



Queen Mary  
University of London

# Person perception from voices

## How do we make sense of who we are talking to?

EPS Nottingham | 11th April 2024  
Nadine Lavan | [n.lavan@qmul.ac.uk](mailto:n.lavan@qmul.ac.uk)

# Voices convey a lot of information

- What is being said?
- How is it being said?
- Who is talking?



# Voices convey a lot of information

- What is being said?
- How is it being said?
- Who is talking?



## **Part 1.**

**Identity perception from voices**

## **Part 2.**

**Person perception from voices**

# Part 1.

## Identity perception from voices

**Voices are extremely  
variable and flexible**

# You are listening to one person



Voices are variable

Identity is stable

# **Research Question**

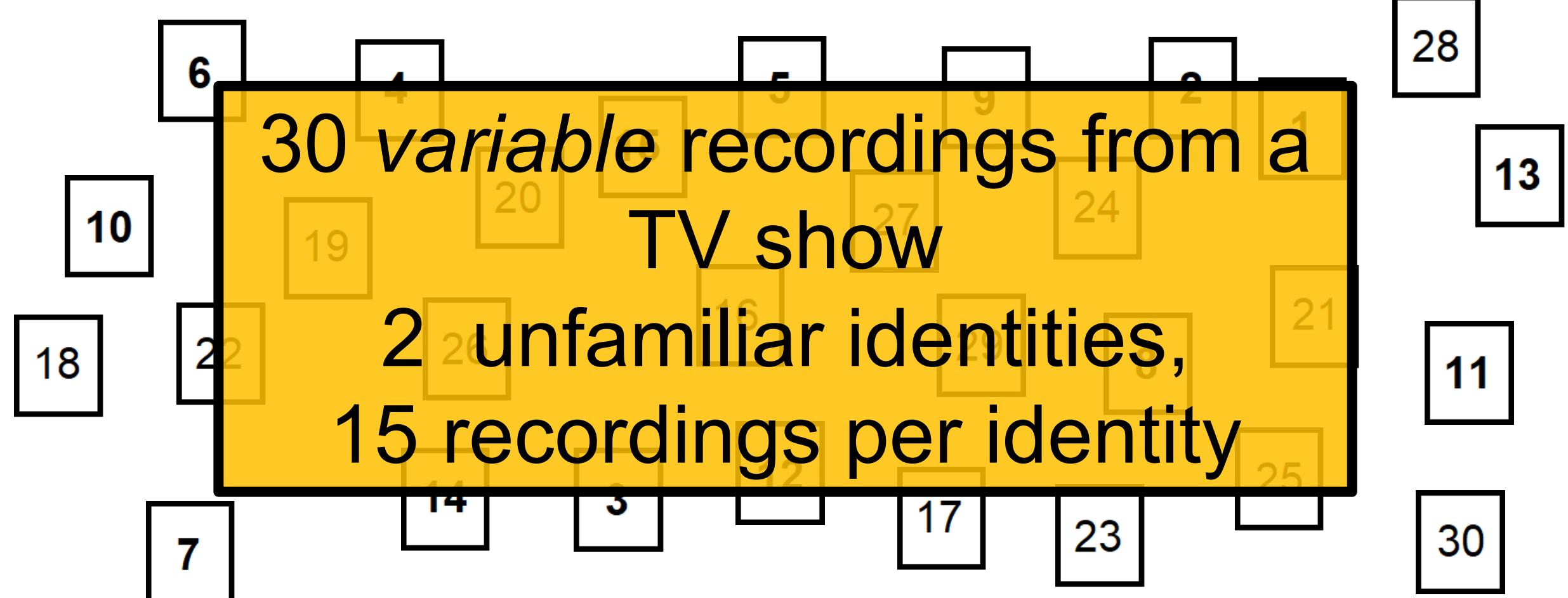
**What does this variability do  
to identity perception?**

Evidence from  
“voice sorting” tasks

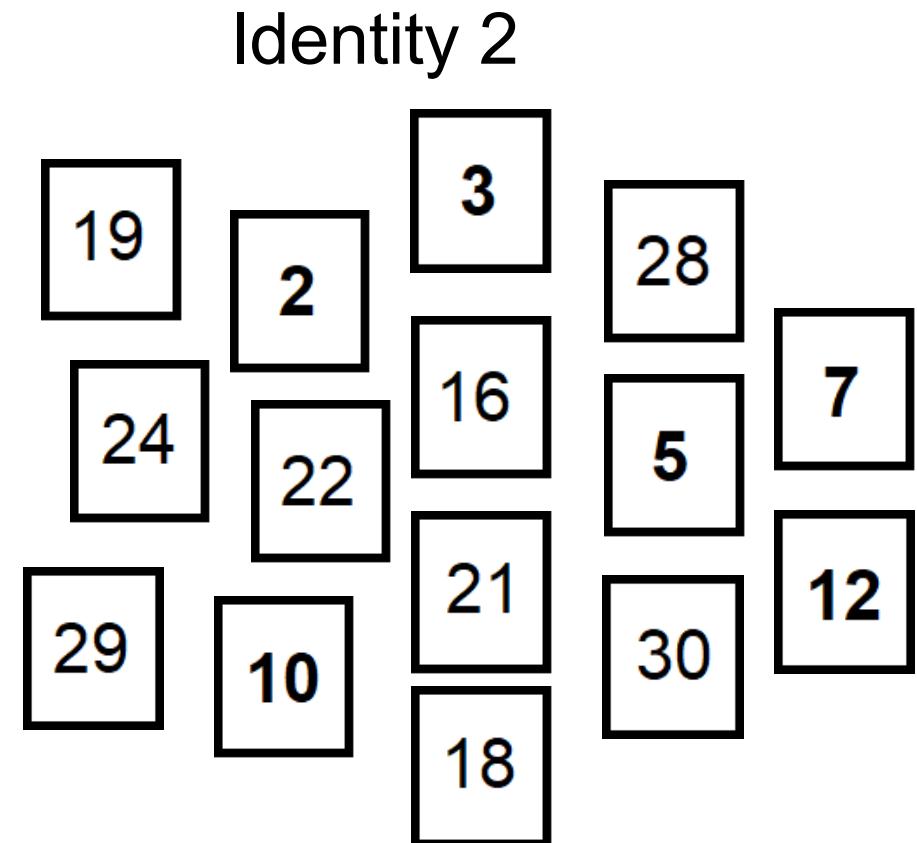
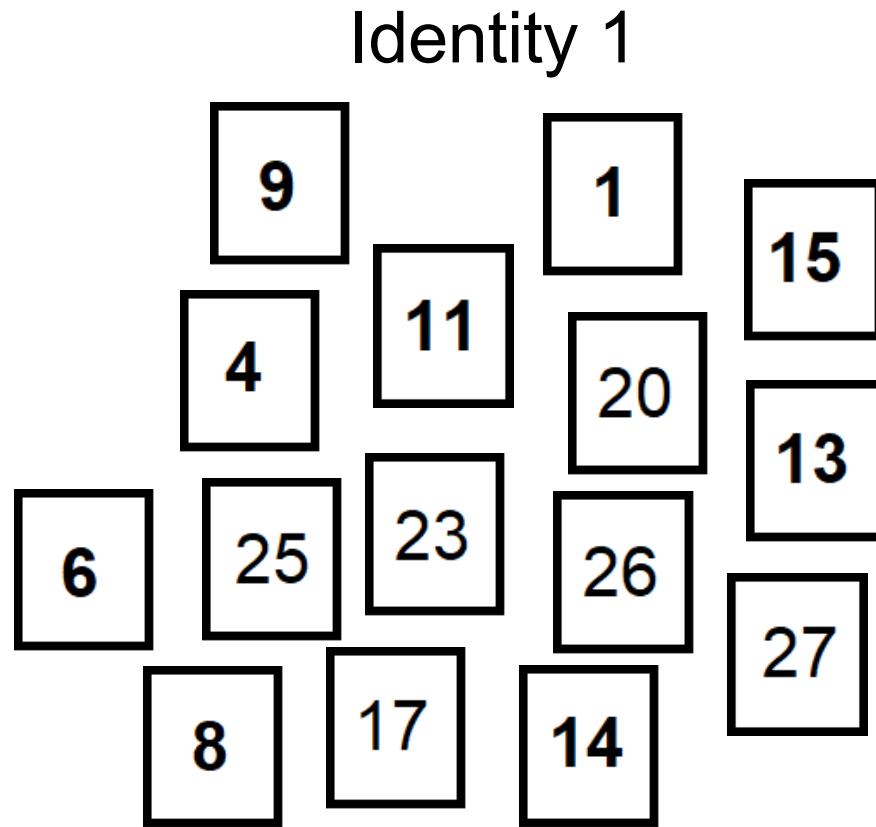
# “Please sort these recordings into different identities”

30 variable recordings from a TV show

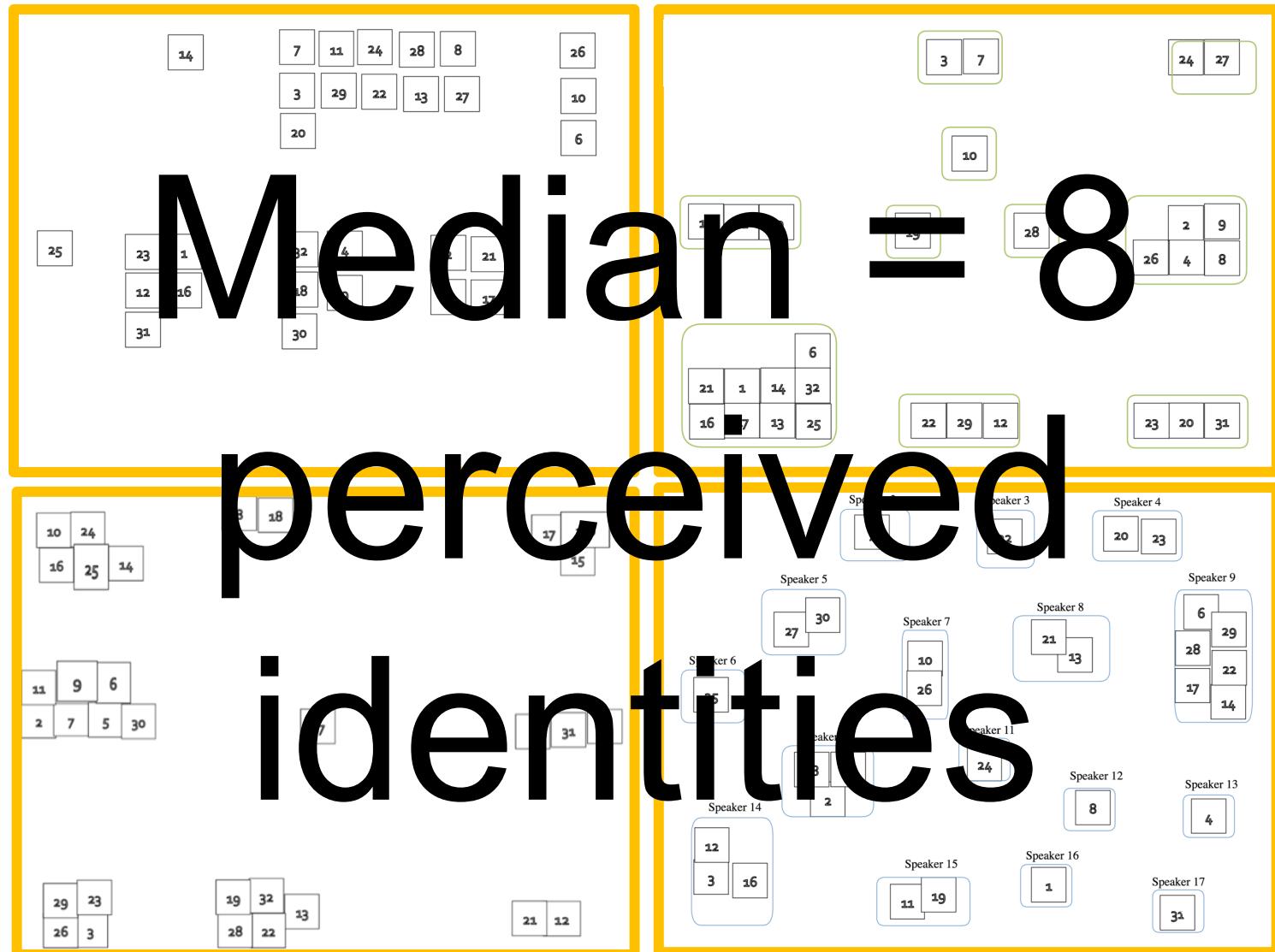
2 unfamiliar identities,  
15 recordings per identity



# “Please sort these recordings into different identities”



# Example data



Lavan et al., 2019, *British Journal of Psychology*; Lavan et al., 2019, *Quarterly Journal of Experimental Psychology*; Johnson, McGettigan & Lavan, 2020, *Quarterly Journal of Experimental Psychology*

# How does variability in voices affect identity perception?

- Vocal variability disrupts voice identity perception
- Listeners cannot make sense of the same person sounding different across variable recordings.
  - Listeners perceive one identity as many people and fail to “*tell people together*”
  - Listeners only rarely perceived two identities as the same person: Listeners are good at “*telling people apart*”

But listeners are not  
always *this* bad at identity  
perception

# Research Question

Can familiarity with the  
voices help?

# Voice identity sorting

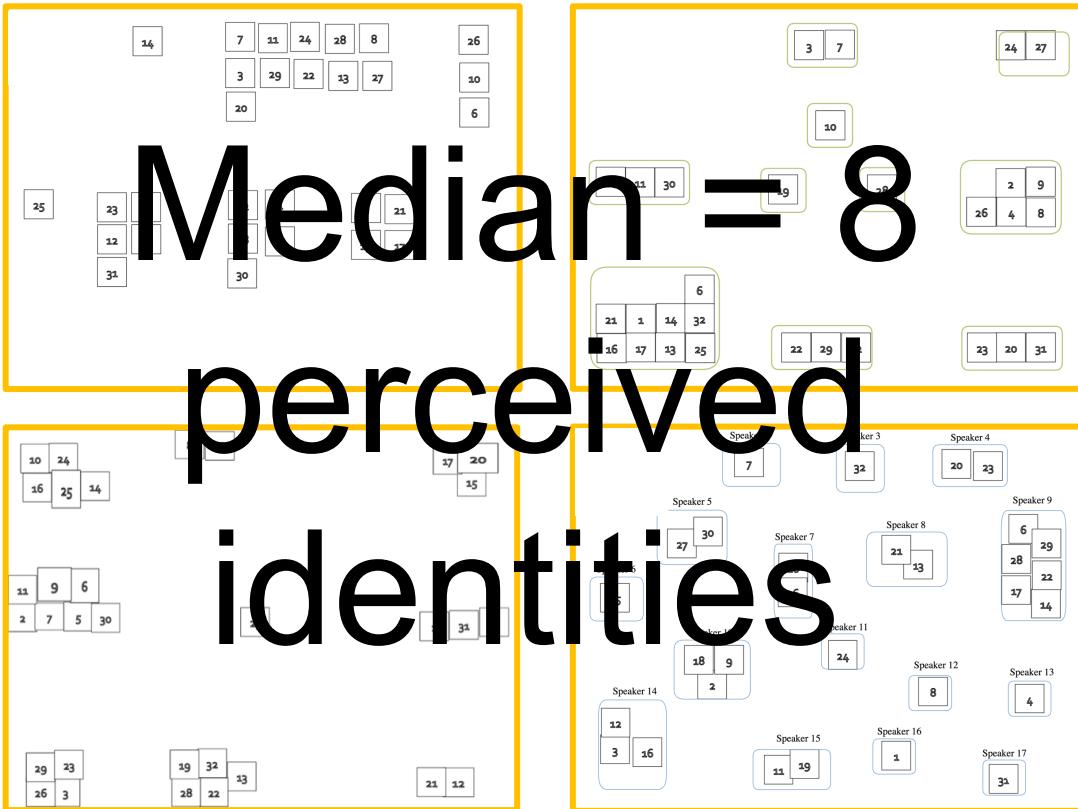


## Participants

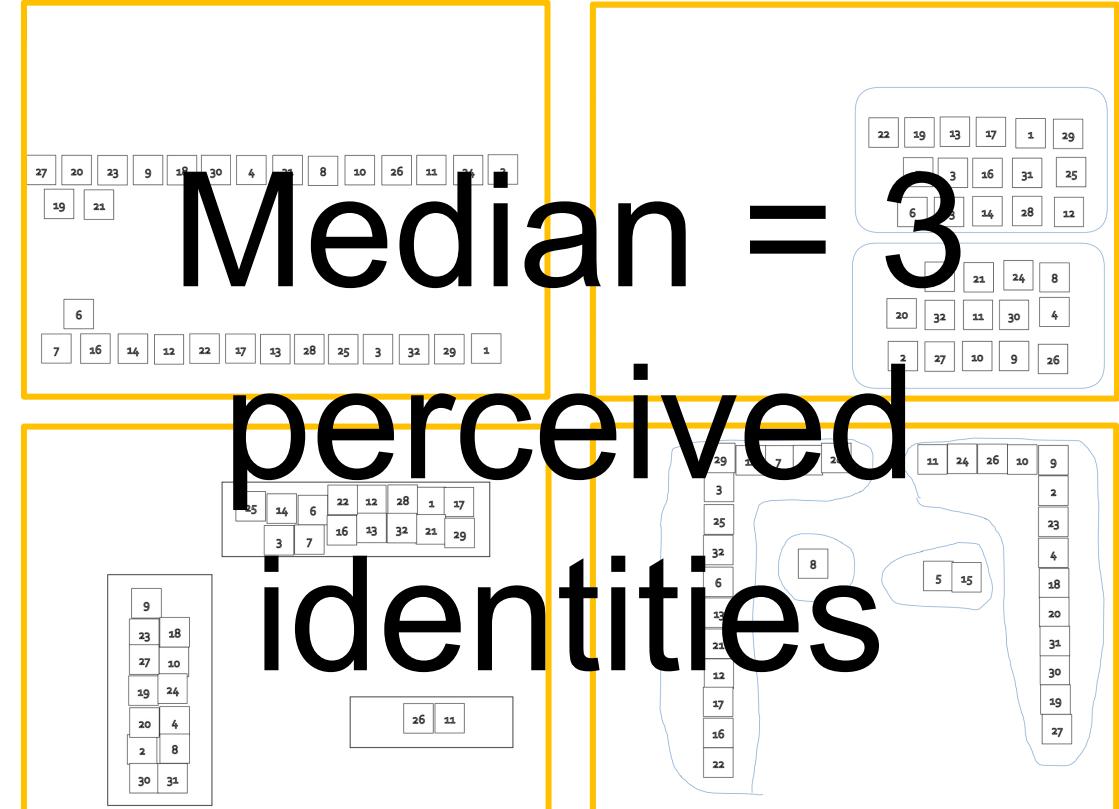
- Unfamiliar: Listeners who had not watched *Breaking Bad*
- Familiar: Listeners who had watched *Breaking Bad*

# Example data

Unfamiliar voices



Familiar voices



# How does variability in voices affect identity perception?

- Vocal variability disrupts voice identity perception – especially if listeners are not familiar with the voices
- If listeners are familiar with voices, they can cope with variability!

Unfamiliar voice



Familiar voice

Listeners form a  
representation of the voice

# Research Question

Becoming familiar:  
How do we form  
representations from  
variable voices?

Based on  
exemplars?

Based on summary  
statistics, e.g.,  
averages?

# How do we form representations from variable voices?

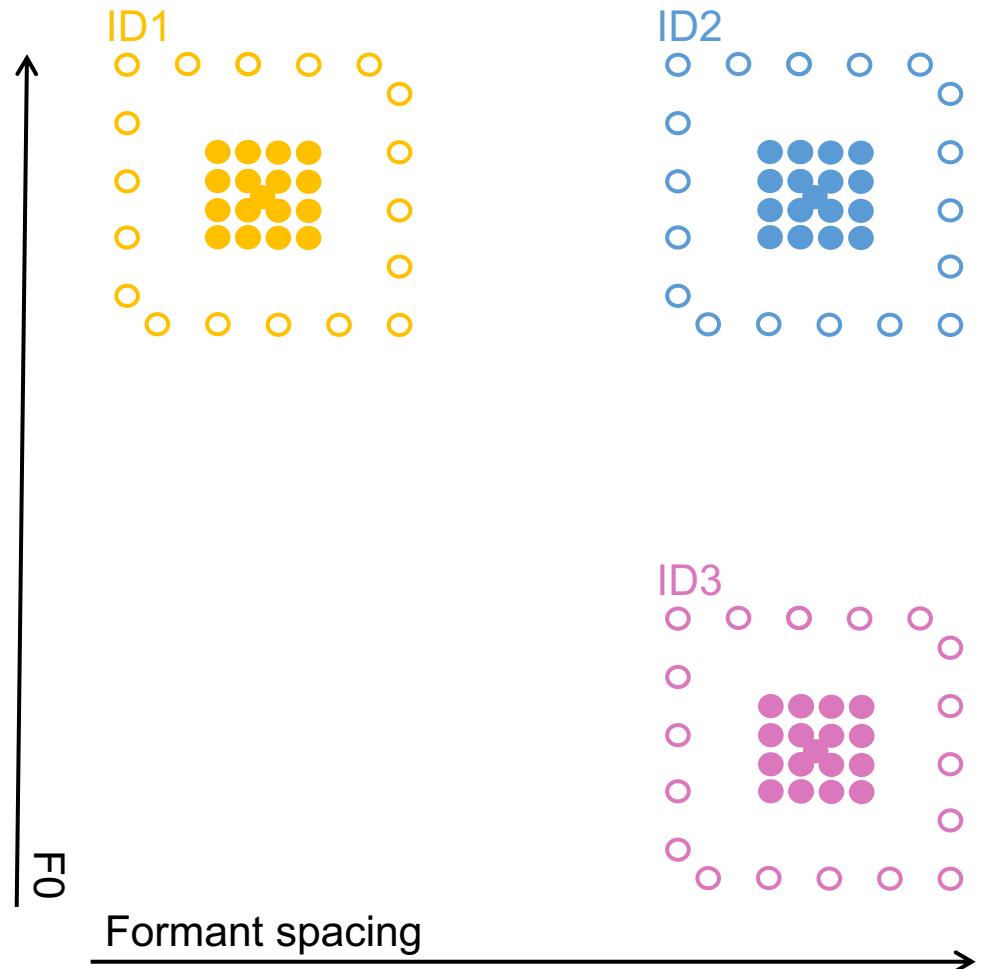
## General approach:

1. Train listeners to recognise voices based on a *distribution of variable stimuli*
2. Test listeners on recognition of the learned voices

## Prediction:

If the recognition accuracy is higher for the average of the distribution, then averages are abstracted

# How do we form representations from variable voices?



## Training:

Train listeners on variable “ring-shaped” distributions that are missing their centres/averages

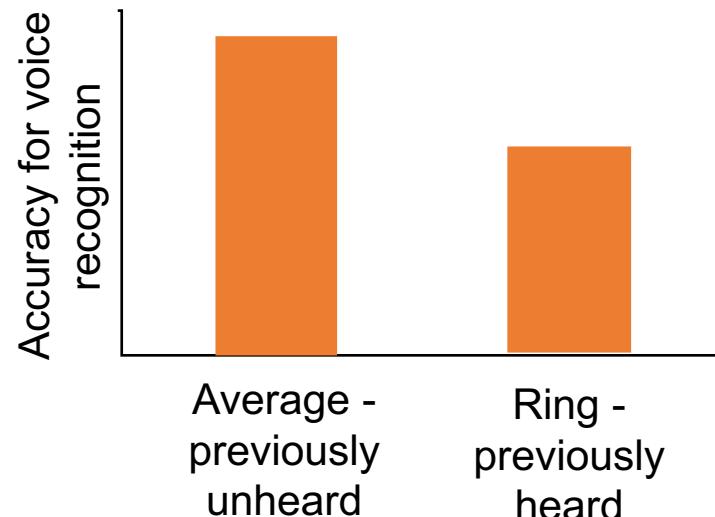
## Test:

Test listeners on the ring-shaped distribution AND the centre distribution

# Are averages abstracted when forming representations?

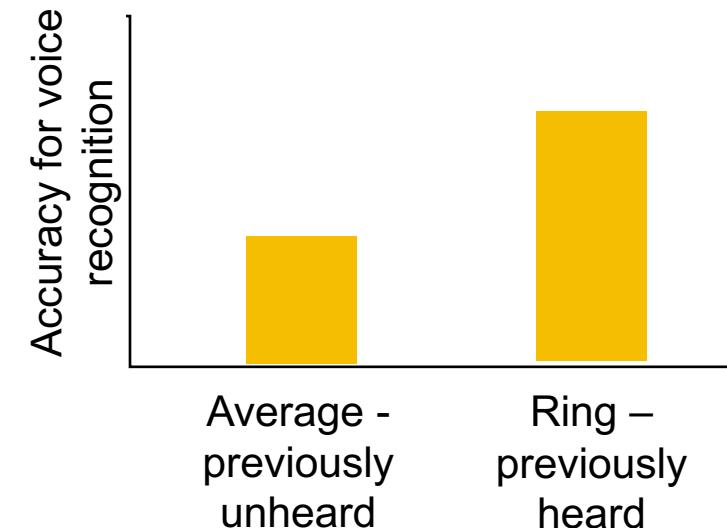
## Hypothesis 1: YES

Accuracy for average is higher than for the ring.



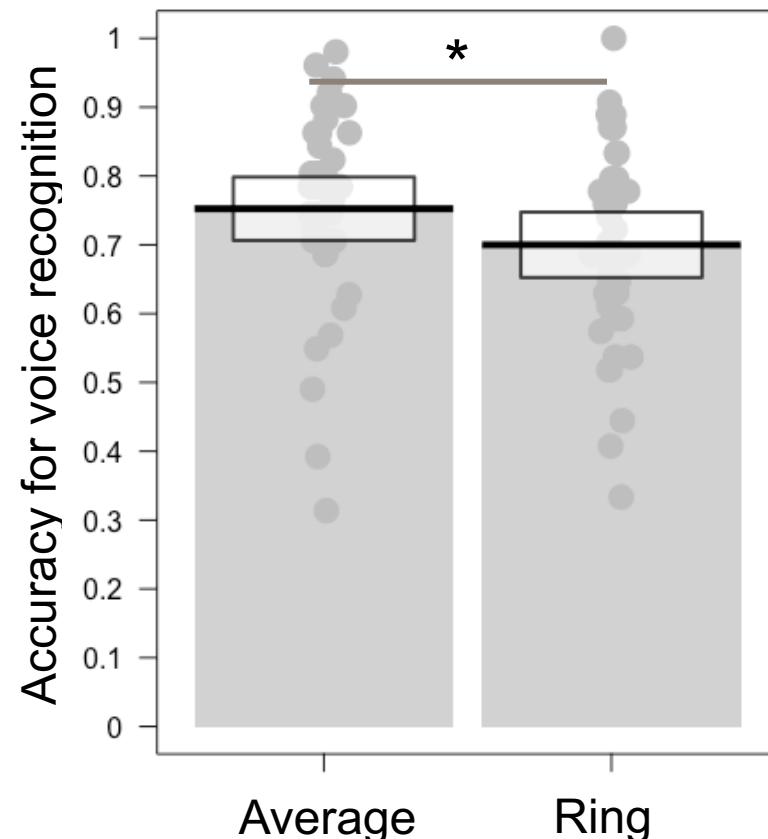
## Hypothesis 2: NO

Accuracy for the average is lower than for the ring.

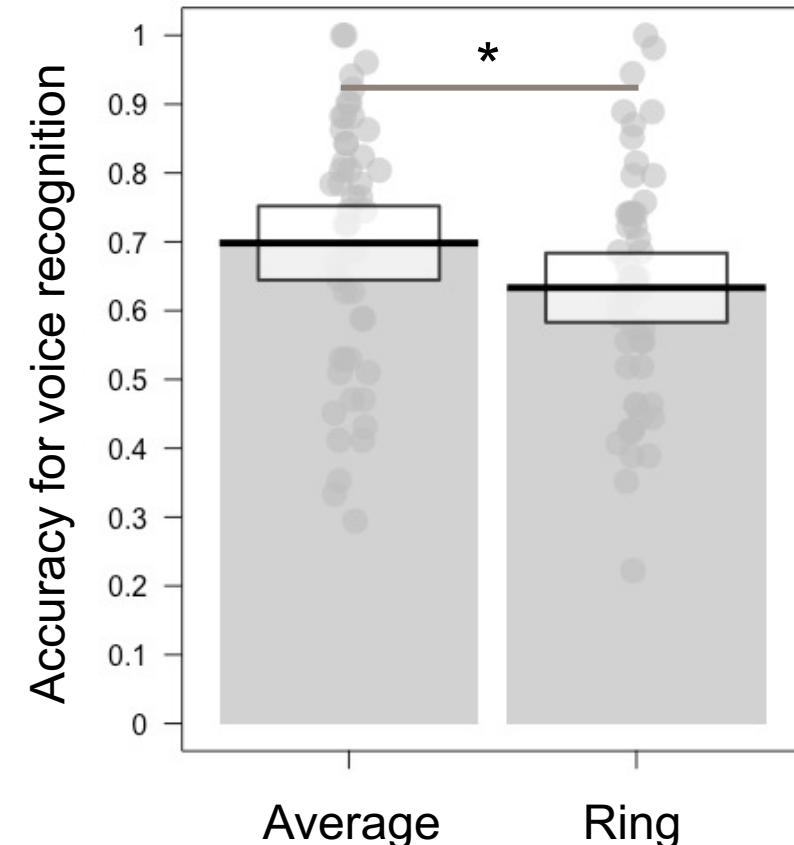


# How do we form representations from variable voices?

Original Experiment



Replication



# How do we form representations from variable voices?

- Listeners form a representation that is based on abstracted averages
- These averages are meaningful: Recognition accuracy is higher for averages!
- This, however, means that we've just discarded all variability.
- But do we also encode and use the variability at any point?

# **Research Question**

Becoming familiar: Can  
variability be useful?

# Can variability be useful?

So far, variability is either bad news for successful voice identity perception or it gets discarded

**But in other fields...**

- High-variability training > low-variability training
  - For faces (e.g., Ritchie & Burton, 2017; Murphy et al., 2015)
  - For vocabulary learning (e.g., Barcroft & Sommers, 2005)
  - And more (see TICS review by Raviv et al., 2022)!

# Can variability be useful?

## General approach:

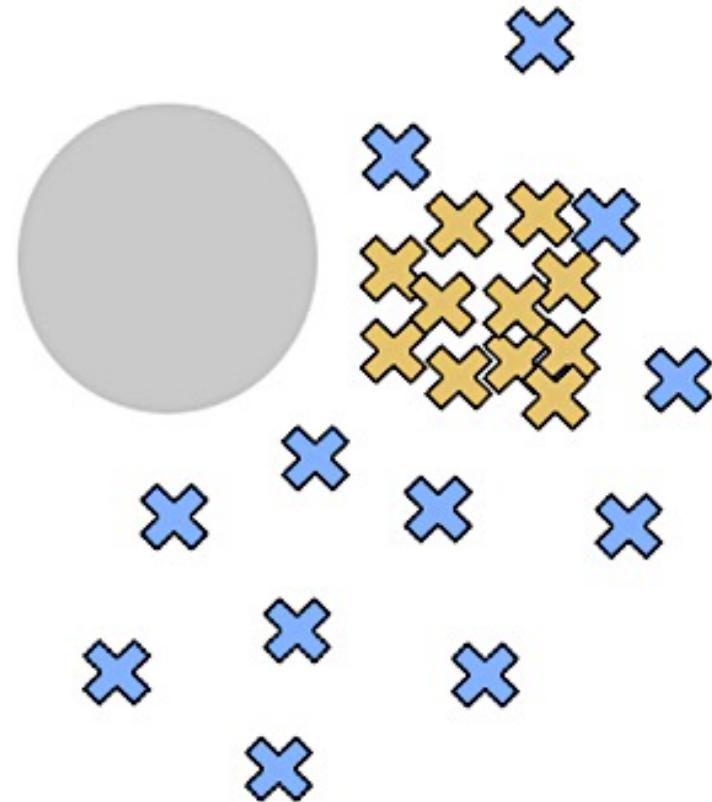
1. Train listeners to recognise voices based on *high vs low variability training*
2. Test listeners on recognition of the learned voices

## Prediction:

Recognition accuracy is higher after high variability training compared to low variability training

# Can variability be useful?

- **High variability training:**
  - Stimuli sampled from *different speaking styles*
- **Low variability training:**
  - Stimuli sampled from *1 speaking style*



**Test stimuli:** Sampled from 1 speaking style  
*not included in the training*

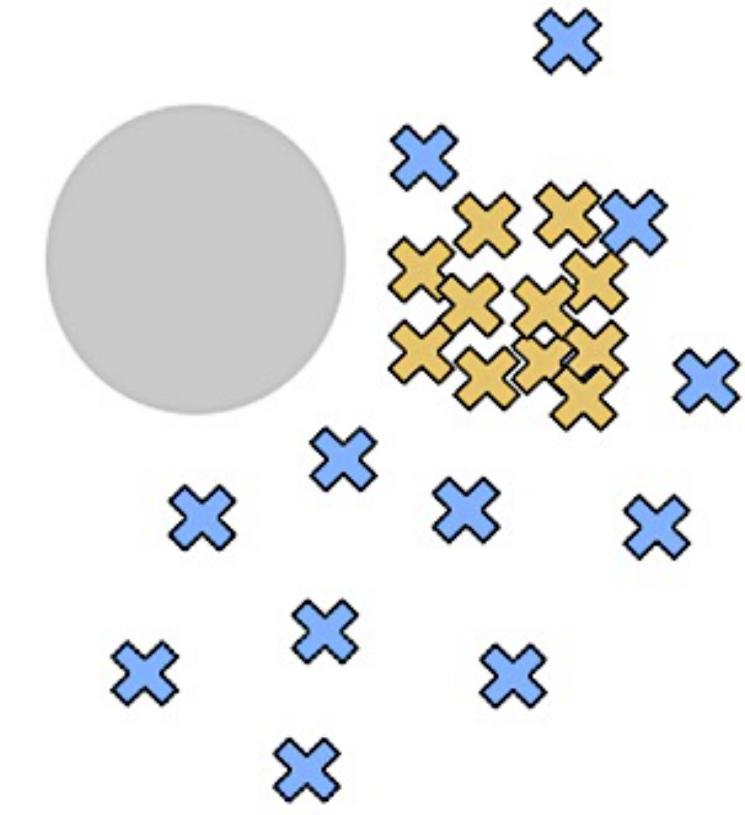
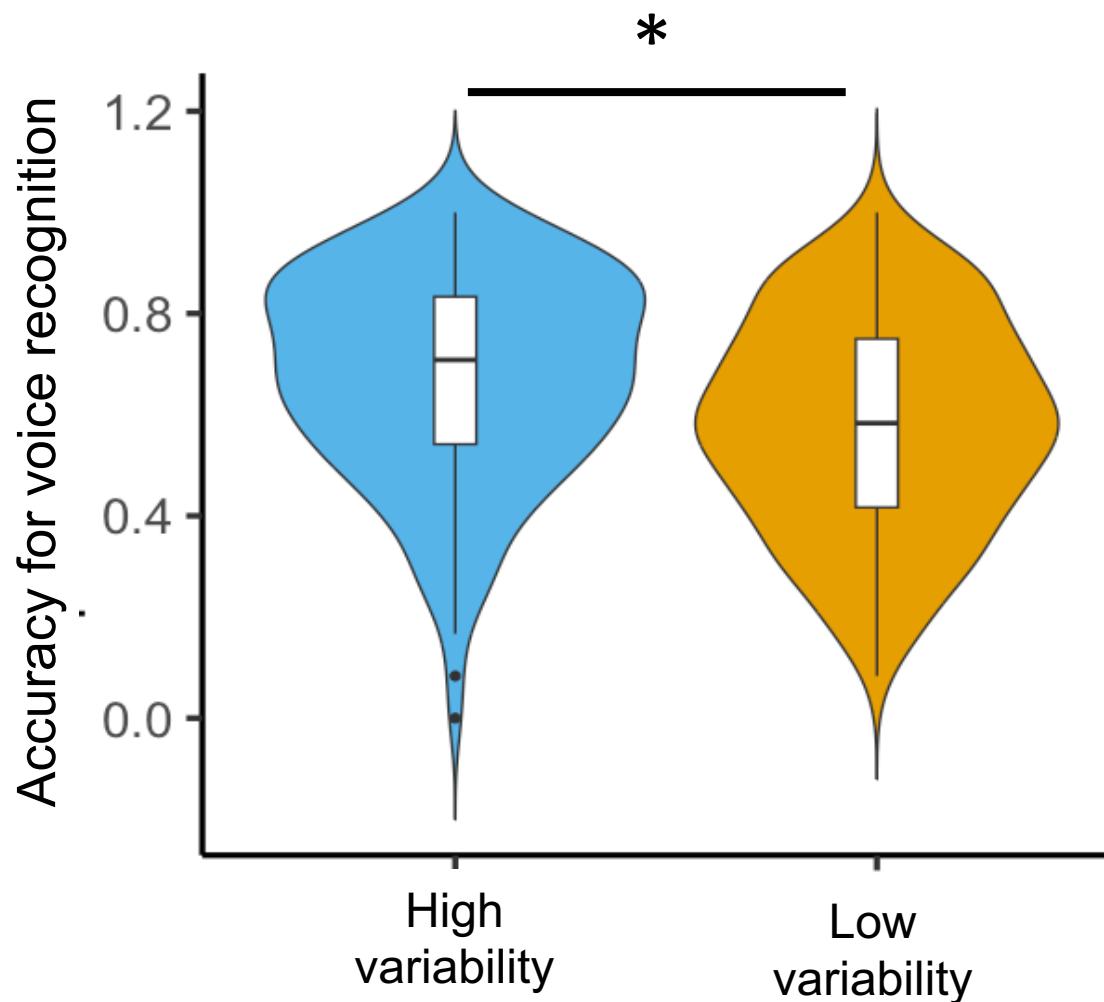
✖ High variability training stimuli

✖ Low variability training stimuli

● Voice space occupied by test stimuli

# Can variability be useful?

There is a high variability advantage!



# Can variability be useful?

- **Big caveat, though.**
- This is 1 of 3 experiments we ran, which all included different definitions of high vs low variability (and training vs test stimuli).
  - High variability advantages only emerge when listeners need to generalise beyond what they already know.
  - In other settings, low variability advantages can arise.
- So, variability affects voice learning but the direction of effects can vary

# Part 1.

## Identity perception from voices

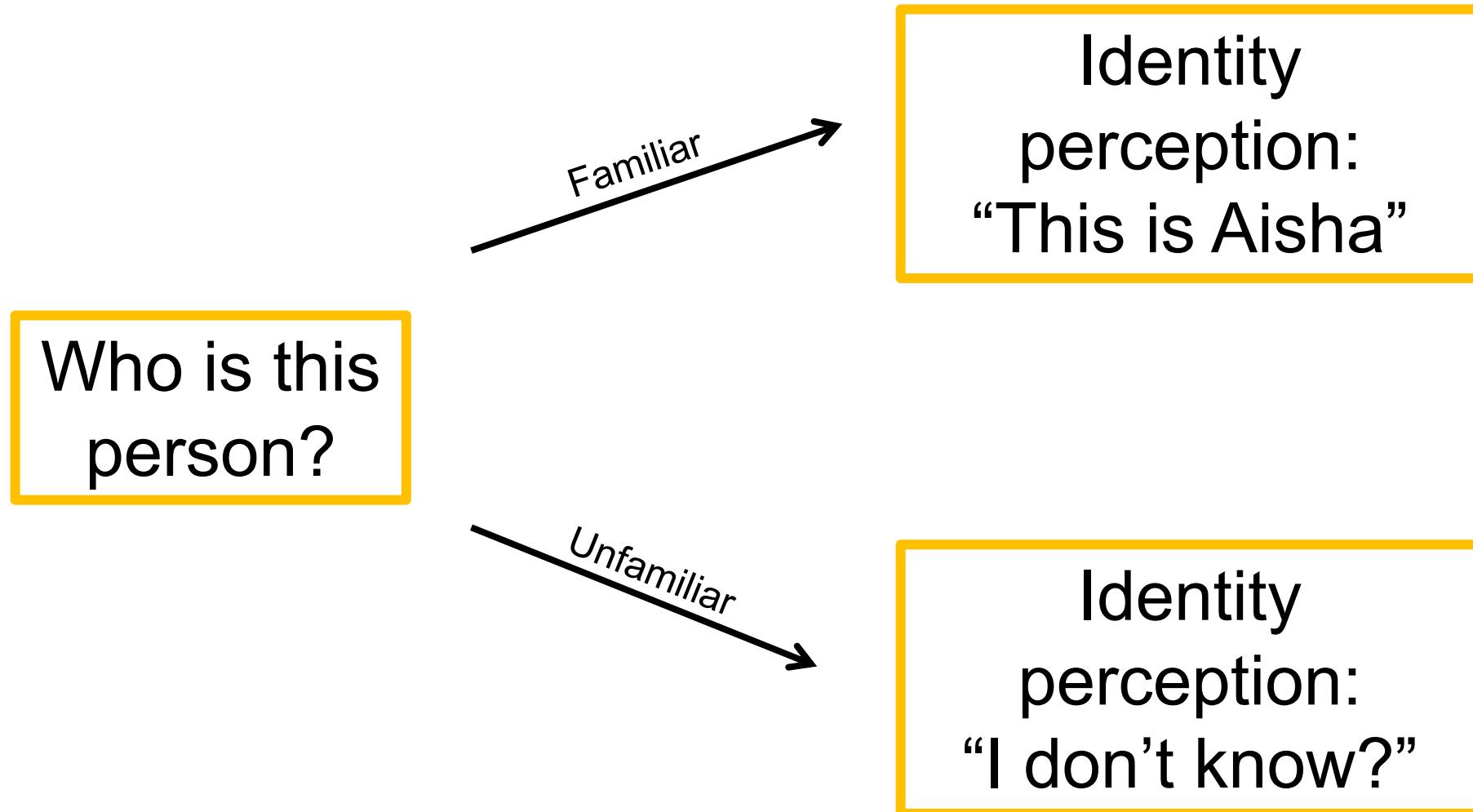
# **Identity perception from voices**

- **Study 1: Voice sorting**
  - Vocal variability disrupts voice identity perception when voices are unfamiliar
  - We don't struggle with variability when voices are familiar
- **Study 2: Averages**
  - When we become familiar with voices, we form representations that are based on abstracted averages
- **Study 3: High variability training**
  - Exposure to variability can help us learn voices – but only when we need to generalize beyond what we know.

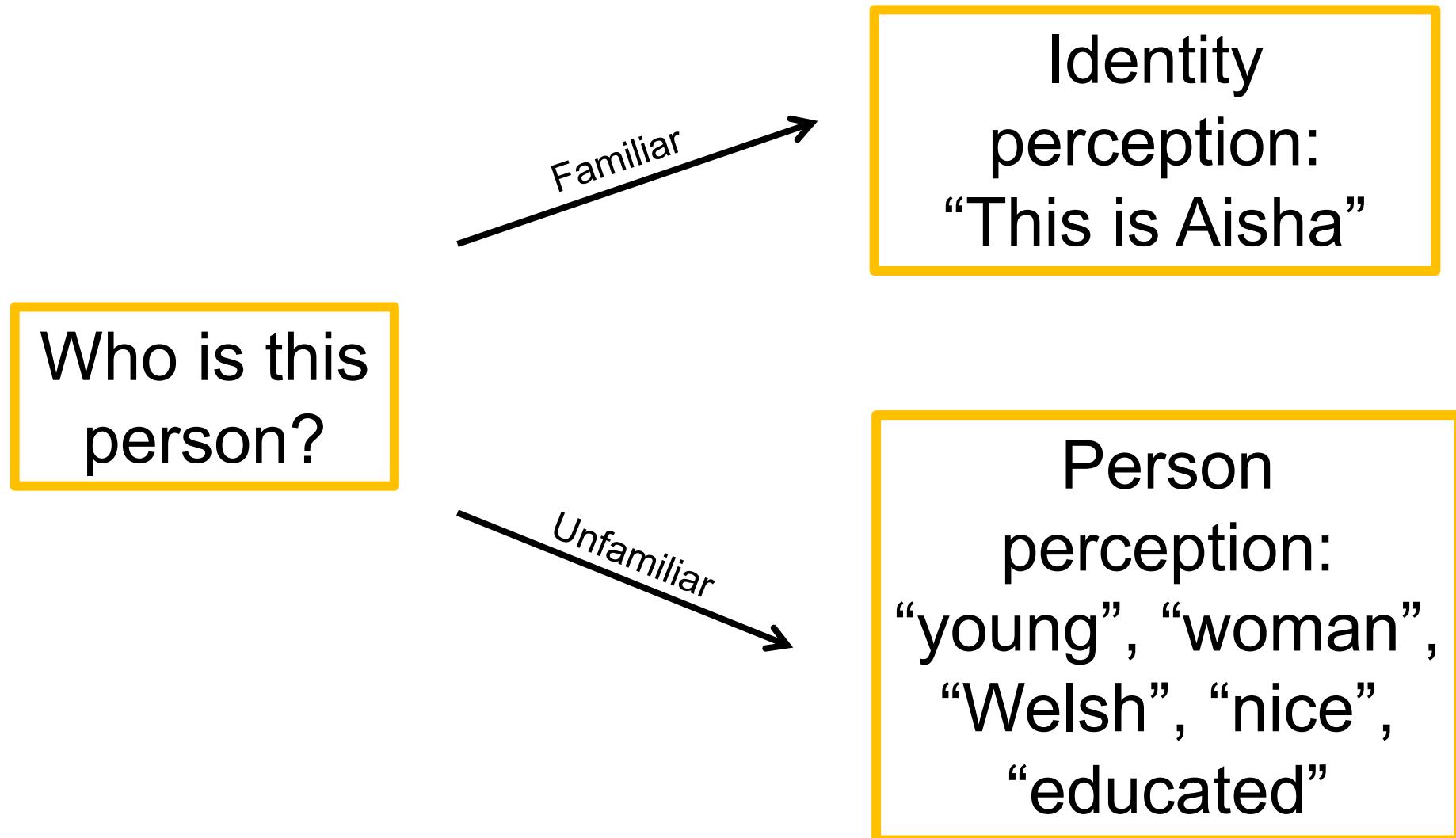
# Part 2.

## Person perception from voices

# Beyond identity: Person perception from voices



# Beyond identity: Person perception from voices

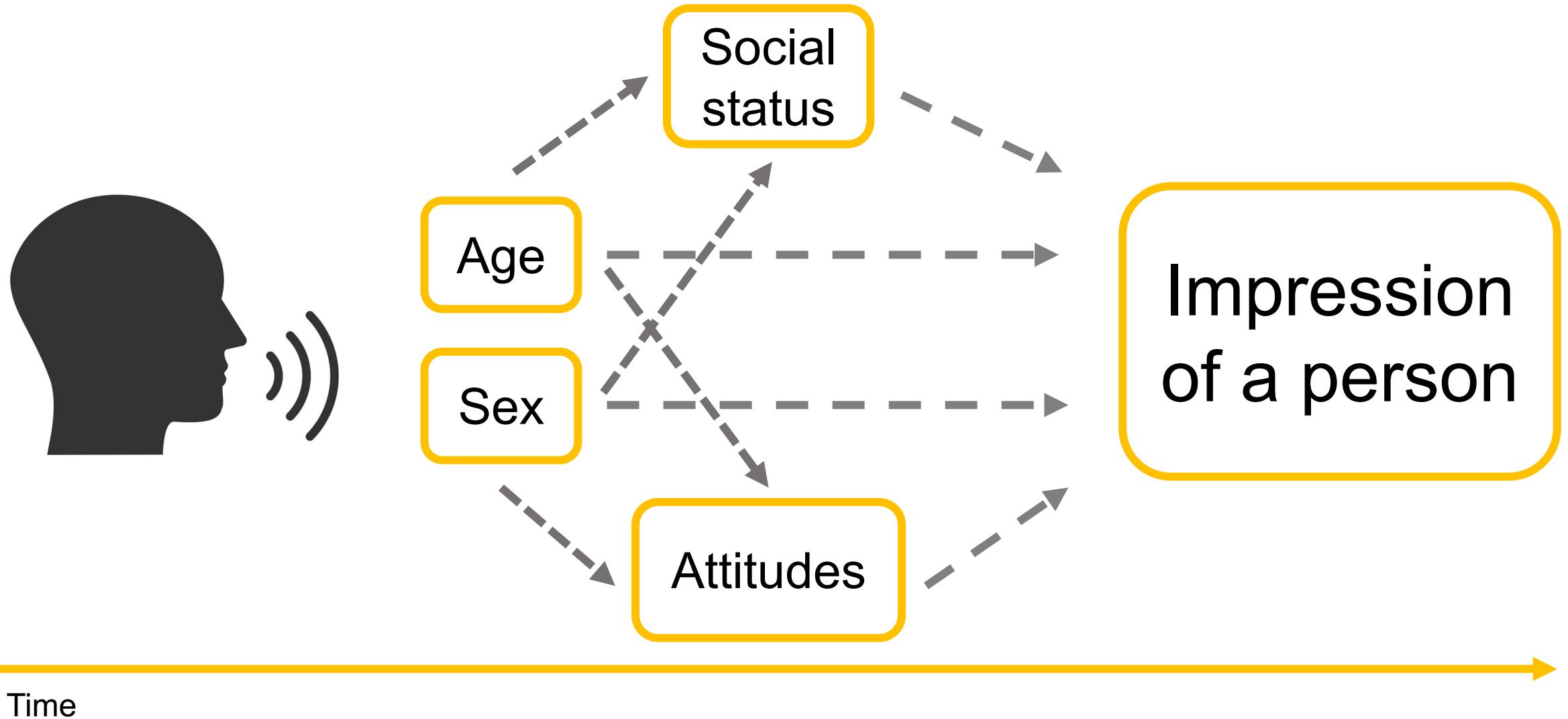




While some of these impressions may be (partially) accurate, others are entirely subjective

understanding  
professional interesting  
**well-spoken** amicable affluent  
**southern**  
english clear white british  
distinct clever nice feminine  
friendly eloquent calm soft accent happy  
caring emotionless kind quiet clipped  
annoyed measured gentle  
female bored concise  
precise young confident attractive  
relaxing good easy pleasant  
smooth posh educated polite  
articulate middle-class smart  
polite reserved intelligent  
well-educated

**Are some  
characteristics  
perceived more quickly  
than others?**



# Research Question

How much exposure to a voice  
do we need to form an  
impression?

# How much exposure to a voice do we need to form an impression?

- We collected evaluations of voices presented for 25ms, 50ms, 100ms, 200ms, 400ms, 800ms respectively.



800ms of exposure



25ms of exposure

# How much exposure to a voice do we need to form an impression?

- **Person characteristics evaluated:**
  - “Physical”: Sex, age, health
  - “Trait”: Attractiveness, dominance, trustworthiness
  - “Social”: Educatedness, “poshness”, professionalism
- **Why these particular characteristics?**
  - They were among the most frequently mentioned characteristics in a study where people freely labelled voices

# **Problem:** How can we assess whether and when a person characteristic has been perceived?

- Accuracy is not the way forward here
- Impressions are to some degree shared among listeners (McAleer et al., 2015).
  - On average, listeners agree which voices sound more or less trustworthy/dominant/attractive.
- Use inter-rater agreement as an index of whether impressions have been formed or not.

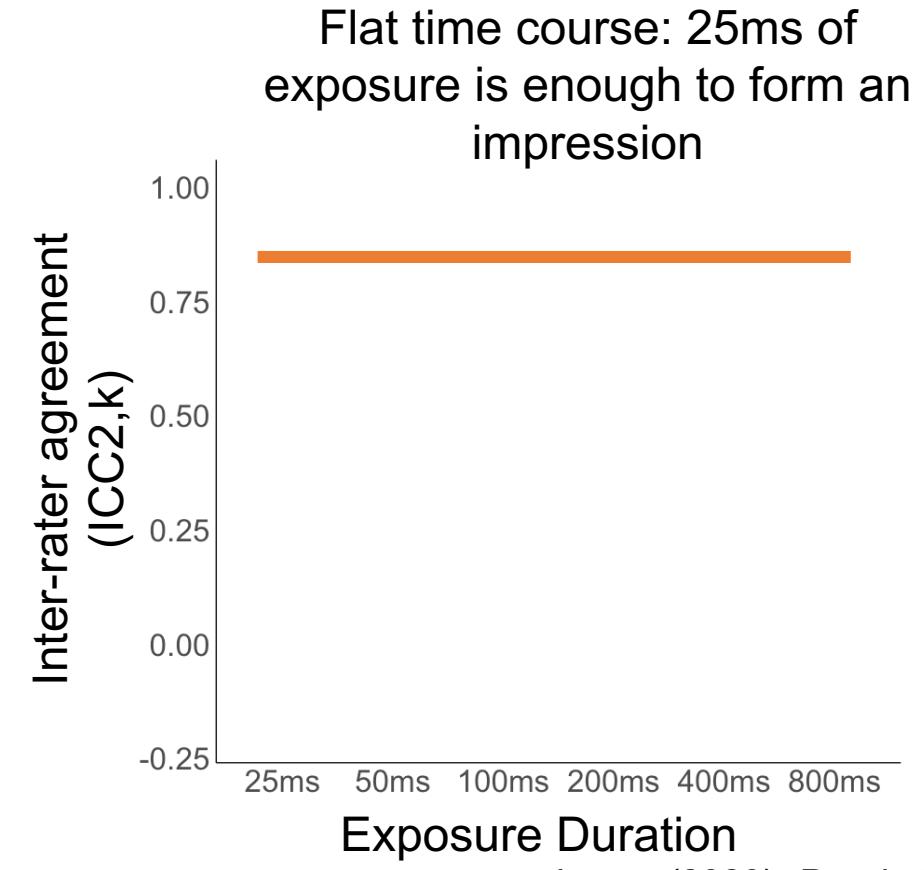
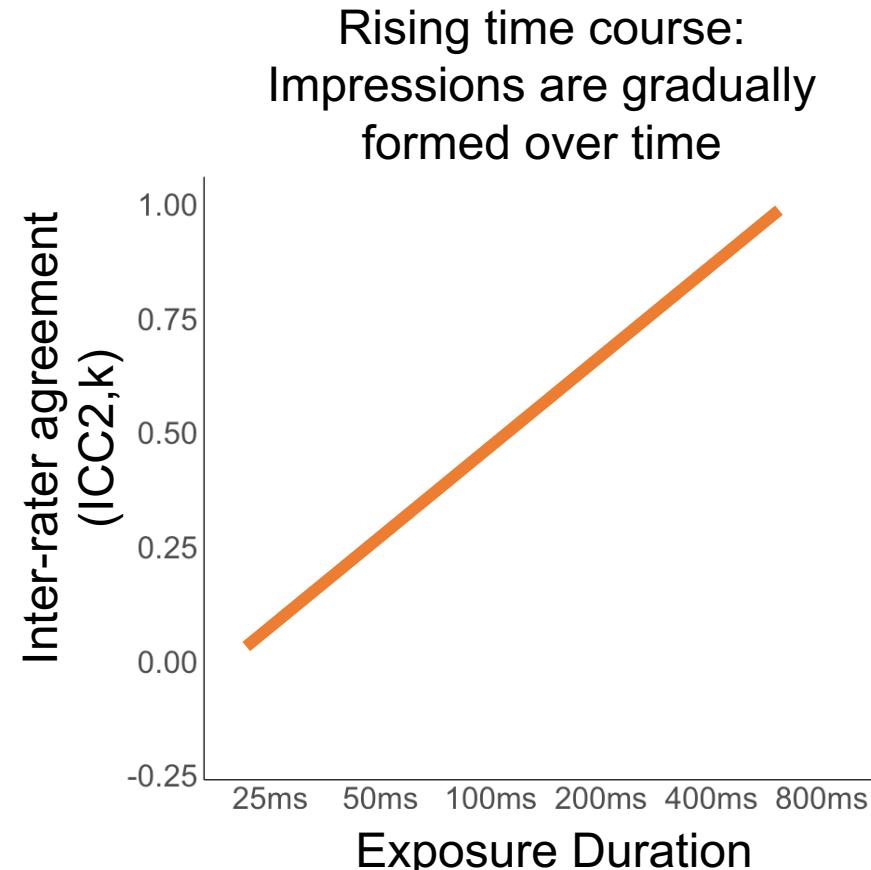
# Problem: How can we assess whether and when a person characteristic has been perceived?

High agreement:  
Impression has been  
formed!

No/low agreement:  
Impression has not  
(yet) been formed!

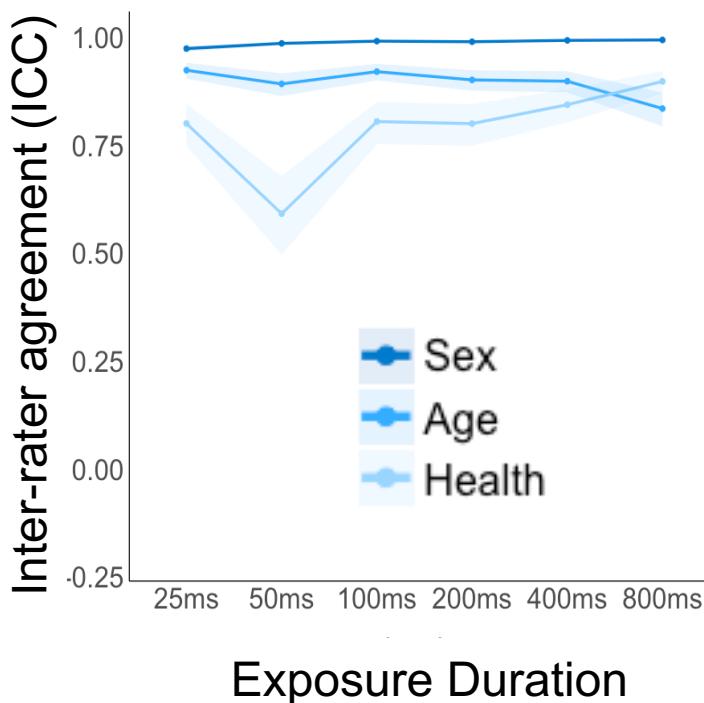
# How much exposure to a voice do we need to form an impression?

What kind of time courses might be observe?

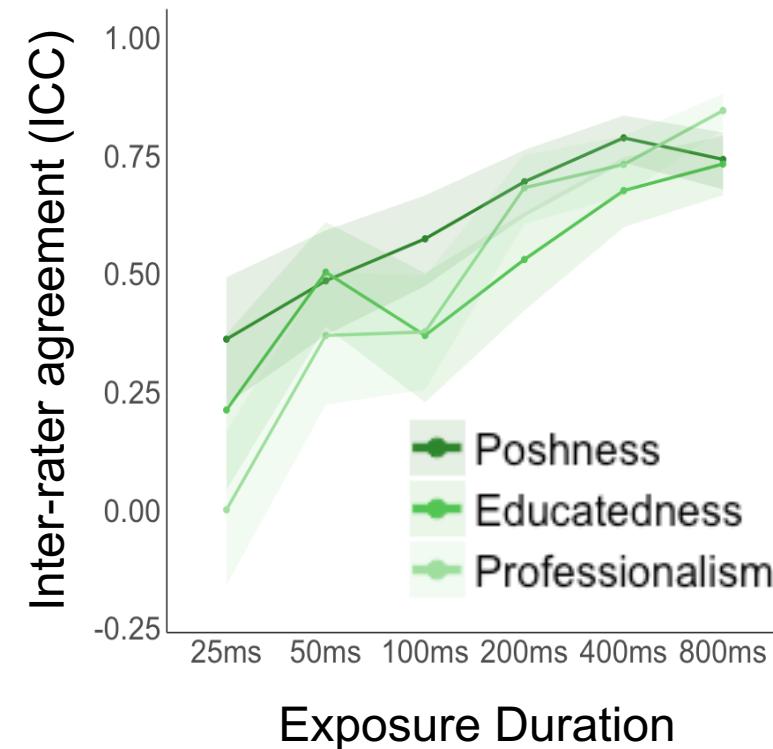


# How much exposure to a voice do we need to form an impression?

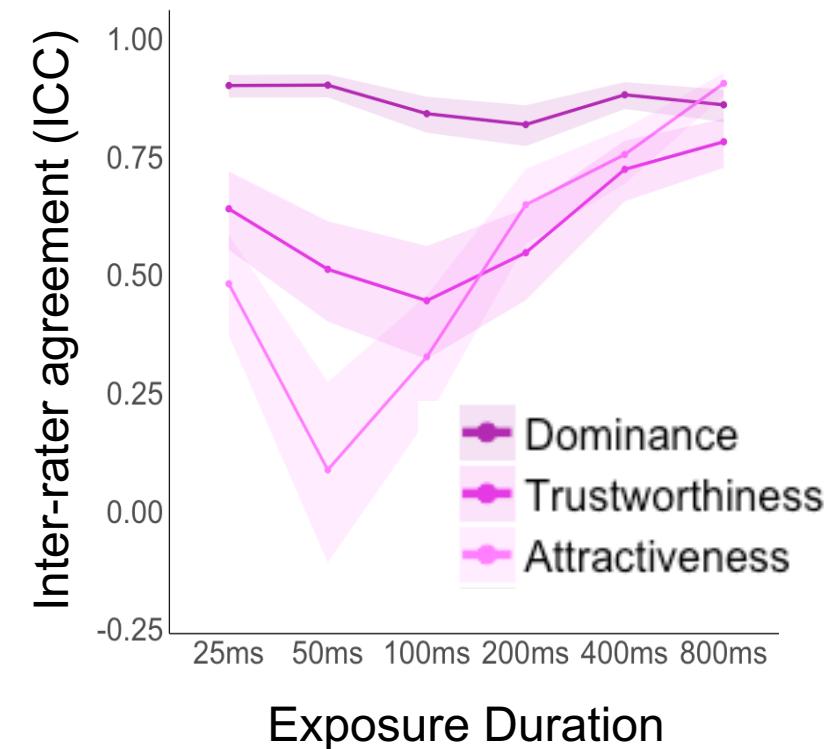
## Physical characteristics



## Social characteristics



## Trait characteristics



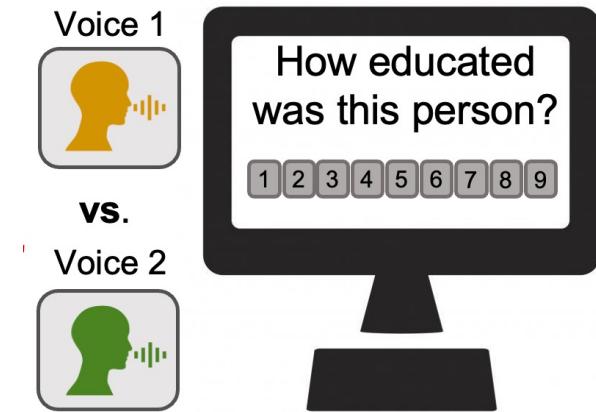
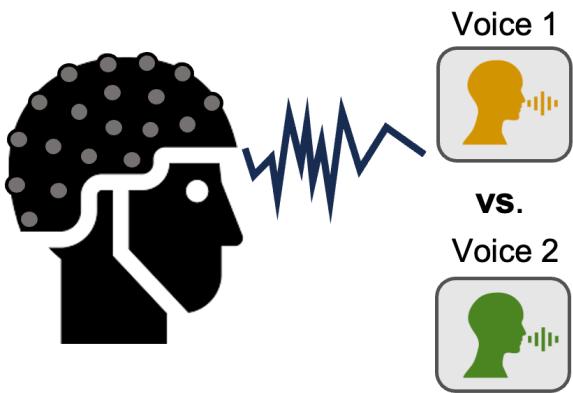
# How much exposure to a voice do we need to form an impression?

- Person perception from voice is **rapid!**
  - 800ms of exposure (and often less) is enough
- Physical characteristics (and dominance) can be perceived from even minimal exposure.
- (Most) trait and social characteristics are perceived more gradually and need a bit more exposure.

# Research Question

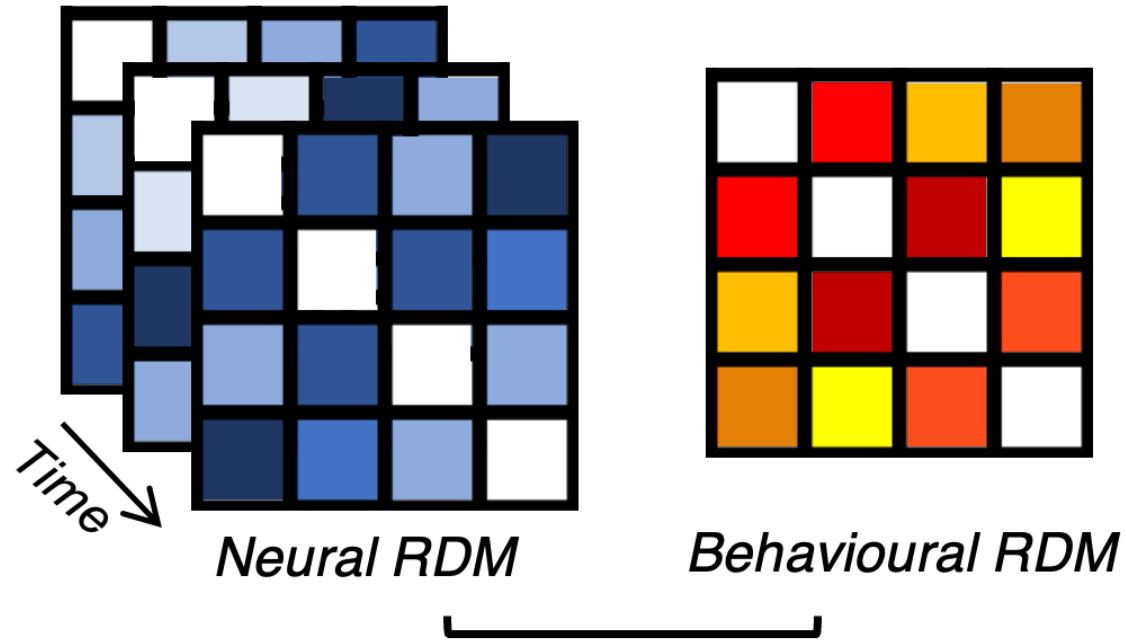
What's the timecourse of  
impression formation in the  
brain?

# Design

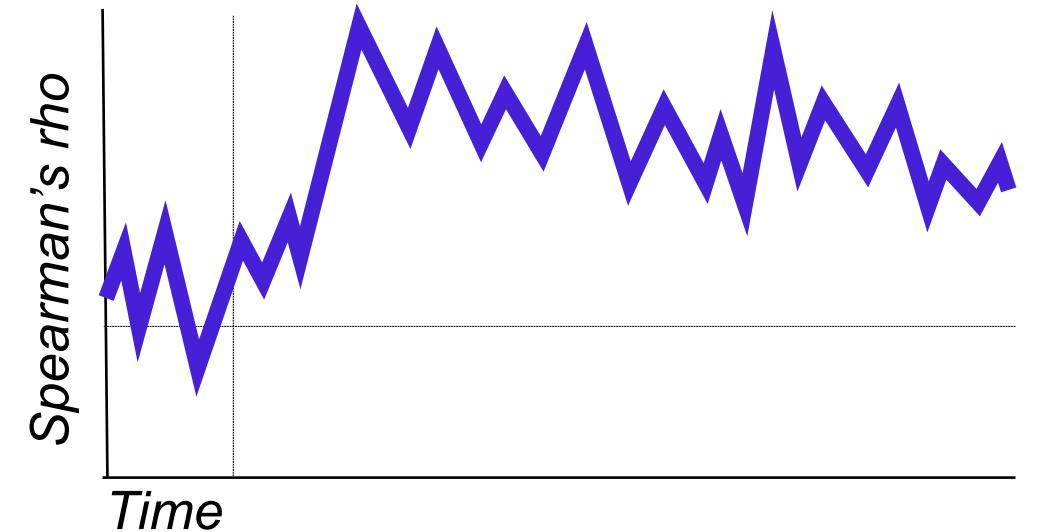


- Record EEG data while participants listen to short recordings of voices
  - Create dissimilarity matrices, i.e., a matrix indexing how dissimilar all possible pairs of voices are from each other
- Get behavioural ratings for voices for different person characteristics (see previous experiment)

# Representational similarity analysis

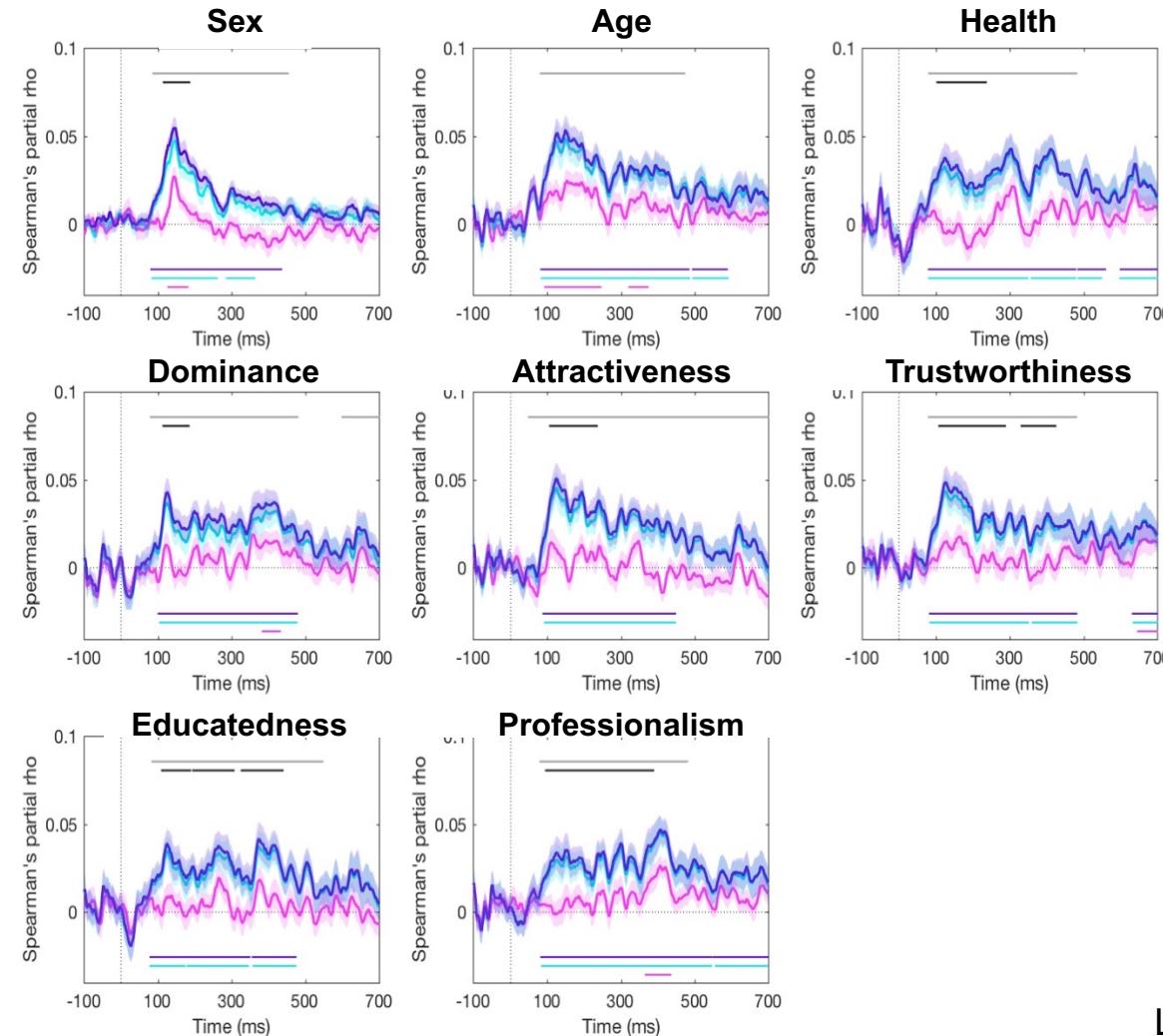


*Partial correlation for each time point (1ms resolution)*

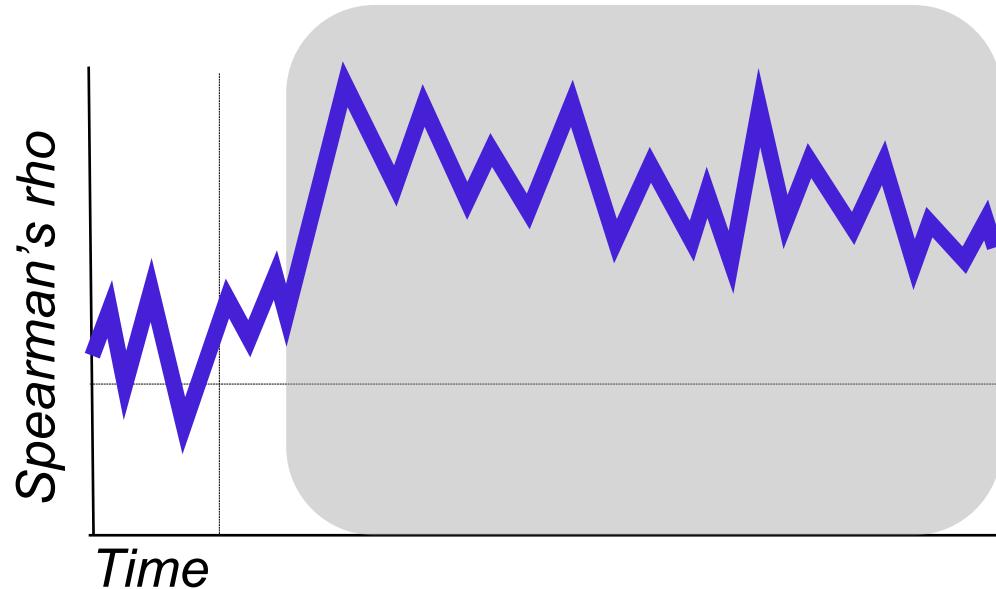


**If correlations are significant, person information is represented in the neural data at that time point.**

# What's the timecourse of impression formation in the brain?



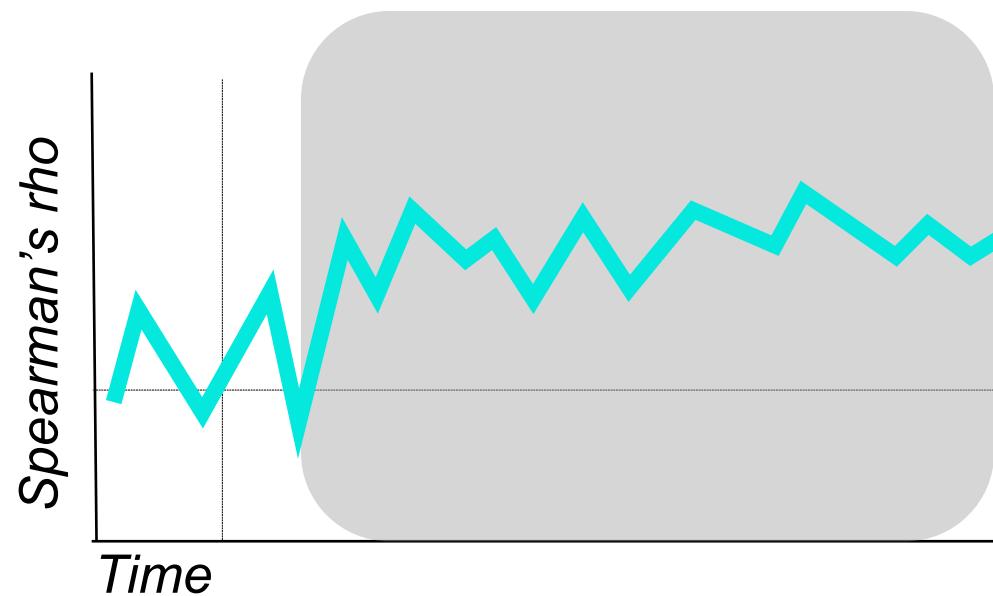
# What's the timecourse of impression formation in the brain?



We find representations of each person characteristic from 80ms after onset

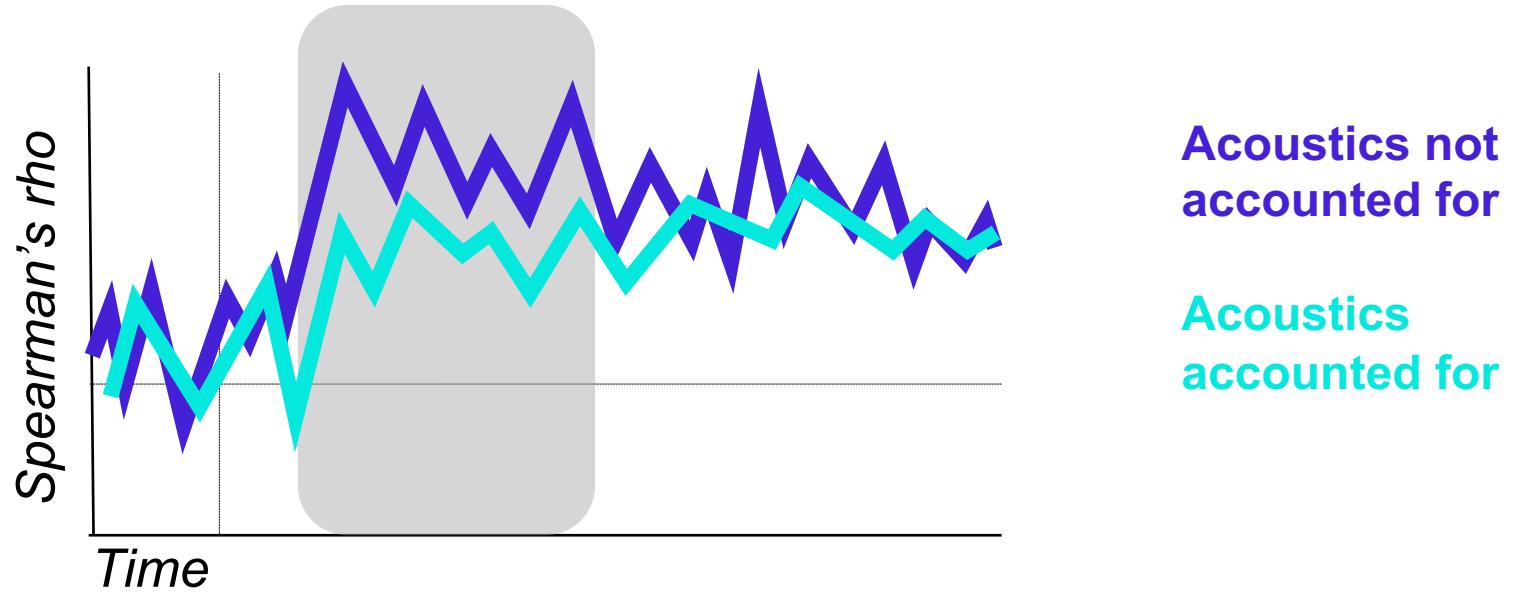
**What happens when we  
account for acoustic  
contributions?**

# What's the timecourse of impression formation in the brain?



We find still representations of each person characteristic from 80ms after onset

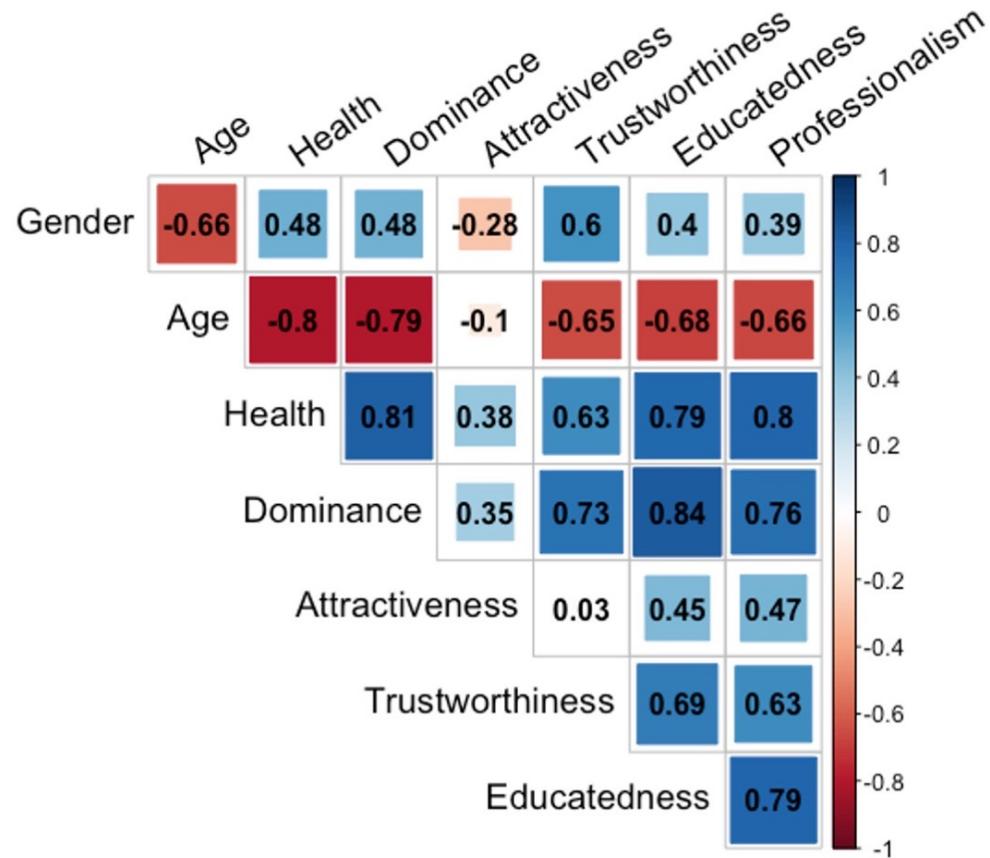
# What's the timecourse of impression formation in the brain?



Representations are sensitive to stimulus properties up 300ms after onset and become invariant after

**But characteristics are  
not perceived  
independently of one  
another**

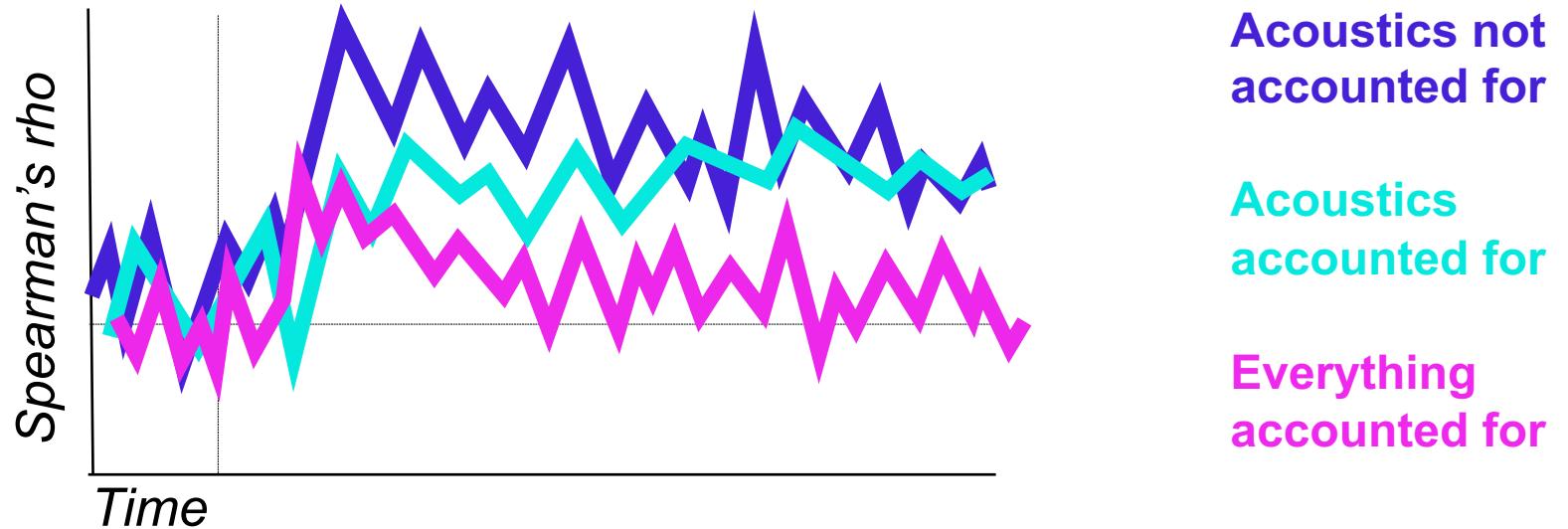
# What's the timecourse of impression formation in the brain?



Person characteristics are highly intercorrelated!

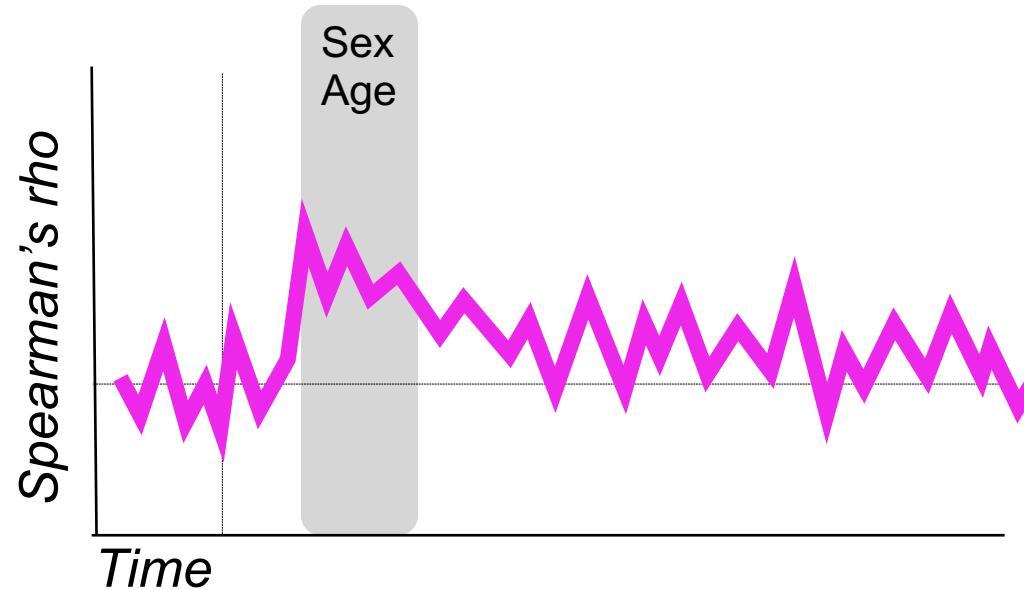
**What happens when we  
also account for all other  
person characteristics?**

# What's the timecourse of impression formation in the brain?



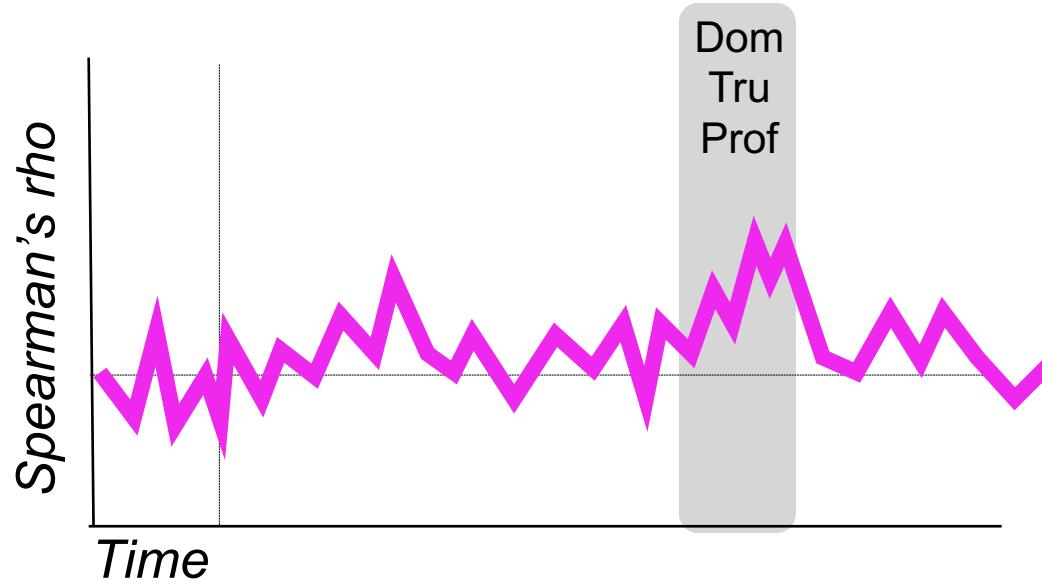
We can then look for “independent” representations

# What's the timecourse of impression formation in the brain?



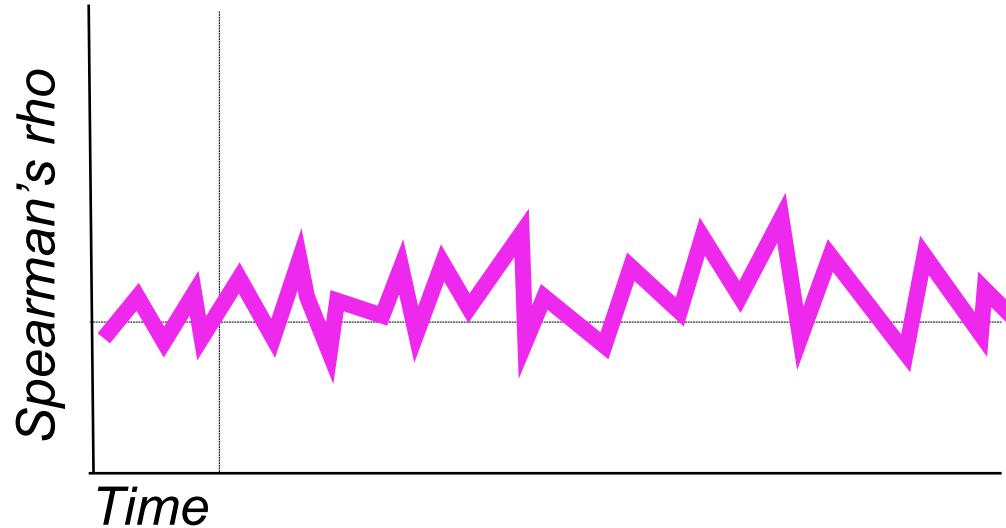
Independent representations of sex and age can be found early (starting pre-150ms)

# What's the timecourse of impression formation in the brain?



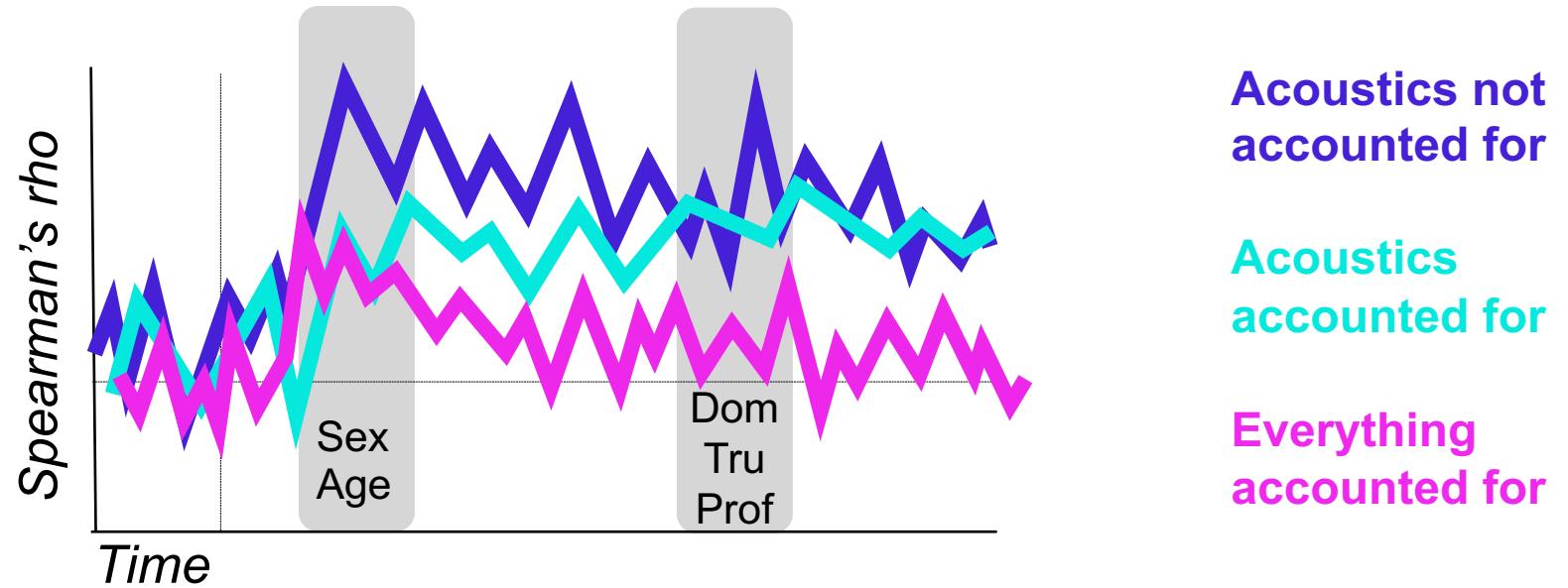
Independent representations of dominance, trustworthiness, and professionalism can be found later (starting post 350ms).

# What's the timecourse of impression formation in the brain?



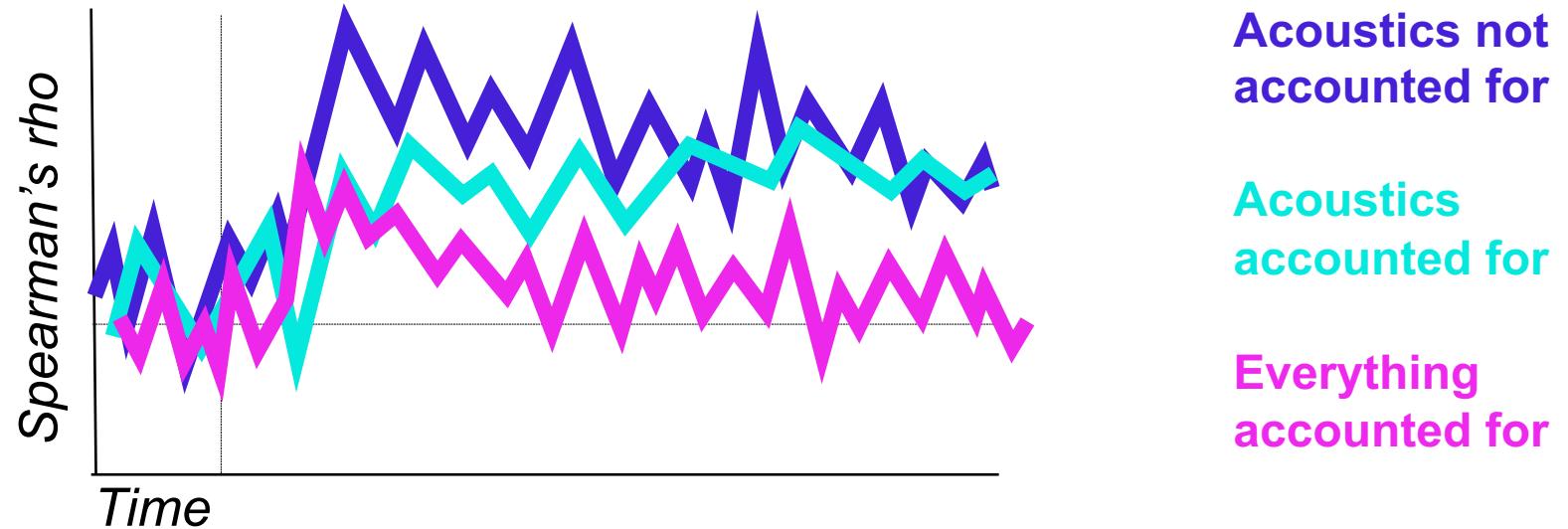
No independent representations of health, attractiveness, and educatedness.

# What's the timecourse of impression formation in the brain?



Some evidence for a staggered time course?

# What's the timecourse of impression formation in the brain?



But the main take-away is that characteristics are not perceived independently of one another

# Part 2.

## Person perception from voices

# Person perception from voices

- We quickly form an impression of a person based on their voice
  - The impressions do not need to be accurate!
- We need less than a second of exposure to form these impressions
- We find representations of the impressions in the brain from ~80ms of hearing a voice
- But impressions of individual person characteristics are highly intercorrelated
  - Halo or overgeneralisation effects? Or not very precise/well-resolved perception?

# **Part 3.**

# **Wrapping up**

**When we hear a voice,  
we hear a person**

**And there's so much  
more to find out...**

What about  
atypical voices?

What about  
atypical listeners?

What about  
development?

Personal taste  
in impressions?

Integration of non-  
linguistic and linguistic  
information?

Integration of face  
and voice  
information?

What about  
AI voices?

# Thank you



Carolyn  
McGettigan



Sophie  
Scott



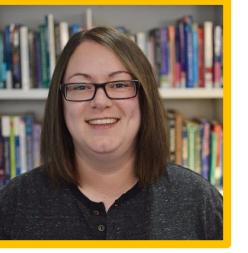
Mila  
Mileva



Lucia  
Garrido



Harriet  
Smith



Kay  
Ritchie



Clare  
Sutherland



Mike  
Burton



Sarah  
Knight



Isabelle  
Mareschal



Mathias  
Scharinger



Clare  
Lally



Bryony  
Payne



Elise  
Kanber



Paula  
Rinke



Andy  
Young



Valerie  
Hazan



Jens  
Kreitewolf



Michel  
Belyk



Stella  
Guldner

And many many more!

LEVERHULME  
TRUST

