Effects of event structure on retrospective duration judgments

MARILYN G. BOLTZ
Haverford College, Haverford, Pennsylvania

Two experiments examined whether varying degrees of event coherence influence the remembering of an event's actual duration. Relying on musical compositions (Experiment 1) or filmed narratives (Experiment 2) as experimental stimuli, the underlying hierarchy of information within events (i.e., melodic intervals or story elements) was either attentionally highlighted or obscured by placing a varying number of accents (i.e., prolonged notes or commercial breaks) at locations that either coincided or conflicted with grammatical phrase boundaries. When subjects were unexpectedly asked to judge the actual duration of events, through a reproduction (Experiment 1) or verbal estimation (Experiment 2) task, duration estimates became more accurate and less variable when the pattern of accentuation increasingly outlined the events' nested relationships. Conversely, when the events' organization was increasingly obscured through accentuation, time judgments not only became less accurate and more variable, but were consistently overestimated. These findings support a theoretical framework emphasizing the effects of event structure on attending and remembering activities.

Although the ubiquity of time estimation behavior often goes unnoticed, it nonetheless mediates a wide variety of cognitive experience. In some cases, time judgments may be involved in extrapolation tasks, as, for example, in passing or braking for another car, or in temporal-order tasks, as in reproducing the sequence of a given event. Similarly, people are also routinely required to make relative duration judgments, temporal discriminations, and timed motor responses. Of particular interest here is still another situation—one requiring an estimation from memory of an event's actual duration. For example, after attending a lecture, movie, or concert, or after driving to a particular destination, one will often have some sense of the event's total duration. The questions explored in the present research concern, first, the set of factors that determine the accuracy of such judgments, and second, the types of processing mechanisms that may be responsible for these effects. Both relate to more general issues within the time estimation literature and to those theoretical perspectives guiding this research.

The Study of Remembered Duration

Empirical investigations of temporal judgments have relied on a wide variety of tasks and measurement techniques whose applications have been extensively de-
scribed in past review papers (e.g., Allan, 1979; Zakay, 1990). However, one of the most general distinctions found in most time estimation research is that of prospective versus retrospective designs. In part, these refer to the types of instructions administered to subjects during an experimental context, but, in addition, they refer to different cognitive processes assumed to mediate subjects' responses. In particular, subjects within a prospective design know in advance that a time judgment will be required, and so these estimates are assumed to reflect experienced duration or time-in-passing. The drawback of this procedure, of course, is that subjects may base their judgments on conscious time-keeping strategies (e.g., counting, tapping) that are artifactual to the experimental situation. Retrospective designs, on the other hand, reduce this potential concern, in that the subjects do not know that a time judgment will be required until after an event has transpired. Given that estimates are measured in an incidental fashion, this technique is therefore assumed to assess remembered duration.

In reviewing the time estimation literature, it is clear that prospective designs have dominated much of the past research. To a large extent, this bias can be attributed to the psychophysical tradition and questions concerning the relationship between perceived duration and stimulus duration. In addition, these types of designs are also more efficient because unlike the retrospective technique, wherein only one judgment per subject can be obtained, prospective studies enable one to obtain multiple time estimates from an individual subject. Nevertheless, retrospective measurements have become quite popular in recent years as a means with which to test the predictions of various cognitive models of time estimation behavior. These models were initially developed to address

relative duration judgments and those processes involved in comparing the remembered duration of one time span relative to another. In general, while all models concur that those intervals requiring more processing activity will seem relatively longer, they differ in terms of the presumed referent for duration judgments. According to some approaches, relative duration is inferred from the amount of information (i.e., chunks) stored in memory (e.g., Ornstein, 1969), while in others, it is inferred from the number of cognitive changes associated with each interval (e.g., Block, 1985). These theoretical frameworks, in turn, have generated a substantial amount of research on retrospective duration judgments and those factors contributing to the relative (in)accuracy of such behavior.

Judgments of Events' Actual Duration

Although the demands of everyday behavior often require one to compare the duration of one event relative to another, it is also the case that people must frequently recall an event's actual duration from memory. This may, for example, involve estimating how much time was spent performing a given activity or recalling the characteristic duration of an event, such as the boot time of a computer, the average length of a commercial break, or the time between subway stops. Despite the frequency with which these occur, relatively little research has investigated the accuracy with which people are able to remember event durations or those cognitive mechanisms at play. To date, there are two primary areas of research directed toward these issues. The first is concerned with those situations in which people are required to reproduce or verbally estimate the duration of a past activity. For example, subjects may be asked to judge the duration of an experimental session that was spent proofreading a prose passage in the presence or absence of background music (Brown & Stubbs, 1992). In general, results of these studies are quite consistent in showing that as task difficulty increases, time judgments become more inaccurate and unreliable (Brown, 1985; Brown & Stubbs, 1992; Brown & West, 1990; Hicks, Miller, & Kinsbourne, 1976; Zakay, Nitzan, & Glicksohn, 1983). This effect is typically intepreted in terms of resource alloca-

tion theory (Kahneman, 1973; Navon & Gopher, 1979). Given that increased task demands tax limited processing resources, time estimates suffer because less attention is devoted to durational information.

A second class of studies has examined effects due to the intrinsic structure of the event to be judged. This research stems from a theoretical model claiming that people use the inherent structure of an event for attending and remembering activities, and that the degree of event coherence determines how well these activities are performed (Jones & Boltz, 1989). The term *event coherence* derives from the idea that all natural events contain an array of both temporal and nontemporal information that varies in the extent to which they are interrelated. In many events, this relationship is a highly lawful one that can often be described in terms of a hierarchical scheme. For example, as seen in Figure 1, most Western music can be characterized as a sequence of notes nested within higher order melodic phrases whose ending points are nonarbitrarily marked by tonic triad members. These intervals, denoted as (h) in the figure, consist of the tonic, mediant, and dominant, and function to reaffirm the melody's underlying tonality scheme as well as to mark points of resolution within the melody's unfolding structure (Berry, 1976; Cone, 1968; Kramer, 1982; Meyer, 1956; Piston, 1978). They also contribute to a hierarchical arrangement, in that pairs of melodic phrases are nested within higher order phrases of a longer time span (i.e., 16 beats).

Although the nontemporal, melodic information within this tune is in itself highly rule governed, the accompanying pattern of temporal structure reinforces this arrangement to yield a coherent melody. The temporal structure of music arises from the interplay of meter and rhythm. Meter refers to the referent beat of a melody and the number of beats within successive measures. The melody in Figure 1, for example, is based on a 4/4 meter, in which the second (lower) value of this time signature indicates that a quarter note ($\quarternote$) is the referent beat and the first (upper) value indicates there are 4 beats per measure. Rhythm, on the other hand, refers to the pattern of durational change within a melody. These changes are
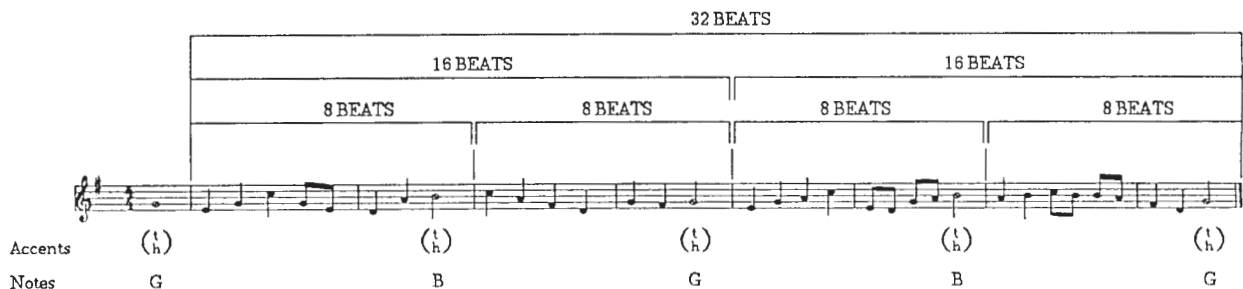

Figure 1. Example of a musical composition displaying a hierarchical arrangement of melodic/temporal structure (excerpt from "There's Music in the Air," by G. Root, adapted from The International Library of Music [1964], *Album of the World's Best Home Songs*, p. 124, by the Editorial Board of The University Society, The University Society, New York). Within this melody's total time span, sequences of notes are nested within melodic phrases marked by tonic triad members (here, the tonic, G, and the mediant, B). These phrase-ending points (h) are attentionally highlighted by temporal accents (t) arising from lengthened notes, and both regularly recur in time to outline higher order phrase relationships.

often characterized in terms of a hierarchical tree structure in which notes that are longer or shorter than the referent beat (half notes, ♩, and eighth notes, ♪, respectively) form nested groupings within the underlying metrical framework (Longuet-Higgins, 1976, 1978; Povel, 1981). Rhythm is psychologically important because pauses and prolonged durations (e.g., half note, ♩) give rise to temporal accents that regularly recur in time and attentionally highlight their corresponding notes. In Figure 1, notice that temporal accents (t) regularly recur after eight beats and coincide with phrase boundaries. A joint accent structure thereby results, in which the melody's temporal information is inextricably intertwined with its nontemporal information, and in which both types of information mutually codefine the melody's duration—the hierarchical nesting of temporal periodicities is defined by different levels of pitch change, and the nesting of pitch changes themselves is defined relative to certain temporal periodicities. This type of structural arrangement generalizes to many other natural events, including speech utterances (Martin, 1972) and conversations (Grosjean, Grosjean, & Lane, 1979; Jaffe & Feldstein, 1972), walking gaits (Carlsoo, 1972; Pierson, 1976), behavioral actions (Bower, Black, & Turner, 1979; Newtson, 1973), and environmental sounds (Boltz, 1992a).

According to the theoretical framework initially offered by Jones and Boltz (1989), and later extended by Boltz (1992a), highly coherent events are adaptive for a wide variety of cognitive activities. Use of the regularly recurrent pattern of temporal accents that stand out and capture one's attention establishes an efficient course for attentional tracking that guides one toward the event's underlying hierarchical organization of structural relations. This in turn facilitates both learning and remembering. Just as temporal accents can be used to guide attending, they can also be used as referential landmarks to recapitulate the event's information. Given that temporal and nontemporal relations codefine the event's structure, this includes the total time span of the event as well as the array of nontemporal information (i.e., notes, words, etc.). Hence, the recapitulation of event structure includes both its temporal and its nontemporal information.

Although many natural events are coherent, others are less so. In the case of music and speech, lack of coherence may arise from a novice performer who fails to place pauses at grammatical phrase boundaries and/or produces temporal accents at irregular intervals. The temporal information within these events is thereby dissociated from the nontemporal information, and a high degree of structural unpredictability results. This, in turn, has a deleterious effect on cognitive processing activities. Given their lack of organization, less coherent events are difficult to track over time and require more attentional effort. Learning and subsequent remembering are also impaired, and people have problems recalling not only the array of nontemporal information, but durational information as well.

Several studies have provided support for these predictions. In one set of experiments by Boltz (1991a),

subjects were asked to recall a set of folk tunes in which temporal accents either coincided with melodic phrase boundaries to yield highly coherent tunes or conflicted with this hierarchical arrangement to yield more incoherent melodies. The results showed that coherent melodies produced not only a higher level of performance but superior recall at phrase ending points, suggesting that these acted as anchors for remembering the retrieval of lower order relations among notes. The recall of incoherent melodies, on the other hand, was quite poor and displayed a high incidence of missing notes and fragmented relations. An analogous set of results has been observed for the remembering of films in which temporal accents in the form of commercial breaks either coincided or conflicted with the hierarchical arrangement of episode boundaries (Boltz, 1992b).

Of more relevance here is the effect of event coherence on the remembering of durational information. In two sets of studies, Boltz (1992a, 1994) developed a new technique that allows one to collect multiple time judgments from subjects in a retrospective paradigm. In this procedure, the subjects were first asked to perform a series of perceptual rating tasks on a set of tonal sequences or environmental sounds that varied in their structural coherence. Some sounds contained a single pitch value that regularly recurred in time (coherent), was sustained in duration (continuous), or was irregularly timed (incoherent). After performing the set of rating tasks for a varying number of trials, the subjects were then unexpectedly asked to either recognize or recall the total duration of those sounds heard earlier. The results indicated that the subjects were in fact able to successfully learn and remember the characteristic duration of events, but that performance was strongly influenced by the events' structure. Overall, the durations of coherent and continuous events were well remembered, and performance increased with greater learning experience. In contrast, the durations of incoherent events were less accurately remembered, regardless of the number of learning trials. These findings parallel those found by Poynter (1983), who asked subjects to reproduce the duration of word lists containing accents in the form of U.S. presidents' names. The results showed that duration estimates were more accurate when accents recurred in a predictable fashion than when they recurred haphazardly.

In summary, the prior research indicates that remembering an event's actual duration depends on both task difficulty and event structure. The two theoretical perspectives guiding these investigations, in fact, are not incompatible with one another. The learning and remembering of coherent events require less attentional effort because their temporal information is inextricably intertwined with their nontemporal information, and the two mutually codefine one another. Hence, these types of events are relatively easy to track over time, and yield a more accurate recapitulation of event structure. Incoherent events, on the other hand, lack structural regularity because their temporal information is dissociated from their nontemporal information. These events therefore

require more attentional effort, and just as the remembering of nontemporal information is impaired, so, too, is the remembering of event duration.

The purpose of the present set of experiments is to extend this previous work by examining some additional ways in which event structure may influence remembered duration. In particular, one prediction stems from the idea that people recapitulate the total time span of an event by using temporal accents as referents for remembering. In coherent events, accents lawfully recur and reliably guide one toward natural phrase boundaries and the hierarchical nesting of event relations. Hence, people are able to well remember both the event's durational information and its nontemporal array. Conversely, incoherent events display either irregular temporal accents or accents that fail to coincide with the event's inherent organization; event relations thereby become less obvious, and this in turn results in poor remembering. Given these assumptions, it is reasonable to expect that variations in the number of temporal accents will differentially affect the remembering of coherent and incoherent events.

For example, consider the melody depicted in Figure 1. Although this tune displays a hierarchical arrangement of melodic relations, this inherent organization could attentionally be either highlighted or obscured to a listener by its accompanying pattern of temporal accents. Relative to an isochronous rhythm in which all notes are of equiduration, those rhythms giving rise to temporal accents (i.e., from a relatively longer half note, ♩) at phrase boundaries will serve to attentionally outline the hierarchical structure to listeners. In addition, while the presence of one accent (after 16 beats) will draw attention to the two higher order melodic phrases, the occurrence of three accents (after every 8 beats) will highlight both higher and lower order nestings of phrase relations. Hence, given that the inherent hierarchical organization of the tune is clearly demarcated, this in turn should facilitate learning. Similarly, the presence of additional accents should establish a more pronounced rhythm for recapitulating the melody's total time span. In sum, it is predicted that a greater number of accents coinciding with phrase boundaries will increase the accuracy of remembered duration.

Conversely, the opposite pattern of results should arise when temporal accents conflict with the inherent organization of melodic structure. Relative to an isochronous rhythm, an increasing number of temporal accents will further obscure the presence of melodic phrase boundaries as well as the relationship between melodic intervals. The tune will therefore appear increasingly fragmented and as a series of disconnected chunks. This not only should disrupt the initial learning of the melody, but should also result in highly unreliable estimates of the event's remembered duration.

EXPERIMENT 1

This predicted interaction was investigated here in a set of two experiments, the first relying on musical se-

quences as experimental stimuli, and the second on filmed narratives. In Experiment 1, the subjects were first asked to rate a series of folk tunes on various perceptual dimensions until they were well learned. These tunes all possessed their own characteristic duration but varied in their accent compatibility and number of temporal accents. Following this initial learning phase, the subjects were given a surprise memory task in which they were asked to recall, via a buttonpress, the total duration of those tunes heard earlier.

Method

Design. The design was a $3 \times 2 \times 2/4$ nested factorial. The subjects listened to a series of six folk tunes that varied in rhythmic accent structure (compatible, incompatible, isochronous) and total number of notes (24, 36). A third manipulation, number of temporal accents assigned to accent-compatible and accent-incompatible melodies (three, six), was nested within a between-subjects variable of melody set (A, B, C, D).

Subjects. Forty-eight subjects from an introductory psychology course at Haverford College participated in the experiment for course credit. Each had normal hearing but no formal training in music. Twelve subjects were randomly assigned to each of the four melody sets.

Stimulus materials. A set of folk tunes from other countries was selected from a music composition book.[1] To ensure that these melodies were unfamiliar to the subjects, a preliminary rating study was conducted with an independent group of 12 experienced musicians; any tune judged even slightly familiar to listeners was eliminated from the experiment. The eight melodies selected as experimental stimuli are depicted in Figure 2. Half of these tunes contain a total of 36 notes, while the remaining contain 24 notes, and within each set, the melodies vary in key (C major, G major, F major, D major) and overall pitch contour (i.e., in terms of the "ups and downs" of pitch changes).

Each of the eight melodies shown in Figure 2 displays a hierarchical melodic structure in which phrase ending points regularly recur after an invariant number of notes and are marked by a tonic triad member (tonic, dominant, or mediant). However, the locations of these phrase ending points vary across the different tunes. For those containing 36 notes, notice that phrase ending points (P) recur after every sixth note in two of the melodies ("Why, Oh Mother," "Look Out, How It's Raining"), but after every ninth note in the remaining two melodies ("Spite," "I'll Have No Other"). Similarly, two of the 24-note melodies display phrase endings after every eighth note ("Fond Regrets," "Lithuanian National Hymn"), or after every sixth note ("The Rabbit and the Hare," "Oberek").

To achieve variations in the melodic/rhythmic compatibility of these melodies as well as a varying number of temporal accents, two types of rhythms were then constructed for tunes of a given length. Within each rhythm, temporal accents were produced with a prolonged note (♩.) that was three times the duration of the preceding context beat (♩). Across the two rhythms, the location of these temporal accents varied. For the 36-note melodies, prolonged notes either occurred on every sixth note (♩♩♩♩♩.) or on every twelfth note (♩♩♩♩♩♩♩♩♩♩♩.) to yield a total of six versus three temporal accents, respectively, within a melody's total time span. As shown in Figure 2, the placement of these temporal accents thereby resulted in a compatible accent structure for melodies displaying 6-note phrases but an incompatible structure for melodies containing 9-note phrases. A similar set of rhythms was created for the 24-note melodies. Here, prolonged notes occurred on every fourth note (♩♩♩.) to yield six temporal accents, or on every eighth note (♩♩♩♩♩♩♩.) to yield three temporal accents. As Figure 2 illustrates, these two rhythms were thereby compatible with the melodic structure of tunes containing 8-note phrases but incom-

Thirty-Six Note Melodies

"Why, Oh Mother" by E. St. Vincent Millay - 6 beat phrases   Poland

(a)

"Look Out, How It's Raining" by L. Speyer - 6 beat phrases   Austria

(b)

"Spite" by A. Mathewson - 9 beat phrases   Italy

(c)

"I'll Have No Other" by J. Mokrejs - 9 beat phrases   Czechia

(d)

Twenty-Four Note Melodies

"Fond Regrets" by E. Markham - 8 beat phrases   The Azores

(e)

"Lithuanian National Hymn" by W. Griffith - 8 beat phrases   Lithuania

(f)

"The Rabbit and the Hare" by G.H. Conkling - 6 beat phrases   Japan

(g)

"Oberek" by A. Guiterman - 6 beat phrases   Poland

(h)

Figure 2. The set of folk tunes used as experimental stimuli in Experiment 1. In the 36-note melodies, tonic triad members marking melodic phrase boundaries (P) occurred either on every 6th note or on every 9th note. Given that temporal accents were subsequently imposed on every 6th or 12th note of these tunes to create a total of six versus three accents, respectively, these thereby resulted in a compatible-accent structure for melodies displaying 6-note phrases (Figures 2a and 2b) but an incompatible structure for melodies containing 9-note phrases (Figures 2c and 2d). A similar set of manipulations was created for the 24-note tunes. Here, temporal accents either occurred on every 4th or every 8th note to yield six versus three accents, respectively. These accent structures were therefore compatible with melodies containing 8-note phrases (Figures 2e and 2f) but incompatible with those containing 6-note phrases (Figures 2g and 2h).

patible with tunes containing 6-note phrases.[2] Finally, isochronous versions of all of the tunes were created, in which all notes were of an equal duration ($\downarrow$), yielding no temporal accents at all.[3]

Across all rhythmic conditions, tonal duration (*on time*) and intertone interval (*off time*) were held constant, such that the value of a quarter note ($\downarrow$) always corresponded to 200 msec on and 100 msec off, and the value of a dotted half note ($\downarrow.$) to 600 msec on and 300 msec off. When these tonal values were applied to melodies that varied in their rhythmic structure and total number of notes, this in turn yielded melodies that varied in their total duration. In particular, the durations of 36-note melodies were 10.7, 12.3, and 14.1 sec when these displayed 0, 3, and 6 temporal accents, respectively. In contrast, the 24-note melodies unfolded over a total time span of 7.1, 8.7, and 10.5 sec when these contained 0, 3, and 6 temporal accents, respectively.

For the initial learning phase of the experiment, the melodies were sorted into four different sets (A, B, C, D), each containing six melodies; the melodies assigned to each set are listed in Table 1. The six melodies within a set were then presented as a block of trials for each of four rating tasks. Melodies within a block were randomly ordered and this randomization varied across the different tasks. Given that the four rating tasks were repeated a second time, there were a total of 48 learning trials—meaning that the subjects heard each of their six tunes on eight different occasions.

Within each of the four melody sets (A, B, C, D), two of the six melodies displayed a compatible accent structure, two an incompatible structure, and two an isochronous rhythm; and for each,

Table 1
The Four Melody Sets of Experiment 1

Rhythm	Number of Notes	Number of Accents	Tune
		Set A	
Compatible	36	3	"Look Out..."
Compatible	24	6	"Fond Regrets"
Incompatible	36	6	"Spite"
Incompatible	24	3	"The Rabbit..."
Isochronous	36		"Why, Oh Mother"
Isochronous	24		"Oberek"
		Set B	
Compatible	36	6	"Look Out..."
Compatible	24	3	"Fond Regrets"
Incompatible	36	3	"Spite"
Incompatible	24	6	"The Rabbit..."
Isochronous	36		"I'll Have..."
Isochronous	24		"Lithuanian..."
		Set C	
Compatible	36	3	"Why, Oh Mother"
Compatible	24	6	"Lithuanian..."
Incompatible	36	6	"I'll Have..."
Incompatible	24	3	"Oberek"
Isochronous	36		"Look Out..."
Isochronous	24		"The Rabbit..."
		Set D	
Compatible	36	6	"Why, Oh Mother"
Compatible	24	3	"Lithuanian..."
Incompatible	36	3	"I'll Have..."
Incompatible	24	6	"Oberek"
Isochronous	36		"Spite"
Isochronous	24		"Fond Regrets"

Note—The six melodies within each set reflect a factorial crossing of rhythmic accent structure (compatible, incompatible, isochronous) and total number of notes (24, 36). However, the particular melody assigned to a given rhythm and the number of temporal accents (3, 6) associated with accent-compatible and accent-incompatible melodies are nested within the four sets.

one was a 36-note tune and the other a 24-note tune. As seen in Table 1, the four sets differed from one another in terms of the particular melody assigned to a given rhythm and the number of accents associated with the compatible/incompatible-accent structures. In Sets A and C, the incompatible and compatible 36-note melodies contained six and three temporal accents, respectively, while the incompatible and compatible 24-note melodies contained three and six accents, respectively. The reverse was true of Sets B and D, in which the (in)compatible 36-note tunes displayed three and six accents and the (in)compatible 24-note tunes displayed six and three accents.[4] Finally, in order to enhance the distinctiveness of melodies, each of the six tunes within a set was assigned a different timbre during presentation (i.e., harp, percussive flute, skool piano, bell pad, breath organ, synthesized strings), and across the four sets, these were evenly distributed among the different rhythmic conditions.

Apparatus. All melodies were constructed and generated with the Midilab software system (Todd, Boltz, & Jones, 1989). During each experimental session, the melodies were presented on line with a Yamaha TX81Z FM tone generator controlled by an IBM AT computer with a Roland MPU-401 MIDI (musical instrument digital interface) unit. Sequences of tones were amplified by a Kenwood KR-4010 receiver and played over Koss-Pro 4AAA Plus headphones at a comfortable listening level (approximately 60 dB).

Procedure. Recorded instructions informed the subjects of pattern-presentation details and task requirements. For the initial learning phase, the subjects were told that the experiment involved a study of perceptual judgments and that they would be asked to rate a set of melodies on four different dimensions. Each rating task was performed separately and the set of instructions for each was administered before the presentation of a block of six trials. The upcoming duration-recall task was not mentioned.

On each trial, a 1-sec warning tone (C_7–C in the seventh octave) preceded the presentation of a melody by 2 sec. Two seconds after the melody's offset, this same tune was presented again. During a 5-sec response period that followed, the subjects were asked to provide a rating from a 10-point scale by circling an appropriate value in their response booklet. For ratings of familiarity (1 = *very familiar*), the subjects were asked to decide whether the tune seemed at all familiar and one they had heard in the past. The second task of pleasantness ratings required subjects to decide whether each melody was pleasing to listen to or aversive in its sound qualities (1 = *very pleasant*). Ratings of predictability directed attention toward the pitch and timing qualities of each melody, with the subjects being asked to judge whether these seemed very regular, such that the melody was easy to track and its future course easy to anticipate (1 = *very predictable*), or unpredictable in nature, such that it was very difficult to follow the tune's ongoing and future course. Finally, ratings of mood required the subjects to determine whether each melody expressed an overall positive emotion, such as happiness or serenity (1 = *very positive*), or a more negative emotion, such as sadness or anger. All of the subjects performed these tasks in the same order and were then asked to perform this set of ratings a second time with a separate response booklet. A previous study by Boltz (1992a) has shown that this degree of learning experience is sufficient for producing relatively accurate levels of duration memory.

Following this initial phase, the subjects were given a surprise recall task for melody duration. On each trial, a 1-sec warning tone (C_7) preceded the first three notes of a melody by 2 sec. At the offset of the third note, the subjects were asked to "mentally continue the duration of the melody until it reaches its original ending point." This ending point was indicated by a hand-held buttonpress that was automatically recorded by the computer. The subjects were asked to recall each of the six melodies from the initial learning phase on three successive blocks of trials, with the six melodies within each block showing a different randomized order.

The subjects were tested in groups of 3–4 individuals, and the duration of an entire experimental session was approximately 50 min.

Results

Data from the rating and duration-recall tasks were analyzed separately through a series of analyses of variance (ANOVAs). For each analysis, a 2 × 2 × 2/4 nested factorial was used to assess the manipulations of accent compatibility (compatible, incompatible), total number of notes (24, 36), and number of temporal accents (3, 6), the latter being nested within the between-subjects variable of melody set (A, B, C, D). An independent set of Dunnett comparisons was used to evaluate the isochronous control melodies relative to these conditions.

Rating data. Table 2 depicts the mean ratings of familiarity, pleasantness, predictability, and mood as a function of the melodies' accent compatibility, number of notes, and number of temporal accents.

For ratings of familiarity, the overall ANOVA revealed nonsignificant effects and indicated that the set of melodies was not recognizable by subjects. ANOVAs also revealed nonsignificant effects for the pleasantness and mood ratings. As seen in Table 2, these sets of means all cluster around the midrange of the scale. The set of predictability ratings, however, does vary as a function of the melodies' accent compatibility [$F(1,46) = 79.4$, $MS_e = 1.58$, $p < .001$]. Relative to the isochronous control sequences, a set of Dunnett comparisons ($MS_e = 1.77$, $p < .05$) indicates that accent compatibility enhances judged regularity, while accent incompatibility reduces a melody's predictability. In addition, there is a significant interaction between accent compatibility and number of temporal accents [$F(1,46) = 9.7$, $MS_e = 2.32$, $p < .01$]. In the compatible-accent conditions of Table 2, notice that melodies containing six accents are judged to be more predictable than those containing three accents. However, when melodies display an incompatible-accent

structure, the reverse is true. Here, melodies containing a greater number of accents are rated as more unpredictable than those containing fewer accents. These differences were confirmed by a set of Tukey post hoc comparisons ($p = .05$) and generalized across both the 24- and 36-note melodies.

In sum, these data indicate that the melodies' familiarity, pleasantness, and mood were comparable across the different experimental conditions. Melodies within the different temporal accent conditions, however, were judged to vary in their degree of internal regularity, which thereby validates the main manipulations of this experiment.

Absolute error scores. In order to assess reproduction accuracy while adjusting for differences in the melodies' total duration, all duration judgments (shown in the first column of Table 3) were transformed into absolute error scores. These scores represent the proportional or percentage difference between a melody's judged and actual duration and are obtained by subtracting a subject's reproduced time from the melody's actual duration, dividing this absolute difference by the melody's actual duration, and multiplying by 100. Several researchers (e.g., Brown, 1985; Michon, 1972) have argued that absolute error is one of the most sensitive measures of reproduction accuracy because it assesses the extent to which subjects' responses vary from an event's actual duration, irrespective of the deviation's sign (i.e., regardless of whether it is an over- or an underreproduction). These values are depicted in the second column of Table 3 as a function of accent compatibility, melody duration, and number of temporal accents.

The most notable finding is that temporal accent structure exerts a significant influence upon reproduction accuracy. Overall, a main effect for accent compatibility indicates that when temporal accents coincide with melodic phrase endings, tune duration is more accurately recalled than when temporal/melodic accent

Table 2
Mean Ratings of Familiarity, Pleasantness, Mood, and Predictability for Melodies in Experiment 1 as a Function of Accent Compatibility, Melody Length, and Number of Temporal Accents

Temporal Accent Structure	Familiarity	Pleasantness	Mood	Predictability
Compatible				
36 notes—6 accents	8.73	5.40	5.68	2.50
24 notes—6 accents	9.01	5.32	5.80	2.54
36 notes—3 accents	8.80	5.12	5.79	3.68
24 notes—3 accents	8.64	5.16	5.97	3.60
Incompatible				
36 notes—6 accents	8.92	5.67	5.93	8.33
24 notes—6 accents	9.10	5.61	5.95	8.21
36 notes—3 accents	8.81	5.70	6.16	7.22
24 notes—3 accents	8.89	5.86	6.06	7.44
Isochronous				
36 notes	8.94	5.91	5.60	5.01
24 notes	8.64	5.99	5.70	4.87

Note—Lower values indicate a higher degree of each perceptual dimension.

Table 3
Mean Performance in the Duration Reproduction Task of Experiment 1 as a Function of
the Melodies' Accent Compatibility, Total Duration, and Number of Temporal Accents

Temporal Accent Structure	Mean Reproductions (sec)		Absolute Error		Standard Deviations Within Subjects (msec)		Ratio Scores	
	M	SD	M	SD	M	SD	M	SD
Compatible								
36 notes—6 accents (14.1 sec)	13.654	.57	3.16	1.58	539	291	.97	.05
24 notes—6 accents (10.5 sec)	10.175	.75	3.10	1.86	551	299	.97	.03
36 notes—3 accents (12.3 sec)	11.685	.97	5.00	4.17	770	299	.95	.08
24 notes—3 accents (8.7 sec)	8.081	1.09	7.12	3.95	782	289	.93	.04
Incompatible								
36 notes—6 accents (14.1 sec)	18.030	6.89	27.87	22.01	2,052	867	1.28	.27
24 notes—6 accents (10.5 sec)	13.727	5.63	30.73	19.05	2,098	831	1.31	.35
36 notes—3 accents (12.3 sec)	14.775	5.74	20.12	9.40	1,769	792	1.20	.20
24 notes—3 accents (8.7 sec)	10.701	5.41	23.00	11.2	1,689	738	1.23	.28
Isochronous								
36 notes (10.7 sec)	11.770	2.45	10.00	7.36	1,033	458	1.10	.09
24 notes (7.1 sec)	7.739	2.86	9.00	7.14	989	492	1.09	.13

structures conflict with one another [$F(1,46) = 158.7$, $MS_e = 100.16$, $p < .001$].

Second, the results indicate that the number of temporal accents exerts a differential effect on conditions of accent compatibility [$F(1,46) = 4.97$, $MS_e = 120.76$, $p < .03$]. Table 3 shows that the accuracy of duration recall significantly improves when an increasing number of temporal accents coincides with melodic phrase structure. Conversely, the opposite effect occurs in the incompatible-accent conditions. Here, an increasing number of temporal accents serves to further obscure the melody's organization and duration recall becomes more inaccurate. These differences were confirmed with a set of Tukey post hoc comparisons ($p < .05$). Finally, a set of Dunnett comparisons ($MS_e = 118.52$, $p < .05$) indicates that the attentional highlighting of melodic phrase structure produces more accurate recall of those melodies than of those containing no temporal accents at all. Performance in the isochronous condition, however, is superior to that of the two incompatible-accent conditions, in which temporal accents obscure the melody's underlying organization of pitch relationships.

In sum, the reproduction of a melody's duration is significantly influenced by the extent to which its hierarchical arrangement of pitch relationships is attentionally highlighted to listeners. This joint effect of accent compatibility and number of temporal accents was a highly robust one that applied to melodies of varying lengths (24- vs. 36-note melodies), as well as to 39 of the 48 subjects in the experiment. By looking at the standard deviations of absolute error scores in Table 3, one can also see that recall accuracy was much less variable across subjects in the compatible-accent conditions than it was in the accent-incompatible ones.

Standard deviations. A second way in which to evaluate the reliability of reproduction performance is to assess the degree of variability not *across* subjects, but *within* individual subjects. That is, each subject reproduced the duration of a given melody on three separate trials, and by determining the standard deviation of these three values, this in turn suggests the extent to which subjects' reproduced times are reliable indicators of duration memory. If the attentional highlighting of a melody's organization does in fact facilitate the recapitulation of a tune's total time span, subjects' responses should not only be more accurate, as shown in the previous analysis, but more consistent with one another in the compatible- versus incompatible-accent conditions. The mean standard deviations of subjects' reproduced times (in milliseconds) are shown in the third column of Table 3.

Overall, the results converge with the absolute error data. Once again, a main effect for accent compatibility is found, indicating that lower variability is associated with accent-compatible structures than with accent-incompatible ones [$F(1,46) = 116.15$, $MS_e = 476,578$, $p < .001$]. In addition, accent compatibility and number of temporal accents significantly interact [$F(1,46) = 18.01$, $MS_e = 175,292$, $p < .003$]. In the accent-compatible conditions, it can be seen that the variability of responses significantly decreases as an increasing number of accents demarcates the hierarchical arrangement of melodic structure. Conversely, as the number of accents increases in the incompatibly structured melodies, response variability becomes greater. These differences among means were confirmed by a set of Tukey post hoc comparisons ($p < .05$). Finally, as revealed by a set of Dunnett comparisons ($MS_e = 246,387$, $p < .05$), the presence of temporal accents serves to increase the reliability of temporal judgments relative to the isochronous melodies, but only when these coincide and do not conflict with melodic structure.

In sum, the temporal accent structure of melodies determines not only how accurately their total time spans will later be recalled, but also the reliability of these judgments. As was true of reproduction accuracy, these effects were relatively robust ones that generalized to 41 of the 48 subjects and to both the 24- and the 36-note melodies.

Ratio scores. A final analysis was conducted to assess the directionality of subjects' time judgments. Recall that the absolute error scores measure reproduction accuracy by ignoring the signed difference between a subject's response and the actual duration of a melody. However, it may be the case that the manipulations of temporal accent structure in this experiment exert a systematic influence upon the direction of temporal judgments such that some conditions are more apt than others to yield over- or underreproductions of duration recall. Hence, a set of ratio scores was calculated for each subject that evaluated reproduced duration relative to the melody's actual duration (RD/AD). Relative to a ratio of 1 (representing no bias), ratios smaller than 1 indicate relative underreproductions, while ratios greater than 1 indicate relative overreproductions. These mean ratio scores are depicted in the fourth column of Table 3.

The results from an overall ANOVA reveal a main effect for accent compatibility [$F(1,46) = 9.03$, $MS_e = .11$, $p < .01$], in which accent-incompatible melodies are judged significantly longer than accent-compatible ones. In addition, accent number and compatibility yield a significant interaction [$F(1,46) = 7.67$, $MS_e = .14$, $p < .01$]. In contrast to the accent-compatible conditions, which show no effect of accent number, the degree of overestimation in accent-incompatible melodies increases as the number of temporal accents increases. As confirmed by a set of Dunnett comparisons ($MS_e = .16$, $p < .05$), this overestimation of a melody's duration is strictly confined to the accent-incompatible conditions. The isochronous and accent-compatible melodies, in fact, both yield ratios that are not significantly different from a value of 1. Finally, it is informative to note that the variability of responses is also significantly greater in the accent-incompatible conditions than in the accent-compatible ones.

Discussion

As a set, these results support the idea that remembered duration is dependent upon event coherence. Those tunes containing temporal accents highlighting the hierarchical organization of melodic structure yield not only higher levels of accuracy but also lower levels of variability. In addition, these effects are enhanced when both higher and lower order nestings of information are increasingly outlined to listeners through additional accents.

The first finding—that predictable time spans are better remembered than less predictable intervals—is one that replicates and extends an earlier set of studies by Boltz (1992a, 1994). In those experiments, subjects listened to a series of auditory patterns or environmental sounds that were much more simplistic than music, in that a single pitch value merely repeated itself with either a regular (isochronous) or an irregular periodicity. Given that listeners typically impose a subjective accent structure on isochronous sequences (Fraisse, 1956; Povel & Okkerman, 1981; Vos, 1977), this in turn provided a retrieval scheme that enabled these event durations to be more accurately remembered than those with unequal timing. In the present experiment, these findings were generalized to musical sequences with a much more complex structure. In contrast to a repetitive pitch value, melodies are composed from a more extensive pitch range that is not only governed by an underlying tonality scheme but that is arranged into melodic phrases that form a hierarchical arrangement. Music also displays an accompanying rhythm in which certain notes are physically accented through pausing or prolonged duration. Although each structural dimension in itself may be quite lawful, as was true in the present experiment, it is the joint relationship between temporal and nontemporal structure that determines overall coherence. This in turn affects remembered duration. When temporal accents reinforce the nestings of melodic relations, duration estimates produce only a 5% error rate; however, when temporal and nontemporal relations conflict with one another, they produce a 25% rate of error. Hence, even though the structural basis for event coherence can vary across different natural events and be determined by physically or subjectively marked hierarchies, it is the predictability of information within a given time span that determines the retrievability of an event's total duration.

Given the inherent hierarchy of nontemporal relations within music, this experiment also allowed assessment of a related prediction—namely, that varying levels of event coherence exert a systematic impact on remembered duration. This effect was in fact revealed through a significant interaction between accent compatibility and number of temporal accents. Relative to melodies containing no temporal accents, recall accuracy increased by 5% as the underlying array of melodic relations was highlighted to listeners through temporal accentuation. In addition, the six-accent condition yielded not only higher accuracy but more reliable responses, both between and within subjects, than the three-accent melodies. On the other hand, a very different pattern of results emerged in the accent-incompatible conditions. Here, a greater number of temporal accents rendered melodic relations less apparent, and this resulted in an increase of judgment error from 21% to 29%, as well as a significant rise in response variability.

It should be noted that the manipulation of accent number in this experiment was confounded with a melody's total duration: those tunes containing six accents were always longer than those containing three and no accents. This confound was one that could not be avoided in the construction of stimulus materials and was, in fact, one reason for including 24-note melodies in the experimental methodology. Given that the 24-note/six-accent melodies displayed a similar total duration to that of the 36-note/no-accent tunes, one could determine whether results were in any way dependent upon the length of a given time span. This was not the case. The absolute error scores adjusted for differences in melody duration and, beyond this, showed that accent number and compatibility exerted the same effects on both the 24- and 36-note melodies. In general, then, these findings illustrate that the effects of event coherence gener-

alize across different melodies with varying note lengths and total durations.

According to the present approach, these overall results are due to the way in which people use natural event structure in attending and remembering. Temporal accents direct attention to certain melodic intervals whose interrelations may or may not be apparent, depending on where accents occur. This in turn determines the extent to which one can successfully track the event and later use temporal accents to reliably recapitulate event relations. This includes both temporal and nontemporal information, and just as variations in event coherence influence the remembering of a melody's sequence of notes (Boltz, 1991a), they also affect the accuracy and reliability of retrospective duration estimates.

The last notable finding of this experiment is a systematic bias observed in the accent-incompatible conditions. Relative to coherent events, incoherent melodies yielded significant overestimates that increased in magnitude with a greater number of accents. One interpretation of this result comes from the relative duration models mentioned earlier (i.e., Block, 1985; Ornstein, 1969). Given that accent-incompatible melodies are not integrated by a common scheme and appear to contain a series of disconnected chunks, the subjects may simply base their judgments on the relative number of such groupings. Hence, those melodies containing six temporal accents should seem longer than those with three accents, which, in turn, should appear longer than isochronous tunes. This, in fact, was the case. This type of strategy has also been observed in an earlier study, in which it was found that subjects comparing the relative duration of two incoherent melodies based their judgments on the relative number of melodic groupings (Boltz, 1991b). When subjects are unable to reliably use temporal accents to recapitulate a melody, they infer their time judgments from whatever salient information is at hand—namely, the number of independent chunks within the total time span.

EXPERIMENT 2

Experiment 2 was designed as a converging operation to determine whether the previous set of results generalizes to a different type of naturalistic event, as well as to a different time estimation task. In lieu of folk tunes, this second study relied on filmed narratives. Although some researchers have suggested that visual events are less susceptible to rhythmic manipulations (e.g., Glenberg & Jona, 1991), films nevertheless display an underlying hierarchical arrangement of story relationships that is very analogous to that of musical compositions.

Several different accounts of story grammar have been offered in the literature (Mandler & Johnson, 1977; Rumelhart, 1975; Thorndyke, 1977), but all concur that stories consist of four major components—a theme, a setting, a plot, and a resolution. The theme specifies the overarching goal of the story (e.g., determining the culprit—"whodunit"—in a murder mystery) that is pre-

sented against a backdrop of a particular time, place, and cast of characters. A story's plot, on the other hand, provides the overt surface structure of a story by relating a series of episodes in which actions are oriented toward various subgoals critical to the story's resolution. These four components reflect an underlying hierarchical organization of structure in which characters' actions are nested within higher level episodes that vary in their relative importance to the overarching theme. Several studies have shown that subjects not only are sensitive to the hierarchical positioning of episodes (e.g., Johnson, 1970; Kintsch, Kosminsky, Streby, McKoon, & Keenan, 1975; Meyer & McConkie, 1973; Thorndyke, 1977), but also perceive episodes as having psychological independence (Black & Bower, 1979; Glenn, 1978; Haberlandt, 1980; Thorndyke, 1977).

In addition to this underlying grammatical framework of a story, films can also be argued to contain a series of accents that arise from the relative placement of commercial breaks. Just as pauses and prolonged durations function as accents within speech and music, so commercial breaks also provide a type of external rhythm that attentionally highlights the corresponding story elements. Further, films may possess varying degrees of coherence depending on *where* various commercial breaks occur. For example, a film may seem more coherent and comprehensible if commercials occur between episode boundaries. This draws attention to the underlying organization of the story and the hierarchical arrangement of story elements. In addition, an increasing number of commercials coinciding with both higher and lower order nestings of episodes should improve learning and remembering because a greater portion of the story's structure is attentionally highlighted. Viewers should be better able to infer the relationship between various patterns of activity and the overarching theme of the story, and subsequently establish an effective course of retrieval. Conversely, performance should decline when commercials conflict with the hierarchical organization of story elements by occurring *within* episodes. In this case, a greater number of commercial breaks should result in a relatively incoherent film in which the relationship among story elements is no longer apparent. Viewers are confronted with a seemingly disjointed array of information that cannot be integrated into a unified scheme, and just as comprehension will suffer, so, too, will the recapitulation of story information.

These ideas have received some empirical support in a study by Boltz (1992b). In this experiment, subjects viewed a film in which three or six commercial breaks either occurred between higher level episodes, to accentuate the story's underlying organization, or within episodes, to obscure this organization. In addition, a control group watched the same film containing no commercial breaks at all. Following this viewing, all of the subjects were given a surprise series of memory tasks. The results showed that the manipulations of commercial-break placement and number both exerted a significant impact on performance. Relative to the results of the control

group, the attentional highlighting of episode boundaries yielded higher recall and recognition performance, and better memory for temporal-order information and details from the story's plot. These effects were more pronounced when a greater number of commercial breaks occurred within the film. On the other hand, the various indices of memory performance all significantly declined when commercials conflicted with the story's organization, and this was particularly true when the number of commercial breaks increased.

In sum, a similar set of cognitive processes appears to apply to both music and films. In both cases, the events' structure approximates a hierarchical scheme in which lower level information (i.e., notes, actions) is nested within higher order grammatical phrases (i.e., melodic phrases, episodes) related to an overarching theme (i.e., melody's key, story's gist). By highlighting these nested relationships through temporal accentuation, this in turn facilitates learning and attending and provides people with an effective retrieval scheme with which to reperceive the event.

The purpose of this second experiment was to investigate whether the placement and number of commercial breaks also influence the overall accuracy of a film's remembered duration. It relied on the same set of stimuli as used in Boltz (1992b)—namely, two 40- to 45-min installments from a televised miniseries entitled "The Perfect Spy." No, three, or six commercial breaks were then inserted at locations that a preliminary group of subjects had identified as *breakpoint* or *nonbreakpoint* locations. The former represent boundaries between higher level episodes that are particularly important to the story's theme and are further reinforced by shifts in scene setting and new classes of activity (Bordwell & Thompson, 1979; Gianetti, 1982). Nonbreakpoints, on the other hand, occur between lower level actions within an episode and thereby obscure the story's underlying structure. After viewing one of these different film versions, the subjects were unexpectedly asked to estimate how much time, to the nearest minute, seemed to have transpired between the film's beginning and its end.

Method

Design. The design was a 2 × 2 × 2 between-subjects factorial. The subjects viewed one of two films that varied in both the number (three, six) and location (between vs. within episodes) of commercial breaks. Two control groups, to whom a version of each film containing no commercials at all was presented, were also included.

Subjects. One hundred subjects from an introductory psychology course at Haverford College participated in the experiment for course credit. Ten subjects were randomly assigned to each of the 10 between-subjects conditions.

Stimulus materials. The films selected as experimental stimuli were two installments (IV and VI) from "The Perfect Spy" miniseries, based on the book by John LeCarré.[5] As seen in the Appendix, both were relatively equal in plot complexity, in that each contained 20 different episodes marked by distinct shifts in scene setting, and 2–3 subplots within the overall story. Although each film was part of a longer miniseries, each qualified as a story

in that the two installments began at different points in the protagonist's life and showed resolution of the plot's conflicts. In their original version, the two films did not contain any commercials.

To determine the location of breakpoints and nonbreakpoints within each film, a preliminary rating study was conducted with an independent group of 12 subjects. As they watched one of the two films and later reviewed it, they were asked to select at least six episode boundaries that represented "major shifts in the story's plot, i.e. shifts in major idea units," and to rate the certainty of each selection on a seven-point scale (7 = very certain). Those six locations most consistently and confidently selected as episode shifts were identified as breakpoints. Nonbreakpoints represented shifts between actions *within* an episode, and were selected from those locations never identified in the preliminary rating study.

Twenty-four different commercials were recorded from U.S. network television and arranged into six sets, each set containing four commercials with a total duration of 1–2 min. These commercial sets were then inserted into each film such that they coincided with the three or six breakpoint or nonbreakpoint locations (see the Appendix). This in turn resulted in different versions of a film that varied in their total duration. For Installment IV of the miniseries, the durations were 44, 50, and 57 min when they contained no, three, and six commercial breaks, respectively. Conversely, the total duration of Installment VI was 43, 48, and 54 min when the different versions contained no, three, and six commercial breaks, respectively. Within each experimental condition, the time interval between successive commercial breaks was a fairly regular one: breaks occurred every 6–8 min in the six-commercial conditions and every 10–12 min in the three-commercial conditions.

Apparatus. A Panasonic AG-1950 editing system was used for the construction of stimulus materials. During the actual experiment, films were presented on a JVC HR-D320 video cassette recorder interfaced with a Toshiba C990 television.

Procedure. The subjects were instructed that they would be watching an installment from a filmed miniseries entitled "The Perfect Spy," and that afterward, they would be asked to perform several different tasks. The nature of these tasks was not disclosed, and those subjects who were familiar with the miniseries (or book) were dismissed from the experiment.

Immediately following the presentation of a given film version, the subjects were asked to estimate how much time, to the nearest minute, seemed to have transpired between the film's beginning and its end. For those subjects in the commercial conditions, it was clarified that this estimate should include the duration of the film's plot as well as all commercial breaks. They were then asked to perform a series of memory tasks that included story recall, scene recognition, and recall of commercial locations. The latter tasks and their results are described in Boltz (1992b).

The subjects were tested in small groups of 2–4 individuals. Before the beginning of an experimental session, they were asked to remove all watches and jewelry so that they would not provide any source of distraction during the film's viewing.

Results

The verbal estimation data (shown in Column 1 of Table 4) were analyzed for both their overall accuracy (absolute error scores) and their directionality (ratio scores). Unlike Experiment 1, it was not possible to measure variability within subjects as well, because each subject provided only a single duration estimate of a particular installment.

Absolute error scores. The second column of Table 4 depicts the mean absolute error scores as a function of commercial location, the films' total duration, and the number of commercial breaks.

Table 4
Mean Performance in the Verbal Estimation Task of Experiment 2 as a Function of
Commercial Placement, Films' Total Duration, and Number of Commercial Breaks

Commercial Placement	Mean Estimations (sec)		Absolute Error		Ratio Scores	
	M	SD	M	SD	M	SD
Breakpoints						
Inst. IV—6 comm. (57 min)	61.95	5.75	8.68	12.63	1.09	.05
Inst. VI—6 comm. (54 min)	59.53	5.10	10.24	8.27	1.10	.09
Inst. IV—3 comm. (50 min)	58.68	6.92	17.35	11.03	1.17	.10
Inst. VI—3 comm. (48 min)	54.67	7.04	13.89	8.41	1.14	.06
Nonbreakpoints						
Inst. IV—6 comm. (57 min)	82.26	31.25	44.31	42.70	1.44	.28
Inst. VI—6 comm. (54 min)	75.55	35.73	39.91	41.52	1.40	.36
Inst. IV—3 comm. (50 min)	64.27	25.40	28.54	20.76	1.29	.33
Inst. VI—3 comm. (48 min)	63.17	22.67	31.60	26.68	1.32	.23
No commercials						
Inst. IV (44 min)	53.15	9.87	20.80	14.79	1.21	.13
Inst. VI (43 min)	53.97	10.11	25.50	20.21	1.26	.07

Note—Inst., installment; comm., commercials.

An ANOVA was conducted to assess the effects of both the number (three, six) and the location (between vs. within-episodes) of commercial breaks for each of the two films. Overall, its results are very similar to those of Experiment 1. A main effect for commercial location [$F(1,72) = 18.37$, $MS_e = 205.87$, $p < .001$] indicates that duration estimates are significantly more accurate when commercials occur between episodes to attentionally highlight the story's structure. Conversely, when the film's organizational framework is obscured by commercials appearing within lower level episodes, remembered duration is highly inaccurate.

The second major finding is that the number of commercial breaks exerts a differential effect across conditions of commercial location [$F(1,72) = 17.51$, $MS_e = 205.87$, $p < .001$]. A set of Tukey post hoc comparisons ($p = .05$) shows that a greater number of commercial breaks increases the accuracy of judgments in the breakpoint conditions. The opposite effect, however, occurs in the nonbreakpoint conditions. Here, duration estimates become more inaccurate as a greater number of commercial breaks serves to further obscure the story's underlying grammar. Lastly, a set of Dunnett comparisons ($MS_e = 187.2$, $p < .05$) reveals that commercial breaks coinciding with episode boundaries lead to more accurate estimates than no commercials at all. Although the latter condition also displays a high degree of error, the magnitude of this effect is significantly less than that of the nonbreakpoint conditions, in which commercials clash with the story's structure.

In sum, the results of this experiment demonstrate that the effects of temporal accent structure on remembered duration generalize to the visual medium of filmed stories. Overall, accent-compatible conditions yield greater accuracy and, as indicated by the standard deviations of the absolute error scores in Table 4, a lower variability of duration estimates than accent-incompatible ones.

Ratio scores. The mean ratio of subjects' estimated durations relative to a film's actual duration are depicted in the third column of Table 4. A main effect of commercial location [$F(1,72)=14.28$, $MS_e = .082$, $p < .001$] indicates that although significant overestimates occur in both the breakpoint and nonbreakpoint conditions, the magnitude of this bias is significantly greater in the nonbreakpoint conditions. In addition, an interaction between location and number of commercial breaks [$F(1,72) = 6.31$, $MS_e = .12$, $p < .05$] further reveals that overestimates in the accent-incompatible conditions markedly increase with a greater number of commercial breaks. Lastly, Dunnett comparisons ($MS_e = .10$, $p < .01$) show that significant overestimates also occur in the no-commercial conditions. Although the magnitude of this effect is significantly less than that of the nonbreakpoint conditions, films with no commercials are nonetheless judged to be longer than those containing commercials at between-episode locations.

Discussion

The results of this second experiment provide further evidence that the degree of structural predictability within events is a major determinant of the overall accuracy and reliability of remembered duration. In the melodic stimuli of Experiment 1, the observed effects of temporal accent structure may not be particularly surprising, because sound necessarily unfolds in a serial fashion and the timing of this information is important for attending and remembering. In Experiment 2, however, similar effects were found to apply to the processing of filmed narratives, a result that is somewhat inconsistent with the previous literature. Some researchers have reported that rhythm exerts a negligible impact on the cognitive processing of visual events and does not aid in the remembering of durational information (Glenberg & Jona, 1991; Glenberg, Mann, Altman, Forman, & Procise, 1989). However, it is important to note that in contrast to the flashing-light sequences that these studies have typically investigated, films display an inherent organization of information that is remarkably analogous

to that of music. Just as music contains an underlying tonality scheme that integrates melodic phrase relations, so stories display a sequence of characters' activities nested into higher level episodes that vary in their relative importance to an overarching theme. When in a televised form, stories also contain accents, in the form of commercial breaks, that act as a type of external rhythm; and as the results of Experiment 2 indicate, the placement of commercials at between- or within-episode locations exerts the same effects on remembered duration as the occurrence of prolonged notes that are located between or within melodic phrases.

In sum, manipulations of accent structure yield highly consistent effects across the auditory and visual modalities and different instances of music and film. A final source of generalizability comes from the use of different time estimation tasks across the two experiments. In Experiment 1, remembered-duration estimates were obtained from a reproduction task that required subjects to actively recapitulate from memory, via a buttonpress, an event's time span. Relative to other methodologies, this type of task typically produces the most accurate and the least variable estimates (Carlson & Feinberg, 1970; Zakay, 1990). Verbal estimation tasks, on the other hand, tend to be more variable because they require subjects not only to recapitulate an event's duration from memory, but to then translate this time span into clock units of minutes or hours. Given our society's dependence on standarized time keepers, people do have ample opportunity to learn such units of time, but nonetheless, this second stage of estimation affords the potential for greater error. Despite these differences, however, both tasks yielded the same pattern of results, thereby supporting their overall reliability and generality.

The discussion of task differences and, in particular, the nature of the verbal estimation method, raises one possibility that merits consideration. That is, given that this second experiment relied on televised films that normally occur within a 1-h time slot and then required subjects to judge these in terms of clock units, one might ask whether this biased the results in any fashion. For example, were subjects more apt to produce an estimate of 60 min because most televised programs are of this duration? Although this seems to be a very reasonable hypothesis, the data in Table 4 do not support this idea. If true, the two longer, six-commercial conditions should yield more accurate estimates than the shorter three- and no-commercial conditions. In addition, the no-commercial condition should show greater overestimates of the films' duration. This overall pattern of results did not emerge. Instead, a significant interaction between accent number and compatibility was found, in which the accuracy and directionality of duration judgments were influenced by the events' degree of structural predictability. The occurrence of six accents within a film only led to greater accuracy when these coincided with episode boundaries. On the other hand, when a greater number of commercials appeared in the middle of a narrative episode, this in turn produced more errors, as well as longer overesti-

mates of the duration of the films. In sum, the results of the present experiment are more consistent with an interpretation that emphasizes the role of event coherence on remembered duration.

GENERAL DISCUSSION

The type of time estimation behavior addressed in the present set of experiments is one that has not been extensively investigated in the previous literature. Most studies of remembered duration have focused on relative duration estimates and have rarely considered those processes involved in remembering an event's actual duration. One exception is a set of experiments that has investigated the influence of task difficulty on the accuracy of remembering the total duration of the task (e.g., Brown, 1985; Brown & Stubbs, 1992; Hicks et al., 1976). The general finding is that as task difficulty increases, time judgments become more inaccurate and unreliable, presumably because fewer attentional resources are devoted to durational information. However, even when an event's duration is assumed to be successfully encoded into memory, it is never clear exactly *how* this process occurs or what underlying mechanisms are mediating the retrieval of the event's total time span.

The approach that has been outlined here provides one such account by emphasizing the effects of event structure on cognitive processing activities. By examining an event for its overall degree of structural coherence, one is able to predict the extent to which it will subsequently be remembered. Those events containing a regularly recurring pattern of accents coinciding with grammatical phrase relations provide cognitive referents for recapitulating those structural relations within the event's total time span. Again, an increase in the number of accents provides more referents for remembering and facilitates the retrieval of both higher and lower order event relations, as well as a more accurate reproduction of the event's duration. Incoherent events, on the other hand, lack such referents, and this in turn results in highly inaccurate and unreliable estimates of the event's total duration.

One key assumption within this approach concerns the inherent relationship between the temporal and nontemporal information of an event. In coherent events, these two levels of structure codefine one another, and the recapitulation of an event is assumed to include both its temporal and its nontemporal information. Conversely, incoherent events display an array of nontemporal information that is dissociated from its temporal structure, and this, too, is presumably reflected in memory performance. Some empirical support for these notions was, in fact, found in the present set of experiments. The manipulations of accent structure in Experiment 2 are identical to those of an earlier study (Boltz, 1992b) in which subjects were asked to recall and recognize scene information from the same set of filmed narratives. The results showed that as well as leading to more accurate estimates of remembered duration, an increasing number

of commercial breaks coinciding with episode boundaries also yielded better memory for a film's sequence of episodes, the temporal ordering of individual scenes, and more accurate details from the story's plot. In addition, films whose commercial breaks occurred at within-episode locations not only produced inaccurate duration estimates, they also impaired memory for the story's plot. A parallel set of findings can be found when comparing the results of Experiment 1 with those of a music-recall study by Boltz (1991a). Together, then, these overall findings lend validity to the notion that the structural relationship between an event's temporal and nontemporal information is one that is also reflected in remembering.

This idea, however, is one that is quite contrary to those of other time estimation models. Most models claim that the temporal duration of an event is processed independently of its nontemporal information. This, for example, is implicit in those models developed to address relative duration judgments. Although the temporal duration of an event may be encoded in a prospective design via conscious time-keeping strategies, remembered duration is inferred from the amount of nontemporal information alone. This may involve the number of chunks within an interval (Ornstein, 1969) or the number of changes (Block, 1985), but in either case, the more nontemporal information is available, the longer the event appears in retrospect. This strategy is useful when one is attempting to remember a structurally unpredictable event, because one is merely left with groups of disconnected items from which to infer event duration. This situation was revealed here both in Experiment 1 and in Experiment 2, in which it was found that as the number of accents increased to further obscure event relations, remembered duration was increasingly overestimated. Problems arise, however, when one considers that the remembering of an event's actual duration improves with greater structural predictability. These effects cannot be explained by relative duration models because these approaches assume that remembered duration is strictly inferred from the amount of nontemporal information alone, thus failing to reflect the intrinsic timing of the event or the total time span itself. In effect, then, this means that duration judgments are in no way related to an event's actual duration, and there is no construct to explain the accuracy and reliability of retrospective time estimates.

The independence of temporal and nontemporal information is also reflected in those models based on resource allocation theory (e.g., Thomas & Weaver, 1975). Again, these approaches were developed to explain the inaccuracy of time judgments as task difficulty increases (Brown, 1985; Brown & Stubbs, 1992; Brown & West, 1990; Zakay et al., 1983). Since increased task demands require greater attentional resources for the processing of nontemporal information, this in turn reduces the amount of processing devoted to temporal information and hence causes the decline in accuracy for judged duration. As was true of the relative duration models, these

notions are not necessarily incompatible with the present view, but the difference lies in specifying those situations in which temporal and nontemporal information are more apt to become disassociated from one another. Just as music, film, and other natural events can vary in their internal predictability, the same can apply to tasks or activities performed by a given individual. For example, the definition of task difficulty may depend on, among other things, the extent to which a selected strategy can reliably be used to complete a task, the arrangement of the material to be processed, or whether the amount of time or activity required to complete a task conforms to one's initial expectancies. The latter idea, in fact, was examined in an earlier study by Boltz (1993), in which it was found that duration estimates of an experimental session were significantly more accurate when subjects' expectancies for the number of experimental trials in a detection task were confirmed. Conversely, the accuracy and reliability of estimates markedly declined when expectancies were either absent or systematically violated.

In other words, it is suggested that manipulations of task difficulty may covary with those of task predictability, such that the less difficult (i.e., the more predictable) tasks are those whose total time span can be recapitulated with greater accuracy. In many ways, this notion is analogous to the assumption concerning the effect of event coherence on judged duration. Coherent events require less attentional effort during learning, due to their mutually defined arrays of temporal and nontemporal information, and both levels of structure can therefore be easily retrieved from memory. Conversely, incoherent events display an array of nontemporal information that is dissociated from its temporal structure, and these events are more difficult not only to attentionally track, but to subsequently remember. In sum, the temporal and nontemporal information within an event is assumed not to be independently processed across all task and event situations, but to depend on the degree of overall predictability within the total time span.

In closing, the approach described here is one that links variations in natural event structure to attending and remembering activities. By acknowledging the inherent temporal structure of an event and its reciprocal relationship with nontemporal structure, this provides not only a means with which to explain the remembering of an event's actual duration, but a more parsimonious framework that incorporates strategies proposed by alternative models. There is no longer a need to postulate separate processing mechanisms for temporal and nontemporal information, nor a redintegration mechanism to explain their interactive influence on one another. Instead, an analysis of an event's structure will indicate the extent to which temporal and nontemporal information mutually define one another and the subsequent effect on both the accuracy and reliability of duration estimates, as well as any systematic bias. This is not to say that factors other than event structure do not influence retrospective time judgments—to be sure, there are a variety of other envi-

ronmental factors and variables within an individual that contribute to temporal experience. Nevertheless, the present approach indicates that future investigations may benefit from an analysis of the events to be judged and how these in turn may influence those strategies used in time estimation behavior.

REFERENCES

ALLAN, L. G. (1979). The perception of time. *Perception & Psychophysics*, **26**, 340-354.

BERRY, W. (1976). *Structural functions in music*. Englewood Cliffs, NJ: Prentice-Hall.

BLACK, J., & BOWER, G. (1979). Episodes as chunks in memory. *Journal of Verbal Learning & Verbal Behavior*, **18**, 309-318.

BLOCK, R. A. (1985). Contextual coding in memory: Studies of remembered duration. In J. Michon & J. Jackson (Eds.), *Time, action, and behavior* (pp. 169-178). Heidelberg: Springer-Verlag.

BOLTZ, M. [G.] (1991a). Some structural determinants of melody recall. *Memory & Cognition*, **19**, 239-251.

BOLTZ, M. [G.] (1991b). Time estimation and attentional perspective. *Perception & Psychophysics*, **49**, 422-433.

BOLTZ, M. G. (1992a). The remembering of auditory event durations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 938-956.

BOLTZ, M. G. (1992b). Temporal accent structure and the remembering of filmed narratives. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 90-105.

BOLTZ, M. G. (1993). Time estimation and expectancies. *Memory & Cognition*, **21**, 853-863.

BOLTZ, M. G. (1994). Changes in internal tempo and effects on the learning and remembering of event durations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1154-1171.

BORDWELL, D., & THOMPSON, K. (1979). *Film art: An introduction*. Reading, MA: Addison-Wesley.

BOTSFORD, F. H. (Ed.) (1921). *Folk songs of many peoples* (Vol. 1). New York: Woman's Press.

BOWER, G., BLACK, J., & TURNER, T. (1979). Scripts in memory for text. *Cognitive Psychology*, **11**, 177-220.

BROWN, S. W. (1985). Time perception and attention: The effects of prospective versus retrospective paradigms and task demands on perceived duration. *Perception & Psychophysics*, **38**, 115-124.

BROWN, S. W., & STUBBS, D. A. (1992). Attention and interference in prospective and retrospective timing. *Perception*, **21**, 545-557.

BROWN, S. W., & WEST, A. N. (1990). Multiple timing and the allocation of attention. *Acta Psychologica*, **75**, 103-121.

CARLSON, V. R., & FEINBERG, I. (1970). Time judgment as a function of method, practice, and sex. *Journal of Experimental Psychology*, **85**, 171-180.

CARLSOO, S. (1972). *How man moves* (W. P. Michael, Trans.). New York: Cran, Russak.

CONE, E. (1968). *Musical form and musical performance*. New York: Norton.

FRAISSE, P. (1956). *Les structures rythmiques*. Louvain, Belgium: Editions Universitaires.

GIANETTI, L. (1982). *Understanding movies*. Englewood Cliffs, NJ: Prentice-Hall.

GLENBERG, A. M., & JONA, M. (1991). Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition*, **19**, 514-522.

GLENBERG, A. M., MANN, S., ALTMAN, L., FORMAN, T., & PROCISE, S. (1989). Modality effects in the coding and reproduction of rhythms. *Memory & Cognition*, **17**, 373-383.

GLENN, C. (1978). The role of episodic structure and story length in children's recall of simple stories. *Journal of Verbal Learning & Verbal Behavior*, **17**, 229-247.

GROSJEAN, R., GROSJEAN, L., & LANE, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, **11**, 58-81.

HABERLANDT, K. (1980). Story grammars and the reading time of story constituents. *Poetics*, **9**, 99-116.

HICKS, R. E., MILLER, G. W., & KINSBOURNE, M. (1976). Prospective and retrospective judgments of time as a function of amount of information processed. *American Journal of Psychology*, **89**, 719-730.

JAFFE, J., & FELDSTEIN, S. (1972). *Rhythms of dialogue*. New York: Academic Press.

JOHNSON, R. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning & Verbal Behavior*, **9**, 12-20.

JONES, M. R., & BOLTZ, M. G. (1989). Dynamic attending and responses to time. *Psychological Review*, **96**, 459-491.

KAHNEMAN, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

KINTSCH, W., KOZMINSKY, E., STREBY, W., MCKOON, G., & KEENAN, J. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning & Verbal Behavior*, **14**, 196-214.

KRAMER, J. (1982). Beginnings and endings in Western art music. *Canadian University Music Review*, **3**, 1-14.

LONGUET-HIGGINS, H. (1976). Perception of melodies. *Nature*, **263**, 646-653.

LONGUET-HIGGINS, H. (1978). The perception of music. *Interdisciplinary Science Reviews*, **3**, 148-156.

MANDLER, J., & JOHNSON, N. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, **9**, 111-151.

MARTIN, J. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, **79**, 487-509.

MEYER, D., & MCCONKIE, G. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology*, **65**, 109-117.

MEYER, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

MICHON, J. (1972). Processing of temporal information and the cognitive theory of time experience. In J. T. Fraser, F. C. Haber, & G. W. Muller (Eds.), *The study of time* (pp. 224-258). New York: Springer-Verlag.

NAVON, D., & GOPHER, D. (1979). On the economy of the human-processing system. *Psychological Review*, **86**, 214-253.

NEWTSON, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality & Social Psychology*, **28**, 28-38.

ORNSTEIN, R. (1969). *On the experience of time*. Baltimore, MD: Penguin.

PIERSON, K. (1976, December). Control of walking. *Scientific American*, **235**, 72-86.

PISTON, W. (1978). *Harmony* (4th ed., revised and expanded by M. DeVoto). New York: Norton.

POVEL, D.-J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 3-18.

POVEL, D.-J., & OKKERMAN, H. (1981). Accents in equitone sequences. *Perception & Psychophysics*, **30**, 565-572.

POYNTER, W. D. (1983). Duration judgment and the segmentation of experience. *Memory & Cognition*, **11**, 77-82.

RUMELHART, D. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 265-303). New York: Academic Press.

THOMAS, E. A. C., & WEAVER, W. B. (1975). Cognitive processing and time perception. *Perception & Psychophysics*, **17**, 363-367.

THORNDYKE, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, **9**, 77-110.

TODD, R., BOLTZ, M. G., & JONES, M. R. (1989). The MIDILAB auditory research system. *Psychomusicology*, **8**, 83-96.

VOS, P. (1977). Temporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics*, **1**, 183-199.

ZAKAY, D. (1990). The evasive art of subjective time measurement: Some methodological dilemmas. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 59-84). Hillsdale, NJ: Erlbaum.

ZAKAY, D., NITZAN, D., & GLICKSOHN, J. (1983). The influence of task difficulty and external tempo on subjective time estimation. *Perception & Psychophysics*, **34**, 451-456.

NOTES

1. The original versions of these folk tunes were selected from Botsford (1921).

2. Notice that the folk tunes comprising the accent-compatible conditions are different from those within the accent-incompatible conditions. This confound was unavoidable because in order to achieve an accent-number manipulation that was constant (three vs. six accents) across all accent-(in)compatible conditions, it was necessary to find tunes that varied in the periodicity of their melodic phrase boundaries (i.e., 6- vs. 9-beat phrases in the 36-note melodies and 8- vs. 6-beat phrases in the 24-note melodies). This confound was not a serious one because across all conditions of the experiment, there were always four different melodies in each of the accent-(in)compatible conditions, to ensure the overall generality of results.

3. It should be noted that these different rhythmic patterns resulted in different musical meters that varied in their relative familiarity. For example, the 6- and 4-note rhythms suggested the familiar 4/4 (8/4) and 3/4 (6/4) meters, respectively, while the 12- and 8-note rhythms implied the less-common 7/4 (14/4) and 5/4 (10/4) meters. These differences in metric familiarity, however, do not represent a confounding of experimental stimuli. The different rhythms were applied to melodies with both compatible and incompatible melodic phrasing and were simply designed to create tunes that varied in their relative number of temporal accents.

4. The primary reason for introducing this nesting arrangement was to reduce subjects' memory load during the initial learning phase. If subjects had been asked to learn and later remember a set of tunes that reflected all of the accent manipulations in the experiment, this would have resulted in a set of 10 melodies {[accent (in)compatibility × 3 vs. 6 accents × 24- vs. 36-note melodies] + [a 24- and 36-note isochronous melody]} that would have severely taxed subjects' memory resources. Hence, the nesting procedure was used to yield a set of six tunes for a given subject—a quantity that can more easily be encoded and later retrieved from memory.

5. "The Perfect Spy" miniseries was produced in 1989 by BBC-TV Productions. The use of all filmed material in the experiment conformed to the specifications of the house report on piracy and counterfeiting amendments (H.R. 97–495, pp. 8–9).

APPENDIX
Film Structure of "The Perfect Spy"

The Cast of Main Characters

Magnus Pym—The story's main protagonist, who is an agent for British Intelligence but also supplies information to Czech Intelligence.

Axel—A spy in Czech Intelligence who is Magnus's friend and contact

Belinda—Magnus's first wife

Mary—Magnus's second wife

Rick—Magnus's father, who makes his living as a con artist

Tom—The son of Magnus and Mary

Jack—Magnus's supervisor, who is a section director in British Intelligence

Sid—Rick's closest friend and business associate

Seth—A friend from Magnus's childhood

Installment IV

Summary. This particular installment focuses on Magnus's role in British Intelligence. He is extremely successful in obtaining information about Soviet activities which, unknown by his superiors, is supplied by Axel. Also unknown by his superiors is the fact that Magnus is providing Axel with information from British Intelligence. The other major theme concerns Magnus's personal relationships, his marriage to Belinda, and his relationship with his father, Rick.

Setting. The story opens in London during the 1960s and then later shifts to Prague and Berlin, where Magnus is stationed on assignment, during the 1970s.

Episodes From the Story's Plot

1. Magnus and Jack are walking in a park and discussing Magnus's upcoming marriage to Belinda.

2. Magnus is questioned by a British Intelligence committee about his past history, values, and patriotic beliefs. He is promoted as an active agent and told he will soon be sent to Prague on assignment.

Breakpoint One (mean rating of 5.33, selected by 4 of the 6 raters).

3. Magnus and Belinda discuss his career and future life.

4. Magnus and Belinda are at their wedding reception. During a speech by Belinda's father, Rick makes an uninvited appearance.

*****Nonbreakpoint One**. Rick gives the couple as a wedding present a new car, which later turns out to be stolen. After the wedding, Magnus and Belinda are in a taxi, where the driver gives them a bouquet of poppies. This is a signal for Magnus to meet Axel.

*****Breakpoint Two** (mean rating of 6.5, selected by 6 of the 6 raters).

5. Magnus is in a small town near Prague. He walks to a stone fence near some abandoned buildings, removes a stone, and extracts a small package.

6. Magnus returns to his hotel room in Prague. Axel enters with a gun and, for the benefit of his colleagues and bugged microphones, pretends to arrest Magnus for espionage.

Nonbreakpoint Two. Magnus protests his innocence as he is led away by Axel and two other Czech agents.

7. Axel, who is responsible for questioning him, takes Magnus to an empty house belonging to Axel's aunt. Axel tells Magnus that the Czechs suspect him of being a British spy and, if convicted, what the consequences will be.

*****Nonbreakpoint Three**. As they later walk in a meadow, Axel promises to protect Magnus and says they must quit spying and defect to the United States.

*****Breakpoint Three** (mean rating of 6.83, selected by 6 of the 6 raters).

8. Magnus presents Jack and other superiors with photos of a Soviet nuclear facility. He is praised for his work and promoted to director of a spy network in Germany.

9. Belinda's father questions Magnus about his job. Magnus insists that he is an intepreter, but the father suspects him of being a spy.

10. Axel and Magnus are in a car, en route to Czechoslovakia, where they are to meet Axel's superiors. Magnus worries that he will be revealed as a British spy and Axel reassures him that he will be safe.

11. Meanwhile, Belinda is alone at home, crying. She is frustrated that Magnus is always working and never spends time with her.

12. Magnus and Axel are in the house of Axel's aunt. Axel says he will introduce Magnus to a new contact who has access to information about the Czech's industrial activities.

Nonbreakpoint Four. This contact turns out to be a Sabina, a woman whom Magnus met long ago in the military and who is his ex-lover.

Breakpoint Four (mean rating of 5.33, selected by 6 of the 6 raters).

13. Jack and Magnus are in London, walking in a park. They discuss his new assignment in Berlin, where Magnus will pose as a British ambassador.

14. Magnus asks Belinda to go to Berlin with him. She refuses and accuses him of ignoring both her and his father, Rick. She then asks for a divorce.

15. Magnus is at an outdoor cafe, where he meets with Jack.

16. Magnus and Axel are at a deserted restaurant. Magnus relates his earlier conversation with Jack concerning Soviet activities. They also discuss the woman, Mary, that Magnus is now dating. Magnus then discretely gives Axel a cassette tape.

*Breakpoint Five (mean rating of 5.83, selected by 5 of the 6 raters).

17. Magnus is sleeping and awakes to the sound of German police approaching his apartment. He quickly destroys all evidence linking him to espionage activities before admitting them to his home. The police tell him that the police commander wishes to speak with him downtown.

*Nonbreakpoint Five. Magnus enters his bedroom to get dressed.

18. Magnus talks with the police commander, who says that a man convicted of a major crime claims to be an acquaintance of Magnus. Magnus, fearing that it is Axel, discovers that it is Rick who is in a jail cell.

Nonbreakpoint Six. Magnus and Rick are reunited after several years.

19. Magnus and Rick are having dinner at a restaurant when Rick asks Magnus to get him some gold that he can resell. Magnus refuses and has Rick promise he will never again be sent to prison.

Breakpoint Six (mean rating of 5.5, selected by 5 of the 6 raters).

20. Magnus and Axel meet in a deserted bar and Magnus presents pictures of Mary. Axel encourages him to court Mary, since her father is a general in the U.S. military and may thereby facilitate their defection to America.

Installment VI

Summary. In this next-to-last installment, Magnus is experiencing increasing anxiety that he will be exposed as a double-agent, as well as guilt from the various betrayals of his past. He flees to an isolated village, where he begins to write a journal of his past life and tries to explain why he betrayed his country and various loved ones in his personal life.

Setting. The story opens in the 1980s on the Greek island of Lesbos, where Magnus and his family are on vacation. They soon return to their home in Vienna, but shortly thereafter, Magnus leaves alone for London to take care of some business. He then goes to a small coastal town in Devon where he rents a room from an elderly woman.

Episodes From the Story's Plot

1. Scenes from Lesbos, a Greek island in the Aegean Sea, where Magnus, Tom, and Mary are on vacation.

2. Magnus and Tom are watching a cricket match when Axel and some of his agents appear. Although Tom is concerned about who they are, Axel and Magnus go to the parking lot to talk. Axel insists that they must quit now before they are both caught.

3. Magnus, Mary, and Tom are at the beach. As Tom swims, Mary asks about the man (Axel) at the cricket match and demands to know who he is. Magnus lies and says he's a retired spy from British Intelligence.

4. As the family returns from the beach to their cottage, Tom once again sees the men from the cricket match. Magnus says he's mistaken.

*Nonbreakpoint One. They all go inside for drinks.

5. Axel and Magnus meet at a cafe, where they discuss Magnus's family. Once again, Axel insists that they must quit spying and defect to the United States.

*Breakpoint One (mean rating of 5.4, selected by 5 of the 6 raters).

6. At the airport, Magnus and Mary send Tom back to school.

7. Back at the cottage, Magnus begins to write his memoirs. He goes to Mary and they talk about the novel he's presumably writing. Mary asks Magnus what's worrying him; he denies that anything is wrong, and leaves for a walk.

8. Mary goes to Magnus's den and reads what he's been writing. She reads an excerpt addressed to Axel which discusses the concept of betrayal. Magnus walks in and announces that a telegram from Jack has ordered him back to his post in Vienna.

Breakpoint Two (mean rating of 6.17, selected by 6 of the 6 raters).

9. Magnus and Mary are hosting a large dinner party.

Nonbreakpoint Two. Guests leave. Afterward, Magnus begins to act strangely and says that during the party, he received a phone call informing him that Rick had died.

10. Magnus goes to London, to the home of two prostitutes who have Rick stashed in their bathtub, on ice. Magnus gives them money and thanks them for their trouble.

11. Rick's casket is cremated and Magnus, alone, watches.

*Breakpoint Three (mean rating of 5.4, selected by 5 of the 6 raters).

12. Mary, home in Vienna, calls Jack in London and asks where Magnus is.

13. Magnus visits Sid and they reminisce about Rick.

*Nonbreakpoint Three. Sid asks Magnus why he has come and Magnus asks for Rick's old papers and files.

Breakpoint Four (mean rating of 5.0, selected by 4 of the 6 raters).

14. In London, Jack is on the phone with Mary and says that Magnus was last seen at Heathrow Airport.

15. Magnus visits Seth, whom he has not seen in years. They greet one another and a valet brings them drinks.

Nonbreakpoint Four. Magnus then begins crying and apologizing for a prank he pulled on Seth in grade school. Magnus speaks of betrayal and says this was his first.

*Breakpoint Five (mean rating of 5.67, selected by 6 of the 6 raters).

16. At the Vienna home, Jack and Mary discuss Magnus's disappearance and whether he ever mentioned defection.

*Nonbreakpoint Five. Jack searches the room.

Breakpoint Six (mean rating of 5.5, selected by 4 of the 6 raters).

17. At a small coastal town in Devon, Magnus gives a present to an elderly lady. He knows her from past stays and is renting a room from her.

18. Back at the Vienna home, Jack reads Magnus's appointment book.

Nonbreakpoint Six. Jack questions Mary.

19. Magnus and the elderly woman are having tea.

20. Jack and Mary go upstairs, where others have discovered a hidden camera and a Czech codebook. Mary defends Magnus and fights with Jack.

*Locations in the three commercial-break conditions.