Speaker Verification using i-Vectors

CAST-Förderpreis IT-Sicherheit 2014

Andreas Nautsch

Hochschule Darmstadt, atip GmbH, CASED, da/sec Security Research Group

Darmstadt, 20.11.2014

atip
advanced technologies for information processing

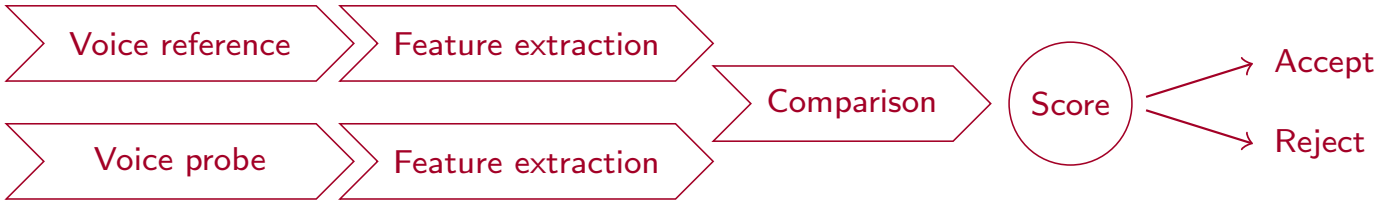
h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES
fbi
FACULTY OF COMPUTER SCIENCE

Outline

- Motivation with research questions
- Speaker verification and i-vectors
- Research and development
- Conclusion and future perspectives

Biometric IT-security & forensic applications

- Authentication and recognition by voice
- Advantages to knowledge-/token-based approaches:
  - Cannot be forgotten
  - Cannot be incorporated
- Application fields and scenarios e.g.:
  - Mobile device authentication: random PINs, short duration
  - Call-center user validation: free speech, variant duration
  - Suspect tracking: various contents & signal qualities

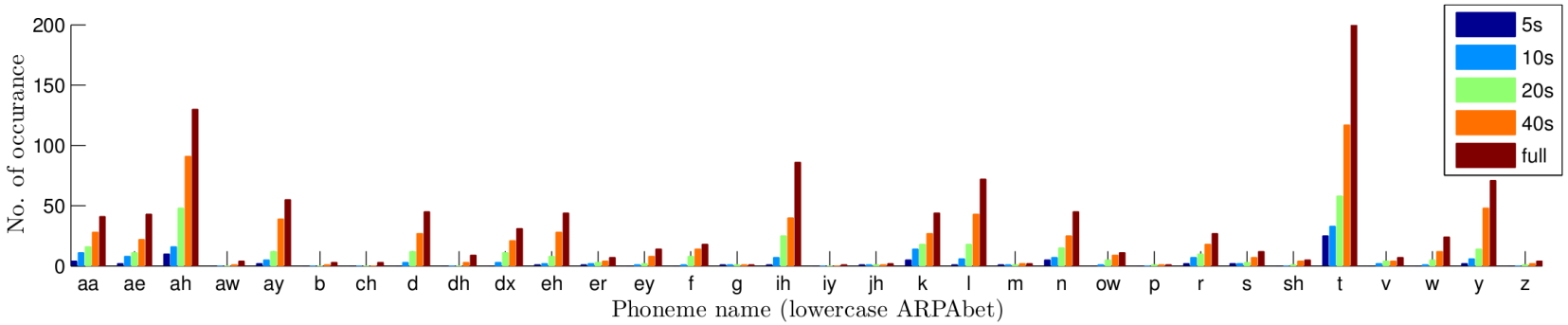


Assessment of Speaker Recognition

- Known technology: modeling acoustic features
  - ✓ Very accurate due to detailed modeling
  - ✓ Fast processing on short duration scenarios
  - ✗ High computational effort on text-independent scenarios
- State-of-the-Art: identity vector (i-vector) features, 2011
  - ✓ Fully text- & language-independent
  - ✓ Fast computation & scoring independent of duration
  - ✗ Unknown behavior in commercial voice biometric scenarios

Voice biometrics: speech duration & sample completeness

- Sound unit (phoneme) distribution by duration

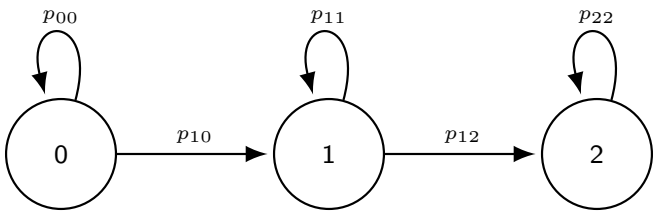


From [T. Hasan et al., 2013]

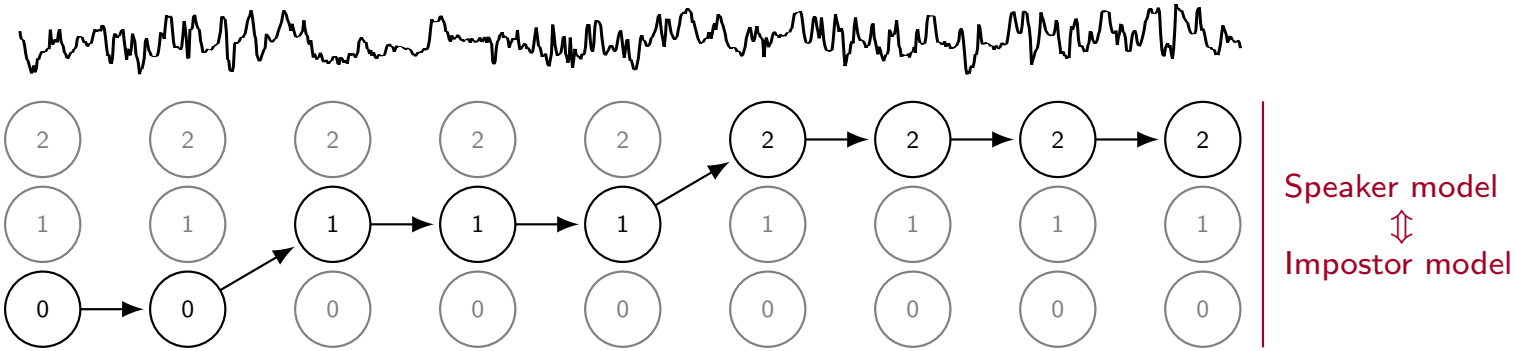
- Text-independent case: content varies from sample to sample
  - Long duration ⇒ stable distribution
  - Short duration ⇒ insufficient data

Baseline speaker recognition approach

- Hidden-Markov-Models (HMMs)
  - State-based model
  - States can represent articulation phases, phonemes, ...



- Detailed, but extensive computation of optimal path

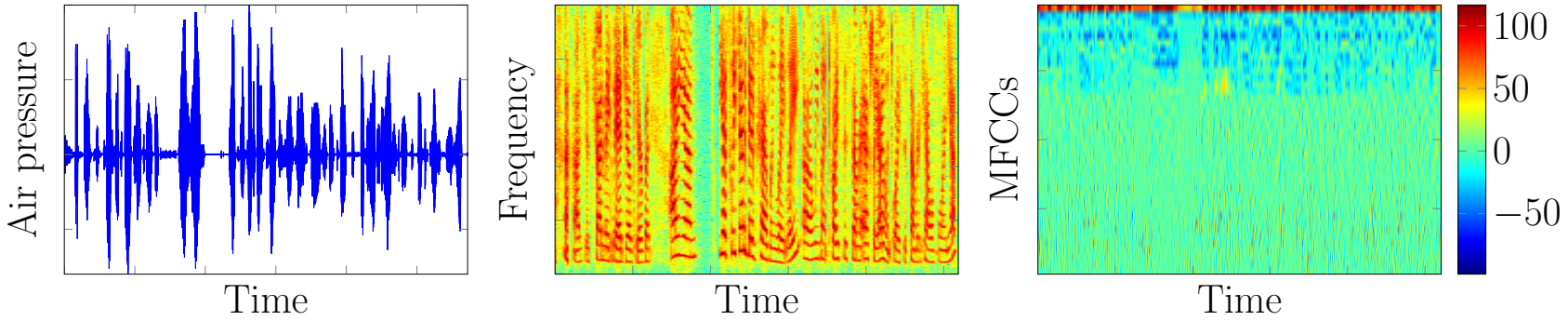


Research questions

1. Is the i-vector approach extensible on short duration scenarios with applicable performances?
2. Do i-vector systems deliver new information to HMM systems?
3. Are duration-depending performance mismatches compensable?

Speech processing

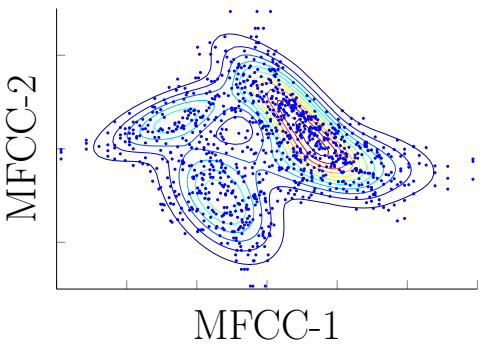
1. Raw speech signal as air pressure changes
2. Frequency analysis: spectral representation
3. Short-time acoustic features, e.g. Mel-Frequency Cepstral Coefficients (MFCCs)

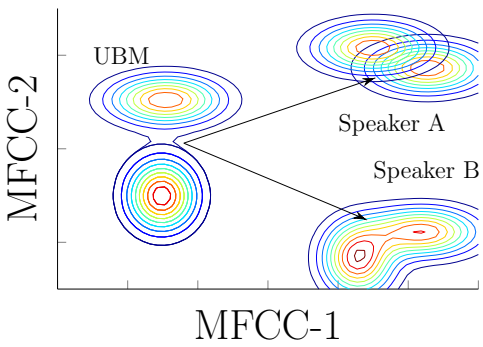


Statistic model: i-vector extraction

4. Gaussian modeling
5. Speaker sub-spaces from Universal Background Model (UBM)
6. *identity vectors* (i-vectors) as characteristic offset

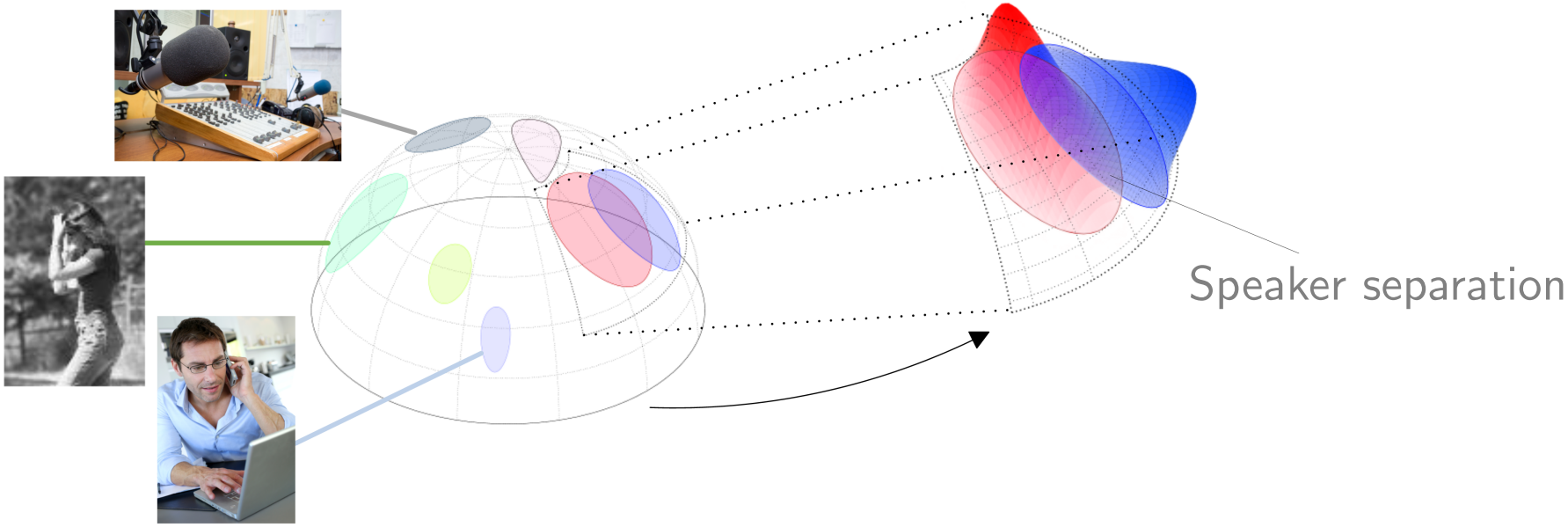
   Mapping by *total variability matrix**





* iteratively in order to optimize the model fit: UBM offset $\mapsto$ i-vectors on development data

How to imagine i-vectors?



- Relevant parameters
  - UBM size: detail of the acoustic space
  - # iterations: adaptation depth of total variability
  - # characteristic factors: i-vector dimension

Extending the baseline, exploring new technologies

- Research on new technologies
  - Experimental evaluations on i-vectors
  - Commercial and academic scenarios
  - Participation in international research evaluation

- Developing more robust approaches
  - Extending state-of-the-art i-vector score normalization
  - Implementation of speaker verification framework in Matlab according to ISO/IEC IS 19795-1
    Information technology – Biometric performance testing and reporting – Part 1: Principles and framework $\Rightarrow$ reproducible research

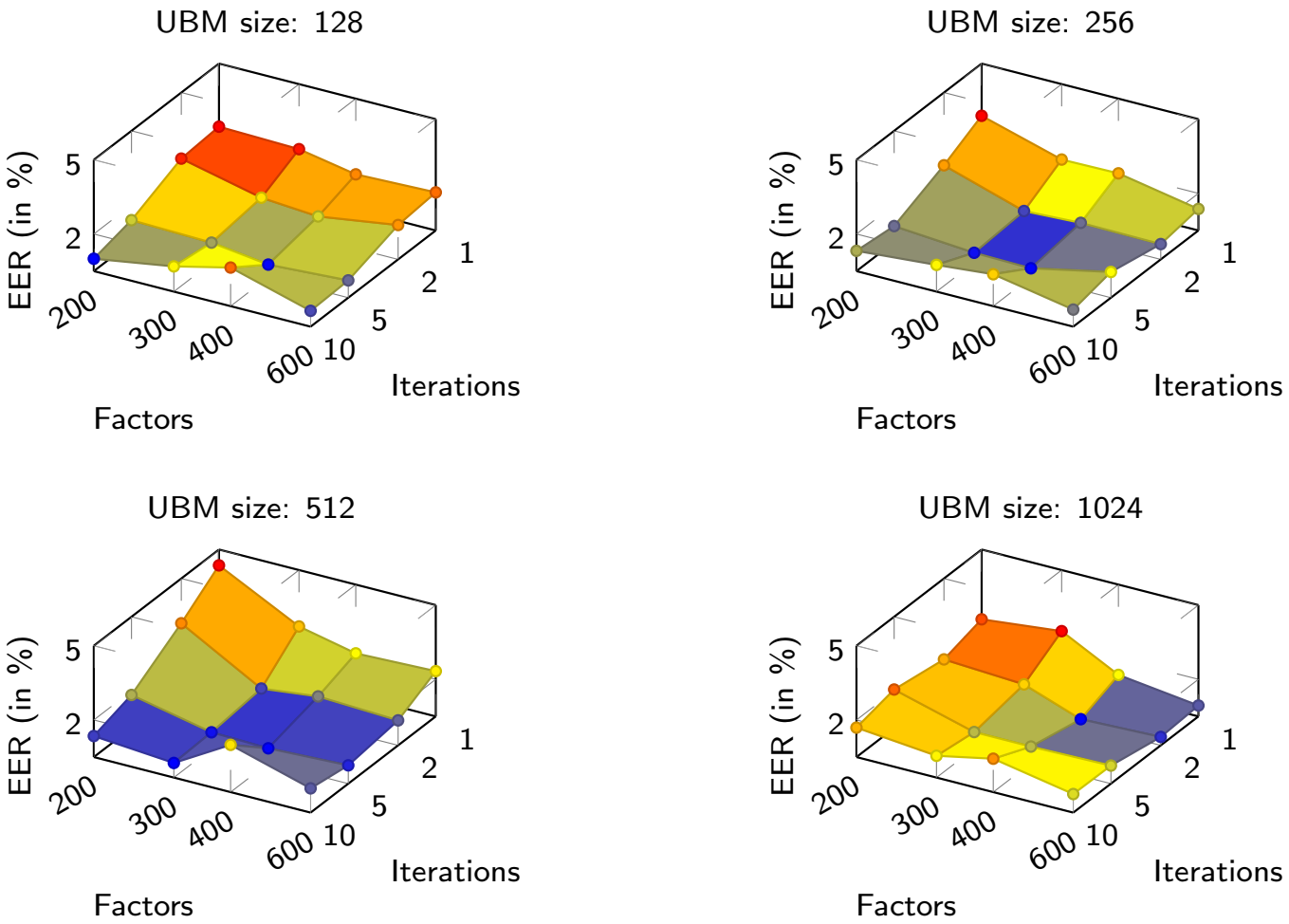
Commercial scenario: experimental set-up

- Aiming at research questions
  1. i-vectors on short duration scenarios?
  2. New information to HMMs by i-vectors?

- Short but fix duration scenario
  - In-house database: 3 – 5 German digits
  - Text-independent: random sequences
  - 362 / 56 / 300 subjects (development, calibration, evaluation)
  - 30 – 34 reference / 2 probe samples

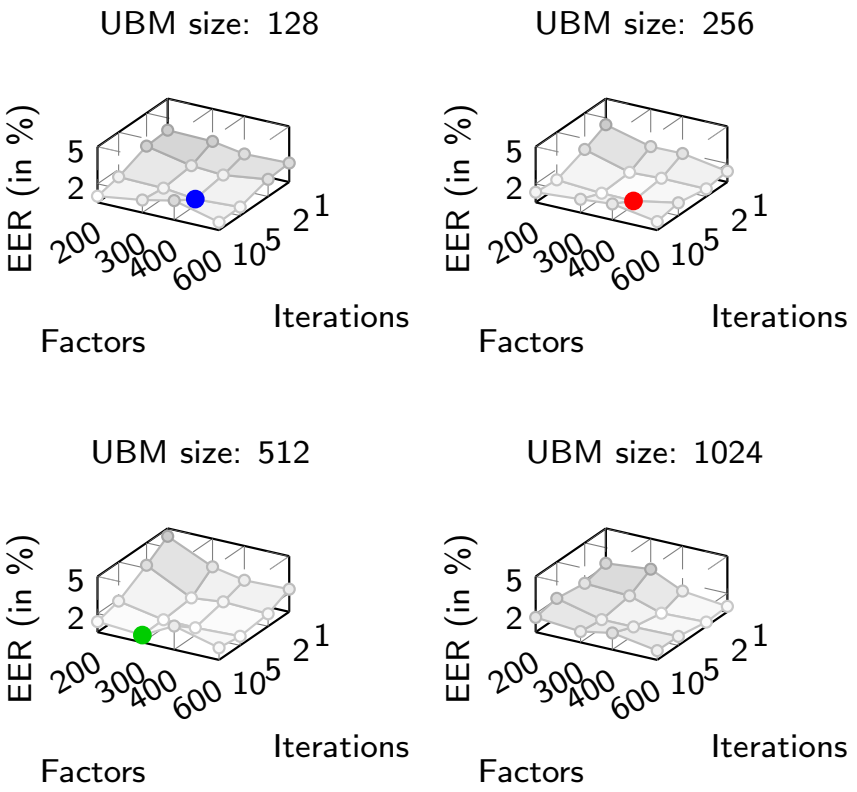
    $\approx$ 200,000 comparisons

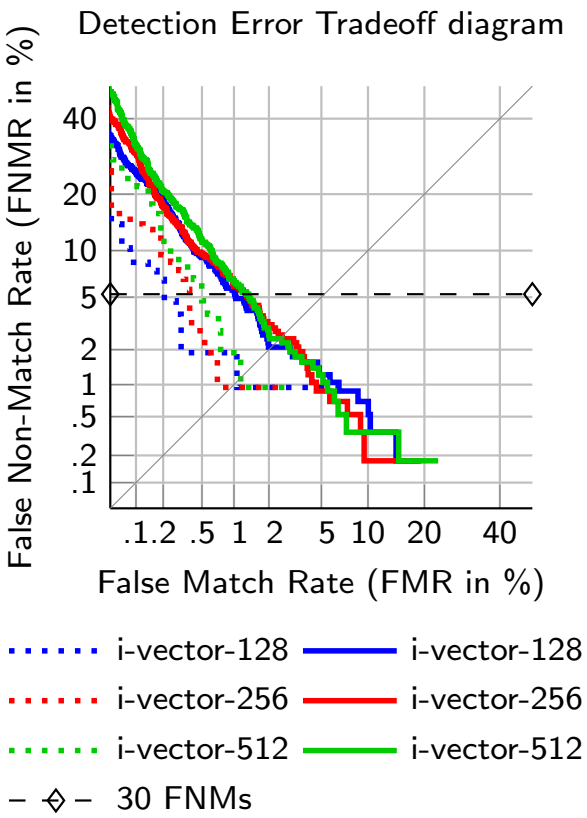
Examining i-vector parameters



Equal Error Rate (EER): % impostor match = % genuine non-match → lower error = better

Performance analysis

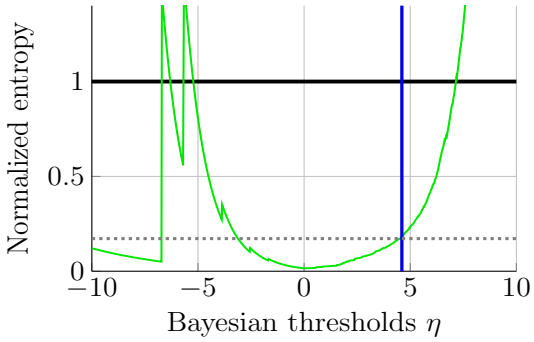




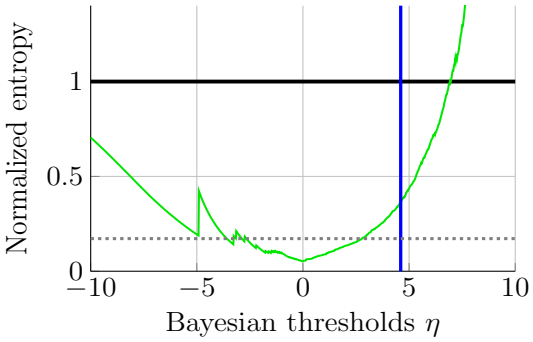
Equal Error Rate (EER): % impostor match = % genuine non-match → lower error = better

Information analysis: cross-entropy of system fusion

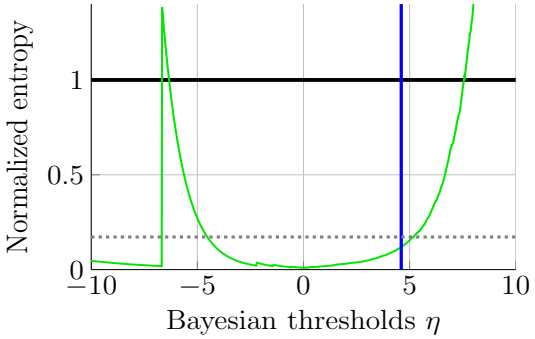
HMM



i-vector-256

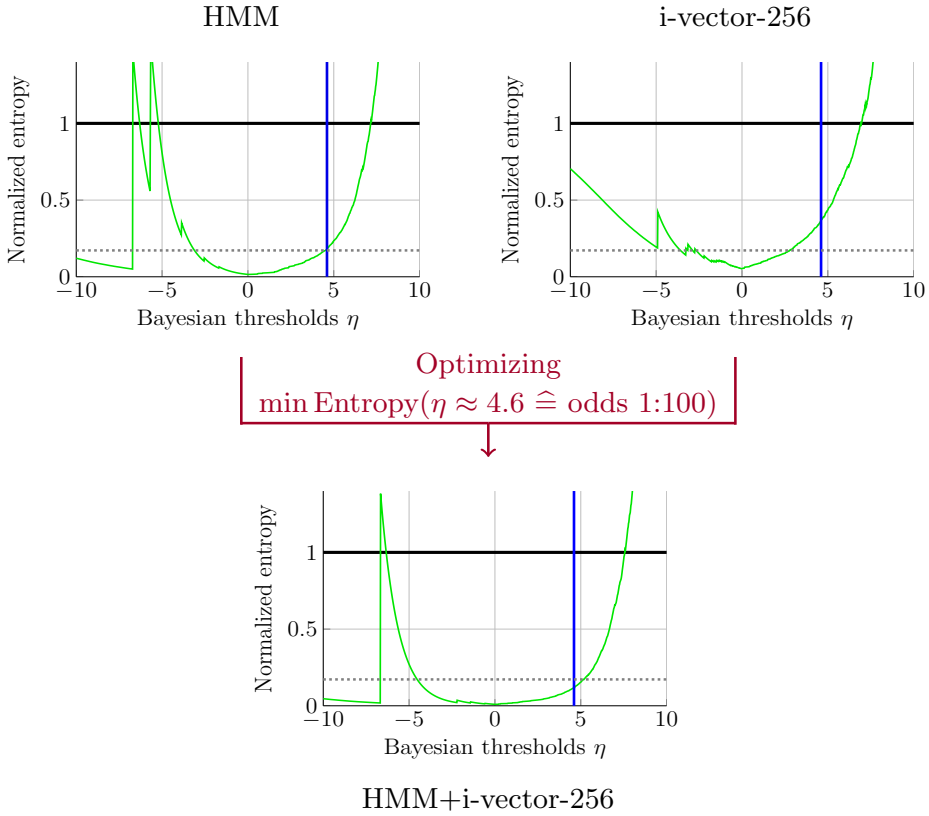


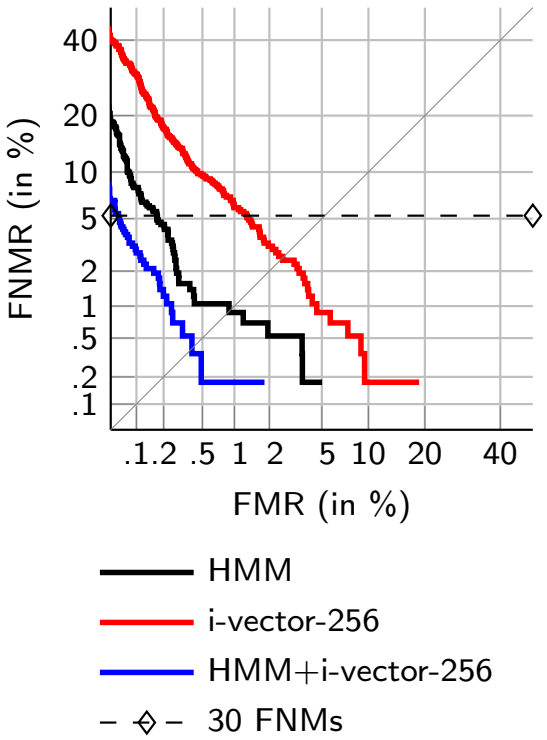
Optimizing
$\min \text{Entropy}(\eta \approx 4.6 \mathrel{\widehat{=}} \text{odds } 1{:}100)$



HMM+i-vector-256

Performance gains by fusion





Academic scenario: experimental set-up

- Aiming at research question
  3. Compensation of duration mismatches on i-vectors?

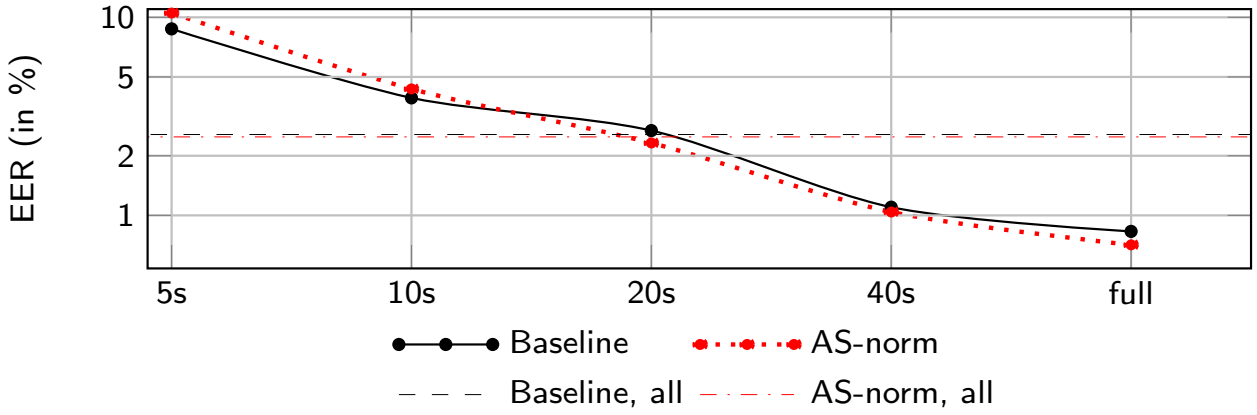
- Variable duration scenario
  - Data of 2013 – 2014 NIST i-vector Machine Learning challenge
  - Text-independent, multi-lingual scenario
  - 4,781 / 1,306 subjects (development, evaluation)
  - 5 reference samples / 9,634 probe i-vectors

    $\approx$ 12,000,000 comparisons

Analysis of a variant duration scenario

- 10 offline 5-fold cross-validations
- NIST baseline system
  2048-component UBM, 60 MFCCs, 600-dim i-vectors
- Adaptive Symmetric (AS) score-normalization

$$S' = \frac{1}{2}\left(\frac{S-\mu_{\text{reference}}}{\sigma_{\text{reference}}} + \frac{S-\mu_{\text{probe}}}{\sigma_{\text{probe}}}\right),$$ each from top-100 scores of comparisons to dev-set



Proposing duration-invariant AS score-normalization

- Observation: independent i-vector factors ⇒ poor comparisons

dev-set	full	40s	20s	10s	5s
5s	141	91	66	**44**	**0**
10s	230	140	70	**0**	
20s	246	118	**0**		
40s	180	**0**			
full	**0**				

(Student t-test)

- Idea:
  - On probes: dev-set duration > 60s
  - On references: dev-set duration $\approx$ probe duration

System	EER (in %)	all	5s	10s	20s	40s	full
Baseline	2.56	0.89	0.95	0.93	0.92	0.89	0.86
AS-norm	2.49	0.17	1.18	0.41	0.18	0.08	**0.05**
dAS-norm	**2.06**	**0.10**	**0.35**	**0.20**	**0.11**	**0.07**	0.07

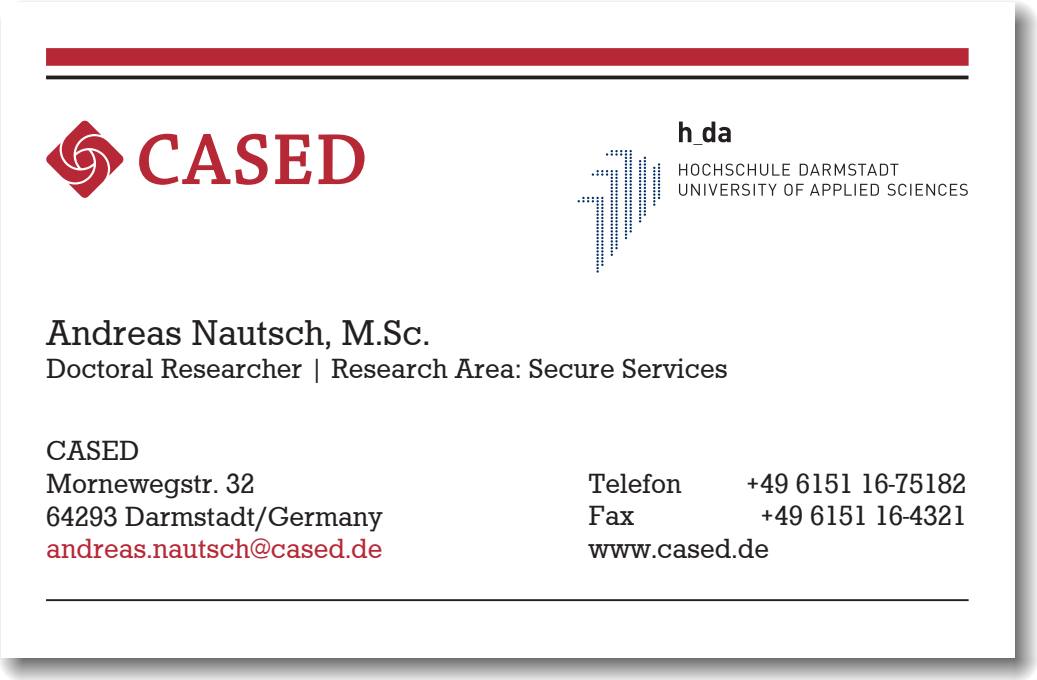
Conclusion and future perspectives

- Research questions positively confirmed
  1. i-vectors are applicable to short duration scenarios
  2. Relative cross-entropy gain: 48% to HMM baseline
  3. Performance mismatch compensation: 89% to baseline

- Examining probabilistic i-vector comparators e.g., Probabilistic Linear Discriminant Analysis (PLDA)
- Analyse effects on other speech signal features

Community impact

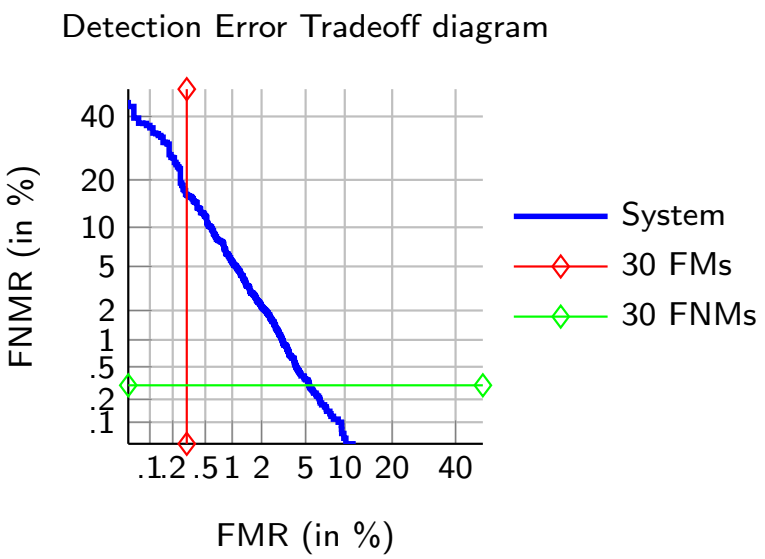
- XI IAPR/IEEE Int.l Summer school for advanced studies on biometrics: Biometrics for Forensics, Security and beyond, Alghero, Italy, June 9 – 13, 2014
  Presentation to well-established biometric researchers
- ISCA Odyssey 2014: The Speaker and Language Recognition Workshop, Joensuu, Finland, June 16 – 19, 2014
  A. Nautsch, C. Rathgeb, C. Busch, H. Reininger, and K. Kasper: *Towards Duration Invariance of i-Vector-based Adaptive Score Normalization*, ISCA Odyssey, 2014.

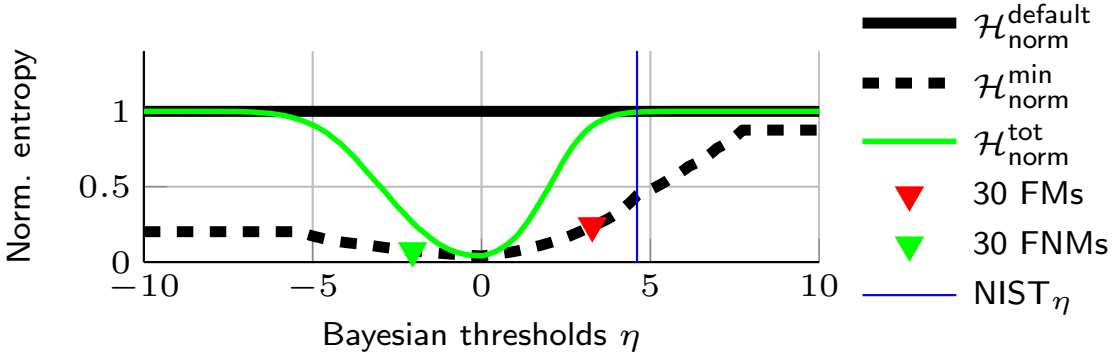


Acknowledgment: atip GmbH, associated company during B.Sc. and M.Sc. studies, for founding and education.

System parameters & metrics

- Variable parameters
  - UBM size
  - # i-vector factors
  - # training interations of *total variability matrix*
- Performance metrics
  - Equal Error Rate (EER)
  - False Match Rate (FMR)
  - False Non-Match Rate (FNMR)
  - Score-cross entropy





Real-time evaluation

Table: Enroll/Verify
HMM ⇔ i-vector

System	Enroll	Verify
HMM	206.2%	5.5%
i-vector-128	6.1%	3.2%
i-vector-256	9.9%	3.1%
i-vector-512	16.2%	3.3%

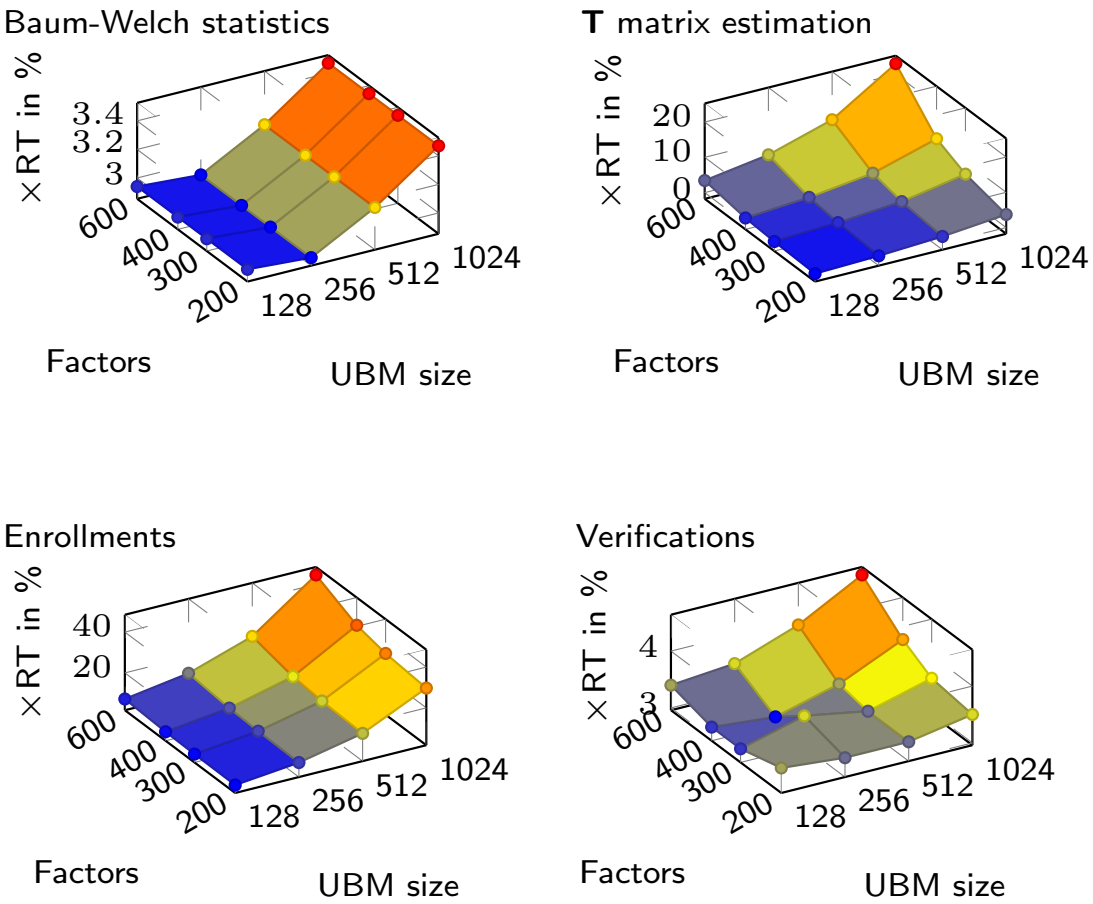
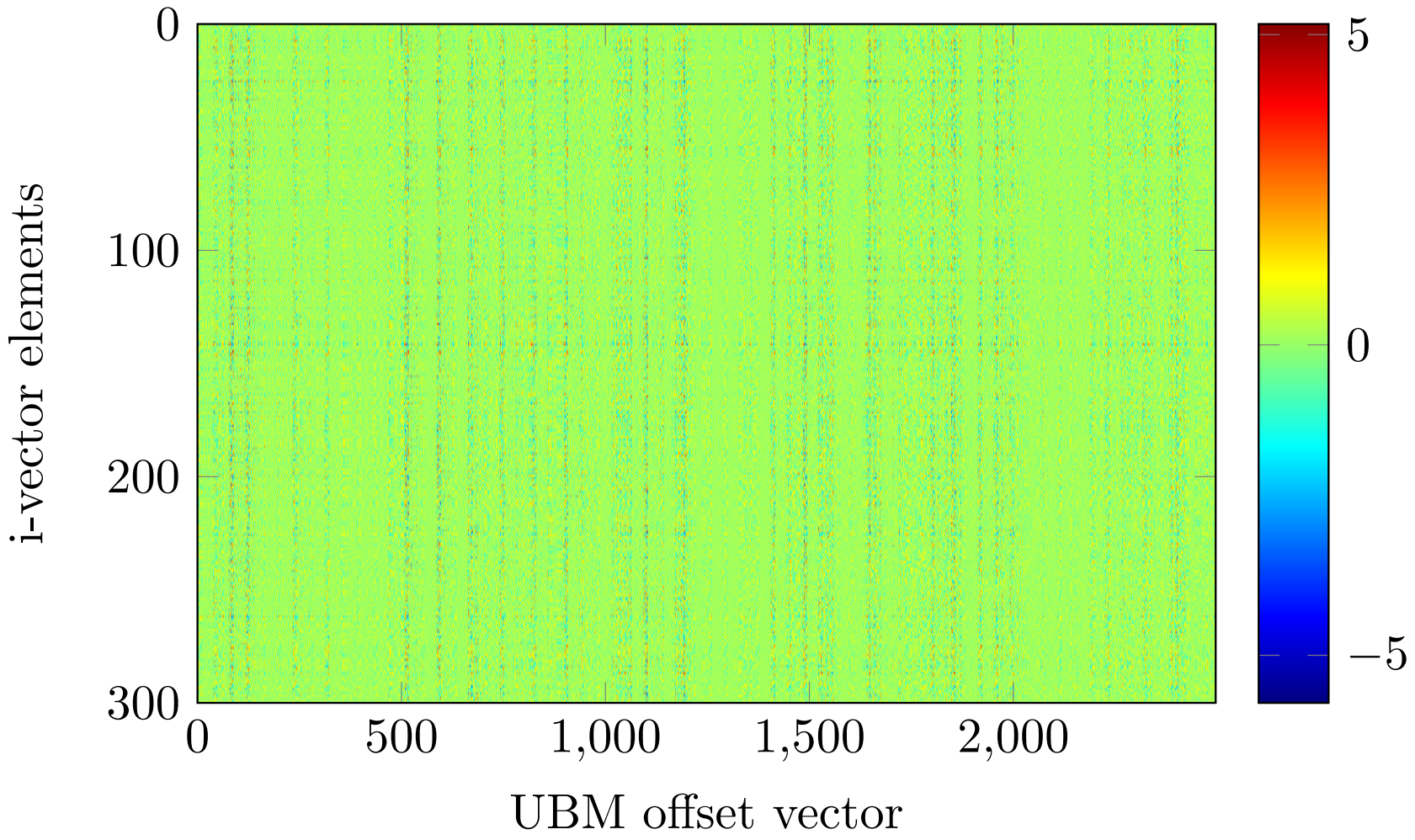
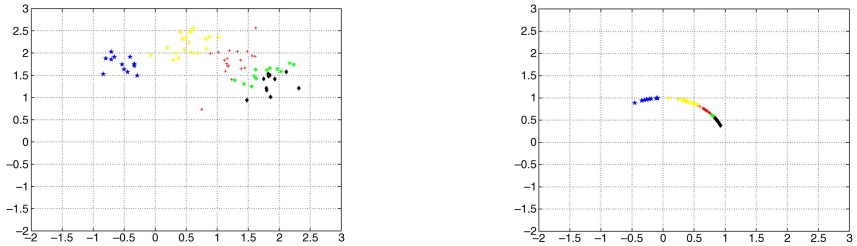


Illustration of total variability matrix



Analyzing i-vector processing steps

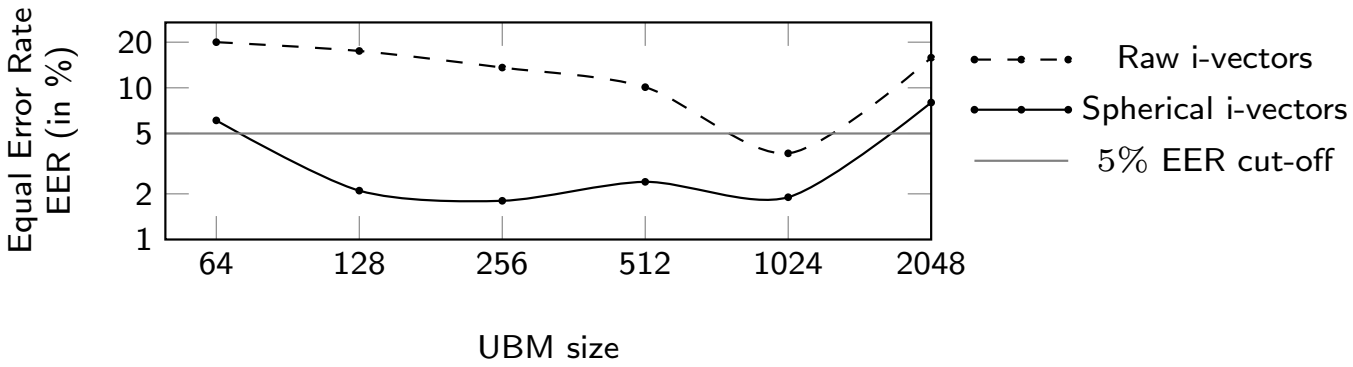
- State-of-the-Art: spherical space projection



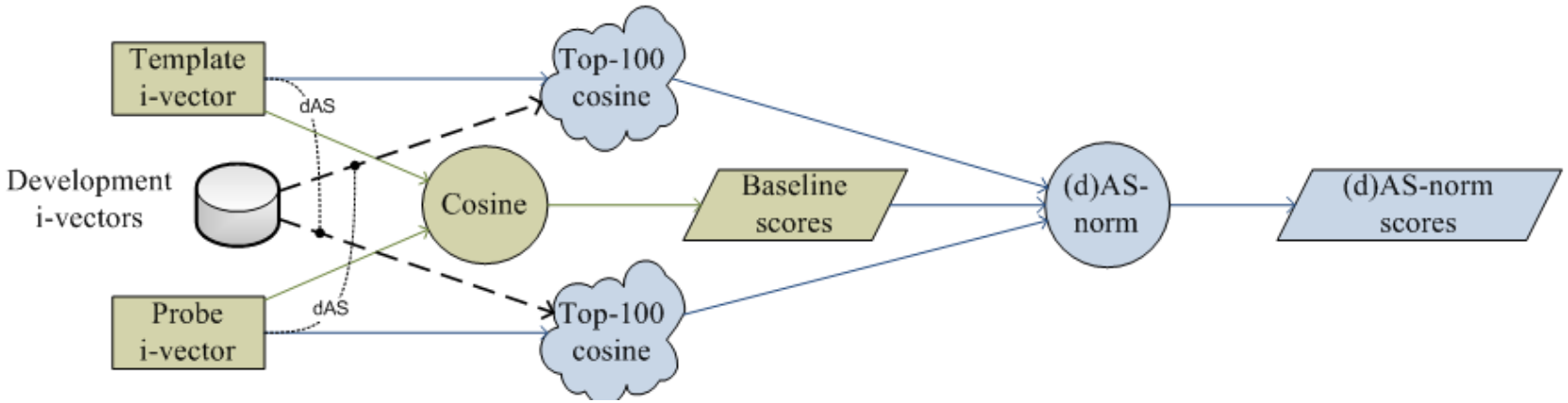
Raw i-vectors $\mapsto$ Spherical i-vectors

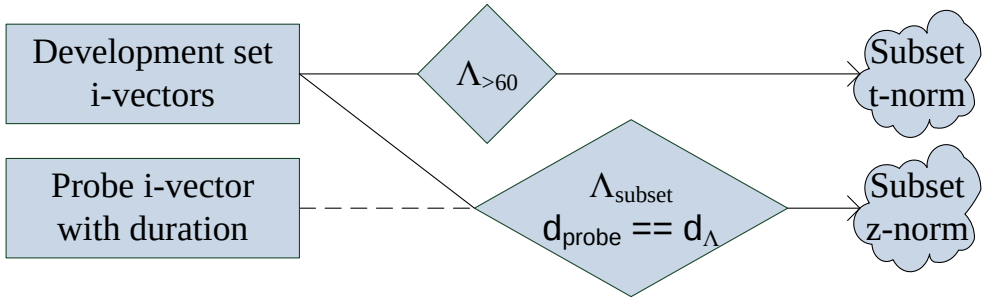
From [N. Dehak et al., 2011]

- On short-duration scenario (calibration set)

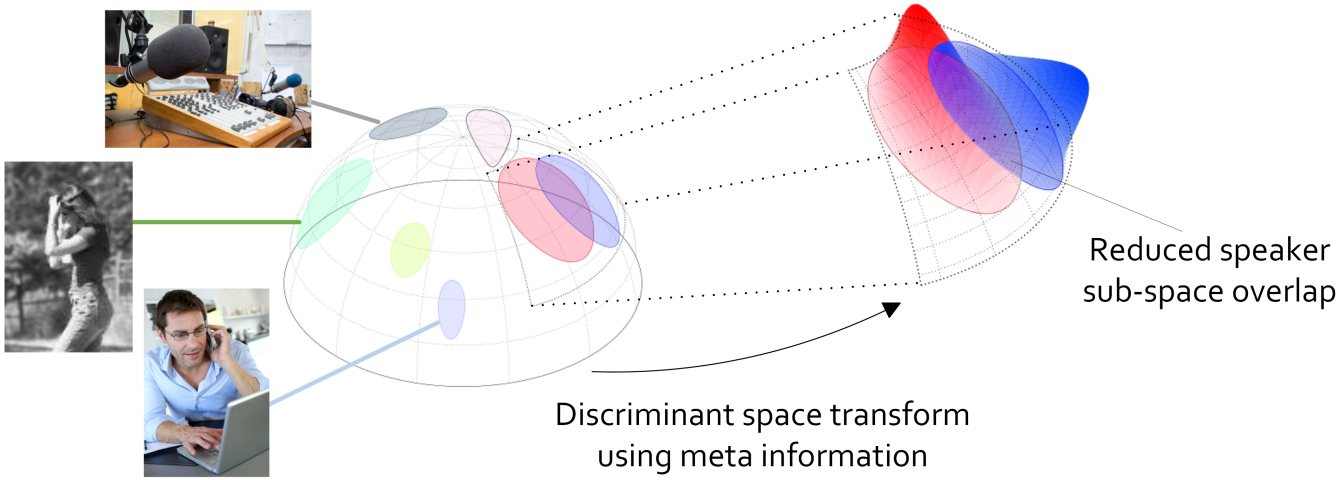


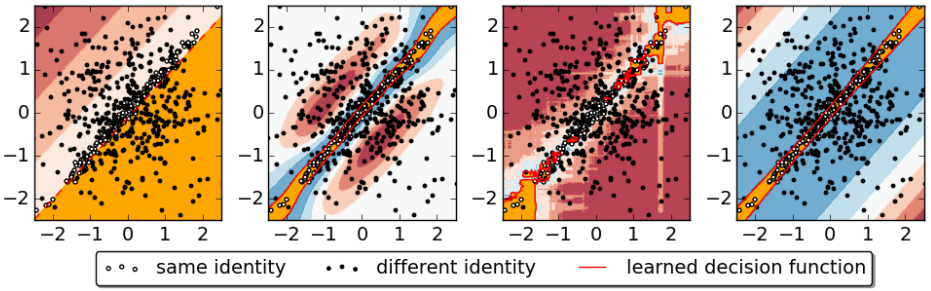
Duration-invariant Adaptive Symmetric score normalization





Probabilistic Linear Discriminant Analysis (PLDA)





Evaluation framework in Matlab

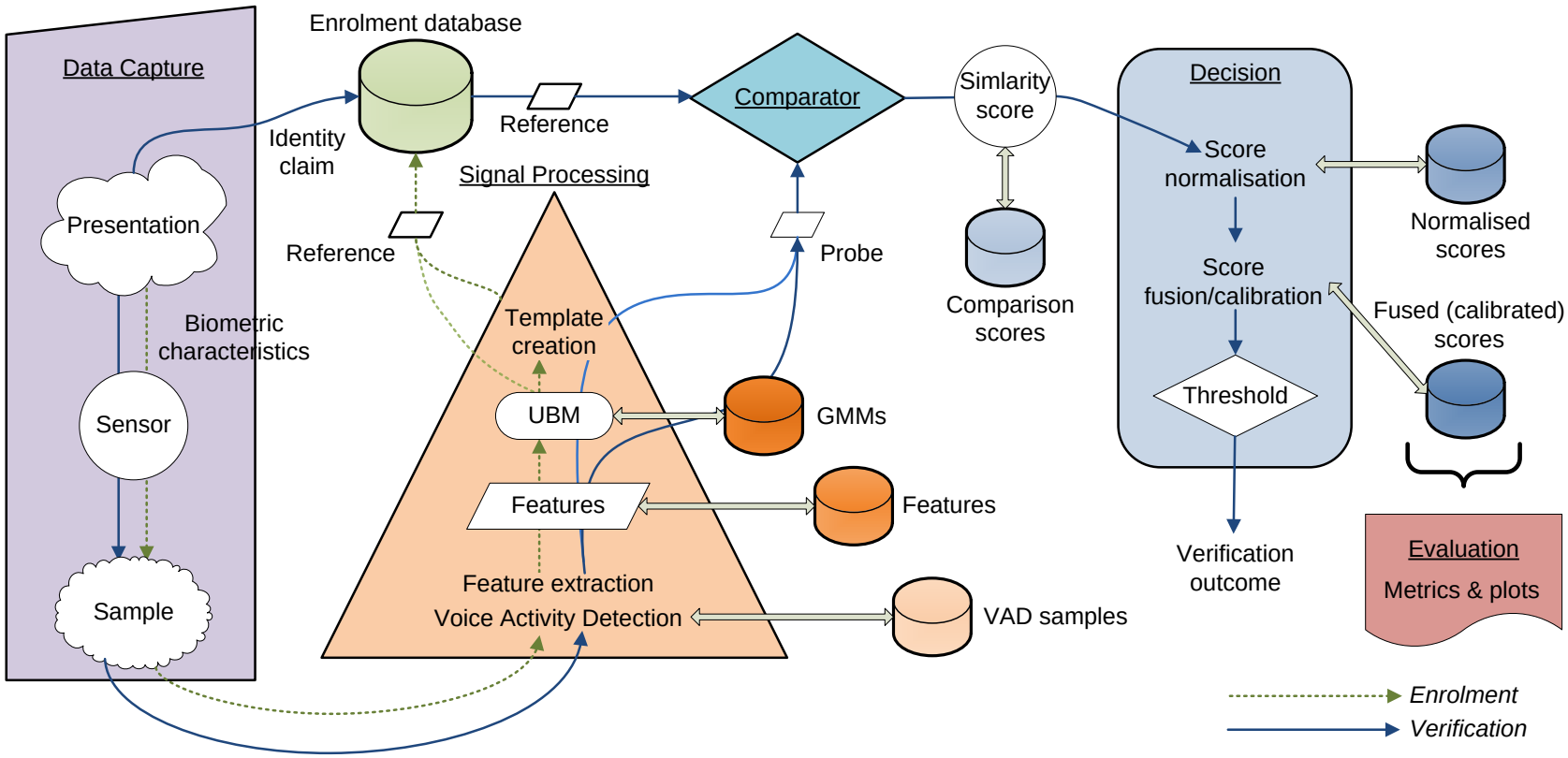


Figure: Framework design according to ISO/IEC 19795-1

Evaluation framework in Matlab

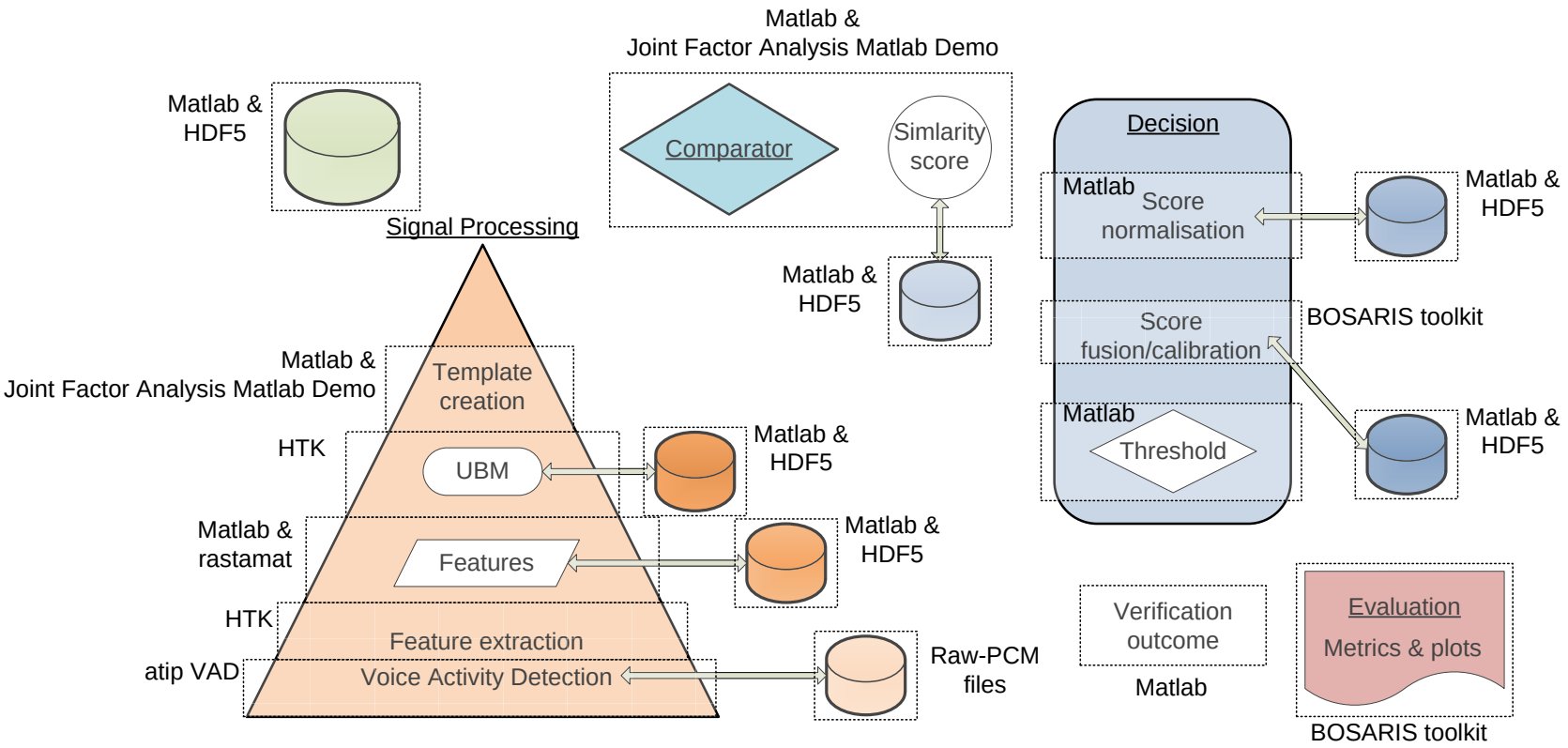


Figure: Framework & Toolboxes