

Earwitness Memory: Effects of Facial Concealment on the Face Overshadowing Effect

Amanda J. Heath

*School of Psychology, University of Central Lancashire
AJHeath@uclan.ac.uk*

Keeley Moore

*School of Psychology, University of Central Lancashire
KMoore2@uclan.ac.uk*

Abstract

The face overshadowing effect (FOE) has been noted in cases where recognition of voices is impaired if they are presented simultaneous to a face at encoding. The current study investigated the effect of facial concealment (with and without wearing a balaclava) and emotionality of vocal tone on the face overshadowing effect in voice identification. It was predicted that the FOE would be reduced in the case of presentation of a concealed face along with voices, as the lack of facial feature information would result in greater attention being paid to the voice. It was further anticipated that angry voices would attract more attention and result in better voice recognition and reductions in the FOE than neutral voices, as hostile voices represent a level of threat that captures attention analogous to the weapon focus effect in eyewitness memory. Results replicated the FOE in a voice plus face video presentation but, contrary to expectations; a concealed face presentation also demonstrated a FOE, with highest accuracy of voice identification in the voice only condition. Angry vocal tone had a slight tendency to result in better recognition of voices across groups and somewhat improve performance in the visual conditions. It was concluded that voice identification is as fallible and prone to error as eyewitness identification but that conditions where the voice is made salient and visual information is absent result in higher accuracy. Implications for the criminal justice system are discussed.

Keywords: Earwitness, FOE, voice identification

1. Introduction

Earwitness memory, the recall and recognition of auditory information by witnesses, has traditionally received much less attention in the psychological literature than eyewitness memory. While the latter, in particular perpetrator identification has been well researched and factors impacting upon it extensively documented, there remain a number of unknowns about witness memory for auditory information. It may be that some of the factors known to affect eyewitness memory similarly affect earwitness memory, but this is yet to be fully established and there may be crucial differences. The current study is concerned specifically with investigating the interaction of visual and auditory information on memory of speakers' voices.

Though the vast majority of real forensic cases involving witness testimony implicate eyewitness memory, there are notable exceptions where memories of auditory events and information, especially perpetrator voice, provide pivotal evidence in criminal cases [1]. An oft reported case in the literature is that of Bruno Richard Hauptmann, who was convicted of the kidnap and murder of aviator Charles Lindbergh's baby son and executed for the crime in the United States in 1936 [2]. Acting as a key prosecution witness, Lindbergh identified the voice of Hauptmann as the same one he had heard roughly three years earlier shout from the bushes at the ransom drop at a cemetery [3]. Since that time voice identification evidence has featured controversially in a number of criminal prosecutions [2], [4]. In 1983 Clifford, in a less than exhaustive survey of British legal cases, found 188 cases in which voice identification evidence was used [5]. Memory of a perpetrator's voice may be the only identification evidence where conditions of visibility are poor, where perpetrators conceal their appearance or blindfold the victim, or where it involves a telephone conversation (such as a ransom demand or a bomb threat). Given this, and the finding that juries find witness testimony a particularly convincing form of evidence regardless of its well-established fallibility [6], [7], the practical implications of research in this area become obvious.

If there has been a relative dearth in research for witness memory of auditory information, then studies on the interaction between visual and auditory information at encoding have been even more scarce [8]. Some studies have focused on the effects on speech content [9] while others on voice identification itself [8], [10], [11]. Of this last group, Yarmey presented slides of a mock rape and a simultaneous voice recording. The level of illumination presented on the slides was manipulated with photographic filters to simulate different times of day and night. Yarmey speculated auditory memory would increase as the level of light decreased because in the low light conditions participants would attend more to the voice. However, he found no such relationship between decreasing visibility and voice recognition. There were some limitations with Yarmey's method, as the author recognises; the use of slides and a separate audio taped voice is not naturalistic, but more importantly there was no audio only condition, and without an appropriate control condition the investigation was unable to fully test the experimental hypothesis. There was also something of a floor effect with performance very poor overall, and Yarmey suggested that this may have minimised the influence of the illumination manipulation [12].

In their paper on cross-modal effects, McAllister and colleagues specifically set out to address some of the methodological issues with Yarmey's study [8]. They used a mock crime method where participants were given photographs of the supposed perpetrator simultaneous to hearing a voice recording and compared recognition performance to an audio only condition. The researchers found that the presence of visual information interfered with the processing of auditory information resulting in poorer voice recognition, with voice recognition best in the audio only condition.

The findings of McAllister and colleagues lend weight to the idea of attention capture and dominance in the visual modality. Attention research has demonstrated that if both visual and auditory information are presented together, attention will be captured in the visual modality which is the dominant processing channel [13]. Given that humans show particular preferences for attending to faces, demonstrated even in newborns [14], we would therefore expect a particular interference effect with face stimuli. It also follows that conditions in which attention is oriented to the voice should improve recognition.

Indeed, Cook and Wilding [10], [11] conducted several naturalistic studies using audio and video of speakers that addressed specifically the presence of a face on unfamiliar voice identification. Cook and Wilding coined the term 'face overshadowing effect' (FOE) to describe their results, whereby the simultaneous presentation of a face during encoding of a

voice impairs later voice identification. In one of their studies [11], the researchers found that hearing an unfamiliar voice three times improved recognition but explicit instructions to attend to the voice did not, with face overshadowing unaffected by instructions. They concluded that the FOE is not due to voluntary attention to face over voice but involuntary attention to the *initial exposure* to the face. In a further experiment they demonstrated that pre-exposure to the face prior to the voice did indeed reduce the FOE effect. Their findings also indicated that memory for speech is processed primarily for meaning as they found an absence of FOE on memory for words uttered while voice identity was impaired.

If visual information at encoding impairs memory and later identification of voices, and if faces in particular capture attentional processing, then a pertinent question to ask is what effect facial concealment might have on voice memory. Yarmey's [12] study varied levels of illumination and found low rates of recognition with even poor visual conditions at encoding. McAllister and colleagues [8] and Cook and Wilding [10, [11] found that audio/voice only encoding conditions were superior to those where visual and face stimuli were presented. Arguably disguise or concealment of the face (which is common in robberies and sexual assaults) might reduce attention capture and increase focus on the voice thereby reducing or removing the interference of visual information. However, it could also be argued on the basis of Yarmey's work that any visual stimulus presented with a voice would produce interference, though his study has some acknowledged limitations [12].

The current research sought to examine the effect of the absence/presence of a face at encoding, and in particular the effect of facial concealment on overshadowing in voice identification. In addition it sought to elucidate further the idea of attention capture by making voice information more salient, thereby reducing the interference effect. Explicit instructions to attend to the voice have not reduced the effect in previous work [11], and it may be the case that dominance of the visual modality involves *automatic* processing that is not easy to consciously override. The current research, therefore, manipulated emotionality of vocal tone, comparing angry and neutral voices, as an implicit method of orienting attention to the voice. There is a wealth of evidence in the eyewitness literature to show that the presence of a weapon at encoding interferes with memory for other aspects of the event due to attention narrowing on the threatening stimulus [15], a phenomenon known as weapon focus [16]. Arguably angry vocal tone represents a level of threat that is somewhat analogous to that found in weapon focus. Indeed research in the face literature has found attention is selectively orientated to angry faces, which are better remembered than neutral ones [17].

On the basis of this previous work a number of predictions were made. A face overshadowing effect was expected to occur in a voice plus face video presentation, with correct recognition being lower than in a voice only audio group. It was expected that the presence of a concealed face alongside a voice, however, would show much less of a face overshadowing effect on identification, as more attention here should be paid to the voice in the absence of distracting facial features. Angry vocal tone was expected to somewhat reduce the face overshadowing effect as it draws attention to the voice in something akin to the weapon focus effect in eyewitness memory. Consequently, the highest rate of accuracy was expected with a combination of voice only and angry vocal tone.

2. Method

2.1 Design

The study employed a 2 (emotionality of voice) X 3 (presentation group) mixed-factorial design. Emotionality of voice was varied within groups with all participants exposed to a line-

up of 6 males speaking short utterances, three angry voices and three neutral voices. The method in which voices were presented was varied between groups and participants were randomly allocated to one of three conditions: 1) a Voice only group where participants heard an auditory line-up of the 6 male voices with no visual input, 2) a Face and voice group where participants viewed a videotape line-up of males speaking the utterance with visible face and shoulders, or 3) a Concealed face group where participants viewed a videotape of each male speaking, as in the Face group, but in this case wearing a full balaclava to hide all facial features except eyes and mouth.

After a short distraction task, participants performed an old/new (yes/no) auditory recognition test containing four previously presented targets (two angry and two neutral) and four foils (two angry and two neutral) not previously presented. There were no target-absent line-ups. This yielded data in the form of correct recognitions and incorrect recognitions or false alarms where participants picked foils as old rather than new.

2.2 Participants

Fifty four undergraduate students were recruited using opportunity sampling from around the School of Psychology building, UCLAN university campus in Preston. The participants were randomly allocated to experimental group with equal numbers of 18 in each presentation condition. The sample consisted of 23 males and 31 females, and the age range was 18-31 years old ($M = 21$, $SD = 2.65$). None of the participants knew the speakers in the line-ups.

2.3. Materials

2.3.1 Video: Ten volunteers were recruited using opportunity sampling from around the university campus. All were aged between 19 and 24 years old, most had mild Lancashire English regional accents. Each participant was taken to a quiet room and videotaped in colour, facing the camera with visible head and shoulder view, speaking two short sentences in a neutral tone of voice and again in an angry tone of voice and with and without wearing a full-face balaclava that covered all facial features except for eyes and mouth. They were asked to repeat this twice so that the best versions (most authentic in terms of emotionality) could be chosen for the experiment.

The two phrases used in the study were (phrase A) "If everyone does exactly as I say no-one will get hurt" and (phrase B) "I'm not the type of guy you want to be messing with right now." These were chosen to roughly match the 15-syllable utterances used in Cook and Wilding's study [10] and to convey a degree of threat relevant to a forensically real situation.

2.3.2 Audio: The test materials for the Voice only group and the recognition line-ups were constructed by ripping the audio from the video using Adobe Premier Movie Editing software and played to participants using a laptop computer.

2.3.3 Presentation line-ups: Six voices (and faces depending on group) formed the line-up in the presentation phase, 3 neutral and 3 angry, and all speakers within a line-up spoke the same sentence. There was a 5-second gap between each voice or voice/face. Order of voices in the line-up was fully randomised across the sample and half the participants in each experimental group heard one sentence, phrase A, and half the other, phrase B.

2.3.4 Recognition line-ups: During the recognition phase, participants were presented with a line-up of 8 voices, 4 old (2 neutral and 2 angry) and 4 new, randomly selected foils (2 neutral and 2 angry). No target-absent line-ups were used. The same phrases were used at recognition

as had been heard by participants in the presentation phase. The order of presentation of targets and foils at recognition was fully randomised across the sample.

An auditory distracter was employed to provide some delay between presentation and test and prevent participants from auditory rehearsal of the voices. This took the form of a music quiz using short clips from 10 popular songs. There was a 5-second gap between each song for participants to write their answer to the previous clip and they were required to write artist and song title. The presentation lasted 3.5 minutes in total and was created with Windows Movie Maker software.

Windows Media Player and a laptop computer were used to present all audio and video materials.

2.4 Procedure

Participants were tested individually. On arrival they were randomly allocated to presentation condition and informed that the experiment tested memories of things seen and heard. They were given the appropriate set of written instructions for their group which informed them that they would be required to see and hear speakers, or just hear in the Voice only group, and would later be tested on the auditory information remembered from the video. They were not, however, told to explicitly remember the voices. Participants listened (or listened and watched in the face/video conditions) to all 6 voices sequentially. At the end of the presentation phase participants were given an auditory distracter task to prevent rehearsal of the voices. This took the form of a music quiz and required participants to listen to short clips of popular music songs, writing down artist and song title for each one on a provided answer sheet. The recognition test followed the distracter task immediately, and all participants were played an audio line-up containing some previously presented target voices and foil voices not previously heard. The 'old' previously presented voices used in the recognition phase were the same vocal recordings used in the target phase. For each voice in the line-up they were asked to make a yes/no judgement about whether they had heard it previously in the presentation phase and this was noted down by the experimenter.

3. Results

Overall participants in the study recalled 2.14 voices on average out of 4 presented and so were far from perfect in voice identification after a short delay. Target recognition was not at ceiling in the any of the conditions but was best in the voice only condition showing the expected Face Overshadowing Effect (see Figure 1). Participants in the voice only group recalled 2.83 voices out of 4, participants in the voice and face group only 1.83 voices and those in the voice and concealed face group 1.78 voices. Emotionality itself showed a small effect on recognition overall, 1.20 Angry voices were correctly recognised and 0.94 of Neutral out of 2.

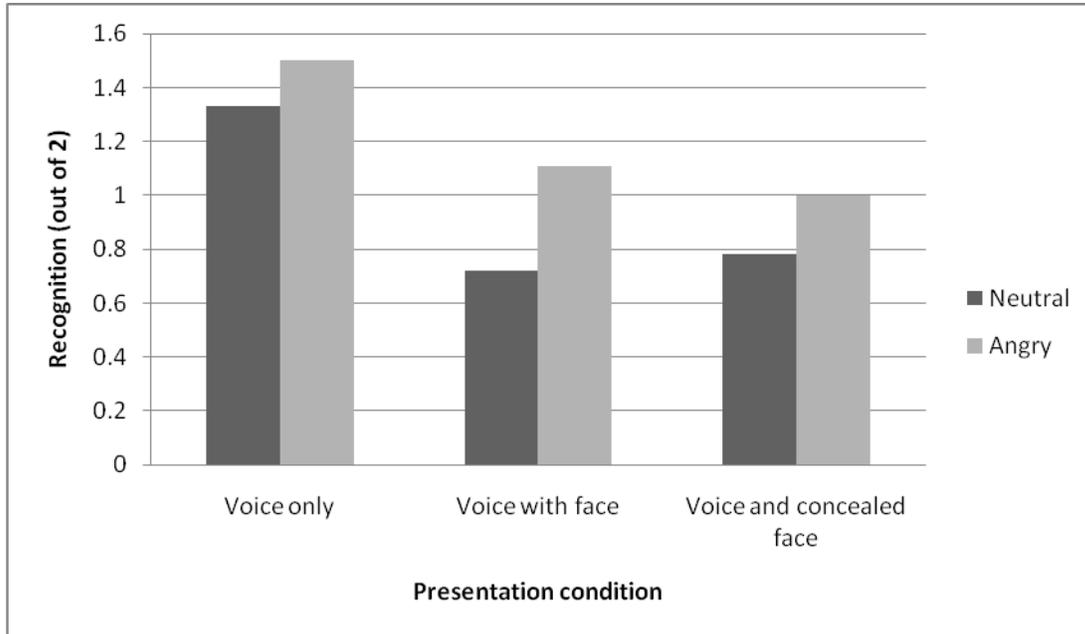


Figure 1. Mean recognition rates (out of 2) of angry and neutral voices in each condition

A mixed-factor ANOVA was used to analyse correct recognition scores. Though the range of possible scores was only 0-2 ANOVA has been justified in such instances where the degrees of freedom in the error term is above 40 [18], as was the case with the current data set. A 2 (emotion of voice: neutral, angry) X 3 (presentation condition: voice only, voice with face, voice with concealed face) mixed-factor ANOVA revealed no effect of emotionality of target voice, $F(1,51) = 3.87$, $p = 0.06$, η_p^2 (partial eta-squared) = .070, showing that neutral and angry voices were as likely to be correctly recognised overall. However, this test figure was very close to significance ($p = 0.055$), and the means indicated that angry voices ($M = 1.20$, $SD = 0.66$) were recognised at a slightly higher rate than voices of neutral tone ($M = 0.94$, $SD = 0.71$). The face overshadowing effect revealed itself in the significant effect of presentation condition, $F(2, 51) = 8.50$, $p = 0.001$, $\eta_p^2 = .250$; post hoc Bonferroni pairwise comparisons showed that the voice only condition had higher rates of correct recognition than both the face condition ($p = 0.003$) and the concealed face condition ($p = 0.002$) supporting predictions of a face overshadowing effect, but neither face nor concealed face conditions differed from one another ($p = 1.0$) which did not support the hypothesis that concealment would influence the face overshadowing effect. Finally, the interaction between emotionality and presentation condition did not reach significance, $F(2,51) = 0.26$, $p = 0.78$, $\eta_p^2 = .010$, and so the pattern of effects was consistent.

The yes/no recognition task yielded figures for false alarms, where participants incorrectly identified foils in the line-up as having been previously presented (see Figure 2). False alarm rates tended to be low when averaged across the sample ($M = 0.48$ out of 4). They were slightly higher with angry voices than with the neutral ones. A 2 (emotion of voice: neutral, angry) X 3 (presentation condition: voice only, voice with face, voice with concealed face) mixed-factor ANOVA revealed no effect of emotionality on false alarms, $F(1, 51) = 2.30$, $p = 0.14$, $\eta_p^2 = .043$, or effect of presentation condition, $F(2, 51) = 0.04$, $p = 0.96$, $\eta_p^2 = .002$, and also no interaction between the two factors, $F(2, 51) = .46$, $p = 0.63$, $\eta_p^2 = .018$.

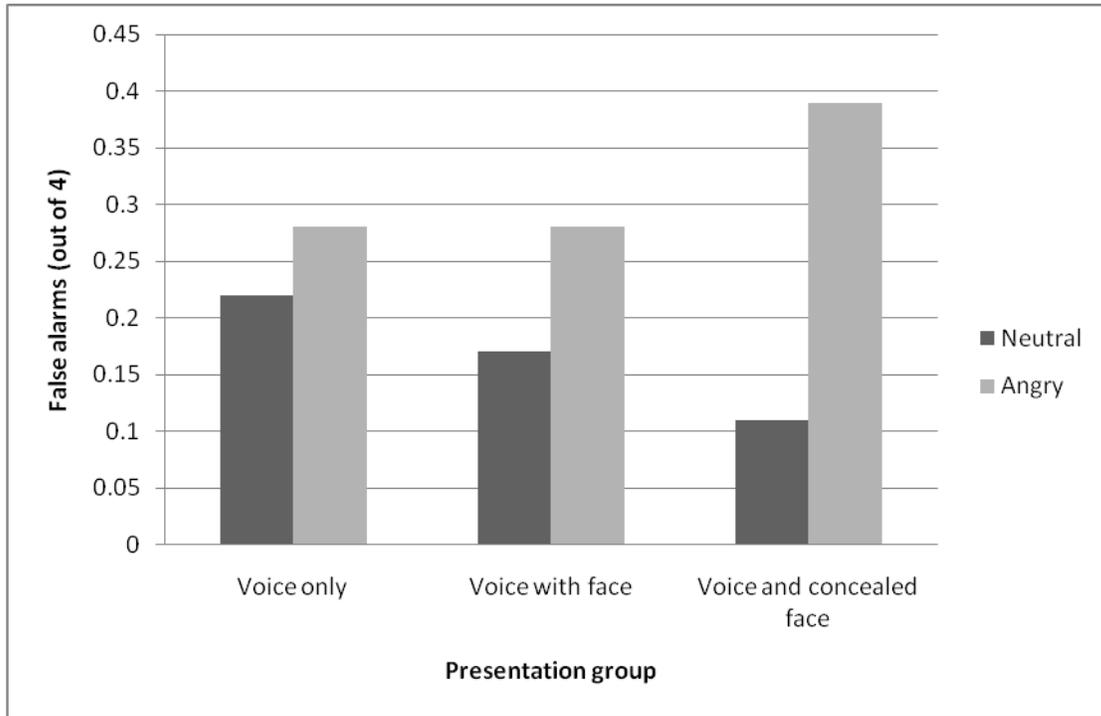


Figure 2. Mean false alarm rates (out of 4) of angry and neutral foil voices in each condition.

4. Discussion

A clear face overshadowing effect was observed in the current study, as predicted, replicating Cook and Wilding's studies showing that the presence of a face simultaneous to voice at encoding impairs later ability to recognise the voice. However, the magnitude of the effect had been expected to differ between the face group and the concealed face group and this was not the case, both resulted in a FOE and did not differ from one another, with poorer hit rates in these groups than the voice alone presentation group.

There might be several explanations for this finding. It may be that any visual stimulus present at encoding simultaneous to a voice impairs memory for the voice, as might be suggested by Yarmey's study that showed no increase in voice recognition with decreasing amounts of visual information [12]. This may be a premature conclusion and the first author has conducted other (unpublished) research with a blank disk covering the face that did not find a face overshadowing effect compared to an unconcealed face. The current stimuli, balaclava-concealed faces, still display some facial feature information; eyes and mouth are visible, as is perhaps something of the shape of the face and the spatial configuration of eyes and mouth. This may be enough face information to grab attention in the same way as an uncovered face. However, arguably we would still expect to see less of a FOE effect than in the unconcealed face condition given the impoverished feature information. Another explanation is that a face concealed by a balaclava may represent a level of personal threat that captures attention similar to the weapon focus effect in eyewitness memory [15],[16], thus reducing attentional resources deployed to the voice [13]. Indeed, angry faces have been found to be especially well remembered in short-term memory and it has been suggested that

memory for the identity of angry people might have a particular behavioural relevance enabling us to deal rapidly and effectively with potential threats in our environment [17]. Of course an alternative explanation may be that a face concealed with a balaclava simply represents a level of unusualness or bizarreness that is distracting in a way that covering the face with a blank disk is not. A future study directly needs to vary just how the face is concealed, using a variety of methods to elucidate this clearly, or perhaps use pre-exposure to the visual stimulus type or prior warnings.

In terms of vocal emotion, there was some evidence that angry voices were slightly better remembered than neutral ones (the effect was very close to significance) and there was a trend for the predicted better voice recognition to occur with the combination of voice only and angry vocal tone. This provides support for the notion that angry voice tone captures attention because it is threatening and therefore somewhat analogous to the weapon focus effect in eyewitness memory. It may be that humans have a bias to attend to angry people, and these findings confirm work that shows better short-term memory for angry faces over neutral and happy ones [17]. It is possible that, even though utterances were identical for neutral and angry voices, anger made the voices more distinctive in some way, perhaps through greater variation in prosodic features, such as more marked intonation, pitch rises, more vocal intensity and so on. There is existing evidence suggesting greater vowel variety in utterance results in better recognition of voices when multiple targets are present at encoding, presumably as the variation aids discrimination between the voices [19]. Whether it is threat or simply vocal artifacts specific to angry faces that are responsible for this effect can be determined in future work by varying these aspects specifically, matching the prosodic variation across speaker-groups. No objective measure of emotion was used in the current study and we cannot therefore be certain that the angry voices were indeed angry. However, as the results showed evidence of superiority in recognition for angry voices it does suggest these conveyed the intended emotion. Of course a future study should use independent ratings of emotion to ensure stimulus validity.

There are two additional limitations to note with the current stimuli set. Firstly, emotionality of facial expression was confounded with emotionality of vocal tone in the voice and face condition, as it follows that enunciating a word with anger requires the concomitant facial expression to be produced in the speaker and is visible in the video. However, as performance in the face group and concealed face group did not differ it does not suggest this had a bearing on results. Secondly, recognition line-ups were not constructed using an objective measure of fairness, and future work needs to ensure that all the foil voices are reasonably similar to the targets. Line-up fairness is perhaps harder to achieve with voices than with faces, but one way to do so would be to use the checklist of Handkins and Cross described by Yarmey [20] which comprises a set of scales to rate voices on characteristics such as pitch and pace. However, again, the current results showed that correct recognition was only around 50%, and given the short delay between exposure and testing this does not suggest that the targets were very distinctive in comparison to foils.

In the current study participants were not told specifically to remember voices, but were informed that they would be tested on the auditory information. Arguably participants might choose to focus on the *content* of the speech rather than the sound of the speaker's voice [21]. However, the use of fairly short utterances and the same phrase being repeated by all speakers made the content of the speech easy to remember and should arguably not have demanded particular attention to the speech content. In addition Cook and Wilding found no effect of explicit instructions to remember voice on the FOE [11], so this seems an unlikely explanation for current findings.

The relatively low rate of false alarms was notable in the current data-set, and perhaps not surprising under conditions of such short delay between presentation and test, and also due to the lack of a target-absent line-up [22], [23]. However, voice recognition was far from perfect overall and especially where the face or a concealed face was present, and this highlights the fallibility of earwitness recognition. The present study did use a high number of voices at presentation and test which arguably made the task more difficult and less forensically valid, and Cook and Wilding used only 2 targets (one male and one female) in their studies [24] compared to 8 males in the present study. However, the results do suggest memory for voices can be poor and alarmingly so after just short delays between exposure and recognition. Taken along with other published findings that demonstrate memory for speakers declines over time [3], [25], [26], this is cause for concern in criminal cases where voice evidence is presented to jurors who are biased to place unmerited confidence in witness testimony.

The current findings suggest earwitness evidence can be unreliable. Future work needs to determine with more precision just what circumstances lead to more or less accurate memory for voices and indeed explore post-event methods of presenting line-ups that facilitate accurate memory. Some work has already been done on the latter, but disappointingly cuing and context reinstatement methods have so far failed to show reliable improvements [10] though it may be that an established memory enhancing technique such as the cognitive interview [27] will improve voice memory, especially as aspects of the technique encourage witnesses to use multiple retrieval pathways by focusing attention on each sensory modality individually.

In conclusion, the current findings provide support for a face overshadowing effect first identified by Cook and Wilding, and suggest that memory for voices will be more reliable where the face of a perpetrator is not seen at all during exposure, as in telephone conversations or conditions of extremely poor light, than in situations where the face of a perpetrator is visible or concealed with a mask or disguise. The results further suggest that research on earwitness memory in the absence of visual information may not generalise to situations where visual information is also present, and, as McAllister and colleagues point out [8], arguably eyewitness and earwitness research should not be conducted in isolation from each other. Voice identification is clearly as fallible and prone to error as eyewitness identification, though the present results suggest that conditions where the voice is made implicitly more salient will result in better identification accuracy.

References

- [1] B.R. Clifford. Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior*, 4(4), 1980, 373-394.
- [2] D. Read, and F.I.M. Craik. Earwitness identification: some influences on voice recognition. *Journal of Experimental Psychology: Applied*, 1(1), 1995, 6-18.
- [3] F. McGehee. The reliability of the identification of the human voice. *The Journal of General Psychology*, 17, 1937, 249-271.
- [4] A.D. Yarmey, A.L. Yarmey, M.J. Yarmey, & L. Parliament.. Commonsense beliefs and the identification of familiar voices. *Applied Cognitive Psychology*, 15(3), 2001, 283-299.
- [5] B.R. Clifford. Memory for voices: the feasibility of quality of earwitness evidence. In S.A. Lloyd-Bostock & B.R. Clifford (Eds.), *Evaluating witness evidence* (pp.189-218). Chichester: Wiley, 1983.
- [6] B.E. Bell and E.F. Loftus. Trivial presentation in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology*, 56(5), 1989, 669-679.
- [7] H. Hollien, G. Bennett., and M.P. Gelfer. Criminal identification comparison: Aural versus visual identifications resulted from a simulated crime. *Journal of Forensic Science*, 28(1), 1983, 208-221.

- [8] H.A. McAllister, R.H. Dale, N.J. Bregman, A. McCabe and C.R. Cotton. When eyewitnesses are also earwitnesses: Effects on visual and voice identifications. *Basic and Applied Social Psychology*. 14(2), 1993, 161-170.
- [9] D. B. Wright and G. Wareham. Mixing sound and vision: The interaction of auditory and visual information for earwitnesses of a crime scene. *Legal and Criminological Psychology*. 10(1), 2005, 103-108.
- [10] S. Cook and J. Wilding. Earwitness Testimony 2: Voices, Faces and Context. *Applied Cognitive Psychology*. 11(6), 1997, 527-541.
- [11] S. Cook and J. Wilding. Earwitness testimony: Effects of exposure and attention on the Face Overshadowing Effect. *British Journal of Psychology*. 92(4), 2001, 617-629.
- [12] A.D. Yarmey. Verbal, Visual, and Voice Identification of a Rape Suspect Under Different Levels of Illumination. *Journal of Applied Psychology*, 71(3), 1986, 363-370.
- [13] S. Sinnett, C. Spence and S. Soto-Faraco. Visual dominance and attention: The Colavita effect revisited. *Perception and Psychophysics*. 69(5), 2007, 673-686.
- [14] E. Valenza, F. Simion, V.M. Cassia and C. Umiltà. Face preference at birth. *Journal of Experimental Psychology Human Perception and Performance*, 22(4), 1996, 892-903.
- [15] J.A. Easterbrook. The effect of emotion on cue utilization and organization of behavior. *Psychological Review*, 66(3), 1959, 183-201.
- [16] N. M. Steblay. A Meta-Analytic Review of the Weapon Focus Effect. *Law and Human Behaviour*. 16(4), 1992, 413-424.
- [17] M.C. Jackson, C.Y. Wu, D.E.J. Linden, J.E. Raymond. Enhanced Visual Short-Term Memory for Angry Faces. *Journal of Experimental Psychology: Human Perception and Performance*. 35(2), 2009, 363-374.
- [18] G.H. Lunney. Using analysis of variance with a dichotomous dependent variable: An empirical test study. *Journal of Educational Measurement*, 7(4), 1978, 263-269.
- [19] R. Roebuck and J. Wilding. Effects of vowel variety and sample length on identification of a speaker in a line-up. *Applied Cognitive Psychology*, 7(6), 1993, 475-481.
- [20] A.D. Yarmey. Descriptions of distinctive and non-distinctive voices over time. *Journal of the Forensic Science Society*, 31(4), 1991, 421-428.
- [21] H.A. Armstrong, and S.J. McKelvie. Effect of face context on recognition memory for voices. *The Journal of General Psychology*, 123(3), 1986, 259-270.
- [22] J.H. Kerstholt, N.J.M. Jansen, A.G. Van Amelsvoort and A.P.A. Broeders. Earwitnesses: effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology*, 18(3), 2004, 327-336.
- [23] A.D. Yarmey and E. Matthys. Voice identification of an abductor. *Applied Cognitive Psychology*, 6(5), 1992, 367-377.
- [24] S. Cook and J. Wilding. Earwitness Testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology*. 11(2), 1997, 95-111.
- [25] B.R. Clifford, H. Rathborn, and R. Bull. The effects of delay on voice recognition accuracy. *Law and Human Behavior*, 5(2/3), 1981, 201-208.
- [26] R. Hammersley and J.D. Read. The effect of participation in a conversation on recognition and identification of speakers' voices. *Law and Human Behavior*, 9(1), 1985, 71-81.
- [27] R.P. Fisher, R.E. Geiselman. *Memory enhancing techniques for investigative interviewing: the cognitive interview*. Springfield, IL: Charles C. Thomas, 1992.