Analysis of mutual duration and noise effects in speaker recognition: benefits of condition-matched cohort selection in score normalization

da/sec BIOMETRICS AND INTERNET-SECURITY RESEARCH GROUP

Aalto University

Andreas Nautsch* Rahim Saeidi†
Christian Rathgeb* 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

h_da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES

CASED

LOEWE

Motivation

- Real-life conditions in commercial and forensic applications
- Vast broadness of conditions e.g.:
  - Sample completeness (duration)
  - Ambient noise (AC / CROWD)
- True sample condition remains unknown due to e.g., SNR estimations
- Unified Audio Characteristics seem promising for condition-matching score-normalization

Research Questions

Q1: How extensive are mutual effects to the performance of a State-of-the-Art system?

Q2: Does condition-informed cohort selection benefit from statistic approaches rather than from condition-matched cohorts?

Q3: Do conditions affect basic i-vector properties, i.e. mean i-vectors?

Related Work

Unified Audio Characteristics

- Single multivariate Gaussian models in original (raw) i-vector space
  $\Lambda_i \sim \mathcal{N}(\mu_i, \Sigma), i = 1, \ldots, 55$
  - Condition-dependent μ_i
  - Shared full Σ by pooling
- Condition quality vector (q-vector) $\vec{q}$ as posterior probabilities for i-vector $\vec{w}$:

$$\vec{q}(i) = \frac{P(\vec{w} \,|\, \Lambda_i)}{\sum_{i=1}^{55} P(\vec{w} \,|\, \Lambda_i)}$$

Conventional AS Score Normalization

- Adaptive & Symmetric
  - Refs vs. probe-alike cohort set
  - Prbs vs. reference-alike cohort set
  - Computation of: $\mu_{ref}, \sigma_{ref}, \mu_{prb}, \sigma_{prb}$
- AS-normalized scores S_{AS} by averaged symmetric zero-norm of score S:

$$S_{AS} = \frac{1}{2}\left(\frac{S - \mu_{ref}}{\sigma_{ref}} + \frac{S - \mu_{prb}}{\sigma_{prb}}\right)$$

Cohort Selection Criterion

- Symmetric Kullback-Leibler divergence:

$$\frac{1}{2}\sum_{i=1}^{55} \vec{q_a}(i) \log \frac{\vec{q_a}(i)}{\vec{q_b}(i)} + \vec{q_b}(i) \log \frac{\vec{q_b}(i)}{\vec{q_a}(i)}$$

- Selecting top-c cohort q-vectors

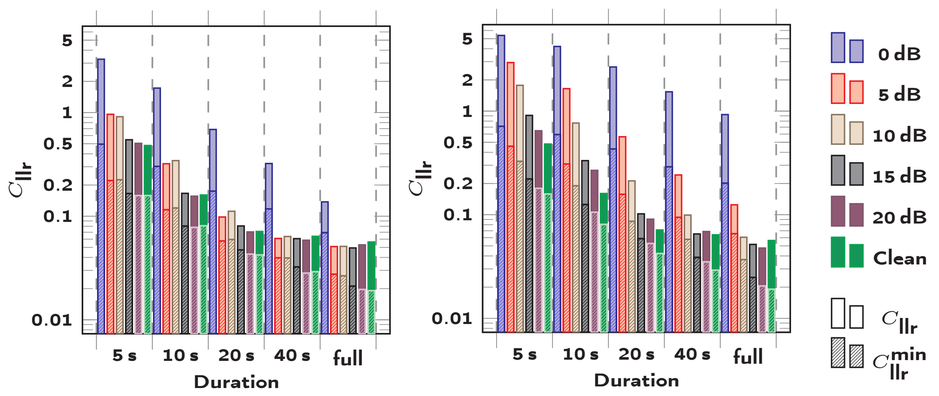
Experimental Set-up & Results

Condition	1	2	3	4	5	6	7	8	9	10	11 ... 15	16 ... 30	31 ... 55
Duration	5 s	10 s	20 s	40 s	full	5 s					10 s	20 s ... full	5 s ... full
Noise	clean					Air Conditioner (AC)							CROWD
SNR						0 dB	5 dB	10 dB	15 dB	20 dB	0 ... 20 dB	0 dB ... 20 dB	0 dB ... 20 dB

I4U Data of NIST SRE'12

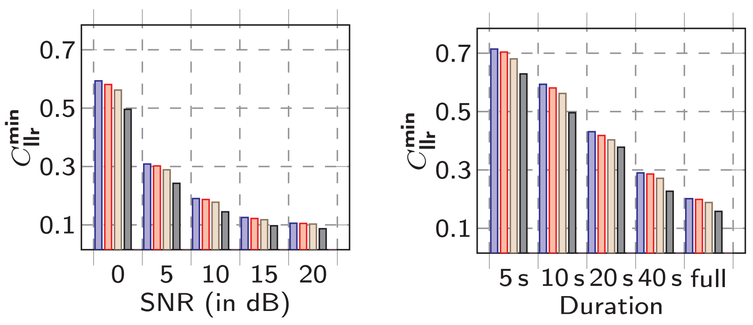
- Condition-dependent sample versions from long-duration & clean samples
- VAD labels of clean samples
- LDA: 400 to 200
- PLDA: 200 factors
- No calibration

Baseline Results: impact of mutual conditions

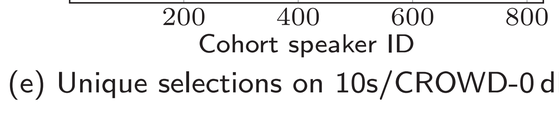


(a) AC noise (b) CROWD noise

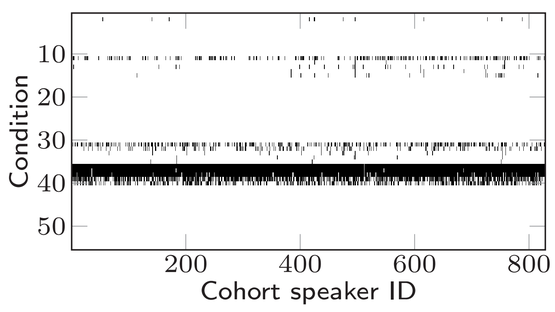
Comparison of AS-norm approaches



(c) 10s/CROWD (d) CROWD-0 dB



Automatic cohort pre-selection



(e) Unique selections on 10s/CROWD-0 dB

Cohort pre-selection by:

- Similar conditions
- SNR more relevant than duration

Analysis: i-vector pool mean shift

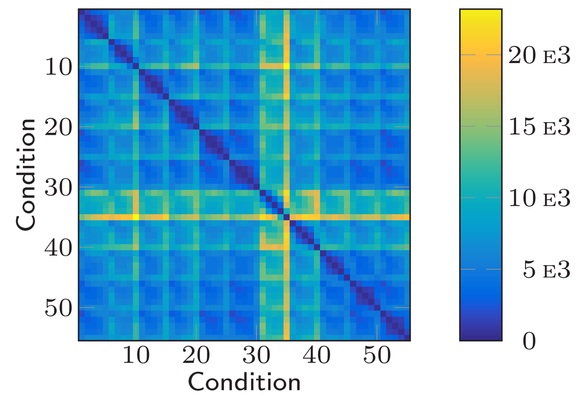
- Multi-variate Student t-test:

$$t^2 = \frac{n_x\, n_y}{n_x + n_y}(\vec{\bar{x}} - \vec{\bar{y}})' \mathbf{W}^{-1} (\vec{\bar{x}} - \vec{\bar{y}}), \textit{with:}$$

$$\mathbf{W} = \frac{\sum_{i=1}^{n_x} (\vec{x}_i - \vec{\bar{x}})(\vec{x}_i - \vec{\bar{x}})' + \sum_{i=1}^{n_y} (\vec{y}_i - \vec{\bar{y}})(\vec{y}_i - \vec{\bar{y}})'}{n_x + n_y - 2}$$

- P-values by cdf F of χ^2 ($D = 200$):

$$t^2 \sim \chi^2_D, \qquad p = 1 - F_{\chi^2_D}(t^2)$$



(f) t^2 scores of Hotelling's T^2-test

$\Rightarrow$ High significance: $p \begin{cases} = 1 & \textit{same condition: same mean i-vector} \\ < 10^{-13} & \textit{cross condition: different mean i-vector} \end{cases}$

Conclusion

- Mutual effects with performance impacts by log-duration and log-SNR
- Quality-based cohort pre-selection yields significant gains
- Performance gap between degraded and non-degraded samples still as open challenge
- Means of cross-condition i-vectors differ

Acknowledgements

We would like to thank the I4U consortium for database sharing.

This work has been partially funded by the Center for Advanced Security Research Darmstadt (CASED), the Hesse government (project no. 467/15-09, BioMobile) and the Academy of Finland (project no. 256961 and 284671).