Representation of Phrase Structure in the Perception of Music
Author(s): Thomas H. Stoffer
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 3, No. 2 (Winter, 1985), pp. 191-220

Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40285332

Music Perception
Winter 1985, Vol. 3, No. 2, 191–220

Representation of Phrase Structure in the Perception of Music

THOMAS H. STOFFER
University of Bielefeld, Federal Republic of Germany

The degree of correspondence between the formal description of phrase structure in music and its cognitive representation as revealed by its effects on responses to clicks superimposed on melodies is explored in two experiments. In Experiment 1, using a click localization task, a high degree of correspondence was established, providing evidence for a hierarchical format of the cognitive representation of phrase structure. In Experiment 2, click detection time was measured for a binary and trinary phrase structure of the same length, using two groups of subjects differing in musical competence. The results support a schema–theoretic description of tacit knowledge of phrase structure in music.

Introduction

One of the general principles in theoretical accounts of cognitive processing that is now emerging seems to be the hierarchical structuring of internal representations of objects (Cooper, 1980). Despite substantial disagreements between the different schools of music analysis (e.g., Lerdahl & Jackendoff, 1983; Narmour, 1977; Nattiez, 1975; Sundberg & Lindblom, 1976) there is growing agreement among music theorists and cognitive psychologists that perceived musical structure, also, is internally represented in the form of hierarchies (e.g., Deutsch & Feroe, 1981; Jackendoff & Lerdahl, 1981; Jones, 1978; Longuet-Higgins, 1976; Narmour, 1984; Stoffer, 1985).

The present research is intended to examine further this hypothesis. In particular, it focuses on two issues. First, it considers the question of a structural correspondence between the formal description of music as a manifestation of certain phrase structures and the description of their cognitive representation as a manifestation of similar internal structures of vertically and horizontally organized functional units (Experiments 1 and 2). Second, it explores the possibility that listeners use tacit knowledge about phrase

Requests for reprints may be sent to Dr. T. H. Stoffer, Abteilung für Psychologie, Universität Bielefeld, Postfach 8640, 4800 Bielefeld 1, West Germany.

structures that are to be expected in certain musical forms. The organization of this kind of knowledge may depend on the amount of experience listeners may have had in the past with different phrase structures in music. It is proposed that this knowledge structure may be organized in the form of hierarchies of representations of musical concepts. The evaluation of this hypothesis is the subject of Experiment 2.

The Format of a Representation of Musical Structures

There is some empirical evidence in favor of the hypothesis that the representation of perceived music can be described in hierarchical form. Using a memory task with recall in musical notation, Deutsch (1980) found that the percentage of tones in a sequence that were correctly recalled was higher for sequences that were structured hierarchically by the application of transformations on subsequences of three tones than for sequences that were either not structured in that fashion, or that were so structured but incompatibly segmented. (Incompatible segmentation was accomplished by the insertion of pauses after every four elements instead of three.)

Stoffer (1981) has shown in a same–different reaction time experiment that the time for memory access to the interval configuration formed by all the four tones of a sequence is significantly shorter than the time it takes to access the constituent configuration of the first or the last pair of tones. This result is in line with the hypothesis of Deutsch and Feroe (1981) that retrieval in a coded hierarchy should be in a top-down fashion.

Apart from research on rhythm perception and production (e.g., Deutsch, 1983; Fraisse, 1982; Martin, 1972; Povel, 1981), another line of evidence comes from research on perceptual grouping and segmentation. Grouping and segmentation are important functional principles in establishing the hierarchical format of a representation. Perceptual grouping has been demonstrated by Garner and associates (Preusser, 1972; Preusser, Garner, & Gottwald, 1970; Royer & Garner, 1966, 1970) and Handel and associates (Handel, 1973, 1974; Handel & Pickens, 1977; Handel & Yoder, 1975).

Dowling (1973) investigated the effect of chunking on memory for short tone sequences. Subjects were asked to recognize a short tone sequence that was either a temporally coherent group marked by a pause at its end, or assembled from consecutive parts of two adjacent groups of a sequence that had been heard before. Recognition of an entire chunk was better than of a sequence that was assembled from two adjacent groups. Grouping and chunking in these experiments formed a representational hierarchy at two levels: the level of chunks or groups and the level of single tones.

Chunking on the basis of harmonic progressions was demonstrated by Tan, Aiello, and Bever (1981). Subjects had to indicate where a two-tone probe occurred in a sequence that had been immediately heard before. Probes were either the two tones that ended the first harmonically defined phrase, or two tones straddling the phrase boundary, or the two that began the second phrase. The results showed that the probe straddling the phrase boundary was more difficult to recognize than either of thee within-phrase probes.

In addition to these findings, the data of Tan et al. (1981) point to a special sensitivity of musically trained listeners to harmonic structure: The effect of probe position was larger for trained than for untrained listeners. Further, whereas untrained listeners did not exhibit a difference in the probe position effect between full cadences and semicadences, trained listeners showed a more marked effect of probe position for full cadences than for semicadences.

This pattern of results suggests that there may not only be an effect of perceptual grouping on the basis of a bottom-up analysis of tonal relationships, but also an effect of a segmentational process operating in a top-down fashion. Musically untrained subjects may not take a semicadence as an indication of a phrase structure boundary. Lacking a definite temporal reference point, they may have difficulty in locating the correct memory unit for comparison with the probe. But musically trained subjects may provide this kind of reference point by chunking all the tones in a phrase into a higher order representational unit so that memorial subunits for comparison with a probe can easily be located relative to a chunk boundary.

Localization of Clicks in Music

An experimental approach focusing on the interaction of perceptual grouping and segmentation is the click localization method. Ladefoged and Broadbent (1960) observed that subjects reported the position of a click superimposed on a spoken sentence as having occurred one or two syllables earlier. They interpreted this migration of the click as an expression of the tendency of discrete perceptual units to maintain their integrity. This, they hypothesized, causes clicks to be perceived only at those positions marked by boundaries between speech units. The question of structural correspondence between the units into which a sentence can be segmented for the purpose of linguistic description, and the perceptual units involved in speech recognition, was first raised by Fodor and Bever (1965). They found that a click embedded in a spoken sentence tended to be located closer to the nearest major phrase structure boundary than to its actual position. This result points to a correspondence between syntactic descriptions of linguistic segments and functional units in speech perception. In particular, the fact that displacement toward phrase structure boundaries is greater, the higher the order of the boundary (e.g., Holmes & Forster, 1970), indicating a hierarchical structure of units in speech perception, qualifies his paradigm as a tool for exploring representational structures in auditory perception of other types of sequential patterns also.

The first experiment of this kind with tonal sequences was conducted by Gregory (1978). The subjective temporal location of a click presented during a six-note sequence of tones depended much upon the way the tonal sequence was visually segmented on the score presented to the subjects while listening to the sequence. When the sequence was notated as two groups of three eighth notes, a click on the third note migrated into the phrase structure boundary between the third and the fourth notes. When the sequence was notated instead as three groups of two notes, a click at the same position migrated in the opposite direction between the second and the third note.

There are three methodological objections to the procedure used by Gregory (1978). First, there was a visual segmentation on the score that might have been the only cue for cognitive segmentation. So this experiment did not demonstrate that listeners can detect segmentational cues in musical structures just by listening. Second, the sequences did not exhibit any kind of musical regularity that could have been recognized by a listener. Without that, no top-down controlled cognitive segmentation is possible that will go beyond the mere detection of chunk boundaries formed by a bottom-up analysis of the sequence. Third, the click localization method is susceptible to response biases (Bertelson & Tisseyre, 1970; Bever, 1973; Seitz & Weber, 1974). A systematic displacement that is dependent on syntactic structure and independent of response biases occurs only when subjects are primarily attentive to the stimulus pattern in the one ear and not to the click in the opposite ear.[1] Attentional focusing on musical structure may be achieved by a primary task to be performed on the basis of information extracted from the music. Localization of the click has to be a secondary task. If it were the other way round, the click would be located relative to the structure of the representation formed only after the click was detected and attention was switched to the musical structure. In that case, click displacement would always have to be in the direction of a late localization. That is exactly what may have happened in the experiment of Gregory (1978): On the average, the click was perceived later than its actual time of occurrence,

1. There seems to be no complete agreement on the issue whether one can legitimately say that listeners actually perceive the clicks at the displaced positions. The alternative view is that they do not perceive any displacements, but rather produce some kind of a response bias that does not concern the form of the task as discussed here. There is no direct positive evidence for either view, but at least there is some evidence in the results of Dalrymple-Alford's (1976) experiment that rules out explanations completely relying on response bias effects (see Discussion section).

whereas in speech it is usually perceived earlier (Fodor & Bever, 1965).

If these objections are correct, Gregory's (1978) results show only an effect of visual grouping of the notes on click localization. But that is not the kind of result the experiment was aiming to produce.

The first and the last objection apply equally to another click localization experiment, by Sloboda and Gregory (1980). Again, the score could be visually inspected all the time during listening. Each melody consisted of a four–measure phrase composed of two subphrases. Boundaries between subphrases were marked either by a relatively long note and/or by concluding each subphrase with a cadence. In one condition, neither kind of phrase marker was used. Melodies in conditions with a physical marker were apt to be influenced by the visual structure of the score, so it was not surprising to find that in this case click localization was quite accurate. The click migration effects in the condition with only a structural marker may be interpreted as demonstrating a segmentational process relying on the recognition of phrase structure. Because all subjects were proficient music readers, effects, which were interpreted in terms of a hypothesis of structurally guided cognitive segmentation, might instead have been visually mediated.

In order for a click localization experiment to be immune from the foregoing objections, it must fulfill the following requirements: (1) In order to rule out any effect of visual segmentation, subjects should not be allowed to read the score until the click is detected. (2) In order to reduce the effect of response bias, attention must be focused primarily on the musical structure. This can be achieved by asking subjects to perform a task that forces them to attend to the music. (3) The musical material used should exhibit genuine musical regularities that can then function as phrase markers. Phrase markers may, for example, be changes in melodic contour (Dowling, (1978), melodic regularities formally described as transformations (cf. Deutsch & Feroe, 1981; Jones, 1978), harmonic progressions (Cuddy, Cohen, & Mewhort, 1981), especially cadences (Tan, et al., 1981), rhythmic regularities (Martin, 1972), patterns of pauses (Handel, 1973), and relative note durations (Sloboda & Gregory, 1980).

Correspondence between Description and Representation of Music

Because the object of a musical syntax is to describe music without the necessary consideration of a potential listener, it may be sufficient for musicology to judge a musical syntax exclusively by its own criteria of descriptive adequacy.[2] But as soon as a model for the description of music claims to

2. A syntax is said to be descriptively adequate if it assigns exactly one structural description to a musical score (analytical adequacy) or if it assigns exactly one musical score to a structural description (generative adequacy).

incorporate psychological aspects of musical listening, for instance, if it attempts to model the kind of structure a listener hears and the kind of knowledge he is using to represent that structure, then it is not sufficient to prove its descriptive adequacy; it must also establish its cognitive adequacy (Watt, 1973).

A musical syntax may be said to be cognitively adequate to the degree that it is compatible with the existing psychological evidence. For instance, Jackendoff & Lerdahl's (1981) reference to parallels between their "Grouping Preferences Rules" and Koffka's (1935) account of grouping in vision is an attempt to establish some degree of cognitive adequacy. One promising cooperation between musicology and the cognitive psychology of music is therefore to establish correspondences between music–syntactive descriptions on the one hand and structural aspects of cognitive processing on the other, by relating properties of the musical structure to properties of the representation. We may suggest that each structural unit in the formal description of a piece of music corresponds to a potential unit in a listener's representation. This implies not only a common format, but also a common organization of structural units at each level of the hierarchy.

Analytically one may parse a piece of music into a large number of levels, but a listener may only be able to attend to a few of these. He may not be able to go up beyond a certain representational level, because he may not know how to integrate units into higher ones, or he may be unable to represent a lower level because he does not know how to achieve the parsing required to attend to the units at that level. So it may be unrealistic to hope that for every listener one could find evidence of a cognitive representation of all the structural units the syntactical analysis has produced. How complete a representation is established depends primarily on the segmentational capabilities of the listener and the task requirements of the experimental procedures. A syntactical analysis can only establish structural boundary conditions for cognitive processing, that is, it may describe *potential* representational units, but it cannot give any hints as to those units actually being processed by a listener at a certain time and in the course of certain action. This does not impair the heuristic value of establishing correspondences as long as the number of functional units is a genuine subset of the number of structural units the syntactic description model enumerates.

To summarize, in the course of a review of the experimental evidence in favor of the hypothesis that there is a common format for syntactic descriptions and cognitive representations of musical structure, we gave a critical appraisal of research on the representation of tonal sequences employing the click localization method. This review has revealed some methodological shortcomings in previous applications of the click localization method to music perception that may be avoided by certain proposed modifications. It is argued that correspondences between the syntactic description of music and the representational structure of perceived music can be proposed that lead to extensions of the common–format hypothesis. Certain hypotheses concerning such correspondences were subjected to empirical test in Experiment 1, which used an improved version of the click localization method.

Experiment 1: Localization of Clicks Superimposed on Melodies

The main purpose of the experiment was an extended replication and test of the generality of the results obtained by Gregory (1978) and Sloboda and Gregory (1980) using the click localization task. The extensions related to the following two points: First, in both studies only a single phrase structure boundary was used, and this divided the tonal sequence into halves. Click migration in the direction of his boundary produced evidence for only two representational levels, one level consisting of the two halves and the other representing the sequence as a whole. The design of our experiment was extended with respect to the number and order of phrase structure boundaries tested, so that comparisons of click localizations at phrase structure boundaries of different orders could be made. Second, because both previous studies did not make explicit the complete phrase structure of the melodic sequences used, they were unable to answer the question to what extent a correspondence exists between the phrase structure of a melody as described syntactically and the structure of its representation as revealed by click migration tendencies. Using melodies with a phrase structure that can be described objectively and explicitly at all levels of the structural hierarchy, we hoped to approach an answer to this question.

Experimental Design

The design of the experiment consisted of a click localization task in which one of 17 different click positions centered around the middle of an eight–measure melody was tested with the presentation of each melody. The melodies were short artificial "children's songs" generated by using a transformational syntax of German folk songs (Stoffer, 1979) with a regular binary phrase structure such as the one shown in Figure 1. They covered all possible orders of phrase structure boundaries: First-order boundaries are those between the two tones of a foot (all uneven numbered positions), second-order boundaries are those between two feet of a measure (positions 10, 14, 18, and 22), third-order boundaries are those between the two measures of each subphrase (positions 12 and 20), fourth-order boundaries are those between subphrases at positions 8 and 24, and the highest-

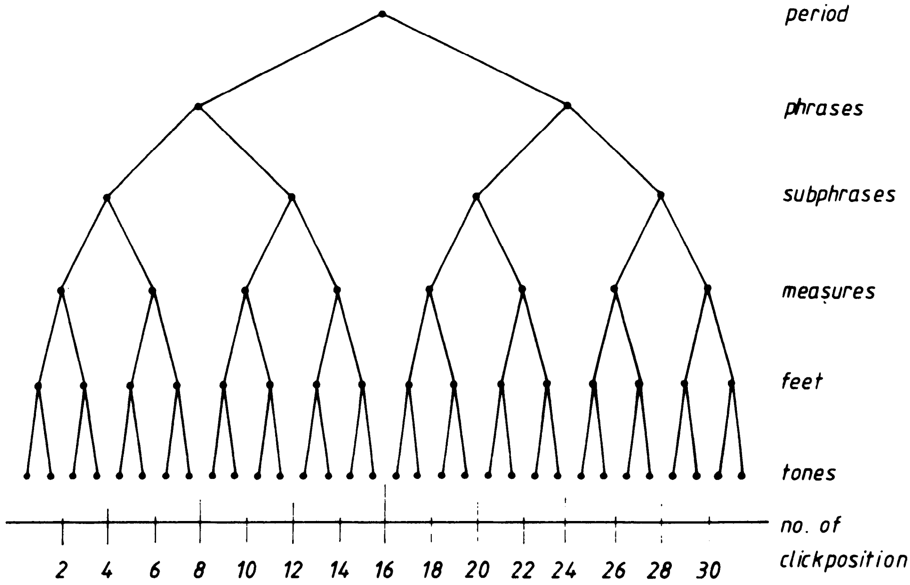

Fig. 1. Phrase structure of melodies used in Experiment 1 with click positions numbered from 1 to 31.

order boundary is at position 16, dividing the melodies into two phrases of equal length.

The melodies were constructed so that each phrase boundary was structurally marked: Structural boundaries within and between measures were marked rhythmically by the distribution of accents. The first tone carried the main accent, the third a minor accent, and the second and fourth were without accent. In addition, the minor boundaries were marked in most cases by changes in melodic contour. Between measures, boundaries were often marked by changes in implied harmony. Boundaries between measures of a subphrase and between subphrases were always marked either by a change in melodic contour, sometimes being a transformation (for instance a mirror image), and/or by changes in implied harmony. The main phrase boundary in the middle of the melodies was always marked by a semicadence. In addition, it was sometimes also marked melodically by the identity of the third and fourth tone in the fourth measure and in most cases by a change in implied harmony between the fourth and fifth measures.

The main concern of the experiment was to determine the amount and direction of click migration as a function of the order of the phrase structure boundary at the position of its actual occurrence. Because one would expect the apparent position of a click to vary in a similar manner to that of a click during speech, the subjective location of clicks superimposed on melodies should shift in such a way as to minimize the number of processing units that the click could interrupt. The larger the number of overlapping processing units in which a click is embedded, the stronger should be the tendency for a listener to displace it to the common boundary of those units. For example, the tendency of a click at the position of the main phrase boundary (position 16 in Figure 1) to be displaced from its actual position of occurrence should be minimal, because a click at that position does not interrupt any higher order representational unit. But a click at position 14 should be displaced to the major phrase structure boundary at position 16, because it would otherwise interrupt the representational units of the last measure and subphrase. At the apparent position between phrases, the number of representational units interrupted is minimal, whereas at an apparent position having the same distance from position 16 as from position 14, but in the opposite direction (which is position 12 at a third-order phrase boundary), the click would interrupt the representational unit of the last subphrase. So the migration of a click at position 14 should be in the direction of the major phrase structure boundary.

The experiment was expected to display a distribution of click localizations in which the probability of locating a click at a certain position was directly proportional to the order of the phrase structure boundary at that position. An analysis of the amount and direction of click migration would further show a more complex pattern of results, because the direction of click migration is a function of two variables: First, the relative distance of the objective click position from the main phrase structure boundary and the nearest fourth-order boundary[3] (at positions 8 and 24; see Figure 1) and second, the number of interrupted phrase structure units at the position of localization. For a click positioned exactly halfway between the main and a fourth-order boundary, the principle of minimization of interrupted units should produce a migration tendency toward the main boundary (Fodor & Bever, 1965). For a click position slightly more in the direction of the fourth-order boundary, there should be some rivalry between the principle of minimization of interrupted units (the minimum being at the position of the main boundary) and the principle of minimization of the migration distance (the minimum being at the position of the fourth-order boundary). So, eventually even for clicks that were slightly nearer to a fourth-order phrase boundary than to the main phrase boundary, migration could be in the direction of the main boundary. At those points where the two minimization principles are of equal effect, there should be maxima of click dis-

3. We are neglecting possible effects of phrase structure boundaries of even lower order. Discussion of their effects, which are of the same kind but of smaller amount as the effects of higher order boundaries, would unnecessarily complicate the picture.

placement, one near position 12, and the other near position 20 (see Figure 1), these being equally distant from the main and the nearest fourth-order phrase structure boundary.

Method

Subjects. Forty-four subjects from the Ruhr-University at Bochum (FRG) participated in the experiment. They were selected on the basis of high performance in a test (see below) following their participation in four instruction sessions on the syntactic structure of German folk songs.[4] The subjects had had at least 3 years of instruction in playing an instrument. All subjects reported normal hearing and none reported absolute pitch. Those subjects who were not psychology majors were paid for their participation (32 subjects); the 12 psychology majors participated in return for course credit.

Apparatus. Melodies were generated on-line by a Dietz 621/8 computer programmed to control a Rohde & Schwarz function generator. The melodies were played through Sennheiser HD 240 earphones. Click production was accomplished by interruption of one channel of the earphones using a Reed relay switched on and off by the computer.

Stimulus Material. Each melody consisted of 32 sine wave tones of 225-msec duration with an interstimulus interval (ISI) of 225 msec. Frequencies were selected from the range between 440 and 1760 Hz. The tones were heard at approximately 78 dB SPL for the first tone of the measure, 74 dB for the third tone, and 70 dB for the second and fourth tone. The peak amplitude of the click was 83 dB, but its subjective loudness was equal to that of the third tone of the measure.[5] The duration of the click was approximately 7 msec. The click always appeared exactly in the middle of the 225-msec ISI.

For each of the 17 click positions, four melodies were presented, resulting in 68 different melodies. These were generated according to the rules of syntax (Stoffer, 1979). Only those obligatory phrase structure rules resulting in a binary phrase structure tree (see Figure 1) were used. All tones were of equal duration, and no pauses intervened, except for the last two beats of the last measure.

Procedure. Before participating in the experiment, subjects had four sessions of instruction on the syntax of the melodies used. Each session lasted approximately 3 hr. and was run as a self-instruction lesson supplemented with listening to musical examples and tests, in which subjects were asked to recognize structural attributes of a melody. Subjects were taught to read a score, to discriminate between different phrase structures, and to recognize musical phrase structure markers (i.e., cadences and melodic transformations). Finally, they received a multiple-choice test with 58 items. Only those subjects with scores greater than 85% correct participated in the experiment.

In the experiment, subjects were instructed that one of their tasks would be to recognize melodies and to judge each one as "old" or "new" depending on whether or not they had heard it before. The melodies to be judged as "old" were 17 melodies played three times in succession at the beginning of the experimental session. Subjects were told that there was another task that asked for localization of the position of a click. In order to mark the apparent click position, subjects had a booklet in front of them containing, on separate pages, transcripts of the melodies with marks below click positions 1 to 31 that could be checked at

4. See Procedure section. A more detailed description of the training and test procedures is given in Stoffer (1981). The melodies used in the experiment were generated by a subset of those rules the subjects learned during the training.

5. The loudness of the click was matched to that of the third tone of a measure by paired comparisons performed by three subjects.

the apparent click position. Subjects were asked not to turn the pages to the transcript of the melody they were listening to until they detected the click. The compliance with this request was closely observed by the experimenter who could monitor the clicks and melodies on his earphones.

Each trial began with a warning tone of 200-msec duration, followed by a pause of 1 sec, after which the melody began. When detecting the click, subjects turned the page of their booklet and first marked the click position before giving the recognition judgment. At the end of each melody there was a pause of 20 sec before the next trial began.

There were six additional trials at the beginning of the experiment, which were regarded as training trials and were excluded from the data analysis. Each click position was combined twice with an "old" and twice with a "new" melody. Melodies were always presented to the left ear and the clicks to the right ear. Using this arrangement, Gregory (1978) observed the largest amount of displacement. The experimental session lasted approximately 75 min, including a short break of 5 min following presentation of the melodies to be recognized. Subjects were tested either in pairs or alone. When tested in pairs, they could not see each other.

Results

The frequencies of click localizations at each position averaged over subjects are shown in Figure 2. Not included are those 0.8% falling outside of the range of the objective click positions (positions less than 8 and higher than 24).[6] A repeated measures analysis of variance revealed significant differences between click localization frequencies[7] [$F(43,688) = 102.7$, $p < .001$].

The histogram in Figure 2 shows a symmetrical distribution of the click localizations. The highest frequency was at position 16 (position of the main phrase structure boundary), the second highest frequency was at position 24 (fourth-order boundary), and this was not significantly different from that at position 8 (also a fourth-order boundary). Differences between the frequencies at positions 16 and 8, and the frequencies at positions 16 and 24 (comparisons between fifth- and fourth-order boundaries) were highly significant ($p < .01$). The third highest frequencies were to be found at positions 12 and 20 (each being a position of a third-order phrase boundary). These did not differ significantly from each other, but pairwise differences with those at the next higher phrase boundaries at positions 8 and 24 were highly significant (Scheffé-test, $p < .01$). Paired comparisons between frequencies at the third-order phrase boundaries (positions 12 and 20) and

6. Because there were no differences in the distributions of the click localization data for "old" and "new" melodies, both sets of data were analyzed neglecting this variable.

7. Localization frequencies for a subject at each position are not independent, so one cannot test deviations from a rectangular distribution by application of a chi-square test. Because the frequency distribution of the localization frequencies did not deviate much from a normal distribution, tests on differences between mean localization frequencies were made with a repeated measures analysis of variance. As a control, a Friedman test was applied, also showing the same level of significance as the analysis of variance ($p < .001$).

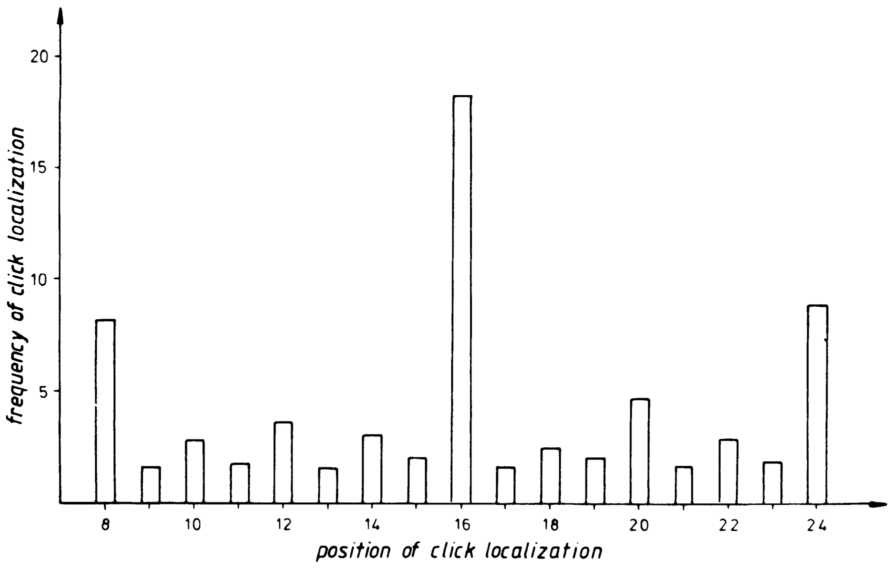

Fig. 2. Mean frequencies of click localization at each click position.

frequencies at all the lower-order boundaries were not significant ($.10 < p < .05$). Higher-order Scheffé-contrasts revealed significant differences between all the mean frequencies at third-order phrase boundaries (positions 12 and 20) and all those at first-order boundaries (uneven numbered positions) ($p < .01$). The only nonsignificant difference between frequencies at phrase structure boundaries of different orders was that between first- and second-order boundaries, all of these being phrase boundaries within measures.

The amount and direction of click displacement was analyzed on the basis of differences between the objective click positions and the click positions marked by the subjects. These differences were summed over the four click localizations for each position and divided by four. The resulting mean displacement score was scaled in units of click positions being equivalent to the duration of a single tone and the pause following it (~ 450 msec). A positive score, that is, a positive direction of displacement, indicates locating the click subjectively later than its actual time of occurrence; a negative score, that is, a negative direction of displacement, indicates locating it subjectively earlier.

The graph of the mean click displacements as a function of the objective click position (Figure 3) displays a tendency of the clicks to be displaced in

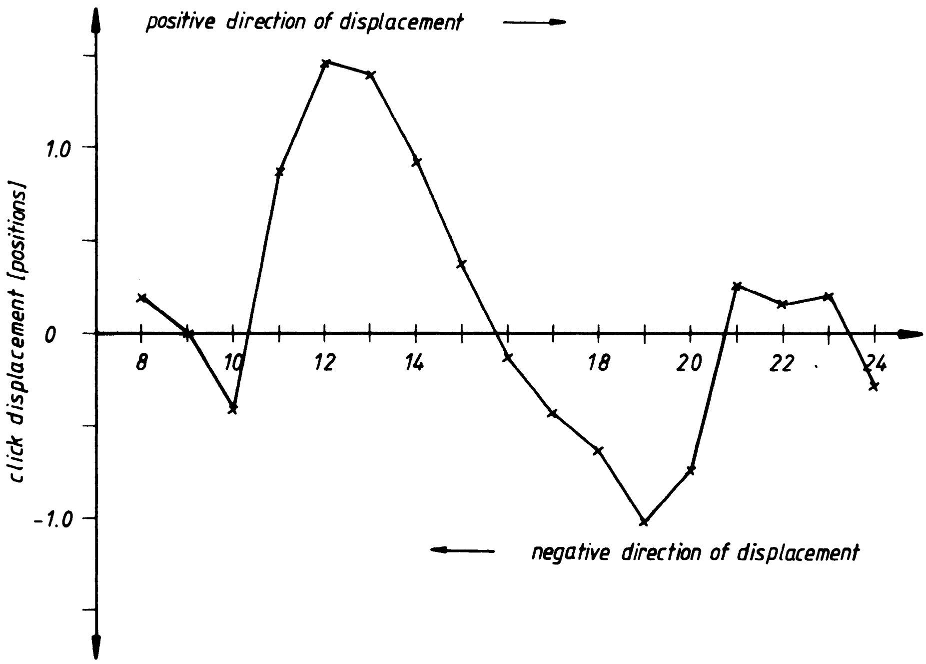

Fig. 3. Mean click displacement scores (scaled in units of click positions being equivalent to a 450-msec time segment) as a function of the objective click position.

the direction of the main phrase structure boundary at position 16.[8] The overall effect is highly significant [$F(43,688) = 21.04, p < .001$).[9] Multiple comparisons with Scheffé-contrasts revealed significant differences that establish the location of the relative maxima at positions 12 and 19: The differences between the means at positions 12 and 19 ($p < .001$) are significant, as are the higher order contrasts between the mean of the values at positions 8 to 10 contrasted with the mean at position 12 ($p < .01$) and between the mean at position 19 contrasted with the mean of the values at positions 21 to 24 ($p < .05$). This means that the amount of click displacement in the direction of the main phrase structure boundary at position 16 was larger than that in the direction of the phrase structure boundaries of the next lower order at positions 8 and 24. The difference between all the

8. Again, there were no differences in the click displacement scores for "old" and "new" melodies, so both sets of data were analyzed neglecting this variable. In contrast to the analysis of the click localization data, clicks displaced outside the range of objective click positions were not excluded from data analysis.

9. The analysis of the data by a Friedman test showed a slightly lower level of significance than the analysis of variance ($p < .005$).

positive and all the negative displacement scores was .12, which is not significantly different from a zero displacement (Scheffé-test, $p>.10$).

Discussion

The main question to be answered by the results is how much correspondence was found between the formal description of phrase structure and the representational structure of the perceived music. At least for the simple phrase structure used, all the higher order boundaries coincided with corresponding boundaries of representational units. The only exceptions were the boundaries within a measure. There was a trend for decreasing frequency differences between phrase boundaries of adjacent order, indicating dominance of the principle of minimization of interrupted units over the principle of minimization of the migration distance. The insignificance of the effect of low-order boundaries may be attributed to some limits in the temporal resolution of the processes involved in locating clicks, but not to the existence of a fixed "minimal representational unit" corresponding to a measure.

This interpretation is corroborated by experimental results with short tonal sequences, which are easily interpreted as demonstrating that access to smaller representational units is harder, the lower the level of the units in the representational hierarchy (Deutsch, 1980; Dowling, 1973; Stoffer, 1981). Locating a click in a melody can be described by a process that progressively narrows down the representational unit in which the click occurred, while passing top-down through the representational hierarchy (Stoffer, 1981). The process starts by focusing on the most recent representational unit at the highest level in the hierarchy that can be represented at the time of click detection. It is assumed that on each level the only decision possible concerning the location of a click is between the alternatives that the click is either within a unit or between the boundaries of two adjacent units. If it is detected between the boundaries of two adjacent units, the listener can look for the corresponding phrase structure boundary in the musical score. But if the click is embedded within a representational unit, a process of parsing the unit into its functional subunits at the next-lower level is initiated. This is done by focusing attention on the internal structure of that unit. Then a decision on the location of the click is repeated in the same way as at the higher level. This cyclic process continues as long as the click cannot be located between the boundaries of two representational units at the same hierarchical level.

The time needed for this process is longer, the deeper the click is embedded in the representational hierarchy. The lower the level of representation, the lower the stability of the representation in memory (Attneave & Olson, 1971; Deutsch, 1969; Deutsch & Feroe, 1981), because more time is required to access lower representational levels than higher ones. The memorial instability of low-level structural information should reduce the effectiveness of localizations relative to low-level boundaries of representational units. So the localization process has to take as reference points those higher order boundaries that remain relatively stable within the time needed for processing.

Because of the greater memory stability of higher order phrase structure boundaries relative to those of lower order, the described processing should produce frequencies of correct localizations that are higher, the higher the order of the phrase structure boundary at the objective click position. In addition, at any position, the frequencies of displaced clicks should be higher, the higher the phrase structure boundary at that position, because those boundaries that are of the next-higher order are potential reference points in the representational hierarchy for clicks at lower order boundaries. The joint distributions of correct click localizations and displaced localizations should therefore show exactly the pattern of results that is presented in Figure 2.

One of the maxima of the click displacement scores is, as expected, at position 12 (see Figure 3). The other one is at position 19, just one position nearer to the fourth-order boundary on the right than to the main boundary on the left. This insignificant deviation from the expected position may be an indication of a small amount of asymmetry in the click displacements, favoring a positive direction of displacement. This was observed by Gregory (1978) as well as Sloboda and Gregory (1980).

The click localization method is sometimes suspected of being susceptible to the influence of response biases that may question the validity of an interpretation in terms of structural properties of cognitive representations. A response bias was demonstrated by Ladefoged (1967) when catch-trials without a click were introduced. Even nonexisting clicks were located by many subjects in the vicinity of the major phrase structure boundaries. However, Bever (1973) and Bertelson and Tisseyre (1970) showed that, in addition to a small effect that can be attributed to response bias, there is a stronger effect of click displacement that is not susceptible to response bias. Dalrymple-Alford (1976) argued that if there really were a marked response bias, click displacements onto higher phrase structure boundaries should be more frequent the more uncertain subjects are concerning the localizations of the clicks. The data show exactly the opposite results: The more certain a subject is concerning the localization of a click, the greater is his/her tendency to displace a click to a major phrase structure boundary.

There are no comparable experimental results using musical stimuli. But there is no reason to believe that there should be a more marked effect of response bias in our data then in studies using speech stimuli. The experiments of Bever (1973) and Bertelson and Tisseyre (1970) show a much stronger effect of response bias in those conditions where attention is not properly focused on the speech, but can be focused on the opposite ear where the click is expected. As far as the control of attentional focusing on the melodies is concerned, there may even be some reason to believe that response biases may have had a less marked effect on our data than those of Gregory (1978) and Sloboda and Gregory (1980). Overall, the results are in accordance with the conclusions of Gregory (1978) and Sloboda and Gregory (1980). They show a detailed picture of the representational structure that can unambiguously be accounted for by the listener's segmentational processes in the course of the auditory perception of attributes that mark the phrase structure of the music. The high degree of correspondence between the musicological description of phrase structure and the cognitive representational structure as indicated here establishes some degree of cognitive adequacy for music–syntactic models that give a description of musical phrase structures on the basis of joint distributions of several musical attributes.

Experiment 2: Detection of Clicks Superimposed on Melodies Varying in Phrase Structure by Listeners of Different Musical Competence

The purpose of Experiment 2 was to test the influence of musical knowledge involved in the construction of a cognitive representation of phrase structure in music. The point of departure was the proposal of Riemann (1889) that in music, the binary partitioning of any structural unit into subunits on the immediately subordinate level can be regarded as prototypical, and that all the other phrase structures can be regarded as systematic derivations from this prototypical phrase structure.

Our general hypothesis was that if Riemann's proposal may also be a cognitively adequate description of the representation of musical knowledge, the binary phrase structure should be represented as the central concept of that part of musical knowledge that is involved in the construction of an internal representation of phrase structure. One likely prediction from this hypothesis is that subjects who are not familiar with more complex phrase structures than the prototypical one would show a tendency to parse complex phrase structures according to a binary subdivision heuristic that may be regarded as the functional equivalent of the prototypical knowledge.

Experimental Design

In Experiment 2, a comparison was attempted between subjects who were highly competent musically, and trained to recognize complex phrase structures, and subjects who were low in musical competence. It was expected that the following differences between trained and untrained listeners would be seen when comparing their segmentational behavior at different times in the course of an experiment embodying different phrase structure conditions: (1) Regardless of the actual phrase structure, subjects low in musical competence should exhibit a segmentational behavior at the beginning of the experiment that is to be expected on the assumption that these subjects have at their disposal only the knowledge of the prototypical phrase structure. On the other hand, subjects high in musical competence should exhibit a segmentational behavior that can be accounted for only by assuming that they are able to construct a representation of phrase structure that matches the differences between the actual phrase structures used. (2) In the course of the experiment, subjects low in competence should form a refined concept of phrase structure incorporating those structures that deviate from the prototypical binary one. So, at the end of the experiment, there should be no (or at least a reduced) difference between the two groups of subjects.

In order to test these hypotheses, we performed an experiment with a click-detection task (Abrams & Bever, 1969; Holmes & Forster, 1970). The task requires subjects to press a key each time they detect a click superimposed on a melody. This method gives both groups of subjects equal opportunity to master the task with comparable efficiency, whereas the localization procedure used in Experiment 1 would require a training of the low-competence group in sight-reading of musical scores that probably would reduce any difference expected between groups.

Melodies were cases of the two phrase structures shown in Figure 4. Both structures incorporated the same number of measures, so that there was no confounding of number of measures and type of phrase structure. Figure 4a shows a structure that consists of three phrases of four measures each: an opening phrase, a middle phrase, and a closing phrase (OMC structure). The OMC structure was identical to the prototypical binary phrase structure within the first eight measures. Figure 4b shows a phrase structure that is derived by extension of the opening phrase as well as the closing phrase by a sequence of the first pair of measures of a phrase (extended OC structure). Both structures are characteristic of phrase structures in German children's and folk songs. The extended OC structure is extremely rare in children's songs and more frequent in folk songs, the OMC structure is the most frequent of both, being of about equal frequency in children's and in folk songs.

As one can see from Figure 4, there are marked differences in the order of some phrase structure boundaries at the same position in the two versions of phrase structure. At position 2 (see Figure 4) there is the main boundary of the OMC structure, but in the extended OC structure the same position

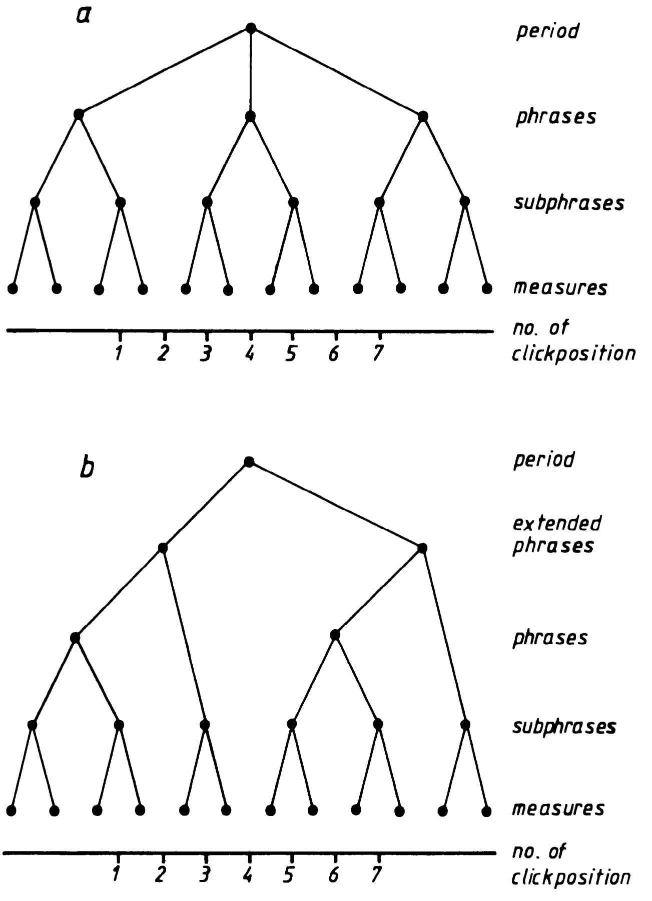

Fig. 4. Phrase structure of melodies used in Experiment 2. Figure 4a shows the regular binary structure below the level of phrases (OMC structure); Figure 4b shows the variant with extended opening and closing phrase (extended OC structure).

marks only the second-highest boundary. It is just the other way round at position 4, which is the position of the main boundary of the extended OC structure, but a boundary of one order less in the OMC structure. At position 6, there is the second main boundary of the OMC structure, but a boundary of two orders less in the extended OC structure. At these click positions, one can expect some differences in the reaction times to clicks between phrase structures if listeners actually represent the two types of phrase structure according to the description given in Figure 4. By analogy with the results of Abrams and Bever (1969) and Holmes and Forster (1970), we predicted, as a general trend, that reaction times would be lower, the higher the phrase structure boundary at a given click position. At least for the musically trained subjects, we expected differences in reaction times between the two phrase structure versions to be directly proportional to the difference in the order of the phrase structure boundaries at a certain position. But for the untrained subjects we did not expect these differences at the beginning of the experiment, because they would be expected to try to segment a melody according to the binary division heuristic regardless of the actual phrase structure. We predicted that at the end of the experiment, differences between groups would be reduced, due to learning by the subjects low in musical competence how to parse adequately a phrase structure that is not of the binary type.

Method

Subjects. Twenty-six subjects per group from the Ruhr-University at Bochum and the University of Osnabrück participated in the experiment. The subjects in the group high in musical competence were selected on the basis of having had at least 3 years of instruction in playing a musical instrument. They participated in a training session of about 1-hr. duration, in which they were trained to discriminate between the two versions of phrase structure. Afterwards their discrimination ability was tested with a presentation of 40 melodies in a multiple-choice test. Only those subjects that produced less than 10% errors participated in the experiment. The subjects of the group low in music competence did not play any instrument and they did not participate in the training procedure.

Apparatus. Melodies were generated on-line either by a Dietz 621/8 computer programmed to control a Rohde & Schwartz function generator (in the laboratory at the Ruhr-University) or by an Eltec System 7000 microcomputer programmed to control a Wavetek Model 164 function generator (in the laboratory at the University of Osnabrück). The melodies were played through Sennheiser HD 420 earphones. Click production wa accomplished by the same method as in Experiment 1.

Stimulus Material. Melodies consisted of 48 sine wave tones of 150-msec duration with an ISI of 150 msec. Frequencies were selected from the range between 440 and 1720 Hz. The tones were heard at approximately 72 dB SPL without any rhythmic accent. The peak amplitude of the click was 81 dB and the click was equal in subjective loudness to the tones. The duration of the click was the same as in Experiment 1.

For each of the seven click positions between measures shown in Figure 4, and for each of the phrase structure conditions, 16 melodies were presented with a click superimposed. This resulted in 224 different melodies. The same melodies were repeated once without a click superimposed (see above). Melodies were generated obeying strictly the rules of syntax (Stoffer, 1979). All tones were of equal duration, and no pauses intervened, with the exception of the last two beats of the last measure.

Procedure. As in Experiment 1, subjects performed a recognition task to focus their attention on the structure of the melody. Because of this task, all the 224 melodies were repeated once during the experiment. Besides that, we introduced 50% catch trials with no clicks superimposed on melodies, in order to reduce the serial position effect in the reaction times observed by Holmes and Forster (1970) and Stoffer (1981). This effect consists of a reduction in reaction time (RT) from the first to the last position, independent of the order of the phrase structure boundary at any given position. It is probably caused by anticipation of the click, which is more effective the later the actual position of the click. In 50% of the cases a "new" (not yet repeated) melody had a click superimposed, and in 50% an old melody also had a click superimposed, so that detection of a click would not be a cue as to whether a melody was "old" or "new."

The experiment was run in four sessions. In each session, there were eight trials for each of the combinations of the two phrase structures and the seven click positions. Four of these eight trials had a click superimposed and four had no click superimposed. This totalled 112 trials per session. Within a session, melodies were presented randomly for each subject. Each session lasted about 60 min. Subjects participated once a week and were always tested individually.

Subjects were asked to perform two tasks: (1) Detection of a click superimposed on a melody and (2) recognition of the melody as new or old. They were instructed to perform the tasks in exactly that order. When detecting the click, they were told to press the key as quickly as possible, but to avoid errors. Their second task was to check on a list if the melody just heard was an old or a new one. It was stressed that they should avoid errors in this task, also.

Each trial began with a 500-msec warning tone of 2500 Hz followed by a pause of 1 sec, after which the melody began. At the end of each melody, there was a pause of 10 sec before the next trial started.

There were 10 additional trials at the beginning of each session. These were regarded as training trials and were excluded from the analysis. The melodies used for these trials were new ones, half of them having a click superimposed on a randomly selected position. As in Experiment 1, melodies were presented to the left ear and clicks to the right ear.

Results

There were only 2.3% errors in click detections, the majority of these being false alarms (83.8%). The proportion of errors in the recognition task was 8.4%, most of them being false old decisions (67.2%). There were no significant differences between the two groups of subjects in either type of error.

At first we analyzed the RT data by a 2 (groups) × 2 (phrase structures) × 7 (click positions) analysis of variance, collapsing data over the four experimental sessions in order to look for general effects that could not be accounted for by learning between sessions. This analysis revealed a general difference between groups, showing that the trained subjects were on the average 62 msec faster in detecting clicks than the untrained subjects [$F(1,50) = 17.8$, $p < .01$]. The strong effect of click position is uninteresting, because it only shows that there were differences independent of structural variations between phrase structures.

A strong phrase-structure-by-click-position interaction [$F(6,300) = 19.6$, $p < .01$] indicated differences between click positions depending on the phrase structure, but independent of the difference in musical competence between groups. Figure 5 shows the means of the RTs for the two phrase structures at each click position as a function of the rank of a phrase structure boundary within the structure hierarchy. The rank of a phrase structure boundary is inversely related to its order; that is, the highest order boundary in a phrase structure is assigned a rank of 1, the next lower order boundary is assigned a rank of 2, etc. For the OMC structure, Scheffé tests revealed a significantly faster RT at the position of the first main boundary

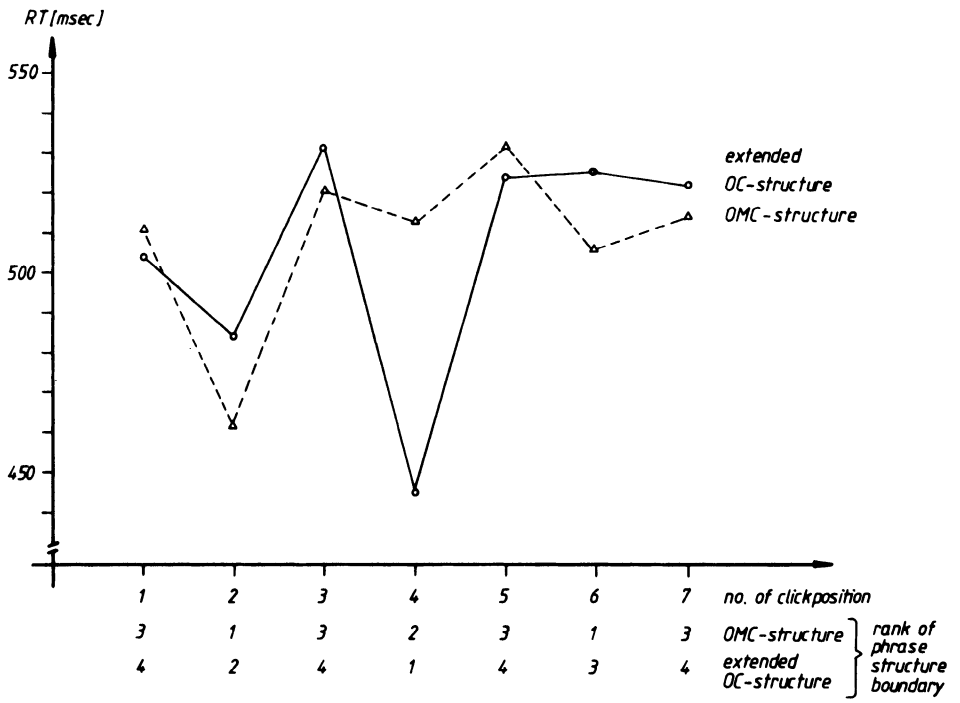

Fig. 5. Mean click detection RTs for the extended OC structure and the OMC structure as a function of click position or of the rank of phrase structure boundary at a certain position.

(position 2, rank 1) than at any other position ($p < .01$), but no difference between the mean at the position of the second main boundary (position 6, rank 1) and the means at the other positions with lower order boundaries ($p > .10$). There were no other significant differences between means for the OMC structure.

For the extended OC structure, Scheffé tests revealed a significantly shorter RT at the position of the main phrase structure boundary (position 4, rank 1, see Figure 5) than at the next lower order boundary (position 2, rank 2) ($p < .01$), and a significantly shorter RT at the latter position than at the position of the boundary with rank 3 (position 6) ($p < .01$). But there was no significant difference between all the positions with phrase structure boundaries of rank 4 (all the uneven numbered positions) and the one with a boundary of rank 3 ($p > .10$). Most of the interaction can be accounted for by the large difference in slope between the graphs of the two phrase structures at positions 3 to 5 (see Figure 5), where there was a rank order difference of 1 for the OMC structure, but a difference of 3 for the extended OC structure. This difference between means is in accordance with these rank order differences. Overall, these results show that there is some correlation between the orders of phrase structure boundaries and the click detection

RTs, at least at the positions at the beginning and the middle of the range studied (for the extended OC structure positions 1 to 5, for the OMC structure positions 1 to 3).

More interesting, as far as our predictions are concerned, is the significant three-factor interaction between the factors "group," "phrase structure," and "click position" [$F(6,300) = 2.17, p < .05$]. Figure 6 shows the graphs of mean RTs as a function of click position for the two groups and the two phrase structures. A comparison of the slopes of the graphs between groups at positions 2 to 5 reveals that for the untrained subjects the difference between the RT at the main boundary of the extended OC structure (position 4) and the next lower order boundary at position 2 is less than the corresponding difference for the trained subjects. In fact, the difference is just about 20 msec for the untrained subjects, but about 50 msec for the trained subjects. This kind of difference between groups must be the main reason for the significance of the four-factor interaction that includes a factor for the first and last session in the analysis of variance [$F(6,300) = 2.81, p < .05$].[10] In the first session, the untrained subjects showed a slightly shorter RT at the position of the boundary of the second-highest order in the extended OC structure (which also is the position of the main boundary of the OMC structure) than at the position of the main boundary (525 and 538 msec, respectively), whereas the trained subjects show a difference just the other way round (454 and 426 msec, respectively). This pattern of differences between mean RTs is shown in Figure 7. In the fourth session there was very little difference between the two groups. For the OMC structure there were no differences between experimental sessions.[11]

10. We excluded the data of the second and third experimental session from the analysis of variance to simplify presentation of results. This procedure is justified by the impenetrability of the four-factor interaction when presented for all the sessions. When excluding the data from the first session from consideration, the four-factor interaction with sessions two to four does not reach significance anymore, demonstrating that the interaction is due to an effect manifesting itself early in the experiment. We choose to compare sessions one and four because in that case the interaction was the strongest, but in all cases of paired comparisons of session one with any other session, the four-factor interaction was significant at the 5% level.

11. The four-factor analysis of variance also revealed further significant effects that are not of much interest in the context of an evaluation of our hypotheses. There was a main effect of experimental sessions [$F(1,50) = 11.8, p < .01$] showing slightly faster reactions to clicks in the last session (482 msec) rather than the first session (534 msec); a main effect of click position [$F(6,300) = 21.6, p < .01$], and an interaction of phrase structure by click position [$F(6,300) = 13.7, p < .01$], not deviating much from the equivalent effect already described in the context of the three-factor analysis of variance.

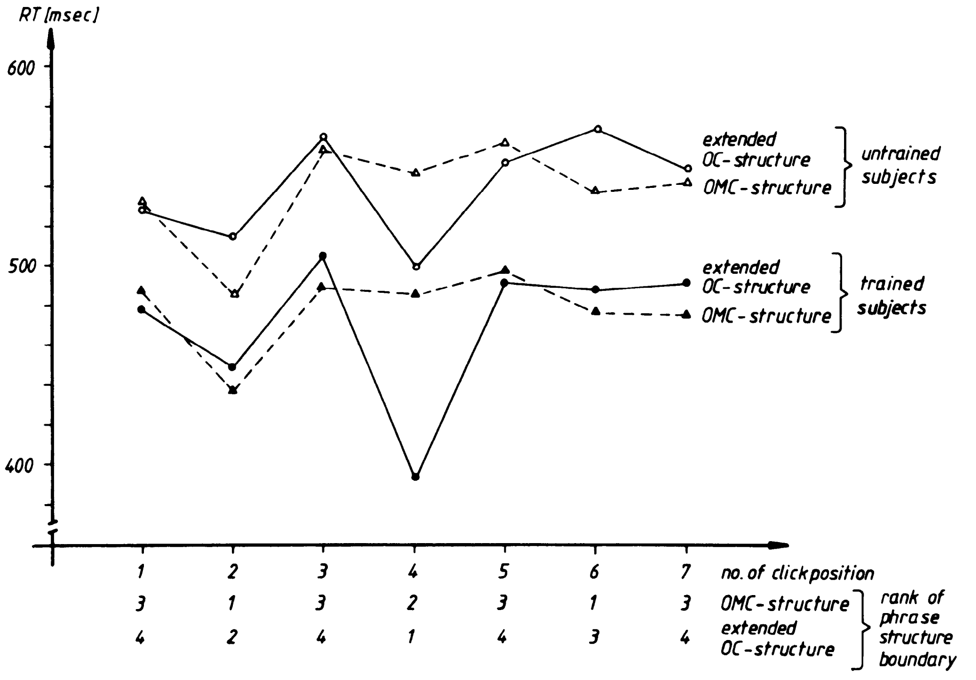

Fig. 6. Mean click detection RTs as a function of click position for the two phrase structure conditions and the two groups of subjects.

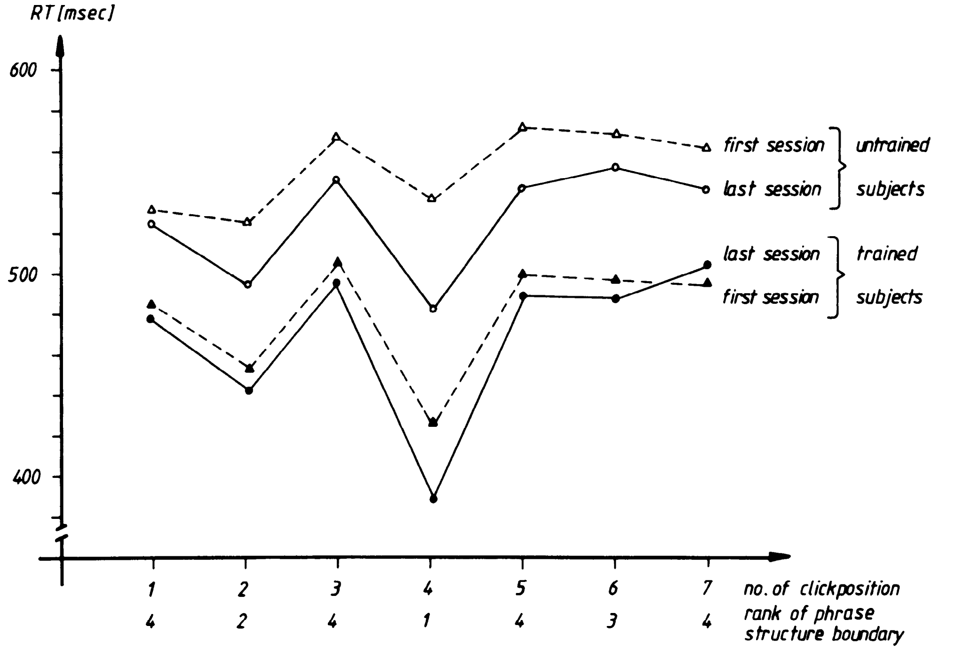

Fig. 7. Mean click detection RTs as a function of click position for the extended OC structure in the first and the last experimental session for both groups of subjects.

Discussion

There are three aspects of the data presented that need to be evaluated with reference to the hypotheses suggested. (1) The change in the click detection times of the low-competence group from the first to the last experimental session, indicating a change in content and structure of the listeners musical knowledge. (2) The degree of correspondence between the structural description of the phrase structures used and the structure of the representation as indicated by the click detection times. (3) The obvious lack of substantial variation in click detection time as a function of click position or level of phrase structure boundary in the second half of the melodies.

The most important result of the click detection experiment is the change in the segmentational behavior of the subjects that were low in musical competence, demonstrated for the extended OC structure. We proposed that these listeners would not have adequate knowledge as to how to segment a melody whose phrase structure does not conform to the heuristic principle of binary subdivision. Consequently, these listeners would at first try to segment a melody in accordance with the principle of binary subdivision. For a melody with an extended OC structure this suggestion means that they would represent the main phrase structure boundary as between measures four and five despite the fact that the actual main phrase structure boundary is between measures six and seven.

A comparison of the click detection times at these two positions (2 and 4 in Figure 7) shows that in the first experimental session the low-competence group detected clicks between measures four and five slightly faster than clicks between measures six and seven, whereas for the high-competence group it was just the other way round. For the low-competence group this difference was only small, probably because they formed a conceptual representation of the actual phrase structure quite early in the experiment. There might have been some segmentations corresponding to the actual phrase structure already within the first session, thereby decreasing the difference between mean reaction times at these positions. This interpretation is supported by the fact that this difference was almost as large when comparing the first and second session as when comparing the first and the last (cf. footnote 10). Contrary to the low-competence group, the high-competence group segmented the extended OC structure right from the start according to the phrase structure description given by the syntax.

This pattern of results can be accounted for by suggesting that the development of knowledge particularly relevant for the construction of representations of musical phrase structure starts with a core concept representing the most general heuristic of cognitive segmentation: namely, the principle of binary subdivision of units. This representation is gradually differentiated by forming concepts subsidiary to the core concept, enabling the listener to construct representations of phrase structures being regular derivations of the prototypical binary structure. These concepts are embedded within the structure of musical knowledge forming a hierarchy of cognitive schemata (for a description of the schema concept see Rumelhart, 1980). The core concept is contained in one of several subschemata of tacit knowledge of musical form, and the concepts concerning regular derivations from binary phrase structures are represented as subschemata of the core schema.

In Experiment 2, the degree of correspondence between the structural description of the phrase structures used and their representations as indicated by click detection times is more difficult to determine than in Experiment 1, because here in the second half of the melodies, click detection times did not vary in accordance with the variation of the order of the phrase structure boundaries at the position of a click. There was a high degree of correspondence at the first three positions for the OMC structure and at the first five positions for the extended OC structure (see Figure 5), but this was much less at later positions, where there were no significant differences between mean click detection times at different positions. The lack of effect of phrase structure at the late positions of the melodies can be accounted for by a decrease in active segmentation by the subjects, and/or by the influence of an additional factor on click detection times that tends to obscure some of the variation depending on the order of the phrase structure boundaries.

There may be at least two reasons for this decrease in segmentational activity. The first one may be the ineffectiveness of the recognition task in ensuring that attention is focused on encoding the structure of a melody. One might argue that recognition can be accomplished in this kind of experimental situation just by attending to the first half of a melody. In fact, the first half of all the melodies differed from each other in at least one structural detail.

A possible objection to this explanation might be that it does not account for the obvious difference between the two versions of phrase structure, showing a better degree of correspondence for the extended OC structure than for the OMC structure. This difference is accounted for by a second possible reason for the decrease in segmentational activity. The first critical position for differentiating between the two phrase structures is at the first main boundary. When reaching it without detecting a click, the subject can decide on the type of phrase structure, and at the same time he or she can recognize the melody. Only after that, attention may be switched to the other ear where the click can now be detected without any delay due to ongoing segmentational processes which require attention.

An additional factor that may have obscured some of the phrase structure-dependent variation of click detection times is the effect of the aging foreperiod. It has been a general observation in reaction time experiments that the longer the foreperiod (the time interval between the warning signal and the stimulus to be reacted to) the shorter the reaction time (e.g., Näätänen, 1970; Telford, 1931). In order to reduce the effect of the foreperiod in our experiment, we introduced 50% catch trials. The effect of catch trials is to compensate for the usually observed decrease in reaction time by raising it slightly for the longer foreperiods (Näätänen, 1971). This effect is clearly seen at positions 5 to 7 (see Figure 5).

However, the introduction of catch trials does not make the foreperiods non-aging. The elimination of the effect of aging foreperiods is usually accomplished by selecting the frequencies of different foreperiods so that they follow a Bernoulli process (Näätänen, 1970). In the context of Experiment 2, this procedure probably would interfere with the segmentational processes at the beginning of the melodies, because relatively short foreperiods (clicks at early positions) would have to be much more frequent than long ones. In that case, it would be a good strategy for the subjects to detect the click first and to attend to the melody afterwards. Together with the effect of attention distribution, which probably reduced segmentational effects at late positions, making foreperiods non-aging would have obscured the effect of segmentation on click detection time all over the range of positions tested.

Similar difficulties with the reduced sensitivity of the click detection task in comparison to the click location task were observed by Abrams and Bever (1969) with linguistic material. They presented sentences three times to their subjects, but not until the second presentation were detection times for clicks at the main phrase structure boundary shorter than at any other position. Even then the difference did not reach statistical significance. On the other hand, Holmes and Forster (1970) demonstrated a small but significant effect of phrase structure on click detection time with linguistic material, even in the presence of a typical effect of the aging foreperiod.

In addition to the above methodological considerations, it may also be that listeners lose track at later positions when trying to establish a multilevel representation of musical structure (cf. Stoffer, 1981). In that case, the processes engaged in detecting the click may not be influenced by ongoing segmentational activity. In this interpretation, we have to admit that the resulting representational hierarchy is incomplete. But this is not critical for our interpretation of segmentation utilizing a cognitive structure that represents a processing heuristic. Using a binary segmentation heuristic does not guarantee a complete representation of all the structural levels possible at any point in time. The heuristic merely sets structural boundary conditions for the formation of a representation in a top-down fashion. How this abstract representational schema is actually filled may be critically dependent on lower-order schemata and bottom-up processes that may be of different effectiveness in different subjects.

In the light of the methodological problems encountered, the established effects of phrase structure on click detection times at the early positions of a melody and the differences demonstrated between the two types of phrase structure permit the conclusion that the syntactic description of phrase structure, on which the hypotheses were based, may be regarded as showing at least some degree of cognitive adequacy.

Krumhansl and Castellano (1983) recently proposed a schema-theoretic account of the recognition of interval relations, especially for the establishment of harmonic function of a tone within the predominant key. They assumed that tonal elements are evaluated primarily in terms of their relationships to temporally adjacent elements. Although they admit that higher order structural attributes may modify the processing of temporally separated elements, their model does not make any provisions for the integration of higher order syntactic relationships. Their model may be regarded as the description of the utilization of a frame of reference for the evaluation of tonal functions in relation to the tonic. It may be part of a low-level schema that interacts with a higher level schema representing knowledge of phrase structure in music. The phrase structure schema may be regarded as receiving information on tonal functions, not only for single elements, but also by the way of intervening subschemata that process higher order moments of variations in tonal functions. This may allow one to recognize degrees of changes that mark phrase structure boundaries of different order: for example, a harmonic progression from the tonic to the dominant or a modulation from one key to another. This example may demonstrate that the theoretical framework given by schema theory of knowledge-dependent processing (e.g., Neisser, 1976; Rumelhart, 1980) may be of some heuristic value for a theoretical integration of models describing different aspects of the processing of music.

Conclusion

Top-down processing while listening attentively to music may be characterized by an interaction between several cognitive schemata that form a hierarchy of musical concepts according to abstractness or generality. A representation of phrase structure in music is formed by analysis of higher order moments in the variation of physical attributes that mark phrase structure boundaries on a certain level of structure. Musical knowledge activated in the course of this analysis is organized as a high-level core schema representing a binary subdivision heuristic. In addition, lower level subschemata representing variants of this heuristic can be applied in order to construct representations of phrase structures that are not of the binary subdivision type.

In part, this theoretical account of a listener's cognitive segmentation of musical structure is the result of a strategy that leads off by establishing correspondences between some aspects of a theory of the description of musical structure and some structural aspects of a psychological model of the processing and representation of perceived music. These correspondences are regarded as a useful strategy for modelling those aspects of a theory of music perception that need to incorporate typical musical structures. The results described are not only in accordance with the cognitive model presented but also strengthen the heuristic value of the syntax model to establish these correspondences.[12]

References

Abrams, K., Bever, T. G. Syntactic structure modifies attention during speech perception and recognition. *Quarterly Journal of Experimental Psychology*, 1969, ***21***, 280–290.

Attneave, F., & Olson, R. K. Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology*, 1971, ***84***, 147–164.

Bertelson, P., & Tisseyre, F. Perceiving the sequence of speech and non-speech stimuli. *Quarterly Journal of Experimental Psychology*, 1970, ***22***, 653–662.

Bever, T. G. Serial position and response biases do not account for the effect of syntactic structure on the location of brief noises during sentences. *Journal of Psycholinguistic Research*, 1973, ***2***, 287–288.

Cooper, L. A. Recent themes in visual information processing: A selected overview. In R. S. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, NJ: Erlbaum, 1980.

Cuddy, L. L. Cohen, A. J., & Mewhort, D. J. K. Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception & Performance*, 1981, ***7***, 869–883.

Dalrymple-Alford, E. C. Response bias and judgement of the location of clicks in sentences. *Perception & Psychophysics*, 1976, ***19***, 303–308.

Deutsch, D. Music recognition. *Psychological Review*, 1969, ***76***, 300–308.

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

Deutsch, D. The generation of two isochronous sequences in parallel. *Perception & Psychophysics*, 1983, ***34***, 331–337.

Deutsch, D., & Feroe, J. The internal representation of pitch sequences in tonal music. *Psychological Review*, 1981, ***88***, 503–522.

Dowling, W. J. Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, 1973, ***14***, 37–40.

Dowling, W. J. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, 1978, ***85***, 342–354.

12. Experiment 1 was reported at the "1. Bochumer kognitionspsychologisches Kolloquium", Altena, Federal Republic of Germany, October 31 to November 2, 1980. Experiment 2 was reported at the "16. Treffen der Gruppe Menschliche Informationsverarbeitung", University of Darmstadt, Federal Republic of Germany, October 7 to 9, 1983. The research reported here was supported by a grant from the "Minister für Wissenschaft und Forschung des Landes Nordrhein-Westfalen" (Az. II B7 - FA 8636).

Fodor, J. A., & Bever, T. G. The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, 1965, ***4***, 414–420.
Fraisse, P. Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic Press, 1982.
Gregory, A. H. Perception of clicks in music. *Perception & Psychophysics*, 1978, ***24***, 171–174.
Handel, S. Temporal segmentation of repeating auditory patterns. *Journal of Experimental Psychology*, 1973, ***101***, 46–54.
Handel, S. Perceiving melodic and rhythmic auditory patterns. *Journal of Experimental Psychology*, 1974, ***103***, 922–933.
Handel, S., & Pickens, J. D. The effect of response constraints on the perception of repeating auditory patterns. *Bulletin of the Psychonomic Society*, 1977, ***9***, 253–255.
Handel, S., & Yoder, D. The effects of intensity and interval rhythms on the perception of auditory and visual temporal patterns. *Quarterly Journal of Experimental Psychology*, 1975, ***27***, 111–122.
Holmes, V. M., & Forster, K. I. Detection of extraneous signals during sentence recognition. *Perception & Psychophysics*, 1970, ***7***, 297–301.
Jackendoff, R., & Lerdahl, F. Generative music theory and its relation to psychology. *Journal of Music Theory*, 1981, ***25***, 45–90.
Jones, M. R. Auditory patterns: Studies in the perception of structure. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception, Vol. VIII*. New York: Academic Press, 1978.
Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
Krumhansl, C. L., & Catellano, M. A. Dynamic processes in music perception. *Memory & Cognition*, 1983, ***11***, 325–334.
Ladefoged, P. *Three areas of experimental phonetics*. London: Oxford University Press, 1967.
Ladefoged, P., & Broadbent, D. E. Perception of sequences in auditory events. *Quarterly Journal of Experimental Psychology*, 1960, ***12***, 162–170.
Lerdahl, F., & Jackendoff, R. *A generative theory of tonal music*. Cambridge, MA: MIT Press, 1983.
Longuet-Higgins, H. C. Perception of melodies. *Nature*, 1976, ***263***, 646–653.
Martin, J. G. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 1972, ***79***, 487–509.
Näätänen, R. The diminishing time-uncertainty with lapse of time after the warning signal in reaction-time experiments with varying foreperiods. *Acta Psychologica*, 1970, ***34***, 399–419.
Näätänen, R. Non-aging foreperiods and simple reaction time. *Acta Psychologica*, 1971, ***35***, 316–327.
Narmour, E. *Beyond Schenkerism: The need for alternatives in music analysis*. Chicago: University of Chicago Press, 1977.
Narmour, E. Some major theoretical problems concerning the concept of hierarchy in the analysis of tonal music. *Music Perception*, 1984, ***1***, 129–199.
Nattiez, J. J. *Fondements d'une sémiologie de la musique*. Paris, Seuil, 1975.
Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.
Povel, D. J. The internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception & Performance*, 1981, ***7***, 3–18.
Preusser, D. The effect of structure and rate on the recognition and description of auditory temporal patterns. *Perception & Psychophysics*, 1972, ***11***, 233–240
Preusser, D., Garner, W. R., & Gottwald, R. L. Perceptual organization of two-element temporal patterns as a function of their component one-element patterns. *American Journal of Psychology*, 1970, ***83***, 151–170.
Riemann, H. *Musikalische Formenlehre*. Berlin: Hesse, 1889.
Royer, F. L., & Garner, W. R. Response uncertainty and perceptual difficulty of auditory temporal patterns. *Perception & Psychophysics*, 1966, ***1***, 41–47.

Royer, F. L., & Garner, W. R. Perceptual organization of nine-element auditory temporal patterns. *Perception & Psychophysics,* 1970, *7,* 115–120.

Rumelhart, D. E. Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, NJ: Erlbaum, 1980.

Seitz, M. R. & Weber, B. A. Effects of response requirements on the location of clicks superimposed on sentences. *Memory & Cognition,* 1974, *2,* 43–46.

Sloboda, J. A., & Gregory, A. H. The psychological reality of musical segments. *Canadian Journal of Psychology,* 1980, *34,* 274–280.

Stoffer, T. H. *Aspekte einer generativen Syntax zur Beschreibung musikalischer Strukturen für eine kognitive Musikpsychologie.* Report no. 11/1979, Psychologisches Institut der Ruhr-Universität Bochum (FRG), Arbeitseinheit Kognitionspsychologie.

Stoffer, T. H. *Wahrnehmung und Repräsentation musikalischer Strukturen: Funktionale und strukturelle Aspekte eines kognitiven Modells des Musikhörens.* Doctoral dissertation, Ruhr-Universatät Bochum (FRG), 1981.

Stoffer, T. H. Modelle der kognitiven Verarbeitung und Repräsentation musikalischer Strukturen. In O. Neumann (Ed.), *Perspektiven der Kognitionspsychologie.* Berlin: Springer, 1985.

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

Tan, N., Aiello, R., & Bever, T. G. Harmonic structure as a determinant of melodic organization. *Memory & Cognition,* 1981, *9,* 533–539.

Telford, C. W. The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology,* 1931, *14,* 1–36.

Watt, W. C. Autonome Linguistik, Psycholinguistik, kognitive Linguistik. In H. Leuninger, M. H. Miller, & F. Müller (Eds.), *Linguistik und Psychologie.* Frankfurt: Athenäum, 1973.

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION, 12/1/85

Title: *Music Perception.* Pub. No. 07307829. Frequency: Quarterly. Four issues published annually. Subscription price: $30.00 individuals, $60.00 institutions. Location of office of publication: 2120 Berkeley Way, Berkeley, Alameda County, CA 94720. Headquarters of publishers: Same. Publisher: University of California Press, 2120 Berkeley Way, Berkeley, CA 94720. Editor: Diana Deutsch, Dept. of Psychology, University of California, San Diego, La Jolla, CA 92093. Owner: The Regents of the University of California, Berkeley, CA 94720. The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes have not changed during the preceding 12 months.

Extent and nature of circulation	Av. no. copies each issues preceding 12 months	Actual no. copies of single issue pub. nearest to filing date.
A. Total no. copies printed	972	935
B. Paid circulation, Mail subscriptions	505	525
C. Total paid circulation	505	525
D. Free distribution	71	71
E. Total distribution	576	596
F. Copies not distributed	396	339
G. Total	972	935

I certify that the statements made by me above are correct and complete. Ellie Young, Journals Manager.