1982

Musical Expertise and Melodic Structure in Memory for Musical Notation

Andrea Halpern
Bucknell University, ahalpern@bucknell.edu

Recommended Citation

Halpern, Andrea. "Musical Expertise and Melodic Structure in Memory for Musical Notation." *The American Journal of Psychology* (1982) : 31-50.

UNIVERSITY OF ILLINOIS PRESS

Musical Expertise and Melodic Structure in Memory for Musical Notation
Author(s): Andrea R. Halpern and Gordon H. Bower
Source: *The American Journal of Psychology*, Vol. 95, No. 1 (Spring, 1982), pp. 31-50
Published by: University of Illinois Press
Stable URL: http://www.jstor.org/stable/1422658

Accessed: 20/05/2010 10:10

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=illinois.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

University of Illinois Press is collaborating with JSTOR to digitize, preserve and extend access to *The American Journal of Psychology.*

STOR®
http://www.jstor.org

AMERICAN JOURNAL OF PSYCHOLOGY
Spring 1982, Vol. 95, No. 1, pp. 31-50

Musical expertise and melodic structure in memory for musical notation

ANDREA R. HALPERN AND GORDON H. BOWER
Stanford University

Two experiments plus a pilot investigated the role of melodic structure on short-term memory for musical notation by musicians and nonmusicians. In the pilot experiment, visually similar melodies that had been rated as either "good" or "bad" were presented briefly, followed by a 15-sec retention interval and then recall. Musicians remembered good melodies better than they remembered bad ones: nonmusicians did not distinguish between them. In the second experiment, good, bad, and random melodies were briefly presented, followed by immediate recall. The advantage of musicians over nonmusicians decreased as the melody type progressed from good to bad to random. In the third experiment, musicians and nonmusicians divided the stimulus melodies into groups. For each melody, the consistency of grouping was correlated with memory performance in the first two experiments. Evidence was found for use of musical groupings by musicians and for use of a simple visual strategy by nonmusicians. The nature of these musical groupings and how they may be learned are considered. The relation of this work to other studies of comprehension of symbolic diagrams is also discussed.

To the musically illiterate, musical notation is simply a series of dots on lines. To the cognitive psychologist, the perception, comprehension, and memory for notation pose some interesting questions. The object of this study was to examine how the experienced music reader takes advantage of his or her expert knowledge to convert those "dots on lines" into a meaningful, unambiguous message about the piece of music.

First, let us describe notation in a little more detail, to formulate the problem for the cognitive psychologist in understanding the comprehension of notation. Notation is a conventional visual representation of auditory entities—the sounds the composer intends the performer to produce and the audience to hear. The mapping from spa-

0002-9556/82/0100-0031$02.00/0

tial arrangement to pitch is ordinally analogue in that the spacing of notes on the lines approximately reflects the spacing of tones in pitch. Notes ascending or descending visually on the five line staff represent notes ascending or descending in pitch. Of course, a musical composition is more than an ordered list of pitches: other aspects of notation tell the musician about rhythm, phrasing, loudness, and expression. However, even the essentially analogue representation of pitch (about which this paper is concerned) has some abstract components that complicate a simple mapping from pitch to notation. For instance, although the visual distance between any two lines on the staff is constant, the musical distance between notes on the first and second lines (E to G, the interval of a minor third) is smaller than the musical distance between the notes on the second and third lines (G to B, a major third).

This paper is concerned with the role that musical structure has in memory for notation. Given two sequences that are visually similar, will the musician be able to remember the passage with a good musical structure ("good melodies") better than he or she remembers one with a poor structure ("bad melodies")? If so, what aspects of the good pattern would enable the superior performance?

Similar questions have been extensively studied in memory for chess positions. In his studies of chess players, de Groot (1965) found few differences between chess masters and novices in the basic reasoning associated with playing, such as the number of moves considered per play, or the depth of search for moves. However, he did find a difference between the better and poorer player in a short-term memory task. Between 20 and 25 chess pieces arranged on a chessboard were exposed for 5 sec, then the player tried to reconstruct the board from memory. The expert was far superior to weaker players, but only when the chess pieces formed a pattern that could legally and "sensibly" be found in a real game. When the chess pieces were placed randomly on the board, then memory performance was equally poor for the two kinds of player. This result suggests that chess experts exceed novices only in chess-specific patterns, not in general memory capacity.

In extending this work, Chase and Simon (1973a, 1973b) described how the expert perceives a chess position. They noted that the expert's recall in the de Groot task followed a particular time course. He would rapidly recall (place on the board) chess pieces that were related either physically (same color or type of piece) or functionally (attack and defense relations). After a longer pause, new clusters of related pieces would be placed down. Chase and Simon proposed that the expert was segmenting the chessboard into familiar subconfigurations (or "chunks"). Experts supposedly included more pieces in each chunk than novices did, thus remembering more individual pieces at recall.

Other research has confirmed that an expert's memory is superior to a nonexpert's, but generally only for meaningful groupings of material (given that the nonexpert at least has some familarity with the stimulus material). In further studies of chess players (Charness, 1976; Frey & Adesman, 1976), of players of the game *Go* (Reitman, 1976), and of readers of electronic circuit diagrams (Egan & Schwartz, 1979), it has been found that experts differ from novices in memory for meaningful but not random patterns. The experiments presented here were an attempt to extend this work to the reading of musical notation.

The meaningfulness of a musical pattern is harder to determine that that of a chess pattern. Certain chess patterns are clearly nonsense to a person familiar with the rules. But in music, almost any sequence or combination of notes may be judged as "good music" by some listeners. Therefore, in our experiments, we avoided the strategy of presenting random strings of notes as the "bad" melodies and familiar tunes as the "good" melodies. Rather, we composed a large number of melodies that had similar visual patterns and then presented them to a panel of judges for rating on quality. The main stimulus variable was how good or bad those melodies were judged to be. Simple, short melodies were used in order to exclude rhythmic and other nonmelodic factors for the purposes of this experiment. "Experts" were taken to be musicians with at least 10 years of music reading experience; "nonexperts" were nonmusicians who had seen notation before but who were not able to read it.

Stimulus melodies

As the precise nature of the stimulus melodies is quite important, they will be described prior to discussion of the individual experiments.

All melodies were 10 notes long; each note was a quarter note (so all were equal in temporal value), and there were no sharps or flats. No adjacent notes were identical. A professional musician generated 24 "good" melodies as an initial pool for normative ratings. He freely adapted extant melodies from many sources to conform to the above constraints. The "bad" melodies were generated by taking the intervals between notes of a given good melody and permuting their order to violate as many features as possible of a good melody. Examples of
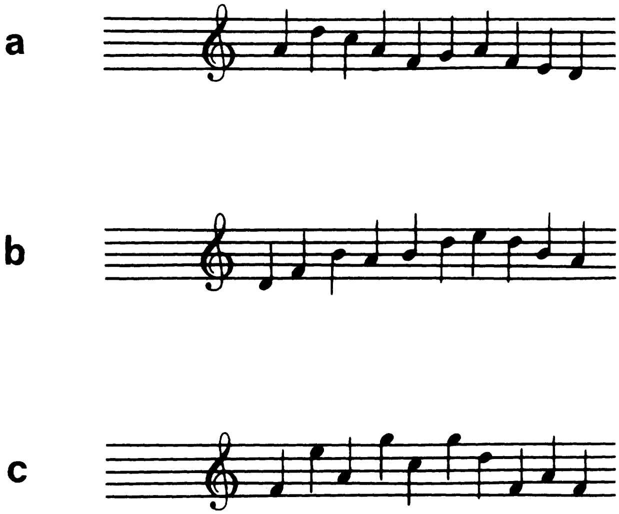
Figure 1. An example of a good melody (A), the bad "mate" to that good melody (B), and a random melody (C)

these permutations are positioning a four line-and-space interval so the first note is an F and the second a B (forming a tritone, a dissonant interval) or positioning the melody in the "wrong" key (i.e., where a sharp or flat would be needed to make a meaningful melody). An example of a good melody is shown in Figure 1(A) and its bad "mate" is shown in Figure 1(B). The 24 good melodies and their bad mates (48 in all) were presented on music paper in random order to six Stanford graduate students in music for sorting into equal sized categories of "better" or "poorer" melodies. To qualify for use in the later memory experiments, a given pattern had to be correctly classified by at least four of the six judges. Of these patterns, 42 met this criterion; the other 6 and their yoked good or bad patterns were eliminated, leaving a final stimulus set of 18 good patterns and their 18 bad mates. Each bad mate had an overall visual configuration that was similar to its good melody. The permutation algorithm insured that average interval sizes were identical, as were the average number of pitch contour changes (3.61 for good, 3.66 for bad melodies).

In addition to "good" and "bad" melodies, 18 "random" sequences were composed. The random melodies were generated by sampling with replacement from the notes E above middle C to G in the next octave and randomly determining the left to right sequence of notes. As was true for the good and bad melodies, no two consecutive notes were identical, and a pattern was eliminated if by chance it appeared to have the structure of a good or bad melody (only one pattern was so eliminated. Average step size was far greater than for the good and bad melodies, and average number of contour changes equaled 6.20. An example of a random melody is shown in Figure 1(C).

EXPERIMENT 1: PILOT

This initial experiment explored the suitability of using the stimuli described above in a memory task. A secondary purpose of the experiment was to investigate possible ways in which musicians and nonmusicians encode the notation.

The method was the same as that more fully described in Experiments 2 and 3 below. Subjects were 12 musicians and 12 nonmusicians. Only good and bad melodies were used. Subjects saw a slide of a melody for 5 sec and recalled it in written form 15 sec after slide offset. The 15-sec retention interval was either unfilled or filled with a distracting auditory or visual task. Due to technical difficulties, the analysis of the distraction tasks was not possible and will not be described further.

However, collapsing across the tasks revealed the expected pattern of results for the other two factors: musicians performed better than nonmusicians, good melodies were remembered better than bad melodies, and the melody type interacted with subject type. The musicians remembered more of the good than bad melodies, whereas the nonmusicians performed equally poorly on the good and bad melodies. This last result validated the attempt to make the two sets of melodies equally complex visual patterns. In addition, although musicians recalled bad melodies more poorly than they remembered good melodies, even for bad melodies their recall exceeded that of the nonmusicians.

Given that the two types of melodies looked similar, one may ask what it was about the good melodies that the musicians could use to good advantage. One question is whether notated melodies rated as good or bad would also sound good or bad to both kinds of subjects. To check this, a rating follow-up was carried out with a separate group of subjects. All the good and bad melodies were played on an electric organ in a random order at a rate of two notes per sec. The melodies were recorded on audio tape with a 7-sec silent period between each one. The tape was presented to 13 musicians and 9 nonmusicians for rating on a 9-point scale of melodic goodness. Both groups rated the good melodies as being significantly better than the bad melodies (t = 4.50 for musicians, t = 4.95 for nonmusicians,

$p < .05$). Thus the stimulus set classification was confirmed both for auditory and visual presentation.

EXPERIMENT 2

The pilot study revealed the appropriateness of using our particular stimuli in a memory task. The next experiment further investigated the effect of musical structure on remembering notation. Like many folk and popular tunes, the good melodies had many small vertical distances between adjacent notes (corresponding to small differences in pitch). To form the bad melodies, we permuted the order of those notes. But these melodies still appear very regular largely because they still have the many small internote distances of the good melodies. This gives them the appearance of a smoothly varying envelope, a simple visual pattern to recall. Nevertheless, the musicians recalled those bad but visually simple melodies better than the nonmusicians did. For what kind of melodies might the musicians and nonmusicians perform equally?

Chase and Simon showed that the chess expert's recall was equal to the novice's only for random chessboards. A closer analogy to random chessboards was obtained in our Experiment 2 by generating truly random melodies. Random placement of notes produces irregular visual patterns, since large intervals between adjacent notes are as likely to occur as small ones. This increase in visual complexity from bad to random melodies should adversely affect recall of both musicians and nonmusicians, although perhaps not to the same extent. Musicians presumably use both visual regularity and musical regularity to aid recall. However, nonmusicians presumably will only use visual regularity. Recall for musicians should be best for good melodies (good visual and musical structure), intermediate for bad melodies (good visual, less musical structure) and worst for random melodies (poor visual and musical structure). For nonmusicians, recall should be moderate and equal for the good and bad melodies since these are equated in visual complexity and the musical differences have no meaning for them. Nonmusician's recall should be worst for the random melodies, which have little visual structure.

METHOD

Subjects

The musicians were six Stanford students who had at least 10 years of experience in reading music (range: 10 - 18 years, median: 12.5) and were paid for their participation. Six Introductory Psychology students with no experience in reading music received course credit for their participation.

Stimuli

The good, bad, and random melodies were drawn on a music staff as in Figure 1, mounted on 3 × 5 in. (7.6 × 12.7 cm) white cards. A treble cleff appeared at the far left of the staff, followed by a space of 3/8 in. (9.5 mm), then the 10 notes separated by 1/4 in. (6.4 mm). Photographic slides of each card were prepared.

Procedure

One to three subjects were tested at a time. The format of the melodies was described to the subjects (e.g., being all quarter notes), and, for the recall test, the nonmusicians were specifically instructed to place notes only on lines or spaces of the staff. Subjects sat at a table placed 6 ft. (1.8 m) away from the viewing surface. Each melody subtended a horizontal visual angle of about 17 degrees.

A trial began with the experimenter saying "ready." She then activated a timer, which caused the stimulus slide to be projected for 4 sec. Immediately after slide offset, subjects wrote down the melody by penciling notes (dots) on music paper. Recall of exactly 10 notes was encouraged. Unlimited time was allowed for recall, but it rarely exceeded 1 min. Subjects first received 6 practice trials, then 18 good, 18 bad, and 18 random melodies in a random sequence. Half the subjects saw the sequence in reverse order. A 2-min. rest occurred midway in the 40-min. session, during which subjects filled out a questionnaire about their musical background.

Scoring

Initially, each melody was scored on an "absolute" basis: every *n*th note of the subject's recall was compared to the *n*th stimulus note, and the response was counted correct only if these notes were identical. In the event that the subjects produced more than 10 notes, only the first 10 notes were scored. A recall of less than 10 notes was scored as the first *n* notes of the melody. These proportions correct were transformed to arcsine scores following the suggestion of Winer (1971).

The absolute scoring system, while providing one estimate of accuracy, is inadequate for capturing many subtleties of "near misses" in the recalled pattern. For example, a subject who recalled the entire melodic pattern, but displaced it one line or space, would receive a recall of 0, and the scoring method would miss the fact that a considerable amount of information had in fact been transmitted by the response. Strict scoring also ignores "musically meaningful" mistakes such as misrecalling a note by an octave. For these reasons, the recall patterns were also scored by a similarity judgment procedure.

Two musically knowledgeable judges compared each response of three musicians and three nonmusicians to the correct answer and assigned it a

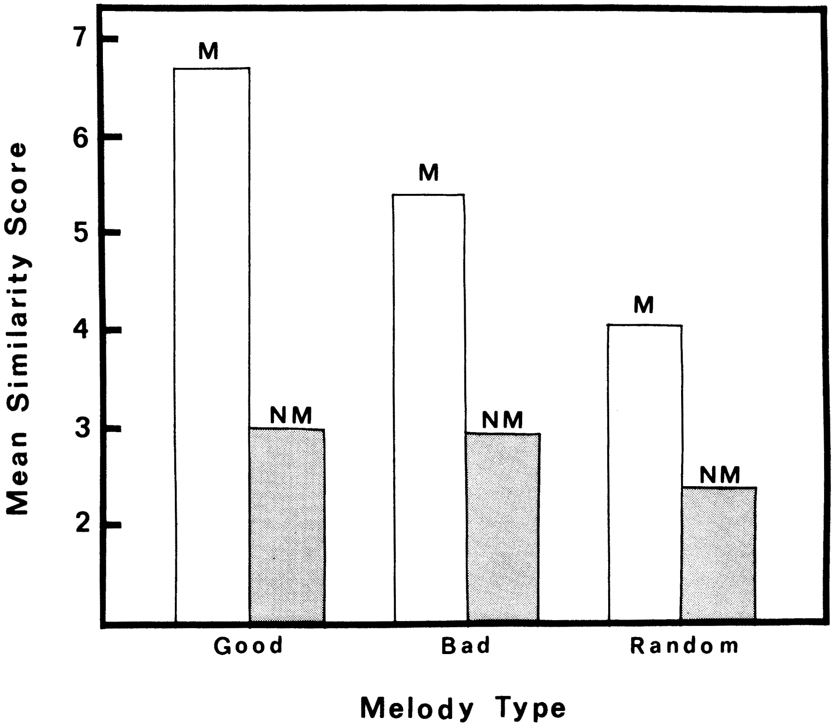

Figure 2. Mean similarity scores for musicians (M) and nonmusicians (NM) for each melody type in Experiment 2

similarity score from 1 (low) to 10 (high). Judges were allowed to use any criteria of visual and/or musical similarity they felt were relevant. However, some "ground rules" were established beforehand. Judges agreed to award a "10" only to perfect responses, and a "1" if the subject's response seemed totally unrelated to the stimulus pattern. To check the judges' reliability, both of them scored protocols for three subjects chosen at random. Their similarity ratings were correlated for these data. Agreement was nearly perfect, $r = .97$.

Both scoring systems revealed the same pattern of results, except where otherwise mentioned below. Because of their greater sensitivity, only the similarity scores will be reported here.

RESULTS

The task was difficult but not impossible for the nonmusicians. By absolute scoring, chance level would be 8% correct. The nonmusicians' scores ranged from about 15% to 25% correct by that scoring method. The task was also difficult for the musicians; recall rarely exceeded 60% correct. By similarity scoring, nonmusicians' scores averaged between 2 and 3 out of a possible 10, reflecting some recall ability. Musicians' scores were somewhat more wide-ranging, but no subject even approached perfect performance in any condition. Therefore, the recall scores fell in a range sensitive to experimental influence.

The main recall results are shown in Figure 2. A two-way analysis of variance (2 Groups × 3 Melody Types) revealed both main effects to be significant. Musicians recalled more than nonmusicians did, $F(1, 10) = 17.49$, and melody types differed from each other, $F(2, 20) = 52.40$. In addition, melody type interacted with group, $F(2, 20) = 17.28$, $p < .001$.

The nature of the interaction was more fully explored in post-hoc tests. A Newman-Keuls test revealed that nonmusicians performed equally on good and bad melodies and more poorly than either on the random melodies ($p < .05$ for all comparisons). In absolute scoring, the bad vs. random comparison was not significant.

As predicted, the musicians' performances were best with the good, intermediate with the bad, and worst with the random melodies ($p < .01$). Comparisons between musicians and nonmusicians showed a reliably decreasing advantage of musicians as the melodies changed from good to bad to random. For good melodies, musicians scored an average of 3.67 similarity points higher than nonmusicians; for bad, 2.67; and for random, 1.68. This last (smallest) difference is, however, significant, $t(17) = 6.85$, $p < .01$. The continued advantage that musicians had over nonmusicians for all stimulus patterns is contrary to the results found for chess experts vs. novices by Chase and Simon (1973a). This is not surprising considering that musicians have had 10 or more years of perceptual differentiation training with the notes as well as note names to aid their encoding and recall.

Qualitative aspects of the responses

Several qualitative aspects of performance can be noted informally. Subjects usually recalled the melody temporally in a serial, left-to-right fashion without any explicit instructions to do so. A few nonmusicians sometimes recalled the last note first before recalling the beginning of the melody. Overall, no serial position effects were apparent for either group. Both groups made many visual pattern errors, such as displacing a pattern either horizontally or vertically. Musicians committed many "musical" errors, such as writing down an incorrect but melodically good ending to a melody. As another example, in one melody the first five notes were taken from the F major triad in the order C-A-F-A-C. Several musicians permuted the order of these notes in their recall, as if they had grouped these into a chunk labeled as "F-major chord," but had then lost order information within the chunk. Many fewer such mistakes occurred in recalling the bad melodies, since the unstructured patterns probably suggested no melodic alternative for a particular note.

DISCUSSION

Experiment 2 confirms several findings of the pilot experiment. First, musicians recalled good melodies better than they recalled bad ones, even when recall was immediate. Thus, the influence of melodic structure occurs quite early in the music reading process and is not just an effect arising from a long retention period. Second, nonmusicians recalled good and bad melodies at the same moderate level, demonstrating again the equal visual complexity of these patterns in the absence of music reading skills.

As predicted, both groups recalled the random melodies more poorly than they recalled the good or bad melodies. The nonmusicians, who were at least able to draw the contour of the good and bad melodies, were less able to reproduce the jagged, irregular shape of the random melodies. The musicians, presented with poor musical structure as well as a complex visual pattern, probably had to rely on strategies such as naming individual notes to produce recall superior to that of the nonmusicians. Use of minimally trained musicians (who at least know note names) instead of nonreaders of music in the unskilled subject group may have led to equal performance in the random stimulus condition.

This experiment is analogous to the conditions under which a musician would be asked to deal with rapidly presented lines of notation, as in sight-reading piano music. Under conditions resembling a sight-reading task, Sloboda (1976a, 1977) found that musicians process written notes as a series of groups. These groups can be either musical ones (such as phrases) or structural ones (places where musical typography requires a vertical line or blank space on the staff). Furthermore, Sloboda (1978a) found that when short music patterns were exposed very briefly, musicians were superior to nonmusicians only for immediate recall of relatively "simple" patterns, i.e., those with few changes of visual direction. Indeed, he reported that some subjects claimed that these simple musical groups seemed to "spring up out of the card" into recognizable patterns (p. 330).

Following these suggestions, Experiment 3 was designed to assess the importance that ease of note grouping has for music reading. As the original stimulus patterns were constructed using only subjective estimates of their goodness, a way of determining the relative presence of recognizable subgroups was needed before this factor could be related to memory performance.

EXPERIMENT 3

When considering the kinds of note arrangements common to subjectively good melodies, two characteristics come to mind that would also be helpful in memory of those patterns. One is the ease of grouping a melody into subpatterns. To investigate this factor, ease of grouping was operationalized as the across-subjects consistency with which a melody is divided into groups. A second characteristic of a melody is the average number of groups it is typically divided into.

How might these two features of a melody affect memory for it? If musicians use a grouping strategy to remember the melodies, then melodies that are easily grouped should be remembered better, especially when study time is limited as in these experiments. Second, melodies that can be divided into fewer groupings should be more easily remembered since memory need only code these few "chunks" and not each individual note. This prediction is supported by a recent study by Deutsch (1980). Auditory sequences described by only a few generative rules were recalled better than sequences needing complex descriptive rules. Thus it was predicted that, for musicians, good melodies should contain fewer groupings than bad melodies and that good melodies would be more consistently grouped than bad melodies. In addition, the average number of chunks in a melody should be negatively correlated with its memorability, and the degree to which a melody is consistently grouped should be positively correlated with its memorability.

Nonmusicians would of course be insensitive to musical groupings, as there were no obvious visual cues to differentiate the notes. They were not expected to differentiate good from bad melodies by either of these measures, nor were their measures predicted to correlate positively or negatively with their recall of the melodies.

METHOD

Subjects

Musicians were seven students and professional musicians, all fluent music readers. Nonmusicians were seven students with no knowledge of music reading. Subjects were paid for 10 to 15 min. of participation.

Stimuli

The good and bad melodies from the previous experiments served as stimuli.

Procedure

Each subject was presented with a sheet of music paper upon which were listed the 18 good and 18 bad melodies in a random order. The instructions attached to the sheet read as follows: "On the following page you will find 36 melodies of 10 notes each. For each melody, I would like you to divide the notes into groupings by marking the melody with zero, one, or more vertical slashes. Use any criteria that you feel will help you segment the melodies—put notes that seem to go together into the same group. You are free to use as many groupings as you like."

Scoring

The first aspect scored was the consistency across subjects with which each melody was segmented. To assign a "consistency score" to each melody, the following procedure was used (analyses were done separately for musicians and nonmusicians). First, for a given melody, it was noted how many subjects out of seven had placed a dividing mark between the first and second notes, the second and third notes, etc. If at least six of the seven subjects agreed about the presence or absence of a division between two notes in the melody, that position was counted as "consistently classified" by the subjects. A melody's total consistency score was the number of its consistent positions out of nine.

In addition, the number of groupings each subject chose as being appropriate for each melody was noted, and the average number of groupings for each melody was computed.

RESULTS

Consistency of grouping will be considered first (see Table 1).

In fact, musicians did not exceed the nonmusicians in consistency of segmenting the melodies. There were no significant differences between the two subject groups for the good melodies; surprisingly, nonmusicians were actually more consistent than the musicians in grouping the bad melodies, $t(17) = 4.70$, $p < .001$. However, inspection of the groupings used suggested a simple explanation: most nonmusicians grouped the melodies according to whether the "stem" of the note was pointing upward or downward. By convention, notes whose "head" rests below the middle line of the five-line staff have stems pointing up. Notes above the middle line have stems pointing down; a note on the middle line can point in either direction. Stem direction has nothing to do with the melodic or rhythmic aspects of the

Table 1. Mean consistency score and number of groupings for good and bad melodies

Group	Melody type	
	Good	Bad
Consistency		
Musicians	4.50	3.50
Nonmusicians	5.16	6.56
Number of groupings		
Musicians	3.61	3.83
Nonmusicians	3.29	2.93

Note: Maximum score = 9 for consistency and 10 for number of groupings.

pattern. Yet stem direction enables a nonmusician to make a consistent, if nonmusical, grouping of the notes.

For neither subject group were the good melodies more consistently grouped than the bad. This refutes the hypothesis that part of what makes a melody "good" to a musician is some obvious, consistent grouping of that melody into chunks.

Although good melodies were not chunked more consistently than bad ones were chunked, one may still ask whether easily chunked melodies are easily remembered, regardless of their rating as good or bad. To check for this, all the melodies were rank-ordered by their consistency scores separately for each subject group. For each melody, again separately for musicians and nonmusicians, the average recall scores from Experiment 2 (immediate recall) and from the 15-sec unfilled condition of the pilot experiment were computed and rank-ordered. A significant correlation would suggest that the ease with which a melody can be grouped determines its level of recall, thus arguing for the existence of chunking in this recall task. It was predicted that nonmusicians would show no significant correlations, as their strategies in the memory task (drawing the contours) and in the grouping task (grouping the same-stemmed notes) are unrelated.

As shown in Table 2, consistency was not correlated with memory performance in either condition for the nonmusicians. This is as predicted. However, a significant correlation was obtained for musicians between consistency of a melody and its immediate recall. This suggests that when musicians must rely on a quick glance at a melody for later recall, they use the "chunks" inherent in that melody to reduce memory load.

An alternative hypothesis is that recall is aided not by ease of chunking a good melody but by the reduction in the number of chunks in a good melody. If a chunking strategy were being used for encoding and storage, one would expect a lower recall the greater the average number of chunks for a given melody, but this effect should occur only for the musicians. In addition, one would expect that there would be fewer chunks in good melodies than in bad ones, but again, only for the musicians.

To test these predictions, the average number of groupings assigned to each melody was determined, again separately for musicians and nonmusicians. As shown in Table 1, musicians did group the good melodies into significantly fewer chunks than they grouped the bad melodies, $t(34) = 5.55$. Interestingly, nonmusicians grouped the bad melodies into fewer divisions than they grouped good melodies, $t(34) = -4.54$, $p < .001$ for both comparisons. To explain this latter result, remember that nonmusicians often segmented the notes by the direction of their stem. To check whether this strategy alone would differentiate the good from the bad melodies, each melody was divided into groups by direction of stem, and the number of groupings was recorded. Indeed, good melodies averaged 3.66 groups per melody, and bad melodies averaged 2.78. Thus, nonmusicians, following this strategy, would group good melodies into more chunks than bad melodies. The fact that musicians grouped good melodies into significantly fewer chunks than bad melodies in spite of an easy visual strategy to do otherwise suggests the importance for them of higher musical structures in comprehending these melodies.

To further indicate how much nonmusicians were following the rule for segmentation, the objective number of visual groupings by stem direction was correlated with the mean number of groupings given for each melody. For good melodies, the correlation was .81 for nonmusicians but only .24 for musicians. This shows the high reliance the nonmusicians were placing on the visual strategy as well as the independence from that strategy by the musicians.

To find out if the average number of groupings in a melody was related to its recall, a set of rank order correlations similar to that done with the consistency measures was performed between recall performance on, (in Experiments 1 and 2) and number of groupings in, each melody. Once again, no relation was found between recall performance and number of groupings for nonmusicians (see Table 2). Although the stem direction strategy helped them group the melodies, the nonmusicians were unable to exploit that strategy in recall.

Table 2. Spearman rank order correlations between consistency and grouping scores for each melody and memory for that melody in Experiments 1 and 2

Group	Recall conditions Delayed[a]	 Immediate[b]
	Consistency	
Musicians	.14	.36*
Nonmusicians	.30	–.14
	Number of groupings	
Musicians	–.34*	–.45**
Nonmusicians	–.24	–.01

Note: Degrees of freedom = 34.

[a] Retention interval of 15 sec from Experiment 1.

[b] Immediate recall from Experiment 2.

*$p < .05$.

**$p < .01$.

For musicians, a significant negative correlation was found between number of groupings and performance in the 15-sec unfilled condition, and a highly significant negative correlation was found between number of groupings and performance in immediate recall. Not only did the musicians group the melodies in a more abstract way than the nonmusicians did, but they could take advantage of these imposed groupings in recall.

DISCUSSION

Contrary to prediction, musicians did not group good melodies more consistently than they grouped bad melodies. In fact, they tended to group both kinds of melodies idiosyncratically. Why might musicians vary in segmenting the good melodies? One likely explanation is the impoverished notation of our stimulus melodies. In fully notated music, the particular grouping a composer desires is indicated by metrical information, such as bar lines and time signatures. Since these melodies contained no rhythmic or other metrical divisions, musicians were free to imagine rhythms by which two or more different grouping schemes made good musical sense.

The unexpected finding that nonmusicians were more consistent than the musicians for bad melodies was explained by their use of a simple visual cue for grouping. The fact that the musicians ignored this cue, presumably in an effort to make musically proper groupings, led to their many solutions to this (for them) ambiguous task.

The average number of groupings assigned to each melody revealed the expected pattern: musicians saw fewer groupings in a good melody than in a bad one. Again, the nonmusicians were shown to have relied on the visual grouping strategy, which happened to have distinguished the good from bad melodies.

Even though two different subject samples were used in these correlations, the procedure showed for musicians both the expected negative relation between the number of chunks in a melody and its recall and the expected positive relation between consistency of chunking a melody and its recall. These relationships are stronger for the immediate recall condition: this finding, together with the results of Experiment 2, suggests the use of music-specific encoding very early in the music reading process. Also, as predicted, nonmusicians showed no evidence of using these strategies.

GENERAL DISCUSSION

Despite representing a pattern laid out in time, notation shares some characteristics with other symbolic diagrams, like chess and circuit drawings, especially in notation's analogue representation of pitch. Can a relation be found between the current study and other work that examines an expert's knowledge representation?

We may suppose that a musician has implicit or explicit knowledge of the "grammar" of his or her musical idiom. A grammar is usually defined as the rules determining admissible and inadmissible constructions in some medium. However, it is difficult to set boundaries on the legality of structures in music. Perhaps the closest analogy to an "illegal utterance" in music would be a passage of music that is inappropriate for a particular style of music. For instance, serial (twelve-tone) music must follow well-defined rules. Nevertheless, a serial passage would sound nonsensical in the middle of a traditional tonal symphony and thereby would violate the grammar for a symphony.

Formalizing the grammar of a language is a difficult enough task. Even more so is describing a grammar for an artistic medium, because of its tolerance for, and even encouragement of, unusual "creative" constructions. Nevertheless, there have been several recent proposals of a grammar for certain classes of music. Sundberg and Lindblom (1976) were able to abstract the generative rules for a set of Swedish folk songs. Using these rules, they created new songs that listeners agreed were typical of that genre. Lehrdahl and Jackendoff (1977) have presented a more complete generative grammar for tonal music, quite explicitly based on models of generative grammar for language.

Thus, there is evidence that there is a well-defined grammar of Western music that both trained musicians and casual listeners of music may have learned implicitly simply by exposure to music. This process may be analogous to the chess expert's "repertoire" of good chess patterns that he or she uses to increase memory performance. When the melodies in these experiments were rated for goodness by nonmusicians in the auditory rating task, the intersubject reliability was a respectable .65 (for musicians, .83). Therefore, the nonmusicians were fairly consistent in distinguishing the good from bad melodies, even though the bad melodies were not blatantly discordant.

In addition to this implicit knowledge, the musician may acquire the grammar more explicitly by several means. First, he or she simply hears more music in social, academic, and professional settings. Second, as a performer, the musician gains a deeper understanding of music both at the microscopic level when practicing individual phrases and at a more global level when fluency in performing the whole piece is reached. This greater amount and depth of exposure to music of a particular style will strengthen expectations about probable melodic, harmonic (chordal), and rhythmic sequences. Third, most trained musicians have learned some formal music theory. Music theory explicitly sets out musical structures and gives names to both the elements of music (such as chords) and the relations between elements. Examples include the language of relations between notes in a key (tonic, dominant) and the names of sequential relations of chords of a particular melody-ending pattern (Plagel cadence). Fourth, musicians learn to depict and read these relationships in the language of musical notation. Part of this learning results from reading and playing many pieces. In addition, music students practice "sight-singing"; that is, singing a written-down melody without instrumental help. They also practice the inverse operation: they hear a melody (or chord progression) and must notate it correctly after only a few hearings.

Because the music reader knows how a good pattern sounds and has practiced associating that sound with its visual representation, he or she can recognize certain notated groupings right away—such as the notes in the first three spaces of a treble staff being an "F major chord." Experiment 3 suggested that when many notes can be grouped into a few well-learned chunks, musicians will remember those passages better (also see Deutsch, 1980). Another ability the music reader develops is generating good guesses about the continuation of a sequence once the initial elements are understood. This top-down strategy predicts the musicians' better recall of the good compared with bad melodies under the following scheme. The notation recognition process probably involves a constant interaction of long-term and working memories. Certain patterns at the beginning of a piece may be directly recognized from long-term memory. Once those patterns are labeled or identified as to their usual function, working memory can generate hypotheses about the next group of notes to be encountered. Approximate confirmation of the hypotheses (probably the usual situation in these simple good melodies) will lead to further routine expectations, while an unexpected development can cause delight or distress, depending on the extent of the deviation and the preferences of the listener. The melody following a good structure will be remembered or recognized because of the accuracy of the hypotheses, since accurate hypotheses are confirmed quickly and do not require radical reevaluations of the musician's mental structure of the melody.

How may these pattern labels and hypotheses about the next patterns be represented in memory? Work by Sloboda (1976b) suggests that the information is not stored in a simple auditory, verbal, or visual form. Rather, a more abstract memory code is implicated. This code would probably record the contour of pitches (Dowling, 1978) and the rhythm of the line (Longuet-Higgins, 1978). Also important is tonality, which in Western music determines what the most appropriate next notes are in a melodic sequence. Tonality, or the functional relation of each note to the home note of the key, has been shown to be both a well-defined mental structure especially for musicians and an important factor in melody recognition (Dowling, 1978; Krumhansl & Shepard, 1979; Krumhansl, 1979). Sloboda (1978b) suggests one scheme that would use tonality to code a note sequence; for example, C-D-E-F-G-D-B-C would be coded as C major (tonic rising scale > dominant, and falling dominant chord > tonic); i.e., using functional relationships. In terms of our experiments, this scheme defines only two "chunks" (the rising and falling portions). A novice musician without awareness of key relations might encode that passage as C (up scale to G > D > B > C), which would increase memory load to four separate groups with no connecting scheme between them. The nonmusician, of course, could only code contour or approximate location of the notes, with no abstract code to help.

Because of the familiarity of notation and the learning of names for these patterns, it is unnecessary to suppose that a musician "hears" the pitches represented by the notation. In fact, many musician subjects in these experiments claimed that they did not, even could not, do so. Analogously, some text readers apparently translate text (also a written representation of an auditory pattern) directly from sight to meaning (Kolers, 1970; Kleiman, 1975). In a real sight-reading situation, actually "hearing" the notation as auditory imagery may cause interference, since musicians are always looking ahead in the score while simultaneously executing the music (Sloboda, 1977).

Notes

This research was supported by a National Science Foundation Graduate Fellowship to Andrea R. Halpern and by National Institute of Mental Health Grant #13950 to Gordon H. Bower. Correspondence and requests for offprints may be sent to Andrea R. Halpern, Psychology Department, Stanford University, Stanford, CA 94305. Received for publication March 31, 1981; revision received July 13, 1981.

References

Charness, N. Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning & Memory,* 1976, *2,* 641-653.

Chase, W. G., & Simon, H. A. The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973.(a)

Chase, W. G., & Simon, H.A. Perception in chess. *Cognitive Psychology,* 1973, *4,* 55-81. (b)

de Groot, A. D. *Thought and choice in chess.* The Hague: Mouton, 1965.

Deutsch, D. The processing of structured and unstructured tonal sequences. *Perception & Psychophysics,* 1980, *28,* 381-389.

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

Egan, D. E., & Schwartz, B. J. Chunking in recall of symbolic drawings. *Memory & Cognition,* 1979, *7,* 149-158.

Frey, P. W., & Adesman, P. Recall memory for visually presented chess positions. *Memory & Cognition,* 1976, *4,* 541-547.

Kleiman, G. M. Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior,* 1975, *14,* 323-339.

Kolers, P. A. Three stages of reading. In H. Levin & J. P. Williams (Ed.), *Basic studies on reading.* New York: Basic Books, 1970.

Krumhansl, C. L. The psychological representation of musical pitch in a tonal context. *Cognitive Psychology,* 1979, *11,* 346-374.

Krumhansl, C. L., & Shepard, R. N. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception & Performance,* 1979, *5,* 579-594.

Lehrdahl, F., & Jackendoff, R. Toward a formal theory of tonal music. *Journal of Music Theory,* 1977, *21,* 111-171.

Longuet-Higgins, H. C. The perception of music. *Interdisciplinary Science Review,* 1978, *3,* 148-156.

Reitman, J. S. Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology,* 1976, *8,* 336-356.

Sloboda, J. A. The effect of item position on the likelihood of identification by inference in prose reading and music reading. *Canadian Journal of Psychology,* 1976, *30,* 228-237. (a)

Sloboda, J. A. Visual perception of musical notation: Registering pitch symbols in memory. *Quarterly Journal of Experimental Psychology,* 1976, *28,* 1-16. (b)

Sloboda, J. A. Phrase units as determinants of visual processing in music reading. *British Journal of Psychology,* 1977, *68,* 117-124.

Sloboda, J. A. Perception of contour in music reading. *Perception,* 1978, *7,* 323-331. (a)

Sloboda, J. A. The psychology of music reading. *Psychology of Music,* 1978, *6,* 3-20. (b)

Sundberg, J., & Lindblom, B. Generative theories in language and music descriptions. *Cognition,* 1976, *4,* 99-122.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.