Abstract

Purpose To investigate the potential susceptibility of eyewitness memory to the presence of extraneous background speech that comprises a description consistent with, or at odds with, a target face.

Design/methodology/approach A between-participants design was deployed whereby participants viewed an unfamiliar target face either in the presence of quiet, extraneous to-be-ignored speech that comprised a verbal description that was congruent with the target face or to-be-ignored speech that comprised a verbal description that was incongruent with that face. After a short distractor task, participants were asked to describe the target face and to construct a composite of the face using PRO-fit software. Further participants rated the likeness of the composites to the target.

Findings Recall of correct facial descriptors was facilitated by congruent to-be-ignored speech and inhibited by incongruent to-be-ignored speech as compared to quiet. Moreover, incorrect facial descriptors were reported more often in the incongruent speech condition as compared with the congruent speech and quiet conditions. Composites constructed after exposure to incongruent speech were rated as poorer likenesses to the target than those created after exposure to congruent speech and quiet. Whether congruent speech facilitated or impaired composite construction was found to depend on the distinctiveness of the target face.

Practical implications The results suggest that the nature of to-be-ignored background speech has powerful effects on the accuracy of information that is verbally reported from having witnessed a face. Incongruent speech appears to disrupt the recognition processes that underpin face construction while congruent speech may have facilitative or detrimental effects on this process, depending on the distinctiveness of the target face.

Originality/Value This is one of the first studies to demonstrate that extraneous speech can produce adverse effects on the recall and recognition of complex visual information: in this case the appearance of a human face.

Keywords Eyewitness memory, auditory distraction, composites

Paper type Empirical paper

To avoid processing sensory information within the visual modality we can simply close our eyes or avert our gaze. Within the auditory modality, there is no easy way to achieve this feat: even though one may not be attending to the auditory world around us, extraneous background sound is processed obligatorily (Hughes & Jones, 2001). Therefore, despite one's best efforts to ignore it, it is inevitable that background sound will sometimes disrupt cognitive processes even when it is irrelevant to the task at hand. Although much is known about the impact of background sound on visual-verbal serial short-term memory (Hughes & Jones, 2001), little is known about whether, and how, background sound impairs memory for complex visual information (cf. Wais & Gazzaley, 2011) especially within applied settings (Hodgetts, Vachon, & Tremblay, 2014). This appears remiss especially in forensic settings since the auditory environment, being one component of multiple sources of environmental information, could potentially have an influence on the accuracy and therefore reliability of information recalled about a crime and its perpetrator(s) (Marsh et al., 2014). Following a crime, the accuracy of eyewitness memory can be pivotal to the effectiveness of the ensuing criminal investigation. Eyewitness accounts can be used to create new leads which can result in the arrest and conviction of the suspect(s) (Loftus, 1975; Samaha, 2005). The current study investigates whether the accompaniment of extraneous background speech influences participants' memory for a target as measured through the accuracy of the free recall of facial descriptors, and the likeness of the composite constructed, in relation to the target.

Eyewitness testimony can be influential when it comes to the administration of justice. If this evidence is misinformed or inaccurate then severe consequences can occur for the trial outcome. Coupled with other forensic evidence in court, testimonies based on eyewitness memories are an important part of a criminal investigation (Kebbell & Milne, 1998). Therefore, it is imperative that eyewitness testimonies are as precise as possible. Many studies, however, demonstrate that human memory can be far from accurate (e.g., Koriat & Goldsmith, 1994). For example, eyewitness memory has been shown to be impaired in situations in which a witness verbally describes the suspect's appearance before attempting to select them from a visually-based lineup: the so-called verbal overshadowing effect (Schooler & Engstler-Schooler, 1990; see also Vanags, Carroll, & Perfect, 2005). According to Wickham and Swift (2006), verbal overshadowing occurs because the verbal description interferes with the verbal description spontaneously produced for an unfamiliar target face during study (descriptions of faces that are henceforth termed "verbal codes"). That interference between verbal codes impacts on later recognition of a face has important

implications for the susceptibility of face memory to impairment by *concurrent* distraction by background speech.

Verbal Coding and Face Learning

It is well accepted that participants spontaneously create verbal descriptions of unfamiliar faces (e.g., “pointy-nose”, “thin lips”; Schooler, 2002). This may be used in an attempt to facilitate visual discriminability of faces that are all highly similar to one another: verbalization may direct processing toward information that can differentiate a face from others that have been encountered (Nakabayashi, Burton, Brandimonte, & Lloyd-Jones, 2012). Evidence suggesting that verbalization is required for face learning has been gleaned from studies that have used articulatory suppression. This requires that participants utter a repeated sound or word (e.g., “the”) continuously during face learning, thus preventing the participant from using subvocalization. Several studies (e.g., Nakabayashi & Burton, 2008; Wickham & Swift, 2006) have shown that articulatory suppression impairs face recognition. For example, Wickham and Swift (2006) demonstrated that preventing subvocalization throughout the 5 seconds an unfamiliar target face was displayed later impaired the ability to identify the target face out of a lineup comprising the target and 9 visually similar faces. This suggests that the spontaneous use of verbalization ordinarily facilitates face recognition. Moreover, articulatory suppression was shown to remove the verbal overshadowing effect: Accuracy at identifying the target face from the lineup was only impaired by the production of a facial description, if subvocal verbalization was possible when the target face was viewed earlier. Wickham and Swift (2006) suggest that articulatory suppression prevents spontaneous generation of a verbal code and thus there is no conflict with the verbal code generated during description.

If subvocal verbalization facilitates face learning as the foregoing suggests (Nakabayashi & Burton, 2008; Wickham & Swift, 2006), then presenting participants with a to-be-ignored auditory description of a different identity when the target face is being viewed may interfere with the verbal code generated for a target face, thereby impairing face recall and face recognition performance.

Irrelevant Sound Impairs Recall Accuracy

It is well-known that distraction effects can arise that are attributable to the similarity in semantic content between the to-be-remembered material and background speech. The semantic similarity between the to-be-ignored speech material and the to-be-remembered material (and presumably the “content” of rehearsal) impairs performance on tasks that make heavy demands on semantic processing (e.g., Bell, Buchner, & Mund, 2008; Marsh, Hughes,

& Jones, 2008). For example, Marsh and colleagues (2008) presented participants with categorized lists of words for free recall whereby all members of a visually-presented list were taken from the same semantic category (e.g., Fruit: *Banana, Grape, Pineapple*). When to-be-ignored distractors were taken from the same category as the visually-presented words (e.g., *Apple, Orange, Cherry*) as opposed to a different category (e.g., *Lorry, Bus, Bicycle*), fewer correct items were produced and more of the distractor items were erroneously recalled. Moreover, participants were often confident that the falsely recalled distractors had been visually presented when in fact they had an auditory origin. This suggests that the content of the to-be-ignored speech can impair memory for what is visually-presented and under some circumstances, can distort memory.

Current Study

The results of the research reported above allow us to derive the prediction that participants presented with to-be-ignored speech comprising a face description that is congruent with an unfamiliar to-be-learnt target face should produce more-accurate verbal descriptions of the target face, whereas participants presented with to-be-ignored speech comprising a face description that is incongruent with that target should produce less-accurate verbal descriptions of the target face. Specifically, an incongruent description could be expected to reduce correct recall of facial descriptors and to increase recall of inaccurate facial descriptors as compared to the congruent and quiet conditions. This would be synonymous with the finding that ignoring category-exemplars “Dog”, “Horse”) drawn from the same category as visually-presented items (“Cat”, “Donkey”) impairs correct recall of the visually-presented information and increases erroneous recall of the ignored exemplars (e.g., Marsh et al., 2008). Congruent speech, however, should not impair—and may even facilitate—recall of accurate facial descriptors since the auditory and visual information do not conflict. Expanding the empirical compass further to the applied domain, impairment in face recall and face recognition could also have transfer effects on the accuracy of facial composites.

Facial composites are computer-generated likenesses of a target that are constructed by a witness working with an interviewer typically following a cognitive interview (Fisher & Geiselman, 19992; see Fodarella, Kuivaniemi-Smith, & Frowd, 2015). Several methods of creating composite images of suspects have been devised. PRO-fit uses a computer interface wherein the witness selects individual face features. It is therefore described as a feature-based system and is thought to bias face processing toward recognition of the individual elements of a face rather than their configuration (spacing between facial features). Other

systems such as EvoFIT are described as holistic systems and rely more on configural processing (Frowd, Hancock, & Carson, 2004). For the current study, PRO-fit was selected because it was considered more likely to capture interference at the featural level and we considered that featural processing of the target face to be the most vulnerable to interference.

In the current study therefore, we examined the potential impact of to-be-ignored speech on face recall and construction. Akin to the real life situation wherein the witness does not typically know the identity of a perpetrator, participants were presented with one of two unfamiliar faces to learn. In the critical conditions, to-be-ignored speech was presented whilst the witness (participant) viewed the face. The relationship between the to-be-ignored speech and the target face was manipulated such that in the congruent condition the to-be-ignored speech comprised a face description that matched the target. In the incongruent speech condition the to-be-ignored speech comprised a description of a different, dissimilar identity. Participants then undertook a face-recall interview whereby they freely recalled as much information as possible about the face and following a two-minute distractor task, constructed a composite of the face using PRO-fit. Assuming our hypotheses that to-be-ignored speech comprising an incongruent verbal description reduces accuracy (i.e., reduces correct recall and increased erroneous recall of facial descriptors) we further hypothesized that the composites created may also be less accurate, as indexed by independent ratings of composite-to-target similarity, compared to the composites produced following exposure to a congruent verbal description or quiet. To examine whether familiarity with the target modulated the later composite-to-target similarity ratings we presented German celebrities as the unfamiliar targets to UK students and required both UK students (unfamiliar with the identity of the targets) and German students (familiar with the identity of the targets) to rate composite accuracy in the secondary stage of the study.

Primary Stage: Construction of Composites

Method

Participants

Forty-eight students (24 males and 24 females) aged between 18 and 24 (M = 21; SD = 1.5) years were recruited via opportunity sample at the University of Central Lancashire. All participants spoke English as their first language and reported normal (or corrected-to-normal) vision and normal hearing.

Design

A between-participants design was employed for the face construction phase. The independent variable was Sound Condition with three levels: quiet, congruent to-be-ignored speech, and incongruent to-be-ignored speech. For the face recall task the dependent variable was Facial Descriptor Category which had three levels: correct details, subjective details and incorrect details (see later). Participants were randomly assigned to one of the three sound conditions.

Apparatus and Materials

Photographs of the faces of two German celebrities who are unknown in the UK—Helge Schneider and Jürgen Vogel—were chosen on the basis that they were dissimilar to one another on overall appearance. Similarity and dissimilarity was measured by presenting a sample of 10 German celebrities to participants on a pairwise basis and asking the participants to rate their similarity on a seven point Likert scale (1 = not at all similar, 6 = highly similar). Schneider and Vogel were chosen because they were rated as the most dissimilar pair. A verbal description of both faces was generated by two staff members of the University of Central Lancashire that were naïve to the purpose of the experiment and the objective featural descriptors (e.g., “long hair”, “beard”, “bald”, “light eyebrows”) that were consistent with one identity but not the other were chosen following discussion. The facial descriptors chosen were chosen due to their opposing nature (e.g., “shaven” vs. beard”). The number of features (12) was therefore the same for Schneider and Vogel. Descriptions were digitally recorded by a male speaker (that was not one of the researchers) and sampled with a 16-bit resolution at a sampling rate of 44.1 kHz using Sound Forge 5 (Sonic Inc., Madison, WI, 2000). The speech was presented at a rate of approximately one word per 750 ms. The descriptions were looped to create sequences that were 30 seconds long. The recorded face description of one of the celebrities served as the incongruent sound for the other. Microsoft PowerPoint software was used to display the target faces and concurrently present the to-be-ignored speech in congruent and incongruent conditions. Target faces were presented centrally on a computer screen with dimensions of approximately 8 cm wide by 10 cm high for 30 seconds. Participants sat at approximately 50 cm from the screen. The faces therefore sustained a vertical visual angle of 11.42° and a horizontal angle of 9.15°. The face was cropped from its background context using Adobe Photoshop software and presented against a white background. Participants were informed by the Experimenter that they would be presented with a photograph of an unfamiliar face and that they were to study and memorise that face in as much detail as possible for a later memory test. A face recall sheet was used to

notate participants' freely recalled details about the target face. This included the headings: "overall", "shape", "hair", "eyebrows", "eyes", "nose", "mouth", "ears" and "other."

Procedure

Participants were recruited on the basis of being unfamiliar with German celebrities and were tested individually. They were first provided with a briefing that explained the broad aim of the study. No mention was made that a facial composite was required. Participants were randomly allocated, with equal sampling, to one of the three sound conditions: quiet, congruent to-be-ignored speech, and incongruent to-be-ignored speech. For each condition, the participant was given 30 seconds to visually inspect the target. Within each condition, Vogel was used half the time (8) and Schneider was used half the time (8). When a participant was assigned to the quiet condition no speech was presented. When a participant was assigned to the congruent condition, the to-be-ignored speech describing the target face was presented while the face was shown. When a participant was assigned to the incongruent condition the to-be-ignored speech describing the alternate identity was presented simultaneously with the target face (Vogel for Schneider, and Schneider for Vogel). Participants were told to ignore the speech. Immediately after viewing the target face, participants were asked to complete a distractor task—a crossword puzzle—for two minutes in quiet. Participants were then told that they would first describe the face seen prior to the distractor task and then to construct a composite of it. No sound was played while participants completed these tasks. The procedure for the face recall interview and face construction using PRO-fit is reported in detail in Fodarella et al. (2015, this volume). The Experimenter conducting the interview was blind as to the sound condition any given participant was in. Subsequent to composite construction, participants were directly asked whether they knew the identity of the target face (none reported that they did). The duration of the whole procedure was about 40 minutes.

Results for Primary Stage

Verbal Description. The quantity of information provided by participants relating to the target face in each condition was analyzed in accordance with the procedure set out by Meissner, Brigham, and Kelley (2001). A correct description was generated by two raters for the target face and a decision was reached between these two raters with regards to the details that would be classified as correct, subjective, or incorrect for each face. The Experimenter and another individual (rater) then coded the details in the descriptions. The two raters were

blind to ´the condition from which a description was obtained and to each other's ratings. Subjective details were responses that related to perceived personality or similarity to a well-known celebrity or family member: Descriptions that could not, therefore, be verified directly. Inter-rater agreement was high [Cohen's $K(48) = .87$, $p < .001$]. The experimenter and the rater resolved any disagreement between ratings through discussion. The mean number of correct, subjective and incorrect facial descriptors for each sound condition can be seen in Table 1. Generally, the pattern of means suggest that participants in the congruent condition produced more correct, and fewer incorrect, facial descriptors than those in the quiet and incongruent conditions. Moreover, participants in the incongruent condition produced more incorrect facial descriptors compared to participants in the quiet and congruent conditions.

Table 1. Mean number of facial descriptors recalled as a function of sound condition and description type. Standard deviations are presented in parentheses. *Significantly less than Congruent Speech, $p < .001$, and Quiet, $p < .001$. †Significantly greater than Incongruent Speech, $p < .001$, and Quiet, $p = .001$. †Significantly less than Congruent Speech, $p = .012$. ◊Significantly greater than Congruent Speech, $p < .001$, and Quiet, $p = .001$.

	Sound Condition		
	Quiet	**Congruent Speech**	**Incongruent Speech**
Verbal Description Type			
Correct	9.50†	11.63†	8.06*
	(1.37)	(1.93)	(1.77)
Subjective	2.63	2.94	2.56
	(1.15)	(1.06)	(1.82)
Incorrect	2.25	1.50	4.06◊
	(1.84)	(1.32)	(1.34)

A 3 (Sound Condition) × 3 (Facial Descriptor Category, within-subjects: correct, subjective, incorrect) × 2 (Target Identity) mixed factor Analysis of Variance (ANOVA)

revealed a main effect of Facial Descriptor Category, $F(2, 84) = 336.79$, $MSE = 2.38$, $p < .001$, $\eta_p^2 = .89$. Follow up pairwise comparisons showed that significantly more correct descriptors were recalled than subjective ($p < .001$ [$CI_{.95} = 6.43, 7.57$], $d = 3.9$) and incorrect ($p < .001$ [$CI_{.95} = 6.39, 7.86$)], $d = 3.5$) descriptors. Furthermore, the ANOVA revealed a main effect of Sound Condition, $F(2, 42) = 3.52$, $MSE = 1.22$, $p = .039$, $\eta_p^2 = .14$. Follow up pairwise comparisons revealed that significantly more information was recalled in the congruent condition than in the quiet condition ($p = .017$ [$CI_{.95} = .11, 1.02$], $d = 0.13$) and the incongruent condition ($p = .049$ [$CI_{.95} = .003, .91$], $d = 0.12$).

The ANOVA also revealed a significant interaction between Facial Descriptor Category and Sound Condition, $F(4, 84) = 15.88$, $MSE = 2.38$, $p < .001$, $\eta_p^2 = .43$. Follow up simple effects analysis (LSD) revealed that significantly more correct descriptors were produced in the congruent condition as compared with the quiet condition ($p < .001$; [$CI_{.95} = 1.01, 3.24$], $d = 1.27$) and the incongruent condition ($p < .001$; [$CI_{.95} = 2.45, 4.67$], $d = 1.93$). Significantly more correct descriptors were also produced in the quiet condition compared with the incongruent condition ($p = .012$; [$CI_{.95} = .33, 2.55$], $d = 0.91$). A significantly greater number of incorrect facial descriptors were produced in the incongruent condition compared to the congruent condition ($p < .001$ [$CI_{.95} = 1.51, 3.61$], $d = 1.93$) and the quiet condition ($p = .001$; [$CI_{.95} = [.76, 2.86]$, $d = 1.12$).

There was also main effect of Target Identity, $F(1, 42) = 11.00$, $MSE = 1.22$, $p = .002$, $h_p^2 = .21$, whereby the number of facial descriptors provided for Schneider ($M = 5.32$, $SD = 4.09$) exceeded those provided for Vogel ($M = 4.71$, $SD = 3.52$; $p = .002$, [$CI_{.95} = .24, .98$] $d = 0.16$). An interaction between Facial Descriptor Category and Target Identity was also obtained, $F(2, 84) = 7.33$, $MSE = 2.38$, $p = .001$, $\eta_p^2 = .15$. Generally, significantly more correct facial descriptors were provided for Schneider ($M = 10.50$; $SD = 2.30$) than Vogel ($M = 8.96$; $SD = 1.90$), $p = .001$ [$CI_{.95} = .64, 2.45$] $d = 0.73$. However, this was also the case for incorrect facial descriptors ($M = 3.13$; $SD = 1.60$; Schneider; $M = 2.08$, $SD = 1.95$, Vogel), $p = .018$ $CI_{.95} = .19, 1.90$] $d = 0.59$). Subjective facial descriptors were significantly more common for Vogel ($M = 3.08$; $SD = 1.41$) than for Schneider ($M = 2.33$; $SD = 1.24$; $p = .034$ $CI_{.95} = .06, 1.44$] $d = 0.57$).

Secondary Stage: Rating of Composite Likeness

Method

Participants. One group of participants that were unfamiliar with the identities of the targets comprised an opportunity sample of 8 male participants, 18 female participants aged between 19 and 53 (M = 31; SD = 11.2) years from the University of Central Lancashire, UK. Another group of participants familiar with the identities of the targets[1] comprised an opportunity sample of 12 male and 11 female student and staff members aged between 22 and 44 (M = 27, SD = 6) years from Heinrich Heine University, Germany. The purpose of using one group unfamiliar and one group familiar with the target identities was to mimic the situation in real-life whereby people may recognize the similarity between a composite and a familiar person (e.g., a family member, or neighbour) or between the composite and an unfamiliar person (e.g., a person on the street, or a novel customer).

Design. The independent variables were Target Familiarity (unfamiliar, familiar; between subjects) and Sound Condition under which the faces had been encountered in Experiment 1 with three levels (quiet, congruent, incongruent) as well as Target Identity with two levels (Schneider, Vogel). The dependent variables were participants' likeness and distinctiveness ratings (see below).

[1] The data of four German participants who indicated that they did not recognize one of the target faces were not analyzed.

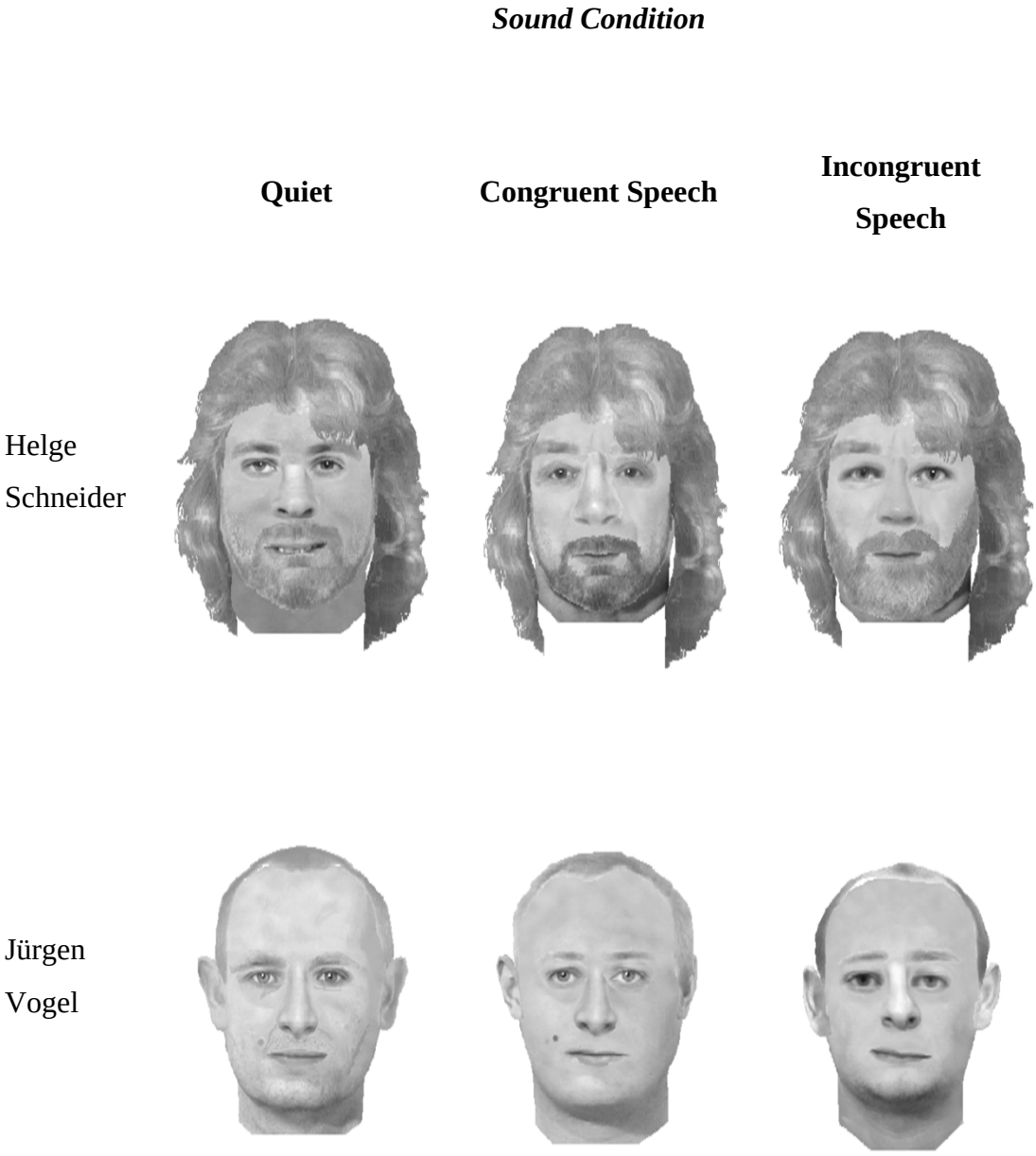

Figure 1: Examples of the male targets Helge Schneider (top row) and Jürgen Vogel (bottom row) constructed in the three conditions of the experiment (displayed are those composites that have received the highest ratings in the German sample). For reasons of copyright, we are unable to reproduce the target photograph used in the experiment.

Materials. Materials were the forty-eight composites produced in the Primary Stage of the study. There were sixteen composites in each of the three sound conditions and of these, eight were of Jürgen Vogel and eight were of Helge Schneider.

Procedure. Participants were tested individually. A computer program was written in LiveCode (http://livecode.com/download/) which displayed each composite against the target face. Participants were required to give a likeness rating for each target-composite pair using a 6-point scale (from 1 = "very poor likeness" to 6 = "very good likeness"). Targets were presented to the left of the screen with composites presented to the right. All forty-eight composites were presented (24 for each assigned identity) in an order that was random for each participant. When all of the target-composite pairs had been presented participants were asked if they were familiar with the identity of the two targets and if possible to provide a name or any semantic information that would individuate them. Because visual inspection of the two photographs suggested that the two identities may differ in distinctiveness, participants were then asked to rate the targets' distinctiveness on a six-point scale (from 1 = "very typical" to 6 = "very distinct"); composites were not presented. The rating task took approximately five minutes.

Results for secondary stage

Table 2 shows the means for the likeness scores given by the 26 English raters and the 23 German raters to composites derived from each of the three sound conditions. The pattern of means suggest that the composites created after viewing the target in the presence of incongruent speech were poorer likenesses of the target than those constructed after exposure to congruent speech and quiet.

Table 2: Mean likeness ratings for composite images in the three sound conditions (standard deviations are presented in parentheses).

	Schneider		**Vogel**	
Presentation Type	Unfamiliar	Familiar	Unfamiliar	Familiar
Quiet	2.53	2.36	2.41	2.14
	(0.66)	(0.85)	(0.61)	(0.82)
Congruent	2.74	2.49	2.14	2.09
Speech	(0.71)	(0.92)	(0.57)	(0.83)
Incongruent	2.21	2.16	1.93	1.88
Speech	(0.61)	(0.91)	(0.50)	(0.75)

A 3 (Sound Condition in Experiment 1) × 2 (Target Identity) × (Target Familiarity) ANOVA revealed a main effect of Sound Condition in Experiment 1, $F(2, 94) = 35.00$, $MSE = 0.096$, $p < .001$, $h_p^2 = .43$, whereby ratings given for quiet ($M = 2.36$, $SD = 0.74$) were significantly different from those given for incongruent speech ($M = 2.04$, $SD = 0.71$; $p < .001$ [$CI_{.95} = 0.23, 0.41$, $d = 0.45$], and those given for congruent speech ($M = 2.37$, $SD = 0.80$) were significantly different from those given for incongruent speech ($p < .001$ [$CI_{.95} = 0.23, 0.41$], $d = 0.43$), but there was no difference between quiet and congruent speech ($p = .96$). There was also a main effect of Target Identity, $F(1, 47) = 15.78$, $MSE = 0.47$, $p < .001$, $\eta_p^2 = .25$, whereby ratings given to Schneider ($M = 2.42$, $SD = 0.79$) were higher than those given to Vogel ($M = 2.10$, $SD = 0.70$; $p < .001$ [$CI_{.95} = 0.16, 0.48$], $d = 0.43$. There was no main effect of Target Familiarity, $F(1, 47) = 0.62$, $MSE = 2.39$, $p = .43$, $\eta_p^2 = .01$.

The interaction between Sound Condition in Experiment 1 and Target Identity was significant, $F(2, 94) = 8.33$, $MSE = 0.09$, $p < .001$, $\eta_p^2 = .15$. A Simple effects analysis (LSD) revealed that, for Schneider, likeness ratings were significantly higher for congruent speech than for quiet ($M = 2.62$, $SD = 0.81$, congruent speech; $M = 2.45$, $SD = 0.72$ quiet; $p = .016$ [$CI_{.95} = 0.03, 0.31$], $d = 0.22$). Moreover ratings were significantly higher for quiet than for incongruent speech ($M = 2.19$, $SD = 0.76$; $p < .001$, [$CI_{.95} = 0.14, 0.38$], $d = 0.20$), and higher for congruent speech than incongruent speech ($p < .001$, [$CI_{.95} = 0.30, 0.57$]).

For Vogel, ratings were significantly higher for quiet ($M = 2.28$, $SD = 0.72$) than for incongruent speech ($M = 1.90$, $SD = 0.62$, $p < .001$ [$CI_{.95} = 0.26, 0.49$], $d = 0.57$) and higher for congruent speech ($M = 2.11$, $SD = 0.70$) than incongruent speech, $p < .001$ [$CI_{.95} = 0.11, 0.31$], $d = 0.32$. However, ratings were higher for quiet than for congruent speech ($p = .01$, [$CI_{.95} = 0.04, 0.29$], $d = 0.24$.

To examine whether the two identities differed in terms of typicality, the distinctiveness ratings assigned to the two target photographs (from which the composites were constructed) were subjected to a 2 (Target Identity) × 2 (Target Familiarity) ANOVA. This revealed a main effect of Target Identity, $F(1, 47) = 4.51$, $MSE = 0.75$, $p = .034$, $\eta_p^2 = .09$, whereby Schneider ($M = 3.86$; $SD = 1.04$) was rated as more distinctive than Vogel ($M = 3.47$, $SD = 1.06$; $p = .039$ [$CI_{.95} = 0.02, 0.73$] $d = 0.37$). However, there was no main effect of Target Familiarity, $F(1, 47) = 1.62$, $MSE = 1.43$, $p = .21$, $\eta_p^2 = .03$, nor any interaction between Target Identity and Target Familiarity, $F(1, 47) = 1.44$, $MSE = 0.75$, $p = .17$, $\eta_p^2 = .04$.

Discussion

In summary, this study has shown that incongruent background speech can impact negatively on the encoding of facial details concerning an unfamiliar target: Incongruent impaired recall of correct facial descriptors and exacerbated recall of incorrect descriptors. Moreover, incongruent speech led to the construction of composites that were poorer likenesses of the target face as compared to the congruent and quiet conditions.

The finding that fewer correct, and more incorrect, facial descriptors were recalled in the incongruent condition can be viewed as analogous to the finding that background speech that is categorically-related to to-be-remembered lists of words reduces correct recall of visually-presented items and also leads to the erroneous recall of distractors (e.g., Marsh et al., 2008). Further work is encouraged to determine whether incongruent speech—which induces recall of incorrect facial descriptors—is leading to memory-blending, whereby the irrelevant verbal features presented as part of the auditory stream somehow combine with some of the visual features from the target face creating a novel trace within memory (cf. Busey & Tunnicliff, 1999).

From a theoretical point of view, one might think that both incongruent and congruent facial descriptions presented as background speech should damage face memory. For example, within the context of the verbal overshadowing effect, the description provided by the participant after facial encoding is an attempt to be as true to the face that they have just seen as possible and yet this congruent description impairs subsequent recognition performance (e.g., Wickham & Swift, 2006). This begs the question of why, in the context of our study, a congruent description presented concurrently while viewing the target can (depending on the target) facilitate face recall and composite production. One possibility as to why congruent speech improved recall of correct facial descriptors is that the repeated presentation of the congruent verbal description reinforces the content of the participant's rehearsal. However, if this was the case, then one might expect both composites constructed in the congruent condition to be better likenesses of the target which was not the case. Our suggestion as to why this may occur is shaped by the idea that the match between the verbal code spontaneously generated at study and produced during description of the target face determines the magnitude of interference and the concomitant disruption of face memory: the verbal overshadowing effect (cf. Wickham & Swift, 2006). The data demonstrate that congruent speech was beneficial to the composite construction of the face with the higher level of distinctiveness (Schneider), but detrimental to the composite construction of the face with the lower level of distinctiveness (Vogel). It is possible that the congruent speech

matched, and reinforced, the verbalizable, objective features of Schneider, thereby improving memory for the objective (correct) descriptors, For Vogel, participants have relied more on subjective descriptors as is evident within the data for the facial descriptors recalled. The congruent speech could, therefore, interfere with the subjective descriptors that may be more useful for the construction of the less distinctive composite. This notion coheres rather nicely with the idea that verbalization impairs memory for typical rather than distinctive faces (Wickham & Swift, 2006). For example, in their study of verbal overshadowing, Wickham and Swift (2006) found that post-encoding descriptions of faces that were independently rated as typical, as compared to distinctive, interfered more with the spontaneous verbal code created during the face learning, thereby impairing recognition of the face from a lineup. This suggests that potentially useful subjective labels created during facial encoding may be susceptible to interference from the description given subsequent to the encoding. Future work that seeks to manipulate the distinctiveness of the faces from the outset is required to add weight to the foregoing conclusions.

If, as we have outlined in the foregoing, to-be-ignored featural face descriptions impact on face memory at a featural rather than configural level, this has implications for the system selected for composite constructions. Composites constructed using PRO-fit, a system weighted toward feature recognition, should be more susceptible to the disruptive effects of incongruent description than EvoFIT, since EvoFIT is a holistic system that places greater weight on the configural organisation of features than the features themselves (Frowd et al., 2004). EvoFIT might, for example, attenuate the potentially disruptive effect of exposure to incongruent facial descriptions. In a similar vein, using a holistic cognitive interview that encourages the witness to remember the face holistically could also reduce the distraction effect (Frowd, Bruce, Smith, & Hancock, 2008).

Other factors that should be addressed within future work include the duration of exposure to the target face and the delay between face learning and face construction since thirty seconds is a rather long time to encode an unfamiliar person's face: In reality, witnesses may only get a matter of seconds. Moreover, composites are seldom constructed on the day the crime occurred and it would therefore appear necessary to compare composite construction after a few minutes or a couple of hours to that after a forensically valid 24-hour delay. To our knowledge this is one of the first studies to demonstrate that background speech can produce adverse effects on the recall and recognition of complex visual information and the only one that concerns memory for the appearance of a human face. Certainly, there are parallels to be drawn here with other work that reveals the fallibility of eyewitness memory.

Previous studies have shown that individual's memory for visual details is susceptible to misinformation presented after the event (Loftus, Miller, & Burns, 1978). For example, co-witness information about a target's appearance can impair performance on a line-up identification task (Zajac & Henderson, 2008). Moreover, other research has shown that the misinformation effect is modulated by credibility: if participants perceive the source of information as incredible, or having an intention to mislead, they will effectively resist suggestion (Underwood & Pezdek, 1998). Future research should therefore address whether the disruption produced by incongruent description triggers more incorrect descriptions and the congruent description triggers more correct descriptors because they are perceived as coming from an authoritative and credible source.

In terms of the practical implications of the foregoing research, on the whole it would seem that distraction should be reduced as much as possible. The results demonstrate that it is very difficult to overcome interference even if the information is known to be irrelevant, and even if the information is presented in a different modality. Although our results demonstrate that congruent background speech was less damaging to composite construction than incongruent background speech and actually proffered an advantage, this was for a face rated as more distinctive, and the possibility of witnessing a distinctive face and hearing such a detailed and accurate background speech description, we feel, is unlikely in the applied context. However, this is not to detract from the real possibility that eyewitnesses can be passively exposed to background speech either during the witnessing of a crime or post-event within the delay between witnessing the event and prior to the opportunity to recall the event at interview (Gabbert, Hope, Fisher, & Jamieson, 2012). Moreover, there are applied circumstances in which background speech may affect the ability to identify a suspect. For example, in the context of security surveillance, the closed circuit television operator in the control room is often tasked with identifying suspects from physical descriptions. It is not uncommon that this task must be undertaken against a background of conversation that may include descriptions of other suspects (Tremblay, personal communication). Although, in this context the task of the operator would seem more of a "matching" task, it nonetheless demonstrates that the cognitive limitations of the operator in terms of distractibility can have consequences for the efficiency of security surveillance and monitoring work. Generally then, the present findings demonstrate that inconsistent background speech information impairs performance even if participants know that the information is irrelevant and should be ignored. Therefore, attempts should be made to minimise exposure to background speech whenever a task critical depends on the accuracy of person identification.

Implications for practice

- Background speech affects face recall and composite construction even if participants are told to ignore the speech.
- The detrimental effects of to-be-ignored facial descriptions is determined by congruity: Only incongruent descriptions produced clearly detrimental effects to face recall and composite construction.
- Attempts should be made to minimize exposure to background speech, and in particular exposure to incongruent facial descriptions, whenever a task critically depends on the accuracy of person identification.
- Exposure to incongruent facial descriptions should be taken into account as a potentially negative influence on the accuracy of eyewitness accounts.

References

Bell, R., Buchner, A., & Mund, I. (2008), "Age-related differences in irrelevant speech effects", *Psychology & Aging*, Vol. 23, pp. 377-391.

Busey, T.A., & Tunnicliff, J.L. (1999), "Accounts of blending, distinctiveness, and typicality in the false recognition of faces", *Journal of Experimental Psychology: Learning, Memory, & Cognition*, Vol. 25, pp. 1210-1235.

Gabbert, F., Hope, L., Fisher, R.P., & Jamieson, K.J. (2012), "Protecting against misleading post-event information with a self-administered interview", *Applied Cognitive Psychology*, Vol. 26, pp. 568-575.

Fisher, R.P., & Geiselman, R.E. (1992), Memory enhancing techniques for investigative interviewing: *The cognitive interview*, Charles C. Thomas, Springfield, IL.

Fodarella, C., Kuivaniemi-Smith, H.J., & Frowd, C.D. (2015), "Detailed procedures for forensic face construction", *Journal of Forensic Practice*.

Frowd, C.D., Hancock, P.J.B., & Carson, D. (2004), "EvoFIT: A holistic, evolutionary facial imaging technique for creating composites", *ACM Transactions on Applied Psychology*, Vol. 1, pp. 1-21.

Frowd, C.D., Bruce, V., Smith, A.J., & Hancock, P.J.B. (2008), "Improving the quality of facial composites using a holistic cognitive interview", *Journal of Experimental Psychology: Applied*, Vol. 14, pp. 276-287.

Hodgetts, H.M., Vachon, F., & Tremblay, S. (2014), "Background sound impairs interruption recovery in dynamic tasks: Procedural conflict?", *Applied Cognitive Psychology*, Vol. 28, pp. 10-21.

Hughes, R.W, & Jones, D.M. (2001), "The intrusiveness of sound: Laboratory findings and their implications for noise abatement", *Noise & Health*, Vol. 4, pp. 51-70.

Kebbell, M., & Milne, R. (1998), "Police officers perceptions of eyewitness factors in Forensic investigations", *Journal of Social Psychology*, Vol. 138, pp. 323–330.

Koriat, A., & Goldsmith, M. (1994), "Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment", *Journal of Experimental Psychology: General*, Vol. 123, pp. 297-315.

Loftus, E.F. (1975), "Reconstructing memory: The incredible eyewitness", *Jurimetrics*, Vol. 15, pp. 188-193.

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978), "Semantic integration of verbal information into a visual memory", *Journal of Experimental Psychology: Human Learning & Memory*, Vol. 4, pp. 19-31.

Marsh, J.E., Patel, K., Skelton, F., Vachon, F., Fergus, J., & Frowd, C.D. (2014), Chatting in the face of the eyewitness: The impact of extraneous cell-phone conversations on memory for a perpetrator. Manuscript submitted for publication.

Marsh, J.E., Hughes, R.W., & Jones, D.M. (2008), "Auditory distraction in semantic memory: A process-based approach", *Journal of Memory & Language*, Vol. 58, pp. 682-700.

Meissner, C.A., Brigham, J.A., & Kelley, C.M. (2001). "The influence of retrieval processes in verbal overshadowing", *Memory & Cognition*, Vol. 29, pp. 176-186.

Nakabayashi, K., & Burton, A.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*, Vol. 20, pp. 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*, Vol. 38, pp. 376-390.

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

Schooler, J.W., & Engstler-Schooler, T.Y. (1990), "Verbal overshadowing of visual memories: Some things are better left unsaid", *Cognitive Psychology*, Vol. 22, pp. 36-71.

Underwood, J., & Pezdek, K. (1998), "Memory suggestibility as an example of the sleeper effect", *Psychonomic Bulletin & Review*, Vol. 5, pp. 449-453.

Vanags, T., Carroll, M., & Perfect, T.J. (2005), "Verbal overshadowing: A sound theory in voice recognition", *Applied Cognitive Psychology*, Vol. 19, pp. 1127-1144.

Wais, P.E., Gazzaley, A. (2011), "The impact of auditory distraction on retrieval of visual memories", *Psychonomic Bulletin & Review*, Vol. 18, pp. 1090–1097.

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

Zajac, R., & Henderson, N. (2009), "Don't it make my brown eyes blue: Co-witness information about a target's appearance can impair target-absent line-up performance", *Memory*, Vol. 17, pp. 266-278.