Remembering environmental sounds: The role of verbalization at input

JAMES CRAIG BARTLETT

*Program in Psychology and Human Development, University of Texas at Dallas,
Richardson, Texas 75080*

Two experiments investigated the role of verbalization in memory for environmental sounds. Experiment 1 extended earlier research (Bower & Holyoak, 1973) showing that sound recognition is highly dependent upon consistent verbal interpretation at input and test. While such a finding implies an important role for verbalization, Experiment 2 suggested that verbalization is not the only efficacious strategy for encoding environmental sounds. Recognition after presentation of sounds was shown to differ qualitatively from recognition after presentation of sounds accompanied with interpretative verbal labels and from recognition after presentation of verbal labels alone. The results also suggest that encoding physical information about sounds is of greater importance for sound recognition than for verbal free recall, and that verbalization is of greater importance for free recall than for recognition. Several alternative frameworks for the results are presented, and separate retrieval and discrimination processes in recognition are proposed.

The role of verbalization in memory for nonverbal stimuli has received sporadic attention in psychology, and several different methodological approaches have been taken to the problem (e.g., Bahrick & Boucher, 1968; Bartlett, 1932; Brown & Lenneberg, 1954; Ellis, 1974; Kurtz & Hovland, 1953; Paivio & Csapo, 1969, 1973; Price & Slive, 1970; Wyant, Banks, Berger, & Wright, 1972). The present experiments take as a starting point a study by Bower and Holyoak (1973) which examined the relationship between the content of verbalization at input and the content of verbalization at test. In one condition of that study, subjects heard a list of environmental sounds, preselected for ambiguity of interpretation, and attempted to label each sound as it was presented. One week later, the subjects took a recognition test, again labeling each sound as it was presented. When sounds were consistently labeled at input and at test, they received much higher recognition ratings than when they were inconsistently labeled (the same effect was found when the experimenter labeled sounds at input and at test, ruling out item selection artifacts). Bower and Holyoak concluded that sound recognition is based largely upon the storage and retrieval of "perceptual interpretations" (Bower & Holyoak, 1973, p. 366). While the nature of these interpretations is unclear, the dependence between recognition and consistent verbal labeling indicates

that the interpretation formed for a sound is highly correlated with the verbal label assigned to it (e.g., if a sound was labeled "cat meow" on one occasion and "baby crying" on another, it is assumed that the interpretation of the sound differed). Hence, the Bower and Holyoak study suggests that sound recognition is based on verbalizable interpretations which might differ from other possible types of interpretation, especially those related to mental imagery.

Experiment 1 of the present research extends that of Bower and Holyoak, supporting their conclusion that verbalizable interpretations can dominate recognition of previously presented sounds. However, the methodology of both studies involves explicit verbalizations of sounds at input (by either the subject or the experimenter), a procedure that might bias subjects toward the use of interpretations. Experiment 2 examines the effects of such explicit verbalization on the use of interpretations in sound recognition as well as on the accuracy of sound recognition. Experiment 2 also examines the manner in which physical information contributes to the recognition of sounds and the recall of their labels.

EXPERIMENT 1

This study was in progress when the Bower and Holyoak (1973) study was published, and serves to replicate and extend that investigation in two respects. First, the 1-week retention interval used by Bower and Holyoak invites the hypothesis that verbalizable interpretations dominate sound recognition only at relatively long retention intervals, with more "peripheral" physical or sensory codes dominating at shorter delays. Experiment 1 assesses the dependence

This report is based on a dissertation submitted to Yale University in 1975. I wish to thank Ruth S. Day, Robert G. Crowder, Michael Kubovy, Robert Abelson, and Jerome Singer for countless helpful comments and suggestions. Requests for reprints should be addressed to James Bartlett, Program in Psychology and Human Development, University of Texas at Dallas, Box 688, Richardson, Texas 75080.

between recognition and labeling at retention intervals of 5 and 65 min, providing a test of a "sensory decay" hypothesis. A second extension made in Experiment 1 concerns the relationship between recognition of sounds and free recall of their labels. Several recent studies with visual materials (Bahrick, Bahrick, & Wittlinger, 1975; Bahrick & Boucher, 1968; Ellis, 1974; Tversky, 1973) suggest "independence" between the information used for recognition and that used for free recall, but the recognition-labeling dependence observed by Bower and Holyoak implies that a recognition-recall dependence might occur with sound stimuli. Experiment 1 includes a free recall test to gain more evidence on this question.

Method

Procedure. The experimental session included four stages: list presentation, retention interval, free recall, and recognition. The experimenter stated that his purpose was to obtain labels for 40 environmental sounds, and instructed subjects to "listen to each sound and attempt to label it so that another person, seeing only your labels, could form as accurate an idea as possible of what each sound was like." Subjects were allowed to leave some items unlabeled, but were encouraged to produce a label if at all possible.

There were two experimental conditions, which differed in respect to the retention interval. In the immediate condition, list presentation was followed by a 5-min delay filled by instructions and the collection and distribution of response sheets. In the delay condition, there was a 65-min delay which included the same activities plus a 1-h verbal learning experiment.

Following the retention interval, subjects took a 5-min free recall test in which they attempted to reproduce the labels they had given to the input sounds. After free recall, subjects took a recognition test which consisted of the 40 input-list targets and 40 lure sounds not previously presented in the experiment. Subjects attempted to label each sound as it was presented, and also rated their confidence that each sound was "old" or "new." They indicated their confidence with a 6-point scale, on which 6 represented certainty that a sound was old and 1 represented certainty that it was new.

Materials. The stimuli were 80 environmental sounds, some taken from sound effects records and others recorded by the author. They were heterogeneous in nature, including 9 human sounds, 19 animate but nonhuman sounds, 9 musical sounds, and 53 inanimate sounds. The stimuli varied in duration from 1 to 1.4 sec, and were adjusted to approximately equal subjective loudness. The Pulse-Code-Modulation system at the Haskins Laboratories (Cooper & Mattingly, 1969) was used to store and manipulate the stimuli, and to record them onto audio tapes.

There were four input lists, each containing 40 environmental sounds. For construction of the lists, the 80 sounds were divided into two sets of 40 items, Set A and Set B. The division was random with the constraint that each set contained sounds from each of the four categories (animate-human, animate-nonhuman, musical, and inanimate) and sounds of high and low "interpretability" (as judged by the author). Input Lists 1 and 2 each contained the 40 Set A sounds, and Lists 3 and 4 each contained the 40 Set B sounds. Lists 1 and 2 differed only in the ordering of the two 20-item blocks making up a list (items in Serial Positions 1-20 in List 1 occupied Positions 21-40 in List 2), and Lists 3 and 4 differed in the same way. The sounds were recorded at a presentation rate of one sound per 6.4 sec. The intervals between successive sounds ranged from 5 to 5.5 sec, depending on stimulus duration.

The two test lists differed only in the ordering of the two 40-item blocks (items in Serial Positions 1-40 in one list occupied Positions 41-80 in the other list). Test list was purposely confounded with input list so that items in the first 20 serial positions of an input list were always tested in the final 40 serial positions of the test list, and items in the second half of an input list were tested in the first half of the recognition test. Thus, items in the second half of an input list were tested with substantially shorter input-test "lags" than items in the first half of an input list.

Subjects and design. There were 32 subjects, 16 serving in the immediate condition and 16 in the delay condition. Within each retention interval group, each of the four input lists was presented to a subset of four subjects. Thus, the design was a 4 by 2 between-subjects factorial. The subjects were undergraduate students enrolled in the introductory psychology course at Yale University. Participation in the study was partial fulfillment of course requirements.

Results

Of central interest in this study is a comparison of the recognition ratings given to targets labeled consistently at input and at test, and to targets inconsistently labeled on the two occasions. Two scoring criteria for consistent labeling were used: a strict criterion, by which a consistently labeled sound was one labeled with the same word or words at input and at test, and a loose criterion, by which an input and a test label were considered equivalent if they appeared "substantially the same in meaning" to the author. The results with the two scoring criteria did not differ in any important way, and only the loose-criterion data are reported here. Table 1 shows that consistently labeled sounds received substantially higher average confidence ratings than sounds which were not consistently labeled (note that 1 indicated a judgment

Table 1
Recognition Ratings Given to Lures and Targets, Conditionalized Upon Success or Failure
of Consistent Labeling and Free Recall in Experiment 1

Type of Item and Fate in Free Recall	Condition and Labeling at Input and at Test											
	Immediate						Delay					
	Consistent		Not Consistent		Total		Consistent		Not Consistent		Total	
	P	R	P	R	P	R	P	R	P	R	P	R
Targets Recalled	.352	5.64	.084	3.22	.436	5.17	.341	5.77	.070	2.89	.411	5.27
Targets Not Recalled	.319	5.57	.245	3.55	.564	4.69	.400	5.64	.189	3.34	.589	4.90
Total	.671	5.61	.329	3.46	1.000	4.90	.741	5.70	.259	3.22	1.000	5.06
Lures						1.88						2.03

Note—Recognition ratings varied from 1 ("sure new") to 6 ("sure old"). P = proportion of items; R = recognition rating.

of "sure new" and 6 a judgment of "sure old"). This trend was shown by all 32 subjects ($p < .01$ by sign test). However, targets not consistently labeled at input and at test were nevertheless given higher recognition ratings than lures which had not been previously presented in the experiment. Twenty-nine out of 32 subjects showed this trend for discrimination between lures and targets not consistently labeled ($p < .01$ by sign test).

The class of sounds that were not consistently labeled included two subcategories of items: inconsistently labeled sounds which were labeled at input but labeled differently at test (by the loose criterion), and nonlabeled sounds which were not given labels at input (subjects failed to label an average of 5.2 of the 40 sounds presented at input). Recognition confidence given to inconsistently labeled items was 3.35 and 3.28 in the immediate and delay conditions, respectively, and recognition confidence given to nonlabeled items was 3.62 and 3.15 in the two conditions. Clearly, the two subcategories of items were treated similarly by subjects.

Relationship between recognition and free recall.

Free recall scoring was based on a lenient criterion of "equivalence of meaning" between labels produced at input and at the recall test (only 1.1% of the recall responses were counted as intrusions by this criterion, so overall performance estimates would not change significantly if subjects were simply given credit for every recall response). Table 1 shows that free recalled targets were given higher recognition ratings than nonrecalled targets, but this relationship disappeared when recognition was made conditional upon the presence or absence of consistent labeling at input and at test. Thus, the dependence between recognition and free recall appeared attributable to a more direct dependence between (1) free recall and consistent interpretation¹ and (2) consistent interpretation and recognition rating. Free recalled items were consistently labeled with a probability of .82, while nonrecalled items were consistently labeled with a substantially smaller probability of .62. Since the probability of consistent labeling was higher for recalled items, and since consistently labeled items were well recognized, free recalled items showed higher recognition than nonrecalled items.

Subsidiary findings. On some occasions, a subject responded to a lure sound in the recognition test with a label he had previously given to one or more targets in the input list. While not a frequent occurrence (the percentage of lures receiving old target labels was 15% and 20% in the immediate and delay conditions, respectively), it is noteworthy that lures were given higher recognition ratings when given old labels. Old-labeled lures received recognition ratings of 3.35 and 3.42 in the immediate and delay conditions, respectively, while lures not given old labels received ratings of 1.62 and 1.70, respectively. However, the old-labeled lures did not receive ratings as high as

those given to consistently labeled targets (Table 1). Thus, recognition ratings were related both to the retrieval of input labels and to prior presentation of sounds (in agreement with findings by Bower & Holyoak, 1973).

Table 1 indicates that the effects of delay of testing were minimal if not absent in this study; t tests revealed no significant difference between the immediate and delay conditions in respect to the recognition ratings given to targets or to lures. There was perhaps a small effect of delay on free recall. While the overall level of recall did not differ significantly between the immediate and delay conditions, the former did show a slightly greater tendency toward positive recency. Recall of initially labeled sounds from the final eight serial positions was .676 and .528 in the immediate and delay conditions [$t(30) = 2.22$, $p < .05$].

Discussion

Experiment 1 showed a strong dependence between sound recognition and equivalent labeling at input and at test, replicating the Bower and Holyoak (1973) results and supporting the conclusion that sound recognition can be based on the storage and retrieval of verbalizable interpretations. In addition, Experiment 1 provided no indication that the importance of consistent labeling varies with the delay between input and test within the range examined (5-65 min). By itself, such a dependence between recognition and consistent labeling does not show that consistent interpretation has a causal effect on sound recognition. However, Bower and Holyoak found the same relationship between recognition and consistent labeling when they experimentally manipulated labeling consistency (through presentation of labels with sounds at input and at test), suggesting that the relationship is indeed causal. Thus, the present data reinforce those of Bower and Holyoak and suggest that sound recognition depends upon storage and retrieval of verbalizable interpretations with lists as short as 40 items and retention intervals as short as 5 min.

The relationship observed between recognition and free recall extends our knowledge of the retrieval characteristics of sound representations. Targets whose labels were retrieved in free recall were given higher recognition ratings than targets whose labels were not retrieved, but the effect appeared entirely attributable to the following two facts: The set of recalled items consisted largely of sounds that were consistently labeled at input and at test, and consistently labeled sounds were well recognized. The pattern of results suggests that a dependence between recognition and label recall will occur only if there is (1) a dependence between free recall and consistent labeling, and (2) a dependence between consistent labeling and recognition.

In contrast to the present results with environmental sounds, several studies with picture stimuli have suggested that no relationship holds between recognition and free recall (Bahrick & Bahrick, 1971; Bahrick,

Bahrack, & Wittlinger, 1975; Bahrack & Boucher, 1968; Ellis, 1974; Tversky, 1973, 1974). However, this discrepancy is probably the result of differences in testing procedures, rather than of intrinsic differences between memory for sounds and memory for pictures.

In the present study (and in Bower & Holyoak, 1973), the verbal labels recalled in the free recall test encode information more or less specific to a single target sound (this is not to deny that some labels were appropriate to more than one sound). Thus, it is plausible that the accessibility of input labels should aid in the discrimination of targets from lures. Indeed, Experiment 2 shows that subjects given only verbal labels for sounds at input have some ability to subsequently recognize target sounds and to reject lure sounds. On the other hand, all of the relevant picture memory studies used labels that were equally appropriate to targets and lures (e.g., subjects discriminated between a previously seen drawing of a coffee cup, and several similar drawings of coffee cups). Thus, the accumulated data suggest that a relationship between recognition and free recall can exist when information encoded by the to-be-recalled labels is more appropriate to targets than to lures.

EXPERIMENT 2

Experiment 1 and the Bower and Holyoak (1973) study support the following conclusion: If subjects label sounds at input (or attend to labels presented with the sounds), subsequent recognition is highly dependent upon consistent interpretation at test. Experiment 2 examines the effects which verbal labeling itself has upon subsequent memory performance. The experiment includes conditions in which sounds are presented alone or with appropriate labels and incorporates tests of free recall and recognition, the latter designed to allow a complete signal detection analysis of discrimination performance as well as criterion effects. The results suggest that verbal labeling has quantitative effects on recognition discrimination as well as qualitative effects upon the nature of the information stored in memory and how it is retrieved.

While no previous research has investigated the effects of labeling on memory for environmental sounds, such labeling effects have been examined with visual materials. Bartlett (1932) noted two distinctive strategies for remembering faces. His "vocalizers" analytically described the faces to themselves, and adopted a cautious response criterion at test, seldom intruding false details. His "visualizers," on the other hand, appeared to respond to the faces more holistically, and adopted a looser criterion at test, producing many more intrusions. Kurtz and Hovland (1953) attempted to manipulate experimentally verbalization and visualization strategies with school children in the fifth, sixth, and seventh grades. While viewing an array of objects, children in the verbalization condition matched

objects to their names, while children in the visualization condition merely matched each object to its representation in a photograph. Kurtz and Hovland found no significant difference between the hit rates of the two conditions, but did find weak evidence for lower false alarm rates in the verbalization condition. Thus, the study provided preliminary evidence that verbalizers possess a greater ability to discriminate targets from lures, and, consistent with Bartlett's observations, a tendency to adopt relatively strict criterion levels. In addition, Kurtz and Hovland found that the verbalization group showed higher free recall.

More recent experiments have shown that presentation of labels along with pictures can improve subsequent recognition performance when the labels are semantically appropriate to the stimuli and the stimuli themselves are relatively complex and difficult to label (e.g., Ellis, 1974; Price & Slive, 1970). However, these studies concentrate on discrimination measures and provide no information on the criterion levels adopted by subjects. Furthermore, they do not compare presentation of labels with presentation of other types of information, especially physical or "imaginal" information (presumably more available from a nonverbal stimulus than from its label, Paivio, 1971). Experiment 2 does both of these things, providing more analytical information on how labels, as distinct from physical information, alter the nature of mnemonic codes.

Recognition of environmental sounds is compared after list presentation in one of three conditions: the sound condition, in which environmental sounds are presented alone at input; the sound + label condition, in which environmental sounds are presented along with appropriate interpretive labels; and the label condition, in which only verbal labels for the sounds are presented at input. Since the recognition test contains only environmental sounds, comparison of the sound + label and label conditions indicates the importance of encoding physical information about sounds for subsequent recognition. Comparison of performance in the sound + label and sound conditions reveals the effects of labeling at input upon subsequent recognition. Free recall performance is also examined after presentation in the three conditions, to further characterize the effects of verbalization upon sound memory.

Method

There were three major groups of subjects which differed with respect to the stimuli heard at input. The 16 subjects in the sound group heard a list of 32 environmental sounds, the 16 label group subjects heard a list of 32 labels, and the 32 sound + label group subjects heard a list of 32 sound-label pairs, each pair consisting of an environmental sound accompanied by an appropriate label. All sound + label subjects heard half of the sound-label pairs with the labels preceding their respective sounds and half with the labels following their respective sounds. Sixteen of the sound + label subjects received label-before presentation in the first list block and label-after presentation

in the second, while the remaining 16 subjects received label-after presentation first and label-before presentation second.

Procedure. Subjects in all conditions were told in advance the nature of the stimuli they were to hear, and were advised of a subsequent recognition test in which old list sounds were to be discriminated from new lure sounds not previously presented. It was emphasized, especially in the label group, that the recognition test would contain only environmental sounds, presented without labels (note that subjects in the label group attempted to recognize sounds on the basis of labels presented at input).

Immediately after list presentation, all subjects took a 5-min free recall test (of which they had not been forewarned). The label and sound + label groups simply recalled the labels which had been presented at input, while the sound group produced labels or descriptions for as many sounds as they could remember. Immediately following free recall, the recognition test was given. Subjects listened to each sound and rated their confidence that it was old or new, using a 6-point confidence scale ranging from "sure old" (6) to "sure new" (1).

In the final phase of the experiment, the sound + label and label subjects were given a cued recall test for the labels presented at input. Subjects in the sound + label group heard the 32 input sounds, presented in an order different from that of the original list, and attempted to recall the label which previously accompanied each sound. Subjects in the label group heard the same 32 sounds, and attempted to recall the input label which corresponded to each sound. Subjects in the sound condition were given their free recall protocols and completed a free recall matching task. Following the procedure of Shiffrin (1973), these subjects listened to the 32 input sounds and attempted to indicate which sound corresponded to each of their free recall responses.

Materials. The stimuli were 64 environmental sounds and 64 labels, each label corresponding to a sound. The sounds were selected from the 80 items used in Experiment 1, discarding several whose most common label corresponded closely to that of another sound.

The labels were chosen in a two-stage procedure. First, the label or labels most frequently given to each sound in Experiment 1 were selected. This procedure was inadequate in itself as some sounds had two labels "tied" for first place, and the label most frequently given to several sounds seemed (to the author) less congruent with that sound than the label of the object or event which actually produced it. For these reasons, five subjects listened to the entire set of 80 sounds, choosing the most "congruent" label for each sound having more than one plausible label. The preferred label for each sound was that used in Experiment 2. The labels were recorded, stored, and manipulated using the Pulse-Code-Modulation system at Haskins Laboratories. They consisted of one or two words and varied in duration from approximately .5 to 1.0 sec.

To reduce the importance of individual-item effects, eight different presentation lists and two different recognition tests were used in the sound group, the label group, and each of the two sound + label groups (label-after/label-before and label-before/label-after). Each of the 16 combinations of list and recognition test was presented to one subject in each stimulus group.

Presentation lists in the sound, sound + label, and label conditions corresponded in respect to nominal identity of items and the ordering of these items; for any sound + label list, there was a sound list with an identical ordering of sounds and a label list with an identical ordering of labels. Lists were constructed so that each item served as a member of four of the eight lists used with a group of subjects, each item occurring in the first half of two lists and the second half of the remaining two lists.

Each of the two recognition tests contained a randomized ordering of the entire set of 64 sounds used in the experiment.

The two tests differed only in the ordering of the two 32-item blocks making up a test, so that each sound stimulus was tested equally often in the first and second list blocks.

For lists presented to the sound group, the environmental sound stimuli were recorded at a presentation rate of approximately 6.4 sec per item, and an interstimulus interval (ISI) of 5.0-5.4 sec, as in Experiment 1. The sound + label lists were recorded with the same 5.0-5.4-sec interval between successive environmental sounds, so that presentation rate was approximately equal to that of the sound condition (6.4 sec per sound-label pair). The verbal labels of sound + label lists were simply placed within the 5.0-5.4-sec interval separating successive sounds. With both label-before and label-after presentations, an interval of 1.0-1.7 sec separated a label from the sound with which it was paired, and a longer interval of 3.0-3.7 sec separated a sound-label pair from the preceding or following sound-label pair. The label lists were likewise constructed so that presentation rate remained approximately 6.4 sec per item. Since the labels were of shorter duration than the sounds, somewhat longer ISIs were required for these lists, 5.4-5.9 sec.

Subjects. The 64 subjects were undergraduates at Yale University. They participated in the study in partial fulfillment of course requirements.

Results

Free recall. In the sound + label and label conditions, a response was counted as a correct recall if it contained at least one word in common with a label presented at input. By this measure, free recall performance was .550 and .535 in the sound + label and label conditions, respectively, and this difference was not significant by *t* test. Estimates of recall in the sound group are complicated by the fact that some recall responses were not matched to a single input sound in the final matching phase of the experiment. If sound group subjects are given credit only for recall responses which they subsequently matched to a single list sound, their recall performance is .318. If these subjects are simply given credit for all recall responses, their performance is .379. It is noteworthy that "recall" in the sound group fell significantly below that in the sound + label group [$t(46) = 4.78$, $p < .01$] and the label group [$t(30) = 3.24$, $p < .01$], even with the more liberal scoring criterion.

Relationship between free recall, cued recall, and recognition. Table 2 shows the relationship between free recall, cued recall, and recognition confidence for sounds in the sound + label and label conditions, as well as the relationship between free recall and recognition in the sound condition. It reveals two noteworthy results. First, in the sound + label and label conditions, sounds that supported cued recall were recognized with higher confidence than sounds that did not support cued recall. Furthermore, this dependence between recognition and cued recall was as large in the sound + label condition as in the label condition (where the recognition/cued recall dependence is a foregone conclusion). Comparing the sound + label and label conditions, it appears that sound presentation did not reduce the dependence between sound recognition and cued recall. Second, there was a

Table 2
Recognition Ratings Given to Lures and Targets, Conditionalized Upon Success or Failure
of Cued Recall and Free Recall in the Three Conditions of Experiment 2

Type of Item and Fate in Free Recall	Condition and Fate in Cued Recall													
	Sound + Label						Label							
	Recalled		Not Recalled		Total		Recalled		Not Recalled		Total		Sound*	
	P	R	P	R	P	R	P	R	P	R	P	R	P	R
Targets Recalled	.457	5.51	.093	3.99	.550	5.25	.367	5.42	.168	3.52	.535	4.83	.318	5.56
Targets Not Recalled	.278	5.43	.172	3.62	.450	4.74	.238	4.86	.227	3.60	.465	4.25	.682	4.67
Total	.735	5.48	.265	3.75	1.000	5.02	.605	5.20	.395	3.57	1.000	4.56	1.000	4.95
Lures						2.13						3.00		2.53

Note—P = proportion of items; R = recognition rating. *Free recall in the sound condition is based on responses which were matched to a single list sound in the final phase of the experiment.

noticeable dependence between recognition and free recall. Sounds whose labels were retrieved in the free recall test received higher recognition ratings than those whose labels were not retrieved. However, this recognition/free recall relationship diminishes when recognition is conditionalized upon success or failure of cued recall. Especially important is the strong dependence between recognition and free recall in the sound condition, shown by all 16 subjects ($p < .01$ by sign test). The data from Experiment 1 suggested that the recognition/free recall relationship reflects a more direct relationship between (1) free recall and consistent interpretation and (2) consistent interpretation and recognition. If so, the present results suggest a dependence of recognition upon consistent interpretation in the sound condition.

Signal detection analysis of recognition. Figure 1 shows memory operating characteristics (MOCs) for recognition in the sound + label, sound, and label groups. While this graph is based on the overall number of hits and false alarms per condition (an inappropriate procedure from the viewpoint of signal detection theory), Tables 3 and 4 show that the major trends remain when signal detection parameters are estimated separately for each individual subject, prior to averaging. Table 3 displays average area under the MOC and average slope of the MOC on normal-normal coordinates. Table 4 displays average criterion levels. Each of these measures will be discussed in turn.

Discrimination of targets from lures can be measured by the area under the MOC², and Figure 1 suggests that this area is smaller in the label group than in the sound + label and sound groups. T tests revealed a significant difference between the label and sound + label groups [$t(46) = 6.30$, $p < .001$] and between the label and sound groups [$t(30) = 3.73$, $p < .001$]. Figure 1 also suggests a smaller difference between the sound + label and sound groups, and this difference is likewise supported by t test [$t(46) = 2.02$, $p < .05$].

In addition to purely quantitative differences in discrimination performance, Figure 1 suggests two qualitative effects. First, there is a suggestion of differences in the slopes of the MOCs. In order to test

the reliability of these trends, an estimate of MOC slope was made for each subject using the Grey and Morgan (1972) program (the average slope values are shown in Table 3), and t tests were performed. The slope difference between the sound + label and sound groups was highly significant [$t(46) = 3.75$, $p < .001$], and that between the sound and label groups was significant as well [$t(30) = 2.44$, $p < .05$].

The MOC plots also suggest that the sound + label group subjects adopted stricter criterion levels than the sound group subjects. At the middle 123/456 cut, for example, Figure 1 shows that the sound group is further toward the upper right than the sound + label group. The criterion data are more clearly represented in Table 4, which shows the average values of β^3 for subjects in each condition at each criterion cut (the 6-point confidence scale allows five criterion cuts). The difference between the sound + label and sound

Figure 1. MOCs for recognition performance in the sound + label, sound, and label conditions of Experiment 2.

Table 3
Average Area Under the MOC and Average Slope of the MOC
(On Normal-Normal Coordinates) for Recognition Performance
in the Sound + Label, Sound, and Label Conditions
of Experiment 2

Condition	Measure	
	Area	Slope
Sound + Label	.885	.774
SD	.053	.317
Sound	.852	1.166
SD	.055	.388
Label	.734	.837
SD	.114	.377

Table 4
Average Levels of β' for Recognition Performance at Each
Criterion Cut in the Sound + Label, Sound, and Label
Conditions of Experiment 2

Condition	Criterion Cut				
	1/23456	12/3456	123/456	1234/56	12345/6
Sound + Label	-.777	-.319	.059	.442	.794
SD	.186	.404	.364	.254	.177
Sound	-.901	-.628	-.251	.248	.708
SD	.103	.245	.399	.395	.342
Label	-.668	-.449	-.035	.266	.545
SD	.316	.229	.170	.258	.277

groups was significant by *t* test ($p < .05$) at four of the five criterion cuts (the 12345/6 cut being the only exception). Like the sound + label group, the label group showed relatively strict criterion levels at the 1/23456, 12/3456, and 123/456 cuts, suggesting that label presentation induces relatively stringent criteria, whether or not sounds are also presented. However, the label group showed relatively low criterion levels at the 12345/6 cut, suggesting that these subjects are reluctant (or unable) to adopt extremely strict response criteria.

The observed slope and criterion effects are *not* direct consequences of differences in discrimination performance between conditions. Subjects in each stimulus condition were dichotomized into "good" and "poor" performers on the basis of area-under-MOC scores, and the average area, slope, and criterion scores were calculated for each group of subjects. The general pattern of slope and criterion effects held for both good and poor performers. Especially informative is a comparison of poor performers from the sound + label condition with good performers from the sound condition. The two groups showed approximately equal discrimination performance (area scores were .842 and .897 for poor sound + label subjects and good sound subjects, respectively). However, the poor sound + label subjects still showed stricter criterion levels and more shallow slopes than the good sound subjects (β' scores at the 123/456 cut were .048 and

-.309 for poor sound + label subjects and good sound subjects, respectively, and average slope was .782 and 1.163 for poor sound + label subjects and good sound subjects, respectively).⁴

Discussion

Recognition of environmental sounds is highly dependent upon consistent labeling at input and at test in either of two conditions: when labels are generated by subjects hearing sounds (Experiment 1 and Bower & Holyoak, 1973) or when labels are presented to subjects by the experimenter (Bower & Holyoak, 1973). In either case, a change in verbal labeling apparently reflects a change in the representation used for sound recognition, while consistency of labeling reflects a consistency in the internal representation. These observations suggest that sound recognition can be based on the storage and subsequent retrieval of interpretations which are closely related to verbal labels. Although the precise nature of such interpretations remains obscure, it is tempting to equate them with word "meanings" and word "senses," thought by some researchers to underlie long-term retention of words (e.g., Martin, 1975; Reder, Anderson, & Bjork, 1974).

Free recalled items are more likely to be consistently interpreted than nonrecalled items, and free recall appears to predict sound recognition for this reason (Experiment 1). The strong relationship between recognition and free recall in the sound condition of Experiment 2 suggests that recognition depends upon consistent interpretation even when the task does not require subjects to explicitly label sounds at input or to process labels presented by the experimenter.

While the evidence argues clearly that interpretations are used in sound recognition, Experiment 2 suggests that explicit verbalization of these interpretations is not the only strategy for processing sounds, nor necessarily the best strategy in all situations. Comparison of the sound + label and sound groups suggested positive effects of labeling on free recall and on recognition as indicated by an overall measure of discrimination (area under the MOC). In addition, sound group subjects adopted looser recognition criteria than sound + label subjects and provided MOCs of steeper slope. These differences support two conclusions. First, while sound group subjects probably implicitly labeled at least some input sounds, their labeling was either less frequent than that of sound + label group subjects, or different in some qualitative way. Second, the difference in verbalization between the sound + label and sound conditions had a qualitative effect on the way in which the recognition task was performed.

Figure 1 suggests a description of the labeling effects with interesting theoretical implications. When the false alarm rate was low (e.g., .05), the hit rate in the sound + label group was higher than that in the sound group. However, when the false alarm rate was higher

(e.g., .50), the superiority of the sound + label group was smaller, if not absent altogether. Thus, the data are compatible with the conclusion that verbalization at input is efficacious for verbal free recall and for recognition at strict criterion levels where the false alarm rate is low, but not for low-criterion recognition where the false alarm rate is high. This conclusion is further supported by the tendency for sound + label subjects to spontaneously adopt stricter levels of criterion than subjects in the sound group (Table 4).

The effects of sound presentation differed from those of label presentation. Sound presentation (comparing the sound + label and label groups) had relatively small effects (if any) on verbal free recall and relatively large effects on recognition, raising the hit rate at all levels of false alarm rate (see Figure 1). Further, while label presentation tightened response criteria at all five criterion cuts in recognition, sound presentation did so only at the highest 12345/6 cut (Table 3). At the very least, these data show that the puzzling effects of label presentation are not merely artifactual consequences of an improvement in overall performance. Sound presentation also improved recognition performance, but did not decrease MOC slope or cause a general increase in response criteria.

GENERAL DISCUSSION

It is useful at this point to summarize several possible accounts of the present findings, as they raise important theoretical questions regarding the role which verbalization might play in memory for nonverbal stimuli. One class of theories assumes two or more qualitatively different types of mnemonic information in long-term memory situations, and it might be proposed that label presentation facilitates storage of one information type while having no positive effect on storage of the other(s). For example, a dual code theorist (Paivio, 1971) might suggest that label presentation increases the probability that a verbal code is stored for each sound, while having no effect on the storage of a nonverbal or imaginal code. Assuming that the verbal code facilitates free recall and high-criterion recognition, while the nonverbal code is the sole support of low-criterion recognition, the present results could be accommodated. Even the criterion-level data fall into place with the further assumption that subjects have some insight into the mnemonic characteristics of the information they have stored. Subjects who have stored more verbal codes might realize that these are useful for high-criterion recognition, and adopt relatively strict criterion levels for that reason.

Experiment 2 converges with other evidence for a functional difference between verbal and imaginal codes in memory for environmental sounds. Comparing free recall and serial recall after presentation of environmental sounds and their labels, Paivio, Philipchalk, and Rowe (1975) and Philipchalk and Rowe (1971)

have found evidence that (1) dual imaginal and verbal codes support memory for sounds and (2) imaginal codes are specialized for the retention of item information, as opposed to order information. The present research is consistent with a specialization of imaginal codes for recognition at loose criterion levels as well as for item information.

Also within the class of "dual component" theories are the suggestions of Bahrack and Boucher (1968) and Tversky (1973). According to this viewpoint, label presentation might improve or organize the storage of "class" features of stimuli, those features which determine the label or concept to which a sound is assimilated, while not affecting the storage of within-class features, those features which allow discrimination of one category member from another (e.g., the discrimination between two "cat meows" and two "squeaky doors"). On the assumption that class features improve free recall and high-criterion recognition, it follows that label presentation would improve performance in these situations. If low-criterion recognition were based solely on within-class features, label presentation would have no positive effect upon it.

Still a third variant of the dual component view has been proposed by Mandler (1972) and recently expanded by Humphreys (1976). Both authors distinguish between information specific to individual to-be-remembered items, and "relational" information, responsible for the effects of list organization and context on recognition memory. Humphreys has provided evidence that the relational information underlying context effects is used only at high criterion levels (or only by subjects who spontaneously adopt high criterion levels). Thus, it appears plausible that verbal labels accompanying sounds provide relational information of the same type as that investigated by Humphreys with verbal materials.

The various dual component theories are actually quite similar, and make identical predictions in many situations. All suggest some type of independence between the two classes of information they delineate, and thus all are consistent with the major result of Experiment 1: Sound recognition is strongly related to consistent interpretation, but sounds which are not consistently labeled are nevertheless discriminated from lures. It is conveniently assumed that the relationship between recognition and consistent labeling reflects the use of one type of information (e.g., verbal codes), while the discrimination of nonlabeled and inconsistently labeled sounds from lures reflects the use of another (e.g., imaginal codes). Similarly, all of these theories are compatible with the relationship between recognition and cued recall in the sound + label and label conditions of Experiment 2. Recognition in the sound + label group was as strongly related to cued recall as recognition in the label group, although the sound + label group showed substantially higher discrimination performance. It is plausibly argued that the relationship between recognition and cued

recall reflects (again) the use of one type of information (that available from labels), while the superior performance of the sound + label group reflects the use of the other (that which depends upon presentation of sounds). If this assumption is made, the data suggest that the two types of information are additive in their effects on recognition. Choosing among the various dual component theories is difficult, and perhaps depends upon a more direct assessment of the information processing activities which subjects perform with different stimuli and tasks (e.g., Kosslyn, Holyoak, & Huffman, 1976).

A second class of theory is a single component view which does not postulate two qualitatively different types of mnemonic information. Rather, a single component view suggests that label presentation merely improves the representations formed for sounds. For example, the labels supplied in the sound + label condition might be, on the average, more congruent with the sounds than labels sound group subjects generated on their own. We might further assume (post hoc) that "good" vs. "poor" interpretations are more important for free recall and high-criterion recognition than for recognition with a loose criterion. Posner (1973) has distinguished between "tight" and "loose" concepts in his discussion of concept attainment. Tight concepts are associated with a strict definition of positive instances while loose concepts, often difficult to verbalize, are more broadly defined. In the present context, we might speculate that label presentation induces the formation of tighter concepts for the stimulus sounds, and therefore favors a strict criterion where the false alarms to lure sounds (often similar to targets) must be kept to a minimum. Furthermore, if looser concepts are indeed more difficult to verbalize than tighter concepts, they should be associated with relatively poor recall performance, as was found in the sound group.

Such a single component view has no particular difficulty with the results of Experiment 1. The dependence between recognition and consistent labeling presumably reflects the fact that the interpretations formed for sounds vary with the label applied to them. However, as long as interpretations are not identified with verbal labels, there need not be a *perfect* correlation between consistency of interpretation and consistency of labeling. Hence, inconsistently *labeled* sounds might occasionally be consistently *interpreted* and recognized.

However, a single component theory appears to have more difficulty with two other results, both from Experiment 2. First, it fails to predict the invariant relationship between recognition and cued recall in the sound + label and label conditions. Why does sound presentation merely improve discrimination performance without reducing the relationship between recognition and cued recall? Dual component theories propose a rather simple answer to this question, but a

single component view does not. A second difficulty for the single component view is the fact that label presentation (comparing the sound + label and sound conditions) affects performance differently than sound presentation (comparing the sound + label and label conditions). While label presentation decreases MOC slope and raises criterion levels, sound presentation has no consistent effect on either measure. If both sound presentation and label presentation merely improve or "tighten" interpretations, these qualitative differences are difficult to explain.

A better test of the dual component vs. single component theories would be a comparison of the sound + label and sound conditions with a sound + sound condition in which each sound is presented twice in a list. According to a single component theory, repetition of a sound should improve or tighten its interpretation just as labeling does. Therefore, repetition and label presentation should have the same type of effect on MOC slope and criterion levels. A dual component theory, as developed here, predicts that only labeling should decrease slope and raise criterion levels. Repetition should have either no effect on these measures or an effect in the opposite direction.

Neither the dual component nor single component theory does much to clarify the question of why labeling or any variable affecting initial encoding should raise criterion levels or decrease slope of the MOC. One plausible account for such effects is based on signal detection theory. According to this framework, the slope of an MOC reflects the relative variance of the noise and signal + noise distributions assumed to underlie recognition performance. The shallow slopes of the sound + label and label conditions reflect a relatively large variance in the signal + noise distribution, while the steeper slopes of the sound condition reflect relatively less variance in the signal + noise distribution. Thus, a signal detection theory analysis would suggest that label presentation increases the variance of the signal + noise distribution relative to the noise distribution.

The signal detection view is not incompatible with the dual component view. For example, if label presentation at input increases the probability that recognition will be based on dual verbal and imaginal codes, it might plausibly increase the variance of the signal + noise distribution, especially if the codes are retrieved independently. Thus, the storage of verbal codes at input might decrease the slope of the MOC, as well as improve discrimination performance.

A second possible approach to the slope and criterion effects is based on the distinction between recognition (awareness of an object as one that has been presented before) and discrimination (the perception of differences), of which Thomson (1972) has recently reminded us. While the recognition of some objects and not others logically entails discrimination, such discrimination might be an indirect consequence of the fact that some

stimuli "feel familiar" while others do not. Psychologically, the "perception of differences" need not be involved in recognition.

In a free choice recognition test such as that used here, performance at low criterion levels might reflect only the *psychological* process of recognition. If a test sound does not access the appropriate mnemonic information, the subject feels certain that it is a lure, and responds with a recognition rating of 1 ("sure new") or 2 ("think new"). However, if the test sound succeeds in accessing the appropriate information in memory, the subject treats the item as possibly a target, responding with a 3, 4, 5, or 6. Thus, discrimination between targets and lures at loose criterion levels would depend upon the fact that some test sounds access appropriate information, while others do not.

If a sound has accessed appropriate mnemonic information, a subject's response of 3 or 4 vs. 5 or 6 would depend upon the psychological process of discrimination, perhaps involving comparison of a test sound with mnemonic information retrieved from memory. Performance at strict criterion levels would therefore depend upon the "perception of differences" between "recognized" sounds and information in memory traces formed at input. Thus, regardless of which theory or framework is accepted, it might be argued that label presentation at input does not improve subjects' ability to recognize sounds when repeated subsequently, but does improve their ability to discriminate between "recognized" targets and "recognized" lures.

REFERENCES

- BAHRICK, H. P., & BAHRICK, P. Independence of verbal and visual codes of the same stimuli. *Journal of Experimental Psychology*, 1971, **91**, 344-346.
- BAHRICK, H. P., BAHRICK, P. O., & WITTLINGER, R. P. Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, 1975, **104**, 54-75.
- BAHRICK, H. P., & BOUCHER, B. Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 1968, **78**, 417-422.
- BARTLETT, F. C. *Remembering: A study in experimental and social psychology*. Cambridge, England: The University Press, 1932.
- BOWER, G. H., & HOLYOAK, K. Encoding and recognition memory for naturalistic sounds. *Journal of Experimental Psychology*, 1973, **101**, 360-366.
- BROWN, R., & LENNEBERG, E. H. A study in language and cognition. *The Journal of Abnormal and Social Psychology*, 1954, **49**, 454-462.
- COOPER, F. S., & MATTINGLY, I. G. Computer-controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America*, 1969, **46**, 115. (Abstract)
- ELLIS, H. C. Stimulus encoding processes in human learning and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7). New York: Academic Press, 1974.
- GREY, D. R., & MORGAN, B. J. T. Some aspects of ROC curve-fitting: Normal and logistic models. *Journal of Mathematical Psychology*, 1972, **9**, 128-139.
- GRIER, J. B. Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 1971, **75**, 424-429.
- HODOS, W. Nonparametric index of responses bias for use in detection and recognition experiments. *Psychological Bulletin*, 1970, **74**, 351-354.
- HUMPHREYS, M. S. Relational information and the context effect in recognition memory. *Memory & Cognition*, 1976, **4**, 221-232.
- KOSSLYN, S. M., HOLYOAK, K. J., & HUFFMAN, C. S. A processing approach to the dual coding hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, **2**, 223-233.
- KURTZ, K. H., & HOVLAND, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *Journal of Experimental Psychology*, 1953, **45**, 157-164.
- MANDLER, G. Organization and recognition. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.
- MARTIN, E. Generation-recognition theory and the encoding specificity principle. *Psychological Review*, 1975, **82**, 150-153.
- PAIVIO, A. *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston, 1971.
- PAIVIO, A., & CSAPO, K. Concrete image and verbal memory codes. *Journal of Experimental Psychology*, 1969, **80**, 279-285.
- PAIVIO, A., & CSAPO, K. Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, 1973, **5**, 176-206.
- PAIVIO, A., PHILIPCHALK, R., & ROWE, E. J. Free and serial recall of pictures, sounds, and words. *Memory & Cognition*, 1975, **3**, 586-590.
- PHILIPCHALK, R. P., & ROWE, E. J. Sequential and nonsequential memory for verbal and nonverbal auditory stimuli. *Journal of Experimental Psychology*, 1971, **91**, 341-343.
- POSNER, M. I. *Cognition: An introduction*. Glenview, Ill: Scott, Foresman, 1973.
- PRICE, R. H., & SLIVE, A. B. Verbal processes in shape recognition. *Journal of Experimental Psychology*, 1970, **83**, 373-389.
- REDER, L. M., ANDERSON, J. R., & BJORK, R. A. A semantic interpretation of encoding specificity. *Journal of Experimental Psychology*, 1974, **102**, 648-656.
- SHIFFRIN, R. M. Visual free recall. *Science*, 1973, **180**, 980-981.
- THOMSON, D. M. Context effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 497-511.
- TVERSKY, B. Encoding processes in recognition and recall. *Cognitive Psychology*, 1973, **5**, 275-287.
- TVERSKY, B. Eye fixations in prediction of recognition and recall. *Memory & Cognition*, 1974, **2**, 275-278.
- WYANT, S., BANKS, W. P., BERGER, D., & WRIGHT, P. W. Verbal and pictorial similarity in recognition of pictures. *Perception & Psychophysics*, 1972, **12**, 151-153.

NOTES

1. The relationship between free recall and consistent labeling is due in part to the fact that subjects did not label all items at input. Such nonlabeled items cannot be recalled nor consistently labeled. However, Experiment 2 shows a powerful relationship

between free recall and cued recall in the sound + label condition, where all sounds are labeled at input. Thus, there is a clear relationship between the accessibility of a label in free recall and its accessibility in response to an appropriate sound in a recognition or cued recall test.

2. The area measure is the total area under the MOC for a subject divided by the total area in an MOC chart. While it is possible to have an area score of less than .50 (the score obtained if all points fall along the positive diagonal), as a practical matter, the area measure varies from .50 (chance performance) to 1.00 (perfect performance).

3. β' is a nonparametric measure of response bias reported by Hodos (1970) and developed by Grier (1971). It varies from -1.00 (loosest possible criterion) to 1.00 (strictest possible criterion), with a score of 0.00 representing points falling on the

negative diagonal of an MOC chart. According to Hodos, the measure assumes an MOC which is symmetrical about the negative diagonal, a condition which does not hold in the present data since slopes are not equal to unity. However, the measure is as free of assumptions as any known to the author and does reflect the extent to which data points fall to the left (positive numbers) or to the right (negative numbers) of the negative diagonal.

4. The effects of labeling on both slope and criterion levels have been replicated in other research. However, statistically significant outcomes appear to depend on a minimum of 512 observations per point on the MOC (as was used here).

(Received for publication May 3, 1977.)