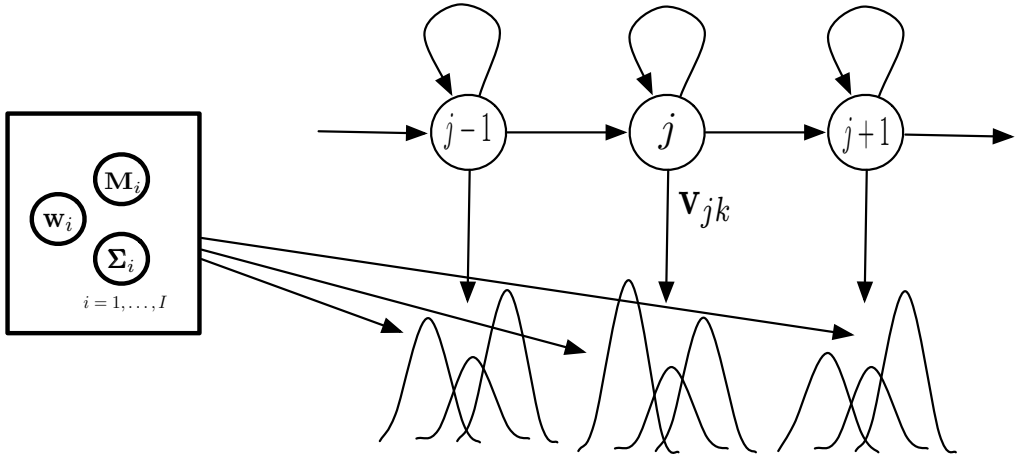
Noise Adaptive Training for Subspace Gaussian Mixture Models

Liang Lu, Arnab Ghoshal, Steve Renals
University of Edinburgh

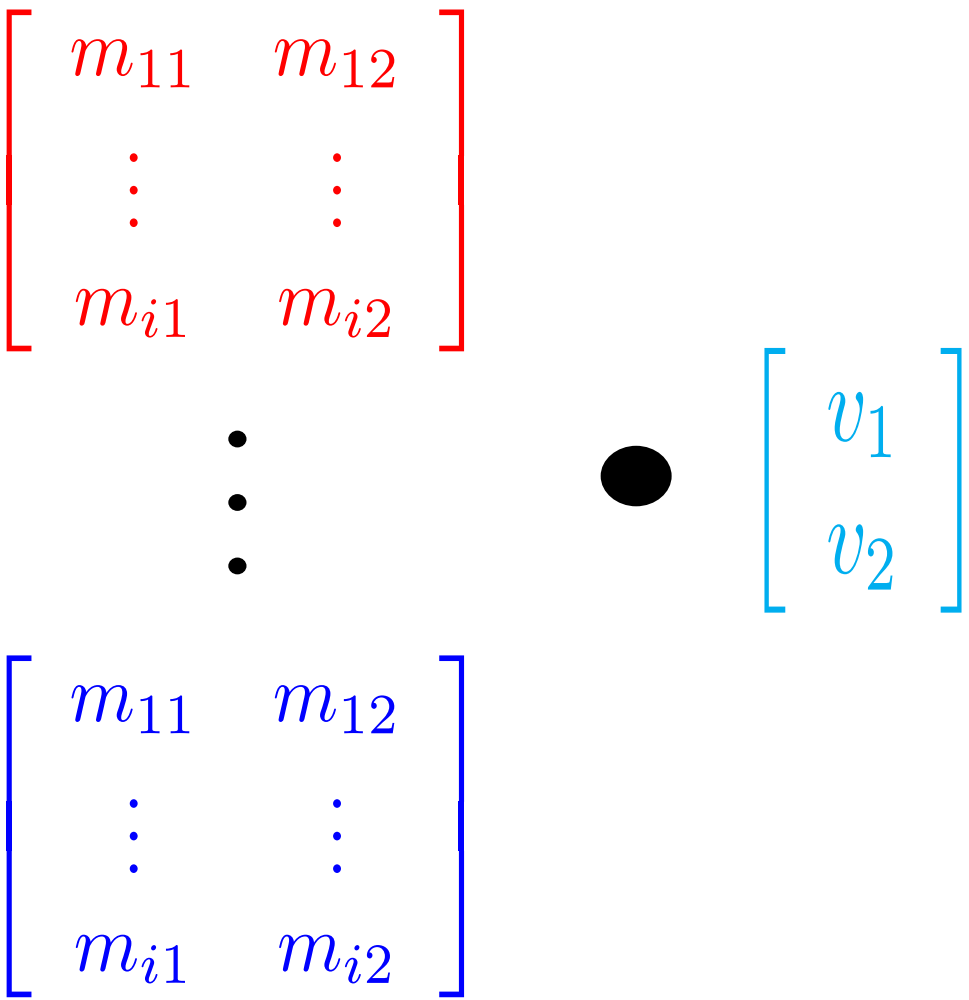
- Introduction
  - Subspace GMM (SGMM) acoustic model
- Noise adaptive training
  - Motivation
  - Adaptive training method
  - Experimental results
- Conclusion

Subspace Gaussian Mixture Models [Povey, et al. 2010]

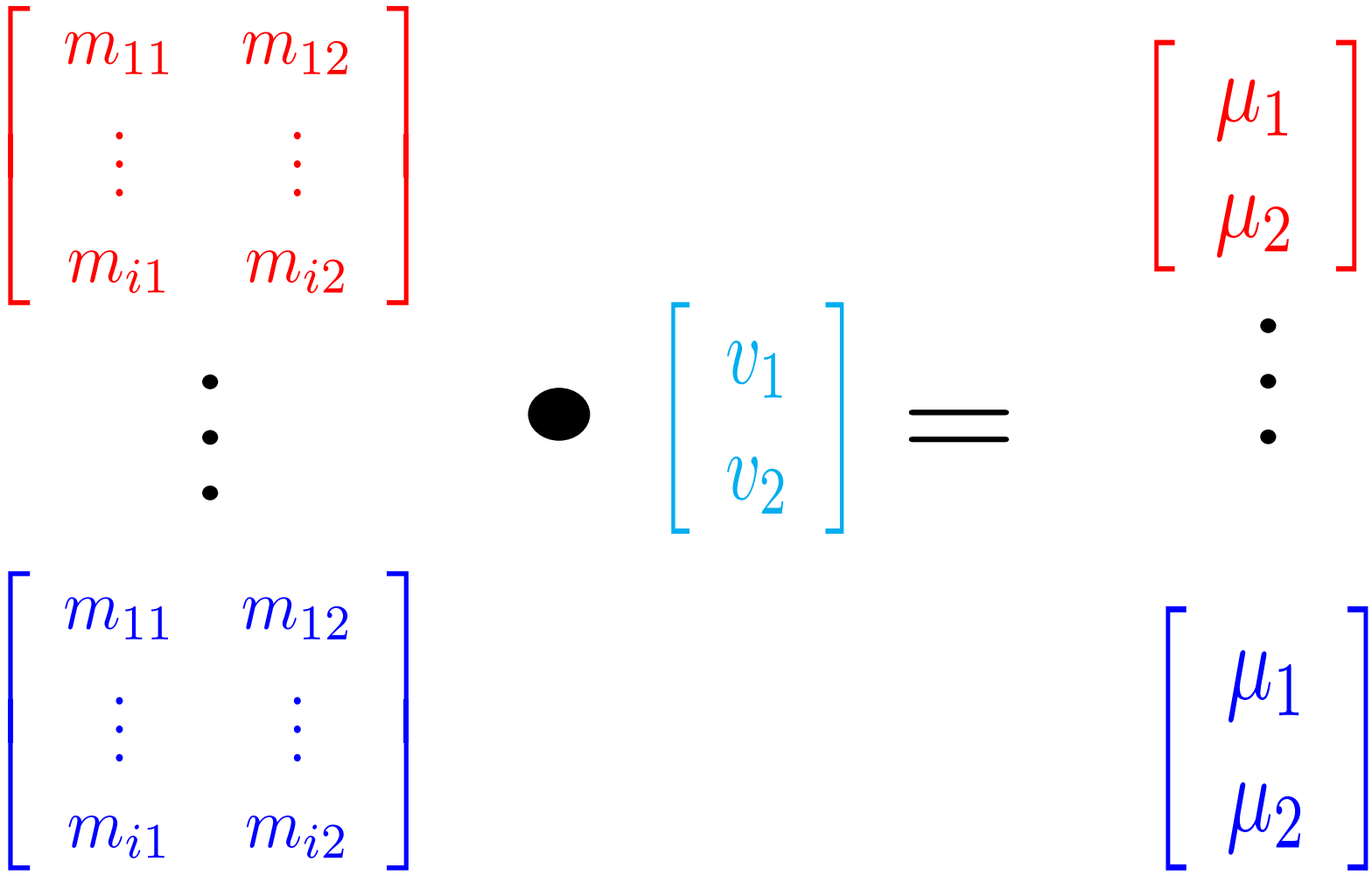


- Globally shared
  - $\mathbf{M}_i$ is the projection matrix for means
  - $\mathbf{w}_i$ is the projection vector for weights
  - $\mathbf{\Sigma}_i$ is the covariance matrix
  - i is the subspace component index
- State-dependent
  - $\mathbf{v}_{jk}$ is low dimensional sub-state vectors (e.g. 40dim)
  - Gaussian mean: $\boldsymbol{\mu}_{jki} = \mathbf{M}_i \mathbf{v}_{jk}$

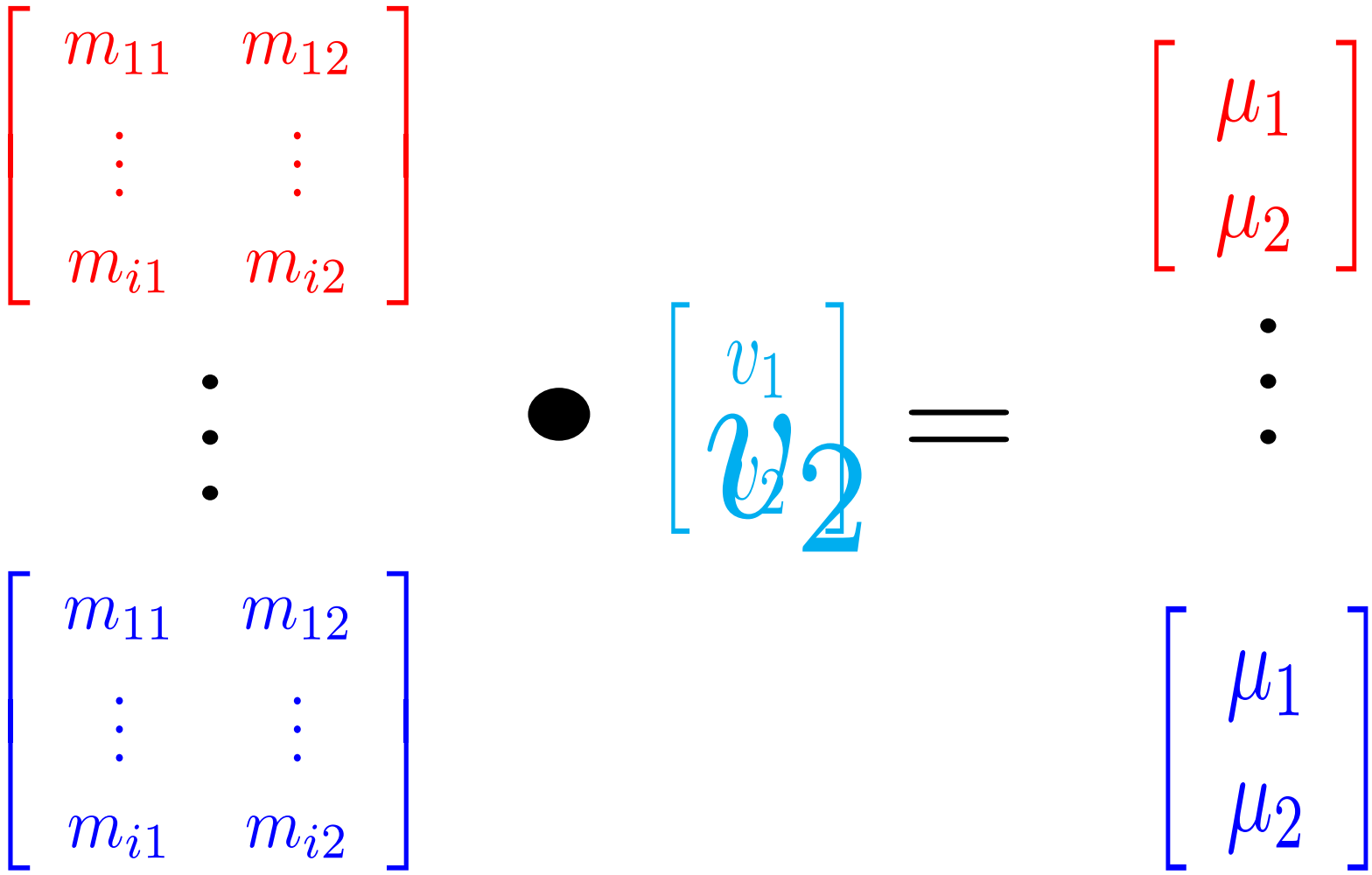
Subspace Gaussian Mixture Models



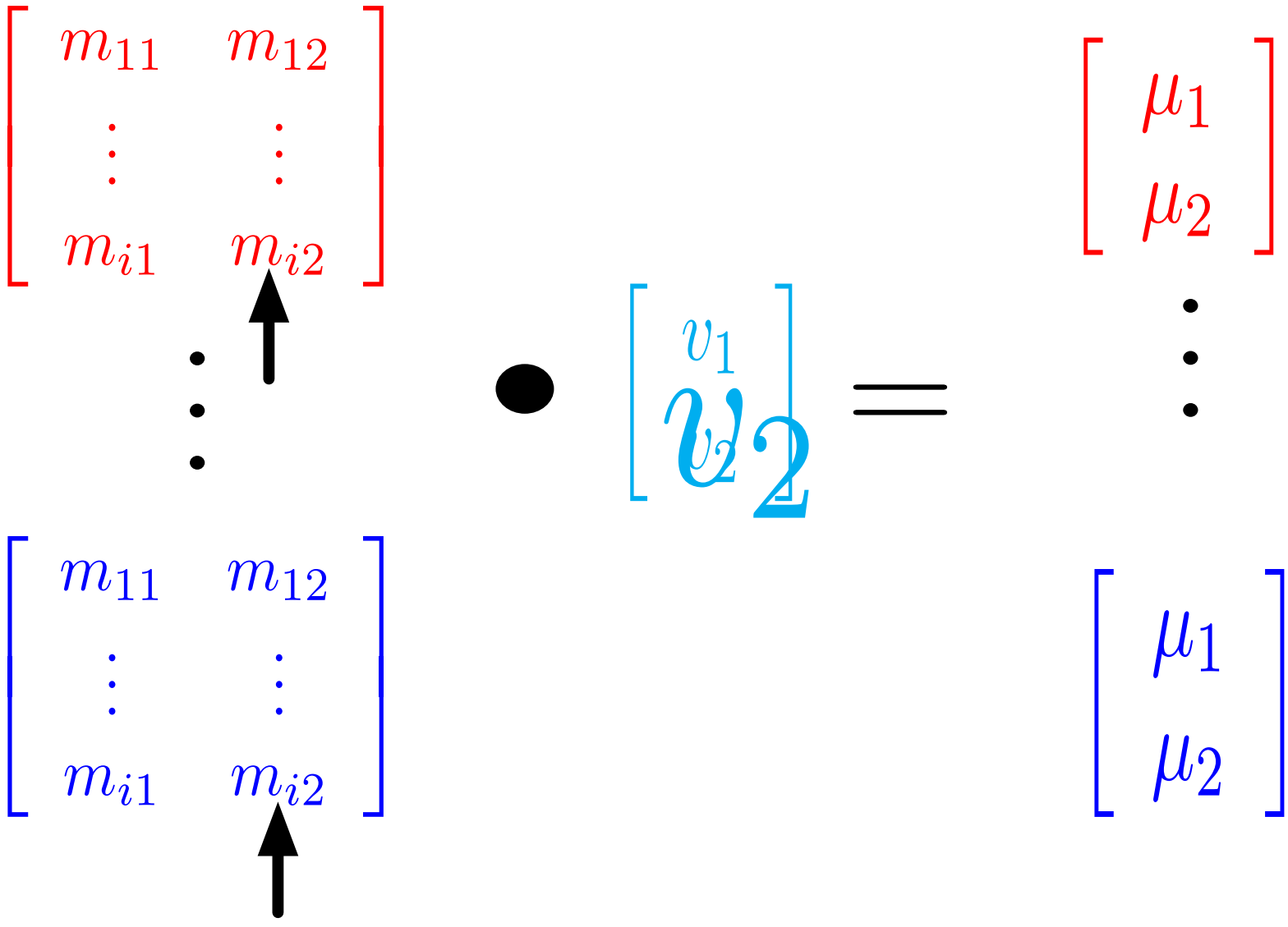
Subspace Gaussian Mixture Models



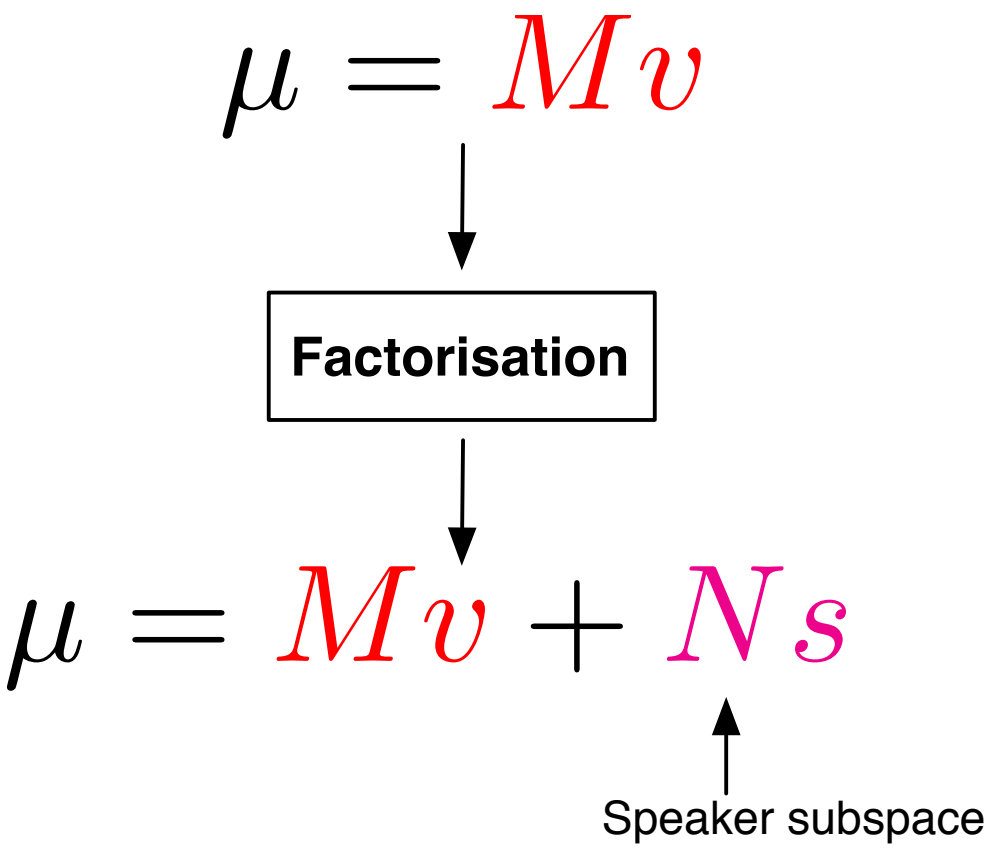
Subspace Gaussian Mixture Models



Subspace Gaussian Mixture Models



Subspace Gaussian Mixture Models

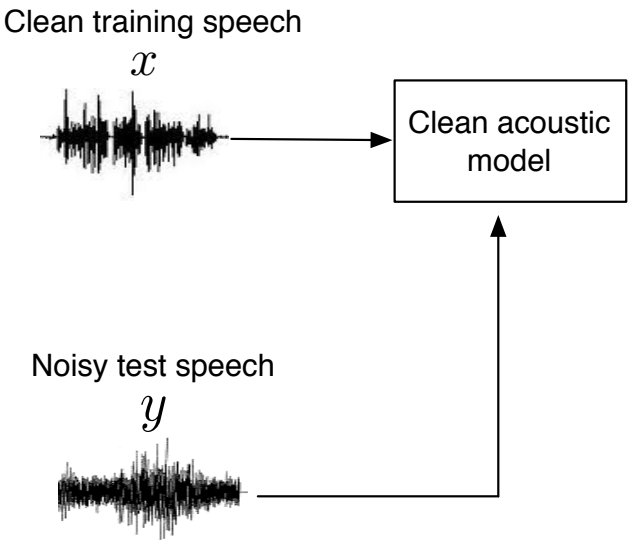


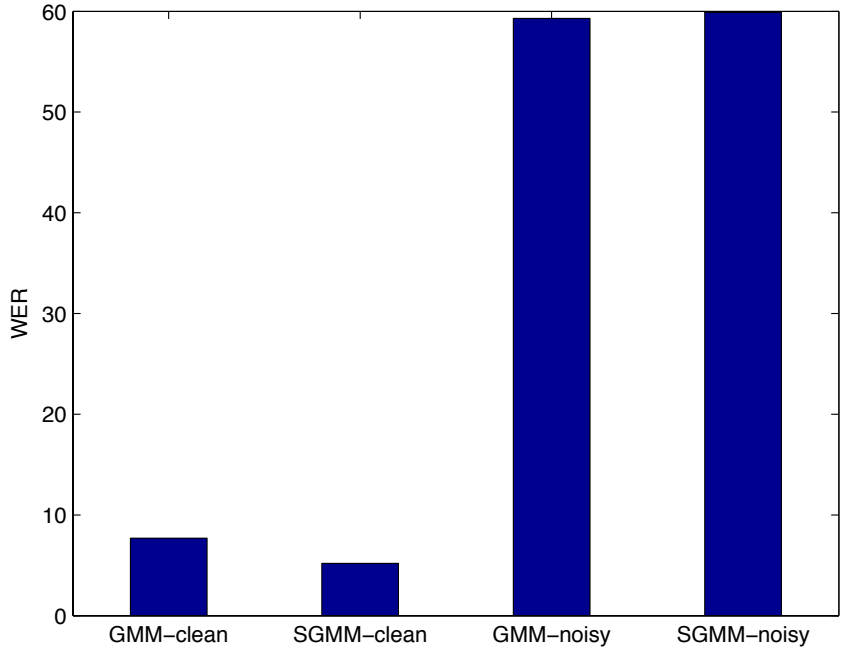
Subspace Gaussian Mixture Models

- Typical features
  - Re-structure the HMM-GMM model paramaters
  - Smaller number of free model parameters
  - Large number of Gaussian components
  - Factorize the phonetic and speaker variability
- Outperforms GMM-based systems on several tasks, e.g. [D. Povey 2010, L. Burget 2010, L. Lu 2011]

Noise robustness

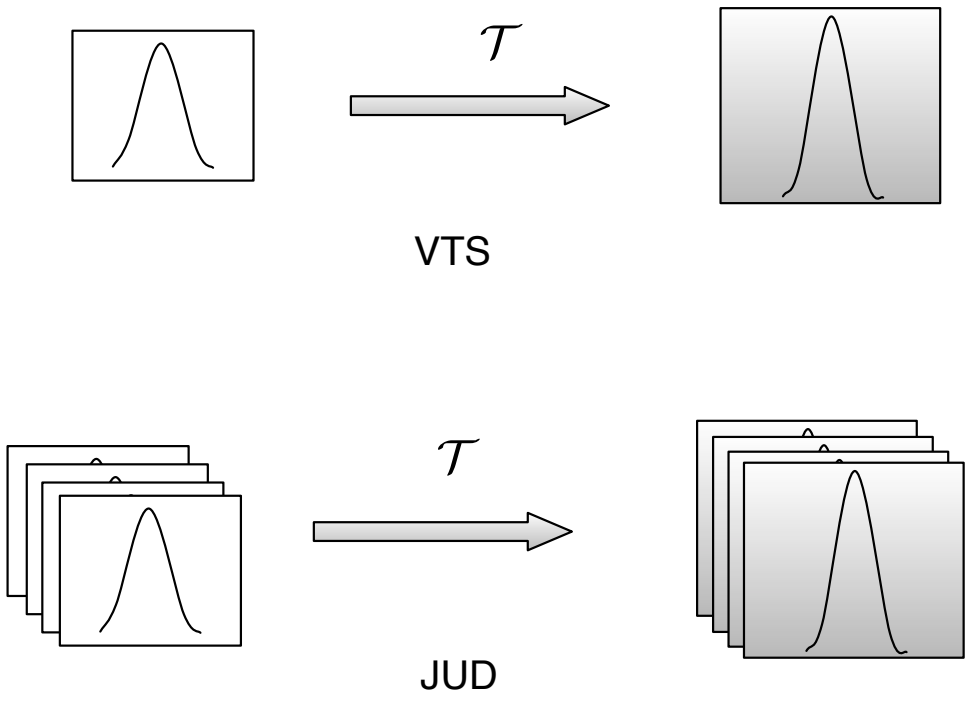
- Aurora 4 dataset
- GMM with 50K components
- SGMM with 6.4M components





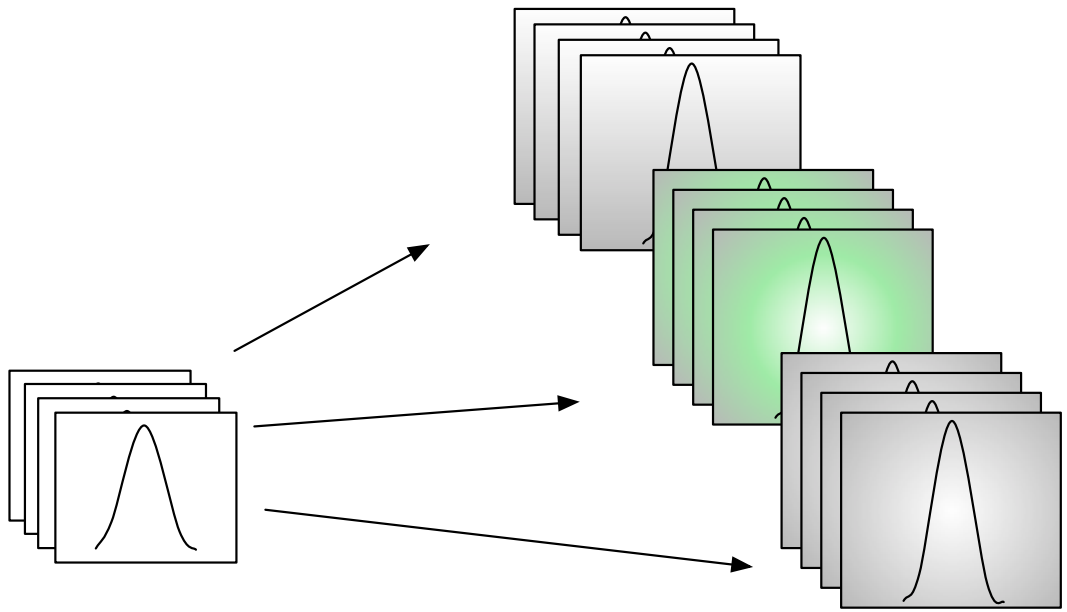
Noise robustness

- SGMM with Joint uncertainty decoding (JUD [H. Liao, 2005])



Noise adaptation of clean speech model

- Adaptation with noise dependent transform for a specific noise condition



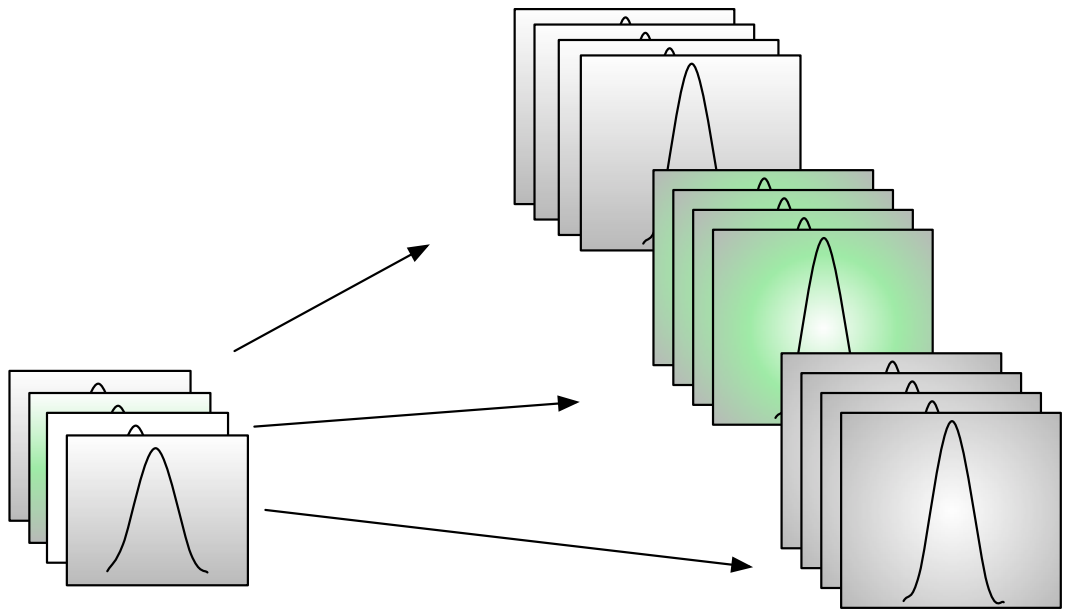
Noise adaptation of clean speech model

- Aurora 4 dataset
- A, B, C and D denote different noise conditions.

Methods	A	B	C	D	Avg
Clean model	5.2	58.2	50.7	72.1	59.9
+JUD	5.1	13.1	12.0	23.2	16.8

Noise adaptation of multi-condition model

- If the training data is from the same types of noise condition



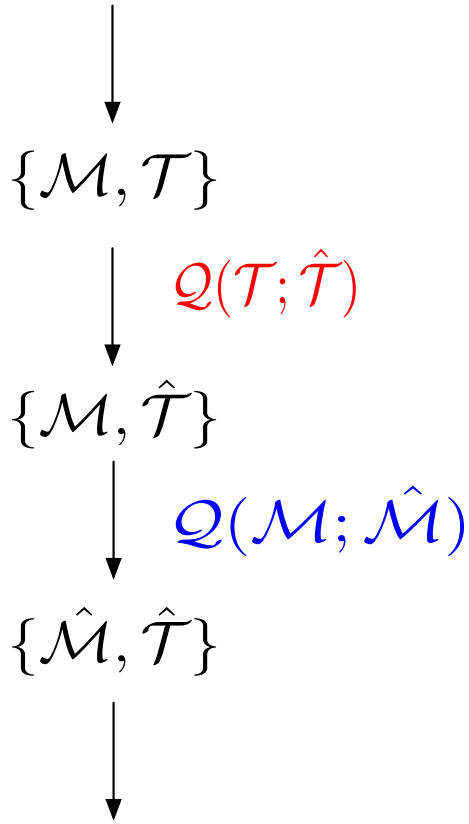
Noise adaptation of multi-condition model

- We obtain better baseline system
- However, we obtain worse results with adaptation

Methods	A	B	C	D	Avg
Clean model	5.2	58.2	50.7	72.1	**59.9**
+JUD	5.1	13.1	12.0	23.2	16.8
MST model	6.8	15.2	18.6	32.3	**22.2**
+JUD	7.4	13.3	14.7	24.1	17.6

Noise adaptive training scheme

- Iterative update of acoustic models $\mathcal{M}$ and noise transforms $\mathcal{T}$
- Optimization of $\mathcal{Q}(\mathcal{T};\hat{\mathcal{T}})$ in [Lu, et al, 2013]
- Optimization of $\mathcal{Q}(\mathcal{M};\hat{\mathcal{M}})$ in this paper



Lu, et al. "Joint Uncertainty Decoding for Noise-robust Subspace Gaussian Mixture Models", IEEE TASLP 2013.

Noise adaptive training - optimization

Optimization of $\mathcal{Q}(\mathcal{M}; \hat{\mathcal{M}})$

- Gradient-based approach: for θ in $\mathcal{M}$ [Liao, et al 2007, Kalinli et al, 2010]

$$\theta = \tilde{\theta} - \zeta \left[\left(\frac{\partial^2 \mathcal{Q}(\cdot)}{\partial^2 \theta} \right)^{-1} \left(\frac{\partial \mathcal{Q}(\cdot)}{\partial \theta} \right) \right]_{\theta=\tilde{\theta}} \quad (1)$$

- EM-based approach, e.g. noisy-CMLLR [Kim, et al 2011]

$$\mathbf{y}_t = \mathbf{H}^{(r)} \mathbf{x}_t + \mathbf{g}^{(r)} + \mathbf{e}_t^{(r)} \longrightarrow P(\mathbf{x}_t | \mathbf{y}_t, r) \quad (2)$$

- We used the EM-based approach for simplicity

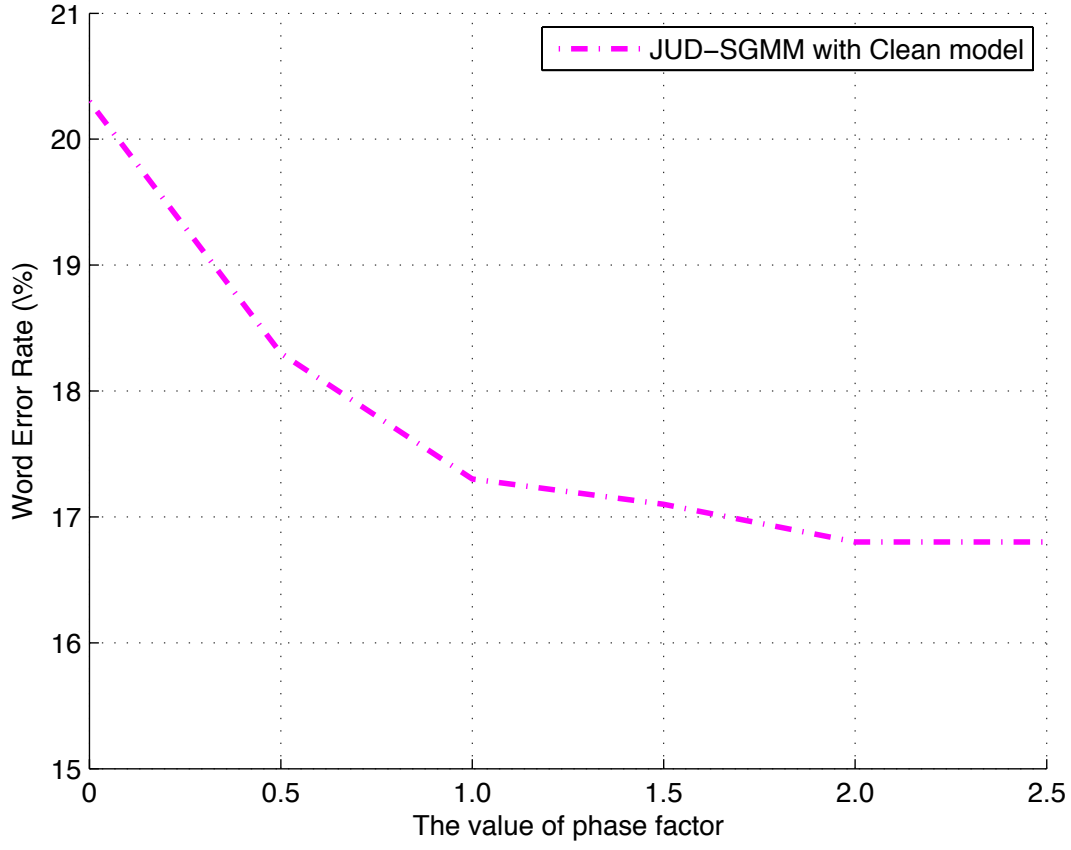
Experiments - noise adaptive training

- With adaptive training, we obtained better results compared to the clean acoustic model

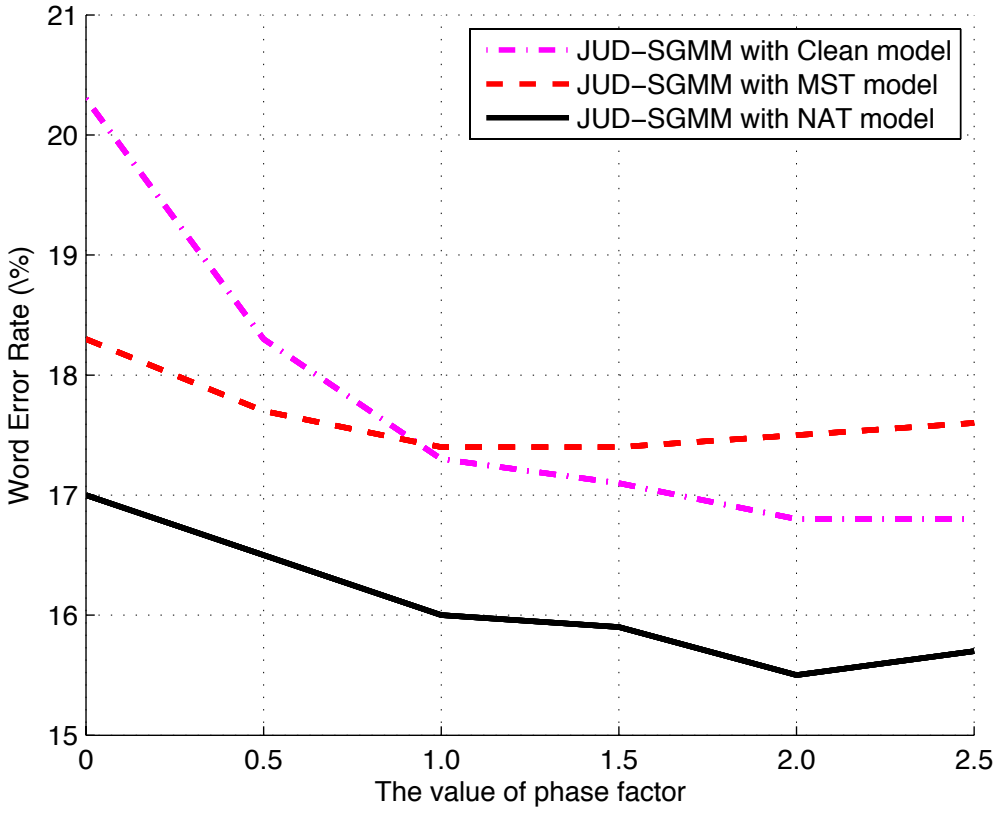
Methods	A	B	C	D	Avg
Clean model	5.2	58.2	50.7	72.1	59.9
+JUD	5.1	13.1	12.0	23.2	16.8
MST model	6.8	15.2	18.6	32.3	22.2
+JUD	7.4	13.3	14.7	24.1	17.6
NAT model	6.5	20.3	19.8	39.7	27.6
+JUD	6.1	11.3	11.9	22.4	15.7

Experiments - noise adaptive training

- Effect of phase factor in the extended mismatch function $\mathbf{y} = f\{\mathbf{x}, \mathbf{n}, \mathbf{h}, \boldsymbol{\alpha}\}$ [Deng, et al, 2004]



Experiments - noise adaptive training



Summary

- Overview of subspace Gaussian mixture models
- Joint uncertainty decoding for noise robustness
- Adaptive training for multi-condition training data
- Experimental results demonstrate the effectiveness of this approach
- To integrate the noise robustness technique with more advanced system

Thanks for your attention!