A method for automatically detecting problematic tongue traces

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

[1]Department of Linguistics, University of Arizona

[2]Department of Linguistics, University of Hong Kong

Current work

- ***Background***: Fasel and Berry [1] introduced a deep belief network approach to automated labeling of tongue contours
- ***Problem***: Errors along the anterior and root edges of the tongue are common [2, 3]

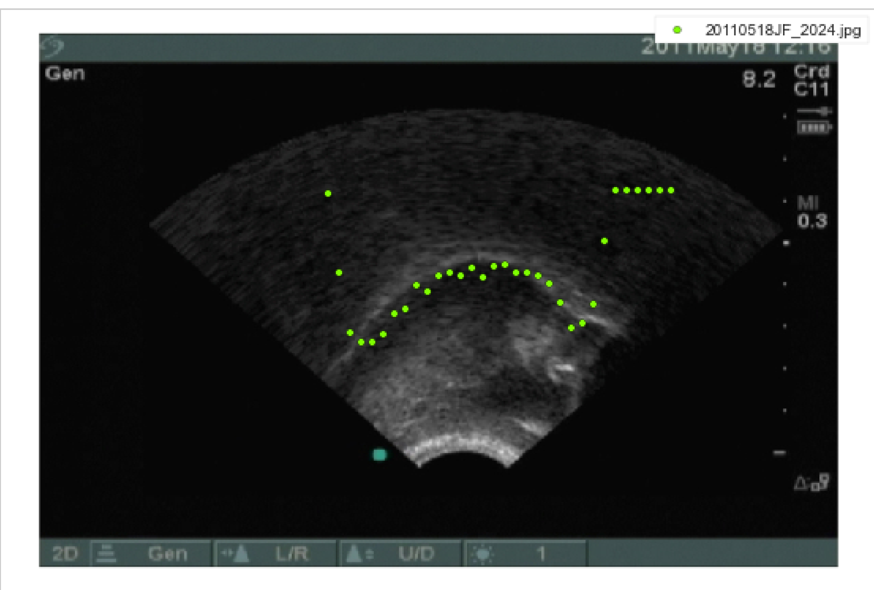


Figure 1: A common error

- ***Goal***: Reduce errors by modifying *training* and *post-processing* procedures

Training data

- 850 images from the Harvard Sentences [4]
  - selected **800** most + **50** least diverse

Why Harvard Sentences?

- **Large dataset**
  - Image pool of 33K+
- **Diverse**
  - Phonetically balanced dataset
    - 1906 distinct words
  - 11 speakers
  - Many tongue configurations

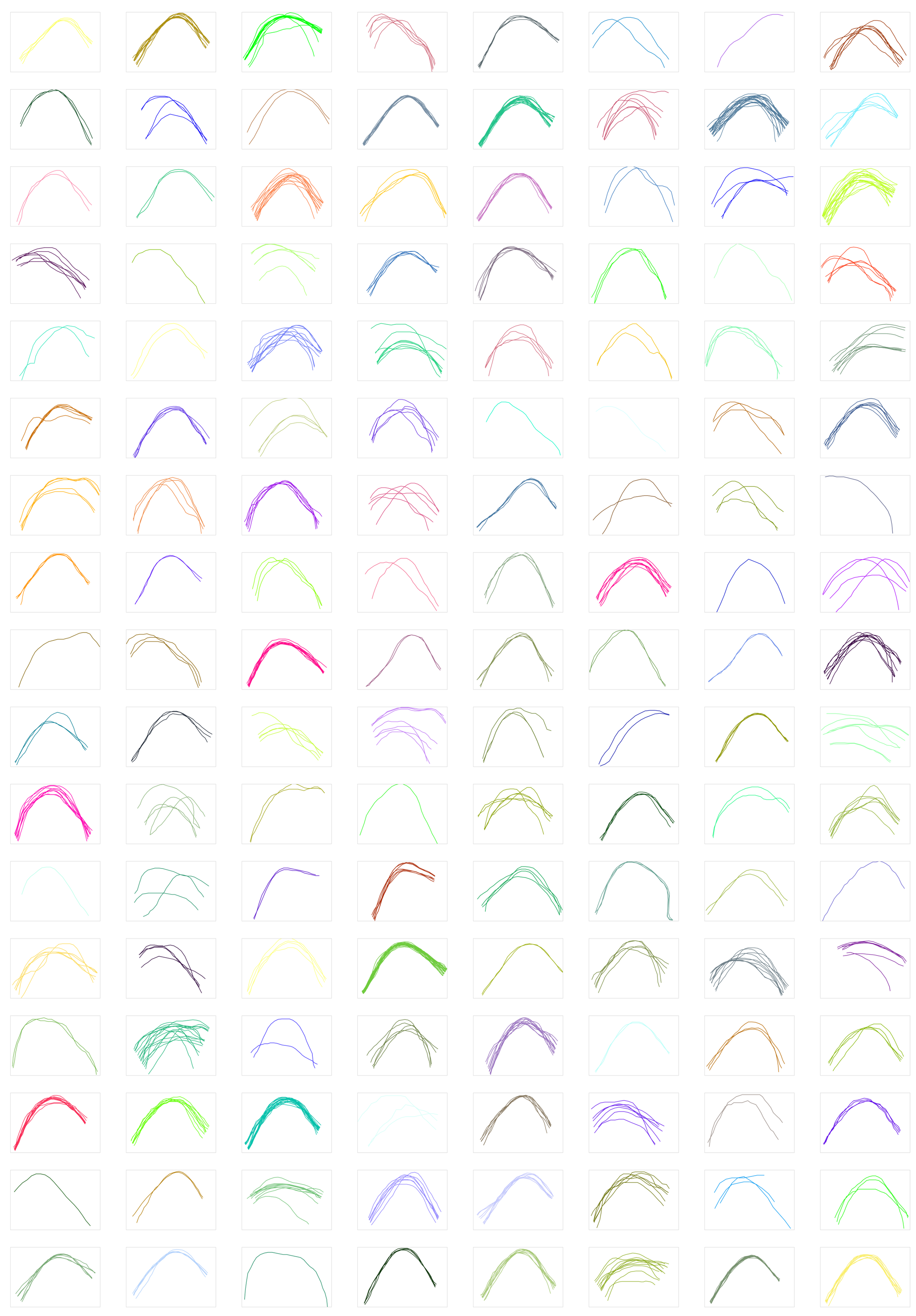


Figure 2: 180 k-means clusters of tongue contours in the training data

Network architecture

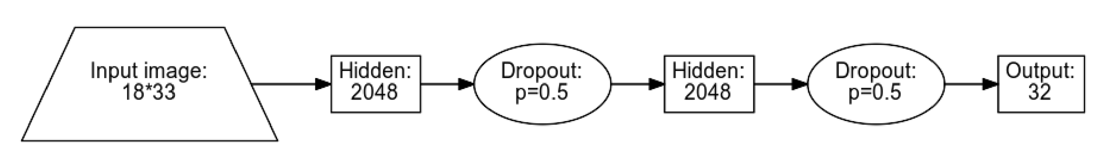


Figure 3: Autotres architecture comprised of two hidden layers with 2048 units each and dropout with probability 0.5.

Training modifications

Preprocessing images

- Add Gaussian blur to reduce image noise

Scaling images

- Decrease image size to simplify training
  - fewer pixels = fewer units

Regularization

- Add ℓ^1 and ℓ^2 regularization [5] to avoid overfitting

$$\mathbf{R}_{\ell^p}(\theta) = \|\theta\|_p = \left(\sum_{j=0}^{|\theta|} \left|\theta_j\right|^p\right)^{\frac{1}{p}}$$

Equation 1: Generalized equation for the ℓ^p-norm

Post-processing

Constraining regional change

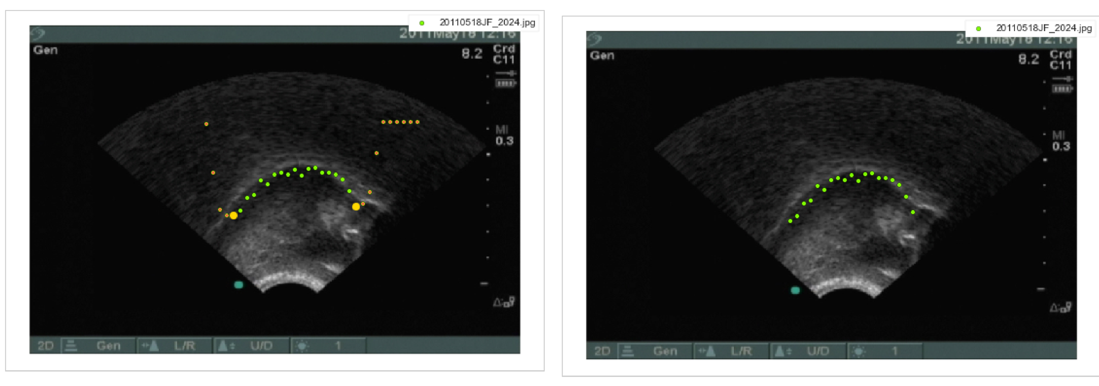
- Detect physiological impossiblities

Figure 4: Use the expected average positional change between adjacent points in three tongue regions [6] to detect and constrain erroneous coordinates.

Figure 5: Corrected through derivative-based constraint

Trimming from local minima

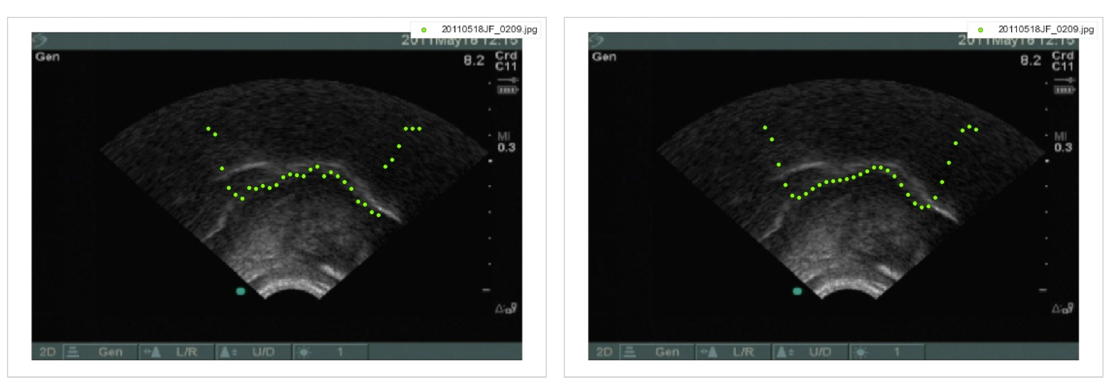
- Find regional y-axis minima



(a) Highlighted regional minima **(b)** Trimmed from local minima

Smoothing via polynomial fit

- Use Savitzky-Golay filtering [7, 8]



(a) Autotres output **(b)** Smoothed

Figure 7: 9^{th} order polynomial fit

Results

$$MSD(U,V) = \frac{1}{2n}\left(\sum_{i=1}^{n} \min_j \left|v_i - u_j\right| + \sum_{i=1}^{n} \min_j \left|u_i - v_j\right|\right)$$

Equation 2: Mean Sum of Distances (MSD) method of Li et al. [9]

Method	MSD
AutoTrace (original)	
Default settings	29.4
(S-G + *minima*)	25.19
Autotres (new)	
.1 scaling	42.6
Regularization	42.5
Savitzky-Golay smoothing	42.9
Constrained regions	40
Trimming minima	16.64
(*Regularization* + S-G + *minima*)	**15.71**

Future work

- New architectures
  - convolutions
  - multiple networks for tongue regions
  - layer modeling whether or not a point should be produced
    - some images don't have a contour
- Unsupervised pretraining

Code

`Autotres`

- third-generation software
- Python 3
- fast training on the GPU
- open source and freely available
- includes utilities for training, tracing, visualization, and evaluation

`https://github.com/arizona-phonological-imaging-lab/autotres`

Acknowledgements

This work was funded by National Science Foundation grants 1059266 and 1244687

References

[1] Ian Fasel and Jeff Berry. 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, 2010.

[2] Berry J. Cooper M. Hahn-Powell G. Sung, J.-H. and D. Archangeli. “testing autotrace: A machine-learning approach to automated tongue contour data extraction”. Edinburgh, 2013. UltraFest VI.

[3] Tamás Gábor Csapó and Steven M Lulich. Error analysis of extracted tongue contours from 2d ultrasound images. In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[4] EH Rothauser, WD Chapman, N Guttman, KS Nordby, HR Silbiger, GE Urbanek, and M Weinstock. Ieee recommended practice for speech quality measurements. *Audio and Electroacoustics, IEEE Transactions on*, 17(3):225–246, 1969. doi: 10.1109/TAU.1969.1162058.

[5] Andrew Y Ng. 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, 2004.

[6] Lisa Davidson. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of Acoustical Society of America*, 120:407–415, 2006.

[7] Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[9] Min Li, Chandra Kambhamettu, and Maureen Stone. Automatic contour tracking in ultrasound images. *Clinical linguistics & phonetics*, 19(6-7):545–554, 2005.