Experimental Evaluation of Latent Variable Models for Dimensionality Reduction

Miguel Á. Carreira-Perpiñán and Steve Renals[a]

Dept. of Computer Science, University of Sheffield

{M.Carreira,S.Renals}@dcs.shef.ac.uk

9th IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing (NNSP98)

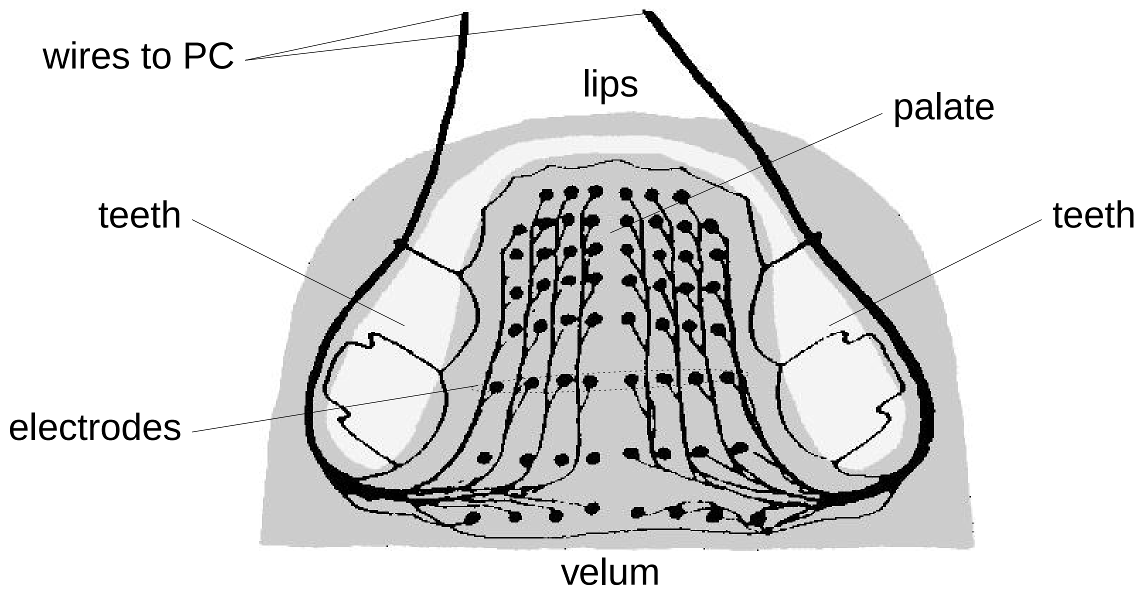
Aug. 31–Sep. 3, 1998, Cambridge, UK

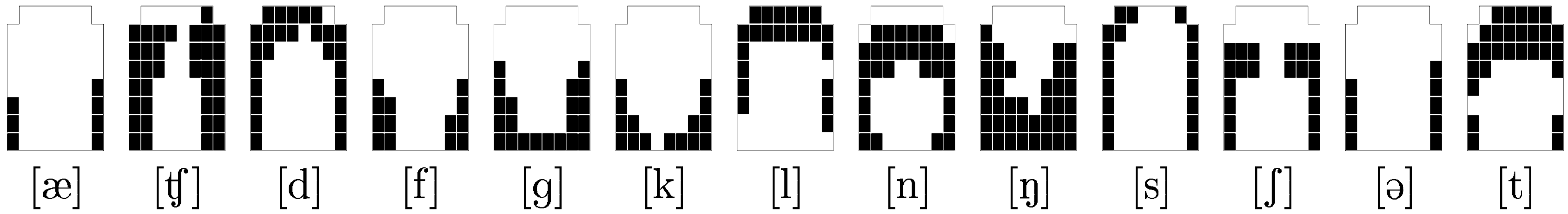
[a]This work has been supported by a scholarship from the Spanish Ministry of Education and Science, by a ESPRIT Long Term Research Project SPRACH (20077) and by an award from the Nuffield Foundation.

Electropalatography (EPG)

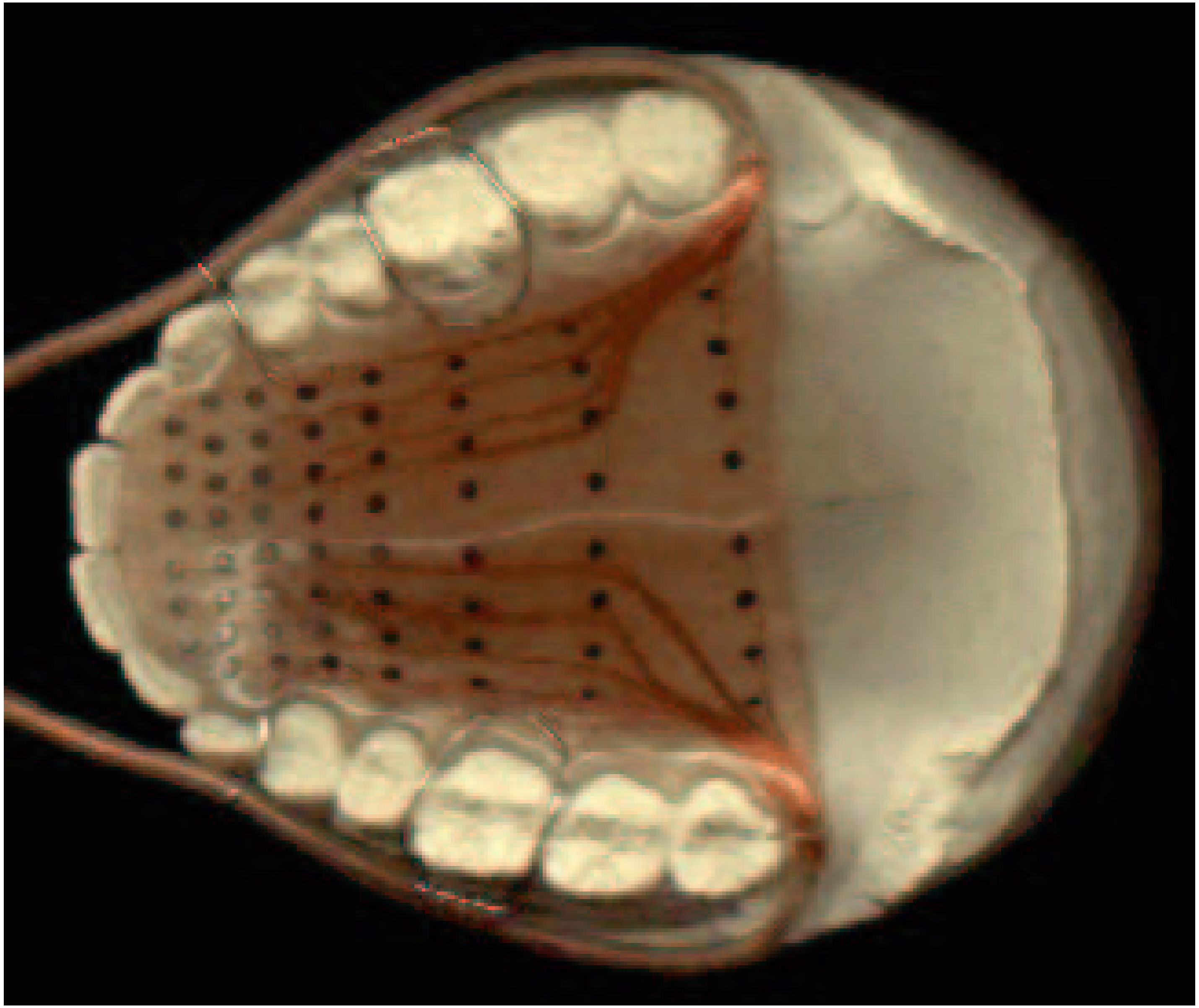
- ▷ A plastic pseudopalate fitted to a person's mouth detects the presence or absence of contact between the tongue and the palate in 62 different locations during an utterance (sampled at 200 Hz).
- ▷ Result: sequence of 62-dimensional binary EPG frames.
- ▷ Data reduction necessary, traditionally via fixed linear indices.
- ▷ ACCOR-II database: synchronised data (EPG, acoustic, etc.) for different utterances and speakers.
- ▷ The mapping phoneme-to-EPG is not one-to-one, e.g. [æ]/[ə] or [d]/[t].

Electropalatography (cont.)



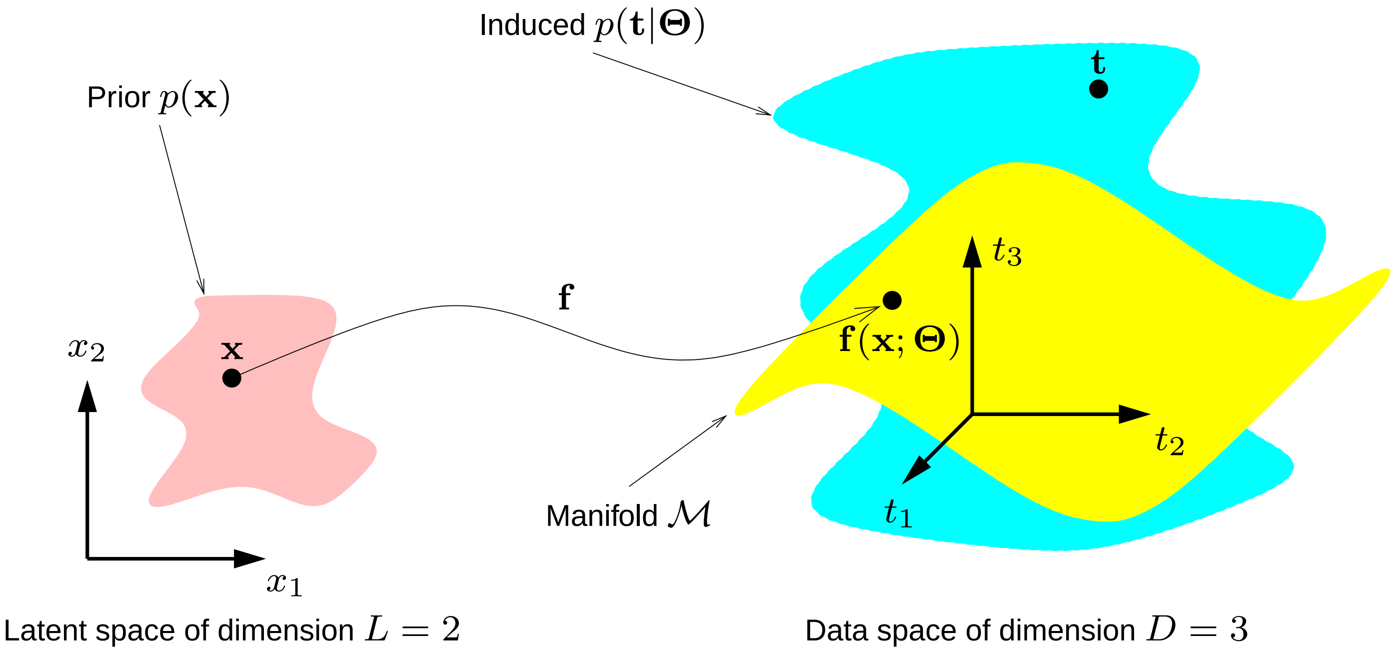


Pseudopalate and representative EPGs for the typical stable phase of different phonemes.



The Reading pseudopalate.

Latent variable models



- ▷ Marginalisation in latent space: $p(\mathbf{t}) = \int p(\mathbf{t}|\mathbf{x})p(\mathbf{x})\, d\mathbf{x}$.
- ▷ Maximum likelihood parameter estimation: $l(\boldsymbol{\Theta}) = \sum_{n=1}^{N} \log p(\mathbf{t}_n|\boldsymbol{\Theta})$.
- ▷ Inverse mapping given by informative point (mean, mode) of posterior: $p(\mathbf{x}|\mathbf{t}) = \frac{p(\mathbf{t}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{t})}$.

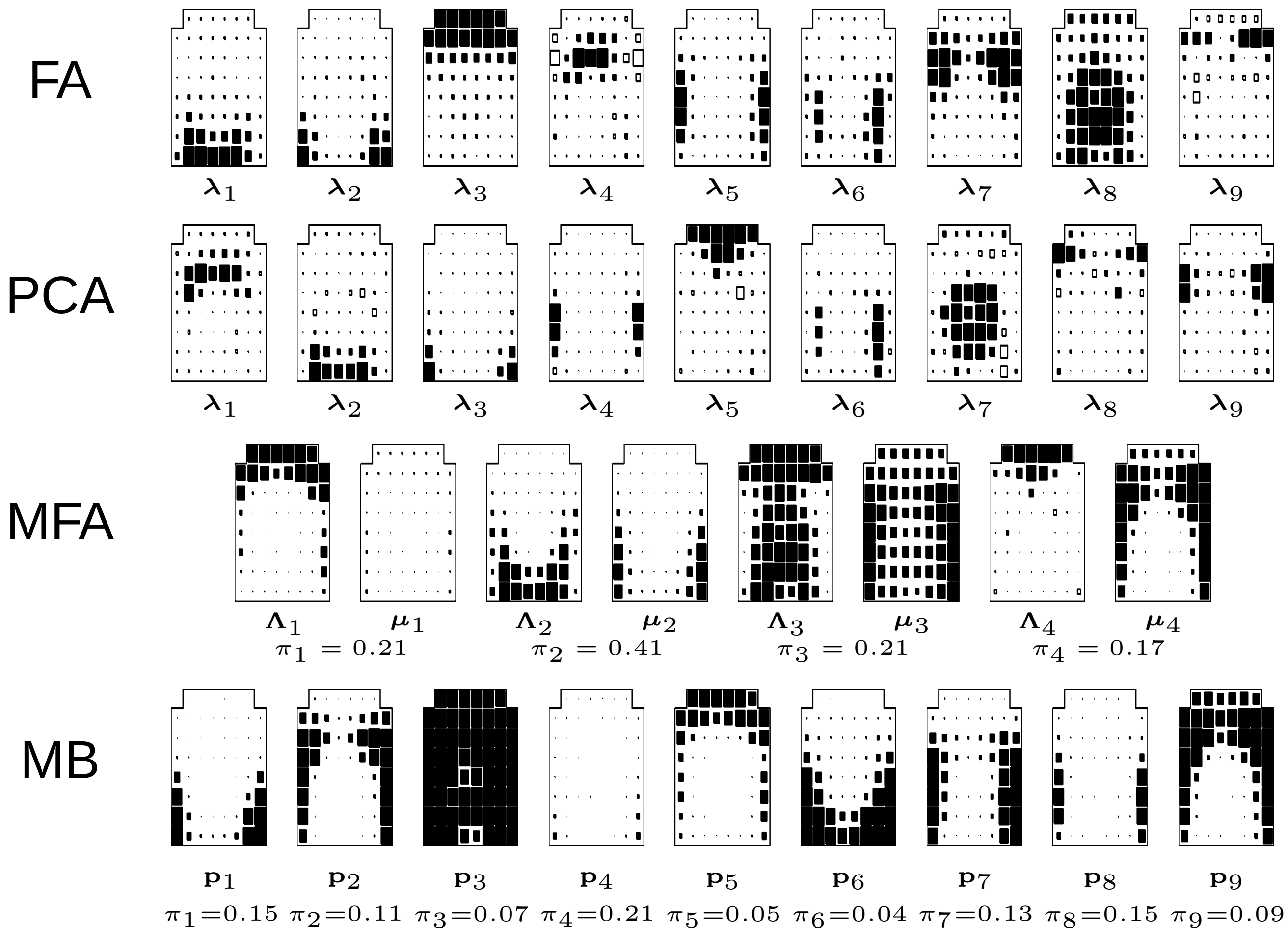
Examples of latent variable models

- Factor analysis: prior is normal $\mathcal{N}(\mathbf{0}, \mathbf{I})$, mapping is linear, noise model is normal with diagonal covariance matrix.
- PCA: like FA but noise model has isotropic covariance.
- GTM: prior is uniform over discrete latent grid, mapping is a generalised linear model, noise model is normal with isotropic covariance matrix.
- Mixtures of factor analysers (one mean parameter and one factor per analyser, noise model covariance matrix common to all analysers).

We also tried mixtures of multivariate Bernoulli distributions (not really a latent variable model).

All these models can be trained via an EM algorithm.

Factors / prototypes (speaker RK)

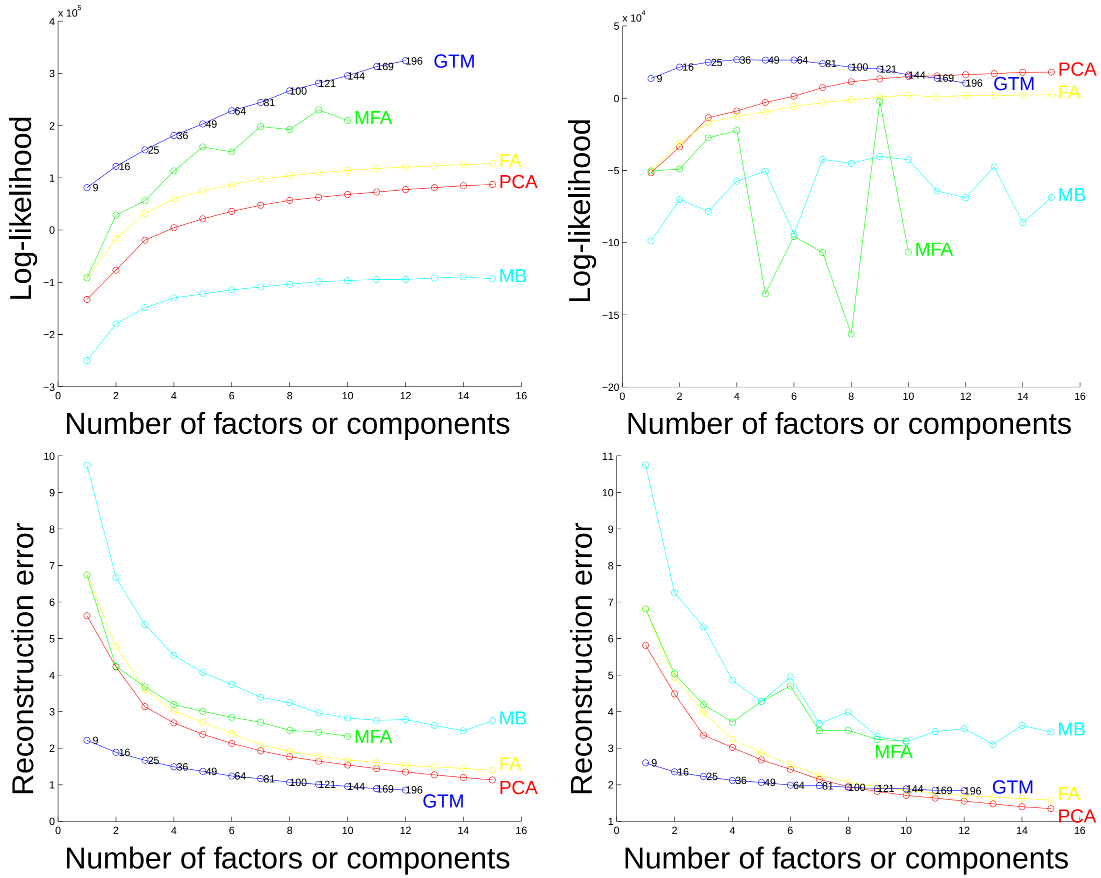


Log-likelihood and reconstruction error (speaker RK)

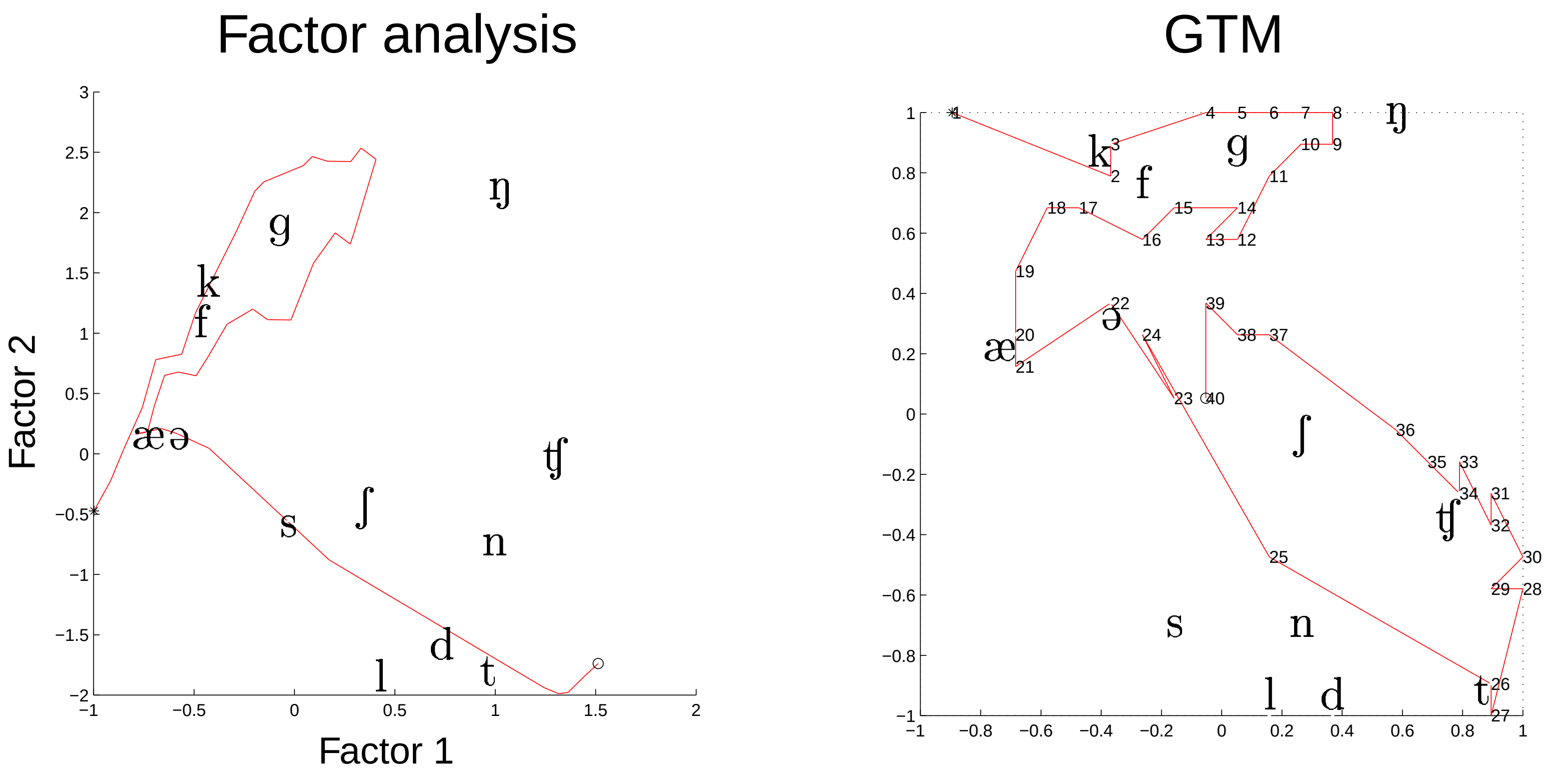
Training set | Test set

Log-likelihood

Squared reconstruction error



Two-dimensional representation (speaker RK)



Trajectory in latent space of the highlighted utterance fragment “I prefer **Kant to** Hobbes for a good bedtime book” (/aɪ prɪ'fɜ 'kænt tə 'hɒbz fər ə 'gʊd 'bedtaɪm 'bʊk/).

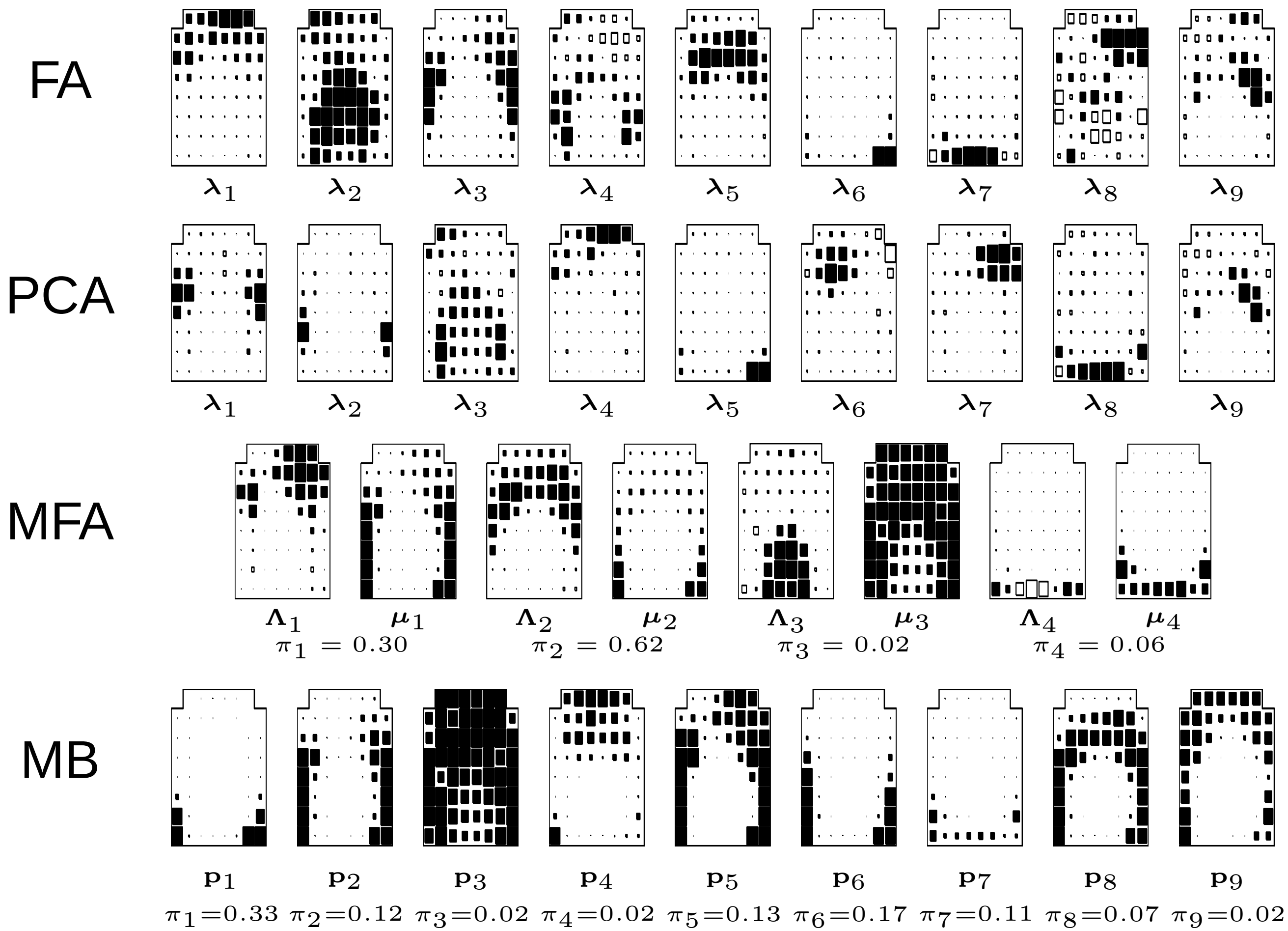
Conclusions

- ▷ Adaptive methods outperform fixed data reduction indices.
- ▷ PCA and FA performed similarly in terms of likelihood.
- ▷ Mixtures of factor analysers and multivariate Bernoulli distributions did not perform well.
- ▷ Two-dimensional GTM outperformed all other methods in terms of likelihood and error reconstruction and reveals nonlinear structure in the data.
- ▷ This suggests a low intrinsic dimensionality for the EPG data.

Additional results available via the web at

`http://www.dcs.shef.ac.uk/~miguel/research/epg.html`

Factors / prototypes (speaker HD)

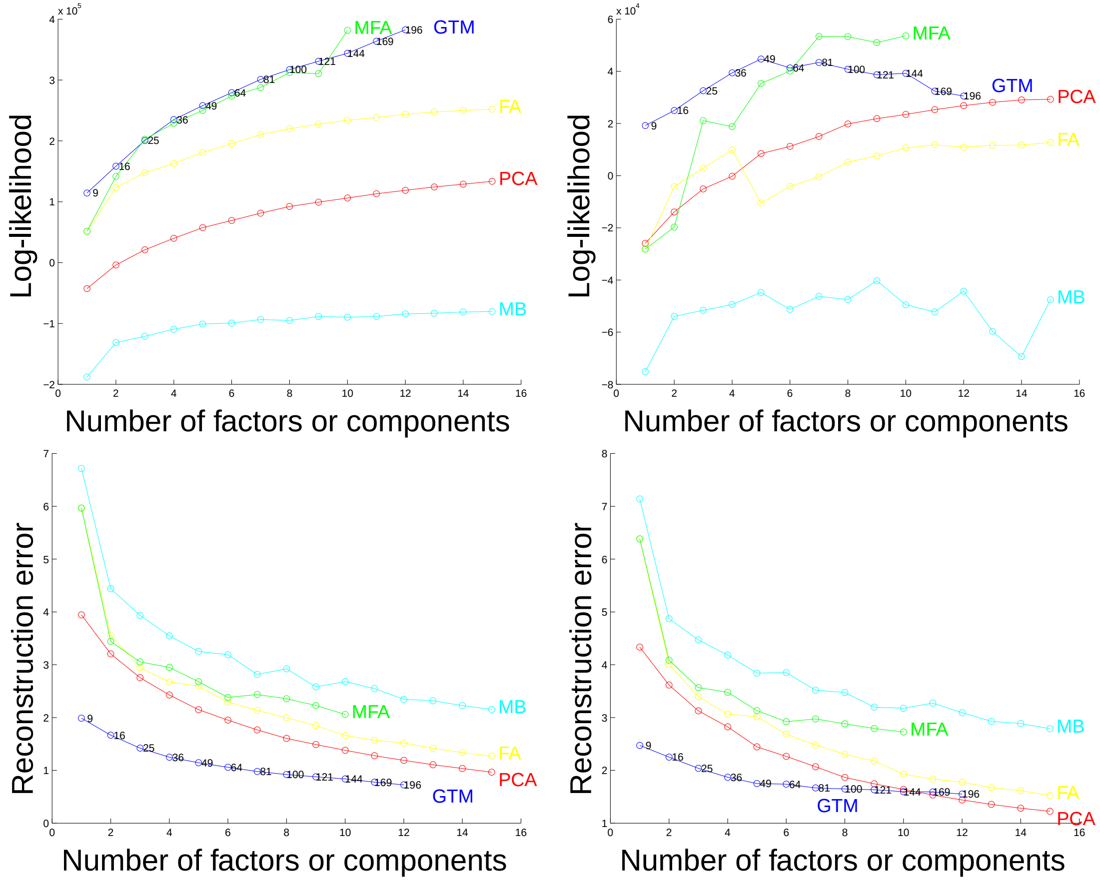


Log-likelihood and reconstruction error (speaker HD)

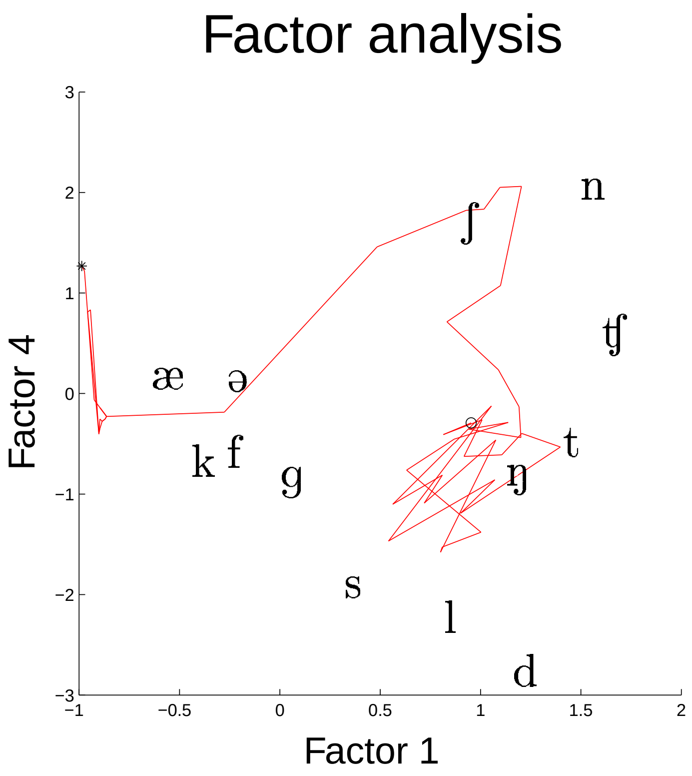
Training set | Test set

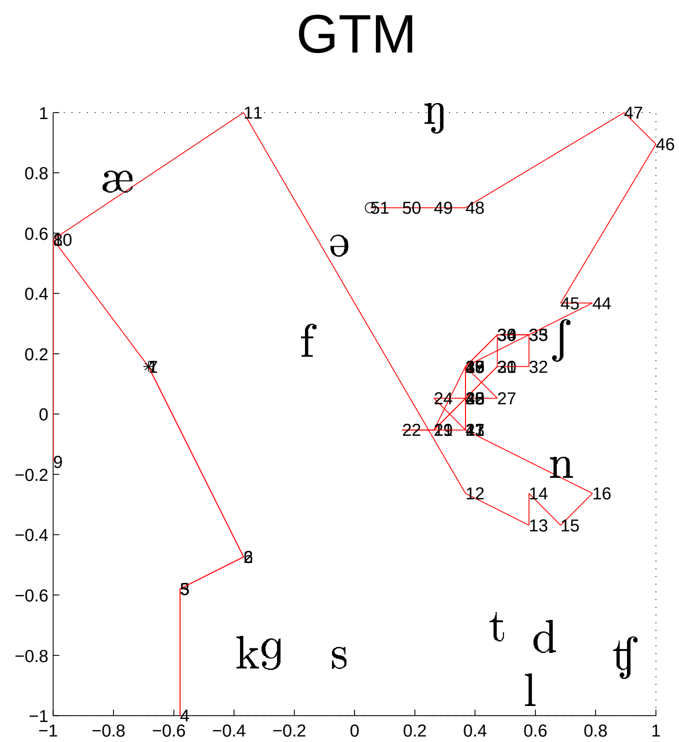
Log-likelihood

Squared reconstruction error



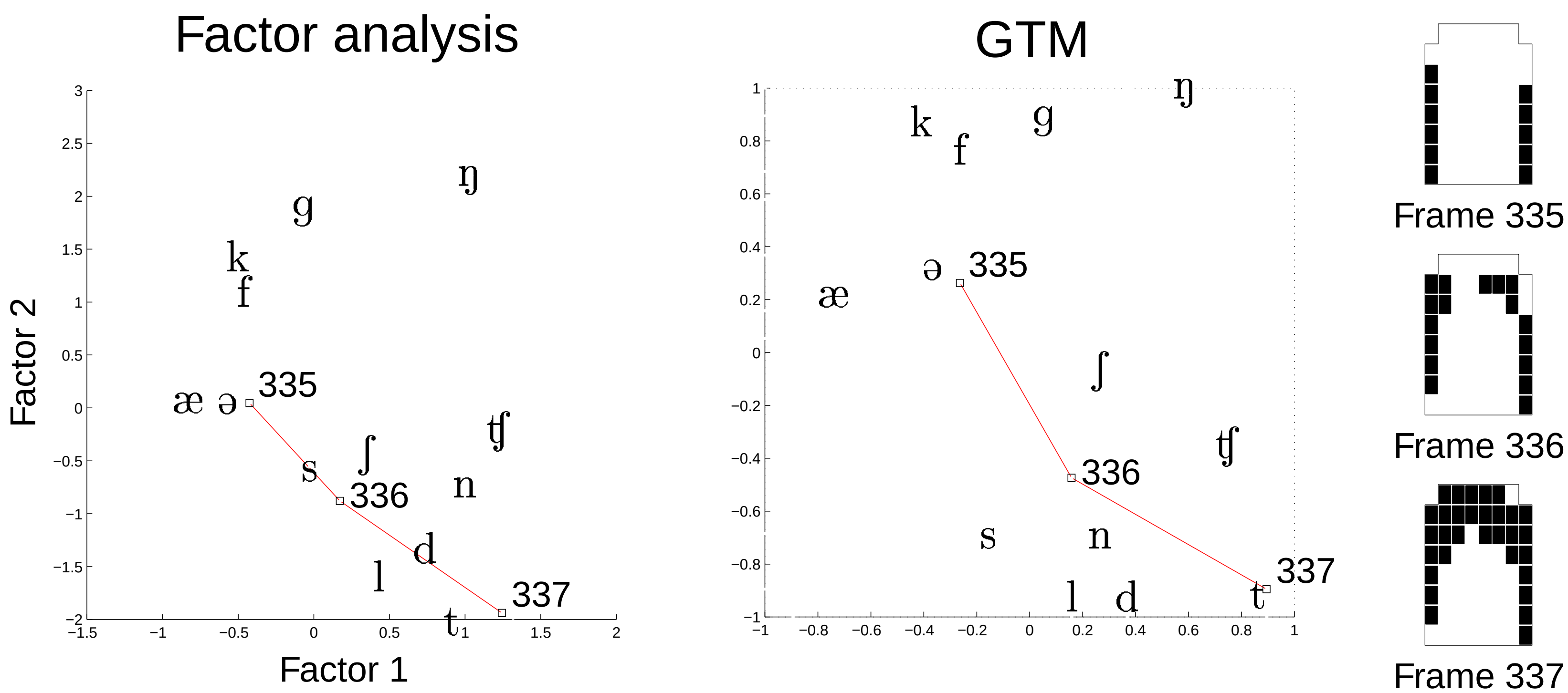
Two-dimensional representation (speaker HD)





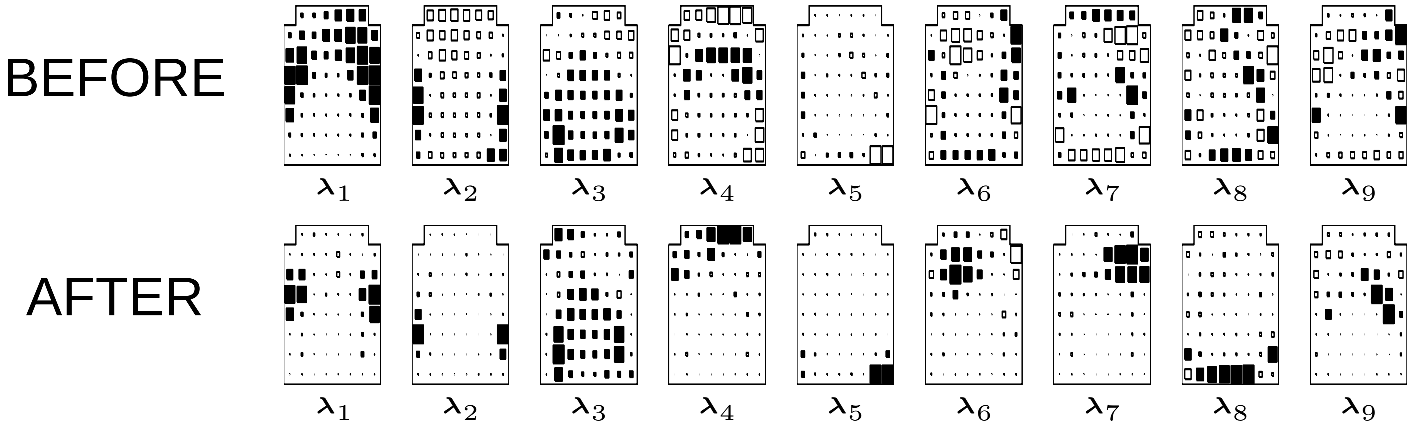
Trajectory in latent space of the highlighted utterance fragment “I prefer **Kant to** Hobbes for a good bedtime book” (/aɪ prɪ’fɜ ’kænt tə ’hɒbz fər ə ’gʊd ’bedtaɪm ’bʊk/).

Discontinuities in latent space (speaker RK)



Selected subsequence of the utterance fragment “I prefer **Kant to** Hobbes for a good bedtime book” (/aɪ prɪ’fɜ ’kænt tə ’hɒbz fər ə ’gʊd ’bedtaɪm ’bʊk/). The abrupt transition from /æ/ to /nt/ (frames 335–336) produces a discontinuity in latent space.

PCA factors before and after varimax rotation (speaker HD)



Both sets of components span the same linear subspace, but the varimax-rotated one is more easily interpretable.