Persay systems for NIST SRE 2006

Ran Gazit & Nir Krause

Persay Ltd., Israel

PerSay

Classifiers

1. **SGM:** **S**VM classification in the **G**MM **M**odels space
2. **PHN:** SGM constrained to n-grams of broad phonetic classes
3. **GMM**

Systems

SGM is the core classifier on all submitted tasks.

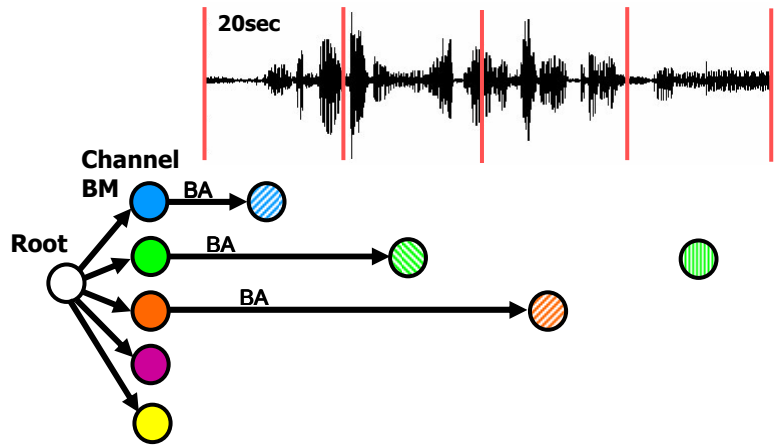
Extensions:

1conv-1conv: SGM + PHN

10sec-10sec: SGM + GMM

PerSay

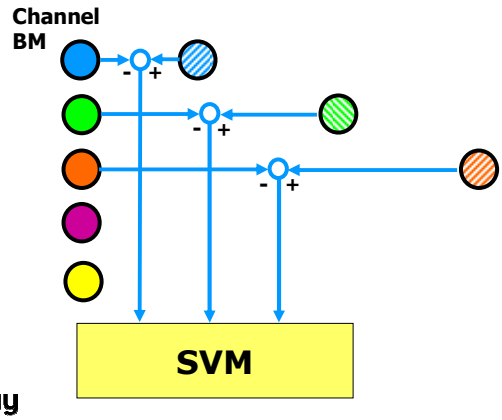
SGM: GMM training



PerSay

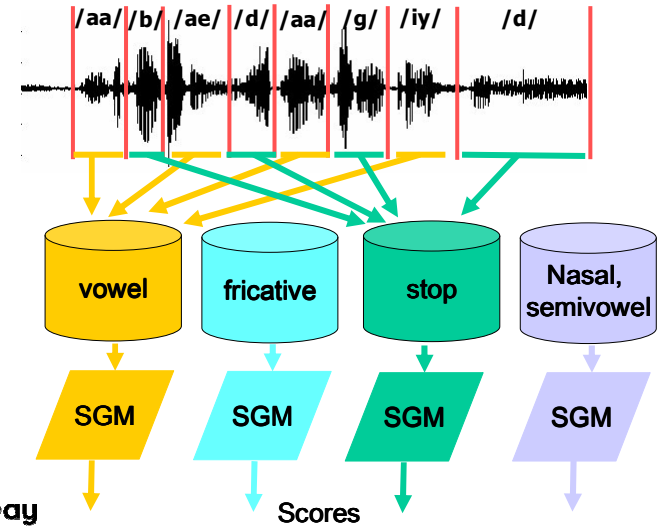
SGM: SVM training

- Positive examples from training segment
- Negative examples from first 20sec of each file in dev set
- All examples are rank normalized



PerSay

PHN: SGM constrained to broad phonetic classes



PerSay

PHN: details…

- Use BRNO phoneme recognizer
- Divide into 4 broad phonetic classes
- Extract audio of each broad phonetic class, and of trigrams of broad phonetic classes
- Create an SGM classifier for each broad phonetic class and each trigram
- Discard classifiers without enough audio for classification of most NIST05 tests

PerSay

SGM Baseline configuration

Features:

- **Echo suppression using 4wire data**
- **Energy-based VAD**
- **2 sets of features:**
  - **20 LPCC + 20 Delta**
  - **19 MFCC + 19 Delta + RASTA**
- **CMS, unit-variance**

Models:

- **5 UBM per gender** (cdma, gsm, carb, elec, nist05)
- **Means-only adaptation from best UBM**

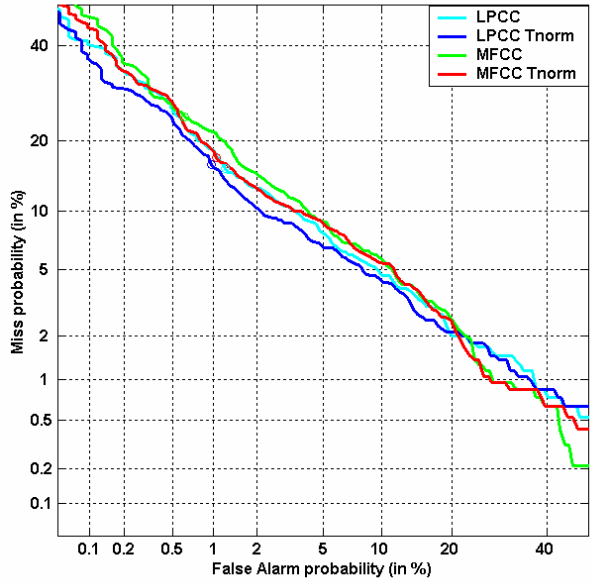
PerSay

Development data

			male	female
Background models	**99**	carb	389	306
		elec	644	1049
	03	**GSM**	221	229
		CDMA	560	819
	05		80	110
SGM negative examples	**99,03,04**		2572	3414
Tnorm	**04**		100	100
Parameter optimization & threshold setting	**05** **English only**	Tar	1099	1366
		Imp	10009	13416

PerSay

T-norm & features effect



1conv4w-1conv4w,
English, Males

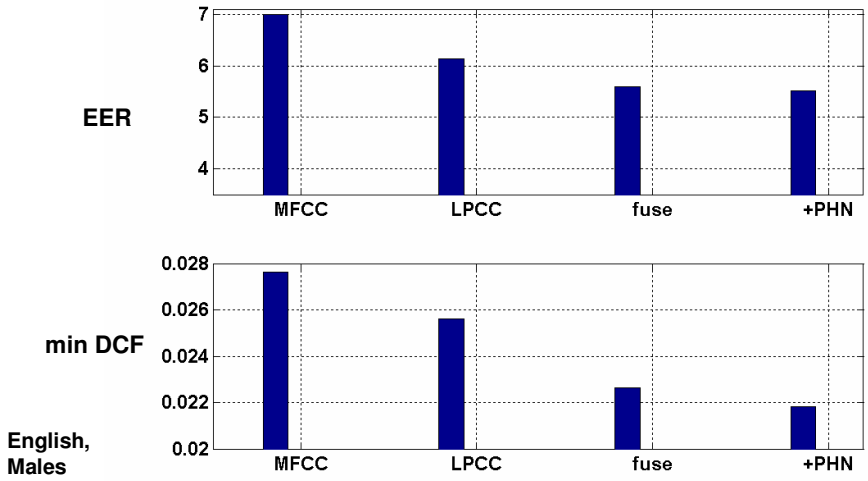
PerSay

Classifier fusion

- SGM MFCC + SGM LPCC: **linear** — all tasks
- SGMs + GMM: **linear** — 10sec-10sec
- SGMs + PHN: **SVM** — 1conv-1conv

- Training: NIST05 split

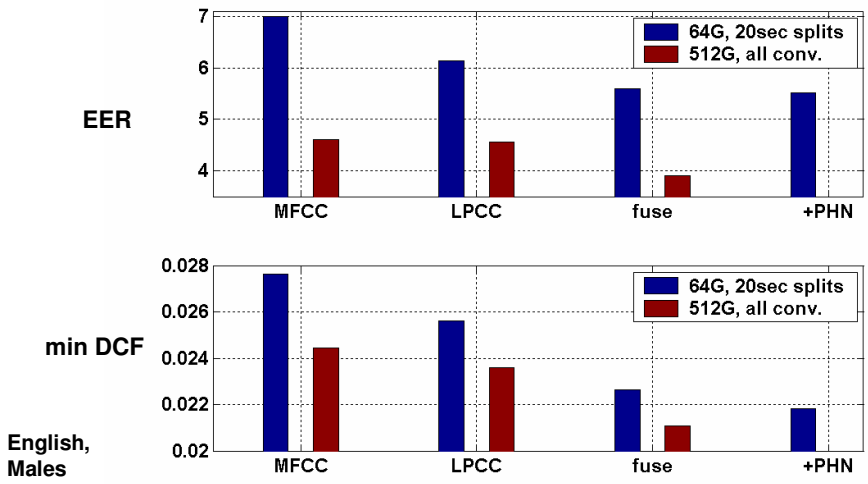
1conv – 1conv Common condition



PerSay

11

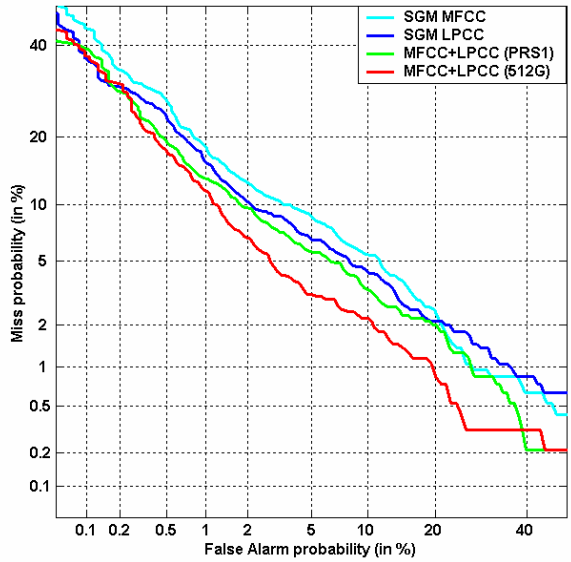
1conv – 1conv Common condition



PerSay

12

Fusion results

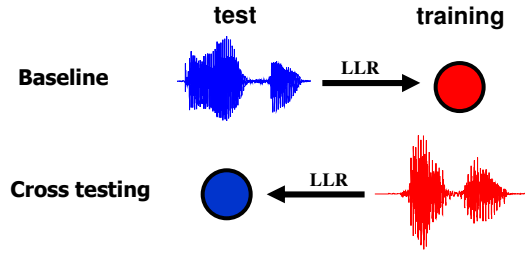


1conv4w-1conv4w,
English, Males

PerSay

GMM system

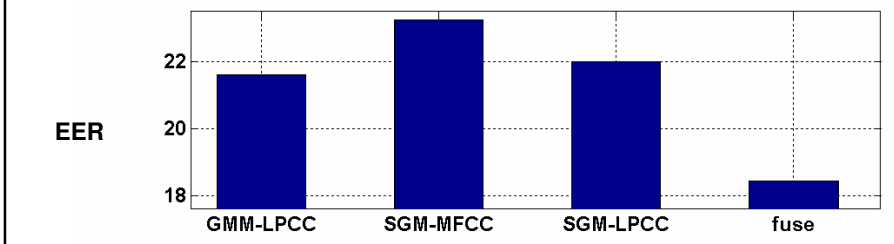
- Channel & gender dependent background models
- 256 Gaussians
- Bayesian adaptation from best matching BM
- Cross – testing:

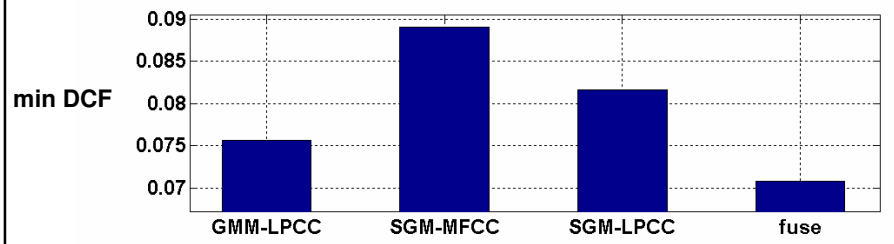


PerSay

14

10sec – 10sec





PerSay

Inter site fusion

Data & configuration:

- **38 systems**
- **Training on NIST05, test on NIST06**
- **1conv4w-1conv4w, all trials, key version 0**
- **Using raw scores (attempts to normalize scores didn't yield significant improvements)**

Algorithms:

1. **Logistic regression using sequential and parallel steps,** Collins et. al., Machine Learning '01
2. **Online Passive-Aggressive,** Crammer et. al., NIPS03
3. **Logistic regression,** Niko's Focal toolkit
4. **SVM light,** Joachims

PerSay

Inter site fusion Results

DCF	EER	algorithm
0.263	4.9	Best system on test set
0.179	**3.0**	**Logistic regression- using sequential and parallel steps, used only 22 systems**
0.199	3.5	Passive aggressive
0.188	3.3	Logistic regression (Niko's Focal)
0.196	3.3	Svm light, kernel degree=1
0.198	3.4	Svm light, kernel degree=2
0.197	3.5	Svm light, kernel degree=3

PerSay

Conclusion

What we had learned:

- **Using 512 Gaussians, without splitting the audio is better**
- **Fusion of LPCC system with MFCC system decreases error rates**
- **The phonetic system is still in early stages, and needs further improvement**

PerSay