Recursive Whitening Transform for Speaker Recognition on Language Mismatched Condition

Suwon Shon, Seongkyu Mun*, Hanseok Ko*

MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA

*School of Electrical Engineering, Korea university, Seoul, South Korea**

Abstract

- Main issue of SRE 16 is ***language mismatch compensation***
- On conventional i-vector extraction, whitening transformation is primary step for normalizing i-vector
- While the whitening transformation is done using in-domain dataset on both training and test dataset, the training dataset always remained as un-whitened because of language mismatches
- We propose recursive whitening transformation approach to remove the un-whitened residual components by using sub-corpora dataset
- Conventional domain mismatch compensation techniques were used to compare (IDVC, DICN)
- For experiments, 4 different i-vectors (GMM, DNN, Sup-GMM, BNF-GMM)
- While state-of-art i-vector extraction based on phonetically aware model does not showed advantages on language mismatched condition, proposed approach shows effectiveness on evaluation of SRE16

SRE 16 dataset

- Statistics

Language set	Category	Labels	Numbers of Utt.	Numbers of Spk.
English	Training	Available	64000>	6400>
Minor	Training	X	200	20
	Enrollment	Available	120	20
	Test	Available	1207	20

English : SRE04~10, SWB
Minor : Cebuano, Mandarin

- Performance evaluation
  - ➢Equal Error Rate (EER)
  - ➢$_{min}$DCF16-1
  - ➢$_{min}$DCF16-2

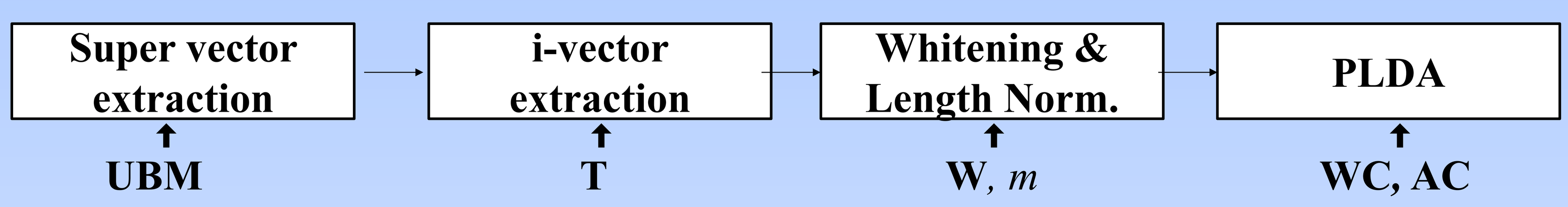
ID	C_{Miss}	$C_{FalseAlarm}$	P_{Target}
1	1	1	0.01
2	1	1	0.005

<SRE16 Cost parameters>

➢$_{min}C_{primary}$ = ($_{min}$DCF16-1 + $_{min}$DCF16-2) / 2

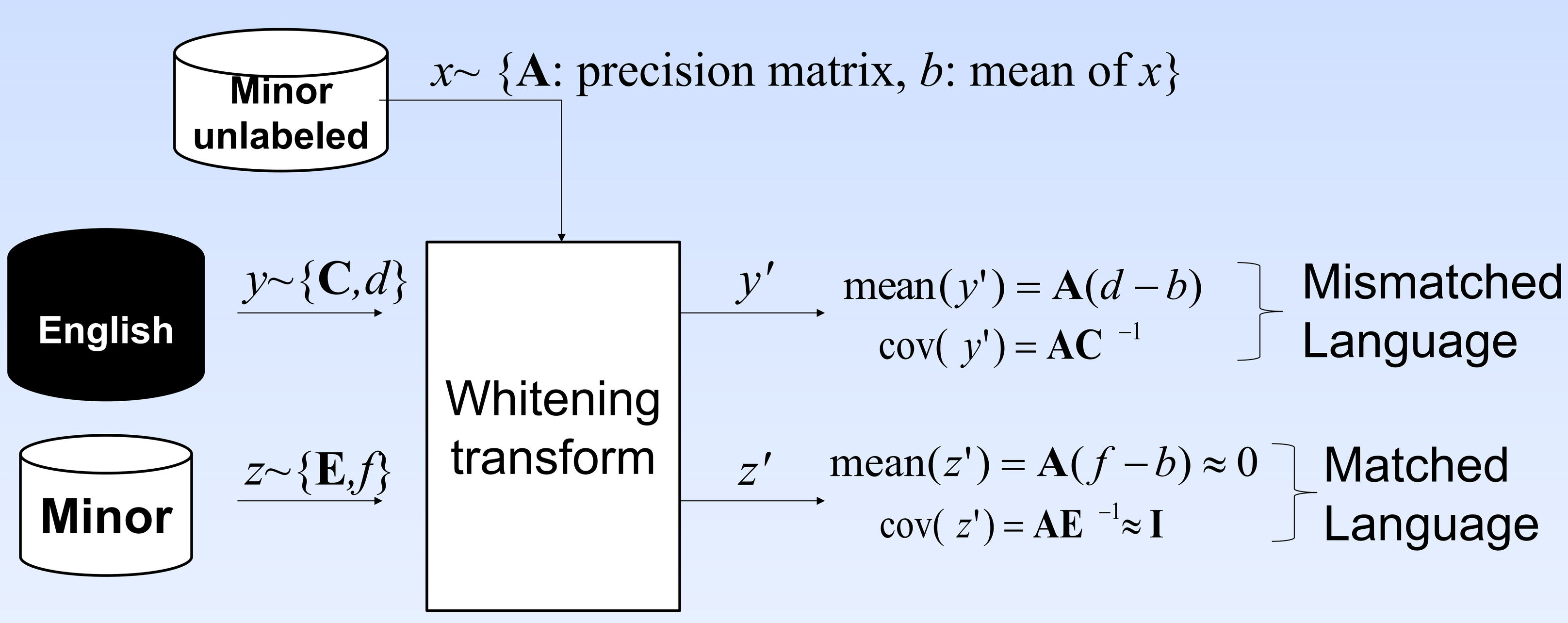
Conventional approach

- i-vector extraction



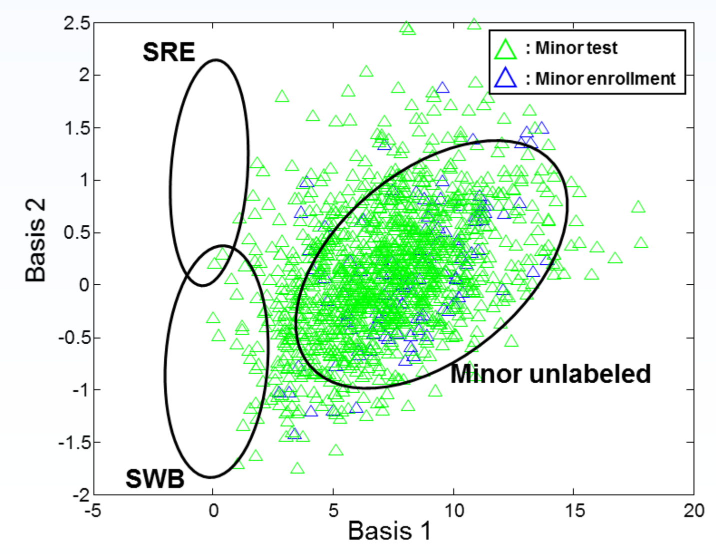
<Block diagram of i-vector extraction>

- Whitening transform and residual component



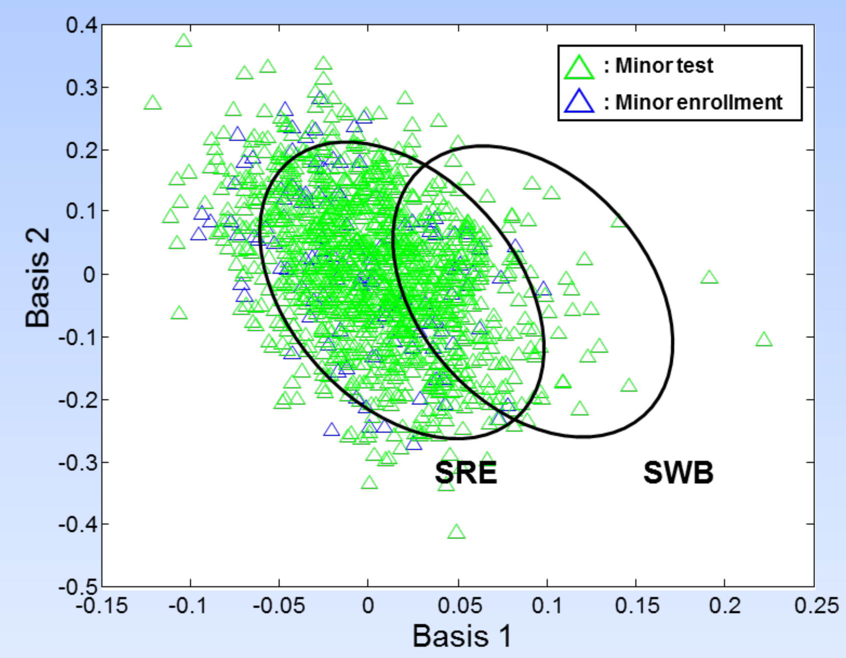
Whitening transform	Compensation techniques	EER	DCF16-1	DCF16-2	$_{min}C_{primary}$
minor	-	21.0232	0.8217	0.8598	0.8407
	IDVC	21.0957	0.8223	0.839	0.8306
	DICN	21.0853	0.8193	0.8722	0.8457
English	-	19.361	0.866	0.8978	0.8819
	IDVC	19.0452	0.8605	0.8841	0.8723
	DICN	19.5112	0.8630	0.9036	0.8833

<Baseline performance on SRE16 minor test>

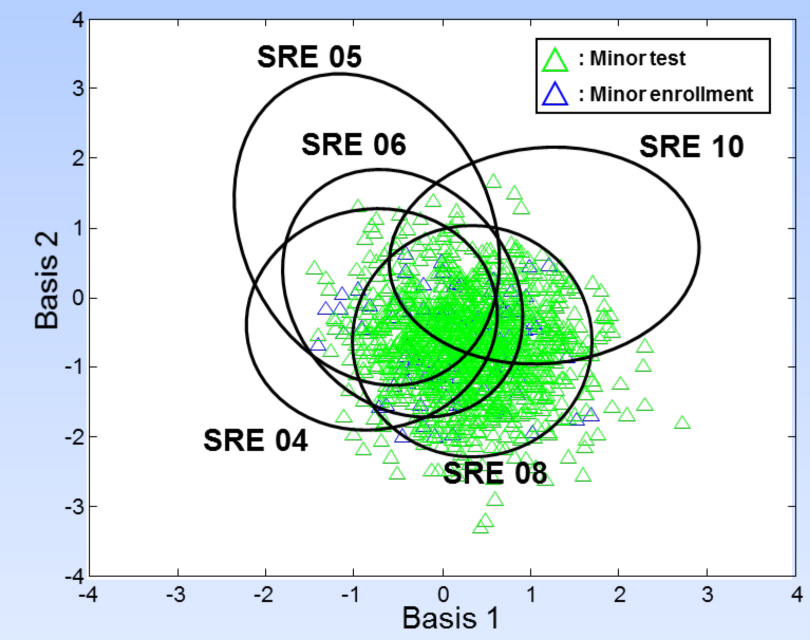


<Projection of minor i-vectors on PCA subspace of SRE, SWB and Minor unlabled dataset from training category>

Recursive whitening



<After whitening transform with minor unlabeled: Projection of minor i-vectors on PCA subspace of SRE and SWB



<After whitening transform twice with minor unlabeled and SRE: Projection of minor i-vectors on PCA subspace of SRE04~10>

- Recursive Whitening transformation using sub-corpora

Sub-corpora Level i	Sub-corpora (sub-corpora index j)	Recursively whitened i-vector at each level i
0	Minor unlabeled dataset (*1*)	$f_0(\omega)$
1	SRE(*1*), SWB(*2*)	$f_1(\omega)$
2	SRE04(*1*), SRE05(*2*), SRE06(*3*), SRE08(*4*), SRE10(*5*), SWB2 p1~p3(*6~8*), SWB2 c1~c2(*9,10*)	$f_2(\omega)$

➢Determine which sub-corpora is closest

$$J_i = \arg\max_{j\in(1,\dots,K)} p(f_{i-1}(\omega)\,|\,\boldsymbol{\theta}_{ij})$$

➢Whitening with the closest sub-corpora

$$f_i(\omega) = \eta\left(S_i(J_i)\cdot f_{i-1}(\omega) - \mu_i(J_i)\right)$$

Experiment

- i-vector extraction systems
  - ➢**GMM-UBM**
    - 2048 comp. GMM-UBM, 60-dim MFCC
  - ➢**DNN-UBM**
    - Time delay NN for acoustic model
    - DNN-UBM is estimated with 5567 comp.
  - ➢**Supervised GMM-UBM (SGMM-UBM)**
    - Phonetically-aware supervised GMM-UBM with 5567 comp. using TDNN post.
  - ➢**Bottleneck feature based GMM-UBM (BNF-UBM)**
    - DNN layer structure for 1500-1500-80-1500
    - 3rd layer(80 dim) for GMM-UBM estimation
- Common Back-end
  - ➢All i-vector was extracted in 600 dim.
  - ➢PLDA parameter were estimated using SRE04~10 with 400 eigenvoice

	Sub-corpora for whitening: Level 0	Level 1	Level 2	Compen-sation	EER	$_{min}$DCF16-1	$_{min}$DCF16-2	$_{min}C_{primary}$
Conventional (Level 0 recursive whitening)	Minor			-	21.02	0.8217	0.8598	0.8407
	Minor			IDVC	21.10	0.8223	0.839	0.8306
	Minor			DICN	21.08	0.8193	0.8722	0.8457
Level 1 recursive whitening	Minor	SRE		-	17.48	0.7358	0.7556	0.7457
	Minor	SRE		IDVC	**17.01**	0.7198	0.7504	0.7351
	Minor	SRE		DICN	17.33	0.7204	0.7518	0.7361
Level 2 recursive whitening	Minor	SRE	SRE-08	-	17.92	**0.7085**	**0.7447**	**0.7266**
	Minor	SRE	SRE-08	IDVC	17.21	0.7123	0.7474	0.7298
	Minor	SRE	SRE-08	DICN	17.33	0.7233	0.7465	0.7349

<Performance evaluation on SRE16 minor language using DNN-UBM i-vector>

i-vector Extraction System Name	Conventional (level 0): EER	$_{min}C_{primary}$	$_{act}C_{primary}$	Level 1 recursive whitening: EER	$_{min}C_{primary}$	$_{act}C_{primary}$
GMM-UBM	21.91	**0.8068**	**0.8271**	**18.93**	0.7155	0.7293
DNN-UBM	**21.21**	0.8267	0.8428	19.12	**0.6862**	**0.7043**
SGMM –UBM	21.23	0.8099	0.8426	20.05	0.7251	0.7461
BNF-UBM	23.94	0.8973	0.9215	20.19	0.7557	0.7824
Fusion of 4 sub-systems	17.01	0.7179	0.7313	15.67	0.6478	0.6727

<Performance evaluation on SRE16 minor language using multiple i-vector system>

Conclusion

- Recursive whitening transformation is relatively simple, but powerful approach to deal with language mismatched condition
- By result, the approach gradually remove un-whitened residual component
- Robustness on challenge condition where in-domain dataset is extremely small