Changing the face of criminal identification

Charlie Frowd*[1], Vicki Bruce[2] and Peter J.B. Hancock[3]

[1]Department of Psychology, University of Central Lancashire, Preston, UK
[2]College of Humanities and Social Science, University of Edinburgh, Edinburgh, UK
[3]Department of Psychology, University of Stirling, Stirling, UK

*Corresponding author: Dr Charlie Frowd, Department of Psychology, University of Central Lancashire, Preston PR1 2HE. Email: cfrowd@uclan.ac.uk. Phone: (01772) 893439.

Keywords: facial composite, witness, cognitive interview, face recognition

Acknowledgement
The research was supported by four EPSRC research grants (GR/L88627, GR/N09701, GR/R86034 and EP/C522893).

Imagine you are a witness to a crime: you saw a young man running from a bank; it all happened very quickly, but you were able to have a good look at his face as he removed his balaclava. Would you be able to describe his face? There might not be useful CCTV footage from the bank, because of the mask, so would you also be able to make a recognisable image of his face? Research suggests not. However, a 10 year programme of research at the Universities of Stirling and Central Lancashire is changing this.

There are currently two main systems that the police use to construct 'facial composites' of criminal suspects. (Later, we will mention a third kind of system, under development.) First, there are forensic artists. These people have skills in portraiture and use pencils or crayons to draw the face by hand. Secondly, there are software systems. The UK has two, E-FIT and PRO-fit, and each contains a selection of ready made parts to build a face. For both artists and these software systems witness select individual facial features – a hairstyle, face shape, eyes, brows, nose, etc. The result is a *composite* of facial parts. See Figure 1 for examples.

Considerable attention to detail is needed to evaluate the efficacy of a composite system. This is described in our 'gold standard' procedure (Frowd et al., 2005b). In brief, about 10 participants would first be shown an unfamiliar target face. Sometime later, these 'witnesses' would describe the face to an artist or computer technician, who would use cognitive interviewing techniques to help them recall, and construct the best possible composite. Technicians must have expertise with the composite system plus good artistic skills to enhance the face (e.g. addition of stubble and wrinkles). Other people who know the targets would attempt to name the composites.

Using this procedure, when the delay to construction is brief (no more than a few hours), E-FIT and PRO-fit composites are typically named about 20 percent of the time on average (e.g. Davies et al., 2000; Frowd et al., 2005b, 2007b). After two days, the minimum for real witnesses, correct naming is normally just a few per cent correct! (e.g. Frowd et al., 2005a, 2007c.) For artist-composites, it is about 10 percent and independent of delay (Frowd et al., 2005a, 2005b). These data suggest that identification is unlikely from composites made using current procedures and systems.

Figure 1. Some of the more recognisable celebrity composites produced in our studies. Each face was made by a different person from their memory. Those on the top two rows were from the E-FIT and PRO-fit ‘feature’ systems; on the third row, from our EvoFIT ‘evolving’ system (described later); and bottom row, from a sketch artist. Can you guess the identities of the famous people? The answers are listed at the end of the article.

The difficulty of constructing faces from our memory has been known for over 30 years (e.g. Davies, 1978). Essentially, we are not good at the tasks required: describing and selecting individual facial features. Instead, we process faces 'holistically', more as a complete entity (e.g. Young et al., 1987). For example, the perception of facial features change in the presence of other features (e.g. Tanaka & Farah, 1993), and so the features and their position on the face are both important. Modern composite systems apply this idea to some extent since witnesses choose facial features in the context of a complete face.

As part of ongoing research, we have designed and evaluated improvements to each stage in the process: to the interview, to the system and to the presentation. As discussed below, one research thread has successfully enhanced the interview; another, provided an alternative system; and a third, caricatured the face at presentation. In time, some of these techniques may also help witnesses elsewhere, for example when they try to identify a criminal from a line-up.

Improving the interview

Berman and Cutler (1989) found that participants recognise a face much better after they made a few personality judgements about it – such as intelligence and attractiveness. If the principle extends to witnesses, then their ability to select facial features may also improve. In Frowd et al. (2005c, 2007c), instead of describing a face, participants made seven trait judgements of a target prior to composite construction. Example traits included intelligence, friendliness and arrogance. We found, as predicted, that trait attribution improved the quality of the composites.

Witnesses normally receive a Cognitive Interview (CI), however, to help them recall details of an event and a suspect's face. The 'revised' Cognitive Interview, as used currently by practitioners, is quite an involved procedure – see Wells et al., 2006 for a review – and includes: rapport building, to help a witness relax; context reinstatement, to assist recall; free recall, whereby a witness freely describes people and events; and cued recall, to allow clarification and elaboration of specific details. The revised interview also allows for repeated recall from different perspectives and from different temporal orders. Not that for composite construction, the interview is important for locating a subset of facial features within a composite system (or via catalogues of features for sketch artists).

Our solution was for witnesses to first receive a CI, as above, followed by a recognition-enhancing 'holistic' interview. The latter involves them thinking about the personality of the face, which is somewhat like the free recall stage of the CI. Witnesses then make seven holistic judgements by rating each on a 3 point Likert scale (low / medium / high). In an evaluation by Frowd et al. (under revision), composites constructed after the combined interview were much better those than after the CI alone; Figure 2 presents a few examples from the study.

Police composite operators in the UK are now being trained on the combined interview.

Figure 2. Composites constructed of the Eastenders' actor Perry Fenwick (stage name Billy Mitchell) and actress June Brown (Dot Branning). The left image was made after a Cognitive Interview, the right after a Holistic plus Cognitive Interview.

Enhancing an existing system

Modern computerised 'feature' systems can produce very good quality composites. In 2000, for example, Graham Davies and his colleagues found an average naming level of 49% from participants constructing E-FIT composites. Unfortunately, in order to achieve this performance, participants were both familiar with the target and had a photograph available to refer to. While the result does not mirror real life, it does suggest that feature selection might be improved were witnesses to process the face as if familiar.

Considerable research has shown that we process familiar and unfamiliar faces differently (e.g. Ellis et al., 1979; Young et al., 1985). For familiar faces, the *internal features* are the most important: the region containing the eyes, brows, nose and mouth. For faces seen a few times, or just once, the *external features* are more salient, especially the hair and face shape, and these tend to dominate our perception. Consequently, we may misidentify an unfamiliar person when their hairstyle changes. Frowd et al. (2007b) have also found that the external features of facial composites are better likenesses of a target than internal features.

The perceptual impact of the external features can be reduced by processing them with a Gaussian filter – see Figure 3. This 'blurring' technique allows the inner face to appear more prominent whilst maintaining a complete face context (important for holistic face processing). The procedure does appear to help participants select facial features: in two small-scale studies, we found that the quality of the *internal features* was significantly better in composites made with blurring than without. Ongoing research is exploring the potential of the technique in a larger, more realistic composite study, as well as a general aid to unfamiliar face recognition.

Figure 3. The application of a Gaussian (blur) filter to the external facial features. Witnesses first select a hairstyle (left), which is then blurred (right) while the remaining features are chosen. Notice how the inner part of the composite face appears more prominent in the right image.

An alternative system

As noted above, a detailed description of the face is a normal prerequisite to composite construction. Even with a good view of a criminal's face and cognitive interviewing techniques, many witnesses are unable to provide a satisfactory description. Sadly, these witnesses may be denied the opportunity of constructing a composite, in spite of feeling confident that they could recognise the person in future.

Several research labs are designing composite systems that are based more on recognition than recall (Gibson et al., 2003; Tredoux et al., 2006). Ours is called EvoFIT and presents sets of whole faces, 18 per screen (Frowd et al., 2004). Witnesses select faces that look something like the criminal's and EvoFIT 'breeds' them together to produce another set. While the initial faces have random characteristics, repeating the selection and breeding procedure a few times normally allows a good likeness to be 'evolved' – see Figure 4. In practice, the process is improved by first choosing facial shapes, then facial textures.

Figure 4. Evolving the face of footballer, David Beckham. Images are example likenesses produced from EvoFIT at the end of generations 1, 2 and 3 respectively. The image from generation 3 was used as the facial composite.

At the heart of EvoFIT is a ‘face model’ that can generate a very large number of synthetic, but realistic-looking faces; the model is built from the statistics of about 70 complete faces. A Genetic Algorithm is used to search the space of possible faces, but converging on a good likeness is sometimes difficult. We now to ask users to select the best match of shape and texture at the end of each generation, since this combined face can be given a greater emphasis during breeding to accelerate the search.

About this time, EvoFIT was used in a criminal investigation (Frowd et al., 2006c). As Figure 5 illustrates, a very good likeness was produced of the criminal.

Figure 5. The Sketch, EvoFIT and E-FIT composites constructed in the ‘Beast of Bozeat’ police investigation. The case involves a series of sexual assaults carried out in Southern England in the late 1990s. Far right, a photograph of the person convicted of this offence.

As part of a recent two year EPSRC grant, four new white male face models were designed, each covering a different age range and spaced apart by about 10 years. This was done since judgements of age are fairly accurate (e.g. Bruce & Young, 1986), enabling faces to be evolved with roughly the correct age. Thus, witnesses now have the fairly easy task of estimating the age of the criminal, to load the appropriate model. Even so, the models are still fairly general and EvoFITs constructed using the above gold standard procedure and a two day delay are named at only 10-15 percent correct – although this is still better than the feature systems used without the above enhancements (Frowd et al., 2007c). The development of EvoFIT is on-going and plans are in place for the construction and evaluation of white female and models of other races.

A general problem of evolving systems is what to do when the search does not converge on the correct region of face space? One way is to simply evolve again using a fresh set of initial faces, a procedure which appears to be quite effective (Frowd et al., 2006b). There is however the associated risk that increasing the number of faces may interfere with a witness's memory, of which there is some evidence for using a popular US system called FACES (Wells et al., 2005). Our evidence is that the construction of a single EvoFIT does not appear to interfere with a user's memory any differently (if at all) from the construction of an E-FIT or PRO-fit, although limiting the number of faces presented is no doubt sensible.

More recent work has built models which match the target on age plus a few distinctive features recalled by a witness. This 'tailoring' approach avoids generating inappropriate characteristics – e.g. wide faces when narrow is required – and the final faces are much more identifiable (e.g. Frowd et al., 2007a).

In spite of such improvements, the apparent age and other 'holistic' aspects of an EvoFIT can be sometimes inaccurate. A set of 'holistic' scales (tools) were therefore designed to allow witnesses to improve the likeness of their evolved face. There are eight scales in total – including age, facial weight and masculinity – and each changes the face along the relevant dimension. The tools can be quite effective, as Figure 6 illustrates (Frowd et al., 2006a).

age	10%
facial weight	-40%
attractiveness	-60%
extroversion	100%
health	20%
honesty	-100%
masculinity	80%
threatening	80%

Figure 6. EvoFIT's holistic tools for face enhancement. In this example, a composite was evolved from memory (left) and then re-worked using the tools (centre). Slider settings for each holistic transform are also presented (right). Can you name the TV celebrity? The answer is at the end.

A version of EvoFIT with blurring, white male face models, holistic tools and caricature is currently being used by the UK police.

Enhancing an existing composite

The above discussion suggests that improvements can be made to both interview and system, but can anything be done to once a composite has been constructed? Research suggests two things: composites can be morphed together, or caricatured.

It is sometimes the case that there is more than one witness to a particular crime. When this happens, the police normally give different tasks to different witnesses, and one person constructs a composite. Asking several witnesses to do this produces very different looking faces. As Figure 7 illustrates, one may even question whether everyone was making the same face?! Currently, there is no reliable test to predict who will produce the best composite; so, all other factors being equal, which witness should the police choose? Our answer is all of them!

Figure 7. Top row are individual composites made by four different laboratory-witnesses using PRO-fit. Below, the morph, which Bruce et al. found to be a better representation of a target (bottom right) than an 'average' composite.

In 2002, Bruce et al. asked groups of four laboratory-witnesses to each construct a single composite of a target. They combined the images together by averaging to produce a *morphed* composite. Bruce et al.'s data suggested that the morph was better at conveying identity than a typical composite produced by an individual witness, and always at least as good as the best individual witness composite. Hayley Ness has shown that a morphed composite can still be effective when composed of faces from different systems (Ness, 2003); a similar result was also found for composites made of the Beast of Bozeat described in Figure 5 (Frowd et al., 2006c).

A morphed image is effective because the consistent parts of the individual composites tend be reinforced, and errors averaged out. The work has prompted a change in police policy to permit construction of multiple composites (of the same face) for the production of a morph.

An alternative is to ask a single witness to construct more than one face (as mentioned above for EvoFIT.) While the norm is to construct at a front-view, there is evidence that unfamiliar faces are processed better in half profile (e.g. Bruce et al., 1987). Thus, Ness et al. (in press) asked participants to construct composites in both front and three-quarter view using an enhanced version of PRO-fit. The three-quarter view images were sometimes better, but best performance was found when trying to guess identity with both views visible.

Another approach can be used when only a single composite is available. This is based on the observation that composites often appear fairly bland, but making them more different to each other may increase identification. We tried this by exaggerating the feature shapes of a composite from an average face to produce a *caricature*. Results indicated that while one level of caricature helped one person to identify a face, for another, a different level was required. The solution was to present a range of exaggerations!

Figure 8. An example caricature sequence. The faces ranging from a 50% negative caricature (far left) to the original composite (middle) to a 50% positive caricature (far right). A sequence of 21 such images (the above plus intermediate states) was found to increase correct naming, in this example from 40% to 80%.

In Frowd et al. (2007d), we presented participants with sequences of images, ranging from a strong positive caricature to a strong negative one, where the feature shapes were deemphasised to look more like the average. Refer to Figure 8 for an example. As shown in Figure 9, three different types of system benefited from the presentation format. We also found that the best gain occurred for the worst quality composites, and therefore should maximally help composites produced in the field. Presenting the caricature sequence in the form of an animated GIF would appear appropriate for TV crime programs; for an example, see https://dspace.stir.ac.uk/dspace/bitstream/1893/720/2/AnimatedBlair.gif.

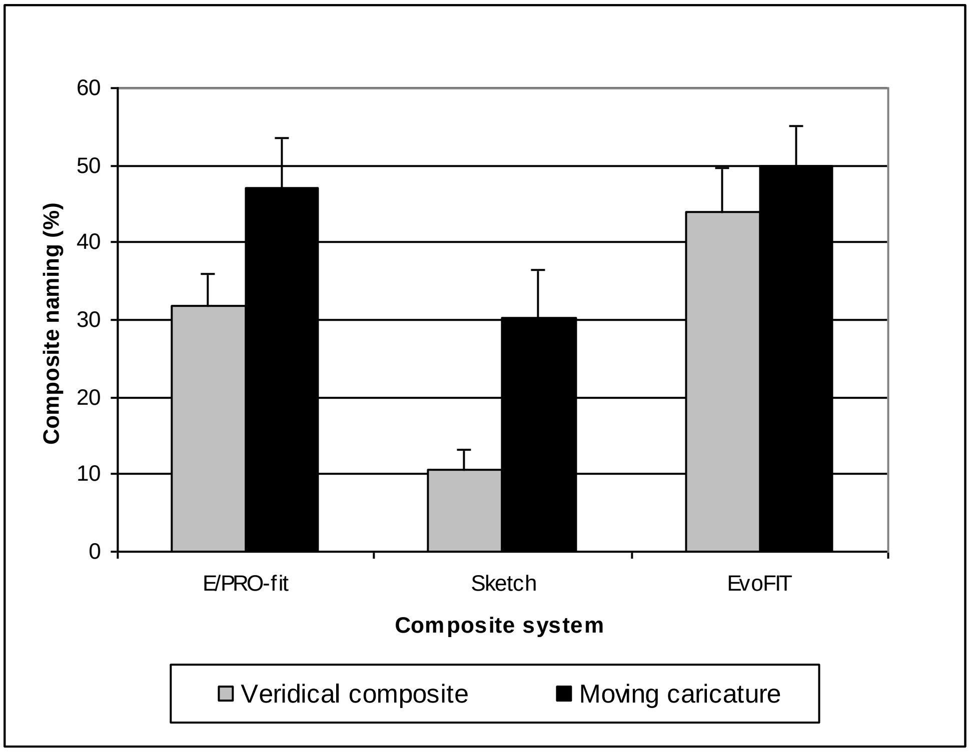

Figure 9. The benefit found by Frowd et al. (2007d) for presenting an animated caricature of composites constructed using the E-FIT and PRO-fit (E/PRO-fit), Sketch and EvoFIT methods.

Concluding remarks: putting it all together

There is good evidence that composites are not very identifiable when produced using standard police procedures and systems. We have found that significant improvements can be made by changing the interview, system and/or presentation format. Current work is looking at which combination of above techniques works the best – which is likely to be blurring, holistic tools, EvoFIT and caricature – and whether any of these could also help witnesses carry out other identification tasks.

Answers

Figure 1: The composites are (from left to right, top to bottom) of Brad Pitt, Graham Norton, Nicholas Cage, Michael Owen, Robbie Williams, Anthony (Ant) McPartlin, David Beckham and Noel Gallagher.
Figure 6. TV celebrity, Simon Cowell.
Figure 8: The former British Prime minister, Tony Blair.

References

Berman, G.L., & Cutler, B.L. (1998). The influence of processing instructions at encoding and retrieval on face recognition accuracy. *Psychology, Crime & Law*, 4, 89-106.

Bruce, V., Ness, H., Hancock, P.J.B, Newman, C., & Rarity, J. (2002). Four heads are better than one. Combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, 87, 894-902.
Bruce, V., Valentine, T., & Baddeley, A.D. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, 1, 109-120.
Bruce, V., & Young, A.W. (1986). Understanding face recognition. *British Journal of Psychology*, 77, 305-327.
Davies, G.M. (1978). Face recognition: issues and theories. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.). *Practical aspects of memory*. New York: Academic Press.
Davies, G.M., van der Willik, P., & Morrison, L.J. (2000). Facial Composite Production: A Comparison of Mechanical and Computer-Driven Systems. *Journal of Applied Psychology*, 85, 119-124.
Ellis, H.D., Shepherd, J., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception*, 8, 431-439.
Frowd, C.D., Bruce, V., Gannon, C., Robinson, M., Tredoux, C., Park., J., McIntyre, A., & Hancock, P.J.B. (2007a). Evolving the face of a criminal: how to search a face space more effectively. Paper to appear in the conference proceedings of BLISS 2007, Edinburgh.
Frowd, C.D., Bruce, V., McIntyre, A. & Hancock, P.J.B. (2007b). The relative importance of external and internal features of facial composites. *British Journal of Psychology*, 98, 61-77.
Frowd, C.D., Bruce, V., McIntyre, A., Ross, D., Fields, S., Plenderleith, Y., & Hancock, P.J.B (2006a). Implementing Holistic Dimensions for a Facial Composite Syst*em. Journal of Multimedia*, 1, 42-51.
Frowd, C.D., Bruce, V., Ness, H., Bowie, L., Thomson-Bogner, C., Paterson, J., McIntyre, A. & Hancock, P.J.B. (2007c). Parallel approaches to composite production. *Ergonomics*, 50, 562-585.
Frowd, C.D., Bruce, V., Plenderleith, Y., & Hancock, P.J.B. (2006b). Improving target identification using pairs of composite faces constructed by the same person. *IEE Conference on Crime and Security*, (pp. 386-395). London:IET.
Frowd, C.D., Bruce, V., Ross, D., McIntyre, A. & Hancock, P.J.B. (2007d). An application of caricature: how to improve the recognition of facial composites. *Visual Cognition*, 15, 1-31.
Frowd, C.D., Bruce, V., Smith, A., & Hancock, P.J.B (under revision). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*.
Frowd, C.D., Bruce, V., Storås, K., Spick, P. & Hancock, P.J.B. (2006c). *An evaluation of morphed composites constructed in a criminal investigation*. Proceedings of the 16th Conference of the European Association of Psychology and Law (pp. 59-66). London: IP-PA Publishing.
Frowd, C.D., Carson, D., Ness, H., McQuiston, D., Richardson, J., Baldwin, H., & Hancock, P.J.B. (2005a). Contemporary Composite Techniques: the impact of a forensically-relevant target delay. *Legal & Criminological Psychology*, 10, 63-81.

Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S. & Hancock, P.J.B. (2005b). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, 11, 33-52.
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 (TAP)*, 1, 1-21.
Frowd, C.D., McQuiston-Surrett, D., Kirkland, I. & Hancock, P.J.B. (2005c). The process of facial composite production. In A. Czerederecka, T. Jaskiewicz-Obydzinska, R. Roesch & J. Wojcikiewicz (Eds.). *Forensic Psychology and Law* (pp. 140-152). Krakow: Institute of Forensic Research Publishers.
Gibson., S.J., Solomon, C.J., & Pallares-Bejarano, A. (2003). *Synthesis of photographic quality facial composites using evolutionary algorithms*. In R. Harvey and J.A. Bangham (Eds.) Proceedings of the British Machine Vision Conference (pp. 221-230).
Ness, H. (2003). *Improving facial composites produced by eyewitnesses.* Unpublished PhD thesis, University of Stirling.
Ness, H., Hancock, P.J.B., Bowie, L., & Bruce, V. (in press). Are two views better than one? A study examining recognition of three-quarter view and full-face composites. *Applied Cognitive Psychology*.
Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition, *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 46A, 225-245.
Tredoux, C.G., Nunez, D.T., Oxtoby, O., & Prag, B. (2006). An evaluation of ID: an eigenface based construction system. *South African Computer Journal*, 37, 1-9.
Young, A.W., Hay, D.C., McWeeny, K.H., Flude, B.M., & Ellis, A.W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, 14, 737-746.
Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configural information in face perception, *Perception*, 16, 747-759.
Wells, G.L., Charman, S.D., & Olson, E.A. (2005). Building face composites can harm line-up identification performance. *Journal of Experimental Psychology: Applied*, 11, 147-157.
Wells, G., Memon, A., & Penrod, S.D. (2007). Eyewitness evidence: improving its probative value. *Psychological sciences in the public interest*, 7, 45-75.