Chatting in the Face of the Eyewitness: The Impact of Extraneous Cell-Phone Conversation on Memory for a Perpetrator

John E. Marsh[1,2], Krupali Patel[1], Katherine Labonté[3], Emma Threadgold[1], Faye C. Skelton[4], Cristina Fodarella[1], Rachel Thorley[1], Kirsty Battersby[1], Charlie D. Frowd[1], Linden J. Ball[1], & Francois Vachon[3]

[1] School of Psychology, University of Central Lancashire, Preston, UK

[2] Department of Building, Energy and Environmental Engineering, University of Gävle, Gävle, Sweden

[3] Université Laval, Québec, Canada

[4]Edinburgh Napier University, Edinburgh, UK

RUNNING HEAD: Impact of Cell-Phone Conversation on Memory

Correspondence: John E. Marsh, School of Psychology, Darwin Building, University of Central Lancashire, Preston, Lancashire, United Kingdom, PR1 2HE.

Phone (+44) 1772 893754, Fax (+44) 1772 892925

E-mail: JEMarsh@uclan.ac.uk

Chatting in the Face of the Eyewitness: Impact of Extraneous Cell-Phone Conversation on the Memory for a Perpetrator

Abstract

Cell-phone conversation is ubiquitous within public spaces. The current study investigates whether ignored cell-phone conversation impairs eyewitness memory for a perpetrator. Participants viewed a video of a staged-crime in the presence of one side of a comprehensible cell-phone conversation (meaningful halfalogue), two sides of a comprehensible cell-phone conversation (meaningful dialogue), one side of an incomprehensible cell-phone conversation (meaningless halfalogue) or quiet. Between 24 and 28 hours later participants freely described the perpetrator's face, constructed a single composite image of the perpetrator from memory, and attempted to identify the perpetrator from a sequential lineup. Further participants rated the likeness of the composites to the perpetrator. Face recall and lineup identification were impaired when participants witnessed the staged-crime in the presence of a meaningful halfalogue compared to a meaningless halfalogue, meaningful dialogue or quiet. Moreover, likeness ratings showed that the composites constructed after ignoring the meaningful halfalogue resembled the perpetrator less than those constructed after experiencing quiet, or ignoring a meaningfulness halfalogue or a meaningful dialogue. The unpredictability of the meaningful content of the halfalogue, rather than its acoustic unexpectedness, produces distraction. The results are novel in that they suggest that an everyday distraction, even when presented in a different modality to target information, can impair the long-term memory of an eyewitness.

Keywords: Distraction, cell-phones, eyewitness memory, dialogue, halfalogue.

Personal accounts and perceptions of how an event under investigation unfolds is a vital element in police investigations. Indeed, the apprehension of criminal suspects is often aided by descriptions of crimes and their perpetrators (Cutler & Kovera, 2010). Accounts provided from eyewitness memory offer valuable information that can contribute to the arrest and conviction of offenders (Samaha, 2005), especially in cases wherein the "hard evidence" needed for a conviction is lacking (Ainsworth, 2002). Eyewitness memory is therefore a domain in which accuracy is crucial and given its importance, investigations of the various factors that may moderate eyewitness error are vital. The auditory environment is just one component of a myriad of complex information that one may experience when witnessing an event such as a crime. Little is known, however, about the influence of the auditory scene on what is perceived or encoded from complex visual scenes that one would experience when witnessing a crime. In this study we investigate the potential impact of extraneous cell-phone conversations—an omnipresent facet of the auditory environment in public areas—on the capability of an eyewitness: (1) to recall detailed and accurate information about a perpetrator's face; (2) to construct a composite of accurate likeness to that face; and (3) to identify the perpetrator from a sequential lineup of visually similar identities.

Within modern society, engaging in cell-phone conversation is known to have adverse consequences on cognition, particularly in relation to driver accuracy (Strayer & Johnston, 2001) and pedestrian behavior (Stavrinos, Byington, & Schwebel, 2011). As a passive bystander, others' halfalogues (halves of conversations such as a cell-phone conversation whereby only one speaker can be heard) are rated as more noticeable and intrusive than dialogues (e.g., both sides of the conversation; Monk, Fellas, & Ley, 2004). Moreover, cognitive performance

can be differentially affected by halfalogues and dialogues. For example, Emberson, Lupyan, Goldstein, and Spivey (2010; see also Galvan, Vessal, & Golley, 2013) found that ignoring a halfalogue as compared with a dialogue produced disruption to performance on a visual monitoring (tracking) task and a choice reaction task. While the existing evidence suggests that overhearing half of a cell-phone conversation is enough to reduce performance on a concurrent, attentionally-demanding task, there has been no attempt to investigate the potential impact of ignoring cell-phone conversations on the recall of complex visual information in more applied tasks such as following the witnessing of a (staged) crime.

Typically, existing work on distraction via background sound has found impairment of short-term memory (STM) for sequences of visually-presented items (e.g., Hughes, Vachon, & Jones, 2005) but no study has shown impairment of long-term memory (LTM), when the sound is presented during the encoding of visual material. Certainly, from what is known about auditory distraction, it should be the case that background sounds that cause attention to be withdrawn from the prevailing task will impair encoding of visual events and therefore the later ability to recall those events from LTM.

One type of auditory distraction has been attributed to attentional diversion and occurs when the sound draws the attentional focus away from the prevailing mental activity (such as when an unexpected acoustic deviation is detected; for example, the “m” in the irrelevant sequence “k k k k k k m k k; Hughes, Vachon, & Jones, 2007). Another type of auditory distraction is attributable to interference-by-process (Jones & Tremblay, 2000). Essentially, performance impairment ensues

when there is a conflict between processes engaged to perform the focal task and processes applied involuntarily to the sound.

According to the attentional diversion standpoint, overhearing half of a conversation during study could impair encoding and therefore later recall from LTM at test because attention is directed involuntarily toward the sound due to a "need-to-listen." This “need-to-listen” is driven by the tendency to predict the semantic content of the inaudible half of the conversation (Monk et al., 2004; Norman & Bennett, 2014). Attentional diversion can also occur due to rudimentary processing of the acoustic features of the ignored speech (Hughes et al., 2007): the unexpected onset and offset of the voice within one side of a phone conversation could produce a violation of the expectancy of auditory events within the sound stream, causing a disengagement of attention away from the focal task and impoverished recall of visual events. This “attentional capture” produced by the unpredictable onsets and offsets of a cell-phone conversation would be synonymous with the finding that unexpected changes in the pattern of auditory stimulation (e.g., the “m” in the irrelevant sequence “k k k k k k m k k”) impairs STM for a sequence of visually-presented items (e.g., Hughes et al., 2005, 2007; Vachon, Hughes, & Jones, 2012). Therefore, both the “need-to-listen” and attentional capture accounts suggest that distraction is produced via attentional diversion.

On the interference-by-process view, only tasks that require retention of serial order information should be vulnerable to distraction via changing-state sound (sound sequences that demonstrate abrupt changes in their acoustic properties [e.g., “c t g u”] Beaman & Jones, 1997). However, in contrast with the distraction produced by interference-by-process, attentional diversion effects occur regardless of the task processes involved (Hughes et al., 2007; Vachon, Labonté, & Marsh,

2016). Therefore, if a half-conversation produces an attentional diversion effect, then disruption should manifest in complex cognitive tasks regardless of whether it involves serial STM. Witnessing and remembering an event is an example of such a task: witnesses encode complex visual and/or auditory information which must be maintained so that it may later be recalled. Any distraction during the event may prevent an eyewitness from encoding details that would later help to retrieve information from LTM, impacting negatively on their memory for event and person details.

Experiment

The current study's primary aim was to determine whether a to-be-ignored halfalogue negatively impacts on the LTM of an eyewitness to a staged-crime. Attention was manipulated during the encoding of the crime event. Participants witnessed a video of a staged-crime while told to ignore either a full conversation (meaningful dialogue) or a cell-phone conversation (meaningful halfalogue) in a language they spoke, a spectrally-rotated cell-phone conversation (incomprehensible to the participant and hence a "meaningless halfalogue") or no sound (quiet). Between 24 to 28 hours later, the same participants described the perpetrator's face from the staged-crime video in as much detail as possible and constructed a computer-generated likeness of the perpetrator (a composite). Finally, the participants were presented with a sequential lineup (cf. Steblay, Dysart, Fulero, & Lindsay, 2001) of nine static facial photographs that included the perpetrator and eight distractor faces that were similar to the perpetrator in overall visual appearance. For each facial photograph, the participants were required to rate on a scale of 1–7 how certain they were that the identity depicted was the person they witnessed in the staged-crime video they viewed the previous day. These tasks were

selected due to their ready use within police investigation (Frowd et al., 2013). Following this initial wave of experimentation, a set of independent judges rated the similarity of composites generated in each of the conditions (meaningful dialogue, meaningful halfalogue, meaningless halfalogue and quiet) to the perpetrator.

Given the demonstrable effect that unexpected auditory stimulation can have on simple attentional tasks (Emberson et al., 2010) regardless of the processes that underpin performance of the primary task (Hughes et al., 2007), it was expected that ignoring a halfalogue would result in greater distraction than ignoring a dialogue (and witnessing the staged-crime in quiet; e.g., Emberson et al., 2010). Within this setting, distraction could manifest via recall of fewer correct facial details about the perpetrator, impaired ability to identify the perpetrator from the sequential lineup, and the production of composites that bear weak resemblance to the perpetrator. Importantly, our inclusion of a meaningless halfalogue offered an opportunity to tease out whether any unique distraction produced by the halfalogue could be attributable to a "need-to-listen"—whereby the semantic properties of the task-irrelevant speech draws attention from the primary task (Monk et al., 2004; Norman & Bennett, 2014)—or to attentional capture—whereby an unexpected physical change in the auditory environment (such as the sudden onset of speech) is responsible for the withdrawal of attention from the focal task (e.g., Hughes et al., 2005, 2007).

Method

Participants. Ninety-six students at the University of Central Lancashire (71 females) aged between 20 and 31 years (M = 23.5; SD = 3.21) took part in the main empirical study. Participants were recruited via opportunity sample. All

participants spoke English as their first language and reported normal (or corrected-to-normal) vision and normal hearing. Twenty-four participants were allocated to each of the four sound conditions in the experiment. Nine participants did not return for the second part of the study and were replaced. A further twenty participants (14 females) aged between 21 and 37 years (M = 25.9; SD = 4.9) were recruited for the rating phase.

Apparatus and Materials. Four versions of the same video of a staged-crime were used which differed only with regards to the auditory background. This comprised quiet, a meaningful halfalogue (one side of a cell-phone conversation between two female speakers presented in the participants' native language) a meaningless halfalogue (the sound presented for the meaningless halfalogue but spectrally rotated to render it incomprehensible), and a meaningful dialogue (two sides of the same cell-phone conversation presented as meaningful halfalogue). The same cell-phone conversation was therefore used for both the meaningful halfalogue and the meaningful dialogue conditions with the former being created by deleting one of the speaker's voices. In the halfalogue version, there were nine pauses that ranged between 1.4 and 7.7 seconds (M = 3.14, SD = 2.08). The video and the cell-phone conversation lasted for one minute and the onset of this conversation coincided with the onset of the video. The video depicted a man in his early twenties entering a corner shop, attempting to steal money from an unoccupied cash register—which could not be forced open—before making good his escape with several packets of cigarettes.

The topic of the phone conversation was based around a BBC news article about the nation's favorite children's book and was digitally recorded and sampled with a 16-bit resolution at a sampling rate of 44.1 kHz using a broadcast quality

Dictaphone in an anechoic chamber. Halfalogues were created by silencing the voice of one of the speakers within the auditory file. The spectrally-rotated halfalogue was created by spectrally inverting the speech recording around 2 kHz (as in Scott, Rosen, Beaman, Davis, & Wise, 2009). Spectrally rotating speech involves transforming the high-frequency energy into low-frequency energy and vice versa. Spectrally-rotated speech is almost identical to normal speech (Scott et al., 2009). For example, variations in sound pressure level across time and the duration of pauses between words and sentences are fairly equal. However, rotated speech is meaningless because it is incomprehensible.

The four versions of the same video (with different audio backgrounds) were created by embedding the audio onto the video using Windows Live Movie Maker. Both normal speech and rotated speech were presented over stereo headphones at approximately 69 dB (LAeq) as measured with an artificial ear.

The computer program PRO-fit (Version 3.5) was used to generate the facial composites. PRO-fit is a feature-based system that involves presenting the witness with facial features (e.g., hair, eyes, nose, mouth, etc.) that match the face that the witness has previously described (for an overview, see Frowd et al., 2014). This stage is described in more detail below.

Procedure. In the first session participants viewed a staged-crime video in the context of one of the four sound conditions that they were randomly allocated to with equal sampling. They were seated at a distance of approximately 60 cm from the PC monitor in a testing cubicle and wore headphones. They were instructed to ignore any background sound, that they would not be asked anything about the sounds during the experiment, and to focus on studying the video. Participants were

asked to return between 24-28 hours later but the nature of the second visit was not revealed at this time.

In the second session, it was revealed that a composite of the perpetrator witnessed in the staged-crime video would be required. It was explained to participants that the goal of creating the composite was to produce an accurate portrayal of the perpetrator's face so that another person could recognize the face as such. Participants were told that they would first describe the appearance of the face and then construct a composite of it. They were also told that there was no time limit to complete the face composite construction procedure. The procedure for undertaking the face-recall interview and PRO-fit construction is detailed and the reader is referred to existing articles for specific details (e.g., Frowd et al., 2013). In brief, participants were asked to think back to the time when the perpetrator had been seen, visualize the face and then to try to recall as much detail about it as possible, without guessing. The experimenter wrote down information that the participants recalled in relation to the face in this free-recall format. Participants were then informed that a composite would be constructed of the face using PRO-fit. The experimenter entered details from the face-recall phase into the description details of PRO-fit. This generated the different features for the described face. If participants were not satisfied with a feature then its size or location was adjusted or it was exchanged for another feature. Once participants reported that the best likeness had been achieved the face was saved to disk as the composite.

Following completion of the composite, participants undertook the sequential lineup task. They were given a sequential presentation of facial photographs of nine identities that comprised the target (perpetrator) and eight foils that resembled the target in overall appearance. Using a 7-point Likert scale

whereby ‘1’ represented guess and ‘7’ certain, participants were asked to indicate the certainty with which they considered that each facial photograph was the same identity as the person they witnessed in the staged-crime video they had viewed. The order in which the facial photographs were presented was pseudo-random: While the foils were presented in a random order for each participant, the target was either presented in Position 4 or 5 within the sequence. Participants were reminded that there was no time limit to complete the sequential lineup task. The time taken to complete the face composite construction and sequential lineup task varied between 25 to 45 minutes.

Once all of the composites had been constructed, further participants were asked to rate the likeness of each of the composites compared to a frontal shot of the target (perpetrator) using a 7 point Likert scale (1 = ’very-poor likeness‘ and 7 = ’very-good likeness‘). Participants provided ratings for 96 composites (the 24 composites generated from within each sound condition). Composites were presented individually, each one next to the photograph of the target on a page in an A4 booklet. The presentation order of the composites was random for each participant.

Design. The main empirical study (as compared to the composite rating task) employed a between-subjects design whereby the independent variable was "Sound Condition" with four levels: quiet, meaningless halfalogue, meaningful halfalogue and meaningful dialogue. For the face-recall part of the study (usually undertaken as part of a cognitive interview), the dependent variable was "Facial Descriptor Type" which had three levels: Correct details, incorrect details, and subjective details; see later). For the sequential lineup component of the task, the independent variable was "Identity" and had two levels: target (i.e., perpetrator) or

foil, and the dependent variable was the confidence rating given to the target face and the mean rating given to the eight foils (collapsed). Finally, for the set of participants that independently rated the similarity of the composites to the target, the design was fully repeated measures, whereby the within-participant factor was Sound Condition (again quiet, meaningless halfalogue, meaningful halfalogue and meaningful dialogue) and the dependent variable was the similarity of each composite to the target rated on a scale of 1-7 (described above).

Results

Verbal Recall. The quality of the face descriptions given by the participants within each Sound Condition was analyzed by two individuals. Following the procedure used by Meissner, Brigham, and Kelley (2001) a correct description was generated by the two raters for the perpetrator's face and a decision was reached between the two raters as to which details would be classed as correct. Details in the descriptions were coded as correct, incorrect, or subjective. Subjective details were those that could not be verified directly (e.g., inferences about personality, or similarity to a well-known celebrity or family member). Inter-rater agreement was high [Cohen's K (72) = .87, $p < .001$; Cohen, 1988]. Contradictory scorings were resolved through discussion. The mean number of correct and incorrect features listed per condition can be seen in Figure 1. The mean number of correct descriptors provided was lower in the Meaningful Halfalogue condition as compared to the Meaningless Halfalogue, Meaningful Dialogue and Quiet conditions. No difference between means was apparent for incorrect descriptors. Only five details were classified as subjective descriptors across all four conditions and because of this, subjective descriptors were excluded in the further analysis.

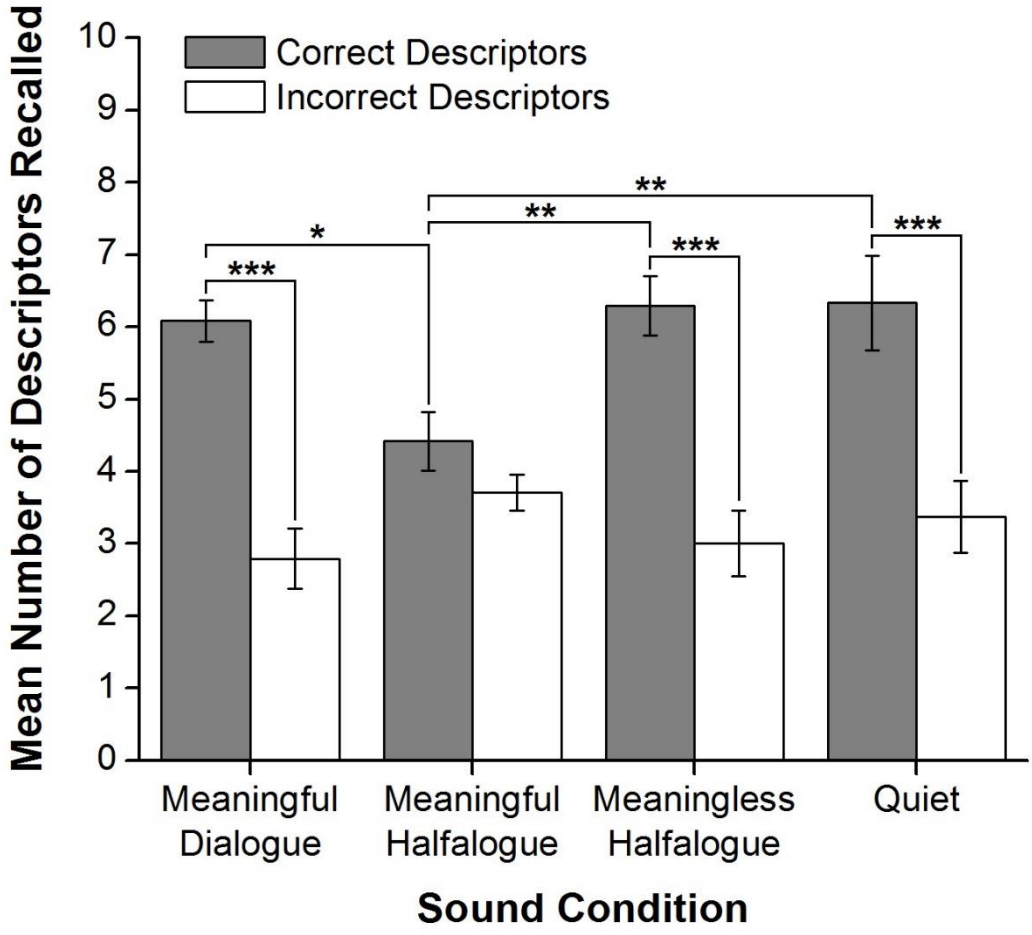

Figure 1. Mean number of face descriptors recalled as a function of descriptor type and sound condition. Error bars represent the standard error of the mean.

A 4 (Sound Condition: meaningful dialogue *vs*. meaningful halfalogue *vs*. meaningless halfalogue *vs*. quiet) × 2 (Facial Descriptor Type: correct response *vs*. incorrect response) mixed factor Analysis of Variance (ANOVA) carried out on the mean number of face descriptors recalled revealed a main effect of Facial Descriptor Type, $F(1, 92) = 47.70$, $MSE = 6.61$, $p < .001$ with more correct than incorrect descriptors recalled, $\eta_p^2 = .34$, but no main effect of Sound Condition, $F(3, 92) = 2.09$, $MSE = 2.62$, $p = .11$, $\eta_p^2 = .06$. The interaction between Facial Descriptor Type and Sound Condition was significant, $F(3, 92) = 2.80$, $MSE = 6.61$,

p = .043, η_p^2 = .084. A simple effects analysis (LSD) revealed that correct facial descriptors were more frequent than incorrect facial descriptors for the quiet condition (p < .001), meaningful dialogue condition (p < .001) and meaningless halfalogue condition (p < .001), but not for the meaningful halfalogue condition (p = .35). Moreover, correct descriptors were less frequent in the meaningful halfalogue condition as compared with the quiet condition (p = .004), meaningful dialogue condition (p = .012) and the meaningless halfalogue condition (p = .005). There was no difference between the means for the quiet and meaningless halfalogue conditions (p = .95), quiet and meaningful dialogue conditions (p = .70), and meaningless halfalogue and meaningful dialogue conditions (p = .75). Moreover, there was no difference between conditions with respect to the frequency of incorrect information provided (p > .1, all comparisons). Therefore, a to-be-ignored halfalogue, provided it is meaningful, presented during the witnessing of the staged-crime video diminished the quality of face description given the next day.

Sequential Lineup Task. For the lineup task, the ratings reflecting the certainty that the identity was the same as the target in the video previously were addressed by comparing the mean rating given to the foil faces with the rating given to the target. Figure 2 shows the mean certainty ratings for the foil identities (collapsed across identities) and the target for each of the four sound conditions. The confidence ratings were clearly greater for the target in the quiet, meaningful dialogue and meaningless halfalogue conditions as compared to the meaningful halfalogue condition. However, confidence ratings assigned to foil identities appears to differ little between conditions.

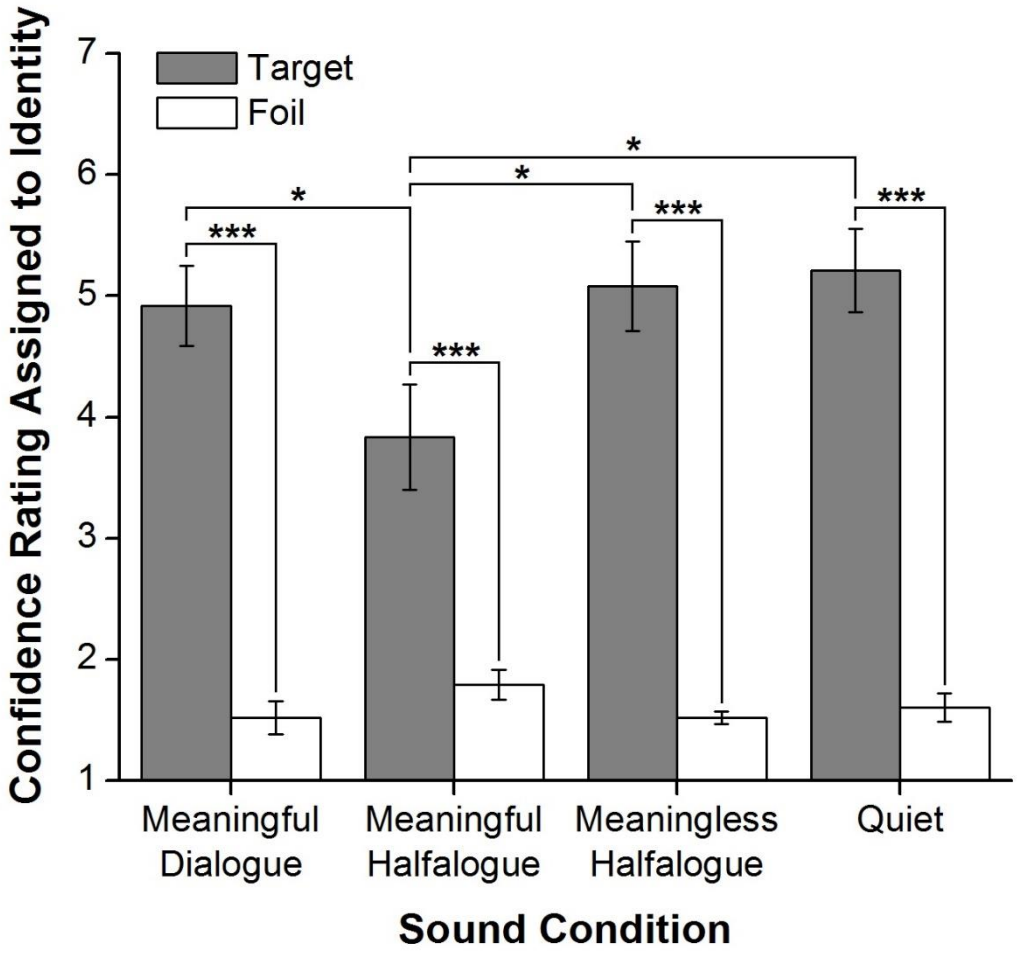

Figure 2. Mean confidence ratings as a function of sound condition in the context of the lineup task. These relate to whether the Target or one of the Foils was viewed earlier in the context of the mock crime video. The mean ratings given to the eight foils are collapsed. 1 = guess and 7 = certain that the identity was seen earlier. Note therefore that a rating of 7 given to the target would essentially be a "hit", whereas a rating of 1 given to the target would be a "miss". Similarly, a rating of 1 given to a foil would be a "correct rejection", whereas a rating of 7 to a foil would be a "false alarm". Error bars represent the standard error of the mean.

A 4 (Sound Condition) × 2 (Identity: target or foil) mixed-factorial ANOVA performed on mean confidence ratings revealed a main effect of Identity with higher confidence ratings for the target than for foils, $F(1, 92) = 250.12$, $MSE = 1.91$, $p < .001$, $\eta_p^2 = .73$, but no main effect of Sound Condition, $F(3, 92) = 1.90$, $MSE = 1.70$, $p = .14$, $\eta_p^2 = .06$. However, there was a significant interaction between Sound Condition and Identity, $F(3, 92) = 3.50$, $MSE = 1.91$, $p = .019$, $\eta_p^2 = .10$. A simple

effects analysis (LSD) revealed that the mean confidence rating given to the target was lower in the meaningful halfalogue condition compared to the quiet condition (p = .010), the meaningful dialogue condition (p = .042), and the meaningfulness halfalogue condition (p = .019). There was no significant difference between the quiet and meaningful dialogue conditions (p = .58), quiet and meaningless halfalogue conditions (p = .81) or between the meaningful dialogue condition and the meaningless halfalogue condition (p = .75). Therefore, a meaningful to-be-ignored halfalogue presented concurrently with the mock-crime video reduced the confidence with which the target is chosen from a lineup the next day.

Composite Likeness Ratings. Figure 3 shows the means for the likeness scores given by the raters for the 24 composites within each of the four sound conditions. The mean values indicate that the raters considered that the composites generated in the quiet, meaningful dialogue condition and meaningless halfalogue conditions looked more similar to the target face than the composites generated in the meaningful halfalogue condition.

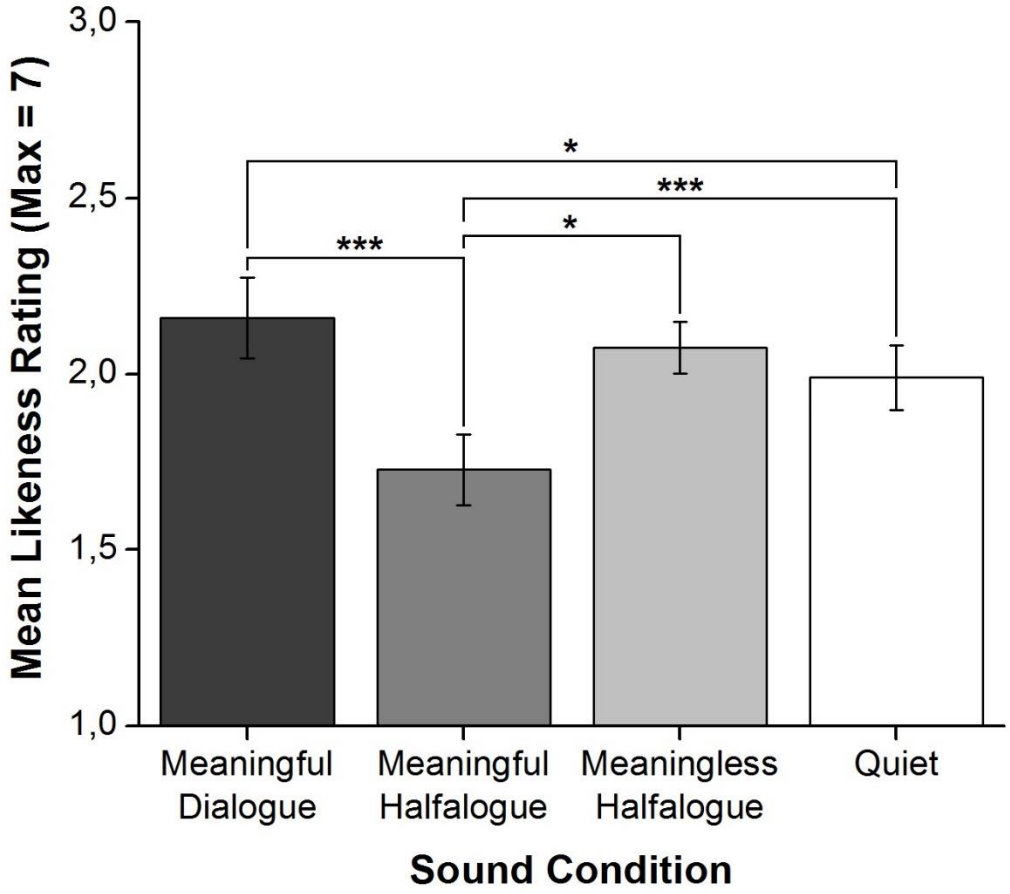

Figure 3. Mean likeness ratings awarded to the composites in the presence of a photograph of the target as a function of sound condition. 1 = very poor likeness, 7 = very good likeness. Error bars represent the standard error of the mean.

A one-way repeated-measures ANOVA demonstrated a significant effect of Sound Condition on composite likeness, $F(3, 57) = 5.31$, $MSE = .132$, $p = .003$, $\eta_p^2 = .22$. Planned repeated contrasts revealed that composites in the meaningful halfalogue condition bore less resemblance to the perpetrator than those for the quiet condition ($p = .001$), meaningless halfalogue condition ($p = .029$), and meaningful dialogue condition ($p < .001$). Additionally, those created in the meaningful dialogue condition were rated as better likenesses of the target face than those created in the quiet condition ($p = .023$, no other comparisons were significant). Therefore, a meaningful to-be-ignored halfalogue presented concurrently with the mock-crime video resulted in facial composites that were

rated poorer likenesses to the target. Figure 4 show examples of the male target constructed in each of the sound conditions.

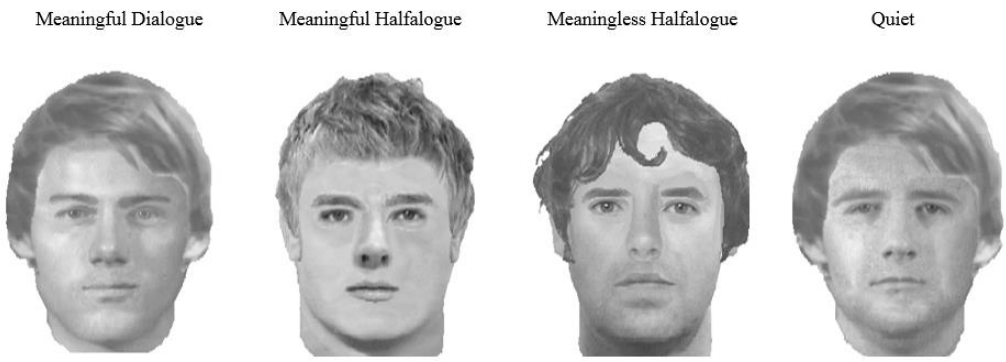

Figure 4. Examples of the male target constructed in the four conditions of the experiment (displayed are those composites that have received the highest ratings for each sound condition). For reasons of copyright, we are unable to reproduce the target photograph or stills from the video used in the experiment.

Discussion

To summarize, ignoring half of a cell-phone conversation, providing it is meaningful, was shown to impair the LTM of the participant eyewitnesses. That the accuracy of eyewitness LTM—as measured through recall of facial descriptors, identification from a lineup and composite accuracy—is susceptible to disruption via the presence of intermittent conversational background speech is important to acknowledge given the prominent role that eyewitnesses play in many criminal cases. Composite images serve two purposes. On presentation within the media, they can generate leads from the general public to aid criminal investigations. They are also used as a reference from which criminal investigators can narrow likely suspects that may already be on file. Therefore, inaccuracies with eyewitness memory—and subsequent composite quality—can potentially lead to false identifications (and arrests) and the pursuit of erroneous leads.

It is emerging that extraneous background speech can impair face memory in several ways. One way, for example, is through disruption of subvocal verbalization. It has become reasonably well accepted that spontaneous verbal codes are created for faces (Schooler, 2002). Indirect evidence that participants verbally rehearse descriptions of faces within STM, and that such rehearsal ordinarily facilitates face recognition performance, comes from studies preventing subvocal verbalization by the use of articulatory suppression: a technique that requires participants to utter some repeated sounds (e.g., ba ba ba ba). Articulatory suppression impairs face recognition (Nakabayashi & Burton, 2008, Experiment 1; Nakabayashi, Burton, Brandimonte, & Lloyd-Jones, 2012; Wickham & Swift, 2006), whereas manual tapping—assumed to be as attentionally demanding as articulatory suppression without preventing verbalization—does not (e.g., Nakabayashi & Burton, 2008, Experiment 3; Wickham & Swift, 2006). While articulatory suppression potentially eliminates the use of subvocal rehearsal, extraneous changing-state speech (sound sequences that are acoustically changing [e.g., "c t g u"] as compared to unchanging, steady-state speech [e.g., "c c c c"]) disrupts subvocal rehearsal due to processing conflict (see Jones et al., 1992). Consistent with the view that changing-state speech disrupts subvocal rehearsal, and that subvocal rehearsal is used spontaneously to facilitate unfamiliar face learning, Marsh et al. (2016) have found that extraneous changing-state speech (randomly presented strings of letters), as compared to steady-state speech (a string of the same letter repeated), presented during a 6-s exposure to a target face impairs recognition of that face from a lineup. However, that such interference is entirely independent of the semantic content of the speech, suggests that the disruption is consistent with an interference-by-process view of distraction (Jones et al., 1992).

Here, the preattentive processing of the serial order of changes within sound interferes with the similar, deliberate process of subvocally rehearsing information derived from the visual modality in serial order.

In the context of the current study however, we favor an attentional diversion account (Hughes et al., 2007; Monk et al., 2004) over the disruption of subvocal rehearsal account for three reasons. First, participants did not know in advance that face recall, composite construction and lineup identification would be required subsequently. Therefore, the participants may not have rehearsed facial details explicitly. Second, perhaps counterintuitively, the subvocal rehearsal process appears to utilize configural as opposed to featural information (Nakabayashi, Lloyd-Jones, Butcher, & Liu, 2012) which, according to Schooler (2002), involves information concerning the face's global percept including the spatial layout amongst its facial features. If disruption of subvocal rehearsal was the cause of face memory impairment then it would appear quite counterintuitive that PRO-fit, a feature-based system (due to its requirement for recall of individual, isolated features, and recognition of features in the context of the whole-face), could capture the distraction effect. Third, since to-be-ignored meaningful dialogue speech—which presumably contains sufficient changing-state information to disrupt serial rehearsal (Jones et al., 1992; and, in fact, more change than within halfalogues)—failed to produce disruption, it is unlikely that the action of the meaningful to-be-ignored halfalogue speech is attributable to the disruption of subvocal rehearsal.

Moreover, in the context of attentional diversion accounts (e.g., Monk et al., 2004; Hughes et al., 2007) the results of the experiment were unequivocal in providing support for the “need-to-listen” account of the halfalogue effect (Monk

et al., 2004; Norman & Bennett, 2014) over an attentional capture account (cf. Hughes et al., 2005, 2007). The halfalogue effect only appeared when the background speech material was meaningful. Since both the meaningful and meaningless (rotated) halfalogue speech were equated in terms of their acoustic complexity and temporal unpredictability, that only the meaningful halfalogue produces impairment refutes the idea that the halfalogue produces disruption due to the acoustic unexpectedness (and hence attentional capture) attributable to the physical characteristics of sound (cf. Hughes et al., 2005). That the halfalogue effect is dependent upon the presence of semantic properties within the sound demonstrates that it is a form of distraction that differs from that attributable to acoustic unexpectedness (Hughes et al., 2005, 2007; Vachon et al., 2012). In the context of the current study, it appears that the meaningful halfalogue produces attentional diversion whereby the "need-to-listen" engendered by the tendency to want to predict/complete the missing part of the conversation causes an impoverished encoding of details about the perpetrator, thereby impairing face recall and recognition. While the task of face description, face construction and target identification from a lineup are usually carried out in this sequence in the real world, it is possible that these tasks may influence each other. For example, describing the target could have influenced the composite construction, and the composite construction may have influenced target identification in the lineup. Therefore, impoverished memory for the target produced by the meaningful halfalogue could have knock on effects at several loci within the procedures undertaken with the eyewitness.

While it is perhaps intuitive that masking or otherwise interfering effects of additional environmental sounds such as voices may impede recognition and recall

of a perpetrator's voice (cf. Stevenage, Neil, Barlow, Dyson, Eaton-Brown, & Parsons, 2013), it is perhaps less intuitive that stimulation from a specific modality (auditory) should impair processing of information that is derived from another (visual). However, the present findings unequivocally demonstrate that cell-phone conversation (meaningful halfalogue) breaks through selective attention and impairs LTM even if participants know that the sounds contain no information that is relevant to the prevailing task (cf. Marsh, Demaine, et al., 2016), and therefore should be ignored.

To our knowledge the current results are novel in demonstrating that extraneous speech presented during encoding, can produce adverse effects on LTM for complex visual information: the appearance of a human face. Therefore, the findings illustrate the importance of considering the auditory environment when assessing the reliability of eyewitness memory. Moreover, these findings have implications far beyond the forensic context. Exposure to half of a conversation is a common occurrence, and can impact negatively on our memory for complex visual information. Our results show that this irrelevant auditory information cannot simply be ignored, and as such has the potential to interfere with our processing of information in a wide range of daily activities.

References

Ainsworth, P. B. (2002). *Psychology and policing*. Portland, Oregon: Willan Publishing.

Beaman, C. P., & Jones, D. M. (1997). The role of serial order in the irrelevant speech effect: Tests of the changing state hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 459–471.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cutler, B. L., & Kovera, M. B. (2010). *Evaluating eyewitness identification*. New York: Oxford University Press.

Emberson, L. L., Lupyan, G., Goldstein, M. H. & Spivey, M. J. (2010). Overheard cell-phone conversations: When less speech is more distracting. *Psychological Science*, *21*, 1383-1388.

Frowd, C. D, Jones, S., Fodarella, C., Skelton, F. C., Fields, S., Williams, A., Marsh, J. E., Thorley, R., Nelson, L., Greenwood, L., Date, L., Kearley, K., McIntyre, A. H., & Hancock, P. J. B. (2014). Configural and featural information in facial-composite images. *Science & Justice*, *54*, 215-227.

Frowd, C. D., Skelton F. C., Hepton, G., Holden, L., Minahil, S., Pitchford, M., McIntyre, A., Brown, C., & Hancock, P. J. B. (2013). Whole-face procedures for recovering facial images from memory. *Science & Justice*, *53*, 89-97.

Galvan, V. V., Vessal, R. S., & Golley, M. T. (2013). The effects of cell phone conversations on the attention and memory of bystanders. *PLOS ONE*, *8*, 1-10.

Hughes, R. W., & Jones, D. M. (2005). The impact of order incongruence between a task-irrelevant auditory sequence and a task-relevant visual sequence. *Journal of Experimental Psychology: Human Perception & Performance, 31*, 316–327.

Hughes, R. W., Vachon, F., & Jones, D. M. (2005). Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? *Journal of Experimental Psychology: Learning, Memory & Cognition*. *31*, 736-749.

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 33*, 1050-1061.

Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology, 44*, 645-669.

Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review, 7*, 550-558.

Marsh, J. E., Demaine, J., Bell, R., Skelton, F. C., Frowd, C. D., Röer, J. P., & Buchner, A. (2015). The impact of irrelevant auditory facial descriptions on memory for target faces: Implications for eyewitness memory. *The Journal of Forensic Practice, 17*, 271-280.

Marsh, J. E., Skelton, F. C., Vachon, F., Thorley, R., & Frowd, C. D., & Ball, L. J. (2016, in preparation). *In the face of distraction: Irrelevant speech impairs person identification*.

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval

processes in verbal overshadowing. *Memory & Cognition*, *29*, 176–186.

Monk, A., Fellas, E., & Ley, E. (2004). Hearing only one side of normal and mobile phone conversations. *Behaviour and Information Technology, 23*, 301-305.

Nakabayashi, K., & Burton, M. (2008). The role of verbal processing at different stages of recognition memory for faces. *European Journal of Cognitive Psychology: a special issue of verbalizing visual memories, 20*, 478-496.

Nakabayashi, K., Burton, A. M., Brandimonte, M. A. & Lloyd-Jones, T. J. (2012). Dissociating positive and negative influences of verbal processing on the recognition of pictures of faces and objects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*, 376-390.

Nakabayashi, K., Lloyd-Jones, T. J., Butcher, N. & Liu, C. H. (2012). Independent influences of verbalization and race on the configural and featural processing of faces: a behavioral and eye movement study. *Journal of Experimental Psychology: Learning Memory & Cognition*, *38*, 61-77.

Norman, B., & Bennett, D. (2014). Are mobile phone conversations always so annoying? The ‘need-to-listen' effect revisited. *Behavior & Information Technology, 33*, 1294-1305.

Samaha, J. (2005). *Criminal Justice*. Belmont, CA: Wadsworth.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology*, *16*, 989-997.

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P and Wise, R. J. S. (2009). The neural processing of masked speech: Evidence for different mechanisms in the right and left temporal lobes. *Journal of Acoustic Society America, 125*, 1737-1743.

Stavrinos, D., Byington, K. W., Schwebel, D. C. (2011). Distracted walking: cell

phones increase injury risk for college pedestrians. *Journal of Safety Research, 42,* 101-7.

Steblay, N. M., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law & Human Behavior, 25*, 459-474.

Stevenage, S. V., Neil, G. J., Barlow, J., Dyson, A., Eaton-Brown, C., & Parsons, B. (2013). The effect of distraction on face and voice recognition. *Psychological Research, 77*, 167-175.

Strayer, D. L., & Johnson, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*, 462-466.

Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*, 164-177.

Vachon, F., Labonté, K., & Marsh, J. E. (2016, manuscript in press). Attentional capture by deviant sounds: A non-contingent form of auditory distraction? *Journal of Experimental Psychology: Learning, Memory, & Cognition.*

Wickham, L. H. V. & Swift, H. (2006). Articulatory suppression attenuates the verbal overshadowing effect: A role for verbal encoding in face identification. *Applied Cognitive Psychology, 20*, 157-169.

Author note

John E. Marsh, Krupali Patel, Emma Threadgold, Cristina Fodarella, Rachel Thorley, Kirsty Battersby, Charlie D. Frowd and Linden J. Ball, University of Central Lancashire, Preston, UK. Katherine Labonté and Francois Vachon, Université Laval, Québec, Canada. Faye C. Skelton, Edinburgh Napier University, Edinburgh, UK. John E. Marsh is also at the Department of Building, Energy and Environmental Engineering, University of Gävle, Gävle, Sweden. The research reported in this paper was financially supported by a British Academy grant (SG122309) awarded to John E. Marsh, Faye C. Skelton and Charlie D. Frowd.