ERGONOMICS, 1991, VOL. 34, NO. 4, 459–468

Face composite construction: In-view and from-memory quality and improvement with practice

MICHAEL S. WOGALTER

Department of Psychology, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

and

D. BRADLEY MARWITZ

Department of Psychology, University of Richmond, Richmond, VA 23173, USA

Keywords: Face; Memory; Learning.

Following a crime, witnesses are often asked by the police to construct a composite likeness of the perpetrator. However, previous research indicates that the quality of the likenesses produced by commercial composite systems (e.g., Photofit and Identi-kit) are very poor. This study investigated whether the face composites produced using the Mac-A-Mug Pro composite system from memory and while viewing the target differ in quality, and whether quality improves with practice. Subjects used the system over two sessions during which they constructed 11 composites (six from memory and five with the face in view). Composites produced while viewing the target face were better than the composites produced from memory, and both improved with practice. Independent judges matched all but the initial composites from memory at levels better than chance. Implications for future face recall research and the training of potential witnesses are discussed.

1. Introduction

Following a crime, police often ask witnesses to construct a pictorial likeness of the perpetrator. This construction is usually done with the aid of an artist or an operator using one of several commercially available kits, for example, Photofit and Identi-kit. These composite kits contain different depictions of features of the face, for example, different kinds and shapes of eyebrows. The witness usually describes the person in question verbally and the operator selects features to achieve a composite that supposedly resembles the target person.

Virtually all of the research examining composite quality show that these techniques do not produce good likenesses of the target face (Davies *et al.* 1978a, Ellis *et al.* 1975, Ellis *et al.* 1978, Laughery and Fowler 1980). This failure of current procedures warrants investigation to improve the quality, hence accuracy, of the resulting composites.

1.1. *Composite quality*

Ellis *et al.* (1975, 1978) found that having lengthy exposures to a target face, having the face present during construction, and using an experienced operator did not significantly improve the quality of Photofit composite likenesses (as rated by independent judges). They concluded that the failure to produce good composites was due to the lack of precision in the Photofit itself, and that it would have limited utility in law enforcement. Similarly, Laughery and associates (Laughery and Smith 1978, Laughery and Fowler 1980) examined the accuracy of likenesses produced by sketch artist and Identi-kit systems. The

0014–0139/91 $3.00 © 1991 Taylor & Francis Ltd.

Identi-kit composites were judged to be poor and produced a low rate of identification. The sketches were better, but also were of poor quality.

Most composite systems require the witness to interact with an artist or operator to produce a face representation. The witness must describe the target while the artist or operator translates that description and produces a likeness. Research (e.g., Davies *et al.* 1978b, Laughery *et al.* 1986) has indicated that people are not fluent in describing faces, probably due to the limited vocabulary of the English language in this domain.

To reduce the problem associated with verbally describing a face, a composite system that a witness can use directly with little assistance would be desirable. One such system is the Field Identification System (FIS). Laughery *et al.* (1980) examined the accuracy of the FIS productions by comparing the results to those from a study by Laughery and Smith (1978) which examined the accuracy of artist sketches and Identi-kit composites. It was concluded that resulting quality was best with sketches, intermediate with Identi-kit composites, and poorest with FIS composites. Laughery *et al.* (1980) suggested that the absence of an expert familiar with face recall procedures was a possible reason for the poor quality FIS composites. Subjects were constructing composites for the first time and they were not very good at it. Alternatively, differences in quality might be due to the limited number of alternatives available to construct a likeness. While sketch artists presumably can produce infinite varieties of a face feature, composite systems have limited numbers of feature exemplars available for selection. Furthermore, some systems are more limited than others. For example, the FIS has fewer alternatives than the Identi-kit.

Given the problems associated with producing likenesses, the present research examined whether use of another composite system, Mac-A-Mug Pro, would be superior. Mac-A-Mug Pro is a computerized system having a large library of feature alternatives. Recently, Cutler *et al.* (1988) reported that expert operators using the Mac-A-Mug Pro system were able to produce composites (with the target face in view) that allowed subjects to discriminate them from non-target photographs and composites. This result, and the fact that it can be used directly by a witness, leads to the suggestion that this system might allow witnesses to produce more accurate composites than those produced by other composite systems.

1.2. *Training*

Training is another potential way to improve composite quality. However, research on face training has been limited to recognition performance. Woodhead *et al.* (1979) attempted to improve memory for faces by having subjects participate in activities which focused on the analysis of individual facial features. Training consisted of three days of intensive instruction and practice, using lectures, slides, films. They found that subjects receiving training never showed significantly better recognition performance than subjects receiving no training. Other studies show similar results with the exception of learning to differentiate faces of a different race than that of the learners (Elliott *et al.* 1973, Malpass *et al.* 1973, Lavrakas *et al.* 1976). Evidently, face recognition is a well-developed skill for which additional training is superfluous (Malpass 1981). While skill at recognizing faces may be near ceiling levels, previous results lead to the suggestion that skill at producing composites is near floor levels.

Consequently, composite production may benefit from training and practice. No research has examined, however, the effect of training on composite quality.

1.3. *Rationale*

Subjects of this study constructed composites of six different target faces using Mac-A-Mug Pro. First, they produced a composite from memory of a face photograph, and second, they produced a composite while the photograph was in view (except the first target which was only produced from memory). Subsequently, performance (as rated by independent judges or determined by ability to match composites to target photographs) could be assessed as a function of (*a*) from memory or in view, and (*b*) practice across production trials. The in-view procedure also provided feedback on the quality of subjects' from-memory composites and thus served as additional training and practice.

Additionally, the present study examined the effect of (*a*) target exposure, and (*b*) initial composite instruction. Usually, witnesses come to the composite task not knowing what procedure to expect, and then must become familiar with the task demands for the first time. Furthermore, memory of the target may be impaired by the intervening instruction. If witnesses were familiar with the composite system, they would know the kind of procedure to expect, they could start composite production with less delay, and they would not experience potentially interfering instructions before starting the composite. In the present study, groups of subjects received initial composite instruction and viewed the first target in one of two orders. One group, the Face-First subjects, viewed the first target before receiving any instructions on the composite system (which is analogous to the order actual witnesses experience these activities). Another group, the Instructions-First Subjects, was given brief instructions on how to use the composite system before they viewed the first target. It was hypothesized that Instructions-First subjects would produce more accurate composites.

A third group, the Recognition-First subjects, assessed the effect of a prior recognition task (mug file search) on subsequent composite quality. While research exists on the influence of composite production on subsequent recognition performance (Davies *et al.* 1978a, Hall 1977, Mauldin and Laughery 1981, Wogalter *et al.* 1989), no study has examined the converse order. After viewing the target, the Recognition-First subjects participated in a recognition test before receiving composite system training and producing their first composite. Composite quality and recognition performance of these subjects were compared to subjects who produced a composite before the recognition test.

2. Method

2.1 *Subjects*

Fifty-four (30 females, 24 males) University of Richmond, Virginia, undergraduates were assigned randomly to one of the three groups (18 per group) and participated individually in the tasks. Later, a group of 10 Rice University, Texas, graduate students took part in either a rating or matching task (five judges per task) to acquire measures of composite accuracy.

2.2 *Apparatus and materials*

Target photographs and recognition test distractors were taken from college yearbooks and converted into slides. Six white male faces were used as targets. Face presentation order was balanced using a Latin Square so that each face appeared in all six positions an equal number of times across subjects.

Subjects generated composites using the Mac-A-Mug Pro software (Shaherazam, Milwaukee, WI) on an Apple Macintosh computer with hard disk drive. This composite system allows one to select and place different face features on the computer screen using a mouse pointer, pull-down windows, and keyboard commands. This program is easy to use and is accompanied by a manual of instructions as well as replicas of the features that the program can access. After brief training, subjects were able to use the system to quickly generate and revise composites with little or no help from the experimenter.

2.3 *Procedure*

Subjects were assigned to one of three groups: Instructions-First, Face-First, and Recognition-First.

2.3.1. *Instructions-First group:* Subjects assigned to the Instructions-First condition initially received instructions and a demonstration on how to use the computer, software, and accompanying manual. Subjects were given instructions on how to locate, access, manipulate and edit features. Finally, the experimenter demonstrated the construction of a sample face of random features. Then subjects were given 10 min to use the system freely to familiarize themselves with its operation. After this initial instruction, a target face was projected for 8 s and then removed from view. Subjects then had 20 min to construct the composite, which was saved and labelled (coded with subject number, face number, face position, and whether the face was constructed from memory or in view). Twenty min provided ample time to complete the from-memory composites.

After the first composite, subjects were given a recognition test in which they were shown a series of 80 slides of white male faces. The target face (the first composite) always appeared in position 76. For every face in the slide sequence, subjects indicated whether it was the target face they saw earlier and their degree of confidence.

Following the recognition test, subjects were shown a second target face for 8 s and given 20 min to generate a composite (which was subsequently labelled and saved). At this point, the second target face was brought back into view and the subject was encouraged to make revisions on a duplicate of their second from-memory composite to improve its quality. Ten min were allowed for the in-view construction and the resulting composite was labelled and saved. The same procedure was repeated for the third target face: (1) subjects viewed the target for 8 s and constructed a composite from memory; (2) the composite was saved; (3) the target photograph was returned into view and a duplicate composite was revised by the subject; and (4) the in-view composite was saved. After the procedure for the third target face was completed, subjects were dismissed from the first session and then returned 6–8 days later for a second session consisting of composite productions of three additional targets. The procedure for the fourth, fifth, and sixth targets was identical to the second and third targets of the first session. After the second session was completed, subjects were debriefed and thanked for their participation.

The experimenter allowed subjects to ask questions about the software and the construction process but took efforts not to influence them in any of their feature selections. Generally, they asked very few questions after the first composite.

2.3.2. *Face-First group:* The procedure for the Face-First condition was identical to the Instructions-First condition with one exception: Face-First subjects were exposed to the first target face *before* receiving any composite instructions, whereas Instructions-First subjects were exposed to the first target *after* receiving composite instructions. Once the Face-First subjects started producing the first composite, the remaining procedure (including the second session) was identical to the Instructions-First subjects.

2.3.3. *Recognition-First group:* Subjects in the Recognition-First condition were first exposed to the target face followed by a 20 min irrelevant distractor task (sorting verbal statements). The distractor task was inserted for the purpose of maintaining an equivalent delay period before the start of the recognition test as in the other two groups. After the recognition test, subjects were given the same composite instructions as subjects in the other two conditions followed by the composite production procedure. After completing their first (and only) composite, subjects in the Recognition-First group performed no other tasks.

2.3.4. *Evaluation of composite quality:* Face-First and Instructions-First subjects produced six composites from memory and five from view (for the first target, there was no in-view condition). Recognition-First subjects produced only one composite (from memory). The 414 composites produced in the experiment were printed onto individual sheets and assembled into a random order. To acquire two measures of composite quality, five student judges performed a matching task and another five performed a similarity rating task. In the matching task, judges were presented with photographs of all six target faces mounted on a cardboard display, each labelled with an identification code. The judges examined every composite and for each chose one face from the photographs that they thought was the basis for the composite.

For the similarity ratings, six booklets were constructed containing the 69 composites produced of each target. The judges in the similarity-rating task compared each composite to its corresponding target photograph and rated the 'goodness of fit'. The ratings were based a six-point Likert-type scale anchored at the low (0) and high (5) ends with the terms 'Not At All Similar' and 'Extremely Similar'. After all composites associated with one target were evaluated, subjects rated all composites associated with another target face, and this procedure continued until all composites were rated. A different random order of targets were rated by each judge. All judges worked independently.

3. Results

3.1. *Measures and design*

Two quality (accuracy) measures were derived from the judges' matching and similarity rating scores. The number of correct matches was averaged across judges to produce a mean proportion matching score for each composite (ranging from 0 to 1). A similar procedure was used to yield a mean similarity rating score for every composite (ranging from 0 to 5). Several analyses of variance (ANOVAs) were necessary due to the incomplete factorial design in the experiment.

3.2 *Group conditions*

The matching scores of the first composite were examined across the three groups (Instructions-First, $M=0{\cdot}29$; Face-First, $M=0{\cdot}32$; Recognition-First, $M=0{\cdot}38$) using a one-way between-subjects ANOVA. No significant effect of group was found ($p>0{\cdot}05$). The rating scores (Instructions-First, $M=1{\cdot}52$; Face-First, $M=1{\cdot}53$; Recognition-First, $M=1{\cdot}46$) also failed to show a significant effect of group. Because group had no effect, subsequent consideration of the Recognition-First scores was dropped. The remaining analyses continued to include the group factor with respect to the other two conditions (Instructions-First, Face-First) to insure that this variable did not interact with any other factor. However, because group did not produce significant main effects or interactions, the cell means in table 1 are shown collapsed across this variable.

Table 1. Matching and rating means as a function of face position and target presence.

		Face position					
		1st	2nd	3rd	4th	5th	6th
Matching	Memory	0·31	0·36	0·41	0·38	0·44	0·39
	In-view	—	0·51	0·42	0·41	0·41	0·48
Rating	Memory	1·53	1·57	1·84	1·79	1·79	2·01
	In-view	—	1·80	1·88	2·06	2·09	2·42

Note: These data are averaged across the Face-First and Instructions-First groups.

3.3 *Composites produced from memory*

A 2 (Face-First, Instructions-First)×6 (first through sixth faces) mixed-model ANOVA was used to analyse the matching and rating scores of composites produced from memory. These means can be seen on the first and third rows of table 1. For the matching scores, there were no significant effects ($ps>0.05$). For the rating scores, no main effect of group or interaction was found. However, there was a significant main effect of face position, $F\,(5, 170)=2.49$, $p<0.05$. Table 1 shows that the mean similarity ratings increase from the first to the sixth face. Pairwise comparisons using Fisher's Least Significant Difference test (LSD) showed that composites in the sixth position were rated significantly better than composites in the first and second position (both $ps<0{\cdot}01$). No other differences were significant ($ps>0.05$).

3.4. *Composites produced from memory and in-view*

In the following analyses, the face presence factor (Memory, In-View) was added.

Because no in-view composites were produced for the first face, this position was not included in the analysis. This resulted in a $2 \times 5 \times 2$ mixed-model ANOVA with the factors of group (Face-First, Instructions-First), face position (second through sixth faces), and face presence (Memory, In-View).

For the matching scores, there were no significant main effects for group or face position. However, there was a significant effect of face presence, $F(1, 34)=4\cdot 90$, $p<0\cdot 05$. In-view composites ($M=0\cdot 44$) were correctly matched more often than composites produced from memory ($M=0\cdot 40$). No interactions were significant.

For the similarity ratings, there was no significant main effect of group, but there was a significant main effect of face presence, $F(1, 34)=22\cdot 78$, $p<0\cdot 001$. In-view composites ($M=2\cdot 05$) were rated significantly higher than from-memory composites ($M=1\cdot 80$). The ANOVA also showed a significant main effect of face position, $F(4, 136)=3\cdot 97$, $p<0\cdot 01$. The table shows that both memory and in-view composites received higher similarity ratings as production progressed from the second to the sixth face. Pairwise comparisons using Fisher's LSD revealed that composites in the sixth position were rated significantly higher than composites in other positions (ps$<0\cdot 05$). No interactions were significant.

3.5. *Additional matching analyses*

If judges made their matching choices randomly, they would correctly guess the target once in every six attempts producing an accuracy rate of 0·167. Selection rates above this level would indicate the composites are providing information that is assisting judges in making their selections. Therefore, in the following analyses, matching performance was compared to chance expectation. For the composites produced from memory by the Face-First subjects, all but the first composite were matched at rates significantly higher than chance (ps$<0\cdot 05$). Similarly, for the Instructions-First subjects, all but the first two from-memory composites were matched significantly greater than chance (ps<0.05). All of the in-view composites (examined separately by group and face position) were matched significantly greater than chance (ps$<0\cdot 05$).

The relationship between the matching and rating scores was also examined. For the Instructions-First and Face-First composites generated from memory, the correlation between the matching and rating scores was positive and significant, $r=0\cdot 46$ ($N=36$), $p<0\cdot 01$. A similar relationship was found for the in-view composites, $r=0\cdot 62$ ($N=36$), $p<0\cdot 001$. Thus, the matching and rating scores for the memory and in-view conditions have 22% and 39% of their variance in common, respectively. These results indicate that both measures are, in part, measuring the same thing, which presumably is composite quality or accuracy. Apparently, the matching scores lacked sufficient sensitivity to detect the practice effect that was shown by the similarity-ratings. More matching judges might have provided greater power to detect a practice effect.

3.6. *Recognition performance*

The last set of analyses examined recognition performance between the three groups. Hit, false alarm, discrimination, and confidence measures failed to show any significant effects (ps$>0\cdot 05$). Correlations between recognition performance and quality measures also were not significant. The failure to find recognition effects is probably attributable to the easiness of the test: all groups had hit rates over 80% and false alarm rates under 3%.

4. Discussion

Previous face recall research has been characterized by the failure of composite systems to produce good likenesses of target faces (Davies *et al.* 1978a, Ellis *et al.* 1975, 1978, Laughery and Fowler 1980). Indeed, performance has been reported to be so poor that: (1) composites produced from memory and those done while the face was in view failed to differ, and (2) composite matching scores were no better than chance (Ellis *et al.* 1975, 1978, Laughery and Fowler 1980). The present results showed that composites constructed with the target face in-view were better than those produced from memory. In addition, composites were matched to photographs by independent judges at levels greater than chance (except for the first one or two positions from memory). Cutler *et al.* (1988) also found that Mac-a-Mug Pro composites produced by an expert operator while viewing the target face provided adequate cues which allowed judges to discriminate them from distractor photographs and composites. Together these findings suggest that the Mac-A-Mug Pro system is superior to previously researched systems.

Prior research on face training failed to show successful recognition improvement except for faces of a different race than that of the subjects (e.g., Elliott *et al.* 1973, Woodhead *et al.* 1979). The present study examined whether practice with the Mac-A-Mug Pro composite system would lead to higher quality face composites. The results showed that judgments of similarity to photographs increased from early to later composites, a finding not previously reported. Given that people are seldom required to perform a composite generation task, it is reasonable that practice facilitated this skill. Face recognition performance may show little improvement with training, however, because face recognition skill may be already near its maximum level.

There were no significant differences among the three groups (Instructions-First, Face-First, and Recognition-First) on initial composite quality and recognition performance, nor was there a relation between these measures. The failure to find a difference between the Instructions-First and the Face-First groups can be taken as mild evidence that initial familiarization with a composite system after viewing the target does not interfere with face memory. However, it was apparent in other results that familiarization over multiple trials of practice leads to better composite quality. In addition, the failure to find a difference between the Face-First and Recognition-only composites suggests that a prior recognition test does not interfere with subsequent composite production. Because recognition performance was quite good for all three groups, it is unclear at this point whether differences might have been apparent had the recognition test been more difficult.

It is important to note that the current procedure may not have trained subjects to remember faces any better than they could before coming to the laboratory. Subjects might have learned how to use the composite system more effectively. This interpretation is supported by the finding that the quality of in-view composites increased with practice. Here, memory is not a factor, so the effect is apparently due to subjects becoming more adept at using the composite system. Alternatively, it is possible that practice with the system did influence subjects' encoding strategy, perhaps leading them to pay more attention to individual features than they did at the start.

The results have implications for future research on composite production. Most previous research has investigated composite quality using systems requiring the assistance of an expert operator to build composites from the witness's verbal description. This intervening step is likely to add 'noise' to the resulting composite because verbal descriptions do not adequately communicate face features. The present study shows that subjects quickly can learn the basic operations of the Mac-A-Mug Pro system and can use it directly to build competent composites with minimum assistance. This system may be useful in future research investigating factors influencing composite quality.

Besides the capability to produce competent likenesses, the results showed that familiarization and practice with a composite system leads to the production of higher quality composites. Persons who are likely to witness a crime (e.g., bank tellers, security guards, and convenience store clerks) could undergo training and practice with the composite system. If they were to view an assailant and subsequently were requested to generate a composite, they are apt to generate a more accurate likeness than witnesses who had not been trained.

Acknowledgement

The helpful comments of Kenneth R. Laughery and Larry D. Reid were much appreciated.

References

Cutler, B. L., Stocklein, C. J. and Penrod, S. D. 1988, Empirical examination of a computerized facial composite production system, *Forensic Reports,* **1,** 207–218.

Davies, G. M., Ellis, H. D. and Shepherd, J. W. 1978a, Face identification: the influence of delay upon accuracy of Photofit construction, *Journal of Police Science and Administration,* **6,** 35–42.

Davies, G. M., Shepherd, J. W. and Ellis, H. D. 1978b, Remembering faces: acknowledging our limitations, *Journal of the Forensic Science Society,* **18,** 19–24.

Elliott, E. S., Wills, E. J. and Goldstein, A. G. 1973, The effects of discrimination training on the recognition of white and oriental faces, *Bulletin of the Psychonomic Society,* **2,** 71–73.

Ellis, H. D., Shepherd, J. W. and Davies, G. M. 1975, An investigation of the use of the Photofit technique for recalling faces, *British Journal of Psychology,* **66,** 29–37.

Ellis, H. D., Davies, G. M. and Shepherd, J. W. 1978, A critical examination of the Photofit system for recalling faces, *Ergonomics,* **21,** 297–307.

Hall, D. F. 1977, Obtaining eyewitness identifications in criminal identifications: two experiments and some comments on the *Zeitgeist* in forensic psychology, paper presented at the meeting of the American Psychology-Law Society.

Laughery, K. R. and Smith, V. L. 1978, Suspect identification following exposure to sketches and Identi-kit composites, in *Proceedings of the Human Factors Society 22nd Annual Meeting* (Human Factors Society, Santa Monica, CA), 631–635.

Laughery, K. R. and Fowler, R. F. 1980, Sketch artist and Identi-kit procedures for recalling faces, *Journal of Applied Psychology,* **65,** 307–316.

Laughery, K. R., Smith, V. L. and Yount, M. B. 1980, Visual support devices: evaluation of a new technique for constructing facial images, in *Proceedings of the Human Factors Society 24th Annual Meeting* (Human Factors Society, Santa Monica, CA), 302–305.

Laughery, K. R., Duval, G. and Wogalter, M. S. 1986, Dynamics of face recall, in *Aspects of face processing,* ed. H. D. Ellis, M. A. Jeeves, F. Newcombe and A. W. Young (Martinus Nijhoff, Dordrecht), 373–387.

LAVRAKAS, P. J., BURI, J. R. and MAYZNER, M. S. 1976, A perspective on the recognition of other race faces, *Perception & Psychophysics,* **20**, 475–481.

MALPASS, R. 1981, Training in face recognition, in *Perceiving and remembering faces,* ed. G. M. DAVIES, H. D. ELLIS and J. W. SHEPHERD (Academic Press, London), 299–309.

MALPASS, R. S., LAVIGUEUR, H. and WELDON, D. E. 1973, Verbal and visual training in face recognition, *Perception & Psychophysics,* **14**, 285–292.

MAULDIN, M. M. and LAUGHERY, K. R. 1981, Composite production effects on subsequent facial recognition, *Journal of Appled Psychology,* **66**, 351–357.

WOGALTER, M. S., LAUGHERY, K. R. and THOMPSON, B. G. 1989, Eyewitness identification: composite construction on subsequent recognition performance, unpublished manuscript, Rensselaer Polytechnic Institute, Troy, NY.

WOODHEAD, M. M., BADDELEY, A. D. and SIMMONDS, D. C. V. 1979, On training people to recognize faces, *Ergonomics,* **22**, 333–343.

Accepted 24 September 1990.