Entropy Analysis of i-vector Feature Spaces in Duration-sensitive Speaker Recognition

da/sec BIOMETRICS AND INTERNET-SECURITY RESEARCH GROUP

Aalto University

h_da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES

CASED

ANDREAS NAUTSCH* CHRISTIAN RATHGEB* RAHIM SAEIDI† CHRISTOPH BUSCH*

* da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany

† Department of Signal Processing and Acoustics, Aalto University, Finland

andreas.nautsch@{cased|h-da}.de

Motivation

Research Questions

Q1: How do speaker sub-spaces accumulate by increasing sample duration?

Q2: Can the biometric disrimination potential of i-vector feature spaces be measured?

While on score level cross-entropy analyses examine scalar speaker separation, feature space analyses place emphasis on e.g., i-vector distributions. By measuring the expected regularized Kullback-Leibler divergence between subjects, sub-space information is derived regarding:

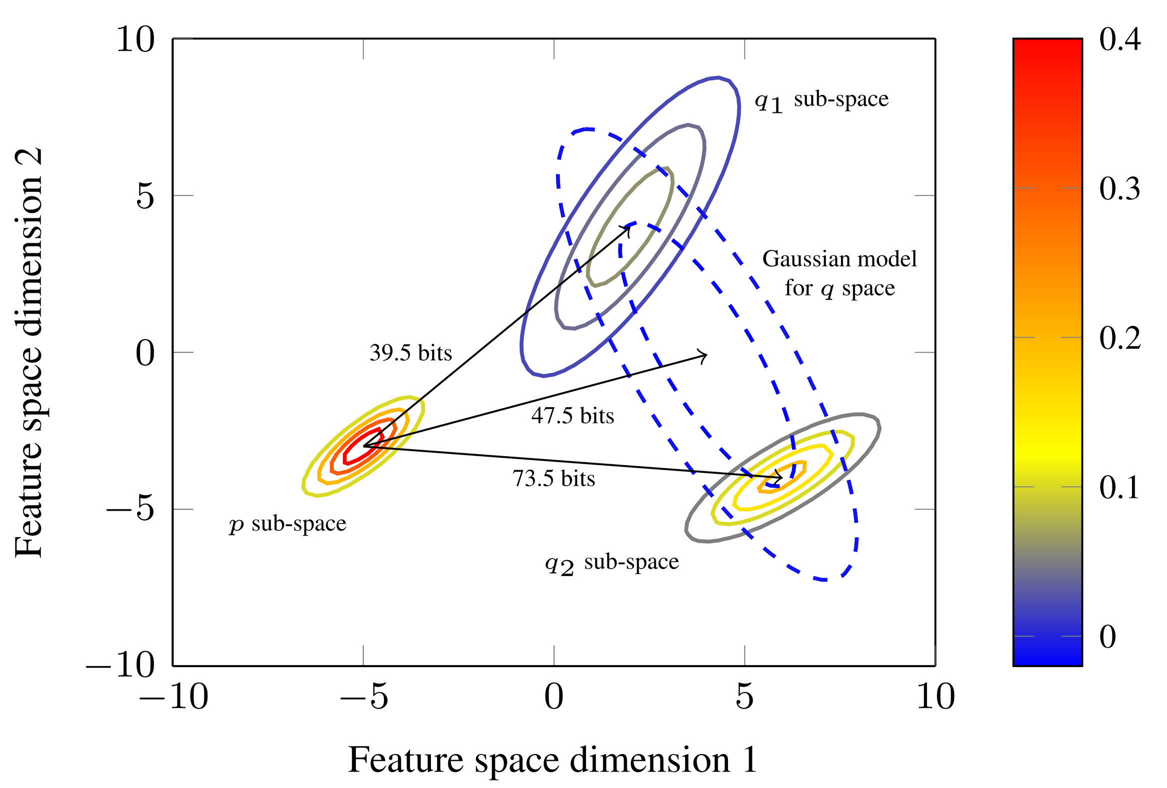
- Collision probability of subjects, and
- Biometric uniqueness in given feature spaces

Feature Space Entropy

Relative Entropy

Biometric information addresses inter-subject distances of features $\mathbf{x}$: measuring the Kullback-Leibler divergence of the intra-subject disctribution $p(\mathbf{x})$ and the inter-subject distribution $q(\mathbf{x})$

Lower Bound by single Gaussians



Estimating KL-divergence

Conventional Approach

Assuming, p and q follow a Gaussian distribution with $p(\mathbf{x}) \sim \mathcal{N}(\vec{\mu}_p, \mathbf{\Sigma}_p)$ and $q(\mathbf{x}) \sim \mathcal{N}(\vec{\mu}_q, \mathbf{\Sigma}_q)$:

$$D(p||q) = k(\lambda + \mathit{trace}((\mathbf{\Sigma}_p + \mathbf{T})\mathbf{\Sigma}_q^{-1} - \mathcal{I})).$$

Where regularization is necessary, when feature spaces have more dimensions than samples are observed per subject.

Adler et al.: *Towards a Measure of Biometric Information*, IEEE CCECE, 2006.

1) Regularization for degenerated features

- PCA transform by q space scatter
- Truncate low eigenvalues $< 10^{-10}[S_q]_1$

2) Regularization for insufficient data

- Ill-disposed $\mathbf{\Sigma}_P$ lead to $D(p||q) \to \infty$
- Consider diag-$\mathbf{\Sigma}_P$ as well
- On N_p samples of subject p: non-diagonal $[\mathbf{\Sigma}_P]_{i,j} = 0, \quad i,j \geq N_p$

In our experiments, we extended 1 & 2 with:

3) Regularization for ill-conditioned PCA space

- Setting iteratively non-diagonal $[\mathbf{\Sigma}_P]_{i,j}$ to 0, until $|\mathbf{\Sigma}_P|$ becomes positive

4) Regularization for insufficient samples

- Minimum samples / subject: 10

Pros / Cons

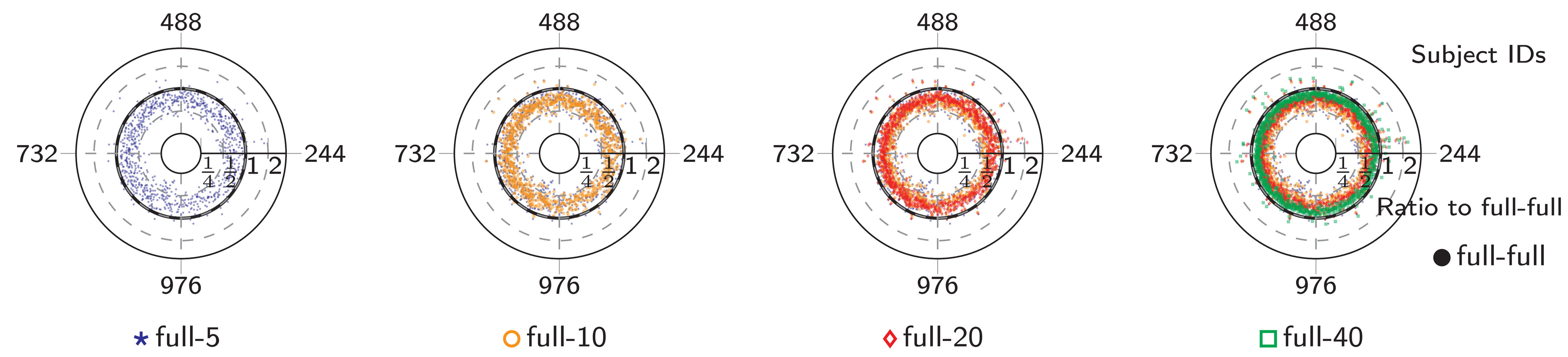
Metric: biometric uniqueness	⇔	gross regularization
Insight: space behavior	⇔	lower bound metric
Perspective: 2-class	⇔	subject-to-subject

Results on I4U NIST SRE'12 train list

Experimental Set-Up

- Pooled male/female I4U i-vectors
- Pooled dev/eval files
- 551 female & 425 male subjects (min. 10 samples)
- 400-dim i-vectors
- Processing: whitening & length-norm
- Analyses regarding 5s, 10s, 20s, 40s, full samples

Speaker sub-space accumulation

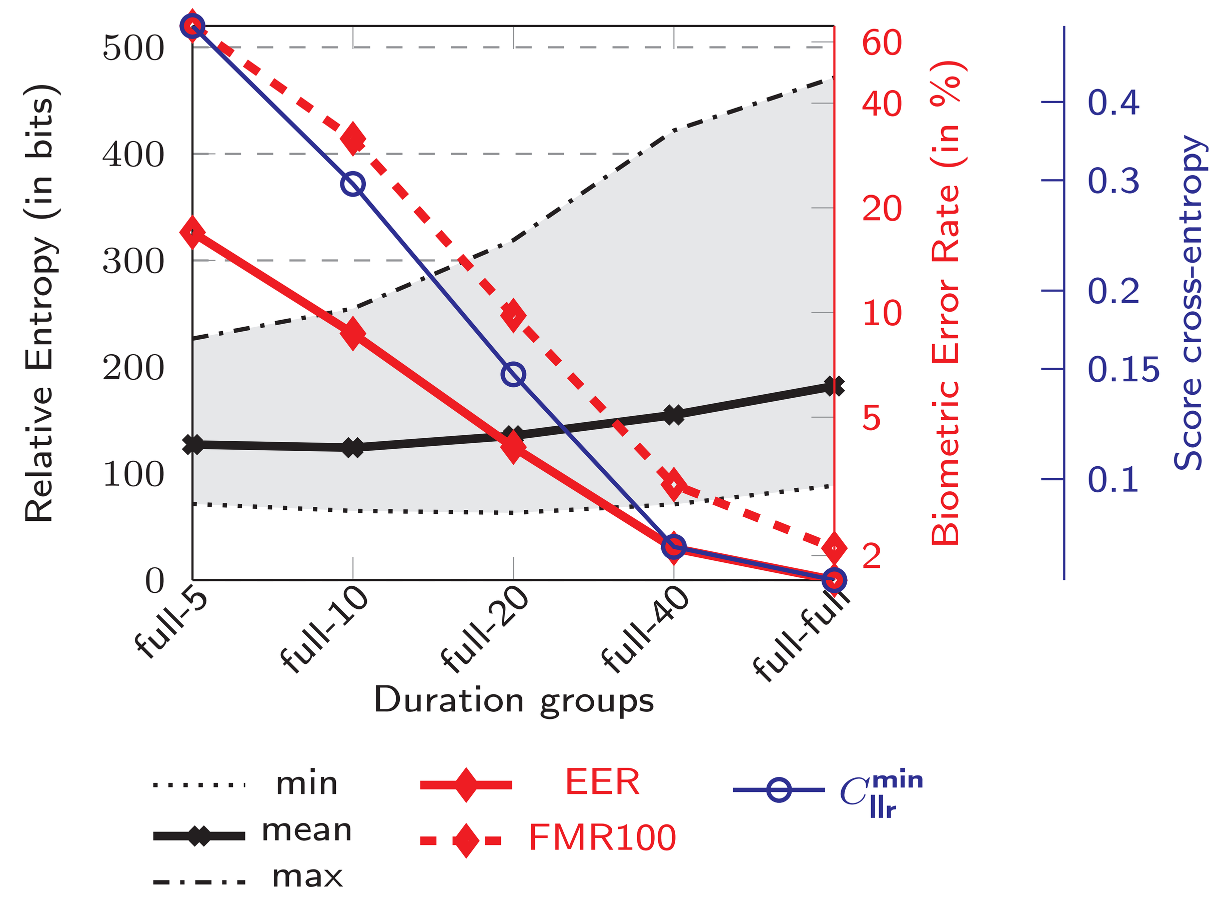


Expected Relative Entropy

Duration group	Entropy (in bits) μ	σ	min	max
full-5	127.2	24.0	71.5	226.6
full-10	124.3	28.1	65.0	254.8
full-20	135.5	35.3	63.2	319.0
full-40	155.0	43.1	71.1	421.9
full-full	182.1	50.0	88.7	471.6

- Collision probabilities diminish by duration: 5×10^{-10} on short duration to 2×10^{-55} on full
- Cross-Modality Comparisons on Feature-Level: 2D-Face (56 bit), Fingerprint (85 bit), Iris (249 bit), and Password (128 bit)

Comparison to performance metrics PLDA(400)



Conclusion

Speaker Recognition is viable for biometric recognition in either forensic or commercial applications. The i-vector feature space conducts comparatively high subject discrimination, where current classifiers need more attention towards robustness against probe sample duration variance.