Perceiving Musical Time

ERIC F. CLARKE
City University, London

CAROL L. KRUMHANSL
Cornell University

Three experiments are described that investigate listeners’ perceptions of the segmentation of a piece of atonal piano music, the location of segments extracted from the piece, and the duration and structural qualities of each segment. The experiments showed that listeners segmented the music in broad agreement with the grouping principles proposed by Lerdahl and Jackendoff (1983) and perceived the location of randomly presented segments of the music in a strongly veridical manner. Listeners’ location judgments did, however, show systematic departures from veridicality, segments towards the beginning and end of the piece appearing to be located closer to the center of the piece than was actually the case. Judgments of the duration of extracted segments also were strongly veridical and were unaffected by concurrent ratings of structural properties of the segments. In order to assess possible effects of the unfamiliar musical style, the same three experiments were carried out on a piece of tonal piano music of comparable length, yielding essentially identical results. It is argued that the pattern of departures from veridicality in the location judgments for both pieces may indicate systematic changes in attention in the course of listening to the music, linked to large-scale properties of musical structure that are found in music from a variety of styles and periods. The independence of the segmental duration judgments from structural properties of the music may be a consequence of the performance skills of the musically trained listeners used in this study (a sense of absolute tempo is one of the abilities that a performer must acquire) and/or the particular methods used in the experiments.

Introduction

Few studies in the music perception literature have investigated listeners’ experiences during relatively extended passages of music. Thus, we have at
present very little empirical data describing listeners' perception of large-scale musical organization. The main reason for this is that the psychology of music has focused primarily on the way in which the fundamental building blocks of the Western musical tradition are organized into the cognitive frameworks of tonality and meter. The most appropriate and effective way to investigate these structures empirically has been to use comparatively brief and somewhat artificial sequences specially constructed for the purposes of the experiments, which can then be used to pinpoint particular properties in a systematic fashion, and, in the interests of clarity, usually manipulate no more than one or two parameters at a time.

There are certain obvious advantages in this very controlled kind of approach, and it has proved extremely powerful and productive for advancing our understanding of tonal and metric hierarchies. However, it has left untouched a range of issues concerned with listeners' understanding of more extended and elaborate structures in which a considerable degree of interaction between different parameters can be expected. This paper is an exploratory study of three related issues, all connected with the perception of large-scale musical form:

1. the manner in which listeners perceive the segmentation of a whole piece of music, and the musical factors that influence that segmentation;
2. the extent to which listeners develop a mental plan of the piece that they can then use to identify the original location of an extract taken from the piece;
3. the perceived duration of segments from the piece and the relationship between their apparent duration and structural characteristics of the music.

The rest of this section will review briefly the general literature on musical segmentation, memory for temporal organization, and time perception relevant to the present study.

SEGMENATION IN MUSIC

The most coherent and systematic account of the factors promoting musical segmentation is contained within Lerdahl and Jackendoff's (1983) generative music theory. Although their work has no empirical component and is primarily a contribution to music theory and analysis, it clearly embodies a number of cognitive concerns and has been regarded both by the authors themselves and by others (e.g. Sloboda, 1986) as a contribution to the cognitive psychology of music. Segmentation (or "grouping" structure as they call it) is one of four hierarchical components of musical structure
(the other three being metrical structure and two kinds of analytic pitch reduction), each of which is treated in the theory by means of a set of explicit and reasonably formal generative rules. The purpose of these rules is to set out with some precision the conditions under which musical structures are created in the mind of a listener.

Grouping structure is initially specified by a set of well-formedness rules, which simply establish the criteria for strict hierarchical structure. The subsequent grouping preference rules (GPRs), which identify the criteria for deciding which of a number of possible well-formed structures a listener is likely to select, are essentially of three types:

1. Preference rules based on the Gestalt principles of proximity and similarity (GPRs 2 and 3), the hierarchic level to which the rule applies being determined by the strength of the Gestalt feature (GPR 4). These are essentially responsive to surface features of the music.
2. A preference rule based on the grouping effects of pitch structure, based on the disposition of stable and unstable harmonic elements within the framework of the tonal system (GPR 7). This rule is responsive to relatively deeper structural features of the music.
3. Preference rules based on the more abstract principles of symmetry and motivic similarity, or “parallelism” (GPRs 5 and 6).

A recent study (Deliège, 1987) empirically assessed the operation of the rules based on Gestalt principles and the relative strength of these rules when they conflict. Using extracts of recordings from the standard musical repertoire as well as specially constructed short test sequences, Deliège demonstrated that the segmentation points predicted by the rules were largely borne out by the experimental results for both kinds of material. She found significant differences in the saliences of the different rules (measured both in terms of the number of times a boundary determined by a particular rule was chosen and the number of times the extract had to be repeated before subjects decided on a boundary location), and in the pattern of relative rule saliences over the eight rules tested for musicians and nonmusicians. However, for both groups of subjects, boundary decisions in accordance with the predictions of Lerdahl and Jackendoff’s theory were significantly greater than chance, with the musicians’ responses conforming with the rules significantly more than the nonmusicians’. Thus, the theoretical proposals embodied in Lerdahl and Jackendoff’s rules are largely borne out by the empirical results and appear to apply to a considerable extent to untrained listeners as well as to the “experienced listeners” envisaged in the
original theory. The only major difference between the two groups was that context appeared to influence the effectiveness of the rules for musicians, but not for nonmusicians. The musicians’ responses were significantly more in accordance with the theory when the extract was preceded by a section of the same music than when it was preceded by different music, or nothing at all (no context).

Deliège’s study does not present a systematic account of the relative strengths of all possible pairwise rule conflicts, but the summary results of the rules preferred by the musicians and nonmusicians allow some general indications to be seen. For musicians, the most powerful rules seem to be those based on changes in timbre and dynamics and on the existence of gaps in the music (the slur/rest rule GPR2a), and the weakest are those based on changes in melodic contour and notated duration. For nonmusicians, changes in timbre and register are the strongest rules together with attack-point interval (GPR2b), and changes in melodic contour and notated duration are the weakest. A problem with these results is that it is not clear how quantities in different parameters should be compared. It is only sensible to consider the relative strength or weakness of different rules if some kind of quantitative comparison can be made across different parameters, since a larger change in notated duration may outweigh a change in timbre or dynamics. At present no such interparametric metric exists.

A more recent study by Deliège (1989), using a methodology similar to that of the present study, was concerned with the recognition of form in reasonably extended musical structures. It looked at the segmentation of whole pieces of music and focused to a considerable extent on the effects of musical expertise on listeners’ ability to identify elements of formal structure in two pieces of contemporary music. The failure of the earlier study (Deliège, 1987) to find any significant differences in the manner of segmentation between the different groups of subjects was attributed to the fact that grouping structure in these short and somewhat artificial sequences was largely determined by surface features of the music (acoustical and/or temporal properties of the stimuli), which made no demands on the musical competence of either the experienced or the inexperienced listeners. A question addressed in the more recent study was therefore whether the same result would be obtained with more complex and extended musical material, in which listeners’ musical competence would be engaged and which would involve an increased memory component.

Deliège’s method was to play commercial recordings of a performance of Sequenza VI for solo viola by the contemporary Italian composer Luciano Berio and of the orchestral work Eclat by Pierre Boulez. In the course of three complete hearings, subjects indicated by pressing a key on a computer keyboard the points at which they heard group boundaries, responding
only on the second and third hearings. The results can be summarized very briefly as follows:

1. No differences were found between the segmentations produced by musicians and nonmusicians, although there was a slight tendency for nonmusicians to make a greater number of boundary responses. Two composers who produced segmentations and analyses of the Berio piece appeared to use a smaller number of more synthetic groups than either of the two groups of subjects.

2. Pauses appeared to be the main grouping indicators in the piece, although only, the author asserts, when they occur in conjunction with a structural differentiation of some sort. A pause on its own is not sufficient to establish a boundary.

MEMORY FOR TEMPORAL ORGANIZATION IN MUSIC

A follow-up experiment by Deliège (1989) using the Berio Sequenza investigated listeners’ ability to recognize the original locations in the piece of a number of short excerpts. The same listeners as had participated in the experiment just described heard a complete performance of the Sequenza that contained marker tones indicating the position of the five principal boundaries in the music. They subsequently heard a total of 40 short extracts from the piece and were required to indicate from which of the six sections each extract came. Again the results can be very briefly summarized as follows:

1. Extracts were correctly located in relation to the six main sections, with between 41% and 75% accuracy by the musicians, depending on the section from which the extracts came, and between 37% and 63% accuracy by the nonmusicians. Thus, musicians tended to be slightly more accurate than nonmusicians in this location task.

2. The least accurately located extracts were, as one might expect, those which consisted of material that is fairly widely distributed through the piece—in other words, material which is not specific to a particular section. Conversely, extracts that contain material that is confined to a single section (i.e., that is particularly idiosyncratic) were the most accurately located.

Deliège’s results suggest that listeners can make location judgments with reasonable accuracy, although they are influenced by the distribution of the material in the piece, as one would expect. She does not theorize about the
internal representations and processes underlying this performance, but we can consider two different strategies that might be used to locate musical segments within a piece as a whole. The first can be thought of as a “rerun” strategy, in which the listener runs through the piece in his/her mind in order to discover where a particular segment belongs and, having “found” it in the rerun version, estimates the relative position of the extract in relation to the whole piece. This is a strategy that receives empirical support on a smaller time scale from Halpern (1988a). She asked subjects to judge the relative positions, or relative pitch height, of items within well-known tunes that they were asked to remember, but did not actually hear. The results could be understood if listeners were assumed to sing through the tunes in their heads in order to perform the task. Decision times were greater for items spaced further apart and for pairs of items further into the tune. Although a strategy such as this seems perfectly plausible for short musical sequences like the tunes used in Halpern’s study, it seems very unlikely for a piece of music of greater length and complexity. Halpern (1988b) has shown that when subjects imagine a tune they do so at a tempo close to that at which they would actually sing it and that even when asked to imagine it at a faster tempo, they cannot rerun it faster than slightly less than double speed. Thus, it would not seem a likely strategy for locating segments within whole pieces of music of any significant length.

An alternative is that listeners form a much more abstract and symbolic formal representation of the piece, which gives faster, but less detailed, access to the musical characteristics of identifiable sections of the piece. These sections may also be labeled with their functions (e.g., “ends the piece,” “develops the first idea”) in a way that allows a listener to judge relatively quickly where the section comes from when it is presented in isolation. If an isolated extract does not carry this kind of functional and positional identification for a listener, then it may be necessary for the listener to scan through the abstract formal representation of the piece, dipping into each of the sections to see whether its musical characteristics match those of the presented extract. This may have some of the rerun characteristics of the “sing in the head” method, but in a highly condensed and streamlined form that allows the whole piece to be scanned at a coarse level in only a fraction of its real running time.

A review by Jackson (1985) of studies of memory for temporal information in verbal materials demonstrates some interesting parallels, although it raises certain basic issues about the possible differences between temporal aspects of memory in verbal materials and music. The studies all make use of word lists whose items relate to one another in different ways and to different extents, ranging from unrelated word lists, through lists with similar sounding words, to lists consisting of words derived from scripts, which
have a strong intrinsic temporal structure. The experimenters used a number of closely related tasks, all of which required subjects to inspect these lists and subsequently to recall either the absolute position or the relative order of named items in the original list or to recall the number of items intervening between a named pair. Analysis of verbal protocols by the subjects concurrently with the task revealed that a number of different kinds of strategy were used spontaneously in both the encoding and retrieval stages of the task. Subjects who used various kinds of elaborative strategy, involving the linking or "narratizing" of the items in the list, performed considerably better at recall than those who used a simple repetitive strategy. In addition, lists that contained stronger intrinsic cues for temporal relations gave rise to better performance than those that were more temporally neutral.

These findings suggest a number of questions about the manner in which temporal relations in a continuous piece of complex music, such as that used in the study by Deliège and the present study, are encoded. Music differs from the kinds of material used in the studies reviewed by Jackson in a number of ways. First, the notion of "item" in music is somewhat difficult: music consists of a continuous flow of information in which the idea of individual, discrete items must be treated with great care. It is too easy for musical notation, which indicates clearly distinguishable discrete events (notes), to be uncritically assumed to be directly equivalent to the way the music sounds, while in reality the identifiability of "notes" may be radically affected by the effects of grace notes, trills, pedalling, registral extremes, dynamics, and so on. Some of the same arguments apply to language, but there is a certain legitimacy in asserting the reality of words as discrete entities. Second, while the items to be located in Jackson's reported studies were single words, the "items" to be located in Deliège's study (and in the present study) were complex musical segments. Subjects must therefore locate a musical region in relation to the whole piece, not just a point. Third, musical pieces, such as those used in these studies, consist of hundreds or thousands of events (depending on how "event" is specified). Although one of the strategies with the verbal experiments was simply to use repetitive rehearsal as a way of maintaining the original list in mind, a similar strategy for whole musical pieces is clearly out of the question. Listeners are obliged to form a representation in long-term memory, and therefore to arrive at some kind of overall conception of the piece—however faulty, idiosyncratic, or incomplete. Finally, no equivalent exists in music to the kind of semantic elaboration that some subjects use to consolidate the relative positions of items within word lists (Jackson, 1985). If listeners use some kind of narratized representation to remember long musical pieces, that representation must depend on intrinsic structural properties of the piece, such
as motivic links and developments, pitch collections and their relations, rhythmic transformations, and so on. All of these are properties derived directly from the piece itself, rather than imposed upon it from outside. As a consequence it is likely that the performance of listeners in locating musical extracts will be influenced primarily by the structural context from which the extract has been taken and rather less by differences in encoding and retrieval strategies. It may be partly the artificiality of the word lists used in the studies reviewed by Jackson that makes these differences in strategy possible, by contrast with the more natural activity of listening to a piece of music from beginning to end.

**TIME PERCEPTION AND MUSIC**

Empirical studies of the temporal structure of music have been confined almost exclusively to the study of its rhythmic and metric organization. Larger scale temporal organization has been largely ignored [although see Clynes & Walker (1986) for a study of large-scale temporal organization in performance], and little attempt has been made to coordinate the experience of time in music with theories of time perception of a more general sort. Theories of time perception are themselves divided rather clearly into two models: those based on the idea of an internal clock or pacemaker (e.g., Treisman, 1963; Luce, 1972; Kristofferson, 1980) and those based on the idea that perceived duration depends on the amount of information processed or stored (e.g., Fraisse, 1963; Ornstein, 1969; Michon, 1972). The continuing debate (Michon & Jackson, 1985) over the relative merits of these two kinds of theory arises from the fact that different aspects of human behavior seem to fit one or the other of the two models. Under certain circumstances, such as highly skilled motor tasks, temporal control can be remarkably precise and seems most easily explained in terms of the operation of an internal clock. In the context of music performance, for example, performers can maintain a tempo, or return to it after varying periods of time, with an accuracy of 99% or better (Clynes & Walker, 1982; Shaffer, Clarke & Todd, 1985). Under other circumstances, however, subjects' judgments of the apparent duration of an interval are strongly affected by the events occurring during that period and by the circumstances in which the subjects interact with those events (Ornstein, 1969; Block, 1985; Michon, 1985).

It may be that rather than viewing these as competing theories, it is more useful to consider duration as the result of convergent information from different sources, with one or the other source dominating in different conditions. Taking this approach, Thomas and Cantor (1978) developed a model for duration judgments combining input from an internal clock with input from an information processor in a weighted function. The weights
reflect the extent to which a perceiver’s attention is focused on the task or on the passage of time. Attention can be switched between the task and the clock, so that when the subject is extremely task oriented, duration judgments will be primarily based on a measure of information and when the subject pays little attention to the task, duration judgments are based on the output of the clock.

As far as music is concerned, the two models seem applicable to different levels of musical structure. Essentially, clock models seem most appropriate for musical behaviors such as performing and conducting, which involve a motor component (see Vorberg & Hambuch, 1978; Shaffer, 1981), and for the perception of the short durations that make up surface-level rhythmic units (e.g., Longuet-Higgins & Lee, 1982; Povel & Essens, 1985). The perception of longer durations in music has not been studied empirically, but informal observation suggests that information processing models are more appropriate, because listeners seem to experience dramatic changes in the rate at which time appears to pass, or in the apparent duration of a passage, depending on factors influencing the complexity or familiarity of the music (Grisey, 1987; Reynolds, 1987). A rather direct parallel seems to exist between these experiences and the results that Ornstein (1969) obtained with durations ranging from 30 sec to 9.5 min, when he found that perceived durations depended on the objective complexity of the stimuli to which subjects were exposed and differences in the efficiency of the coding strategy which they learned in the experiment.

An important attempt to put this kind of theory into practice in a musical context is Stockhausen’s (1958) analysis of the projected temporal experience of a listener hearing the opening of the second movement of the Webern string quartet, op. 28. The analysis considers the predictability or surprise value of events as they occur through the movement, based on a consideration of a number of parameters of musical structure, including the mode of attack, the number of notes in a chord, the registral spread and interval content of a chord, and the dynamic level. From these a composite measure of the “degree of alteration” of an event in relation to what precedes it is constructed and used in an informal way to convey the momentary sense of tempo, or temporal passage, at that point in the music. Stockhausen’s analysis, although interesting as an illustration of the way in which the idea of information content and its influence on temporal experience might be directly applied to music, contains no empirical component and is based on a piece of music with a very unusual texture. The quartet movement consists of notes of only one duration (quarter notes), which facilitates the task of deciding on the predictability of successive events: with complete temporal predictability, the surprise value of an event can be entirely specified in terms of the other parameters (register, articulation, dynamic, etc.) upon which Stockhausen’s analysis is based. With the variety of
durational values normally found in music, the estimation of surprise value (or information content) is extremely problematic, since it is necessary to consider not only the predictability of the event characteristics, but also the predictability of their temporal location. In the light of this difficulty, one of the aims of the present study is to investigate empirically whether listeners’ duration judgments are related in any systematic way to the perceived complexity of the music they hear.

Experiments 1–3: Stockhausen’s Klavierstück IX

The purpose of these experiments was to bring together the related issues of segmentation, remembered location, and perceived duration so as to gain insight into the formal and temporal experiences of listeners as they listen to a complete piece of music. The first three experiments all make use of the same piece of music—Karlheinz Stockhausen’s Klavierstück IX for solo piano. A number of factors influenced the choice of this work: it consists of a single movement of substantial but manageable length (about 10 min); it contains a variety of different musical ideas, as well as an element of development and continuity; it contains material of a metrical nature as well as entirely nonmetrical passages—a distinction that might have an interesting effect on listeners’ perception of both form and duration; and it encompasses different notated tempi—another possible factor influencing time judgments. It was also important that a professional performer was available who knew the piece and from whom we could obtain the specialized recordings required by the experiment (see below).

The piece, completed in 1961, is atonal with a pitch structure organized according to the principles of 12-note serialism (Perle, 1980). Its rhythmic structure is extremely varied, and is based on proportions derived from the Fibonacci series—a technique Stockhausen has used in a number of other works. The single continuous movement is divided into three broad regions: measures 1–16 focus on multiple, isochronic repetitions of a single chord, interrupted at measure 3 by a brief interlude of slower, linear music; measures 17–116 introduce irregularly spaced chords, linear material related to the interlude of measure 3, and trills, all of which are developed separately and together in different ways, incorporating references back to the repeated chords of the first region; measures 117 to the end (measure 153) constitute a kind of coda characterized by rapid, nonmetrical and virtually exclusively linear (or “melodic”) material played at a very high register on the piano. As this brief summary makes clear, the middle region is the most developmental. Figure 1 shows the opening page of the score to give an idea of the kind of musical texture involved. Notice the large number of
Fig. 1. The opening page of Stockhausen's *Klavierstück IX*. Reprinted with permission from Universal Edition (London) Ltd.
repetitions of the opening chord, the time signatures derived from the Fibonacci series, the abundant dynamic indications and the abrupt juxtaposition of radically different dynamic levels, the alternating sections at different tempi, and the change in rhythmic texture at measure 17 where the second region starts.

EXPERIMENT 1: IDENTIFYING BOUNDARIES

The first of the experiments reported here is concerned with the perceptual segmentation of Klavierstück IX. In designing this experiment, three considerations were of primary importance:

1. Listeners should perform the segmentation task in a way that interfered as little as possible with the normal pattern of continuous listening.
2. Listeners should perform the segmentation initially without reference to the notated score, so that their judgments were based as much as possible on auditory rather than visual information.
3. Listeners should position their boundary judgments as accurately as possible in the music, so as to provide precise information about the structural features causing the boundary.

This last consideration conflicts somewhat with the first two, because a judgment based purely on auditory information during continuous listening will frequently be reactive—a retrospective response to a change in the music that is only recognized as a boundary some time later. The position at which listeners first make their response will therefore be located a variable period of time after the true position of the boundary. A method was required that would allow these variably retrospective judgments to be relocated by the subject to the actual point of change in the music. These three considerations led to the three-part procedure described below.

Methods

Apparatus and Stimulus Materials

The stimulus materials were based on a performance of Karlheinz Stockhausen's Klavierstück IV by Pierre-Laurent Aimard of the Ensemble InterContemporain. The performance was played on a Yamaha KX-88 keyboard with a piano timbre produced on Yamaha TX-816 and DX-7 synthesizers, amplified and played over loudspeakers. A Macintosh Plus computer (with MIDI interface) recorded the timing (onset and duration) and velocity of each key press using Performer software. Three complete recordings of the piece were made, and the performance judged most satisfactory by the performer was used in the experiment. Its total duration was 10.22 min. The performance was played back during the experiment.
using a MacIntosh Plus computer, MIDI interface, and a TX-802 synthesizer producing a piano timbre similar to that used during the recording session. The analog output of the synthesizer was amplified and played at a comfortable listening level over loudspeakers. A foot pedal connected to a Yamaha MCS2 recorded the responses during the second part of the experiment.

**Subjects**

The seven volunteer participants were working in various capacities at I.R.C.A.M. (Institut de Recherche et Coordination Acoustique/Musique). Four were primarily researchers in music psychology, psychoacoustics, or music acoustics. All but one of these listeners had extensive instrumental training and performance experience and had also studied theory and composition. The three remaining participants were composers. All but two of the participants had heard the piece before, but none had played it. One participant had studied it analytically in depth, and two others had studied it briefly.

**Procedure**

Listeners were told that the experiment investigated the perception of temporal organization in a piece of contemporary music, Stockhausen’s Klavierstück IX, and that the experiment consisted of three parts. In the first part they heard the entire piece played without interruption. This was to ensure familiarity with the piece; no responses were required. In the second part they again heard the entire piece played without interruption and were asked to indicate where segment boundaries occurred by pressing a foot pedal after hearing a boundary. They were told that the study was concerned with relatively large-scale segments of which there might be anywhere from 5 to 15 in this piece. However, they also were told they could be quite liberal in this second part of the experiment, because in the third part they would have an opportunity to remove any boundaries about which they had changed their minds. (They would not have an opportunity to add boundaries, however.) In the third part of the experiment, the listeners were given a copy of the score on which the experimenter had marked the approximate location of each boundary that the listener had indicated in part two. The piece was then played from the beginning, stopping at each successive boundary. For each stopping point, four judgments were required. The first was to indicate on the score the precise position of the boundary by drawing a vertical line through the stave. The second was to rate on a seven-point scale the strength of the boundary (1 = very weak boundary; 7 = very strong boundary). The third was to rate on a seven-point scale how easy it was to locate the boundary precisely in time (1 = very difficult to locate; 7 = very easy to locate). The fourth and final response was to describe briefly the features of the music that helped form the boundary. These last three responses were made on a separate response form. Listeners were tested individually, and the duration of the experimental sessions ranged from approximately 1 hr to 2.5 hr. At the end of the session, the participants described their musical backgrounds on a short questionnaire.

**Results and Discussion**

Listeners varied in the number of boundaries they indicated, ranging from 6 to 21. The number of boundaries averaged 11.29, but this value is skewed by one listener who indicated a large number of weak boundaries. In general there was considerable agreement between listeners in the placement and relative strength of boundaries. The ten boundaries that were
agreed on by a majority of the listeners and that had the greatest average judged boundary strengths are shown in Figure 2. The figure shows the average judged boundary strengths as a function of the time at which the boundary occurred; the measure numbers also are indicated in the figure. This summary indicates that the piece divides perceptually into three main sections. The first section extends from the beginning of the piece to the end of measure 16; this section includes a number of well-marked subsections. The second section goes from measure 17 to the end of measure 116; it includes a number of relatively weakly marked subsections. The final section begins at measure 117 and continues to the end of the piece; it weakly divides into two subsections. No interesting patterns were found in the listeners' judgments about how easy or difficult boundaries were to locate. The majority of the judgments were that the boundaries were very easy to locate; listeners gave ratings of six or seven on the seven-point scale for over 70% of the boundaries identified. For six of the seven listeners, there was no correlation between ease of localization and boundary strength, whereas a positive correlation was found for the remaining subject.

Table 1 lists the musical characteristics that listeners described as contributing to the formation of the 10 strongest boundaries. Although various characteristics were identified, they fall into four general categories. The first includes silences and long pauses. The second includes contrasts within musical parameters, such as dynamics, register, texture, and rhythm. The
### TABLE 1
Musical Characteristics Contributing to the 10 Strongest Boundaries in Stockhausen’s *Klavierstück IX*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Strength</th>
<th>Musical Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.429</td>
<td>Pause (silence) (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Return of material (chordal) (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic (2)</td>
</tr>
<tr>
<td>3</td>
<td>4.000</td>
<td>New material (chords changing to melody) (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pause (silence) (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of rhythm (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of pitch content (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of articulation (1)</td>
</tr>
<tr>
<td>4</td>
<td>3.857</td>
<td>Return of first material (chordal) (5)</td>
</tr>
<tr>
<td>17</td>
<td>5.857</td>
<td>New material (change of pitch content) (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Start of Development (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of rhythm (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of articulation (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of register (expansion) (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic contour (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of texture (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pause (1)</td>
</tr>
<tr>
<td>25</td>
<td>2.143</td>
<td>Return of material (chromatic run) (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relaxation of tension (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of register (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic (1)</td>
</tr>
<tr>
<td>51</td>
<td>3.000</td>
<td>Return of material (chordal) (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic (1)</td>
</tr>
<tr>
<td>79</td>
<td>2.000</td>
<td>Return of material (chordal with new pitches) (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Introduction of trill (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pause (1)</td>
</tr>
<tr>
<td>94</td>
<td>2.714</td>
<td>New material (isolated block chords) (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of tempo (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of register (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of pitch content (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of tone (due to pedal) (1)</td>
</tr>
<tr>
<td>117</td>
<td>6.000</td>
<td>New material (unmeasured high, fast notes) (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Arrival of Coda (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of register (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of rhythm (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic (1)</td>
</tr>
<tr>
<td>145</td>
<td>2.286</td>
<td>Isolated low note (as part of chord) (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fragment of earlier material (2)</td>
</tr>
</tbody>
</table>

**NOTE:** The numbers in parentheses indicate the number of listeners (out of seven) noting each characteristic.
third includes changes in pitch content, or melodic contour, or shifts between vertical and horizontal organization. The final category includes the restatement or repetition of previously heard material. These categories relate quite directly to Lerdahl and Jackendoff's (1983) grouping preference rules. The first corresponds to their grouping preference rule 2 (temporal gaps in the music induce boundaries); the second corresponds to their rule 3 (changes in register, dynamic, articulation or note length induce boundaries); and the fourth corresponds to their rule 6 (segments of music that can be construed as repetitions, or variant repetitions, of one another form parallel groups). The third category does not relate to Lerdahl & Jackendoff's rules quite so clearly, but it is interesting to note that melodic contour change also is proposed as an additional rule by Deliège (1987) and that changes in pitch content are an atonal equivalent of rule 7 (prefer a grouping structure that ties in with the harmonic structure of the music).

**EXPERIMENT 2: LOCATING SEGMENTS**

One of the main purposes of the first experiment was to provide a perceptual segmentation of the music that could be used in the two experiments that follow. The first of these investigated how successful listeners were in identifying the original location of an extract from the piece. Segments of the piece, extracted on the basis of the perceptual boundaries identified in Experiment 1, were presented in isolation to listeners who were asked to indicate the original location of each of the segments in relation to the whole piece. The aim of this study was to investigate memory for large-scale musical form by analyzing the pattern of listeners' location judgments.

A number of structural features of the music may influence the location judgments. First, listeners may identify a structural feature that changes systematically throughout the entire course of the piece, which can be used as a key to locate a given segment. In the case of Klavierstück IX, there is a progressive (although not strictly monotonic) change from a comparatively dense, low-register, chordal texture at the start to a thin, high-register, single-line texture at the end. It may be that listeners can first locate the approximate position of a given segment simply on this textural basis alone and then pinpoint it more precisely in some other way. This would result in quite veridical location judgments.

Second, the formal structure of the piece is organized into the three sections (exposition, development, coda) mentioned in the introduction. The characteristic of these three sections is that the first and third present relatively clearly stated and distinct musical ideas, while the middle section develops ideas in a much more fluid and less declarative manner. This may result in more accurate location judgments for segments drawn from the first and last parts than for segments drawn from the middle part.
Finally, an extract might be located according to its relation to nearby segment boundaries. An extract beginning or ending at a segment boundary should have a greater degree of closure than an extract straddling a boundary, making it more self-contained and hence harder to locate in relation to the piece as a whole. By contrast, an extract spanning a boundary should be more open, and hence less self-contained as a musical entity, and furthermore, by including material from more than one part of the piece, might signal its location in the overall scheme more clearly. Thus, as the following methods section makes clear, extracts conforming to each of these three types were prepared.

Methods

Apparatus and Stimulus Materials

The stimulus materials were based on the same performance of Stockhausen’s *Klavierstück IX* used in the first experiment. The stimuli were played using a Macintosh Plus computer (running *Performer* software), MIDI interface, and an Akai S-900 Digital Sampler (producing a piano sample); the analog output was amplified and played at a comfortable listening level over loudspeakers. Eighteen segments, of duration equal to approximately 30 sec, were extracted from the piece. The durations averaged 29.84 sec, with a range from 24.56 to 32.37 sec. They were selected as follows. The six strongest boundaries (excluding the first boundary) were identified; these are boundaries number two, three, four, six, eight, and nine (see Figure 2). (The first boundary was not included because the materials before and after the boundary are essentially identical and thus would be extremely difficult to distinguish in the task.) Six segments ended immediately before these six boundaries. Six segments began immediately after these boundaries. Six segments spanned the boundaries with the middle of the segment occurring as close to the boundary as possible. Figure 3 shows the segments ending at boundary number four, beginning at boundary number four, and spanning boundary number four as examples.

Subjects

The 23 listeners who participated in the experiment were paid £3.00 each. They were all music students at City University, London. On average, they had received 16.4 years tuition on various musical instruments. Nine listeners were first-year students, twelve were second-year students, and two were third-year students. None of them was familiar with the piece before the experiment.

Procedure

Listeners were told that the experiment investigated the perception of the location of short musical segments in relation to the whole of Stockhausen’s *Klavierstück IX*. At the beginning of the experimental session, the task was described to the listeners and they heard the whole piece played through twice. After this, they were presented with three practice trials, and then the 18 experimental trials corresponding to the 18 segments. For each trial, they were asked to indicate where the segment occurred in the piece. They made their response on a horizontal line of length 14.5 cm, where the left end represented the beginning of the piece and the right end, the end of the piece. They were instructed to indicate where they thought the segment began by drawing a vertical line, and to indicate where they thought the segment ended by drawing a second vertical line. The segments were presented in random order, and listeners were tested in groups.
Fig. 3. Examples of segments extracted from Stockhausen’s *Klavierstück IX*, and used in Experiment 2. The segments shown (A) and at the fourth boundary, (B) begin at the fourth boundary, and (C) span the fourth boundary.
Results and Discussion

The midpoint of the time interval indicated by each listener for each segment was measured. (The length of the time interval, which can be taken as an indirect measure of the perceived duration of the segment, also was measured; these data will be considered later in connection with the third experiment.) In order to examine the correspondence between listeners' judgments of the position of each segment and its actual position in the piece, we determined the median judgment for each segment across the 23 listeners. These median values are plotted in Figure 4 as a function of the actual midpoint of the segments. The diagonal line in the figure indicates perfect accuracy. The listeners' judgments deviate from this line, particularly for segments near the middle of the piece, where there is a tendency for the location judgments to be displaced toward the center of the piece. Despite these deviations, there was a strong effect of the actual location of the segment; the median values correlated significantly with the actual locations of the segments [$r(16) = .856, p < .0001$]. The accuracy of individual subjects' judgments (assessed by correlating their judgments with the actual locations) did not vary systematically with quantifiable aspects of their music backgrounds (years' tuition on musical instruments or year at university).

Fig. 4. The judged location of segments from Stockhausen's Klavierstück IX plotted against their actual position in the piece. The curved line shows the third-order polynomial regression solution for the data.
These individual subject correlations ranged from .269 to .893, with an average of .635. The veridicality of these judgments may reflect the way in which the music progresses from a low register chordal texture at the start to a high linear texture at the end, as mentioned earlier.

Nonetheless, a nonlinear regression of the median location judgments, whose function is shown as the curved line in Figure 4, showed that significant contributions were made by cubic and quadratic components. The nonlinear regression gave an \( r = .942 \) \[ F(3,14) = 37.060, \ p < .0001 \], with significant weights for both the cubic component \( t(14) = 4.020, \ p = .0013 \) and the quadratic component \( t(14) = 3.927, \ p = .0015 \). When the individuals’ judgments were entered into a nonlinear regression with the actual locations, 21 of the 24 listeners showed the same pattern as the group data: a negatively accelerating function changing to a positively accelerating function near the middle of the piece. Thus, the pattern of deviations from veridical judgments is characteristic of the individual subject data.

It may be that the steeper functions at the start and end of Figure 4 are the result of an increased sense of musical progress through the piece, brought about by the relatively clear presentation of distinct musical ideas that advance the musical argument. Listeners overestimate the relative time intervals between extracts from these two sections because they have a stronger sense of the succession of musical ideas at the start and at the transition to the coda than elsewhere. By contrast, the central part of the piece is something of a mixture, where different ideas are combined and juxtaposed, so that the sense of goal-directed musical progress is weakened. In this section listeners regard all extracts as being closely located in relation to one another, because of their strong interactions, resulting in the flat gradient in the middle part of Figure 4.

A closer look at the figure shows that segments 1–3 and 7–9 account for the steep gradient at the start of the piece; all these segments come from the opening exposition section. (Segments 4–6, which do not demonstrate the same effect, are segments that, although originating in the exposition section, signal a return to material from the start of the piece and hence do not contribute to a sense of goal-directed progress.) Similarly, segments 16–18 account for the steep gradient at the end of Figure 4, and all relate to the clear transition from the central development section to the final coda. The six segments (10–15) that occupy the flat middle part of the graph all come from the central development section.

There is an alternative kind of explanation for the S-shaped pattern found. It might be that segments from the beginning and end of the piece are well-anchored to a temporal frame, but that listeners are simply less certain about segments from the middle of the piece. As a consequence, there is a tendency for the latter type of segment to “regress” toward the center of the
piece. This account would predict smaller variability in judgments of extracts near the beginning and end of the piece than of extracts from the middle. However, a regression of the standard deviation of the judgments against actual segment locations showed no systematic low-order polynomial components. Thus, we prefer an account of the S-shaped curve involving the kinds of structural properties discussed above.

Lastly, there is no effect of segment type on either the absolute accuracy, or the variability, of subjects’ judgments of the location of a segment. A repeated measures analysis of variance performed on the absolute difference between the judged position of a segment and its actual position (both expressed as a percentage of the length of the whole piece), averaged across all subjects for each of the segments, showed no effect of segment type \( F(2,10) = 0.969, p = .412 \). The variability of subjects’ location judgments, as expressed by their standard deviation, can be taken as an indication of the uncertainty with which a judgment is made. The same analysis with the standard deviation of the judged position across all subjects for the 18 segments as the dependent variable similarly showed no effect of segment type \( F(2,10) = 0.139, p = .872 \). It may be that the relatively long duration of each segment, giving the listeners access to a fair amount of musical information, eliminates any noticeable effect that the presence or absence of a boundary within the segment might have had.

**EXPERIMENT 3: JUDGING SEGMENT DURATIONS AND QUALITIES**

Having discovered something of how segments from the piece are coded in relation to the whole, the purpose of the third experiment was to investigate properties of the individual segments themselves, focusing on perceived duration. The hypothesis under test was that the perceived duration of segments would be affected by structural properties of the segments, both as a result of their intrinsic structure and as a consequence of the way they were extracted from the piece. According to an information-processing approach to duration perception (e.g., Ornstein, 1969; Michon, 1972, 1985), segments with a greater degree of closure, which are therefore easier to encode and are stored more economically, should be perceived as shorter than segments that are incomplete and lacking in unity. For the 18 segments used in the second experiment, this should mean that segments that start or finish at a boundary will be perceived as shorter than those that straddle a boundary, because the latter both begin and end at structurally arbitrary points in the music and should consequently possess very little unity or closure.

The segments also differ in structural complexity depending on where in the piece they come from. Some parts of the piece have a low level of complexity, either because they have a high level of redundancy (such as the
opening section, where a single chord is repeated 229 times), or because they are constructed out of very basic materials: the third measure, for example, which is intended to last for 42 sec, consists simply of a chromatic scale from middle C to the B above it. By contrast other parts of the piece, in particular the developmental middle section, have a high level of complexity, consisting of diverse and rapidly changing materials that incorporate sudden changes in pitch height, dynamic level, event density, attack, and duration. The distinction between segments drawn from these different areas of the piece also should affect their perceived duration according to an information-based model, along the lines of Stockhausen’s (1958) analysis of the Webern string quartet movement, as discussed in the introduction. In order to get a measure of the perceived complexity of the segments used in the experiment, we therefore asked listeners to rate each segment on a number of structural qualities to see whether these ratings bear any relationship to perceived duration. Two differences between our approach and Stockhausen’s are important here. First, Stockhausen’s ratings of complexity are based purely on an analysis of the score, whereas ours are based on empirical evidence from the same listeners that make the time judgments. Second, Stockhausen’s analysis is concerned with a listener’s sense of temporal passage, or momentary tempo, whereas our study focuses on retrospective judgments of elapsed duration. Thus, the experimental situation is not strictly comparable with that considered in Stockhausen’s theoretical analysis.

Methods

Apparatus and Stimulus Materials

These were the same as those used in Experiment 2.

Subjects

The 24 listeners who participated in the experiment were paid £3.00 each. They were all music students at City University, London, and had received an average of 16.8 years tuition on various musical instruments. Nine were first-year students, 13 were second-year students, and two were third-year students. Eleven had participated in Experiment 2 (which was run before this experiment); otherwise listeners indicated no experience with the piece.

Procedure

Listeners were told that the experiment investigated some properties of segments of music extracted from Stockhausen’s Klavierstück IX. In order to familiarize them with the variety of materials contained in the piece, we began the experimental session by playing the entire piece through once. After this, they heard one practice trial and the 18 experimental trials that used the 18 segments from Experiment 2. They were asked to rate each segment on
eight scales. The first seven were seven-point scales with endpoints labeled: Simple—Complex; Static—Active (dynamic); Not at all unified—Very unified; Unchanging material—Varied material; No sense of development—Strong sense of development; Incomplete—Complete; and Not very distinctive—Very distinctive. The last of the eight judgments was of the duration of the segment relative to a 30-sec standard. They were told that the segments they would hear averaged 30 sec in length and that the practice trial would present a segment of approximately that duration. They were to make their duration judgment by drawing a vertical line through a continuous response scale (of length 10.4 cm) on which 30 sec was indicated at the midpoint. The segments were presented in a random order, and listeners were tested in groups.

Results and Discussion

The actual durations of the segments ranged from 24.56 to 32.37 sec (with an average of 29.84 sec), measured from the onset of the first note to the onset of the last note. These values correlated significantly with the average of the listeners’ duration judgments for the 18 segments \( r(16) = 0.808, p < 0.001 \). The correlations computed for individual listeners averaged 0.370, which is significantly greater than zero \( t(23) = 7.013, p < 0.001 \). These individual subject correlations did not covary with any aspect of the listeners’ music backgrounds (years of tuition on musical instruments, year at university, or whether or not they had participated in the previous experiment).

To examine whether the veridicality of the duration judgments depended on the fact that listeners were explicitly instructed to judge durations, we also considered data from Experiment 2. In that experiment, listeners were instructed to judge the location of the beginning and end of the segments, not their durations per se. However, the length of the interval between these two judgments can be taken as an indirect measure of perceived duration. These values correlated significantly with the actual durations \( r(16) = 0.568, p = 0.014 \), and the correlations for individual listeners averaged 0.295, which is significantly greater than zero \( t(22) = 6.312, p < 0.001 \). Once again these correlations did not covary with music background variables (years tuition, year at university), but did covary weakly with the accuracy of their location judgments \( r = 0.386, p = 0.069 \). Thus, for both direct and indirect measures, listeners were quite veridical in judging relative durations—a striking result given the small range of actual durations and the extremely varied nature of the musical material contained in the excerpts.

The next analyses examined the deviations of the direct and indirect duration judgments from veridical values. These analyses correlated the residuals of the regression equations (the direct and indirect measures predicted by the actual durations) with the seven other quality judgments (complex, active, unified, varied, sense of development, complete, and distinctive). No
significant correlation was found with any of these variables, nor any consistent effect of segment type (whether a segment began at a boundary, ended at a boundary, or spanned a boundary). To summarize the analyses of the duration judgments (both direct and indirect), the listeners were quite veridical in judging the relative durations of the segments and, to the extent that their judgments deviated from veridicality, the deviations could not be traced to musical attributes reflected in any of the other quality judgments made in this experiment.

Of subsidiary interest were the quality judgments themselves. To investigate the relationships between these variables, a factor analysis was computed, which revealed two underlying factors. The first factor showed a heavy weighting from four variables: active, complex, varied, and sense of development. The second factor received heavy weightings from the remaining variables: distinctive, unified, and complete. The segments that received low ratings on the four variables loading on the first factor were those that might be characterized as low in information. Relatively low ratings went to segments containing repeated chords (measures 2 and 4–14), the ascending chromatic scale (measure 3), and slow passages containing relatively few events (measures 95–104 and 114–116). Passages with more rapid and diverse events received high ratings (measures 46–56, 51–62, 82–94, and 115–120). The second factor corresponded to materials that are typical or characteristic of the piece: the repeated chord (measures 2 and 4–14), the ascending chromatic scale (measure 3), and the unmeasured, high melodic material from the end of the piece (measures 117–126). None of the quality judgments correlated with the number of attacks in the segments or with the number of different vertical pitch-class collections. When variables measuring accuracy of location and duration judgments were entered into the factor analysis, they did not load on either of these factors, but constituted an additional third factor.

The idea that perceived segment duration might be related to structural characteristics of the music seems not to be borne out by the results of this experiment. Neither the direct duration judgments from this experiment nor the indirect judgments from Experiment 2 correlate with any of the seven measures of perceived structural quality. One possibility is that the subjects were unable to use the seven scales in a coherent fashion and that the quality judgments themselves are unreliable as a consequence. However this seems not to be the case, since the factor analysis shows that the seven quality judgments themselves group into two coherent factors in a readily interpretable manner. A more likely possibility, to be discussed in more detail later, is that judgments of duration are made, at least by musicians with considerable performing experience, with respect to an internal clock of the sort described earlier. It may be that reliance on an internal clock, rather than on musical information, is also in part a consequence of the unfamiliar
style. Thomas and Cantor's (1978) model of the relationship between information content and the output of an internal clock in determining a duration judgment predicts that the less task-oriented the subject is, the greater is the influence of the internal clock. If listeners find the music difficult and unfamiliar, they may engage with the task only at a rather superficial level and base their duration judgments primarily on clock output. This possibility is considered in three subsequent experiments that use a piece written in a very different musical style.

Experiments 4–6: Mozart's Fantasie in C minor, K. 475

The three experiments that follow were conducted to investigate possible effects of the musical style of the Klavierstück IX on the results obtained. For this purpose we used a comparison piece which, while sharing certain of the global characteristics of the Stockhausen piece, was conventionally tonal and metrical in its structure. In order to use the same experimental tasks had been used with Klavierstück IX, the piece had to have the following properties:

1. A total duration of about 10 min.
2. A reasonable number of different musical ideas.
3. As little literal repetition as possible.
4. A phrase structure that was not broken up into a large number of short units.
5. A piece that would be unknown to the majority of our listening subjects.

These criteria, in particular the fourth, seemed to suggest a piece of eighteenth or nineteenth century piano music. With the additional constraint of being confined to the current repertoire of a cooperative performer who would be prepared to play the piece under conditions similar to the recording of the Klavierstück IX, we decided to use the Fantasie in C minor, K. 475, by W. A. Mozart. The piece, which comes from fairly late in Mozart's output, is striking in its tonal diversity and the multiplicity of different musical ideas, these being contained within five sections differentiated by tempo and meter (adagio and allegro in \( \frac{3}{4} \); andantino and piu allegro in \( \frac{1}{2} \); and adagio in \( \frac{2}{4} \)). The last of these sections consists of a varied repetition of material presented in the first 20 measures of the work and in this respect conforms to the idea of recapitulation that is central to the sonata form of music of this period. In other respects, however, the piece is rather unusual.
Fig. 5. The opening page of Mozart's Fantasie in C minor, K. 475 (20 Mai 1785, Wien). Reprinted with permission from Wiener Urtext Edition, Musikverlag Ges. m. b. H. & Co., K. G. Wien.
and reflects the "fantasy" or improvisation that its title suggests. There is no straightforward linear development of material through the piece, nor much evidence of the systematic development of previously presented material that is typical of development sections in sonata movements. Rather, the piece contains a number of sudden contrasts and surprising transitions from one kind of material to another. These features make the piece particularly suitable for our purposes in two respects: first, the absence of repeated material other than at the very beginning and end of the piece eliminates the possibility that listeners' location judgments might be affected by material with more than one section of origin; second, the diversity of musical material in the piece allows us to investigate the possible influence of structural features on duration judgments in a variety of contexts. Figure 5 shows the opening page of the score to give an idea of the variety of musical textures within even this single section.

**EXPERIMENT 4: IDENTIFYING BOUNDARIES**

The method used here to identify boundaries in the *Fantasie* is similar to that used with the *Klavierstück IX* (Experiment 1). The listeners had comparable musical backgrounds, and performed a similar task, although with a more familiar musical style. We anticipated that tonal structure, and contrasts of meter and tempo, might play a significant part in determining listeners' boundary judgments in a manner precluded by the atonal and ametrical style of the *Klavierstück*.

**Methods**

**Apparatus and Stimulus Materials**

The stimulus materials were based on a performance of W. A. Mozart's *Fantasie* (K. 475) by Mark Lockett, a freelance professional pianist. The performance was played on a Yamaha KX-88 keyboard with a piano timbre produced on an Akai S-900 digital sampler, whose audio output was amplified and played over loudspeakers. A Macintosh Plus computer (with MIDI interface) recorded the timing (onset and duration) and velocity of each key press using the *Performer* software. Two complete recordings of the piece were made, and the performance judged more satisfactory by the performer was used in the experiment. Its total duration was 10.51 min. The performance was played back during the experiment with the same equipment; a footpedal connected to the KX-88 keyboard was used to record subjects' responses during the second part of the experiment.

**Subjects**

The four volunteer participants were postgraduate students in composition at the Music Department, City University, London. All four listeners had heard the piece before the experiment; one also had some experience analyzing and playing it.
Procedure

The participants were told that the experiment investigated the perception of structural organization in a piece of classical music, Mozart's Fantasie. The procedure was in all other respects the same as in Experiment 1.

Results and Discussion

Listeners varied in the number of boundaries they indicated, ranging from 8 to 37. However, they showed considerable agreement on the boundaries they considered the strongest. Six boundaries were agreed on by all four listeners; four other boundaries were identified by three of the four listeners. The average judged boundary strengths for these 10 boundaries are shown in Figure 6. The ratings show a number of relatively strong boundaries located throughout the piece. In general, the boundaries were judged as relatively easy to locate; ratings of six or seven on the seven-point scale were given 75% of the time for all boundaries indicated. The primary purpose of this experiment was to obtain a basis for segmenting the piece for the experiments that follow. However, it may be worth mentioning in passing that all the main divisions indicated in Mozart’s score with double bars (at measures 26, 36, 85, 124, and 160) are also perceived as boundaries by the four subjects. These divisions correspond to changes in tonality (e.g., measure 26), tempo (e.g., measure 36), or meter (e.g., measure 85) and in most cases by combinations of two or more of these. This suggests
that at the largest scale, the formal structure of the music from a compositional perspective coincides very well with its perceived formal structure. It is also worth noting that, as with Experiment 1, there is a high level of inter-subject agreement about the strongest boundaries in the piece but that there is a more even distribution of strong boundaries throughout the piece than was found in the Klavierstück IX, where there was an absence of strong boundaries in the middle of the piece.

EXPERIMENT 5: LOCATING SEGMENTS

Experiment 5 is equivalent to Experiment 2 and is motivated by the same general questions. The only difference in the procedure for this experiment was the inclusion of an additional task in which the subjects indicated the strength of each of the boundaries used in the experiment, and the kinds of features that were responsible for establishing the boundary. The purpose of this additional task was to obtain information concerning the judged strength and nature of each boundary from the same subjects as had made the location judgments, in order to establish whether any systematic relationship existed between the two sets of judgments. Before turning to the experiment itself, we should consider the possible consequences of the different structure of the Fantasie as compared with Klavierstück IX. We might expect the more familiar style of the music, and in particular its metrical structure, to make it easier for listeners to keep track of the music and subsequently locate extracted segments. However, as observed earlier, the music has no continuous progression or linear development of the kind that exists in Klavierstück IX, and this may make it more difficult to make veridical location judgments. The basic progression from chordal to linear material in Klavierstück IX, and the division of the piece into three main sections, may allow listeners in the location judgment task to get their approximate bearings in the piece in a way that is not possible in the Fantasie.

Methods

Apparatus and Stimulus Materials

The apparatus was the same as in the previous experiment. Eighteen segments, of approximate duration equal to 30 sec, were extracted from the piece. The segment durations averaged 29.62 sec, and ranged from 22.60 to 34.56 sec. The segments were selected as follows. The six strongest boundaries identified in the previous experiment were considered. One of these boundaries (at measure 56) was eliminated because a segment beginning at the previous boundary (measure 56) would extend up to measure 56. The next strongest boundary (at measure 101) was therefore selected. As in Experiment 2, segments of three types were constructed around each of the six boundaries, ending at each boundary, beginning at each boundary, and spanning each boundary.
Subjects

The 24 listeners who participated in the experiment were paid £3.00 each. They were all music students at City University, London. On average, they had received 16.5 years tuition on various instruments. Five listeners were first-year students, 12 were second-year students, five were third-year students, and two were postgraduate students. Five knew the piece previously as listeners, including three who had also played it.

Procedure

Listeners were told that the experiment investigated the perception of the location of short musical segments in relation to the whole of Mozart's Fantasie. The procedure was the same as in Experiment 2 with the following exception. After completing the location judgment task, the piece was played from the beginning, stopping shortly after each of the six boundaries around which the segments had been constructed. For each stopping point, listeners were asked to rate the strength of the boundary on a seven-point scale. They also were asked to describe the features that helped form the boundary.

Results and Discussion

The boundary strength judgments for the six boundaries included in this experiment, which are shown in Table 2, were similar to those in the previous experiment. The two sets of judgments correlated significantly \( r(4) = .925, p = .0081 \). The table also lists the characteristics identified as contributing to the formation of these six boundaries. The characteristics listed are compiled from the descriptions of the six participants in this experiment whose ratings correlated most highly with those of the previous experiment. In addition to the kinds of characteristics noted in Experiment 1, listeners in this experiment noted properties specific to tonal-harmonic music (change of key, change of harmony). As with Experiment 1, the characteristics can be related in a general way to the grouping preference rules of Lerdahl and Jackendoff (1983). Three classes of features are indicated by our subjects:

1. Changes in the surface characteristics of the music (changes of texture, dynamic, register, tempo). These are equivalent to the Gestalt-based grouping preference rule 3 of Lerdahl and Jackendoff.

2. Changes of a structural kind (changes of meter or key, or the introduction of new material). Changes of key are embodied in grouping preference rule 7 (prefer a grouping structure that ties in with the harmonic structure of the music), but changes in meter are not dealt with in Lerdahl and Jackendoff's theory. The "introduction of new material" can in theory be broken down into a number of simultaneous and coordinated changes in the surface characteristics of the music, such as texture, register, dynamic, and melodic shape and hence could be included in the category above. However, if we are to treat this as a response of a different kind, indicating a more significant and structural change, then this cannot be related directly
### TABLE 2
Musical Characteristics Contributing to the Six Boundaries
in Mozart’s Fantasie

<table>
<thead>
<tr>
<th>Measure</th>
<th>Strength</th>
<th>Musical Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>5.250</td>
<td>Change of key (minor to major) (6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of texture (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New material (lyrical) (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic (2)</td>
</tr>
<tr>
<td>36</td>
<td>4.917</td>
<td>Change of key (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of texture (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of register (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New material (dramatic) (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of tempo (3)</td>
</tr>
<tr>
<td>85</td>
<td>4.542</td>
<td>Change of key (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of tempo (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of register (more restricted) (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>New material (end of cadenza) (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of meter (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamics (1)</td>
</tr>
<tr>
<td>101</td>
<td>3.792</td>
<td>Change of register (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of dynamic (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of texture (thicker) (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of rhythm (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of melody (1)</td>
</tr>
<tr>
<td>124</td>
<td>5.167</td>
<td>Change of dynamic (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of tempo (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of register (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of texture (4)</td>
</tr>
<tr>
<td>160</td>
<td>5.826</td>
<td>Return of previous material (6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of meter (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of register (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of tempo (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Change of harmony (cadence) (4)</td>
</tr>
</tbody>
</table>

**Note.** The numbers in parentheses indicate the number of listeners (out of six) noting each characteristic.

This content is attributed to Lerdahl and Jackendoff’s rules because it deals with an associational, and not a hierarchical property, in the music. Lerdahl and Jackendoff state quite clearly in their book that their theory does not attempt to deal with this kind of property (Lerdahl & Jackendoff, 1983, p. 17).

3. The return of previously heard material. This is equivalent to the sixth preference rule (parallelism), which states that segments of music that can be construed as repetitions, or variant repetitions, of one another form parallel groups. This is mentioned only rarely by our listeners. These three categories correspond to three of the four categories identified in Experiment 1.
(Table 1), the additional category in Experiment 1 being concerned with silences or pauses in the music. The absence of this category in the Fantasie judgments simply reflects the more continuous style of this music.

Turning now to the location judgments, the midpoint of the time interval indicated by each listener for each segment was measured, as in Experiment 2. The median judgments for each segment are shown in Figure 7, plotted as a function of the actual midpoints. The diagonal line indicates perfect accuracy, and the figure shows that listeners' judgments deviated from this line. Despite these deviations, there is a strong effect of the actual location of the segment; the median values correlated significantly with the actual locations of the segments \[ r(16) = .906, \ p < .0001 \]. The accuracy of individual subjects' judgments (assessed by correlating their judgments with the actual locations) did not vary systematically with quantifiable aspects of their musical backgrounds (years tuition on musical instruments or year at university). These individual subject correlations ranged from .000 to .933, with an average of .490.

The group data were entered into a nonlinear regression to evaluate

![Graph](Fig. 7. The judged location of segments from Mozart's Fantasie in C minor plotted against their actual position in the piece. The curved line shows the third-order polynomial regression solution for the data.)
whether the deviations from veridicality were systematically related to segment location. Significant contributions were made by cubic and quadratic components, the regression giving an \( r = .933 \) \( [F(3,14) = 31.579, \ p < .0001] \), with significant weights for both the cubic \( [t(14) = 2.518, \ p = .0246] \) and quadratic \( [t(14) = 2.457, \ p = .0276] \) components. The curved line in Figure 7 shows the regression equation. An identical analysis of the individuals’ judgments showed the same pattern as for the group data: a negatively accelerating function changing to a positively accelerating function near the middle of the piece. Thus, the pattern of deviations from veridical judgments is typical of the individual subject data. As in Experiment 2, the standard deviations of the judgments did not depend systematically on segment location.

Comparison of Figures 4 and 7 shows similar functions relating judged location to actual location for the two pieces; both figures show the same type of S-shaped curve. Our explanation for the pattern of location judgments is therefore the same as for Klavierstück IX, namely that the judgments reflect changes in the sense of goal-directed progress in the music. However, the nonlinearity is considerably more pronounced for the Stockhausen piece than for the Mozart. There is a variety of differences between the pieces, but two features mentioned in the introduction to this experiment may have a particular bearing on the location judgments. One is the continuously metrical structure of the Fantasie by contrast with Klavierstück IX. This metrical structure, despite tempo and time signature changes, may provide the music with an internal temporal reference framework which allows listeners to maintain a reasonably veridical sense of temporal location. By contrast, in the Klavierstück the absence of a continuous metrical framework may mean that a listener’s sense of location is far more directly determined by a more global sense of “how much has happened” without the corrective of the loosely chronometric frame that even a changing meter provides. The second feature that may account for the greater veridicality of the location judgments is the more even distribution of strong boundaries throughout the Mozart, by comparison with Klavierstück IX. If we ignore the somewhat weaker boundary at measure 101, the five remaining boundaries in the Fantasie (which are all of strength 6.25 or more in Experiment 4, out of a maximum of 7) divide the piece into six sections of very similar duration. By contrast, only two boundaries in the Klavierstück exceed a value of 5 (at measures 17 and 117), dividing the piece into three large sections. It seems likely that the greater number of shorter sections in the Mozart may help listeners to identify, in at least an approximate fashion, the original location of an extract.

Lastly, there is no evidence that either the accuracy or the variability of subjects’ location judgments are affected by whether a segment begins at a boundary, ends at a boundary, or spans a boundary. A repeated measures
analysis of variance with the absolute difference between the judged location of the segment and its actual location (both midpoints expressed as a percentage of the length of the piece) as the dependent variable, averaged across all subjects for each of the 18 segments, showed no effect of segment type \[ F(2,10) = 0.592; p = .572 \]. The same analysis with the standard deviations of the midpoint judgments as the dependent variable was similarly nonsignificant \[ F(2,10) = 0.706; p = .517 \]. This is essentially the same result as that for Experiment 2 with Klavierstück IX, so that the failure to find an effect of segment type on the accuracy of listeners' location judgments cannot be simply attributed to stylistic unfamiliarity.

**EXPERIMENT 6: JUDGING SEGMENT DURATIONS AND QUALITIES**

The final experiment in this study used the same techniques as Experiment 3 to investigate segmental duration judgments and structural qualities in the Mozart Fantasie. One rather surprising result of Experiment 3 with Klavierstück IX was the apparent independence of duration judgments from information content. A possible interpretation of this result was that the unfamiliar musical style made coding information content difficult. As a consequence, only veridical relationships were found between the actual duration and the judged duration of the segments, presumably based on the output of some kind of internal clock. With a more familiar musical style, more attention might be directed toward the information content of the segments, with a corresponding reduction in the veridicality of the judgments and increased evidence of the effect of segmental structure.

**Methods**

*Apparatus and Stimulus Materials*

These were the same as those used in the previous experiment.

*Subjects*

The 25 listeners who participated in the experiment were paid £3.00 each. They were all music students at City University, London, and had received an average of 16.1 years tuition on various musical instruments. Six were first-year students, eleven were second-year students, four were third-year students, and four were postgraduate students. Seventeen had participated in Experiment 5 (which was run before this experiment). Five indicated that they were previously familiar with the piece as listeners, including three who had also played it (although none had played it recently).

*Procedure*

Listeners were told that the experiment investigated some properties of segments of music extracted from Mozart's Fantasie. The procedure was the same as in Experiment 3.
Results and Discussion

The actual durations of the segments ranged from 22.60 to 34.56 sec (with an average of 29.62 sec), measured from the onset of the first note to the onset of the last note. The actual durations of the segments correlated significantly with the average of the listeners' duration judgments for the 18 segments \[ r(16) = .727, p = .0006 \]. The correlations computed for individual listeners averaged .431, which is significantly greater than zero \[ t(24) = 10.627, p < .0001 \]. These individual subject correlations did not covary with any aspect of the listeners' backgrounds (years of tuition on musical instruments, year at university, or whether or not they had participated in the previous experiment).

To examine whether the veridicality of the duration judgments depended on the fact that listeners were explicitly instructed to judge durations, we also considered the indirect duration judgments from Experiment 5 (the length of the interval between the judged beginning and ending of the segments). These values also correlated significantly with the actual durations \[ r(16) = .684, p = .0017 \]. The correlations for individual listeners averaged .205, which is significantly greater than zero \[ t(2.3) = 4.008, p = .0006 \]. These correlations did not covary with music background variables (years tuition, year at university), nor with the accuracy of their location judgments. Thus, for both direct and indirect measures, listeners were quite accurate in judging relative durations.

The next analyses examined the deviations of the direct and indirect duration judgments from veridical values. These analyses used the residuals of the regression equations (the direct and indirect measures predicted by the actual durations). These residuals were correlated with the seven other quality judgments (complex, active, unified, varied, sense of development, complete, and distinctive). Of the 12 correlations, only two were significant: the correlation between the residuals of the indirect duration judgments and ratings of how varied segments were \[ r(16) = .604, p = .0079 \] and the correlation between the indirect judgments and ratings of distinctiveness \[ r(16) = .519, p = .0274 \]. For both direct and indirect measures, there was a significant effect of whether a segment ended before a boundary, began after a boundary, or spanned a boundary \[ F(2,15) = 4.007, p = .0403 \], and \[ F(2,15) = 8.179, p = .0040 \], respectively]. However, the effects were different in the two cases; for the direct measures, the residuals were smallest for the segments spanning boundaries, whereas for the indirect measures, the residuals were largest for the segments spanning boundaries. In fact, the residuals for the two measures tended to show opposite effects in a number of cases and correlated slightly negatively with one another \( r = -.129 \).

Of subsidiary interest were the relationships between the quality judgments themselves, which were examined using a factor analysis. This analy-
sis revealed two underlying factors. Four variables weighted heavily on the first factor: active, complex, sense of development, and distinctive. The remaining variables weighted heavily on the second factor: unified, not varied, and complete. Four objective measures of characteristics of the segments also were measured: the number of attack points, the number of harmonic changes, the total number of different harmonies, and the number of implied keys. These four characteristics loaded on the first factor when entered into a factor analysis with the quality judgments. When variables measuring accuracy of location and indirect duration judgments were entered into the factor analysis, they did not load on either of these factors, but constituted an additional, third factor. The accuracy of the direct duration judgments, however, loaded on the second factor.

The duration judgments, both the direct judgments from this experiment and the indirect judgments from Experiment 5, can be summarized as follows: the listeners were quite accurate in judging the relative durations of the segments, as the correlations with actual duration showed. To the extent that their judgments deviated from veridicality, the deviations could not be traced consistently to musical attributes reflected in any of the quality judgments in this experiment. The quality judgments themselves, as in Experiment 3, seemed to cluster together in a readily interpretable manner, suggesting that the failure to find a relationship between the quality and duration judgments is not simply a result of unfamiliarity with the style. The fact that in the present experiment the four objective measures of musical structure load on the first of these two factors supports the idea that listeners are making the quality judgments in accordance with objective properties of the music: all four objective measures are different aspects of structural complexity and all four load on to the factor with perceived activity, development, variety and distinctiveness. The somewhat surprising result, therefore, is that in the context of both the indirect duration judgments of Experiments 2 and 5 and the direct judgments of Experiments 3 and 6, listeners are quite veridical and appear not to be affected by structural attributes of the segments.

General Discussion

Taking the general conclusions of this study in the order of the experiments themselves, the two segmentation experiments (Experiments 1 and 4) showed that with contrasting styles of music, listeners used segmentation criteria that were broadly consistent with the predictions of Lerdahl and Jackendoff (1983). Furthermore, although Lerdahl and Jackendoff’s predictions are primarily aimed at comparatively low levels of grouping structure, our results show that listeners continue to use the same kinds of crite-
ria at higher levels of structure where groups have a duration of the order of 30–50 sec or more. It is striking (and supportive of Lerdahl and Jackendoff's claims for the universality of many of their grouping rules) that with the two very different pieces used in this study, listeners cite similar features as indicators of segment boundaries. There are only two differences between Tables 1 and 2: the greater emphasis on pauses and silences as grouping factors in the Klavierstück than in the Fantasie, which is a direct result of the more continuous nature of the music in the Mozart, and the greater importance of repetition as a boundary indicator in the Stockhausen.

Turning to the location judgments (Experiments 2 and 5), a striking feature of the results is their similarity for the two pieces. Not only did listeners demonstrate a high level of veridicality in their judgments for both pieces, but the functions plotting their departures from veridicality also were remarkably similar. This suggests either that listeners' memory representations for the two pieces have similar large-scale characteristics or that the two pieces share some high-level property on which listeners focus. Given the notable absence of information on the nature of listeners' memory representations for large-scale musical structure, and the limited nature of the evidence available from this study, it is impossible to develop specific proposals here. However, we suggest that a combination of our listeners' attentional strategy and a very general property of the musical structure may underlie our results. The S-shaped function that describes the departures from veridicality of our listeners' judgments implies that for both pieces listeners experience the music as advancing more rapidly at the beginning and the end than in the middle: the function describing their results accelerates at these points in both figures. This could be explained by a heightened sensitivity on the part of our listeners to the presentation of new musical information at the start of the piece and as the end approaches, as the result of more focused attending. This ties in with an objective structural property of the two pieces—and of a great deal of Western music: the music starts with a section in which relatively well-defined ideas are introduced in a clearly distinguishable manner, proceeds to a more developmental section where the separation and identity of ideas is attenuated through techniques of interaction, elaboration and fragmentation, and concludes with a section in which identity and clarity are reestablished. Despite the use of different stylistic techniques, both the pieces of music used in this study project this fundamental structural outline, and it is the basis of most of the forms used in Western music, ranging from sonata form to jazz and popular songs. It may be more appropriate, therefore, to regard attentional strategy and formal structure as an example of mutual development and to acknowledge the possibility that our results are as much a function of our listeners' long-term exposure to a widespread pattern of formal organization as an expression of an intrinsic attentional strategy.
Turning now to perceived segment duration, the failure to find any coherent relationship between the structural quality judgments and the duration judgments of Experiments 2, 3, 5, and 6 is not inconsistent with the interpretation of the location judgments given above. Our understanding of the pattern of results for the location judgments is based on a view of changes in information, and concurrent changes in attention, across whole pieces of music. It thus emphasizes the global context in which musical information processing is carried out. By contrast, the duration and quality judgments of Experiments 3 and 6 are essentially local judgments, focused on the approximately 30-sec segments themselves, rather than the total context from which they come. Furthermore, the task itself eliminates any possible effects originating from changes in listeners' perceptual strategies, since segments were presented in random order and in isolation. The task thus encourages listeners to adopt the same perceptual approach to each segment in turn, rather than reflecting the kind of longer term changes that we have proposed for the more realistic activity of continuous listening to a complete piece. In order to obtain quality and duration judgments that might reflect the informational component of the location judgments of Experiments 2 and 5, a different experimental method would have to be devised that could somehow combine a global listening approach with the need to obtain a more focused segmental judgment. Finally, it may also be significant that our listeners were musicians with a considerable amount of performing experience. One of the skills that a performer must develop is the ability to maintain a tempo despite changes in musical material, and from one occasion to another. Empirical research into this ability has focused on a performer's ability to make reference to an extremely stable internal clock (e.g., Clynes & Walker, 1982; 1986; Shaffer, 1981), and it may be that our listeners thus possess a sense of absolute time that is not representative of nonperforming listeners.1

References


1. The research reported in this paper was supported by grants from The British Academy and The Worshipful Company of Musicians (Allcard Grant) to the first-named author and by the National Science Foundation (INT - 8701948) to the second-named. It was conducted, in part, while the two authors were visiting researchers at IRCAM (Institut de Recherche et Coordination Acoustique/Musique), Paris.
Perceiving Musical Time

Jackson, J. Is the processing of temporal information automatic or controlled? In J. A. Michon & J. L. Jackson (Eds.), *Time, mind and behaviour*. Heidelberg: Springer Verlag, 1985.
Michon, J. A. Processing of temporal information and the cognitive theory of time experience. In J. T. Fraser, F. C. Haber, & G. H. Müller (Eds.), *The study of time*. Heidelberg: Springer Verlag, 1972.
Michon, J. A. The compleat time experiencer. In J. A. Michon & J. L. Jackson (Eds.), *Time, mind and behaviour*. Heidelberg: Springer Verlag, 1985.