



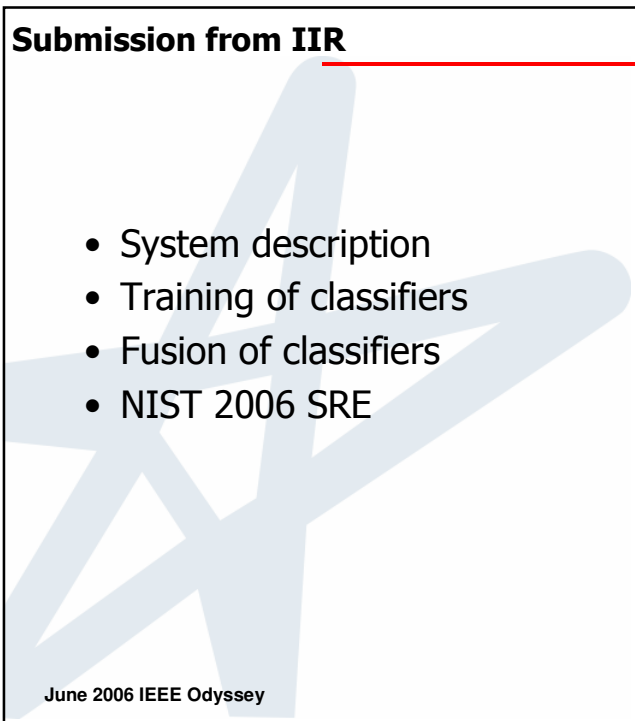
NIST 2006 Speaker Recognition Evaluation – IIR submission

Haizhou Li
Institute for Infocomm Research
Singapore



Submission from IIR

- System description
- Training of classifiers
- Fusion of classifiers
- NIST 2006 SRE



June 2006 IEEE Odyssey

2



IIR Submissions

		Test segment condition			
		10sec 2chan	1conv 2chan	1conv summed chan	1 conv aux mic
Training condition	10sec 2chan	10sec4w- 10sec4w			
	1conv 2chan	1conv4w- 10sec4w	1conv4w- 1conv4w		
	3conv 2chan	3conv4w- 10sec4w	3conv4w- 1conv4w		
	8conv 2chan	8conv4w- 10sec4w	8conv4w- 1conv4w		
	3conv summed chan				

June 2006 IEEE Odyssey

3



System Description - Preprocessing

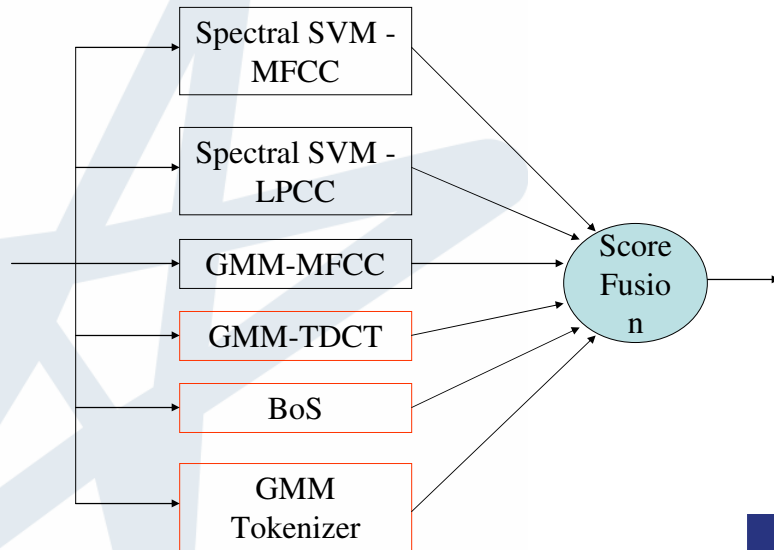
- VAD: voice activities detection with GMM resulting in 38% speech vs. 62% nonspeech in terms of frames
- Removal of cross-talk
- MFCC, LPCC, Temporal DCT feature extraction
- Conversation-based CMN

June 2006 IEEE Odyssey

4



System Description



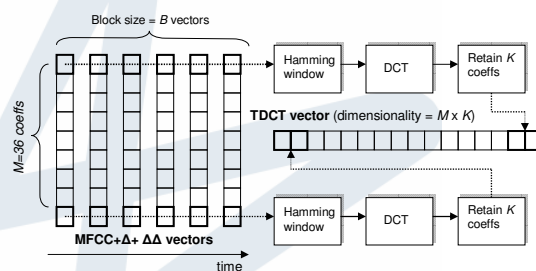
June 2006 IEEE Odyssey

5



Temporal Discrete Cosine Transform (TDCT) Classifier

- Modeling of temporal information spanning longer time than delta/delta-deltas



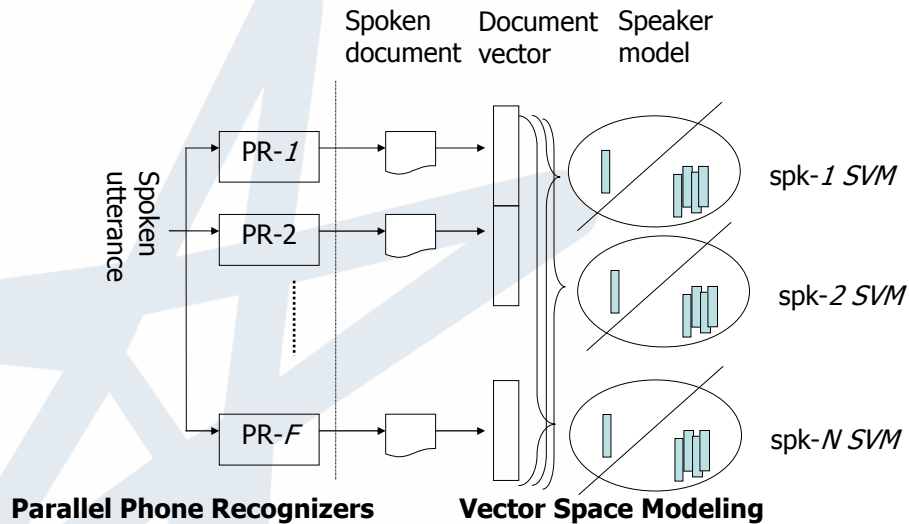
- Block size $B = 8$ vectors (250 ms, ~syllable length unit)
- Retain $K = 3$ lowest DCT coefficients → Dimensionality = $3 \times 36 = 108$
- Applying VAD indicator + mean/variance normalization after feature extraction
- Standard GMM-UBM recognizer using gender-dependent background models of 256 components each

June 2006 IEEE Odyssey

6



Bag-of-sounds Vector Space Classifier* - VSM by language

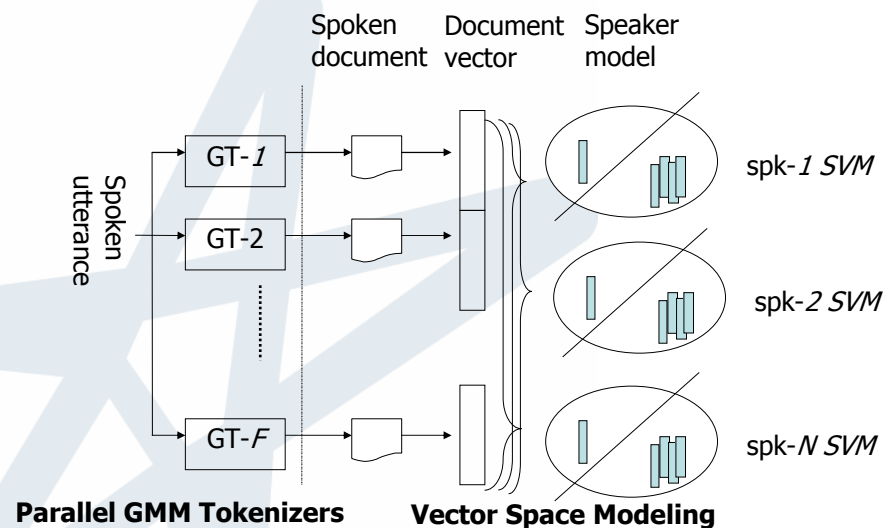


*H. Li and B. Ma, "A Phonotactic Language Model for Spoken Language Identification",
43rd Annual Meeting of the Association for Computational Linguistics (ACL05), 2005
June 2006 IEEE Odyssey

7



GMM Tokenizer (Vector Space) Classifier*- VSM by speaker group



*B. Ma, D. Zhu, R. Tong, H. Li, "Speaker cluster based GMM tokenization
for speaker recognition", *Interspeech'06*
June 2006 IEEE Odyssey

8



Training of Classifiers

	Spectral SVM-MFCC	Spectral SVM-LPCC	GMM-MFCC	GMM-TDCT	Bag-of-sounds	GMM tokenizer
Tokenizer training data	N.A.				1. IIR-LID corpus 2. LDC Korean corpus (LDC2003S03) 3. MAT corpus 4. OGI-TS corpus	NIST SRE 2002 corpus
Background speaker data (UBM)	2 gender-dependant background data sw3p1, sw3p2, sw2p2 and sw2p3		NIST2004 1side training files	NIST2004 1side training files	NIST SRE 2002 corpus	NIST SRE 2004 corpus
Cohort data (Tnorm)	Evaluation set of the NIST SRE 2004 corpus		N.A.			
Dev. Data (train/test)	NIST SRE 2005 corpus					
1conv4w-1conv4w EER%	11.02	9.55	12.71	13.15	18.99	18.67

June 2006 IEEE Odyssey

9



Fusion of Classifiers

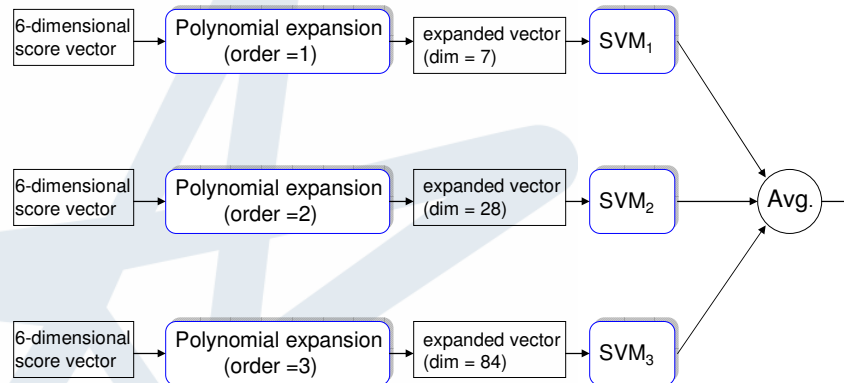
- Each sample is encoded in a vector of 6 elements
- SVM classifier is trained by using *true* vectors as positive data, *false* vectors as negative data
- SVM output is used as the fused score
- The threshold is decided based on NIST 2005 SRE

June 2006 IEEE Odyssey

10



Fusion by SVM



SVM training:

- Score vectors labeled "+1" for genuine speakers and "-1" for impostors
- Linear SVM, trained using SVMTool package

June 2006 IEEE Odyssey

11



Final Submissions

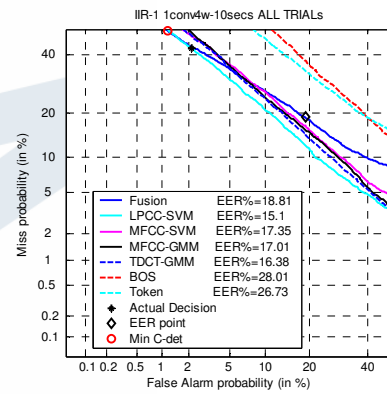
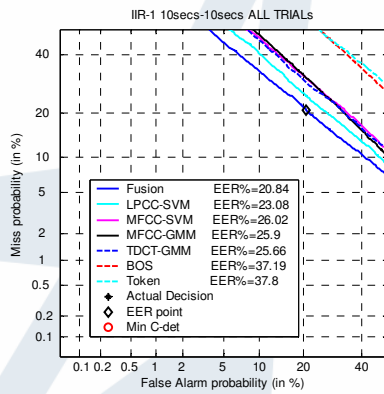
- IIR1 – Fusion by SVM (EER), fused score decision (Actual decision)
- IIR2 – Fusion by SVM (EER), majority vote (Actual decision)
- IIR3 – Spectral SVM LPCC

June 2006 IEEE Odyssey

12



NIST 2006 SRE (4 wire: 10sec-10sec/1conv-10sec)

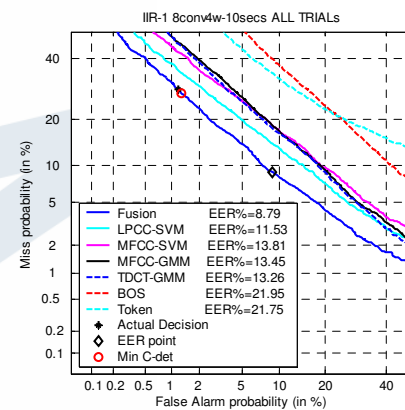
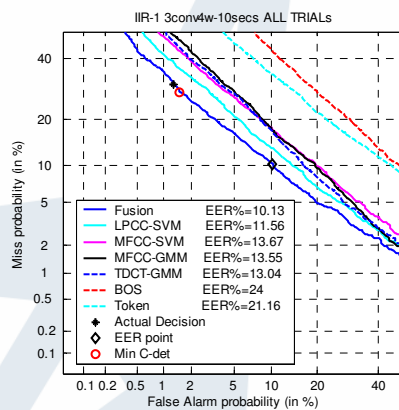


June 2006 IEEE Odyssey

13



NIST 2006 SRE (4 wire: 3conv-10sec/8conv-10sec)

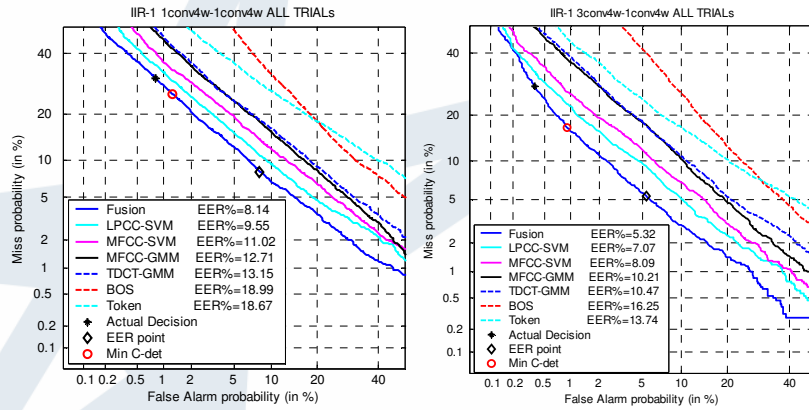


June 2006 IEEE Odyssey

14



NIST 2006 SRE (4 wire: 1conv-1conv/3conv-1conv)

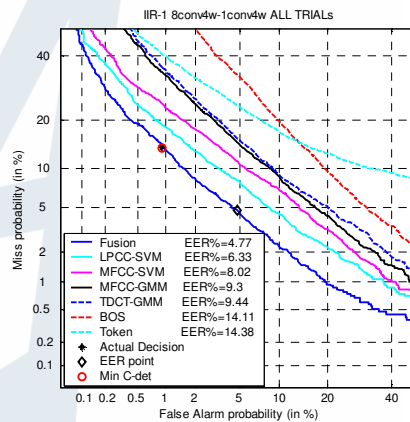


June 2006 IEEE Odyssey

15



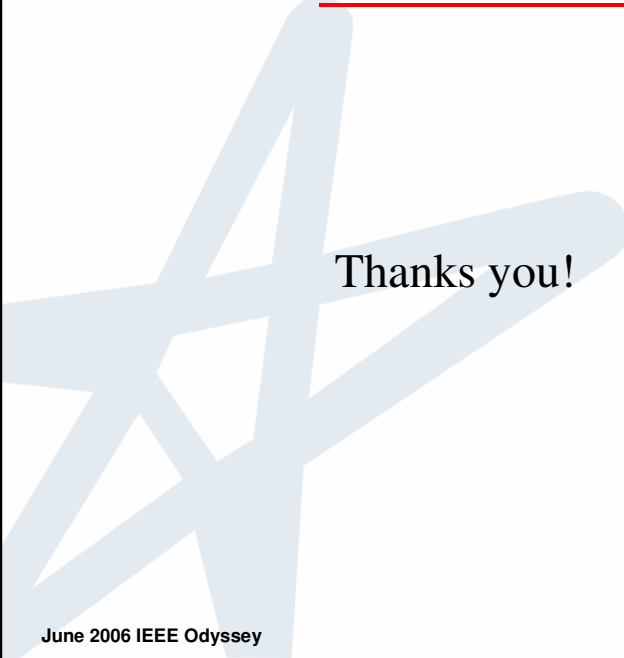
NIST 2006 SRE (4 wire: 8conv-1conv)



June 2006 IEEE Odyssey

16





Thanks you!

June 2006 IEEE Odyssey

17

