Thesis Final Manuscript: Earwitness Identification Memory

Taylor Smith

April 10, 2015
Abstract

Earwitness identification can be used as evidence in criminal trials. Judges and jurors may weigh earwitness testimonies heavily, though research indicates inconsistencies in accuracy rates. Earwitness memory can be impacted by the presence of facial cues and by whether or not the perpetrator’s voice is in the lineup. I examined voice identification accuracy by presenting participants with either criminal or non-criminal audio scenarios followed by either present or absent five voice lineups. Participants were told that the perpetrator's voice may or may not be in the voice line-up and then listened to all voices in the lineup and decided who, if anyone, was the perpetrator. There was an increase in accuracy of picking the perpetrator's voice when it was present but did not differ between the criminal and non-criminal scenarios. This is consistent with some of the existing earwitness research, which finds high accuracy rates.
Earwitness Identification Memory

Eyewitness research has dominated the field of forensic psychology mainly due to the fact that it is possibly the most reliable way to identify a person, besides physical evidence. Since vision is developmentally more advance than our other senses, it seems logical to assume that we would be more accurate at identifying a criminal with vision versus audition. Most research focuses on eyewitness identification. Although, the literature is dominated by eyewitness research, other cues (e.g. voices) are often presented during crimes that are worth examining.

Although few studies have looked at face and voice memory together, it is common for a witness to both see and hear a crime (Cook & Wilding, 1997a). McAllister, Dale, and Keay (1993) also investigated voice and face recognition. Participants viewed a photo and listened to a tape recording simultaneously. They found that seeing the photo and hearing the voice did not reduce facial identification accuracy. They also found that accuracy decreased in voice identification. These studies suggest voice identification accuracy seems to be impacted by facial presence more than face identification accuracy is impacted by auditory presence.

Similarly, Cook and Wilding (1997a) looked at the connection between voice, face, and context. They wanted to examine the effect that context had on the accuracy of identifying the voice. All participants listened to the same audiotape. Participants in the context group were given a copy of a context sentence or facial context to view while they listened to the voice line-up (Cook & Wilding, 1997a). There was a significant effect of face presence on voice identification of the target voice. They found that accuracy decreased when facial context (pictures were
shown along with the audio clip) was included. Also, accuracy increased when personal information context (information of the person speaking) or both types of context were included. This suggests that face presence can effect voice identification depending on the stimuli presented. Facial presence seems to interfere with ability to accurately identify a voice. This idea is connected to the phenomena of face overshadowing. Face overshadowing is when facial context influences the witness's ability to accurately remember a voice because the facial context is overshadowing the vocal context being heard.

Witnesses often experience both visual and auditory stimuli from the crime. McAllister, Dale, Bregman, McCabe and Cotton (1993) conducted a study with both visual and auditory stimuli presented. The purpose of the first experiment was to examine if auditory information interfered with the process of learning visual information. They found that participants were more accurate at identifying a guilty suspect with just visual information compared to having both visual and auditory information. McAllister et al. (1993) also looked at the interference of visual information on auditory information. They found that participants were more accurate at identifying a voice from a line-up when only auditory information was presented during witnessing. Stevenage, Howland and Tipplet (2011) also examined interference of eye and earwitness identification. They found that there was significant interference of voice recognition when a face was presented. Stevenage et al. (2011) found that face recognition was able to withstand the presence of a voice (as a distractor), compared to voice recognition that was affected by the
presence of a face (as a distractor). These findings are consistent with results from Cook and Wildings (1997a), as mentioned above.

As discussed above, research has proposed that facial presence effects learning of a voice. This is possibly because voice recognition is suggested to have a weaker pathway than face recognition (Stevenage et al., 2013). Humans depend on the visual system for a lot cues about the environment or emotions. We have developed a stronger ability to focus on visual stimuli even when other stimuli are presented. Therefore, voice recognition would be more susceptible to distraction or interference than face recognition. Stevenage et al. (2013) demonstrated that face recognition stayed constant when presented with a distractor, whereas voice recognition decreased as the distraction level increased.

Another factor often examined in voice identification research is time-delay. By the time witnesses are interviewed, asked to give identification evidence, or give a testimony in court weeks, months or even years have passed. This time-delay between the crime and the trial can have detrimental effects on voice recognition accuracy. Saslove and Yarmey (1980) examined long-term auditory memory in relation to speaker identification. They found that speaker identification is only accurate under specific learning and retention intervals. Therefore, this suggests that earwitnesses are inaccurate at voice identification. However, high levels of accuracy were found in the informed group (they knew they were being tested on the target voice). They also found that if the voice was constant throughout both trials, accuracy increased compared to when the voice was changed between each trial. Saslove and Yarmey (1980) concluded that participants were more accurate at
identifying a voice after a delay period if it is the same voice and not changed (e.g. tone).

Clifford, Rathborn, and Bull (1981) also examined the effect of time-delay on voice identification. Participants were required to listen to recordings and then asked to identify the target voice. Before identifying the target voice, participants were asked to wait a 10, 40, 100, or 130- minutes (thus, different time-delays). They found that time-delay had little effect on voice recognition. Further the less distinctive the target’s voice the more time-delay had an effect on voice recognition. Clifford et al. (1981), also examined longer retention intervals of 10 minutes, 24 hours, 7 days and 14 days. They found that the shortest interval revealed the most accurate results and the highest recognition performance. They also found that voices with less distinctive features decreased accuracy, where as the more distinctive features for example, accent, tone, speech pathology increased accuracy over the delay periods. Yarmey and Matthey (1992) also examined time-delay and found that the accuracy was not affected by the delay period but error rates were. Time-delay did not affect how many participants correctly identified the target voice, but it did effect how many participants incorrectly identified the target voice. They examined the influence of retention interval and distribution exposure to the target voice, and found that false alarms were consistently higher at each exposure condition (Yarmey & Matthey, 1992).

Van Wallendel, Surace, Parsons and Brown (1994) also examined the effect of time-delay in voice identification, just in a different context. They had participants hear a sales pitch followed by a voice line-up presented at 0, 7, or 14 day time-
delays. Longer time intervals did not have a significant impact between the 0 and 14-day delay in terms of accuracy and discrimination, although the 7-day interval did show high scores on discrimination compared to the immediate and 14-day delay conditions. Although higher scores were found for the 7-day interval, this was not consistent across all four voices.

Reid and Craik (1995) examined the effects of voice passages and tone over a time-delay of 17-days. Tone is important to examine because most witnesses will be asked to identify a voice from a conversational sample and not from the original phrase or sentence heard. They found that voice identification accuracy was not better than chance when the target spoke a different passage. They also found that if participants heard the original recording, identification significantly compared to when the target spoke a different passage. Kerholt, Jansen, Van Amelsvoort and Broeders (2004) also examined time-delay and speech duration. In the target absent condition they found that there was not effect of time-delay when the speech duration was short. However, longer duration resulted in fewer errors over a week time-delay compared to a longer duration with an immediate time-delay.

Yarmey (1993) reviewed research that had already been conducted in the duration of voice sample. Yarmey (1993) found that accuracy of identification is better when there is a longer duration of a person speaking. Although longer duration seemed to help with accuracy of identification, it was also increasing error rate as well. Previously Legge et al. (1984) found that participant’s accuracy did not differ significantly from chance when the stimulus was presented for 6 seconds. However, when a voice stimulus was presented for 1 minute, accuracy scores
improved (Cook & Wilding, 1997b). Researchers have also examined the impact of variety and length of voice samples used in voice recognition. Cook and Wilding (1997b) examined how the variation a length of a sentence would affect the accuracy of voice recognition in known and unknown voices. They found that the target voice was more often positively identified when high variety sentences were longer compared to shorter in duration. Cook and Wilding (1997b) found a significant effect of length on enhancing long-term memory of a once heard voice, rather than variety, which was concluded in a previous study. These results show that the length utterance of a target voice can be of significance when trying to accurately identify a voice.

An important methodological aspect of identification research is the inclusion of both target present and target absent line-up conditions when the suspect is innocent versus guilty. It is common for both ear and eyewitnesses to pick a person from a line-up whether or not they are the actual perpetrator. This is because many people tend to believe that when police say the suspect is in custody that the suspect is guilty. Researchers have confirmed that earwitnesses are more accurate at identifying the target voice when the suspect is present in the line-up compared to when the suspect is absent (Philippon, Cherryman, Bull & Vriji. 2007). Philippon et al. (2007) examined voice identification accuracy of familiar versus unfamiliar language in target absent versus target present line-ups. Earwitnesses were more likely to positively identify a voice that was spoken in a familiar language and when the target was presented in the line-up (Philippon et al., 2007). As in previous literature, they found that participants would be less accurate when the
target was absent compared to when the target was present (Kerstholt, Jansen, Van Amelsvoort and Broeders, 2006).

Though there seems to be a difference between the present/absent line-ups, the number of suspects within a line-up also seems to be a contributing factor in voice recognition accuracy. Yarmey, Yarmey and Yarmey (1994) examined face and voice recognition in both show ups and line-ups. They exposed participants to a target voice in person for a 15 second duration. Once the target had walked away, a female investigator approached the participant, and asked them to recall the target that had just spoken to them. Participants were asked to recall both auditory and visual characteristic to be able to identify the target. They found that participants were more accurate at identification when it was a 6-person line-up compared to a one-person show up. However the 6-person line-ups were more susceptible to errors than the one-person show up (Yarmey et al., 1994).

Reviewing all the studies above, there are many factors that affect the accuracy of positive identification. Since there is minimal research in the area of ear witness testimonies, it is important to investigate this area. Looking at the inconsistencies in evidence, I plan on adding to the research by examining target present/absent line-ups and both criminal and non-criminal scenarios using a voice identification paradigm. By examining scenarios (criminal and non-criminal), I plan to determine the impact that context has on accuracy of voice identification. This is relevant to the legal system because some crimes are more impactful than others (e.g. a telephone scam versus an abduction). Investigating the influence of certain scenarios can be informative when trying to decide how much weight should be given to a piece of evidence.
Method

Participants

Seventy-two participants (54 women, 14 men, 2 other and 1 chose not to fill out the socio-demographic sheet μ age= 20.5) were randomly assigned to one of four groups, (18 participants per group). Participants were recruited from introductory psychology courses at Algoma University and received class credit for their participation.

Materials

**Voices.** A total of seven voices were recorded that met the following criteria: all speakers were Caucasian, between the ages of 18 and 30, raised in Northern Ontario, and without any speech pathology as identified by the experimenter. There were 6 male voices and 1 female voice recorded.

**Audio passages.** The male speakers all read the same two passages. One was a criminal passage, which had the male speakers act as a criminal and one was a non-criminal passage, which had the male speaker act as themselves. The experimenter created both of these passages by examining common crimes and controversial topics at the time. The criminal passage included a male asking for directions and then stealing the woman’s purse at the end, whereas the non-criminal passage included the male and the female talking about the recent shooting that occurred in Ottawa. One male speaker was recruited as the target voice, whereas the other males were recruited to be part of the suspect line-up. The suspect male speaker recorded his passage with the female speaker. For the line-up, all 6 male speakers (including the target speaker) were told to read the lines for both passages.
and speak normally (for example without exaggeration). All recordings took place in a sound proof room at Algoma University. All voices were recorded on the Zoom H4N (Zoom corporation, Japan) and transferred to a computer.

**Equipment.** The recordings were open played on Quick Time Player (version 10.3 copyright 2009-2013 Apple Inc). Participants listened to the audio clip and 5-voice line-up via headphones (Klipsch/Mode M40).

**Procedure**

Participants were run individually in a sound proof room at Algoma University. There were four conditions that participants could be placed in: criminal scene (target present), criminal scene (target absent), non-criminal scene (target present), non-criminal scene (target absent). Participant’s placement into these conditions was randomized.

Participants each listened to one audio recording (criminal or non-criminal). All recordings of the relevant male/female conversation were approximately 30 seconds long. After they had listened to the recording they were presented with a voice lineup for the purpose of identifying the suspect voice among the 5-voices presented. The voice recording and each voice within the line-up were only played once for the participant. The order of the line-up was randomized when the target was absent or present. When the target was present, the recording was placed at either the beginning, middle or end of the line-up. The suspects voice position was also randomized to be at the beginning, middle or end of the line-up when target voice was present. Participants were asked to circle “yes” if the voice was the male voice from the scene or circle “no” if the voice was not the male voice from the
scene. They were told that there was a possibility that the suspect voice from the audio recording was not in the line-up.

**Results**

A chi-square was used to examine the effect of scenario type (Criminal vs. Non-criminal) and line-up (target voice present vs. target voice absent). An alpha level of $p < .05$ was set for the statistical analysis.

**Main Effects of Line-up Type**

There was a statistically significant main effect of target present and target absent line-up. Participants were more accurate at identifying the target voice in the present line-up, $X^2 (1, N=36) = 3.50, p < 0.05$, participants were significantly more accurate at identifying the target voice whether in the criminal or non-criminal scenario (80.6% correct identification and 19.4% error rate). Compared to the target absent line-up $X^2 (1, N=36) = 8.50, p < 0.05$, participants were inaccurate or inconsistent at being able to identify the target voice (52.8% correct identification and 47.2% error rate) compared to the present line-up condition (Refer to figure 1 and figure 2).

**Main Effects of Scenario Type**

The type of scenario had no significant influence on participant’s identification accuracy. Specifically, participant’s correct identification accuracy was similar in both the criminal and non-criminal scenario. Participants in the criminal scenario made 77.8% correct identifications compared to the 83.3% of correct Identification made in the non-criminal scenario when presented with a present line-up condition. When participants were presented with absent line-ups,
within the criminal scenario they correctly identified 44.4% versus 61.1% in the non-criminal scenario. Although there seems to be a larger gap between the correct identification in the absent line-up condition compared to the present line-up conditions it was non-significant (Refer to figure 1 and figure 2).

Interactions

No significant interactions were found between scenario type and line-up type. Participants correctly identified the target voice 80.6% of the time in the present line-up compared to the 52.8% of correct rejections made in the absent line-up but this difference was not statistically significant.

Discussion

Judges and jurors use earwitness testimonies and identifications frequently in court, despite research showing people are often inaccurate at identifying another person’s voice. My research examined the effect of scenario type and line-up type on voice identification accuracy. I hypnotized that participant’s would be more accurate at identifying the target voice in the criminal scenario or when presented with the target present line-up.

Although no effects were found in the scenario condition, significant effects were found in the line-up conditions. Previous studies have found that witnesses are more accurate at identifying a target voice in a present line-up compared to an absent line-up (Philippon, Cherryman, Bull & Vrij. 2007). My results showed that scenario type did not affect identification accuracy, but line-up type did. I presented participants with a scenario that either included a crime or not followed by either a target present or target absent line-up. My results were consistent with previous
research in finding that there was a significant effect of line-up type (Kerstholt, Jansen, Van Amelsvoort and Broeders 2006).

There were a total of five suspects that could have been chosen instead of the target voice in an absent line-up, but only four in a present line-up. Innocent suspects were randomly ordered in each trial, so there was no order effect. We found that suspect one (19.4%) and suspect two (11.1 %) were more likely to be identified as the target voice in the absent condition (Refer to chart 1). Each participant was told that the male voice from the clip may or not be in the line-up. This shows that innocent voices that were chosen as the target voice must have had some similar qualities as the target voice. If that is not the case possibly participants were guessing to make identification even if the target voice was absent (feeling pressured to make a identification).

Order of the voices may have also had an influence on results in the present line-up. I did control for an order effect, but the order of the target voice in the present line-up may have a significant result with more participants. I randomly placed the target voice in three locations within the line-up (beginning, middle, or end). Examining whether or not the position in which the target voice is presented affects the accuracy of correct identifications would be useful.

The findings in this study can have numerous implications within law enforcement and criminal trials. I found that participants were as likely to make correct identifications as error in the criminal and non-criminal scenarios. This implies that witnesses may have no better chance in identifying the voices of a criminal in the absent line-up, which implies a danger of false identification/
wrongful convictions. Compared to that of a person they talked to in more formal contact. Also for law enforcement these findings can help improve identification rates. By apprehending the correct person and placing them in the line-up witnesses have a better chance of identifying the person. Although this can be difficult to do for law enforcement, it would possibly result in higher correct identification rates.

**Limitations**

Even though we did find a significant result of line-up, there were some limitations to the study. For instance the criminal scenario that was chosen was of a male asking for directions and then mugging the female at the end of the audio clip. If the crime would have been more extreme (e.g. assault) there may have been an effect of scenario. Also the target voice that was chosen may have been distinct compared to the innocent voices chosen. With more time and resources, a line-up with more similar voices may have impacted the correct identification rate and been more applicable (since we had a high correct identification rate in the present line-up).

**Acknowledgements**

This research was conducted at Algoma University and I would like to thank the university and the psychology department for the support and materials they provided to conduct this study.
References


Figure 1. The absent line-up condition shows more error in both scenario types. The criminal scenario however shows a higher error rate, but this difference is non-significant 52.8% correct identification and 47.2% error rate.

Figure 2. The present line-up conditions a higher accuracy rate compared to the absent line-up. Participant’s correct identification were higher in the non-criminal scenario, but this result is non-significant 80.6% correct identification and 19.4% error rate.
<table>
<thead>
<tr>
<th>Suspect accuracy</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid suspect 1</td>
<td>7</td>
<td>19.4</td>
<td>41.2</td>
<td>41.2</td>
</tr>
<tr>
<td>suspect 2</td>
<td>4</td>
<td>11.1</td>
<td>23.5</td>
<td>64.7</td>
</tr>
<tr>
<td>suspect 3</td>
<td>1</td>
<td>2.8</td>
<td>5.9</td>
<td>70.6</td>
</tr>
<tr>
<td>suspect 4</td>
<td>3</td>
<td>8.3</td>
<td>17.6</td>
<td>88.2</td>
</tr>
<tr>
<td>suspect 6</td>
<td>2</td>
<td>5.6</td>
<td>11.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>47.2</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Missing System</td>
<td>19</td>
<td>52.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>100.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chart 1. This chart describes the percentage at which innocent suspects were chosen in the absent line-up instead of making the correct decision. Suspect 1 was chosen the most as the innocent target (19.4 %) when the male target voice was not presented. This chart shows that some of the innocent voices were more likely to be chosen than others when the target voice was not presented.