Speaker identification by listeners compared to expert forensic voice comparison based on state-of-the-art automatic-speaker-recognition technology

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

[1] Forensic Data Science Laboratory, Aston University
[2] Forensic Evaluation Ltd
[3] School of Psychology, University of New South Wales
[4] School of Law, Society & Criminology, University of New South Wales
[5] Instituto de Lingüística y Literatura, Universidad Austral de Chile

Aston University
BIRMINGHAM UK

$$\frac{p(E|H_p)}{p(E|H_d)}$$

SCIENTIA
MANU ET MENTE

UNSW
SYDNEY

LIBERTAS CAPITUR
UNIVERSIDAD AUSTRAL DE CHILE

Language

- Sorry, I don’t speak Portuguese.
- This presentation will be in English.
- If you want to speak with me later, I’m fluent in English and Spanish.

Acknowledgements

- This research was supported by Research England's Expanding Excellence in England Fund as part of funding for the Aston Institute for Forensic Linguistics 2019–2024.

Disclaimer

- All opinions expressed are those of the presenter and, unless explicitly stated otherwise, should not be construed as representing the policies or positions of any organizations with which the presenter is associated.

Slides

- http://geoff-morrison.net/

Publications

- Basu N., Bali A.S., Weber P., Rosas-Aguilar C., Edmond G., Martire K.A., Morrison G.S. (2022). **Speaker identification in courtroom contexts – Part I: Individual listeners compared to forensic voice comparison based on automatic-speaker-recognition technology**. *Forensic Science International*, 111499. https://doi.org/10.1016/j.forsciint.2022.111499

- Basu N., Weber P., Bali A.S., Rosas-Aguilar C., Edmond G., Martire K.A., Morrison G.S. (2023). **Speaker identification in courtroom contexts – Part II: Investigation of bias in individual listeners' responses**. *Forensic Science International*, 349, 111768. https://doi.org/10.1016/j.forsciint.2023.111768

- Bali A.S., Basu N., Weber P., Rosas-Aguilar C., Edmond G., Martire K.A., Morrison G.S. (2023). **Speaker identification in courtroom contexts – Part III: Groups of collaborating listeners compared to forensic voice comparison based on automatic-speaker-recognition technology**. Manuscript submitted for publication.

- Supplementary material:

  https://forensic-voice-comparison.net/speaker-recognition-by-humans/

Vocabulary

- **Speaker identification** by lay listeners refers to situations where a **listener who is unfamiliar with the speaker or speakers** listens to:
    - a voice they hear on one occasion (e.g., while a crime is being committed) and a voice that they hear on another occasion (e.g., during a voice lineup);
    - two voice recordings (e.g., recording of a crime being committed and a recording of a police interview);
    - or one voice recording (e.g., a recording of a crime being committed) and a live speaker (e.g., a defendant speaking in court);
- and **attempts to determine whether they are the same speaker or different speakers**.

Vocabulary

- **Speaker identification** by lay listeners refers to situations where a **listener who is unfamiliar with the speaker or speakers** listens to:
    - a voice they hear on one occasion (e.g., while a crime is being committed) and a voice that they hear on another occasion (e.g., during a voice lineup);
    - two voice recordings (e.g., recording of a crime being committed and a recording of a police interview);
    - or one voice recording (e.g., a recording of a crime being committed) and a live speaker (e.g., a defendant speaking in court);
- and **attempts to determine whether they are the same speaker or different speakers**.

Research questions

- **Expert testimony is only admissible in common-law jurisdictions if it will potentially assist the trier of fact to make a decision.**

    - Is **speaker identification by a judge listening alone** more or less accurate than the output of a **forensic-voice-comparison system** that is based on state-of-the-art automatic-speaker-recognition technology?

    - Is **speaker identification by jury members listening and collaboratively making a judgement** more or less accurate than the output of a **forensic-voice-comparison system** that is based on state-of-the-art automatic-speaker-recognition technology?

Research questions

- **Expert testimony is only admissible in common-law jurisdictions if it will potentially assist the trier of fact to make a decision.**

  - Is **speaker identification by a judge listening alone** more or less accurate than the output of a **forensic-voice-comparison system** that is based on state-of-the-art automatic-speaker-recognition technology?

  - Is **speaker identification by jury members listening and collaboratively making a judgement** more or less accurate than the output of a **forensic-voice-comparison system** that is based on state-of-the-art automatic-speaker-recognition technology?

Research questions

- Triers of fact sometimes attempt to perform speaker identification on speech that is in an accent that is unfamiliar to them or even a language that is unfamiliar to them.

  - Is the accuracy of a judge's speaker identification better or worse when the speech is in an **unfamiliar accent**?

  - Is the accuracy of a judge's speaker identification better or worse when the speech is in an **unfamiliar language**?

Stimuli

- Pairs of recordings:
    - 31 same-speaker pairs
    - 30 different-speaker pairs

Stimuli

- Pairs of recordings:
    - 31 same-speaker pairs
    - 30 different-speaker pairs
    - each recording
        - ~15 s long
        - adult male speaker of Australian English

Stimuli

- Pairs of recordings reflect the conditions of a real forensic case:
    - Questioned-speaker condition
        - landline-telephone call
        - background babble noise
        - saved using lossy compression
    - Known-speaker condition
        - interview recorded in a reverberant room
        - background ventilation-system noise

Listeners

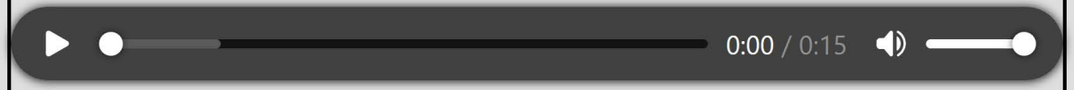
- Australian-English listeners (53)
- North-American-English listeners (61, 57)
- Spanish-language listeners (55)

Procedures for listeners

Instructions

Recording Pair 1 of 66

Questioned Speaker Recording:



Known Speaker Recording:

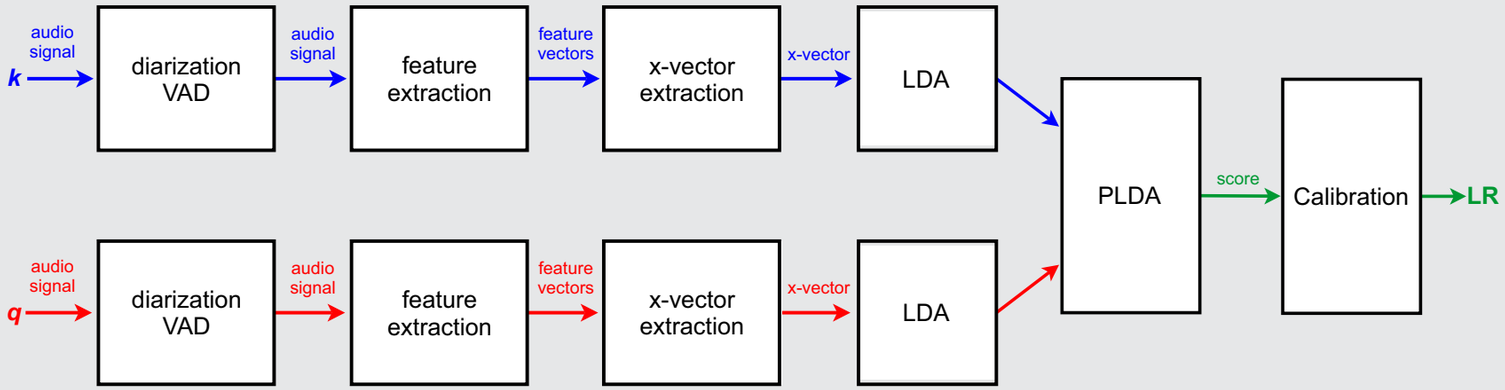
I think the properties of the recordings are [ ] times **more likely if they are both recordings of the same adult male Australian-English speaker** than if they are recordings of two different adult male Australian-English speakers.

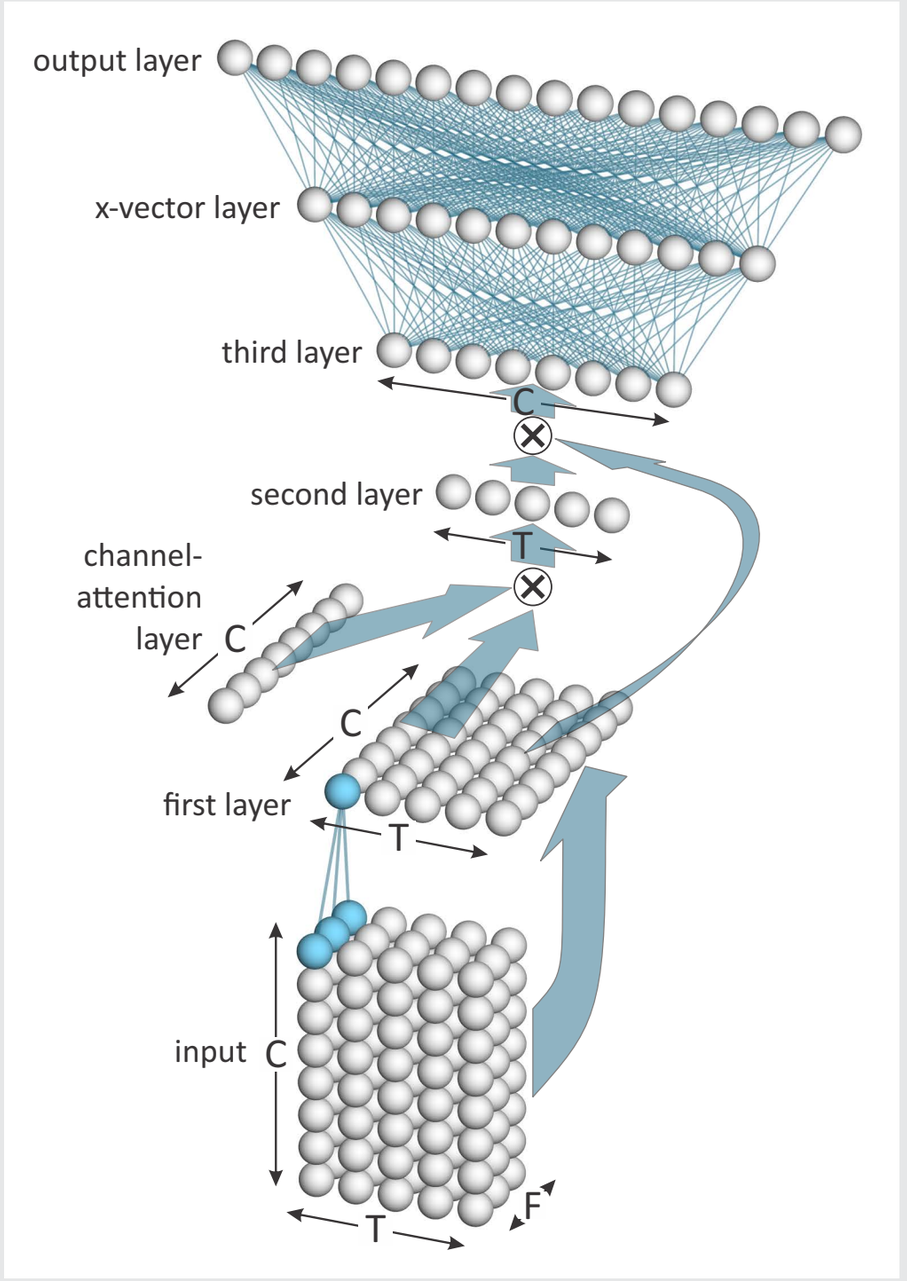
I think the properties of the recordings are [ ] times **more likely if they are recordings of two different adult male Australian-English speakers** than if they are both recordings of the same adult male Australian-English speaker.

NEXT

Forensic-voice-comparison system

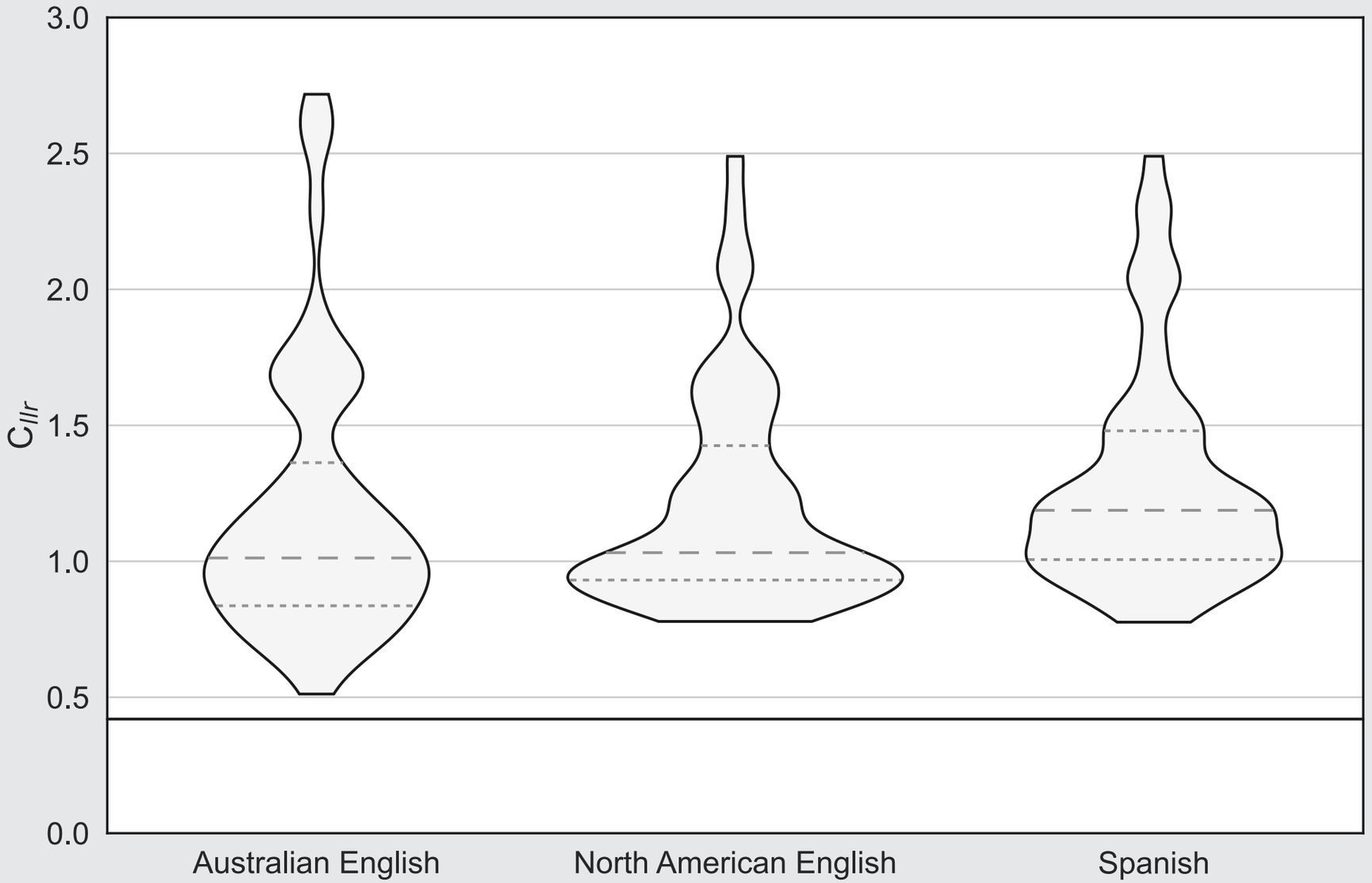
- E^3 Forensic Speech Science System (E^3FS^3)
  - x-vector (DNN-embedding) based
  - **calibrated** under casework conditions





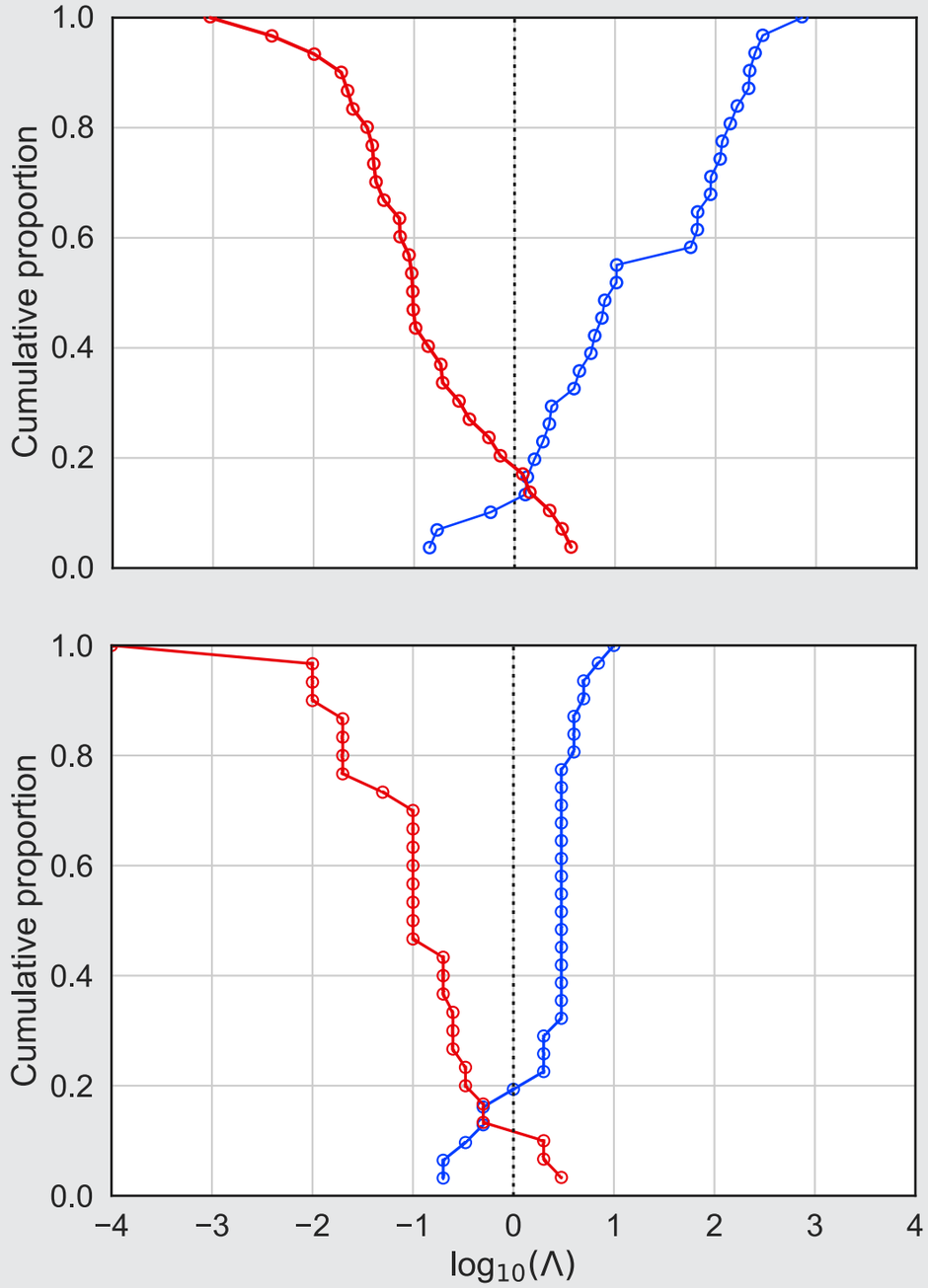
Results

- Accuracy
  - log-likelihood-ratio cost
  - C_{llr}



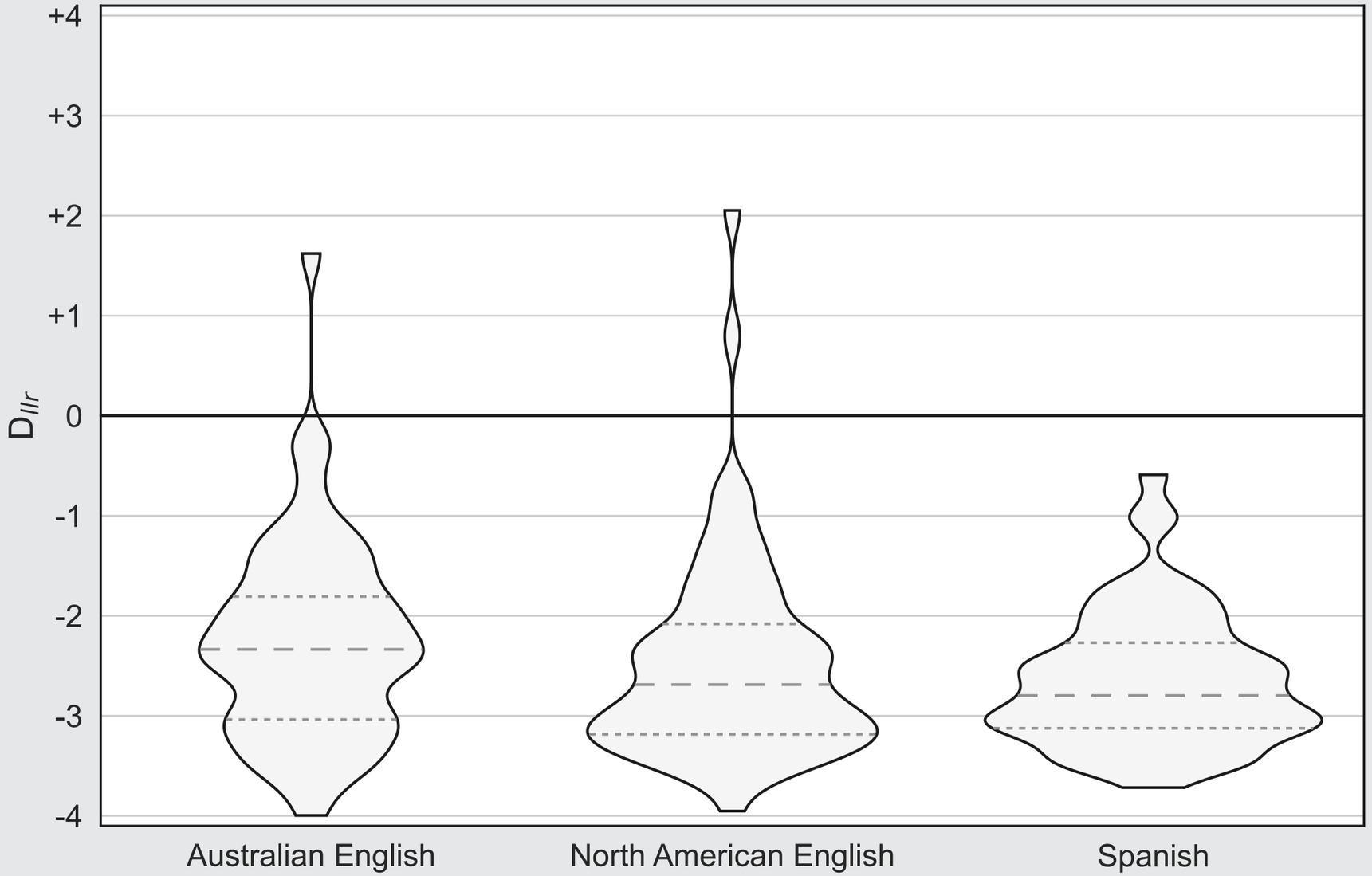
Results

- Forensic-voice-comparison system
  - $C_{\text{llr}} = \mathbf{0.42}$
- Best listener
  - $C_{\text{llr}} = \mathbf{0.51}$
  - $D_{\text{llr}} = -1.3$
  - $B_{\text{llr}} = -1.5$



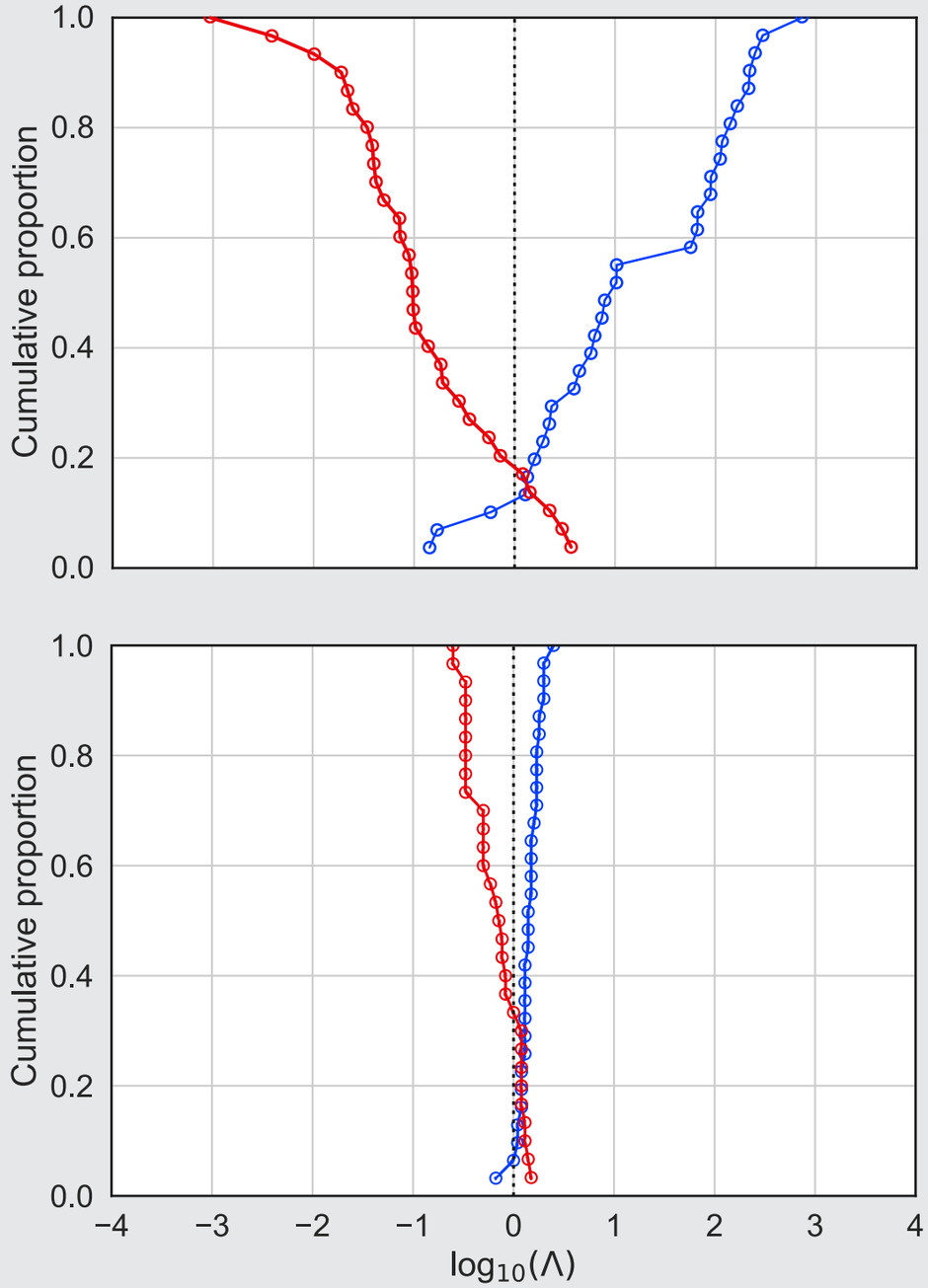
Results

- Discrimination relative to FVC system
  - D_{llr}



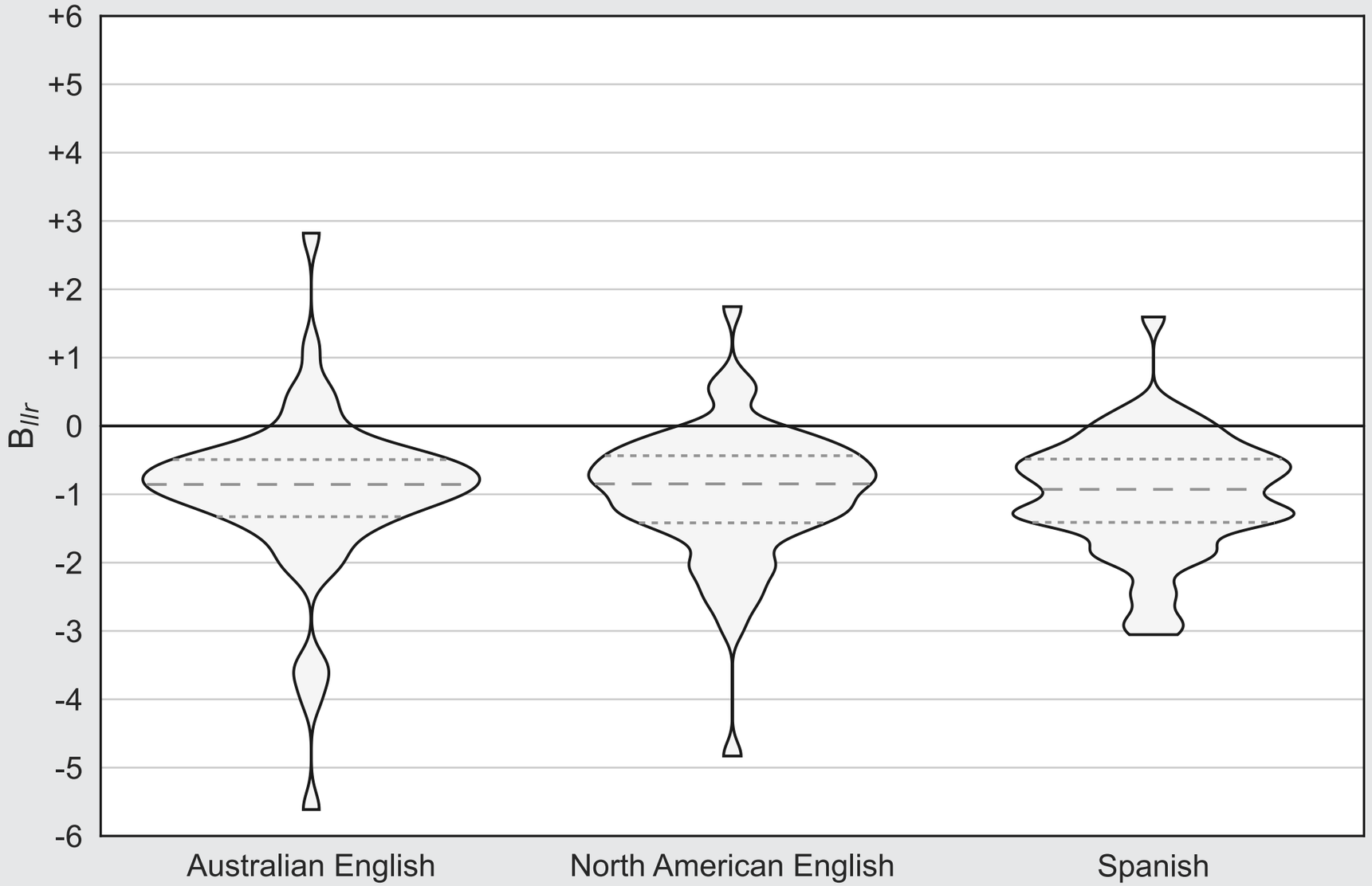
Results

- Forensic-voice-comparison system
  - $C_{\text{llr}} = 0.42$
- Example of listener with poor discrimination
  - $C_{\text{llr}} = 0.77$
  - $\boldsymbol{D_{\text{llr}} = -2.9}$
  - $B_{\text{llr}} = -0.5$



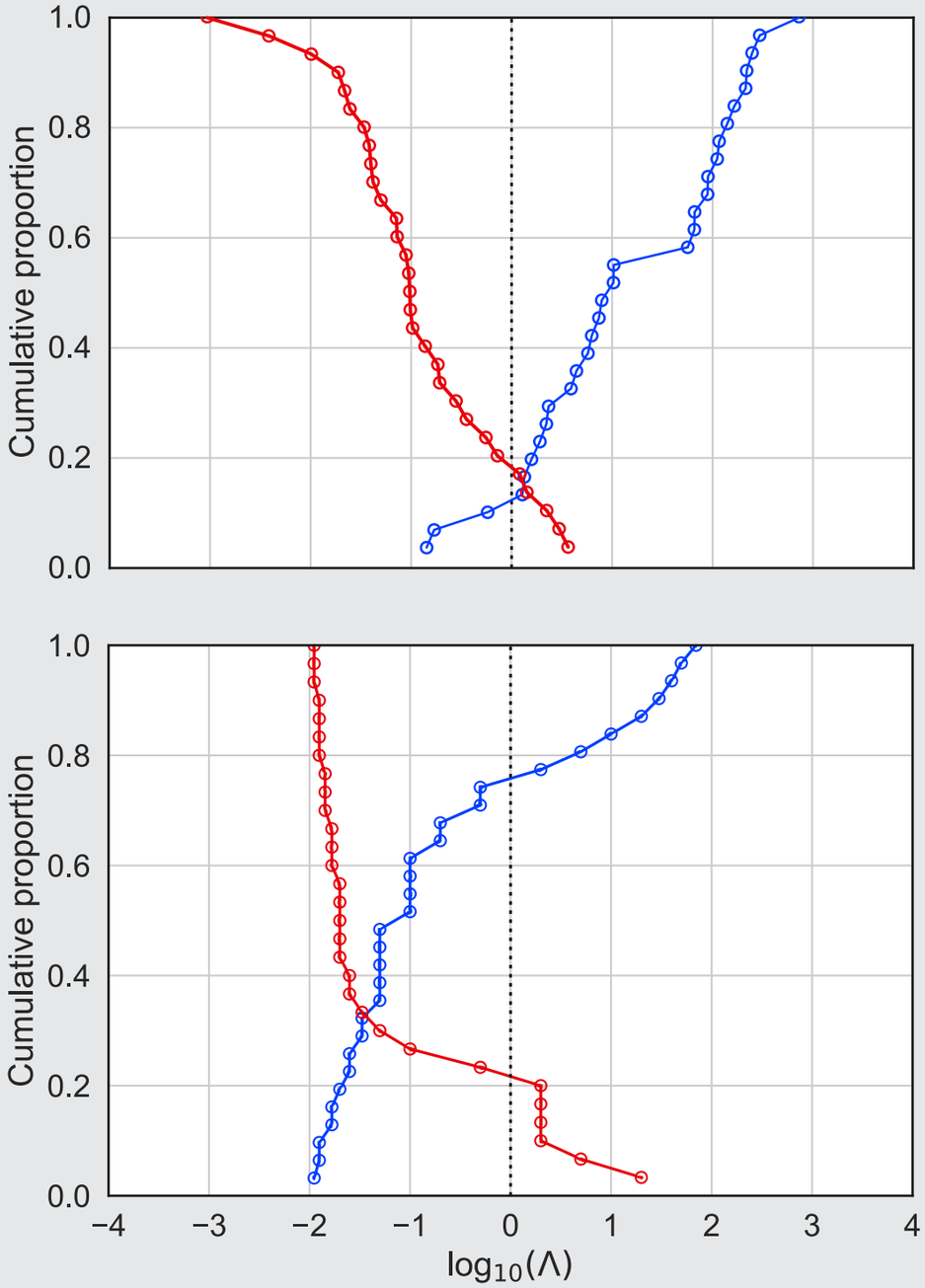
Results

- Bias relative to FVC system
  - B_{llr}



Results

- Forensic-voice-comparison system
  - $C_{\text{llr}} = 0.42$
- Example of listener with strong bias toward the different-speaker hypothesis
  - $C_{\text{llr}} = 1.90$
  - $D_{\text{llr}} = -2.5$
  - $\boldsymbol{B_{\text{llr}} = -3.5}$



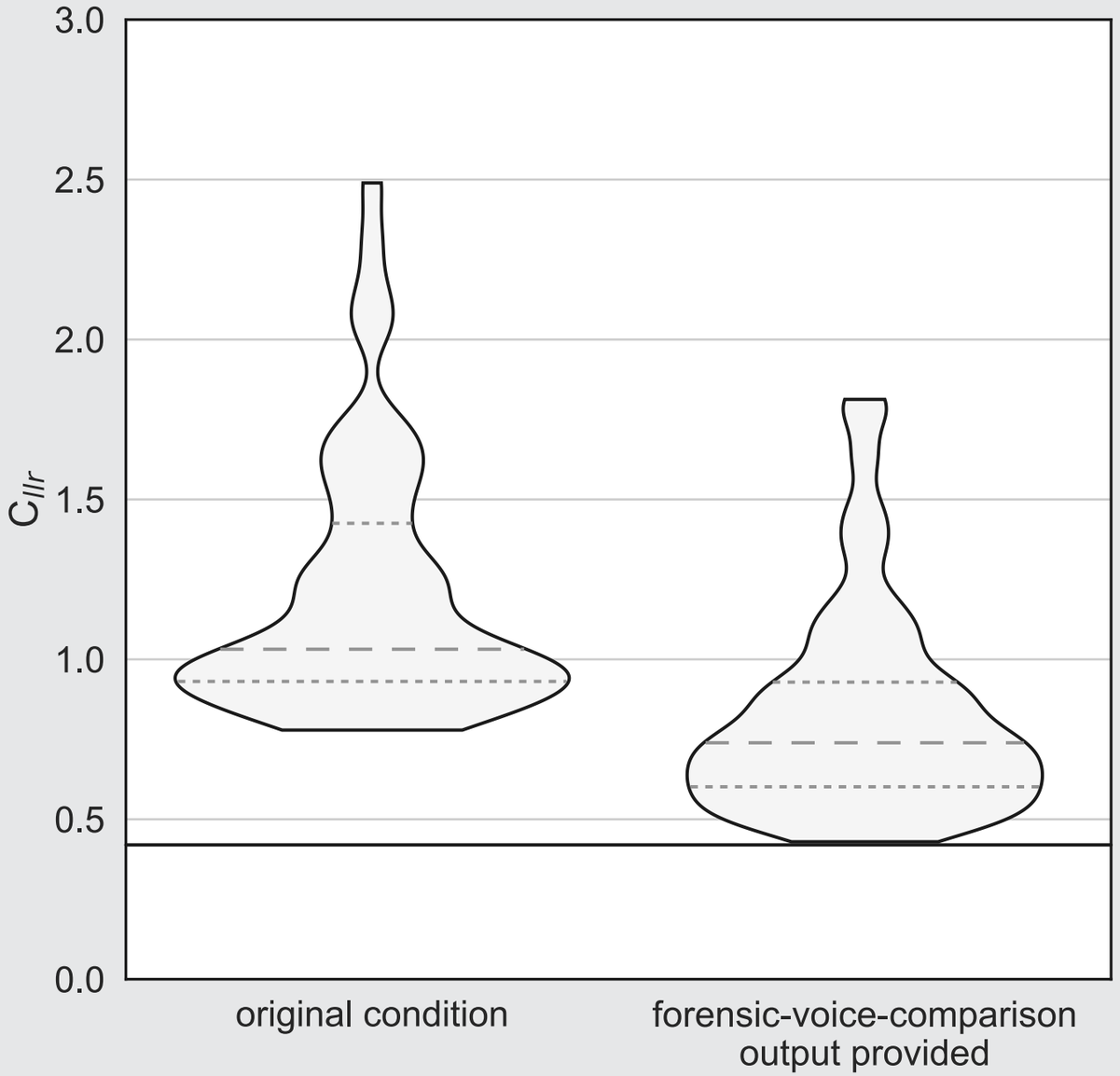
Research question

- When presented with expert evidence on forensic voice comparison, triers of fact usually also listen to the recordings and also attempt to perform their own speaker identification.

  - Is **speaker identification by a judge who both listens to the recordings and considers the output of the forensic-voice-comparison system** more or less accurate than the output of a **forensic-voice-comparison system alone**?

Results

- Accuracy
  - log-likelihood-ratio cost
  - C_{llr}



Research questions

- **Expert testimony is only admissible in common-law jurisdictions if it will potentially assist the trier of fact to make a decision.**

  - Is **speaker identification by a judge listening alone** more or less accurate than the output of a **forensic-voice-comparison system** that is based on state-of-the-art automatic-speaker-recognition technology?

  - Is **speaker identification by jury members listening and collaboratively making a judgement** more or less accurate than the output of a **forensic-voice-comparison system** that is based on state-of-the-art automatic-speaker-recognition technology?

Stimuli

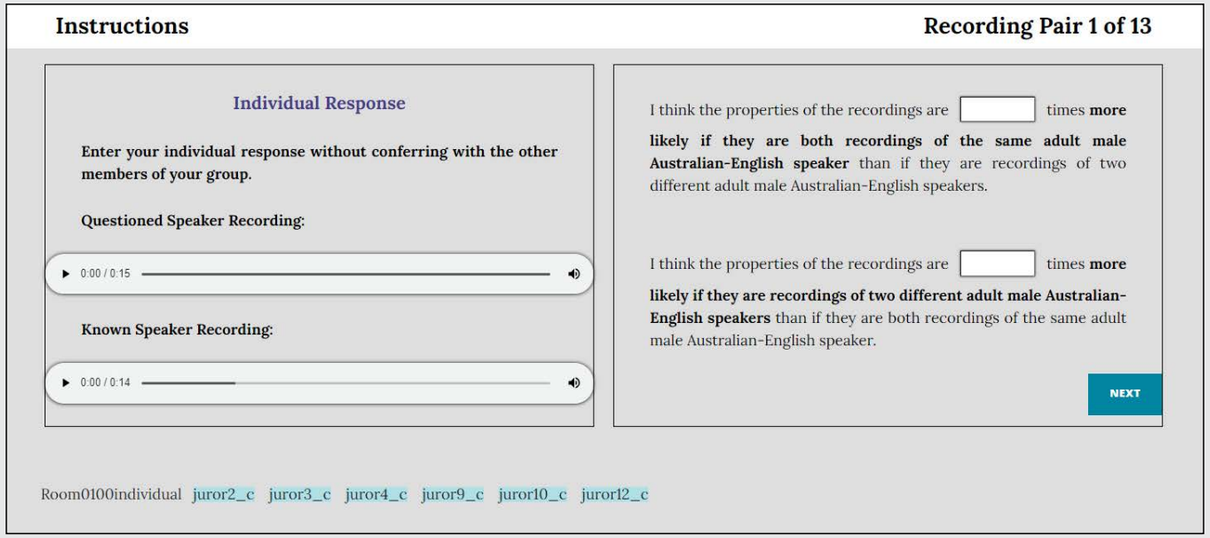
- Pairs of recordings:
    - 12 same-speaker pairs
    - 12 different-speaker pairs

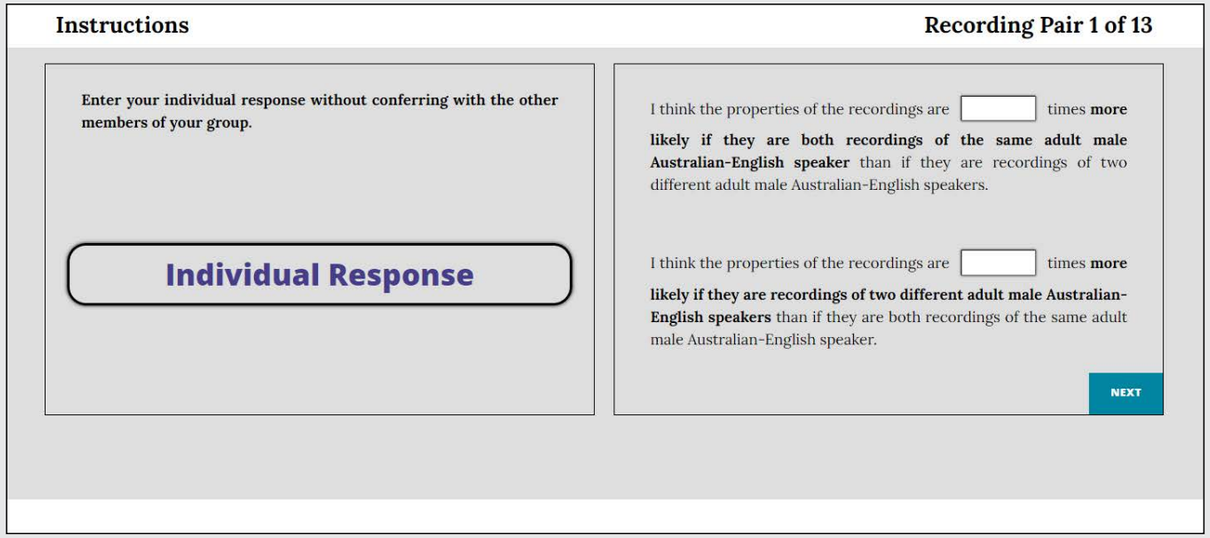
Listeners

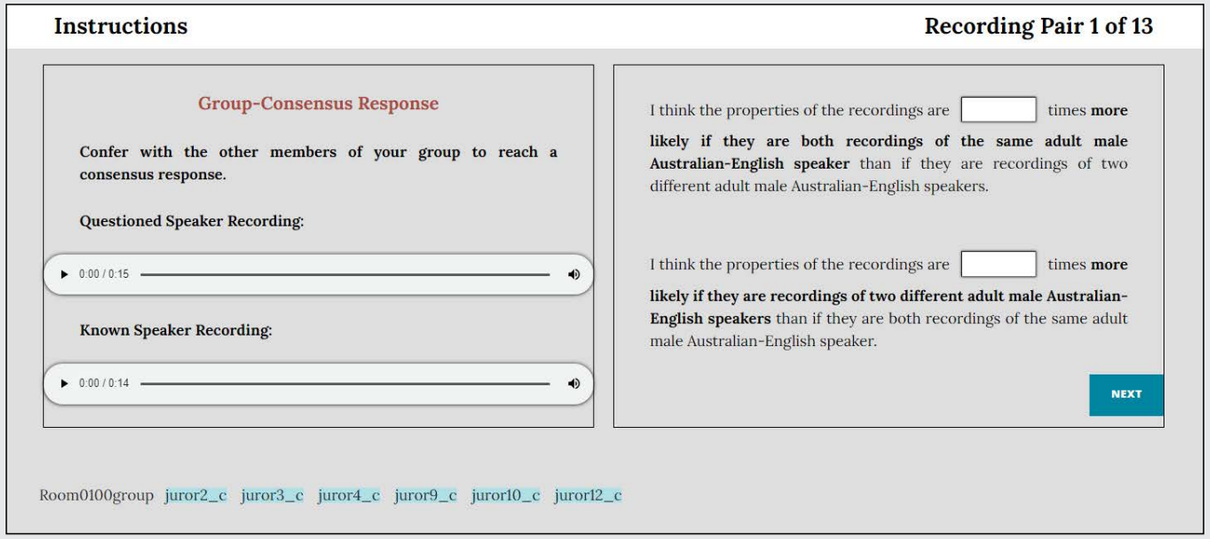
- Australian-English listeners

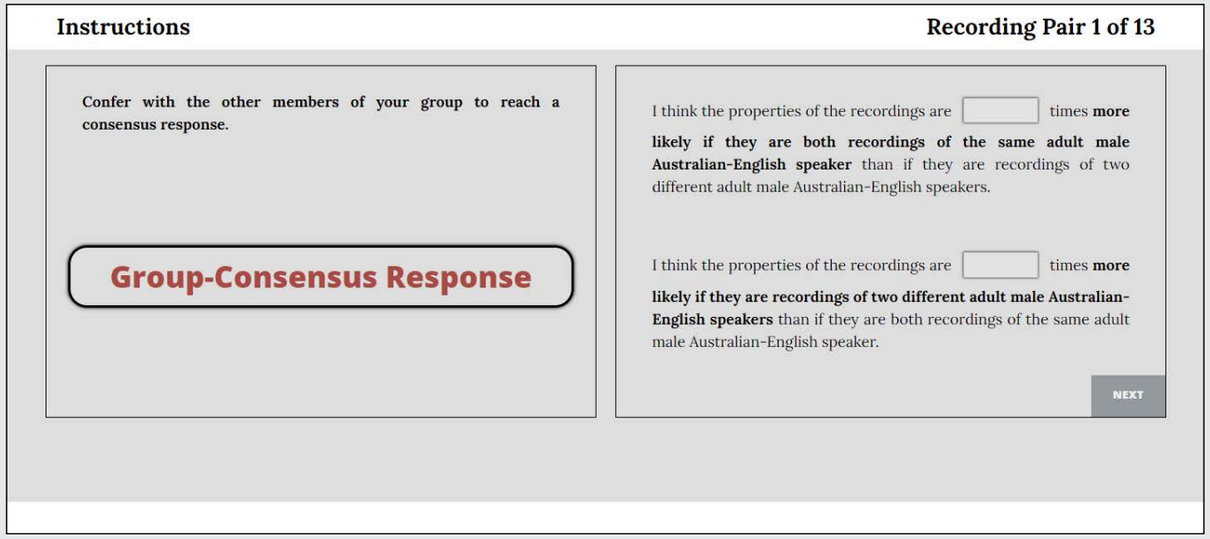
- 23 groups ranging in size from 5 to 12 listeners

Procedures for listeners



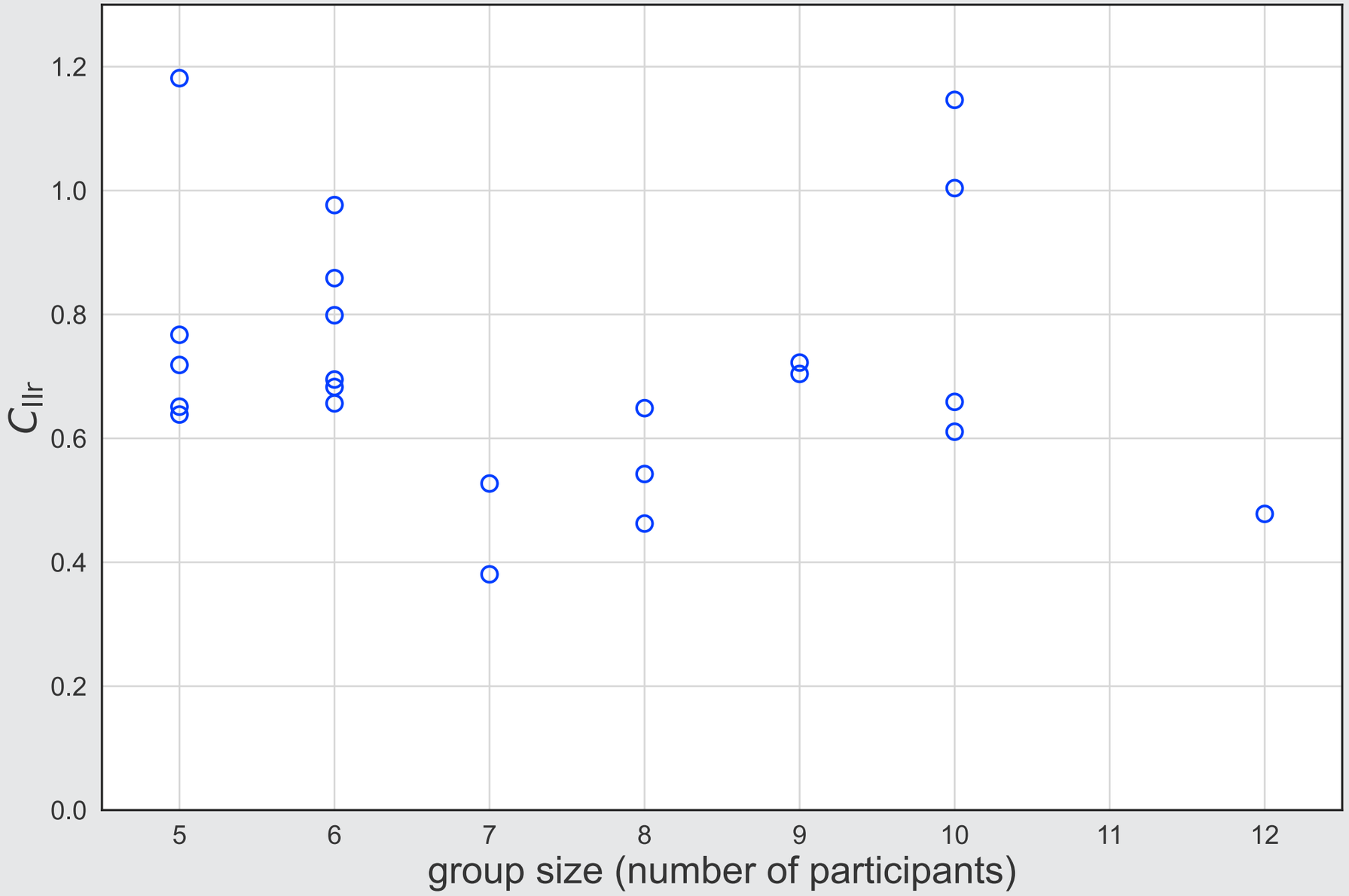






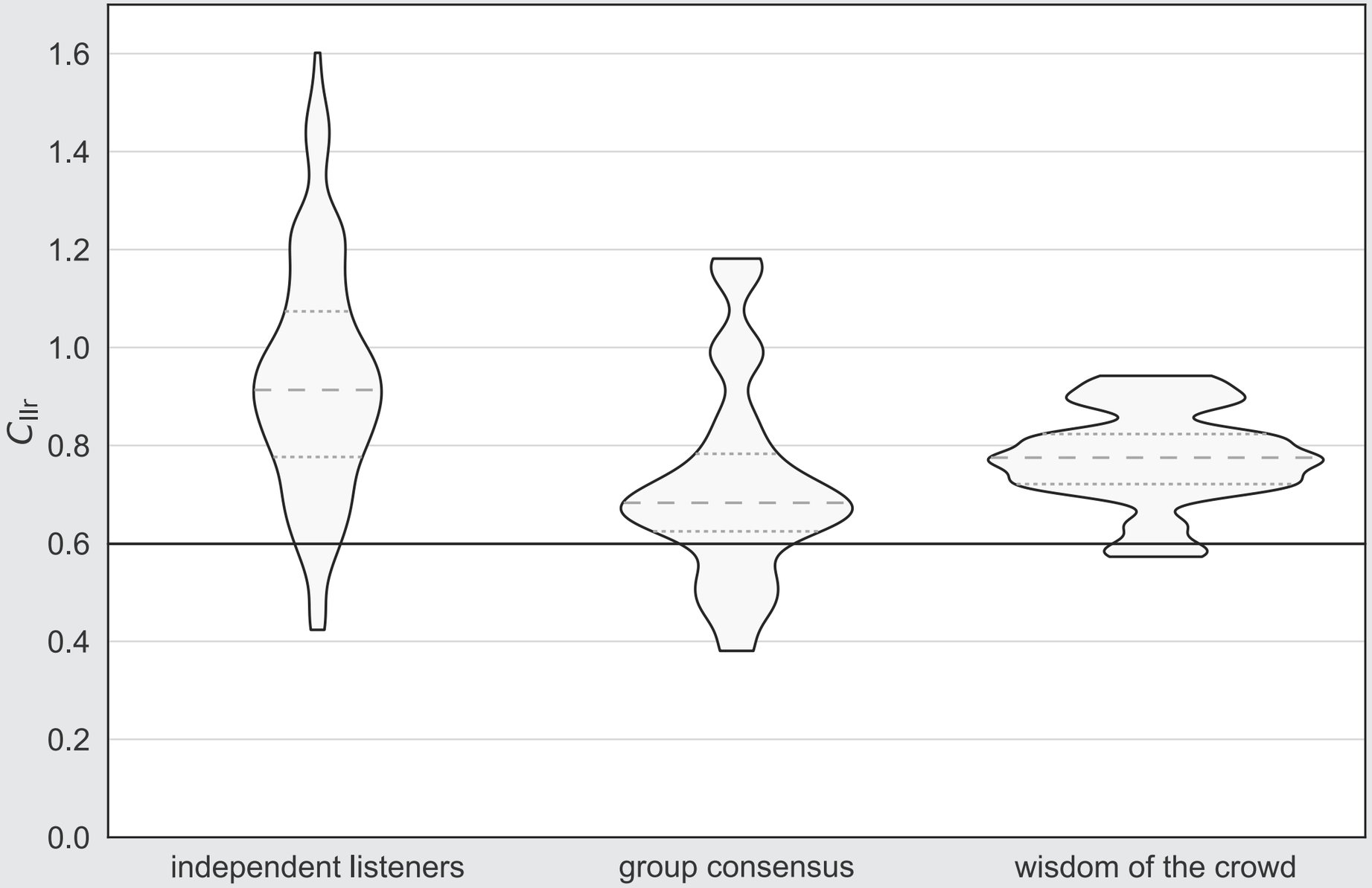
Results

- Accuracy
  - log-likelihood-ratio cost
  - C_{llr}



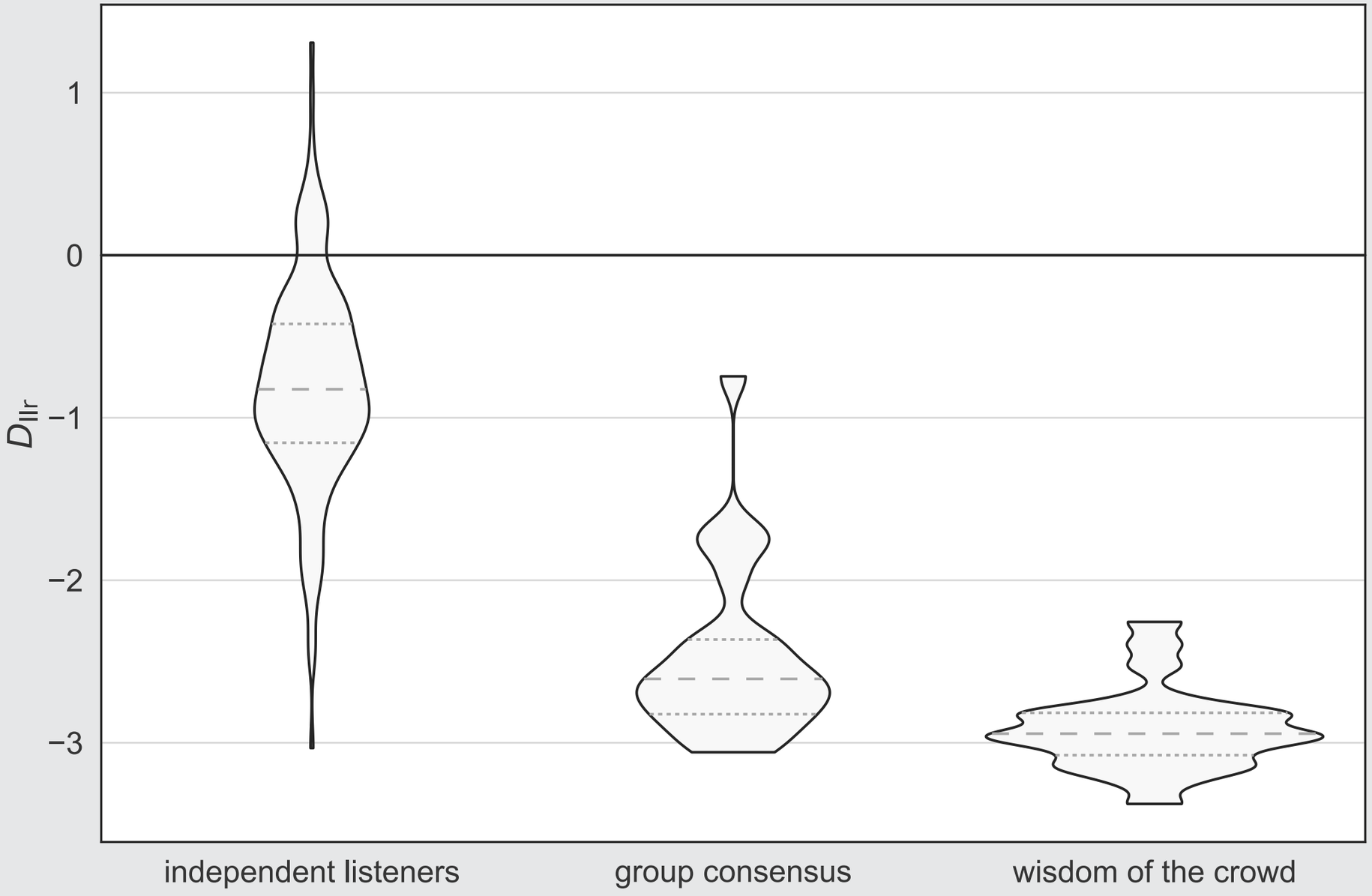
Results

- Accuracy
  - log-likelihood-ratio cost
  - C_{llr}



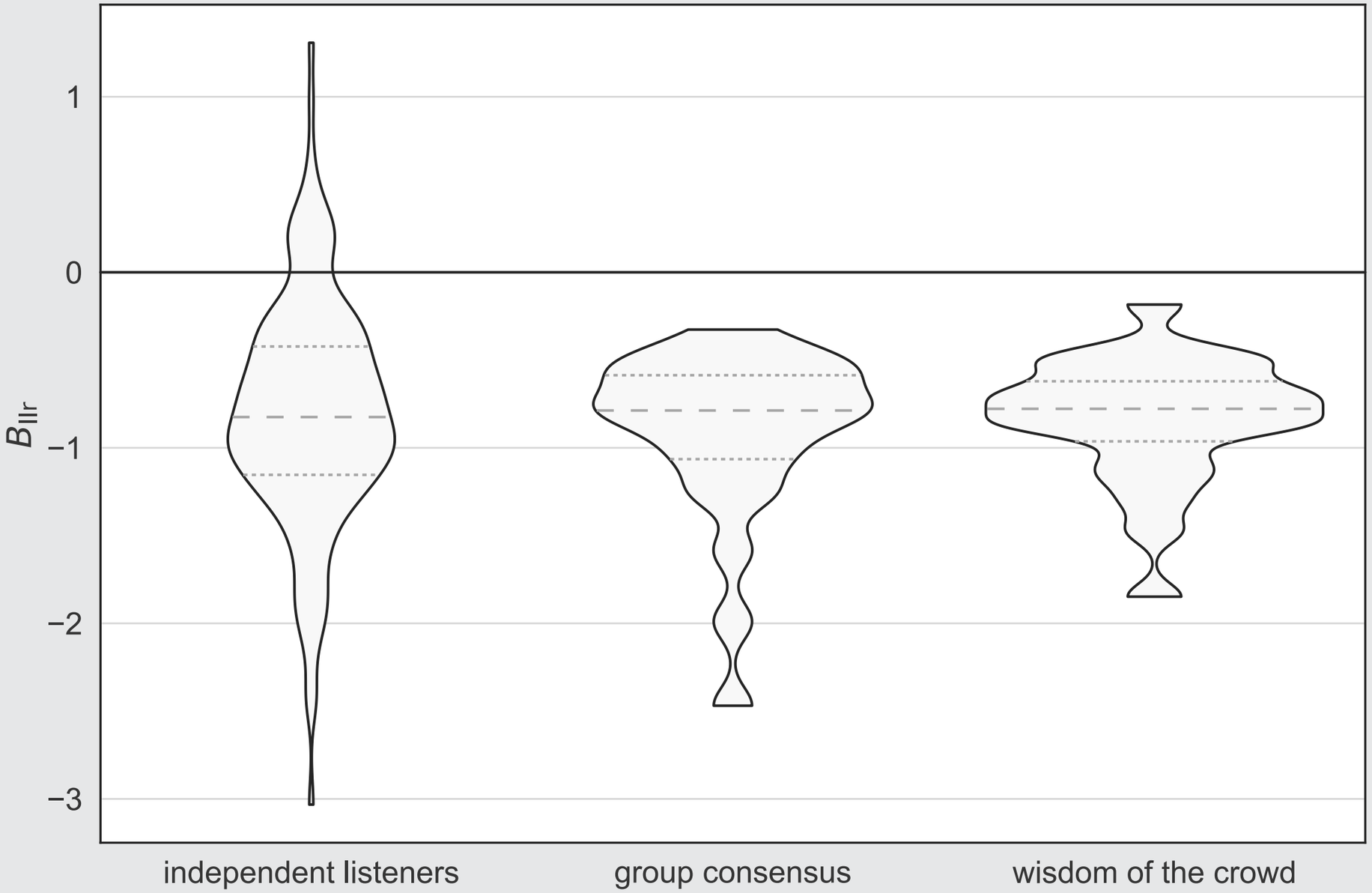
Results

- Discrimination relative to FVC system
  - D_{llr}



Results

- Bias relative to FVC system
  - B_{llr}



Conclusions

- Is forensic voice comparison based on state-of-the-art automatic-speaker-recognition technology more accurate than speaker identification by individual lay listeners?
    - Yes

Conclusions

- Is forensic voice comparison based on state-of-the-art automatic-speaker-recognition technology more accurate than speaker identification by individual lay listeners?
    - Yes
- Is forensic voice comparison based on state-of-the-art automatic-speaker-recognition technology more accurate than speaker identification by groups of listening whose members collaboratively make a speaker-identification judgement?
    - Yes

Conclusions

- Is the accuracy of individual lay listeners' speaker identification worse when the speech is in an unfamiliar accent and even worse when it is in an unfamiliar language?
    - Yes

Conclusions

- Is the accuracy of individual lay listeners' speaker identification worse when the speech is in an unfamiliar accent and even worse when it is in an unfamiliar language?
    - Yes
- Can individual lay listeners outperform forensic voice comparison based on state-of-the-art automatic-speaker-recognition technology by considering the likelihood ratio output by the forensic-voice-comparison system and also performing their own speaker identification?
    - No

Recommendations

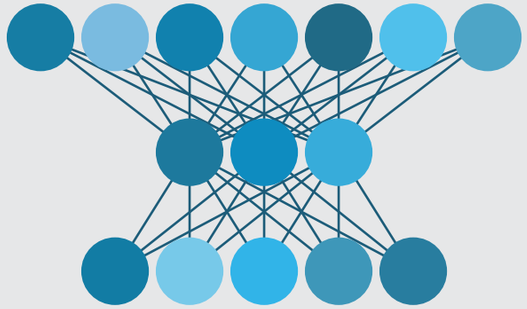
- Should judges and juries attempt to perform their own speaker identification?
    - No.
    - They should rely on expert testimony based on a calibrated and validated forensic-voice-comparison system.

Recommendations

- Should judges and juries attempt to perform their own speaker identification in addition to considering the likelihood ratio output by a forensic-voice-comparison system?
    - No.
    - They should rely exclusively on expert testimony based on a calibrated and validated forensic-voice-comparison system.

Thank You

http://forensic-data-science.net/

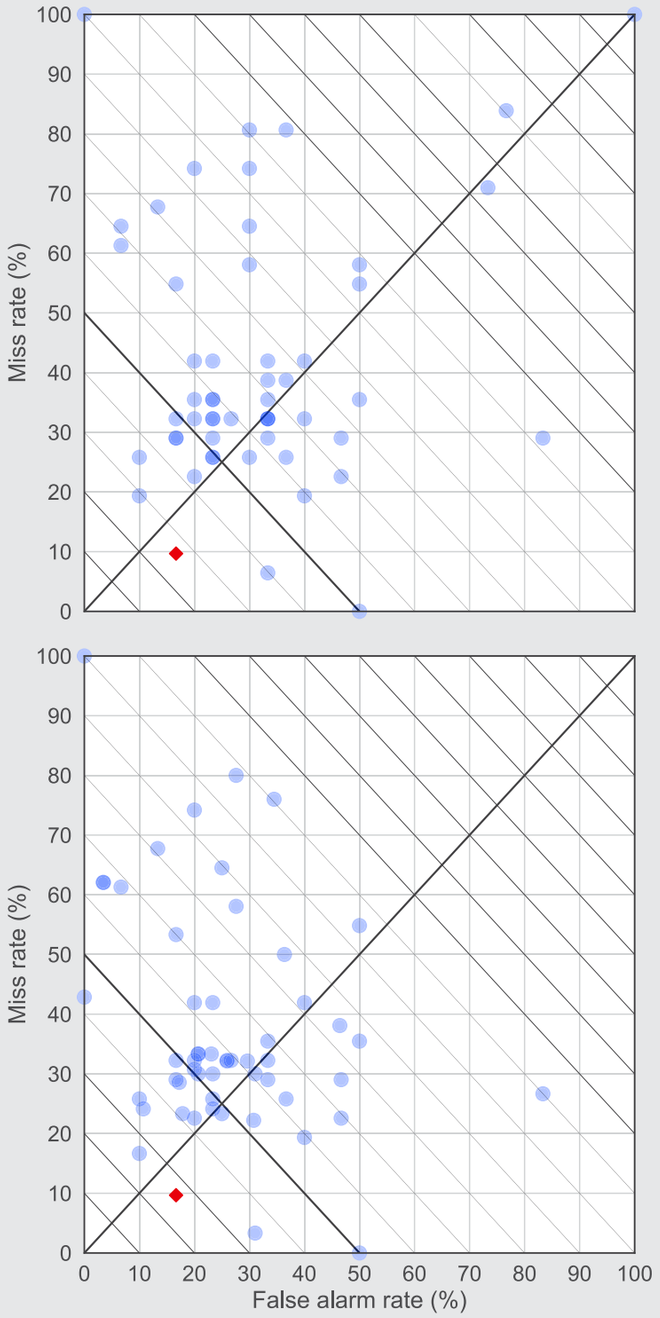


Results - individuals

- Accuracy
    - miss rate
    - false-alarm rate

top: “1” treated as error

bottom: “1” ignored



Results - group concensus

- Accuracy
    - miss rate
    - false-alarm rate

top: “1” treated as error

bottom: “1” ignored