Psychological Review
1981, Vol. 88, No. 6, 503-522

0033-295X/81/8806-0503$00.75

The Internal Representation of Pitch Sequences in Tonal Music

Diana Deutsch
University of California, San Diego

John Feroe
Department of Mathematics
Vassar College

A model for the internal representation of pitch sequences in tonal music is advanced. This model assumes that pitch sequences are retained as hierarchical networks. At each level of the hierarchy, elements are organized as structural units in accordance with laws of figural goodness, such as proximity and good continuation. Further, elements that are present at each hierarchical level are elaborated by further elements so as to form structural units at the next-lower level, until the lowest level is reached. Processing advantages of the system are discussed.

It may generally be stated that we tend to encode and retain information in the form of hierarchies when given the opportunity to do so. For example, programs of behavior tend to be retained as hierarchies (Miller, Galanter, & Pribram, 1960) and goals in problem solving as hierarchies of subgoals (Ernst & Newell, 1969). Visual scenes appear to be encoded as hierarchies of subscenes (Hanson & Riseman, 1978; Navon, 1977; Palmer, 1977; Winston, 1973). The phrase structure of a sentence lends itself readily to hierarchical interpretations (Chomsky, 1963; Miller & Chomsky, 1963; Yngve, 1960). When presented with artificial serial patterns that may be hierarchically encoded, we readily form encodings that reflect pattern structure (Bjork, 1968; Kotovsky & Simon, 1973; Restle, 1970; Restle & Brown, 1970; Simon & Kotovsky, 1963; Vitz & Todd, 1967, 1969).

In considering how we form hierarchies, however, theories have generally been constrained by the nature of the stimulus material under consideration. For example, visually perceived objects are naturally formed out of parts and subparts. The hierarchical structure of language must necessarily be constrained by the logical structure of events in the world. The attainment of a goal is generally arrived at by an optimal system of subgoals, and so on.

An analogous situation exists for theories based on experiments utilizing serial patterns that were devised by the experimenter. To take a concrete example, Restle's (1970) theory of hierarchical representation of serial patterns evolved from findings based on the following experimental paradigm. Subjects were presented with a row of six lights, which turned on and off in repetitive sequence, and they were required on each trial to predict which light would come on next. The sequences were structured as hierarchies of operators. For instance, given the basic sequence X = (1, 2), the operation R ('repeat of X') produces the sequence 1 2 1 2, the operation M ('mirror-image of X') produces the sequence 1 2 6 5, and the operation T ('transposition +1 of X') produces the sequence 1 2 2 3. Through recursive application of such operations, long sequences can be produced that have compact structural descriptions. Thus M(T(R(T(1)))) describes the sequence 1 2 1 2 2 3 2 3 6 5 6 5 5 4 5 4. Restle and Brown (1970), using sequences constructed in this fashion, found compelling evidence that subjects were encoding these patterns in accordance with their hierarchical structure. However, each pattern was constructed so as to allow for only one parsimonious interpretation. Thus it is difficult to estimate the generalizability of this model to situations where alternative hierarchical realizations are possible.

Preparation of this paper was supported by United States Public Health Service Grant MH-21001.

Requests for reprints should be sent to Diana Deutsch, Department of Psychology, C-009, University of California, San Diego, La Jolla, California 92093.

In contrast, the hierarchical structure of tonal music provides us with a unique opportunity to examine how we optimally form hierarchies, since such music is solely the product of human processing mechanisms, unfettered by external constraints. Further, tonal music can reasonably be considered to have evolved so as to capitalize on these mechanisms.

In this article we propose a model of how the observer represents the pitch sequences of tonal music in abstract form. This model falls into the class of those developed by Leewenberg (1971), Restle (1970; Restle & Brown, 1970), Simon and his colleagues (Simon, 1972; Simon & Kotovsky, 1963; Simon & Summer, 1968; Greeno & Simon, 1974), and Vitz and Todd (1967, 1969), among others; in that it proposes a specific language or notation for describing serial patterns, and this language is considered to reflect specific encodings. Indeed, many of the concepts and certain notations are owed to this previous work, as will be described below. However, our model differs from earlier ones in its basic architecture. In essence it may be characterized as a hierarchical network, at each level of which structural units are represented as an organized set of elements. Elements that are present at any given level are elaborated by further elements so as to form structural units at the next-lower level. It is further proposed that gestalt principles such as proximity and good continuation contribute to organization at each hierarchical level.

Before embarking on a formal description of the model, it should be noted that this concerns the representation of pitch information at the highest stage of abstraction, and that such information is assumed to be represented in parallel at lower stages also. At the lowest stage absolute pitch values are held to be represented, and interactions in storage that occur at this stage have been described elsewhere (Deutsch, 1975, in press-a). The next-higher stage is concerned with abstracted intervals and chords (Deutsch, 1969, 1978b). At the highest stage pitch information is further mapped onto a set of highly overlearned alphabets (Cuddy & Cohen, 1976; Cuddy, Cohen, & Miller, 1979; Deutsch, 1977, 1980; Dowling, 1978; Francès, 1958; Krumhansl, 1979; Krumhansl & Shepard, 1979).

Craik and Lockhart (1972) have argued that the higher the stage of abstraction of information, the longer its persistence in memory. This may well be true of the system retaining pitch information. Memory for melodic and harmonic intervals clearly persists longer than memory for absolute pitch values (Attneave & Olson, 1971; Deutsch, 1969). It appears plausible that memory for higher order abstractions persists longer still, but this hypothesis requires experimental verification.

The Model

Our model can best be introduced by musical example. Let us consider the pitch sequence shown on Figure 1(a). One way to represent this sequence is in terms of steps traversing the chromatic scale. We may say that a basic subsequence consisting of a step up this scale is presented four times in succession, the second presentation being four steps up from the first, the third being three steps up from the second, and the fourth being five steps up from the third. This type of analysis assigns prominence to the basic subsequence, and does not relate the successive transpositions to each other in any meaningful way. If the observer did indeed encode the pitch sequence in such a fashion, we may expect the basic subsequence to be well remembered, but the exact positions at which it is realized to be only poorly remembered.

The above analysis does not accord with musical intuitions. A musical analysis of this sequence would instead describe it as on the two structural levels shown on Figure 1. We can see that the basic relationship expressed here is that of the elaboration of a higher-level subsequence by a lower-level subse-

Figure 1. Pitch sequence represented on two hierarchical levels. (Panel a: Lower level. Panel b: Higher level.)

quence. At the higher level, shown on Figure 1 (b), there is an arpeggiation that ascends through the C major triad (C-E-G-C). At the lower level each note of this triad is preceded by a neighbor embellishment, thus forming a two-note pattern. We may represent this hierarchical structure in tree form as on Figure 2. (In this example, as is often the case, the elements of the higher-level subsequence are given metrical stress, to emphasize their prominence.)

Various points should be observed here. The first is that in this representation, a specific sequence of notes is realized at each structural level. This contrasts with representations in which specific events are realized only at the lowest structural level, the elements at higher levels being rule systems. We may also observe that in this representation, notes (or sequences of notes) that are present at any given level are also present at all lower levels. Thus the higher up a note (or sequence of notes) is represented in this hierarchy, the larger the number of its representations. This analysis therefore assigns prominence to elements at higher rather than at lower structural levels. In contrast, representations of serial patterns that are based on the concept of a subsequence that is repeatedly presented under transformation assigns greater prominence to elements at lower structural levels.

Another point illustrated by this example is that when a note at a higher level is elaborated by a sequence of notes at a lower level, the dominant note in the lower-level sequence (i.e., the note that also occurs at the higher level) need not be the first note of this sequence. In the present example the dominant note is the second of the two lower-level notes, the first being a submetrical embellishment of the second.

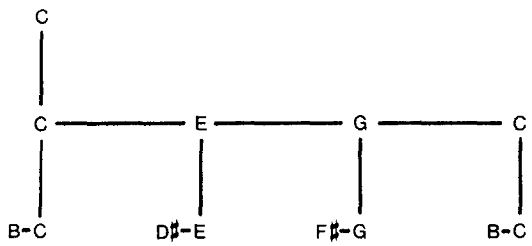

Figure 2. Tree diagram of pitch sequence shown on Figure 1.

Finally, we can see that in this example, distinct pitch alphabets are employed at different structural levels: The alphabet of the major triad is employed at the higher level, and the chromatic alphabet at the lower level. Such use of multiple alphabets occurs very commonly in music and, as we shall see, confers several processing advantages.

Formal Rules for the Representation[1]

Elementary Operators

1. An *alphabet* α is a linearly ordered set of symbols $\alpha = \{\ldots, e_1, e_2, \ldots\}$ which may be finite or extend infinitely in either direction. Common pitch alphabets in sequences of tonal music are the chromatic scale, the major and minor diatonic scales, and arpeggiated chords. These will be described below.

2. With respect to an element e_k in an alphabet α the *elementary operators* s (same), n (next), p (predecessor), n^i, p^i are defined as follows:[2]

$$s(e_k) = e_k$$

$$n(e_k) = e_{k+1}$$

$$p(e_k) = e_{k-1}$$

$$n^i(e_k) = e_{k+i}$$

$$p^i(e_k) = e_{k-i}$$

3. A *structure* A of length n is notated as

$$A = (A_0, A_1, \ldots, A_{\ell-1}, *, A_{\ell+1}, \ldots, A_{n-1})$$

where for each integer j with $0 \leq j \leq n - 1$, $j \neq \ell$, the symbol A_j is an elementary operator. The symbol * provides a reference point for the other operators. It appears exactly once in position ℓ where $0 \leq \ell \leq n - 1$. We note the particular cases $(*, A_1, \ldots, A_{n-1})$, $(A_0, \ldots, A_{n-2}, *)$ and $(*)$.

4. A *sequence* A is notated as $\{A; \alpha\}$ where A is a structure and α is an alphabet.

[1] Simon (1972) gives a detailed description of related formalisms.

[2] The symbols s (same), n (next), p (predecessor), n^i, and p^i are due to Simon (1972).

A sequence A together with a *reference element* $r \in \alpha$ produces a *sequence of notes*

$$S = \{A; r\} = (a_0, a_1, \ldots, a_{n-1})$$

where each $a_k \in \alpha$ is as follows:

$$a_k = \begin{cases} r & \text{if} \quad k = \ell \\ A_k(A_{k-1}(\ldots A_{l+1}(r)\ldots)) & \text{if} \quad k > \ell \\ A_k(A_{k+1}(\ldots A_{l-1}(r)\ldots)) & \text{if} \quad k < \ell \end{cases}$$

If the alphabet for a sequence is understood, the explicit reference to it in the notation may be omitted.

5. In any structure, the occurrence of a string of length k of an elementary operator A will be abbreviated kA. For example,

$$\{(*, n^2, n^2, n^2, p, p); \alpha\} = \{(*, 3n^2, 2p); \alpha\}.^3$$

Here we give some simple examples to illustrate the system so far presented. The representation $\{\{(*, 4n); C\}c\}$, where C represents the C major scale and c the reference element, corresponds to the sequence of notes C-D-E-F-G shown on Figure 3 (a). When the structure and reference element are held constant, but the alphabet of the C major triad is substituted for that of the C major scale, we have the representation $\{\{(*, 4n); C_{tr}\}c\}$, which corresponds to the sequence of notes C-E-G-C-E shown in Figure 3 (b). When the alphabet of the chromatic scale is substituted instead, we have $\{\{(*, 4n); Cr\}c\}$, which corresponds to the sequence of notes C-C#-D-D#-E shown on Figure 3 (c).

For a given structure and alphabet, a different sequence of notes is produced when the reference element is altered. Thus $\{\{(*, 4n); C\}e\}$ corresponds to the sequence of notes E-F-G-A-B shown on Figure 3 (d). Similarly, $\{\{(*, 4n); C_{tr}\}e\}$ corresponds to the sequence of notes E-G-C-E-G shown on Figure 3 (e); and $\{\{(*, 4n)Cr\}e\}$ corresponds to the sequence of notes E-F-F#-G-G# shown on Figure 3 (f).

It should be observed that the identical sequence of notes may be represented in terms of a number of alternative structures, depending on the placement of the reference element. Thus the sequence shown on Figure 3 (a) may be represented alternatively as $\{\{(p, *, 3n); C\}d\}$; as $\{\{(2p, *, 2n); C\}e\}$; or as $\{\{(4p, *); C\}g\}$; and so on. This flexibility in placement of the reference element is important and reflects the fact that the dominant element in a sequence will vary depending on the context in which this sequence occurs. For example, the lower-level sequences B-C, D#-D, F#-G, B-C shown on Figure 1 should be represented as $\{(p, *); Cr\}$, since in each case the second of the two notes is dominant, and these second notes combine to form a sequence at a higher level. However, given a different context, any of these two-note sequences might be represented as $\{(*, n); Cr\}$ instead. Thus whereas a large number of alternative representations may in principle be constructed for many sequences, the constraints imposed by the hierarchical organization of tonal music greatly reduce the number of alternative representations that the listener will produce. These constraints will be discussed in detail below.

Figure 3. Simple examples to illustrate the system.

Sequence Operators

1. A *compound sequence* is produced by the combination of two or more sequences under the action of a *sequence operator*. The central sequence operator is pr (prime), with two others, ret (retrograde) and inv (inversion) defined as elaborations of pr. As is the case for sequences, the designation of a ref-

[3] The coding of a run of identical symbols in such a fashion has been proposed by others (e.g., Leewenberg, 1971; Restle, 1970; Simon, 1972; Vitz & Todd, 1969).

erence element r for a compound sequence produces a sequence of notes.

2. Consider two sequences: $A = \{(A_0, \ldots, *, \ldots, A_{n-1}); \alpha\}$ and $B = \{(B_0, \ldots, *, \ldots, B_{m-1}); \beta\}$ for not necessarily distinct alphabets α and β. Observe that for $r \in \alpha$, $\{A; r\}$ is a sequence of notes $(a_0, \ldots, a_{n-1})$, and that for each i, $0 \leq i \leq n - 1$ such that $a_i \in \beta$, $\{B; a_i\}$ is a sequence of notes $(b_{i0}, \ldots, b_{i(m-1)})$. The compound sequence $A[\mathrm{pr}]B$ together with the reference element r produces the sequence of notes of length $n \times m$.

$$\begin{aligned}\{A[\mathrm{pr}]B; r\} &= \{B; a_0\}, \{B; a_1\}, \ldots, \{B; a_{n-1}\} \\ &= (b_{00}, b_{01}, \ldots, b_{0(m-1)}, b_{10}, \ldots, \\ &\qquad b_{1(m-1)}, \ldots, b_{(n-1)0}, \ldots, b_{(n-1)(m-1)}).\end{aligned}$$

Note that this is possible only if $a_i \in \beta$ for $0 \leq i \leq n - 1$. (Thus there are constraints on the alphabet of B imposed by the alphabet of A.)

This process is reversible; that is, the sequence of notes

$$(b_{00}, \ldots, b_{0(m-1)}, b_{10}, \ldots, b_{(n-1)(m-1)})$$

and the sequence $B = \{B; \beta\}$ produce the representation

$$\{B; a_0\}\{B; a_1\}, \ldots, \{B, a_{n-1}\}$$

and therefore the higher-level sequence of notes $(a_0, \ldots, a_{n-1})$.

The example shown on Figure 1 provides a simple illustration of the use of the operator pr (prime). This has the representation:

$$A = \{(*, 3n); C_{tr}\}$$
$$B = \{(p, *); Cr\}$$
$$S = \{A[\mathrm{pr}]B; c\}$$

where C_{tr} represents the C major triad, Cr the chromatic scale, and c the reference element.

3. For any sequence $B = \{(B_0, \ldots, B_{\ell-1}, *, B_{\ell+1}, \ldots, B_{m-1}); \beta\}$ define the retrograde sequence $\bar{B} = \{(B_{m-1}, \ldots, B_{\ell+1}, *, B_{\ell-1}, \ldots, B_0); \beta\}$. The compound sequence $A[\mathrm{ret}]B$ together with the reference element r produces the sequence of notes

$$\{A[\mathrm{ret}]B; r\} = \{A[\mathrm{pr}]\bar{B}; r\}$$

4. For any sequence $B = \{(B_0, \ldots, B_{\ell-1}, *, B_{\ell+1}, \ldots, B_{m-1}); \beta\}$ define the inverted sequence $\hat{B} = \{(\hat{B}_0, \ldots, \hat{B}_{\ell-1}, *, \hat{B}_{\ell+1}, \ldots, \hat{B}_{m-1}); \beta\}$ where

$$\hat{B}_i = \begin{cases} n & \text{if } B_i = p \\ n^i & \text{if } B_i = p^i \\ s & \text{if } B_i = s \\ p & \text{if } B_i = n \\ p^i & \text{if } B_i = n^i \end{cases}$$

The compound sequence $A[\mathrm{inv}]B$ together with the reference element r produces the sequences of notes $\{A[\mathrm{inv}]B; r\} = \{A[\mathrm{pr}]\hat{B}; r\}$.

5. Recognizing that a structure might be invoked with different alphabets, define a sequence with multiple alphabets as $B = \{B; \beta_0, \beta_1, \ldots, \beta_{n-1}\}$ where each β_i is an alphabet not necessarily distinct from the others. A compound sequence, say $A[\mathrm{pr}]B$ where $A = \{A; \alpha\}$ for $A = (A_0, \ldots, *, \ldots, A_{\ell-1})$, is realized for a reference element $r \in \alpha$ as a sequence of notes

$$\begin{aligned}\{A[\mathrm{pr}]B; r\} = &\{\{B; \beta_{0(\mathrm{mod}\, n)}\}; a_0\}, \\ &\{\{B; \beta_{1(\mathrm{mod}\, n)}\}; a_1\}, \ldots, \\ &\{\{B; \beta_{\ell-1(\mathrm{mod}\, n)}-1(\mathrm{mod}\ n)\}; a_{l-1}\}\end{aligned}$$

Similar definitions hold for ret and inv. Note that the single alphabet case given in Rule 2, that is, the case $n = 1$, is simply a particular case of this rule.

6. The power of the sequence operators is extended by allowing for a string of operators to act on a string of sequences. For any sequence $A = \{(A_0, \ldots, *, \ldots, A_{n-1}); \alpha\}$, sequences $B_0, B_1, \ldots, B_{M-1}$, and sequence operators $\mathrm{op}_0, \mathrm{op}_1, \ldots, \mathrm{op}_{N-1}$ the compound sequence

$$A[\mathrm{op}_0, \mathrm{op}_1, \ldots, \mathrm{op}_{N-1}](B_0, \ldots, B_{M-1})$$

together with the reference element r produces the sequence of notes

$$\begin{aligned}\{A[\mathrm{op}_0, \ldots, \mathrm{op}_{N-1}](B_0, \ldots, B_{M-1}); r\} \\ = \{C_0; a_0\}, \{C_1; a_1\}, \ldots, \{C_{n-1}; a_{n-1}\}\end{aligned}$$

where $\{A; r\} = (a_0, \ldots, a_{n-1})$ and

$$C_i = \begin{cases} B_{i(\text{mod } M)} & \text{if} \quad op_{i(\text{mod } N)} = \text{pr} \\ \bar{B}_{i(\text{mod } M)} & \text{if} \quad op_{i(\text{mod } N)} = \text{ret} \\ \hat{B}_{i(\text{mod } M)} & \text{if} \quad op_{i(\text{mod } N)} = \text{inv} \end{cases}$$

Note that for N = M = 1, the definitions in Rules 2, 3, and 4 are simply special cases of this rule.

A simple example of the use of pr (prime) together with ret (retrograde) is illustrated on Figure 4 (a). This may be represented as

$$A = \{(*, s); C\}$$
$$B = \{(*, 2n); C\}$$
$$S = \{A[\text{pr, ret}]B; c\}$$

where C represents the C major scale and c the reference element.

A simple example of the use of pr (prime) together with inv (inversion) is illustrated on Figure 4 (b). This may be represented as

$$A = \{(*, p); C_{tr}\}$$
$$B = \{(*, n, p); C_{tr}\}$$
$$S = \{A[\text{pr, inv}]B; c\}$$

where C_{tr} represents the C major triad and c the reference element.

7. In any compound sequence $A[op_0, \ldots, op_{N-1}](B_0, \ldots, B_{M-1})$ it is permissible for one or more of $A, B_0, \ldots, B_{M-1}$ to be compound sequences. In this case Rules 2, 3, 4, 5, and 6 above apply as stated, subject to the restriction imposed by the fact that $\bar{B}$ and $\hat{B}_i$ are not defined if B_i is a compound sequence.

8. In any compound sequence the occurrence of a string of length k of a sequence A or sequence operator op will be abbreviated kA and kop respectively.

a.

b.

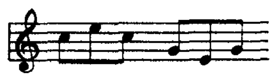

Figure 4. Simple examples to illustrate the use of sequence operators ret (retrograde) and inv (inversion).

Alternation

Consider two sequences of notes,

$S = (a_0, a_1, \ldots, a_{n-1})$ of length n, and

$T = (b_0, b_1, \ldots, b_{m-1})$ of length m.

For integers i and j, subject to the constraint that $n/i = m/j$ is an integer k, define the sequence of notes of length n + m

$$S[\text{alt } i, j]T$$
$$= (a_0, a_1, \ldots, a_{i-1})(b_0, b_1, \ldots, b_{j-1})$$
$$(a_i, \ldots, a_{2i-1})(b_j, \ldots, b_{2j-1}) \ldots$$
$$(a_{(k-1)i}, \ldots, a_{ki-1})(b_{(k-1)j}, \ldots, b_{kj-1})$$

Note that if n = m and i = j = 1 this results in the simple alternation of the elements of two equally long sequences of notes.

A relatively simple example illustrating the use of the alternation operation is given below, illustrated on Figure 11.

Choice of Sequence Operators

The sequence operators pr (prime), ret (retrograde), and inv (inversion) were chosen from considerations of musical analysis. Prime is the basic operation that produces a compound sequence from a set of sequences. The term prime is employed in the theory of twelve tone music to refer to the presentation of a row of tones without transformation (Perle, 1972). The term is borrowed here, but no other assumptions from twelve-tone theory are implied. The term retrograde is used in music theory to refer to the presentation of a sequence of notes in reverse order. Similarly, the term inversion is employed to refer to the presentation of a sequence of notes in such a way that all ascending steps become descending steps, and vice versa. Retrogression and inversion are frequently used as compositional devices in both traditional and contemporary music.

It should be noted, however, that in tonal music, inversion takes place along a given pitch alphabet (such as a diatonic scale or a triad) with the result that interval sizes are typically altered. This is captured in our formalism. (In atonal music based on the twelve-tone chromatic scale, inversion results in the preservation of interval sizes also; we do not assume that this is necessary). The frequent use of retrogression and inversion in music provides strong evidence that we employ these operations with ease.

A further advantage of these operations is that they considerably reduce the number of structures that are required. For example, given the basic structure (*, n), its retrograde is (n, *) and its inversion is (*, p). These operations therefore allow for a considerable reduction in memory load.

Pitch Alphabets

The system as so far described specifies the pitch alphabet associated with each structure in absolute terms. This device was employed in order to simplify the exposition of other parts of the system; however it is unsatisfactory on several grounds. First, in order to represent even simple melodic lines in absolute terms we would need to call on a huge repertory of alphabets. Second, it is evident that we encode pitch alphabets in relational terms rather than absolute, since we retain segments of music in transposable form. Third, in tonal music there are certain well-defined rules governing relationships between pitch alphabets, and these rules considerably restrict the number of alphabets that can be invoked in combination. These rules are captured in notational devices used by musicians, and we propose that they reflect the ways in which relationships between alphabets are encoded.

One may think of the twelve-tone chromatic scale as the parent alphabet from which families of alphabets are derived. The alphabets most commonly employed in tonal music are the major and minor diatonic scales (e.g., C major, D minor). All these can be expressed in terms of steps traversing the chromatic scale. Thus one octave of the ascending C major scale may be notated as

$$\{\{(*, 2n^2, n, 3n^2, n); Cr\}; c\}$$

where Cr refers to the chromatic alphabet and *c* is the reference element. (The descending major scale is the ascending scale in retrograde form.) Similarly, one octave of the ascending D major scale may be notated as

$$\{\{(*, 2n^2, n, 3n^2, n)Cr\}; d\}$$

Observe that these two representations differ only in the identities of their reference elements. Thus we may assume that we have encoded in long-term memory the sequence

$$\{(*, 2n^2, n, 3n^2, n); Cr\}$$

which specifies any major diatonic scale.

The harmonic form of the minor scale may be notated as

$$\{(*, n^2, n, 2n^2, n, n^3, n); Cr\}$$

The natural form may be notated as

$$\{(*, n^2, n, 2n^2, n, 2n^2); Cr\}$$

In both cases the ascending form is the retrograde of the descending form. The melodic minor scale has two different representations depending on whether it is in ascending or descending form. In ascending form it may be notated as

$$\{(*, n^2, n, 4n^2, n); Cr\}$$

and in descending form as

$$\{(2n^2, n, 2n^2, n, n^2, *); Cr\}$$

Again we assume that these sequences are retained in long-term memory.

The term "key" is used to refer to the collection of notes forming a particular diatonic scale. Thus the term "key of C major" refers to the collection of notes (C, D, E, F, G, A, B). The term "key of D major" refers to the collection (D, E, F#, G, A, B, C#), and so on. Any segment of tonal music is held to be in one of the 12 possible major or minor keys. This will be reflected in our notation.

Another common alphabet employed in tonal music is the arpeggiation of a triad. Triads can be constructed on each note or degree of a diatonic scale. As shown on Figure 5, each triad has the structure

$$(*, 2n^2, n^3)$$

which is realized upon specifying as alphabet the diatonic scale on which it is based, and as reference element the diatonic position of its fundamental note. Thus in the key of C major the C major triad may be notated as

$$\{\{(*, 2n^2, n^3); C\}; 1\}$$

where C denotes the alphabet of the C major scale, and 1 specifies the diatonic position of its fundamental note. Similarly, again in the key of C major, the D minor triad may be notated as

$$\{\{(*, 2n^2, n^3); C\}; 2\}$$

Observe that the intervals comprising the different triads vary, so that they may be major, minor or diminished, depending on the scale degrees on which they are based. However this difference can be ignored in musical notation which may simply specify a triad by its scale degree. We assume that this reflects a simplicity of encoding, that is, that all triads are encoded in terms of the same overlearned structure.

Another alphabet that is traversed in tonal music is the arpeggiation of a seventh chord. Such a chord is formed by the addition to a triad of a note that forms an interval of a seventh with the fundamental note. As shown on Figure 6, each seventh chord therefore has the structure

$$(*, 3n^2, n)$$

regardless of the intervals formed by its components. Other chords may also serve as alphabets, and can be notated in analogous fashion.

We can take advantage of such relationships to simplify our notation and at the same time enable it to reflect more accurately the ways in which pitch alphabets and their relationships are encoded. We will dispense with specifying alphabets in absolute terms, with the exception of the chromatic scale (Cr). Instead, for each sequence of notes we shall specify a key, such as G (G major) or c (C minor). If the alphabet associated with a structure is diatonic it will not be specified further (as provided for in Rule 4). If it is triadic, we will only specify the scale degree on which it is based (I, II, etc.). Similarly if it is a seventh chord we will specify it as I^7, II^7, etc. The reference element (r) is also specified as a scale degree (Arabic numerals are used here to differentiate the specification of the reference element from the specification of a chord arpeggiation).

Thus the example on Figure 1 may be notated as

$$A = \{(*, 3n); I\}$$
$$B = \{(p, *); Cr\}$$
$$S = \{A[pr]B; 1\}C$$

where I indicates the triad on the first de-

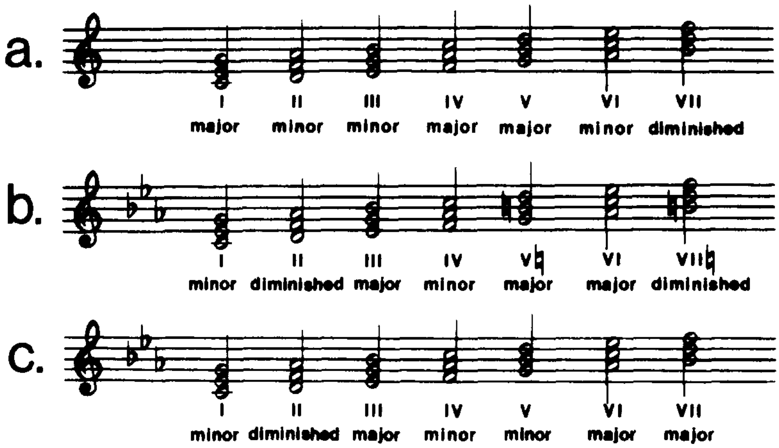

Figure 5. The diatonic traids. (Panel a: Major. Panel b: Minor in harmonic form. Panel c: Minor in natural form.)

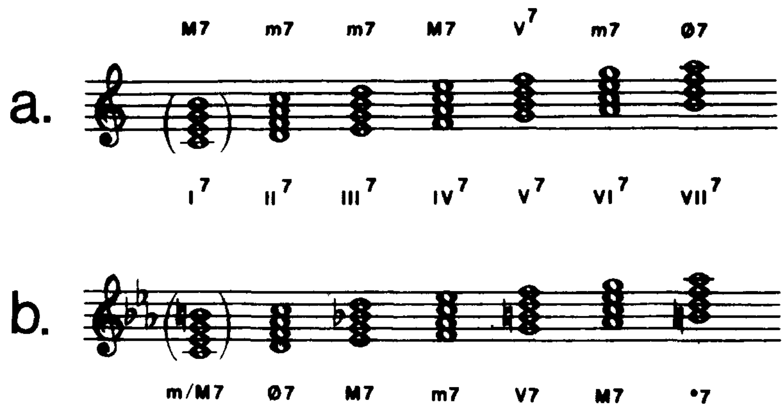

Figure 6. The diatonic seventh chords. (Panel a: Major. Panel b: Harmonic minor.)

gree, Cr the chromatic alphabet, 1 the reference element and C the key of C major.

Observe that if a sequence of notes is transposed to a different key, only one symbol in this notation is changed. (For example if the above sequence were transposed to the key of G major, the C would change to G.) Further if a sequence of notes is modulated between major and minor, again only one symbol is changed. (If the above sequence were modulated to C minor, the C would change to c.) Thus these new notational devices capture the ready transposability of melodic segments and their easy modulation: In each case the representation is barely altered.

It can be seen that in specifying a sequence that has an arpeggiated chord as alphabet, we are in effect specifying a structure that has as alphabet another structure (such as (*, $2n^2$, n^3), the structure for the triad), which has in turn as alphabet another structure (such as (*, $2n^2$, n, $3n^2$, n), the structure for the major diatonic scale), which in turn is based on the fundamental alphabet Cr. Thus in place of a substantially large number of alphabets we now have a very small number of highly overlearned structures that act on each other in hierarchical fashion. This system allows for the production of melodic segments of enormous variety through the invocation of a very small set of basic structures. To give a concrete idea of this encoding parsimony, let us restrict ourselves to tonal music that is composed of the following alphabets: the 12 major scales, the 12 ascending and 12 descending melodic minor scales, the 12 harmonic minor scales, the 12 major triads, the 12 minor triads, the 12 diminished triads, the 12 major seventh chords, the 12 minor seventh chords, the 12 diminished seventh chords, the 12 half-diminished seventh chords, the 12 dominant seventh chords, and the chromatic scale. This gives us a total of 145 possible alphabets, specified in absolute terms, that the listener would be required to invoke. However, in the present system the listener need only retain seven overlearned structures to obtain the same result (the structure for the major scale, the ascending and descending melodic minor scales, the harmonic minor scale, the triad, the seventh chord and the chromatic scale). Adding a further arpeggiated chord to the repertory of alphabets would analogously lead to the addition of a large number of alphabets as specified in absolute terms, but only one additional structure on the present system. This encoding parsimony is achieved through the superposition of one unequal-interval scale on another. In a musical system that was composed instead of equal-interval scales, the advantage of such a hierarchy would be greatly reduced (Figure 7).

We should also observe that in the present system there is a restriction on the number of alphabets, as specified in absolute terms, that are allowed to be combined to form a sequence of notes. This is in accordance with musical intuitions. If we were to pick a combination of alphabets at random, the resultant sequence of notes would be likely to sound incorrect to a listener who is familiar primarily with tonal music. We propose that this is because the listener would be unable to fit such a combination into the coding system proposed here. At the same time, tonal music is enormously versatile, and we are not generally conscious of these restrictions.

We are not assuming that the proposed system is hardwired in any way; clearly from consideration of other types of music it is not. However, it is likely that any musical tradition would have evolved its own system of rules that restrict the number of allowable combinations of alphabets, because without such restrictions the processing load would be too heavy.

Chord Progressions

A different type of generative process also occurs in tonal music. This concerns underlying harmony. The degree to which harmonic structure influences melodic structure has been the subject of considerable debate among music theorists; some asserting that melody can be understood only in terms of implied harmony (Schenker, 1956, 1973) and others hypothesizing a relative independence (Meyer, 1973; Narmour, 1977). It would seem that the degree to which one process depends on the other is a function both of the type of music and also of the tendencies of the individual listener.

Chord progressions are strongly hierarchical in nature. In tonal music the tonic triad predominates over the other triads in a key. It serves as a point of departure for harmonic progressions, and also as the ultimate goal of a harmonic progression. Thus the tonic triad may generate a progression that itself generates other progressions; and so on, in hierarchical fashion. A detailed analysis of chord progressions is, however, outside the scope of the present paper.

The generation of a sequence of chords differs in an important respect from the generation of a sequence of notes. A note has only one realization; however, a chord is an abstraction that can be realized in a number of different ways. Thus I in the key of C major may be realized as any combination of Cs, Es or Gs. V in the key of C major may be realized as any combination of Gs,

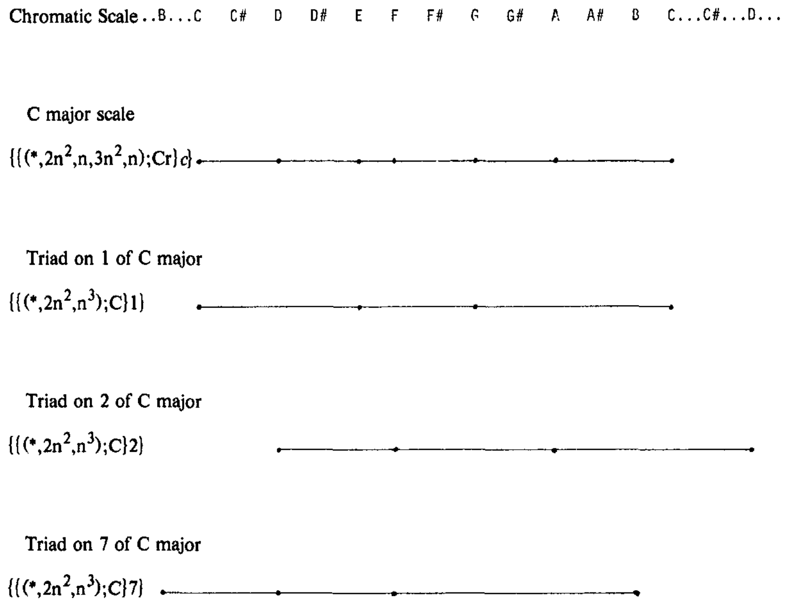

Figure 7. Parsimony of encoding achieved by embedding of alphabets. (The intervals composing the different triads vary depending on the scale degrees on which they are based. However, they have the identical abstract structure when encoded in terms of the diatonic scale, rather than in terms of the equal-interval chromatic scale.)

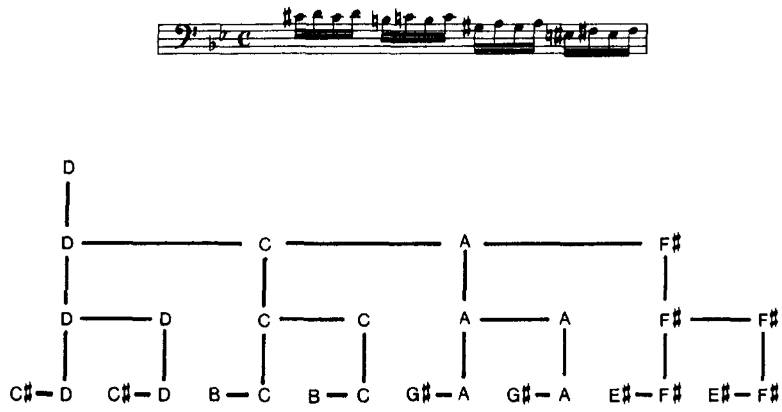

Figure 8. Example to illustrate the system. (From Beethoven, Sonata, op. 22.)

Bs or Ds. A harmonic progression therefore results from the generation of one abstraction by another. This type of generation is more similar to that found in transformational linguistics, where a grammatical category such as "noun phrase" may, through the application of a rewriting rule, produce other grammatical categories such as "determiner" and "noun." It may be observed, however, that the choice of a lexical formative ultimately depends on the grammatical category to which it belongs. In contrast, a given note can in principle serve as the realization of any part of a sequence of chords.

We assume that the generation of a sequence of chords may occur in parallel with the generation of a sequence of notes. The sequence of notes that is realized at any structural level is always compatible with the sets of alternative notes determined by the sequence of chords at that level. We assume that the listener is aware of this compatibility, which provides redundant information for use in retrieval.

A chord progression may be stated in horizontal form, and so may serve as a string of alphabets associated with a structure. This is illustrated in the third musical example below (see Figure 10 below).

Some Musical Examples

Here we give three examples to illustrate the system. The example on Figure 8 may be represented as on three structural levels.

$$A = \{(*, 3\mathrm{p});\ \mathrm{V}^7\}$$
$$B = \{(*, \mathrm{s})\}$$
$$C = \{(\mathrm{p}, *);\ \mathrm{Cr}\}$$
$$S = \{A[\mathrm{pr}]B[\mathrm{pr}]C;\ 5\}\mathrm{g}$$

This example illustrates the use of a chord arpeggiation as alphabet at the highest level and the chromatic scale at the lowest level. It also illustrates flexibility in placement of the reference element; in sequences A and B the reference element occurs first in the structure; in sequence C it occurs second.

The example on Figure 9 may also be represented as on three structural levels.

$$A = \{(*, 4\mathrm{p});\ 1\}$$
$$B = \{(*, \mathrm{n}, \mathrm{p});\ 1\}$$
$$S = \{A[\mathrm{pr}](B, 4\{(*)\})[\mathrm{inv}, 5\mathrm{pr}]\ (B, \{(*)\});\ 3\}\mathrm{b}$$

This example illustrates the use of the operator inv together with pr.

The example on Figure 10 may be represented as on four structural levels.

$$A = \{(*, \mathrm{p})\}$$
$$B = \{(*, \mathrm{n});\ 1, \mathrm{V}^7\}$$
$$C = \{(*, \mathrm{s})\}$$
$$D = \{(*, \mathrm{n})\}$$
$$S = \{A[\mathrm{pr}]B[\mathrm{pr}]C[\mathrm{pr}](D, \{(*)\},\ 2\{(*)\});\ 3\}\mathrm{A}$$

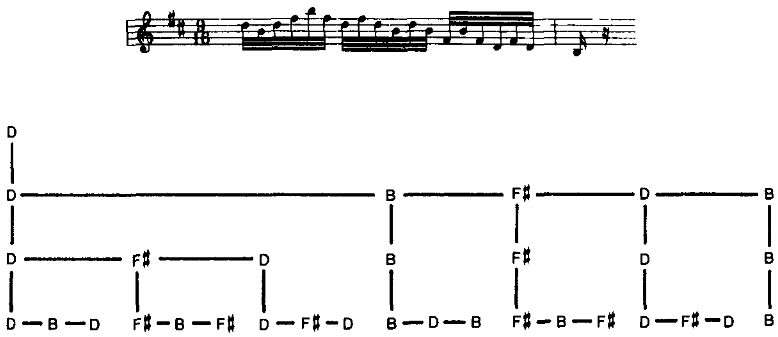

Figure 9. Example to illustrate the system. (From Bach, Sinfonia 15, BWV 801.)

This example illustrates the use of a chord progression as alphabet.

The example shown on Figure 11 may be represented as two interleaved sequences of notes, each consisting of two structural levels.

$$A = \{(*, 2p)\}$$
$$B = \{(*, n, p)\}$$
$$C = \{(*, 2s)\}$$
$$S_1 = \{A[2pr, inv]B; 3\}$$
$$S_2 = \{A[pr]C; 5\}$$
$$S = S_1[alt\ 1, 1]S_2, D$$

This example illustrates the use of the alternation operation.

Generation of a Pitch Sequence From its Stored Representation

It is assumed that, as reflected in the above formalism, sequence structures and their associated alphabets are retained in parallel at different hierarchical levels. It is further assumed that the observer most commonly generates a sequence of notes from its stored representation in a "top-down" fashion. The reference element is first applied to the highest level, thus realizing a sequence of notes at this level. These notes in turn serve as reference elements for the realization of a sequence of notes at the next-lower level (through the action of a compound operator or operators). This process is continued until the sequence of notes at the lowest level is realized.

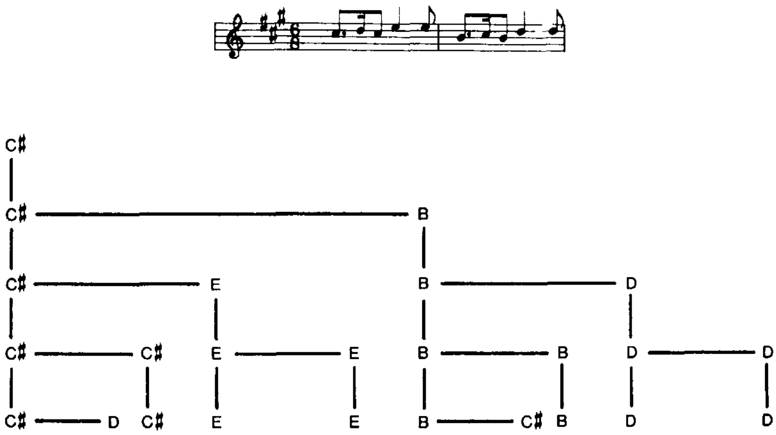

Figure 10. Example to illustrate the system. (From Mozart, Sonata, K. 300[f].)

Figure 11. Example to illustrate the system. (Sequences S_1 and S_2 are presented interleaved in time. From Beethoven, Six Variations on the Duet "Nel cor piu non mi sento" from Paisello's La Molinara.)

This system has the consequence that the notes occurring at the highest level should be recalled best and those occurring only at the lowest level should be recalled least well. This is for two main reasons. First, when retrieval occurs in a top-down fashion, in order to retrieve notes at a lower level, the higher-level notes must already have been retrieved. Thus, if a retrieval error occurs at a higher level, this will be reflected in further errors at all lower levels. Second, regardless of the direction of the retrieval process, once the full sequence of notes has been retrieved, the higher up a note is represented, the more often it is represented. This redundancy again increases the probability of accurate recall for the higher-level notes. Such an emphasis on the higher-level notes, which is a consequence of this system, is in accordance with musical intuitions and also with assumptions generally made by music theorists.

The Acquisition of a Representation

This section is concerned with the processes whereby the listener acquires a representation from the pattern of sounds that he or she hears. It is assumed that an initial set of groupings is formed on the basis of simple perceptual mechanisms, following which more complex mechanisms are invoked.

One of the most powerful principles involved in grouping a sequence of items is temporal proximity. This has been shown using a variety of stimulus materials (Bower & Springston, 1970; Bower & Winzenz, 1969; Dowling, 1973a; Handel, 1973; McLean & Gregg, 1967; Mueller & Schumann, 1894; Restle, 1972). A study addressed to this issue specifically with regard to pitch sequences was performed by Deutsch (1980). Subjects were presented with sequences of 12 notes which they recalled in musical notation. In the first experiment half of the sequences were structured in accordance with the present model, such that a higher-level subsequence of four elements acted on a lower-level subsequence of three elements. The remaining sequences were unstructured. Sequences were presented either with no temporal segmentation, with segmentation in groups of three (i.e., in accordance with tonal structure), or with segmentation in groups of four (i.e., in conflict with tonal structure). It was found that the level of recall for the structured sequences was very high in the absence of temporal segmentation, and even higher when segmentation was in accordance with tonal structure. The recall level was much lower for structured sequences that were segmented in conflict with tonal structure, as it was for the unstructured sequences. Analyses of serial position curves and transition shift probabilities demonstrated that the subjects were grouping these pitch sequences in accordance with temporal proximity rather than tonal structure when the two were placed in conflict. A second experiment examined the effects of compatible and incompatible segmentation for sequences that were hierarchically structured such that the lower-level subsequences consisted of either groups of three or groups of four. In both cases recall was excellent when temporal segmentation was in accordance with tonal structure, and poor when temporal segmentation conflicted with tonal structure.

This study demonstrates that even with sequences whose tonal structure is so clear as to produce a very high level of recall in the absence of temporal segmentation, segmentation in conflict with tonal structure essentially obliterates the listener's ability to exploit this structure to produce a parsimonious representation. This emphasizes the importance of low-level perceptual grouping in the induction of a sequence representation.

Another principle involved here is proximity along the pitch dimension. There is a strong tendency to group together elements that are proximal in pitch, and to separate those that are farther apart. For this reason, pitch separation between two melodic lines is required to achieve the perception of pseudopolyphony (Dowling, 1973b), and so for the encoding of a representation involving the alternation operation.

Processing difficulties have also been shown to occur between temporally adjacent notes that are widely separated in pitch (Bregman, 1978; Bregman & Campbell, 1971; Deutsch, 1972; Van Noorden, 1975). Thus, a prerequisite for the formation of coherent perceptual groupings is that the pitch separation between the notes within a group should not be too large. If a sequence of notes is presented such that adjacent subsequences are in different pitch ranges, the listener will tend very strongly to form groupings in accordance with pitch proximity. If the pitch ranges of adjacent subsequences are too far apart, the listener may be unable to integrate the key elements of these subsequences, and so be unable to form higher-order linkages between them. Optimally for our purposes, therefore, adjacent subsequences should be composed of notes that differ somewhat in pitch range, but not so much as to prevent the formation of higher-order linkages between the key elements of these subsequences.[4]

Perceptual groupings are also likely to be formed on the basis of loudness, timbre, or spatial location. As discussed above for the case of pitch, substantial differences along such dimensions will act as powerful grouping principles; however, if the differences are too large, the listener may be unable to integrate the key elements of the different subsequences. So again, the optimal perceptual condition here is some difference along the given dimension, but not too large a difference.

In considering low-level perceptual factors that lead the listener to choose an element in a subsequence as the dominant element, a similar argument applies. If this element differs from the others along some dimension (e.g., if it is higher, louder, or has a distinctive timbre), it will assume prominence. However, if this difference is too large, it will instead be dissociated from the other elements in the group. (This point has been made by Cooper & Meyer, 1960.) In general, patterns of metrical stress provide strong cues for the formation of tonal hierarchies.

Other simple perceptual principles are also involved. For example, sequences whose components combine to form a unidirectional pitch change are likely to be perceived as a group. This may be regarded as an example of the principle of good continuation (Divenyi & Hirsh, 1974; McNally & Handel, 1977; Nickerson & Freeman, 1974; Van Noorden, 1975; Warren & Byrnes, 1975). Further, if the pitch contour is repeatedly

[4] This issue is a thorny one, and the reader is referred to Deutsch (in press-a, in press-b) for a discussion of the conditions under which perceptual integration of sequences of notes that are far apart in pitch is made possible. This also includes a discussion of octave equivalence effects.

presented, the listener will tend to form groupings on the basis of this identity of contour. Dowling (1978) has made the point that contour, independent of either interval size or number of steps along a scale, is treated as a perceptual attribute in music.

In addition to the general, rather primitive perceptual principles we have described, the encoding of a sequence representation must also involve complex processes, in which the listener draws on his expectations about a given musical style. Perhaps the most important of these is the process of key attribution. An extended discussion of how this is achieved is beyond the scope of the present paper. It is sufficient to note here the experimental evidence that key attribution is readily and quickly accomplished, and on the basis of very little information (Cohen, Note 1; Cuddy, Cohen, & Miller, 1979).

Once a key has been attributed, it is assumed that the listener searches for notes that are prominent within the key as candidates for inclusion in higher-level subsequences. It is generally accepted in the theory of tonal music that the first, third, fifth, and eighth scale degrees, forming the tonic triad, have prominence or conceptual priority over the other scale degrees. (Thus, for instance, in the key of C the notes C, E, and G have prominence over the other notes.) The remaining notes in the diatonic scale in turn have prominence over the rest of the notes in the chromatic scale (see also Krumhansl, 1979). Thus, in representations of simple tonal music, the tonic triad is most likely to be traversed at the highest structural level, and the chromatic scale at the lowest level. It is assumed that the listener makes use of this knowledge in assigning notes to different structural levels.

Once such a preliminary mapping has taken place, it is assumed that the listener attempts to form representations in which sequence structures are repeated at any hierarchical level. That is, the more often a structure is perceived as repeating at a given level, the greater the probability that it will be encoded at that level (see also Simon & Sumner, 1968).

So far we have been viewing the listener as generating a single representation for each pitch sequence. However, segments of music are often amenable to more than one analysis, and it can be shown that composers exploit such ambiguities. For example, two adjacent notes may clearly belong to separate groupings when a theme is first presented, but later the relationship between these notes may assume importance. Indeed, it has been argued that such structural ambiguity contributes importantly to interest in music (Lerdahl & Jackendorff, 1977; Meyer, 1973; Narmour, 1977). Given such evidence we assume that the listener often sets up multiple representations in parallel. At any one time, the representation that is most parsimonious, or that is most in accordance with perceptual grouping mechanisms, is most likely to be realized. However, given a change in the stimulus configuration (for example in its temporal patterning) an alternative representation may be realized instead.

Meyer (1973) presents a good example of such structural ambiguity. The theme of the first movement of Mozart's Sonata in A major (part of which is notated above) is given in Figure 12 (a). Meyer observes that as this theme is presented, the descending fourths E-B and D-A are not perceived by the listener, since these melodic intervals cross perceptual group boundaries as determined by the configuration as a whole. However, the potential for a representation that exploits the repeated descending fourths is present and is actualized in a late variation, shown in Figure 12 (b). Here the pattern of temporal relationships is such as to induce the listener to perceive an alternative representation instead.

This discussion of multiple representations may be related to the model proposed

a.

b.

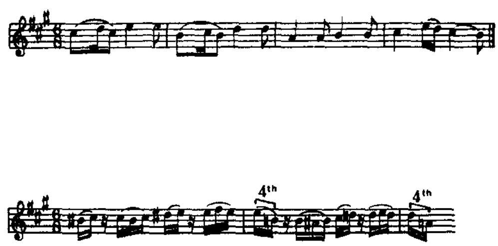

Figure 12. Panel a: Theme of first movement of Mozart, Sonata K. 300i. Panel b: Variation exploiting an alternative representation. (Adapted from Meyer, 1973.)

by Restle (1979) of the perception of motion configurations. Restle points out that a given display may potentially be represented in a large number of different ways, and can be thought of as ambiguous in principle. However many interpretations, though possible, are not seen. Restle argues that the observer will actualize the interpretation that has the minimum information load. If two or more interpretations have equal and minimal information loads, then both these interpretations will be seen, and the display will be ambiguous in practice.

One may also view the acquisition of a sequence representation as an ongoing process in which the listener, when presented with an initial sequence of pitch events, generates a set of alternative representations, some of which are confirmed by later pitch events and others of which are discarded. The later events in turn combine with earlier events to form the basis for a set of more elaborate representations; again, some of which are confirmed by later pitch events and others discarded. This process of generating successively more elaborate representations and eliminating earlier ones may be quite prolonged; but ultimately the listener achieves a set of alternative representations and their preference weightings.

This view is in accordance with the "implication–realization" model of Meyer (1973). Meyer argues that an implicative relationship is one in which a pitch event, called the generative event, is patterned in such a way that reasonable inferences can be made as to how the event is to be continued. A pitch event that is implied by a generative event may itself become a generative event at a higher level. Meyer argues that in forming such implications the listener relies in large part on principles such as good continuation at each hierarchical level. For example, a linear pattern (i.e., based on a diatonic scale) at one level may imply a further event, which when realized in turn implies a continuation of an arpeggiated triad at a higher level.

Processing Advantages of the System

The system proposed here has several processing advantages. The first involves redundancy of representation. It has been shown by Restle (1970) and Restle and Brown (1970) that when a sequence of elements is composed of subsequences that are linked together only by rule systems, recall is best for elements at the lowest level, and progressively poorer for elements at progressively higher levels. It follows that if no higher level sequences of notes were realized in music, we should expect musical segments to be recalled in fragmentary fashion: The listener would be most likely to make errors at the highest-level locations. The present system avoids this problem, since the higher up a note or sequence of notes is represented, the more often it is represented. This has the consequence that higher-level sequences serve to cement lower-level sequences together.

A second processing advantage involves the ability to invoke distinct alphabets at different structural levels. This primarily concerns the process whereby the listener acquires a representation from the pattern of sounds that he hears. The presence of distinct alphabets at different structural levels helps to separate out the sequences of notes associated with each level. Such an advantage is implied in statements by music theorists who advise, for example, that lower-level notes should be chromatically altered under certain circumstances to disambiguate the hierarchical structure of a melody. In the example in Figure 13 (a), for instance, the chromatic alterations make the hierarchical structure easy to perceive. However, if these notes were not chromatically altered, as shown in Figure 13 (b), the structure of the passage would become ambiguous (Forte, 1974). A similar line of reasoning applies to the chromatic alterations in the example on Figure 1 (a).

A third advantage to be gained from this system is that it enables sequence structures together with their associated alphabets to be encoded as chunks. Several investigators have shown that for serial recall of a string of items, performance levels are optimal when such a string is grouped by the observer into chunks of three or four items each (Estes, 1972; Wickelgren, 1967). Thus on the present system if a string of operators together with an alphabet were grouped together in chunks of three or four, superior

Figure 13. Clarification of hierarchical structure by chromatic alterations. (Panel a: From Mozart, Symphony in D major, K. 385. Panel b: Same passage with chromatic alterations removed [adapted from Forte, 1974].)

recall would be expected, in comparison with a system in which each operator is encoded independently. When segments of tonal music are notated on the present system, there emerges a very high proportion of chunks of three or four items each (e.g., {(*, p);1} or {(*, p, n); Cr}). This is exemplified by the examples in the present paper. As pitch sequences become more elaborate, they are represented as on a larger number of hierarchical levels, but the basic chunk size does not appear to vary with changes in sequence complexity. This chunking feature therefore serves to reduce memory load.

A further processing advantage that arises from a system in which strings of operators are chunked together, is that it enables representations to be created whose parts form configurations that are in accordance with laws of figural goodness (Wertheimer, 1923). For example, a structure consisting of operators of the same type (e.g., n, n^2, n) will produce a sequence that exhibits good continuation. Evidence has been obtained that pitch sequences are more efficiently perceived when their components combine to produce unidirectional pitch changes than when they do not (Divenyi & Hirsh, 1974; McNally & Handel, 1977; Nickerson & Freeman, 1974; Van Noorden, 1975; Warren & Byrnes, 1975).

Similarly, it has been shown in a number of contexts that sequences are more efficiently perceived when their components are proximal in pitch than when they are spaced farther apart (Bregman, 1978; Bregman & Campbell, 1971; Deutsch, 1975, 1978a; Dowling, 1973b; Van Noorden, 1975). This is a manifestation of the principle of proximity. In all the above work, however, only proximity along a single pitch scale (corresponding to log frequency) was considered. This principle may be extended to the use of scales based on abstract alphabets as well (see also Longuet-Higgins, 1978).

When segments of tonal music are represented in the present system, there emerges a very large proportion of single steps (n's or p's) in the representations. This is exemplified by the examples in the present paper. Double steps (n^2's or p^2's) also sometimes occur; but steps larger than these are rare. This is made possible only through the use of multiple pitch alphabets. For example, if only one alphabet were allowed, the pitch sequence shown on Figure 1 would have to be represented as

$$\{\{(*, n, n^3, n, n^2, n, n^4, n); Cr\}; 7\}C$$

However, with the use of the triadic alphabet in conjunction with the chromatic alphabet, this pitch sequence may be represented as

$$A = \{(*, 3n); 1\}$$

$$B = \{(p, *); Cr\}$$

$$S = \{A[pr]B; 1\}C$$

It can be seen that only single steps are employed in this second representation. The present system, therefore, by providing for the simultaneous invocation of distinct alphabets at different structural levels, enables the listener to be presented with melodic patterns of considerable richness and variety, while at the same time enabling an encoding mainly in terms of proximal relationships.

In addition to conforming with the principle of proximity, the fact that one step size is used much more frequently than others also acts to reduce processing load. This point has been made in a related context by Dowling (1978).

Discussion

The present model may be related to representations of hierarchical structure proposed by music theorists. The most influential work in this field is that of Heinrich

Schenker (1868–1935) who proposed a hierarchical system for tonal music that has points of similarity to the system proposed by Chomsky for linguistics (Chomsky, 1963). (In fact Schenker acknowledged that his ideas were inspired by the work of C. P. E. Bach [1714–1788] who in his *Essay on the True Art of Playing Keyboard Instruments* detailed the processes by which a simple musical event may be replaced by a more elaborate musical event that expresses the same basic content [Bach, 1949].) In Schenker's system music is regarded as a hierarchy in which pitch events at any given level are considered "prolonged" by sequences of pitch events at the next-lower level. Three basic levels are distinguished (though several hierarchical orderings may be found within each level). First there is the *foreground*, or surface representation; second there is the *middleground*; and third there is the *background*, or *Ursatz*. The *Ursatz* is considered to be a prolongation of the triad (Schenker, 1956, 1973).

Schenker's theory is based primarily on the concept that harmonic structure determines melodic structure; as such it has been criticized by other theorists who argue that melody often acts independently of harmony. A further criticism is of the rigid and 'a priori' nature of the *Ursatz*, which Schenker considered immune to change. In addition, his critics have argued that by positing only one structural possibility for a piece, Schenker's scheme is too inflexible, since multiple interpretations are often indicated. Another criticism is that Schenkerian analysis does not consider the importance of relationships formed within groupings: Only the hierarchical nature of the representation is considered. It should be noted that the present system does not run into any of the above difficulties. For literature related to Schenkerian analysis, see particularly Lerdahl and Jackendorff (1977), Meehan (1979), Meyer (1973), Narmour (1977), Salzer (1962), and Yeston (1977).

The general characteristics of the present hierarchical system may also be compared with those of systems proposed by others for the representation of visual arrays. Winston (1975) has proposed that visual scenes are represented as structures consisting of many embedded levels of organization. Restle (1979) has argued that certain moving configurations can best be represented as hierarchies, in which the motion of a point or points is described with reference to another set of points, which are themselves in motion with reference to a third set of points, and so on.

Bower and Glass (1976) have proposed that pictures are represented as structural hierarchies that are composed of related parts, each part corresponding to relationships among features at a lower level of analysis. They further assume that relationships within each part follow gestalt rules such as proximity and good continuation. As evidence for this, they showed that fragments of a picture that formed good patterns served as strong retrieval cues for redintegrating memory for the entire picture; whereas equally large fragments that did not form good patterns served as weak retrieval cues. Further, memory confusions occurred more often between patterns containing the same structural units, than between patterns containing different structural units.

Palmer (1977) has also proposed that visual shapes are represented as a hierarchy of structures, whose parts serve as structures at the next level down in the hierarchy. He also obtained evidence that we tend to form representations in which elements at each structural level are organized in accordance with laws of figural goodness such as proximity. When subjects were asked to divide figures into parts, they chose organizations that were most in accordance with the principle of proximity. Further, verification that a part was contained in a figure was faster, the greater the degree of goodness of the part within the figure. In addition, the time taken to synthesize a figure from two parts was shorter when there was a high degree of goodness of the parts within the figure.

The above findings in the case of vision lead us to speculate that the type of model proposed here may be applied to the internal representation of patterns beyond those of tonal music. For example, an analogous model could be proposed for the internal representation of the environment (Lynch, 1960); Chase & Chi, in press). It is unlikely that tonal music has evolved to accord with an arbitrary set of rules; rather it would be expected to reflect general principles of cognitive organization.

Reference Note

1. Cohen, A. Inferred sets of pitches in melodic perception. *Cognitive structure of musical pitch.* Symposium presented at the meeting of the Western Psychological Association, San Francisco, April 1978.

References

Attneave, F., & Olson, R. K. Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology*, 1971, *84*, 147-165.

Bach, C. P. E. *Essay on the true art of playing keyboard instruments* (W. J. Mitchell, Ed. and trans.). New York: Norton, 1949.

Bjork, R. A. All-or-none subprocesses in the learning of complex sequences. *Journal of Mathematical Psychology*, 1968, *5*, 182-195.

Bower, G. H., & Glass, A. L. Structural units and the redintegrative power of picture fragments. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 456-466.

Bower, G. H., & Springston, F. Pauses as recoding points in letter series. *Journal of Experimental Psychology*, 1970, *83*, 421-430.

Bower, G. H., & Winzenz, D. Group structure, coding, and memory for digit series. *Journal of Experimental Psychology Monograph*, 1969, *80*(2, Pt. 2).

Bregman, A. S. The information of auditory streams. In J. Requin (Ed.), *Attention and performance VIII.* Hillsdale, N.J.: Erlbaum, 1978.

Bregman, A. S., & Campbell, J. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, 1971, *89*, 244-249.

Chase, W. G., & Chi, M. T. H. Cognitive skill: Implications for spatial skill in large-scale environments. In J. Harvey (Ed.), *Cognition, social behavior and the environment.* Potomac, Md.: Erlbaum, in press.

Chomsky, N. Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2). New York: Wiley, 1963.

Cooper, G., & Meyer, L. *The rhythmic structure of music.* Chicago: University of Chicago Press, 1960.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 671-684.

Cuddy, L. L., & Cohen, A. J. Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology*, 1976, *28*, 255-270.

Cuddy, L. L., Cohen, A. J., & Miller, J. Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology*, 1979, *33*, 148-157.

Deutsch, D. Music recognition. *Psychological Review*, 1969, *76*, 300-307.

Deutsch. D. Octave generalization and tune recognition. *Perception & Psychophysics*, 1972, *11*, 411-412.

Deutsch, D. Musical illusions. *Scientific American*, 1975, *233*, 92-104.

Deutsch, D. Memory and attention in music. In M. Critchley & R. A. Henson (Eds.), *Music and the brain.* London: Heinemann, 1977.

Deutsch, D. Delayed pitch comparisons and the principle of proximity. *Perception & Psychophysics*, 1978, *23*, 227-230. (a)

Deutsch, D. Interactive effects in memory for harmonic intervals. *Perception & Psychophysics*, 1978, *24*, 7-10. (b)

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

Deutsch, D. The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press, in press. (a)

Deutsch, D. Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press, in press. (b)

Divenyi, P. L., & Hirsh, I. J. Identification of temporal order in three-tone sequences. *Journal of the Acoustical Society of America*, 1974, *56*, 144-151.

Dowling, W. J. Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, 1973, *4*, 37-40. (a)

Dowling, W. J. The perception of interleaved melodies. *Cognitive Psychology*, 1973, *5*, 322-337. (b)

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

Ernst, G. W., & Newell, A. *GPS: A case study in generality and problem solving.* New York: Academic Press, 1969.

Estes, W. K. An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, D.C: Winston, 1972.

Forte, A. *Tonal harmony in concept and practice* (2nd ed.). New York: Holt, Rinehart & Winston, 1974.

Francès, R. *La perception de la musique.* Paris: Vrin, 1958.

Greeno, J. G., & Simon, H. A. Processes for sequence production. *Psychological Review*, 1974, *81*, 187-196.

Handel, S. Temporal segmentation of repeating auditory patterns. *Journal of Experimental Psychology*, 1973, *101*, 46-54.

Hanson, A. R., & Riseman, E. M. (Eds.). *Computer vision systems.* New York: Academic Press, 1978.

Kotovsky, K., & Simon, H. A. Empirical tests of a theory of human acquisition of concepts of sequential events. *Cognitive Psychology*, 1973, *4*, 399-424.

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

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

Leewenberg, E. L. A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, 1971, *84*, 307-349.

Lerdahl, F., & Jackendorff, R. Toward a formal theory of music. *Journal of Music Theory*, 1977, *21*, 111-172.

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

Lynch, K. *The image of the city*. Cambridge, Mass.: Harvard University Press, 1960.

McLean, R. S., & Gregg, L. W. Effects of induced chunking on temporal aspects of serial retention. *Journal of Experimental Psychology*, 1967, *74*, 455-459.

McNally, K. A., & Handel, S. Effect of element composition on streaming and the ordering of repeating sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 451-460.

Meehan, J. R. An artifical intelligence approach to tonal music theory. *Proceeding of the Annual Conference: Association for Computing Machinery*, 1979, 116-120.

Meyer, L. B. *Explaining music: Essays and explorations*. Berkeley: University of California Press, 1973.

Miller, G. A., & Chomsky, N. Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2). New York: Wiley, 1963.

Miller, G. A., Galanter, E. H., & Pribram, K. H. *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston, 1960.

Mueller, G. E., & Schumann, F. Experimentelle Beitrage zur Untersuchung des Gedachtnisses. *Zeitschrift fur Psychologie und Physiologie der Sinnesorgane*, 1894, *6*, 81-190; 257-339.

Narmour, E. *Beyond Schenkerism*. Chicago: University of Chicago Press, 1977.

Navon, D. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 1977, *9*, 353-383.

Nickerson, R. S., & Freeman, B. Discrimination of the order of the components of repeating tone sequences: Effects of frequency separation and extensive practice. *Perception & Psychophysics*, 1974, *16*, 471-477.

Palmer, S. E. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977, *9*, 441-474.

Perle, G. *Serial composition and atonality* (3rd ed.). Berkeley: University of California Press, 1972.

Restle, F. Theory of serial pattern learning: Structural trees. *Psychological Review*, 1970, *77*, 481-495.

Restle, F. Serial patterns: The role of phrasing. *Journal of Experimental Psychology*, 1972, *92*, 385-390.

Restle, F. Coding theory of the perception of motion configurations. *Psychological Review*, 1979, *86*, 1-24.

Restle, F., & Brown, E. Organization of serial pattern learning. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 4). New York: Academic Press, 1970.

Salzer, F. *Structural hearing*. New York: Dover, 1962.

Schenker, H. *Neue musikalische Theorien and Phantasien: Der freie Satz*. Universal Edition, Vienna, 1956.

Schenker, H. *Harmony* (O. Jonas, Ed. and annotator; E. M. Borgese, trans.). Cambridge, Mass.: MIT Press, 1973.

Simon, H. A. Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 1972, *79*, 369-382.

Simon, H. A., & Kotovsky, K. Human acquisition of concepts for sequential patterns. *Psychological Review*, 1963, *70*, 534-546.

Simon, H. A., & Sumner, R. K. Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley, 1968.

Van Noorden, L. P. A. S. *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation. Technische Hogeschool, Eindhoven, The Netherlands, 1975.

Vitz, P. C., & Todd, T. C. A model of learning for simple repeating binary patterns. *Journal of Experimental Psychology*, 1967, *75*, 108-117.

Vitz, P. C., & Todd, T. C. A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, 1969, *76*, 433-449.

Warren, R. M., & Byrnes, D. L. Temporal discrimination of recycled tonal sequences: Pattern matching and naming of order by untrained listeners.*Journal of the Acoustical Society of America*, 1975, *18*, 273-280.

Wertheimer, M. Untersuchung zur Lehre von der Gestalt II. *Psychologische Forschung*, 1923, *4*, 301-350.

Wickelgren, W. A. Rehearsal grouping and the hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, 1967, *19*, 97-102.

Winston, P. H. Learning to identify toy block structures. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology: The Loyola Symposium*. Washington, D.C.: Winston, 1973.

Winston, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Yeston, M. (Ed.). *Readings in Schenker analysis and other approaches*. New Haven, Conn: Yale University Press, 1977.

Yngve, V. H. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 1960, *104*, 444-466.

Received December 15, 1980 ■