The dynamics of dynamics: A model of musical expression

Neil P. McAngus Todd
Department of Psychology, University of Exeter, Exeter EX4 4QG, United Kingdom

(Received 7 December 1990; revised 27 January 1992; accepted 2 March 1992)

A computational model of musical dynamics is proposed that complements an earlier model of expressive timing. The model, implemented in the artificial intelligence language LISP, is based on the observation that a musical phrase is often indicated by a *crescendo/decrescendo* shape. The functional form of this shape is derived by making two main assumptions. First, that musical dynamics and tempo are coupled, that is, "the faster the louder, the slower the softer." This tempo/dynamics coupling, it is suggested, may be a characteristic of some classical and romantic styles perhaps exemplified by performances of Chopin. Second, that the tempo change is governed by analogy to physical movement. The allusion of musical expression to physical motion is further extended by the introduction of the concepts of energy and mass. The utility of the model, in addition to giving an insight into the nature of musical expression, is that it provides a basis for a method of performance style analysis.

PACS numbers: 43.75.St

Hence, again, it becomes possible for motion in music to imitate the peculiar characteristics of motive forces in space, that is, to form an image of the various impulses and forces which lie at the root of motion. And on this, as I believe, essentially depends the power of music to picture emotion. [Helmholtz, 1863]

INTRODUCTION

Although our understanding of musical expression has made some advances since the publications of Seashore and co-workers (Seashore, 1938) we are still essentially in a state of ignorance. What is meant by this is that despite the work that has been done in the last decade or so (Shaffer, 1981; Sloboda, 1983; Sundberg, 1988; Clarke, 1988; Gabrielsson, 1987; Todd, 1989c) there still exists the vast corpus of recorded performances from the earliest pianola rolls to the latest CD that we have barely touched in terms of analysis [a few exceptions apart (Repp, 1990)]. Even with the progress that has been made in instrument technology, both in the form of hardware and software, the number of analyses of performances by skilled musicians using direct measurement is very small. With such a small empirical base we are not in any position to answer many basic questions such as the effects of style, individual differences, instrumentation, etc. The goal of this paper therefore, is first to develop a working model of musical expression and second to demonstrate in principle how this model could provide the basis for a method of performance analysis.

I. DEFINITIONS OF BASIC TERMS

Before proceeding further it is useful to define some basic terms. In this paper we will be considering two main variables namely, *tempo*, which we denote by v, and *dynamic*, which we denote by I.

A. Tempo

In research on expressive timing, tempo, as such, is a fictitious variable since it cannot be measured directly. What is in fact measured is the onset time t_i of an event such as a note or chord in a series of events. In the case of metrical music it is possible to assign a number to the event according to its position in the metrical grid. This number we refer to here as *metrical distance* x_i. Thus, for any sequence of events the basic empirical relation is a series of pairs $\{(t,x)_i | i = 1 \cdots L\}$. In the simplifying case that $\{x_{i+1} - x_i = 1 | i = 1 \cdots L\}$ we may drop the subscript i since $x \equiv i$ so that we can write the series as $\{t_x | x = 1 \cdots L\}$.

If we let $\Delta t_x = t_{x_2} - t_{x_1}$ and let $\Delta x = x_2 - x_1$ then by the *forward difference* method the tempo can be estimated by

$$v_x \doteq \frac{\Delta x}{\Delta t_x} = \frac{x_2 - x_1}{t_{x_2} - t_{x_1}}, \quad (1)$$

which reduces to $v_x \doteq 1/\Delta t_x$ if $\Delta x = 1$ (cf. Sundberg and Verillo, 1980).

B. Intensity

The piano system (Shaffer, 1981) used to obtain the data shown here not only measures the onset times of individual notes, but also the hammer flight time T for each keypress. The simplest estimate of note intensity, which is adopted here, is to assume $I \propto 1/T$. For a grouping of n notes the mean intensity is taken to be representative, i.e., $I \propto \Sigma_{i=1}^{n} 1/nT_i$.

II. TEMPO AND DYNAMICS

The model of musical expression presented here is an extension of an earlier model of expressive tempo variation (Todd, 1985, 1989a, 1989b, 1989c). These earlier models of

 0001-4966/92/063540-11$00.80

tempo were based on the idea that musical phrasing has its origin in the kinematic and dynamic variations involved in single motor actions (Stetson, 1905; Sachs, 1943). Typically, such actions have a characteristic velocity/force profile that involves an acceleration/deceleration in velocity and a corresponding rise and fall in tension. That a phrase is often marked by an *accelerando/ritardando* shape has now been well established (Seashore, 1938; Todd, 1985; Shaffer and Todd, 1987; Repp, 1990). A corresponding marking by a crescendo/decrescendo shape, which the motor hypothesis would suggest, also appears to have some empirical support.

In C. E. Seashore's *The Psychology of Music* (Seashore, 1938) an intensive study by H. Seashore was reported that involved the use of eight singers. He listed a comprehensive "inventory of factors in rhythmic expression in singing," distinguishing in each case between "composition" and "performance" factors. Amongst these many observations it was noted that associated with a phrase was "a tendency of tonal power to rise to a peak and then fall away."

Gabrielsson (1987) carried out a study in which the dynamics and *rubati* from performances of the theme from the Mozart A-Major Sonata K331 were compared (see Fig. 1). The use of a *crescendo/decrescendo* to mark a phrase is quite clearly observed by Gabrielsson

> Considered as a whole, the amplitude profile within each phrase shows an increase toward a maximum at, or close to, the transition from the next last to the last measure and then falls steeply. The termination of each phrase is thus associated with diminishing amplitude. [p. 98]

Gabrielsson further notes that the degree of *crescendo* at the beginning of the phrase is often a function of the structural importance of the phrase.

> In most cases there is a *crescendo* before the maximum. The maximum in the second phrase is louder than in the first phrase, and on the whole the dynamic range is larger in the second phrase than the first.

Consider now the data presented in Fig. 2, obtained from the Bechstein piano at Exeter (Shaffer, 1981), from which we may make a number of observations. First, both beat intensity (I_x), obtained by taking the average note intensity in a beat, and tempo (v_x), obtained by taking inverse

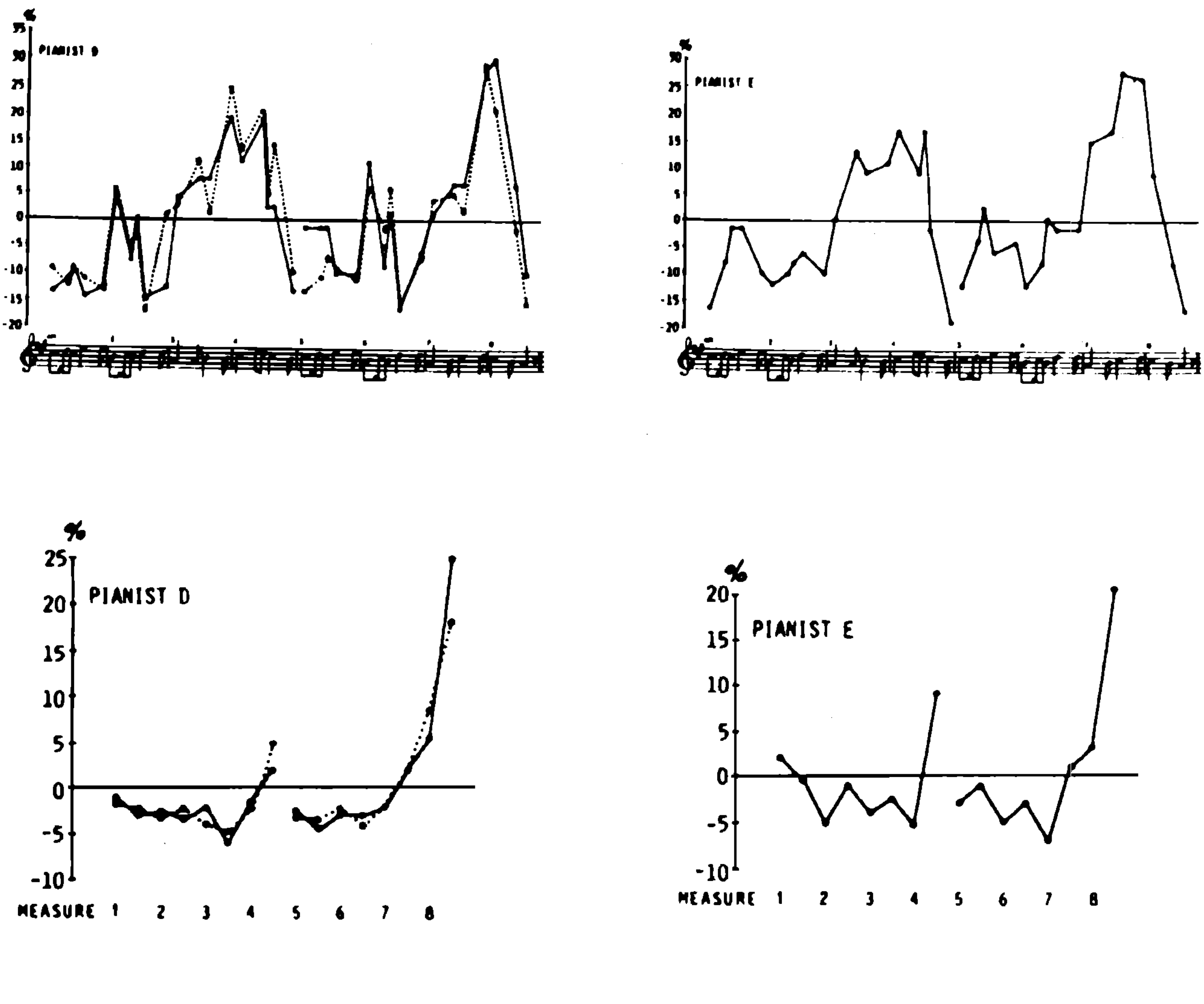

FIG. 1. Note amplitudes (upper) and measure durations (lower) from performances of the theme from Mozart's Sonata K331 by pianists D and E. [Adapted from A. Gabrielsson, in *Action and Perception in Rhythm and Music*, edited by A. Gabrielsson (Royal Swedish Academy of Music, Stockholm)].

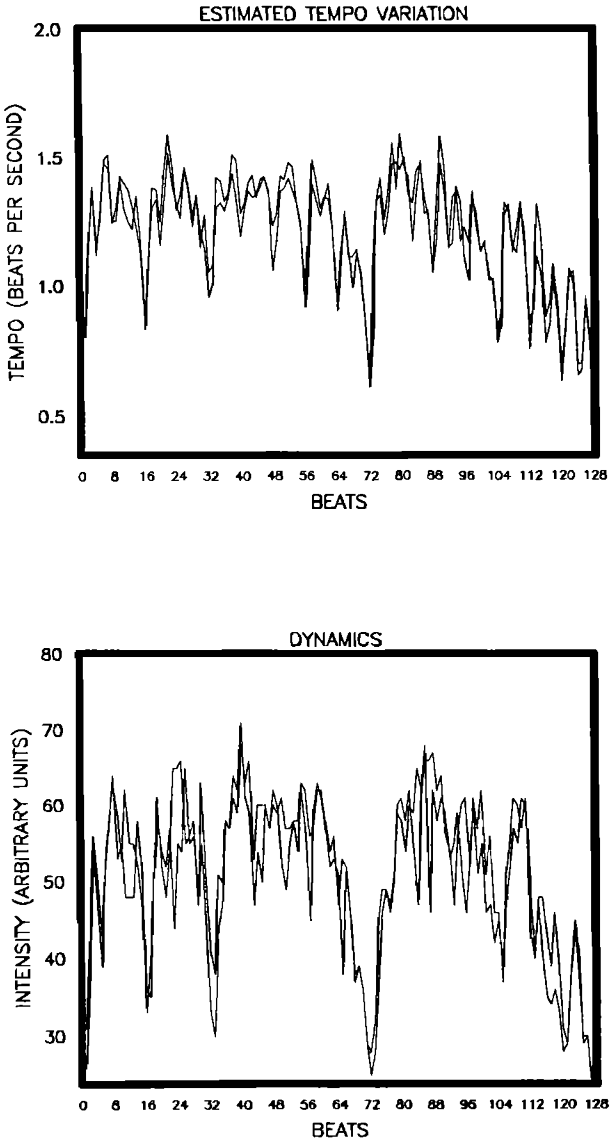

FIG. 2. Estimated tempo (top) and dynamics (bottom) from two performances of the prelude in F-sharp minor, Chopin.

beat duration [see Eq. (1)], show a high degree of reproducibility from one performance to the next ($r_{I_1,I_2} = 0.845$, $r_{v_1,v_2} = 0.973$). Whilst the ability of skilled performers to reproduce timing has been known since Seashore (1938), the corresponding reproducibility of dynamics has been rather less commented upon. Second, there is a high correlation between I_x and v_x ($r_1 = 0.689$, $r_2 = 0.728$), i.e., there appears to be a tendency to "the faster the louder the slower the softer." Third, often there is no direct relationship between dynamic markings in the score and actual performance. For example, although a *diminuendo* is indicated at beat 96 in the score the performer actually makes a small crescendo in one performance. This suggests that the expression marks in a score are used only as a rough guide by performers. Fourth, the shape of both tempo and musical dynamics seems, as in the Gabrielsson data, to be a function of structural importance, i.e., the more important the boundary the greater the *decrescendo/ritardando*.

On the basis of the above observations then we make the following proposition.

Proposition: (a) a group is phrased by a *crescendo/decrescendo* shape; (b) the particular shape is a function of structural importance; (c) musical dynamics is coupled to tempo.

In putting forward this proposition we are not suggesting that this is a hard and fast rule for musical dynamics in performance. Indeed, there are numerous cases where the dynamic increases at the end of a section of phrase. What is being suggested, however, is that the style of musical dynamics embodied in the proposition may be a kind of normative default mode of performance that a performer will adopt in the absence of any alternative instructions in the score.

III. A MODEL OF MUSICAL DYNAMICS

A. Framework and assumptions

The question now arises of how can we implement the proposition as a working model which we can compare with performance data? In Todd (1989c) a theoretical framework, called an *abstract expression system* (AES), was proposed that specifies that an expressive device has a particular set of objects that have a certain logical relationship to each other. An abstract expression system works on the same level of explanation as do *machines* and *formal languages* in computer science. That is, they describe a class of objects that have certain properties.

The utility of an AES is at least twofold. First, it lays down the basis for the construction of models of expression in that it defines what kind of objects are needed. So, rather than start from scratch, the prospective modeler has some idea what to look for. Whilst an AES is not a complete specification for the modeling of any particular device, since knowledge of the particular computational theory [in the sense of Marr (1982)] is required, the device's component parts will have a well-defined logical relationship irrespective of its computational theory. The second, but related, utility of an AES is that it provides a language with which to describe the various objects encountered in musical expression. So, in short, an abstract expression system provides a way of looking at and talking about expression.

The concept of an AES was developed by generalizing from the series of models of tempo the idea of structural to serial mappings and the recovery of structure from the series. The tempo models have in common at least two things with other computational systems for the production of expressively modulated output (Clynes, 1987; Sundberg, 1988; Longuet-Higgins and Lisle, 1989). First, they share the same overall form that is **score input → internal process → expressive output**. Second, to varying degrees of explicitness, they utilize two kinds of object: (1) some aspect of structure; (2) functions for encoding structure into serial output. Some systems have the further facility for taking data from performance measurements and attempting to reconstruct a representation (Longuet-Higgins and Lisle, 1989; Todd, 1989b). More formally, we may say that a model of expression requires the following components:

(a) a *representation of structure*, usually a tree or vector space;

(b) a *performance procedure*;

(c) an *encoding function*;

(d) a main independent variable, either *metrical distance* x or *time* t;

(e) a set of *structure variables*, which describe the structure; and

(f) a set of *style parameters*.

In order to build a particular model then two steps are required (Todd, 1990). First, we need to specify the computational theory of the device. This will involve choosing a representation of the structure, an encoding function, and parameters. Second, the computational theory must be implemented as an algorithm. The output from the algorithm can then be evaluated by either producing numerical output that can be compared to performance measurements or by producing sound that can be listened to.

In the Todd (1989a) model the representation that formed the input was taken to be a grouping structure. Proposition (b) suggests that grouping should also form the input to the musical dynamics model. In the previous models it was found that grouping requires two variables for its description—*group length L* and *boundary strength S*—which we may also carry over.

The previous encoding function for duration was a parabola. There are, however, at least two reasons why we should reject this function. First, to obtain a function for musical dynamics from this would involve a quite complicated transformation. Second, there are empirical grounds for believing that there is a better alternative anyway, which is that tempo change is linear as a function of real time [Eq. (2b)]. I will discuss this further in the next section. The style parameters that form arguments to the function will depend on what the encoding function is and the way in which structure is coupled to encoding function.

Finally, we may also assume that the performance procedure, previously a recursive algorithm involving procedures of look-ahead and planning (Todd, 1989a), can be carried over. This procedure will of course have to be modified to include the new variable of intensity and the new encoding function, but the main algorithm will be the same.

B. Linear tempo (LT)

As mentioned above there is evidence to suggest that expressive tempo variation is linear in time. According to LT the variation of tempo is governed in a manner analogous to velocity in the equations of elementary mechanics which are the following:

$$a(t) = a, \tag{2a}$$

$$v(t) = u + at, \tag{2b}$$

$$x(t) = ut + 1/2at^2, \tag{2c}$$

$$a(x) = a, \tag{3a}$$

$$v(x) = (u^2 + 2ax)^{1/2}, \tag{3b}$$

$$t(x) = (1/a)[(u^2 + 2ax)^{1/2} - u], \tag{3c}$$

where a is *acceleration*, u is *initial tempo*, v is *tempo*, x is *metrical distance* (measured in units of beats or bars) and t is *time*. The duration between two events at x_2 and x_1 can be obtained by

$$\Delta t = \int_{x_1}^{x_2} \frac{1}{v(x)}\, dx \tag{4a}$$

for the continuous case, which can be estimated by

$$\Delta t_x \doteq \Delta x / v(x), \tag{4b}$$

which gives rise to an error approximately $-(\Delta x)^3 a/12$.

Kronman and Sundberg (1987) attempted to apply the LT system to the final *ritardandi* from performances of 24 pieces, mostly by J. S. Bach. In this study regression analyses were carried out on the 24 final *ritardandi* using the square-root function for velocity as a function of metrical distance [Eq. (3b)]. They found that this gave a reasonable approximation to the data. This analysis, however, is inadequate for the modeling of tempo throughout a whole performance. The main reason is that a performance consists of a whole series of *accelerandi* and *ritardandi*. So, it is not clear how one should model the *accelerandi* or how the *accelerandi* should be connected to the *ritardandos*. Also there is no indication of how the acceleration should vary from one phrase to the next. Further, given that it is clear that the tempo in many styles is composed of the superposition of a number of timing components, the single function is incomplete. Finally, the simple assumption that tempo is equal to inverse duration [Eq. (1)] is problematic (Todd, submitted).

In recent work on the synthesis of performances Longuet-Higgins and Lisle (1989) also concluded that tempo should be linear in time within a single *accelerando* or *ritardando* [Eq. (2b)] since it produces the most naturally sounding tempo. They argued (private communication) that the best means of connecting the series of *accelerandi* and *ritardandi* was to make them piecewise continuous. So that if one plotted the series a kind of sawtooth pattern would emerge.

In a recent study (Todd, submitted) the *accelerandi* and *ritardandi* from complete piano performances were examined using regression analysis. The main conclusion was that the best account of metrical distance as a function of time was given by a series of linear, second-order polynomials [see Eq. (2c)].

C. Energy, tempo, and intensity

The above studies then provide convincing evidence that the equations of elementary mechanics can be used to model an expressive change of tempo. Rather than taking a purely *kinematic* approach, however, it is useful also to consider the *dynamics* of the situation. That is to consider the sequence of forces which go with any given time sequence of motions.

In order to build a new encoding function on this basis then, let us consider the jth group in a piece to which corresponds a stereotypical *accelerando/ritardando* shape. Associated with this shape we may further imagine the movement of a particle of mass m in a V-shaped potential well of length L_j (see Fig. 3). This corresponds to an attractive force directed towards the center of the well so that in the region $0 \leqslant X_j \leqslant \delta_j$ the particle accelerates and in the region $\delta_j \leqslant X_j \leqslant 1$

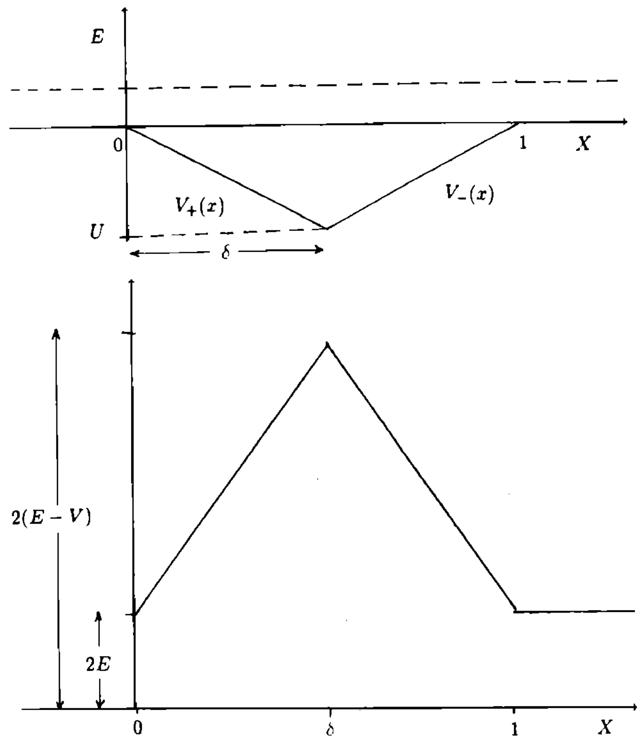

FIG. 3. A V-shaped potential well of normalized length 1 (top) and the corresponding velocity/tempo/intensity profile (bottom).

the particle decelerates where $X_j = x/L_j$ is normalized distance. Outside the well the particle moves with constant velocity.

Let us assume also that the total energy E of the system is constant and is given by

$$E = T + V, \tag{5}$$

where T is the kinetic energy and is given by

$$T = \tfrac{1}{2}mv^2 \tag{6}$$

and V_j is the potential energy given by

$$V_{+j}(x) = U(X_j/\delta_j), \quad 0 \leqslant X_j \leqslant \delta_j, \tag{7a}$$

$$V_{-j}(x) = U\frac{(1-X_j)}{(1-\delta_j)}, \quad \delta_j \leqslant X_j \leqslant 1, \tag{7b}$$

where U is depth of potential.

From (5) and (6) the velocity (tempo) in the well is given by

$$v_{\pm}(x) = \sqrt{(2/m)(E - V_{\pm})}. \tag{8}$$

Proposition (c) suggests that intensity and velocity (tempo) are coupled. There are many physical systems in which intensity is proportional to the square of velocity such as a hammer/string interaction or a wind/surface interaction. If we assume this relationship then

$$I_x = Kv_x^2, \tag{9}$$

which in terms of energy gives

$$I_{\pm} = (2K/m)(E - V_{\pm}). \tag{10}$$

This linear relationship is of course an oversimplification and examination of Figs. 1 and 2 show that the musical dynamics/tempo coupling is much more subtle than suggested by (9). However, for the sake of a first model we shall adopt this simplification.

D. Strategies for coupling the structure variable *S*

We now come to the question of how to couple the structure variable S to the potential function. The function given in Eqs. (7) contains three variables δ, U and X each of which may be made to be a linear function of S such that

$$\delta_j = \delta_0 + c_\delta S_j, \tag{11a}$$

$$U_j = U_0 + c_U S_j, \tag{11b}$$

$$X_j = X_{0j} + c_X S_j, \tag{11c}$$

where $X_{0j} = x/L_j$. The three transforms given above may be applied in any combination but the simplest combinations are the following:

(a) $c_\delta = c_U = c_X = 0$, no coupling;
(b) $c_\delta = c_X = 0$, potential depth shift;
(c) $c_U = c_X = 0$, offset shift;
(d) $c_\delta = c_U = 0$, coordinate shift.

With strategy (a) the structure variable S is not coupled in, so that the only variable which affects the expression is phrase length L. Strategy (d), the coordinate shift, was applied in the earlier models of tempo (Todd, 1985, 1989a). With musical dynamics, however, at least two different strategies appear to be used. Strategy (b), in which the position of the maxima of intensity is fixed but the magnitude of crescendo is varied, appears to be used in the Mozart data of Gabrielsson. Strategy (c), which is to vary the position of the maxima, keep the height of the maxima constant but adjust the slope of the *crescendo/decrescendo* so that the connected segments remain piecewise continuous, appears to be used in the Chopin data.

E. The encoding function with *N* components

In most performance data there are usually a number of components, from global variation over the whole piece to local fluctuations at the note level. These components are superimposed onto each other (Todd, 1989b, 1989c). Thus the complete function that generates the intensity series is given by the following:

$$I_x = \sum_{l=1}^{N} \frac{2K}{m_l}(E - V_{j_l}), \tag{12}$$

where the subscript l refers to *structural level*. For each component the potential V_{j_l} is such that

$$V_{+j_l} = U_{j_l}\frac{X_{j_l}}{\delta_{j_l}}, \quad 0 \leqslant X_{0j_l} \leqslant \delta_{j_l}, \tag{13a}$$

$$V_{-j_l} = U_{j_l}\frac{(1-X_{j_l})}{(1-\delta_{j_l})}, \quad \delta_{j_l} \leqslant X_{0j_l} \leqslant 1, \tag{13b}$$

where

$$\delta_{j_l} = \delta_0 + c_\delta S_{j_l}, \tag{14a}$$

$$U_{j_l} = U_0 + c_U S_{j_l}, \tag{14b}$$

$$X_{j_l} = X_{0j_l} + c_X S_{j_l}, \tag{14c}$$

and normalization requires that

$$\sum_{l=1}^{N} \frac{1}{m_l} = 1. \tag{15}$$

IV. EXAMPLES OF MODEL OUTPUT

In order to see the model working more clearly let us consider a specific example.

A. Input structure

The simplest three-component structure is a forest of binary trees in which each member tree has only two branches (see Fig. 4).

In LISP this structure can be represented by

```
(setq tree '(A B))
(setq A '(a a))
(setq B '(b b))
(setq a '(g4 g4))
(setq b '(g4 g4))
(setq g4 '(1 1 1 1)).
```

The values of the structure variables L and S corresponding to this structure are given by

$(L,S)_{j_1} = (16,0),(16,1),$

$(L,S)_{j_2} = (8,0),(8,1),(8,0),(8,1),$

$(L,S)_{j_3} = (4,0),(4,1),(4,0),(4,1),(4,0),(4,1),(4,0),(4,1).$

B. Parameter values

To scale E and U for a particular data set we assume initially that $m_l = 1$ so that

$$E = I_{\min}/2K, \tag{16a}$$

$$U_0 = (I_{\min} - I_{\max})/2K, \tag{16b}$$

$$K = I_{\max}/v_{\max}^2, \tag{16c}$$

where $I_{\min}$ and $I_{\max}$ are the minimum and maximum intensities over the whole piece and $v_{\max}$ is the maximum tempo. (Obviously, there is a degree of arbitraryness in this scaling method. For example, an alternative way of finding a value for U_0 is to use the interquartile range.) For the sake of the example we let the range of intensities and maximum tempo be $I_{\min} = 0$, $I_{\max} = 100$, $v_{\max} = 1$ which implies that $E = 0$, $U_0 = -\frac{1}{2}$, $K = 100$. The model also restricts the number of degrees of freedom by the following equalities:

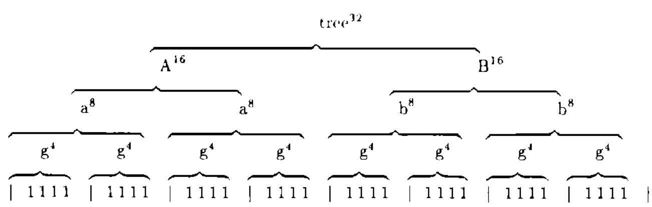

FIG. 4. A simple three-layer tree structure for the example.

$$U_{0_1} = U_{0_2} = U_{0_3} = U_0,$$

$$\delta_{0_1} = \delta_{0_2} = \delta_{0_3} = \delta_0,$$

$$c_{U_1} = c_{U_2} = c_{U_3} = c_U,$$

$$c_{\delta_1} = c_{\delta_2} = c_{\delta_3} = c_\delta,$$

$$c_{X_1} = c_{X_2} = c_{X_3} = c_X.$$

C. Model output for different coupling strategies

In order to demonstrate the range of different outputs, the same structure is input for the first three coupling strategies with the following parameter values:

(a) $c_U = 0$; $c_\delta = 0$,

(b) $c_U = -0.5$; $c_\delta = 0$,

(c) $c_U = 0$; $c_\delta = 0.1$,

and for each of the three strategies, three values of δ_0 (0.3,0.5,0.7) and three different component weightings with ratios [(6,3,2);(1,1,1);(2,3,6)] are input. Thus, for each strategy there are nine example outputs shown in Figs. 5–7.

V. ANALYTIC METHOD

A. Method

Having built a basic model for musical dynamics we are now in a position to produce an analysis of performance data. The method adopted here is essentially a kind of analysis/synthesis but using performance data as the starting point (Todd, 1989c) (see Fig. 8). The basic idea is that from the data a guess of the structure and parameters is made. These can then be fed into the performance algorithm to produce a simulation of the original data. The simulation can then be compared to the original data using regression analysis (Draper and Smith, 1981). This cycle is repeated until the variance accounted for by the regression reaches a criterion of acceptabililty, which in this case is the variance accounted for by a repeat performance. Obviously, some intuition is required in guessing a structure and initial parameter values.

The products of the analysis are the following:

(a) A structure and corresponding values of the structure variables $\{(L,S)_{jl} | l = 1\cdots3\}$;

(b) The style parameter estimates E, (U_0,c_U), (δ_0,c_δ), and c_X and component masses m_l, which are obtained by regressing

$$I_x = \beta_0 + \sum_{l=1}^{3} \beta_l I_l, \tag{17}$$

where $m_l = 1/\beta_l$ and $I_l = 2K(E - V_l)$;

(c) R^2, the variance accounted for by the regression.

B. Analysis of Chopin prelude

When the above model and method using strategy (c) was applied to the first performance from the Chopin Prelude the structure given in Fig. 9 was obtained. The style parameter estimates were $K = 31$, $E = 0.40$, $U_0 = -0.73$, $\delta_0 = 0.58$, $c_\delta = -1/3.8$, $c_U = 0$, $c_X = 0$, whilst regression gave the following:

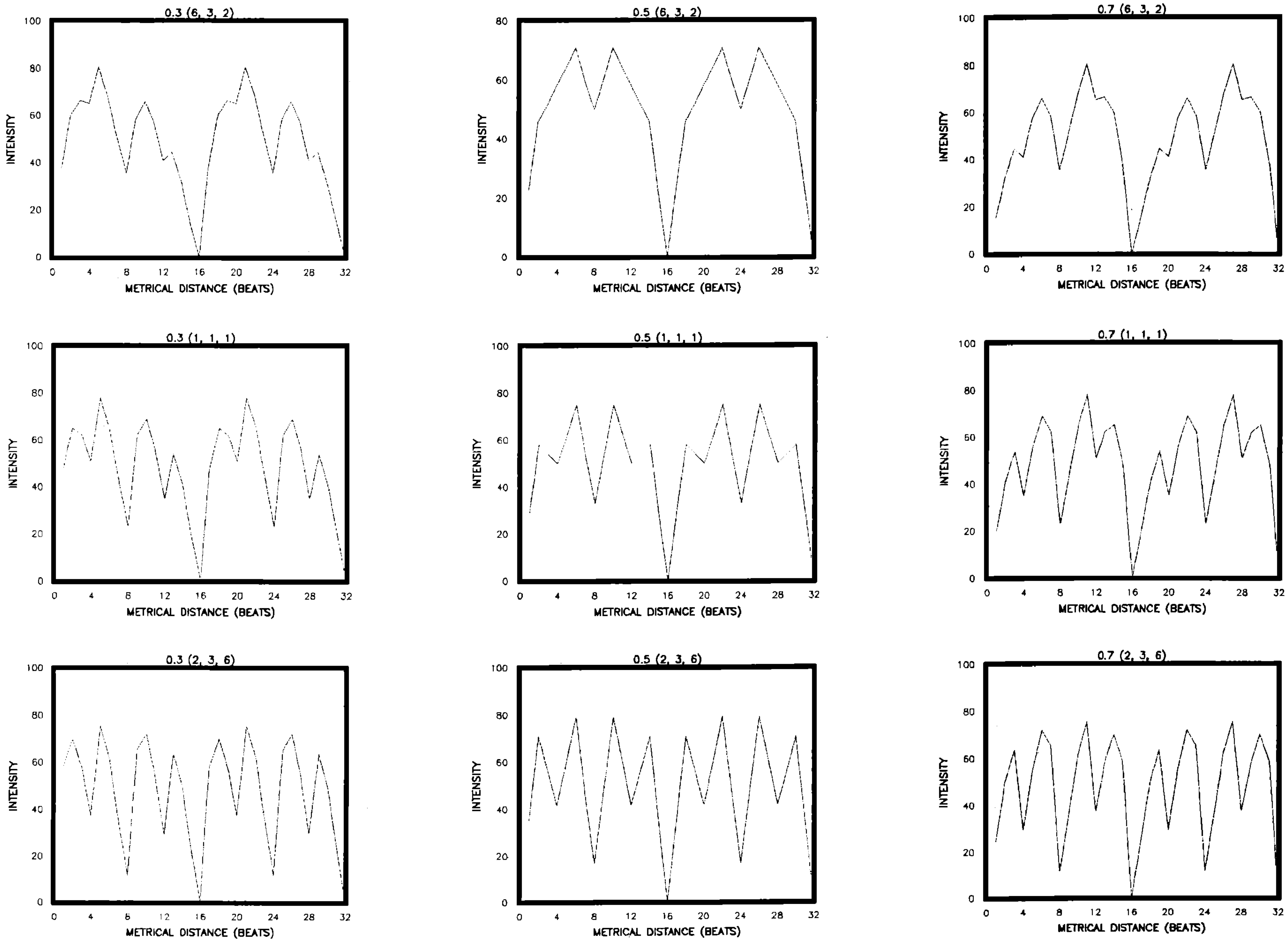

FIG. 5. Example outputs from the model with coupling strategy (a); component weightings ratio of (6,3,2) (top row), (1,1,1) (middle row) and (2,3,6) (bottom row); and with offset values δ_0 of 0.3 (first column), 0.5 (middle column), and 0.7 (third column).

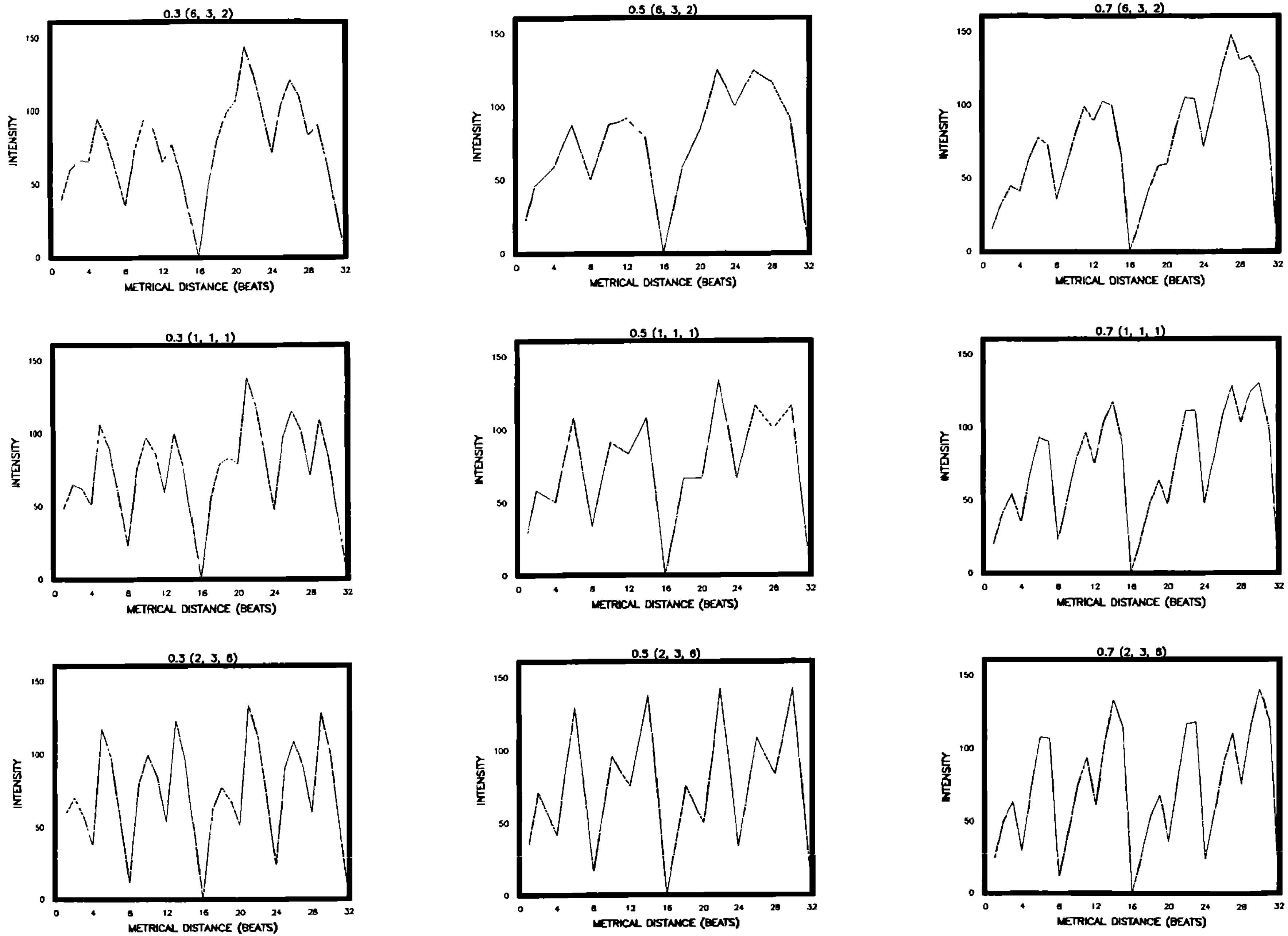

FIG. 6. Example outputs from the model with coupling strategy (b) with parameter values as in Fig. 5.

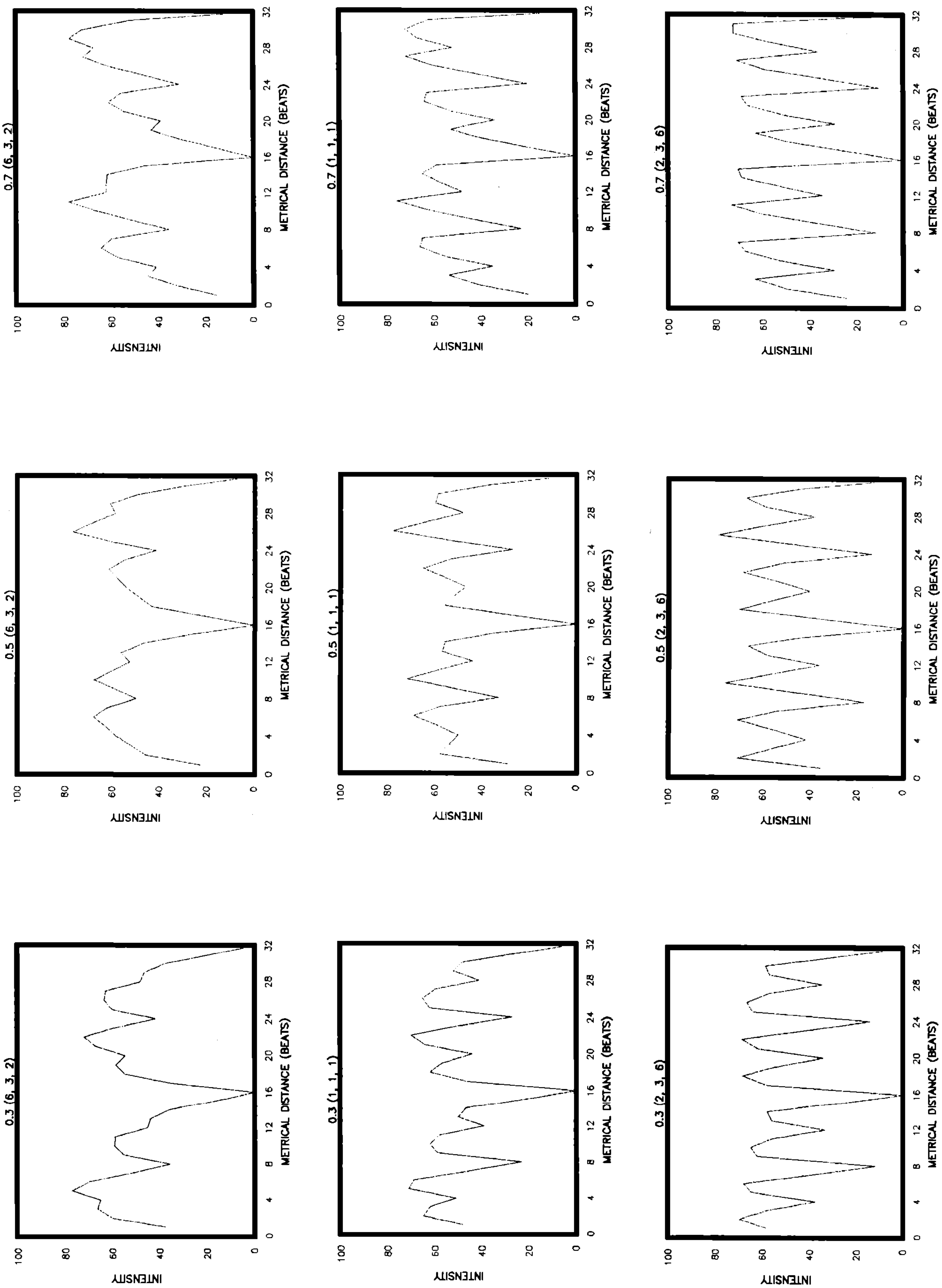

FIG. 7. Example outputs from the model with coupling strategy (c) with parameter values as in Fig. 5.

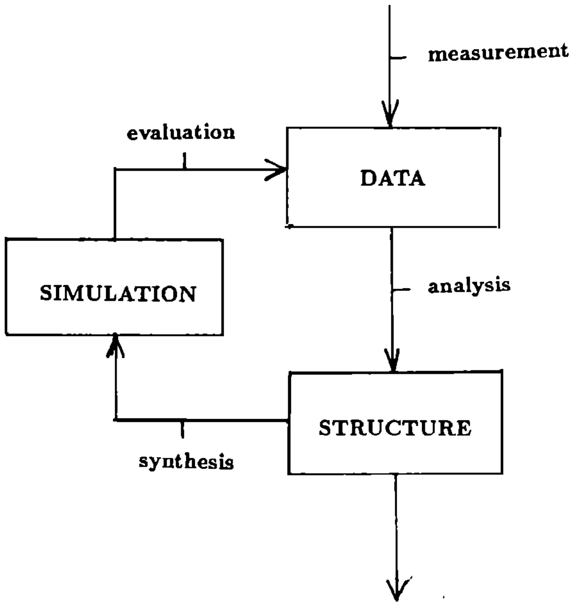

FIG. 8. The analysis/synthesis/evaluation method.

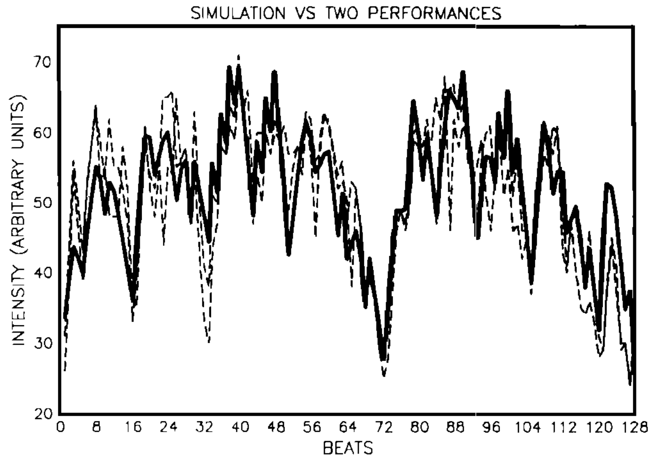

FIG. 10. The simulation (bold line) corresponding to the input structure as in Fig. 9 compared with the beat intensities from the two performances (dotted lines).

$$I = 2.70 + 0.196 I_1 + 0.379 I_2 + 0.466 I_3$$

so that the three masses are $m_1 = 5.10$, $m_2 = 2.64$, $m_3 = 2.15$. The regression satisfies the acceptability criterion with $R^2 = 74.0\%$ and $F(3,124) = 117.42$. The resultant simulation is shown in Fig. 10. The simulation is computed using a recursive algorithm written in LISP (Todd, 1989c).

```
(setq g2 '(1 1))
(setq g3 '(1 1 1))
(setq g4 '(1 1 1 1))
(setq g5 '(1 1 1 1 1))
(setq g6 '(1 1 1 1 1 1))
(setq tsr '(A B))
(setq A '((G1 G2) G3 G4 G5))
(setq B '((G6 G7 G8 G9) G10))
(setq G1 '(g5 g5 g6))
(setq G2 '(g5 g5 g3 g4))
(setq G3 '(g2 g2 g2 g4))
(setq G4 '(g2 g2 g4))
(setq G5 '(g6 g5 g2 g4 g4))
(setq G6 '(g5 g4 g3))
(setq G7 '(g5 g4))
(setq G8 '(g4 g2 g2 g4))
(setq G9 '(g5 g3 g4 g3))
(setq G10 '(g6 g2))
```

FIG. 9. The LISP implementation of grouping that corresponds to the simulation in Fig. 10.

VI. DISCUSSION

The model of musical dynamics presented in this paper was based on two basic principles. First, that musical expression has its origins in simple motor actions and that the performance and perception of tempo/musical dynamics is based on an internal sense of motion. Second, that this internal movement is organized in a hierarchical manner corresponding to how the grouping or phrase structure is organized in the performer's memory. The model was then used to develop a method for the analysis of actual performance data which enabled a compact description of the data.

It is interesting to speculate on possible psychological/neurophysiological interpretations of this model. In particular, why does artificial expression based on motion under constant acceleration sound natural? One possible answer lies in the idea that the other five organs of the inner ear, in addition to the *cochlear*, namely the *succule* and *utricle* in the vestibule and the *ampulla* in the semicircular canals, play an important role in the perception of expressive sounds. Traditionally, it has been thought that these organs are sensitive to gravity, linear and rotational acceleration to produce a percept of self-motion on the vestibular cortex. However, there is now compelling evidence (Lackner and Graybiel, 1974; Young *et al.*, 1977; Kalmijn, 1989; Hudspeth, 1989; Todd, in press) to support the view that these organs may also be sensitive to vibrational phenomena. In other words it may be the case that expressive sounds can induce a percept of self-motion in the listener and that the internal sense of motion referred to above may have its origin in the central vestibular system. Thus, according to this theory, the reason why expression based on the equations of elementary mechanics sounds natural is that the vestibular system evolved to deal with precisely these kinds of motions.

VII. CONCLUSION

Whilst the analysis carried out in this paper would appear to be convincing, the model and analytic method cannot be properly evaluated until it is carried out on a large number of performances for both tempo and musical dynamics. These further analyses should examine the effects of tempo, composer and performer style, individual differences, and instrument choice. Finally, in addition to the use of regression, the resultant simulations needs to be tested by producing synthetic performances.

ACKNOWLEDGMENTS

The research presented in this paper was made possible by a Science and Engineering Research Council post-doctoral fellowship. I am also indebted to Christopher Longuet-Higgins for his ideas on linear tempo.

Clarke, E. F. (1988). "Generative principles in music performance," *Generative Processes in Music: The Psychology of Performance, Improvisation and Composition,* edited by J. Sloboda (Claredon, Oxford).

Clynes, M. (1987). "What can a musician learn about music performance from newly discovered microstructure principles (PM and PAS)?," in *Action and Perception in Rhythm and Music,* edited by A. Gabrielsson (Royal Swedish Academy of Music, Stockholm), Vol. 55, pp. 201–237.

Draper, N. R., and Smith, H. (1981). *Applied Regression Analysis* (Wiley, New York).

Gabrielsson, A. (1987). "Once again: The theme from Mozart's piano Sonata in A Major (K.331)," in *Action and Perception in Rhythm and Music,* edited by A. Gabrielsson (Royal Swedish Academy of Music, Stockholm), Vol. 55, pp. 81–103.

Helmholtz, H. L. (1954). *On the Sensations of Tone as a Physiological Basis for a Theory of Music* (Dover, New York), 2nd English ed. of 3rd German ed. (1st ed. was 1863).

Hudspeth, A. J. (1989). "How the ear's works work," Nature 341, 397–404.

Kalmijn, A. J. (1989). "Functional evolution of lateral line and inner ear sensory systems," in *The Mechanosensory Lateral Line: Neurobiology and Evolution,* edited by S. Coombes, P. Görner, and H. Münz (Springer-Verlag, New York).

Kronman, U., and Sundberg, J. (1987). "Is the musical ritard an allusion to physical motion?," in *Action and Perception in Rhythm and Music,* edited by A. Gabrielsson (Royal Swedish Academy of Music, Stockholm), Vol. 55, pp. 57–68.

Lackner, J. R., and Graybiel, A. (1974). "Elicitation of vestibular side effects by regional vibration of the head," Aerosp. Med. 45, 1267–1272.

Longuet-Higgins, H. C., and Lisle, E. R. (1989). "Modelling music cognition." Contemp. Music Rev. 3, 15–27.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco).

Repp, B. H. (1990). "Patterns of expressive timing in performances of a Beethoven minuet by 19 famous pianists," J. Acoust. Soc. Am. 88, 622–641.

Sachs, C. (1943). *The Rise of Music in the Ancient World East and West* (Dent, London).

Seashore, C. E. (1938). *Psychology of Music* (McGraw-Hill, New York).

Shaffer, L. H. (1981). "Performances of Chopin, Bach and Beethoven: Studies in motor programming," Cognitive Sci. 13, 326–376.

Shaffer, L. H., and Todd, N. P. (1987). "The interpretive component in musical performance," in *Action and Perception in Rhythm and Music,* edited by A. Gabrielsson (Royal Swedish Academy of Music, Stockholm), Vol. 55, pp. 139–152.

Sloboda, J. A. (1983). "The communication of musical meter in piano performance," Q. J. Exp. Psychol. 88, 377–396.

Stetson, R. (1905). "A motor theory of discrete succession," Psychol. Rev. 12, 250–350.

Sundberg, J. (1988). "Computer synthesis of music performance," in *Generative Processes in Music: The Psychology of Performance, Improvisation and Composition,* edited by J. Sloboda (Clarendon, Oxford).

Sundberg, J., and Verillo, V. (1980). "On the anatomy of the retard: A study of timing in music," J. Acoust. Soc. Am. 68, 772–779.

Todd, N. P. McA. (1985). "A model of expressive timing in tonal music," Music Percep. 3, 33–58.

Todd, N. P. McA. (1989a). "A computational model of tempo," Contemp. Music Rev. 3, 69–88.

Todd, N. P. McA. (1989b). "Towards a cognitive theory of expression: The performance and perception of tempo," Contemp. Music Rev. 4, 405–416.

Todd, N. P. McA. (1989c). "Computational Theory and Implementations of an Abstract Expression System: A Contribution to Computational Psychomusicology," Ph.D. thesis, University of Exeter.

Todd, N. P. McA. (1990). "Computational modelling of musical expression," *Proceedings of the International Computer Music Conference* (Computer Music Association, San Francisco), pp. 373–375.

Todd, N. P. McA. (submitted). "The kinematics of musical expression," J. Acoust. Soc. Am.

Todd, N. P. McA. (in press). "The communication of self-motion in musical expression," *Proceedings of the International Workshop on Man-Machine Interaction in Live Performance* (Computer Music Department of CNUCE/CNR, Pisa, Italy).

Young, E. D., Fernandez, C., and Goldberg, J. M. (1977). "Responses of squirrel monkey vestibular neurons to audio-frequency sound and head vibration," Acta. Otolaryngol. 84, 352–360.