

Understanding earwitness testimony

Most students are familiar with eyewitness testimony, but earwitness testimony might be valuable too. Cognitive specialist **Andrew K. Dunn** considers the evidence

Balink Radu/Fotolia

Earwitness and eyewitness testimony (what someone reports hearing or seeing in relation to an alleged crime) is both an essential element of many detective dramas, and of great real-world value in legal settings. For some crimes, the only useful and available evidence is earwitness/eyewitness testimony (Yarmey 1995). This kind of evidence is highly prized by law enforcement agencies, not least because juries are often greatly persuaded by this form of evidence.

Unfortunately, our memory for events can often be unreliable (e.g. Loftus and Palmer 1974), even when (and sometimes because) we are confident about what we have seen or heard. Therefore, it is of real-world importance that we improve our understanding of the way we process, represent and recollect the world around us.

Earwitness testimony

There has been a lot of focus on eyewitness testimony but very little on earwitness testimony. This is odd since witnesses might actually see very little if, for example, the criminal is masked or the witness is hiding. Here we will focus on earwitness testimony, specifically the ability to recognise and identify individual voices.

Some basic problems

Recognising and identifying a voice and working out what is being said is far more difficult than it first appears, not least because there is a remarkable amount of natural variation within and between individual speech patterns (see Box 1). Additionally, most earwitnesses are asked to describe or identify an unfamiliar and possibly disguised voice (and what was said) long after hearing it having only heard it for a few brief but stressful moments (Yarmey 1995). Even when voices are very familiar to us, we can be easily mistaken about who it is that is speaking simply because the voice is out of context.

Box 1 How do voices differ?

Individual voices differ in a number of important ways (not including contextual differences and speech disorders) that can be used to identify a speaker and what is being said. These include:

- vocal quality (timbre and clarity)
- dynamics and style (intonation, rhythm, stress, timing and speed)
- accent and dialect
- loudness (intensity – dBels) and pitch (frequency – Hertz)
- vocabulary (formal, regional and slang)

Explorations in the lab

In our lab-based studies, we focus on three main variants:

- voice recognition — being able to say if you've heard a voice before
- voice recall (identification) — saying who said it
- content recall — remembering what was said but not necessarily who said it

Voice recognition is at a lower level than voice recall/identification. It's a bit like saying 'I know that voice [recognition], but I can't tell you who it is [recall/identification]'.

Many lab-based studies on earwitness testimony employ some form of voice line-up methodology. This is because it offers a great deal of control over the variables that might influence earwitness performance. Typically, voice line-ups involve participants having to first listen to a spoken word or sentence, from one or more speakers, learn the voice or content, and then after a delay (from a few minutes to several days or longer), identify one or more voices from amongst never before heard voices. There is variation in these methods and

they are by no means ideal, but they have produced some very interesting and unexpected findings.

Voice recognition and identification accuracy

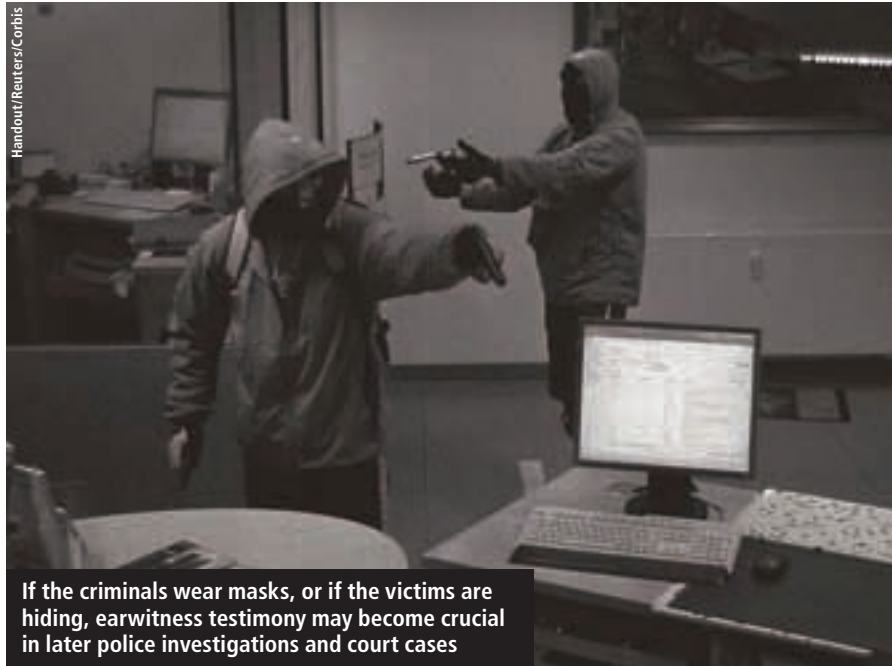
Some early work predicted that (unfamiliar) voice line-up accuracy should be similar to that of eyewitness line-up accuracy — around 55–60% accurate (Yarmey et al. 1994). In fact, performance was very much lower than expected; between 9% and 28% depending on the number of voices present in the line-up. Familiar voice recognition can be better, with performance accuracy ranging from 33% to 98%, depending on the level of familiarity (Yarmey 1995).

So why is performance often so bad?

We simply don't know why earwitness performance is often so poor. It has been suggested that there might not be enough contextual information at recall, especially after a delay, to access the memory of the originally heard voice. Certainly learning (encoding) voice content when the face is also present, can improve content recall when asked to recall the content with the face present (Campos and Alonso-Quecuty 2006). Presumably this is because there is more contextual information from the original context available to help access memory.

However, visual information at the point of learning might actually disrupt voice encoding and hence disrupt subsequent ability to recall voice. Consider these further findings:

- a Performance should be better when recalling a voice after learning the voice with speaker's face present. There are two conditions: learning voice + face or learning voice only. Participants are subsequently asked to identify the voice. We might expect participants to do better in the former condition because context information is available. However research shows that performance is about the same in both conditions (Stevenage et al. 2011)
- b This leads us to believe that there should be no difference between face recall/recognition and voice recall/recognition after



If the criminals wear masks, or if the victims are hiding, earwitness testimony may become crucial in later police investigations and court cases

learning face + voice together. In fact, voice recall is significantly worse than face recall under these conditions (Stevenage et al. 2011). The face overshadows the recollection process and impairs performance overall, sometimes called the 'face overshadowing effect'.

This suggests that the problem is at the level of encoding (learning), possibly because voices (especially if unfamiliar or less distinctive) are more weakly encoded than faces and/or have a more fragile representation in the brain.

Real-world implications

Such research is useful to know theoretically (in terms of how we process voices or don't) but also very important in practical settings.

For example, imagine that you had an innocent bystander/earwitness — someone who didn't see the criminal because the criminal wore a mask or the bystander was hiding and only overheard

Box 2 Some practical problems in gathering earwitness and eyewitness testimony

An obvious way of finding out what someone sounds (or looks) like is to get a witness to describe the person or persons concerned. Unfortunately, most people are very bad at doing this (try it yourself) because we simply don't have the vocabulary that we need. Worse, our descriptions can be unintentionally biased by natural memory processes and/or by being accidentally misled by the interviewer during the interview.

Problems with using live line-ups

Another method used for gathering earwitness and eyewitness testimony is a line-up procedure. In a line-up, witnesses might be presented with real people (live line-up), recordings of voices, videos or photographs, and then asked to select the possible target.

This seems like a sensible method but, despite a long history of use, there is no generally agreed or standardised formal procedure amongst law enforcement agencies. Thus there is no agreement on where it should take place, how many people, faces or voices should be used in the line-up, or how similar/different they should be from the target. Nor is it clear how best to present them — sequentially (one at a time) or in parallel (all at once).

Live-line ups can be particularly stressful for a witness and can yield false positive reports (identification of someone even when the proposed culprit is absent from the line-up).

The effects of line-up problems

It is not clear how many people may have been wrongly accused or convicted because of problems associated with live line-ups. However, a study of eyewitness performance in 314 real, live line-ups shows a worrying rate of 20% false positive identifications and 40% no selection at all (Valentine et al. 2003). Live line-ups are rarely used these days because they are so problematic.

what was said (this is like the experimental condition of voice only followed by voice + face). Next, imagine that the police show them a video of people talking (e.g. CCTV footage of the event; a sequential line-up of possible culprits) and then ask them to select the person they had heard but did not see. There is a real possibility that they may not be able to identify anyone or, worse, identify the wrong person because the memory of the voice and the speaker are disrupted by the presence of the face even if what was said is unaffected.

An obvious way around this would be to get the witness to close their eyes (to remove the face information). However, if you wanted to know what was said (or if it matches with what they thought they had heard) as well, then closing their eyes would be less productive. So, you have to be careful about what you ask and how you ask it or try to get at it.

Concluding remarks

Earwitness testimony is highly prized and highly persuasive, but it can also be highly unreliable. It is therefore important for practical (and theoretical) reasons that we understand the factors that contribute to inaccurate earwitness performance and voice processing more generally, if we are to avoid miscarriages of justice.

There are many relevant factors (e.g. duration of exposure, familiarity and distinctiveness of the voice) that we simply cannot manipulate or control in the real world but we can manipulate and control in a psychology lab. As a result, lab-based research has made significant inroads to better understanding earwitness performance that can be applied in real-world situations.

Further reading



Campos, L. and Alonso-Quecuty, M. (2006) 'Remembering a criminal conversation: beyond eyewitness testimony', *Memory*, Vol. 14, No. 1, pp.27–36.

Stevenage, S. V., Howland, A. and Tippelt, A. (2011) 'Interference in eyewitness and earwitness recognition', *Applied Cognitive Psychology*, Vol. 25, No. 1, pp.112–118.

Valentine, T., Pickering, A. and Darling, S. (2003) 'Characteristics of eyewitness identification that predict the outcome of real line-ups', *Applied Cognitive Psychology*, Vol. 17, No. 8, pp.969–993.

Yarmey, A. D., Yarmey, A. L. and Yarmey, M. J. (1994) 'Face and voice identifications in showups and line-ups', *Applied Cognitive Psychology*, Vol. 8, No. 5, pp.453–464.

Yarmey, A. D. (1995) 'Earwitness speaker identification', *Psychology, Public Policy, and Law*, Vol. 1, No. 4, p.792–816.

PsychologyReviewExtras



Go to the **PSYCHOLOGY REVIEW Extras** page (see back cover) for a PowerPoint presentation to accompany this column.

Andrew Dunn is a senior lecturer at Nottingham Trent University, with interests in perception, attention, memory and evolution. He is particularly interested in face, voice and body-shape processing, and location memory.

PHILIP ALLAN FOR
HODDER
EDUCATION

Getting the most from your magazine?

magazines
extras

Extra resources online

Get free revision exercises, weblinks, podcasts and lots more, linked to articles in each issue of every A-level magazine.

Simply go to www.hoddereducation.co.uk/magazineextras and select resources to view for each magazine.

You can also subscribe to any Philip Allan magazine at www.hoddereducation.co.uk/magazines

Subscribe to the online archives

Subscriptions to the online archives of back issues are available for all our A-level magazines. Go to www.hoddereducation.co.uk/magazines for:

- 100s of articles from digital versions of back issues
- Simple, accessible and comprehensive search
- Unlimited access for staff and students, with instant weblinks from your VLE
- Annual subscriptions for everyone to share, **from as little as £70!**

Subscribe today at www.hoddereducation.co.uk/magazines

online archives