

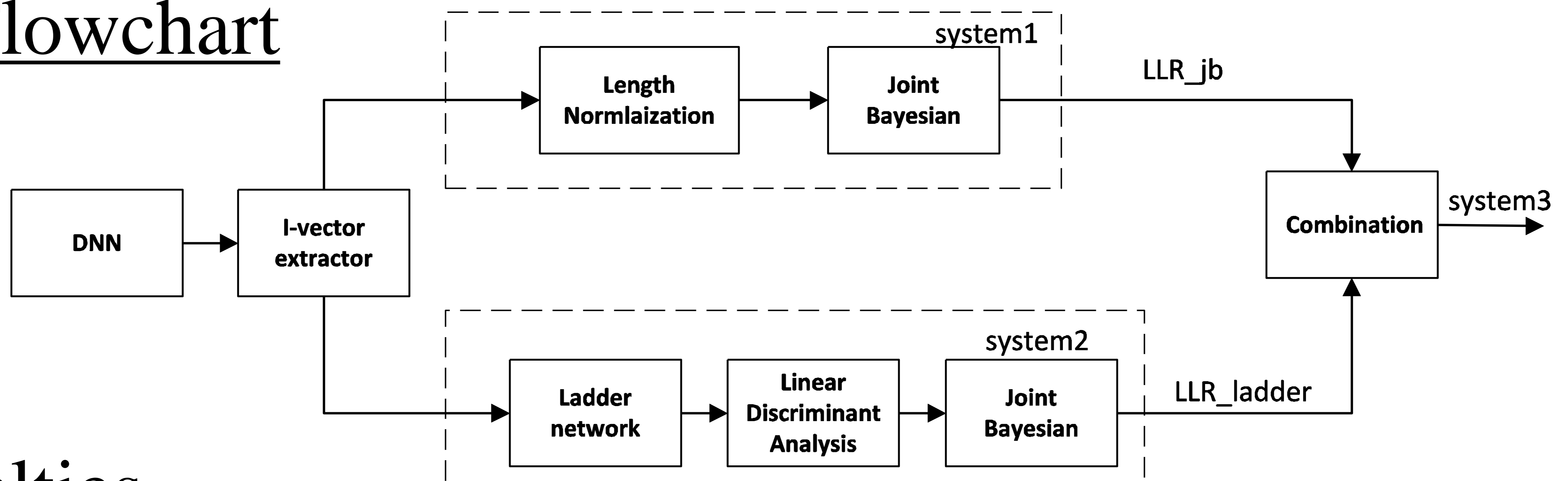
The THU-SPMI SRE-16 System with Joint Bayesian Scoring and Ladder Network based Feature Learning



Yiyan Wang, Haotian Xu, Zhijian Ou

Speech Processing and Machine Intelligence (SPMI) Lab, Tsinghua University, Beijing, China

System Flowchart



Our Novelties

1. Joint Bayesian Scoring

$$x_{ij} = \mu_i + \epsilon_{ij}$$

Experiment results on SRE10

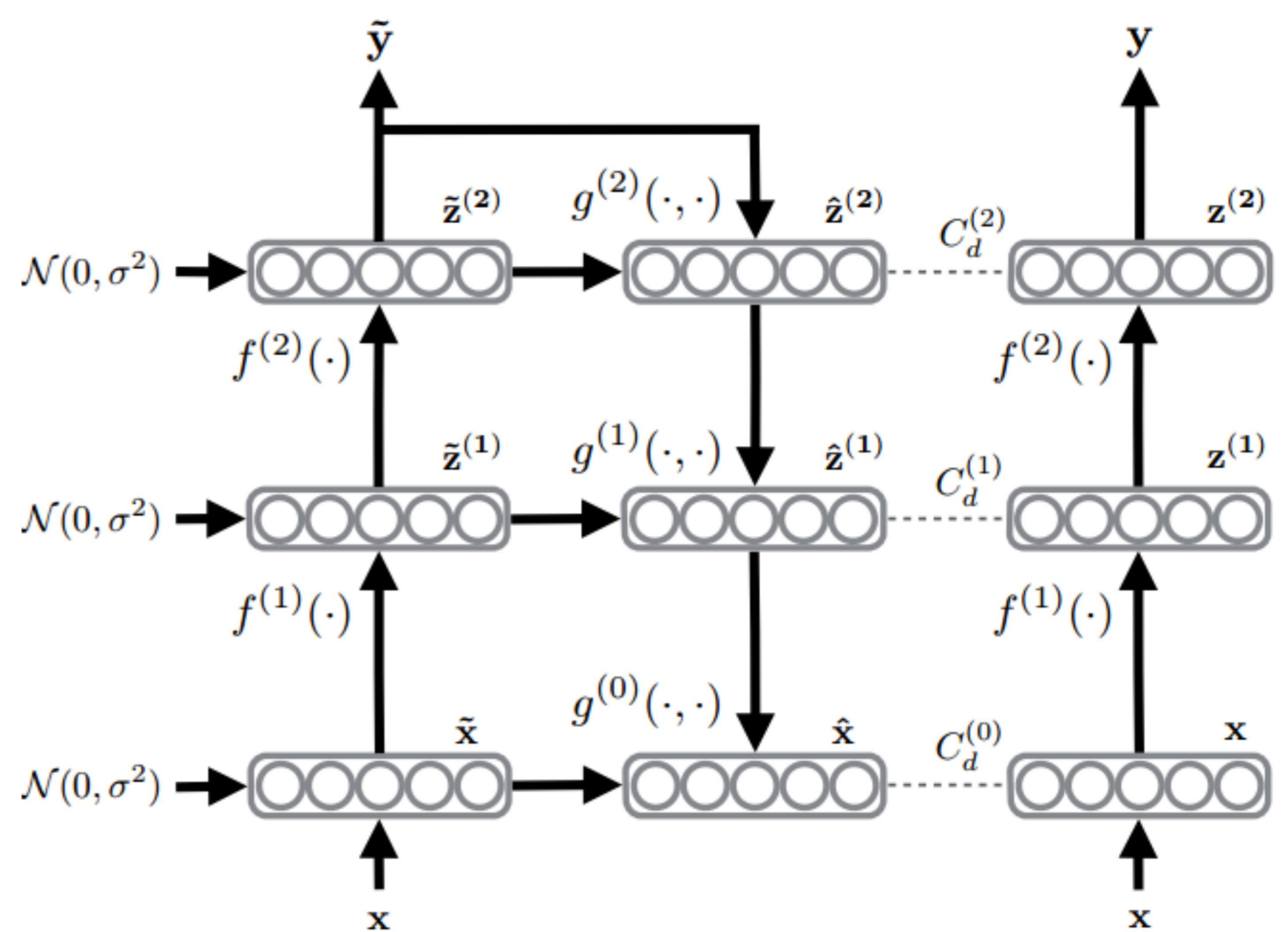
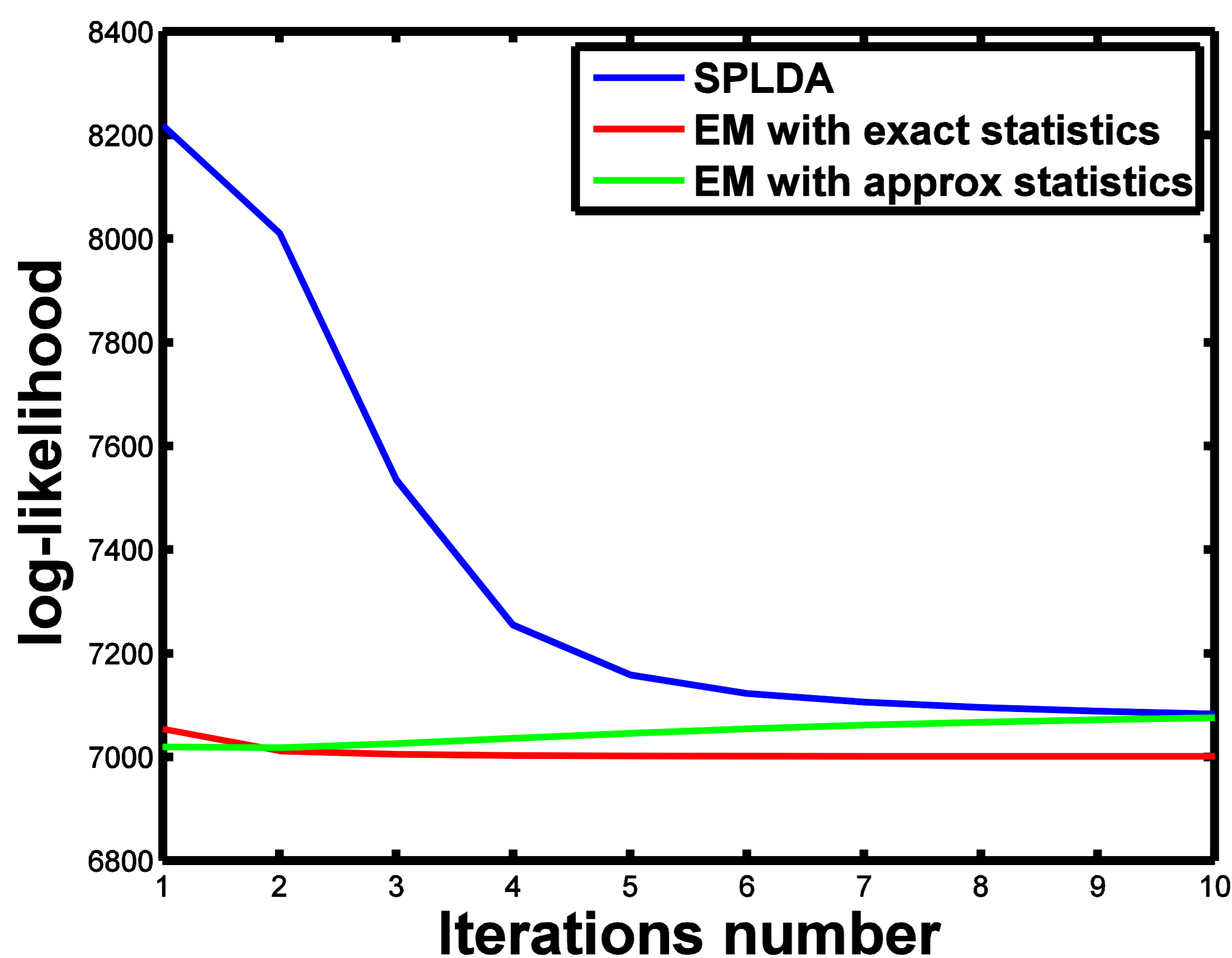
System	SRE10 MALE			SRE10 FEMALE		
	EER	DCF10	DCF08	EER	DCF10	DCF08
LDA+COS	1.905	0.292	0.091	2.619	0.399	0.126
Kaldi PLDA	1.299	0.284	0.079	1.944	0.345	0.102
SPLDA	1.010	0.217	0.055	1.621	0.287	0.079
JB	0.894	0.188	0.048	1.485	0.245	0.069

2. Ladder Network based Feature Learning leveraging both labeled and un-labeled i-vectors

$$L(x, y) = -\log p(\tilde{y}|x) + \sum_{i=1}^N \|z^{(i)} - \hat{z}_{BN}^{(i)}\|^2 + \alpha \|h^{N-1} - c_y\|^2$$

Center loss for verification

$$h^{N-1} = f(\gamma^{N-1}(\tilde{z}^{N-1} + \beta^{N-1}))$$



Reference

Experiment results

1. Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," ECCV, 2016.
2. A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," NIPS, 2015.
3. Y. Wang, H. Xu, and Z. Ou, "Joint Bayesian Gaussian discriminant analysis for speaker verification," ICASSP 2017, accepted.

	System 1(secondary)	System 2	System 3(primary)
dev			
Equalized			
eer	20.29	19.13	19.18
min_Cprimary	0.8243	0.8225	0.8080
act_Cprimary	0.9095	0.9992	0.9986
Un-equalized			
eer	19.41	19.78	19.55
min_Cprimary	0.8170	0.8213	0.8045
act_Cprimary	0.9209	0.9994	0.9989
eval			
Equalized			
eer	15.39	17.24	14.71
min_Cprimary	0.7826	0.8570	0.7747
act_Cprimary	0.8993	0.9154	0.8860
Un-equalized			
eer	15.19	17.90	14.58
min_Cprimary	0.8025	0.8795	0.7949
act_Cprimary	0.9205	0.9474	0.9089