The process of facial composite production

To appear in A. Czerederecka, T. Jaskiewicz-Obydzinska, R. Roesch & J. Wojcikiewicz (Eds.). *Forensic Psychology and Law* (pp. 140-152). Krakow: Institute of Forensic Research Publishers.

Charlie Frowd*
Department of Psychology, University of Stirling, UK.

Dawn McQuiston-Surrett
Department of Psychology, Arizona State University West, USA.

Imran Kirkland
Department of Psychology, Arizona State University West, USA.

Peter Hancock
Department of Psychology, University of Stirling, UK.

*Corresponding author (email: cdf1@stir.ac.uk)

Facial composites remain an important tool for the apprehension of suspects. While research has led to an improvement in composite quality, contemporary systems rarely produce highly identifiable renditions in realistic studies. The current work manipulated the type of interview administered, as well as the composite face first presented to witnesses during construction. In the study, witnesses received a short exposure to a target face, then 2 days later were given a Cognitive Interview, a Holistic Interview (a set of personality ratings) or no interview. They then constructed a composite with the modern PROfit system in the normal way, by first setting the feature descriptors to obtain an 'initial' composite, or by proceeding from the 'default' composite, the first face displayed when PROfit is run. Although the composites were not named spontaneously, the result of an identification task revealed an advantage for the Holistic Interview when used with an 'initial' composite.
(150 words)

Keywords: Facial composite, witness, interview, holistic

Facial composites are used by law enforcement agencies to identify individuals who commit crime. To construct a composite, a witness or victim normally describes a suspect's face and, with the assistance of a skilled computer technician or 'operator', uses computer software to select an appropriate set of facial features. The resultant face is then circulated within a police force, or distributed more generally in the media, to generate lines of enquiry.

The computer software available for composite production is now quite sophisticated, containing a large number of facial features that can be resized and positioned as required. There are also software painting tools available to improve the likeness, such as the addition of ageing lines and wrinkles. Many such systems are available worldwide, including Identikit 2000 and FACES in the U.S.; and E-FIT and PROfit in the UK.

These contemporary systems have been the subject of some empirical evaluation. The general methods used to evaluate them have largely evolved from research carried out in the 1970s and 1980s on the non-computerised Photofit and Identikit systems (e.g. Davies, Ellis and Shepherd, 1978). This lab-based work involved recruiting 'mock' witnesses who first would be exposed to an 'assailant', normally in the form of a photograph, then describe his face and construct a composite using a composite system. Composites from a number of such witnesses would then be evaluated by 'judges' using likeness ratings, or by matching the composites to the target photographs (a 'sorting' task). In fact, the systems have improved to such an extent that composites may now be named spontaneously about 20% of the time (Brace et al., 2000; Bruce et al., 2002; Davies et al., 2000; Frowd et al., 2004c).

While it is encouraging to find a fair level of composite naming, a different story appears to be emerging when a more realistic paradigm is sought: naming levels fall to a few percent when a witness is required to wait a couple of days between seeing a target face and constructing a composite (Frowd et al., 2004a, 2004b, 2004d; Koehn and Fisher, 1997). This is a typical delay found in police work, a time that is unlikely to be reduced due to witness trauma and other practical constraints (e.g. availability of police resources).

It would seem that a witness's memory becomes appreciably weaker after several days, resulting in a relatively poor composite. As a consequence, a Cognitive Interview is used with witnesses to assist with the recall of information (e.g. Geiselman et al., 1986). In the realistic studies of Frowd et al. (2004b, 2004c, 2004d), several aspects of the Cognitive Interview were employed – including rapport building, mental reinstatement of context, free recall and cued recall – yet naming remained low. However, composite construction contains elements of recall – describing a face, and selecting facial features – and recognition, deciding when a composite face may be recognisable to another person. Therefore, enhancing recognition, or the holistic aspects of a witness's memory, may be beneficial.

Frowd et al. (2004a) investigated this possibility through the use of a Holistic Interview, a procedure that required their witnesses to make a number of personality judgements. Although utilising a minimal delay between inspecting a target face and composite construction, they found better quality composites using the Holistic Interview than a Cognitive Interview. The current study extends their work by comparing the effectiveness of these 2 interviews over a longer retention interval; comparison was also made to composite construction without an interview, a situation where the worst performance was expected.

The current work also manipulated a second factor: the composite face first presented to a witness. As contemporary systems contain too many features for a witness to reasonably inspect, an operator uses the witness's verbal description to identify a subset of features; achieved by setting the 'feature descriptors' (in PROfit or E-FIT). The operator then presents the witness with an 'initial' composite, a face that best matches the description. It is from this representation that a witness suggests amendments to create the best possible likeness of the assailant. It has been established that verbal and visual memories are different (e.g. Davies, 1983), and so the use of a verbal memory (description) to localise visual items (facial features) may introduce error – perhaps emerging as a misunderstanding (on the part of the operator) from what is perhaps best described as the 'vagaries of language'.

Frowd et al. (2004a) provide an alternative construction procedure that does not depend on a verbal description. In one of their experiments, construction commenced by presenting witnesses with a 'default' composite, the composite face that is displayed when the PROfit system is first run; it contains example facial features listed first in the database. An operator then clicked through several example features in the presence of the witness, up to about 2 dozen, to establish a general pattern for each feature, and then used the feature descriptors to narrow the search. Thus, the procedure enabled construction to proceed without the dependence on the witness's verbal description. We adopted this approach in our study and compared it with the current UK approach that uses feature descriptors as the *first* stage in construction (a procedure that gives rise to the initial composite).

In summary, the current study recruited 'mock' witnesses to view a target face and construct a composite 2 days later using PROfit software in one of 6 conditions: witnesses underwent a Cognitive Interview, a Holistic Interview or no interview prior to composite construction; they then constructed a composite starting with either the 'default' composite, a face containing features listed first in the PROfit database, or the 'initial' composite, a face produced after setting PROfit's feature descriptors. Famous faces were used as targets to allow the composites to be given to third persons to be named (the targets were also unfamiliar to witnesses). It was expected that the Holistic Interview coupled with the default face condition would be the best.

Method

The design opted for a realistic methodology, being as close to real life construction as possible in the Cognitive Interview / initial face condition (baseline). In general, witnesses were shown a photograph of a celebrity face, then 2 days later constructed a composite using PROfit software with the assistance of an experienced PROfit operator. As in real life, the composite construction session was open-ended and under the control of the witness; software tools were also provided to allow composites to be artistically enhanced if required by the witness.

The design involved construction in the US and evaluation in the UK, with the celebrity targets well-known in the UK but not in the US. This allowed unfamiliar face processing at construction and familiar face processing at test, the correct real life model. The targets were 8 celebrities taken under good lighting conditions in a full-face pose and a neutral expression (as far as possible). In addition, the celebrities were clean shaven (or had minor stubble) and none wore spectacles. Included were 3 TV presenters (Ant McPartlin, Paul Merton and Alan Titchmarsh), 2 pop stars (Noel Gallagher and Bob Geldof), 2 sportsmen (Tim Henman and Stephen Hendry) and a politician (William Hague).

Composite construction

Participants

Forty-eight undergraduates at Arizona State University West received a course credit to act as witnesses and construct a composite.

Procedure

A procedure that follows UK police guidelines was used to construct the composites. This procedure is given in brief here but may be found in full elsewhere (e.g. Frowd et al., 2004b, 2004c). Participants (witnesses) visited the laboratory twice, first to inspect a celebrity photograph and 2 days later to construct a composite. In the first visit, they inspected one of eight randomly-selected celebrity faces for 1 minute; each person was aware that a facial composite would be required. A check was made to verify that the selected celebrity was not recognised (and none of the witnesses reported knowing their selected celebrity). Each target was constructed 6 times, once for each of the experimental conditions.

In the second visit, witnesses were given a Cognitive Interview, a Holistic Interview or no interview. The Cognitive Interview involved context reinstatement, and one cycle of free recall and cued recall. For the Holistic Interview, witnesses were asked to make a series of personality judgements about their celebrity face in the form of 7-point ratings (1 = *low* / 7 = *high*). Prior to each

rating, witnesses were asked to form a mental image of their target face. The traits requested were: honesty, intelligence, friendliness, kind\ness, excitability, selfish and arrogance. Witnesses skipped this part if assigned to the no interview condition.

The session moved on to composite construction. For witnesses in the default face condition, the operator initially displayed the 'default' composite, the first face available when PROfit is run. Using this face as a reference, the operator worked with witnesses to select facial features, normally starting with hair and face shape, but this order was determined by witnesses. For each feature, the operator presented the first 2 dozen examples in the database, and witnesses identified those that were most similar. The operator then used the feature descriptors to present a more focussed set and witnesses selected the best, with resizing and repositioning as required. The artistic paint package within PROfit was made available to allow enhancement of selected features.

Witnesses in the initial face condition were presented with a composite face with features identified using PROfit's feature descriptors. The setting of these descriptors varied by interview condition. For witnesses receiving a Cognitive Interview, the operator directly set the descriptors to match the verbal description given during the free and cued recall part of the Cognitive Interview. As a verbal description was not available for the other interview conditions, the operator introduced the feature descriptors and witnesses were invited to set them appropriately. When the descriptors had been set, the initial composite face was displayed and construction proceed as for the default face condition.

Composite evaluation

Examples of the composites produced in the study may be seen in Figure 1. The set of 48 composites were printed (to a size roughly 5cm x 6cm) and evaluated by a new set of participants, or judges, first by composite naming and then by sorting and identification tasks.

Figure 1 about here.

Composite naming

As participants inspected all 48 composites, composite naming is a repeated-measures design for type of interview (Cognitive Interview / Holistic Interview / No Interview) and starting face (default / initial).

Participants

A total of 8 staff and students named the composites.

Procedure

Participants were tested individually. Each person was told that a set of facial composites had been constructed of well-known British celebrities. They were asked to provide a name for each composite, even if they were not certain. Thus, each of the composites was presented sequentially and participants provided a name where possible. When all 48 composites had been inspected, naming was repeated for the target photographs. An *a priori* rule was applied such that participants were excluded if they did not know at least a half of the target photographs (otherwise few composites would be named). In practice, only one of the twelve participants was omitted on this basis. The order of presentation for composites and target photographs was randomised after each person.

Results

In spite of participants correctly naming at least 6 of the target photographs, and therefore exhibiting good familiarity with the target set, none of the composites were correctly named.

Discussion

Unexpectedly, the level of naming was very low across all conditions. This result is clearly similar to other research involving composite construction after a 2 day delay (Frowd et al., 2004a, 2004b, 2004d; Koehn and Fisher, 1997). We administered a composite sorting task as an alternative.

Composite sorting

This task required another group of participants to match the composites to their target photographs. As for composite naming, sorting is a repeated measure design for interview type and starting face.

Participants

Seventeen undergraduates were given a course credit to sort the composites.

Procedure

Participants were tested individually. Each person was informed that they would be evaluating a set of facial composites constructed in a realistic study. The celebrity photographs were introduced and laid out in front of the participants (in 2 rows of 5). A pile containing all 48 composites were given to each person and they were told to match the composites to the target photographs. It was also mentioned that as there were more than one composite of each celebrity, participants should make a pile in front of each photograph. In addition, they were told to match the composites in the order given, taking as long as necessary, but to try to avoid making exchanges once placed on the table. The order of presentation for composites and targets was randomised after each person.

Results

The mean sorting accuracy was the same for composites constructed by the Cognitive Interview (58.6%) and the Holistic Interview (58.6%), and only slightly better than those without an interview (57.0%). Similarly, sorting accuracy in the initial face condition (57.3%) was very similar to the default condition (58.9%). Accordingly, a repeated-measures Analysis of Variance (ANOVA) of the participant accuracy scores was not significant for interview, $F(2, 30) = 0.02$, $p = .901$, nor starting face, $F(1, 15) = 0.05$, $p = .594$; the interaction between these factors was not significant either, $F(2, 30) = 1.04$, $p = .366$.

Discussion

Overall, composites were sorted quite well (58.1%) but, unfortunately, neither the interview type nor the starting face affected composite quality. Even though the sort task has been found to perform reasonably well compared with composite naming – Davies et al. (2000), Frowd et al. (2004b, 2004c) – it is not a perfect match and so we administered a ‘cued’ naming task (identification).

Identification

One problem with the sorting task is that the target photographs are present and so participants may carry out a feature-by-feature match rather than engage in familiar face processing (necessary for naming). A better task involves participants selecting from a list of celebrity names, prompting recognition. We administered such a task, providing a list of written names for each composite. Once again, a repeated-measures design was employed as participants inspected all composites.

Participants

Seventeen students at Stirling University volunteered for the identification task.

Procedure

Participants were tested individually. They were told that they would be evaluating composites of famous faces constructed in a realistic study. This time, for each composite, they were asked to select a name from a written list of alternatives (the names of the celebrities). They worked through the composites sequentially, selecting a name from the list (a forced choice task). Afterwards, they repeated the identification exercise for the target photographs. As before, the order of presentation was randomised for composites and targets.

Results

The overall identification in the cued naming task was 43.2%. Performance by interview type and starting face can be seen in Figure 2. Composites were identified best after a Cognitive Interview (47.3%), followed by a Holistic Interview (43.0%); they were worst without an interview (39.5%). Varying the starting face affected identification to a lesser degree; 43.0% from the default face and 43.4% from the initial face condition. A repeated-measures ANOVA of participant accuracy scores was significant for interview type, $F(2, 30) = 3.48$, $p = .044$, but the starting face was not significant, $F(1, 15) = 0.03$, $p = .870$, and neither was the interaction, $F(2, 30) = 0.82$, $p = .449$. Simple contrasts of the ANOVA revealed that a Cognitive Interview was better than no interview, $p < .05$, but there was no difference between a Cognitive Interview and a Holistic Interview.

Figure 2 about here.

Discussion

The result of the identification task, with selection made from a list of written names, revealed a general benefit for the presence of an interview during composite construction. Once again, there appeared to be no difference in quality when the starting face was changed and there was no benefit of the holistic procedure relative to a Cognitive Interview.

Before discussing the implications of these data, we present a final evaluation. We received reports from some participants that several composites could be identified by their hairstyle. This notion appears to be at least in part correct: for example, William Hague (politician) has a distinctive hairstyle (bald pate) and his composites were correctly identified 77% of the time, by far the best (the next best were composites of Stephen Hendry, at 49%). Interestingly, the data also reveals that composites of Ant McPartlin (TV presenter) were (mis)identified most often as Tim Henman (sportsman), which is understandable given the similarity of their hairstyles. We therefore re-ran the identification task with the hair removed, with the aim of promoting a closer match to naming (dominated by the internal features of a face, e.g. Bruce and Young, 1998).

Identification: internal features

The 48 composites were reworked in Adobe Photoshop to remove the external facial features. This was achieved using the *elliptical marquee tool*, to select an overall shape for the face, and the *rectangular marquee tool*, to unselect most of the forehead (often containing hair); the result being a u-shaped image. In this way, the composite faces had little or no hair present.

Participants

Seventeen students at Stirling University volunteered for the identification task.

Method and Procedure

The previous identification task was repeated using the internal feature composites.

Results

Overall performance was much lower at 26.8% for the internal feature composites, reflecting the increased difficulty of the task; note that this level is still appreciably above chance (12.5%). Performance by interview type and starting face can be seen in Figure 3. This time, the Holistic Interview performed best (30.5%), followed by the Cognitive Interview (26.5%) and no interview (23.5%). This time, the difference between the default and initial face condition was quite large (12.5%), with the initial face superior (33.1% vs. 20.6%). The ANOVA was marginally significant for interview type, $F(2, 30) = 3.24$, $p = 0.053$, but was significant for starting face, $F(1, 15) = 32.2$, $p < .001$, and the interaction was also significant, $F(3, 30) = 9.33$, $p < .001$. Simple-main effects revealed that the composites in the initial face condition constructed with a Holistic Interview were better than a Cognitive Interview, $p < .01$, or no interview, $p < .001$; for the default face condition, there was an approaching advantage in the opposite direction, with a Cognitive Interview

outperforming the Holistic Interview, p = .072. In addition, composites constructed using a Holistic Interview were much better in the initial condition than the default condition. No other significant differences were observed.

Figure 3 about here.

Discussion

Although the identification task was much more difficult for the judges with just internal features, a somewhat different pattern emerged. This time, a clear benefit emerged for the Holistic Interview, but only when construction followed the initial face route; when construction took place with a default face, a Cognitive Interview was better (approached significance). Holistic Interviews aside, there was no difference between the starting faces and there was also no evidence of a difference between a Cognitive Interview and no interview. In addition, the data suggest that the previous identification task was greatly influenced by the presence of the external facial features (most likely hair).

General Discussion

The current study sought to improve the poor level of naming emerging from facial composites in use today. A design was employed that followed current UK police procedures as far as possible. This involved 'mock' witnesses who inspected a celebrity face and then constructed a composite after a 2 day delay. Witnesses were given one of 3 types of interview: a Cognitive Interview (baseline), a Holistic Interview or no interview. Composite construction proceeded from the 'default' composite, the first face available when PROfit is started, or the 'initial' composite, obtained by initially setting PROfit's feature descriptors. Forty-eight composites were constructed, one per witness, with 8 each being made in the 6 cells of the design.

The quality of the composites was initially assessed by spontaneous naming, though this yielded no correct names. In the next part, fresh participants matched the composites to their target photographs. Performance on this sorting task was much better at 58.1% correct, but indicated that the composites were affected by neither interview type nor starting face. We next administered an identification task, requiring further participants to select from a list of written names. This was a more difficult exercise to complete (43.2% correct) but suggested benefit for the administration of an interview (rather than no interview). Finally, given evidence that participants were basing responses on hairstyle, the identification task was re-run with the external facial features removed. The data indicated that the best composites emerged from the use of a Holistic Interview when used in conjunction with an initial composite.

In spite of our attempt to improve composite quality, naming remained very poor, a result observed elsewhere after a similar retention interval (Frowd et al., 2004a, 2004b, 2004d; Koehn and Fisher, 1997). In real life, additional information supplied with the composite – e.g. the *Modus Operandi* – may serve to improve identification. Our alternatives to un-cued naming proved interesting. The sort task is believed to be a general measure of feature quality, as participants tend to compare facial features between targets and composites (e.g. Frowd et al., 2004b, 2004c), but this task suggested that the composites contained similar features across interviews and construction treatments. Identification tasks were run as a better measure, since the names of the celebrities themselves could serve as a cue to promote familiar face processing and thereby recognition. The first identification task suggested that more recognisable composites emerged when either of the 2 interviews were employed (cf. no interview). Such a finding fits with previous research reporting a benefit for a Cognitive Interview (e.g. Luu and Geiselman, 1993) and our Holistic Interview (Frowd et al., 2004a). However, it turned out that participants were influenced by hair, thus weakening the task as a measure of familiar face perception. Removal of the external facial features suggested that the most identifiable composites emerged from a Holistic Interview and initial face construction. Also, there was no difference between a Cognitive Interview and no interview, nor between composites constructed after the default or initial face (these data suggest that there did not appear to

be a problem with an operator interpreting a witness's description and preparing an initial composite).

The current study therefore replicates the basic finding of Frowd et al. (2004a), demonstrating a benefit for a Holistic Interview with composite construction. This is in spite of several differences between the 2 studies: their work involved immediate construction and a sort task (their targets were not famous). These 2 studies suggest that, relative to a Cognitive Interview, the Holistic Interview can produce more identifiable composites irrespective of retention interval. We also note that this effect has been replicated a second time now, in a study involving immediate construction and the FACES composite system (Frowd et al., in prep).

For the (internal feature) identification task, one might ask why there was a benefit for a Holistic Interview only in the initial face condition? Composites in the Holistic Interview / default face condition were worse than the baseline (Cognitive Interview / initial face). It would appear that the attribution of a number of personality traits in the Holistic Interview is appropriately enhancing the recognition ability of a witness during construction. However, it is likely that the default face condition may be interfering with a witness's memory to a greater extent than the initial face condition, since relatively more faces were presented using the 'default' procedure. Recall that, in the default condition, each witness initially inspected up to 2 dozen examples per feature, starting from the default face, to establish a general 'description' of features. This information would be used to set the feature descriptors and another 2 dozen examples would be shown (to identify the best). In contrast, construction by initial composites proceeded with setting the feature descriptors, thereby presenting fewer faces.

The Holistic Interview involves a simple procedure, one we believe could be used with 'real' witnesses: in place of describing a face, witness make 7 personality traits in the form of ratings on a 7-point Likert scale (they also form a mental image of their 'assailant' prior to each judgement). The composite operator then directs witnesses to the composite system's feature descriptors and the most appropriate labels for each feature are selected (normally, the operator would set these directly from the witness's verbal description). An 'initial' composite is then displayed containing the features that match the description, and the witness follows the normal construction procedure (selecting the best features, adjusting the distance between features, etc.)

List of figures

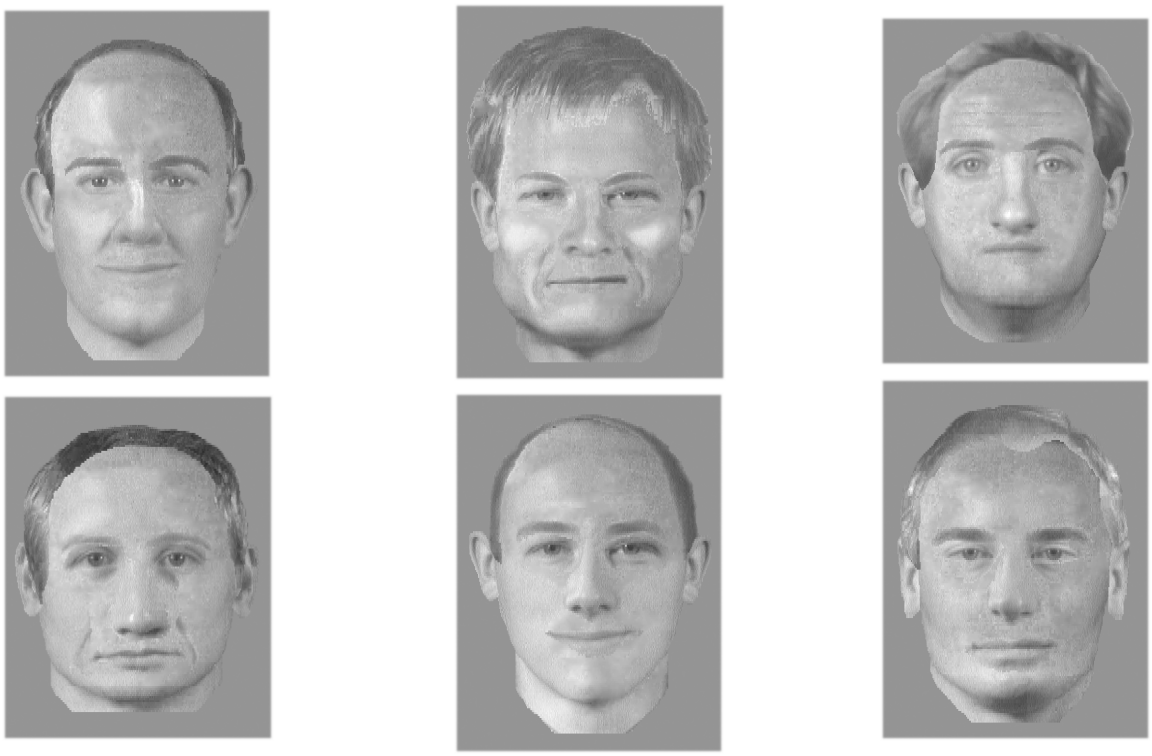

Figure 1. Example composites constructed by witnesses in the study. They are of the UK politician William Hague.

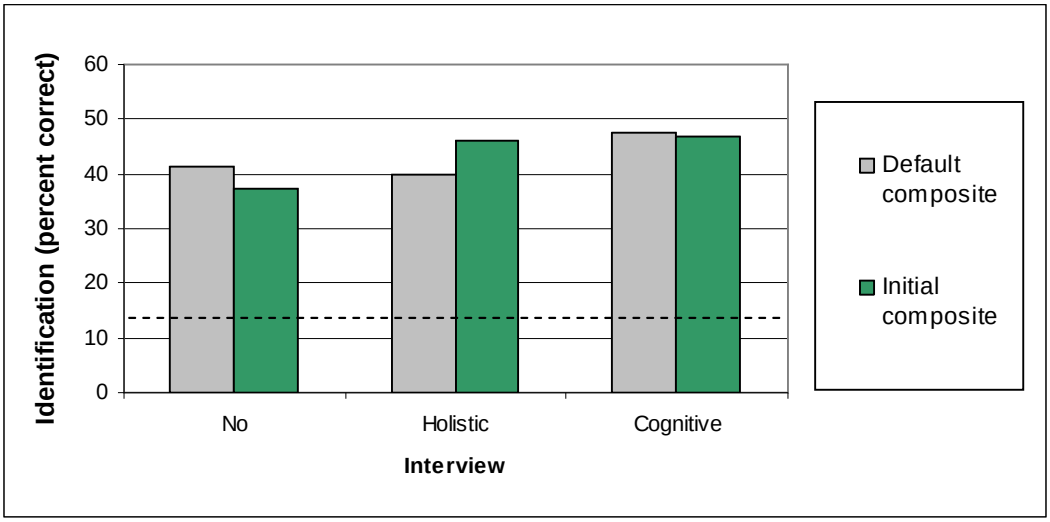

Figure 2. The quality of composites as assessed by the first Identification task (complete composites).

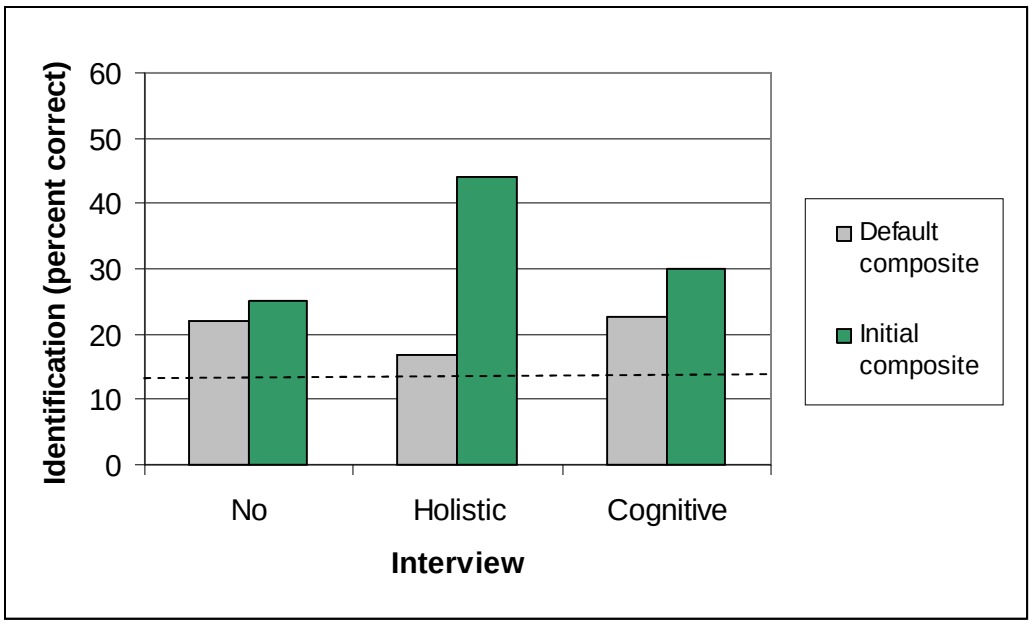

Figure 3. Internal feature quality measured by an Identification task. An appreciable benefit emerged for the Holistic Interview when used in conjunction with initial face construction.

References

Brace, N., Pike, G., and Kemp, R., Investigating E-FIT using famous faces, (in:) Forensic Psychology and Law, A. Czerederecka, T. Jaskiewicz-Obydzinska and J. Wojcikiewicz (Eds.), Krakow: Institute of Forensic Research Publishers 2000, pp. 272-276.

Bruce, V., Ness, H., Hancock, P., Newman, C., and Rarity, J., Four heads are better than one: Combining face composites yields improvements in face likeness, Journal of Applied Psychology 2002, vol. 87, pp. 894-902.

Bruce, V. and Young, A., The eye of the beholder, Oxford University Press 1998.

Davies, G., Forensic face recall: the role of visual and verbal information, (in:) Evaluating witness evidence, S. Lloyd-Bostock and B. Clifford (Eds.). John Wiley and Sons Ltd. 1983, Chapter 6.

Davies, G., Ellis, H., and Shepherd, J., Face identification: The influence of delay upon accuracy of photofit construction, Journal of Police Science and Administration 1978, vol. 6, pp. 35-42.

Davies, G., van der Willik, P. and Morrison L., Facial composite production: a comparison of mechanical and computer-driven systems, Journal of Applied Psychology 2000, vol. 85, pp. 119-124.

Frowd, C., Bruce, V., Ness, H., Thomson-Bogner, C., Peterson, J., Mcintyre, A. and Hancock, P., Parallel approaches to composite production, Psychology Crime and Law 2004a, under review.

Frowd, C., Carson, D., Ness, H., McQuiston, D., Richardson, J., Baldwin, H., and Hancock, P., Contemporary Composite Techniques: the impact of a forensically-relevant target delay, Legal and Criminological Psychology 2004b, in press.

Frowd, C., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S. and Hancock, P., A forensically valid comparison of facial composite systems, Psychology, Crime and Law 2004c, in press.

Frowd, C., Hancock, P., and Carson, D., EvoFIT: A Holistic, Evolutionary Facial Imaging Technique for Creating Composites, ACM Transactions on Applied Perception (TAP) 2004d, vol. 1, pp. 1-21.

Frowd, C., McQuiston-Surrett, D., Kirkland, I., and Hancock, P., Target and process factors in facial composite production, Journal of Applied Psychology, in prep.

Geiselman, R., Fisher, R., MacKinnon, D. and Holland, H., Eyewitness memory enhancement with the cognitive interview, American Journal of Psychology 1986, vol. 99, pp. 385-401.

Koehn, C., and Fisher R., Constructing facial composites with the Mac-a-Mug Pro system, Psychology, Crime and Law 1997, vol. 3, pp. 215-224.

Luu, T. and Geiselman, R., Cognitive retrieval techniques and order of feature construction in the formation of composite facial images, Journal of Police and Criminal Psychology 1993, vol. 9, pp. 34-39.