A method for automatically detecting problematic tongue traces

Gus Hahn-Powell[1], **Benjamin Martin**[1], and **Diana Archangeli**[1,2]

[1]*Department of Linguistics, University of Arizona*
[2]*Department of Linguistics, University of Hong Kong*

While ultrasound provides a remarkable tool for tracking the tongue's movements during speech, it has yet to emerge as the powerful research tool it could be. A major roadblock is that the means of appropriately labeling images is a laborious, time-intensive undertaking. In work reported at ICPR in 2010, Fasel and Berry (2010) introduced a "translational" deep belief network (tDBN) approach to automated labeling of ultrasound images. The current work extends that methodology with a modification of the training procedure to reduce reported errors (Sung and Archangeli, 2013) along the anterior and root edges of the tongue by altering the network's loss function and incorporating ℓ^1 and ℓ^2 regularization (Ng, 2004) to avoid overfitting. This training-internal approach to error reduction is compared to an independent post-processing procedure which uses the expected average positional change between adjacent points in three tongue regions (Davidson, 2006) to detect and constrain erroneous coordinates. Positional variance was calculated using the 800 most diverse and 50 least diverse tongue configurations by image pixel intensity across multiple subjects from a recitation of the phonetically balanced Harvard sentences (Rothauser et al., 1969).
Index Terms: articulatory phonetics, ultrasound imaging, tongue imaging, speech processing, deep belief networks, regularization, computer vision

References

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of Acoustical Society of America*, 120:407–415.

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*, pages 1493–1496.

Ng, A. Y. (2004). Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78. ACM.

Rothauser, E., Chapman, W., Guttman, N., Nordby, K., Silbiger, H., Urbanek, G., and Weinstock, M. (1969). Ieee recommended practice for speech quality measurements. *IEEE Trans. Audio Electroacoust*, 17(3):225–246.

Sung, J.-H., B. J. C. M. H.-P. G. and Archangeli, D. (2013). "testing autotrace: A machine-learning approach to automated tongue contour data extraction". Edinburgh. UltraFest VI.