How to ask without speech?
On quantifying zero-evidence speech

“The Privacy ZEBRA”

Andreas Nautsch
Context: VoicePrivacy 2020 Challenge

- **Task**: audio pseudonymisation ⇒ modify raw audio
  - Voice biometrics should fail
    “Same person or different person?”
  - Speech recognition should work
    “What was said?”

- **Metric**: Zero-Evidence Biometric Recognition Assessment
Intuition: benefit to decision making?

- Motivation in forensic sciences
  - What is the benefit of evidence reporting to decision making?
  - How to validate?

- Empirical cross-entropy (ECE)
  - Less uncertainty with evidence than without?

- Strength-of-evidence: likelihood ratios
  - Which decision is more supported?
Textbook: empirical cross-entropy (ECE)

- The ECE step-by-step
  - Prior entropy in making yes/no decision
  - Posterior entropy based on scores/evidence
  - Issue: no theoretical foundation for reference likelihoods
  - Remedy: cross-entropy, law of large numbers & priors are external to the classifier

$$\Theta = \{ A: \text{“same person”}, \ B: \text{“different person”} \}$$

$$H_P(\Theta) = - \sum_{\theta \in \Theta} P(\theta) \log_2 P(\theta)$$

$$H_P(\Theta | S) = - \sum_{\theta \in \Theta} P(\theta) \int_{s \in S} P(s | \theta) \log_2 P(\theta | s) \, ds$$

$$H_{P|P}(\Theta | S) = \sum_{\theta \in \Theta} \tilde{P}(\theta) \int_{s \in S} P(s | \theta) \log_2 \tilde{P}(\theta | s) \, ds$$

$$P(s | \theta) \approx |S_\theta|^{-1}$$

Strength-of-evidence

score: $P(\text{voice} | \text{same}) / P(\text{voice} | \text{different})$

Ramos, Franco Pedroso, Lozano-Diez, Gonzalez-Rodriguez: Deconstructing Cross-Entropy for Probabilistic Binary Classifiers, Entropy 20(3), 2018

Disclosure: worst-case?

- **Motivation:** privacy for the individual; not for the average only

- **Analogue from forensic sciences to privacy preservation**
  - Prosecutor & defendant in a tug of war \( \Rightarrow \) i.e. *strength-of-evidence*
  - Decision maker: the adversary \( \Rightarrow \) i.e. what is the worst case?

- **Categorical tags**

<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 10,000 to 100,000</td>
</tr>
<tr>
<td>E</td>
<td>( 10^5 \leq l &lt; 10^6 )</td>
<td>one wrong in 100,000 to 1,000,000</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>

Categorical scale of privacy disclosure (adapted from forensic sciences)

Willis et al.: Guideline for evaluative reporting in forensic science, European Network of Forensic Science Institutes (ENFSI), 2016
ZEBRA framework, an example

- VoicePrivacy 2020 Challenge — audio pseudonymisation
  - Task: speech recognition should work — voice biometrics should fail
  - Unprotected data: state-of-the-art voice biometrics
  - B1: DNN baseline
  - B2: signal processing baseline

![Expected privacy disclosure (population)](image)

- Perfect privacy (0.00, 0.00, 0)
- B1 (0.11, 2.27, C)
- B2 (0.36, 3.58, C)
- Unprotected data (0.58, 3.98, C)

> https://gitlab.eurecom.fr/nautsch/zebra
The Privacy ZEBRA