Changing the face of criminal identification

Charlie Frowd, Vicki Bruce and **Peter J.B. Hancock** describe the latest techniques to construct the face of a criminal

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? Until recently the answer to both questions would probably have been 'no'. But now a 10-year programme of research at the Universities of Stirling, Edinburgh and Central Lancashire is changing things.

This article describes this ongoing work, which designs and evaluates improvements to each stage in the process: to the interview, to the system and to the presentation. The police in the UK are now using these techniques.

questions

How well can you describe someone's face?

Could you select parts of a face to make a recognisable whole?

The difficulty of constructing faces from our memory has been known for over 30 years (e.g. Davies, 1978). 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 changes 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 facial composite systems, where witnesses choose individual features in the context of a complete face, apply this idea to some extent.

There are currently two established systems that the police use to construct 'facial composites' of criminal suspects. (Later, we will mention a third kind of system that is currently being introduced.) Firstly, 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 packages, of which the UK has two, namely E-FIT and PRO-fit, which offer a selection of ready-made parts to build a face. In both the artist and computerised systems witnesses select individual facial features – 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, Carson et al., 2005a). 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 per cent of the time on average (e.g. Davies et al., 2000; Frowd, Bruce, McIntyre et al., 2007; Frowd, Carson et al., 2005a). After two days – which is typically the minimum for real witnesses – correct naming is normally just a few per cent correct! (e.g. Frowd, Bruce, Ness et al., 2007; Frowd, Carson et al., 2005b). For artist-composites, it is about 10 per cent and independent of delay (Frowd, Carson et al., 2005a, 2005b). These data suggest that identification is unlikely from composites made using current procedures and systems.

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 has provided an alternative system; a third has 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 have 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 two of our experiments (Frowd, Bruce,

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. et al. (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. et al. (2007). Evolving the face of a criminal. Paper to appear in the conference proceedings of BLISS 2007, Edinburgh.

Frowd, C.D., Bruce, V., McIntyre, A. et al. (2006). Implementing holistic dimensions for a facial composite
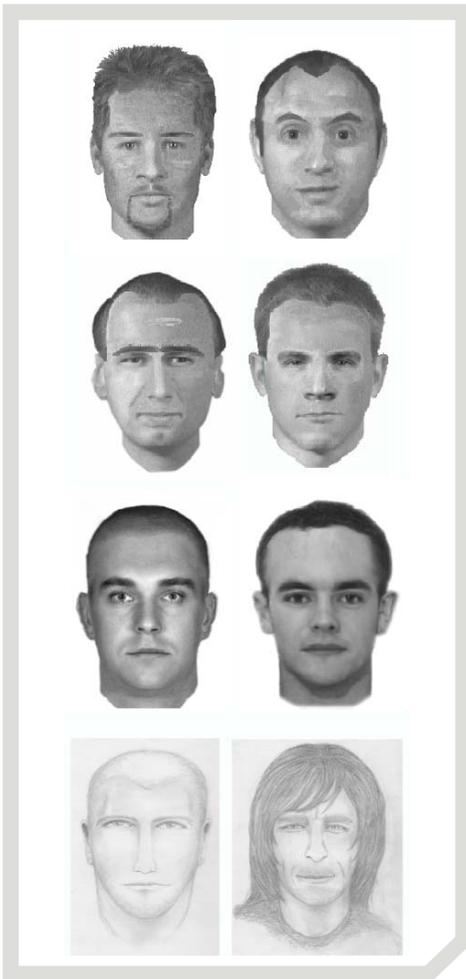
Figure 1. Celebrity composites, each made by a different person from memory. Those on the top two rows were from the E-FIT and PRO-fit 'feature' systems; on the third row, from our EvoFIT 'evolving' system (described later); and bottom row, from a sketch artist. Can you guess the identities of the famous people? The answers are listed at the end of the article.

Ness et al., 2007; Frowd, McQuiston-Surrett et al., 2005), 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., 2007 for a review), including: 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. This interview is important for locating a subset of facial features within a composite system, or via catalogues of features for sketch artists.

We thought to combine this CI with a recognition-enhancing 'holistic' interview. The latter involves witnesses thinking about the personality of the face, which is somewhat like the free recall stage of the CI. They then make seven personality judgements by rating each on a three-point Likert scale (low/medium/high). In an evaluation by Frowd, Bruce, Smith et al. (in press), composites constructed after the combined interview were much better than those after the CI alone. Police composite operators in the UK are now being trained on the combined 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
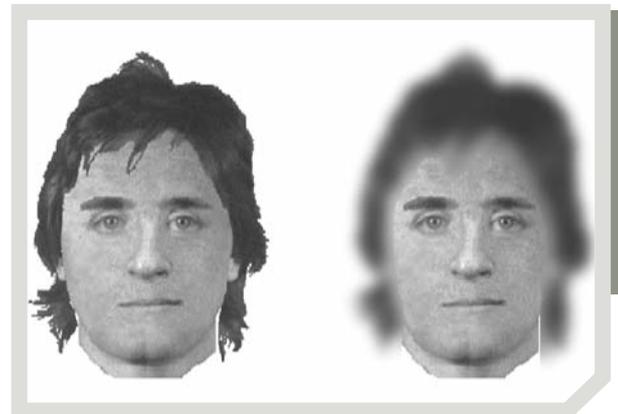

Figure 2. Applying a Gaussian (blur) filter to the external facial features. Witnesses select a hairstyle, which is then blurred while the remaining features are chosen.

an average naming level of 49 per cent from participants constructing E-FIT composites. Unfortunately, in order to achieve this level of 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 if witnesses processed 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, Bruce, McIntyre et al. (2007) have also found that the external features of facial composites are better likenesses of a target than the internal features.

It is possible, however, to lessen the perceptual impact of the external features by processing them with a Gaussian filter. This 'blurring' technique (see Figure 2) allows the inner face to appear more

system. *Journal of Multimedia, 1*, 42–51.

Frowd, C.D., Bruce, V., McIntyre, A. & Hancock, P.J.B. (2007). The relative importance of external and internal features of facial composites. *British Journal of Psychology, 98*, 61–77.

Frowd, C.D., Bruce, V., Ness, H. et al. (2007). Parallel approaches to composite production. *Ergonomics, 50*, 562–585.

Frowd, C.D., Bruce, V., Plenderleith, Y. & Hancock, P.J.B. (2006). 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. et al. (2007). 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. (in press). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*.

Frowd, C.D., Bruce, V., Storås, K. et al. (2006). 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. et al. (2005a). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law, 11*, 33–52.

Frowd, C.D., Carson, D., Ness, H. et al. (2005b). Contemporary composite techniques: The impact of a forensically-relevant target delay.

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 (e.g. Frowd, Park et al., in press). 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.

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, Hancock 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 3. In practice, the process is improved by first choosing facial shapes, then facial textures.

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 ask users to

Figure 3. 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.

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.

At this stage in development EvoFIT was used in a criminal investigation, the 'Beast of Bozeat' case (Frowd, Bruce, Storås et al., 2006). As the image on the contents page of this issue illustrates, a good likeness was produced of the criminal.

As part of a recent project, four new white male face models were designed, each covering a different age range and spaced apart by about 10 years. Judgements of age are fairly accurate (e.g. Bruce & Young, 1986), so this enables 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.

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 that appears to be quite effective (Frowd, Bruce, Plenderleith et al., 2006).

Figure 4. An EvoFIT constructed in Lancashire during the pilot scheme. The case involved an indecent assault in Blackpool. The EvoFIT was identified by a member of the public, which led to the arrest and conviction of the person shown on the right.

Legal & Criminological Psychology, 10, 63–81.

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. (2005). 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.

Frowd, C.D., Park, J. et al. (in press). *Effecting an improvement to the fitness function: How to evolve a more identifiable face.* To appear in 2008 ECSIS Symposium on Bio-inspired,

Learning, and Intelligent Systems for Security. (Edinburgh).

Gibson., S.J., Solomon, C.J. & Pallares-Bejarano, A. (2003). Synthesis of photographic quality facial composites using evolutionary algorithms. In R. Harvey & 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? *Applied Cognitive Psychology.*

Tanaka, J.W. & Farah, M.J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: Human Experimental*

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 from the use of 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 that 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, Bruce, Gannon et al., 2007).

In spite of such improvements, the apparent age and other 'holistic' aspects of an evolved face can sometimes be inaccurate. A set of 'holistic' scales (tools) was therefore designed to allow witnesses to improve the likeness of their evolved face. There are 10 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 5 shows (Frowd, Bruce, McIntyre et al., 2006).

Developments to EvoFIT have largely been successful, with composite naming increasing from only a few percent correct (Frowd, Carson et al., 2005b) to 10–15 percent (Frowd, Bruce, Ness et al., 2007) after a two-day delay. The most recent version of EvoFIT includes blur, holistic tools and an improved control of the facial aspect ratio (facial width to height). This has just been evaluated using the gold standard and a two-day delay, and composites were correctly named at 25 per cent, compared to 5 per cent for a traditional 'feature' system.

EvoFIT was recently involved in a six-month pilot scheme by Lancashire police. About 30 composites were constructed and the system was reported to be of value in 20 per cent of cases. (An example is shown in Figure 4.)

The development of EvoFIT is clearly ongoing. The system can now evolve white female faces, and models of other races are currently being built and evaluated.

Enhancing an existing composite

The above discussion suggests that improvements can be made to both interview and system, but can anything be done 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. 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!

In 2002 Bruce and colleagues asked groups of four laboratory-witnesses to each construct a single composite of a target. They combined the images by averaging to produce a morphed composite. The 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 in the Beast of Bozeat case (Frowd, Bruce, Storås et

al., 2006). 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-

Figure 5. 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.

AGE	10%
FACIAL WEIGHT	40%
ATTRACTIVENESS	-60%
EXTRAVERSION	100%
HEALTH	20%
HONESTY	-100%
MASCULINITY	80%
THREATENING	80%

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 from 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!

In Frowd, Bruce, Ross et al. (2007), 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 (see Figure 6). Whether using the E-FIT, PRO-fit, sketch or EvoFIT system, an animated composite led to significantly better naming than an

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. et al. (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.

Figure 6. An example caricature sequence. The faces range from a 50 per cent negative caricature (far left) to the original composite (middle) to a 50 per cent 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 per cent to 80 per cent.

unaltered (veridical) one. We also found that the best gain occurred for the worst-quality images, which suggests that the animation will be maximally beneficial for composites produced in the field. Presenting the caricature sequence in the form of an animated GIF would appear appropriate for TV crime programmes. This animated format is currently being used by several UK constabularies (examples from Lancashire constabulary can be found at tinyurl.com/5tpc74).

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 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 5. *Simon Cowell.*
Figure 6: *The former British Prime Minister, Tony Blair.*

I **Charlie Frowd**
   is in the School of Psychology, University of Central Lancashire cfrowd@uclan.ac.uk

I **Vicki Bruce**
   is in the School of Psychology, Newcastle University
I **Peter J.B. Hancock**
   is in the Department of Psychology, University of Stirling