DOCUMENT RESUME

ED 369 128   CS 508 566

AUTHOR   Walter, Todd J.
TITLE   Voice Identification: Levels-of-Processing and the
                Relationship between Prior Description Accuracy and
                Recognition Accuracy.
PUB DATE   Aug 92
NOTE   51p.; Paper presented at the Annual Meeting of the
                American Psychological Association (100th,
                Washington, DC, August 14-18, 1992).
PUB TYPE   Speeches/Conference Papers (150) -- Reports -
                Research/Technical (143)

EDRS PRICE   MF01/PC03 Plus Postage.
DESCRIPTORS   Communication Research; Higher Education; Research
                Needs; *Speech Communication; Undergraduate
                Students
IDENTIFIERS   Levels of Processing; Research Suggestions; Speech
                Identification; Speech Research; Supreme Court;
                *Voice Recognition

ABSTRACT
                A study examined whether a person's ability to
accurately identify a voice is influenced by factors similar to those
proposed by the Supreme Court for eyewitness identification accuracy.
In particular, the Supreme Court has suggested that a person's prior
description accuracy of a suspect, degree of attention to a suspect,
and confidence in identifying a suspect, are reliable predictors for
accurately identifying a suspect. Subjects, 18 males and 42 females
from an undergraduate psychology course or volunteers from the local
community, listed to a voice and later described the voice on a
speech characteristic checklist. Later they were asked to identify
the voice from a lineup and denote how certain they were of their
choice. Results indicated no relationship between voice description
accuracy and identification accuracy, or between degree of confidence
and identification accuracy. Moreover, depth of processing had no
effect on description accuracy, identification accuracy, or the
relationship between the two. Future "earwitness" research should:
employ a voice lineup in which the target voice is either present or
absent; use longer retention delay between target presentation and
voice identification; and develop a valid descriptive measure.
(Contains 21 references, 2 tables, and 1 figure of data. The voice
description checklist is attached.) (RS)

ED 369 128

Voice Identification: Levels-of-Processing and the

Relationship Between Prior Description Accuracy

and Recognition Accuracy

Todd J. Walter

Niagara University

Running Head: VOICE IDENTIFICATION

Abstract

The purpose of the present study was to examine whether a person's ability to accurately identify a voice is influenced by factors similar to those proposed by the Supreme Court for eyewitness identification accuracy. In particular, the Supreme Court has suggested that a person's prior description accuracy of a suspect, degree of attention to a suspect, and confidence in identifying a suspect, are reliable predictors for accurately identifying a suspect. The present study empirically explored these concepts, relative to voices, by varying the levels of processing subjects devoted to voices. Sixty subjects were asked to listen to a voice and later describe the voice on a speech characteristic checklist. Later they were asked to identify the voice from a lineup and denote how certain they were of their choice. Results indicated no relationship between voice description accuracy and identification accuracy, or between degree of confidence and identification accuracy. Moreover, depth of processing had no effect on description accuracy, identification accuracy, or the relationship between the two. Practical and theoretical implications of the present findings are discussed.

Voice Identification: Levels-of-Processing and the

Relationship Between Prior Description Accuracy

and Recognition Accuracy

Within society each day numerous crimes are

committed in which a victim or bystander may be witness

to a perpetrator engaging in a criminal act. In such

instances, witnesses will most often rely on their

visual senses to gather and store information about the

criminal offender. However, there may also exist

instances, for reasons such as visual impairment, in

which the witness will be unable to see the perpetrator

and instead have to rely on his or her auditory senses

for acquiring information. Bomb threats, muggings,

rapes and assaults committed in darkness, and obscene

phone calls, are just some examples of crimes in which

the victim's only bit of evidence may be the sound of

the perpetrator's voice. Consequently, the accuracy of

correctly identifying a suspect's voice may be the most

determinant factor in criminal court procedures that

deal with such crimes. In this regard, the judicial

profession tends to take such evidence as admissible

and at face value, without substantiating whether a

witness' auditory identification of a suspect is

reliable (Deffenbacher et al., 1989).

Despite the significant role that voice recognition evidence may play in various crimes, not much research has been devoted to it. Rather, the majority of studies that have been conducted so far have overwhelmingly dealt with eyewitness identification accuracy (Bull, 1978; Clifford, 1980; Saslove & Yarmey, 1980). A literature search in two of the major professional outlets for this research, the Journal of Applied Psychology and Law and Human Behavior, over the last decade support the preceding findings. Research publications within these journals have found that 98 facial recognition studies have been published since 1980, in comparison to only five for voice recognition. Although some studies have shown facial identification accuracy to be high, other research has also shown it can be extremely fallible by being influenced by such factors as the witness' opportunity to see a person, whether the witness expected to see a person, and the delay between seeing and identifying a person (Clifford, 1980; Clifford, Rathborn, & Bull, 1981; Saslove & Yarmey, 1980; Yarmey, 1986).

The United States Supreme Court has even cited five such factors as criteria for evaluating the reliability

of a visual identification: (1) the opportunity to witness the criminal, at the time of the crime; (2) the witness' degree of attention; (3) the accuracy of the witness' prior description of the criminal; (4) the level of certainty demonstrated by the witness at the confrontation; (5) the length of time between the crime and the confrontation (Neil v. Biggers, 1972, p.199). Although these criteria were initially stated for visual identifications, they have since been accepted to apply to earwitness identifications as well (Deffenbacher et al., 1989). Nevertheless, some of the criteria have not been supported by research. For example, Deffenbacher et al. reviewed the research on the criterion of the degree of certainty a witness demonstrates during identification procedures. They found that most studies indicate little or no correlation between the witness' confidence in being able to identify the suspect and subsequent identification accuracy within both eyewitness and earwitness research. In other words, there is no evidence that witnesses who are confident that they can identify a face or voice are more accurate identifiers than those expressing lesser confidence. Also, in examining the question of whether there is a

relationship between description accuracy and
recognition accuracy of a suspect, Deffenbacher et al.
found that, in general, little or no relationship
exists between the two.

Despite no known earwitness studies regarding the
effect of prior description accuracy on identification
accuracy exist, in a review of past research on facial
identification, Deffenbacher et al. (1989) states that
there is no evidence that one's ability to describe a
suspect is strongly correlated with one's ability to
identify the suspect later. For example, Pigott and
Brigham (1985) asked subjects to describe the physical
characteristics of a live target person and later
identify the target person from a photographic lineup.
They found no relationship between the accuracy of the
subject's description of the target person and whether
they later identified the target correctly. Jenkins and
Davies (1985) reported two experiments in which
subjects viewed a filmed incident that included an
individual disrupting a class experiment. Following the
video subjects were exposed to either an accurate or
misleading composite of the individual they had
witnessed. Subjects then participated in an adjective
checklist description task and later attempted to

identify the target person. They found no relationship
between description and identification accuracy for
either those in the misleading or those in the accurate
composite conditions.

One plausible explanation that has been suggested
for such results, at least in studies using live
persons, is that a live stimulus may be so involving to
a subject that it overrides differences in orienting
instructions (Pigott & Brigham, 1985). In other words,
although subjects may be given prior directions to
focus on a target person's physical characteristics,
exposure to a live situation may cause the subject to
become more attentive to irrelevant factors (e.g., what
the person is doing) rather than to the more important
physical traits of the person. Consequently, subjects'
inability to focus properly on the qualities of a
target person may thus explain their failure to be
perceptually accurate in subsequent description and
identification tasks. Although such explanations and
the lack of empirical support for a relationship
between description and identification accuracy have
been confined to eyewitness research thus far, it has
been presumed that there would not be a significant
correlation between these two factors in earwitness

8

studies either (Deffenbacher et al., 1989). In light of this, the present study is one of the first to empirically assess whether a relationship between prior voice description accuracy and subsequent voice identification accuracy exist. Moreover, this study was conducted within a controlled setting and with pre-recorded voices in order to maximize earwitness accuracy in both description and identification tasks.

One final criterion from Neil v. Biggers (1972) that has important implications in both eyewitness and earwitness research is the witness' degree of attention towards the suspect at confrontation. Past studies of memory for pictures have suggested that, at least for recognition purposes, a person's attention to a stimulus is related to its later identification accuracy (Pigott & Brigham, 1985). For instance, one theory that has been used to explain how attention and recognition are related is the levels-of-processing perspective devised by Craik and Lockhart (1972). According to this theory, the type of attention one gives to a particular stimulus is related to the degree or depth of memory processing of that stimulus. Consequently, this theory stresses that the deeper the processing of a stimulus the better the memory of it

later (Klatzky, 1980). For instance, relatively deep processing in pictures of faces is caused by focusing subjects' attention on the personality traits (e.g., friendliness, honesty) of the pictured individual. Shallow processing is engendered by focusing attention on specific physical characteristics such as the size of an individual's nose (Pigott & Brigham, 1985). In one study that examined this phenomena, Bower and Karlin (1974) found that subjects asked to judge the likableness of a pictured face performed much better on a later recognition test than those asked to judge the sex of the person pictured. Also, in a series of experiments, Winograd (1981) found that persons who focus on specific features of a face do more poorly on later recognition tests than those who attempt to attribute more generalized personality traits (e.g., assigning personality features such as friendliness) to a pictured individual.

Although research studies such as these seemingly confirm that memory and subsequent performance are related to the processing level of a stimulus, there exists some disagreement with Craik and Lockhart's explanation of how such processing comes about. In other words, Craik and Lockhart's levels-of-processing

theory attributes differen: processing levels to the "semantic meaning," or number of unique associations a person attaches to a stimulus. For example, persons exposed to a pictured individual may be reminded of someone they are familiar with. In doing so the person makes an association between the pictured person and the individual they know. Thus in face recognition tasks, this theory would attribute better performance to an increase in depth-of-processing by way of an increase in the number of unique associations to the target face. However, recent research has produced more convincing evidence that manipulating the depth of processing faces into memory may produce better recognition performance as a result of what Winograd (1981) called the "elaboration hypothesis" (or what is also referred to as the "feature quantity hypothesis"). According to this theory, when one examines the physical qualities of a stimulus, only a few features of the stimulus need to be viewed, whereas in judging a certain personality trait of a stimulus (target face) one needs to view a greater number of features of the stimulus. Consequently, this explanation argues that as the number of encoded features increases, the level-of-processing increases, and as a result, increases

recognition accuracy (Bloom & Mudd, 1991). The present study attempted to apply Winograd's elaboration hypothesis in order to examine its effect on the recognition accuracy of voices. This was accomplished by asking subjects to either judge the sincerity and friendliness of a voice (deep processing) or the rate and gender of the speaker (shallow processing).

One final consideration on this issue of attention and depth-of-processing relates to instances when no instruction is given about characteristics to pay attention to. Evidence in recent facial recognition research has shown that when subjects are told only to try to remember a face, this produces an intermediate level of processing that is superior in performance to those who focus on specific physical features (shallow processing), but lower than those who attribute generalized traits to faces (deep processing) on recognition tasks (Sporer, 1991). These findings were applied to voice recognition accuracy in the present study by having a third group of subjects just listen to and comprehend what a speaker said (intermediate processing). The voice recognition accuracy of these subjects was then compared to those in deep and shallow processing conditions in order to examine whether

recognition accuracy across conditions were similar to those found for faces above.

In considering the cumulative findings and explanations for levels-of-processing discussed above, I suggest that just as there exists a difference in memory processing for pictures (non-live stimuli), so too may there exist a difference in processing voices. In other words, as with pictures, a deep level-of-processing may be produced by focusing attention on the general personality traits of a voice, such as sincerity and friendliness. Shallow processing will be produced by focussing attention on only specific traits of the voice (i.e., rate, gender, etc.). Moreover, I suggest that those who are simply given tasks to listen to and remember a voice will produce an intermediate level of processing that is similar in performance accuracy to that found in facial recognition research. Although a few studies by Clifford and McCardle (cited in Bull & Clifford, 1984) and Hammersley and Read (1985) contradict these assumptions by finding no significant differences between levels of processing single voices and recognition accuracy (Deffenbacher et al., 1989), no other studies have been found to confirm this. In light of the fact that little research has

examined the effect of depth of processing on voice
recognition accuracy, the present study sought to
examine whether a person's type of attention to a voice
affects their ability to both describe and later
recognize it. In other words, just as performance on
facial recognition tasks has been found to be affected
by the number of features encoded, I believe that one's
performance on descriptive recall tasks will similarly
be contingent on the level-of-processing. Consequently,
I suggest that a positive correlation exists between
voice description and voice identification accuracy
across shallow, intermediate, and deep processing
conditions.

Therefore, the hypothesis of this experiment is
that voice description and recognition accuracy, as
well as a relationship between the two, will be a
function of the amount of attention subjects pay to the
features of a target voice. I predict, in accordance
with facial recognition research, that subjects in the
shallow processing condition will be less accurate
in their descriptions and subsequent identifications of
a target voice than subjects in the intermediate and
deep processing conditions. Further, I predict that the
correlation between description accuracy and

identification accuracy will increase with the depth of processing. The relationship between confidence and voice description accuracy and voice identification accuracy will also be examined.

Method

Subjects

A total 18 males and 42 females, with an average age of 21.8 years, participated as subjects. Subjects were composed of students who received course credit from undergraduate psychology courses as well as volunteers from the local community.

Materials

The voices used in this experiment were from male volunteers from a local construction products manufacturing company. Volunteers were all Caucasians and ranged in age from 27 to 35 ($M = 31.4$) years old. Of 10 original voices chosen, eight were selected to minimize unusual vocal characteristics between them (as confirmed by 6 independent raters).

The volunteers were recorded individually in an enclosed office. Voices were recorded (and later played to subjects) via a stereo cassette recorder. Controls for volume, tone, and bass were constant for all procedures. Each volunteer's voice was recorded twice:

once as a "target" in which each volunteer was recorded
on his own separate audio cassette, as well as together
with the other seven voices on a single "voice lineup"
tape. Voices on the individual target cassettes calmly
uttered the same statement: "Everyone get down on the
floor and be quiet, this is a hold up. If everyone
cooperates no one will get hurt." Similarly, statements
by each voice on the voice lineup cassette recorded a
paraphrased version of the target statement: "This is a
hold up. Everyone get down and shut up, no one will get
hurt." The presentation order of each voice on the
voice lineup cassette was randomly determined.

To examine the relationship between a witness'
description accuracy and recognition accuracy, a voice
description checklist was developed. This recall task
employed a five-point scale of eight different speech
characteristics: intensity, rhythm, pitch, accent,
rate, inflection, clarity, and nasality. Beside each
speech quality was a brief description of the
characteristic in order to clarify and prevent any
misunderstandings. Scales ranged from 1 (low/none) to 5
(high/noticeable).

In order to obtain the voice description accuracy
of subjects, 22 independent description raters were

16

employed to describe each of the eight target voices.
Raters were composed of 10 male and 12 female
volunteers of mean age 24.3 years. Raters were given a
voice description checklist for each of the eight
target voices and were instructed to describe each
voice as best as possible in accordance with the eight
speech characteristic scales. Each target voice was
presented eight times (once for each speech
characteristic) in order to gain more reliable
descriptions. Subjects were instructed to rate one
speech characteristic for each time a target voice was
presented. Completion order of the scaled speech
characteristics and target voice presentation was
randomized for each rating trial. Description accuracy
scores were based on the modal rating (on a five-point
scale) given by the independent description raters for
each speech characteristic of each target voice.
Subjects' description scores were thus calculated by
subtracting the subject's speech characteristic ratings
from the modal ratings obtained from the independent
raters. For instance, a subject who indicated a rating
of 2 (on a five-point scale) in describing the rate of
speech for a voice would receive a score of 2 had the
modal rating given by the independent raters been 4.

Scores for each individual speech characteristic were
then summed in order to obtain a subject's overall
voice description score. Accurate description was thus
indicated by low description checklist scores since
they were close to the independent modal ratings.

Also employed in this study was the use of a
stopwatch in order to gauge the amount of time subjects
spent on various tasks within a trial. The purpose of
this was to ensure that each subject experienced the
same constant duration between voice description and
voice identification procedures.

Procedure

Subjects were told that they were participating
in a multi-task study designed to examine how well
persons can complete certain assignments within a
specified time frame. The assignments they were told
they would be exposed to included a word search puzzle
and voice description tasks. Subjects were seated at a
table, 3 feet from an audio cassette recorder.

Upon introducing the first task subjects were told
that they would hear a recording of a person speaking a
sentence. They were then assigned by block
randomization to either shallow processing ($\underline{n}$=19),
intermediate processing ($\underline{n}$=21), or deep processing

conditions (n=20). Subjects were given directions in accordance with the condition to which they were assigned. Those in the deep processing (high attention) condition were told to "judge how sincere and friendly the person who spoke was." Those in the intermediate memory processing (no instructions) condition were simply told to "remember and comprehend what the person said," without emphasis of what qualities to focus on. Subjects in the shallow memory processing condition were instructed to judge the rate, or how fast the speaker spoke, and to determine the person's gender. Following these directions subjects heard the target voice. Selection of the target voice for each subject was randomly determined by a random numbers table (each of the eight voices was assigned a number prior to the study). The remaining seven voices that were not selected for target presentation served as lineup foils in the subsequent recognition task.

After exposure to the target voice, subjects completed a voice description survey (which was used for time delay purposes) in which they were instructed to freely recall as much information as possible about the voice they had just heard and write it on the paper provided. All participants were told they had 3 minutes

to complete this task. If subjects finished early they were asked to turn their survey over and wait patiently until the entire 3 minutes had expired.

After the free descriptive recall task, subjects completed a voice description checklist (See Appendix). Subjects were instructed that they had 2 minutes to rate the target voice they had heard on eight different voice characteristics along a five-point scale. Subjects were told to turn their checklist over when completed and wait for further instructions (after the two minutes elapsed). This procedure was used to measure voice description accuracy.

After completion of the voice description checklists subjects were then given a word-find puzzle that included a list of 41 words. Subjects were instructed to locate and circle as many of the words listed as possible within a 10 minute time period. Due to the number of words listed, none of the subjects were able to complete this task within the 10 minutes provided. Like the voice description survey this measure served as a retention interval between voice description and voice recognition procedures.

At the conclusion of the word-find puzzle subjects were given an identification data sheet. They were told

that they would hear eight different voices which may
or may not include the target voice heard previously.
Subjects were instructed to try to identify the voice
they believed was the one they were exposed to earlier
by checking the corresponding voice number (or voice
not in lineup space) on the data sheet provided.
Although an option for voice not in lineup was
provided, target voices were always present in the
lineup. Participants were further directed not to
choose an answer until they had listened to all of the
lineup voices. Although the order of the voices in the
lineup remained constant for each trial, voice
positioning was randomized by the fact that it was
based on the number of the target voice which was
randomly chosen. Consequently, the target voice did not
necessarily appear in the same position of the lineup
for each subject.

Following exposure to the voice lineup and
selection of recognition choice, subjects rated their
confidence in their selection on a scale ranging from 1
(not sure) to 5 (very sure). After this, subjects were
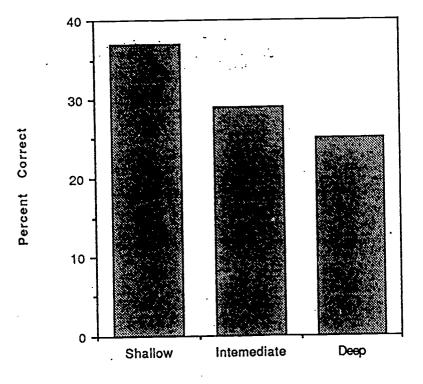completely debriefed about the study.

Results

Although only 18 out of 60 subjects were able to
correctly identify the target voice in the lineup, hit
rates were still above chance (1 out of 9), z = 4.75,
p<.001. The distribution of hits between conditions was
5 for deep processing (N = 20), 6 for intermediate
processing (N = 21), and 7 for shallow processing
(N = 19). An analysis between conditions for
identification accuracy yielded no significant
differences [x^2(2, N = 60) = .68, p>.05]. In fact, as
is illustrated in Figure 1, results somewhat
contradicted the depth of processing findings for
facial recognition from which the predictions for this
study were based.

In examining whether a relationship exists between
a subject's confidence in recognition choice and actual
identification accuracy, a statistical analysis showed
that no significant relationship was found r(58) = .19,
p>.05. Moreover, Table 1 shows that most subjects
indicated that they were fairly confident of their
recognition choice despite actual recognition accuracy
(M = 3.55, SD = .96).

In analyzing subjects description accuracy across
all three conditions, results showed that subjects

22

Figure 1. Identification accuracy in levels-of-
processing conditions.

Table 1

Frequencies of Confidence Level Ratings for

Identification Accuracy

Accuracy	Uncertain			Certain	
Level	1	2	3	4	5

Correct Frequency	1	0	5	7	5
Incorrect Frequency	0	7	15	15	5

description ratings for all eight speech characteristics were on average less than 1 point off those obtained by independent raters for each characteristic ($\underline{M}$ = 6.6, $\underline{SD}$ = 2.53). There were no statistically significant differences in description score means between level-of-processing conditions, $\underline{F}$(2, 57) = .08, $\underline{p}$>.05. Thus, like the recognition results, subjects' depth of processing does not appear to affect voice description accuracy. Also, the correlation between description accuracy and subjects confidence in recognition was not significant $\underline{r}$(58) = -.02, $\underline{p}$>.05.

Lastly, statistical analyses were conducted in order to ascertain whether subjects accuracy in describing a target voice was related to their accuracy in recognizing the voice from a lineup later. Although subjects mean description scores in deep ($\underline{M}$ = 6.75, $\underline{SD}$ = 2.63), intermediate ($\underline{M}$ = 6.62, $\underline{SD}$ = 2.80), and shallow ($\underline{M}$ = 6.42, $\underline{SD}$ = 2.19) processing conditions tended to correspond with identification accuracy results, no statistically reliable relationship was found, $\underline{r}$(58) = .21, $\underline{p}$>.05. Moreover, correlations between description and identification accuracy within each level-of-processing condition were calculated.

Description score means within conditions and as a
function of identification accuracy are presented in
Table 2. Although a statistically significant
relationship was nearly obtained for subjects in the
deep processing condition [$r(18) = .41$, $p > .05$], no
significant correlation of description and
identification accuracy was found for subjects in
intermediate [$r(19) = .20$, $p > .05$] or shallow
[$r(17) = .003$, $p > .05$] processing conditions either.

Table 2

Description Score Means As a Function of Levels of

Processing and Recognition Accuracy

Level of Processing	Description Means for Correct Recognition	Description Means for Incorrect Recognition
Deep	8.60, N=5	6.13, N=15
Intermediate	7.50, N=6	6.27, N=15
Shallow	6.43, N=7	6.42, N=12

Note. Lower description means reflect better accuracy.

Discussion

The results of the present study have both practical and theoretical implications for crimes in which earwitness accuracy is important.

Practical Implications

In regard to practical implications, this study failed to provide any support for several of the criteria outlined by the Supreme Court in Neil v. Biggers (1972). For instance, the present experiment was one of the first to empirically test the Supreme Court's guideline that prior description accuracy of voices is related to subsequent recognition accuracy. The present findings do not substantiate such a claim. Similar to research that has failed to find a relationship between description accuracy and identification accuracy for faces (Pigott & Brigham, 1985; Jenkins & Davies, 1985), the results of the current study found no correlation between subjects accuracy in describing a voice and their accuracy in later identifying it from a lineup.

The current study also had as one of its purposes to minimize factors that may have plagued past facial description and identification accuracy studies. For instance, studies that have failed to find a

relationship between facial description and recognition accuracy by employing a live stimulus (e.g., Pigott & Brigham, 1985) have attributed such results to subjects being distracted from focussing on a person's physical characteristics because of other situational cues (e.g., attending to the behavior of the target). The present study sought to reduce such elements by presenting voices under highly controlled experimental conditions. However, the present failure to obtain a significant relationship between voice description and identification accuracy even under highly controlled conditions, coupled with the negative findings in facial recognition research when using a live stimulus, suggest that the possibility of obtaining a significant description-recognition relationship under a forensically relevant situation is doubtful. Consequently, in presuming the Supreme Court's assumptions regarding eyewitness identification accuracy are applicable to earwitness evidence, the present results imply that the Supreme Court may have been incorrect in its assumption that persons who are accurate in describing another person's voice will also be accurate in recognizing that voice later.

29

Another Supreme Court criterion examined in the
present study was whether a witness' confidence in
identifying a voice was related to actual
identification accuracy. Considering that jurors have
tended to strongly rely on this criterion in
determining the reliability of an eyewitness' testimony
(Deffenbacher et al., 1989; Yarmey, 1986), this
guideline can be of considerable importance to
earwitness criminal proceedings. The present
investigation, however, found no correlation between a
witness' degree of certainty and their identification
accuracy of a voice. Moreover, no relationship was
observed between voice description accuracy and
subsequent certainty in identification procedures.
Taken collectively, these findings suggest that a
person's professed confidence in being able to
recognize a voice does not predict how accurate their
description or identification of a voice was. These
conclusions agree with previous eyewitness and
earwitness studies which have tended to find small,
insignificant correlations between confidence and
recognition accuracy (Bull & Clifford, 1984;
Deffenbacher et al., 1989; Saslove & Yarmey, 1986).
Thus, like the criterion for a relationship between

voice description and identification accuracy discussed previously, the present investigation further casts doubt on the Supreme Court's assumption that a witness' confidence is associated with the witness' ability to identify a voice.

The present study also examined one interpretation of the Supreme Court's criterion that a witness' degree of attention to a voice is important for later recognition accuracy. This guideline has been validated in voice recognition research (e.g., Legge, Grosmann, & Pieper, 1984) which has found that persons are less accurate in identifying a voice from a lineup when the number of targets needed to be identified is increased (i.e., the more voices exposed to, the less attention a person can give to remembering any one voice). The present study was aimed at discerning whether attention given to different aspects of one particular voice affects subsequent description and recognition accuracy. The results found that subjects who focussed their attention on judging the personality traits of a voice were no more accurate than those who simply listened to a voice or those who focussed their attention on superficial characteristics (e.g., the sex and speech rate). Such findings are further reinforced

by other voice recognition studies which have failed to obtain significant differences in recognition accuracy by manipulating levels of attention to a single voice (Deffenbacher et al., 1989). Thus again, contrary to the Supreme Court's guideline that a person's degree of attention influences recognition accuracy, the present findings suggest that the qualities a person attends to when hearing a person speak does not affect how accurately they can later describe or recognize a voice.

One final finding from the present study which I believe has important practical implications is the low voice identification accuracy for subjects. Although 30% of the subjects correctly identified the target voice they were exposed to (which was above what could have been predicted by chance), slightly less than 29% of subjects in the intermediate processing condition were able to do so. The significance of this is that in attempting to generalize the results of the present study to actual crime scenarios, the fact that no prior instructions were given to subjects in the intermediate condition perhaps most closely reflects the attention level of actual earwitnesses. In other words, just as persons who are witness to a voice related crime will

tend to listen to and comprehend what is taking place, subjects in the intermediate (no instruction) condition of the present study did likewise. As a result it can be assumed, based on the voice identification rates for intermediate condition subjects, that identification rates in actual situations would be no better. Furthermore, it is important to consider that the present investigation provided only a 15 minute retention interval for voice recognition tasks and was conducted under controlled conditions so as to allow for maximum earwitness performance. It is doubtful that had the present study been conducted under more realistic conditions, by employing factors such as stress and a longer retention interval, that voice recognition rates would have been the same or higher. Although other factors such as the length of a speech sample or number of foils in a lineup can strongly influence earwitness identification accuracy (Bull & Clifford, 1984), it seems apparent, based on the current study's somewhat low recognition rates, that instances of prosecuting persons solely on voice identification evidence should be critically questioned.

Theoretical Implications

Having discussed some of the legal ramifications of the present study, it is also important to consider some of the theoretical implications of the present findings.

In examining whether voice identification accuracy is influenced by varying levels of processing, the present results do not confirm such a hypothesis. Subjects who were instructed to judge the sincerity and friendliness of a voice (deep processing) were no more accurate in identifying the voice than those subjects who were told to judge the speech rate and gender of the voice (shallow processing) or those instructed to only listen to the voice (intermediate processing). Although such results support the findings of previous depth-of-processing research in voice recognition accuracy (e.g., Clifford & McCardle, cited in Bull & Clifford, 1984; Hammersley & Read, 1985), they sharply contrast with those found in studies of facial recognition accuracy (Bloom & Mudd, 1991; Bower & Karlin, 1974; Sporer, 1991; Winograd, 1981).

One possible explanation for such results is that the present encoding instructions for subjects in differing conditions may not have been appropriate for

producing comparable effects. In other words, unlike
facial recognition studies, research in voice
recognition studies within the levels-of-processing
framework may fail to employ a proper independent
measure of depth. Consequently, the present study's
inability to find significant differences in voice
recognition accuracy between conditions may have been
due to inadequate deep and shallow encoding
instructions. However, it is important to consider that
past voice recognition experiments have used different
encoding tasks other than those employed in the present
study (Bull & Clifford, 1984). For instance, the
present study was based primarily on the depth of
processing framework of Winograd (1981), who theorized
that deep memory processing is produced by judging the
personality traits of a face. In order to make such a
judgement, Winograd argued that persons need to
incorporate numerous features of a face. Such encoding
would in turn produce better recognition accuracy.
However, the present results fail to confirm the
effectiveness of this theoretical application in regard
to voice recognition accuracy. Moreover, the present
findings are in agreement with those found in
experiments by McCardle and Clifford (cited in Bull and

Clifford, 1984) who similarly detected no differences
in voice recognition accuracy for subjects who judged
the warmth of a speaker (deep processing) and those who
judged the speaker's age and sex (shallow processing).
These studies also employed other encoding tasks such
as instructing subjects to judge whether a voice
reminds subjects of someone they know. Such directions
assessed the levels-of-processing approach devised by
Craik and Lockhart (1972) by discerning whether
attaching semantic meaning to a voice (i.e., by
associating the voice to some familiar person) would
produce deeper voice processing and hence better
recognition accuracy. However, results were no better
for this encoding task than for those that judged
personality characteristics or those that judged the
gender of a voice (Bull & Clifford, 1984). These
results, combined with those found in the present
study, are significant in that both failed to produce
superior voice recognition accuracy by employing a wide
range of deep encoding strategies (i.e., neither making
associations nor attending to personality traits of a
voice produced better identification accuracy).
Consequently, it is doubtful that employing different
encoding strategies, such as having subjects judge

other personality characteristics not yet examined in

voice recognition research, will produce significantly

better voice identification accuracy for deep

processing subjects. Thus, the interpretation that the

failure to find differing effects for levels of

processing in earwitness research is caused by

inappropriate encoding instructions is not favored.

A possible alternative explanation for the failure

of varying depths of processing to affect voice

recognition accuracy is that a longer retention

interval may be needed. The depth of processing

approach theorizes that deep processing of information

will create better memories (and consequently better

memory performance), whereas shallow processing will

engender only superficial short-term memory.

Consequently, it is plausible that the 15 minute

interval between target voice exposure and recognition

procedures was not long enough to differentiate voice

retention abilities, and voice recognition performance,

between shallow processing and deep processing

subjects.

Another aspect of the present findings that has

important psychological implications regards the depth

of processing voices on voice description accuracy, and

its relation to voice recognition accuracy. In

examining the effects of the levels of voice processing

on description accuracy, the present study found no

significant differences between conditions.

Consequently, there is no support for the hypothesis

that subjects in the deep processing condition would be

more accurate in describing a voice than those in the

intermediate and shallow processing conditions.

However, a comparison of descriptive score means

between conditions showed that, similar to the results

for depth of processing on identification accuracy,

shallow processing subjects were slightly (but

insignificantly) more accurate than subjects in the

intermediate processing and deep processing conditions,

respectively. One possible explanation for such

findings regards the process of retrieving previously

encoded information. More specifically, studies have

shown that the depth of processing leads to potentially

better memory performance when the task to retrieve

information is most similar to the level of processing

in which the material was encoded (Klatzky, 1980). For

example, suppose one person is given a list of words

and asked to assess whether each word is appropriate in

a sentence completion exercise, while another person is

given the same list but is instead instructed to determine whether each word rhymes with a corresponding word. In this example, the person in the sentence completion exercise is engaging in deep memory processing (because of having to think about the meaning of the word) while the other person is taking part in more superficial processing (since rhyming does not include semantic meaning). However, the individual participating in shallow memory processing would most likely remember more listed words if both subjects were asked to recall any words that rhymed with, for example, "toy" (Matlin, 1989). The reason proposed for better memory performance for the shallow processing subject is that the material was encoded into memory according to its sound (i.e., its acoustic characteristics) and was retrieved from memory by an acoustic cue. Although the current study found no statistically reliable difference in description accuracy between conditions, it is conceivable that the insignificant tendency for shallow processing subjects to have more accurate description score means than other conditions is because their encoding instructions (to judge the rate and gender of the voice) were more closely related to the actual description measure (to

rate different speech characteristics).

Also of psychological importance in the present study was the inability to discover an effect for the level of voice processing between description and recognition accuracy. Results of the present study showed that no statistically significant correlation existed between subjects voice description accuracy and voice identification accuracy within each condition. Consequently, contrary to my predictions, subjects in the deep processing condition were no more accurate in describing a voice and later identifying it than subjects in either intermediate or shallow processing conditions. One possible explanation for such results is that, similar to the interpretation for description accuracy between conditions, the lack of a description-identification relationship may be due to an incongruence between encoding and retrieval strategies. For instance, Wells and Hryciw (1984), in examining their findings for memory of faces, argued that subjects in deep encoding conditions (e.g., those who made personality trait judgements) were more accurate in identification procedures than in facial Identi-kit reconstructions because the identification retrieval test more closely corresponded with subjects encoding

instructions. In turn, subjects in superficial
encoding conditions (e.g., those told to judge features
such as a nose or eyes) yielded poor identification
accuracy but good Identi-kit reconstructions. Wells and
Hryciw argued that superficial processing subjects,
like those in the deep encoding condition, were more
accurate in facial reconstruction tests because the
retrieval cues closely resembled the superficial
encoding instructions. In adapting these arguments to
the present earwitness results, it is possible that the
failure to obtain a significant correlation between
voice description and identification accuracy is
because different encoding strategies may facilitate
better performance on different forms of retrieval.
In other words, no description-identification
relationship may have been found since recognition
retrieval tasks may more closely correspond with deep
encoding, whereas description retrieval tasks may more
closely parallel shallow encoding. However, this
explanation is not favored since no significant results
were obtained for accuracy in description or
identification procedures with regard to the encoding
instructions employed.

An alternative and more general explanation for these negative findings is that description and identification may involve independent cognitive processes. For instance, description is a recall task in which at least some of the context is provided (i.e., via instructions to think back to the scene when the voice was initially presented) and the individual must retrieve the target information. However, identification is a recognition task in which the target item is provided and witnesses must retrieve contextual aspects of the original episode (e.g., "was this the person who said...") (Wells, 1985). Thus, two distinct sets of retrieval cues exist for recall and recognition: one that provides the context of an event (description/recall) and one that provides the target item (identification/recognition). Moreover, research has shown that these cues for recall and recognition are uncorrelated (Broadbent & Broadbent, 1977; Pigott & Brigham, 1985; Wells, 1985). In applying such findings to the present study, it is arguable that the retrieval cues available during completion of the voice description checklist were not related to those available to subjects in identifying the target voice from a lineup. As a result, voice description accuracy

42

may not be related to identification accuracy since each procedure uses a different retrieval process. Moreover, the present findings suggest that using a depth of processing method to bridge the gap between recall and recognition processes for voices may not be an appropriate approach to observe a relationship between voice description and voice identification accuracy.

Practical and Theoretical Conclusions

In conclusion, the present results provide no empirical support for the validity of several of the Supreme Court's criteria for eyewitness evidence, as applied to earwitness accuracy. As one of the first studies to examine the relationship between prior voice description accuracy and subsequent identification accuracy, the present findings offer no support for the contention that an earwitness who accurately describes the voice of an offender will be more accurate in identifying the voice than a witness whose description is less accurate. The present study also found that subjects confidence in being able to accurately identify a voice is not related to either voice description or voice identification accuracy. Furthermore, this study failed to find any effect for

the type of attention or depth of processing voices on

description accuracy, identification accuracy, or

the relationship between the two. Consequently, the

utility of a levels-of-processing framework for

producing different levels of accuracy in voice recall

or recognition seems questionable.

Future Considerations

One suggestion for further earwitness studies is to

employ a voice lineup in which the target voice is

either present or absent. Omitting the target voice

from some lineups could produce not only a more

accurate measure of voice recognition accuracy, but

could perhaps better assess the description-

identification relationship by determining if subjects

would choose the voice that best fit their description,

even when the target voice was not present.

Another suggestion for future earwitness

investigations, particularly in regard to assessing the

effect of varying levels-of-processing, is the use of a

longer retention delay between target presentation and

voice identification. It is possible that the null

results in voice recognition accuracy for deep,

intermediate, and shallow processing subjects were in

part due to a short retention interval. Since deep

encoding is thought to engender better information
retention while shallow processing encompasses only
superficial, short-term memory, a longer delay (e.g.,
one or two days) might produce reliable differences due
to better preservation of information for deep
processing persons. Consequently, deep processing
subjects could produce superior recognition performance
over time.

A final consideration for future earwitness
studies regards the development of a valid descriptive
measure. Presently, the research on voice description
accuracy has been most plagued by the lack of a valid
voice description measure (Deffenbacher et al., 1989).
Although the current study's descriptive checklist was
arguably sufficient, the actual validity of this
measure has not been demonstrated. Also, due to the
failure to acquire empirical support for the prediction
that deep voice processing would yield better
description accuracy, it is possible that the present
study failed to provide a proper descriptive measure to
facilitate retrieval from deep processing subjects. As
a result, the possibility of developing a descriptive
measure that is valid, and which corresponds with deep
encoding processes, should be investigated. The results

of such research may help to ascertain whether voice
processing influences description accuracy, and may
serve as a basis for research into whether voice
description accuracy predicts voice identification
accuracy. Should research into a valid description
measure provide evidence for a relationship between
voice description and recognition accuracy, such a
measure could perhaps serve as an important practical
tool in law enforcement procedures.

References

Bloom, L. C., & Mudd, S. A. (1991). Depth of processing approach to face recognition: A test of two theories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 556-565.

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. Journal of Experimental Psychology, 103, 751-757.

Broadbent, D. E., & Broadbent, M. H. P. (1977). Effects of recognition on subsequent recall: Comments on "Determinants of recognition and recall: Accessibility and generation" by Rabinowitz, Mandler, and Patterson. Journal of Experimental Psychology, 106, 330-335.

Bull, R. (1978). Eyewitnesses also have ears. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical Aspects of Memory. New York: 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.

47

Clifford, B. R. (1980). Voice identification by human

listeners. Law and Human Behavior, 4, 373-394.

Clifford, B. R., Rathborn H., & Bull, R. (1981). The

effects of delay on voice recognition accuracy.

Law and Human Behavior, 5, 201-208.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of

processing: A framework for memory research.

Journal of Verbal Learning and Verbal Behavior,

11, 671-684.

Deffenbacher, K. A., Cross, J. F., Handkins, R. E.,

Chance, J. E., Goldstein, A. G., Hammersley, R., &

Read, J. D. (1989). Relevance of voice

identification research to criteria for evaluating

reliability of an identification. Journal of

Psychology, 123, 109-119.

Hammersley, R., & Read, J. D. (1985, May). Assessing

the voice recognition performance of single

witnesses. Paper presented at the Midwestern

Psychological Association meeting, Chicago, IL.

Jenkins, F., & Davies, G. (1985). Contamination of

facial memory through exposure to misleading

composite pictures. Journal of Applied Psychology,

70, 164-176.

Klatzky, R. L. (1980). Human Memory: Structures and
Processes. San Francisco: W. H. Freeman & Co.

Matlin, M. W. (1989). Cognition. New York: Holt,
Rinehart and Winston, Inc.

Neil v. Biggers. (1972). 409 U.S. 188.

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

Saslove, H., & Yarmey, A. D. (1980). Long term auditory
memory: Speaker identification. Journal of Applied
Psychology, 65, 111-116.

Sporer, S. L. (1991). Deep-deeper-deepest? Encoding
strategies and the recognition of human faces.
Journal of Experimental Psychology: Learning,
Memory, and Cognition, 17, 323-333.

Wells, G. L. (1985). Verbal Descriptions of faces from
memory: Are they diagnostic of identification
accuracy? Journal of Applied Psychology, 70, 619-
626.

Wells, G. L., & Hryciw, B. (1984). Memory for faces:
Encoding and retrieval operations. Memory &
Cognition, 12, 338-344.

Winograd, E. (1981). Elaboration and distinctiveness in

memory for faces. Journal of Experimental

Psychology: Human Learning and Memory, 7, 181-190.

Yarmey, A. D. (1986). Verbal, visual, and voice

identification of a rape suspect under different

levels of illumination. Journal of Applied

Psychology, 71, 363-370.

Appendix

Voice Description Checklist

The following is a list of speech characteristics, and their accompanying definition, given to subjects for rating a voice on a five-point scale:

1. Intensity (What was the volume of the voice?)

2. Rhythm (Did the voice proceed flowingly?)

3. Pitch (Was the voice of high or low frequency?)

4. Accent (Did the speaker have a distinct accent?)

5. Rate (How quickly did the speaker talk?)

6. Inflection (How much did the pitch fluctuate?)

7. Clarity (How clear was the pronunciation?)

8. Nasality (What degree did the speaker talk nasally?)