Harry Hollien,[1] *Ph.D.; Gerald Bennett,*[1] *M.A., J.D.; and M. P. Gelfer,*[1] *M.A.*

Criminal Identification Comparison: Aural Versus Visual Identifications Resulting from a Simulated Crime

REFERENCE: Hollien, H., Bennett, G., and Gelfer, M. P., **"Criminal Identification Comparison: Aural Versus Visual Identification Resulting from a Simulated Crime,"** *Journal of Forensic Sciences*, JFSCA, Vol. 28, No. 1, Jan. 1983, pp. 208-221.

ABSTRACT: This research was undertaken to investigate and contrast visual and aural/perceptual identifications of a previously unknown individual from a set of photographic and tape-recorded exemplars following a simulated crime. All participants were volunteers for the "criminal," the victim of an "assault," and all suspects drawn from a Reserve Officer Training Corps (ROTC) class; the "witnesses" were 61 students in a law class. The student/witnesses were divided into four groups. Group A made three identifications (serially) one day, one week, and two weeks after the crime took place. Group B saw the photographs and heard the tapes only once a week after the incident, and Group C only after two weeks had elapsed. Witnesses in a fourth group (D) followed the same schedual as did Group A; however, they were presented with foils similar in appearance and speech to the criminal. Group D was presented both the foil and criminal in the photographic lineup at the final judging session. The results demonstrated that visual identification can be quite accurate although not consistently or predictably so. By contrast, aural/perceptual identifications were relatively poor. No strong trends for latencies were observed, either for repeated trials or for procedures involving different initiation latencies; nor did confidence levels appear to be related to accuracy of judgment. Finally, when a similar looking foil was included in the identification task, there was a weak trend for the foil to be chosen more often in subsequent trials. These results support the position that eyewitness—and especially earwitness—testimony should be viewed by judges and juries with greater caution than has been the case in the past; by appropriate instructions, juries should be given assistance in interpreting and in assigning appropriate weight to this kind of testimony.

KEYWORDS: criminalistics, criminal identification, witnesses

In a substantial number of criminal trials, guilt is determined on the basis of eyewitness or earwitness testimony. The witnesses, who were physically present at the commission of the crime, will clearly and unequivocally identify the defendant seated in the courtroom as the individual whom they saw or heard committing the offense. Such testimony has been held sufficient, in and of itself, to sustain a defendant's conviction. Indeed, it is often considered to be the most highly persuasive evidence that can be presented to a jury. Yet, while the witness

Presented at the 34th Annual Meeting of the American Academy of Forensic Sciences, Orlando, FL, 8-11 Feb. 1982. Received for publication 12 April 1982; accepted for publication 22 June 1982.

[1]Professor of Linguistics, Speech, and Criminal Justice and founding director of the Institute for Advanced Study of the Communication Process, professor of Law at the Holland Law Center, and research assistant, respectively, University of Florida, Gainesville, FL.

may have observed the criminal clearly, or for a relatively long period of time, or both, he or she may not have seen or heard the criminal clearly at all or may not be able to recall those events that were perceived. To further complicate matters, jurors are instructed to evaluate the testimony of the witness by judging his or her opportunity to observe, and ability to recall, the relevant events. Unfortunately, they are provided no specific guidance as to the criteria that they should use in these evaluations.

But what level of accuracy can be expected from these eyewitness and earwitness accounts? Are human observers capable of remembering what they have seen or heard (possibly under adverse circumstances) with sufficient detail to be able to accurately identify a criminal? Certain assumptions have been made by the courts in answer to these questions—assumptions generally based on legal precedent rather than on relevant scientific evidence. The lack of empirical foundation for judging eyewitness or earwitness credibility is unfortunate as the emerging experimental data have not consistently supported the underlying legal assumptions.

Aural/Perceptual Identification

McGehee's study [*1*] shows that the aural/perceptual approach to speaker identification—the identification of a talker from his voice alone—has had a long history of acceptance in the courts. It has been conceded generally that the identification of speakers can be accurate if the listener is previously well acquainted with the speaker's voice. Indeed, during the century preceding McGehee's review, identification of a known individual by voice alone had been admitted into court as evidence many times. In some cases, even the identification of a previously unknown speaker had been ruled admissible, although the weight of such evidence depended upon the jury's assessment of the witness's capacity to make the judgment accurately. Admissibility of aural/perceptual testimony may be traced back to the year 1660, to an instance where voice identification was offered in the case of one William Hulet, and has since been generally accepted in courts both in the United States and Britain [*2*]. In the State of Florida, for example, acceptance of aural/perceptual identification testimony is first noted in the year 1907, in a case in which a hitherto unknown, unseen, cross-racial defendant was identified as a rapist by his having spoken two sentences: "I have got you now," and "I don't want your money." The Court's decision is explained by a sort of innate, experiential logic:

> ...The manner, time and place of his assault upon her threw her instantly into the highest state of terror and alarm, when all of her senses and faculties were at the extreme of alert receptiveness, when there was nothing within her reach by which to identify her assailant but his voice. Who can deny that under these circumstances that voice so indelibly and vividly photographed itself upon the sensitive plate of her memory as that she could forever afterwards promptly and unerringly recognize it on hearing its tones again... [*2*].

This early decision is accepted today as appropriate legal precedent for the admissibility of aural/perceptual testimony in courts in the State of Florida. Similar decisions may be found throughout the United States [*3*].

Contemporary research, however, tends to substantiate McGehee's conclusions and to indicate that reliance on aural/perceptual identifications, particularly those of a previously unknown subject, may be misplaced. For example, Bricker and Pruzansky [*4*] report 98% correct identification of familiar speakers by listeners when sentences were provided as stimuli; identification accuracy fell to only 56%, however, when isolated vowel samples were substituted for the sentences. Iles [*5*] obtained similar results. The Pollack et al [*6*] experiments with speech sample duration suggest that identification accuracy improves with increasing duration up to about 1200 ms; for longer periods, accuracy did not appear to be related to duration, but rather to phonemic repertoire. These researchers also found that when they

degraded the speech signal by increasing the number of speakers or by substituting whispered speech for normally spoken speech, listeners needed longer samples to identify known talkers. In a later study, Hollien et al [*7*] report high levels of correct identification of known talkers under normal (98%) and stressful (97%) speaking conditions, but lowered accuracy for disguise (79%).

When speakers unfamiliar to the listeners are used, the cited relationships are degraded considerably. For example, McGehee [*1*] found that identification performance was reduced as a function of time (that is, identification accuracy fell from 80% the first week to 13% after three months); Brown [*8*] suggests that this decay process is a complex one which is governed by an individual's ability to store information relative to short-term and long-term memory. Several authors also report that speaker disguise, dialects, noncontemporary samples, and a larger number of speakers reduce identification accuracy and yet other investigators agree [*5-7, 9*]. Further, Rothman [*10*] reports data about "sound-alikes" and noncontemporary speech samples. He found that when noncontemporary samples of the same speakers were played, listeners exhibited 58% error rates—accuracy was at the same low level when the speaker's voice was paired with a similar sounding voice. Hollien et al [*7*] also found that listeners could identify a particular unfamiliar voice at only about 40% accuracy, and when stress and disguise were added to their paradigm, accuracy fell to 31 and 21%, respectively. The values were even lower when they studied listeners who were unfamiliar with both the speakers and the language spoken. Both Hollien et al [*7*] and Rosenberg [*11*] found great variability among their listeners' ability to make correct identifications.

Although variation in the methods used in the experiments cited can make comparisons difficult, it does appear that, under ideal conditions, listeners who know talkers well can identify them at reasonably high levels of accuracy. Auditors who do not know the speakers cannot be expected to perform at similarly high levels even after having been exposed to some sort of training procedure. Moreover, with the addition of any type of complicating factor—speaker or system distortions, for example—speaker identification by aural/perceptual means can be expected to drop precipitously.

Finally, it must be stressed that virtually all research on the aural/perceptual approach to speaker identification has been carried out under laboratory conditions—a situation that makes it difficult to generalize to the forensic science model. Almost no research relevant to auditory identifications as they relate to court cases had been carried out. Yet voice identification is being used in the courts, often in a manner similar to the traditional "lineup" of photographs or suspects. Recent literature in this area [*12-14*] has suggested that these auditory or earwitness lineups are difficult to structure properly and can produce results that are confusing or misleading.

Visual/Perceptual Identification

The process of visually identifying a criminal from a pool of suspects is a long established practice. According to Buckhout and Freire [*15*], "the crucial point in many criminal investigations occurs when a suspect is formally identified by a witness in the showup, lineup or photo array" and Loftus [*16*] warns that eyewitness testimony may prove to be of considerable importance to a jury, regardless of the circumstances under which the crime was committed. She cites an experiment in which a simulated crime was committed and a suspect was tried by a student jury. The students "convicted" the defendant 68% of the time even though the single eyewitness, who was shown to have only 20/400 vision, was not wearing glasses at the time of the crime. From this, and other experimental evidence, it would appear that the ability of an individual to recognize a person once seen, even if only briefly and under adverse conditions, carries great weight in American courts, as well as with the general public. Indeed, according to Buckhout [*17*], the ability to visually identify a criminal is one of the foundations of our system of justice.

Research results have cast some doubt on the belief that accurate visual identification is as easily accomplished as would be suggested by judicial precedent. Buckhout et al [*18*] simulated a crime before a class of student witnesses who later were shown two videotaped lineups—the criminal was present in one but not in the other. Witnesses were asked to identify the criminal, or to indicate that he was not in the lineup, and to specify how confident they were in their choice both before and after their judgments. The investigators found that only 13.5% of their subjects identified only the criminal and did not impeach their choice by choosing a second but innocent suspect; they were mistaken 40.3% of the time and there were 19.2% nonidentifications. Pre- and post-viewing confidence levels were about the same when the choices were correct but, strangely, confidence levels increased when witnesses impeached their correct choice. These data appear to be at odds with the commonly held belief that witnesses are able to remember the face of a criminal and identify it with reasonably certainty at a later time, and that confidence levels are directly related to accuracy of choice.

Buckhout and Figueroa [*19*] also investigated the ability of witnesses to identify a criminal following a simulated crime when intentional bias was introduced during the identification process. The experimental situation was similar to that cited above. In this case, however, witnesses were provided photographic spreads of suspects where (1) the photographs of the criminal was misaligned, (2) the verbal instructions were biased toward the criminal, or (3) both; a fourth group served as controls. The results were as follows: the control group chose the correct criminal 40.6% of the time whereas the mean score for those receiving the double bias was 61.3%. The results were somewhat ambiguous for those groups receiving only one of the two biases, that is, 46.9% when the photographs were biased and 37.5% when the instructions were biased. Buckhout and Figueroa interpreted these results to support the hypothesis that biasing in photographic lineups can increase the chances of a particular suspect being identified as the criminal. However, our additional interpretation of these data is that an observer needs substantial biasing if the selection process is to be significantly influenced.

Finally, Williams [*20*] investigated the effect of verbal interrogation versus written descriptions of a simulated crime and how these factors might affect visual identification. She used 121 student witnesses: half responded verbally to questions about a 9-s film of simulated crime while the other half filled out written questionnaires. At this juncture, witnesses were asked to identify the criminal from a photo spread. The results demonstrated that, although the verbally interrogated group remembered a greater number of relevant details, the groups were not significantly different in their correct identifications. Indeed, neither did very well with respect to the identification task—perhaps because of the brevity of the motion picture.

A number of experiments have been carried out which are designed to evaluate those factors that tend to enhance or degrade eyewitness testimony. Some studies attempt to assess the strategies used by the witnesses; many of these factors/strategies can be found summarized in Buckhout [*21*] and Ellis [*22*]. To be more specific, included among the factors that serve to enhance the accuracy of eyewitness testimony are: (1) race: individuals generally are better at identifying members of their own race [*23–25*]; (2) sex: females tend to be better at identifying females whereas males appear not to show a sex bias [*23*]; (3) attractiveness: individuals who are substantially more or less attractive than the general population appear to be easier to identify [*23,26*]; (4) age: older individuals tend to be somewhat more identifiable than do younger subjects [*23*]; and (5) clarity of observational field: as would be expected, poor lighting and poor sight reduce accuracy of identifications [*21*]. Additionally, approval, that is, a positive environment, tends to enhance identification accuracy [*27*] as do longer periods of observation and more immediate presentations [*28*]. However, neither the pose (front, side, portrait) nor the type of photograph (color, black and white) appear to have much effect on the identification task [*28*] and considerable variability in observer performance has been reported [*23*]. While other factors have been reviewed

also, those cited above would appear to provide a general understanding of the complexity of the identification process.

In summary, although eyewitnesses or visual identification testimony is well accepted by the courts in the United States, research has demonstrated that such identifications may not be particularly accurate—and that judgments may be biased by presentation patterns as well as other factors. Further, the confidence of a witness in his or her selection does not always correlate with its accuracy. Finally, no one has compared the ability of the same subjects to make accurate auditory and visual identifications from the same event.

Purpose

This research had several goals. One purpose was to examine, by means of a simulated crime, the accuracy levels of witnesses, who do not know the criminal, for both visual and auditory recognitions under generally favorable conditions of observation. A second purpose was to contrast the effectiveness of the two modalities, thereby providing a bridge between speaker and eyewitness identification. A third, and important, purpose was to study levels of correct identification as a function of time (that is, extinction of the identification response), both within a single group and among different groups. Fourth, we wished to add experimental data to those available relative to whether or not the identification process can be affected by presentations of incorrect stimuli. A final goal was to further study the relationship between confidence levels and the accuracy of judgments.

Methods

The Simulated Crime

Since some of the experimental protocols were basic to all of the questions asked and procedures to be carried out, they will be considered first. In brief, a volunteer "criminal" entered a large law school class and simulated a physical attack upon a volunteer "victim." The criminal spoke a number of aggressive-type phrases, knocked the victim out of his chair, hit and kicked him several times, and then exited the room. The incident lasted approximately 20 s. Immediately following this "crime," the class members were asked to describe, in writing, the criminal and the crime. It should be noted that the students who witnessed this scene had been instructed beforehand that "an event" would take place sometime during the class period, that they were simply to observe it, and under no circumstances leave their seats. These instructions were considered necessary, because, in a separate pilot study carried out in a different class, several students attempted to aid the victim, prematurely terminating the experiment. These instructions may have somewhat biased the experiment; they undoubtedly increased the arousal level of the class members and, perhaps, slightly increased the number of correct identifications. Other aspects of the simulated crime also were manipulated to reduce the impact of those factors previously determined to increase erroneous identification: the lighting was good, all class members had an unobstructed view of the event, the criminal faced the observers for a significant time, he spoke clearly and in a reasonably loud voice, and the classroom was closed off so there would be no distractions. Finally, after the event, the witnesses were asked to write down their descriptions of the criminal before discussing the event with anyone else. Obviously not all such error-producing elements would be missing from an actual crime. However, we judged it legitimate to remove them in order to insure development of a data base against which each of these factors could be compared at a later time.

Subjects

Twenty young adult male volunteers were selected from a Reserve Officer Training Corps (ROTC) class to create a "suspect pool." One of these individuals volunteered to play the role of the criminal; a second was the victim while the remainder of the group provided spoken and photographic exemplars. The observers or auditors were drawn from the law school class in which the crime was simulated. Four groups were selected from among these witnesses; each group contained an equal number of students who sat in the front and back of the room, and on each side.

Exemplars

The speech produced by the criminal was recorded during the simulated attack; a videotape of the event also was made as a backup. However, it was not necessary to use this tape in the experiment. Each suspect provided a spoken exemplar consisting of (1) a code number (not replayed to the observers), (2) several phrases similar to those that had been spoken during the attack by the criminal, and (3) as backup material, a short reading passage 20 to 30 s in length. Each subject was photographed several times and in a variety of clothes.

Experimental Procedure

The tape-recorded exemplars and photographs were presented to the four groups of observers on the basis of the schedule found in Table 1. The criminal is represented as *C*, other suspects as *S*, the same seven were used for Groups A through C. In Group D, a foil (*F*) was included in place of the criminal and different suspects (*S*) used in each trial. In order to follow typical protocols and permit the data to be comparable with those from like projects, the criminal was presented to Group D along with the foil at the last session but for the visual procedure only.

The paradigm involving Group A was included to examine the potentially decaying effects of elapsed time on the accuracy of identification as a function of repeated presentations. Members of Group A saw the photos and heard the taped exemplars, played separately one day after the simulated crime (Day-1), one week later (Week-1), and again two weeks later (Week-2). The question asked was whether or not the identification response would be extinguished, as a function of time, during a period of this admittedly short duration. Groups B, Week-1, and C, Week-2, are associated with Group A, Day-1. They were included to examine the effects of latency on accuracy of identification when there are delays of these magnitudes in task presentation. Finally, the Group D protocols were included in order to discover

TABLE 1—*Schedule of judgments for the four observer groups. All observers completed both the auditory and the visual tasks. Three sets of comparisons were made, that is, among the data found in column A. those found in column D, and diagonally from Group A (Day-1) to Group B (Week-1) to Group C (Week-2).*[a]

Latency	Group A (N = 14)	Group B (N = 18)	Group C (N = 16)	Group D (N = 13)
Day-1	$C + S_{1-7}$	...	...	$F + S_{1-7}$
Week-1	$C + S_{1-7}$	$C + S_{1-7}$	...	$F + S_{8-14}$
Week-2	$C + S_{1-7}$	...	$C + S_{1-7}$	$F + S_{15-21}$[b] $C + F + S_{15-20}$[c]

[a] C = criminal, F = foil, and S = innocent suspects.
[b] Auditory procedure only.
[c] Visual procedure only.

if identifications could be shifted (by means of multiple presentations) to someone who was not the criminal. For the auditory procedure, the foil only was included in the tapes at all three points in time. In the visual procedure, Group D served the same function but the protocols were varied slightly. That is, for the last presentation (that is, Week-2), the photograph of the actual criminal was included along with the foil and six other suspects in order to determine if continued exposure to a foil would result in his being selected as the criminal even when the real criminal was present. Photographs of these two individuals may be seen in Fig. 1.

Specific Procedure for the Judgments

For the aural/perceptual procedure, listeners were requested to identify the criminal after listening to eight recordings of the eight volunteers speaking the same phrase. Observers also were permitted a ninth response, that is, that the criminal was not present among the samples. Multiple choices were permitted but not encouraged. The tape recordings were played to the observers three separate times: (1) to familiarize them with the task, (2) to permit the selection or selections to be made, and (3) to confirm or correct the selection(s). In addition, all observers were requested to rate their confidence; they did so by indicating it on a seven-point scale where one denoted an extremely positive opinion and seven functionally equaled a response in which the observer felt little confidence: a guess. Protocols for the visual procedure paralled those for the aural/perceptual task, using the photographs as materials for judgment.

Results

Aural/Perceptual Data

Table 2 presents the percentage of correct responses and associated confidence levels for all groups. Responses for Groups A through C have been divided into three categories: R_c in which only the correct criminal was chosen; R_{cs} in which the listener chose the correct criminal but impeached his or her choice by choosing an additional innocent suspect or suspects; and the total both as a function of the number of subjects in the group (% correct) and the number of responses (mean %). The data for Group D are presented similarly, with the foil substituted for the criminal.

Consideration of the data for Group A reveals a relatively poor overall performance by the auditors. The best results were obtained for the third set of trials (that is, two weeks after the simulated crime) where a mean correct response of 28% was found or 50% if only the best choice of an observer is considered. For all three sets of trials, the confidence levels associated with correct response ordinarily were higher than the average confidence level of 3.5 for

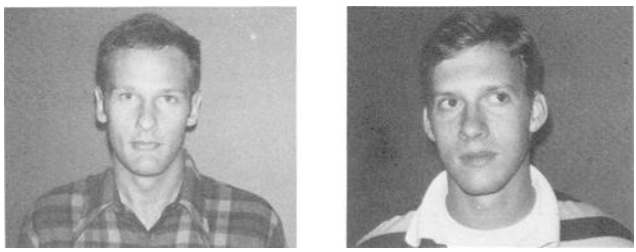

FIG. 1—*Photographs of the two volunteers who served as the criminal and the foil. As can be seen they resemble each other. They are not further identified so as to insure their anonymity.*

TABLE 2—*Aural/perceptual response data for all groups at three different time latencies— Day-1 (the day following the simulated crime), Week-1, and Week-2. % correct represents the number of responses in each category divided by number of listeners in each group; mean % represents the number of responses for the "criminal" divided by all the responses given by that group of listeners. There were 14 listeners in Group A, 18 in Group B, 16 in Group C, and 13 in Group D.*[a]

	Group A			Group B			Group C				Group D		
Latency	% Correct	Mean %	Confidence Level	% Correct	Mean %	Confidence Level	% Correct	Mean %	Confidence Level	Latency	% Correct	Mean %	Confidence Level
						DAY-1							
R_c	7	...	4.0	...	...	...	...	...	...	R_f	0	...	...
R_{cs}	29		4.3	...	...	...	...	...	...	R_{fs}	0	...	...
Total	36	20	...	...	...	...	...	...	...	total	0	0	...
						WEEK-1							
R_c	7	...	2.0	11	...	4.5	...	...	...	R_f	15	...	3.0
R_{cs}	21	...	4.3	28	...	4.2	...	...	...	R_{fs}	0	...	...
Total	28	18	...	39	24	...	...	...	...	total	15	13	...
						WEEK-2							
R_c	7	...	7.0	...	...	...	31	...	4.4	R_f	15	...	2.0
R_{cs}	43	...	4.0	...	...	...	19	...	3.3	R_{fs}	8	...	4.0
Total	50	28	...	...	...	...	50	44	...	total	23	19	...

[a] R_c = listener chose only the correct criminal, R_{cs} = listener chose the correct criminal plus one or more (innocent) suspects, and R_f and R_{fs} are the same but for the foils.

the incorrect responses, indicating that the auditors had less confidence in their correct choice than they did for their incorrect choices.

To test the significance of these results, a chi-square was calculated for the eight suspects and the three observation latencies. None of the chi-squares were significant even at the 10% level (much less the 5% level) indicating that the identifications were so spread out among the eight suspects that there was little chance that the criminal would be selected a significant proportion of the time. To further test this assumption, both asymmetric and symmetric lambdas as well as uncertainty coefficients were calculated. All of these metrics are measures of predictability—unfortunately they were poor (0.05-0.21). Finally, chi-square tests also were carried out for each time period individually; none were significant.

The next step was to compare latencies of various periods relative to the first time any of the observers heard the experimental tapes. Thus, data are compared among Group A: Day-1, Group B: Week-1, and Group C: Week-2. Once again, a relatively poor overall performance by the listeners is revealed. In this case, the subjects in Group B who listened to the exemplars one week after the simulated crime scored slightly better than did those in Group A: Day-1 or the Group C: Week-2 auditors. As with the previously cited data, the level of confidence in the identification was lower for correct responses than for the incorrect responses. Again, the chi-square values calculated were not significant, nor were the measures of predictability. Moreover, the trend, even though not significant, appears to be in the wrong direction with the correct percent values increasing from 36 to 50%—or from 20 to 44% if the mean percent is considered. It would appear that the data are not inconsistent with Brown's [*8*] suggestion that identifications associated with long-term memory may be easier to organize than those associated with short-term memory.

The last column in Table 2 provides the auditory response patterns for the subjects in Group D, the observers who heard the foil but not the criminal. Speakers, except for the foil, were changed for each set of trials, and only the sound-alike foil was common to all three. Auditors' responses have been divided into categories parallel to those used in the previous arrays, except that the f for foil replaces c for criminal. The foil was not chosen at all the first time he was presented (Day-1) but false positive choices increased for the subsequent sets. Finally, "innocent" talkers were selected as the criminal a majority of the time for all trials (not shown on table) and few listeners took the option of indicating that the criminal was not in the group. Confidence levels for those who chose the foil were at about the same level as for those who made other choices. Only individual chi-square tests were carried out on the Group D data, because the innocent subjects for each trial were different individuals each week. The relationships among the Day-1 and Week-1 data were not statistically significant but those for Week-2 were. The fact that a number of false positives occurred, including a significant number for Week-2, may be one of our more important findings—at least for the legal community. These data suggest that innocent suspects can be convicted on the basis of honest, but incorrect, testimony by earwitnesses.

Results of the Visual Procedure

The % correct, mean % responses, and confidence levels associated with the eyewitness judgments for all groups are provided by Table 3. Analysis of the data demonstrates that subjects in Group A could identify the criminal most of the time—and for all latencies—although with decreasing levels of correctness (85, 79, and 64%). As would be expected, the mean values paralleled those for % correct (80, 61, and 60%, respectively). The confidence levels varied widely and with no consistent pattern.

Since there was some evidence that subjects could make relatively accurate visual identifications, the photographic data were evaluated at the 0.05 level of confidence. While neither the overall chi-square test or the measures of predictability, asymmetric and symmetric lambda, and uncertainty coefficient were significant, all individual chi-square tests were, a

TABLE 3—*Visual responses for all groups for three different latencies—Day-1 (following the simulated crime), Week-1, and Week-2. % correct represents the number of responses in each category divided by number of observers in each group; mean % represents the number of responses per subject divided by the total number of responses for that set. There were 14 observers in Group A, 18 in Group B, 16 in Group C, and 13 in Group D.*

	Group A			Group B			Group C				Group D		
Latency	% Correct	Mean %	Confidence Level	% Correct	Mean %	Confidence Level	% Correct	Mean %	Confidence Level	Latency	% Correct	Mean %	Confidence Level
							DAY-1						
R_c	64	...	3.2	...	...	...	...	...	...	R_f	0	...	...
R_{cs}	21	...	4.0	...	...	...	...	...	...	R_{fs}	8	...	2.0
Total	85[a]	80[a]	...	...	...	...	...	...	...	total	8	7	...
							WEEK-1						
R_c	36	...	1.8	94	...	2.0	...	...	...	R_f	15	...	2.5
R_{cs}	43	...	4.7	0	...	...	...	...	...	R_{fs}	0	...	...
Total	79[a]	61[a]	...	94[a]	100[a]	...	...	...	...	total	15	13	...
							WEEK-2						
R_f	...	...	...	...	...	...	...	...	...	R_f	8	...	3.0
R_{fs}	...	...	...	...	...	...	...	...	...	R_{fs}	8	...	5.0
T_f	...	...	...	...	...	...	...	...	...	T_f	16	11	...
R_c	50	...	2.3	...	...	...	75	...	2.5	R_c	31	...	3.0
R_{cs}	14	...	4.5	...	...	...	0	...	...	R_{cs}	8	...	6.0
T_c	64[a]	60[a]	...	...	...	...	75[a]	75[a]	...	T_c	39	28	...

[a]Significant at the 0.05 level of confidence.

relationship suggesting that individuals can accurately and repeatedly identify faces, although at differing levels of efficiency.

In the case of the different latencies before first identification, the pattern is substantially dissimilar from aural/perceptual procedure. In this case, the subjects in Group B who attempted to identify the criminal's picture a full week after the crime did better than did either of the other two groups (see again Table 3). Indeed, they chose only the correct criminal 94% of the time; innocent suspects were not chosen at all and there was only one false negative. Thus, if only the positive responses are considered, this group can be said to be 100% accurate. Members of Group C, who carried out the identification task two weeks after the simulated crime, also chose only the criminal most of the time (75%). Again, all individual chi-square tests were significant beyond the 5% level of confidence, although the overall chi-square was not—a finding that suggests that witnesses can be expected to make correct identifications at reasonably high levels but that the patterns will be variable as a function of delay before presentation.

The last column in Table 3 provides the response patterns for the visual identification procedure when a look-alike foil was included among the subjects. The first day after the simulated crime most observers chose innocent subjects, only one chose the foil, and 38% of the subjects, a fairly large number, indicated that the criminal was not in the group at all. After one week, two observers chose the foil, a few chose innocent suspects, but the majority still indicated that the criminal was not in the group. For the final set of trials both the foil and criminal were included in the lineup. At this time, the largest proportion of observers chose innocent subjects as the criminal with an almost equally large number indicating that the criminal was not included in the group. However, when number-of-times the foil was selected was compared to selections of the actual criminal, it may be seen that the criminal was identified more than twice as often (30/28%) as was the foil (16/11%). It appears that the procedure used tends to confuse the perceptors and prevent them from choosing the criminal a significant number of times when he does appear. On the other hand, it apparently is also rather difficult to shift the identification from the actual criminal to another suspect even though this foil looks like the criminal and has had his picture presented a number of times.

Discussion

One of the purposes of this research was to contrast visual identifications with aural/perceptual judgments. In this particular simulated crime, where observers both saw the criminal and heard him speak, visual identifications unquestionably were more accurate than were the auditory ones. Moreover, the visual procedure resulted in fewer confusions; that is, fewer innocent suspects were implicated in the visual paradigm. There appear to be several factors that could account for this difference in performance levels. First, is it possible that human observers are better able to retain and recall visual stimuli than they can auditory stimuli. Second, in a situation such as this simulated crime, where observers both see and hear the criminal, it is possible that the visual modality may be dominant, that is, visual stimuli act as a distraction, inhibiting attention to the aural aspects of the situation and depressing an individual's ability to retain and recall the auditory stimuli and, in this particular situation, the subjects were somewhat aroused by the potential of an event. While there is no evidence that law students are necessarily more visual than auditory, the situation as developed may have tended to bias memory toward the visual modality. Finally, as has been indicated, the observers were requested to write descriptions of the criminal and the incident immediately after it took place. The immediate recall and writing of these seen events may have aided the visual memory.

The data clearly indicate that visual identification of a previously unknown criminal from a pool of photographs can be surprisingly accurate. The statistical tests for Groups A, B,

and C were uniformly significant, indicating that the responses obtained were probably not the result of chance alone. Accuracy levels of this magnitude seem inconsistent with those reported by Buckhout [*17,21*] and others but these differences may have resulted from our efforts to minimize those factors previously identified as adversly affecting accurate identification. Indeed, it should be noted that substantially reduced accuracy was found for Group D, the one instance where distracting factors were included. It should be noted also that levels of correct response were not consistent and did not appear to change predictably with time. Laws governing response patterns over a two-week period may exist, but they were not evident in this research. Of interest in a forensic science model would be the number of false positives, innocent suspects mistakenly identified as the criminal. Although not reported in the table, it should be pointed out that their number appeared to increase somewhat over time in the visual paradigm, at least for Group A, thus suggesting an increasing danger of choosing the incorrect person after repeated presentations. However, there was little, if any, evidence to suggest that visual recall decays markedly within the time frame of this experiment. The subjects appeared to be just as able to correctly identify the criminal after one or two weeks as they were the day after the crime.

On the other hand, it would appear that aural/perceptual identification of a criminal from a voice lineup is generally quite poor—50% at best in this investigation, a finding that appears to agree with the position taken by Brandt [*12*] and Michel [*29*]. We also observed limited auditory identifications in an actual case [*13*]. It also should be noted that our data demonstrate that the number of innocent suspects identified as the criminal can be relatively high (36 to 57%). However, it was evident that delays of up to two weeks did not cause much deterioration of the auditory identification response. Indeed, the highest number of correct scores were obtained two weeks after the incident; both Groups A and C achieved an accuracy of 50% at the Week-2 set of trials. The reasons for this seeming improvement are not clear; it may be that, for Group A, continued exposure to the criminal's voice reinforced the identification process. However, since none of the chi-square tests were significant, this trend actually may be a chance one. When interpreting these results, it must be stressed that the criminal was both seen and heard, and that the visual perception may have been so powerful that it depressed auditory memory. Perhaps in a situation in which the criminal was heard and not seen, such as over the telephone, aural/perceptual identification would be more accurate and predictable, although data from a number of experiments [*5, 7, 10*] would suggest otherwise.

Further consideration should be given to false positives, or the identification of innocent suspects. For example, when a foil is included in a lineup (the D paradigm), the frequency pattern of the selection process appears to be similar for both the aural/perceptual and the visual series. There appears to be a weak tendency for the foil to be chosen with increasing frequency as a function of time—0, 15, and 23% in the aural/perceptual paradigm, and 0, 15, and 16% in the visual. It is possible that continued exposure to the foil reinforces his selection. It should be noted further that, in the aural/perceptual paradigm, innocent suspects (only) usually were identified as the criminal even when the foil was not counted (100, 77, and 69%), with few listeners choosing the option that the criminal was not present in the suspect pool (0, 8, and 8%). Indeed, the statistical evaluations we carried out indicate that most listeners chose innocent suspects with greater-than-chance frequency. The reason for this tendency is not clear. On the other hand, it was more difficult to confuse the Group D eyewitnesses. A substantial number indicated that the suspect was not in the lineup, that is, 38 to 54% depending on the latency and fewer innocent suspects were chosen than in the auditory procedure (visual = 31 to 50% versus auditory = 69 to 100%). This number of false positives is still unacceptably high. Finally, when the real criminal was included in the last photographic set, he was identified by 39% of the observers. The memory of his facial characteristics apparently was robust enough to counter, in some cases anyway, the "training" that had occurred by continued inclusion of a look-alike foil.

Another conclusion that can be drawn from this investigation is that confidence levels appear not to relate very strongly to the identification process. Observers who chose only innocent suspects often were more confident than were those who chose only the criminal, a relationship most striking for the aural/perceptual paradigm. The confidence levels among the visual procedures were not so orderly; however, since they varied so widely, their net effect was negative. Thus, while it should be noted that none of the cited trends were statistically significant, it must be stressed that, based on this research, confidence estimates appear to add little if anything to the identification process.

Although this experiment was designed to parallel a forensic science situation, mitigating factors cannot be overlooked. First, as has been stated, the observers had been alerted to the possibility that something would happen the day the incident took place. This warning may have increased observer arousal and, hence may explain the higher percentage of correct identifications, especially visual, in this investigation when it is compared to other studies [*18,19*]. Second, those factors that have been identified as adversly affecting identification, brevity of observation, inadequate lighting, inattention of the witnesses, and the like, were purposefully eliminated, insofar as possible, from the observed event. Finally, in a forensic science situation where a witness had seen the criminal, the addition of voice identification probably would not be necessary. Had the observers in this investigation only heard the criminal, the aural/perceptual procedures/results would have more closely paralleled the forensic science model. Nevertheless, the design used may be defended on two bases: (1) there were not enough subjects available to structure more than four groups of auditors and (2) one of the major purposes of this project was to directly compare visual and auditory judgments of related stimuli.

In view of this investigation, and other extant data, it appears that both visual and especially aural/perceptual identifications should be approached more critically and with greater caution than is currently the case. Although visual identification of a previously unknown individual can be accurate, the outcome tends to be somewhat unpredictable. Moreover, the aural/perceptual identification of a previously unknown suspect does not appear to be robust enough, at all, for courtroom use. Even the use of witnesses who know the talker is somewhat suspect. Precedent based acceptance of eyewitness and earwitness testimony, relying on common sense rather than on critical evaluation of emerging data, gives an emphasis to such testimony, perhaps far beyond that which it deserves. Certainly, the courts should accept aural/perceptual identifications only after careful scrutiny—if at all.

Acknowledgment

The authors would like to thank Dr. J. W. Hicks, Jr. for his assistance with the project.

References

[*1*] McGehee, F., "The Reliability of the Identification of the Human Voice," *Journal of General Psychology*, Vol. 17, 1937, pp. 249-271.

[*2*] *Mack v. State*, 54 Fla. 55, 44 So. 706 (1907), citing Howell's State Trials 1186.

[*3*] 70 A.L.R. 2d 995.

[*4*] Bricker, P. and Pruzansky, S., "Effects of Stimulus Content and Duration on Talker Identification," *Journal of the Acoustical Society of America*, Vol. 40, No. 6, Dec. 1966, pp. 1441-1450.

[*5*] Iles, M., "Speaker Identification as a Function of Fundamental Frequency and Resonant Frequencies," unpublished Ph.D. dissertation, University of Florida, Gainesville, 1972.

[*6*] Pollack, I., Pickett, J., and Sumby, W., "On the Identification of Speakers by Voice," *Journal of the Acoustical Society of America*, Vol. 26, No. 6, May 1954, pp. 403-412.

[*7*] Hollien, H., Majewski, W., and Doherty, E. T., "Perceptual Identification of Voices Under Normal, Stress and Disguised Speaking Conditions," *Journal of Phonetics*, Vol. 10, 1982, pp. 139-148.

[8] Brown, R., "Memory and Decision in Speaker Recognition," *International Journal for Man-Machine Studies*, Vol. 11, 1979, pp. 729-942.
[9] Stevens, K., Williams, C., Carbonnell, J., and Woods, B., "Speaker Authentification and Identification: A Comparison of Spectrographic and Auditory Presentation of Speech Materials," *Journal of the Acoustical Society of America*, Vol. 44, No. 6, Dec. 1968, pp. 1596-1607.
[10] Rothman, H., "A Perceptual (Aural) and Spectrographic Identification of Talkers with Similar Sounding Voices," in *Proceedings of the International Conference on Crime Countermeasures, Science and Engineering*, Vol. BV-113, University of Kentucky: Continuing Education, Lexington, KY, 1977, pp. 37-41.
[11] Rosenberg, A., "Listener Performance in Speaker Verification Tasks," *Institute of Electrical and Electronics Engineers, Transactions on Audio Electroacoustics*, Vol. AU-21, 1973, pp. 221-225.
[12] Brandt, J. F., "Can You *Hear* Me?," *Forensic Communication*, Vol. 2, 1977, pp. 9-11.
[13] *Ohio v. Trussell*, Montgomery County Court, Dayton, OH, Honorable John M. Meagher, presiding judge, Feb. 1978.
[14] D'Angelo, F. G., Jr., "Eyewitness Identification," *Social Action and the Law*, Vol. 5, No. 1, March 1979, pp. 18-19.
[15] Buckhout, R. and Freire, V., "Suggestivity in Lineups and Photospreads: A Casebook for Lawyers," Report CR-5, Center for Responsive Psychology, Brooklyn College C.U.N.Y., Brooklyn, June 1975, pp. 1-26.
[16] Loftus, E., "Eyewitness Testimony Does the Malleable Human Memory Interfere with Legal Justice?," *Social Action and the Law*, Vol. 2, No. 2, April 1975, pp. 5-6.
[17] Buckhout, R., "Expert Testimony by Psychologists," *Social Action and the Law*, Vol. 3, No. 4, June 1976, pp. 41-52.
[18] Buckhout, R., Alper, A., Chern, S., Silverberg, G., and Slomovits, M., "Determinants of Eyewitness Performance on a Lineup," *Bulletin of the Psychonomic Society*, Vol. 4, No. 3, Sept. 1974, pp. 191-192.
[19] Buckhout, R. and Figueroa, D., "Eyewitness Identification: Effects of Suggestion and Bias in Identification from Photographs," Report CR-11, Center for Responsive Psychology, Brooklyn College C.U.N.Y., Brooklyn, May 1974, pp. 1-24.
[20] Williams, L., "Application of Signal Detection Parameters in a Test of Eyewitness to a Crime," Report CR-20, Center for Responsive Psychology, Brooklyn College C.U.N.Y., Brooklyn, April 1975, pp. 1-31.
[21] Buckhout, R., "Eyewitness Testimony," *Scientific American*, Vol. 231, No. 6, Dec. 1974, pp. 23-31.
[22] Ellis, H. D., "Recognizing Faces," *British Journal of Psychology*, Vol. 66, No. 4, Nov. 1975, pp. 409-426.
[23] Cross, J. F., Cross, J., and Daly, J., "Sex, Race, Age and Beauty as Factors in Recognition of Faces," *Perception and Psychophysics*, Vol. 10, No. 6, Dec. 1971, pp. 393-396.
[24] Galper, R. E., "Functional Race Membership and Recognition of Faces," *Perceptual and Motor Skills*, Vol. 37, No. 2, Oct. 1973, pp. 455-462.
[25] Malpass, R. S. and Kravitz, J., "Recognition for Faces of Own and Other Race," *Journal of Personality and Social Psychology*, Vol. 13, No. 4, Dec. 1969, pp. 330-334.
[26] Shepherd, J. W. and Ellis, H. D., "The Effect of Attractiveness on the Recognition Memory for Faces," *American Journal of Psychology*, Vol. 86, No. 3, Sept. 1973, pp. 627-633.
[27] Schill, T. R., "Effects of Approval Motivation and Varying Conditions of Verbal Reinforcement on Incidental Memory for Faces," *Psychological Reports*, Vol. 19, No. 1, Aug. 1966, pp. 55-60.
[28] Laughery, K. R., Alexander, J. F., and Lara, A. B., "Recognition of Human Faces," *Journal of Applied Psychology*, Vol. 55, No. 5, Oct. 1971, pp. 477-483.
[29] Michel, J. F., "Use of a Voice Lineup," in *Convention Abstracts*, American Academy of Forensic Sciences, Colorado Springs, CO, 1980, p. 67.

Address requests for reprints or additional information to
Dr. Harry Hollien
Institute for Advanced Study of the Communication Processes
ASB-50
University of Florida
Gainesville, FL 32611