Nottingham Trent University

Does multimodal identity information improve face and voice identity matching accuracy?

Harriet Smith

harriet.smith02@ntu.ac.uk

14/04/2024

Queen Mary
University of London

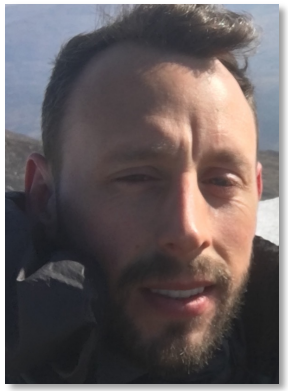
UNIVERSITY OF
LINCOLN

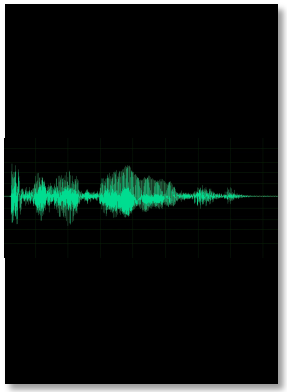
Pre-print

SCAN ME

Face and voice matching can be error-prone

(Bruce et al., 1999; Davis & Valentine, 2009; Megreya & Burton, 2006, 2008; Smith et al., 2018; Stevenage & Neil, 2014)



Same person or different people?

Comparing face and voice identity perception

Voice identity perception is overall more error-prone than face identity perception

In the current study we adapt traditional unimodal matching tasks by including multimodal information in integrated and interacting

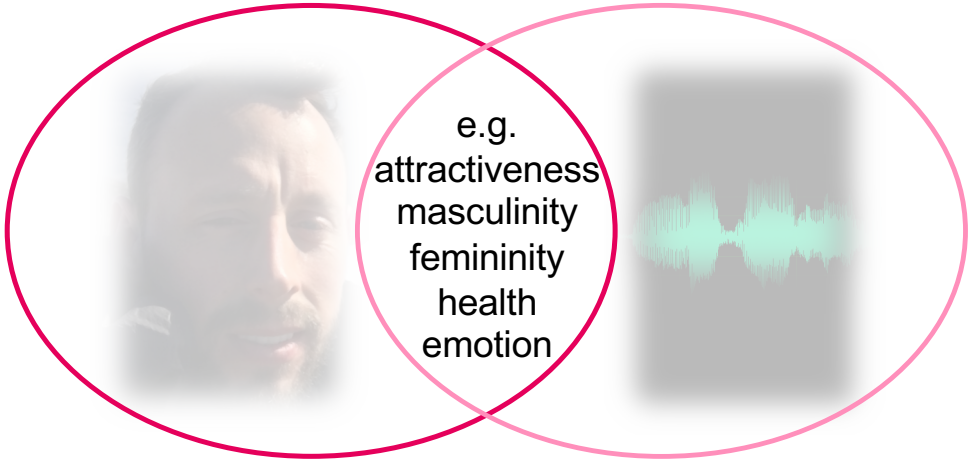
(Belin, 2017; Belin et al., 2004; Maguinness & von Kriegstein, 2017; Young et al., 2020)

NTU

The potential benefit of multimodal information

Faces and voices signal concordant information

(Collins & Missing, 2003; Pourtois & Dhar, 2012; Saxton et al., 2006; Smith et al., 2016)



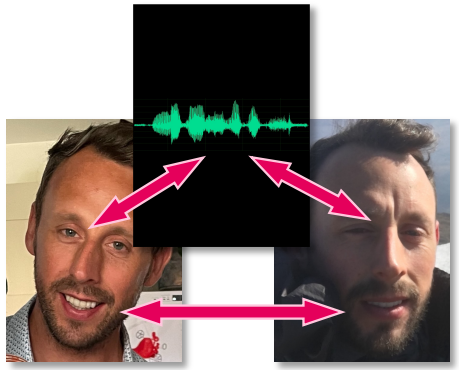
Some studies have even shown that crossmodal matching is possible

(Krauss et al., 2002; Kamachi et al., 2003; Lachs & Pisoni, 2004; Mavica & Barenholtz, 2013; Smith et al., 2016a, 2016b; Stevenage et al., 2017)

NTU

Within- or between-person variability?

Within- or between-person variability?



Might a concordant other-modality stimulus be helpful?

Triangulate the matching decision

Unimodal similarity/dissimilarity

Do the face and voice match?

Three experiments

Experiment 1

Experiment 1

Does having access to identity-related information from both the **face *and* the voice** enhance accuracy for identity matching compared to when information from only one modality is present (i.e., the **face *or* the voice**).

Experiment 1: Stimuli

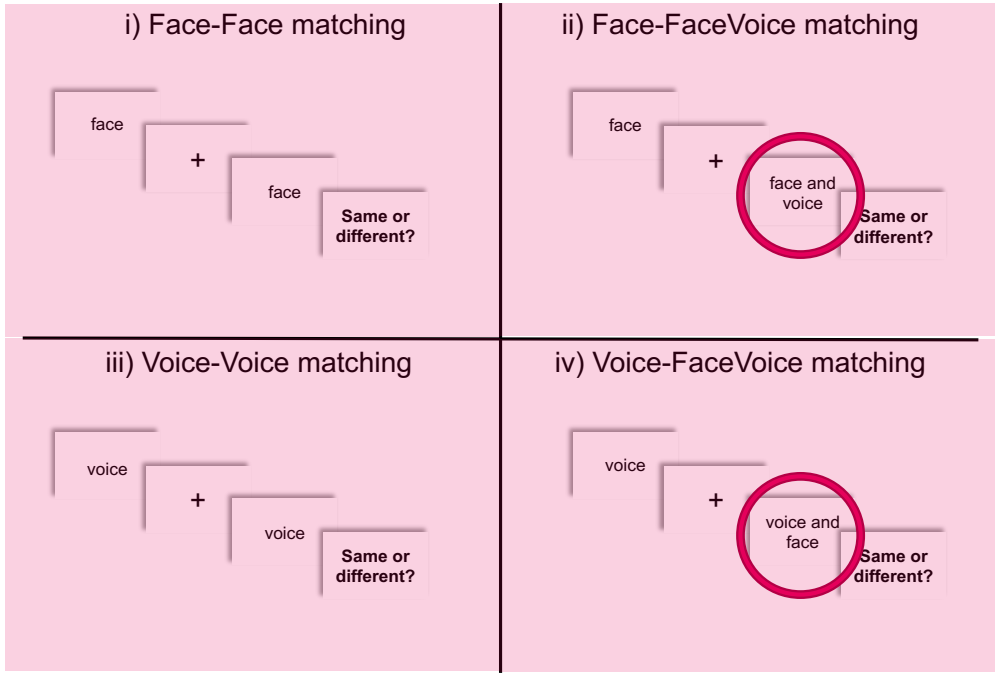
YouTube

- 48 local celebrities from Canada and Australia (24 male; 24 female)
- Video - 2-3 second meaningful utterance
  - **Voice**: audio-only version
  - **Face**: video-only version (audio muted)
  - **Face-voice**: audiovisual version

NTU

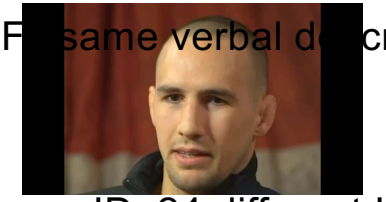
Experiment 1: Procedure

N = 199 (106 male)



Same ID: Different occasions. Variability in lighting, facial expression, hairstyle etc

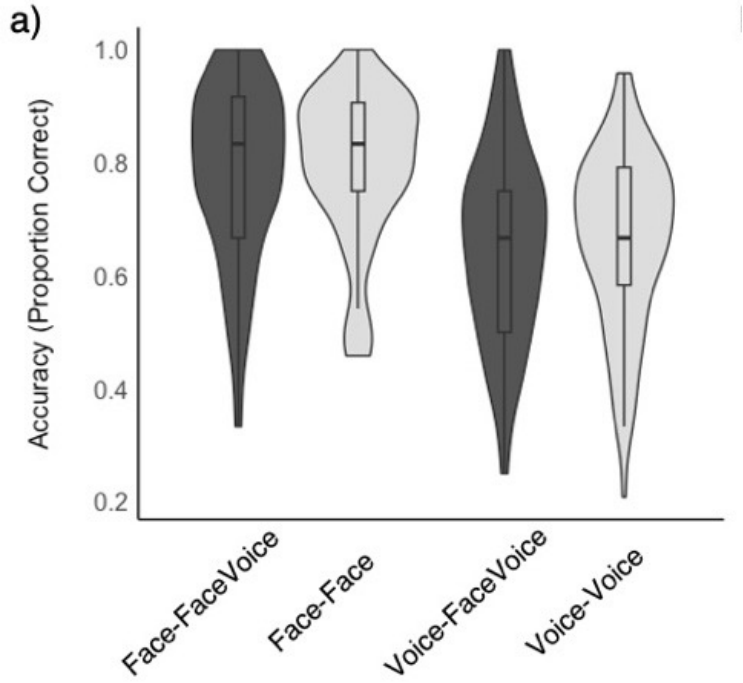
Different ID: F same verbal d cription



48 trials: 24 same ID; 24 different ID

Stimulus order counterbalanced

Experiment 1: Results



Fixed effects

matching task (face vs voice)

other-stimulus modality (unimodal face/voice vs multimodal facevoice)

Random effects

participant

first stimulus in the matching trial

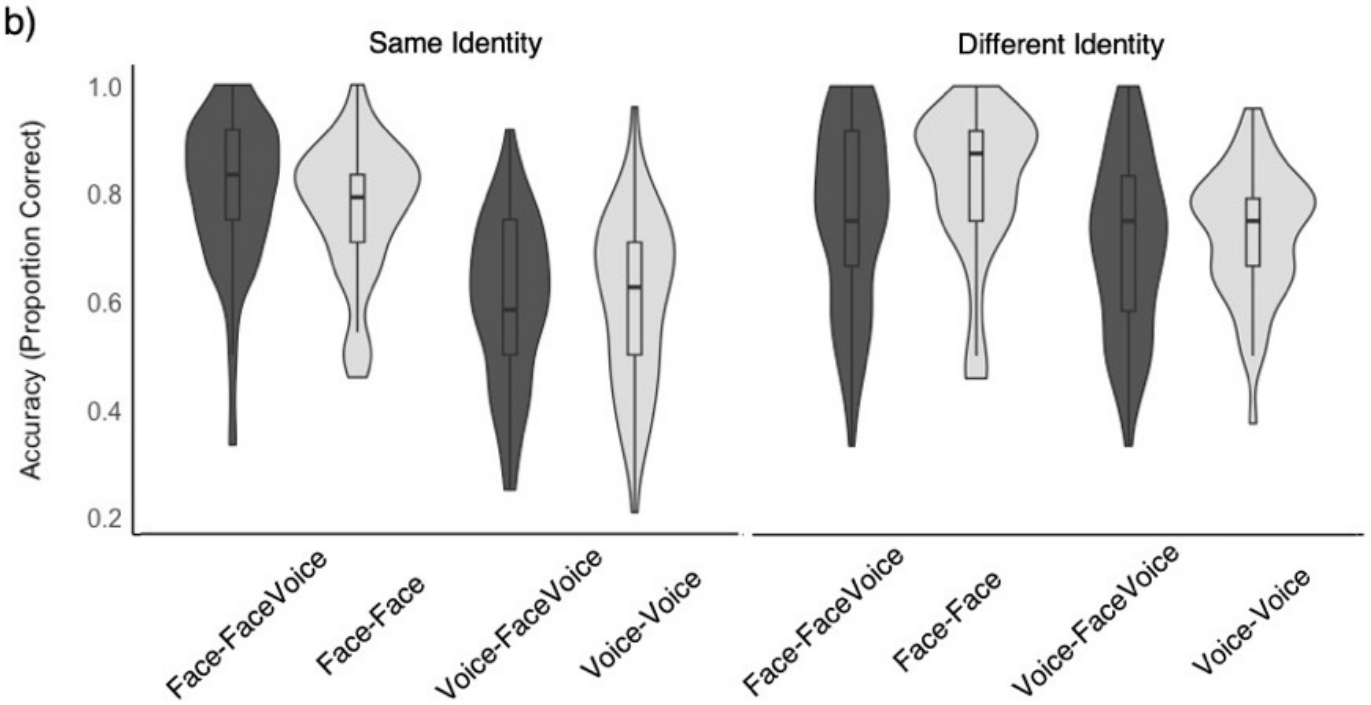
second stimulus in the matching trial

country of origin (Australia vs Canada)

Does having access to multimodal information during face and voice matching tasks improve accuracy?

NO

Experiment 1: Results



Fixed effects

matching task (face vs voice)

other-stimulus modality (unimodal face/voice vs multimodal facevoice)

trial type (same vs different identity within a stimulus pair)

Experiment 1

Conclusions

- No evidence that multimodal information as part of a matching task increases the accuracy of identity matching
- EITHER faces and voices don't provide sufficiently concordant information
- OR, if they do, this information is not used to improve matching accuracy

NTU

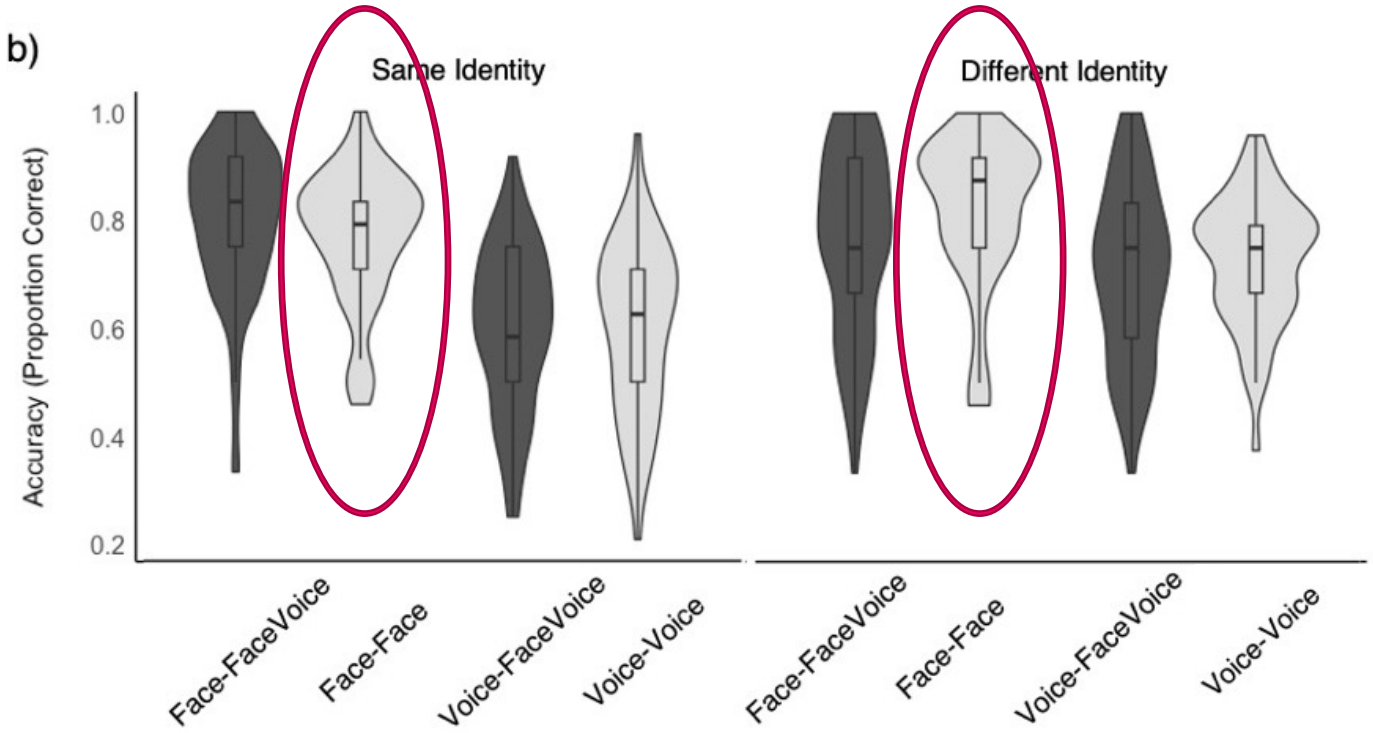
Experiment 2

Experiment 2

Does having access to **multimodal information for *both* stimuli** in a matching trial improve accuracy?

Experiment 2

Comparing new multimodal data (Experiment 2) to the face-face matching data from Experiment 1



Experiment 2: Procedure

N = 51 (24 male)

face and voice

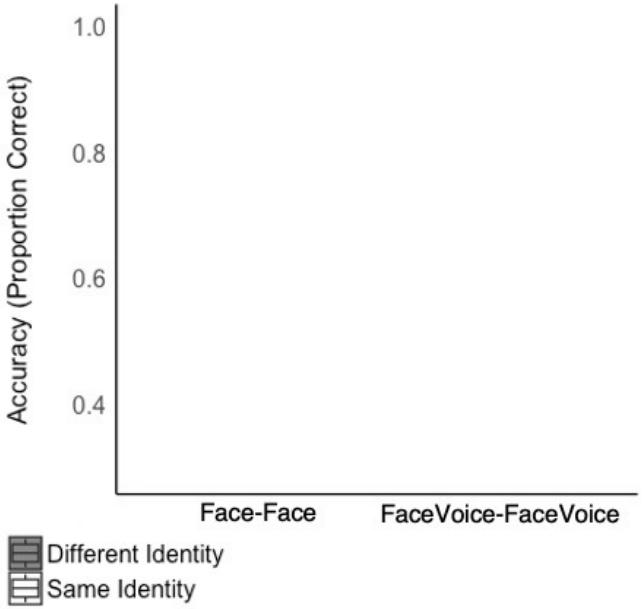
+

face and voice

Same or different?

NTU

Experiment 2: Results



Fixed effects

matching task (face-face vs facevoice-facevoice)

trial type (same vs different)

Does having access to consistent multimodal information – as opposed to only having access to information from the face – improve matching accuracy?

NO

Experiment 2

Conclusions

- Voices provide some identity information
- BUT this information is **not** used by perceivers to enhance their matching performance
- Reliance on face identity information

NTU

Experiment 3

The story so far…

When audiovisual information is involved in identity verification:

- Face-voice concordance is limited or inconsequential
- People don’t really use the voice information anyway!

Experiment 3: Do people look and sound similar?

How well can participants **cross-modally match** faces to voices in our stimulus set?

Can **item effects** can shed more light on the results from Experiments 1 and 2?

Experiment 3: Procedure

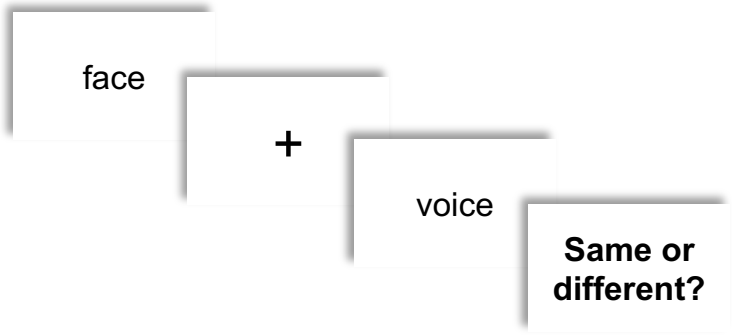
face

+

voice

Same or different?

Experiment 3: Results



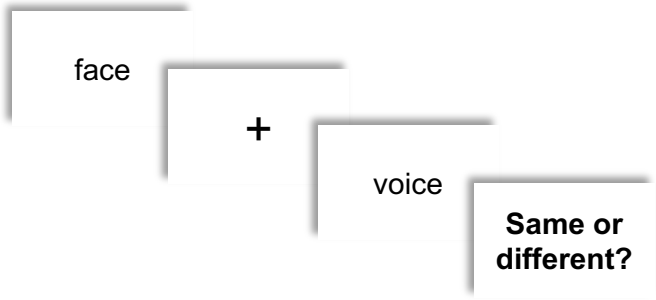
Different identity

45.6%; 95% CIs [34.4%; 56.9%]

Same identity

68.0%; 95% CIs [57.1%; 77.2%]

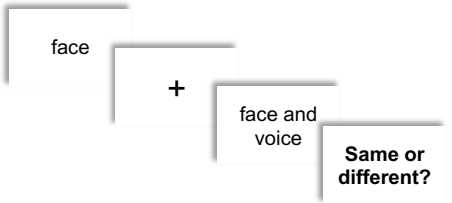
Experiment 3



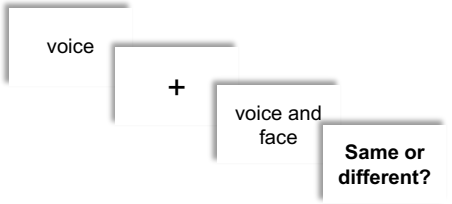
Perceived shared cross-modal identity information (i.e. Do people look and sound similar?)

Experiment 1

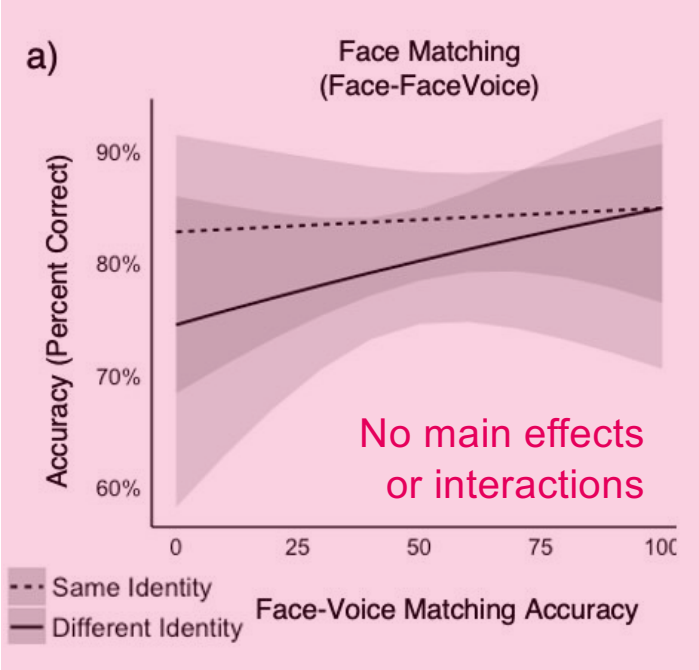
ii) Face-FaceVoice matching

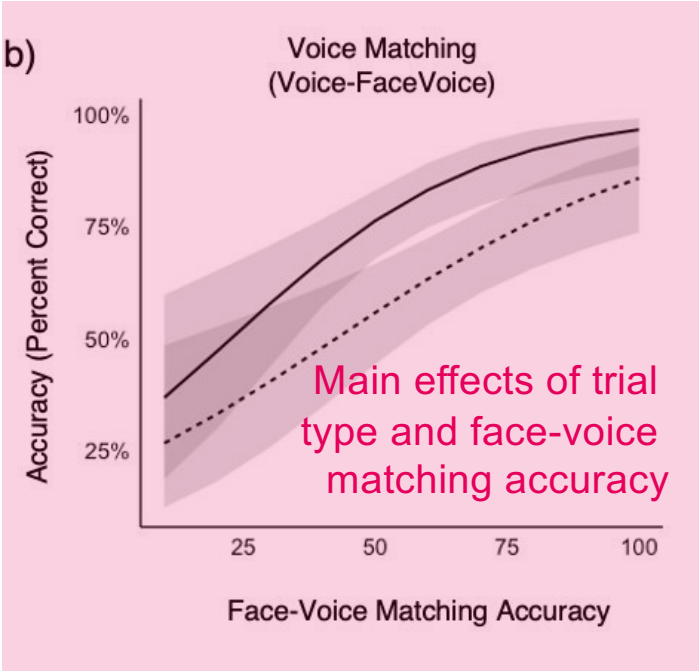


iv) Voice-FaceVoice matching



Experiment 3: Results (compared to multimodal data from Exp 1)





Fixed effects

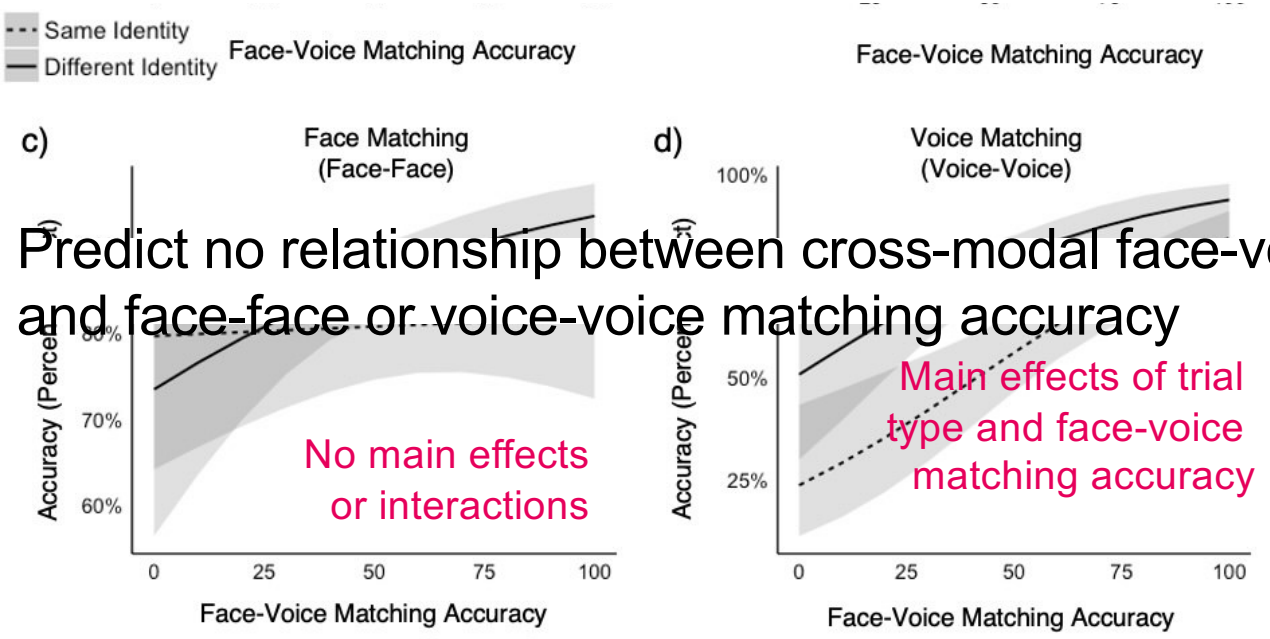
Trial type (same vs different identity)

Voice-face voice matching

Cross-modal face-voice matching accuracy

Matching decisions can be linked back to perceived face-voice concordance

Experiment 3: Results (compared to unimodal data from Exp 1)



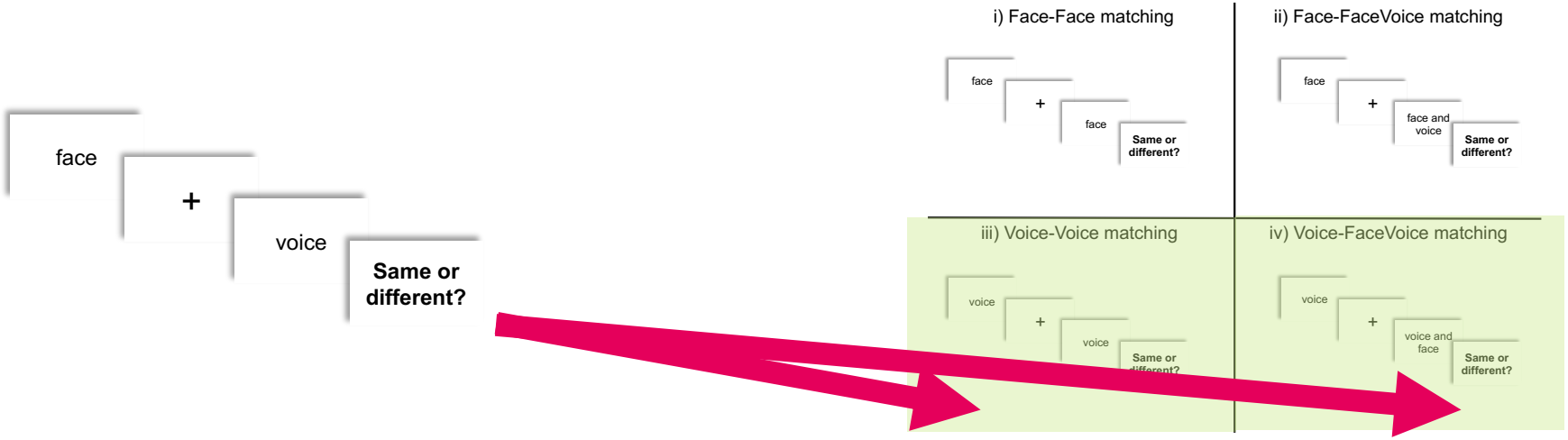
Predict no relationship between cross-modal face-voice matching accuracy and face-face or voice-voice matching accuracy

Voice-voice matching
Matching decisions can be linked back to perceived face-voice concordance

Experiment 3

Conclusions

Some people look and sound more similar than others



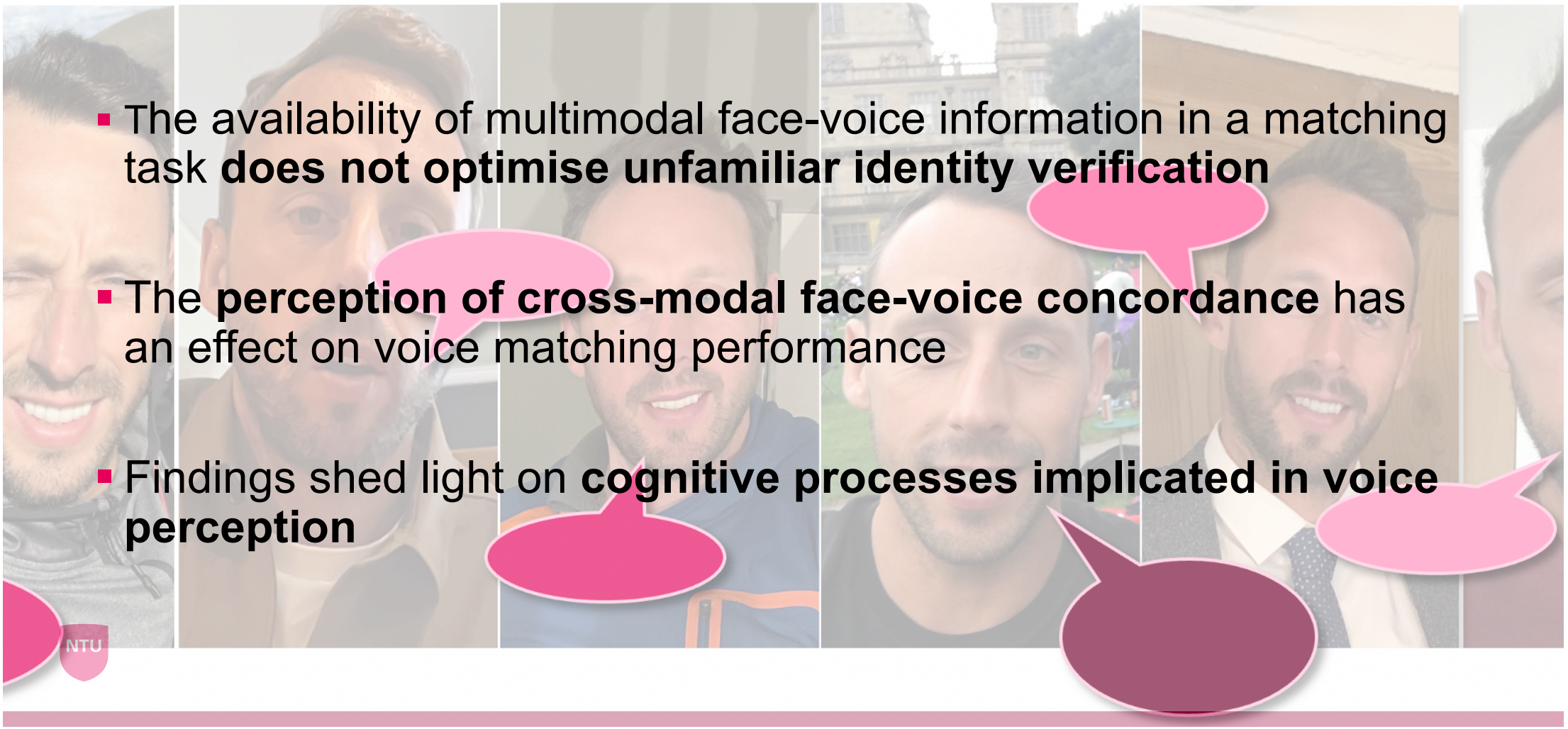
Experiment 3

Conclusions

- Face information is more reliable than voice information
- People might default to attempting to reconcile vocal identities by appealing to a ‘visual’ explanation
- An ‘image’ based on broad demographics or stereotypes while listening to an unfamiliar person speak
- While some face and voice cues are concordant, other cues are likely to be discordant. Shared information might bias people to overgeneralise
- Might account for some of the error in voice matching

NTU

Conclusions



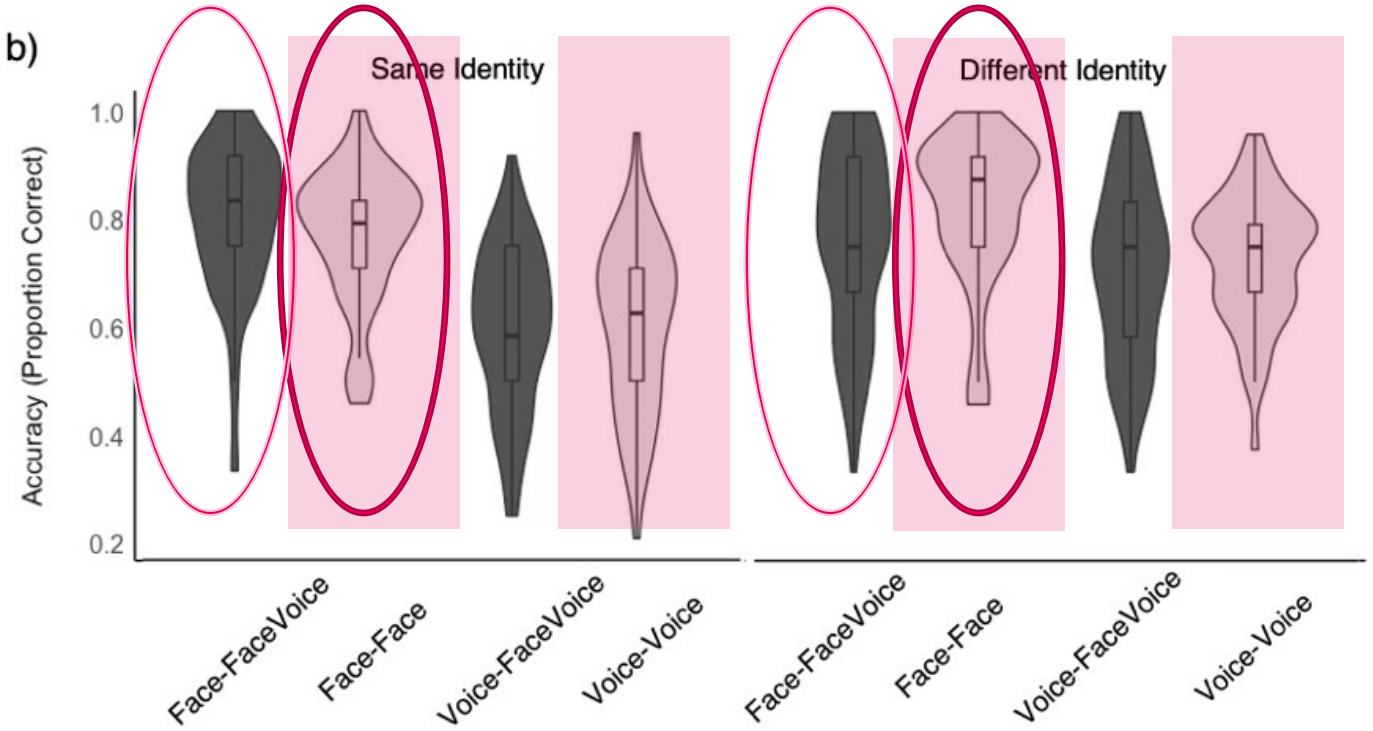
- The availability of multimodal face-voice information in a matching task **does not optimise unfamiliar identity verification**
- The **perception of cross-modal face-voice concordance** has an effect on voice matching performance
- Findings shed light on **cognitive processes implicated in voice perception**

Nottingham Trent University

Thank you

harriet.smith02@ntu.ac.uk

Experiment 1: Results



Fixed effects

matching task (face vs voice)

other-stimulus modality (unimodal face/voice vs multimodal facevoice)

trial type (same vs different identity within a stimulus pair)

NTU