TRAINING POTENTIAL WITNESSES TO PRODUCE HIGHER QUALITY FACE COMPOSITES

D. Bradley Marwitz
Psychology Department
University of Richmond
Richmond, VA 23173

Michael S. Wogalter
Psychology Department
Rensselaer Polytechnic Institute
Troy, NY 12180

ABSTRACT

This study attempted to determine if training and familiarization with a face composite system would improve the quality of the produced composites. Subjects were trained in the use of the Mac-a-Mug Pro system over two sessions during which they constructed eleven composites (six from memory and five with the face in view). The results indicate that the composites produced while the target face was in view were significantly better than the composites produced from memory, and that both improved with practice. Initial training with the composite system prior to exposure to the first face or after the first face did not affect composite quality. These results have implications for the training of personnel at high risk of witnessing a crime.

INTRODUCTION

Following a crime, witnesses are often asked to aid the police investigation by constructing a composite likeness of the assailant, by using a sketch artist or one of several commercial composite kits (e.g., the Identi-kit, the Photofit, and the Field Identification System). However, virtually all of the research examining composite quality show that these production techniques do not produce good likenesses of the target face (Davies, Ellis, & Shepherd, 1978; Ellis, Shepherd, & Davies, 1975; Ellis, Davies, & Shepherd, 1978; Laughery & Fowler, 1980). The failure of composite systems to produce good likenesses would seem to warrant investigation on the procedures that improve the quality and accuracy of the face composites in order to be more useful to law enforcement.

Face Training

One way to improve performance, in general, is to train. However, face training has had mixed success. Woodhead, Baddeley, and Simmonds (1979) attempted to improve face memory by having subjects participate in activities which focused on the analysis of individual features of the face. Training consisted of three days of intensive instruction using lectures, slides, films, and applied practice. Woodhead et al. found that recognition performance of the subjects who received training was never significantly better than subjects who received no training, and in one condition was worse. Despite Woodhead et al.'s failure to find an effect of training, other studies have been able to show a positive effect of training on recognition performance, but only to erase a deficit that often exists for faces of another race.

Elliott, Wills, and Goldstein (1973) investigated the effects of paired associate discrimination training using white or Oriental faces as stimuli. White subjects were originally superior on their recognition and discrimination of white faces compared to Oriental faces. Training with Oriental faces improved their ability to recognize other Oriental faces, but training with white faces did not further improve their ability to recognize other white faces. Several studies have confirmed that training improves recognition of other-race faces, but no study has shown a positive effect of training with own-race faces (Malpass, Lavigueur, & Weldon, 1973; Lavrakas, Buri, & Mayzner, 1976). Malpass (1981) suggests that the failure to find improvement for own-race faces may be due to two factors: 1) face recognition is overlearned to the point that further training would have no effect, and 2) the methods used in training subjects are unnatural and counteract one's own natural memorial strategies.

Composite Production Techniques

While research has examined the effects of face training on recognition, there has been no known research directed at composite construction training. The void in face recall research is probably due to the consistent finding that subjects produce very poor quality composites. Ellis et al., (1975; 1978) found that longer exposures to the face, having the face present during construction, and the use of an experienced operator did not significantly improve the quality of the composite likeness as rated by independent judges. Ellis and colleagues concluded that the composites would be of limited use to law enforcement and suggested that the failure of the subjects to produce accurate composites was due to a lack of precision in the Photofit system itself. Similarly, Laughery and associates (Laughery & Smith, 1978; Laughery & Fowler,1980) examined the accuracy of the sketch artist and Identi-kit systems The Identi-kit composites were judged to be quite poor and had a low rate of identification. The sketches were better, but were poor nonetheless.

Most composite systems require the interaction of the witness with another person (either artist or an experienced operator) to produce a representation of the face. This interaction can potentially cause communication problems. The witness must describe to the artist or operator what the target looks like and this second person must translate that description to produce a likeness. Ideally, a system is desired that the witness can use to produce a likeness directly. One such composite system is the Field Identification System (FIS) which does not require the presence of a second person. Laughery, Smith, and Yount (1980) explored the accuracy of the FIS following procedures identical to those used by Laughery and Smith (1978) using FIS composites. They concluded that composite quality was best with sketches, intermediate with Identi-kit composites, and poorest with FIS composites.

They suggest that the absence of an expert familiar with face recall procedures was responsible for the poor quality composites produced by the FIS. Subjects using the FIS were constructing composites for the first time and were likely not very good at it. Another possible reason for quality differences focuses on the range of features available for the production of a likeness. The sketch artist can produce an infinite range of feature varieties, while commercial composite systems are somewhat more limited in the number of feature exemplars that are available (the FIS is more limited than the Identi-kit).

Thus, prior research on composite system quality has not found much success. The present research sought to determine whether training and familiarization with another composite system, Mac-A-Mug Pro, would enable potential witnesses to construct better quality composites. The Mac-A-Mug system has shown some promise of producing better quality composites than the systems examined thus far (Wogalter, Laughery & Thompson, 1989). By becoming familiar with the system through practice, it was expected that composite quality would improve as subjects continue to generate subsequent composite likenesses. In a sense, the subject becomes his/her own expert operator.

In the present study, subjects constructed composites based on six different target faces. For all targets (except the first) subjects produced a composite while the target was in view after they had produced a composite from memory. This procedure served two purposes. The first was to determine if the composites produced from memory and while in view would differ in quality. Prior research (Ellis et al., 1978; Laughery & Fowler, 1980) has failed to show this difference using the Photofit and Identi-kit systems, indicating that these two systems were unable to produce good likenesses even when memory was not a factor (i.e., the In-View condition). The second purpose for having subjects produce In-View composites is that it served as additional training and practice in composite production.

The present experiment also examined whether the quality of the first composite would be influenced by the presentation order of the initial composite system instructions and the initial view of the first target face. One group, the Instructions-First subjects, was given brief instructions on how to use the composite system before they viewed the first target. Another group, the Face-First subjects, viewed the first target before receiving any instructions on the composite system. The sequence in the Face-First condition is analogous to the order in which actual witnesses would experience these activities. That is, generally the witness would be first exposed to an assailant and then would have to become familiar with a composite production system in order to generate a composite. Because the Instructions-First subjects would be somewhat familiar with the composite system before viewing the first target, they would know (to some extent) the kind of procedure to expect at test. They would also not experience potentially interfering instructions in the period between viewing the target and composite production. Thus, it was hypothesized that subjects in the Instructions-First group would produce more accurate initial composites than subjects in the Face-First group.

A third group, the Recognition-First subjects, was also included in the experiment. These subjects produced only one composite and were included in the experiment to determine what effect a recognition task would have on subsequent composite construction quality. The issues involved with the recognition performance are not directly relevant to the primary purpose of the present research, and thus, will be discussed only briefly in this report.

METHOD

Subjects

Fifty-four (30 females, 24 males) University of Richmond undergraduates were randomly assigned to one of the three groups (18 per group) and participated individually in the composite construction tasks. Later, a group of 10 Rice University graduate students took part in one of two tasks (five judges per task) for the purpose of acquiring measures of composite accuracy.

Apparatus and Materials

The target photographs (and the distractors in the recognition test) 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 a hard disk drive. This composite system allows the individual user to select and place the 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 feature exemplars that the program can access. After brief training on how to use the software, subjects were quickly able to use the system to generate and revise composites with little or no help from the experimenter.

Procedure

Subjects were assigned to one of three groups: Instructions-First, Face-First, and Recognition-First. The specific procedures of these groups are described below.

Composite Training of Instructions-First Subjects. Subjects assigned to the Instructions-First condition initially received instructions and a demonstration on how to use the computer, software, and accompanying manual. Subjects received specific instructions on how to locate, access, manipulate and edit features. Finally, the experimenter demonstrated the construction of a sample face of random features. Subjects were then given 10 min to freely use the system to familiarize them with its controls and operations. After this initial instruction, the subjects were told that they would be exposed to a target face and be asked to construct a composite likeness of that face. The target face was projected for a total of eight sec and then removed. The subject then had 20 min to construct the composite, which was saved and labeled (coded with subject number, face number, face position, and whether the face was constructed from memory or in-view). Twenty minutes proved to be ample time for subjects to complete the composites from memory while ten minutes were allowed for the In-View composites.

The experimenter allowed the subjects to ask questions about the software or the construction process after the first composite and all subsequent composites. After producing 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 in their answer. Recognition results will not be discussed in this report.

Following recognition, subjects were shown a second target face for eight sec and given 20 min to generate a composite of this face, which was subsequently labeled 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 version of their second composite to improve its quality. The resulting In-View version of the composite was also labeled and saved. The same procedure outlined above for the second target face was repeated for the third target face: (1) Viewing the target for 8 sec and constructing a composite of that face from memory, (2) saving the composite, (3) returning the target to view in order to revise the composite and then, (4) saving the In-View composite as well. 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 more target faces. The procedure for the fourth, fifth, and sixth target faces was identical to the second and third face of the first session. Subjects were then debriefed and dismissed.

Composite Training of Face-First Subjects. Subjects in the Face-First condition participated in a procedure identical to that of the Instructions-First subjects, with one exception: Face-First subjects were exposed to the first target face before receiving any composite instructions, whereas Instructions-First subjects saw the first target after receiving composite instructions. Once the Face-First subjects began the first composite construction, the remaining procedure (including the second meeting) was identical to that of the Instructions-First subjects.

Composite Training of Recognition-First Subjects. Subjects in Recognition-First condition initially received no training with the composite system. They 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 a delay period comparable to the other two groups prior to the start of the recognition test. Following recognition, subjects were given the same composite instructions that were given to subjects in the other two groups, followed by composite production. After completing their first (and only) composite, subjects in the Recognition-First group were dismissed. These subjects returned for a second meeting but participated in a set of unrelated tasks.

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 that were produced 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 labeled 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. The judges worked independently.

For the similarity ratings, six booklets were constructed containing the 69 composites produced of each target. The similarity rating judges' task was to compare the composites to its corresponding target photograph and to rate the "goodness of fit." The ratings were based a six-point scale with "0" meaning not at all similar and "5" meaning extremely similar. After all of the composites associated with one target were evaluated, subjects evaluated all of the composites associated with a another target face, and this procedure continued until the composites of all six targets were rated. A different random order of targets were rated by each judge.

RESULTS

Two quality measures were derived from the judges' matching and similarity rating scores. The number of correct matches was averaged across the five judges producing a mean matching score for each composite. The same procedure was followed to yield a mean similarity rating score for each composite. Several analyses of variance (ANOVAs) designs were required to analyze the data due to the incomplete factorial in the experiment (there was no In-View condition for the first target).

Group Conditions. A one-way between-subjects ANOVA was used to examine the matching scores of the first composite across the three conditions (Instructions-First, M = .289; Face-First, M = .322; Recognition-First, M = .378). No significant effect of group was found ($F < 1.0$). The same analysis on the rating scores (Instructions-First, M = 1.522; Face-First, M = 1.533; Recognition-First, M = 1.456) also failed to detect an effect of group ($F < 1.0$). Because group membership had no effect, subsequent consideration of the Recognition-First scores were dropped. The remaining analyses continued to include the group factor with respect to the other two groups (Instructions-First, Face-First) to insure that this variable did not interact with any of the other factors. Because this factor did not show a main effect or interact with other variables in any of the subsequent analyses, the cell means in Table 1 are shown collapsed across this variable.

Face Order for Composites Produced from Memory. A separate 2 (Face-First, Instructions-First) X 6 (first through sixth faces) mixed-model ANOVA was used to analyze the matching and rating scores. Because the first face had no In-View condition, only the composites generated from memory were analyzed in this ANOVA. For the matching scores, no main effects or interaction was found (all F's < 1.0). These scores appear in the first line of results in Table 1. For the rating scores, no main effect of group or significant interaction was found, (F's < 1.0). However, there was a significant main effect found for face order, $F(5,170) = 2.49$, $p < .05$. The memory means in Figure 1 show a general increase in similarity ratings from the first to the sixth faces. Pairwise comparisons using Fisher's Least Significant Difference (LSD) test showed that the sixth face (M = 2.01) was rated significantly better than both the first face (M = 1.53) and the second face (M = 1.57), p's < .01.

Table 1. Overall Cell Means for the Experimental Design Collapsing Across Face- First and Instructions First-Groups.

		Face Order					
		1st	2nd	3rd	4th	5th	6th
Matching	Memory	.306	.361	.406	.378	.444	.394
	In-View	--	.506	.417	.406	.406	.478
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

There were no other significant differences among these means.

Composites Produced from Memory and In-View. In the following analyses, the face presence factor (Memory, In-View) was added. However, because the first face had no In-View condition, this face position was not included in this analysis. This resulted in a 2 X 5 X 2 mixed-model ANOVA with the factors of group (Face-First, Instructions-First), face order (second through sixth faces), and face presence (Memory, In-View). For the matching scores, there were no significant main effects for group or face order (F's < 1.0). However, there was a significant effect of face presence, $F(1,34) = 4.90$, $p < .05$. The composites produced while the targets were in view (M = .44) were more often correctly matched to the target photographs than the composites produced from memory (M = .40). No significant interaction effects were shown.

A 2 X 5 X 2 ANOVA was also performed on the similarity rating scores. Again, there was no significant main effect of group (F< 1.0), but there was a significant main effect for face presence, $F(1,34) = 22.78$, $p < .001$. In-View faces (M = 2.05) were rated significantly better than Memory faces (M = 1.80). The ANOVA also showed a significant main effect for face order, $F(4,136) = 3.97$, $p < .01$. Figure 1 shows that the similarity ratings increase from the second to the sixth faces for both Memory and In-View faces. Pairwise comparisons among the faces using Fisher's LSD revealed that the sixth face composites (M = 2.21) were rated as significantly more similar to the target photographs than the second (M = 1.69), third (M = 1.86), fourth (M = 1.93), and fifth (M = 1.94) faces (p's < .05). No significant interaction effects were shown.

Further Analysis of the Matching Data. Due to the failure of the matching scores to find effects other than the In-View and Memory difference, these data were analyzed further. Matching performance was compared to what would be expected by chance guessing alone. If the judges were randomly guessing in their matching choices, they would have a correct guess once in every six attempts (due to the six possible target face photographs). Thus, chance would be reflected by an accuracy rate of 0.167. For the Face-First composites, all but the first face were matched at a rate significantly higher than chance (p's < .05). Similarly, for the Instructions-First composites, all but the first two

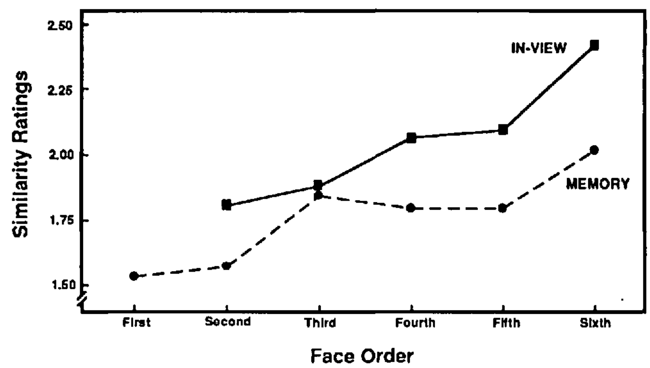

Figure 1. *Mean Similarity Rating as a function of Face Order and Presentation Conditions.*

composites were matched significantly better than chance (p's < .05).

The relationship between the matching and rating scores was also examined. For the Instructions-First and Face-First subjects' composites generated from memory, the correlation between the matching and rating scores was positive and significant, $r = .46$ ($N = 36$), $p < .01$. A similar relationship was found for the In-View composites, $r = .62$ ($N = 36$), $p < .001$. It appears that the matching and rating scores are, in part, measuring the same thing--presumably, composite accuracy. Apparently, the matching scores lack the sensitivity of the rating scores to detect the training effect. Perhaps a larger number of matching judges would have provided greater power.

DISCUSSION

Previous research has shown that current methods of composite construction do not produce accurate likenesses of the target face (Davies et al., 1978; Ellis et al., 1975, 1978; Laughery & Fowler, 1980). In addition, research (Woodhead, et al., 1979) has failed to show any effect of training on face recognition performance with the exception of being able to erase other-race deficits (e.g., Elliott et al., 1973). However, results of the present study show that training subjects on a composite system leads to higher quality composites. This is a finding heretofore unreported.

Woodhead et al. (1979) attributed their failure to find a positive effect of training to the procedure that they used. Their training procedure involved the analysis of individual features and then tested its effect using recognition. It might be speculated that the training procedure used by Woodhead et al. might have been effective, but was simply not shown because they tested their subjects with the inappropriate kind of test. In support of this notion, Wells and Hryciw (1984) present results suggesting that the selection of memory test depends on the congruency of the processes used to study face information. That is, feature training might useful for tests involving individual feature recall like composite production, but not for tests of recognition. Conversely, training to discriminate global aspects of faces might benefit recognition performance, but not tests involving individual features. Had Woodhead et al. used a composite production test, they might have found an effect of face training with the study procedure they used.

The present results showed that faces constructed with the target face in-view were rated and matched significantly better than composites generated from memory. Additional research using the Photofit and Identikit systems (Ellis et al., 1975,1978; Laughery & Fowler, 1980) found no difference in the quality of composites constructed from memory and those done while the face was in view. The present results suggest that the Mac-A-Mug Pro may be superior relative to the composite systems investigated earlier.

There were no significant differences among the three groups (Instructions-First, Face-First, and Recognition-First) on the quality of the initial composite. The failure to find a difference betwen the Instructions-First and the Face-First groups can be taken as mild evidence that learning a composite system after viewing the face does not interfere with face memory. However, the time delay from target exposure and the beginning of the composite generation was very short. Had the delay had been somewhat longer (as would occur in a real eyewitness situation), perhaps an effect of learning the system before or after target exposure would be found. Also, the failure to find a difference between the Recognition-First and the Face-First subjects suggests that a prior recognition test does not interfere with subsequent composite production.

The training effect found in the present study also might be viewed as overcoming a deficit. Face recognition is a skill used daily and is probably overlearned (Malpass, 1981). On the other hand, people are seldom required to perform a composite generation task. This may help to explain why training may facilitate composite performance but have very little or no effect on recognition performance. There is more room to improve composite production skills because performance begins at a very low level, whereas, recognition may be already near its maximum level.

It is important to note that the present study did not specifically train subjects in methods to better remember faces. Rather, it was designed to train subjects to use a composite system. That is, subjects may not have improved their face memory ability but were simply able to use the composite system more effectively to reconstruct what they had encoded. This interpretation is supported by the finding that the quality of the 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. However, it is possible that the use of the system also influenced subjects' encoding strategy to adopt a more useful method (e.g., analysis of individual features). The strategy question could be addressed in future research in two ways. First, having subjects report the encoding strategy that they use as they progress through the composite training sessions can help to determine whether it changes and whether there is any relationship of encoding strategy to composite performance. Second, if composite practice affects encoding strategy then one might expect that practice with one composite system should transfer to another system rather than the effect being system specific.

Future research should be directed at examining the parameters of the memory curve. Although brief instruction prior to exposure to the first target face did not increase the quality of the initial composites, extended practice involving the construction of multiple composites (a total of three hours across two sessions) did have a significant positive effect on quality. For a better understanding of the processes involved in composite production, several basic questions still need answers. Will more than three hours of practice further improve composite quality? For what duration is the effect maintained? Would periodic practice be useful?

Results of the present study may have an implications for law enforcement procedures and the training of potential witnesses. These results suggest that familiarization and practice with a composite system can allow the witness to produce higher quality composites. Persons at high risk for witnessing a crime (e.g., bank tellers, security guards, and convenience store clerks) could undergo training with the composite system and if called upon to generate a composite, they would be prepared to generate a more accurate likeness of the suspect than witnesses who had not been trained.

REFERENCES

Davies, G. M., Ellis, H. D., & Shepherd, J. W. (1978). Face identification: The influence of delay upon accuracy of Photofit construction. *Journal of Police Science and Administration, 6*, 35-42.

Elliott, E. S., Wills, E. J., & 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., Davies, G. M., & Shepherd, J. W. (1978). A critical examination of the Photofit system for recalling faces. *Ergonomics, 21*, 297-307.

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

Laughery, K. R., & 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. (1978). Suspect identification following exposure to sketches and Identi-kit composites. In *Proceedings of the Human Factors Society 22nd Annual Meeting* (pp. 631-635). Santa Monica: Human Factors Society.

Laughery, K. R., Smith, V. L., & 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* (pp. 302-305). Santa Monica: Human Factors Society.

Lavrakas, P. J., Buri, J. R., & Mayzner, M. S. (1976). A perspective on the recognition of other race faces. *Perception and Psychophysics, 20*, 475-481.

Malpass, R. (1981). Training in face recognition. In G. M. Davies, H. D. Ellis, & J. W. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 299-309). London: Academic Press.

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception and Psychophysics, 14*, 285-292.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 406-417.

Wogalter, M. S., Laughery, K. R., & 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., & Simmonds, D. C. V. (1979). On training people to recognize faces. *Ergonomics, 22*, 333-343.