The effect of voice sample duration and lineup size on voice identification performance

Harriet M. J. Smith, Nikolas Pautz, Kirsty E. McDougall, Katrin Mueller-Johnson, Alice Paver, & Francis Nolan
Voice identification evidence

ID evidence can be determinative

In some crimes visual information is not available

Voice identification is admissible evidence in jurisdictions worldwide
What are the main issues?

High false alarm rates (Kerstholt et al., 2004; Stevenage et al., 2012, 2013)

Juries find voice identification evidence extremely persuasive (Van Wallendael et al., 1994)

Unfamiliar voice identification is under-researched, especially when it comes to system variables
Improving voice identification procedures (IVIP)

ESRC-funded project (ES/S015965/1)

Multi-disciplinary approach (psychology, forensic phonetics, linguistics, criminology & law)

4 strands:

Strand 1: What are the optimal parameter values for voice parade procedures?

Strand 2: What are the psycho-phonetic underpinnings of voice distinctiveness?

Strand 3: How do social stereotypes affect voice identification?

Strand 4: How accurate are the normative assumptions of criminal justice practitioners in respect of voice identification?
1. Representative sample of the suspect speaking naturally.

2. Voice samples should be 1 minute long.

3. Voice parade should consist of 9 voices.

4. Witness must be instructed that the voice of the suspect may or may not be present.

5. The witness must listen to each tape at least once before making a selection.

6. The witness must be allowed to listen to the samples as many times as they wish.
Experiment 1

Can sample durations be reduced without a performance cost?

Practical considerations – time consuming for the police

- People can extract basic identity information from much shorter durations (Bestelmeyer et al., 2010; McAleer et al., 2014)
- Voice samples should be 1 minute long
- Voice parade will take at least 9 minutes
- Temporal ratio models of memory (Bjork and Whitten, 1974; Brown et al., 2007)
In ‘real’ voice lineups recordings of the suspect and foil voices are taken from recordings of police interviews.

Forensics-orientated speech databases

- Dynamic Variability in Speech (DYVIS)
- York Variation in Speech (YORVIS)
- West Yorkshire Regional English Database (WYRED)

These databases include recorded telephone calls of a perpetrator discussing a crime and mock police interviews.
Experiment 1 $N=271$ (135 female)

**IVs**
- Voice (60 sec)
- 5 min task
- Voice lineup
- Lineup sample: 15s, 30s or 60s
- Perpetrator: present or absent

**DVs**
- Decision
- Accuracy
- Self-rated confidence (0-10)
## Decision frequency

<table>
<thead>
<tr>
<th>Sample Duration</th>
<th>Target Present</th>
<th>Target Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hits</td>
<td>Foil</td>
</tr>
<tr>
<td>15 seconds</td>
<td>20 (45%)</td>
<td>21 (48%)</td>
</tr>
<tr>
<td>30 seconds</td>
<td>14 (32%)</td>
<td>26 (59%)</td>
</tr>
<tr>
<td>60 seconds</td>
<td>17 (37%)</td>
<td>27 (59%)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>51 (38%)</strong></td>
<td><strong>74 (55%)</strong></td>
</tr>
</tbody>
</table>
**Accuracy**

![Graph showing accuracy over different sample lengths](image)

- **Target Absent**
- **Target Present**

- Percent Correct vs. Parade Sample Length

Monday, 04 October 2021
SDT analyses

Response criterion $c$  
Signal sensitivity $d'$

Posterior density

Signal strength

-3 -2 -1 0 1 2 3 4

15s duration
30s duration
60s duration

15s duration
30s duration
60s duration

Monday, 04 October 2021
Confidence

Target Absent

Target Present

Median Confidence

Parade Sample Length

15s 30s 60s 15s 30s 60s
Experiment 1 conclusions

The results underline the importance of admitting voice identification with caution. Consistent with previous research, performance is low.

Our results highlight the importance of stimulus sampling. Many earwitness studies have been conducted using single targets.

No evidence to suggest that there is any benefit in using lineup samples of 60 s. These preliminary results suggest that the voice identification procedure currently recommended in England and Wales can be safely adapted by reducing the duration to 30 s or even 15 s.
Experiment 2

Can lineup size be reduced without a performance cost?

Voice parades should consist of 9 voices

Practical considerations
Larger lineups offer more protection to innocent suspect?

But erroneous auditory attention is more likely to occur when the demand on resources is high (Zimmerman, Moscovitch & Alain, 2016)
Experiment 2 $N=278$ (136 female)

**IVs**
- Voice (60 sec)
- 5 min task
- Voice lineup

**Lineup sample:** 15s, 30s or 60s

**Perpetrator:** present or absent

**DVs**
- Decision
- Accuracy
- Self-rated confidence (0-10)
## Decision frequency

<table>
<thead>
<tr>
<th>Sample Duration</th>
<th>Target Present</th>
<th></th>
<th></th>
<th>Target Absent</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit</td>
<td>Foil</td>
<td>Reject</td>
<td>Foil</td>
<td>Reject</td>
<td>Reject</td>
</tr>
<tr>
<td>15 seconds</td>
<td>16 (36%)</td>
<td>26 (58%)</td>
<td>3 (7%)</td>
<td>36 (78%)</td>
<td>10 (22%)</td>
<td></td>
</tr>
<tr>
<td>30 seconds</td>
<td>14 (33%)</td>
<td>22 (51%)</td>
<td>7 (16%)</td>
<td>37 (82%)</td>
<td>8 (18%)</td>
<td></td>
</tr>
<tr>
<td>60 seconds</td>
<td>21 (46%)</td>
<td>21 (46%)</td>
<td>4 (9%)</td>
<td>37 (82%)</td>
<td>8 (18%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td><strong>51 (38%)</strong></td>
<td><strong>69 (51%)</strong></td>
<td><strong>14 (10%)</strong></td>
<td><strong>110 (81%)</strong></td>
<td><strong>26 (19%)</strong></td>
<td></td>
</tr>
</tbody>
</table>
Accuracy

[Graph showing accuracy data for different sample lengths with target absent and target present conditions.]
SDT analyses
Confidence

- **Target Absent**
  - 15s: Median Confidence
  - 30s: Median Confidence
  - 60s: Median Confidence

- **Target Present**
  - 15s: Median Confidence
  - 30s: Median Confidence
  - 60s: Median Confidence

*Parade Sample Length*
Experiment 1 and 2 comparison
Conclusions

Poor performance - but the task is not impossible if the target is present

Reduce sample duration? ✓

Reduce the number of foils? ✗
Thank you for listening

@harrietsmith15
Harriet M. J. Smith
harriet.smith02@ntu.ac.uk