Metrics in Audio Security & Privacy

Andreas Nautsch
EURECOM

TReSPAsS-ETN TE2
2021-01-21/22 — Virtual Event
Outline

● Audio Security
  ○ Automatic Speaker Verification Anti-Spoofing Challenge 2019 (ASVspoof 2019)
    https://www.asvspoof.org/

● Audio Privacy
  ○ VoicePrivacy 2020 Challenge
    https://www.voiceprivacychallenge.org/
Audio Security Metric

tandem Decision Cost Function (t-DCF)

Speech $\Leftrightarrow$ SC37 dictionary :)

tar  ~  mated, bona fide, ...
non ~  non-mated, non-attack, ....
spoo  ~  presentation attack,
        logical access spoof, ...
miss ~  FRR, FNMR, BPCER, ...
false alarm (fa) ~  FAR, FMR, APCER, ...
Audio Security: The Setting

- **Anti-Spoofing**
  - “Physical Access” Replay attacks (see Voice PAD)
  - “Logical Access” Voice synthesis/morphing/conversion attacks (not PAD)

- **In-Scope**
  - Tandem operation of countermeasure (CM) and ASV sub-systems
  - Throughout formalised assessment

- **Out-Scope**
  - Informal descriptors by error rates
  - Purely CM-focused performance
Audio Security: Expected Cost as Metric

- **Quantification of beliefs**
  - What is the impact of a decision outcome?
  - How likely is a decision outcome?

- **Expected class discrimination risk**
  - $\mathbb{E} \left[ \text{risk} \mid \text{costs, class priors, classification rates} \right]$
  - Risk minimisation: sweep thresholds, take minimum

\[
t\text{-DCF} = C_{\text{miss}} \cdot \pi_{\text{tar}} \cdot P_a + C_{\text{fa}} \cdot \pi_{\text{non}} \cdot P_b + C_{\text{fa}, \text{spoof}} \cdot \pi_{\text{spoof}} \cdot P_c + C_{\text{miss}} \cdot \pi_{\text{tar}} \cdot P_d
\]
Audio Security: Tandem Classification Rates

\[ P_c(\tau_{cm}, \tau_{asv}) = P_{fa}^c(\tau_{cm}) \times P_{fa,spoof}^a(\tau_{asv}) \]

\[ P_b(\tau_{cm}, \tau_{asv}) = (1 - P_{miss}^c(\tau_{cm})) \times P_{fa}^a(\tau_{asv}) \]

\[ P_a(\tau_{cm}, \tau_{asv}) = (1 - P_{miss}^c(\tau_{cm})) \times P_{miss}^a(\tau_{asv}) \]
Audio Security: Metric Normalisation

- Better comparability (other costs/priors)
- What are the extreme actions?
  - CM & ASV: all-pass
    \[ C_{fa} \cdot \pi_{non} \cdot 1 + C_{fa,spoof} \cdot \pi_{spoof} \cdot 1 \]
  - CM: no-pass
    \[ C_{miss} \cdot \pi_{tar} \cdot 1 \]
  - CM: all-pass & ASV: no-pass
    \[ C_{miss} \cdot \pi_{tar} \cdot 1 \]

\[ t-DCF' (\tau_{cm}, \tau_{asv}) = \frac{t-DCF(\tau_{cm}, \tau_{asv})}{t-DCF_{default}} \]

\[ t-DCF_{min} \frac{t-DCF_{min}}{t-DCF_{default}} \leq \frac{t-DCF_{min}}{t-DCF_{min}} = 1 \]

\[ t-DCF_{default} = \min \{ C_{fa} \cdot \pi_{non} + C_{fa,spoof} \cdot \pi_{spoof} \cdot 1 \} \]
Audio Security: t-DCF Examples

- ASVspoof 2019 Challenge
  - Cost & prior parameters as per challenge
  - Synthetic ASV & CM scores
Audio Privacy Metric

Zero Evidence

Biometric Recognition Assessment
(The Privacy ZEBRA)
Audio Privacy: The Setting

- Pseudomise audio speech data
- Decoupling layers & taking the perspective of an adversary

- Existing metrics do not suffice!
  - Zero-knowledge proofs are unavailable.
  - EER is the worst possible decision policy that an adversary can take for herself.
  - Unlinkability (not devised for this setting) — identity confirmation but not short-listing.
  - Any fixed error rate/cost metric prejudices privacy disclosure impacts to an individual.
Audio Privacy: Zero Evidence as Metric

- **Population level: Empirical Cross-Entropy (ECE)**
  - Idea: prior entropy $\Rightarrow$ evidence $\Rightarrow$ posterior entropy
  - Cross-entropy of classification by scores from ground truth
  - Zero evidence: prior ECE = posterior ECE, regardless of prior $\pi$

- **Individual level: Zero Strength of Evidence**
  - Forensic sciences: likelihood ratio
  - Who is stronger: prosecutor or defendant?

![Coin-tossing simulation: all scores are equal](image-url)

<table>
<thead>
<tr>
<th>Tag</th>
<th>Category</th>
<th>Posterior odds ratio (flat prior)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$l = 1 = 10^0$</td>
<td>50 : 50 (flat posterior)</td>
</tr>
<tr>
<td>A</td>
<td>$10^0 &lt; l &lt; 10^1$</td>
<td>more disclosure than 50 : 50</td>
</tr>
<tr>
<td>B</td>
<td>$10^1 \leq l &lt; 10^2$</td>
<td>one wrong in 10 to 100</td>
</tr>
<tr>
<td>C</td>
<td>$10^2 \leq l &lt; 10^4$</td>
<td>one wrong in 100 to 10000</td>
</tr>
<tr>
<td>D</td>
<td>$10^4 \leq l &lt; 10^5$</td>
<td>one wrong in 10000 to 100000</td>
</tr>
<tr>
<td>E</td>
<td>$10^5 \leq l &lt; 10^6$</td>
<td>one wrong in 100000 to 1000000</td>
</tr>
<tr>
<td>F</td>
<td>$10^6 \leq l$</td>
<td>one wrong in at least 1 000 000</td>
</tr>
</tbody>
</table>

Figure based on [wikimedia.org](https://en.wikipedia.org/wiki/Utility_investigation)
Audio Privacy: ZEBRA Examples

● VoicePrivacy 2020 Challenge
  ○ Task: speech recognition should work — voice biometrics not ⇒ modification of raw audio
  ○ ASV: pre-trained kaldi x-vector recipe
  ○ B1: DNN baseline
  ○ B2: signal processing baseline

Expected privacy disclosure (population)  

Worst-case privacy disclosure (individual)  

Categorical tag

https://gitlab.eurecom.fr/nautsch/zebra
Summary & Conclusion

● Summary
  ○ Audio security: cost-based approach for expected risk minimization
  ○ Audio privacy: relative information & strength of evidence approach

● Conclusion
  ○ Constrained cost as a guide for the CM optimization given a biometric system
    ⇒ taking a holistic perspective
  ○ Audio privacy must achieve privacy for every single one; are we a marginalising society?
    ⇒ expectation & worst-case estimates

● Security vs. Privacy: seek positive-sum solutions; not zero-sum solutions