores (Falahati 2013; Pouplier & Hoole, 2013; Carignan & Mielke 2014).

Results

The research question followed in this study was whether the articulatory signals generated for coronal stops [t d] could distinguish between the tokens with simplified and unsimplified consonant clusters and whether the result was comparable to Falahati (2013). The preliminary results of this study for one subject shows that this method is quite successful for teasing apart the tokens with full alveolar gestures versus the ones which lack it. The results of these token frames traced manually in Falahati (2013) supports the results for LDA class scores here. Figure 2 below illustrates a representative number of tokens with and without coronal gestures.

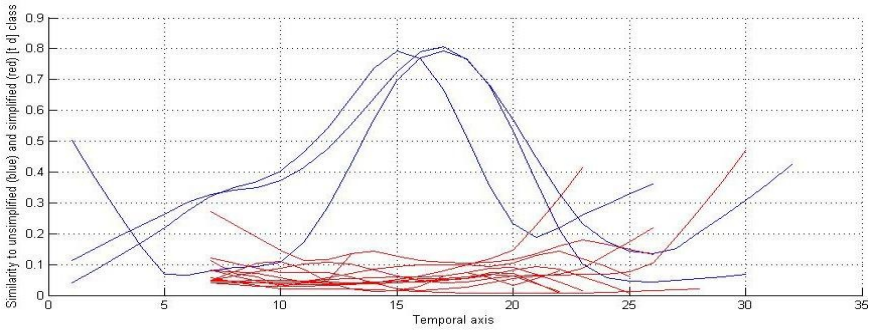

Figure 2: The LDA class scores over time. Tokens with unsimplified coronal stops (blue); Tokens with simplified coronal stops (red).

References

Baker, A. 2005. *Palatoglossatron 1.0*. University of Arizona, Tucson, Arizona. http://dingo.sbs.arizona.edu/~apilab/pdfs/pgman.pdf.

Carignan, C. 2014. TRACTUS (Temporally Resolved Articulatory Configuration Tracking of UltraSound) software suite. *URL: http://phon.chass.ncsu.edu/tractus*.

Carignan, C., & Mielke, J. 2014. Extracting articulatory signals from lingual ultrasound video using principal component analysis. MS.

Falahati, R. 2013. *Gradient and Categorical Consonant Cluster Simplification in Persian: An Ultrasound and Acoustic Study*. Ph.D dissertation, University of Ottawa.

Fasel, I. and Berry, J. 2010. Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. *In Proceedings of the 20th International Conference on Pattern Recognition*, pp. 1493–1496.

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., and Stone, M. 2007. Eigen tongue feature extraction for an ultrasound-based silent speech interface. In *Proceedings of 2007 International Conference on Acoustics, Speech, and Signal Processing*, pp. 1245–1248.

Hueber, T. 2013. Ultraspeech tools: Acquisition, processing and visualization of ultrasound speech data for phonetics and speech therapy. *In Proceedings of Ultrafest VI Conference, pp.* 10-11.

Li, M., Kambhamettu, C., and Stone, M. 2005. Automatic contours tracking in ultrasound images. *Clinical Linguistics and Phonetics*, 19:545–554.

Pouplier, M. and Hoole, P. 2013. Comparing principal component analysis of ultrasound images with contour analyses in a study of tongue body control during German coronals. *In Proceedings of Ultrafest VI Conference,* pp. 25-26.

Sung, J. H., Berry, J., Cooper, M., Hahn-powell, G., and Archangeli, D. 2013. Testing AutoTrace: A machine learning approach to automated tongue contour data extraction. In Proceedings of Ultrafest VI Conference, pages 9-10.

Tang, L., Bressmann, T., and Hamarneh, G. 2012. Tongue contour tracking in dynamic ultrasound via higher order MRFs and efficient fusion moves. *Medical Image Analysis*, 16:1503–1520.