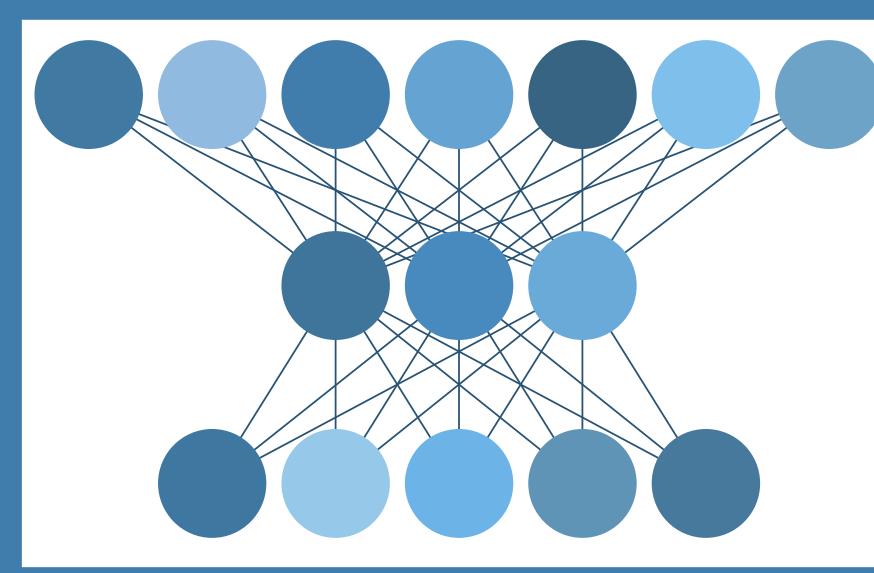# Speaker identification in courtroom contexts:
Individual listeners *vs* automatic forensic voice comparison

Nabanita Basu[1], Agnes S Bali[2], Philip Weber[1], Claudia Rosas-Aguilar[1,3], Gary Edmond[4], Kristy A Martire[2], Geoffrey Stewart Morrison[1,5]

Aston University
BIRMINGHAM UK

Questions : True ✓ or ✗ False?

❝ the **judge** could … **simply listen** to the recordings and make a decision

❝ performance of **automatic-speaker-recognition** systems **not better** than human listeners

❝ the jury is **'perfectly well-equipped'** to listen … compare … draw conclusions"

⟫⟫ **So what?** Affects **admissibility** in court.

Future of this project
- groups of listeners ("jury")
- effect of biasing factors

Future of the lab (FDSL)
- case-specific corpus collection
- state-of-the-art → other biometrics
- case-specific validation and theory
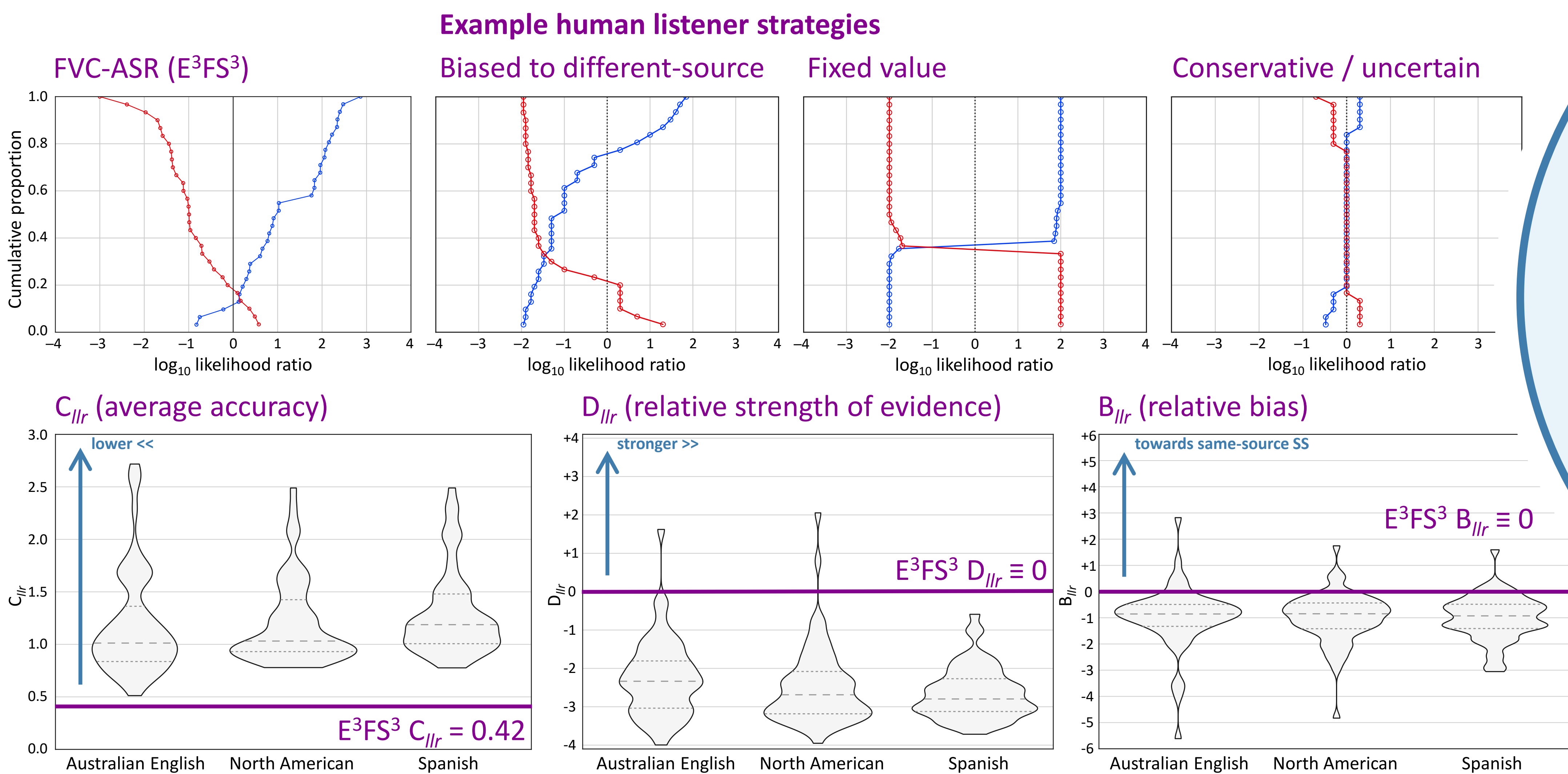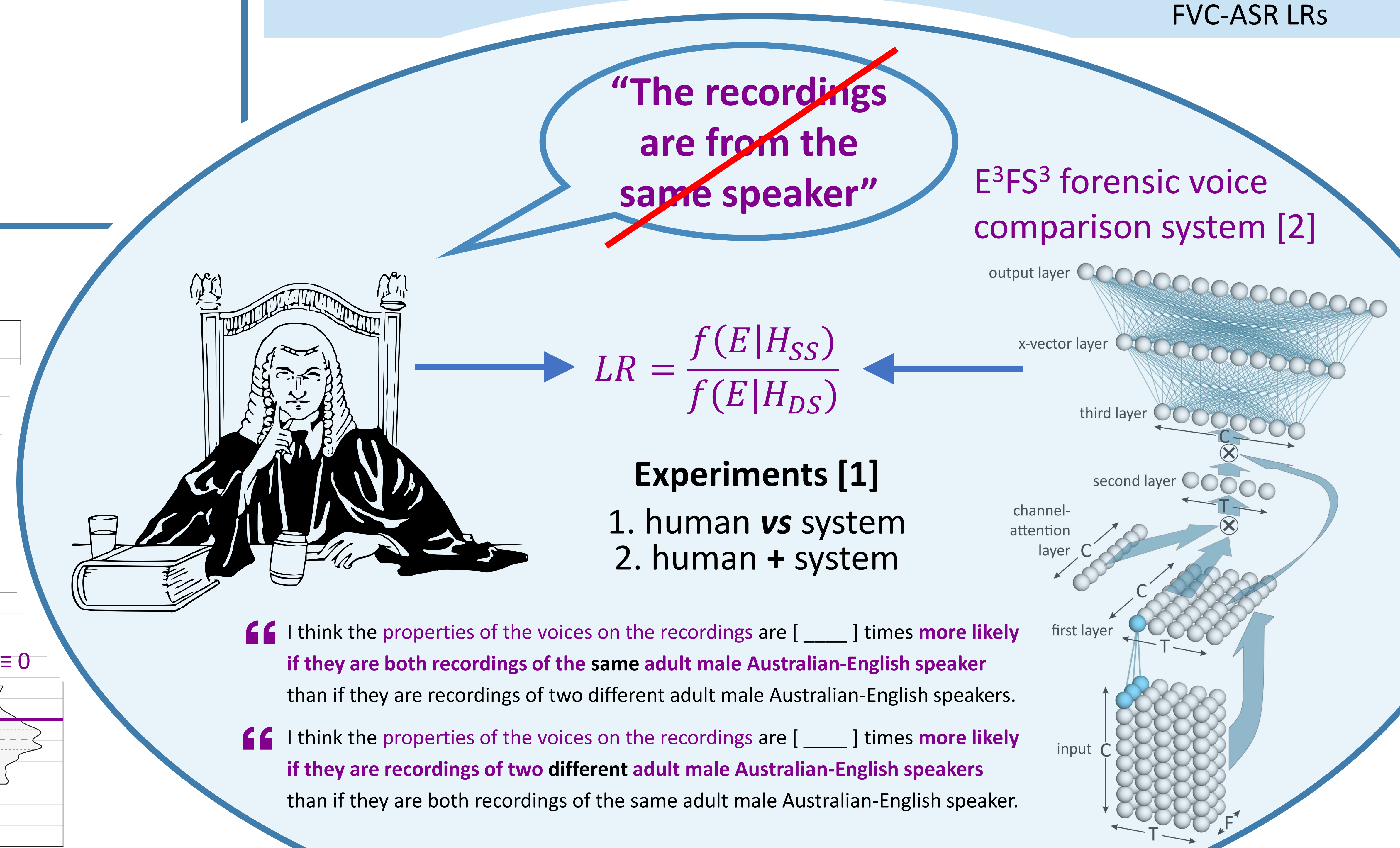
Aston Forensic Data Science Laboratory
- develop methods for evaluation of forensic evidence – based on
- quantitative measurements, statistical models, relevant data;
- apply state-of-the-art machine learning and data analytics in forensic contexts

Findings

✓ forensic voice comparison using **state-of-the-art automatic-speaker-recognition** (FVC-ASR) is **more accurate** 📈 than speaker identification by **individual listeners**

✗ listeners **overestimate** their ability ✗ **knowing** the FVC-ASR output **does not** help

✓ listeners' accuracy 📉 **unfamiliar accent** 📉 📉 **unfamiliar language**

✗ judges should **not attempt** to perform their own speaker identifications

✗ **even** when also considering the validated FVC-ASR system output

✗ **nor** rely on speaker ID by lay or "ad hoc expert" listeners

Method
~ **60 listeners** each of Aus Eng, North American, Spanish, recruited through **Prolific**
61 pairs Aus Eng recordings in **15s** cuts from *forensic_eval_01* in **case-specific conditions**
calibrated FVC-ASR likelihood ratios from **E³FS³** (ResNet-LDA-PLDA-logistic regression)
1. probabilistic judgements on listening only; **2.** additional experiment providing the FVC-ASR LRs

"The recordings are from the same speaker"

E³FS³ forensic voice comparison system [2]

$$LR = \frac{f(E|H_{SS})}{f(E|H_{DS})}$$

output layer
x-vector layer
third layer
second layer
channel-attention layer
first layer
input

Experiments [1]
1. human *vs* system
2. human + system

❝ I think the properties of the voices on the recordings are [____] times **more likely if they are both recordings of the same adult male Australian-English speaker** than if they are recordings of two different adult male Australian-English speakers.

❝ I think the properties of the voices on the recordings are [____] times **more likely if they are recordings of two different adult male Australian-English speakers** than if they are both recordings of the same adult male Australian-English speaker.

Example human listener strategies

FVC-ASR (E³FS³) | Biased to different-source | Fixed value | Conservative / uncertain

Cumulative proportion vs log₁₀ likelihood ratio

C_{llr} (average accuracy) — lower ≪ — E³FS³ C_{llr} = 0.42

D_{llr} (relative strength of evidence) — stronger ≫ — E³FS³ D_{llr} ≡ 0

B_{llr} (relative bias) — towards same-source SS — E³FS³ B_{llr} ≡ 0

Australian English | North American | Spanish

References
[1] Basu N., Bali A.S., Weber P., Rosas-Aguilar C., Edmond G., Martire K.A., Morrison1 G.S. (2022).Speaker identification in courtroom contexts – Part I: Individual listeners compared to forensic voice comparison based on automatic-speaker-recognition technology. Submitted to Forensic Science International (FSI).

[2] Weber P., Enzinger E., Labrador-Serrano B., Lozano-Díez A., Ramos D., González-Rodríguez J., Morrison G.S. (2022). Validation of the alpha version of the E3 Forensic Speech Science System (E³FS³) core software tools. FSI: Synergy, 4, 100223.