Improving the identification of composite images using sketch-like faces
Danielle McKay (Psychology)

Introduction
Facial composites are used in forensic police work to create a visual likeness of a criminal’s face. They are produced by eye-witnesses, with the help of computer software or by a sketch artist. Constructing a facial composite of an offender is sometimes central to a police investigation when the criminal’s identity is unknown and there are no other immediate leads. Many different methods have been used over the years to create facial composites, these include: Identikit, Photofit, E-FIT, PRO-fit, Mac-A-Mug Pro and EvoFIT.

Arguably, the earliest technique is the sketch artist. These people are highly skilled in portraiture and work with eye-witnesses (witnesses and victims) using crayons or pencils to draw out the face (Laughery & Fowler 1980). Later system were devised from the 1970s for those with less artistic skills.

Identikit and Photofit are two of the early ‘mechanical’ systems. Identikit is a US sketch--like system. It uses transparent celluloid sheets that each contains a line drawing of a facial feature. A trained operative constructs a composite with a witness selecting appropriate features that are superimposed to make a face. Features available for selection include: hairlines, noses, mouths, chins and eyes. The police operator also has a marking pencil to make additional changes if needed (Green and Geiselman 1989). Photofit is similar, but this system uses photographic quality facial features to construct a likeness of the face.

Computer software programs have largely replaced the ‘mechanical’ systems: Identikit and Photofit (Davies et al. 2000; McQuiston-Surrett et al. 2006). In the UK, the police now use E-FIT and PRO-fit. These systems contain large databases of photographed individual features that can be selected, resized and positioned freely on the face. As for Photofit and Identikit, they make use of an artwork package to enhance the appearance of the face. Other systems exist worldwide and are similar to these; some of them still use sketch-like facial features; examples include Mac-a-Mug Pro and Identikit 2000.
Frowd et al. (2005a) conducted an experiment to compare the effectiveness of a range of composite systems, past and present. In this work, participants viewed a target face for one minute and two days later constructed a composite with a trained computer operator or a sketch artist. Evaluation of the composites was assessed by asking independent observers to name them. The naming of composites was found to be low, at 3% overall, but the sketch composites were best, at 8%. Also, in this study, sketch composites were more successful than Photofit and the other modern ‘feature’ methods (E-FIT and PRO-fit) tested. This suggests that even the most traditional method of using a sketch artist could produce better results than computer-based programs; it also supports other findings which that sketch-based programs can produce better results (e.g. Frowd et al., 2007). Other research has similarly found important deficits for computerized systems (e.g. Koehn and Fisher 1997).

The problem with all the above methods is that they require a person to synthesize the whole face from individual facial features. This may not be the way in which a person remembers or recognises a face (e.g. Davies and Christie 1982). Witnesses are required to concentrate on facial features when there is evidence that people perceive faces as complete entities. If a feature is changed on a face, such as a nose for example, the person may be able to say that the face looks different, but not be able to specify exactly what has changed. This can make it difficult for witnesses to know how to produce recognisable composites (e.g. Davies and Christie 1982; Frowd, Hancock and Carson, 2004).

Charlie Frowd and his colleagues in Psychology at UCLan (and previously at Stirling University) have been working on a new system to construct a face. The EvoFIT system presents witnesses with sets of complete faces, and a face is ‘evolved’ over time (Frowd et al. 2004). Essentially, they select from arrays of such faces and these are then bred together using a Genetic Algorithm to produce another set of faces for them to select. Repeating this process enables a face to emerge that normally has a very good likeness of the intended target. The system has been the subject of considerable research and development and a recent review may be found in Frowd et al. (2009). As EvoFIT requires users to select complete faces, it uses face recognition ability rather than face recall to construct a face. The process of recognition involves comparing
whether an image of a face/person is the same that has already been seen; recall involves verbalising information about facial appearance and selecting the features of a face. While the latter is a hard task for people to do, the former is much easier.

In the most recent evaluation, EvoFITs were correctly named with an average (mean) of 25% when constructed by people who were unfamiliar with the target faces, and after a two day delay, as is frequently the case with real witnesses. In comparison, PRO-fits constructed in the same way were correctly named at 5% (Frowd et al. 2009, under revision).

While EvoFIT has been shown to produce more identifiable faces than a ‘feature’ system, there is clearly room for improvement. One way to do this may be to simplify the appearance of the faces that are presented to users at construction. For example, using a sketch-like representation similar to that produced by artists, Mac-A-Mug and the Identikit (2000), may be beneficial. This is because the detail in a photographic-type face may be too great for a person to reproduce it accurately. With a simplified representation, there is also potentially less information to be incorrect. In support of this idea, composites from the sketch-like Identikit 2000 have been found to have better likenesses than those from the photographic PRO-fit (Frowd et al. 2007).

Frowd et al. (2008) built a sketch-type database for their EvoFIT system. An example face produced from this can be seen in Figure 1; also shown is an example produced from the normal photographic database. To produce the images in the sketched database, faces were processed using the ‘photocopy’ transform filter in Adobe Photoshop. The image transform maintains both the outline of the features as well as some of the shading information that is important for recognition (Bruce et al. 1992). In pilot work, to explore the potential of this technique, a set of famous faces were constructed by an experienced user using EvoFIT. These were then shown to 24 participants to name, given a list of suitable alternatives. The standard EvoFIT composites were recognised correctly 22.4% of the time, but the sketch-like composites were recognised better, at 32.8% (Frowd et al. 2008). Thus, this study found that simplifying the image ‘mode’ can improve recognition. But this experiment had little ecological validity: it used celebrity faces, had no delay between viewing the target face and constructing a composite, and gave people a list of names as part of the evaluation.
The present study tested whether changing the EvoFIT image mode (database) from the photographic to the sketched version would improve identification using a more realistic design than Frowd et al. (2008). It was hypothesised that the sketched version would be superior.

**Experiment**

There were two stages to the experiment. The first involved participants viewing a target face and constructing a composite; the second recruited further participants to name the resulting composites. The experiment was set up so that participants in the first stage would be unfamiliar with the target faces before viewing, as in real life, but in the second stage they would be familiar to them for naming.

**Overall design**

Photographs of staff from the Psychology department at UCLan were used as the target faces. The reason for this is that it would be fairly easy to recruit participants who had not seen Psychology staff before constructing the composites, and would therefore be unfamiliar with the face, as in police-work, whereas people who had seen them from Psychology would be available to later name them. A time delay of 22-26 hours was used for the participants viewing the target and constructing the face. This was done to mirror real life situations, where a delay of a day or two is typical. As participants constructed a single composite using one of the two EvoFIT databases (photographic / sketch), the design was between-subjects.

**Stage 1: Composite Construction**

**Participants**

Twenty participants were recruited on campus at UCLan using opportunity sampling to create the composites. These were non-Psychology students from UCLan, aged between 18 and 22 years. The mean age was 20.2 years, and the standard deviation (SD) was 1.0 years.
Materials
The target faces shown to the participants were greyscale photographs of Psychology Staff. The photographs were taken full face, and looking straight into the camera. There were five white males and five white females. None wore glasses. The males were clean-shaven. Each photograph was printed in colour, one per page. EvoFIT software was used to construct the composites.

Procedure
Participants were tested individually and worked at their own pace throughout. Each person was randomly assigned to construct a composite using either the normal or sketch version of EvoFIT within the constraint that each of the 10 target photographs was constructed once in each of these two conditions. They first looked at a photograph of the face (selected randomly) for one minute and then waited between 22 and 26 hours.

After the delay, participants described the face to the experimenter using a cognitive interview. This involved the participants freely describing the face whilst the experimenter noted down what they said. Afterwards, the experimenter asked them to attempt to recall more detail about each feature.

A composite was then constructed using EvoFIT. Participants first chose the appropriate age, gender and type of database (sketch or photographic) in EvoFIT. They then selected a hairstyle from the available range. After the hair had been chosen, the participants were shown a random selection of 60 faces, and were asked to choose six faces over three screens that had facial shape similar to the target’s face. The participant was then shown a fourth page where they were able to change some of their chosen facial shapes for better likenesses. EvoFIT then generated another set of 60 faces, and participants then selected facial textures (feature greyscale colourings) that were similar to their target face. Again, they selected six faces over three screens, and reviewed their choices on the fourth. Next, the selected shapes and textures were presented in combination and participants were asked to choose the best overall likeness. EvoFIT then generated (‘bred’) another two sets of 60 faces from the selected items and the procedure was repeated. A set of tools was used to enhance the overall likeness of the final selected face. These tools changed the
apparent age, weight, position of features, happiness and other overall aspects of the face. The experimenter gave participants the option of further editing their composite in PRO-fit, to add, for example, extra shading or earrings.

Example composites constructed are shown in Figure 1 below.

![Example composites produced using the photographic (left) and sketch (centre) EvoFIT database. The target photograph used is shown on the right.](image)

**Figure 1:** Example composites produced using the photographic (left) and sketch (centre) EvoFIT database. The target photograph used is shown on the right.

**Stage 2: Composite Evaluation**

In this part, the composites were evaluated by asking other people to name them. An *a priori* was put in place: these participants were required to know at least eight out of the ten targets for their data to be included in the study, since poor target naming would undesirably result in poor composite naming. (In practice, all of the targets were known by all of the participants.)
Participants were presented with composites from both types of image mode (photographic / sketch) and therefore the design was within subjects.

**Participants**

Staff members from the Psychology department were recruited to evaluate the composites. There were six males and six females, aged between 30 and 52 years. The mean age was 42.0 years and the standard deviation was 6.0 years. None of these participants featured as targets in the construction stage.

**Materials**

Materials were the 20 composites produced in Stage 1. They were printed in greyscale on A4 paper, one face to a sheet to dimensions of 13 cm (width) x 16.5 cm (height). The target photographs from Stage 1 were also used.

**Procedure**

Participants were asked to name the composites of members of staff in Psychology, proving a name where possible. They were shown the composites in sequence and in a different random order for each person. After looking at all of the composites, and attempting to name them, they were shown the target photographs and similarly attempted to name those. Participants worked individually and each at his/her own pace.

**Results**

The mean percentage correct naming by composite-mode for male and female targets are presented in Figure 2. The graph shows that the sketched composites were named somewhat better than the photographic composites, with a greater effect for the female targets, and that male composites were more successfully named than female composites overall.
A 2x2 mixed factorial Analysis of Variance was conducted in the participant correct naming scores. This analysis approached significance for database type (normal and sketch), $F(1, 11) = 4.71$, $p = .053$, effect size $\eta^2 = .300$. Given past work, the direction of database type is known and therefore this factor can be considered as significant in a one way test, $p = .027$. Gender type was also found to be significant, $F(1, 11) = 11.44$, $p = .006$, $\eta^2 = .510$, as composites of male targets were named more successfully than composites of female targets. The interaction between database type and gender type was not significant, $F(1, 11) = 2.37$, $p = .152$. In the by-items analysis, database type was found to approach significance in a one-way t-test, $p = .091$, providing weak evidence for a benefit of the sketched database. Gender type and the interaction were not significant.

An analysis of incorrect naming was carried out, to explore whether participants tended to offer more names in general for sketched or photographic composites. While there was a trend that sketch composites received lower incorrect names, a further ANOVA found that none of the differences approached significance (for neither main effects nor the interaction).
Discussion

The purpose of the study was to find out if a sketched version of EvoFIT would produce more identifiable composites than the normal photographic mode, as indicated in pilot work. It was found that the sketch composites produced better correct naming scores than did the normal composites. There was some evidence of better naming for the male targets. The results thus support the experimental hypothesis.

The study found that the simplified (sketch) composites produced better naming than the more detailed (normal or photographic quality) composites. A reason that the sketch composites were better recognised could be that as there was less information for users to focus on: the amount of ‘texture’ or shading information was considerably reduced. This enabled users to focus on the overall appearance of the face, and is supported by past research. First, in pilot work, Frowd et al. (2008) found a benefit for the simpler version. Secondly, Frowd et al. (2005a) found that the sketch artist was better than Photofit or the other UK systems. Thirdly, Frowd et al. (2007) found that composites from the sketch-like Identikit 2000 were rated significantly better than the photographic PRO-fit. In the present study, sketch composites produced better naming, and with a medium effect size, a benefit that would be useful forensically. Thus, changing the image modality would appear to be valuable.

The experimental hypothesis was that sketch composites would be better recognised than photographic ones. This was found to be true for both male and female targets but, overall, composites of male targets were recognised more successfully than those of female targets. A potential reason for this could be related to the hair that participants chose for the composites. As hair differs greatly between females, this feature could make a composite distinctive, and therefore, easier to recognize, or harder, if the selected hairstyle were to be inaccurate. As most males’ hair tends to be short, this feature may not be a distinctive cue to aid recognition. Also, sketched hair may hide deficiencies in the selected hair, as well as making the internal features less inaccurate. Further research could usefully explore the contribution to naming of the hair from sketched composites, since poor quality hair could be rendered less distracting in this modality.
It is possible that the procedure used to construct the face could be simplified in another way. At present, users are asked to describe the face in detail using a cognitive interview. There is some evidence to suggest that describing a face can interfere with a person’s ability to subsequently recognise a face. This is known as ‘verbal overshadowing’ (e.g. Finger and Pezdek 1999) and could reduce the quality of a person’s composite. A follow-up experiment could test this idea. This would involve the sketch system being used in two conditions, one that asked a witness to construct the face after a cognitive interview and the other to just construct the face. This could be extended to include a second experimental factor, where only half the witnesses interact with a composite operator. Removing the presence of the operator may allow a witness to relax and thereby better engage in the task. Such a procedure, if effective, could also be used in situations where police time was limited, as part of an automated method for face production (and even available over the Internet for witnesses to use in their home environment). Current research is exploring these ideas.

Follow up studies would also be valuable to see if the results replicate for different race databases. In addition to the white male database, there are others in EvoFIT: e.g. Afro-Caribbean, Indian, Eastern European and various mixed race options. It is well known that memory for own-race faces is superior to faces of a race that is unfamiliar (Bothwell, Brigham and Malpass 1989), and therefore a simplified representation could be even more effective in such situations (relative to a more photographic one).

In conclusion, the present study was found to support previous research by Frowd et al. (2008) such that simplifying the image modality of the faces helped users to produce better named composites. The findings suggest that the sketch system would be valuable to the police. Further research could be usefully carried out to see if a sketch database would be beneficial to the construction of other-race faces.

References


Frowd, C. D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., McIntyre, A. and Hancock, P.J.B. Forthcoming. ‘The psychology of face construction: giving evolution a helping hand’. *Applied Cognitive Psychology*.

