Journal of Applied Psychology
1989, Vol. 74, No. 5, 706-713

0021-9010/89/$00.75

Enhancing Lineup Identification Accuracy: Two Codes Are Better Than One

Robert D. Melara
John B. Pierce Foundation Laboratory
and Yale University

Thomas S. DeWitt-Rickards
and Thomas P. O'Brien
University of Virginia

Ways of improving identification accuracy were explored by comparing the conventional visual lineup with an auditory/visual lineup, one that paired color photographs with voice recordings. This bimodal lineup necessitated sequential presentation of lineup members; Experiment 1 showed that performance in sequential lineups was better than performance in traditional simultaneous lineups. In Experiments 2A and 2B unimodal and bimodal lineups were compared by using a multiple-lineup paradigm: Ss viewed 3 videotaped episodes depicting standard police procedures and were tested in 4 sequential lineups. Bimodal lineups were more diagnostic than either visual or auditory lineups alone. The bimodal lineup led to a 126% improvement in number of correct identifications over the conventional visual lineup, with no concomitant increase in number of false identifications. These results imply strongly that bimodal procedures should be adopted in real-world lineups. The nature of memorial processes underlying this bimodal advantage is discussed.

The potential for fallibility in eyewitness identifications has been amply documented (e.g., Baddeley, 1979; Buckout, 1974; Bull & Clifford, 1979; Loftus, 1979; Yarmey, 1979). Recently, however, the efforts of several researchers have been directed toward developing techniques that can improve identification accuracy. Improvement can be expressed directly as an increase in the number of times eyewitnesses correctly identify a perpetrator (*hits;* e.g., Geiselman, Fisher, MacKinnon, & Holland, 1985; Malpass & Devine, 1981a). However, accuracy can also be enhanced by decreasing false identifications of innocent suspects (*false positives;* e.g., Malpass & Devine, 1981b; Wells, 1984). Wells, for example, found that the use of a *blank lineup* (i.e., a lineup that precedes the actual lineup and contains foils known to be innocent) produced no change in hit rates. However, eyewitnesses who were given a blank lineup and did not make a positive identification produced significantly fewer false-positive errors on the actual lineup, relative both to eyewitnesses who made a blank lineup identification and to eyewitnesses not given a blank lineup.

Several of the techniques to improve overall identification performance are based on prior research on memory processes. For example, Malpass and Devine (1981a) tested a guided-memory procedure that itself was inspired by laboratory demonstrations of improved memory performance when the situational context present at the time of retrieval matched that present during encoding (Smith, 1979; Smith, Glenberg, & Bjork, 1978; Tulving & Thomson, 1973; Watkins, Ho, & Tulving, 1976). The interview technique involved asking eyewitnesses to image all aspects of a staged act of vandalism as vividly as possible: the physical environment, the sequence of events, their personal reactions, and so forth. Imaging was used, therefore, to reinstate the original context of the crime without returning the eyewitness to the actual physical environment (see Smith, 1979). Malpass and Devine (1981a) found that subjects interviewed with the imaging technique produced more correct identifications and fewer false positives in the subsequent lineup than noninterviewed control subjects. The Malpass and Devine study thus provides an effective modification of a *system* variable (Wells, 1978), one that can be incorporated fairly easily into the standard criminal investigation repertoire.

The Bimodal Lineup

This study was motivated by similar concerns. First, does current laboratory research offer any suggestions for improving overall identification performance? Second, is it economically feasible to adopt these procedures, that is, to modify techniques presently in use? Our research evolved from the following observation: Currently used identification procedures (e.g., lineups) typically present retrieval cues from only a single sensory modality (e.g., visual), even though information from other sensory modalities (e.g., auditory) was also probably encoded during the criminal episode. Consider, for example, an armed robbery in which the criminal communicated his intentions verbally to his victims. In terms of the victims' memory of this criminal, coding was redundant; retrieval cues from at least two sensory modalities, visual and auditory, are potentially useful

Preparation of this article was supported by National Institutes of Health Postdoctoral Fellowship F32 HD07137-02 to Robert D. Melara and National Institutes of Health Grant NS21326 and by National Science Foundation Grant BNS84-20017 to Lawrence E. Marks.

We wish to thank Lawrence Marks, Daniel Algom, William Howell, Phoebe Ellsworth, and two anonymous reviewers for their very helpful comments on an earlier version of this article. We also wish to thank Theresa Phillips, Hilary Levyne, Liz Stack, Steve Friedman, Amy Harris, and especially Jennifer Hoffman and Holly Harper for their assistance in collecting and analyzing the data reported herein.

Correspondence regarding this article should be addressed to Robert D. Melara, who is now at Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907.

in identifying the individual. Yet the typical lineup may involve only a static visual cue (e.g., a photograph), and this single cue may itself be degraded with respect to the original context (e.g., changed hair style, new facial hair, etc.).

The potential of multimodal retrieval cues for enhancing memory performance is indicated on several theoretical and empirical grounds. First, many theorists view the memory trace as composed of multiple elements (Bower, 1967; Underwood, 1969; Wickens, 1970). In line with this view, evidence shows that the probability of successfully retrieving an event increases with the number of constituent attributes for that event that are presented at the time of testing (Bower, 1967; Martin, 1971, 1973). In the armed robbery example, the perpetrator's voice might provide several additional dimensions (e.g., pronunciation, intonation, etc.) on which an accurate positive identification could be based.

Second, recent investigations of paired-item recognition memory (e.g., memory for a pair of words) have shown that memory for one item is better if tested with its pair member (an intact pair) than if tested with the member of a different pair that was also studied (a rearranged pair; Clark & Shiffrin, 1987; Humphreys, 1976, 1978). With respect to the armed robbery example, the finding of an *intact advantage* implies that the encoding and later familiarity of the criminal's face would not be independent of the criminal's voice. Several current models of recognition memory are explicit attempts to account for the finding of an intact advantage (Gillund & Shiffrin, 1984; Humphreys, 1976, 1978; Mandler, 1980). For instance, according to Humphrey's model (Humphreys, 1976, 1978; Humphreys & Bain, 1983), the encoding process entails the storage of *relational information* (e.g., memory of the co-occurrence of face and voice) over and above the storage of individual-item information (e.g., memory of face or memory of voice). Relational information can serve, on this account, as an additional cue for retrieving the information to be remembered (e.g., the identity of the criminal).

In this regard, cross-modal information may have particularly high relational value. Recent work in the area of cross-modal perception (e.g., McGurk & MacDonald, 1976; Melara, 1989; Melara & O'Brien, 1987) has demonstrated the strong interaction of information from separate sensory modalities. For example, Melara (1989) showed that subjects are unable to attend selectively to one sensory dimension (e.g., color); that is, they suffer interference (in terms of speed and accuracy of identification) from irrelevant variation along a cross-sensory dimension (e.g., pitch).

Finally, multiple retrieval cues may be helpful for reestablishing the original encoding context, which in turn has been shown to enhance memory performance (e.g., Light & Carter-Sobell, 1970; Smith et al., 1978; Tulving & Thomson, 1973). The combination of information from two sensory modalities might be especially effective for reinstating context.

In this study we tested the possible superiority of two sensory cues over a single cue by comparing identification performance in an auditory/visual lineup with performance in either an auditory (voices) or a visual (color photographs) lineup. We hypothesized that the additional cue information accessed by a bimodal (relative to unimodal) recognition probe, either in the form of relational or extra individual-item information, would enhance the likelihood of correctly identifying the criminal without increasing the probability of committing false alarms.

Simultaneous Versus Sequential Lineups

Our use of auditory cues necessitated a variation from the usual lineup procedure. In a standard visual lineup, the eyewitness reviews all lineup members simultaneously. However, for lineups that include voice recordings (see Bull & Clifford, 1984), simultaneous presentation is disruptive and therefore impractical. As a result, suspect presentation in our study was done sequentially.

This modification raises the question of whether identification performance in a sequential lineup differs from performance in the traditional simultaneous lineup, irrespective of number and modality of cues. Research by Lindsay and Wells (1985) suggests that the two lineups do in fact lead to differences in performance (see also Cutler & Penrod, 1988). Indeed, they found that the sequential visual lineup resulted in better overall identification performance because it reduced the percentage of false-positive identifications (from 35% in the simultaneous condition to 18% in the sequential condition). Lindsay and Wells claimed that simultaneous lineups encourage a *relative judgment* cognitive set, that is, a tendency to identify positively the foil who most closely resembles the perpetrator relative to the other foils in the lineup (see also Wells, 1984). The relative judgment set, in turn, fosters proportionally greater false positive identifications. On the other hand, in a sequential lineup, the eyewitness makes a yes/no decision as each suspect is presented. This procedure minimizes relative judgments and so reduces false positives.

The Lindsay and Wells (1985) results imply that our use of a sequential procedure in the multimodal lineup is appropriate. Still, because the sequential lineup involves a fairly marked change from standard police procedure, we believe there was some merit in retesting the Lindsay–Wells findings. Thus, Experiment 1 was a replication of the comparison of simultaneous with sequential visual lineups using different stimulus materials, that is, a videotaped episode rather than a live staged crime. In Experiments 2A and 2B we tested the main hypothesis of this study—that two sensory codes are better than one—by comparing performance in unimodal sequential lineups (photographs only, voices only) with identification accuracy in a bimodal sequential lineup (photographs plus voices). In each experiment, hits and false positives were examined in tandem to ensure that a change in performance did not reflect merely an adjustment in a subject's decision criterion. Moreover, in all experiments the target may or may not be present in any given lineup. This permitted determination of lineup diagnosticity (i.e., a measure of the informativeness provided by a particular lineup procedure; see Lindsay & Wells, 1985; Wells & Lindsay, 1980) by obtaining the ratio of accurate identifications (when the target was present) to false identifications (when the target was absent). The main prediction of this study is that this ratio will be greater for bimodal than for unimodal lineups.

Experiment 1: Retesting the Sequential Lineup

In Experiment 1 we asked the following question: Do subjects commit significantly fewer false-positive errors in a sequential

Table 1
Percentages of Total Correct Identifications (Hits) and False-Positive Errors for Simultaneous and Sequential Lineup Conditions (Experiment 1)

Lineup condition	Hits %	Hits No.	False positives %	False positives No.
Target present				
Simultaneous	25	2	63	5
Sequential	13	1	25	2
Target absent				
Simultaneous			100	8
Sequential			25	2

lineup than in a simultaneous lineup? In line with the Lindsay and Wells study, we expected this change to be independent of the number of correct identifications. With regard to the aims of our study, the main objective of Experiment 1 was to demonstrate that performance in a sequential lineup was, at the very least, no worse (in terms of either hits or false alarms) than performance in a simultaneous lineup.

Method

Subjects. Thirty-two University of Virginia undergraduates volunteered to participate in this experiment.

Stimuli and apparatus. A videotape was prepared with the assistance of the Charlottesville Police Department for the purpose of this experiment. The 6-min audiovisual sequence showed an erratically driven vehicle pulled over by a uniformed Charlottesville police officer in his patrol car. The officer administered routine driving-under-the-influence (DUI) tests to the driver. The officer's face was clearly visible from several camera angles for 4 min 45 s; he spoke throughout the 6-min sequence.

Seven color photographs (6.35 cm × 7.62 cm) of the head and shoulders of uniformed police officers (including the videotaped officer) were used in either a simultaneous or a sequential lineup. The videotaped officer's photograph had been taken 4 years earlier and included a mustache not present in the videotaped sequence. The remaining photographs were selected by members of the police department on the basis of (a) the departmental guidelines for lineup construction (i.e., targets and foils matched in approximate age, facial hair, race, presence of distinguishing features and facial shape) and (b) a facial similarity between these foil photographs and the target's photograph.

A standard lineup carrier holding six photographs was used to present target or foils, or both, in the simultaneous condition. Half of the subjects were given a simultaneous lineup, and half were given a sequential lineup. Within each condition, the target was present in half of the lineups. In the remaining lineups, the target's photograph was replaced with an additional foil photograph. The target never appeared in either the first or last position of either lineup. The target's photograph (and that of his foil substitute) appeared in Positions 2 through 5 an equal number of times across subjects.

Procedure. The videotape sequence was presented to small groups of subjects (1 to 3). An incidental learning paradigm was used: The videotape was described as a training film for new police officers. Subjects were instructed to attend to all cues that might be used to identify intoxicated drivers. Following the videotape, a questionnaire was administered to each subject; it contained a series of questions regarding DUI testing. When 5 min had elapsed, subjects were escorted to separate rooms for individual testing. Once there, subjects were instructed to identify the officer in the videotape and to rate the certainty of their decision on a 1–7 scale. The leftmost pole of the scale (1) was additionally labeled *not at all certain,* the rightmost pole (7) *very certain.* Certainty ratings were obtained for all members of the sequential lineup, but only for positively identified members of the simultaneous lineup. All subjects were given the following no-choice option prior to lineup presentation: "Remember, as in a real identification procedure, the guilty party may or may not be present" (Malpass & Devine, 1981b; see also Lindsay & Wells, 1985).

Experimenters were kept unaware of the correct identification. In the sequential condition, the experimenter held a stack of 12 photographs, showed the top photograph to the subject, and placed that photograph at the bottom of the stack after the subject had responded. Similarly, the subject's response sheet contained 12 yes/no boxes and confidence scales. Thus, subjects had little indication of the actual number of photographs that comprised a sequential lineup. The entire experiment lasted about 20 min.

Results

The percentages of hits and false-positive errors in the simultaneous and sequential lineups appear in Table 1. In the simultaneous condition, subjects wrongly identified a foil on 13 of 16 lineups (81%), whereas subjects given sequential lineups made false identifications on only 4 of 16 lineups (25%), $\chi^2(1, N = 32) = 28.73, p < .05$. The differences were greater when the target was present, but both groups reached statistical significance when analyzed separately: target-present lineups (63% vs. 25%), $\chi^2(1, N = 16) = 15.23, p < .05$; target-absent lineups (100% vs. 25%), $\chi^2(1, N = 16) = 43.81, p < .05$.

Subjects were poor at identifying the target; only 3 of the 16 subjects correctly identified the police officer.[1] Two of these received the simultaneous lineup, and the other received the sequential lineup, $\chi^2(1, N = 16) = 3.52$, *ns.* Moreover, there was no relationship between subjects' accuracy and their confidence ratings (point-biserial $r = -.09$), although there was a slight tendency for subjects receiving the sequential lineup (4.6) to be more confident than those receiving a simultaneous lineup (3.5), $F(1, 30) = 3.63, p = .10$.

Discussion

Experiment 1 replicated the findings of Lindsay and Wells (1985) in demonstrating that eyewitnesses are less likely to falsely accuse a suspect when lineup members are presented serially rather than when the traditional simultaneous method of presentation is used. In terms of our purposes in this study, these results justify the use of the sequential lineup when testing the *bimodal hypothesis,* that is, the hypothesis that two sensory codes are better than one. We tested this hypothesis in Experiments 2A and 2B.

[1] This result reflects very poor identification accuracy indeed, even relative to the performance of subjects in other eyewitness identification studies. For example, across a sample of nine experiments assessing eyewitness accuracy that we cite in this article, 46% of the subjects correctly identified the target (range from 27% to 71%). At present, we have no explanation concerning why our subjects displayed such poor performance.

Experiments 2A and 2B: Testing the Bimodal Hypothesis

In Experiment 1, subjects viewed a single event and made identification judgments in a single lineup. The single-event/single-lineup paradigm is the one most commonly used in lineup investigations (e.g., Brigham, Maass, Snyder, & Spaulding, 1982; Kassin, 1985; Lindsay & Wells, 1985; Malpass & Devine, 1981a; Murray & Wells, 1982; Pigott & Brigham, 1985; Wells, 1984). This paradigm allows a close match between the laboratory lineup situation and real-life lineups. Still, the measurement reliability for individual eyewitnesses is probably less stable in the single-event design than if each eyewitness had been tested in multiple lineups. The latter, statistically more sensitive design may be more desirable for some lineup investigations (e.g., research programs in their early stages). Moreover, the multiple-lineup procedure permits the use of more powerful tests of identification performance, such as parametric techniques or analyses of signal detection theory. Accordingly, in Experiments 2A and 2B we used a multiple-measures design to test the bimodal hypothesis. Each subject was presented with three videotape sequences and was tested with 4 six-member sequential lineups. Thus, each subject made 24 yes/no decisions and provided 24 seven-point confidence ratings. In Experiment 2A we examined performance in the auditory/visual lineup relative to the auditory lineup alone and to the visual lineup alone. In Experiment 2B we replicated the comparison of the standard photograph lineup with a cross-sensory lineup by using a different but comparable set of distractor stimuli.

Method

Subjects. In Experiment 2A we tested 72 University of Virginia undergraduates who participated in partial fulfillment of a course requirement. We tested 80 subjects in Experiment 2B. No subject participated in both Experiments 2A and 2B (or had participated in Experiment 1).

Stimuli and apparatus. The three videotapes were played on a Samsung VT226T video cassette recorder and a Magnavox 8CM515 A/V monitor. All actors in the videotapes were actually officers in the Charlottesville Police Department. The "DUI" videotape from Experiment 1 was used in these experiments. In the 5-min "Holdup" videotape, a plain clothes detective conducted a staged interview with a man who, it was learned, was the victim of an armed robbery at a local fast food restaurant. The two men sat at a 60° angle to each other; both faces were in full view during the entire sequence. In the "Shoplift" videotape, an actor dressed in casual attire pilfered several items of merchandise from a local department store and was apprehended subsequently by the store detective. This film was narrated; the narrator, who identified himself as the shoplifter/actor, detailed a set of police guidelines for apprehending shoplifters.

Four lineups were constructed for each of three conditions (Experiment 2A: visual, auditory, and auditory/visual) or two conditions (Experiment 2B: visual and auditory/visual). Each lineup involved one of four identifications, as follows: (a) the police officer in "DUI," (b) the interviewer in "Holdup," (c) the interviewee in "Holdup," and (d) the shoplifter in "Shoplift." Visual lineups were constructed in the same manner as in Experiment 1: Twenty-four 6.35 × 7.62 cm color photographs of uniformed police officers were selected on the basis of their facial similarity to the four videotaped actors/officers. Each lineup contained six members; one photograph was used to replace the target in target-absent lineups. A different set of visual foils was used in Experiments 2A and 2B. However, the foils' rated similarity to their corresponding targets (as assessed by 10 independent judges) was approximately equal across the two experiments, $t(9) = 1.22$, *ns*.

To construct the auditory lineups, the four actors each produced a voice sample. Each target voice stimulus was matched to six similar-sounding (i.e., in terms of accent, pitch, etc.) foil voice stimuli. With all voice stimuli (target and foil), the speaker created a fictitious name, address, birth date, and social security number (e.g., "My name is John Smith. My address is 123 Park Avenue in New York City. My date of birth was January 1st, 1950. My social security number is 123-12-1234"). Thus, the temporal length of each recording was equivalent, that is, about 15 s. Auditory lineups were recorded on a Quasar VP5430WQ recording system and presented by using a Realistic 2000 cassette player.

The auditory/visual lineups were created by combining (i.e., presenting simultaneously) a photograph with a voice stimulus. The auditory stimulus set was the same in both Experiments 2A and 2B. The target's voice was always matched with his photograph; for foils, voices and photographs were combined randomly.

Procedure. Subjects (in groups of 1 to 3) were told that (a) they would watch a set of three films, (b) each film contained information about standard police procedures, and (c) they would be asked several questions about the films at the conclusion of the set. The order of videotape presentation was counterbalanced across subjects.

Lineups were always presented in the same order as the videotapes. The target of the lineup was referred to by the role he played in the videotape (e.g., "the witness to the robbery in the Holdup film"). Subjects were warned that certain physical characteristics of the target (e.g., hairstyle, age, voice quality) might be different in the lineup relative to the videotape. All subjects were given the no-choice option. For each lineup member, each subject made a yes/no decision and provided a certainty rating on a 12-line response sheet.

Each subject was presented with four lineups, but all of one type (e.g., visual, auditory, or auditory/visual; target-present or target-absent). Thus, in Experiment 2A, one third of the subjects saw only visual lineups, one third heard only auditory lineups, and one third saw and heard only auditory/visual lineups. In Experiment 2B, half of the subjects received four visual lineups, and half received four auditory/visual lineups. In addition, half of the subjects were presented with lineups in which the target was always present, whereas the remaining subjects were given only target-absent lineups. In target-present lineups, the target's photograph or voice, or both, never appeared in the first or last lineup position and was counterbalanced across subjects for the remaining positions. Subjects were allowed to examine each photograph for as long as they wished, but each voice was played only once. No photograph (in the visual condition), voice stimulus (auditory condition), or photo–voice combination (auditory/visual condition) was used in more than one lineup. Each session lasted about 55 min.

Results

For both experiments, a series of chi-square analyses was conducted on frequencies of false-positive identifications in target-absent lineups as a test of the perceived similarity of the foil stimuli. The four lineups (DUI, shoplifter, interviewer, interviewee) did not differ in rates of false-positive errors: In Experiment 2A, $\chi^2(3, N = 72) = 1.61$, *ns*; in Experiment 2B, $\chi^2(3, N = 80) = 7.72$, *ns*. No foil in either visual or auditory lineups was falsely identified more often than other lineup members. Moreover, false identifications were not made more often for any one lineup serial position relative to the other positions. It is reasonable to conclude, therefore, that our lineups were fair and unbiased.

Table 2
Percentages of Total Correct Identifications (Hits) and False-Positive Errors for Visual, Auditory, and Auditory/Visual Lineup Conditions (Experiments 2A and 2B)

	Hits		False positives	
Lineup condition	%	No.	%	No.
Target present				
Experiment 2A				
Visual	29	14	67	32
Auditory	60	29	63	30
Auditory/visual	44	21	38	18
Experiment 2B				
Visual	19	15	54	43
Auditory/visual	59	47	36	31
Target absent				
Experiment 2A				
Visual			44	21
Auditory			85	41
Auditory/visual			69	33
Experiment 2B				
Visual			68	54
Auditory/visual			64	51

Analysis of Experiment 2A

Percentages of false identifications and correct identifications for target-present and target-absent lineups in each of the three conditions (visual, auditory, auditory/visual) appear in Table 2 (for both Experiments 2A and 2B). In Experiment 2A, correct identifications were made in 44% of target-present lineups, and false identifications were made in 61% of all lineups. Inspection of Table 2 reveals a trade-off between hits and false alarms: Improvement in correct identifications was sometimes accompanied by an increase in false identifications, presumably reflecting a low decision criterion. Specifically, the auditory condition produced the largest percentage of hits (60%) but also the largest number of false positives (74%). The presence of different decision criteria indicates that identification accuracy should be considered jointly with identification inaccuracy in the analyses that follow. The low criterion in auditory lineups was accompanied by another finding of interest: On a rather large proportion of auditory target-present lineups (31%), subjects made a false identification in addition to a correct identification. This latter result is appreciated readily in Table 2: The percentage of total identifications in this condition (hits and false alarms) exceeded 100%.

A multivariate analysis of variance (MANOVA) conducted on the percentages of hits and false-positive errors of subjects in the target-present group showed a significant effect of lineup condition, $F(4, 66) = 3.62$, $p < .05$. A univariate analysis of variance (ANOVA) conducted on the percentages of hits also revealed a difference among lineups, $F(2, 33) = 4.47$, $p < .05$. A Newman-Keuls post hoc test of this effect (.05 criterion level) indicated that the auditory lineup led to significantly more hits (60%) than did the visual lineup (29%), with performance in the bimodal lineup falling between the two (44%).

A two-way ANOVA was conducted on percentages of lineups in which a false-positive error was made; Lineup Type (visual, auditory, auditory/visual) and Target Status (present, absent) were factors. There was a significant main effect of lineup, $F(2, 66) = 3.89$, $p < .05$. A Newman-Keuls analysis (.05 level) showed that the auditory condition led to significantly more false-positive lineups (74%) than either the visual (57%) or the auditory/visual (54%) conditions. Target presence had little effect on false identifications, $F(1, 66) = 2.34$, *ns*. However, the Lineup × Target interaction reached significance, $F(2, 66) = 6.12$, $p < .05$. As is evident in Table 2, these factors interacted because subjects in visual lineups produced relatively more identification errors when the target was present than when absent (69% vs. 44%), whereas the reverse was true for subjects in bimodal lineups (38% vs. 69%).

Analysis of Experiment 2B

The results seem more clear-cut in Experiment 2B. A MANOVA of hits and false positives in the target-present group revealed a significant main effect of lineup condition, $F(2, 37) = 13.89$, $p < .05$. Univariate F tests indicated that the two lineups had significantly different hit rates (visual = 19%; auditory/visual = 59%), $F(1, 38) = 28.36$, $p < .05$. A two-way ANOVA of the percentage of false-positive lineups found a main effect of target (target present = 45%; target absent = 66%), $F(1, 76) = 11.78$, $p < .05$; and a marginally significant difference between lineups (visual = 61%; auditory/visual = 50%), $F(1, 76) = 3.16$, $p < .10$. There was no Lineup × Target interaction, $F(1, 76) = 1.34$, *ns*.

In neither Experiment 2A nor 2B were individual differences a salient factor. For example, there was no evidence that some subjects made identifications on all lineups and others made identifications on no lineups. In both experiments, the type of lineup presented appeared to be the best predictor of identification performance.

Data on confidence ratings were largely uninformative. Experiment 2A produced a modest relationship between certainty scores and percentages of false-positive errors (Pearson $r = -.23$). But this finding was not replicated in Experiment 2B ($r = -.02$). Mean confidence and percentage of hits correlated oppositely in Experiment 2A ($r = -.19$; i.e., high confidence was associated with poor accuracy) but appropriately in Experiment 2B ($r = .33$).

Discussion

The main results of Experiments 2A and 2B are easily summarized, as follows: The bimodal lineup led to substantially more correct identifications than the visual lineup, and false alarms were committed on slightly more visual than bimodal lineups. These results emerged more strongly in Experiment 2B than in 2A, a difference that seems best attributed to sampling variation because percentages of both correct and false identifications were, on average, equivalent across the two experiments (suggesting that performance was not altered by differences in the stimulus sets). Nonetheless, a combined analysis of the percentages of hits in Experiments 2A and 2B showed that the auditory/visual condition (52%) led to substantially greater accuracy overall than the visual condition (24%), $F(1, 62) = 22.95$, $p < .05$. Moreover, lineup diagnosticity (i.e., the ratio of hits to target-absent false alarms; see Wells & Lindsay, 1980, and Table 2) was substantially higher for auditory/visual lineups (.77) than for visual lineups (.44). Taken together, these findings provide strong support for our bimodal hypothesis.

The high percentage of hits made in auditory lineups (60%) was obtained at the cost of the greatest number of false identifications. It is important to note that the low criterion found in auditory lineups did not extend to auditory/visual lineups, the latter being the lineups with which we were most concerned in this research. The fact that auditory/visual lineups engendered the fewest false identifications, given this potential for a low criterion from the auditory component, makes our conclusion that a bimodal lineup is superior to a standard visual lineup appear even more compelling.

The low criterion found in auditory lineups is perhaps a function of earwitnessing per se. Bull and Clifford (1984) reported that in their auditory lineups, subjects committed false-positive errors on 87% of target-absent trials, despite the use of a no-choice option. As in our experiment, they found a hit rate of 60%, suggesting that a trade-off was operating. Their results, in conjunction with ours, suggest that a low criterion may be an intrinsic characteristic of the auditory lineup.

In both experiments, the percentage of lineups in which a false-positive error was committed was smaller when the target was present (Experiment 2A = 57%; Experiment 2B = 45%) versus absent (66% in both experiments), a result that might reflect a hypothesis on the part of subjects that the target would be presented. Moreover, when the target was present, there was a tendency for the percentage of lineups having a false positive to be higher in the visual condition (Experiment 2A = 69%; Experiment 2B = 54%) than in the bimodal condition (Experiment 2A = 38%; Experiment 2B = 36%), indicating an additional and unexpected benefit of bimodal over traditional lineups. One anomalous finding was the relatively few false-positive lineups that were obtained in Experiment 2 when the target was absent and the lineup was visual (i.e., 44%; see Table 2). We have no ready explanation for that finding.

In neither experiment was a strong relationship found between subjects' accuracy of identification and their rated confidence. Furthermore, even these small correlations showed little consistency between experiments. Previous research on the accuracy–confidence correlation has been mixed (see Wells & Murray, 1984, for a review); most investigators report slight correlations (i.e., below .40; e.g., Brown, Deffenbacher, & Sturgill, 1977; Clifford & Scott, 1978; Lindsay & Wells, 1985; Lindsay, Wells, & Rumpel, 1981). Yet, some have found moderate correlations (i.e., between .40 and .70; e.g., Brigham et al., 1982; Kassin, 1985; Malpass & Devine, 1981b; Pigott & Brigham, 1985); others have obtained evidence that a substantial relationship exists (i.e., above .70; see Malpass & Devine, 1981a); and still others have reported negative correlations (e.g., Buckout, 1974; Loftus, Miller, & Burns, 1978). The basis of these discrepancies is not clearly understood and is likely complex. Therefore, we resist speculating about the nature of the small, often contradictory results obtained with confidence ratings in these experiments.

Finally, we were surprised by the large number of lineups in which subjects failed to use the no-choice option (i.e., made a positive identification) in these experiments: 66% of the target-absent lineups, compared with only 25% in Experiment 1 and 35% in the Lindsay and Wells (1985) experiment. We suspect that the present finding may arise partially from the nature of our multiple-lineup procedure. Perhaps in such extended decision situations, witnesses find it difficult to avoid making a positive identification. If this interpretation is valid, it could have considerable practical import because, in real identification situations, it is not unusual for an eyewitness to be exposed to several police lineups.

General Discussion

Collectively, these experiments have demonstrated that a sequential bimodal lineup leads to a greater number of correct identifications than does the conventional visual lineup (presented sequentially), with no concomitant increase in the number of false identifications. Indeed, a bimodal lineup was less likely to be marred by a false-positive error than a standard lineup. In the terminology of signal detection theory, the two-code lineup appears to enhance sensitivity to the signal without altering the decision criterion. Furthermore, the use of a sequential lineup format, a necessity for auditory/visual presentations, is not detrimental to accuracy and, in fact, probably reduced false identifications by minimizing relative judgments. It is interesting that the large improvement wrought by this stimulus change (two signals vs. one signal) appeared to go largely unnoticed by subjects, as measured by the absence of reliable differences in confidence ratings. Apparently, therefore, the effects of this powerful stimulus manipulation took place outside subjects' conscious awareness, a finding that may have important legal implications (see, e.g., Wells, Lindsay, & Ferguson, 1979; Wells & Murray, 1983, 1984).

The Applied Use of Bimodal Lineups

The goal of this research was to investigate improvements in the lineup procedure that are both practical and economically feasible. The dual-code lineup has been shown to produce more efficient performance than the conventional single-code lineup: The bimodal lineup led to a 126% improvement in hit rate over the visual lineup (across Experiments 2A and 2B). Such large improvement justifies some expense in modifying currently used procedures. We believe that these additional costs could be easily absorbed by most moderate- and large-sized police departments.

The sampling of auditory (voice) information could be incorporated into the normal booking process. At that time, accused individuals would be instructed to recite a standard phrase in their normal speaking voice (e.g., the name, social security number, etc., sequence used in this study, or a common verse such as the Pledge of Allegiance). Further testing is suggested to determine the optimal length of the auditory signal for witness performance; greater accuracy may result from stimuli that are longer than the 15-s voice samples used in our study (see Bull & Clifford, 1984). Some precaution is needed to guard against the purposeful disguising by the accused of his/her normal speaking voice; an altered voice is the auditory analogue of changed facial features in visual lineups. Finally, cost will be incurred in the effort to establish fair, unbiased lineups whose voice stimuli are similar in important respects. Therefore, voice recordings would need to be catalogued in a tape library, matched according to pitch, accent, tempo, and so on. Larger police departments could conceivably develop such libraries from their own pool of booked individuals. To establish libraries in smaller departments, recordings will probably need to be supplied externally.

The Nature of the Bimodal Advantage

Why are two codes better than one? Previous laboratory research (as reviewed early in this article) suggests two different explanations. On the one hand, improvement might be due exclusively to the mere presence of an additional cue. In the bimodal condition more cues are available, so the probability of a correct identification is enhanced. Thus, according to this view, the cues *act independently* and so would be equally effective whether they were administered separately or jointly.

On the other hand, improvements in recognition memory for the bimodal condition may be due primarily to the presence of relational cues that occur when modal cues are presented coincidentally. According to this account, cues from separate modalities *interact,* and it is the emergent properties arising from this interaction that lead to the bimodal (intact) advantage. Thus, the joint presentation of cross-modal information is critical.

Of course, the two explanations are not mutually exclusive. In fact, it would be surprising if the presence of additional cues, presented either separately or coincidentally, did not improve identification accuracy. The critical question concerns the incremental improvement that is produced by joint presentation. The answer to this question has considerable practical import. For example, Wells (1988) recommended separate presentation of auditory and visual lineups as a way to identify which of these cues led to a hit in any given lineup. Notice, however, that if the bimodal advantage is primarily (or even substantially) due to the presence of relational information, then asynchronous presentation deprives the police investigator of a potentially powerful tool.

In this experiment we cannot answer the question of whether the bimodal improvement we found was due to cue interaction. However, it is clear from previous laboratory investigations that (a) cross-sensory information does interact perceptually (e.g., Melara, 1989; Melara & O'Brien, 1987) and (b) relational information does aid recognition memory (Gillund & Shiffrin, 1984; Humphreys, 1976, 1978). Thus, in lieu of future lineup investigations directed at this question, we believe it is reasonable to conclude that cue interaction was at least partly involved in the bimodal advantage we found.

References

Baddeley, A. (1979). Applied cognitive and cognitive applied psychology: The case of face recognition. In L. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary* (pp. 367–388). Hillsdale, NJ: Erlbaum.

Bower, G. H. (1967). A multi-component theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 1, pp. 229–325). New York: Academic Press.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology, 42,* 673–681.

Brown, E. L., Deffenbacher, K. A., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology, 62,* 311–318.

Buckout, R. (1974). Eyewitness testimony. *Scientific American, 231,* 23–31.

Bull, R., & Clifford, B. R. (1979). Eyewitness memory. In M. M. Gruneberg & P. E. Morris (Eds.), *Applied problems in memory* (pp. 151–184). London, England: Academic Press.

Bull, R., & Clifford, B. R. (1984). Earwitness voice recognition accuracy. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 92–123). New York: Cambridge University Press.

Clark, S. E., & Shiffrin, R. M. (1987). Recognition of multiple-item probes. *Memory & Cognition, 15,* 367–378.

Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology, 63,* 352–359.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73,* 281–290.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology, 70,* 401–412.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Humphreys, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition, 4,* 221–232.

Humphreys, M. S. (1978). Item and relational information: A case for context independent retrieval. *Journal of Verbal Learning and Verbal Behavior, 17,* 175–188.

Humphreys, M. S., & Bain, J. D. (1983). Recognition memory: A cue and information analysis. *Memory & Cognition, 11,* 583–600.

Kassin, S. M. (1985). Eyewitness identification: Retrospective self-awareness and the accuracy–confidence correlation. *Journal of Personality and Social Psychology, 49,* 878–893.

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 1–11.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 3,* 556–564.

Lindsay, R. C. L., Wells, G. L., & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations? *Journal of Applied Psychology, 66,* 79–89.

Loftus, E. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 19–31.

Malpass, R. S., & Devine, P. G. (1981a). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482–489.

Malpass, R. S., & Devine, P. G. (1981b). Guided memory in eyewitness identification. *Journal of Applied Psychology, 66,* 343–350.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

Martin, E. (1971). Verbal learning theory and independent retrieval phenomena. *Psychological Review, 78,* 314–332.

Martin, E. (1973). Memory and negative transfer. *Memory & Cognition, 1,* 495–498.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 69–79.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General, 116,* 323–336.

Murray, D. M., & Wells, G. L. (1982). Does knowledge that a crime was staged affect eyewitness performance? *Journal of Applied Social Psychology, 12,* 42–53.

Pigott, M., & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology, 70,* 547–555.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460–471.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6,* 342–353.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review, 76,* 559–573.

Watkins, M. J., Ho, E., & Tulving, E. (1976). Context effects in recognition memory for faces. *Journal of Verbal Learning and Verbal Behavior, 15,* 505–517.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 12,* 1546–1557.

Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89–103.

Wells, G. L. (1988). *Eyewitness identification: A system handbook.* Carswell Legal Publications.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88,* 776–784.

Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64,* 440–448.

Wells, G. L., & Murray, D. M. (1983). What can psychology say about the *Neil v. Biggers* criteria for judging eyewitness accuracy? *Journal of Applied Psychology, 68,* 347–362.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). New York: Cambridge University Press.

Wickens, D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

Yarmey, A. D. (1979). *The psychology of eyewitness testimony.* New York: Free Press.

Received March 1, 1988
Revision received January 24, 1989
Accepted January 31, 1989 ■