THE HMM ERROR MODEL

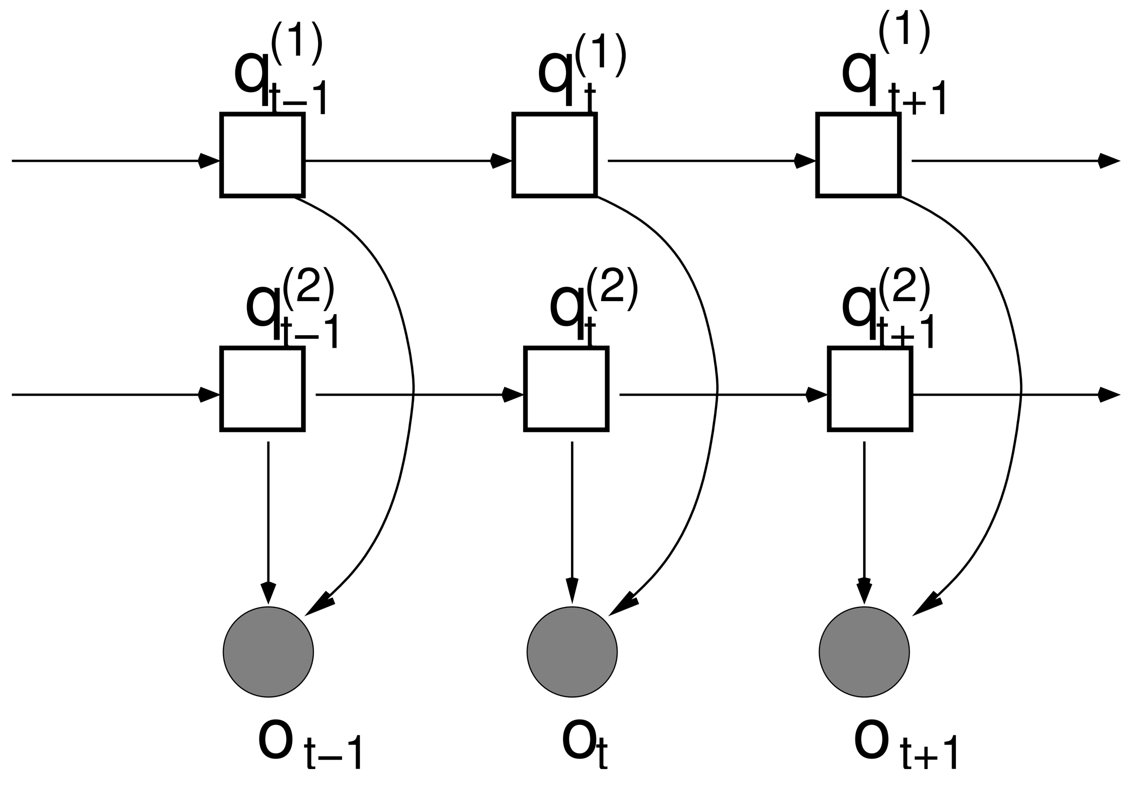
Mark Gales
Cambridge University Engineering Department
mjfg@eng.cam.ac.uk

INTRODUCTION

- A new form of Factorial HMM presented:
  - generalises standard factorial HMMs;
  - introduces transformation streams.
- One example of this form is the HMM Error Model

FACTORIAL HMMS

- A two stream factorial HMM has the form:

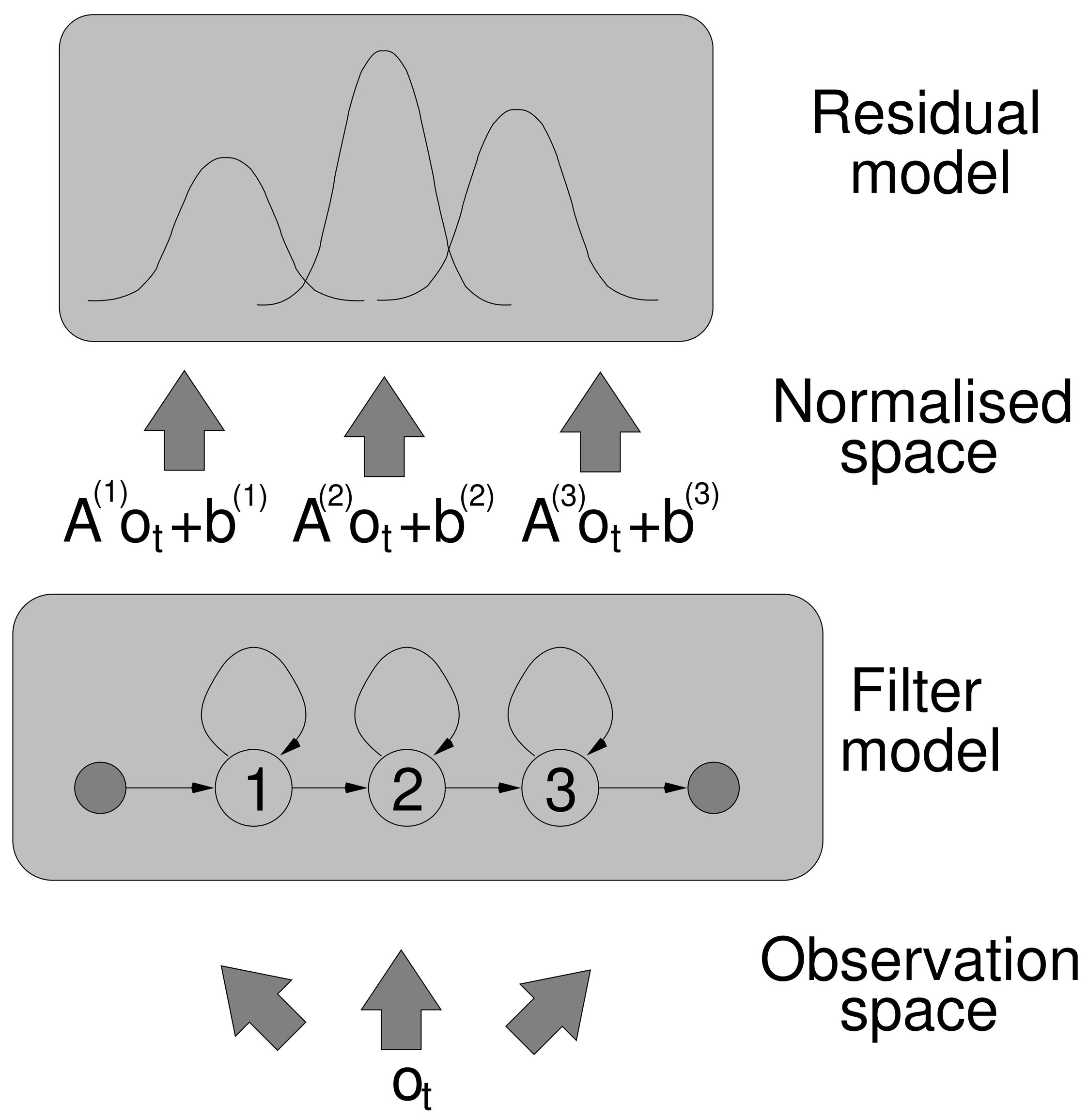


Various forms of interactions between streams possible:

- **Linear stream combination**: the observation is a linear combination of observations from the streams
$$\mathbf{o}_t = \sum_{s=1}^{S} \mathbf{o}_t^{(s)}; \quad p(\mathbf{o}_t|\mathbf{q}_t) = \mathcal{N}(\mathbf{o}_t; \sum_{s=1}^{S} \boldsymbol{\mu}^{(q_t^{(s)})}, \boldsymbol{\Sigma})$$
- **Independent streams**: each stream models an independent subspace of the observation space
$$\mathbf{o}_t = \begin{bmatrix} \mathbf{o}_t^{(1)} \\ \vdots \\ \mathbf{o}_t^{(S)} \end{bmatrix}; \quad p(\mathbf{o}_t|\mathbf{q}_t) = \prod_{s=1}^{S} p(\mathbf{o}_t^{(s)}|q_t^{(s)})$$
- **Loosely coupled streams**: the observations in one stream "influence" those in another. One example, the mixed memory model, uses
$$p(\mathbf{o}_t|\mathbf{q}_t) = \prod_{s=1}^{S} \left(\sum_{u=1}^{S} \lambda_u^{(s)} p(\mathbf{o}_t^{(s)}|q_t^{(u)}) \right)$$
- **Transformation streams**: the features, or model parameters, of one stream are altered by the state of another stream. For 2 streams and a constrained MLLR transform
$$p(\mathbf{o}_t|\mathbf{q}_t) = |\text{det}(\mathbf{A}^{(q_t^{(1)})})| \mathcal{N}(\mathbf{A}^{(q_t^{(1)})}\mathbf{o}_t + \mathbf{b}^{(q_t^{(1)})}; \boldsymbol{\mu}^{(q_t^{(2)})}, \boldsymbol{\Sigma}^{(q_t^{(2)})})$$

THE HMM ERROR MODEL

- A simple implementation of transformation streams



The two streams may be interpreted as:

- **Filter model**: transforms observations to normalised domain with zero mean and:
  - "identity" covariance matrix using *full transforms*
  - "ones" on leading diagonal using *diagonal transforms*

  Mixtures of transforms may be used;
- **Residual model**: models the normalised data from the filter model.

ATTRIBUTES AND TRAINING

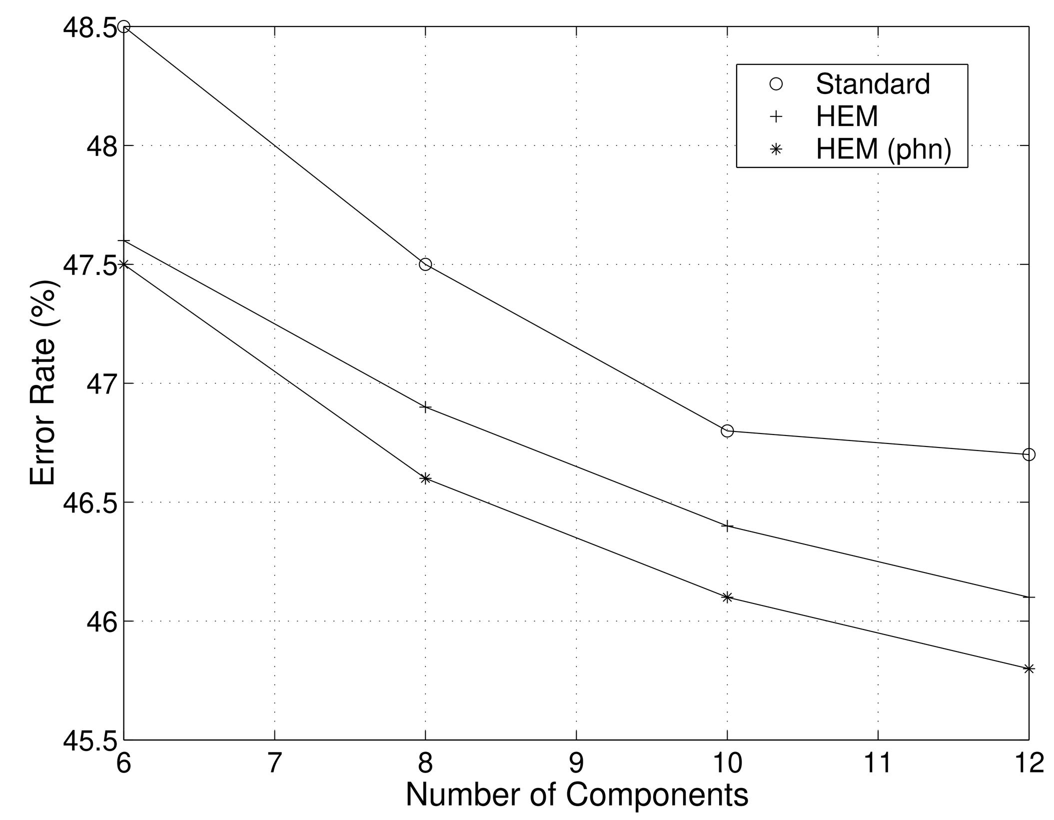
- A standard HMM may be described as
  - single Gaussian component residual model with
  - zero mean, identity covariance matrix;
- More complex residual models (e.g. GMM) can model
  - higher-order moments
  - off-diagonal terms (for diagonal transform case)
- Filter model may be initialised from
$$\mathbf{A}^{(m)} = \mathbf{C}^{(m)-1}; \quad \mathbf{C}^{(m)}\mathbf{C}^{(m)T} = \boldsymbol{\Sigma}^{(m)}$$
$$\mathbf{b}^{(m)} = -\mathbf{A}^{(m)}\boldsymbol{\mu}^{(m)}$$

- Simple iterative training based on Expectation-Maximisation;
- May be viewed as a form of *soft-tying*
- Likelihood calculated as
$$p(\mathbf{o}_t|q_t = q) = \sum_{m=1}^{M^{(q)}} \sum_{k=1}^{K} c^{(m)} c^{(k)} |\text{det}(\mathbf{A}^{(m)})| \mathcal{N}(\mathbf{A}^{(m)}\mathbf{o}_t + \mathbf{b}^{(m)}; \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)})$$

SWITCHBOARD EXPERIMENTS

Standard CU-HTK Switchboard setup:

- 68hour "h5train00sub" training data set, MLE training;
- MF-PLP to give 13 coefficients + delta+delta-delta parameters;
- VTLN + mean and variance side-based normalisation;
- 12 component state-clustered triphone system;
- Trigram language model;
- Global and phone specific residual models;
- Three component GMM residual models used.



Contrast: global semi-tied covariance matrix system gave 46.1%

CONCLUSIONS

- New form of Factorial model described
- Particular version, the HEM, gave significant reductions in word error rate.
- Investigate appropriate complexity of residual and filter models.
- Investigate adaptation/adaptive training of residual model.