The role of loudness in auditory imagery

MARGARET JEAN INTONS-PETERSON
Indiana University, Bloomington, Indiana 47405

Four experiments studied characteristics of auditory images initiated by named but unheard sounds. The sounds varied in their loudness ratings. As the difference between the loudness ratings of the two sound phrases increased, the times to mentally equate the loudness of the two images increased, whereas the times to identify the louder (or softer) of the images decreased. Moreover, congruity effects were found in the comparative judgment task: Times were faster to identify the louder of two loud-rated stimuli than to judge the softer of the same two stimuli, and times were faster to identify the softer of the two soft-rated than two loud-rated stimuli. The loudness ratings did not always influence performance, however, for neither an image generation nor a reading task showed response times that varied with loudness ratings. These results suggest that sensory/perceptual components are optionally represented in auditory images. These components are included when appropriate to a given task. A control experiment showed that the results cannot be considered epiphenomenal.

This paper considers manipulations of verbally named sounds when the sounds themselves are absent. Presumably, the names of the sounds guide the search for associated information in memory. But what information? Are perceptual characteristics such as loudness always included in the auditory representations or are these characteristics included only when relevant to the task? Do the subjects have sufficient knowledge of such imagery processes to be able to predict performance accurately? To what extent do demand characteristics operate in this setting? Answers to these questions are not available in the literature, for few experimenters have studied auditory imagery. The experiments reported in this paper address some of these questions.

Suppose that, in general, memorial representation (images) are closely related to perceptual processes, as suggested by a number of investigators (e.g., Neisser, 1972; Shepard & Podgorny, 1978). Though the evidence for this position is largely based on visual imagery, it implies that auditory representations should encode loudness information, a major dimension of actual auditory experience. It follows that the times to generate auditory images of the sound descriptions might vary (inversely) with an index of loudness, the loudness rating of the sound phrases. This suggestion, however, seems counterintuitive, for while loudness information appears to be an essential, even an integral, feature of actually experienced sounds, it is not an obvious attribute of imagined sounds. For example, it does not seem to take longer to imagine a cat purring than to imagine a door slamming. Under some conditions, then, auditory imagery might not involve loudness information.

Alternatively, one can introspectively equate or compare the loudness of two auditory images. This ability, if experimentally verified, would indicate that loudness information is retrievable from long-term memory, when needed. These considerations suggest that the inclusion of a loudness dimension in auditory imagining is optional, that is, on instruction subjects can include it in their auditory imagery, even though it may often be absent.

These considerations also suggest a tentative framework. Auditory representations include such perceptual information as loudness. The retrieval of this information depends on the search strategy. This strategy targets information required by the task. Hence, when loudness is required by the task, the search strategy is biased toward retrieval of this information. When loudness is not important, the retrieval search will access this information on a random basis.

To investigate the above predictions, we selected tasks that did or did not encourage the retrieval of loudness information. Experiment 1 tests a situation that does not encourage the retrieval of loudness: the simple generation of auditory images to correspond to the sound phrases. One task in Experiment 2 also does not require loudness information: the reading of the sound phrases. In contrast, the other task in Experiment 2 and those in Experiments 3 and 4 encourage the retrieval of loudness by asking the subjects to equate or compare the loudness of two verbally described sounds. We expect the reaction times to vary with the loudness ratings for the latter tasks but not for the former.

The specific relationship between reaction times and the loudness ratings depends on the task. When subjects equate or match the loudness of imaginary sounds, the times to achieve the matches will be directly related to the differences between the loudness ratings of the two described sounds (Experiments 2 and 3). However, when

subjects identify the louder (or softer) of the imaginary sounds, the comparison reaction times will be inversely related to the differences in loudness ratings (Experiment 4).

Even data consistent with these expectations would be subject to other interpretations. One is that people know and can predict their own cognitive processes. Their knowledge of these processes, rather than characteristics of auditory representations per se, could mediate the reaction time performance. Another, related interpretation is that imagery results are epiphenomenal. This interpretation is based on the idea that people are accustomed to making various kinds of perceptual judgments and have a substantial amount of knowledge about the physical relationships that exist in the world. Their laboratory performance, particularly when the comparison stimuli are physically absent, may reflect this knowledge (e.g., Mitchell & Richman, 1980). Experiment 3 addresses these questions.

Briefly, then, the current research investigates the effects of loudness ratings of verbally described sounds on the time required to (a) generate single auditory images, (b) read the descriptions of the sounds, (c) mentally adjust the loudness of one image to equal that of another, and (d) identify the louder (or softer) of the two described sounds. In addition, the contributions of demand characteristics are assessed.

EXPERIMENT 1

The study of apparent loudness of named but unheard sounds poses questions in addition to those cited above. Apparent loudness is likely to be influenced by many factors, such as the actual intensity of the sounds themselves, their frequency, timbre, duration, and so forth. To rate the loudness of named but unheard sounds, the subjects must collapse these factors into a single judgment, perhaps by ignoring some factors, by averaging or integrating over others, or by more complicated combinatorial processes. Despite the complexity of the processes, however, our pilot work shows that subjects make consistent judgments about the loudness of these sound phrases.

As noted above, it seems reasonable to assume that the times to generate auditory images will be unrelated to the loudness ratings of the described sounds. Should loudness always be retrieved as an integral component of auditory images, however, the image generation times and the loudness ratings would covary.

Method

Materials. The basic materials were 30 phrases describing common environmental sounds. The phrases contained one to three words, as shown in Table 1. One group of subjects rated the loudness of each phrase; another group generated auditory images to each phrase.

The loudness ratings used a scale from 1 (very soft) to 7 (very loud). Four randomized orderings of the phrases were prepared. The loudness rating task and the auditory image generation task both presented the phrases twice to obtain stable estimates of the times to rate or to construct the images.

Table 1
Means and Standard Errors for Ratings of Loudness and Times to Generate Auditory Images

	Loudness Rating		Generation Time (in Milliseconds)	
	Mean	SE	Mean	SE
Whisper	1.23	.044	1187	158.543
Champagne Bubbling	1.39	.071	1161	415.434
Alka Seltzer	1.46	.060	995	411.052
Cat Purring	1.54	.065	948	371.838
Sigh	1.54	.094		
Bee Buzzing	1.96	.080	907	202.923
Faucet Dripping	1.96	.089		
Clock Ticking	2.23	.116	1823	278.978
Cricket Chirping	2.31	.093	1260	465.582
Wind Chimes Tinkling	2.38	.093	1066	232.184
Fire Crackling	2.66	.104	1044	290.298
Popcorn Popping	2.97	.082	1491	374.556
Typewriter Typing	3.49	.098	1248	174.575
Laughter	3.76	.103	1329	195.016
Surf Pounding	4.13	.121	825	251.405
Telephone Ringing	4.23	.118	1053	265.112
Vacuum Cleaner Cleaning	4.24	.114	1659	226.053
Clapping	4.50	.120	1276	232.614
Horn Honking	5.16	.107	760	217.962
Slamming a Door	5.33	.115	1490	168.598
Opera Singer Singing	5.38	.113	1004	294.834
Train	5.67	.118	897	194.541
Chain Saw	5.80	.106		
Firecrackers Going Off	5.80	.093		
Gun Firing	5.96	.117	1164	106.437
Crash	6.02	.097	1419	319.286
Police Siren	6.05	.092	995	292.994
Thunder	6.19	.094	1409	217.313
Volcano Erupting	6.46	.089	1349	308.617
Dynamite Exploding	6.83	.047	773	248.442

Four phrases were inadvertently omitted from the computerized listings, so image generation times were collected for 26 of the phrases. The orderings of the 26 phrases within the first and second presentation sets were randomized individually for each subject.

Design and Procedure. The rating tasks were administered to groups of 10 to 50 students. The students were instructed to read each phrase and to judge the phrase on the 7-point scale, taking as much time as they needed.

The subjects assigned to auditory image generation were tested individually. These students were seated at a console, then told that phrases would appear on the screen in front of them. Their task was to imagine hearing the sound. When their image was a clear and distinct as when they actually heard the sound, they were to press a button held in their preferred hand. The intertrial interval was approximately 15 sec. The subjects then were asked if they had any difficulty forming the images, which phrases produced difficulty, if any, and whether they had formed visual images. If they stated they had formed a visual image, they were asked to indicate on a list of phrases those phrases that also elicited a visual image and the order of elicitation of auditory and visual images.

Subjects. The subjects were introductory psychology students who participated as one way to fulfill partial requirements of the course. Ninety-four students rated the loudness of the phrases; 27 generated auditory images.

Results

The primary focus of interest is on the extent to which the loudness ratings predict the times to generate audi-

tory images. Loudness ratings did not reliably predict the times to generate auditory images ($r = -.13$, $p < .05$), consistent with the expectation that loudness information would not invariably influence image generation times. Further, this low association cannot be attributed to two alternative explanations. These explanations are that excessive variability camouflaged the association or that the measures themselves were unreliable. High variability would not be surprising in view of the subjective nature of the tasks. This explanation of the data was contradicted by reasonably small standard errors, as shown in Table 1. Table 1 lists the mean scores and standard errors for each of the measures for each sound phrase.

The question of the reliability of the measures was probed by correlating performance on the first rating or imaging of each phrase with the second performance. These coefficients, then, constitute test-retest estimates of reliability. The coefficients were high and statistically significant for both the loudness ratings (.93) and the image generation times (.92), indicating that the low-rating/image generation association could not be attributed to unreliability of the single measures.

The postexperimental query yielded an intriguing result. Although instructed to form auditory images, subjects claimed that they also generated visual images to 95% of the phrases. In contrast, other subjects who had been instructed to form visual images to the same sound phrases[1] stated that they also formed auditory images 53% of the time. This distribution differed reliably from chance [$\chi^2(1) = 107.80$, $p < .01$]. Subjects typically reported that their visual images preceded the auditory ones.

Discussion

The data uniformly indicate that the times to generate auditory images to named but unheard common environmental sounds do not vary systematically with the loudness ratings of the sounds. The most important implication of these results is that the information retrieved from memory during image generation does not necessarily include an index of auditory perceptual experience, the loudness ratings.

The data also suggest that auditory images may be related to, or even dependent on, visual images to a greater extent than the converse. The information retrieved from memory thus may include multimodality perceptual markers, not just those in the "primary" modality. This view assumes that both kinds of perceptual markers or information are separable, an assumption not explicitly tested by Experiment 1. Experiment 2 considers whether, under certain circumstances, loudness will influence image manipulation times.

EXPERIMENT 2

To establish a condition conducive to the retrieval and inclusion of loudness dimensions in auditory images, in Experiment 2 subjects were asked to equalize (match) the loudness of two auditory images. With the task, we expected the times to match the two images to increase with the distance between the loudness ratings of the named sounds. The sound phrases were selected to sample various distances between the loudness ratings and to compare approximately equidistant stimuli from different locations along the judged loudness continuum.

The times to read the pairs also were tested. This task should not instigate a retrieval search for loudness information. The reading times, then, should be independent of the loudness ratings.

Method

Materials. The materials represent the cells of a 3 (loudness ratings) by 3 (distances between loudness ratings) design, as illustrated in Table 2, plus additional pairs to extend the range of distances. For each of the three levels of loudness ratings (soft, medium, loud), pairs of sound phrases were selected to represent each of the three different ranges of distances (close, intermediate, far). The overall mean loudness ratings ranged from 2.53 for "soft" pairs to 5.38 for "loud" (on a 7-point loudness scale, with 7 as the maximum loudness score). The overall mean distances (defined as the differences between the loudness ratings of the two members of the pairs) increased from .37 for the "close" pairs to 2.98 for the "far" pairs. To extend the ranges of the distances, additional pairs were generated by selecting one member rated soft and one rated loud. These "very far" pairs are indicated at the bottom of Table 2. Table 2 summarizes these relationships and the basic design.

Two sets of pairs were constructed. Each set contained 28 practice pairs and 24 regular pairs. The sounds for each of the regular pairs in the first set were selected to correspond to

Table 2
Experimental Design, Mean Distances (D), Loudness Ratings (LR), and Exemplars of Stimulus Pairs: Experiments 2 and 3

Pair Type	D	Soft LR		Medium LR		Loud LR	
		Exemplar	LR	Exemplar	LR	Loud	LR
Close	.37	Wind Chimes Tinkling	2.30	Telephone Ringing	4.18	Slamming a Door	5.24
		Popcorn Popping	2.82	Clapping	4.50	Train	5.52
Intermediate	1.49	Bee Buzzing	1.75	Laughter	3.76	Clapping	4.50
		Typewriter Typing	3.23	Horn Honking	5.24	Police Siren	6.00
Far	2.98	Champagne Bubbling	1.31	Cricket Chirping	2.31	Vacuum Cleaner Cleaning	4.18
		Surf Pounding	3.81	Thunder	6.10	Dynamite Exploding	6.83
Mean Close-Far			2.53		4.35		5.38
Very Far	5.13	Whisper	1.36	Volcano Erupting	6.49		

Note—Mean distance = difference between the mean loudness ratings for the two members of each pair.

the cells of Table 2. Then another set of 24 regular pairs was chosen in the same way for the second set. All of the pairs were unique, although a given sound phrase appeared more than once. Practice sets were then prepared, representing all combinations of Table 2 and avoiding duplication of the regular pairs.

The orders of the pairs were randomized individually for each subject. Half of the pairs had the louder rated sound on the left; the other half had the louder rated sound on the right.

The materials were shown, one pair at a time, in a display apparatus. Both the presentation of a pair and its removal from view were subject-controlled.

Subjects. Twenty-four naive introductory psychology students were assigned randomly to each of the two sets of materials.

Procedure. The subject's task was to imagine hearing the left-hand sound and then the right-hand sound. Next the subject was to mentally manipulate the right-hand sound to be the same loudness as the left-hand sound. (S)he indicated when the match was achieved by pressing a lever that simultaneously stopped a digital timer and advanced the display to a blank space.

To obtain estimates of the time to read the phrases used in the main experiment, 12 additional students read the phrases from each set, pressing the lever as soon as they finished reading each phrase. No information about imagery or loudness matching was given.

Results

Matching times. If imagined sounds include retrieved loudness information, the times required to match the loudness of the two members of a pair should increase systematically with the distance between the loudness ratings of the sound phrases. And that is what happened. The matching times increased reliably by 163 msec from "close" to "very far" pairs [$F'(3,140) = 10.22$, MSe = 2.7164] (see dashed-triangle line in the lower panel of Figure 1). The linear trend was significant [$F'(1,140)$

$= 10.42$], but the deviations from the linear trends and higher order trends were not.

In addition to varying the distance between the rated loudness of the two members of a pair, pairs representing approximately the same distances were sampled from different locations along the underlying continuum. Another analysis considered these pairs, excluding the very far distance pairs which necessarily contained one sound rated soft and one rated loud. The mean matching times for the soft, medium, and loud pairs were 1,672, 1,658, and 1,680 msec, respectively ($F' < 1$). Thus, the mean loudness ratings of the pairs did not affect matching times, nor did the mean loudness ratings interact with other major variables.

The above analysis collapsed the position of the louder and the softer rated members of the pair. The next analysis considered whether the subjects had to increase or decrease the imagined loudness of the second member of the pair to match the imagined loudness of the first member of the pair, as they were instructed. The matching times to increase the imagined loudness of the second member of the pair were compared, on a pair-by-pair basis, with the matching times to decrease the second member of the pair. The time to match the louder to the softer member (mean = 1,709 msec) was almost the same as the time to match the loudness of the softer to the louder rated member of the pair (1,704 msec), suggesting that it was no more difficult to imagine reducing the loudness of an auditory image than to imagine increasing its loudness.

Reading times. The mean time to silently read the phrases was 1,151 msec. The reading times were not affected by the distance between the loudness ratings of the two sounds of a pair [$F'(2,170) = 2.98$] or the mean loudness ratings of the pairs [$F'(2,170) = 1.15$]. Hence, differences in matching times could not be attributed to differences in the times to read the phrases.

Discussion

Experiment 2 shows that the times to match the imagined loudness of the two named sounds increases with the difference between the loudness ratings of the sounds. The times are not affected by the absolute loudness ratings or by whether the second sound had to be increased or decreased to achieve the match.

These results, combined with those of Experiment 1, demonstrate the discretionary inclusion of loudness information in auditory images. Before examining the implications of these results in greater detail, it is important to reconsider the possibility that the subjects response to some demand characteristics in the situation. The possibility does not seem likely, in view of the postexperimental comments of our subjects, but we sought a more convincing test of this possibility.

EXPERIMENT 3

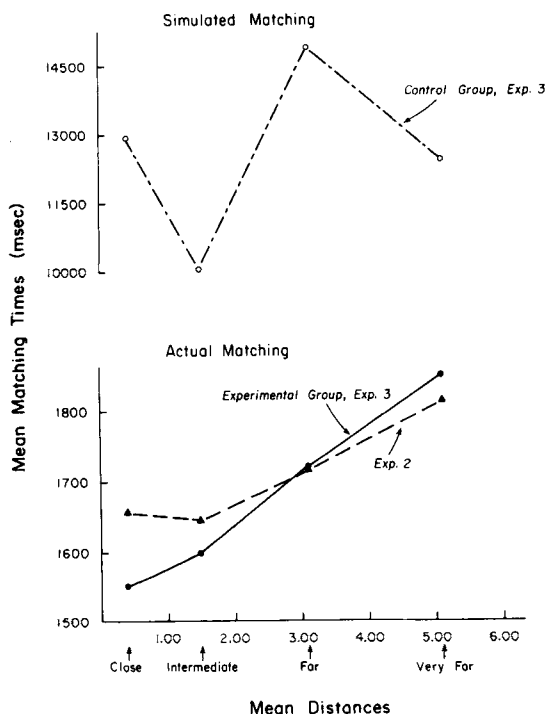
Figure 1. Time to match auditory images and to simulate the matching or equalizing of the loudness of sounds, Experiment 3.

In this experiment, the subjects had two tasks. The first was to predict the length of time it would take

them to match the sound of the pair, assuming that they could control intensity, frequency, and other parameters of the actual sounds. The second task was to perform in the imaginary task situation. The first task did not mention the use of imagery. It should, then, indicate how long the subjects thought it would take to achieve the matches, based on their general knowledge about the loudness of sounds in the environment and on reactions induced by characteristics of the experimental task. This task constituted what has been called a "nonexperiment" by Orne (1969).

One of the major difficulties with the use of "non-experimental" conditions to assess demand characteristics is that the technique does not contrast the subjects' estimates with their actual performances. To avoid this problem, our subjects' other task was to match the loudness of the imaginary sounds. This design provided both a comparison between the predicted and observed times to make loudness matches and a replication of Experiment 2.

Method

Materials. The basic materials were similar to those used in Experiment 2, so that in both the control and experimental conditions, the subjects saw practice pairs and 24 regular pairs of phrases. The ratings of loudness of the phrases were selected to represent the same three levels of loudness ratings (soft, medium, loud) and the levels of difference between the loudness ratings (close, intermediate, far, and very far) used in Experiment 2. One set of materials reversed the order of the two members of the pairs. Separate randomized orderings of the pairs were prepared for each of the two conditions for each subject.

Design and Subjects. Twenty-four naive introductory psychology students were assigned to six replications of the two orders of conditions (experimental first, control second, and vice versa) and the two sets (A and B).

Procedure. In the control conditions, subjects estimated the amount of time required to make loudness matches of the sounds described by the two phrases. The experimental condition replicated the procedure of Experiment 2.

Results

Experimental condition. As predicted, the time to match the loudness of the imagined sounds increased as the distance between the loudness ratings increased from close to very far pairs [$F'(3,140) = 8.82$, MSe = 5.2420]. The filled circles in the lower panel of Figure 1 plot the mean matching time for the close, intermediate, far, and very far pairs. This analysis collapsed the data over the loudness ratings. When the loudness ratings were considered in the analysis and the very far pairs were excluded (because each of these pairs contained one soft- and one loud-rated stimulus), the increase in matching times from close to far also was reliable [$F'(2,380) = 6.18$, MSe = 8.1688]. The loudness ratings did not yield a reliable main effect nor did the ratings interact with the other variables, replicating the results of Experiment 2.

Control condition. If the previous results reflect little more than general knowledge about perceiving environmental sounds, the results should be mimicked by the control condition. In fact, the times did not show significant increases as the distances increased [$F'(3,140) = 1.25$]. Moreover, neither the loudness ratings of the pairs ($F < 1$) nor any of the interactions was significant. Thus the results of the experimental condition and Experiment 2 cannot reasonably be attributed to the use of general world knowledge or to the demand characteristics of the situation.

The most striking aspect of the control data was the overestimation of the times needed to match loudness, as shown by the open circles in the upper panel of Figure 1. The estimated matching times were at least six times the actual matching times for the counterpart experimental conditions, providing additional evidence that contradicts a demand-characteristic explanation of the mental matching results.

Discussion

Experiments 1-3 focus on the effects of a likely perceptual attribute of actual auditory experience, loudness ratings, on the times to perform various operations with described sounds. The time to mentally match the loudness of the imagined sounds increases with increases in the differential between loudness ratings for the two phrases of the pair. The times to generate auditory images, to read the phrases, and to simulate loudness matching times are not systematically related to the loudness ratings. The phenomenon of loudness as an optionally included attribute of auditory images is real, robust, and reliable.

Basic to the contention that the differences in matching times are due to differential loudness retrieval is the assumption that the discriminability between two described sounds increases with loudness-rating differentials. A loud-rated sound is more discriminably different from a soft-rated sound than it is from a medium-rated sound. While this assumption predicts, accurately, that matching times increase with increasing loudness-rating differentials, it makes the contrary prediction for comparative judgments of the louder (or softer) of the two described sounds. In this case, the greater the discriminability (loudness-rating differential), the easier it is for the subject to distinguish between the two sounds. Hence, the greater the discriminability, the faster the judgment times should be. In other words, a symbolic distance effect is predicted. Experiment 4 tests this prediction.

EXPERIMENT 4

The inverse relationship between the distance separating the two members of the pair and the times to make judgments has been observed repeatedly with other types of stimulus dimensions (see Banks, 1977, and Moyer & Dumais, 1978, for reviews). If loudness ratings are used in the same manner as other dimensions, judgment times should decrease with increasing distances between the loudness ratings.

Symbolic distance studies have noted three other major effects. One is a semantic congruity effect, in which judgments are faster when the type of decision is compatible with the characteristics of the stimuli on the assumed underlying dimension (e.g., response times are faster to judge the smaller of two small stimuli than to judge the smaller of two large stimuli, and vice versa) (Banks, 1977; but see Marschark & Paivio, 1979).

Another effect, serial position, is shown by faster response times for representations of stimuli near the ends of the distributions than for stimuli near the middle (e.g., Banks, 1977; Trabasso & Riley, 1975) or by monotonic increases in response times as the position of the pair increases toward one extreme of the underlying dimension (e.g., Lovelace & Snodgrass, 1971).

A fourth effect, judgment preference, occurs when one type of judgment (e.g., identifying the louder of two stimuli) is faster than the other type of judgment (identifying the softer of the two stimuli), even though the same pairs are compared. Statistically, the judgment preference effect corresponds to a main effect for the type of judgment, whereas the semantic congruity effect corresponds to an interaction between the type of judgment and the positions of the stimuli on the underlying dimension. Judgment preference effects have been reported for various types of comparative judgments. For example, "large" digits are judged faster than "small" digits (Parkman, 1971); the brighter of two lights, faster than the dimmer (Audley & Wallis, 1964); arrows pointing up, faster than down (Clark & Brownell, 1976); and the higher of the pitches of two tones, faster than the lower of the pitches (Wallis & Audley, 1964).

If judgments of imaginary loudness parallel those based on other conceptual stimulus dimensions, all four effects should be shown in Experiment 4. This experiment used a modified "infinite" set approach. In the infinite set design, each stimulus pair is considered only once (see Banks, 1977). The extreme version of this approach was not practical, hence the design was modified by testing each pair of sounds no more than once.

Method

Materials. The phrases were the same as those used in Experiment 2. The orders of the pairs were randomized for each subject as part of a computer-generated listing. Half of the pairs had the louder rated phrase on the left side of the pair; the other half presented the louder phrase on the right.

The subject's sheet was mounted in a display apparatus that showed the two phrases of a pair in a viewing window. A digital timer was started by the display of a pair and stopped by the subject's depression of one of two buttons.

Subjects. Thirty-two additional students from introductory psychology classes participated. Eight students were randomly assigned to each of the four cells of the block created by two sets of materials and two type-of-judgment orders.

Procedure. The type of judgment variable was manipulated within subjects, using an ABBA design. Half of the subjects had the sequence judge louder, judge softer, judge softer, judge

louder for the four blocks of trials, respectively. The other half of the subjects had the BAAB sequence. On each trial, the subject indicated the phrase that was the louder (or softer) of the two by pressing the response button on the side of the louder (or softer) sound with the index finger of the preferred hand.

Results

Symbolic distance and judgment preference effects. We expected both symbolic distance and judgment preference effects. These expectations were supported, as shown in Figure 2. Showing the symbolic distance effect, the very far pairs were processed 467 msec faster than the close pairs [$F'(3,196) = 36.89$, MSe = .0712]. The judgment preference effect was numerically smaller but consistent. The louder of the two stimuli was identified 186 msec faster than the softer of the two stimuli, even though the same sounds were described [$F'(1,196) = 7.82$, MSe = .0712]. Errors were infrequent (range = 0 to 5.2%) and showed no speed-accuracy tradeoff. The times for error trials were estimated from the mean times for corresponding correct responses.

Semantic congruity effects and ratings of loudness. The basic assumption behind the congruity prediction is that comparisons will be faster when the type of judgment and the to-be-compared stimuli are compatibly coded (e.g., judging the louder of two sounds rated loud) than when they are not (e.g., judging the louder of two sounds rated soft). Figure 3 shows the predicted interaction between the loudness ratings and the type of judgment [$F'(2,476) = 22.12$, MSe = .1823]. The response times to judge the softer of the two stimuli were 277 msec faster when the stimuli were rated soft than when they were rated loud, whereas the response times for identifying the louder of the pair were 418 msec faster for the loud-rated than for the soft-rated items. The main effect for loudness ratings, however, was not reliable ($F' < 1$).

Serial position effects. Presumably, the characteristics of the underlying stimulus dimensions influence coding. When judging the louder of the phrases, the louder
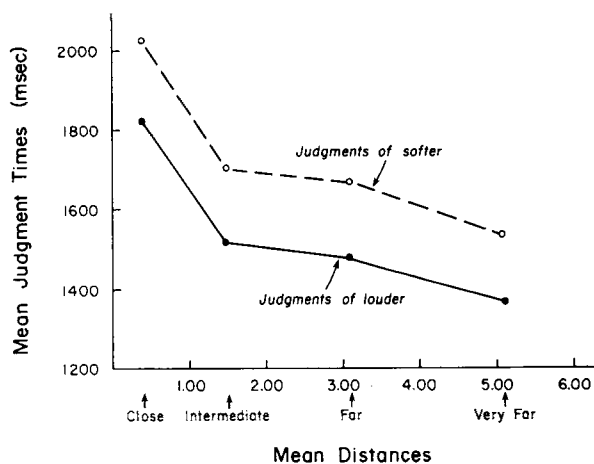

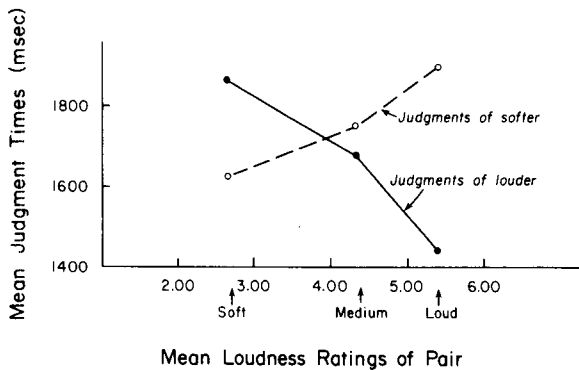Figure 2. Symbolic distance and preference effects, Experiment 4.

Figure 3. Congruity effects, Experiment 4.

rated items are more likely to be discriminably coded than the softer-rated items. Hence, the times to judge the louder should increase as the ratings move toward the soft end of the continuum. The opposite effect would hold for judgments of the softer stimulus. Serial position functions calculated from the close pairs conformed to the predictions.

Discussion

Extending the previous findings, Experiment 4 demonstrates standard symbolic distance, congruity, and serial position effects with named but unheard sounds as the comparison stimuli. Further, judging the louder of the two described sounds is faster than judging the softer, even though the same sounds are compared. In general, the data demonstrate that the greater the differential between the stimuli's loudness ratings, the more readily the subjects make comparative judgments.

GENERAL DISCUSSION

The results indicate that loudness information is an optional component of auditory imagery. Memory for auditory experiences may include correlates of auditory sensory/perceptual experience, such as loudness, when appropriate for the experimental task. This conclusion stems from two sets of converging results: (1) differences between the loudness ratings affect the times to match and to make comparative judgments of described but unheard sounds; (2) loudness ratings do not affect the times to generate auditory images or to read the sound phrases.

The data also showed faster identification of the louder than of the softer of two sounds even though the same pairs were compared (Experiment 4). The "absolute" loudness ratings did not yield a corresponding effect. Neither the matching nor the comparative judgment times differed for the louder rated pairs, when these ratings were considered as a main effect. Hence, instructions to process the louder information produced a bias toward faster processing of "louder" than "softer," but the "absolute" mean loudness ratings of the pairs did not induce the same bias.

Before considering the implications of the findings,

it is important to note that these reliable, robust effects cannot be dismissed as epiphenomenal. They cannot be explained as products of the demand characteristics associated with the experimental setting or with the subjects' knowledge of themselves and the outside world (Experiment 3).

As described in the introduction, one approach treats the memorial representations of sounds as groups of attributes or features, some of which reflect perceptual characteristics. This view is similar to those of Hasher and Zacks (1979), Rosch (1978), Tulving and Watkins (1975), and Underwood (1977). While some attributes may typically have higher cue validities or weights than others, and thus be more often associated with a concept or category, "louder" and "softer" cues do not appear to be differentially weighted. Hence, this perceptual information is available, but the likelihood of retrieving representations associated with louder sounds is no greater than the likelihood of retrieving representations associated with softer sounds, unless the search itself targets a particular kind of information.

Indeed, the likelihood of activating certain features depends on the retrieval strategy, a strategy determined by the task demands. Tasks requiring decisions about loudness induce search strategies for the targeted kind of loudness information, whereas tasks not requiring loudness judgments would not. Once retrieved, the features are compared, matched, or subjected to other manipulations as indicated by the task. Lastly, a response is executed.

How does this view explain the matching and symbolic distance effects? Since both tasks are couched in terms of matching or comparing the loudness of the imaginary sounds, the search strategy focuses on the retrieval of loudness information. Pairs with different or discrepant loudness features will be subjected to decision and response stages quickly, whereas more information must be sought to distinguish between pairs with similar loudness codes or features. In effect, the more distinguishable pairs are compared faster than more similar ones, producing the symbolic distance effect. The more discrepant pairs, however, will require more processing to match characteristics than will similar ones, so that the same basic principles cover both the matching and the symbolic distance effects.

The congruity effect is produced in the following manner. The instructions (or expectations) specify the type of judgment. This information guides the retrieval search. When stimuli representing targeted information are presented (say, loud-rated stimuli under instructions to identify the louder sound), no additional search is necessary before entering the judgment-decision stage. When stimuli representing nontargeted loudness information are shown (e.g., two soft-rated stimuli under instructions to identify the louder), additional searches must occur before the subject can make a judgment, thereby increasing response times.

This view is compatible with two interpretations of

the preference effect. The first is that the preference effect represents a bias toward initiating the search at one end of dimensioned characteristics, the dominant (ecologically valid?) end. The loud end would have to be assumed to be dominant. If this interpretation is correct, a combination of instructions and stimuli with ratings that coincide with the dominant end (i.e., judge the louder of two loud-rated stimuli) should yield faster comparative judgments than other conditions. The data offer some support for this prediction (see Figure 3).

An alternative explanation of the preference effect stems from decision criteria. Because, in perception, more intense sounds are heard more readily than less intense ones, it is reasonable to propose that decisions about representations of louder sounds are based on a more lenient criterion and are processed faster than decisions about representations of softer sounds. Though not explicitly addressed to this interpretation, the current data are in accord with it.

When the search strategy is not biased by the instructions, as presumably is the case with the image generation and reading tasks, the chances of retrieving items will correspond to a probabilistic sampling of the features included in the concept. If this notion is correct, it suggests that the overall cue validities associated with loudness features do not dominate the features in the representational bundle. Supporting this interpretation are the results cited above and the reports of the subjects.

This tentative perspective is, of course, unsatisfactorily sketchy. It does, however, accord reasonably well with the data. Recapitulating, the view treats the representations of sounds as bundles of features, some of which capture sensory/perceptual information. Features representing loudness are among those accessible. The search for these representations is guided by the demands of the task, thus permitting the optional inclusion of loudness information. When targeted for loudness information by the task demands, the search is biased toward loudness features, and the discrepancies between the loudness characteristics of the sound phrases influence performance.

Other models often applied to comparative judgments could be extended to cover some, but not all, of the current results. These models have been reviewed recently (Banks, 1977; Moyer & Dumais, 1978), so they will be discussed briefly. The models are directed toward the symbolic distance, congruity, and serial position effects, with little attention paid to preference effects.

The models fall into two main classifications, discrete and continuous (analog). Paivio's (1978) dual-code model combines elements of both classifications. The discrete models assume that discrete, semantic codes are generated for the stimuli and the instructions (e.g., Banks, 1977; Trabasso & Riley, 1975). According to Banks' (1977) three-stage model, the initial, encoding stage used information in the data base (memory) to associate discrete codes with each stimulus. The second,

choice state processes the stimulus codes until they are distinguishable and then matches the resulting codes with those of the instructions. The third stage produces the response. The symbolic distance effect arises from the longer time needed to generate distinguishable codings for stimuli that are initially assigned similar codes (e.g., "loud" and "loud +"). These items require more time to differentiate than do items with initially different codes (e.g., "loud" and "soft"). The congruity effect occurs because stimuli with codes that match the instruction code (choose the louder of two loud-rated stimuli) will be processed faster than stimuli with codes that do not match the instructions code (choose the softer of two loud-rated) stimuli. Predictions about the serial position effect follow from the considerations given above. These models provide a reasonable account of a substantial literature, as Moyer and Dumais (1978) note, although there are some ambiguities about the data base, in particular.

Analog models (e.g., Holyoak, 1978; Jamieson & Petrusic, 1975; Marks, 1972; Moyer & Dumais, 1978) and the imaginal portion of Paivio's (1978) dual-code model postulate that memorial representations are coded in a continuous, analog manner. The analog representations of the two stimuli are scanned and compared. According to the Moyer & Dumais (1978) model, the difference between the two analog representations is tested against a criterion. If the difference exceeds the criterion, as is likely to occur when distant stimuli are presented, a response is made. If the difference does not exceed the criterion, it is added to an internal accumulator of some kind, and the process is repeated until a response can be made. Because the frequency of repetition of this iterative process increases as the difference between the stimuli decreases, the symbolic distance effect is predicted. The congruity effect occurs because instructions signal the starting point of the memorial search (Moyer & Dumais, 1978) or because the instructions induce expectancies for similarly coded stimuli (Marschark & Paivio, 1979).

As noted previously, all of these models accommodate the symbolic distance and congruity effects, such as those shown in Experiment 4. Moreover, they could be extended to cover the matching results of Experiments 2 and 3. Rather than recoding the representations to achieve different codes, as in comparative judgment tasks, the matching task requires recoding to reconcile the representational codes. It follows that, regardless of whether the representations are coded discretely or continuously, more time would be required to achieve coding matches for discrepant stimuli than for similar ones.

Summarizing the above arguments, all of the models predict comparative judgment and matching data. These results, then, do not distinguish between the models. The comparative judgment models encounter some difficulty with other results, however. Chief among these difficulties are the different effects of

loudness ratings in various experimental tasks. There is no obvious reason why characteristics of the coded representations would be influential in some situations, such as matching and comparative judgment, and not in others, such as the image generation task, particularly when all of the situations rely on memorial representations. Hence, the lack of sensitivity to loudness ratings shown by the image generation, reading, and even the main effects for loudness ratings in the matching and comparative judgment tasks seem inconsistent with the various comparative judgment models.

Another problem is posed by the tendency to generate visual images spontaneously when instructed to construct auditory ones. Although this tendency was not predicted by any of the models, it seems more compatible with our treatment of images as bundles of features retrieved from long-term memory than with any of the other views. Note that the data suggest that the retrieval of the visual components is not optional, in contrast to the retrieval of loudness information.

A third problem for models that assume the generation of linguistic codes is the failure to show a processing advantage of loud-rated over soft-rated pairs of stimuli. The loudness ratings, like the preference effect, would be coded semantically. Clark (1969) suggested that lexically unmarked adjectives, such as "loud," are stored in memory in a simpler form than lexically marked adjectives, such as "soft." The unmarked adjectives are stored and retrieved faster than are unmarked ones. This perspective predicts both the preference effect and an advantage of loud-rated over soft-rated pairs. Only the former was manifested.

The foregoing discussion indicated that models traditionally used to explain comparative judgments account for some, but not all, of the obtained data. Our alternative view seems to fare better in this regard, although the framework is still tentative and incomplete.

In conclusion, the major implication of these results is that imaginal representations are closely related to perceptual experience (e.g., Neisser, 1972; Shepard & Podgorny, 1978). Some perceptual features, such as loudness features, are optionally included in the representations, in accord with task demands.

REFERENCES

AUDLEY, R. J., & WALLIS, C. P. Response instructions and the speed of relative judgments: I. Some experiments on brightness discrimination. *British Journal of Psychology*, 1964, 55, 59-73.

BANKS, W. P. Encoding and processing of semantic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 11). New York: Academic Press, 1977.

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, 76, 187-404.

CLARK, H. H., & BROWNELL, H. H. Position, direction, and their

perceptual integrality. *Perception & Psychophysics*, 1976, 19, 328-334.

HASHER, L., & ZACKS, R. T. Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 1979, 108, 356-388.

HOLYOAK, K. J. Comparative judgments with numerical reference points. *Cognitive Psychology*, 1978, 10, 203-243.

JAMIESON, D. G., & PETRUSIC, W. M. Relational judgments with remembered stimuli. *Perception & Psychophysics*, 1975, 18, 373-378.

LOVELACE, E. A., & SNODGRASS, R. D. Decision times for alphabetic order of letter pairs. *Journal of Experimental Psychology*, 1971, 88, 258-264.

MARKS, D. F. Relative judgments: A phenomenon and a theory. *Perception & Psychophysics*, 1972, 11, 156-160.

MARSCHARK, M., & PAIVIO, A. Semantic congruity and lexical marking in symbolic comparisons: An expectancy hypothesis. *Memory & Cognition*, 1979, 7, 175-184.

MITCHELL, D. B., & RICHMAN, C. L. Confirmed reservations: Mental travel. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, 6, 48-66.

MOYER, R. S., & DUMAIS, S. T. Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12). New York: Academic Press, 1978.

NEISSER, U. Changing conceptions of imagery. In P. W. Sheehan (Ed.), *The function and nature of imagery*. New York: Academic Press, 1972.

ORNE, M. T. Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research*. New York: Academic Press, 1969.

PAIVIO, A. Mental comparisons involving abstract attributes. *Memory & Cognition*, 1978, 6, 199-208.

PARKMAN, J. M. Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, 1971, 91, 191-205.

ROSCH, E. Principles of categorization. In E. Rosch & R. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, N.J: Erlbaum, 1978.

SHEPARD, R. N., & PODGORNY, P. Cognitive processes that resemble perceptual processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes*. Hillsdale, N.J: Erlbaum, 1978.

TRABASSO, T., & RILEY, C. A. The construction and use of representations involving linear order. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, N.J: Erlbaum, 1975.

TULVING, E., & WATKINS, M. J. Structure of memory traces. *Psychological Review*, 1975, 82, 261-275.

UNDERWOOD, B. J. *Temporal codes for memories: Issues and problems*. Hillsdale, N.J: Erlbaum, 1977.

WALLIS, C. P., & AUDLEY, R. J. Response instructions and the speed of relative judgments: II. Pitch discrimination. *British Journal of Psychology*, 1964, 55, 121-132.

NOTE

1. In addition, subjects rated the ease of imagining the auditory or the visual referents of each of the sound phrases, or they were timed while generating visual images. Information about these results may be obtained from the author.