Evaluating Earwitness Identification Procedures: Adapting Pre-Parade Instructions and Parade Procedure

Nikolas Pautz
Nottingham Trent University

Harriet Smith, Jens Roeser, Josh Davis, Jeremy Robson, David Wright, Natalie Braber, Paula Stacey
Voice identification

• “The circumstances in which voice identification can provide crucial evidence in a criminal trial are infinitely variable.” (Robson, 2018, p. 220)
  • A masked perpetrator;
  • A blindfolded witness;
  • Offences committed over the phone or captured using audio surveillance.

• At least 150 cases of voice parades being used in England & Wales
  • E.g., *R v Khan and Bains*, discussed in Nolan, 2007

• While England and Wales does have some guidance, Voice ID parades are largely based on Face ID procedures

• Memory for voices and faces differs (Stevenage et al., 2011; Stevenage & Neil, 2014; Stevenage, 2019).
Why focus on system variables?

• For some unknown reason, most (but not all) previous earwitness research has focused on variables that we can’t actually control (estimator variables).

• The results presented here add to the slowing growing literature focusing on how system variables (variables we can control) can be manipulated to increase voice identification performance.

• We focused on the system variables of:
  • Parade procedures
  • Pre-parade instructions
Parade procedures

- Serial procedure: the witness listens to all nine voices at least once before making a decision (the target is present in positions 1-9 or they are not present at all)
  - This procedure may have high WM demands, contributing to task difficulty
  - Recommended by the Home Office
- Sequential procedure: the witnesses responds either ‘YES’ or ‘NO’ after listening to each voice, potentially reducing cognitive load and task difficulty.
  - Smith et al (2020) found higher hits and lower false alarms when comparing a sequential voice parade to a serial voice parade.
  - Sequential procedures may lead to a more conservative criterion in procedures (as has been found in face ID research) rather than increasing discriminability per se (Ebbesen & Flowe, 2002; Wixted, 2012),
Pre-parade instructions

- Content of pre-parade instructions has been found to influence the decisions of eyewitnesses
  - Stronger warnings reduce false alarms (Brewer & Wells, 2006; Meissner et al., 2005)
- Important to consider the form a warning should take:
  - Complex instructions may not have an effect (Wilcock et al., 2005)
  - Simple, criterion-based instructions have been found to improve discrimination (Meissner et al., 2005; Steblay, 1997)
- Standard warning: “the voice you heard in the original recording may or may not be present”
  - Such ‘unbiased’ instructions are mandatory in England and Wales
- Strong warning: “Please consider your response(s) carefully. In a real case, selecting someone from the lineup when the perpetrator is not present could lead to a wrongful conviction”
Hypotheses

• We expected that the strong warning would make participants less likely to false alarm in both types of parade.

• We expected that accuracy on target-present and target-absent parades would be higher for the sequential compared to the serial procedure.
Experiment 1 $N = 526$

- University of Greenwich Face and Voice Recognition Lab volunteer participant database
- 2 (parade type: serial, sequential) by 2 (parade instructions: strong, standard) by 2 (target presence: absent, present) between-subjects
- Three different target-groups with stimuli taken from the DYVIS forensic speech database (Nolan et al., 2009)

**Flowchart**

1. Voice (60 sec)
2. 5 min task
3. Voice parade
4. Decision

**IVs:**
- strong vs standard warning
- target: present vs absent
- parade: serial vs sequential

**DV:**
- accuracy (0, 1)
- self-rated confidence (0-10)
## Decision frequency

| Parade Type | Target-present | | | | Target-absent |
|-------------|----------------|----------------|----------------|----------------|
|             | Target         | Foil           | Reject         | Foil           | Reject         |
| Pre-parade Instructions | 25 (38%) | 38 (58%) | 3 (5%) | 55 (85%) | 10 (15%) |
| Standard Warning | 30 (46%) | 32 (49%) | 3 (5%) | 48 (73%) | 18 (27%) |
| Strong Warning | 32 (47%) | 30 (44%) | 6 (9%) | 53 (85%) | 9 (15%) |
| Serial       | 28 (45%) | 21 (34%) | 13 (21%) | 52 (72%) | 20 (28%) |

Total: 115 (44%) 121 (46%) 25 (10%) 208 (78%) 57 (22%)
Accuracy

Target absent

Target present

- Standard warning
- Strong warning

Response accuracy

Parade type

Sequential  Serial  Sequential  Serial
Signal Detection Model

The peaks of the posterior distribution represent the most likely parameter.
Experiment 1 conclusions

• Strong warnings improved participants’ ability to distinguish between targets
• Suggests that the wording used is sufficient to produce an effect in both serial and sequential procedures
• Did not observe differences between serial and sequential parades – in contrast with previous results (Smith et al., 2020)
  • This leads us to the second experiment
The number of ‘laps’

• Home Office (2003) guidelines recommend that participants listen to each serial parade voice at least once before making a decision.

• Smith et al. (2020) compared a serial parade with 2 ‘laps’ of the parade against a sequential parade and found a sequential superiority effect (not replicated in Exp1).

• In order to see if this may have been due to the number of passes, we compared identification performance between serial parades with 1 and 2 laps.

• The ‘lap effect’ has not been previously studied in earwitness identification
  • In the eyewitness literature, no evidence of performance benefit, and possibly can cause a more lenient response criterion (Steblay et al., 2011; Horry et al., 2015).
Experiment 2 $N = 225$

- Exactly the same procedure as Exp 1, but two laps of a standard warning serial parade
- 1-lap data were taken from Exp 1, standard warning condition.
- 2 (laps: 1 lap, 2 laps) by 2 (target presence: absent, present) between-subjects
### Decision frequency

<table>
<thead>
<tr>
<th>Number of laps</th>
<th>Target-present</th>
<th></th>
<th>Target-absent</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Target</td>
<td>Foil</td>
<td>Reject</td>
<td>Foil</td>
</tr>
<tr>
<td>1 lap</td>
<td>32 (47%)</td>
<td>30 (44%)</td>
<td>6 (9%)</td>
<td>53 (85%)</td>
</tr>
<tr>
<td>2 lap</td>
<td>24 (45%)</td>
<td>23 (43%)</td>
<td>6 (11%)</td>
<td>50 (93%)</td>
</tr>
<tr>
<td>Total</td>
<td>56 (46%)</td>
<td>53 (44%)</td>
<td>12 (10%)</td>
<td>103 (89%)</td>
</tr>
</tbody>
</table>
Signal Detection Model

Response criterion $c$

Signal sensitivity $d'$
Conclusions

• Our results underline the value of system variable research in voice ID as well as the need for replication and thorough testing before policy recommendations are made.
• The serial procedure recommended by the Home Office can be easily adapted to provide increased levels of protection for innocent suspects by adapting pre-parade instructions that encourage more conservative response behaviour, without decreasing the probability of successful identification.
  • As an extrinsic cue (based on the cue-belief model – Leippe et al., 2009) the strong warning may communicate task difficulty.
• We did not find any accuracy differences between serial and sequential parades.
  • It is unlikely to be due to the number of laps that participants have been exposed to in the serial parade.
Thanks for being earwitnesses to this presentation 😊

Preprint: https://psyarxiv.com/nxr3e/