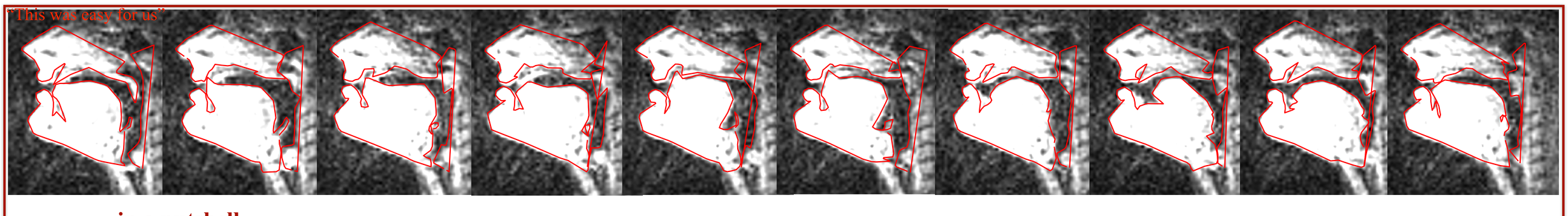
# validating rt-MRI based articulatory representations via articulatory recognition

*A. Katsamanis, E. Bresch, V. Ramanarayanan, S. Narayanan*

Signal Analysis and Interpretation Laboratory, Electrical Engineering, University of Southern California, Los Angeles, CA 90089

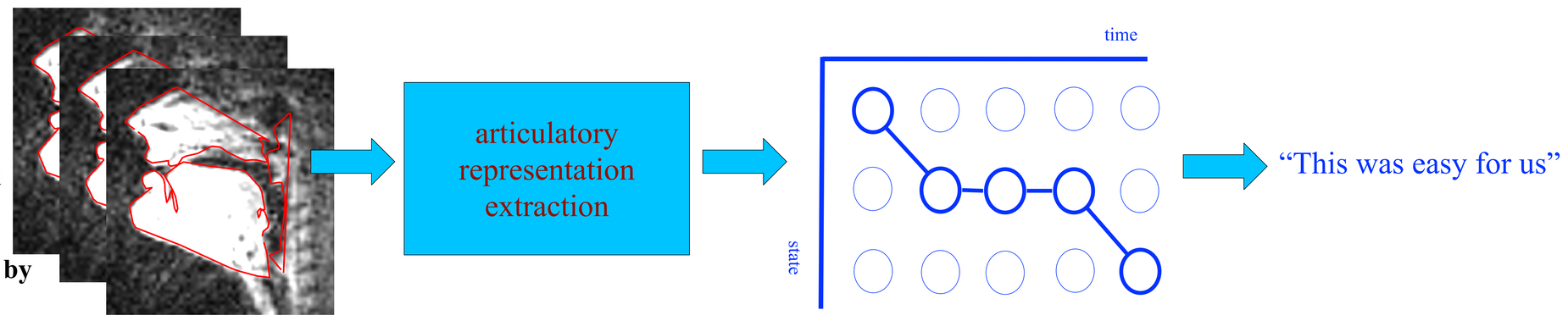
http://sail.usc.edu

Research supported by NIH Grant R01 DC007124-01

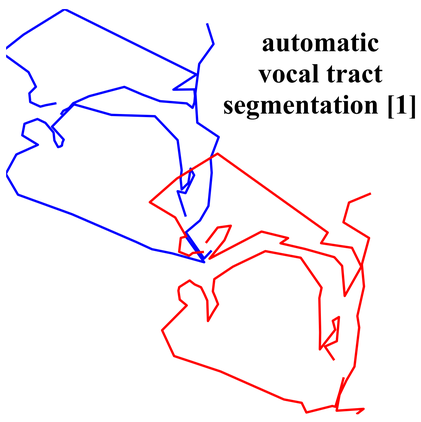


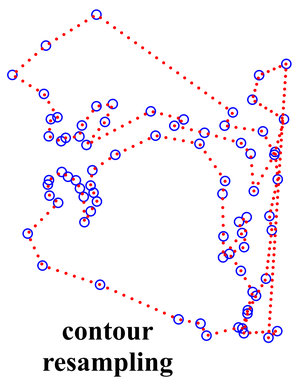

## in a nutshell

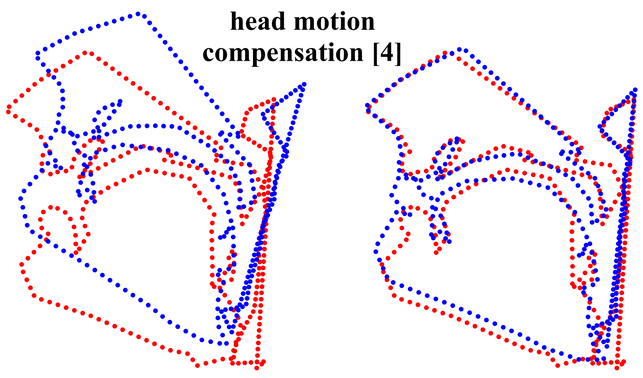
- our goal is to devise an articulatory representation that would allow us to computationally analyze real-time speech production data
- we investigate deformable shape model and linear discriminant analysis based representations
- we validate alternative representations by HMM-based articulatory recognition

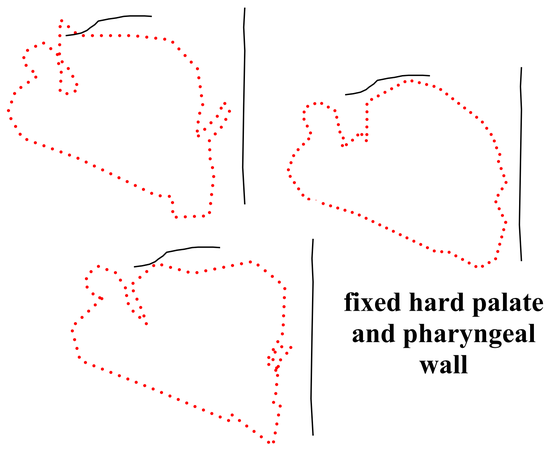



## vocal tract outlines

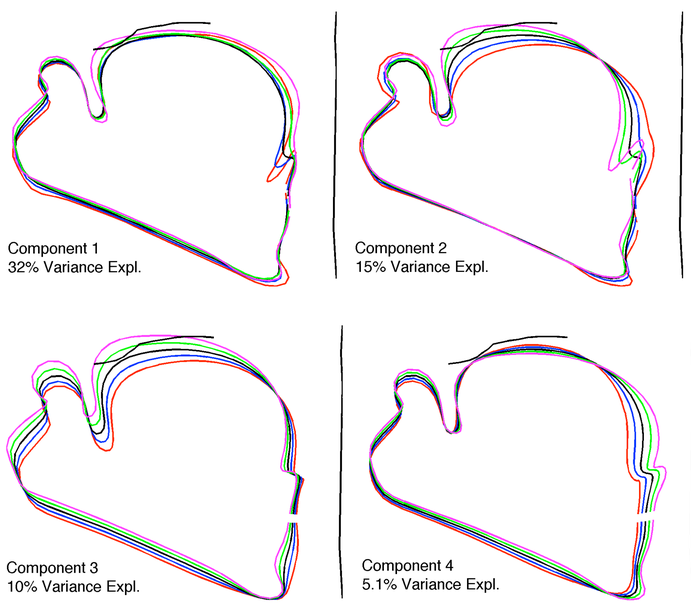



## statistical deformable model using principal component analysis




- we want to model the (co-)variation of the points on the vocal tract contour
- each vocal tract corresponds to a point in an M-dimensional space and all such points form a cloud in the "Allowable Shape Domain" [3]
- it is assumed that this cloud is approximately ellipsoidal and its center and major axes are estimated using principal component analysis
- the eigenvectors of the covariance matrix corresponding to the largest eigenvalues describe the longest axis of the ellipsoid
- 26 components describe more than 95% of the variance

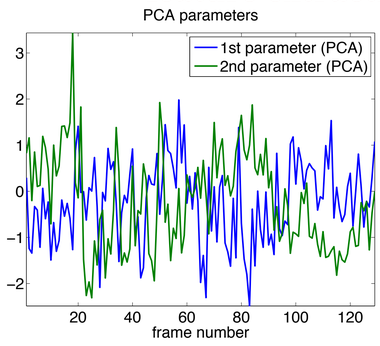
## linear discriminant analysis

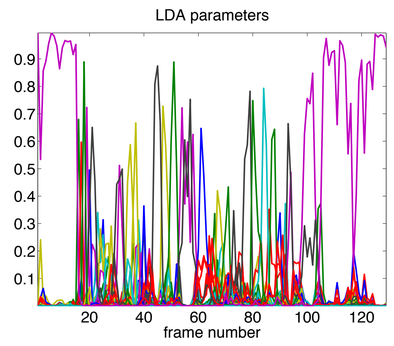
- our goal is to obtain a vocal tract representation that would preserve as much of the discriminative information among classes of shapes (one class per phoneme) as possible

$$\mathbf{b} = W^T \mathbf{s}$$

- assuming unimodal Gaussian distributions of the points for each class the optimal projection matrix can be found via linear discriminant analysis, by maximizing the ratio of between-class to within-class scatter

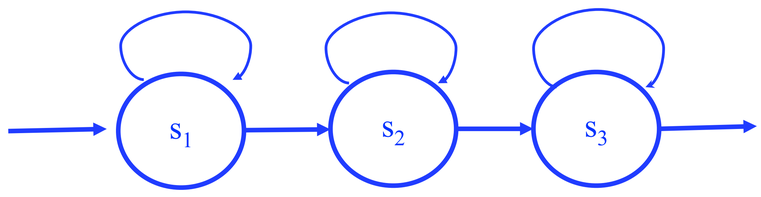
## feature extraction

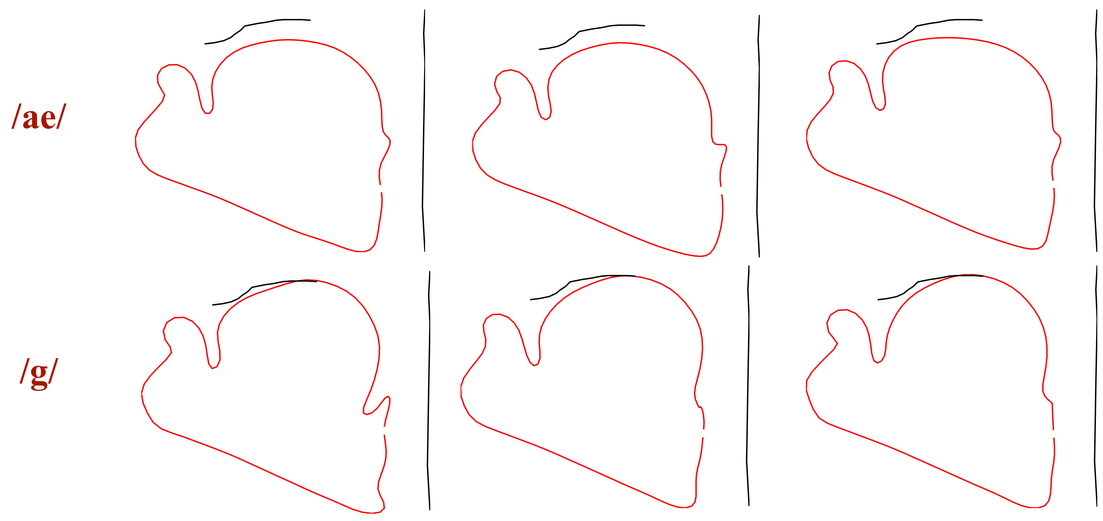



- "Draw every outer line first then fill in the interior"

## 3-state left-right phoneme hidden Markov models




## articulatory recognition results




- 40 LDA / 25 PCA based features
- results using image intensity based discrete cosine transform features are also shown (100 features)

## training

- the models were initialized using the timed phonetic transcriptions of the training sequences
- 4-component Gaussian mixture models as observation probability distributions
- each pair of phonemes whose articulation presumably only varies in terms of voicing are considered to be in the same class, e.g., /z/ and /s/ or /t/ and /d/
- training was performed using the Hidden Markov Model toolkit (HTK)
- finally, the articulatory models are allowed to be asynchronously aligned with the corresponding acoustics

## references

[1] E. Bresch and S. Narayanan, "Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images," *IEEE Trans. Medical Imaging, vol. 28, no. 3, pp.*323–338, Mar 2009.

[2] S. Narayanan, E. Bresch, P. Ghosh, L. Goldstein, A. Katsamanis, Y. Kim, A. Lammert, M. Proctor, V. Ramanarayanan, and Y. Zhou, "A multimodal real-time MRI artculatory corpus for speech research," in Proc. of Interspeech 2011.

[3] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models – their training and application," *Computer Vision and Image Understanding, vol. 61, pp. 38–59,* 1995.

[4] P. Besl and N. McKay, "A method for registration of 3-d shapes," *IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, pp. 239–256,* 1992.