Entropy Analysis in Speaker Recognition

Andreas Nautsch

Hochschule Darmstadt, CASED, da/sec Security Research Group

Berlin, 15.12.2015

Outline

1. Motivation
2. Overview on Speaker Recognition
3. Biometric Strength of State-of-the-Art Voice Features
4. Conclusion

Motivation: Different Approach towards Entropy

- Established: Goodness of (LLR) scores
  Focus: scores values ⇔ expected meaning?

- Proposed: metric for the strength of biometric features
  - Collision probability of subjects within feature spaces
  - Metric towards *biometric uniqueness*
  - Comparability to other modalities on early stages
    Face: 56 bit
    Fingerprint: 84 bit
    Iris: 249 bit

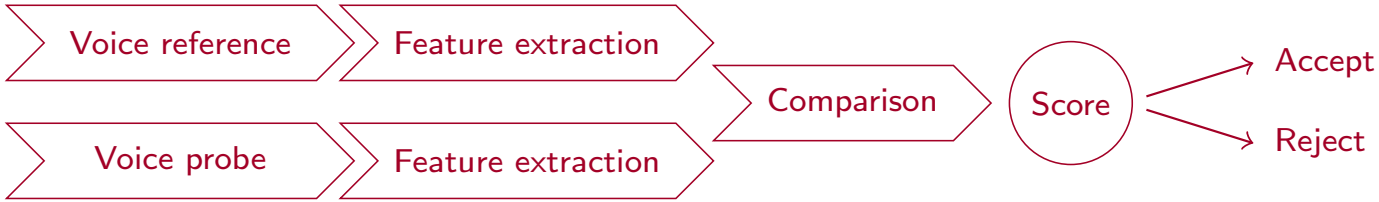
[Buchmann+14] N. Buchmann, C. Rathgeb, H. Baier, C. Busch: *Towards electronic identification and trusted services for biometric authenticated transactions in the Single Euro Payments Area*, APF'14, 2014

Overview on Speaker Recognition

- Voice as biometric characteristic



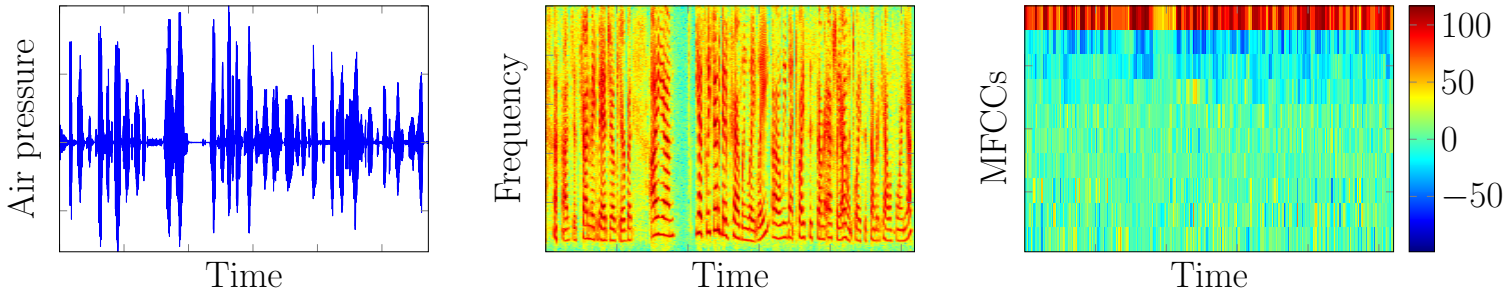

- Application scenarios and challenges (brief excerpt)
  - Call-Center fraud prevention:
    natural free speech, variable duration and content
  - Mobile devices: random PINs, short duration
  - Forensic: various contents and signal qualities



Overview on Speaker Recognition

1. Psycho-acoustic spectral analyses

⇒ 60 Melody-Frequency Cepstral Coefficients (MFCCs)



2. MFCC Clustering by Gaussian Mixture Models (GMMs)

⇒ $2048 \times 60 = 122\,880$ free parameters per voice sample

3. Total Variability Analysis: intermediate-sized vectors

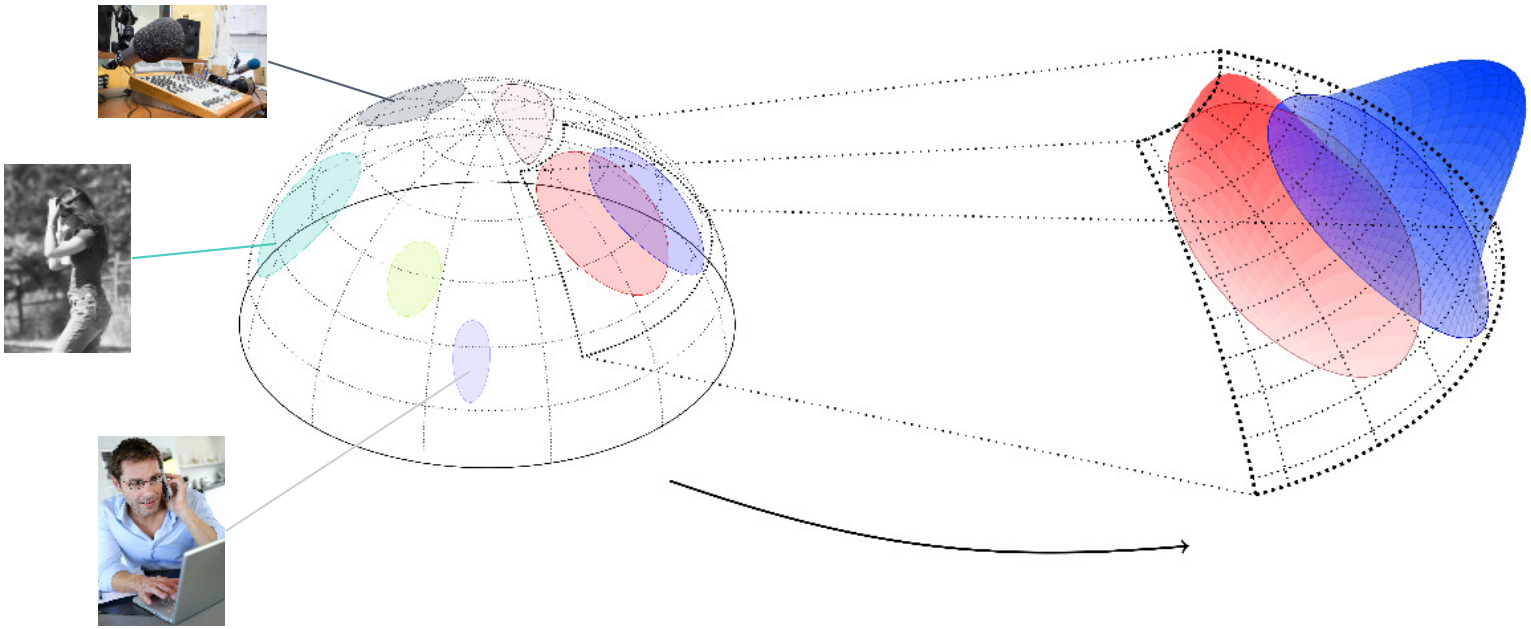
⇒ 400-dimensional i-vectors per sample

Overview on Speaker Recognition

4. Linear Discriminant Analyse (LDA)

⇒ 200-dimensional i-vector

5. Noise reduction by projection into spherical space

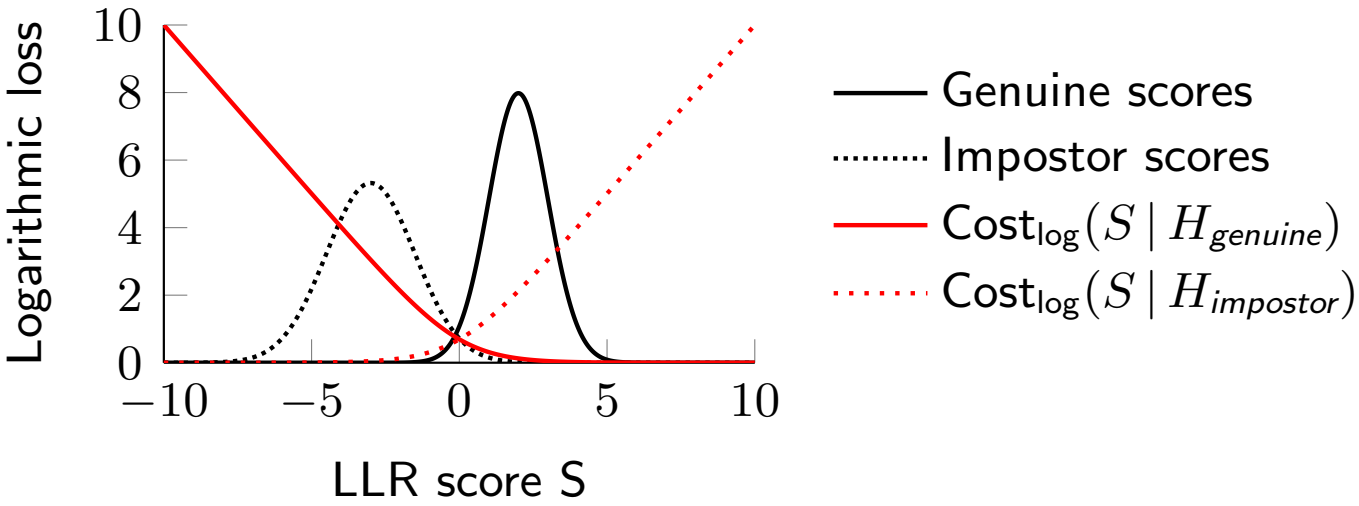


Probabilistic Discriminative speaker sub-spaces

Cross-Entropy in Score-Domain: C_{llr}

Representing the *Goodness of Log-Likelihood Ratio (LLR) scores*:

- Proper scoring rule:
  Costs by too low genuine ⇔ too high impostor scores
- Generalized cross-entropy:
  Information-loss to perfect classification

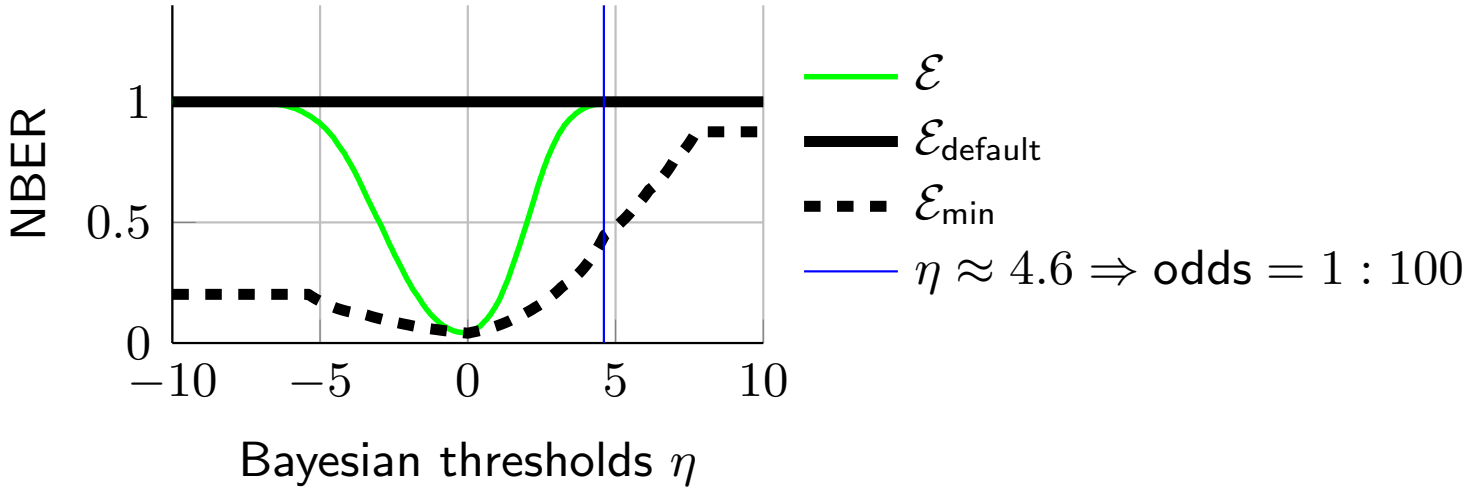


[Brümmer10] N. Brümmer: *Measuring, refining and calibrating speaker and language information extracted from speech.* Ph.D. thesis, University of Stellenbosch, 2010.

Cross-Entropy in Score-Domain: C_{llr}

Examining all operating-points $\tilde{\pi}$

- $\mathcal{E}$: empiric Bayes error $= \tilde{\pi}\mathrm{FNMR}(\eta) + (1-\tilde{\pi})\,\mathrm{FMR}(\eta), \eta = -\mathrm{logit}\,\tilde{\pi}$
- $C_{llr} = \int_{-inf}^{inf} \mathcal{E}\,\mathrm{d}\tilde{\pi}$ as application-independent uncertainty
- Aiming at: minimal information-loss: $C_{\mathrm{llr}}^{\mathrm{min}} = \int_{-inf}^{inf} \mathcal{E}_{\mathrm{min}}\mathrm{d}\tilde{\pi}$



[Brümmer+11] N. Brümmer, E. de Villiers: *The BOSARIS Toolkit User Guide: Theory, Algorithms and Code for Binary Classifier Score Processing.* Tech. Report, AGNITIO Research, 2011.

[Nautsch14] A. Nautsch: *Speaker Verification using i-Vectors.*, M.Sc. thesis, Hochschule Darmstadt, 2014.

Relative Entropy in Feature Spaces

Can the biometric discrimination potential of i-vector feature spaces be measured?

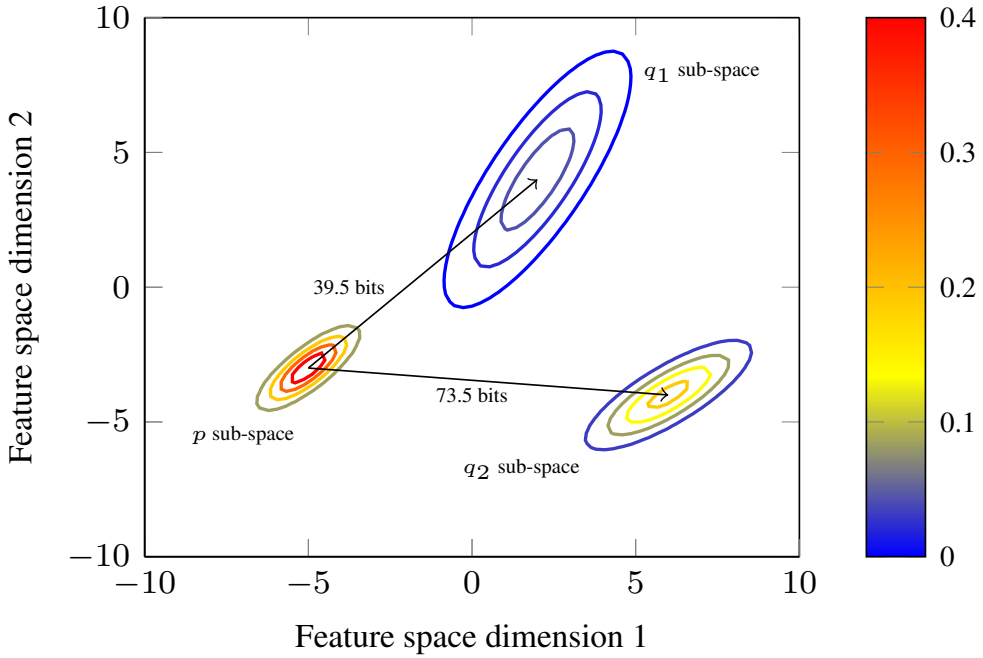
- Measuring biometric uniqueness
- Goal: Comparability of feature extractors
  - of one biometric modality
  - among modalities

- Note: entropy of passwords $\mathcal{H} = L \log_2 N$ [NIST06]
  Example: 4-digit PIN $\mathcal{H} \approx 13.3$ bits

[NIST06] NIST Tech. Rep.: *Electronic authentication guideline, recommendations of the national institute of standards and technology, information security*, 2006.

Estimating relative Entropy

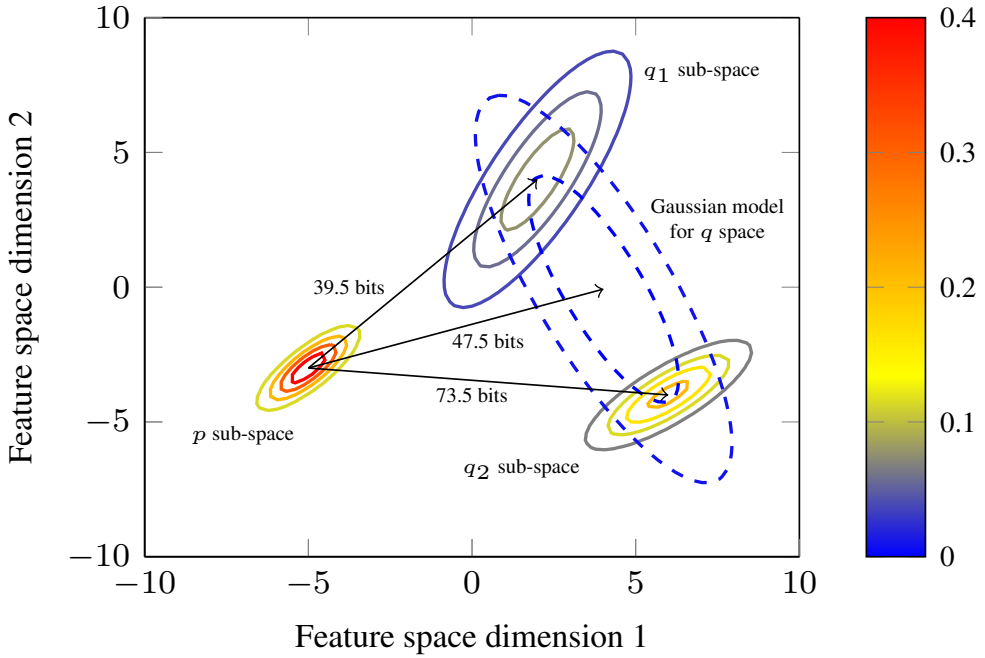
- Two-class problem: p same subject, q other subjects
- Kullback-Leibler Divergence as lower bound [Adler+06]



[Adler+06] A. Adler, R. Youmaran, S. Loyka: *Towards a Measure of Biometric Information*, IEEE CCECE, 2006.
[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Estimating relative Entropy

- Two-class problem: p same subject, q other subjects
- Kullback-Leibler Divergence as lower bound [Adler+06]



[Adler+06] A. Adler, R. Youmaran, S. Loyka: *Towards a Measure of Biometric Information*, IEEE CCECE, 2006.
[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Necessary Regularizations

Challenge: Which information is significant?

- Regularizations according to Adler et al. [Adler+06]:
  1. Degenerated features: $US_qV^t = \mathrm{svd}(\Sigma_q)$
     $\Rightarrow$ Sub-space with: $[S_q]_{i,j} \geq 10^{-10}[S_q]_{1,1}$
     $\Rightarrow$ Subject sub-space as: $S_p = U^t \Sigma_p V$
  2. Insufficient data: $[\Sigma_p]_{i,j} = 0 \quad i,j \geq N_p, i \neq j$

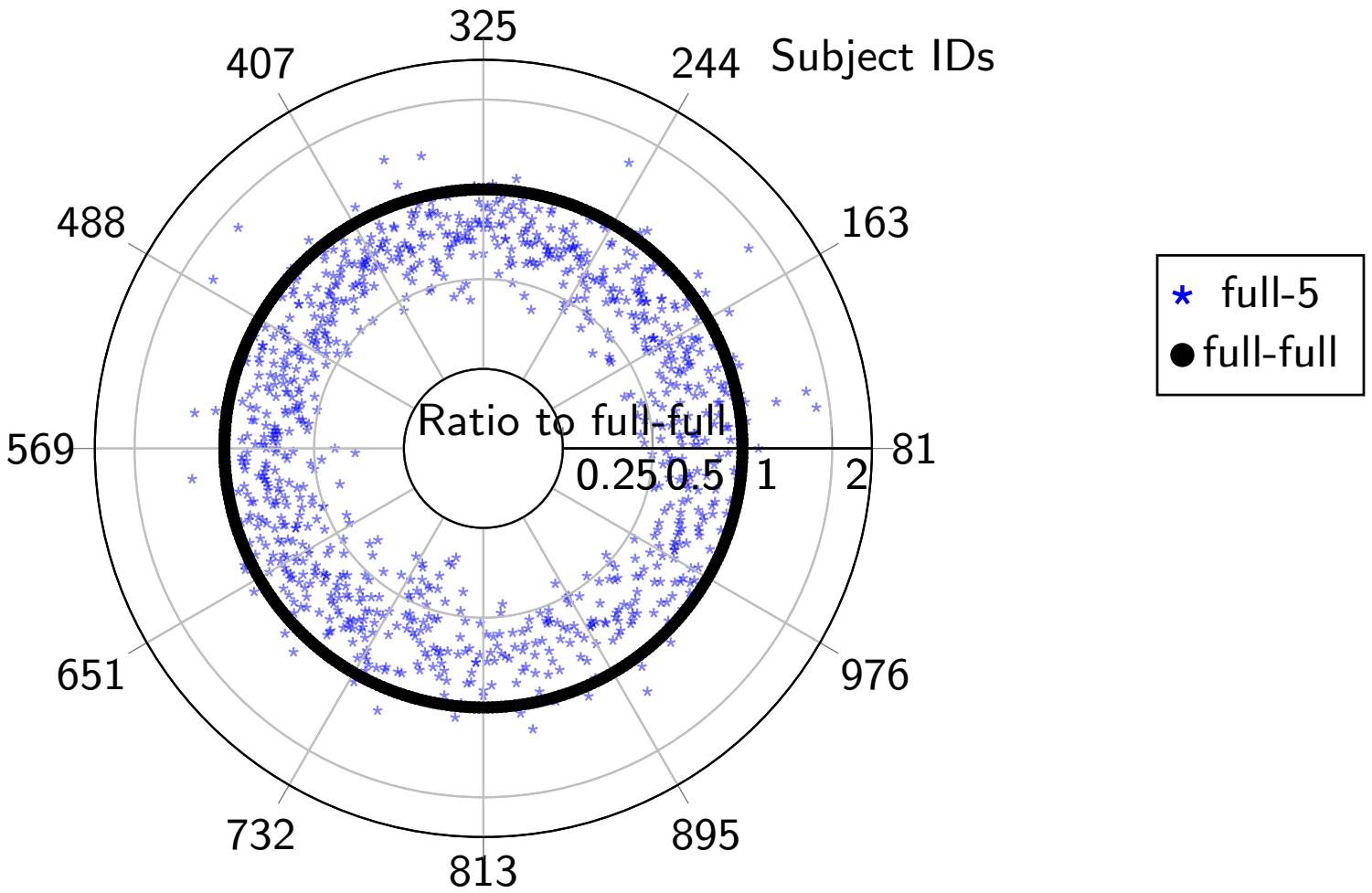
- Extension for variable N_p per subject [Nautsch+15]:
  3. ill-disposed regularized models:
     $\Rightarrow$ iterative $[\Sigma_p]_{i,j} = 0$ until Σ_p is positive-definite
  4. Minimal amount of genuine samples: $N_p \geq 10$

Experimental Set-up

- Database: I4U i-vectors of NIST Speaker Recognition Evaluation 2012 (SRE'12), comprising SREs 2004 – 2010
- 551 female, 425 male subjects

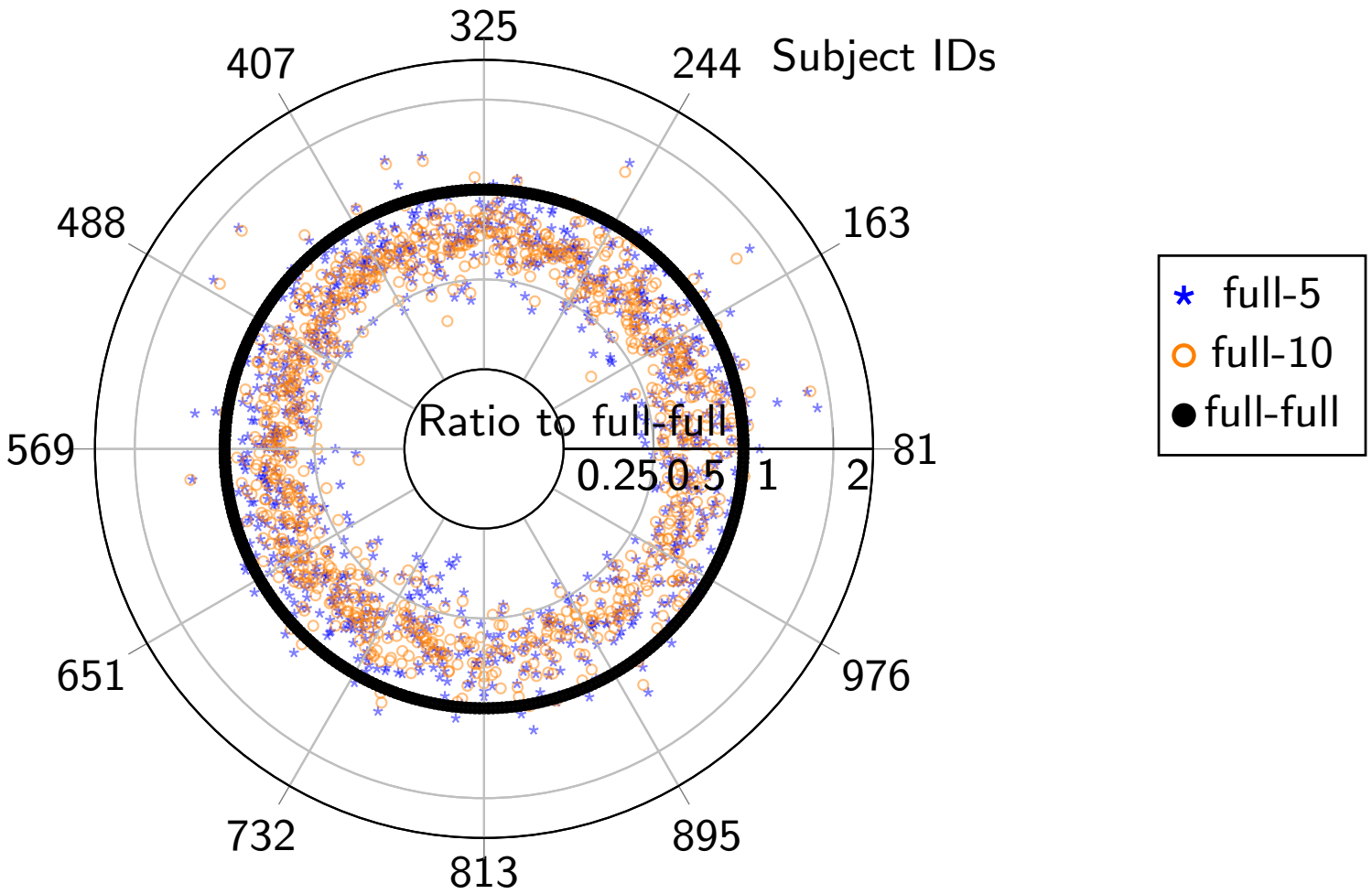
- Focus: incomplete probe samples
  $\Rightarrow$ Duration variations: 5s, 10s, 20s, 40s, full
- *How do speaker sub-spaces accumulate by increasing sample duration?*
- Expectation: correlation to system performance

Accumulation of Voice Templates by Duration



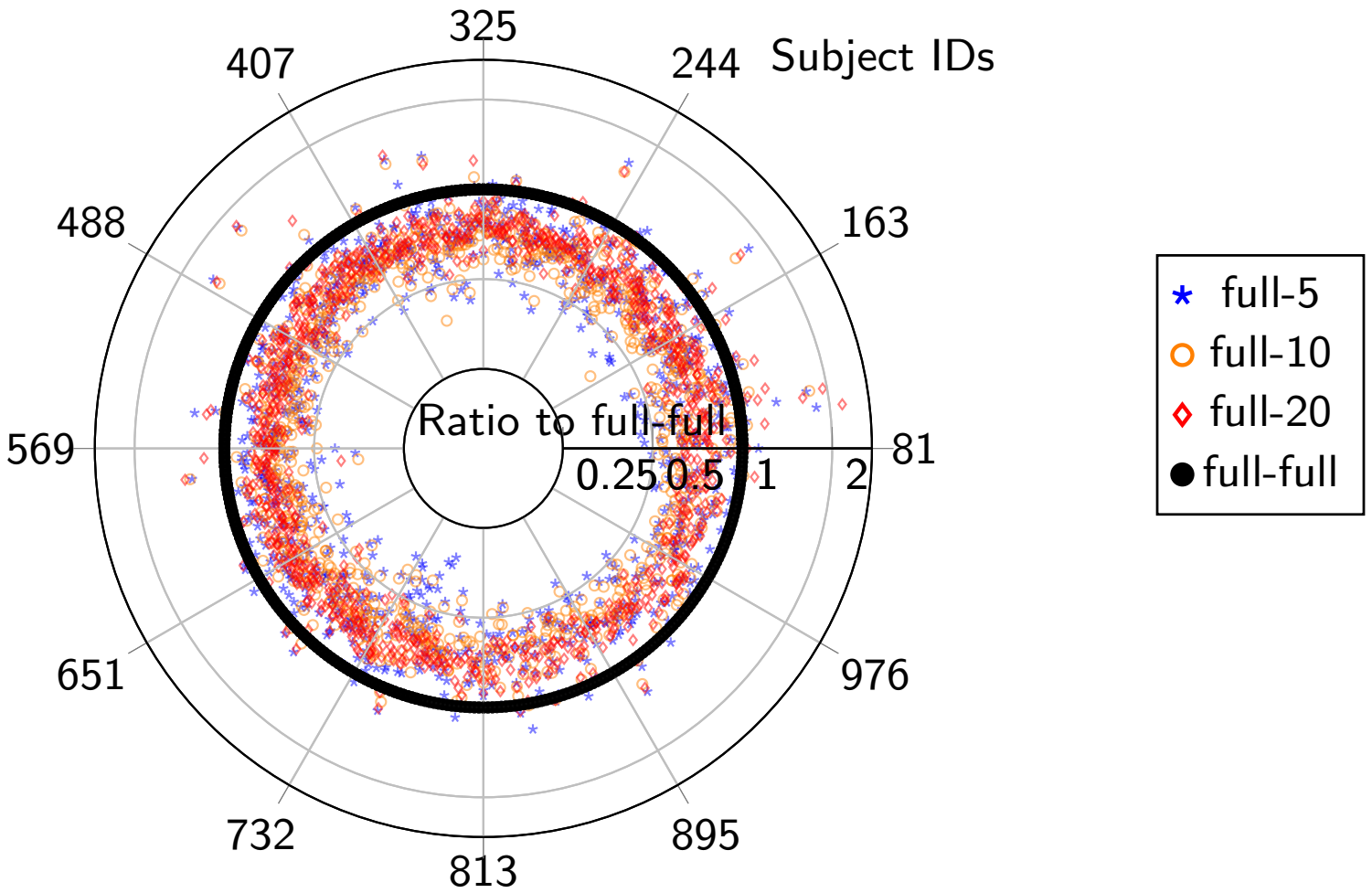
[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Accumulation of Voice Templates by Duration



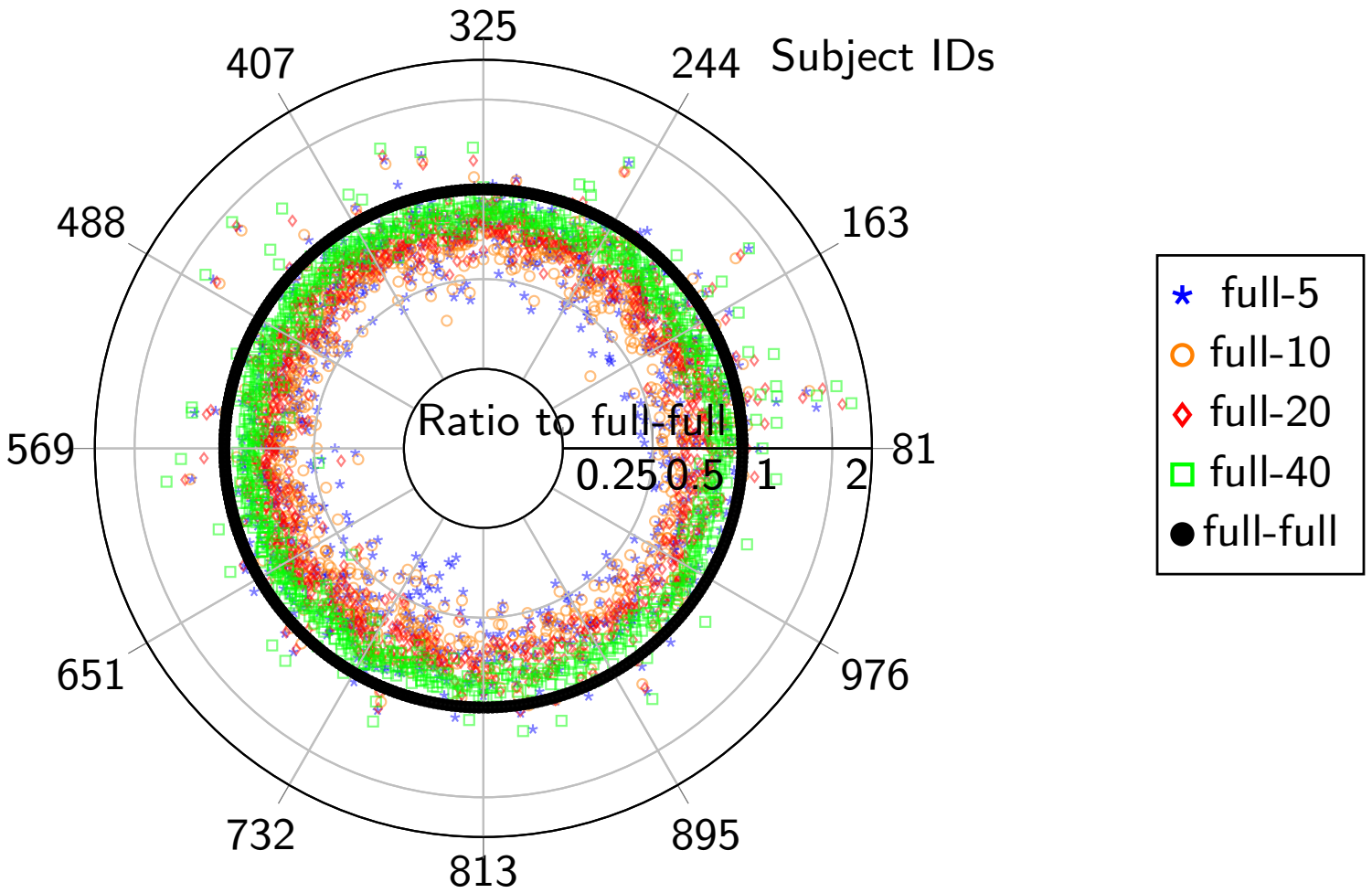
[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Accumulation of Voice Templates by Duration



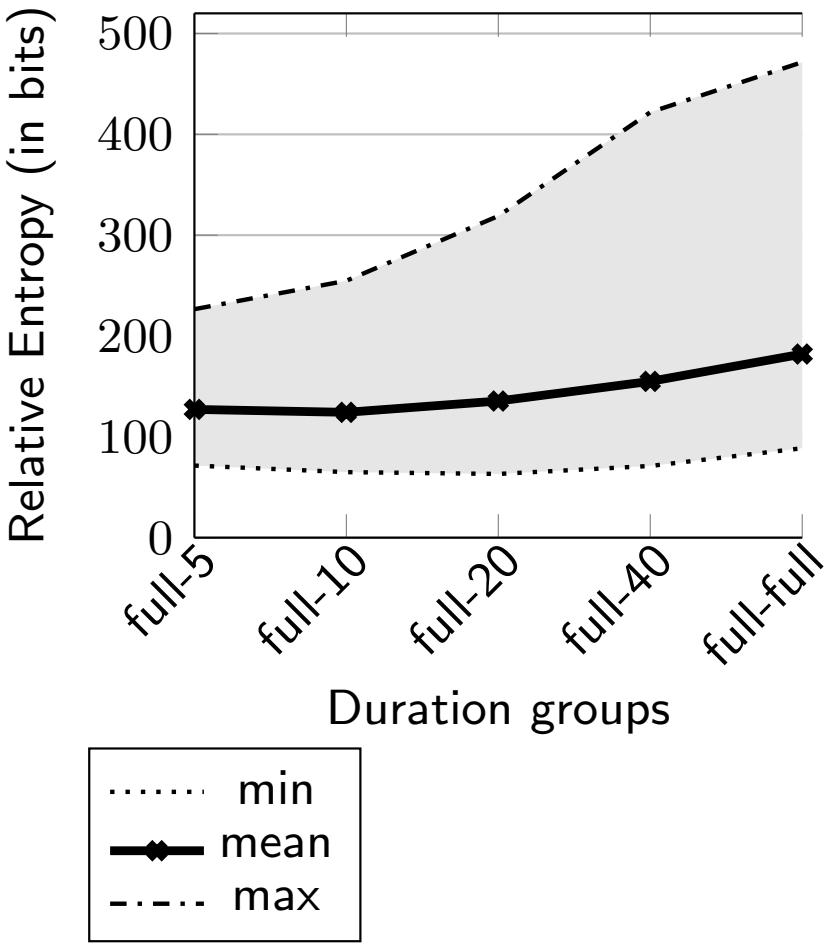
[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Accumulation of Voice Templates by Duration



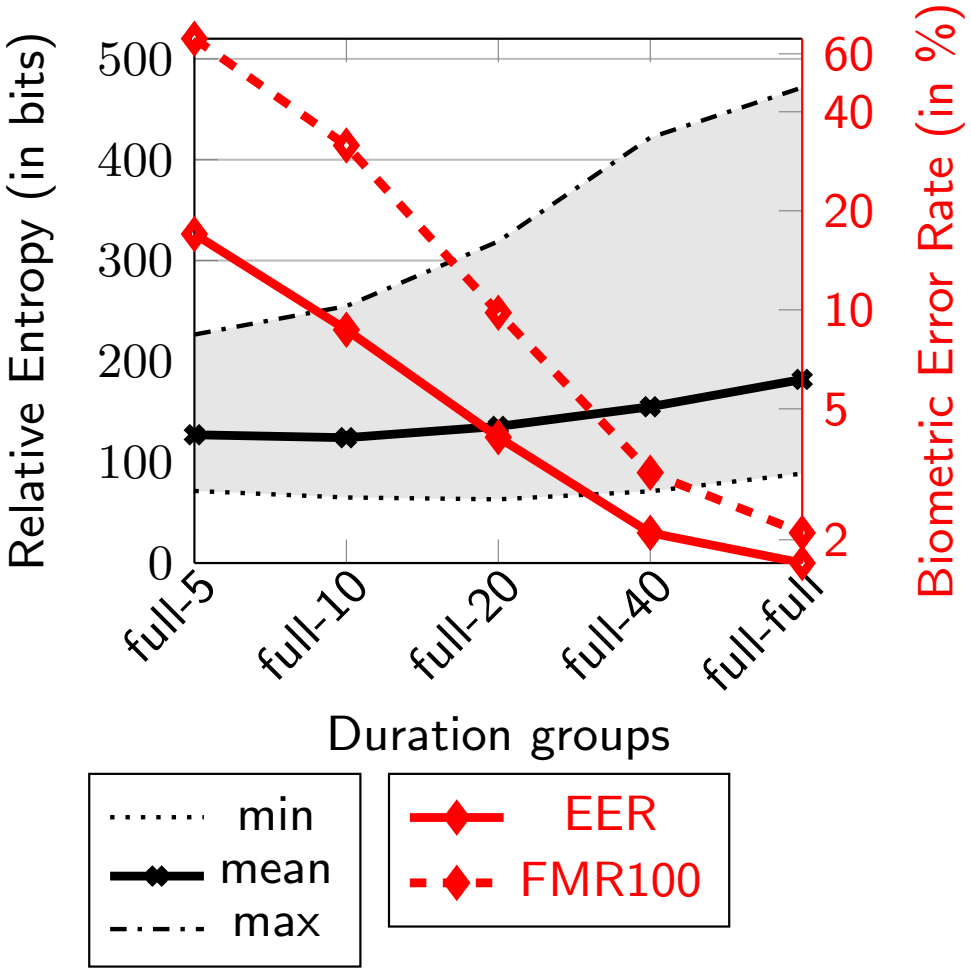
[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Comparison: relative Entropy ⇔ System Performance



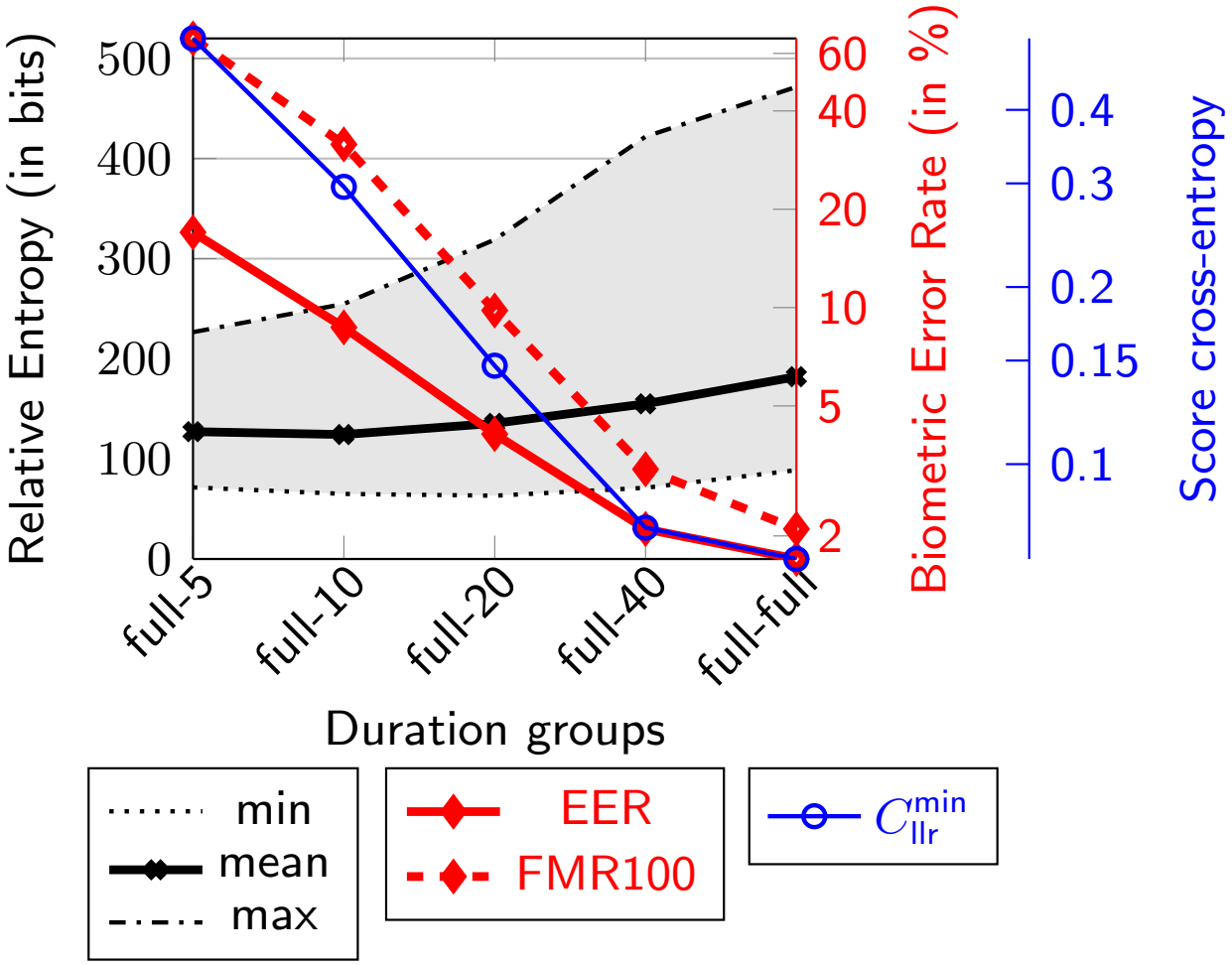
[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Comparison: relative Entropy ⇔ System Performance



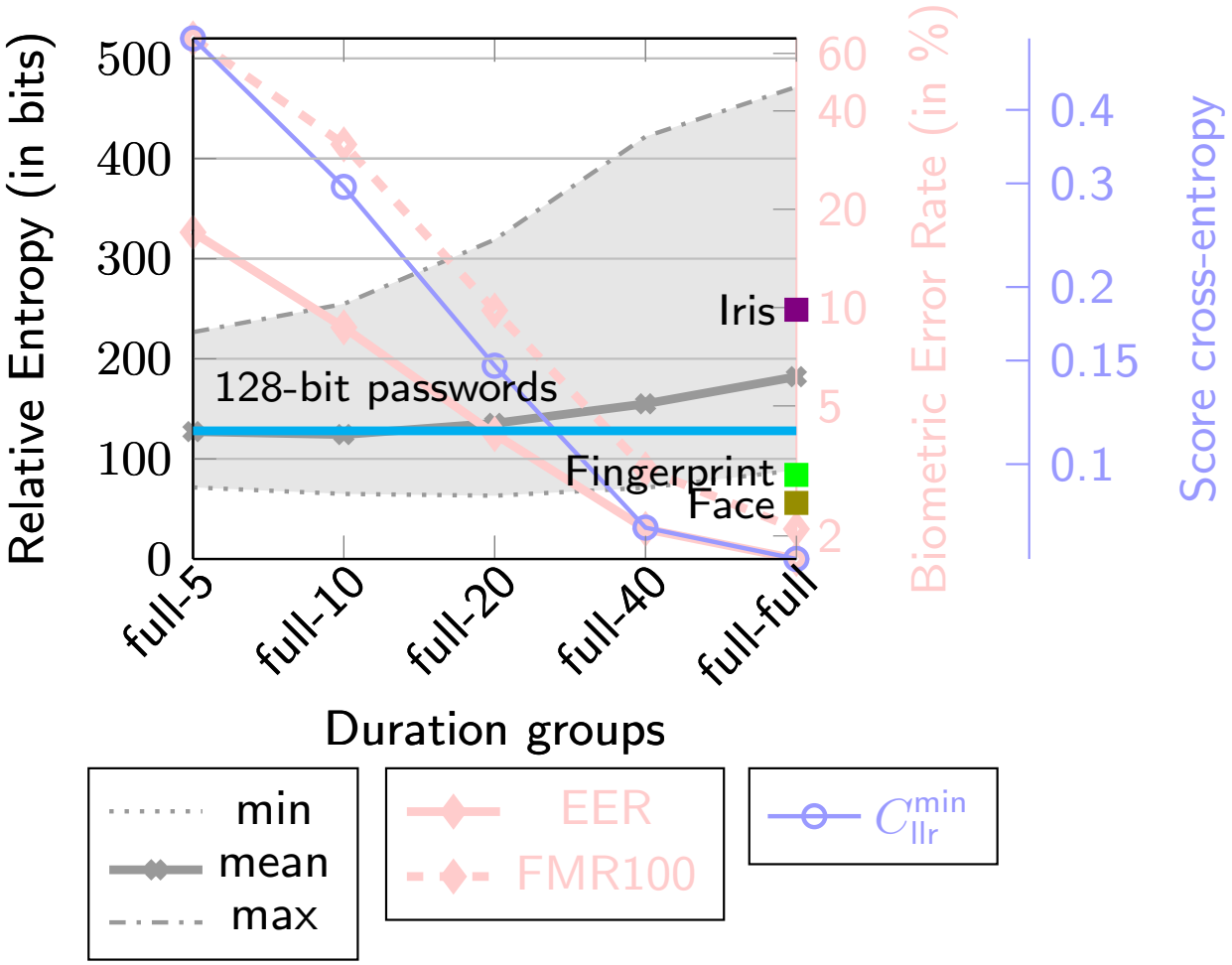
[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Comparison: relative Entropy ⇔ System Performance



[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Comparison: relative Entropy ⇔ System Performance



[Nautsch+15] A. Nautsch, C. Rathgeb, R. Saeidi, C. Busch: *Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition.*, IEEE ICASSP, 2015.

Conclusion

- Relative entropy as *metric for biometric uniqueness*
- Accumulation of voice templates by duration depicted

- Relative Entropy: from 63.2 up to 421.9 bit, $\mu > 124.3$ bit
- Comparability to other biometric modalities
- Comparability to password feature spaces (e.g., 128-bit)

- Subject collision probability: subject- and sample-depending 5×10^{-39} down to 2×10^{-55} (empiric data)