"That Voice Sounds Familiar": Errors in Memory for Disguised Speech

Georgina Gous

Nottingham Trent University

georgina.gous@ntu.ac.uk

Overview of Talk

1. **Idea for Research**
   - Mullenix et al. (2010) and the Accentuation Effect
   - What they found
2. **My Research**
3. **Real World Implications**

   - how errors might arise during earwitness testimony for a suspects voice
4. **Future Work**

Overview of Research

- **Acoustic cues of the voice**
- **What are acoustic cues?**
  - Cues that are **directly measurable** from the speech signal and provide us with **paralinguistic information** about the speaker
  - **HOW** we say something
  - e.g. how loud, how fast or slow, how high or low in frequency/ pitch
- **Recognition performance for these cues**
- **Unfamiliar voice recognition**

Why is this important?

- Ability of listener to correctly recognise speaker depends on **INTER-** and **INTRA- speaker variability** in the voice

Why is this important?

- **INTER- speaker variation**: differences that exist in voice of different speakers (between-speaker variation).

Why is this important?

- **INTRA- speaker variation**: differences that exist in the voice of the same speaker (within-speaker variation).

Why is this important?

- **Natural variation**
- Speakers rarely sound exactly the same even when an utterance is produced in quick succession
- Other factors (e.g. time of day, mood state, emotional state, changes in health, intoxication)
- Robust to these changes

Why is this important?

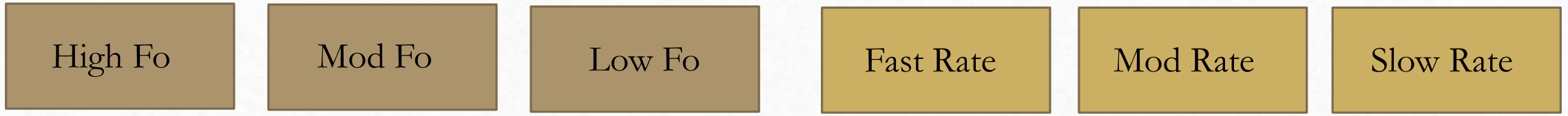
- **Accurate recognition can be problematic**
- Especially if **deliberately** try to alter characteristics of voice (e.g. voice disguise)
- Can provide substantial **acoustic variation** in the voice and fool the listeners ear successfully (Endres, Bambach, & Flosser, 1971)

Mullenix et al. (2010)

- **How accurately are acoustic cues recollected from memory?**
  - Manipulations in fundamental frequency (F0) and speech rate
  - **Fundamental Frequency (F0)** = frequency of vocal fold vibration. Measure of how high or low the frequency of a person's voice sounds (psychological correlate is perceived pitch)
  - **Speech Rate** = how fast or how slow someone is speaking

Mullenix et al. (2010)

- Created **high, moderate,** and **low frequency** voices AND **fast, moderate,** and **slow rate** voices (target voices – i.e. voices of interest)



- For each of these target voices, created distractor voices
- Manipulated versions of the target voice (higher or lower in frequency OR faster or slower in speech rate

Mullenix et al. (2010) - Method

2AFC: Presented with target voice and sequentially paired voices

-Previously heard target voice

-Manipulated version of target voice (higher or lower in F0 OR faster or slower in speech rate)

TARGET VOICE

1 sec

TARGET VOICE

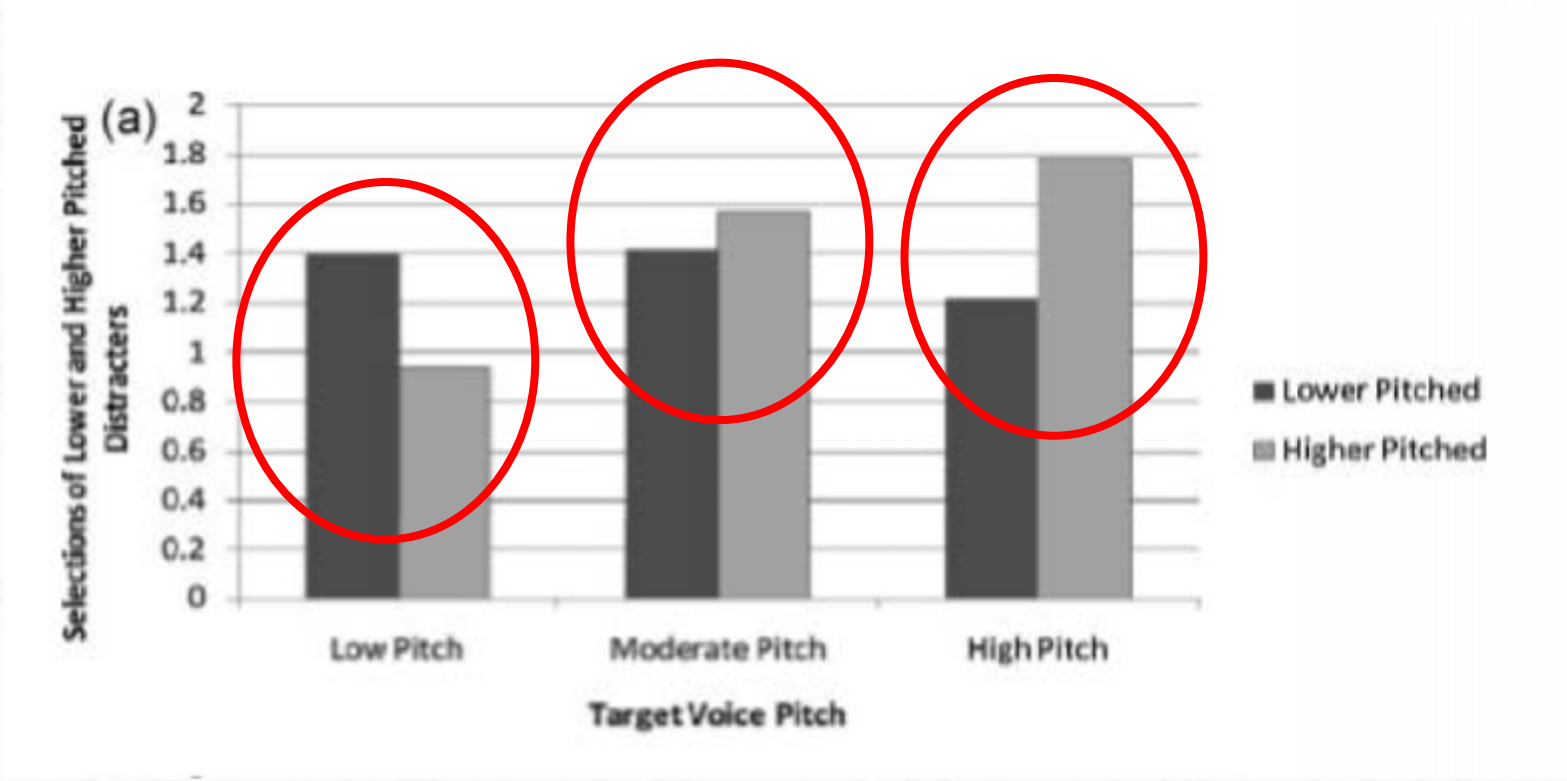
1 sec

MANIPULATED VOICE

'Was the voice you previously heard voice 1 or voice 2?' (key press1/2)

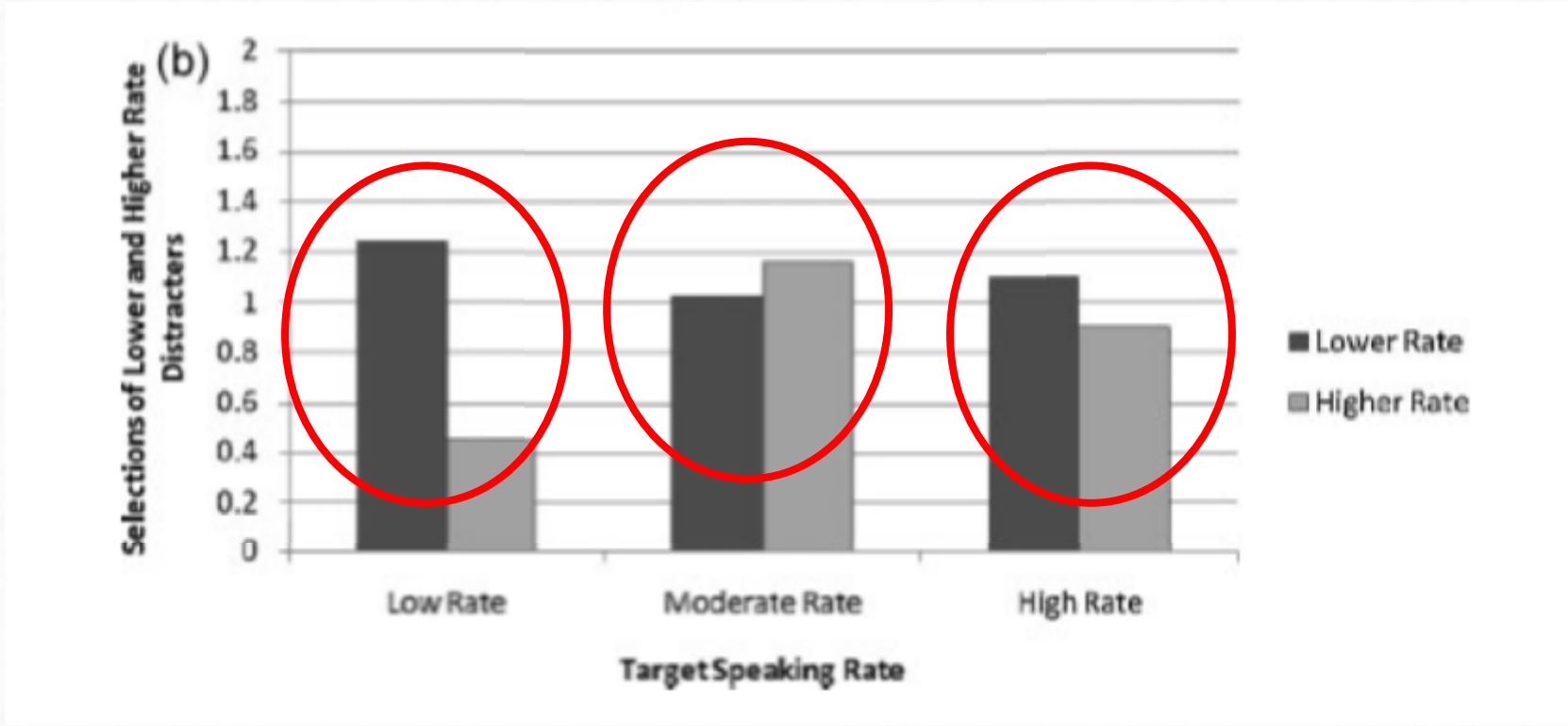
What did they find?

Fundamental Frequency (F0)



What did they find?

Speech Rate



Why does this happen?

ACCENTUATION EFFECT

- **Category based memory distortion**

- **Categorisation?** – cognitive process in which stimuli are recognised, differentiated, and understood
- Stimuli grouped into distinct categories for some specific purpose
- Ideally, this category illuminates a relationship between the stimuli
- Less cognitively effortful

Why does this happen?

ACCENTUATION EFFECT

- **More likely to make errors when remembering details about stimuli**
- Exaggerate similarities between stimuli in the same category
- Stimulus might be remembered as more closely matching that category rather than any individual differences that it actually has

Mullenix et al. (2010)

ACCENTUATION EFFECT

- Place voices into categories using most salient properties

high pitch voice → place into 'high pitch' category → remember this as being higher in pitch than it actually is

- In other words, memory for voice pitch has been **ACCENTUATED** towards **more typical features** of that category

Mullenix et al. (2010)

ACCENTUATION EFFECT

- **NOT** a general biasing process that produces distortions for all properties of voice

  - **Different properties** may be more or less susceptible to category-based memory distortions

  - **Transient** and **stable properties**

Implications

- **EARWITNESS TESTIMONY**
- Earwitness hears a perpetrators voice that is high (or low) in pitch
- Remember voice as being even higher (or evn lower) in pitch than it actually is

Implications

- Inaccurate statements given to police
- Less likely to recognise perpetrator of a crime

Implications

Innocent punished for a crime they did not commit

Perpetrator is released

Problems With Research So Far

- **One Male Voice**

- **Range of Voices Used**

  - target and distractor voices did not remain within typical values observed in population

  - F0: 80-180 Hz (males) and 165-255 Hz (in females)

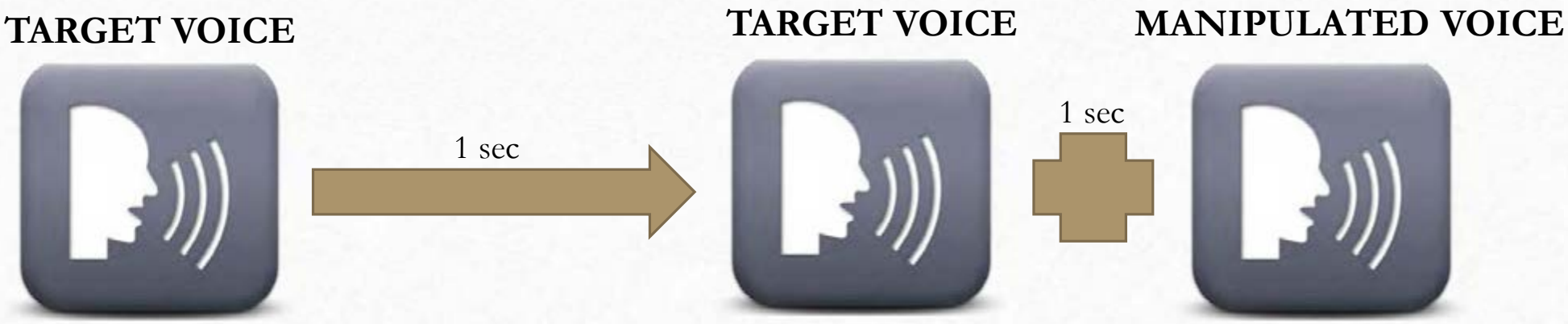
  - Speech Rate: 3.3-5.9 (syll/sec)

Method

2AFC: Presented with target voice and sequentially paired voices

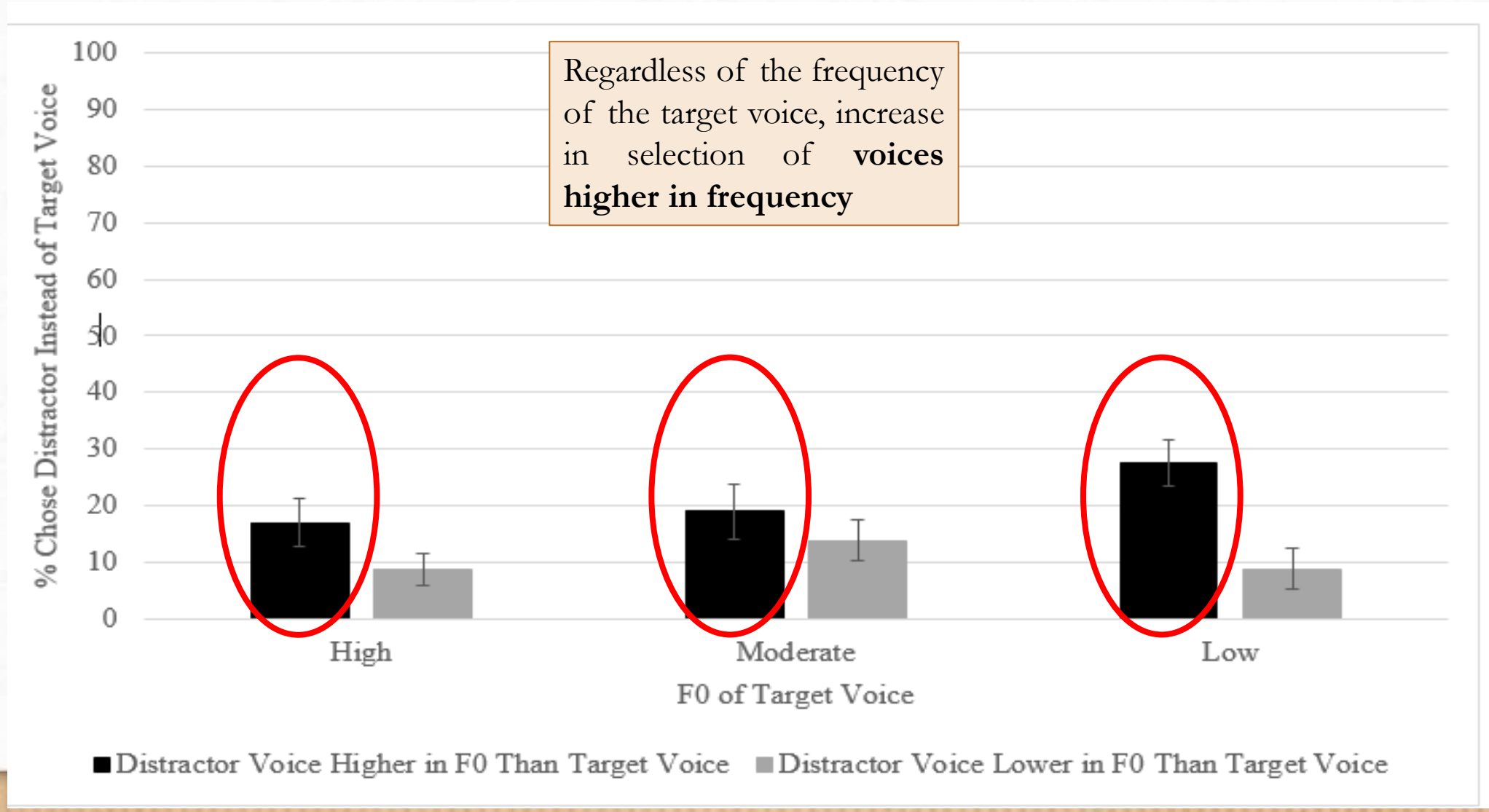
-Previously heard target voice

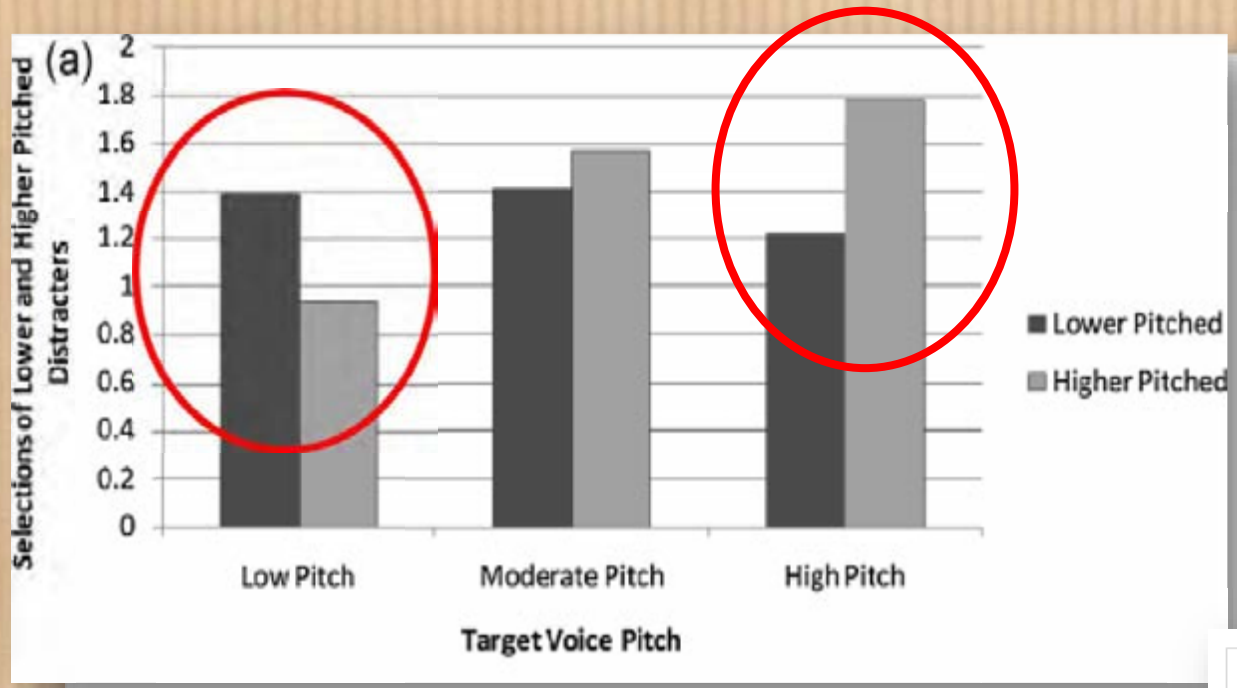
-Manipulated version of target voice (higher or lower in F0 OR faster or slower in speech rate)



'Was the voice you previously heard voice 1 or voice 2?' (key press1/2)

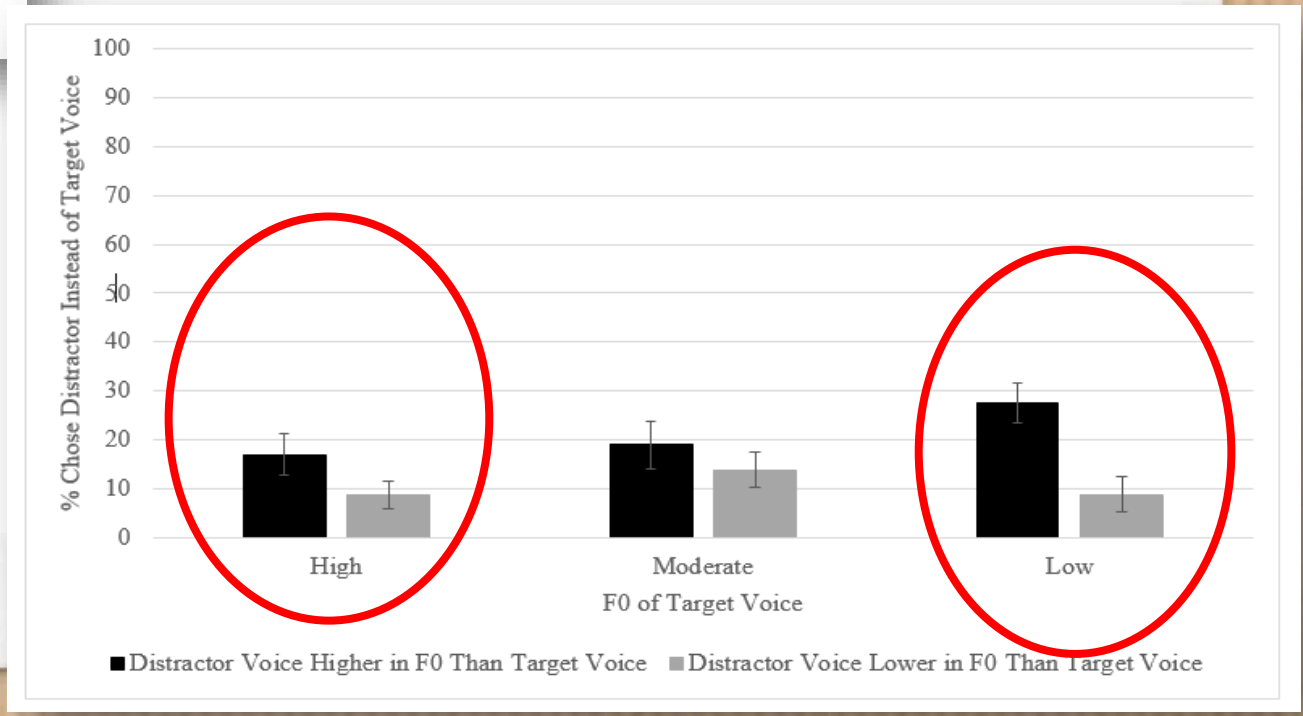
Fundamental Frequency (F0)



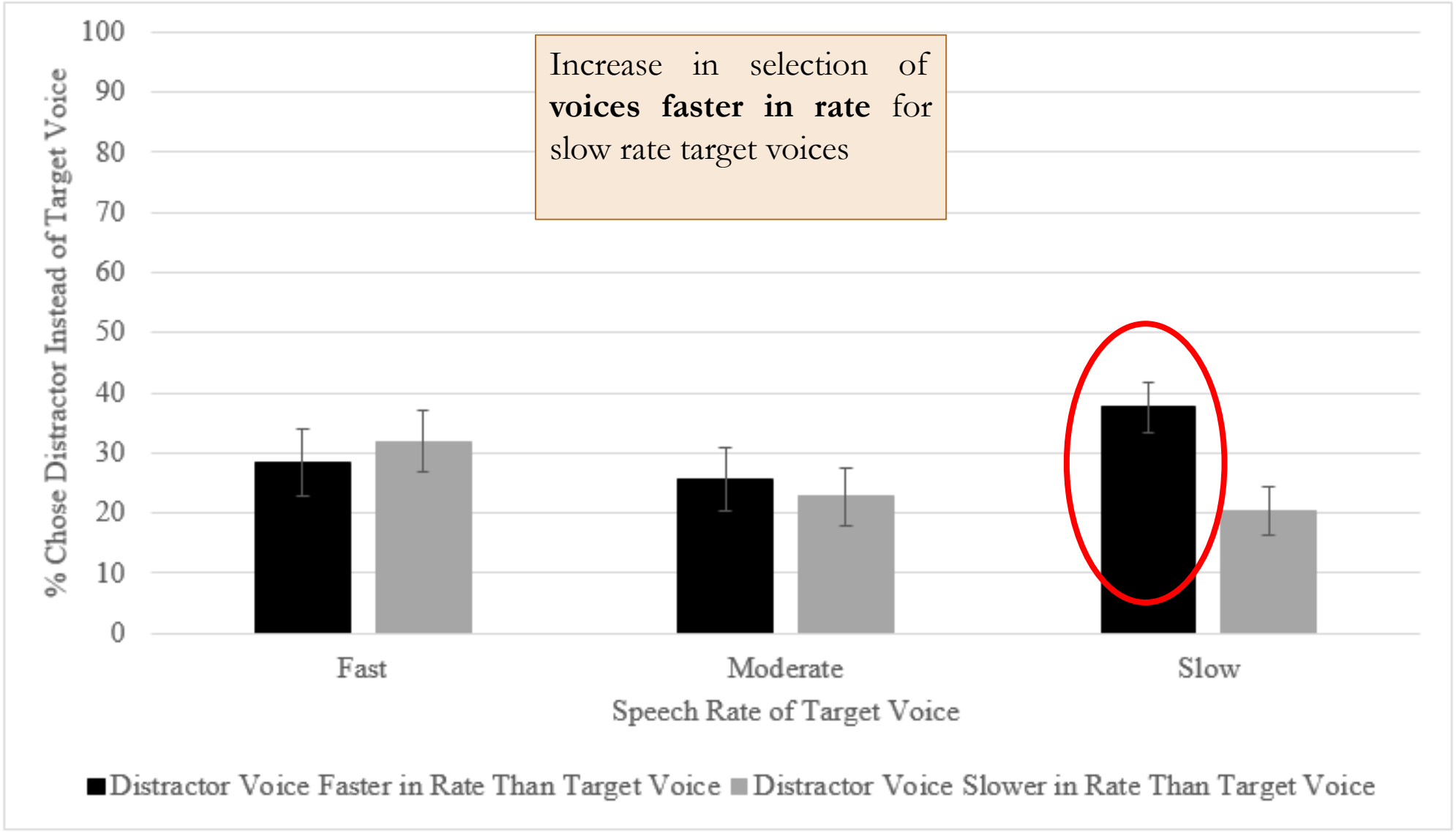


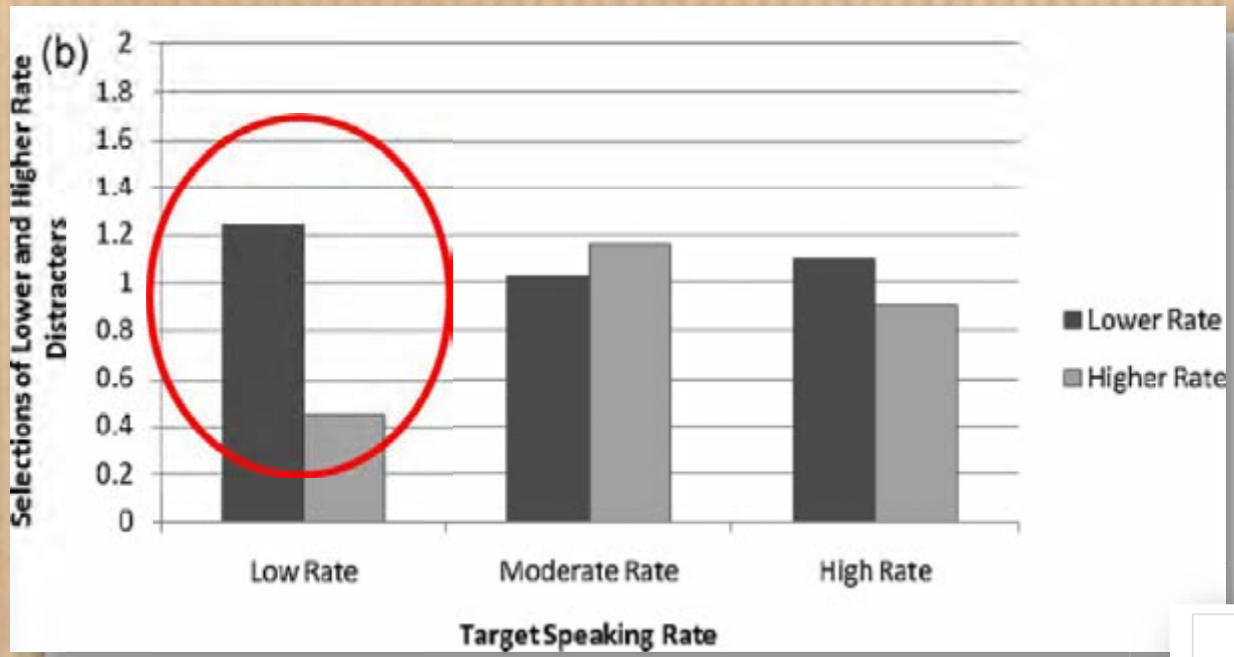
Mullenix et al. (2010)

Our Findings



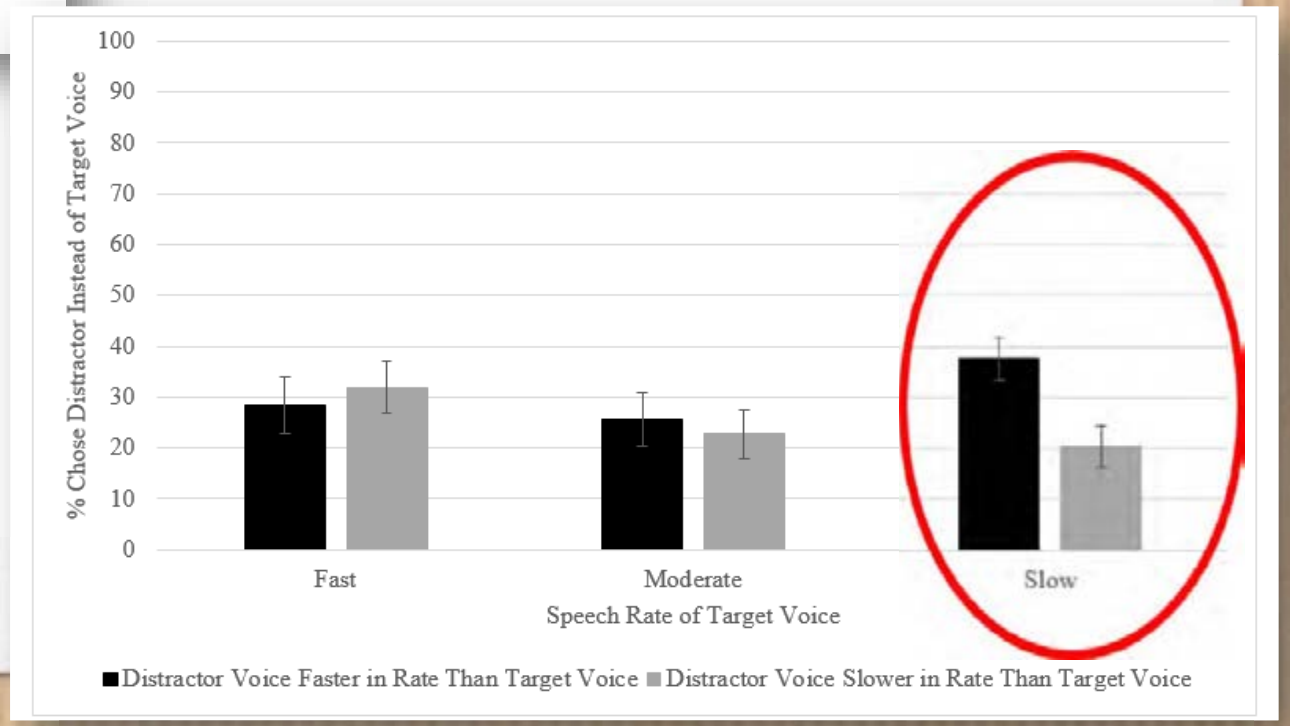
Speech Rate


Increase in selection of **voices faster in rate** for slow rate target voices
% Chose Distractor Instead of Target Voice
100
90
80
70
60
50
40
30
20
10
0
Fast
Moderate
Slow
Speech Rate of Target Voice
Distractor Voice Faster in Rate Than Target Voice
Distractor Voice Slower in Rate Than Target Voice



Mullenix et al. (2010)

Our Findings



Future Work

- **Lineup**
  - Hear target voice then manipulated version of the target voice amongst DIFFERENT voices (like a lineup)
  - 'Voice not present'

  a) in lab setting

  b) make more realistic (e.g. video of crime, hear voice only)

- **Retention Interval**
  - Hear target voice then come back and conduct lineup at a later date (e.g. one week later)

Concluding Comments

- Listeners **ARE** susceptible to distortions in memory for certain properties of voice
- At the very least, for frequency and speech rate
- Accentuation bias does not account for findings

ERRORS ARE OCCURING

Concluding Comments

- Important implications in the real world (accurate earwitness testimony)
- **Future?**
- Development of a useful conceptual tool in determining properties of voice that are more or less affected by intra-individual variation and voice disguise

THANK YOU FOR LISTENING

"That Voice Sounds Familiar":
Errors in Memory for Disguised Speech

Presented by Georgina Gous

Nottingham Trent University

georgina.gous@ntu.ac.uk