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

“The Privacy ZEBRA”

Andreas Nautsch

EURECOM
Sophia Antipolis

Picture taken in
Heidelberg Zoo, 2020

2021-05-21

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


Natalia Tomashenko


Brij M.L. Srivastava


Xin Wang


Emmanuel Vincent


Andreas Nautsch


Junichi Yamagishi


Nicholas Evans


Jose Patino


Jean-François Bonastre


Paul-Gauthier Noé

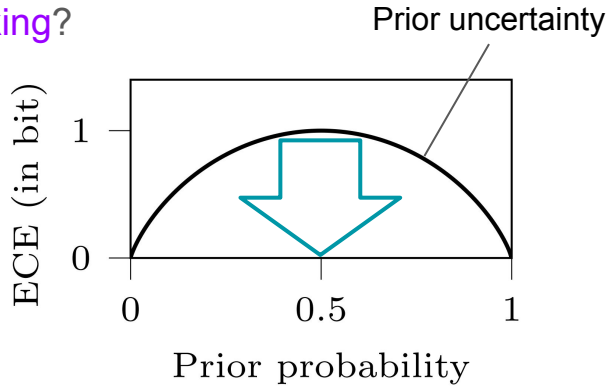

Massimiliano Todisco

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*?



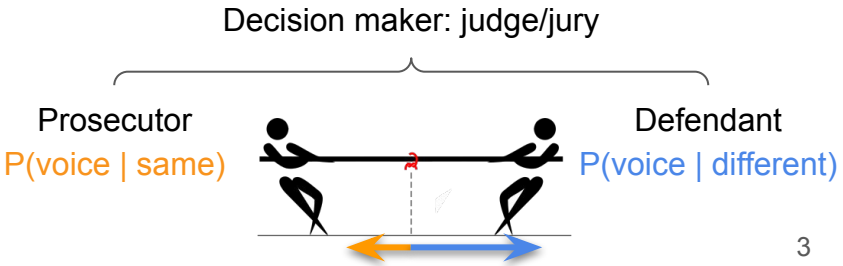
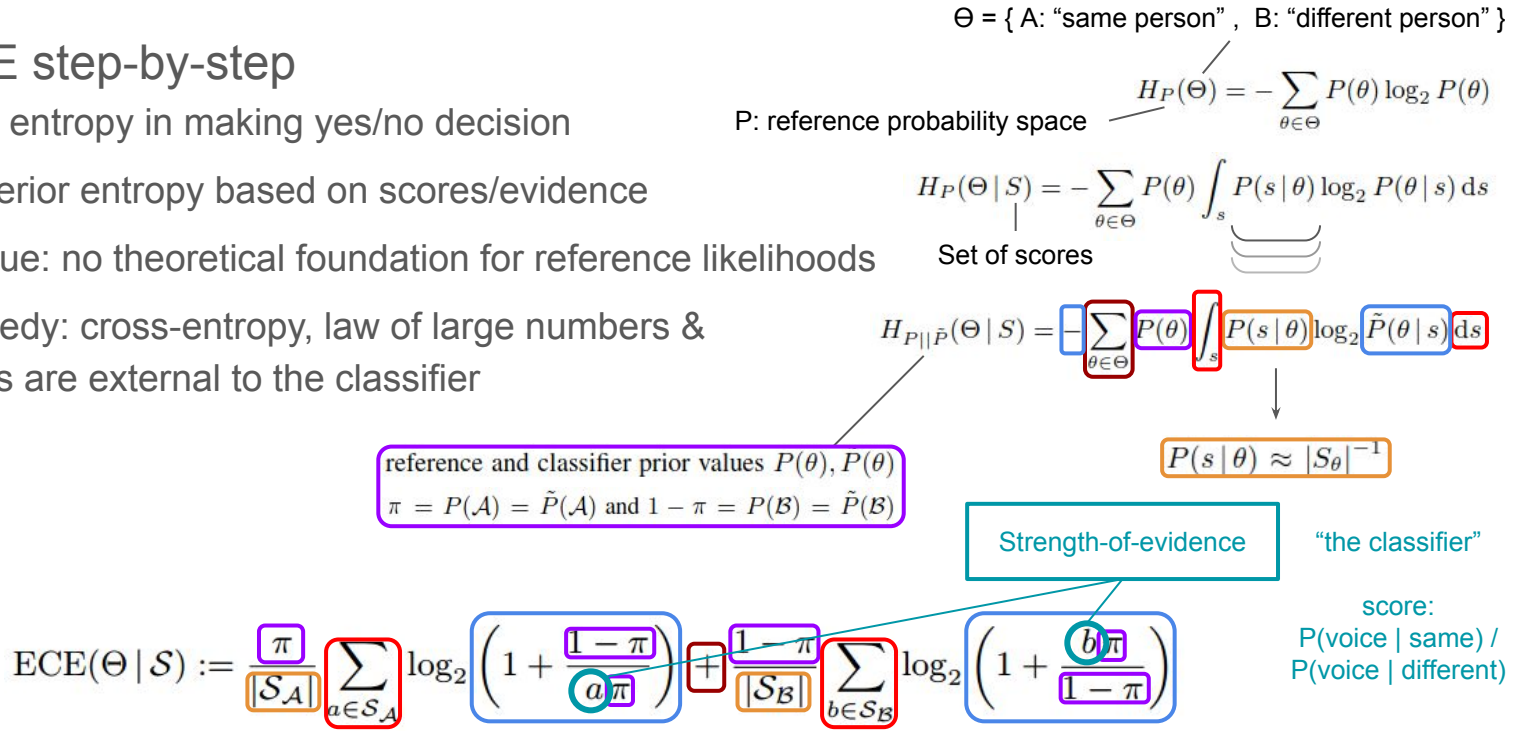


Figure based on wikimedia.org

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



Ramos & Gonzalez-Rodriguez: Cross-entropy Analysis of the Information in Forensic Speaker Recognition, in Proc. Odyssey, 2008
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 ⇒ i.e. *strength-of-evidence*
  - Decision maker: the adversary ⇒ i.e. what is the worst case?
- Categorical tags

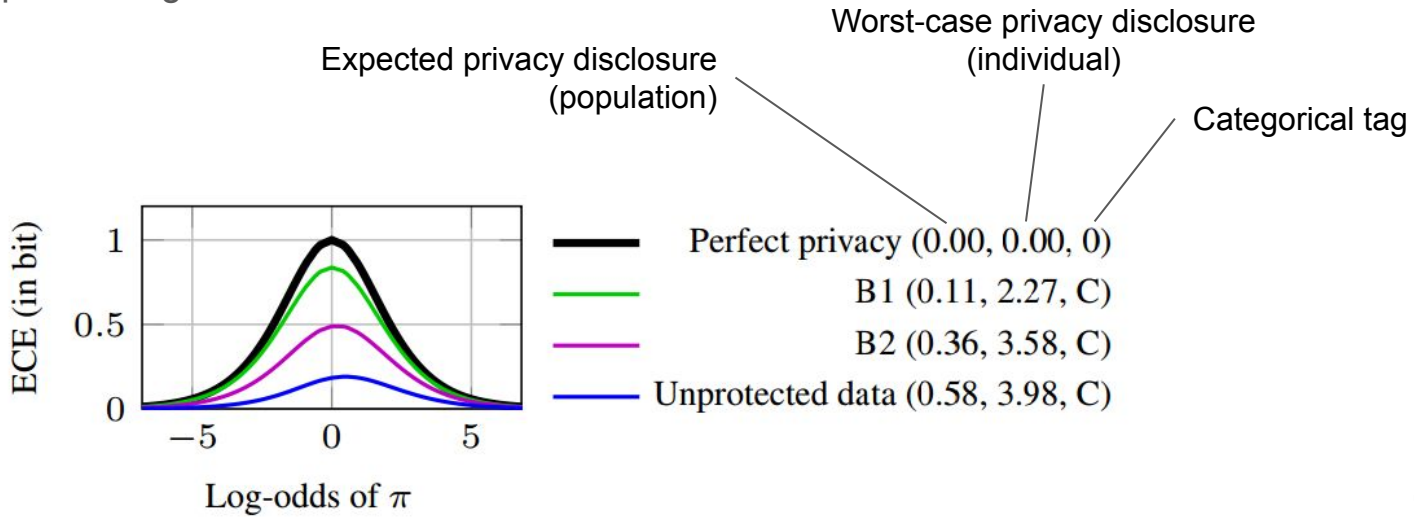
Tag	Category	Posterior odds ratio (flat prior)
0	$l = 1 = 10^0$	50 : 50 (flat posterior)
A	$10^0 < l < 10^1$	more disclosure than 50 : 50
B	$10^1 \leq l < 10^2$	one wrong in 10 to 100
C	$10^2 \leq l < 10^4$	one wrong in 100 to 10 000
D	$10^4 \leq l < 10^5$	one wrong in 10 000 to 100 000
E	$10^5 \leq l < 10^6$	one wrong in 100 000 to 1 000 000
F	$10^6 \leq l$	one wrong in at least 1 000 000

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



https://gitlab.eurecom.fr/nautsch/zebra

The Privacy ZEBRA