

Continuous Profile Models in ASL Syntactic Facial Expression Synthesis

Hernisa Kacorri
Carnegie Mellon University

Matt Huenerfauth
R.I.T

To create accessible content for deaf users, we investigate automatically synthesizing animations of American Sign Language (ASL). Based on recordings of humans performing various types of syntactic face and head movements, we evaluate the efficacy of Continuous Profile Models at identifying an essential “latent trace” of the performance, for use in producing ASL animations. The effectiveness of the approach is supported by a metric-based evaluation and a study with deaf users.

Continuous Profile Model (CPM)

CPM (Listgarten et al., 2004) can align a set of time series data while simultaneously accounting for changes in their amplitude. With the **assumption** that a noisy, stochastic process generates the observed time series data, CPM infers the underlying noiseless representation of the data (latent trace).

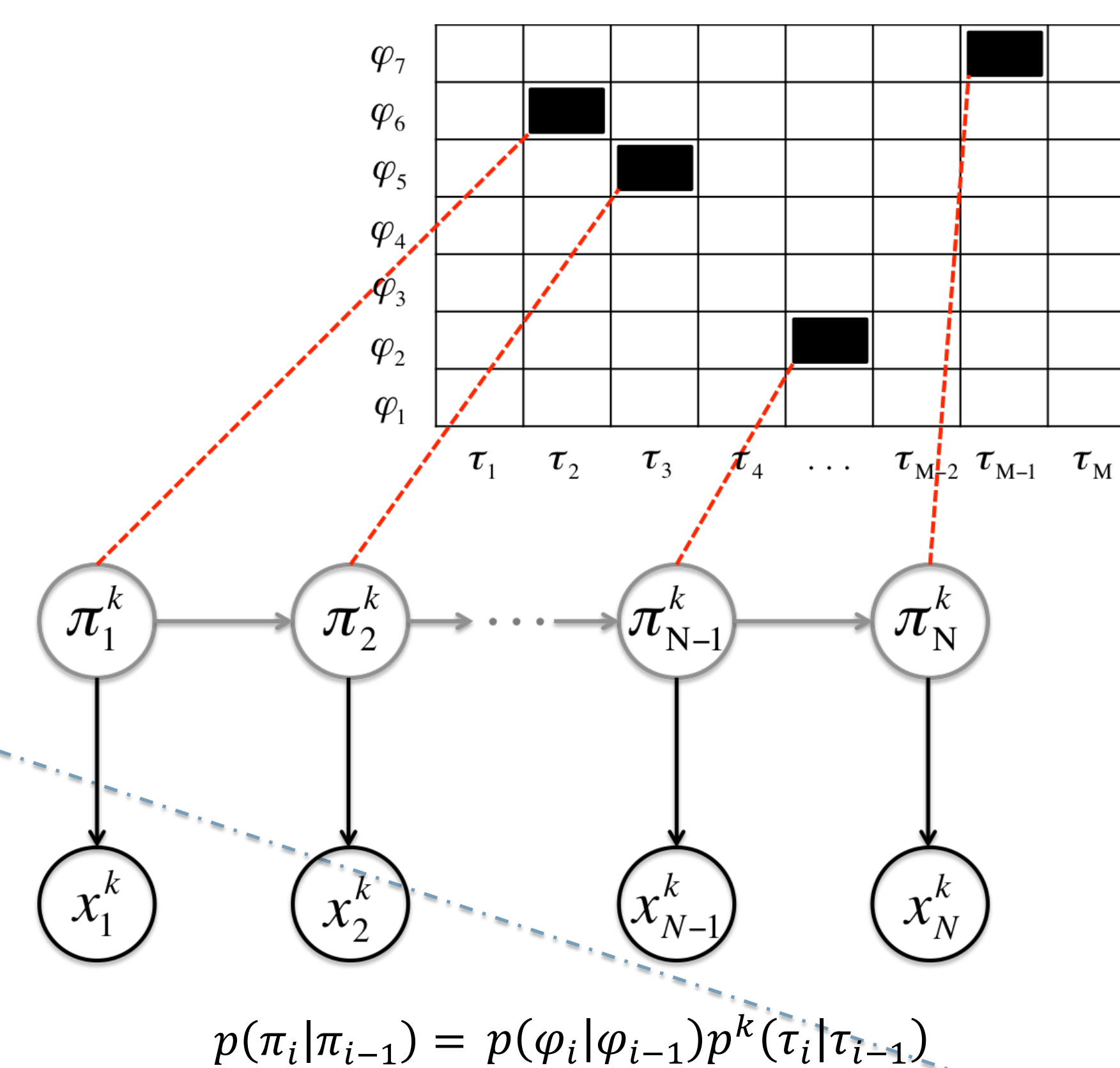
k observed time series: $\vec{x}^k = (x_1^k, x_2^k, \dots, x_N^k)$

Each x_i^k is assumed to be emitted by a Gaussian distribution:

$$\mu_i^k = z_{\tau^k} \phi_i^k u^k \quad \sigma: x_i^k \sim \mathcal{N}(\mu_i^k, \sigma)$$

Assume there is a latent trace: $\vec{z} = (z_1, z_2, \dots, z_M)$

$$M = (2 + \epsilon)N$$

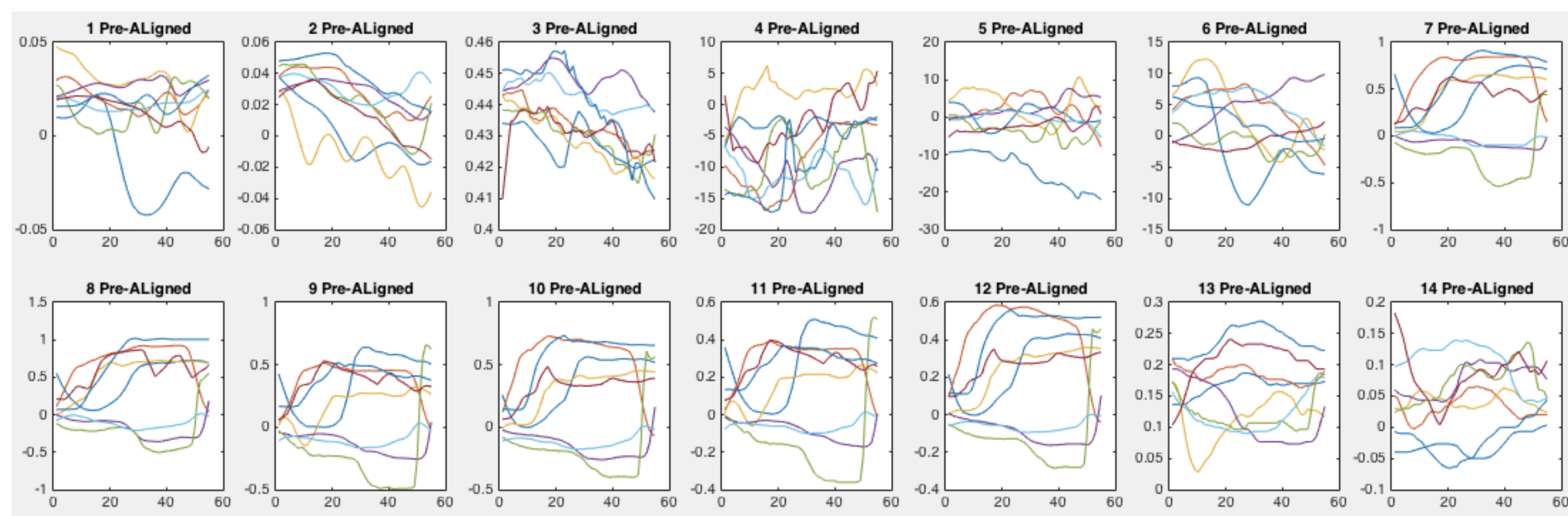


ASL Facial Expressions

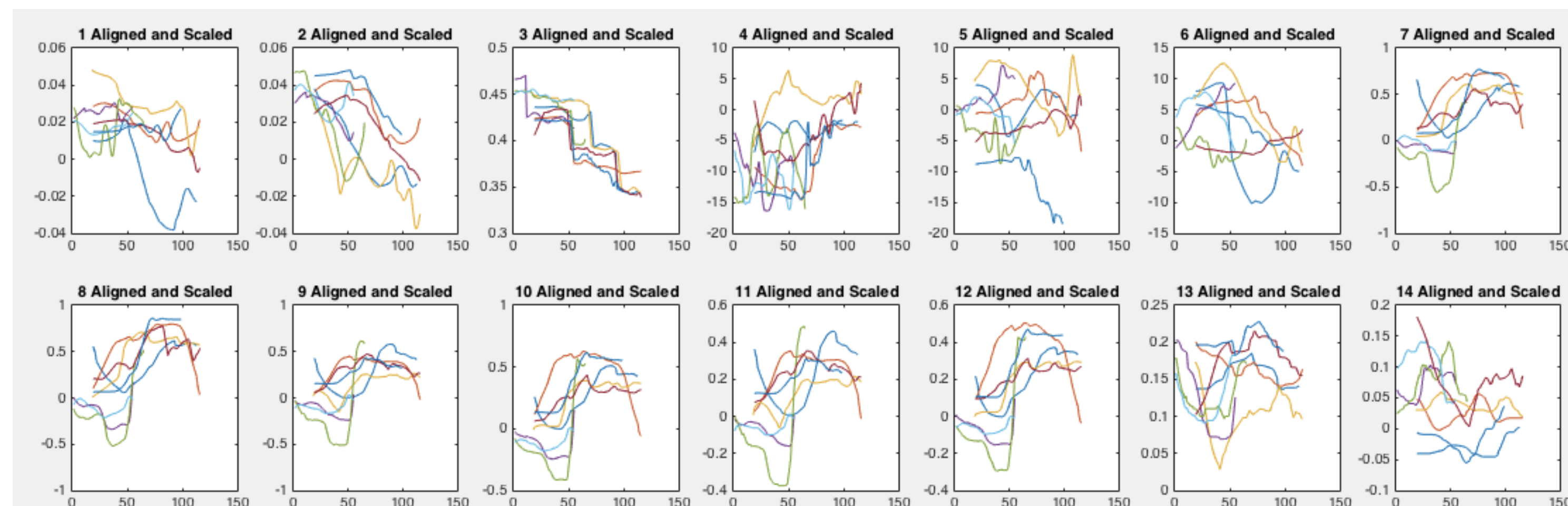
Type	Subgroup “A” (Num. of Videos)	Subgroup “B” (Num. of Videos)
YesNo	Immediately preceded by a facial expression with raised eyebrows, e.g. Topic. (9)	Not immediately preceded by an eyebrow-raising expression. (10)
WhQuestion	Performed during a single word, namely the wh-word (e.g., what, where, when). (4)	Performed during a phrase consisting of multiple words. (8)
Rhetorical	Performed during a single word, namely the wh-word (e.g., what, where, when). (2)	Performed during a phrase consisting of multiple words. (8)
Topic	Performed during a single word. (29)	Performed during a phrase consisting of multiple words. (15)
Negative	Immediately preceded by a facial expression with raised eyebrows, e.g. Topic. (16)	Not immediately preceded by eyebrow-raising expression. (25)

Example of CPM Modeling for Rhetorical_B

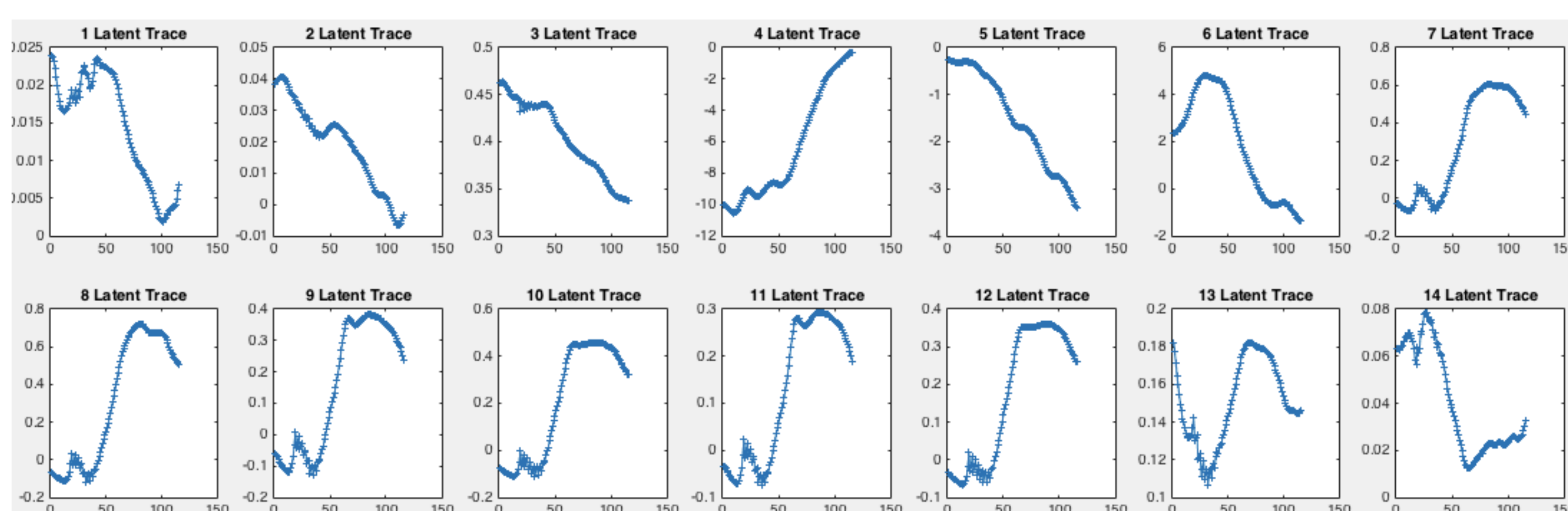
Training Examples Before CPM



Training Examples After CPM



Latent Trace

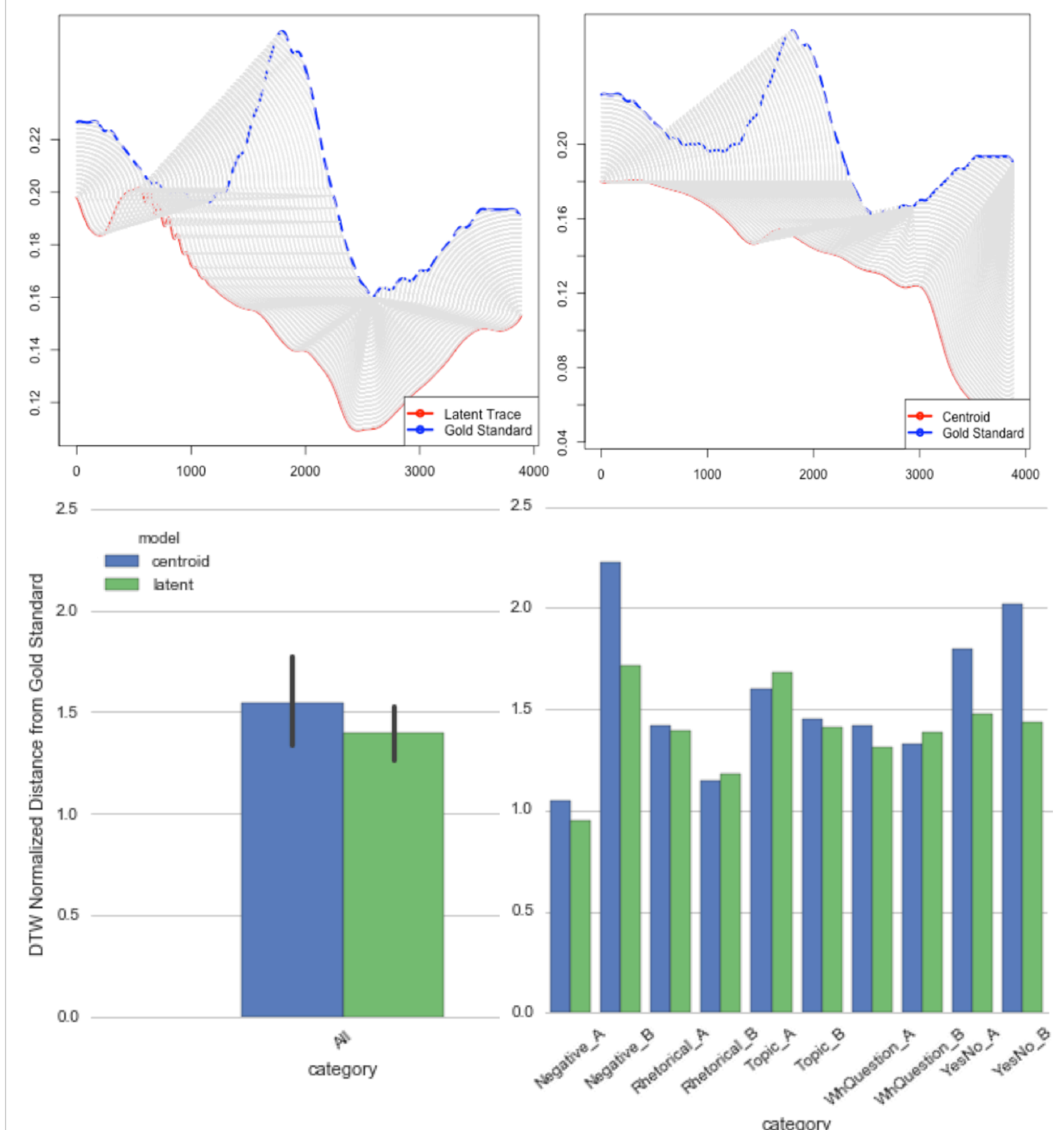


Metric Evaluation

$d_{DTW}(\text{latent trace, gold standard})$ vs. $d_{DTW}(\text{centroid, gold standard})$

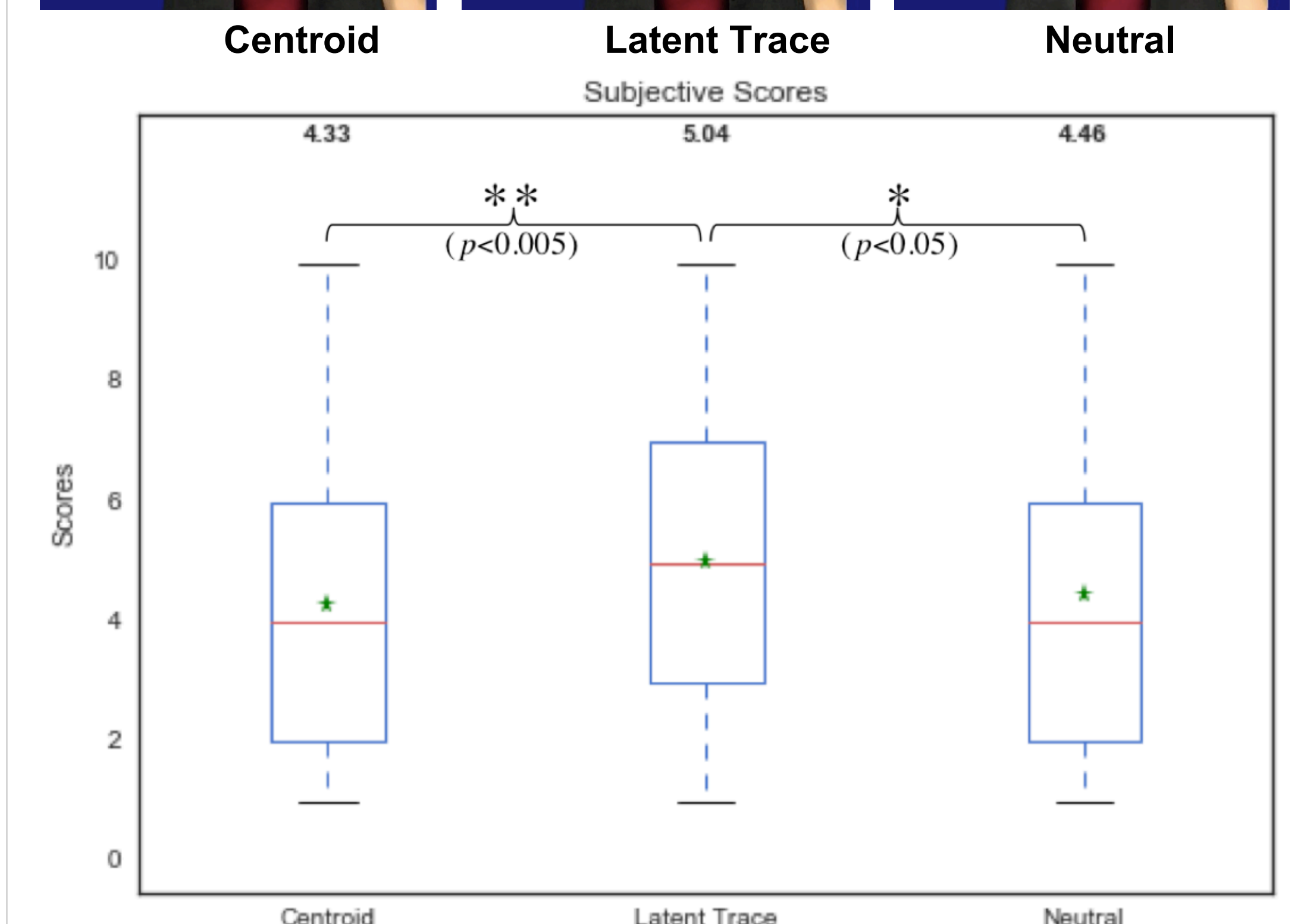
Centroid: example with min cumulative DTW distance.

Gold standard: recordings from another ASL signer.



User Evaluation

Compare ASL signers' subjective responses to neutral-head/face animations and animations driven by latent trace and centroid.



Acknowledgments

This material is based upon work supported by the National Science Foundation under award number 1065009 and 1506786. This material is also based upon work supported by the Science Fellowship and Dissertation Fellowship programs of The Graduate Center, CUNY. We are grateful for support and resources provided by Ali R. Syed at The Graduate Center, CUNY, and by Carol Neidle at Boston University.