Acoustic Factor Analysis

Kaer

Outlines

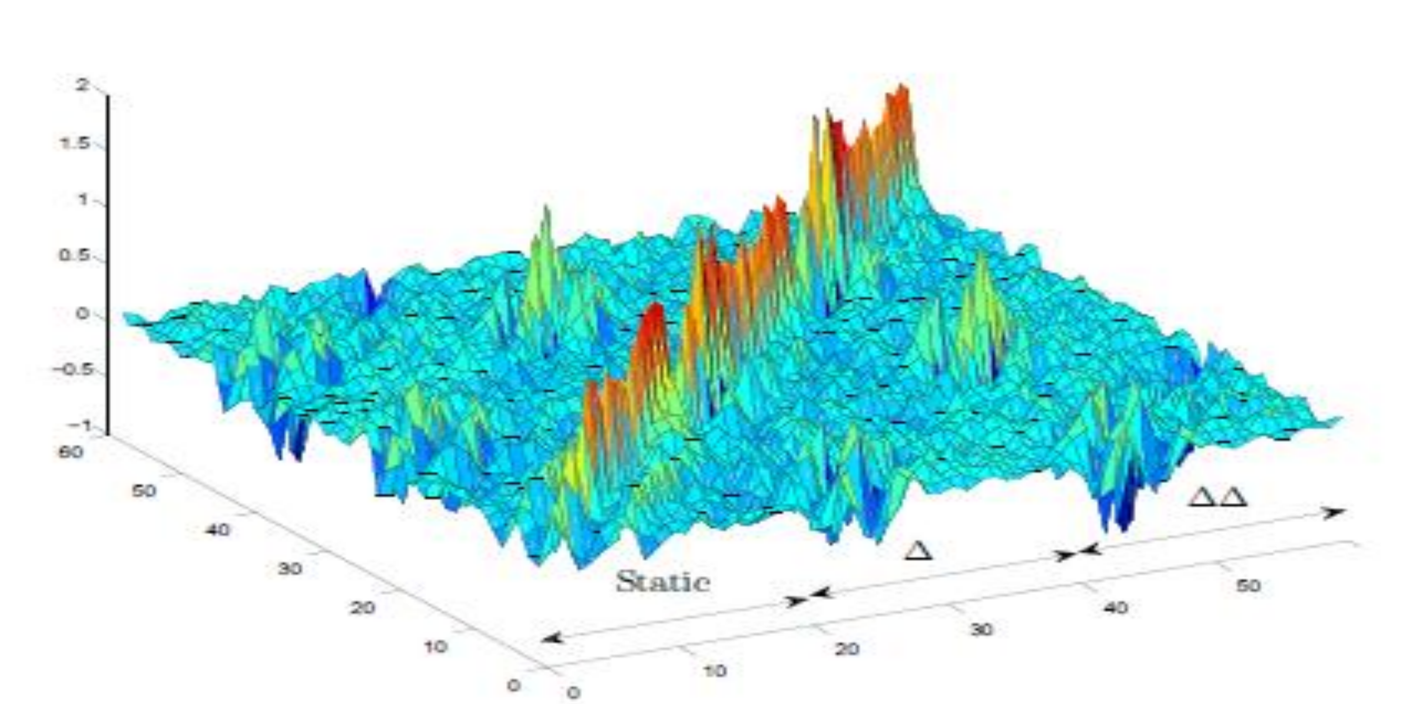
1. Introduction
2. ACOUSTIC FACTOR ANALYSIS
3. Isotropic Residual Noise Model
4. Diagonal Covariance Residual Noise Model
5. *Experimental results*
6. references

1. Introduction

- There is a popular belief that cepstral feature components are uncorrelated and thus can be modeled by a diagonal covariance GMM model.
- Recently, full covariance GMM models have shown to provide advantage in speaker recognition system performance

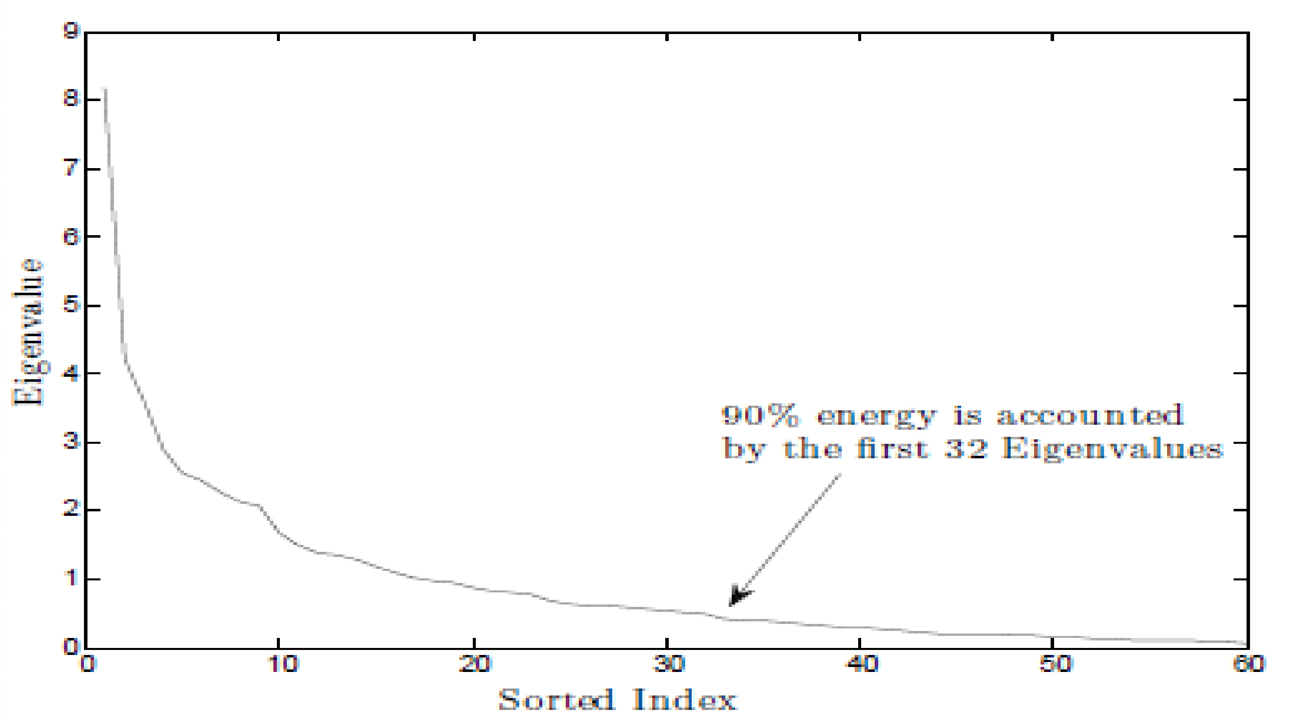
1. Introduction

- The feature coefficients are not fully uncorrelated.



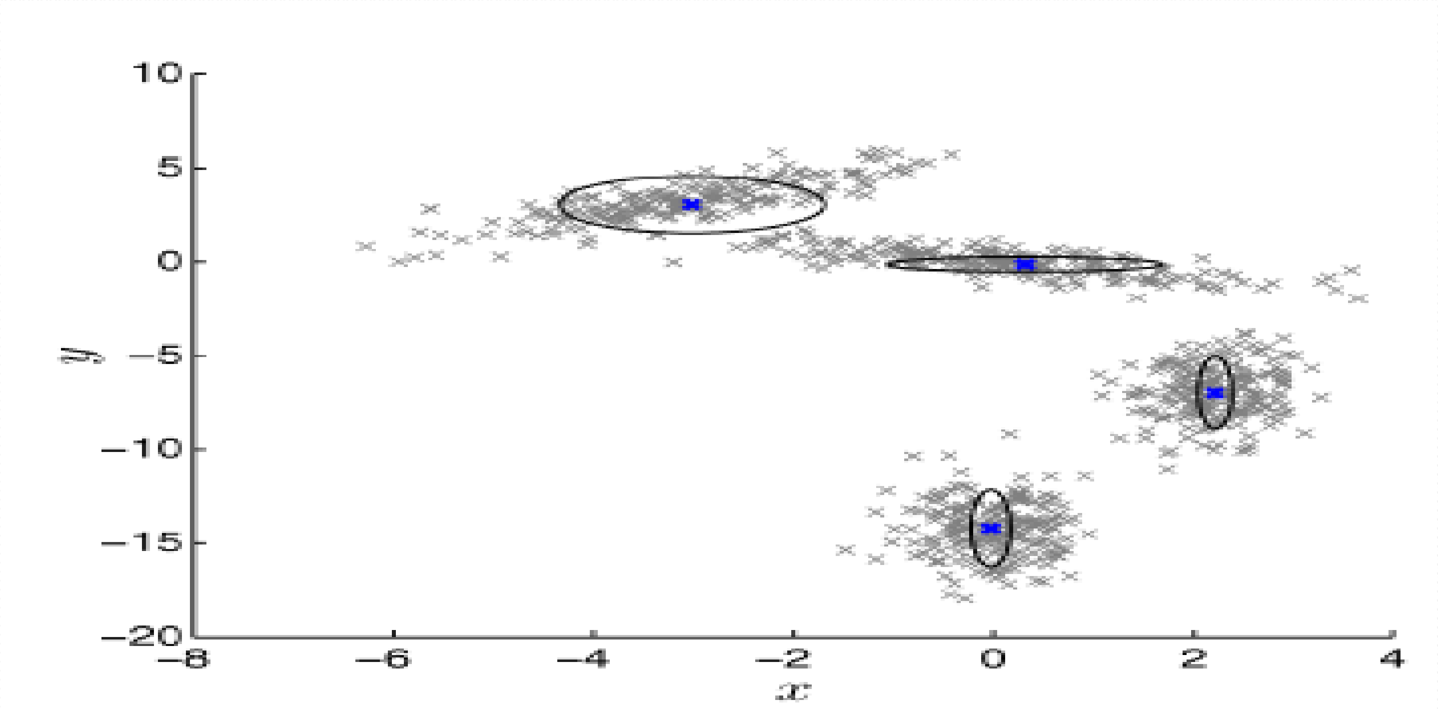
1. Introduction

- Sorted Eigenvalues demonstrating that most of the energy is accounted for by in the first few dimensions



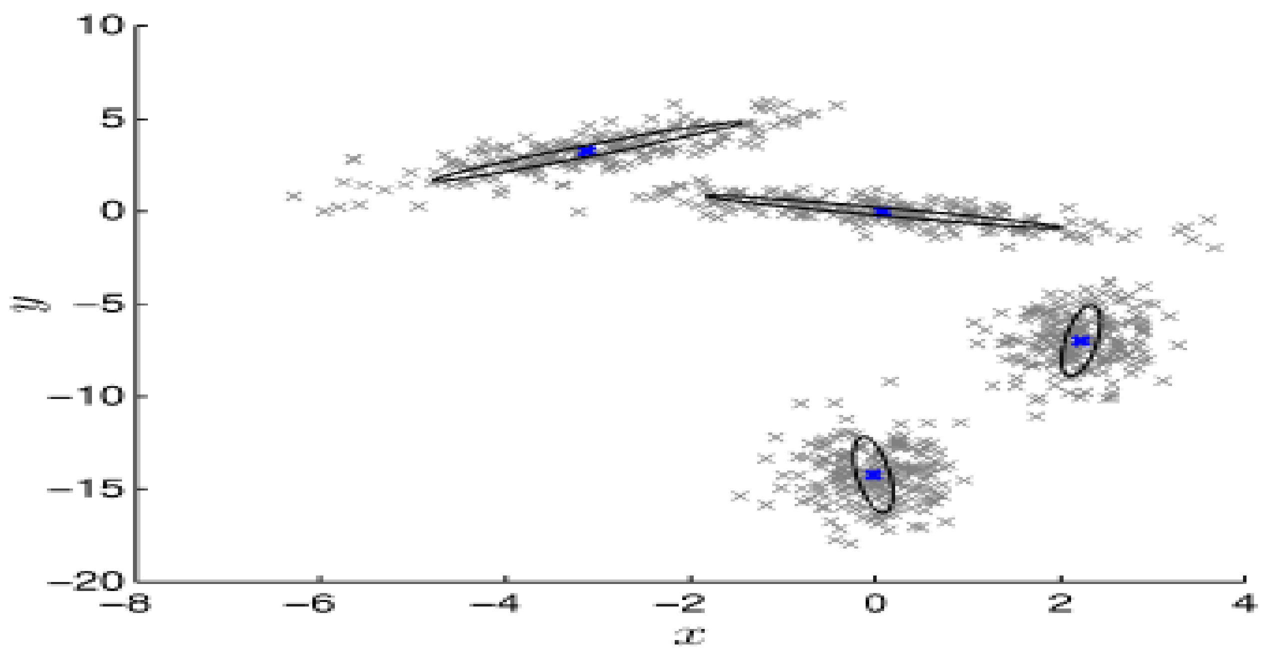
1. Introduction

- diagonal-covariance model is insensitive to the dominant direction of the data



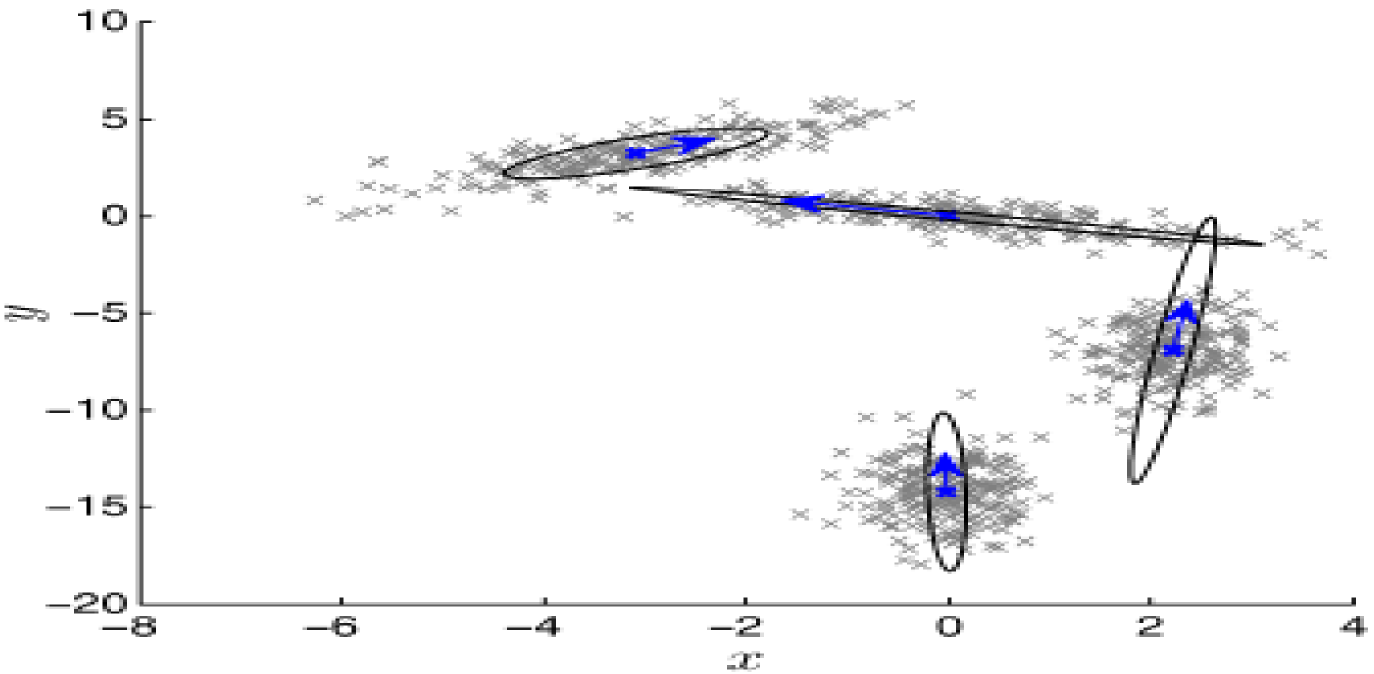
1. Introduction

- full covariance model take the dominant direction into account.



1. Introduction

- Consider the dominant direction(q=1) of the data and considers the other direction as noise.



1. Introduction

- Speaker dependent information resides within the first few dominant directions in the feature space.
- Acoustic Factor Analysis (AFA) only consider the dominant directions of the feature space, providing more robustness to the noisy test data.

2. ACOUSTIC FACTOR ANALYSIS

- Formulation

the d dimensional feature vector, x , can be represented by

$$x = Wy + \mu + \epsilon$$

- W: $d \times q$ factor loading matrix, $q < d$
- μ: $d \times 1$ mean vector
- $\epsilon \sim N(0, \Psi)$ noise component
- $y \sim N(0, I)$ acoustic factor
- $x \sim N(\mu, \Psi + WW^T)$

2. ACOUSTIC FACTOR ANALYSIS

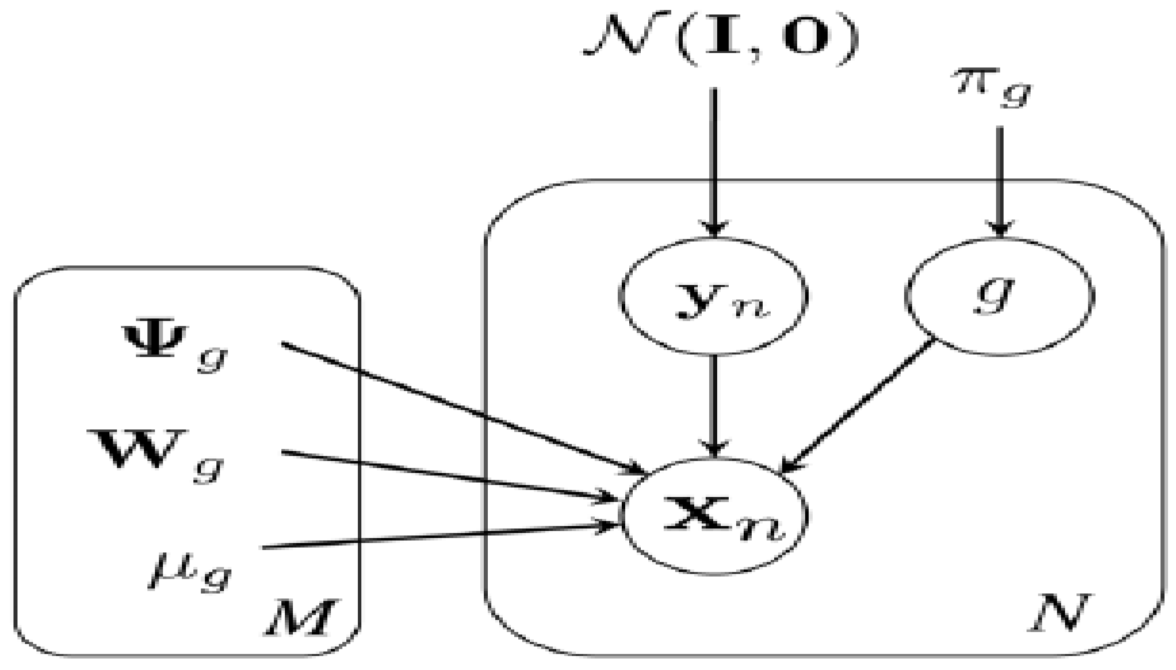
- A mixture of AFA models

$$p(\mathbf{x}_n) = \sum_{g=1}^{M} \pi_g p(\mathbf{x}_n | g),$$

  - π_g: g-th mixture weight
  - M: mixture number
  - $p(x_n|g) \sim N(\mu_g, C_g) \quad C_g = \Psi_g + W_g W_g^T$

2. ACOUSTIC FACTOR ANALYSIS

- Probabilistic graphical model of a Mixture of Factor Analyzer (MFA) model



2. ACOUSTIC FACTOR ANALYSIS

- Feature dimensionality reduction

① Train a full covariance UBM model λ_0

$$p(\mathbf{x}|\lambda_0) = \sum_{i=1}^{M} w_i \mathcal{N}(\mu_i, \Sigma_i)$$

② Estimate noise variance for i-th mixture

$$\sigma_i^2 = \frac{1}{d-q} \sum_{j=q+1}^{d} \lambda_j$$

2. ACOUSTIC FACTOR ANALYSIS

- Feature dimensionality reduction

  ③ Compute maximum likelihood estimation of the factor loading matrix W_i of the i-th mixture

$$\mathbf{W}_i = \mathbf{U}_q^{(i)}(\Lambda_q^{(i)} - \sigma_i^2 \mathbf{I})^{1/2}\mathbf{R}$$

  - $U_q^{(i)}$: $d \times q$ matrix whose columns are the q leading eigenvectors of Σ_i
  - $\Lambda_q^{(i)}$: diagonal matrix, contains the corresponding q eigenvalues
  - R: q$\times$ q arbitrary orthogonal rotation matrix

2. ACOUSTIC FACTOR ANALYSIS

- Feature dimensionality reduction

④ Compute latent factors, the reduced version of x_n can be obtained from the posterior mean of y

$$E\{\mathrm{y}|\mathrm{x}_n, i\} = \langle \mathrm{y}_n^{(i)} \rangle = \mathrm{M}_i^{-1}\mathrm{W}_i^T(\mathrm{x}_n - \mu_i)$$

$$\mathrm{M}_i = (\sigma_i^2 \mathrm{I} + \mathrm{W}_i^T \mathrm{W}_i)$$

2. ACOUSTIC FACTOR ANALYSIS

- Integration within the i-vector system
  - Conventionally, the i-Vectors are extracted using the zero and first order statistics, The zero order statistics for utterance s are extracted as:

$$N_s(g) = \sum_{n \in s} \gamma_n(g), \qquad \gamma_n(g) = p(g|\mathbf{x}_n, \Lambda) = \frac{p(\mathbf{x}_n|g, \Lambda)\pi_g}{p(\mathbf{x}_n|\Lambda)}$$

  - The first order statistics are extracted as

$$\mathbf{F}_s(g) = \sum_{n \in s} \gamma_n(g)\mathbf{x}_n.$$

2. ACOUSTIC FACTOR ANALYSIS

- For the AFA model, the first order statistics are extracted as follows:

$$\hat{\mathbf{F}}_s(g) = \sum_{n \in s} \gamma_n(g) \mathbf{A}_g^T (\mathbf{x}_n - \mu_g)$$

$$= \mathbf{A}_g^T \left[\mathbf{F}_s(g) - N_s(g)\mu_g \right] = \mathbf{A}_g^T \bar{\mathbf{F}}_s(g),$$

3. Isotropic Residual Noise Model

- the noise covariance matrix is $\Psi_g = \sigma_g^2 I$
- In this model, assume Ψ_g in each mixture is isotropic
- The first order statistics

$$\hat{\mathbf{F}}_s(g) = \sum_{n \in s} \gamma_n(g) \mathbf{A}_g^T (\mathbf{x}_n - \mu_g)$$
$$= \mathbf{A}_g^T [\mathbf{F}_s(g) - N_s(g)\mu_g] = \mathbf{A}_g^T \bar{\mathbf{F}}_s(g),$$

where $A_g^T = M_g^{-1} W_g^T$

4. Diagonal Covariance Residual Noise Model

- assume Ψ_g in each mixture is diagonal
- The first order statistics

$$\hat{\mathbf{F}}_s(g) = \sum_{n \in s} \gamma_n(g) \mathbf{A}_g^T (\mathbf{x}_n - \mu_g)$$
$$= \mathbf{A}_g^T [\mathbf{F}_s(g) - N_s(g)\mu_g] = \mathbf{A}_g^T \bar{\mathbf{F}}_s(g),$$

where $A_g^T = M_g^{-1} W_g^T \Psi_g^{-1}$

5. Experimental results

- *Effect of the Modeling Method*

Model	Method	%EER					$\min C_{\text{primary}}$				
		cc-1	cc-2	cc-3	cc-4	cc-5	cc-1	cc-2	cc-3	cc-4	cc-5
GMM-diag	GMM	3.243	2.819	3.127	3.113	3.228	0.264	0.312	0.130	0.271	0.307
GMM-full	GMM	3.302	3.714	3.328	3.770	4.142	0.273	0.378	0.137	0.318	0.354
$\text{ML-AFA}_{\text{iso}}(q=42)$	GMM	3.375	3.924	3.261	3.948	4.348	0.270	0.392	0.130	0.327	0.381
$\text{ML-AFA}_{\text{diag}}(q=42)$	GMM	3.522	3.717	3.213	3.946	4.143	0.271	0.390	0.125	0.317	0.368
$\text{ML-AFA}_{\text{iso}}(q=42)$	AFA	3.298	2.642	3.118	3.007	3.080	0.245	0.304	0.123	0.260	0.294
$\text{ML-AFA}_{\text{diag}}(q=42)$	AFA	2.993	2.655	3.242	2.928	3.027	0.221	0.291	0.107	0.257	0.282

5. Experimental results

- *Variation of Acoustic Factor Dimension*

UBM model		%EER	minC_{primary}	C_{primary}
GMM-diag		3.2428	0.2642	0.3385
GMM-full		3.3020	0.2729	0.3553
Method	*q*	*Absolute/%relative performance*		
ML-AFA$_{\text{iso}}$	42	3.298/-1.7	0.245/7.3	0.336/0.7
	48	**2.779/14.3**	0.241/8.6	0.334/1.4
	54	2.874/11.4	0.236/10.8	0.326/3.8
ML-AFA$_{\text{diag}}$	42	2.993/7.7	**0.221/16.5**	**0.316/6.6**
	48	3.008/7.2	0.242/8.3	0.339/-0.2
	54	2.951/9.0	0.237/10.4	0.331/2.2

5. Experimental results

- System Fusion and Calibration

ID	System	minC_{primary}					C_{primary}				
		cc-1	cc-2	cc-3	cc-4	cc-5	cc-1	cc-2	cc-3	cc-4	cc-5
1	GMM-diag	0.269	0.327	0.132	0.301	0.333	0.345	0.558	0.143	0.456	0.603
2	GMM-full	0.280	0.389	0.139	0.354	0.378	0.363	0.600	0.148	0.496	0.640
3	ML-AFA$_{\text{iso}}^{q=48}$	0.244	0.304	0.115	0.298	0.300	0.344	0.540	0.129	0.452	0.582
4	ML-AFA$_{\text{diag}}^{q=48}$	0.245	0.301	0.130	0.282	0.305	0.348	0.540	0.149	0.450	0.585
Fusion$_{1-3}$	LR *(Abs.)*	0.240	0.286	0.115	0.285	0.284	0.301	0.483	0.120	0.403	0.529
	CLR* *(Abs.)*	0.231	0.261	0.110	0.244	0.247	0.262	0.425	0.115	0.352	0.469
	CLR *(% Rel.)*	16.0	25.5	19.8	23.3	34.6	32.1	31.2	24.6	29.6	28.7
Fusion$_{1,2,4}$	LR *(Abs.)*	0.241	0.285	0.126	0.275	0.289	0.301	0.479	0.128	0.393	0.524
	CLR* *(Abs.)*	0.231	0.267	0.119	0.238	0.246	0.263	0.424	0.121	0.349	0.468
	CLR *(% Rel.)*	16.0	22.5	10.6	26.7	35.1	31.2	31.7	18.2	30.5	28.7
Fusion$_{1-4}$	LR *(Abs.)*	0.238	0.276	0.117	0.274	0.276	0.298	0.471	0.121	0.393	0.516
	CLR* *(Abs.)*	0.231	0.257	0.109	0.236	0.240	0.258	0.416	0.116	0.346	0.459
	CLR *(% Rel.)*	14.1	21.4	17.4	21.6	27.9	25.2	25.4	18.9	24.1	23.9

6. References

[1] T. Hasan and J. H. L. Hansen, “Factor analysis of acoustic features using a mixture of probabilistic principal component analyzers for robust speaker verification,” in Proc. Odyssey, Singapore, Jun. 2012

[2] T. Hasan and J. H. L. Hansen, “Acoustic factor analysis for robust speaker verification,” IEEE Trans. Audio, Speech, Lang. Process., vol. 21, no. 4, pp. 842–853, Apr. 2013.

[3] T. Hasan and J. H. L. Hansen, “Maximum Likelihood Acoustic Factor Analysis Models for Robust Speaker Verification in Noise,” IEEE Trans. Audio, Speech, Lang. Process., vol. 22, no. 2, pp. 381–391, Feb. 2014.

Thank you!