Towards Duration Invariance of i-Vector-based Adaptive Score Normalization

Andreas Nautsch*†‡, Christian Rathgeb†, Christoph Busch†, Herbert Reininger* and Klaus Kasper‡

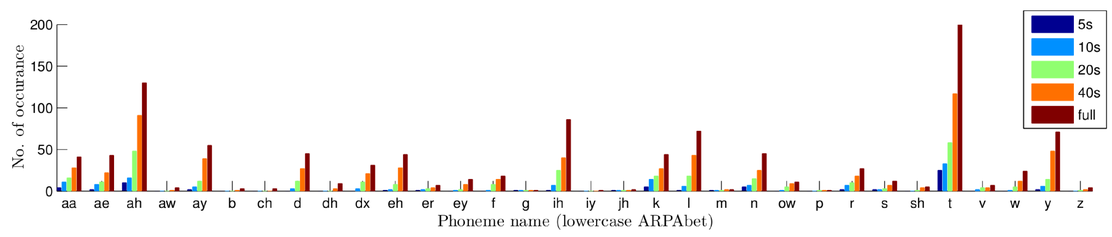
* atip Advanced Technologies for Information Processing GmbH, Frankfurt, Germany
† da/sec Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
‡ Department of Computer Science, Hochschule Darmstadt, Germany

Motivation

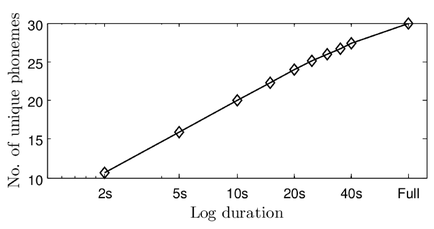
- Presence of noise in speech signals results
- Bad estimated speaker i-vector supspaces
- Sample duration has great performance effects
  - High-sufficient Baum-Welch statistics only on long-duration speech samples
  - Observation of „acoustic holes“ on short-term samples, which lead to differences in i-vector subspaces

Related Work

- Duration mismatches cause acoustic holes by phonetic content variations [1]
- Occurrence quantity of observed phones „reduces exponentially with duration“ [1]



- Duration depending score calibration [2]
- Lowest C_{LLR} observed for probe sample durations among templates from full (> 40s) duration samples



i-Vector subspace analysis

The i-vectors of the NIST i-vector challenge were analyzed w.r.t. durations Λ and i-vector elements:

- Log-normal distributed sample durations
- Independence test of i-vector element means on development set: big subspace variations

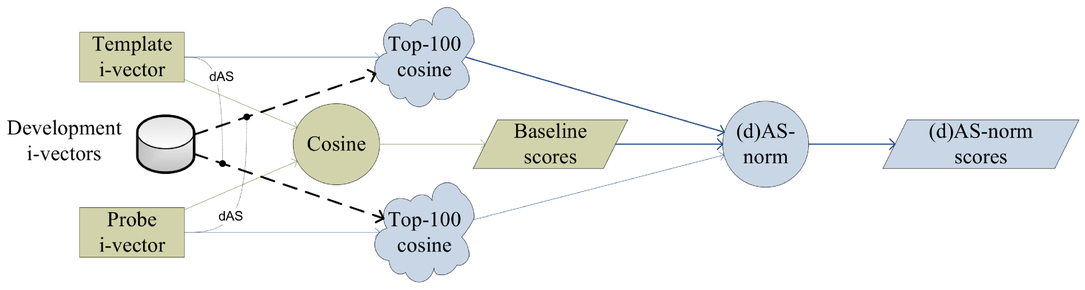
Table Student t-test of independent i-vector elements with respect to sufficiency classes.

dev-set	all	Λ_{full}	Λ_{40}	Λ_{20}	Λ_{10}
Λ_5	84	141	91	66	44
Λ_{10}	142	230	140	70	
Λ_{20}	132	246	118		
Λ_{40}	35	180			
Λ_{full}	172				

Duration-invariant AS-norm

Idea: compare i-vectors of similar sufficiency.

- Same duration group
  - Full duration dev-set i-vectors on templates
  - Same duration dev-set i-vectors on probes
- Similar subspace by cosine similarity



Evaluation: i-Vector challenge

Offline evaluation: 10x 5-fold cross-validations

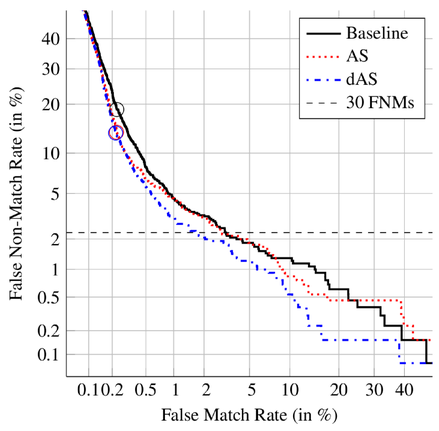
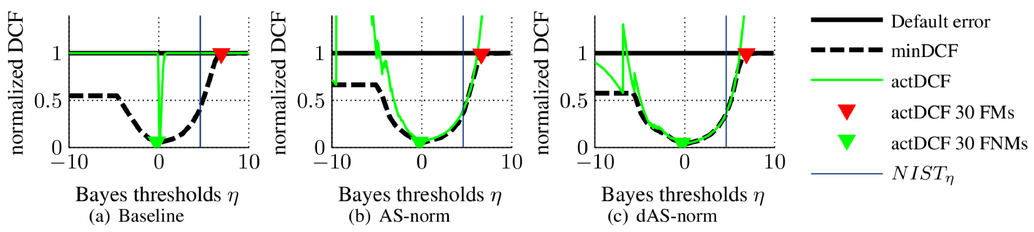


Table Average entropy comparison: all scores & duration-groups.

System	all	Λ_5	Λ_{10}	Λ_{20}	Λ_{40}	Λ_{full}
Baseline	0.89	0.95	0.93	0.92	0.89	0.86
AS-norm	0.17	1.18	0.41	0.18	0.08	**0.05**
dAS-norm	**0.10**	**0.35**	**0.20**	**0.11**	**0.07**	0.07

Table System performances: avg. EER, FMR100, minDCF.

System	EER	FMR100	minDCF	Challenge
Baseline	2.56	5.15	0.428	0.386
AS-norm	2.49	4.48	0.378	0.331
dAS-norm	**2.06**	**3.47**	**0.364**	**0.312**



(a) Baseline (b) AS-norm (c) dAS-norm

Conclusion and Future Work

- 19% relative-gain to baseline system
- Robust evidence strength over various durations
- Transfer to GPLDA & fast scoring approaches

References:
[1] T. Hasan et al.: “Duration Mismatch Compensation for i-Vector based speaker recognition systems”, IEEE ICASSP 2013.
[2] Mandasari et al.: “Quality Measure Functions for Calibration of Speaker Recognition Systems in Variant Duration Conditions”, IEEE/ACM TALSP 2013.

Contact:
andreas.nautsch@{atip|cased}.de
christian.rathgeb@cased.de

TECHNISCHE UNIVERSITÄT DARMSTADT
Fraunhofer SIT
h_da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES
LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz