Normalization of Total Variability Matrix for I-Vector/PLDA Speaker Verification

Wei Rao[1], Man-Wai Mak[1], and Kong-Aik Lee[2]

[1]Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hong Kong SAR.
[2]Human Language Technology Department,
Institute for Infocomm Research, A*STAR, Singapore.

Introduction

- Motivations
  - The i-vectors should be subject to length normalization (LN) for effective Gaussian PLDA modeling.
  - However, length normalization causes difficulty for uncertainly propagation (Kenny et al. 2013) because it is not sure how to perform length normalization on the posterior covariance matrix of an i-vector.
- Goal
  - Avoid performing length normalization on i-vectors in Gaussian PLDA modeling.
- How?
  - Normalizing the column vectors of the total variability matrix.

Outline

- Review of I-Vector/PLDA and Uncertainty Propagation
- Normalization of Total Variability Matrix
- Experimental Setup
- Results
- Conclusion

I-Vector Extraction

- Factor analysis model:

$$\boldsymbol{\mu}_s = \boldsymbol{\mu} + \mathbf{T}\mathbf{w}_s$$

$\boldsymbol{\mu}_s$: speaker- and channel-dependent GMM-supervector
$\boldsymbol{\mu}$: GMM-supervector of the UBM
$\mathbf{T}$: low-rank total variability matrix
$\mathbf{w}_s$: i-vector

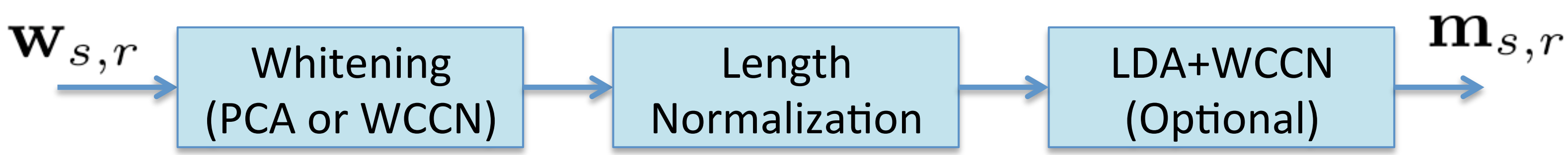
- Given an utterance *X* from speaker *s*,

$$\mathbf{w}_s = \mathbf{L}_s^{-1}\mathbf{T}^{\mathsf{T}}\boldsymbol{\Sigma}^{(b)^{-1}}\tilde{\mathbf{F}}_s \quad \text{where} \quad \mathbf{L}_s = \mathbf{I} + \mathbf{T}^{\mathsf{T}}\boldsymbol{\Sigma}^{(b)^{-1}}\mathbf{N}_s\mathbf{T}$$

i-Vector/PLDA

- Standard PLDA for the *r-th* session of speaker *s*:

  Step 1: I-vector preprocessing



  Step 2: PLDA modeling

$$\mathbf{m}_{s,r} = \mathbf{m} + \mathbf{V}\mathbf{y}_s + \boldsymbol{\epsilon}_{s,r} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$$

- Given enough speakers and sessions per speaker, we can estimate ML solutions of m, V and $\boldsymbol{\Sigma}$

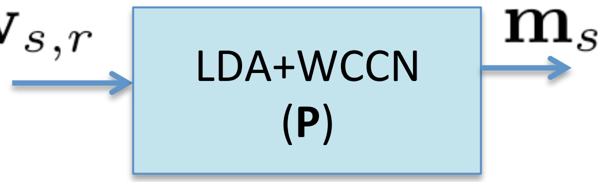
Uncertainty Propagation (UP)

- In standard i-vector/PLDA, i-vectors are point estimates.
- Uncertainty propagation uses the posterior distribution of i-vectors rather than their point estimates.
- The idea is to use the channel factor in the PLDA model to model the uncertainty (posterior cov) of the i-vectors.

Uncertainty Propagation (UP)

- Uncertainty propagation **without** i-vector length normalization:

$\mathbf{w}_{s,r}$



$\mathbf{m}_{s,r} = \mathbf{P}\mathbf{w}_{s,r}$

Projection matrix combining LDA and WCCN

$$\mathbf{m}_{s,r} = \mathbf{m} + \mathbf{U}_{s,r}\mathbf{x}_{s,r} + \mathbf{V}\mathbf{y}_s + \boldsymbol{\epsilon}_{s,r}$$

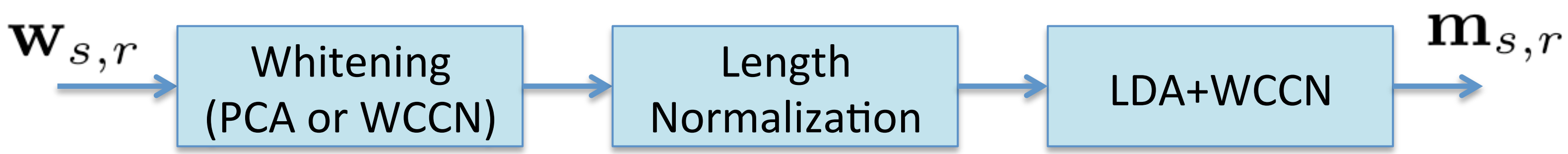
where

$$\mathbf{U}_{s,r}\mathbf{U}_{s,r}^{\mathsf{T}} = \mathrm{cov}(\mathbf{m}_{s,r}, \mathbf{m}_{s,r}) = \mathbf{P}\mathbf{L}_{s,r}^{-1}\mathbf{P}^{\mathsf{T}}$$

$$\mathbf{L}_{s,r}^{-1} = \mathbf{I} + \mathbf{T}^{\mathsf{T}}\boldsymbol{\Sigma}^{(b)^{-1}}\mathbf{N}_{s,r}\mathbf{T}$$

Uncertainty Propagation (UP)

- Uncertainty propagation **with** i-vector length normalization:



$$\mathbf{m}_{s,r} = \mathbf{m} + \tilde{\mathbf{U}}_{s,r}\mathbf{x}_{s,r} + \mathbf{V}\mathbf{y}_s + \boldsymbol{\epsilon}_{s,r}$$

- An **expedient** method to find $\tilde{\mathbf{U}}_{s,r}$:

$$\mathbf{U}_{s,r}\mathbf{U}_{s,r}^{\mathsf{T}} = \mathrm{cov}(\mathbf{w}_{s,r}, \mathbf{w}_{s,r}) = \mathbf{L}_{s,r}^{-1}$$

$$\tilde{\mathbf{U}}_{s,r} = \frac{\mathbf{U}_{s,r}}{\|\mathbf{w}_{s,r}\|}$$

Uncertainty Propagation (UP)

- Another method is to use **unscented** transformation.
- Our goal in this paper is to **avoid** i-vector length normalization so that UP can be applied without using **expedient** methods or **unscented** transformation.

Normalization of Total Variability Matrix

- The total variability matrix $\mathbf{T}$ and i-vectors $\mathbf{w}_s$ cannot be uniquely identified.
- Denote $\mathbf{\Phi}$ as an arbitrary diagonal matrix with non-zero diagonal elements. Then we define:

$$\tilde{\mathbf{T}} = \mathbf{T}\mathbf{\Phi}^{-1} \quad \text{and} \quad \tilde{\mathbf{w}}_s = \mathbf{\Phi}\mathbf{w}_s \quad \text{s.t.} \quad \mathbf{\Phi}^{-1}\mathbf{\Phi} = \mathbf{I}$$

- Then,

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_s &= \boldsymbol{\mu} + \tilde{\mathbf{T}}\tilde{\mathbf{w}}_s \\ &= \boldsymbol{\mu} + \mathbf{T}\mathbf{\Phi}^{-1}\mathbf{\Phi}\mathbf{w}_s \\ &= \boldsymbol{\mu} + \mathbf{T}\mathbf{w}_s \\ &= \boldsymbol{\mu}_s \end{aligned}$$

→ No change in the statistics of supervectors

Normalization of Total Variability Matrix

- Denote $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_R]$ where $\mathbf{t}_r$ is an *MD* x 1 column vector. The normalized total variability matrix is defined as:

$$\mathbf{T}_{norm} = \left[\frac{\mathbf{t}_1}{||\mathbf{t}_1||} \cdots \frac{\mathbf{t}_R}{||\mathbf{t}_R||} \right] = \mathbf{T}\mathbf{\Phi}^{-1}$$

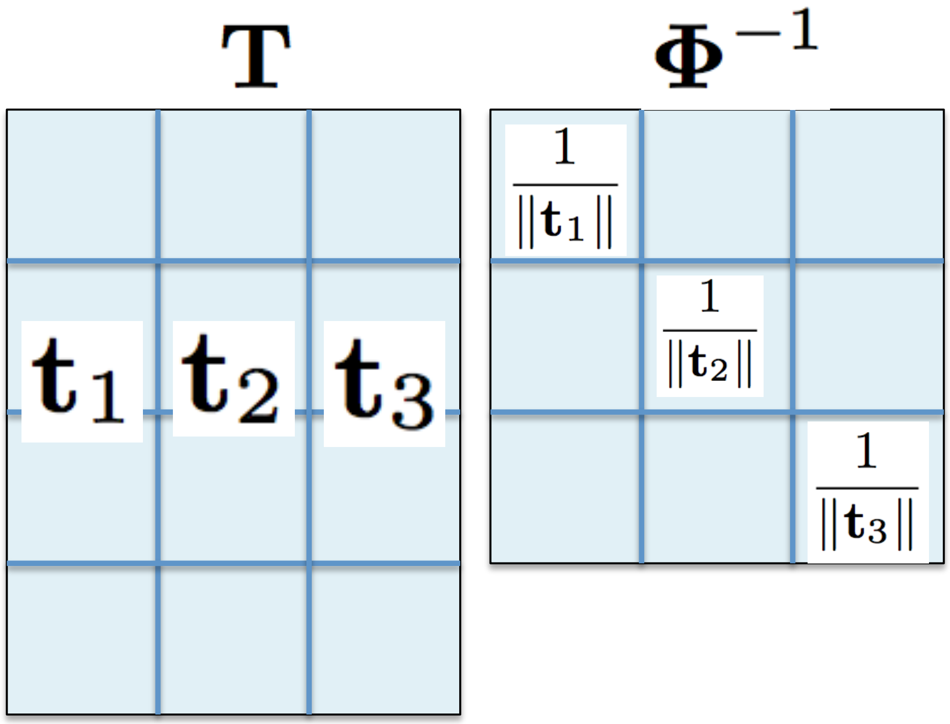
where $\mathbf{\Phi}$ is an *R* x *R* diagonal matrix and its diagonal element $\phi_r = 1/||\mathbf{t}_r||$ is the norm of the *r*-th column of $\mathbf{T}$

T-matrix Normalization

Normalization of Total Variability Matrix

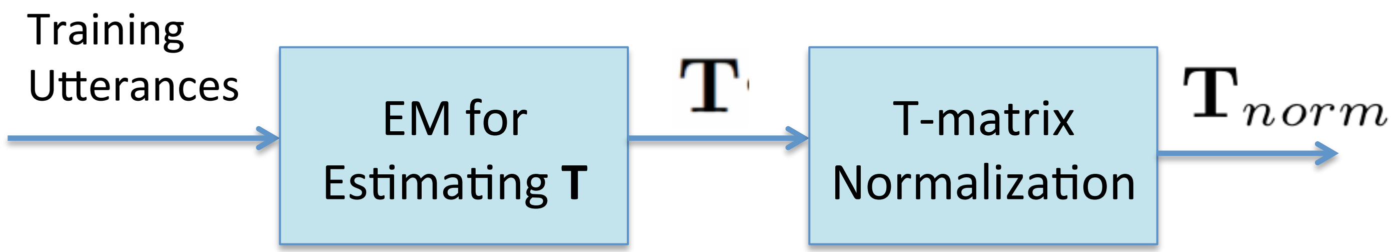
$$\mathbf{T}_{norm} = \left[\frac{\mathbf{t}_1}{||\mathbf{t}_1||} \cdots \frac{\mathbf{t}_R}{||\mathbf{t}_R||} \right] = \mathbf{T}\mathbf{\Phi}^{-1}$$

$$\mathbf{T}_{norm} =$$



Normalization of Total Variability Matrix

- Two ways to perform normalization:
  - **After EM**: After finishing the EM algorithm



  - **Within EM:** At the end of each EM iteration of estimating **T**

T-matrix Normalization within EM

for each **iteration**

E-step:

$$\mathbf{w}_s = \mathbf{L}_s^{-1}\mathbf{T}^\mathsf{T}\boldsymbol{\Sigma}^{(b)^{-1}}\tilde{\mathbf{F}}_s$$

$$\mathbf{L}_s = \mathbf{I} + \mathbf{T}^\mathsf{T}\boldsymbol{\Sigma}^{(b)^{-1}}\mathbf{N}_s\mathbf{T}$$

M-step:

$$\begin{aligned}
\mathbf{C}_i &= \sum_s \tilde{\mathbf{F}}_{s,i}\mathbf{w}_s{}^\mathsf{T}, \\
\mathbf{A}_i &= \sum_s \mathbf{N}_{s,i}\left(\mathbf{L}_s^{-1} + \mathbf{w}_s\mathbf{w}_s{}^\mathsf{T}\right), \\
\mathbf{T}_i &= \mathbf{C}_i\mathbf{A}_i^{-1} \quad i = 1, \ldots, M
\end{aligned}$$

$$\mathbf{T} = \left[\mathbf{T}_1^\mathsf{T} \cdots \mathbf{T}_M^\mathsf{T}\right]^\mathsf{T}$$

T-matrix Normalization

$$\mathbf{T}_{norm} = \left[\frac{\mathbf{t}_1}{\|\mathbf{t}_1\|} \cdots \frac{\mathbf{t}_R}{\|\mathbf{t}_R\|}\right] = \mathbf{T}\boldsymbol{\Phi}^{-1}$$

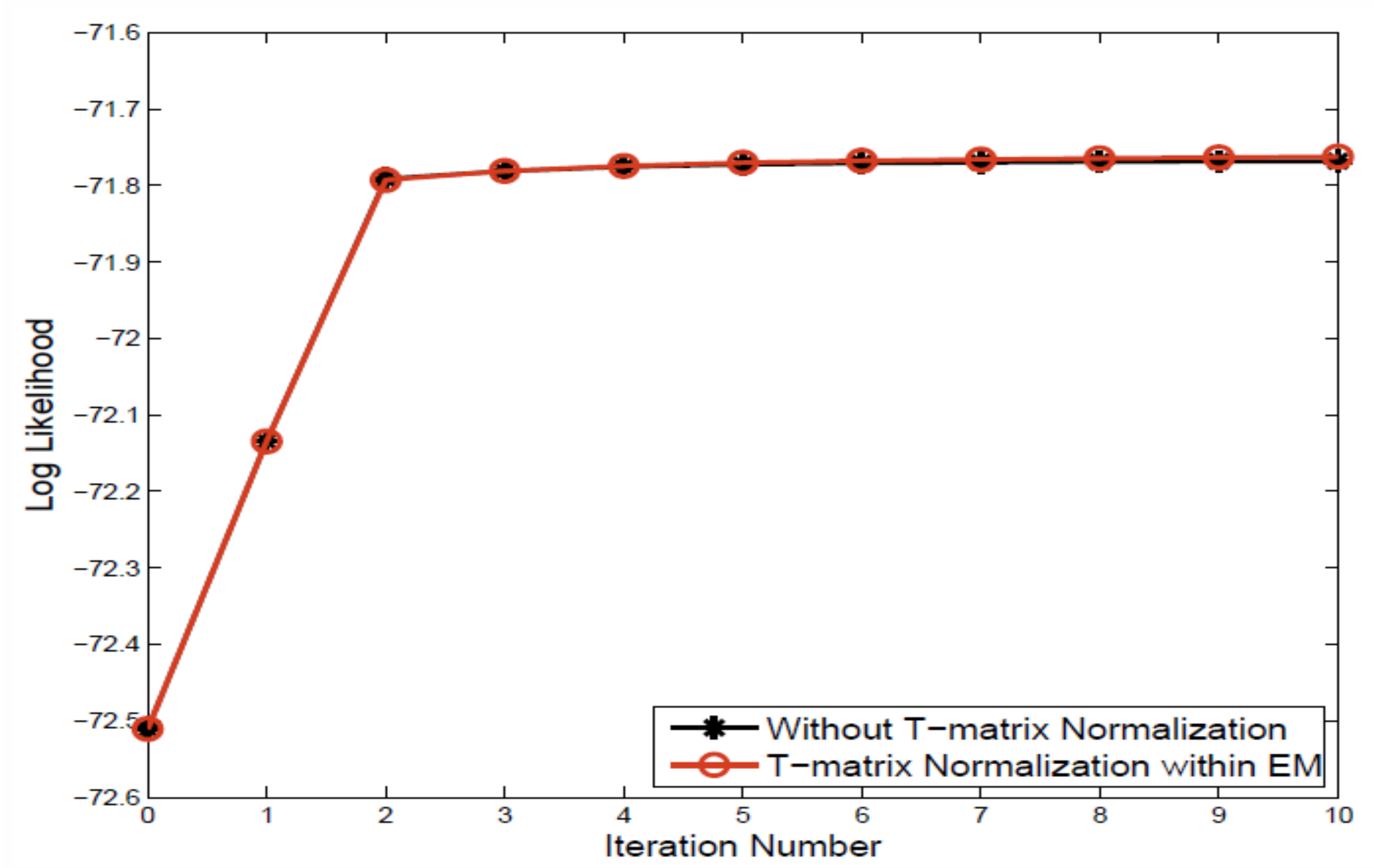
$$\mathbf{T} \leftarrow \mathbf{T}_{norm}$$

end

Experimental Setup

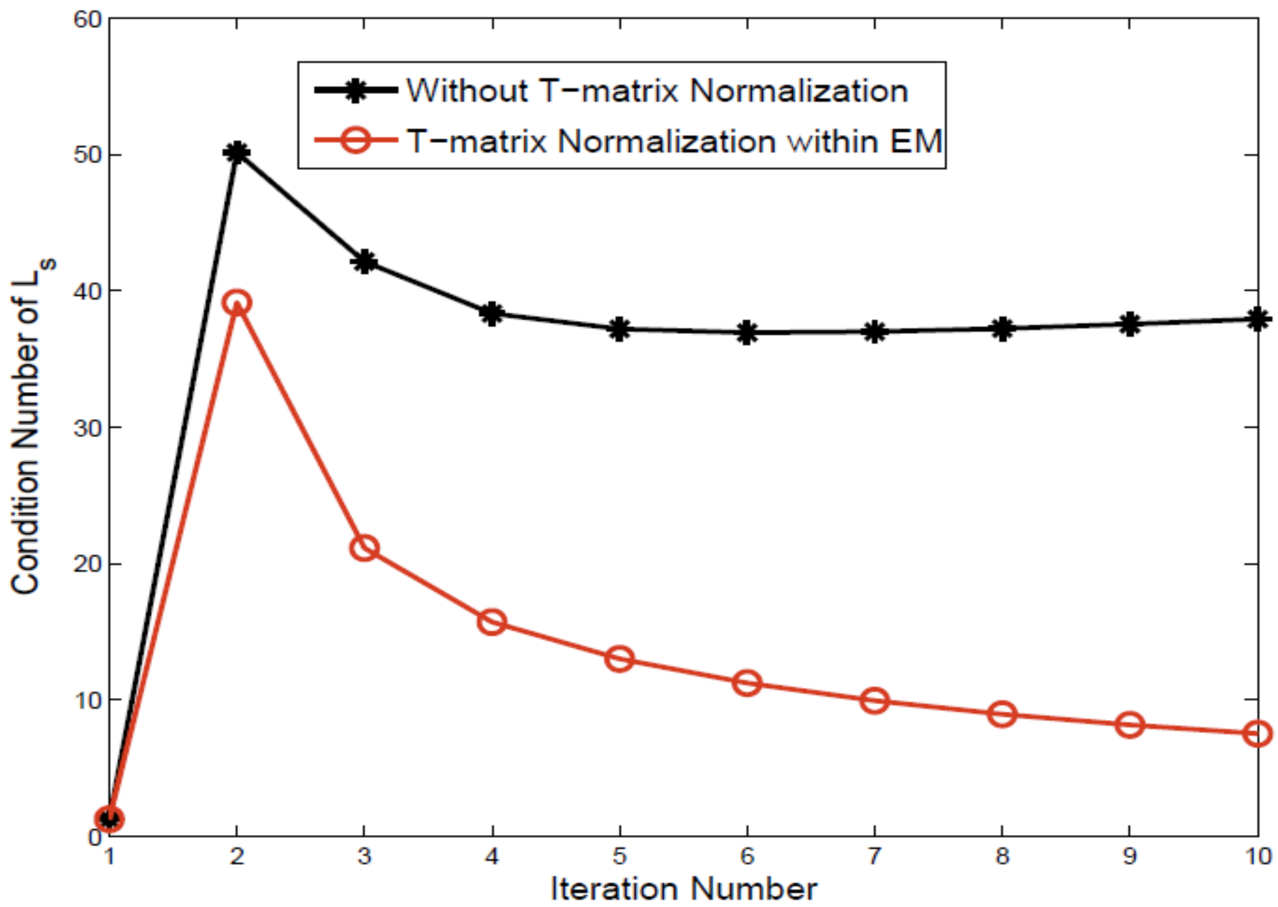
- **Evaluation dataset:** CC1, CC2, and CC4 of NIST 2010 SRE male *extended core* set and CC2 of NIST 2012 SRE male *core* set.
- **Parameterization:** 19 MFCCs together with energy plus their 1st and 2nd derivatives → 60-Dim
- **UBM:** gender-dependent, 1024 mixtures
- **Total Variability Matrix:** gender-dependent, 400 total factors
- **I-Vector Preprocessing:**
  - Whitening by WCCN then length normalization
  - Followed by LDA (400-dim → 200-dim) and WCCN
- **PLDA:** Gaussian PLDA model with 150 latent variables

Property of Normalized T-Matrix



Performing T-matrix normalization within the EM iterations will not affect the convergence of the EM algorithm.

Property of Normalized T-Matrix



The T-matrix normalization helps to improve the numerical stability of the posterior covariance matrices $\mathbf{L}_s^{-1}$.

Within EM vs. After EM

T-matrix Normalization	I-vector Preprocessing	EER (%)			MinNDCF (2010)		
		CC1	CC2	CC4	CC1	CC2	CC4
Within EM	WCCN	1.41	2.43	3.27	0.354	**0.515**	0.522
After EM	WCCN	**1.37**	**2.34**	**3.23**	**0.353**	0.521	**0.520**

NIST 2010 SRE (male)

Performing T-matrix normalization after the EM algorithm is slightly better

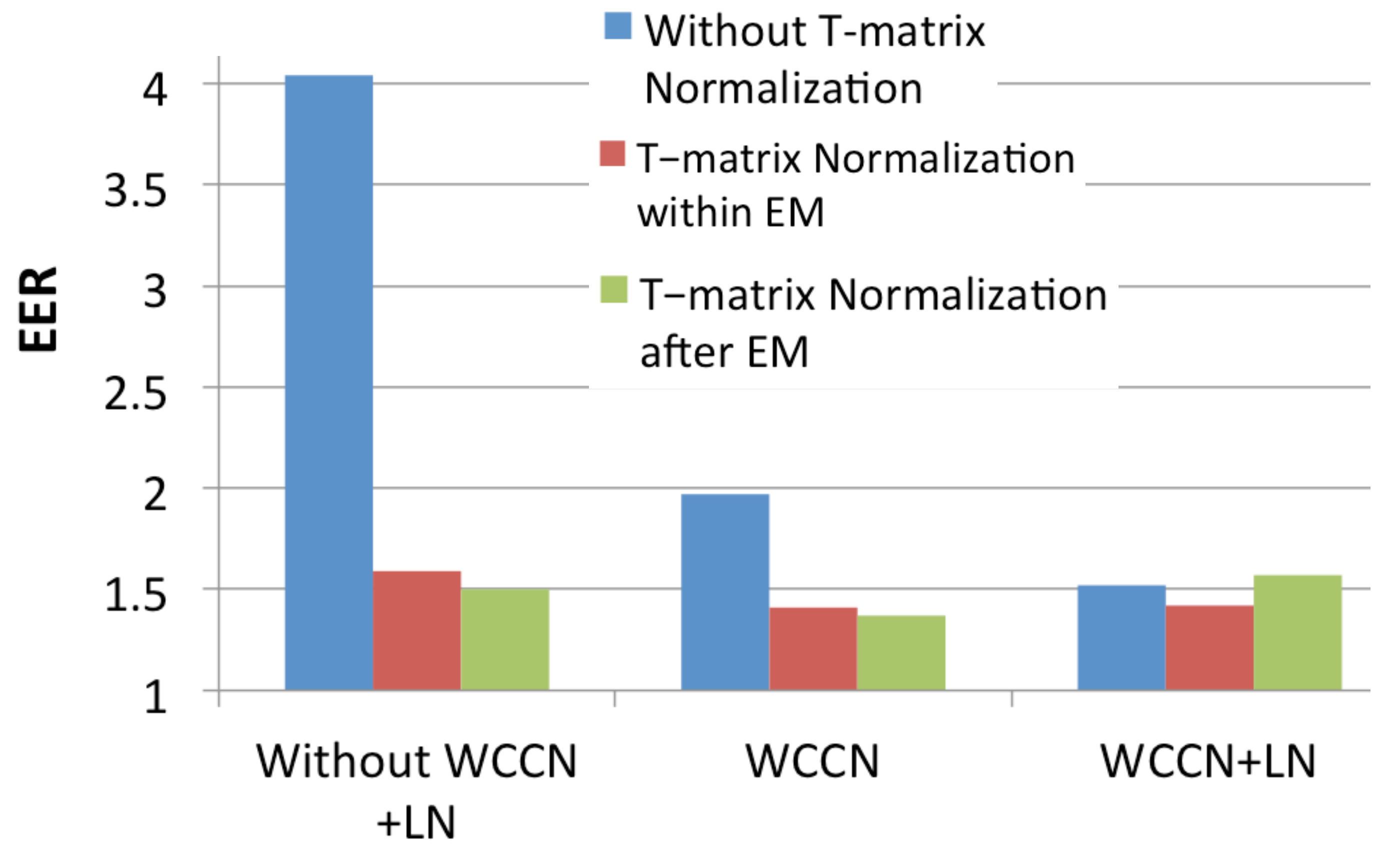
Do We Need Length Norm?

T-matrix Normalization	I-vector Preprocessing	EER (%)			MinNDCF (2010)		
		CC1	CC2	CC4	CC1	CC2	CC4
No	WCCN+LN	1.52	2.41	**2.97**	0.308	0.457	0.442
Yes (After EM)	None	1.50	2.49	3.07	**0.233**	**0.414**	**0.433**
Yes (After EM)	WCCN	**1.37**	**2.34**	3.23	0.353	0.521	0.520

NIST 2010 SRE (male)

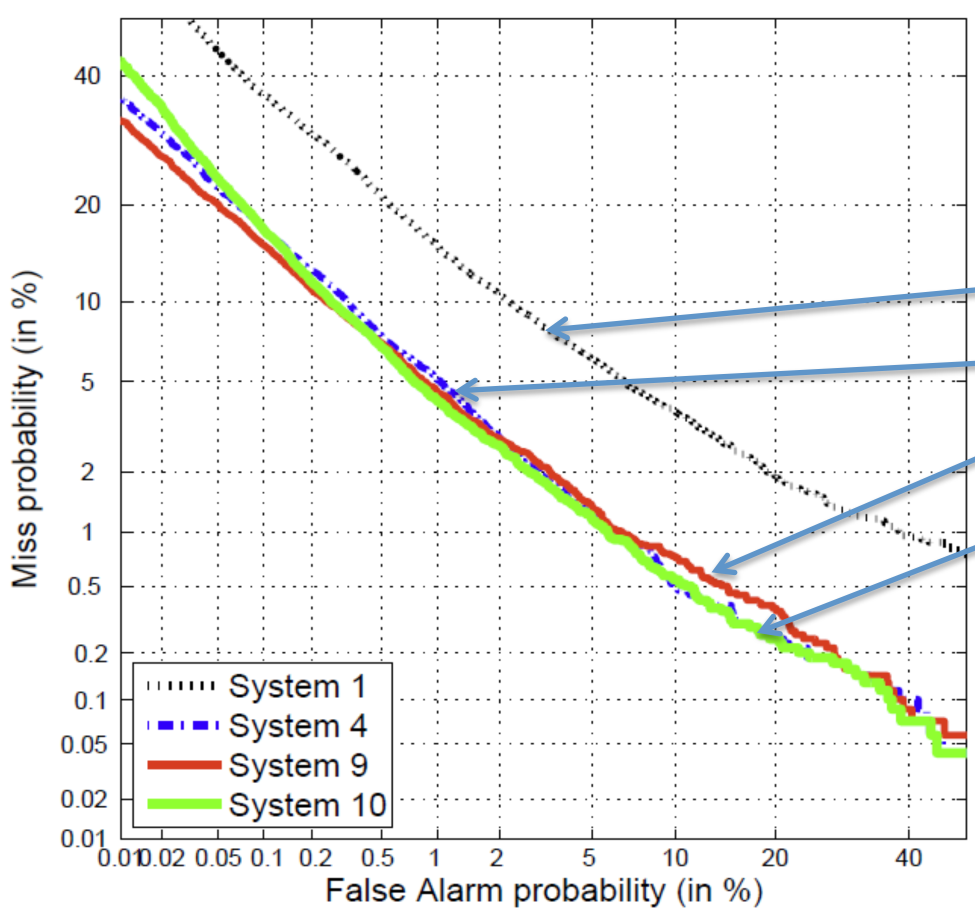
I-vector length normalization (LN) is not necessary whenever the total variability matrix has been normalized.

Do We Need Whitening (WCCN)?



- After T-matrix normalization, whitening (WCCN) can slightly improve performance but length norm hurts a bit.

The DET performance under CC2 in NIST 2010 SRE



Sys	T-matrix Normalization	I-vector Preprocessing
1	No	None
4	No	WCCN+LN
9	Yes (After EM)	None
10	Yes (After EM)	WCCN

Length norm is not necessary if we have T-matrix norm

Performance Comparison for NIST 2012 SRE

Common Condition 2

T-matrix Normalization	I-vector Preprocessing	EER (%)	MinNDCF (2010)
No	WCCN+LN	**2.82**	0.351
Yes (After EM)	None	3.30	0.273
Yes (After EM)	WCCN	3.25	**0.272**

I-vector length normalization is not necessary whenever the total variability matrix has been normalized, which also agrees with the results in NIST 2010 SRE.

Conclusions

- This paper proposes performing T-matrix normalization instead of performing length normalization on i-vectors in i-vector/PLDA speaker verification.
- The proposed method can achieve almost the same performance as i-vector length normalization.
- The proposed method opens up opportunity to enhance the performance of uncertainly propagation because the issue of applying length normalization on the posterior covariance of i-vectors can now be resolved.