Long-term memory for temporal structure: Evidence from the identification of well-known and novel songs

MATTHEW D. SCHULKIND

Duke University, Durham, North Carolina

In three experiments, long-term memory for temporal structure was examined by having participants identify both well-known (e.g., "I've Been Working on the Railroad") and novel songs. The target songs were subjected to a number of rhythmic alterations, to assess the importance of four critical features of identification performance. The four critical features were meter, phrasing, rhythmic contour (ordinal scaling of note durations), and the ratio of successive durations. In contrast with previous work, the unaltered version of each song was identified significantly better than any altered version. This indicates that rhythm is stored in long-term memory. Furthermore, the results demonstrated that all four critical features play a role in the identification of songs. These results held for both well-known and novel tunes.

Many studies have shown that the temporal structure of an event influences cognitive performance (Bekarian & Bowers, 1983; Boltz, 1992, 1995; Bower & Winzenz, 1970; Dowling, 1973; Rubin, 1977). For example, Dowling (1973) has shown that recognition memory for a musical sequence is better when it is phrased the same way at learning and at test; this result is very similar to the work of Bower and Winzenz with memory for digit strings. Unlike previous work, which was directed at understanding how temporal manipulations affect cognitive performance, the present paper focuses on long-term memory for temporal structure itself. This issue was addressed by examining what types of rhythmic manipulations affect identification of previously learned pitch patterns. Musical rhythm was chosen as the object of study because, unlike lists of words or digits, timing is an integral part of music (Jones & Ralston, 1991). Before outlining specific hypotheses regarding the nature of long term-memory for rhythm, a brief overview of the temporal structure of music is provided, to help musically naive readers better understand the experiments that follow.

The most basic level of temporal structure in music is called the *beat*. Beats occur at regular intervals, thereby creating a steady pulse train that underlies most Western music. When we dance, assuming that we are reasonably

competent, we are keenly aware of the beat. By coordinating our movements with the beat, we move in synchrony with it. Some beats in the pulse train tend to stand out or to sound accented, relative to others. The grouping of accented and unaccented (sometimes referred to as strong and weak) beats is called the *meter*. In most Western music, strong beats tend to recur at regular, fixed intervals, which allows the listener to predict when they are likely to occur (Jones & Boltz, 1989). Although the concept of meter may be difficult for nonmusicians to grasp, most listeners are sensitive to the meter of a piece of music. For example, the contrast between the long, lilting sound of a waltz and the more rigid sound of a march results from the different types of meters used by these musical genres. Also, when people clap their hands or tap their feet along with a melody, the claps and/or taps tend to fall on strong, as opposed to weak, beats.

Meter is also important because it constrains the *rhythm* of a piece of music. For the purposes of this paper, rhythm will be defined as the serial pattern of variable note durations in a melody. Although a complete definition of the term rhythm is difficult to specify, its effect on the perception of musical sequences is easy to illustrate. The seriousness of "Pomp and Circumstance" is conveyed via its long, steadily held notes. Imagine how this melody might sound if all of those long notes were broken down into three or four shorter notes. "Pomp and Circumstance" begins to sound like a Sousa march! Rhythm is a more complex form of organization than meter, because it can vary more freely. For example, a composer can fill a simple two-beat pattern with any of the following: one note that is two beats long, two notes that are each one beat long, or one note that is one beat followed by two notes that are each one-half beat. The meter influences the choice of the composer, however, because in Western music, longer notes typically start on stronger beats, whereas

The author thanks Tamara Rahhal, David Rubin, and Marilyn Boltz for helpful comments on an earlier draft. The quality of the paper was greatly improved by the advice of the reviewers, Mari Jones, Emmanuel Bigand, and Robert Lorch. The author also appreciates the contribution of Andrea Halpern, who provided helpful comments regarding the stimuli used in Experiment 2 and made her stimuli available for Experiment 3. Scott Lipscomb, who composed the novel stimuli used in the final experiment, is also thanked. Correspondence concerning this article should be addressed to M. D. Schulkind, Department of Psychology, University of Illinois, 603 E. Daniel Street, Champaign, IL 61820 (e-mail: mschulki@psych.uiuc.edu).

shorter notes typically start on weaker beats. Thus, rhythm and meter are highly dependent on one another, and consequently, some of the rhythmic alterations employed in the current experiments had a significant effect on the meter, as well as on the rhythm. These effects will be noted where relevant.

The present paper was designed to investigate the nature of long-term memory for the temporal structure of music. Some pioneering work in music cognition suggests that listeners retain little temporal information over long retention intervals. This hypothesis, which will be called the *independence* hypothesis (i.e., memory is independent of temporal structure), is supported by data reported by White (1960). He found that recognition memory for standard versions of familiar tunes was not significantly better than memory for rhythmically altered versions in which every note was played for the same duration (isochronous rhythm). In other words, altering the rhythm of a familiar tune had little effect on identification. It is important to note, however, that the null finding reported by White may have resulted from ceiling effects; mean percent correct for the standard and isochronous versions were 94% and 88%, respectively. Recent neuropsychological data also support the independence hypothesis. Peretz and Kolinsky (1993) report data from a patient suffering from bilateral temporal lobe lesions who performed normally when required to make rhythmic discriminations, but at chance levels when required to make melodic discriminations.

However, other research leads one to question the validity of the independence hypothesis. For example, in contrast to White (1960), Hebert and Peretz (1997) found that isochronous versions of familiar melodies were not identified nearly as well as standard versions. Although these data suggest that rhythm is stored in long-term memory, they do not address what aspects of rhythm facilitate identification. One possibility can be drawn from the work of Dowling (1973; see also Deutsch, 1980), who found that recognition memory for 20-tone melodic sequences was significantly worse if the phrasing of the pitch pattern was altered between learning and test. These data run counter to the notion that rhythm plays a minor role in memory for music. Furthermore, they suggest that phrasing is an aspect of temporal structure that listeners retain over long retention intervals. That is, on the basis of Dowling's (1973) work, one might argue that the long-term representation of music consists of groups of elements with little information regarding their individual durations (i.e., Notes 1, 2, and 3 go together; Notes 4, 5, and 6 go together, etc.). Thus, one might predict that altering the durations of individual notes will not affect identification unless the phrasing of the original is altered.

Jones (1987; Jones & Boltz, 1989), however, has argued that phrasing is neither the only nor the most important function of rhythm. Rather, she emphasizes rhythm's ability to guide attention to important pitch elements and argues that the *relationship* between pitch and rhythm is crucial for understanding how rhythm affects memory

and perception of music (see Jones & Boltz, 1989, for a discussion). The relative length of a tone gives the listener a rough estimate of its importance in the evolving melody. Therefore, although Jones would agree that tones located at the ends of phrases are important (because of their relative length), she would argue that the timing of tones within a phrase is important, as well. Her view also implies that regular meters are important because they help the listener to predict when important events will occur. Evidence in support of Jones' framework was reported by Jones, Summerell, and Marshburn (1987), who found that recognition memory for 9-tone melodic sequences was reduced when the rhythm, but not the phrasing, was altered at test. Jones and Ralston (1991) reported similar data, using longer (18-tone) pitch patterns and rhythms that more tightly controlled the meter.

Although these results suggest that listeners retain some information about the durations of individual notes within phrases, they do not specify how much detail is preserved. For example, listeners may retain little more than a local, ordinal scaling of relative note durations—such as Note 1 is longer than Note 2, 2 is shorter than 3, 3 is longer than 4, and so forth. The ordinal ranking of note durations will be called the *rhythmic contour*, because it is analogous to another term from the music cognition literature. *Melodic contour* refers to the ordinal ranking of relative pitch frequency (i.e., Note 1 is higher than Note 2, 2 is lower than 3, 3 is higher than 4, etc.). Dowling (1978) has argued that memory for melodic contour constitutes one of the basic building blocks of memory for pitch patterns. Thus, it is reasonable to predict that rhythmic contour may constitute the basic building block of memory for temporal structure.

Alternatively, one could predict that listeners retain the ratios of successive note durations—such as Note 1 is twice as long as Note 2, 2 is half as long as 3, 3 is three times longer than 4, and so forth. This hypothesis also has a parallel in the pitch domain. Pitch interval refers to the specific distance between notes. For example, in the key of C Major, a leap from C to E represents an interval of two scale degrees (C–D–E); a leap from C to G (C–D–E–F–G) represents an interval of four scale degrees. Research with pitch has shown that memory for interval information improves with the familiarity of the music, whereas memory for contour information does not (Dowling & Bartlett, 1981; Edworthy, 1982, 1985). Thus, maintaining the rhythmic contour, but not the ratio of successive note durations, might have a particularly strong effect on well-known music.

Finally, it is possible that listeners store the actual lengths of notes in long-term memory (Note 1 is 400 msec, 2 is 200 msec, 3 is 600 msec). This prediction implies that listeners not only store the ratios of note durations, but the tempo, or speed, of the melody as well. Levitin and Cook (1996) have shown that listeners can accurately reproduce the tempi of popular songs that are associated with a single, canonical version. Furthermore, Halpern (1988) has shown that the imagined tempi of

songs without canonical versions are consistent across participants. This indicates that listeners generally agree on what is an appropriate tempo for a given song. However, the present paper cannot address whether listeners store information about the specific durations of notes, because few of the experimental songs are associated with a canonical version; that is, the participants have heard many different versions performed at many different tempi.

On the basis of this review, the present paper posits that the temporal structure of music consists of four critical features: meter, phrasing, rhythmic contour, and the ratio of successive note durations. Furthermore, and in contrast with the independence hypothesis, it holds that that each feature is stored in long-term memory and, consequently, contributes to the identification of melodies. This alternative account will be labeled the *interdependence* hypothesis, to reflect the fact that the four critical features of temporal structure are highly interdependent. For example, it was difficult to create a condition that maintained the rhythmic contour, but not either the phrasing or a regular meter. Therefore, no attempt was made to construct completely orthogonal treatment conditions that would isolate the effect of each individual feature. Rather, the temporal manipulations that were employed in the experiments that follow were designed to vary the number of critical features that were altered from the original or standard version. The independence hypothesis predicts no effect of these manipulations, whereas the interdependence hypothesis predicts that identification

performance will increase directly with the number of unaltered critical features.

EXPERIMENT 1

The null effect reported by White (1960) suggests that little temporal information is stored in long-term memory or, at the very least, that temporal information plays a minor role in the identification of melodies. This conclusion was contradicted by Hebert and Peretz (1997). The aim of Experiment 1 was to determine which of these results could be replicated. The participants identified 20 well-known songs (e.g., "I've Been Working on the Railroad") that were presented in one of four rhythmic conditions. In this experiment, as well as in those that follow, the pitch pattern was never manipulated. Therefore, the participants were instructed to identify the tunes on the basis of the pitch information alone, because the rhythm information was frequently altered.

Two tests were administered. First, the participants heard and attempted to identify each song. This test, subsequently referred to as the free identification test, should prove more difficult than White's testing procedure and, therefore, should result in more moderate levels of identification performance. After hearing each song once, the entire list was repeated in a new random order. For the second hearing, the participants chose their response from a list containing an equal number of target and foil song title cues. Compared with the free identification test, the cued identification test more closely matched White's pro-

Standard

I've been work- ing on the rail-road, All the live-long day!

One-note

Iso

Random

Figure 1. The first two lines of "I've Been Working on the Railroad" in the four rhythmic conditions used in Experiment 1. The tempo for the standard, one-note, and random conditions was 112 quarter notes per min. The tempo for the isochronous (Iso) condition was 100 quarter notes per min.

cedure, in that it incorporated multiple exposures to the stimuli and provided the participants with response choices.

Figure 1 shows the first two lines of all four rhythmic conditions, using the pitch pattern of "I've Been Working on the Railroad." In the standard condition, the rhythm was left unaltered. In the isochronous (Iso) condition, each note was set to the same duration; note that the standard and Iso conditions were the ones used both by White (1960) and by Hebert and Peretz (1997). In the one-note condition, the original rhythmic pattern was shifted by one serial position: the second note took the duration of the first note, the third note took the duration of the second note, and so forth (the first note took the duration of the last note). Finally, in the random condition, the durations used in the standard condition were randomly assigned to serial positions.

The experimental conditions can be ordered in terms of how much temporal information each maintained relative to the unaltered standard version (see Table 1). The one-note condition preserved the ratio of successive durations and rhythmic contour and also included a regular meter. However, the phrasing of the standard was altered, because the rhythmic groups were shifted, relative to the pitch pattern. The Iso condition possessed a regular, although slightly altered, meter but completely altered the phrasing, rhythmic contour, and ratio of note durations. Finally, the random condition altered the ratio of successive durations, rhythmic contour, and phrasing and also included an irregular meter.

According to the independence hypothesis, temporal information is not stored in long-term memory. Therefore, identification performance should not vary across the four rhythmic conditions. Alternatively, the interdependence hypothesis holds that temporal structure consists of a constellation of features, all of which contribute to memory. Thus, as can be seen in Table 1, the interdependence hypothesis predicts that identification performance should be ordered from highest to lowest as follows: standard (preserved all four features), one-note (preserved three features), Iso (preserved one feature), and random (preserved no features).

Method

Participants. The experiment used 40 participants (17 undergraduates and 23 graduate students) drawn from the Duke University community. Undergraduates either volunteered or received par-

tial credit toward a course requirement; all the graduate students were volunteers. No participants were music majors or graduate students in music.

Design. The experiment was a one-way within-subjects design. Rhythm (standard, one-note, random, Iso) was the lone independent variable.

Materials. Twenty songs were selected from a corpus generated by a group of 80 undergraduates, who were instructed to write down a list of songs that they thought "everyone would know." A separate group of 97 undergraduates rated how well they knew these songs. The 20 target songs that follow were selected because a majority of the pilot group claimed that they "easily knew most of the words" to the song: "Amazing Grace," "America, the Beautiful," "Clementine (Oh, My Darling)," "Do-Re-Mi (Do a Deer)," "Frosty, the Snowman," "If I Only Had a Brain," "I've Been Working on the Railroad," "Joy to the World," "Lean on Me," "Puff, the Magic Dragon," "The Rose," "Rudolph, the Red-Nosed Reindeer," "Santa Claus Is Coming to Town," "Silent Night," "This Land Is Your Land," "Tomorrow," "When You Wish Upon a Star," "Yankee Doodle Dandee," "Yellow Submarine," and "You Are My Sunshine." The same information was used to select the foil song titles for the cued identification test. An attempt was made to select song titles from similar genres that a large number of undergraduates claimed to know well (e.g., "White Christmas," "Stand by Me," "Somewhere Over the Rainbow").

The standard version of each song consisted of the melody line from a published sheet music arrangement. Three additional versions of each song were constructed in the manner described previously (see Figure 1). None of these versions altered the pitch of any note, nor was the sequence of notes changed; only the durations were manipulated. The four versions of each song were presented by a Macintosh IIsi computer. To ensure that the tempo of each song was "normal," two moderately trained (5–10 years of musical study) independent judges (one was the author) listened to every standard version and adjusted the tempo up or down, as needed. Slight alterations in tempo were sometimes required in the Iso condition, to ensure that total presentation time was equivalent across presentations.¹

The participants heard one version (standard, one-note, Iso, random) of each of the 20 target songs. The songs were presented to each participant in the same random order. Rhythmic condition was counterbalanced across four orders, so that every participant heard the same number of songs in each condition, and, across the experiment, every song appeared an equal number of times in each condition. A given participant heard the same version of each song for both the free and the cued identification tests.

Procedure. The participants were tested individually and sat behind a screen that separated them from the experimenter and the computers used to present the stimuli. The participants were informed that they would be asked to identify the titles of 20 familiar songs and that, to make the task more difficult, the temporal structure or rhythm of some of the songs would be altered. They were told, therefore, that they would need to base their responses on the pitch pattern alone, because it was never altered. The participants were asked if they understood the nature of the task, and further clarification was given, if necessary.

The free identification test was run first. The participants verbally indicated when they had identified the title but were instructed not to respond until they were certain of the title; this precaution reduced the number of incorrect guesses. The experimenter stopped the music and recorded the participant's response on a separate answer sheet. The participants were not encouraged to guess, although they were allowed to do so, if they had not identified the song before it ended.

The cued identification test used the same basic procedure, with the following changes. The participants were given a list of target and foil song titles from which to choose their response. Because a

Table 1
The Critical Features of Temporal Structure
That Were Maintained by the Four Rhythmic Conditions
Used in Experiment 1

Condition	Ratio of Successive Durations	Rhythmic Contour	Phrasing	Regular Meter
Standard	✓	✓	✓	✓
One-note	✓	✓		✓
Iso				✓
Random				

Note—Check marks indicate features that were maintained.

list that contained 20 target and 20 foil titles was considered too cumbersome to use, the cued test was split into two blocks of 10 songs each. The two blocks consisted of the first and second 10 songs heard in the experiment, respectively; a new random order was created for both blocks. Each target and foil song appeared on only one of the two lists. Prior to the start of each block, the participants were given 1 min to study the list of 10 target and 10 foil song titles for that block. The participants were allowed to use the cue list throughout testing for a given block, although the list for the first block was collected before the list for the second block was distributed. For the cued test, the participants were required to guess if no response was made by the end of a song.

After completion of the cued test, the participants gave a simple yes/no familiarity judgment for each of the target songs. Data for songs that participants claimed they did not know were eliminated from subsequent analyses (2.13% of all the responses).

Results

The following procedures were adopted across all the analyses reported in this paper. Every analysis was conducted twice, once with participants (F_1) and once with songs (F_2) as the unit of analysis. Because none of the data met the assumption of sphericity, the p value for every analysis of variance (ANOVA) was calculated, using the Geisser–Greenhouse adjustment to the degrees of freedom (the degrees of freedom reported in the body of the paper are the unadjusted values). Pairwise comparisons were analyzed with Tukey's honestly significant difference (HSD). Because the data failed to meet the assumption of sphericity, a separate HSD value was calculated for each comparison, using the procedure outlined by Hays (1994); as a result, the individual HSD values are not reported. Alpha for all analyses was set at .05 across all the experiments.

Mean percent correct identification (see Table 2) across the four rhythmic conditions was ordered from highest to lowest, as follows: standard, one-note, Iso, random. This ordering conformed to the pattern predicted by the interdependence hypothesis. Separate one-way within-subjects ANOVAs were run for the free and cued identification tests with rhythm (standard, shifted, Iso, random) as the lone independent variable. A significant main effect of rhythm was observed for both the free [$F_1(3,117) = 27.67$, $MS_e = 0.043$; $F_2(3,57) = 19.46$, $MS_e = 0.032$] and the cued [$F_1(3,117) = 10.18$, $MS_e = 0.025$; $F_2(3,57) = 8.65$, $MS_e = 0.015$] identification tests.

Pairwise comparisons on the free identification data showed the following ordering of mean correct identification from highest to lowest: standard > one-note >

Iso = random, although the one-note condition was not reliably different from either the standard [$t(19) = 2.74$, $MS_e = 0.089$, $.05 < p < .10$] or the Iso [$t(19) = 2.47$, $MS_e = 0.045$, $p > .10$] conditions in the item analysis. The corresponding analysis of the cued identification data showed that the standard condition differed from the Iso and random conditions, but not from the one-note condition. The one-note condition differed from the random condition, but not on the item analysis [$t(19) = 2.22$, $MS_e = 0.049$, $p > .10$]. No other comparison for the cued data was significant.

Finally, as would be predicted by the interdependence hypothesis, trend analyses for both the free [$F_1(1,39) = 65.06$, $MS_e = 0.054$; $F_2(1,19) = 25.37$, $MS_e = 0.072$] and the cued [$F_1(1,39) = 45.88$, $MS_e = 0.016$; $F_2(1,19) = 11.59$, $MS_e = 0.032$] identification data indicated that identification performance decreased monotonically with the number of features preserved in each condition.

Discussion

The results of Experiment 1 lead to different conclusions than those reported by White (1960), even though the cued test data closely matched White's data in terms of performance levels. Mean percent correct for the standard and Iso versions in White's study were 94% and 88%, respectively; the corresponding values for the current experiment were 96.4% and 85.8%, nearly identical to those reported by Hebert and Peretz (1997). Despite the numerical similarity of the present data to those reported by White, these differences were significant in the present experiment because of the increased power of the design ($n = 40$ vs. $n = 9$; 20 vs. 10 songs). Further support for the existence of a significant difference comes from the free identification test, on which performance in the standard condition (85.2%) far exceeded performance in the Iso condition (55.5%). Again, these values are in close agreement with those reported by Hebert and Peretz.

In fact, all three altered rhythms resulted in performances that were significantly worse than that for the standard. This finding contradicts the independence hypothesis, which predicts no effect of rhythmic manipulations. However, the results are consistent with the interdependence hypothesis, in that the identification performance decreased as the number of critical temporal cues decreased. This conclusion was strongly supported by the significant linear trend analysis and the significant effect of rhythm in the ANOVAs.

EXPERIMENT 2

The results of Experiment 1 confirmed that temporal structure is retained over long retention intervals (Hebert & Peretz, 1997) and provided evidence consonant with the interdependence hypothesis. Experiment 2 was designed to provide further evidence for the interdependence hypothesis by controlling the presence or absence of the four key features more systematically than was

Table 2
Mean Percent Correct With Standard Errors
by Rhythmic Condition in Experiment 1

Condition	Testing Condition			
	Free		Cued	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Standard	85.2	2.6	96.4	1.5
One-note	68.1	3.3	88.3	2.4
Iso	55.5	4.1	85.8	3.7
Random	45.2	4.1	77.1	2.7

done in Experiment 1. As was mentioned in the introduction, orthogonal contrasts of meter, phrasing, rhythmic contour, and the ratio of successive durations were difficult to achieve. Therefore, the temporal manipulations used in the present experiment decreased the number of key features one by one.

The participants in the present experiment identified familiar songs presented in one of five ways. The five different conditions used in the experiment are listed in Table 3; the first two lines of each version are displayed in Figure 2, using the pitch pattern of "I've Been Working on the Railroad." As in Experiment 1, the *standard* condition was an unaltered version of the target song; thus, the ratio of successive durations, rhythmic contour, phrasing, and meter were all preserved. The second, *rhythmic contour* (RC) condition changed the durations of individual notes but left the ordinal ranking of consecutive notes largely unchanged. On some occasions, two notes that were of different durations in the standard were of the same duration in the RC condition. However, the relative lengths of notes were never reversed. If A was shorter than B in the standard, then A was either less than or equal to B (but not greater than B) in the RC condition. Thus, the ratio of successive durations was altered, although the rhythmic contour was largely (but not perfectly) maintained. Consequently, the RC condition maintained three key features of the standard (rhythmic contour, phrasing, and a regular meter) but not the successive ratio of note durations.

The third, *marked* condition was similar to the *Iso* condition in Experiment 1, in that the duration of every note within a phrase was identical. However, phrase endings were "marked" by doubling the lengths of notes at the ends of phrases. This manipulation, which was similar to the one employed by Jones et al. (1987), preserved two key features of the standard, phrasing and a regular meter, but maintained neither the ratio of successive durations nor the rhythmic contour. The fourth condition was identical to the *Iso* condition employed in Experiment 1. The *Iso* condition altered the ratio of successive durations, the rhythmic contour, and the phrasing of the standard. As such, it maintained only one key feature of temporal structure, a regular meter.

The fifth and final condition in the experiment also preserved a single feature of temporal structure. It maintained

the phrasing of the standard, but altered the ratio of successive durations and the rhythmic contour and possessed an irregular meter. This condition was called the *scrambled rhythmic contour* (SRC) condition because it was constructed by randomly assigning the note durations used in the RC condition with two restrictions. First, note durations were randomly assigned within the same phrase. Second, the lengths of notes at the ends of phrases were never altered. Thus, the SRC condition preserved the relative phrase lengths of the standard and RC conditions, although the ratio of successive durations and rhythmic contour were substantially altered and an irregular meter was introduced.

The interdependence hypothesis predicts that identification performance will vary directly with the number of features preserved in each condition. Thus, as can be seen in Table 3, the interdependence hypothesis predicts that identification performance should be ordered from greatest to lowest as follows: standard (preserved four features), RC (preserved three features), marked (preserved two features), *Iso* (preserved one feature), and SRC (preserved one feature). The SRC condition was predicted to fall below the *Iso* condition because, even though both maintained one critical feature of the standard, the irregular meter in the SRC condition should make it difficult for listeners to predict when phrase boundaries will occur. In other words, even though the SRC condition maintained the phrasing of the standard, the irregular meter made it difficult for listeners to capitalize on this information.

Method

Participants. The experiment used 40 participants (28 undergraduates and 12 graduate students), who were either summer school or fall semester students at Duke University. Participants were either paid \$5 or received partial credit toward a course requirement. No participants were music majors or graduate students in music.

Design. The experiment was a one-way within-subjects design. Rhythm (standard, RC, SRC, marked, *Iso*) was the lone independent variable.

Materials. Experiment 2 used the same songs as those in Experiment 1. The target melodies were played in a new random order and were arranged into five counterbalance orders generated by the same rules as those used in Experiment 1. Figure 2 depicts the first two lines of all five versions of "I've Been Working on the Railroad."

Procedure. The responses were collected using the procedure for the free identification test in Experiment 1. Cued test data were not collected, because the main purpose of this test in Experiment 1 was to mimic White's (1960) procedure. The same postexperimental information was collected. Data were again eliminated if a participant claimed a particular song was unfamiliar (1.8% of total responses).

Results

Mean percent correct identification (see Table 4) across the five rhythmic conditions was ordered from highest to lowest as follows: standard, RC, marked, *Iso*, SRC. This ordering conformed to the pattern predicted by the interdependence hypothesis. One-way ANOVAs, with rhythm (standard, RC, SRC, marked, *Iso*) as the lone in-

Table 3
The Critical Features of Temporal Structure
That Were Maintained by the Five Rhythmic Conditions
Used in Experiments 2 and 3

Condition	Ratio of Successive Durations	Rhythmic Contour	Phrasing	Regular Meter
Standard	✓	✓	✓	✓
RC		✓	✓	✓
Marked			✓	✓
<i>Iso</i>				✓
SRC			✓	

Note—Check marks indicate features that were preserved.

Figure 2. The first two lines of "I've Been Working on the Railroad" in the five rhythmic conditions used in Experiment 2. The tempo for the standard, rhythmic contour (RC), marked, and scrambled rhythmic contour (SRC) conditions was 112 quarter notes per min. The tempo for the Iso condition was 100 quarter notes per min.

dependent variable, revealed a significant effect of rhythm [$F_1(4,116) = 13.24$, $MS_e = 0.054$; $F_2(4,76) = 11.82$, $MS_e = 0.043$]. Pairwise comparisons showed the following ordering of mean correct identification from highest to lowest: standard > RC > marked = Iso = SRC. However, the RC condition did not differ significantly from either the standard [$t(19) = 2.69$, $MS_e = 0.038$, $p > .10$] or the marked [$t(19) = 2.74$, $MS_e = 0.089$, $p > .10$] condition, in the item analysis.

Finally, as was predicted by the interdependence hypothesis, trend analyses indicated that identification performance decreased linearly with the number of features preserved in each condition [$F_1(1,39) = 60.10$, $MS_e = 0.063$; $F_2(1,19) = 60.43$, $MS_e = 0.033$].

Discussion

The overall pattern of results clearly favor the interdependence hypothesis over the independence hypothe-

sis. The standard condition again resulted in significantly better performance than did any altered version (although the comparison with the RC condition was only significant for the subject analysis). Furthermore, the ordering of the various conditions matched that predicted by the interdependence hypothesis and was supported by significant linear trend analyses, as well as significant main effects of rhythm in the ANOVAs. The subsidiary pairwise comparisons generally support the ordering of conditions specified by the interdependence hypothesis. However, it should be noted that some comparisons (e.g., among the marked, Iso, and SRC conditions) were not significant, and others (e.g., standard vs. RC) were only significant for the subject analysis. Still, the overall pattern is highly consistent with the interdependence hypothesis.

EXPERIMENT 3

Although the data from Experiments 1 and 2 point toward some firm conclusions regarding the long-term representation of temporal structure in music, the use of well-known melodies may overestimate how much information is stored in long-term memory, because the stimuli were overlearned. Experiment 3 was conducted to replicate the results of Experiment 2, using a matched set of eight novel and eight well-known melodies taken from Bartlett, Halpern, and Dowling (1995). Each novel

Table 4
Mean Percent Correct With Standard Errors
by Rhythmic Condition in Experiment 2

Condition	<i>M</i>	<i>SE</i>
Standard	85.6	3.1
RC	74.0	3.7
Marked	58.5	4.6
Iso	50.2	2.5
SRC	50.0	4.8

melody was a carefully composed derivation based on one of the well-known tunes. Thus, the novel and well-known melodies possessed similar pitch and rhythmic structures, although the origin of the novel tunes was unrecognizable (Bartlett et al., 1995). The rhythmic manipulations and predicted results for Experiment 3 were identical to those outlined in Experiment 2.

Using novel melodies required several changes in methodology. The experiment was conducted in three phases: familiarity, training, and test. The first two phases gave the participants an opportunity to learn the novel melodies; well-known melodies were not presented until the third (test) phase began. During the familiarity phase, the eight novel melodies were played twice in a specified order. The "title" (nos. 1–8) of each novel melody was read aloud before it was heard; the participants did not make any responses during the familiarity phase. During the training phase, the participants heard each novel melody twice in a random order and were given 8 sec to identify the title. Feedback was given after each trial. Four familiarity and four training blocks were conducted in an alternating sequence (familiarity, training, familiarity, training, etc.) prior to the test phase. During the test phase, the participants attempted to identify five versions (standard, RC, SRC, marked, and Iso) of the eight novel and eight familiar melodies; no feedback was given during the test phase. Unlike Experiments 1 and 2, a recognition paradigm was used. The purpose of the experiment was to determine whether the results obtained in Experiment 2 could be replicated on a recognition test and would generalize to novel tunes.

Method

Participants. The experiment used 42 undergraduate volunteers, who received partial credit toward a course requirement. No participants were music majors.

Design. The experiment was a 2×5 within-subjects design. Song familiarity (novel vs. well known) and rhythm (standard, RC, SRC, marked, Iso) were the two independent variables.

Materials. Both the well-known and the novel melodies were selected from a set used by Bartlett et al. (1995) and consisted of between 10 and 17 tones (each novel tune had the same number of tones as its corresponding well-known melody). The eight well-

known melodies were: "America, the Beautiful," "Deck the Halls," "For He's a Jolly Good Fellow," "The National Anthem," "Row, Row, Row Your Boat," "Rudolph, the Red-Nosed Reindeer," "This Land Is Your Land," and "Yankee Doodle." Each of the novel melodies was a carefully composed derivation of one of the well-known tunes that was constructed by permuting the rhythmic and pitch intervals of the well-known tune (see Bartlett et al., 1995, for more detail on the construction and piloting of the novel tunes). Figure 3 displays the standard version of "The National Anthem," along with its corresponding novel tune.

The well-known melodies were all played in the key of C Major. The tempo was adjusted until two independent raters (including the experimenter) agreed on a suitable tempo. During the test phase, all the novel melodies were played at 90 beats per min in the key of C Major. However, during the familiarity and training phases, novel melodies were played in two keys (C and G Major) and two tempi (90 and 112 beats per min). Multiple versions of the novel melodies were used to encourage the participants to discriminate melodies on the basis of their pitch patterns as a whole, rather than their individual features, such as "Melody No. 3 has the highest tone." The orthogonal combination of key and tempo resulted in four versions of each novel melody. All four versions were randomly assigned and used equally often throughout the familiarity and training phases (to be described fully in the Procedure section).

The rhythmic transformations used in the test phase were constructed in the same manner as those in Experiment 2. As in previous experiments, all five versions of the 16 melodies were digitized onto a Macintosh IIsx computer, using the Songworks software package. The computer files were transferred to audio tape for the running of the experiment.

Procedure. The participants were seated at a table in a large room and were tested either individually or in groups of up to five. A stereo was placed in the front of the room, facing away from the participants. Each participant was given a sheet of paper containing the "titles" (nos. 1–8) of the novel melodies. The participants were informed that, initially, their task would be to learn a set of novel melodies. After learning the melodies, their ability to identify rhythmically transformed versions of these melodies would be tested. The first phase of the experiment was called the familiarity phase. The participants listened to one version of each novel song in a forward order (1–8), followed by a different version in the reverse order (8–1). Successive melodies were separated by 4 sec of silence. The experimenter called out the number of each melody before it was played. No data were collected during the familiarity phase.

The next phase of the experiment was the training phase. The participants heard two versions of each novel melody, played in a random order. At the end of each melody, the experimenter stopped the tape, and the participants were given as long as 8 sec to mark

Figure 3. The standard version of "The National Anthem" and its novel variation, used in Experiment 3. Both the well-known and the novel versions of all the melodies used in the experiment originally appeared in Bartlett, Halpern, and Dowling (1995) and were composed by Scott Lipscomb.

their response on an experimenter-provided answer sheet; the experimenter timed the response interval with a hand-held stopwatch. Once the 8 sec elapsed (or as soon as all the participants recorded a response), the experimenter read the correct answer and restarted the tape. Four familiarity and four training phases were conducted in an alternating sequence.

After the last training phase, the participants were given a new sheet of paper that contained the titles of the both the novel (nos. 1–8) and the well-known songs. The interval between the beginning of the test phase and the final training phrase was approximately 5 min; this time was required to pass out the materials and to read the instructions for the test phase. The participants were given the same instructions as in the test phases of the first two experiments, except that they were told to record their response (a number in the case of the novel tunes or a two-letter abbreviation in the case of the well-known tunes) on the answer sheet. The procedure for collecting responses was identical to that used in the training phase, except that no feedback was given. The test phase consisted of 80 trials (16 songs $\times$ 5 rhythms) that were arranged randomly, so that no song was repeated until all 16 songs had been heard and no rhythm was repeated until all five rhythms had been heard. Two presentation orders (one the reverse of the other) were used to account for possible fatigue effects. Preliminary analysis indicated that performance did not differ for the two presentation orders; thus, this variable was dropped from the analyses reported in the Results section.

Results

Two participants were dropped from the study because their performance during the training phase fell 2.5 standard deviations below the mean. Mean percent correct was tabulated separately for the training and test phases. Identification of the novel melodies improved steadily across the four training sessions. Mean percent correct (and standard errors) for sessions 1, 2, 3, and 4 were 45.8 (2.5), 63.8 (2.8), 73.1 (2.8), and 79.4 (2.4), respectively.

The test phase data are displayed in Table 5. As in Experiment 2, mean percent correct identification across the five rhythmic conditions was ordered from highest to lowest as follows: standard, RC, marked, Iso, SRC. This ordering conformed to the pattern predicted by the interdependence hypothesis.

The data were submitted to a 2×5 within-subjects ANOVA; song familiarity (well known vs. novel) and rhythm (standard, RC, SRC, marked, Iso) were the two independent variables. The analysis revealed significant main effects of both rhythm [$F_1(4,156) = 56.95$, $MS_e = 0.015$; $F_2(4,56) = 12.18$, $MS_e = 0.043$] and familiarity [$F_1(1,39) = 88.48$, $MS_e = 0.086$; $F_2(1,14) = 13.26$, $MS_e = 0.115$]. The two-way interaction was significant

for the subjects analysis [$F_1(4,156) = 3.11$, $MS_e = 0.015$] but did not approach significance for the item analysis [$F_2(4,56) < 1$, $MS_e = 0.043$, $p < .63$].

Although the subject analysis did yield a significant interaction effect, the reliability of this finding is dubious, for several reasons. First, it appears in the subject analysis but not in the item analysis. Second, when separate one-way within-subjects ANOVAs were conducted for each level of familiarity, with rhythm as the lone independent variable, both analyses yielded significant effects of rhythm [well known, $F_1(4,156) = 40.00$, $MS_e = 0.014$; novel, $F_1(4,156) = 21.74$, $MS_e = 0.016$]. Thus, the interaction term did not arise because rhythm had a significant effect at one level of familiarity but not at the other. Finally, the familiar and novel melodies both yielded the same ordering of conditions.

Therefore, pairwise comparisons were conducted for the data, averaged over familiarity. The results of these comparisons were very different for the item and subjects analyses. For the subject analysis, all five rhythmic conditions were significantly different from one another. However, for the item analysis, the standard condition differed from every altered rhythm, and the RC and SRC conditions differed from one another. No other differences were significant (all $ts < 2.39$, $ps > .10$).

Finally, as was predicted by the interdependence hypothesis, trend analyses revealed that identification performance decreased linearly with the number of features preserved by each condition. The results of these analyses were significant for both familiar [$F_1(1,39) = 63.26$, $MS_e = 0.035$; $F_2(1,7) = 22.74$, $MS_e = 0.01$] and novel [$F_1(1,39) = 69.51$, $MS_e = 0.019$; $F_2(1,7) = 22.19$, $MS_e = 0.016$] tunes considered separately, as well as for the data collapsed across familiarity [$F_1(1,39) = 171.59$, $MS_e = 0.01$; $F_2(1,15) = 49.06$, $MS_e = 0.012$].

Discussion

The results of Experiment 3 are very similar to those of Experiment 2 and provide further support for the interdependence hypothesis. Although the results of the subsidiary pairwise comparisons differed across the subject and item analyses, the most crucial comparisons for the interdependence hypotheses—those comparing the standard condition with the four altered conditions—were significant for both sets of analyses. The ordering of conditions matched the predicted pattern and was confirmed by significant linear trends for both the familiar and the novel tunes. Thus, the results again reflected the fact that identification performance increased directly with the number of critical features that were present in a target melody.

In addition, the results of Experiment 2 generalized to novel melodies. This finding indicates that (1) the critical features of temporal structure are abstracted relatively early in the learning process, (2) the results of the first two experiments were not artifacts based on using highly overlearned melodies, and (3) temporal structure is an important component of long-term memory for music.

Table 5
Mean Percent Correct With Standard Errors
by Rhythmic Condition and Familiarity in Experiment 3

Condition	Familiarity			
	Well Known		Novel	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Standard	94.7	1.2	65.3	3.1
RC	89.1	1.9	54.7	3.0
Marked	78.1	2.7	52.5	3.6
Iso	70.6	2.5	49.4	3.4
SRC	66.6	3.2	39.4	3.2

Furthermore, the results also generalized to a recognition paradigm; this indicates that the results of Experiments 1 and 2 were not idiosyncratic to the cued or free identification testing procedures.

GENERAL DISCUSSION

Three experiments were conducted to determine the nature of long-term memory for temporal information in music. Early work in the field (White, 1960) suggested that temporal information plays a minor role in memory. None of the data reported in this paper were consistent with this hypothesis. In fact, every rhythmic manipulation used in the experiments resulted in a decrement in identification performance, relative to an unaltered, standard version. Thus, in accordance with the results of Hebert and Peretz (1997), rhythm clearly plays an important role in memory for music. The present paper extends these results by providing insight into what temporal features are stored in long-term memory. The results were consistent with the idea that the temporal structure of music consists of four interdependent features: meter, phrasing, rhythmic contour, and the successive ratio of note durations. As each of these features was removed from the standard version, identification performance decreased. In other words, the more features that were preserved by a given rhythmic manipulation, the more likely listeners were to identify the song.

Some individual findings bear special attention. For example, identification performance in the one-note condition was significantly worse than in the standard condition, even though both conditions used the same rhythmic pattern. The only difference between the two conditions was that the one-note condition shifted the relationship of the rhythm and pitch patterns by one serial position. The significant decrement in identification performance that resulted from this change suggests that the relationship between the pitch and the rhythm of a melody is crucial for memory and perception. This interpretation is consistent with Jones's (1987) conception of rhythm.

Jones's (1987) framework for understanding rhythm also helps explain why the SRC condition led to the worst performance in Experiments 2 and 3, even though it reproduced the phrasing of the standard. Jones argued that one of the most important aspects of temporal structure is its ability to guide attention to important elements in the melody. Regular meters allow the listener to anticipate when important elements will occur. Thus, even though the SRC condition reproduced the phrasing of the standard, the meter made it difficult for listeners to predict when phrase boundaries would occur. It should also be noted that the random condition, the only other condition that possessed an irregular meter, resulted in the lowest performance in Experiment 1. Thus, meter and rhythm constrain one another, an idea echoed by the work of Monahan and Carterette (1985), who showed that meter influences the perceived similarity of rhythms. More spe-

cifically, the data support Jones's contention that regular temporal structures support learning and memory by helping listeners predict when melodic elements will occur.

One final issue that must be addressed is the discrepancies between the subject and the item analyses. The results of the most critical analyses, those for the linear trend and for the main effect of rhythm in the ANOVAs, were consistent across both sets of analyses. However, there were several discrepancies in the pairwise comparisons between individual conditions, especially in Experiment 3. In general, more significant differences were obtained by the subject analyses. This pattern suggests that the effects of the various rhythmic manipulations differed across target songs. One explanation for the discrepancies between the subject and the item analyses is that the various temporal manipulations might have had different effects on the different pitch patterns. The results of the present paper are consistent with earlier work (Jones & Ralston, 1991; Jones et al., 1987) showing that pitch patterns sound different when combined with different rhythms. However, the data do not provide a means for predicting what types of pitch patterns "go better" with what types of rhythms. This represents a potentially fruitful area for future research.

In sum, the three experiments reported in the present paper help illuminate the nature of long-term memory for temporal structure. First, they show that, contrary to previous work (White, 1960), temporal information is maintained in long-term memory (see, also, Hebert & Peretz, 1997). More specifically, they demonstrate that the information abstracted from musical stimuli includes the ratio of successive note durations, rhythmic contour, phrasing, and some representation of meter. When combined with the results of other papers looking at the effect of temporal structure on cognitive performance (Bekarian & Bowers, 1983; Boltz, 1992; Bower & Winzenz, 1970; Rubin, 1977), these results suggest that contemporary theories of memory need to address the temporal structure of to-be-remembered events.

REFERENCES

- BARTLETT, J. C., HALPERN, A. R., & DOWLING, W. J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Memory & Cognition*, **23**, 531-546.
- BEKARIAN, D. A., & BOWERS, J. M. (1983). Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 139-145.
- BOLTZ, M. G. (1992). The remembering of auditory event duration. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 938-956.
- BOLTZ, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception & Psychophysics*, **57**, 1080-1096.
- BOWER, G. H., & WINZENZ, D. (1970). Group structure, coding and memory for digit series. *Journal of Experimental Psychology Monographs*, **80**(2, Pt. 2).
- DEUTSCH, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, **28**, 381-389.
- DOWLING, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, **14**, 37-40.
- DOWLING, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, **85**, 341-354.

- DOWLING, W. J., & BARTLETT, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology*, **1**, 30-49.
- EDWORTHY, J. (1982). Pitch and contour in music processing. *Psychomusicology*, **2**, 44-46.
- EDWORTHY, J. (1985). Melodic contour and musical structure. In P. Howell, J. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 169-188). New York: Academic Press.
- HALPERN, A. R. (1988). Perceived and imagined tempos of familiar songs. *Music Perception*, **6**, 193-202.
- HAYS, W. L. (1994). *Statistics* (5th ed.). New York: Holt, Rinehart & Winston.
- HEBERT, S., & PERETZ, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory & Cognition*, **25**, 518-533.
- JONES, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, **41**, 621-634.
- JONES, M. R., & BOLTZ, M. (1989). Dynamic attending and reactions to time. *Psychological Review*, **96**, 459-491.
- JONES, M. R., & RALSTON, J. T. (1991). Some influences of accent structure on melody recognition. *Memory & Cognition*, **19**, 8-20.
- JONES, M. R., SUMMERELL, L., & MARSHBURN, E. (1987). Recognizing melodies: A dynamic interpretation. *Quarterly Journal of Experimental Psychology*, **39A**, 89-121.
- LEVITIN, D. J., & COOK, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, **58**, 927-935.
- MONAHAN, C. B., & CARTERETTE, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception*, **3**, 1-32.
- PERETZ, I., & KOLINSKY, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *Quarterly Journal of Experimental Psychology*, **46A**, 301-325.
- RUBIN, D. C. (1977). Very long-term memory for prose and verse. *Journal of Verbal Learning & Verbal Behavior*, **16**, 611-621.
- WHITE, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology*, **73**, 100-107.

NOTE

1. In all three experiments, presentation times were not exactly equivalent across all conditions; however, two versions of the same song never differed by more than 5%.

(Manuscript received April 7, 1997;
revision accepted for publication September 17, 1998.)