THE IMPORTANCE OF INTERVAL INFORMATION IN LONG-TERM MEMORY FOR MELODIES

W. Jay Dowling and James C. Bartlett
The University of Texas at Dallas

Abstract

Four experiments examined the roles of melodic contour and pitch interval information in recognition memory for melodies. In Experiments 1 and 2, subjects heard excerpts from Beethoven String Quartets, and subsequently attempted to detect copies of input melodies (Targets) as well as Related items, which resembled input items with respect to contour and rhythm, but not pitch intervals. Targets were recognized significantly better than Relateds, which were recognized only slightly more often than lures (Lures were drawn from other quartet movements not represented on the input list). Experiment 3 replicated this finding with novel, randomly generated melodies. All three experiments supported substantial retention of information regarding pitch intervals, as well as contour, over a retention interval of several minutes. Using the materials of Experiment 3 in a short-term memory paradigm, Experiment 4 examined short (5 sec) and long (31 sec) retention interval conditions. Contour information dominated performance with the short delay, but not with the long delay, where performance resembled that of the prior long-term memory experiments. While interval information is difficult to encode, it is apparently retained with high efficiency in long-term memory.

When a melody is transposed from one key to another, the pitches of notes are changed, while other perceptual properties remain invariant. In particular, both the melodic contour (the pattern of ups and downs from note to note) and the pitch intervals along a logarithmic scale of frequency remain constant through transposition. In recognizing melodies, listeners can use both types of information, but the relative importance of one versus the other appears to change with the familiarity of the stimuli. With well-learned melodies, interval and/or chroma information plays a dominant role. For example, people are quite good at discriminating between transpositions of well-known songs and lures that have the same contour but different intervals (Bartlett & Dowling, 1980). Further, if a well-known melody is transformed in a way that preserves its contour but destroys the precise intervals between notes, it becomes difficult to recognize (Dowling & Fujitani, 1971; Idson & Massaro, 1978). When the listener has no information about the population of songs being tested, a contour cue alone seems virtually useless (Kallman & Massaro, 1979). In contrast to research with well-learned melodies, research with novel melodies indicates that contour plays a dominant role while information about precise interval sizes contributes little if anything to performance. For example, Dowling (1978) using a short-term memory paradigm demonstrated chance discrimination between transpositions of novel melodies and "tonal answers", melodies sharing contour and key with the original, but not containing the same sequence of intervals. But listeners in this situation are quite good at rejecting melodies whose contour differentiates them from an originally presented stimulus (Bartlett & Dowling, 1980). Thus the data indicate that the initial memory representation of a novel melody contains accurate information about contour and (probably key), but not intervals.

This last result along with the practices of some composers led us to a question about the cognition of actual pieces of music. A structural arrangement typical of (but not restricted to) western music of the past 750 years is that a melodic theme will appear several times in a piece. In its several appearances, the theme might undergo various transformations of pitch level and interval sizes among its notes while retaining its contour. Our question was whether a listener can "pick up" these contour repetitions when hearing a piece for the first time. Such an ability might play a large role in the listener's initial perception of the work, facilitating the detection of its organization and those invariant features that distinguish the work from others by the same or different composers.

In the first two experiments reported here we aimed at a greater degree of ecological validity than in our previous work by using excerpts from Beethoven's String Quartets as the to-be-remembered stimulus material. The Quartets have several desirable properties. First, they constitute an extensive body of material produced by the same composer using the same instrumentation and recorded by the same musicians under the same conditions. Second, though some pieces by Beethoven were known to almost all our subjects, the Quartets were all but completely unfamiliar. Third, Beethoven often wrote in the style just alluded to, in which a relatively brief melodic theme is repeated over and over again throughout a piece, appearing at various pitch levels with various changes in interval sizes. (The most famous example of Beethoven's use of this approach is the opening movement of his Fifth Symphony). In these experiments, listeners first heard an input list containing brief excerpts from such passages. They then took a recognition test consisting of Targets, copies of previously heard excerpts; Related Items, excerpts of the original compositions imitating input items in contour but differing in interval sizes; and Lures, excerpts from movements of the Quartets other than those used in the input list. Figure 1 shows examples of excerpts used in Experiments 1 and 2. Note that each excerpt contains exactly one iteration of a thematic "chunk"—a brief melodic phrase functioning as a unit in the overall structure of the piece and containing within itself little or no internal repetition. We asked listeners to respond positively to Targets and Relateds, and to reject Lures.

In these experiments we examined listeners' discrimination between Targets and Lures, and between Relateds and Lures. We expected both types of discrimination to be better than chance, with little difference between them. Such a result would indicate that listeners can detect the repetition of same-contour chunks while ignoring differences of interval size. This would lay the groundwork for future investigations of how such repetitions might influence the listener's perceptual organization of a piece of music. We were led to these expectations by a series of pilot studies (Dowling & Bartlett, Note 1) in which input-list excerpts drawn from pieces built of short-chunk themes tended to produce positive recognition responses to both identical repetitions and thematically related excerpts from the same pieces. In experiments 1 and 2 we exerted greater control over the stimuli by making each stimulus contain exactly one iteration of a chunk and by making the similarity of Targets and Relateds depend on melodic contour. (Other results we found in the pilot work were: (1) Increasing temporal context of list items from 5 sec to 30 sec of the piece had little effect on recognition. (2) Both high school and college students were able to listen to a list of 20 brief excerpts in an unfamiliar musical genre and perform at better than chance accuracy on a recognition test 10 min later. (3) When lures were drawn from a variety of composers, the false-alarm rates of even non-musicians followed generalization gradients predictable from musicology. For example, when Targets were from a Haydn symphony, we obtained a declining series of false-positive recognitions to lures from the symphonies of Haydn, Mozart, Beethoven and Schubert.)

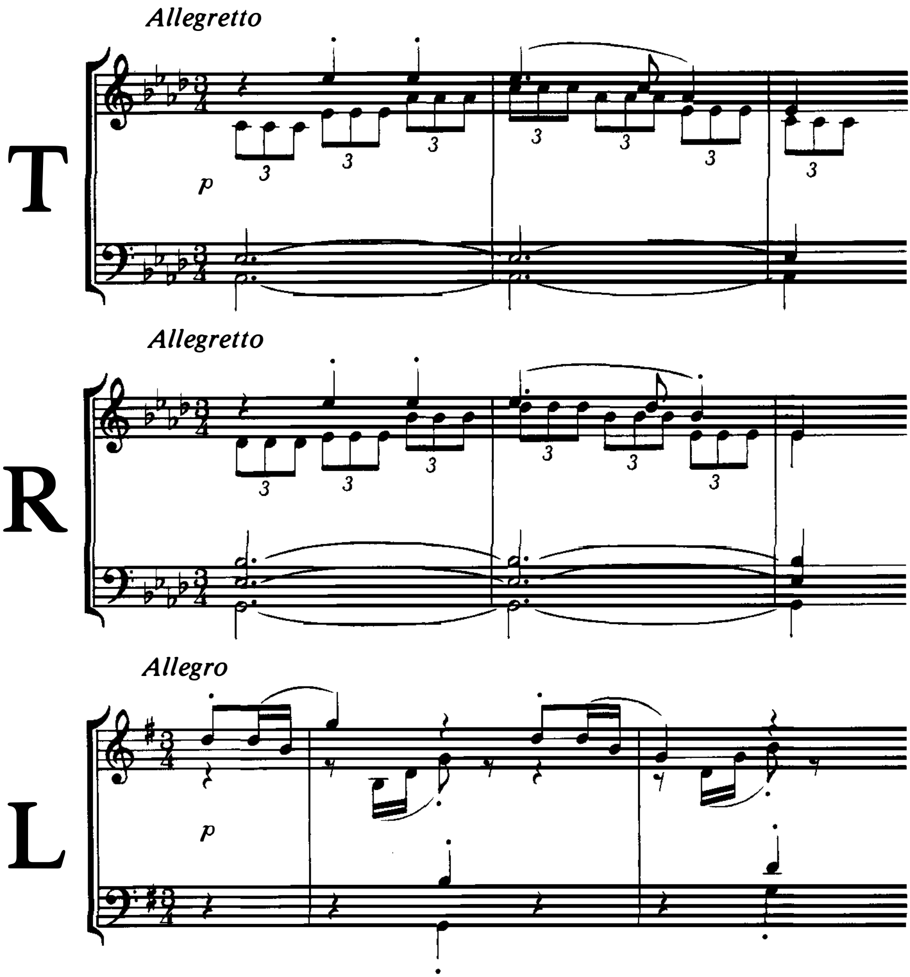

Figure 1. Examples of Target (T), Related (R), and Lure (L) stimuli used in Experiments 1 and 2. Each stimulus contains one melodic "chunk". T and R are from Beethoven's Quartet, Op. 18, No. 4, Trio in Minuet; L is from Op. 18, No. 2, Scherzo. Boldface notation indicates the melody selected as a "chunk".

Frankly, we were surprised at the results of Experiments 1 and 2 of the present report. Listeners showed much less discrimination between Relateds and Lures than we had anticipated. Therefore, we performed Experiment 3, which was an attempt to show that the phenomena we observed with the more naturalistic musical stimuli of Experiments 1 and 2 could also be obtained with the simpler, more artificial materials characteristic of our earlier work. Experiment 4 retained the use of artificial materials, and explored one possible cause of our unexpected results. This final experiment focused on the forgetting of contour versus interval information in a paradigm that contrasted long- and short-term memory.

Experiment 1

In addition to exploring listeners' ability to recognize same-contour imitations of previously heard melodic chunks, Experiment 1 also explored the possibility that the kind of encoding task listeners do when first hearing the stimuli might affect how they are stored in memory and the sorts of confusions the listeners would subsequently make. In particular, we expected that listeners who did an encoding task that called attention to the melodic contour might recognize the contour better later, getting good scores for both Targets and Relateds. Therefore we had one group of listeners respond to each input item, choosing which of two visual contour patterns best described it. (Dowling, 1972, had used a similar encoding task, and it enhanced recognition of melodic transformations at a level that approached significance.) Other groups performed either a rhythmic-pattern choice task or a stimulus-duration estimation task, neither of which encoded the contour. Of these last two tasks, we expected the rhythm task to produce better performance since the rhythmic patterns coded stimulus information common to both the Target and Related items of a pair, while the duration estimation task encoded little that was unique to any set of stimuli.

METHOD

Design

Experiment 1 had the design: 3 Encoding Tasks × 2 Test Item Types × 2 Experience Levels × 2 Test Blocks. Three separate groups of 16 subjects each performed one of the three encoding tasks. The encoding tasks required subjects to respond to each input item as it occurred, categorizing it according to contour, rhythm, or duration. Subjects in the contour condition chose between two up-and-down patterns of dots the one that best represented the contour of the excerpt. Subjects responded by marking an X next to the pattern they chose on an answer sheet containing all 18 items. Subjects in the rhythm condition chose between pairs of Morse-code type dot patterns the one that best represented the rhythm of the excerpt. Subjects in the duration condition wrote down their estimates of the duration of each excerpt in seconds. The two sets of scores for different item types were derived from the comparison of recognition rates for Targets with false-alarm rates to Lures, and of recognition rates for Relateds with false-alarm rates to Lures. These scores were derived by calculating the area under the Memory Operating Characteristic (MOC). There were two test blocks of 24 items each. The order of test blocks was reversed for half the subjects in each encoding condition. Each test block consisted of eight Targets, eight Relateds, and eight Lures. List items tested with Targets in Block 1 were tested with Relateds in Block 2, and vice versa. Lures were different in the two blocks.

Subjects

Forty-eight high school females ages 16 and 17 yr were each paid $2.00 for participation. Subjects had a mean of 4.7 yr musical experience, defined as lessons on an instrument or voice, or playing in an ensemble. We used a criterion of at least 2 yr training to define "experienced". Subjects were assigned blindly to encoding-task conditions. Of the 28 inexperienced and 20 experienced subjects, no fewer than six fell in any of the three conditions. Subjects served in group sessions, and one-third of the subjects in each session were assigned to each encoding condition.

Stimuli

There were 20 items on the input list, of which the first and last two were buffers not otherwise used in the experiment. All stimuli were drawn from Beethoven's String Quartets (recorded by the Guarneri Quartet, RCA albums VCS-6195, -6415, and -6418). Each input item contained exactly one chunk, as defined above, and items varied in length from approximately 2 sec to 8 sec. An ISI of 10 sec separated each item from the next. In selecting the items we first made a list of all the possible items we could find by searching the musical scores of the Quartets. Our search was guided by the necessity of finding Target-Related pairs; that is, pairs in which each member contained the same melodic contour but with different interval sizes. Targets were always the first member of each pair to appear in the piece; Relateds always the second. Twenty of the possible Target stimuli were randomly chosen as list items. Of these 20 the middle 16 were tested. Each was tested twice: once with a Target identical to the list item, and once with a Related stimulus. Eight list items (randomly determined) were tested as Targets in test Block 1 and as Relateds in Block 2. The reverse pattern of testing held for the other eight list items. Lures in the test blocks were randomly selected equally often from potential Targets and Relateds in the initial pool, and if one member of a Target-Related pair appeared as a Lure, its mate did not appear at all. Stimuli were recorded on tape and presented via high quality stereo equipment over loudspeakers at comfortable levels. Each stimulus was bounded by smooth onset and offset ramps approximately 1 sec long.

Procedure

We instructed subjects that they were about to hear a series of brief musical excerpts, and that their task was to listen to them closely so that they would be able to recognize them if they heard them again later. We assigned one-third of the subjects in each session to each of the three encoding tasks. Each subject read silently the instructions for her encoding task. We then presented sample stimuli for which the correct encoding response was already entered on the encoding answer sheet. We then presented the input list, and subjects wrote their encoding responses on the answer sheet provided. Following the input list the encoding responses were concealed for the rest of the experiment.

Between the input list and the test, subjects listened to 5 min of "Ruby, My Dear" by John Coltrane and Thelonious Monk (Milestone M-47011) and rated it for pleasantness on a 10-point scale. Prior to the test we explained the relationship of Target and Related items, that they both came from the same piece and had very similar melodies. We instructed the subjects to respond on a four-level confidence scale with the degree to which each test item reminded them of the list items they had heard ("Strongly reminds", "reminds", "unfamiliar", "very unfamiliar"), and that a given item might in fact strongly remind them of a list item and still be different from it in detail. We also asked subjects to try to make this last distinction by putting a check next to those confidence-level responses they thought pertained to exact copies of list items (Targets as opposed to Relateds). Each test item was followed by an 8-sec response period, and the onset of the succeeding trial was announced by the trial number followed by 2 sec of silence. Subjects responded to all 48 items on a single sheet of paper. Finally, subjects completed a brief questionnaire concerning musical experience.

Data Analysis

For each individual subject we calculated the areas under the MOCs generated by taking responses at each of the confidence levels to Targets or Relateds as hit rates, vs. responses to Lures as false-alarm rates. This gave a *Target vs. Lure and a Related vs. Lure area score for each subject on each* test block, which we fed into the analysis of variance. We took area under the MOC as an estimate of the unbiased proportion of correct responses where chance would be .50 (Swets, 1973). We also converted both "check" and "strongly reminds" responses for Targets (and Relateds) vs. Lures to d' scores for each subject. If subjects had been able to respond flexibly, using the checks to denote just those test items that were identical to input items, a different pattern should have emerged for the two sets of scores. However, discrimination between the Target vs. Lure score and the Related vs. Lure score was just about the same for both response modes, indicating that subjects were incapable of this particular type of flexibility. Both d' analyses showed essentially the same results as the area scores. In the following data analyses we will refer only to the area under the MOC.

Results

We performed analysis of variance on the 3 Encoding Tasks × 2 Item Types × 2 Experience Levels × 2 Blocks design. Table 1 gives a summary of the data.

Table 1

Areas Under the MOC for Experiment 1

Encoding task:	Contour	Rhythm	Duration	Mean
Target	.79	.78	.70	.76
Related	.55	.56	.52	.54
Mean	.67	.66	.61	.65

Only the effect of item type was significant, $F(1,42) = 235.20, p < .001$, with Targets better recognized than Relateds. The difference between Targets and Relateds was substantial, while Relateds were recognized at only slightly better than chance. This latter difference, though small, was significant, $t\ (47) = 2.82, p < .01$.

Performance on the encoding tasks was not very accurate. Inexperienced subjects got 49% correct on the contour task (about chance); while experienced subjects got 58% correct. Both groups performed better on the rhythm task, achieving about 63% correct. We did not score the duration encoding task.

Discussion

Contrary to our expectations, recognition of Targets was much better than recognition of Relateds, and recognition of Relateds was barely better than chance. Target recognition was better than in our pilot studies, perhaps due to the combination of explicit encoding and more coherent stimuli (in that each item contained exactly one melodic chunk, with no superfluous material before or after). But whatever aided subjects in recognizing Targets did not lead to good memory for the more general features that Targets and Relateds have in common, such as the melodic contour and rhythmic pattern, instrumental texture, tempo and meter.

Contrary to expectations, encoding task had no reliable effect upon recognition performance. While there was a trend for target recognition to be higher after contour and rhythm encoding than after duration encoding, there was virtually no tendency for encoding task to influence related-item recognition. These results invite the hypothesis that recognition of related items is always low, regardless of encoding activity at input. However, there are at least two alternative interpretations of the results. First, performance on the separate contour and rhythm tasks was far more difficult than expected, and their difficulty perhaps militated against their effectiveness for encoding.

A second interpretation for the absence of encoding effects might be that all three encoding tasks of Experiment 1 focus attention on physical, "surface" features of the stimuli, as opposed to affective or imaginal responses of the listeners themselves. It is obvious that string quartets can stimulate such responses in a listener (see Brown, 1980), and there is evidence that nonmusical imagery may facilitate memory for musical selections (Delis, Fleer & Kerr, 1978). Since a target item and its related-item mate come from adjacent spots in an original piece, and since they are obviously similar in several respects (e.g., contour, rhythm), it is plausible that they might stimulate the same or similar images or affective reactions. If so, an encoding task which favors attention to such images or affective reactions might facilitate subsequent recognition of related as well as target items.

Reflecting on the results of Experiment 1 led us to believe that there were two ways in which we could make our encoding tasks more effective. First, performance on the separate contour and rhythm tasks was far more than difficult than expected. We therefore decided to combine the features of the contour and rhythm tasks into one task for Experiment 2, a move which in fact led to much better encoding performance. Second, we thought that the contour and rhythm tasks were leading the subjects to encode very specific information about Targets at the expense of general information common to both the Target and Related stimuli in a pair. We thought that perhaps the more general meanings of these musical excerpts might lie in something deeper than the surface texture of the music—in some more affective response subjects could have. Such affective responses might relate in a vague way to verbal characterizations. There was some previous evidence that evoking such affective responses improved memory for music. Delis, Fleer and Kerr (1978) found that vivid verbal images associated musical excerpts led to better recognition of the music. And Brown (1980) found that subjects could agree on which excerpts out of a large selection should be paired as representing similar underlying meanings. This last result suggested to us that since the Target and Related in a pair came from adjacent spots in the original piece, they very likely shared the same sorts of affective meanings. Therefore, evoking those meanings on the initial presentation of the Target item should lead to better generalization to the Related item later.

Experiment 2

In Experiment 2 we combined the contour and rhythm encoding tasks into one and added two new tasks designed to evoke associations with underlying affective meanings. In one of these subjects chose between two words the one most similar to the meaning or affective quality of the piece; for example, between "bubbly" and "morose" for the excerpt from Op. 18, No. 4 in Figure 1. In the other task subjects were asked to write down in a few words an image or feeling that the excerpt brought to mind, avoiding technical musical descriptions of the physical stimulus (such as "pizzicato" or "fast").

Method

Experiment 2 differed from Experiment 1 only in the encoding tasks used in the subject population. Fifty-three students at the University of Texas at Dallas participated for course credit. Over a long series of experiments these subjects have tended toward a mean age of 29.4 yr and a mean of 5.3 yr musical experience. Experience was dichotomized at 2 yr as before. Separate groups of subjects in separate sessions were assigned blindly to the three task conditions, with 15, 13 and 25 subjects in the contour-rhythm, verbal checklist, and affective image conditions respectively. The contour-rhythm task combined the contour and rhythm tasks from Experiment 1 so that each pattern showed both the ups and downs of the melody as well as the relative durations of the notes. In the verbal checklist task subjects chose which of two words best fitted the stimulus. In the affective image task subjects wrote a brief descriptive phrase that related to the meanings of the piece or an image that it evoked, avoiding physical descriptions. In all of these tasks subjects made all their encoding responses on a single sheet of paper which they concealed during the test phase of the session. Input presentation was slowed to an ISI of about 20 sec for the affective image condition to give all subjects time to write their descriptions.

Results

Table 2 gives mean areas under the MOC for the various conditions of Experiment 2. We performed a 3 Encoding Tasks × 2 Item Types × 2 Blocks × 2 Experience Levels analysis of variance. The effect of item type was significant $F(1,47) = 432.54, p < .001$, with Targets better recognized than Relateds. The interaction of Item Type × Encoding Task was significant, $F(2, 47) = 6.17, p < .01$, with Targets best recognized with contour-rhythm encoding and worst with affective image. And the interaction of Item Type × Block was significant, $F(1, 47) = 7.39, p < .01$. Related recognition was lower in block 1 (.50) than in block 2 (.55), while Target recognition showed the opposite trend (.77 vs. .72). No other effects were significant at the .05 level. Related recognition was again significantly better than chance overall, $t(52) = 2.50, p < .05$. As in Experiment 1 we looked for evidence of flexibility in subjects' restriction of "check" responses to Target items only, but found none. The d' analyses of check and confidence-level responses both parallelled the area-score analyses very closely.

Table 2

Areas Under the MOC for Experiment 2

Encoding task	Contour-Rhythm	Verbal	Image	Mean
Target	.80	.76	.71	.76
Related	.53	.53	.53	.53
Mean	.66	.65	.62	.64

Performance on the encoding tasks was better than in Experiment 1. Both experienced and inexperienced subjects achieved about 67% correct on the contour-rhythm task. And though, strictly speaking, there were no "correct" answers to the verbal checklist items, subjects at both experience levels agreed with the experimenters' choice 83% of the time. Responses to stimuli in the image condition were quite varied. For example, the stimulus represented in Figure 1 stimulated affective descriptions (e.g., "a little sad", "comfortable") on the part of some subjects, scenic descriptions (e.g., "I see a countryside", "ballroom party") on the part of others, and relatively unemotional and abstract descriptions (e.g., "smooth and dark") on the part of others. We have not attempted a rigorous analysis of these responses because we suspect that the relative frequency of different types of description is highly dependent upon the detail of encoding task instructions, as well as upon individual-subject and individual-item differences. It can be noted however, that subjects found the task a natural one, although it clearly was not optimal for purposes of subsequent melody recognition.

Discussion

Again, to our surprise, there was little generalization to related items, regardless of encoding task. While encoding task did enter into an interaction with item type, there was clearly no indication that this variable influenced the recognizability of related items (related-versus-lure discrimination was .53 in all three encoding conditions). The contour-rhythm encoding task is apparently a relatively efficacious task for purposes of target recognition, but not for related recognition. The impressive invariance of related-recognition scores, in this experiment as well as Experiment 1, offer little encouragement for the view that appropriate encoding activities at input can facilitate subsequent recognition of related items on the basis of contour or other features (rhythm).

The results of Experiments 1 and 2 clearly contrast with those of short-term memory experiments which have shown a great deal of confusion between targets and same-contour related items. We began to think that the specificity of memory for targets might be the result of using real music as stimulus material rather than the rather artificial five-note melodies of the short-term memory experiments. It would not be the first time, after all, that different results were obtained with actual, meaningful materials that with artificial "nonsense" materials. Therefore we decided to try a long-term memory experiment similar to the foregoing, but using materials similar to those we had been using in the short-term memory experiments (Bartlett & Dowling, 1980). This simplification reduces the number of specific cues that differentiate Target-Related pairs from Lures and from each other; for example, rhythm and instrumental texture. The Target and Related stimulus of a pair had only melodic contour in common, and differed in interval sizes. No Lure had the contour of any Target-Related pair. We expected that the little interval information available in short-term recognition would disappear over time in the subject's memory, and that recognition performance after the several minute delay in this experiment would be based on the more general information contained in the melodic contour. That is, Targets and Relateds should have been more confused in long-term than in short-term memory.

Experiment 3

Method

Experiment 3 applied the structure of the preceding experiments using randomly generated tonal melodies. We used seven-note melodies because there are enough different contours($2^6 = 64$) for each Target-Related pair to have a distinct contour, while there are not enough five-note melodies distinct in contour ($2^4 = 16$). A rhythmic grouping of a seven-note melody of 3 + 3 + 1 was chosen because it sounds natural. We presented stimuli at a rate of 3 notes/sec, with the last note about three times the duration of the short notes, so that each stimulus was about 3 sec long.

In selecting stimuli, we used only 50 of the contours which contained at least two ups and downs. This was because we had some evidence that unidirectional and V-shaped stimuli differ qualitatively from the rest in memorability. Pilot studies showed that interval information is particularly difficult to retain with stimuli with few reversals of direction. And Ortmann (1933), using a recall paradigm, found contours of such stimuli easy to remember. Other investigators have found no evidence for the influence of number of reversals of melodic direction on accuracy of reproduction (Taylor, 1976) or recognition of an excerpted tone (Williams, Note 2). But Taylor's task seems to involve both interval and contour components; and Williams' task involves neither, if the argument of Dowling and Fujitani (1971) regarding direct recognition of pitches in untransposed melodies is to be believed. Therefore, we chose to use the set of 50 seven-note contours having two or more reversals of pitch direction as being most probably a homogeneous set with respect to both contour and interval recognition.

The 50 contours with two ups and downs divide into two sets of 25 such that one set consists of the inversions of the other set. (An inversion of a contour goes up where the original went down, and vice versa.) A set of 25 contours was randomly selected such that no member of the set was the inversion of any other member. Twenty-four of these contours (randomly selected) were divided into three sets of eight. The first two sets of contours functioned as either Targets or Relateds in the two versions of the input list; the third set of eight functioned as Lures. The remaining contour and one of the hitherto unused 14 contours were used for buffer stimuli in the input list.

With the contours selected (plus the additional buffer contour) we generated melodies according to the rules used by Dowling (1978); that is, given the direction of the diatonic step, $P(\pm 1 \text{ step}) = .67$, $P(\pm 2 \text{ steps}) = .33$. We generated a melody beginning on the first step of the major scale and a melody beginning on the sixth step, for each contour, using the same set of diatonic intervals for both. (We should note here that such a melody might accidentally have an implied tonality different from that defined by the initial conditions. For example, the melody: $+1$, -1, -2, -2, $+1$, -1, starting on the first step of C major would be: C, D, C, A, F, G, F, which definitely falls in F major, not C major.) Melodies starting on a sixth scale step constituted "tonal answers" of melodies starting on the first step, and vice versa.

Each list consisted of 18 items. The first and last items were buffers, not tested. The middle 16 items were tested, 8 tested as Targets in block 1 and as Relateds in block 2, and the other 8 tested as Relateds in block 1 and as Targets in block 2. A second list was constructed out of the tonal answers of all the 16 critical items in the first list. Items that were tested as Targets on block 1 for one list were tested as Relateds on block 1 for the other list, and vice versa. Different groups of subjects performed the experiment with the two input lists, using the same test lists. For each input list, the order of test blocks was reversed for half the subjects. This gave four list-test combinations with different groups of subjects. All Related items were tonal answers of list items, either in the one-six relation of melody origin or the six-one relation, equally often. The origin keys of list and test items were randomized, with the constraints that no item was tested in the key of its list presentation, and that none of an item's test or list presentations be in the same key as any other. All starting pitches lay within the ascending octave beginning on middle-C (fundamental frequencies of 262 to 495 Hz). Stimuli were produced on a freshly tuned Steinway piano, and a Davis timer was used as a metronome. Stimuli were recorded and presented as in Experiments 1 and 2.

Forty-one university students (as in Experiment 2) served in Experiment 3, with approximately equal numbers in the four list-plus-test conditions. While subjects in Experiments 1 and 2 made a dual judgment on each trial, attempting to respond positively to both Targets and Relateds with their confidence-level judgements and only to Targets with their checks, we instructed subjects in Experiment 3 to respond positively only to Targets. We gave them examples of Targets, Relateds, and Lures, and explained the differences. Subjects responded on a four-level confidence judgement scale ("sure same", "same", "different", "sure different"). This one-task procedure seemed a useful simplification since (a) subjects in Experiments 1 and 2 were unable to differentiate between the two types of judgment requested and seemed to make both judgments on the same information; and (b) it was subjects' inability to distinguish between Targets and Relateds in short-term memory that led to our surprise at their inability to generalize from Targets to Relateds in Experiments 1 and 2. We wanted to find out if subjects could perform with long-term memory a task they found very difficult with short-term memory.

Results

Table 3 shows mean area under the MOC scores for Experiment

Table 3
Areas Under the MOC for Experiments 3 & 4

	Experiment 3	Experiment 4	
		Inside	Outside
Target	.58	.75	.65
Related	.52	.72	.57

3. We performed a 2 Item Types × 2 Blocks × 2 Experience Levels analysis of variance. The effect of Item Type was significant, $F(1, 39) = 19.85, p < .001$, with Targets better recognized than Relateds. The effect of experience approached significance at the .05 level, $F(1, 39) = 3.63$. No other effects approached significance. We also tested whether overall Related recognition was better than chance, but it was not, $t(40) = 1.63$.

Discussion

At this point the results seemed paradoxical. In our previous short-term memory experiments (Dowling, 1978; Bartlett & Dowling, 1980) subjects were very poor at discriminating between transpositions of previously heard melodies and tonal answers—non-transpositions which shared contour and key with the originals but differed in interval sizes. Those experiments employed retention intervals of only a few seconds. Experiment 3 used retention intervals which averaged several minutes, and yet subjects showed above-chance discrimination between transpositions (Targets) and tonal answers (Relateds). There was another side to the paradox. In our previous experiments subjects had been excellent at discriminating between either transposition cues or tonal-answer cues on the one hand, and different-contour cues on the other. Yet in Experiments 1, 2 and 3 there was very little discrimination between same-contour tonal answer stimuli (Relateds) and the different contour Lures.

A possible resolution of this paradox is that contour information, which has powerful effects in short-term memory tasks, is not as useful in long-term memory tasks. In contrast, interval information, which is not used very effectively in the short-term memory situation, can be fuctional in some degree with longer retention intervals. This idea is compatible with some prior evidence, expecially the observation of Kallman and Massaro (1979) that same-contour different-interval cues are virtually useless in the recognition of octave-scrambled familiar songs. This suggests that contour, by itself, cannot be used effectively to retrieve a song from long-term memory. However, Kallman and Massaro's stimuli were well-learned songs whose interval patterns were thoroughly familiar to subjects. The present research suggests that long-term memory of even once-presented novel songs is based on an interval match, and not simply a match of contour. We know that relatively untrained subjects are poor at extracting precise interval information from novel stimuli (Attneave & Olson, 1971). Yet it might be that when intervals *are* extracted they are retained over long time periods. Contour information, though easily extracted, may be rapidly forgotten.

Since the stimuli of Experiment 3 were longer and more complex than those used in our previous short-term memory experiments, and so may have offered more cues by which subjects could have made better Target vs. Related distinctions, we decided to test the importance of interval information over long and short retention intervals directly. Experiment 4 provides for such a comparison at two retention intervals: an "empty" interval of 5 sec and a "filled" interval of 31 sec. While it is not certain that these two intervals represent a clean separation between short- and long-term memory, it is plausible that performance in the short-interval condition is mediated by information retained in a "maintenance rehearsal loop" of some kind (Craik & Lockhart, 1972). It is unlikely that performance in the long-interval condition should be so mediated. On the basis of Experiments 1, 2 and 3 we expected that contour information would dominate performance in the short-interval condition, but not with the long interval. In contrast, interval information should contribute to performance at least as much with the long as with the short interval.

Experiment 4

We used the same melodies in Experiment 4 as in Experiment 3. On each trial we presented two pairs of stimuli: an inside pair and an outside pair. The time delay between the standard and comparison stimuli in the inside pair was 5 sec. The time delay for the outside pair was 31 sec. On each trial the outside standard was presented first; then there was a 10-sec pause; then the inside standard was presented; then a 5 sec pause; then the inside comparison; then a 10-sec pause during which the subject responded to the inside pair; and finally the outside comparison. The recognition task involving the inside pair served as a distractor that interfered with the continued rehearsal of the outside pair. Such interpolated material has been shown to disrupt short-term retention of pitch (Deutsch, 1975). Hence it seemed reasonable that performance on the outside condition should differ from that on the inside condition.

Method

Stimuli were produced as in Experiment 4. Two lists of 36 trials each were constructed, and different groups of subjects heard the two lists. For the first list we randomly divided the 24 stimuli from Experiment 3 into a set of 12 outside standards and 12 inside standards. Three different random permutations of the 12 outside standards and the 12 inside standards made up the sequence of standard stimuli for the 36 trials. Each standard appeared three times, and was tested with a different type of comparison (Target, Related, or Lure) on each appearance. Test type of comparison was assigned randomly to trials with the constraint that each block of 12 trials contain four of each type. The outside standards in each block of 12 were in different keys, randomly assigned. The key of the inside comparison was always a perfect fourth of fifth (randomly determined) removed from that of the outside standard. All standards and comparisons started within the octave above middle C ($F_0 =$ 262 Hz). Half of the standards started on the first degree of the major scale, and half on the sixth, evenly distributed through the three blocks of 12. Comparison stimuli always began on a pitch a minor third higher or lower than the corresponding standard, inside or outside. If the standard began on the first degree of the scale, the comparison began a minor third lower, and vice versa. Related comparisons were in the same key as the standard. Target comparisons were transpositions of the standard melodies to keys determined by their starting notes. That is, Target transpositions of stimuli beginning on the first degree of the scale were in a key that added three sharps; transpositions from the sixth degree added three flats. Lures were diatonic inversions of Targets.

For the second list inside and outside standards were reversed, and the order of trials on which each appeared was reversed as well.

Twenty-five university students as described above served in Experiment 4: 11 with the first list and 14 with the second. Nine were experienced and 16 were inexperienced, divided approximately equally between the two lists. As in Experiment 3 subjects were instructed to respond positively only to Targets and were given examples of the difference between Targets and Relateds. Subjects were encouraged to rehearse the standard stimuli during the ISIs—we used the phrase "just as you would think through the melody of 'Happy Birthday' without actually singing it" and then hummed a little of it *sotto voce*, following this with instructions that rehearsals had to be silent. (We had found in pilot work with this paradigm that retention over the long interval was little better than chance without active rehearsal.)

Results

Table 3 shows the results for Experiment 4. We did a 2 Item Types × 2 Delays × 2 Experience Levels analysis of variance on the area scores. The effect of item types was significant, $F(1, 23) = 8.00, p < .01$, with Targets better recognized than Relateds. The effect of delay was significant, $F(1, 23) = 12.47, p < .005$, with better recognition on the inside pairs. No other effects were significant at the .05 level. While there was a trend for the difference between Targets and Relateds to be greater on the outside condition, the Item Type × Delay interaction did not approach significance, $F(1, 23) = 1.73$. Note, however, that the trend is opposite to that which would indicate forgetting of interval information over the interval between 5 and 31 sec. The main effect of delay suggests a loss of contour information, but not interval information, in the outside condition. Related recognition in the outside condition was nevertheless better than chance, $t(24) = 3.21, p < .01$.

GENERAL DISCUSSION

Our original hypothesis, which motivated the first three experiments, was that novel melodies could be recognized at the level of contour even after some delay. We expected subjects to be relatively good at discriminating between Related items (that shared contour with previously presented stimuli) and Lures (that did not). On the basis of prior work with short-term memory paradigms (Bartlett & Dowling, 1980) we expected little discrimination between Targets (transpositions or copies of list items) and same-contour Related items. Experiments 1 and 2 obtained exactly the opposite pattern of results. Subjects were virtually at chance in their recognition of Relateds, and this was so whether they were attempting such a discrimination or not. Moreover, discrimination between Targets and Relateds was consistently substantiated by significant main effects of item type. In Experiment 3 we replicated this pattern of results using relatively artificial materials similar to those we had used in our short-term memory research. Thus our results cannot be attributed to the complexities of actual music or to the uncontrolled nature of naturalistic stimuli. Further, Experiment 3 used transpositions, rather than exact copies of list items, as Targets. Therefore we know that the effects of item type in that experiment does not represent retention of absolute pitch information. (And the possibility of using such information over the long retention intervals of Experiments 1 and 2 is quite remote.) Finally, we should remark that the low Related recognition scores of our subjects cannot be explained by the overall low level of performance in Experiments 1 through 3. It is true that even Target recognition scores were relatively low in these experiments. The effects of item type, however, were highly reliable (with all alpha levels less than .001) if not very large in magnitude. Thus the evidence is quite clear that subjects reliably discriminate between Targets and Relateds in this long-term memory situation. In contrast, there was only marginal statistical support for discrimination between Relateds and Lures in Experiments 1 and 2, and none in Experiment 3.

Experiment 4 was intended as a bridge between the long-term memory situations of Experiments 1, 2, and 3 and the short-term transposition-detection paradigm of our previous work. Experiment 4 was completely successful in this respect, replicating in the short-delay (inside) condition the pattern of low Target-versus-Related discrimination and high Related-versus-Lure discrimination. As in earlier studies, subjects in the short-delay condition had great difficulty discriminating between transpositions of standards and tonal answers (Related items), but showed a strong tendency to discriminate between both of these same-contour items and different-contour Lures. In contrast, subjects in the long-delay (outside) condition showed relatively low discrimination between the same-contour Related and Lures, while the difference between Target and Related recognition was approximately equal to that found in the long term memory situation of Experiment 3. As usual, Target-versus-Related discrimination was never very large in magnitude, but it was highly reliable, and showed no tendency to drop from the short- to the long-delay condition. (There was no cue × delay interaction.) Experiment 4 shows that even a 30-sec delay, filled with musical stimuli and tasks, can drastically reduce the contribution of contour information to performance in a transposition detection task. Performance on the transposition detection task was never very high (with our non-musician subjects), indicting that precise interval information is quite difficult to extract from novel melodies upon a single hearing. Nonetheless, the contribution of interval information to performance is reliable, and is not reduced by an interpolated musical task. Indeed, a highly reliable role of interval information was even apparent with a traditional long-term memory paradigm involving lengthy input lists (Experiments 1, 2, and 3). In summary, the present results suggest that contour information is easily extracted from novel musical stimuli, but contributes to performance only with short (and/or unfilled) retention intervals. In contrast, interval (and/or chroma) information is difficult to extract, but contributes more or less equally to performance over a broad range of retention intervals.

It is important to consider here the point that was touched on earlier, that the construction of stimuli in Experiments 3 and 4 was done in such a way as to enhance contour recognition. If all the subject has to do is distinguish between one contour and another, then the two most different contours, in terms of interval-for-interval comparison, are inversions of each other. Any of the six intervals in the seven-note melody will disclose the difference. By comparison, the stimuli used in our previous work (Dowling, 1978; Bartlett & Dowling, 1980) were less conducive to contour recognition, in that fewer intervals were involved and that Relateds differed from Lures in few interval directions. Yet in the long-term memory tasks of Experiments 3 and 4, contour information was not well preserved in memory. Further, the instructions of our earlier short-term memory tasks emphasized responding *positively only to Targets (transpositions) as opposed* to Relateds (same- or near-key imitations). Subjects were generally not able to do that, experiencing great confusion between Targets and Relateds. Yet in Experiments 1 and 2, subjects were instructed to respond positively to both Targets and Relateds, and succeeded mainly in picking out Targets.

A side issue worth commenting on here is that of the possible effects of contour complexity on memory for melodies. If it is true, as we have argued elsewhere (Dowling, 1978; Bartlett & Dowling, 1980), that memory for melodic contour is an important component of memory for melodies, then complexity of contour, in terms of reversals of contour direction, should have some impact on memory. As noted above, Ortmann (1933) found such an effect with a recall task in which musically trained subjects wrote down what they had heard. (Contour was almost never misremembered when there were zero or one reversal of direction.) However, as noted above, it is important to distinguish contour-recognition from interval-recognition components in memory tasks. To the degree that interval recognition is involved, melodies with few contour reversals, and hence few repeated pitches and intervals, will be more difficult to recognize. To the degree that just contour recognition is required, then the simpler contours will be easier. When different-contour Lures differ in every one of six interval directions (as in Experiments 3 and 4), contour changes are easier to detect than when Lures differ in only one or two of four intervals. Contour complexity needs to be explored with a wide variety of tasks, as does interval complexity.

A major question raised by these results is why false recognitions of same-contour-different-interval cues (tonal answers) falls so drastically with a slight increase in retention interval. At an empirical level, the finding converges nicely with that of Kallman and Massaro (1979). Still, it is not obvious how the result should be explained. The explanation cannot be that contour information is not picked up by subjects in the first place, because the short-delay data show that it is. Further, it seems implausible that contour information has a rapid rate of forgetting. First, long-term memory experiments on octave scrambling (Dowling & Hollombe, 1977; Idson & Massaro, 1978) show that contour information *can* contribute to melody recognition if chroma information is also preserved in the retrieval cues. More to the point, retrievability of interval information without contour information seems a logical impossibility. The contour of a melody is clearly implicit in its interval structure (see Dowling and Fujutani, 1971).

It seems to us that the low effectiveness of same-contour-different-interval cues in *long term memory situations could be explained in either of two* ways: one emphasizing the nature of melodic representations in memory, and the other emphasizing the nature of the retrieval process in short-term versus long-term situations. A representational account might begin with the assumption that melodies of the sort we studied here are retained in the form of two separate memory traces. First, there is a short-term trace which explicitly represents information pertaining to absolute pitch levels (Deutsch, 1978; Attneave and Olson, 1976), contour, and (perhaps) key (Bartlett & Dowling, 1980). Second, there is a long-term trace which does *not* contain absolute pitch or key information, but which explicitly represents precise interval information along with (perhaps) mode (Dowling, 1978). This long-term trace would represent contour information, but only implicitly. That is, the listener could retrieve a contour only by first retrieving the melody, rehearsing it, and extracting contour from the rehearsed version. We might assume that only explicitly represented information can play a role in retrieval of a trace. In this case, a same-contour-different-interval cue would not access a long-term trace, although it could access a short-term trace. Thus, these cues would cause many false alarms in short-term situations, such as the inside condition of Experiment 4. Within the framework of such a dual-representation theory, the present data suggest that the short-term trace is truly short term — it can be rendered ineffectual by as little as 31 sec of intervening activity.

The alternative account of our data, one that seems more plausible to us, does not propose two separate memory traces, and makes only minimal assumptions regarding the content and format of the information stored. This account proposes that listeners' retrieval operations in a short-term situation differ qualitatively from those in a long-term situation, and that this difference in retrieval processes is responsible for the low effectiveness of same-contour-different-interval cues in long-term-memory situations. Consider the task of a listener who has just heard a melody and now must decide if a comparison melody is a transposition of it or not. It is not necessary for the comparison stimulus to remind the listener of the standard, he is already "reminded" of it. (That is, if we assume that a melody is stored at a certain location in an immediate testing situation.) In such a situation, the listener might be sensitive to many sources of similarity between the standard and the comparison, including contour (Experiment 4, inside condition) and pitch or key (Bartlett and Dowling, 1980). Since it is difficult to extract interval information from novel melodies, these other "irrelevant" sources of similarity might frequently govern responses. Hence, the high false-alarm rate to same-contour cues. Now consider the task of a listener in a situation like that of Experiment 3, where he hears a test item and must decide if it matches any one of 18 previously presented melodies. Here, the process of *reminding* is crucial. If a cue does not provide access to the appropriate trace, the listener is unable to detect *any* type of similarity between the test cue and the appropriate input item. Psychologists still know little about the process of retrieval from long-term memory, but it is easy to imagine why a same-contour-different-interval cue might be ineffective. First, it is certain that many songs in long-term memory contain the same or similar contours, especially if we restrict attention to seven-note segments. Moreover, even within the context of an 18-item list, such as that used in Experiment 3, many melodies have identical contours up to the third or fourth note. Thus, contour information by itself might be of little use in memory retrieval simply because it is not very discriminating. Interval information might be much better in this regard. While many songs share segments of contours, fewer share precise patterns of intervals for several notes in succession. Hence, subjects might be forced to use interval information in the retrieval process, even if the task instructions discourage such a strategy (Experiments 1 and 2).

One advantage of the retrieval account is that it nicely handles the fact that same-contour-different-interval cues are relatively more useful in long-term memory when the set of response alternatives is small and known to subjects. Both Dowling and Hollombe (1976) and Idson and Massaro (1978) found above-chance recognition of same-contour-different-interval cues when subjects knew that each melody was one of a few well-known songs. In contrast, Kallman and Massaro (1979) found almost no recognition of same-contour cues when subjects were not told in advance the stimulus set. Presumably, a small set of alternatives alters the retrieval process (Deutsch, 1975; Kallman & Massaro, 1979) in some way. It is probable that the retrieval process used with a small set of alternatives is essentially that used with a short, unfilled retention interval. In both cases the subject might be able to compare a test melody with his representation of a previously presented melody, noting many sources of similarity including contour. In the long-term, large-alternative-set situation, the subject might be forced to rely on interval and/or chroma information in order to narrow his search of memory. It should be stressed that this retrieval view makes no claims regarding how melodies are represented in memory. It does not even preclude an extreme ''literal copy'' or ''echo'' view, nor does it imply such a view. Also note that it is not necessary to choose between a representational view and a retrieval view. Without doubt, a truly adequate theory will articulate properties of both representation and retrieval of melodies in short- and long-term situations. The present discussion merely highlights our extreme ignorance about these matters.

REFERENCES NOTES

1. Dowling, W.J., and Bartlett, J.C. *Memory for Beethoven Quartets: Effects of Context and Chunk Size.* Paper presented to the Psychonomic Society, San Antonio, November, 1978.

2. Williams, D.B. *A Study of Tonal Strength and Its Influence on Melodic Memory.* Paper presented to the North-Central MENC Convention, Indianapolis, 1979.

REFERENCES

Attneave, F., & Olson, R.R. Pitch as medium: A new approach to psychophysical scaling. *American Journal of Psychology,* 1971, *84*, 147-166.

Bartlett, J.C., & Dowling, W.J. The recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance,* 1980, *6*, 501-515.

Brown, R. Music and language. *Proceedings of National Symposium on the Applications of Psychology to the Teaching and Learning of Music.* Reston, Va.: Music Educators National Conference, 1980.

Craik, F.I.M., & Lockhart, R.S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11*, 671-684.

Delis, D., Fleer, J., & Kerr, N.H. Memory for music. *Perception & Psychophysics*, 1978, *23*, 215-218.

Deutsch, D. The organization of short-term memory for a single acoustic attribute. In J.A. Deutsch (Ed.), *Short-term memory*. New York: Academic Press, 1975, pp. 107-151.

Deutsch, D. Delayed pitch comparisons and the principle of proximity. *Perception & Psychophysics*, 1978, *23*, 227-230.

Dowling, W.J. Recognition of transformed melodies: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics*, 1972, *12*, 417-421.

Dowling, W.J. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 1978, *85*, 341-354.

Dowling, W.J., & Fujitani, D.S. Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, 1971, *49*, 524-531.

Dowling, W.J., & Hollombe, A.W. The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception & Psychophysics*, 1977, *21*, 60-64.

Idson, W.L., & Massaro, D.W. A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics*, 1978, *24*, 551-565.

Kallman, H.J., & Massaro, D.W. Tone chroma is functional in melody recognition. *Perception & Psychophysics*, 1979, *26*, 32-36.

Ortmann, O. Some tonal determinants of melodic memory. *Journal of Educational Psychology*, 1933, *24*, 454-467.

Swets, J.A. The relative operating characteristic in psychology. *Science*, 1973, *182*, 990-1000.

Taylor, J.A. Perception of tonality in short melodies. *Journal of Research in Music Education*, 1976, *24*, 197-208.

We thank James Carlsen, Lola Cuddy, Darlene Smith, and David Williams for helpful suggestions and comments. Requests for reprints should be addressed to W. Jay Dowling, Program in Psychology and Human Development, University of Texas at Dallas, Richardson, Texas 75080.