

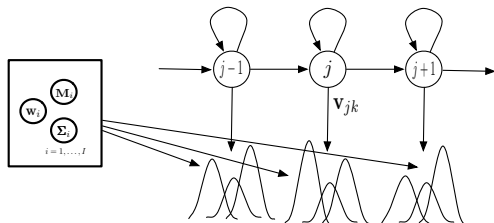
Noise Adaptive Training for Subspace Gaussian Mixture Models

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





► Globally shared

- \mathbf{M}_i is the projection matrix for means
- \mathbf{w}_i is the projection vector for weights
- Σ_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

$$\begin{bmatrix} m_{11} & m_{12} \\ \vdots & \vdots \\ m_{i1} & m_{i2} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

•
•
•

$$\begin{bmatrix} m_{11} & m_{12} \\ \vdots & \vdots \\ m_{i1} & m_{i2} \end{bmatrix}$$

Subspace Gaussian Mixture Models

$$\begin{bmatrix} m_{11} & m_{12} \\ \vdots & \vdots \\ m_{i1} & m_{i2} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \end{bmatrix}$$

The image shows a matrix multiplication equation. The first matrix is red, the second is blue, and the result is red. A second set of the same matrices is shown below in blue. A black dot is placed between the two matrices in the equation.

Subspace Gaussian Mixture Models

$$\begin{bmatrix} m_{11} & m_{12} \\ \vdots & \vdots \\ m_{i1} & m_{i2} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \end{bmatrix}$$

The image shows a matrix multiplication equation. On the left, a red matrix with two columns and multiple rows is multiplied by a blue vector with two elements. The result is a red vector with two elements and a vertical ellipsis below it. A second, identical equation is shown below in blue, with the matrix and vector elements in blue and the result vector elements in blue.

Subspace Gaussian Mixture Models

$$\begin{bmatrix} m_{11} & m_{12} \\ \vdots & \vdots \\ m_{i1} & m_{i2} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \end{bmatrix}$$

The diagram illustrates the relationship between a matrix of means, a vector of subspace basis vectors, and a vector of means. The top row of the matrix and the resulting vector are shown in red, while the bottom row and the basis vector are shown in blue. Arrows point to the m_{i1} and m_{i2} elements in both the red and blue matrices, indicating a specific data point's projection onto the subspace.

$$\mu = Mv$$

↓

Factorisation

↓

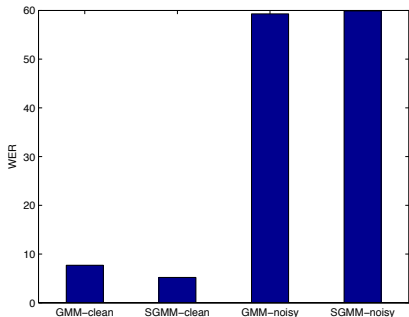
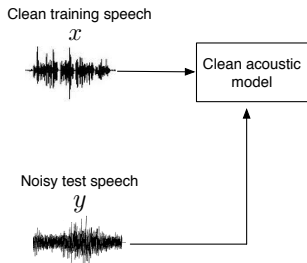
$$\mu = Mv + Ns$$

↑
Speaker subspace

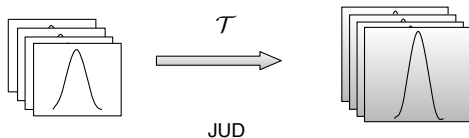
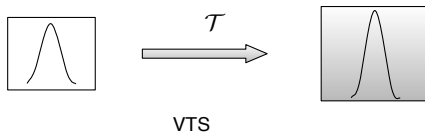
- ▶ Typical features
 - ▶ Re-structure the HMM-GMM model parameters
 - ▶ Smaller number of free model parameters
 - ▶ Large number of Gaussian components
 - ▶ Factorize the phonetic and speaker variability
- ▶ Outperforms GMM-based systems on several tasks, e.g. [D. Povey 2010, L. Burget 2010, L. Lu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



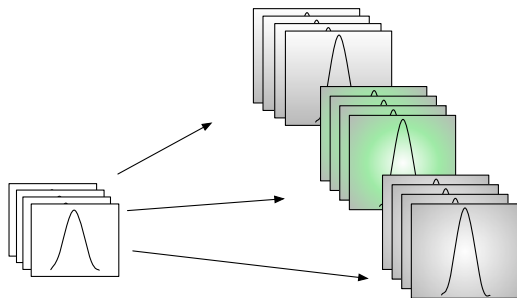
- ▶ Aurora 4 dataset
- ▶ GMM with 50K components
- ▶ SGMM with 6.4M components



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



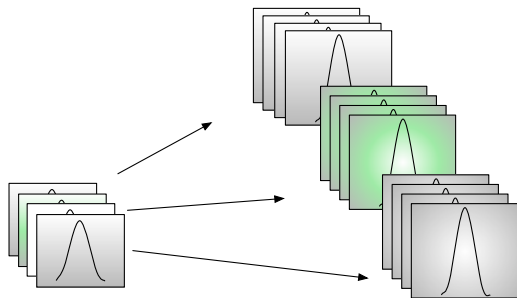
- ▶ Adaptation with noise dependent transform for a specific noise condition



- ▶ 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

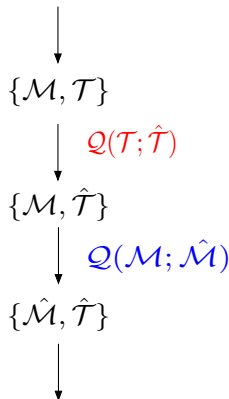
- ▶ If the training data is from the same types of noise condition



- ▶ 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

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



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

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

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

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

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

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

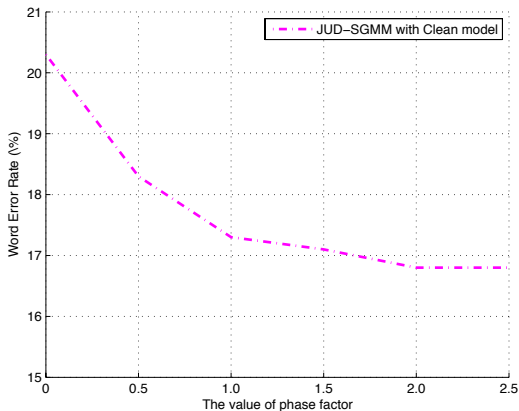
- ▶ We used the EM-based approach for simplicity



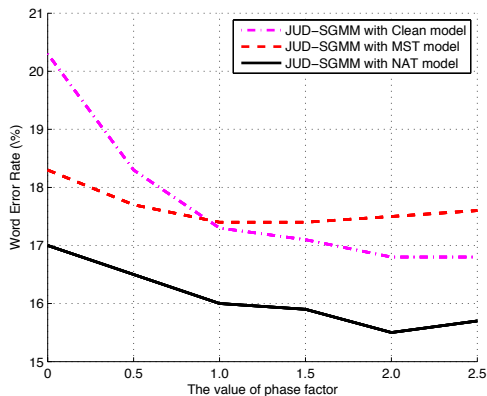
- ▶ With adaptive training, we obtained better results compared to the clean acoustic model

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

- ▶ 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



- ▶ 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!

