A Model of Expressive Timing in Tonal Music
Author(s): Neil Todd
Reviewed work(s):

Source: *Music Perception: An Interdisciplinary Journal*, Vol. 3, No. 1 (Fall, 1985), pp. 33-57
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40285321
Accessed: 15/12/2012 12:04

Music Perception
Fall 1985, Vol. 3, No. 1, 33–58

A Model of Expressive Timing in Tonal Music

NEIL TODD
University of Exeter

Listening to a concert, I often find myself unexpectedly in a foreign country, not knowing how I got there; a modulation had occurred which escaped my comprehension. I am sure that this would not have happened to me in former times, when a performer's education did not differ from a composer's.

(A. Schoenberg)

During a performance, a pianist has direct control over only two variables, duration and intensity (Seashore, 1938). Other factors such as pitch and timbre are determined largely by the composer and the mechanics of the instrument. Thus expressiveness imparted to a performance lies in the departures from metrical rigidity and constant intensity. In this article, the first of the two variables is considered and it is shown how a duration structure can be generated, corresponding to the rubato in a performance, from the musical structure. The main input to the model is the time-span reduction of Lerdahl and Jackendoff's theory (1977, 1983). Also shown is an interesting analogy between this model and the algorithms of Grosjean, Grosjean, and Lane (1979). Thus the hypothesis that expression is largely determined by musical structure, and the formal parallel between time-span reduction and prosodic structure are given empirical support.

Introduction

Music theorists have been concerned with the formulation of theories of abstract structures of music. Schoenberg (1969) was concerned with the relationship of different tonalities within a composition and evoked an abstract tonal space to demonstrate the relatedness of the various tonalities to the main tonality. Thus a composition was thought of as being "monotonal" where modulations within a movement are merely deviations from, and not negation of, the main tonality. Schenker (1906, 1935) produced a

Requests for reprints may be sent to Neil Todd, Department of Psychology, Washington Singer Laboratories, University of Exeter, Exeter EX4 4QG, England.

theory based on the hypothesis that music is organized hierarchically and that the listener attempts to organize the musical surface into a single coherent structure, such that it is heard as a hierarchy of relative importance. As a consequence, one could analyze a musical passage using a step-by-step simplification (or reduction) leaving a structural skeleton of the piece. Thus the music was thought to have several levels, "foreground," "middleground," and "background." Underlying the whole piece was what Schenker called the "Ursatz," which can be thought of as a kind of "deep structure."

These theories have attracted the attention of researchers from other disciplines. Psychologists interested in music perception have concentrated on the idea that perceptual processing involves the abstraction of an underlying structure from the musical surface. In particular, much attention has been paid to hierarchies of stability. Lerdahl and Jackendoff (1977, 1983), starting from Schenker's reduction hypothesis and adopting a theoretical position analogous to that taken in linguistics by Chomsky (1965, 1972), have developed a generative theory of tonal music. This theory is organized along four hierarchical dimensions: grouping structure, metrical structure, time-span reduction, and prolongation reduction. Each component is determined by well-formedness rules and preference rules.

It has been found here that the formalism developed in Lerdahl and Jackendoff's theory is at present the most useful language with which to relate expressive variables to the musical structure. In what follows I will take Lerdahl and Jackendoff's theory of grouping and combine it with the principle of phrase-final lengthening (the tendency to slow at a boundary) to generate a duration structure corresponding to the rubato in a performance. The central thesis here is that the performer uses phrase-final lengthening as a device to reflect some underlying structure abstracted from the musical surface.

Phrase-Final Lengthening

This effect is simply the tendency to slow at the end of a single motor action or sequence. It has been known and well documented for some time. For example, in 1905 R. H. Stetson wrote "There are many reasons for considering the phrase as simply the form of a single act . . . . The tensions of the muscle sets do not cease until the end of the phrase. The dynamic form of the phrase is the form of a movement; there is a rise to a central point of effort and then a decline at the end. Any elaborate, rapid flourish made with a pencil, or with a finger in the air will show just these dynamic variations . . . . In reciting verse, or in singing, a phrase becomes a single act of expiration; indeed, just this movement of breathing is probably the origin of musical phrasing." (p. 315)

It is generally accepted in speech that the main function of pausing is that of grouping (Black, Tosi, Singh, & Takefuta, 1966; Grosjean, 1979; Goldman-Eisler, 1972). Speakers tend to draw breath at the end of large conceptual units such as sentences and clauses. The use of pauses as a major boundary marker between and within sentences seems to be similar across those languages for which there is available data (Black et al., 1966). Also there is a tendency to lengthen the final elements in an utterance before the pause. Cooper (1976) has suggested that slowing down at the end is a natural tendency characterizing all motor sequences: similar patterns can be observed in birdsong and insect chirps. This phenomenon has been observed in music (Seashore, 1938; Gabrielsson, Bengtsson, & Gabrielsson, 1983; Clarke, 1984; Shaffer, Clarke, & Todd, 1985).

Not only does end-slowing seem to act as a boundary marker but, in certain domains, variations in relative lengthening, if perceived, contribute to the recovery of syntactic structure by the listener: the greater the lengthening the more important the syntactic break (O'Malley, Kloker, & Dara-Abrams, 1973; Scott, 1982).

Generative Music Theory

Before examining the requisite aspects of the theory, it is useful to look at some of the assumptions and idealizations Lerdahl and Jackendoff (1983) make, since any performance model must inherit those of its input.

They take the goal of music theory to be a "formal description of the musical intuitions of a listener who is experienced in a musical idiom." The concept of "experienced listener" is meant as an idealization; no two listeners will hear a piece in precisely the same way or detail. Nevertheless, there is usually considerable agreement on what is the most natural way of hearing a piece.

Another idealization is that the structures generated represent the "final state" of the listeners' comprehension of the piece. Clearly, a listener will not know the precise status of an event in the structure as it occurs during a performance. However, these idealizations are rather more justified for the performer. It is assumed here that a trained musician is likely to be a good candidate for the status of "experienced listener" and, if a particular piece has been practised, is likely to have some kind of global understanding of the piece.

A third idealization is that the structures generated represent the "preferred listening." This means that the experienced listener is more likely to attribute some structures to music than others. Thus any performance model generated by a "preferred listening" corresponds in some way to the most likely performance. It would not represent the "correct" performance.

Let us now look at the requisite aspects of the theory in more detail. In the theory groups are defined to be segments in the musical surface such as themes, phrases, and motifs. Thus grouping structure refers to the way in which the various phrases are organized. Metrical structure refers to the regular patterns of strong and weak stress points in the surface. Both grouping and metrical patterns exhibit the fundamental property of hierarchical structuring. Their abstract structures are generated in the formalism with well-formedness rules and preference rules. The rules are such that any one alone is a sufficient condition for choosing an analysis (or way of hearing) and when two or more rules apply, they may either reinforce each other or conflict. This usually results in there being an ambiguity between one structure or another.

Time-span reduction (TSR) is the generalized interaction of grouping structure and metrical structure and represents the rhythmical structure of the music. The elements that make up the TSR are the durations (or time spans) between one beat and another so that every group is a time span in the segmentation of the piece. An important TSR well-formedness rule says that for each time span there is an event that is the most important in that time span, such that all other events are subordinate to it. If a larger time span contains smaller time spans, a most important event must be selected from those of the smaller time spans. Thus a hierarchy of most important events is generated. The preference rules for selecting these most important events (or heads) fall into three basic categories: local rules, global rules, and structural accent rules, which involve the articulation of group boundaries at and above the level of the phrase. Structural accents are events that initiate and terminate arcs of tonal motion and are caused by points of melodic and harmonic stability. A cadenced group is one which at some level of reduction reduces to two structural accents, a structural beginning (denoted [b]) and a structural ending (denoted [c]). The structural accent rules are "top-down" in that they concern the function of the time-span head within the whole structure of the piece. It is the cadenced groups and higher levels that are of particular importance here, since expressive timing seems to operate on several levels. In the following discussion we join the reduction at the level that has been reduced to a set of structural beginnings and endings.

The structural beginning or ending of a phrase must emerge as its structurally most important event in the TSR. Thus one will dominate the other. This is represented by a tree notation as in Figure 1. Once the structural-accent content of a piece has been denumerated it is a matter of applying the preference rules to generate a complete tree structure for the piece, at and above the level of the phrase.

Prolongation reduction (PR) assigns pitches to a hierarchy that expresses the harmonic and melodic tension and relaxation. The elements of the PR are prolongation regions, which represent an overall tensing or relaxing in a progression. Tensings or relaxings internal to each region represent subor-

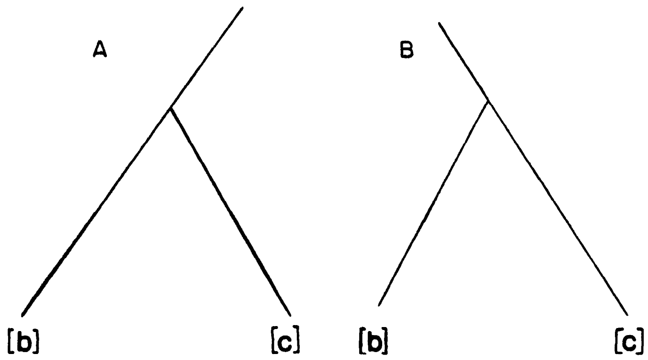

Fig. 1. (A) The structural beginning [b] dominates the structural ending [c]; (B) the structural ending [c] dominates the structural beginning [b].

dinate stages in the overall progression. PR is not used formally here but it is suggested later where it seems to affect timing.

Unit Time Span

In order to examine the real-time properties of a performance, we need to define some unit such that we can compare the real-time duration with metrical duration. This is not as trivial as it might seem. In all but the simplest pieces, the grouping structure and metrical structure are not in step such that a phrase ending may not occur at the end of a bar, for example. If the bar was chosen as a unit it would cross the group boundary and the information contained in the duration of the bar would be ambiguous. So, it would be convenient if we could choose units such that a group could be measured in whole numbers of units. However, it has become clear that expressive timing is organized componentially (Gabrielsson et al., 1983; Todd, in prep.) such that a performance is the resultant of the superposition of these components. So we need to keep in mind which component we wish to examine. It is not always possible to satisfy the whole number of units per group condition because of this.

Briefly, the number of components in the performance depends very much on the structure of the music. If the piece has very regular metrical and grouping structures (e.g., Bach Prelude No. 1) the performance is likely to exhibit a number of components. If the musical surface is irregular and complex it can be almost impossible to distinguish one component from another. For a local component (i.e., duration fluctuation at the note level) we would choose either normalized note length or some fraction of the beat as a unit. For a more intermediate component we might choose the bar or half-bar. As mentioned earlier, I am primarily concerned here with the groups at and above the level of the phrase, since phrase-final lengthening seems to be dominant at this level and hence in the intermediate components. A more complete account is provided in Todd (in prep.), which specifically addresses the issue of componential timing.

Structural Endings and Embedding Depth

Earlier the idea of structural accent was introduced. It is useful to formalize this concept a stage further. That is, we define a set of structural endings by referring to the time spans that contain them.

DEFN. The set C is an ordered set of time spans such that

$$C = (c_1, \ldots, c_n)$$

where c_j is the time span containing the jth structural ending. n is the total number of structural endings.

As mentioned earlier, the structural accents articulate group boundaries at and above the level of the phrase. Thus the set C completely denumerates the phrase content of a piece. The phrases are organized hierarchically so it is useful to formalize this idea too. This we do by defining another set associated with the set C such that it tells us how deeply embedded a particular ending is in the TSR.

DEFN. The set E is an ordered set of numbers with one–one correspondence with the elements of C such that

$$E = (e_1, \ldots, e_n)$$

where e_j is the embedding depth of the jth structural ending. The embedding depth of a particular ending c_j is equal to the number of structural beginnings dominated by c_j (Nb_j) plus the number of events forming the cadence, (Nc_j). That is

$$e_j = \mathrm{Nb}_j + \mathrm{Nc}_j$$

Thus if c_j is a full cadence (i.e., the harmonic progression V–I) then $\mathrm{Nc}_j = 2$. A full cadence is represented by two lines rather than one (Figure 2) and any left branches dominated by the cadence intersect at a node. Nb_j is calculated by counting the number of left branches dominated by the ending. Consider the example shown in Figure 2. Figure 2 represents a time span reduction for a set of four groups, each of which are four time span units (tsu) long. The structural endings are contained in the units 4, 8, 12, 16, hence the structural ending set is given by

$$C = (4, 8, 12, 16)$$

Applying the rules above for calculating embedding depth we get

$$E = [(0+1), (1+2), (0+1), (3+2)]$$

that is, $E = (1, 3, 1, 5)$.

Thus the two sets C and E completely denumerate the TSR at and above the level of the phrase.

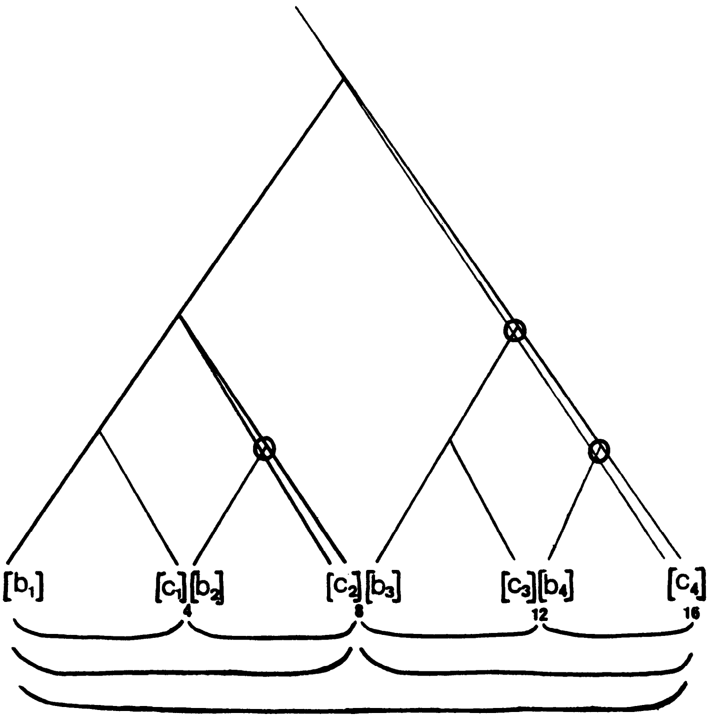

Fig. 2. The tree diagram shows a time-span reduction for a set of four groups of four unit time-spans. Right branches show where structural beginnings dominate structural endings and left branches show where structural endings dominate structural beginnings. If a structural ending contains two elements it is given a double branch, and any left branch intersects at a circle.

Duration

We can consider timing to be organized on about three levels:

1. A global component (GC) which is essentially tempo variation over the whole piece,
2. Intermediate components (IC) which we may regard as rubato, and
3. Local components (LC) which are fluctuations at the note level.

Each component is made up of a series of segments that seem to correspond to segments in the musical surface (groups). They also exhibit a hierarchical structure by the relative slowing of one segment to another.

As mentioned earlier, it is with the ICs that we are primarily concerned here. If a suitable uts is chosen we can ignore LCs for the time being. That is, a uts is chosen such that local fluctuations are not picked up. Suffice it to say that the LCs seem to be determined largely by stress patterns and metrical structure at the surface. The GC cannot be ignored, however, since any global tempo fluctuation is superimposed onto the IC.

With these considerations in mind, we make the following hypothesis about the IC. The task of a performer is

1. To perceive the structure of the music. That is, to organize the musical surface into a hierarchical structure. (As far as the model is concerned, we take this to mean the generation of a TSR and hence the sets *C* and *E*.)
2. To elucidate the rhythmic structure by
   a. slowing at the structural endings and
   b. reflecting hierarchical structure by the degree of slowing at an ending (i.e., the degree of slowing is a function of embedding depth).

This hypothesis agrees, in principle, with Cone (1967), when he says "a valid performance depends primarily on the perception and communication of the rhythmic life of a composition. That is, we must first discover the shape of the piece . . . and then try to make it as clear as possible to our listeners."

A Model (Intermediate Component)

As discussed earlier, the duration structure of a performance is characterized by a series of gestures or segments corresponding to the grouping structure. Here I have chosen a parabola to model these gestures, since it is the simplest function that captures the features of phrase-final lengthening or end-slowing. This is intended only as an approximation. In this section, I will develop this idea mathematically to produce an explicit function, or homeomorphism, which generates a duration structure corresponding to the grouping structure of a piece.

Let us consider the *j*th segment in the duration structure. Then if t_{ij} is the *i*th unit time span in the *j*th segment and $D(t_{ij})$ is the real-time duration of t_{ij} we can define

$$D(t_i^j) = M(t_i^j) + A, \tag{1}$$

where A is a constant and $M(t_{ij})$ is the deviation from strict metricality of t_{ij}.

We need to find an explicit form for $M(t_{ij})$. If we assume the segments to be parabolic we can write

$$M(t_{ij}) = m[t_{ij} - f(e_j)]^2, \tag{2}$$

where m is simply an amplitude and t_i^j starts at zero and counts up to the end of the segment. That is

$$0 \leqslant t_i^j \leqslant l_j$$

where l_j is the length of the *j*th segment.

Examination of performance data shows that the length of segments rarely exceeds 6 uts. The most frequently occurring length is 4 uts, as would be expected. Any phrases longer than 6 uts will break up into smaller segments in the duration structure. Thus

$$2 \leqslant l_j \leqslant 6$$

$f(e_j)$ is some function of the embedding depth e_j and controls the offset of the minimum of the parabola from the origin $t_i^j = 0$, so that increasing embedding depth increases the steepness of slowing in the segment.

It is convenient to normalize t_i^j since l_j is variable. We do this by dividing by l_j. So, if we let $T_i^j = t_i^j / l_j$ then

$$0 \leqslant T_i^j \leqslant 1$$

To see the constraints on the function $f(e_j)$, consider Figure 3. This represents a set of possible segments whose length is 4 uts. In order to arrive at a reasonable form for $f(e_j)$ we must make two assumptions. The first is that at the lowest embedding depth ($e_j = 1$) the minimum of the underlying parabola will be about halfway along the segment. That is,

$$f(1) = 1/2.$$

Now, the maximum embedding depth is likely to be about seven and we would like the upper bound of $f(e_j)$ to be zero. That is,

for $1 \leqslant e_j \leqslant 7$, $1/2 \geqslant f(e_j) \geqslant 0$.

The simplest function that satisfies these constraints is

$$f(e_j) = \frac{(e_j^{-1})}{12} - \frac{1}{2}$$

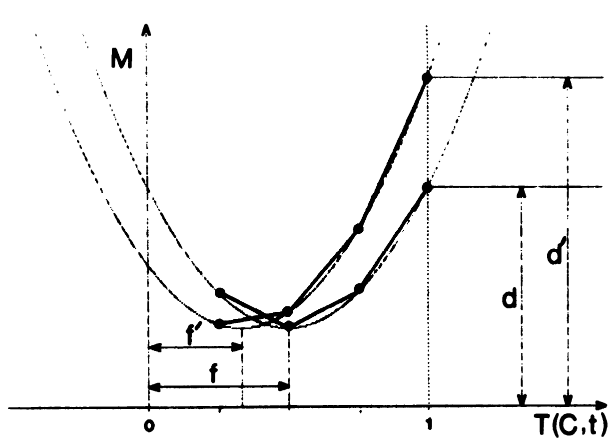

Fig. 3. A set of possible curves of the function M_j for a given set of values of the function f_j. By varying f_j, the value of M_j at $T_j = 1$ is also varied. The interval is normalized to unity, that is $0 \leqslant T_j \leqslant 1$.

So, finally, Eq. (1) becomes

$$D(t_{ij}) = 4m\left(\frac{t_{ij}}{e_j} + \frac{(e_j^{-1})}{12} - \frac{1}{2}\right)^2 + A$$

Eq. 3 satisfies the above conditions. The constants m and A can be thought of as representing *rubato amplitude* and *tempo,* respectively. When comparing the model with actual performances, A is taken to be the average minimum representing a kind of normal tempo to which a performer returns after a boundary slowing. For $e_j = 1$, a peak will have the value 1/4 so Eq. 3 is multiplied through by 4 in order to bring the value to unity. Thus when comparing the model, the constant m represents the real-time difference between the minimum line A and the height of a peak of embedding depth $e_j = 1$.

We are now in a position to see how the model works. Consider again the situation as in Figure 2. In order to make this example more realistic the constants m and A are given values. The uts is taken to be the bar. If the tempo were allegro we would expect the real-time duration of a bar to be about 4000 msec. A typical rubato amplitude might be about 500 msec. Thus we have $A = 4000$ msec and $m = 500$ msec. As in Figure 2 the sets C and E are given by

$$C = (4, 8, 12, 16) \text{ and}$$
$$E = (1, 3, 1, 5)$$

Inputting the constants and the sets C and E into Eq. 3 produces the duration structure shown in Figure 4. The set C determines where the peaks occur and the set E determines the relative heights of the peaks. The duration structure generated in Figure 4 represents an abstract idealization of an actual performance of a piece whose structure is described by the TSR shown.

As mentioned earlier, the prolongation reduction seems also to affect the duration structure. If there is a prolongation boundary dissecting a group, it is possible that the performer will slow at this boundary. The effect is pronounced if the prolongation is a tonic prolongation accompanied by a change in dynamics. At present, however, it is only possible to say this effect may occur and not how likely it is. A more complete model must examine the interaction of TSR and PR in a formal manner. Another factor that seems to affect the duration structure is the tendency of a performer to diminish the real-time duration of a rest from its metrical value, irrespective of its position in the TSR. Thus if a time-span unit contains many rests, its real-time value will be less than the model predicts. This effect can cause havoc in a composition that has rests either side of the bar, since the end of a bar is measured by the onset of notes in the next.

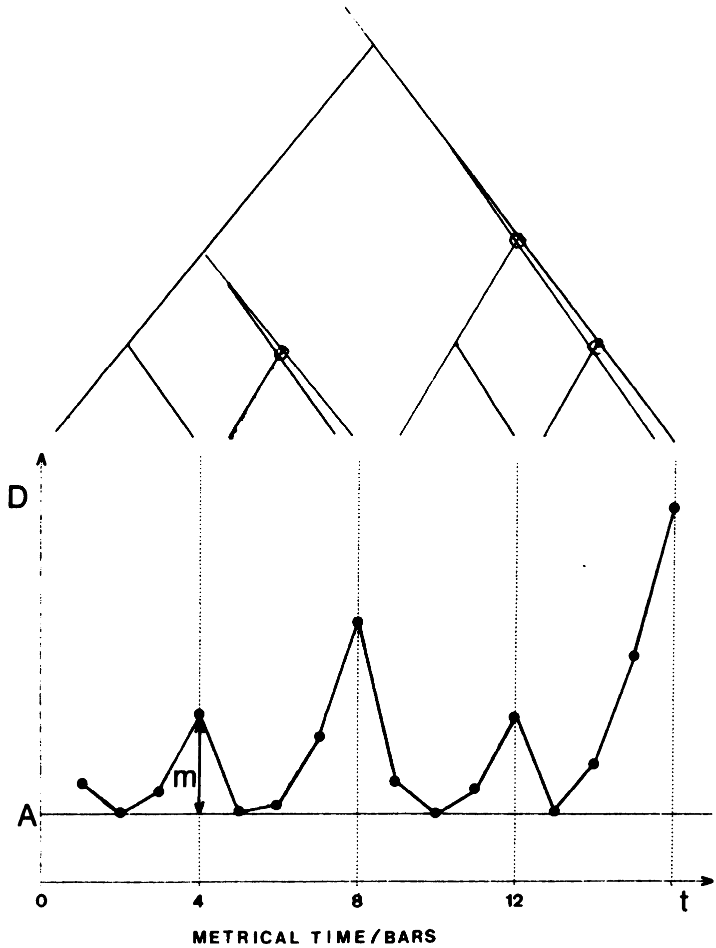

Fig. 4. A hypothetical performance duration structure generated by the model with the given TSR above.

Applications

Presented below are three examples of actual performances compared with the model. The data was obtained on a Bechstein grand piano which has photocells suspended in vertical pairs opposite each hammer in the action. This optical method does not affect the touch or timbre of the piano. A pair of cells detects the moment the key is struck and the moment the key is released. The signals are coded and passed to a minicomputer (PDP12) which assigns clock times and stores the data on tape (Shaffer, 1981).

Two versions are shown for the second example. The first version shows the model without prolongation interaction, the second shows how tonic prolongation within a cadence group can affect the duration surface.

Mozart A Major Sonata K.331. The data for this example are from two performances of the theme with repeats by an undergraduate student. Examination of the score (Appendix I) shows that a suitable uts is the bar, even though the Peters edition, where the GS and MS are slightly out of phase, was used. The TSR shown in Figure 5 indicates that the structural ending set is given by

$$C = (4, 8, 12, 16, 20, 24, 26, 30, 34, 36)$$

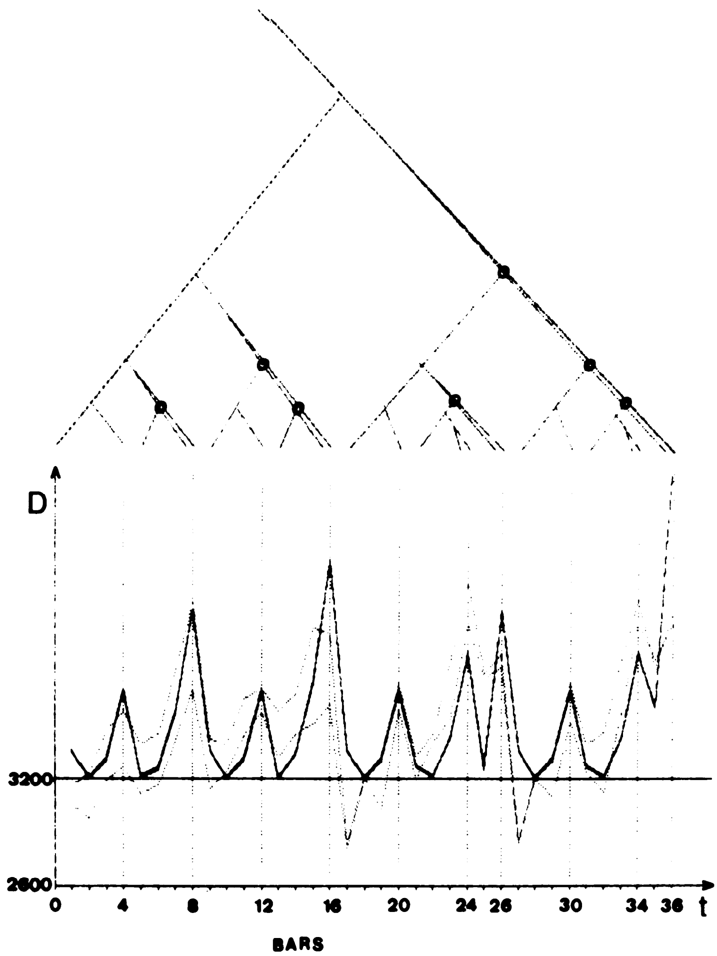

Fig. 5. Two performances of the theme from Mozart's Sonata K.331 compared with the model.

and using the rules for calculating embedding depth, the set E is given by

$$E=(1, 3, 1, 4, 1, 2, 3, 1, 2, 6).$$

The rubato amplitude is $\simeq$450 msec and $A=3200$ msec.

The first thing to note about Figure 5 is that the general features of the duration structure of the performance have been captured by the model. Most of the peaks occur at the structural endings, and the relative heights of the model coincide quite well with the data. There are, however, three ways in which the data depart from the model. First, there is a general slowing from Bars 1–16, 17–26, and 27–36; this is shown in Figure 5 and is the global component (GC) discussed earlier. Second, the performer seems to have endowed the imperfect cadences, at Bars 24 and 34, with greater status in the hierarchy than does the TSR. It would seem then, in this case, that rhythmic closure has more weight than harmonic closure. Third, at Bars 18 and 28, peaks have appeared that were not predicted by the model. Examination of the score shows that there is a tonic prolongation from Bars 17–18 and 27–28. This is an example of prolongation interaction. The fact that only one of the performances showed this effect is interesting because it means that it is an optional expressive device.

Haydn Sonata 59 Adagio. The data for this example are from a concert pianist. Again we take the uts to be the bar. The TSR shown in Figure 6 shows that the sets C and E are given by

$$C=(4, 8, 12, 16, 20, 26, 30, 36, 40, 46, 50, 56)$$
$$E=(1, 2, 1, 4, 1, 2, 1, 4, 1, 2, 1, 7)$$

$m\simeq450$ msec and $A\simeq3300$ msec. Harmonic analysis of this piece (Appendix II) shows that many of the cadenced groups have a return to the tonic within them (of the prevailing tonality, that is). These are indicated in the TSR of Figure 6 by dotted lines, thus indicating where slowing within cadenced groups is possible. In order to accommodate the tonic prolongations within the model, the time-span units containing the prolongation boundary, and the corresponding embedding depths ($e_j=1$), are inserted into the already established sets C and E, although, strictly speaking, the prolongation boundaries are not structural endings. This is, of course, ad hoc and the resultant duration structure, shown in Figure 7, is a curve-fitting exercise. However, it does demonstrate that intrinsically the model is sufficiently general to accommodate a more complex formal input. Nevertheless, the general features are still captured by the model. The global component is essentially a constant here.

The sets for Figure 7 are given by

$$C=(4, 8, 12, 16, 20, 26, (28), 30, (32), 36, 40, 46, (48), 50, (52), 56)$$
$$E=(1, 2, 1, 4, 1, 2, (1), 1, (1), 4, 1, 2, (1), 1, (1), 7)$$

where the numbers in parentheses correspond to the prolongation interaction. It is interesting to note that out of eight tonic prolongations, the per-

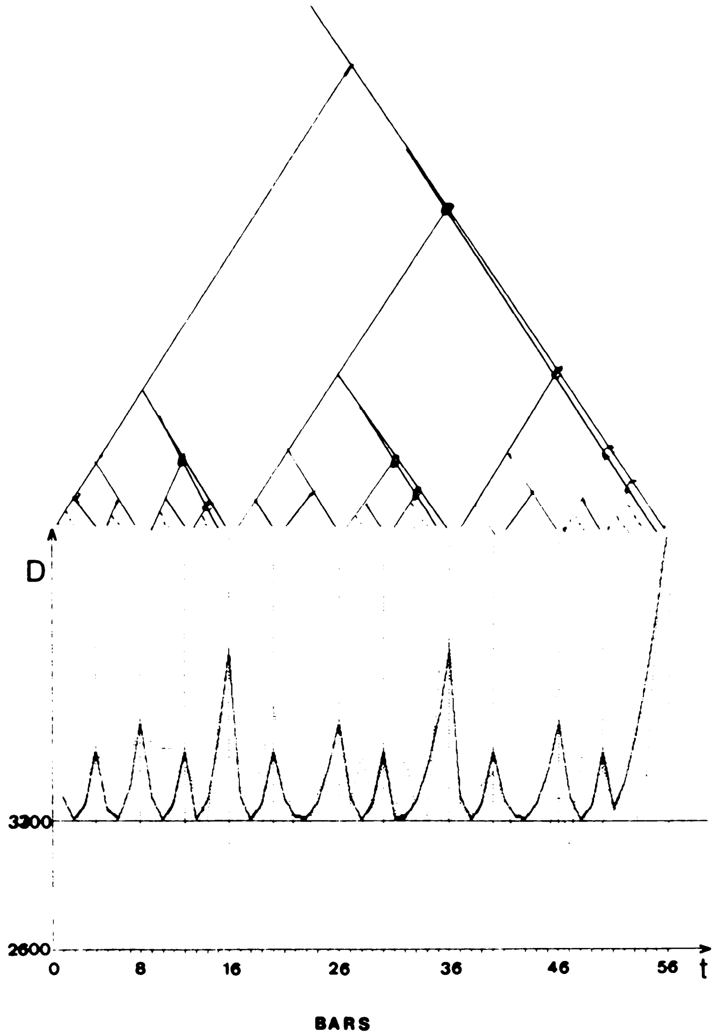

Fig. 6. A performance of the adagio from Haydn's Sonata No. 59 compared with the model.

former chose to emphasize only the last four. Again, at present, it is not really possible to say which prolongation boundary a performer will choose to slow at.

Chopin Trois Nouvelles Etudes No. 3. The data for this last example are also from a concert pianist. Once again the bar is chosen as a uts, even though many of the harmonies are suspended over the bar. A problem associated with this piece is that during the middle section the tonality becomes

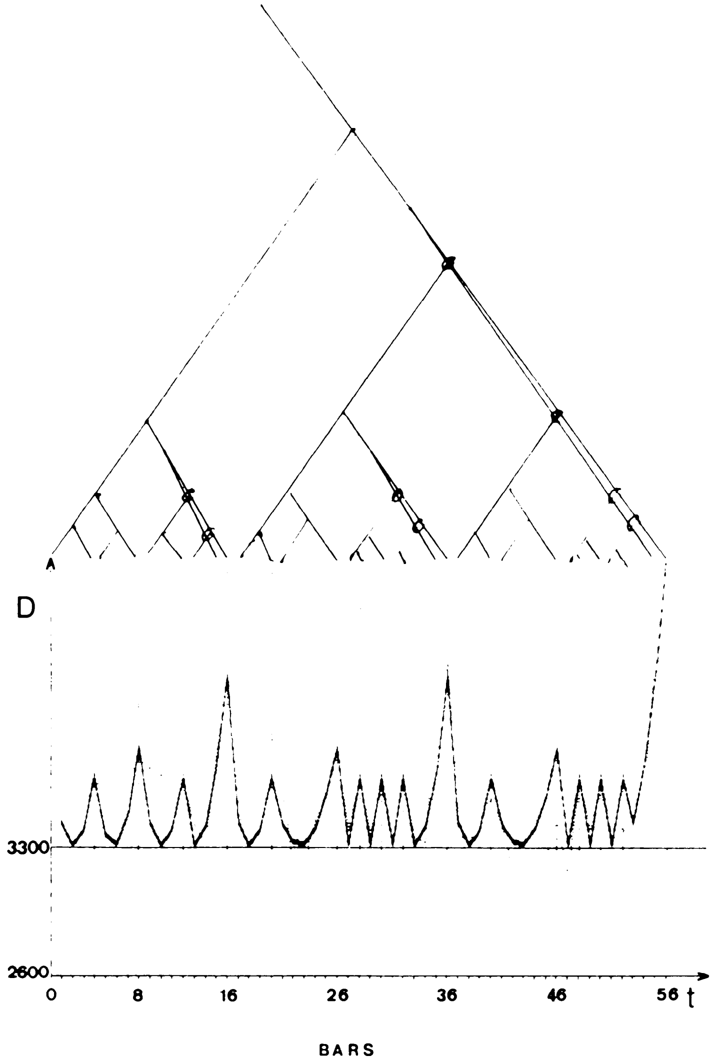

Fig. 7. The same performance of the adagio from Haydn's Sonata No. 59 compared with the model with the prolongation interaction added.

much weakened by frequent modulation (see Appendix III), it thus becomes difficult to choose structural endings. However, since this piece has clearly been phrased in groups of four bars we choose the TSR such that structural endings occur every four bars. Thus the sets C and E are given by

$C = (4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60)$

$E = (1, 2, 1, 3, 1, 2, 1, 2, 1, 4, 1, 2, 1, 4, 6)$

where $m \simeq 650$ msec and $A \simeq 1600$ msec. The resultant model is shown in Figure 8.

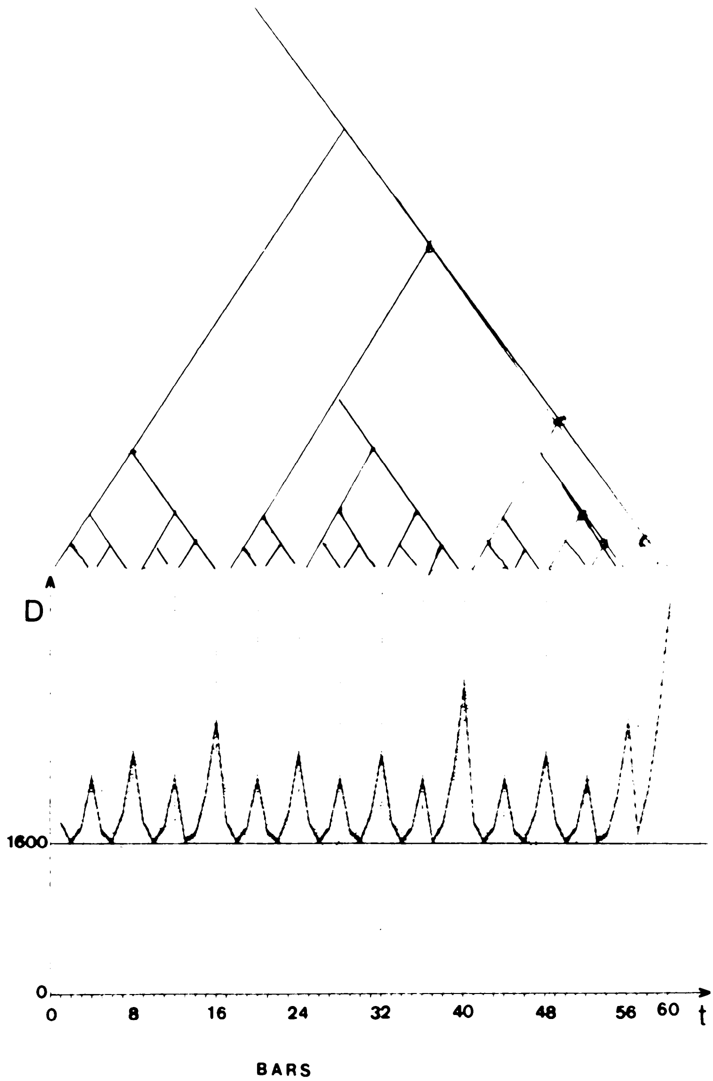

Fig. 8. A performance of the third of Chopin's Trois Nouvelles Etudes compared with the model.

The global component of this piece is fairly constant throughout except for the final 8 bars where there is a ritardando. At Bars 13 and 53, which occupy parallel positions in the piece, peaks have occurred when they were expected at Bars 12 and 52. Examination of the harmonic analysis shows that there is a return to the tonic key here via the dominant, so it seems as if the performer has treated this as a half cadence, thus shifting the structural ending along one. The peak expected at Bar 20 has shifted to Bar 22. This is odd because one would have thought that the half cadence at Bar 20 would have been a more compelling target than the weak tonic prolongation at Bar 22. Finally, the peak at Bar 28 has shifted to Bar 29. This is more understandable since Bars 25–28 act as a prolongation of the dominant of C major, the resolution to C minor occurring at Bar 29.

The important thing to realize about these departures from the expected is that when the tonality is not so well defined, the TSR becomes more ambiguous and the performer has a greater choice of interpretation. Also the hierarchical structuring of the performance will tend to collapse with a weakened tonality because it is tonality that generates points of harmonic stability and hence hierarchical structuring. Thus the peaks from Bars 17–40 are not as high as expected from the TSR, relative to the peaks occurring where the tonality is well defined. This is well illustrated in a performance of Satie's Gnossienne No. 5 (Shaffer, Clarke, & Todd, 1985), where the harmony fluctuates from a major key to its relative minor throughout such that no tonality is ever really established. The duration shows little or no hierarchical structuring as well as having only a small rubato amplitude, about 150 msec about a mean tempo of 2200 msec.

Discussion

It is important that music is organized hierarchically, because it enables the listener to comprehend the complex musical relationships. If it were not so organized all relationships would be local and transient, since the understanding of music places extraordinary demands on the memory of the listener (Meyer, 1973). The hierarchical structures in music are generated by closure—or arrival at relative points of stability—the result of the action and interaction of the various parameters in music. These can act with or against each other, resulting in a potentially ambiguous surface. We have seen that the greater part of the variation of duration in a performance can be accounted for by the rhythmic structure of the music. That is, the performer reflects hierarchical structuring by slowing at points of stability, relative slowing controlled by relative stability. The performer selects a structure from the set of possible structures and elucidates this to the listener. Here we have used the idea of "preferred structure" of Lerdahl and Jackendoff's theory to represent the perception of the performer and generated a corresponding duration structure. It is argued here that hierarchical slowing at points of stability is a kind of parsing device that enables the listener to perceive the hierarchical structure of the music and thus to comprehend the complex musical relationships at the surface.

As stated above, there is an interesting analogy here to the work of Grosjean et al. (1979), Grosjean and Gee (1983), and also Cooper and Paccia-Cooper (1980) in their studies of syntactic to phonetic coding. The main variables used in the study of phonology are duration, intensity, and fundamental frequency which are broadly analogous to those in musical performance. Duration phenomena in speech tend to fall into two main categories, namely, pausing and segmental lengthening. A number of models have been advanced that demonstrate that pausing and segmental lengthening can be accounted for largely by syntactic and prosodic structure.

Grosjean et al. (1979) have developed two alternative algorithms for predicting pausing durations. The data was obtained by asking subjects to read a number of sentences at different rates. The pausing values were averaged and the means were used to make hierarchical representations of the sentences using Johnson's clustering program (1967). For the first algorithm they used the $\overline{X}$-theory of phrase structure (Jackendoff, 1977), although a more traditional theory of phrase structure (Chomsky, 1965) could be used. In the second algorithm the prosodic structure of Liberman and Prince (1977), or its modification by Selkirk (1980), was used. The basic idea of the algorithms is that the pausing structure generally reflects the input structure via some boundary strength metric. There are a number of competing metrics for this. The one used by Grosjean et al. (1979) is the Complexity Index (CI): the numerical value of a boundary between two words is the number of branching nodes dominated by the node dominating the word boundary including the node itself. This is analogous to the Embedding Depth used here.

Cooper and Paccia-Cooper have proposed a general algorithm (1980) to account for segmental lengthening, pausing, and blocking of cross-word conditioning of phonological rules. Like the Grosjean et al. (1979) algorithm, it is based on the surface structure of the whole sentence. It also used a boundary strength metric. However, instead of being a mere reflection of the structural tree as is the CI, this metric contains extra information and is a performance algorithm in itself. Thus the performance structure obtained is often quite different from its input. Cooper also discusses a number of alternative linear metrics.

From this brief survey of these alternative algorithms there emerges a clear analogy to the model proposed here. Although the precise content is different, the general structure of the models is identical. We have an input string from which some hierarchical structure is generated, be it $\overline{X}$-syntax, prosodic structure or TSR. This then acts as an intermediate state from which a performance structure is generated via some metric.

Lerdahl and Jackendoff (1983) have recently made a claim that there is a formal equivalence between the prosodic structure of Liberman and Prince (1977) and their time-span reduction. This is, there is a one-to-one correspondence in the form of their grammars. If this is so it would make the above analogy more striking, since both speech and musical performance use the same empirical principle to map the structural trees into the performance structure, that is phrase-final lengthening.

Conclusion

The model as it stands is only a first approximation. It does not take into account harmonic structure or prolongation reduction in a formal manner, which as we have seen clearly affects the duration structure. The model does not attempt to account for duration fluctuation at the note level, although these local fluctuations must be constrained by the intermediate component. It is thought, however, that note duration is partly due to metrical length (a long note is likely to be construed as a boundary) and the stress patterns at the surface (a note will be lengthened to give it stress). Finally, a complete theory of expression must account for intensity and other secondary expressive variables such as vibrato; and, like the intrinsic variables in the musical structure, these expressive variables are not independent. Clearly more research is needed in this area, not only does it give insight into real-time processing in music, but it is a good testing ground for any theories of musical cognition.[1]

References

Black, J. W., Tosi, O., Singh, S., & Takefuta, Y. A study of pauses in oral reading of one's native language and in English. *Language and Speech,* 1966, *9,* 237–241.

Chomsky, N. *Aspects of the theory of syntax.* Cambridge, MA: MIT Press, 1965.

Chomsky, N. *Language and mind.* Cambridge, MA: MIT Press, 1968/1972.

Clarke, E. *Structure and expression in the rhythm of piano performance.* Ph.D. Thesis, University of Exeter, 1984.

Cone, E. *Musical form and musical performance.* New York: Norton, 1967.

Cooper, W. E. *Syntactic control of timing in speech production.* Ph.D. Thesis, MIT, 1976.

Cooper, W. E., & Paccia-Cooper, J. *Syntax and speech.* Cambridge, MA: Harvard University Press, 1980.

Gabrielsson, A., Bengtsson, I., & Gabrielsson, B. Performance of musical rhythm in $\frac{3}{4}$ and $\frac{6}{8}$ meter. *Scandinavian Journal of Psychology,* 1983, *24,* 193–213.

Goldman-Eisler, F. Pauses, clauses, sentences. *Language and Speech,* 1972, *15,* 103–113.

Grosjean, F., & Collins, M. Breathing, pausing and reading. *Phonetica,* 1979, *36,* 98–114.

Grosjean, I., & Gee, J. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology,* 1983, *15,* 411–458.

1. This research was supported by a grant from the University of Exeter. The pianists were Penelope Blackie, Richard Little, and John McCabe. The data for the performance of the Haydn Sonata was collected by Eric Clarke. I am indebted to Henry Shaffer for his help and advice on this paper.

Grosjean, F., Grosjean, L., & Lane, H. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology,* 1979, ***11***, 58–81.
Jackendoff, R. $\bar{X}$ syntax. *Linguistic Inquiry,* Monograph Two. Cambridge, MA: MIT Press, 1977.
Johnson, S. C. Hierarchical clustering schemes. *Psychometrica,* 1967, ***32***, 241–254.
Lerdahl, F., & Jackendoff, R. Toward a formal theory of tonal music. *Journal of Music Theory,* 1977, ***21***, 111–171.
Lerdahl, F., & Jackendoff, R. *A generative theory of tonal music.* Cambridge, MA: MIT Press, 1983.
Liberman, M., & Prince, A. On stress and linguistic rhythm. *Linguistic Inquiry,* 1977, ***8***, 249–336.
Lindblom, B. Temporal organisation of syllable production. *Speech Transmission Laboratory (Stockholm), Quarterly Progress and Status Report,* 2-1968.
Meyer, L. *Explaining music.* Berkeley: University of California Press, 1973.
O'Malley, M. H., Kloker, D. R., & Dara-Abrams, B. Recovering parentheses from spoken algebraic expressions. *IEEE Transactions on Audio Electroacoustics,* 1973, ***AU-21***, 217–220.
Schenker, H. *Harmony.* Cambridge, MA: MIT Press, 1906/1954.
Schenker, H. *Free composition.* New York: Longman, 1935/1976.
Schoenberg, A. *Structural functions of harmony.* New York: Norton, 1969.
Scott, D. R. Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America,* 1982, ***71***, 996–1007.
Seashore, C. E. *Psychology of music.* New York: McGraw-Hill, 1938.
Selkirk, E. O. The role of prosodic categories in English word stress. *Linguistic Inquiry,* 1980, ***11***, 563–603.
Shaffer, L. H. Performances of Chopin, Bach and Bartok. *Cognitive Psychology,* 1981, ***13***, 327–376.
Shaffer, L. H. Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology,* 1984, ***36***, 577–596.
Shaffer, L. H., Clarke, E., & Todd, N. Meter and rhythm in piano playing. *Cognition,* 1985, ***20***, 61–77.
Stetson, R. A motor theory of rhythm and discrete succession. *Psychological Review,* 1905, ***12***, 250–350.
Todd, N. P. The componential nature of expressive timing in musical performance (in prep.).

Appendix I. Score of the Theme from Mozart's Sonata K.331.

Appendix II. Score and Harmonic Analysis of the Adagio from Haydn's Sonata No. 59.

Haydn Sonata No. 59
Harmonic Analysis

B♭:	1 $\vert$I	$VII^6/I \vert I$	$\vert I^6$	$\vert I^6_4$ V		$\vert$		
	5 $\vert$I	$VII^6/I \vert I$	$\vert V^7$ of V	$\vert$V of V/V		$V^7 \vert$		
	9 $\vert$I	$VII^6/I \vert I$	$\vert I^6$	$\vert I^6_4$	V	$\vert$		
	13 $\vert$I	$VII^6/I \vert I$	$\vert V^7$ of V	$\vert$V of V/V		V $\vert$		
	21 $\vert V^7$	$\vert I^6_4$	$\vert$VII I	IV $\vert$I	V	$\vert$		
	25 $\vert I^6$	$\vert II^6$	$\vert V^{0}_{\substack{6\\5}}$ of IV	$\vert$IV		$\vert V^{0}_{\substack{6\\5}}$ of V	$\vert I^6_4$	V $\vert$
	31 $\vert$I	$VII^6/I \vert I$	$\vert I^6$ IV^6	$\vert I^6_4$	V	$\vert$		
	35 $\vert$I IV^6	$VII^6/I \vert I$	V^7 of IV $\vert IV^6$	$\vert$I		$\vert II^6$ III^6	$V^7 \vert V^7/I$	I^7 $\vert$
	41 $\vert V^7$	$\vert I^6_4$	$\vert V^6_5$ I^7	$II^6 \vert I$	V	$\vert$		
	45 $\vert I^6$	$\vert II^6$	$\vert V^{0}_{\substack{6\\5}}$ of IV	$\vert$IV		$\vert V^{0}_{\substack{6\\5}}$ of V	$\vert$V	$\vert$
	51 $\vert$I	$VII^6/I \vert I$	$\vert I^6$ IV	$II^6 \vert I^6_4$	V	$\vert$		
	55 $\vert V^7$ of IV	$\vert$I	V^7 of IV $\vert IV^6$	$\vert I^6$	I	$\vert II^6$ III^6	$V^7 \vert V/I$	I^7 $\vert$

Appendix III. Score and Harmonic Analysis of Chopin's Trois Nouvelles Etudes No. 3.

Chopin Trois Nouvelles Etudes No. 3 Harmonic Analysis

A♭ :	1	I	V^7	I	V^7	I	V^7	I	D♭ : V^7
	5	I	V^7	I	V^7	I	V^7	I	A♭ : II^7
	9	V^{7} (−9)		I		f : V^{7} (−9)		I	VI^7
A♭ :	13	V^7		I	IV^7	II	V	I	
E :	17	I	V^4_3	I^6	I^6_4	V		V	
C :	21	I	V^4_3	I^6	I^6_4	V		V	
	25	VI	A : $V^{5\sharp}$	I	B♭ : $V^{5\sharp}$	I	B : $V^{5\sharp}$	I	C : $V^{5\sharp}$
C :	29	I	b♭ IV	I	a : IV	I^6_4	V^7	I	
f :	33	I	C : IV^6	I	b♭ : IV	I^6_4	V^7	I	
f :	37	V^7		I^6_4	VI	VI^2	II	II	A♭ : V^4_3
	41	I	V^7	I	V^7	I	V^7	I	D♭ : V^7
	45	I	V^7	I	V^7	I	V^7	I	A♭ : II^7
	49	V^7		I		f : V^{7} (−9)		I	VI^7
A♭ :	53	V^9		I	IV^7	II I^6_4	V^{13}_{7}	III^6	I^6
	57	I		I	V^7	V^7		I	

It should be pointed out that although changes of tonal center have been indicated during Bars 25–34, strictly speaking they are not modulations. The notation has been used simply to illustrate the sequential nature of this passage.