Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists

719

#719

Bruno H. Repp
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

(Received 18 September 1989; revised 23 April 1990; accepted 1 May 1990)

The timing patterns of 19 complete performances of the third movement of Beethoven's Piano Sonata op. 31, No. 3, were measured from oscillograms and analyzed statistically. One purpose of the study was to search for a timing pattern resembling the "Beethoven pulse" [Clynes, in *Studies of Music Performance* (Royal Academy of Music, Stockholm, 1983), pp. 76–181]. No constant pulse was found at the surface in any of the performances. Local patterns could be interpreted as evidence for an "underlying" pulse of the kind described by Clynes, but they could also derive from structural musical factors. On the whole, the artists' timing patterns served to underline the structure of the piece; lengthening at phrase boundaries and at moments of melodic/harmonic tension were the most salient features. A principal components analysis suggested that these timing variations in the Minuet could be described in terms of two orthogonal factors, one capturing mainly phrase-final lengthening, and the other reflecting phrase-internal variation as well as tempo changes. A group of musically experienced listeners evaluated the performances on a number of rating scales. Their judgments showed some significant relations to the measured timing patterns. Principal components analysis of the rating scales yielded four dimensions interpreted as force, individuality, depth, and speed. These preliminary results are encouraging for the development of more precise methods of music performance evaluation.

PACS numbers: 43.75.St, 43.75.Cd, 43.75.Mn

INTRODUCTION

It is generally recognized that competent music performance, especially of Western art music of the past two centuries, must go beyond the written score. Without such "deviations" from the literal notation, the music would sound inexpressive and mechanical, and the art of great interpreters lies largely in using such deviations with skill and taste. To the extent that the musical instrument permits it, variations in intensities, durations, and timbres of notes need to be introduced because traditional notation is not sufficiently precise in this regard as to the composer's intentions. Even those physical aspects that are precisely defined on paper, viz., fundamental frequency and timing of note onsets, require modulation by a performer to make the music come alive. Of these latter two aspects, that of variation in timing is of special interest to the psychomusicologist because it is universal to all instruments (see Gabrielsson, 1987), is a crucial aspect of performance skill, and can be measured without much difficulty.

Systematic studies of timing patterns in instrumental performance, usually on the piano, go back quite a number of years. Extensive work was done at the University of Iowa in the laboratory of Carl Seashore, who devoted a chapter in his classic book to piano performance (Seashore, 1938, pp. 225–253). Seashore and his collaborators used a photographic technique to record the hammer movements of a piano as it was played. At about the same time, Hartmann (1932) reported a detailed analysis of timing measurements derived from piano rolls. After a long hiatus during which little research of this kind seems to have been conducted, there is now renewed activity in several laboratories, especially at the universities of Exeter (Clarke, 1982; Shaffer, 1981, 1984; Shaffer *et al.*, 1985; Sloboda, 1983, 1985) and Uppsala (Bengtsson and Gabrielsson, 1980; Gabrielsson, 1974, 1987; Gabrielsson *et al.*, 1983); see also Povel (1977) and Palmer (1989).

This research has amply confirmed the existence of systematic deviations from strict timing in the performance of experienced keyboard players, and some understanding of the rules governing the timing deviations has begun to emerge (Todd, 1985; Clarke, 1988; Sundberg, 1988). Clarke (1988) has identified three structure-governed principles within the domain of expressive timing: (1) graduated timing changes that indicate grouping of notes, with maxima at group boundaries; (2) lengthening of a note inside a group to add emphasis to the following note; and (3) lengthening of structurally significant notes, especially at the beginnings of groups. Other timing rules based on local musical relationships have been proposed by Sundberg and his colleagues (see Sundberg, 1988), and composer-specific timing patterns have been postulated by Clynes (1983), based on analysis-by-synthesis techniques. Yet, both data and knowledge in this area are still quite limited in view of the diversity of musical compositions and of their possible interpretations. Quantitative analyses of music performance also tend to be laborious; for this reason, earlier studies have employed rather limited samples of music and small numbers of performers, so that their results are not necessarily representative of general principles.

The present study, although limited to a single musical composition, is the first to include a statistically representative sample of performers ($N = 19$). Moreover, whereas

 0001-4966/90/080622-20$00.80

most previous studies analyzed performances recorded in the laboratory, the present research follows Hartmann (1932), Povel (1977), and Gabrielsson (1987) by analyzing commercial recordings of world-famous artists. Thus the performances examined here reflect pianistic skill and interpretive insight at the highest level. The cost of this approach was some loss of measurement accuracy (partially compensated for by replication) and restriction of the investigation to timing variation only, since other measures are very difficult to obtain from sound recordings.

The goals of the study were threefold. One aim was to describe objectively and to compare the expressive timing patterns of famous pianists in relation to the musical structure of the composition; to point out commonalities and individual differences; and to look for instances of expressive features observed in earlier research. A second aim was to obtain musically experienced listeners' impressions and evaluations of the various performances, to uncover the judgmental dimensions used by these listeners, and to examine whether the judgments bear any relation to the objective timing patterns. The third aim was to search for a particular timing pattern, the "Beethoven pulse," which requires some more detailed explanation.

The theory of "composers' inner pulses" was developed by Clynes (1983, 1986, 1987a), following earlier ideas by Becking (1928) and himself (Clynes, 1969). Clynes proposed that the performance of the works of the great composers of the Classical and Romantic periods, if it is to capture the composer's individual personality, requires specific patterns of timing and intensity relationships that convey the composer's individual style of movement, as it were. These living, personal pulses (distinct from a mechanically precise pulse) are said to apply to a composer's works regardless of tempo, mood, and style. The pulse pattern is assumed to be nested within hierarchical metric units and to be repeated cyclically throughout a composition. Several such composer-specific pulse patterns have been "discovered" by Clynes using computer synthesis in conjunction with his own musical judgment, and they have been implemented in computer performances of various compositions using patented software (Clynes, 1987b). Perceptual tests in which listeners of varying musical experience were presented with pieces by four different composers performed with appropriate and inappropriate pulses (Repp, 1989, 1990; Thompson, 1989) have not consistently provided support for the perceptual validity of the pulse concept, although the latest study (Clynes and Patinson, in preparation) did obtain positive results with subjects that included a number of outstanding musicians.

In view of these difficulties with the perceptual validation of Clynes' theory, the complementary question of whether a composer's personal pulse can be found in great artists' interpretations deserves special attention. It is noteworthy that Clynes did not rely on measurements of actual performances in deriving composers' pulse configurations; moreover, he has warned researchers that the requisite measurement accuracy cannot be achieved with current methods (Clynes, 1987a, p. 207). This warning may apply to some very subtle effects; however, many expressive deviations, such as those considered in the present study, are sufficiently large to be measurable with reasonable accuracy even from noisy sound recordings. Although it is clear that the fixed, repetitive timing and intensity ("amplitude") patterns implemented in Clynes' computer performances are an idealization and that real performances are much more variable, Clynes' theory nevertheless implies that the pulse should be found to some extent in an excellent performance, perhaps overlaid on a multitude of structurally determined expressive deviations. Thus, for example, a performance of Beethoven's music by a pianist renowned as a Beethoven interpreter should exhibit the "Beethoven pulse" to some extent, and perhaps more so than a performance by a pianist with special expertise in the music of, say, Chopin. In addition, musical listeners' judgments of the extent to which real Beethoven performances "capture the composer's spirit" should show some positive relationship to a measure of the relative prevalence of the Beethoven pulse in these performances. These predictions were examined in the present study with respect to timing deviations.

The limitations of this enterprise should be recognized. Consideration of a single physical dimension of performance variation carries with it the danger of ignoring interactions with other dimensions, such as intensity variations. Although Clynes (1983) specifies independent timing and amplitude components of composers' pulses, this does not necessarily imply that these components are independent in actual performance; rather, they may be complementary to some extent. Another limitation is the restriction to a single composition by a single composer; clearly, a thorough search for composers' pulses in performances will eventually have to include many compositions by many composers. A third limitation, to be explained in more detail below, was that this investigation, because of the nature and tempo of the composition chosen, concerned primarily the higher level in Clynes' scheme of hierarchically nested pulses, even though the lower level (comprising time spans in the vicinity of 1 s) is considered the more basic one.

Despite these limitations, which diminish the impact of any negative outcome, the present study offered a valuable opportunity to provide an existence proof for Clynes' Beethoven pulse. The music chosen for this investigation was the third movement of Beethoven's Piano Sonata No. 18 in E-flat major, op. 31, No. 3, which is a representative and highly regarded work from Beethoven's early mature period. The movement has two contrasting sections (Minuet and Trio), the first having an expressive melodic line over a steady eighth-note accompaniment, and the second consisting of a kind of dialogue between "questioning" chords and "answering" melodic phrases. The Minuet, which constitutes the focus of this investigation, is well suited to a search for specific timing patterns because of its continuous movement. It might be argued that the traditional form of the Minuet imposed constraints on Beethoven's characteristic expression as well as on performers' realization of it, but this particular piece, which serves as the slow movement of the sonata, is, in fact, not particularly dancelike but highly expressive, at least in the Minuet section. One important consideration in choosing this music was that it was included in the perceptual tests of Repp (1989), where listeners consistently expressed a preference for a computer performance having the Beethoven pulse over performances with different pulse patterns or with none at all. Thus it seemed appropriate to search for a similar pulse in real performances of this music. The piece also offered methodological advantages: Both the Minuet and the Trio sections are divided into two parts with repeats, and the whole Minuet is repeated after the Trio, with the repeats within the Minuet again prescribed by the composer (and obeyed by most performers). Since, in addition, the two sections of the Trio are structurally very similar (if several interpolated bars are ignored) and may be treated as repetitions of each other, a single performance contains four repetitions of the musical material making up most of the composition. This fact was desirable both for a systematic assessment of performance variability across repetitions and for the reduction of measurement error by averaging across repetitions (in the absence of systematic differences).

I. MUSICAL MATERIALS

A. The composition

The third movement of Beethoven's Piano Sonata No. 18 in E-flat major, op. 31, No. 3, is reproduced in Fig. 1. It is entitled *Menuetto, Moderato e grazioso,* and has two contrasting main parts, the Minuet and the Trio. Both Minuet and Trio are in E-flat major and have 3/4 time signature. The Minuet consists of an upbeat (bar 0) followed by two 8-bar sections, labeled bars 1–8 and 9–16, respectively. Each section is repeated, with altered versions of bars 1, 8, and 16. (The two versions are labeled A and B.) There is continuous eighth-note movement in the accompaniment of the principal melody, which is rhythmically more varied and contains some sixteenth-notes. The Trio, too, starts with an upbeat that is followed by two sections with repeats. The first section has 8 bars (bars 17–24), while the second section has 14 bars (bars 25–38). Bars 25–30 are an interpolated ostinato passage, but bars 31–38 are very similar to bars 17–24 of the first section and were treated in the present analyses as if they were a repeat of those bars. The Trio features widely spaced, rising chord sequences followed by faster moving, falling cadences. As usual, the Minuet is repeated after the Trio; contrary to prevailing custom, however, the composer wrote the music out and prescribed repeats for each section. (It is common practice to omit section repeats in the second playing of a Minuet from the classical period, and some artists indeed disobey Beethoven's instructions in that regard.) The piece ends with an 8-bar Coda (bars 39–46) that perpetuates the rhythm of the Minuet upbeat.

For purposes of timing analysis, the piece was divided into three 8-bar sections (bars 1–8, 9–16, and 17–24/31–38), each of which occurred four times (except for those performances that omitted some repeats). The interpolated section in the Trio (bars 25–30, one repetition) and the coda (one occurrence only) were measured but excluded from

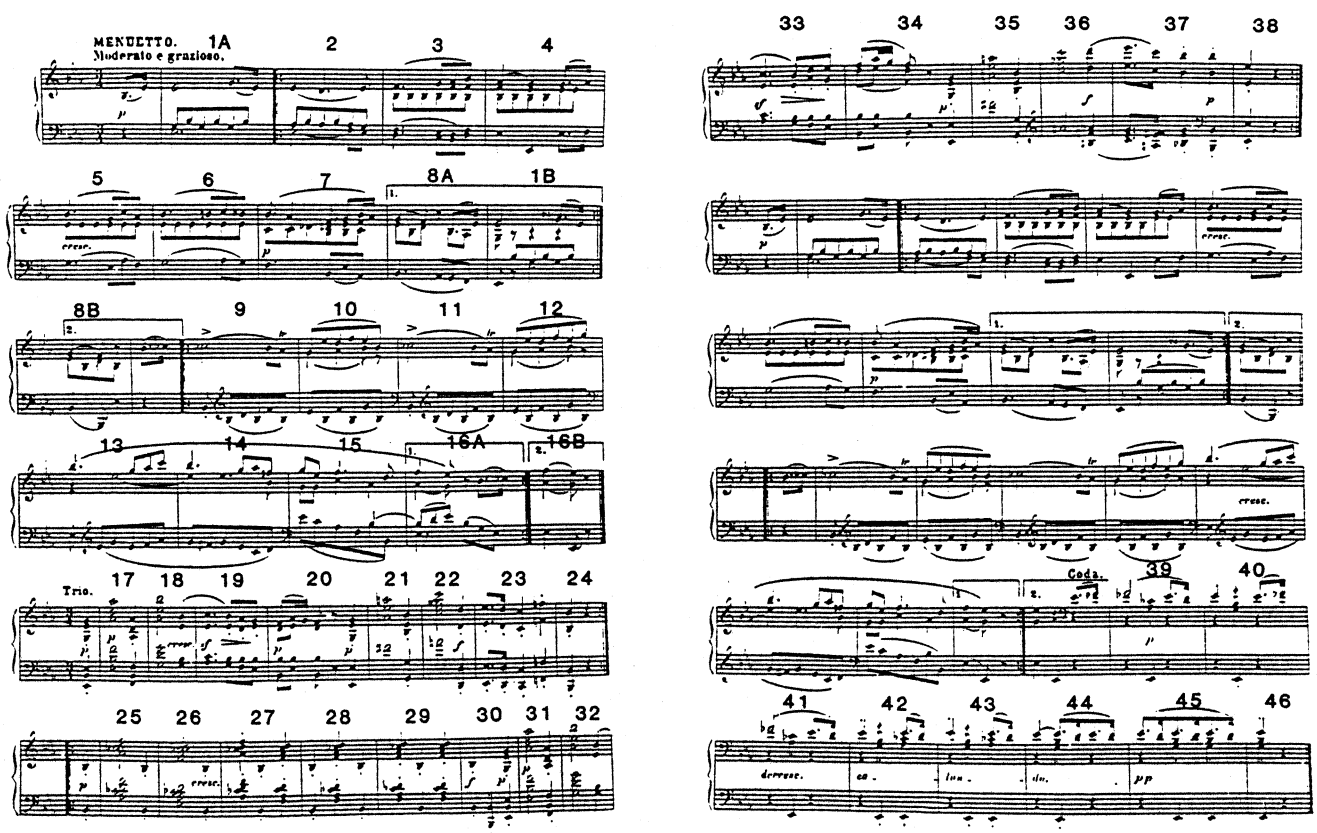

FIG. 1. The third movement of Beethoven's Piano Sonata in E-flat Major, op. 31, No. 3 (Urtext edition, Breitkopf & Härtel, 1898), with added numbering of bars.

most quantitative analyses. Some analyses were conducted only on the Minuet which, because of its steady motion, was more pertinent than the Trio to the goal of detecting a continuous timing pulse.

B. The recordings

Nineteen different recordings of the Beethoven Sonata were obtained from various sources. The artists and the record labels are listed in Table I, together with the total durations (excluding the first upbeat and the final chord) as determined by stopwatch. All except the Perahia performance (a cassette) were on regular long-playing records. Three of the performances (Davidovich, Rubinstein, Solomon) had been recorded originally from radio broadcasts onto reel-to-reel tape. Prior to measurement, all recordings were transferred to cassette tape.[1]

In addition to the 19 human performances, a computer performance of the piece (without section repeats) was available on cassette from the earlier perceptual study (Repp, 1989). This performance had been synthesized by Clynes at the Music Research Institute of the New South Wales Conservatorium of Music in Sydney, Australia, using a special program developed there (Clynes, 1987b) to drive a Roland MKS-20 digital piano sound module. This performance instantiated the "Beethoven pulse" defined by Clynes (1983) and is described in more detail below.

II. TIMING MEASUREMENTS: QUARTER-NOTES

A. Measurement procedure

Each recording was input to a VAX 11/780 computer at a sampling rate of 10 kHz, low-pass filtered at 4.8 kHz. Portions of the digitized waveform were displayed on the large screen of a Tektronix 4010 terminal. A vertical cursor (resolution: 0.1 ms) was placed at the onsets of notes, and the times between successive cursor positions (here called the onset–onset interval durations) were recorded. If the onset of a note was difficult to determine visually, enlargement of the waveform segment on the screen sometimes helped; otherwise, the cursor was moved back in small increments and the waveform up to the cursor was played back at each step until the onset of the sought-after note was no longer audible. Only a small percentage of the measurements was obtained using this perceptual criterion, usually for accompanying notes in the initial bars of the Minuet. When several notes coincided, their individual onsets could not be resolved in the waveform, and the earliest onset was measured. This would normally have been the melody note (cf. Palmer, 1989).

Complete measurements of all recordings were made at the level of quarter-note beats (i.e., three measurements per bar). Because hand measuring so many intervals (well over 6000) was extremely time consuming, some accuracy was sacrificed for speed by using relatively compressed waveform displays (about 5 s per 16-in. screen). An estimate of the average measurement error was available from one recording (Ashkenazy) that was accidentally measured twice. The mean absolute discrepancy between corresponding onset–onset interval durations was 12 ms or about 2%; the correlation coefficient was 0.98. This was considered quite satisfactory, especially since averaging over the quadruple repetitions of most of the music reduced random variability further by a factor of 2. A second estimate of measurement error was obtained from the computer performance, where the two repeats of the Minuet (before and after the Trio,

TABLE I. Alphabetical list of the artists and their recordings with durations timed by stopwatch (from onset of bar 1 to onset of last bar).

Artist	Recording	Duration
Claudio Arrau	Philips PHS 3-914	5 min 6 s
Vladimir Ashkenazy	London CS 7088	3 min 46 s
Wilhelm Backhaus	London CM 9087	3 min 43 s
Lazar Berman	Columbia M3421	4 min 26 s
Stephen Bishop	Philips 6500 392	3 min 43 s
Alfred Brendel[a]	Vox SBVX 5418	3 min 30 s
Bella Davidovich[a]	Philips 9500 665	3 min 39 s
Claude Frank	RCA VICS 9000	4 min 10 s
Walter Gieseking	Angel 35352	3 min 41 s
Emil Gilels	DG 2532 061	4 min 46 s
Glenn Gould[a,b]	CBS Masterworks 7464-39547-1	4 min 12 s
Friedrich Gulda	Orpheus OR B-1225	3 min 38 s
Clara Haskil	Epic LC 1158	4 min 4 s
Wilhelm Kempff	DG 2740 228	4 min 25 s
Murray Perahia	CBS MT 42319 (Cassette)	4 min 8 s
Charles Rosen	Nonesuch NC-78010	4 min 22 s
Artur Rubinstein[a]	RCA LM 2311	3 min 56 s
Arthur Schnabel	Angel GRM 4005	4 min 2 s
Solomon	EMI RLS 704 (probably)	4 min 12 s
Computer[c]	Manfred Clynes (private cassette)	2 min 9 s

[a] Section repeats not taken in second playing of Minuet.
[b] Second section repeat of Trio omitted.
[c] No section repeats at all.

TABLE II. (A) Average quarter-note onset–onset interval durations in milliseconds. (B) The corresponding metronome speeds (quarter-notes per minute). (C) and (D) more accurate metronome speeds for the Minuet and Trio separately (see text for explanation).

Artist	A	B	C	D
Gould	977	61	64	68
Arrau	879	68	72	71
Gilels	822	73	76	78
Rubinstein	787	76	70	89
Berman	764	79	83	82
Kempff	761	79	78	85
Rosen	753	80	85	83
Davidovich	730	82	80	87
Solomon	724	83	83	87
Frank	718	84	90	85
Perahia	713	84	85	91
Brendel	700	86	86	88
Haskil	701	86	82	91
Schnabel	695	86	86	92
Computer	694	86	84	95
Ashkenazy	649	92	95	91
Backhaus	641	94	93	105
Bishop	641	94	93	101
Gieseking	635	94	96	94
Gulda	626	96	99	92

both measured independently) presumably were physically identical. After correcting one large mistake and omitting the values for the final note, whose duration had been deliberately extended by Clynes in synthesis, the average absolute measurement error was 6.5 ms, or less than 1%, and the correlation coefficient between the two sets of measurements was 0.99.

B. Overall tempo

The average quarter-note duration of each performance was calculated by dividing the total duration (see Table I) by the number of quarter-note beats (348 in a performance with all repeats). From this average quarter-note duration, the average metronome speed [quarter-notes per minute (qpm)] of each performance was determined.[2] Both measures are listed in Table II (columns A and B), where the performances have been rearranged from slowest to fastest.

It can be seen that there was a wide range of tempi represented, with the fastest performance (Gulda, 96 qpm) being more than 50% faster than the slowest (Gould, 61 qpm). The average metronome speed was 83 qpm. These values underestimate the underlying tempo somewhat because they include ritards, lengthenings, and pauses at phrase endings. To obtain better estimates, and also to compare the tempi for the Minuet and Trio, separate estimates for these two sections were obtained by computing the average quarter-note durations and corresponding metronome speeds from the detailed timing measurements, after excluding all bars showing conspicuous lengthening of one or more onset–onset in-

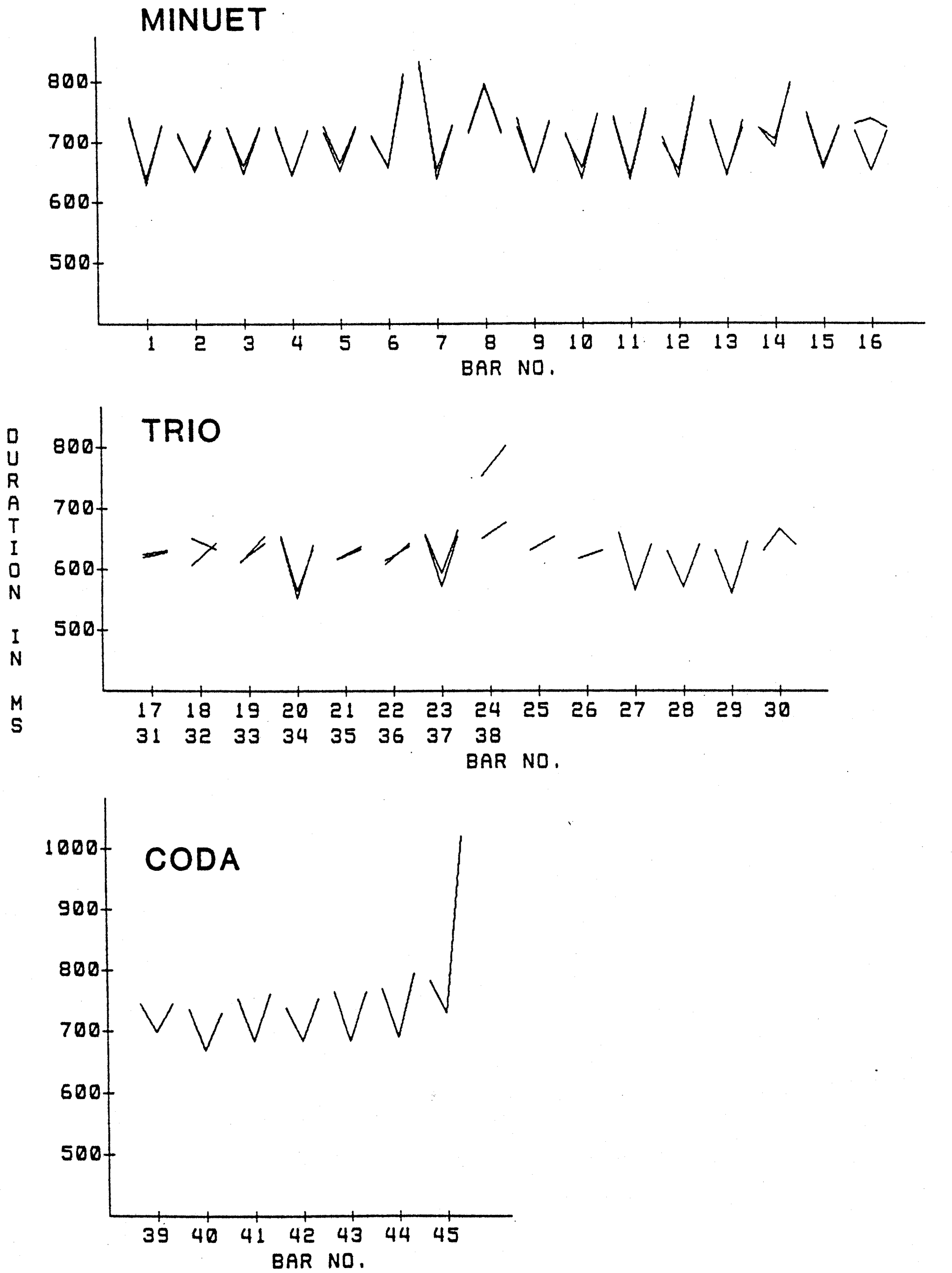

FIG. 2. Quarter-note timing patterns in the computer performance created by Manfred Clynes, as measured by the author.

terval durations in the grand average timing pattern (discussed below). These excluded bars were nos. 1, 6, 7, 8, 15, and 16 in the Minuet, and nos. 24, 30, and 38 in the Trio; the Coda was also excluded. The resulting metronome speeds are shown in columns C and D in Table II. They are indeed somewhat faster than estimated previously, with the average speed of the Minuet being 84 qpm and that of the Trio being 87 qpm.[3]

C. The three-beat Beethoven pulse

Clynes (1983) defined a composer's basic pulse as a particular pattern of time (and amplitude) relationships of the notes within a time unit of approximately 1 s. This pulse is nested within a slower, similar pulse operating on larger time units. In the Beethoven piece under investigation, the faster pulse was defined over the four sixteenth-notes within each quarter-note, whereas the slower pulse was defined over the three quarter-notes within each bar. In the computer performance, the faster pulse thus extended over a time unit of approximately 700 ms duration, whereas the slower pulse extended over about 2 s. The duration (and amplitude) ratios at one level are independent of those at the other.

In this study, the relative paucity of sixteenth-notes in the piece chosen led to a main focus on the slower pulse, defined over three quarter-notes within bars. The three-beat pulse used in generating the computer performance (Clynes, personal communication) had a basic timing pattern of [102.5, 94, 103.5]. This means that the first quarter-note onset–onset interval was 2.5% longer than it would have been in a mechanical performance, the second was 6% shorter, and the third was 3.5% longer.[4] Represented graphically in terms of onset–onset interval durations, this pulse reflects a V-shaped pattern within a bar: The first and third intervals are about equally long, but the middle one is shortened.

The actual timing of the quarter-notes in the computer performance was measured in the same fashion as in the human performances. The results of these measurements are displayed in Fig. 2. The upper panel of Fig. 2 displays the 16 bars of the Minuet, with the two repeats superimposed; the center panel shows the Trio, with bars 31–38 laid on top of bars 17–24; and the bottom panel shows the Coda. The onset–onset interval values of the three quarter-notes within each bar have been connected to reveal the V-shaped pattern. The initial upbeat is not included; other upbeats constitute the last notes of bars 8, 16, 30, and 38, respectively. A number of bars in the Trio contain a half-note followed by a quarter-note; for these, only two values are plotted, the first of which represents half the duration of the first onset–onset interval.

The identity of the two repeats is obvious; any small discrepancies represent measurement error (see above). Large discrepancies occur in bars 16 and 24/38, where the final onset–onset interval of the Minuet preceding the Coda, the final onset–onset interval of the Trio preceding the Minuet repeat, and the upbeat of the Minuet repeat must have been deliberately extended by Clynes; another inconsistency, in bars 18/32, is of uncertain origin. A deliberate elongation of phrase-final intervals (on the second beat) is also evident in bars 8 (both repeats) and 30. Other modifications that Clynes apparently applied "by hand" to improve the musical quality of the computer performance include the prolongation of the last onset–onset interval of bar 6 (an expressive deviation that we will encounter again in many human performances), of the first onset–onset interval of bar 7, and of the second and third intervals of bar 14. The remaining bars (1–5, 9–13, 15, 16 [first repeat], 20/34, 23/37, and 27–29) exhibit the V-shaped onset–onset interval pattern characteristic of the Beethoven pulse, although some variability of the pulse shape is evident. For example, the V is deeper in bar 1 than in bars 2–5, and bars 10 and 12 exhibit an asymmetry not shared by most other bars, some of which show a smaller asymmetry in the opposite direction. This variability may represent additional adjustments made by Clynes in creating the computer performance. The two-note bars of the Trio show a rising pattern, which is consistent with the prescribed pulse, since the duration of the first value is simply the average of the first two pulse beats and thus is expected to be slightly shorter than the value for the third beat.

All in all (including repeats and Coda), there were 34 measured bars with regular V-patterns. The onset–onset interval durations in these bars were expressed as percentages of one-third of the total bar duration, and the averages of these percentages were calculated across the 34 bars. The result was an average pulse of [103.7, 92.4, 103.9], which is reasonably close to (but not identical with) the pulse of [102.5, 94, 103.5] that was reportedly used in generating the computer performance.

The timing pattern of the computer performance may be considered a hypothesis about the timing pattern to be observed in expert human performances. It is not an exact prediction because human performances may be expected to be less precise and also may include additional variations in response to musical structures, that were not implemented in the computer performance. This additional variation would be superimposed on the composer's pulse (if any) to create a more complex and changing timing pattern. However, unless this additional variation is ubiquitous and large in relation to the pulse, the pulse should be detectable in the timing pattern, if it is present. Moreover, being a cyclic repetitive phenomenon, it should be present throughout a performance, although perhaps not with the consistency illustrated in Fig. 2.

D. The "grand average performance"

A "grand average" timing pattern was obtained by averaging the onset–onset interval durations across 15 of the 19 human performances, keeping the repeats separate. The four performances that omitted some repeats (Brendel, Davidovich, Gould, and Rubinstein) were not included; this was just as well, since Gould's and Rubinstein's were the two most deviant performances. (See their discussion by Kaiser, 1975, pp. 340–342.) The result is plotted in Fig. 3 in the format introduced by Fig. 2. The lines for the four repeats are superimposed. The grand average represents those aspects of expressive timing that were common to most performances.

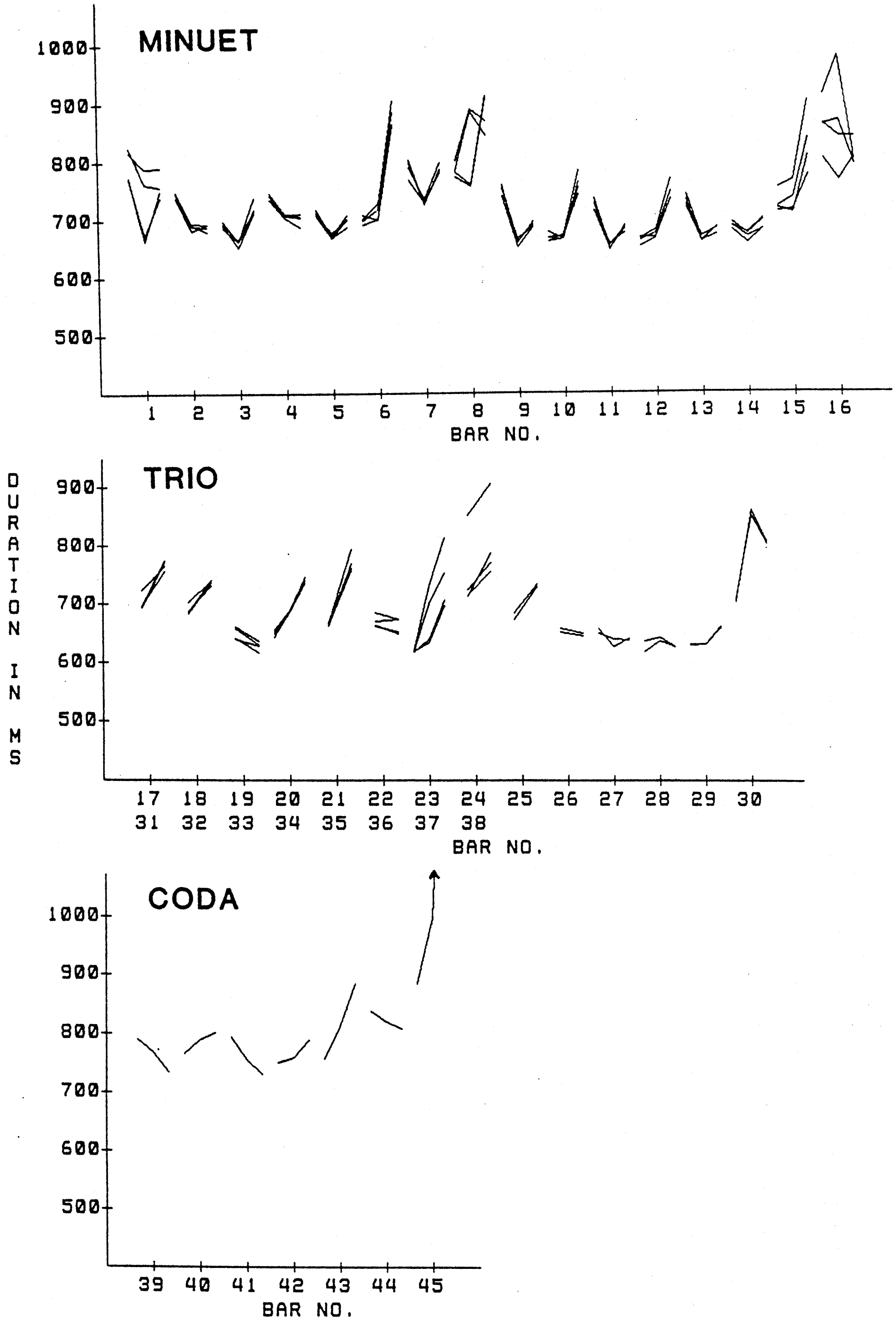

FIG. 3. Grand average timing pattern of 15 human performances, with repeats plotted separately. The last onset–onset interval in the Coda (arrow) is 1538 ms.

1. Repeats

It is evident, first, that repeats of the same material had extremely similar timing patterns. This consistency of professional keyboard players with respect to detailed timing patterns has been noted many times in the literature, beginning with Seashore (1938, p. 244). The only systematic deviations occurred in bar 1 and at phrase endings (bars 8, 15–16, and 23/37–24/38), where the music was, in fact, not identical across repeats (see Fig. 1): In bar 1, Beethoven added an ornament (a turn on E-flat) in the repeat (bar 1B), which was slightly drawn out by most pianists. In bar 8A, which led back to the beginning of the Minuet, the upbeat was prolonged, but in bar 8B, which led into the second section of the Minuet, an additional ritard occurred on the phrase-final (second) beat. Similarly, a uniform ritard was produced in bar 16A, which led back to the beginning of the second Minuet section, and an even stronger ritard occurred on the phrase-final (first and second) notes of bar 16B, which constituted the end of the Minuet, whereas the third note constituted the upbeat to the Trio and was taken shorter. Bar 15 anticipated these changes, which were more pronounced in the second playing of the Minuet, following the Trio. Similarly, bar 37 anticipated the large ritard in bar

38 when, in the second repeat, it concluded the Trio. Another phrase-final phenomenon, consistent across both repeats, occurred in bar 30 of the Trio, at the end of the interpolated section. All these deviations confirm the well-known principle of phrase-final lengthening (e.g., Lindblom, 1978; Todd, 1985; Clarke, 1988). Only one major phrase-internal expressive deviation was evident: the lengthening of the third beat in bar 6. That beat not only carries an important melodic inflection in eighth-notes, but also is followed by an abrupt change in dynamic level (*subito piano*) in the score (see Fig. 1). Thus it is an example of Clarke's (1988) second and third principles: lengthening of a note to underline its own structural significance as well as to enhance the next note (which, being an appoggiatura, needs the enhancement especially because of the sudden reduction in intensity).

The consistency of the grand average timing pattern across repeats in the Minuet was quantified by computing the correlations of onset–onset intervals between (a) immediate repeats (where slight changes in the music occurred) and (b) distant repeats (i.e., of the identical music before and after the Trio). The correlations were computed separately for the two Minuet sections and then averaged. These correlations were (a) 0.85 and (b) 0.95, the former being predictably lower because of the musical and interpretive changes discussed above. Corresponding correlations were also calculated for each of the 19 individual performances, to determine the individual consistency of the different artists. The average individual correlations were (a) 0.66 and (b) 0.79. Only one pianist (Davidovich) showed exceptionally low correlations across repeats (0.36, 0.32); all others showed rather high consistency. The correlations are lower than those for the grand average because they include random as well as perhaps intended but idiosyncratic variation across repeats that canceled out in the grand average.

It can be seen in Fig. 3 that the performances were similarly consistent across repeats in the Trio.

2. Pulselike patterns

Consider now the within-bar onset–onset interval patterns in the grand average. According to Clynes' theory, they should follow a fairly consistent V-shaped pattern, especially in those bars that do not contain major deviations due to phrase boundaries or special emphasis. It is clear from Fig. 3 that in the Minuet only two bars show the nearly symmetric V shape of the Beethoven pulse: bar 1 (first repeat only) and bar 7. Bar 5 shows a shallower V shape, bar 3 has an asymmetric V shape, and so do bars 9, 11, and 13, but with the asymmetry going in the opposite direction. Interestingly, these bars are all odd-numbered ones. The even-numbered bars, discounting those with major expressive deviations (6, 8, 16), do not show V shapes: Bars 2 and 4 show only an elongated first onset–onset interval, bars 10 and 12 an elongated third onset–onset interval, and bar 14 is almost evenly timed.

This curious alternating pattern makes good sense when the melodic and harmonic structure of the Minuet is considered. Moments of tension (relative dissonance, elevated pitch) alternate with relaxation (high consonance, lower pitch). The former occur on the first beats of bars 2, 4, 7, 9, 11, 13, and 16, which are all prolonged; the latter fall on the following beat, which is invariably shortened. The beat preceding a moment of high tension also tends to be prolonged (third beats of bars 1, 3, 6, 8, 10, 12, and 15). Thus it seems that the timing pattern in the Minuet was determined primarily by the expressive requirements of the melody, not by a constant, autonomous pulse.

The results for the Trio and for the Coda reinforce these conclusions. None of the bars with three measured onset–onset intervals shows a V-like shape: Some shapes are rising, others are flat or falling. Of the bars with only two measurements, some show a rising pattern, but others a slightly falling one. The observed patterns, however, again make sense with respect to the musical events. Basically, the pianists tended to take the staccato upbeats preceding the half-notes somewhat longer than notated, perhaps to create a slight suspense, or simply as a physical consequence of the large leap upwards. Other bars in the Trio, including the ostinato in the interpolated section (bars 26–29), were played with very even timing. In the Coda, a tension-relaxation pattern is evident, which follows the melodic contour and echoes that observed in bars 9–12 of the Minuet.

Although the grand average does not show any evidence of a *constant* V-shaped pulse, Fig. 3 does suggest an overall tendency for the second onset–onset interval in a bar to be shorter than the first and the third. To examine whether these within-bar differences in relative onset–onset interval duration were consistent across pianists, a repeated-measures analysis of variance (ANOVA) was conducted on the Minuet onset–onset interval data after converting them to percentage values of the total bar durations (as used in Clynes' pulse specifications), which effectively eliminated tempo differences among different performances, and among bars within performances. The analysis included only the 15 performances with complete repeats, and only the 11 bars (2–5, 7, and 9–14) without major expressive deviations and/or variations across repeats. The fixed factors in the ANOVA were beats (3), bars (11), and repeats (4); the random factor was pianists (15). There was indeed a highly significant main effect of beats [$F(2,28) = 43.92$, $p < 0.0001$], which confirms that, overall, there were reliable differences among the three onset–onset intervals in a bar. The average percentages were [102.2, 96.3, 101.5], showing the predicted reduction of the second onset–onset interval. However, there was also a highly significant interaction of beats with bars [$F(20,280) = 17.96$, $p < 0.0001$], which shows the bar-to-bar variations in the onset–onset interval patterns (cf. Fig. 3) to be reliable across pianists. While the overall "pulse" (the beats main effect) was highly significant with pianists as the sampling variable, it was less reliable, although still significant, when bars were considered the sampling variable—that is, when the size of the beats main effect was compared to that of the beats by bars interaction [$F(2,20) = 7.22$, $p < 0.01$]. These results permit the interpretation that there *is* some underlying constant pulse that, together with local musical requirements, contributes to the surface pattern of onset–onset interval durations. To the extent that the average pattern of [102.2, 96.3, 101.5] is the best estimate of such an underlying pulse (which it may

TABLE III. Individual pianists' average timing patterns (in onset–onset interval percentages) across 11 bars of the Minuet; significance levels of the beats main effect across repeats (R) and across bars (B); and mean-squares term of the beats by bars interaction (MSQ), which provides a measure of the relative bar-to-bar variability of the timing pattern: (***) = $p < 0.0001$, (**) = $p < 0.001$, (*) = $p < 0.01$.

Artist	Onset–onset interval pattern	R	B	MSQ
Arrau	[101.2, 96.0, 102.8]	**		161
Ashkenazy	[98.9, 97.0, 104.1]	**	*	89
Backhaus	[104.1, 97.2, 98.7]	**	*	68
Berman	[101.6, 92.7, 105.7]	***	*	241
Bishop	[102.5, 95.5, 102.0]	**		183
Brendel	[102.9, 96.2, 100.9]	*		89
Davidovich	[102.5, 96.1, 101.4]	**		68
Frank	[101.9, 97.2, 100.9]	**		109
Gieseking	[103.1, 95.3, 101.6]	**	*	76
Gilels	[101.8, 95.7, 102.5]	*	*	67
Gould	[101.0, 99.1, 99.9]			21
Gulda	[102.7, 97.3, 100.0]	**	*	46
Haskil	[101.8, 98.7, 99.5]	**		66
Kempff	[103.4, 96.5, 100.1]	***	*	79
Perahia	[103.1, 95.4, 101.5]	***	*	99
Rosen	[102.6, 94.7, 102.7]	**		202
Rubinstein	[102.1, 95.2, 102.7]	**		259
Schnabel	[101.3, 97.9, 100.8]	*		117
Solomon	[103.6, 96.8, 99.6]	*		124

not be), it is not radically different from the pattern applied by Clynes in the computer performance [102.5, 94, 103.5], although a separate analysis showed the difference to be significant [$F(2,20) = 19.06, p < 0.0001$].

E. Differences in timing patterns among individual pianists

Even though the grand average timing pattern did not show much evidence of a continuous V-shaped pulse, the possibility exists that certain individual pianists did exhibit such a pattern to a greater extent. The discovery of a composer's personal pulse is said by Clynes (1987a) to be restricted to those who are intimately familiar with a composer. Although all of the 19 great artists examined here must be (have been) thoroughly familiar with Beethoven's works, some of them are nevertheless considered greater Beethoven interpreters than others by critics and concert audiences, and they also differ in how often they perform(ed) Beethoven's music. In the overall ANOVA, when repeats (rather than pianists) were considered the random factor, there was, in fact, a significant pianists by beats interaction [$F(28,84) = 10.16, p < 0.0001$], showing that the artists differed in their average within-bar timing patterns. There was also a significant pianists by bars by beats interaction [$F(280,840) = 4.40, p < 0.0001$], indicating that the artists varied their timing patterns across bars in different ways.

Individual analyses of variance were conducted on each pianist's Minuet data, again expressing onset–onset durations as percentages within bars (i.e., eliminating tempo variations across bars and across pianists) and including only the 11 bars without major expressive deviations. Repeats served as the random factor in these analyses, its interaction terms providing the error estimates. The results of these analyses are summarized in Table III. The average onset–onset interval patterns of the individual artists show some striking similarities: All artists but one (Ashkenazy) prolonged the first onset–onset interval somewhat, and all, without exception, reduced the second onset–onset interval by varying amounts. The majority did not change the last onset–onset interval much, although some prolonged it. These consistencies explain the statistical reliability across pianists of the grand average onset–onset interval pattern described above. The individual average onset–onset interval patterns were reliably different from mechanical evenness ([100, 100, 100]) *across repeats* for all pianists but one (Gould), although some pianists showed more consistently expressive patterns than others. (There appears to be no relation to the artists' renown as Beethoven interpreters.) Only eight pianists, however, showed a reliable onset–onset interval pattern *across bars*, and then only at the $p < 0.01$ level. Moreover, all pianists showed a highly significant ($p < 0.0001$) beats by bars interaction; that is, every one of them (even Gould) varied the timing pattern between bars and maintained these variations systematically across repeats. Note that this analysis concerned only those bars that did not have any major expressive deviations to begin with. Thus no single artist showed any constant pulse at the surface, although eight of them might be credited with a possible underlying pulse that was overlaid by expressive variations of a different kind.

The last column in Table III lists the mean-square terms of the beats by bars interaction, which provide a relative numerical measure of how much the onset–onset interval pattern varied from bar to bar. These values are correlated with the "depth" of the average onset–onset interval modulations: Pianists with a shallow-average pattern (e.g., Gould, Haskil, Gulda) also tended to vary less from bar to bar, whereas pianists with a highly modulated average pattern (e.g., Berman, Rosen, Rubinstein) also showed large variations from bar to bar. In other words, the highly expressive pianists in the latter group, whose average timing patterns resembled most that of the postulated Beethoven pulse, were least inclined to maintain a constant pulse throughout. The V-shaped average pattern may well be the consequence of structural musical factors that, on the whole, favored relative lengthening of the first and third beats, rather than the reflection of an underlying autonomous Beethoven pulse.

F. Factor analysis of timing patterns

So far, the analysis has considered only a subset of the bars in the Minuet, those without major expressive deviations, and tempo variations across bars have been disregarded. A more comprehensive analysis was conducted on the complete Minuet quarter-note onset–onset interval data (absolute durations, averaged over repeats) of all 19 pianists. The statistical technique employed was principal components factor analysis with Varimax rotation, which reveals the structure in the matrix of intercorrelations among pianists' timing patterns. One purpose of the analysis was to determine whether the individual differences in timing patterns could be described in terms of a single factor (implying that all pianists follow the same pattern, only in different degrees), or whether several factors would emerge. The sec-

TABLE IV. Factor loadings and communalities (squared multiple correlations) of the various performances in the three rotated factors found by principal component analysis of the Minuet data.

	Factor 1	Factor 2	Factor 3	Communality
Arrau	0.662	0.371	0.341	0.692
Ashkenazy	0.380	0.666	0.335	0.700
Backhaus	0.085	0.884	0.214	0.834
Berman	0.579	0.352	0.630	0.865
Bishop	0.333	0.623	0.426	0.681
Brendel	0.261	0.360	0.749	0.758
Davidovich	0.372	0.517	0.509	0.664
Frank	0.665	0.581	0.289	0.863
Gieseking	0.181	0.704	0.403	0.692
Gilels	0.707	0.506	0.332	0.865
Gould	0.919	− 0.210	0.181	0.921
Gulda	0.569	0.307	0.613	0.793
Haskil	0.567	0.539	− 0.056	0.616
Kempff	0.284	0.618	0.491	0.704
Perahia	0.807	0.395	0.293	0.893
Rosen	0.691	0.450	0.404	0.843
Rubinstein	0.768	0.407	0.228	0.807
Schnabel	0.582	0.644	0.122	0.769
Solomon	0.615	0.452	0.385	0.731
Computer	0.104	0.111	0.807	0.674

ond purpose was to see whether a factor could be extracted that reflects the hypothetical underlying Beethoven pulse. If there is such a pulse that combines additively with timing variations of a different origin, then principal components analysis would seem to be a good method of separating these different sources of variation.

To facilitate extraction of a "pulse" factor, the computer performance, which instantiated the Beethoven pulse, was included in the analysis. This computation on the 20×20 intercorrelation matrix yielded three orthogonal factors considered significant by the criterion that their eigenvalues were greater than one. These three factors together accounted for 77% of the variance in the data. Before rotation, the first factor accounted for most of that variance (63%), with the other two factors adding only 8% and 6% of variance explained, respectively. After Varimax rotation, which aims for a "simple pattern" of factor loadings, the variance accounted for was more evenly divided among the three factors: 31.2%, 26.6%, and 19.0%, respectively. Table IV shows the factor loadings of the individual performances (i.e., the correlations of individual timing patterns with the patterns characterizing each of the three factors) and their communalities (i.e., their squared multiple correlations with all three factors, which represent the variance explained by the factors). Figure 4 shows the factor timing patterns themselves, rescaled into the millisecond domain by multiplying the standardized factor scores with the average standard deviation and adding this product to the grand mean. In interpreting these data, it should be kept in mind that all the performances, with few exceptions, showed substantial positive intercorrelations (0.5–0.8) that were caused by the major expressive excursions and ritards shared by most artists.

The first factor represents primarily the phrase-final lengthenings in bars 8 and 15–16. The timing pattern in bars 9–13 is weakly represented, as is the phrase-final lengthening in bar 4. Most artists have high loadings on this factor, which means that they dutifully marked the major phrase boundaries. The highest correlation is exhibited by Gould (whose performance offered little else), followed by Perahia and Rubinstein; low correlations are shown by Backhaus, Clynes' computer performance, and Gieseking.

The second factor, orthogonal to the first, also represents the phrase-final lengthening in bars 8 and 15–16, although less strongly, and in addition shows a "slow start," the expressive lengthening in bar 6, and a strong tendency for the second half of the Minuet to be faster than the first. This last feature was especially obvious in Backhaus' performance, which also shows the highest loading on this factor, followed by Gieseking and Ashkenazy. Low loadings are exhibited by Gould and Clynes' computer performance.

The third factor, in striking contrast to the other two factors, represents a relatively even, V-shaped timing pattern, although its depth exceeds that of the pulse implemented in the computer performance (cf. Fig. 2). Not surprisingly, the computer performance exhibits the highest loading on this factor, followed by Brendel, Berman, and Gulda. Low loadings are associated with Haskil, Gould, and Schnabel.

The emergence of this third factor is intriguing and might be taken as a confirmation of an underlying Beethoven pulse in at least some of the performances. However, this factor was due to the inclusion of the computer performance in the analysis: When the analysis was repeated with the computer performance excluded, it returned only two factors that together accounted for 73% of the variance, which was split about evenly after rotation. The first factor was quite similar to that of the previous analysis, while the second factor was a conflation of the second and third factors obtained earlier; that is, it exhibited various tilted V-shaped patterns within bars, similar to those seen in the grand average performance (Fig. 3). It is not clear, therefore, how much importance should be attached to the extraction of a separate pulse factor when the computer performance was included. Nevertheless, that analysis provides a description of the individual performances in terms of several independent timing aspects, and it successfully isolates a pulselike aspect from other component patterns. The loadings in factor 3 (Table IV) provide a measure of the degree of presence of an underlying pulse in individual performances, regardless of whether or not the evidence is deemed sufficient for concluding that there is such a pulse. We will examine later whether this measure shows any relation to judgments of the performances as more or less "Beethovenian."

III. TIMING MEASUREMENTS: SIXTEENTH- AND EIGHTH-NOTES

The analyses so far have concerned the temporal microstructure at the level of quarter-notes, which constitutes the higher level in Clynes' pulse hierarchy for this particular piece. This level was relatively easy to access and measure. Within each quarter-note, however, Clynes defined a four-beat Beethoven pulse, which forms the lower level in the pulse hierarchy. This level was more difficult to evaluate because its full expression required sixteenth-notes, which

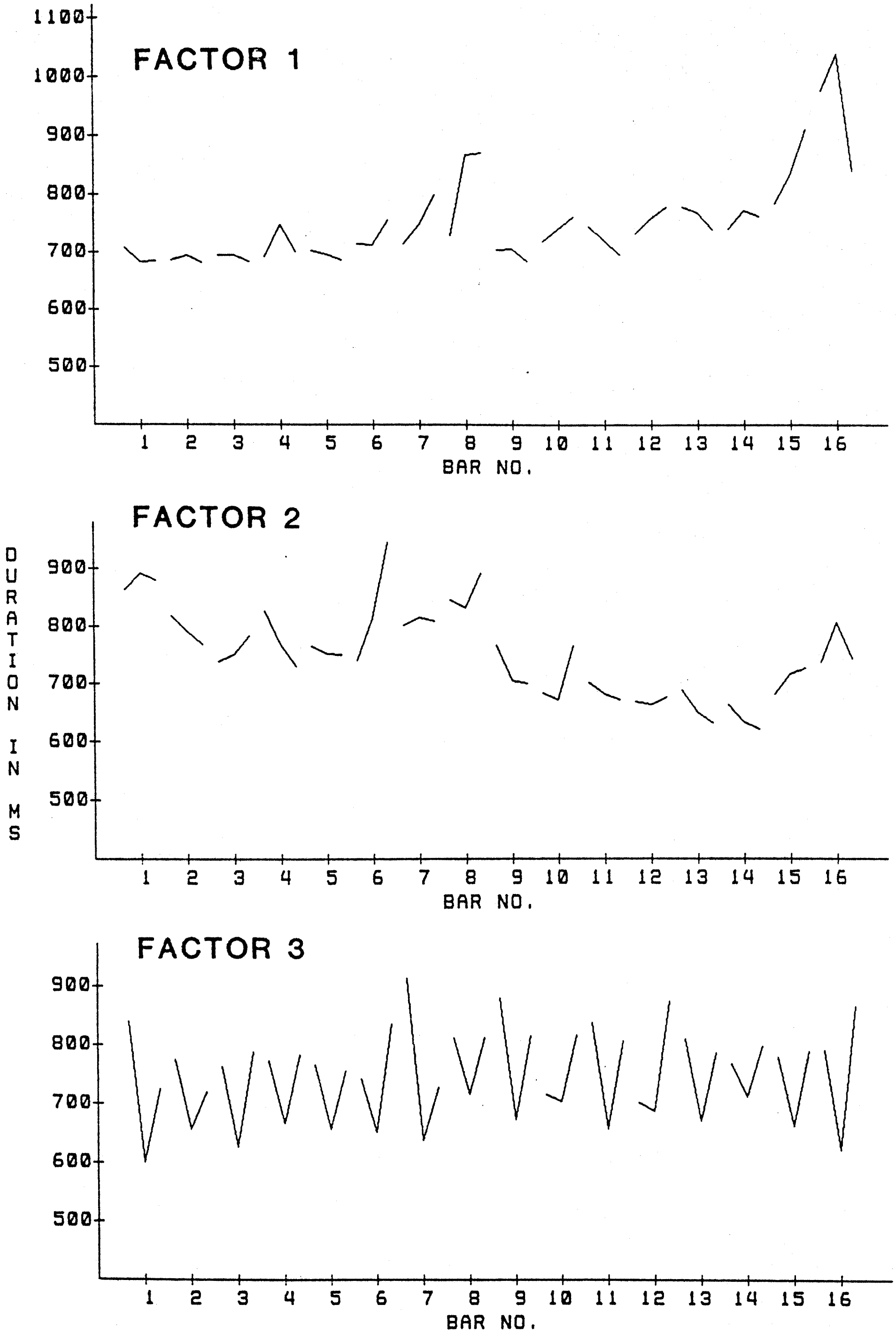

FIG. 4. Factor timing patterns emerging from the principal components analysis including the computer performance.

were quite rare in the present composition. Eighth-notes were common but provided less information, since they reduced the four-beat pulse to a two-beat pulse. Some measurement problems were also encountered. Nevertheless, some data were obtained about the temporal microstructure at this level.

A. Sixteenth-notes

1. Measurement procedures

Sequences of two sixteenth-notes occur in several places (bars 7, 20, and 34), but proved very difficult to measure; the onset of the second note could usually not be found in the acoustic waveform. Therefore, the measurements were restricted to single sixteenth-notes following a dotted eighth-note. Such notes occur in bars 0/8A, 1, 4, and 8B/16A of the Minuet, in bar 23/37 of the Trio, and throughout the Coda. With four repeats of the Minuet and two of the Trio in most performances, there were generally four independent measures available for each of the four sixteenth-note occurrences in the Minuet and for the single occurrence in the Trio (the latter really being two similar occurrences, each repeated twice). For the Coda, of course, only a single set of measurements was available for each artist, but there were 11 occurrences of sixteenth-notes.

The measurements were performed as previously, but with less of the waveform (about 2 s) displayed at a time, to reduce measurement error. In each of the relevant bars, the onset–onset interval durations of the dotted eighth-note and the following sixteenth-note were measured. The relative duration of the sixteenth-note onset–onset interval following a dotted eighth-note was computed as the ratio of its measured duration to its "expected" duration, times 100. The expected duration was one-fourth of the sum of the dotted eight-note and sixteenth-note onset–onset interval durations.

2. The four-beat Beethoven pulse

Clynes' four-beat Beethoven pulse is defined as [106, 89, 96, 111] (Clynes, 1983). Thus it implies that a sixteenth-note following a dotted eighth-note should be 11% longer than its expected value, and that this difference (with due allowance for random variability and other expressive demands of the piece) should be present throughout a Beethoven composition. The aim of the present measurements was to examine these predictions.

To rule out any possible misunderstanding, the computer performance synthesized by Clynes was measured to determine the relative lengths of the sixteenth-notes in it. Since that performance did not contain any repeats, only two independent measurements (on identical renditions) were available for the four occurrences of sixteenth-notes in the Minuet and the one in the Trio. These two measurements were in close agreement; the average absolute difference due to measurement error was 1.2 percentage points. The average relative onset–onset interval durations of the five sixteenth-notes were: 107.5, 108, 103.5, 110.5, 111.5. The last two are in agreement with Clynes' specifications, but the first three are smaller, the third (bar 4) conspicuously so. A similar impression was obtained from the Coda; the 11 values were: 106, 102, 110, 112, 99, 111, 110, 108, 115, 117, 147. The first, second, and fifth values are clearly smaller than expected, and the last three are larger, although the last one obviously includes a manual adjustment, the final ritard. The origin of this variability in Clynes' own creation remains unclear. The measurements nevertheless illustrate the postulated lengthening of sixteenth-notes following dotted eighth-notes.

3. Results

The relative durations of the sixteenth-notes in the 19 human performances are presented in Table V. Each percentage for bars 0/8A, 1, 4, 8B/16A, and 23/37 represents an average across four (sometimes fewer) repeats; for the Coda, a single average was computed from the eight values of bars 16B–44, but separate values are reported for the three occurrences in the final bar (45a, b, c).

Consider first the four occurrences in the Minuet (the first four columns in Table V). There are very large individual differences among the artists, particularly in the initial upbeat, bar 0/8A. Some pianists (Ashkenazy, Berman, Rosen, and Frank) show a considerable elongation of the sixteenth-note in that beat, while Schnabel, the great Beethoven authority, played it much shorter than notated. Many others are close to precise timing. Although not shown in the table, a number of artists (most notably Ashkenazy, Backhaus, Frank, Kempff, Perahia, and Rosen) prolonged the sixteenth-note much more in the repeat (bar 8A) than in the first playing (bar 0). Despite considerable variability, individual differences tended to be maintained throughout the Minuet. Some of the variability, of course, may have been intended by the artist. It is perhaps noteworthy that the pianists with rather short sixteenth-notes are mostly of German extraction, while those with prolonged sixteenth-notes are Russian or American.

TABLE V. Relative onset–onset interval durations of sixteenth-notes following dotted eighth-notes (in percent).

	Minuet				Trio	Coda			
Bar	O/8A	1	4	8B/16A	23/37	16B-44	45a	45b	45c
Arrau	109.5	120.3	111.2	119.0	87.8	89.4	100.0	104.0	144.0
Ashkenazy	145.8	120.8	108.3	126.8	85.8	98.8	107.0	104.0	185.0
Backhaus	101.3	92.8	84.0	94.0	83.3	85.5	87.0	110.0	152.0
Berman	129.8	126.5	110.8	131.8	84.8	87.5	84.0	88.0	127.0
Bishop	101.8	111.3	119.8	108.0	82.0	86.6	108.0	123.0	143.0
Brendel	103.8	90.5	124.8	94.8	79.3	90.6	109.0	115.0	125.0
Davidovich	98.5	102.0	108.3	100.3	85.5	92.1	111.0	104.0	116.0
Frank	125.8	112.5	103.0	123.5	86.5	77.8	86.0	101.0	168.0
Gieseking	101.8	101.3	96.5	96.5	101.3	100.1	113.0	113.0	142.0
Gilels	94.5	104.8	105.3	85.5	81.5	80.6	92.0	106.0	130.0
Gould	105.5	112.8	100.5	102.8	81.0	70.6	77.0	84.0	112.0
Gulda	98.3	106.0	94.8	107.3	84.5	93.9	118.0	106.0	164.0
Haskil	108.5	107.8	114.5	92.5	93.3	87.1	86.0	84.0	133.0
Kempff	98.8	100.8	110.0	108.8	81.8	83.4	94.0	98.0	123.0
Perahia	115.5	120.8	126.8	103.0	92.8	97.8	127.0	134.0	199.0
Rosen	129.0	142.3	134.3	121.8	100.3	86.5	96.0	96.0	119.0
Rubinstein	102.0	104.5	107.0	96.5	79.8	87.9	95.0	100.0	149.0
Schnabel	78.3	94.0	99.5	89.3	90.5	78.8	60.0	69.0	116.0
Solomon	101.5	122.0	108.8	105.0	80.8	89.8	90.0	88.0	135.0
Computer	107.5	108.0	103.5	110.5	111.5	107.3	115.0	117.0	147.0

The Minuet data were subjected to a repeated-measures analysis of variance with two fixed factors, bars and repeats.[5] The main effect of bars was not significant, showing that the artists as a group did not systematically modulate their characteristic sixteenth-note timing patterns within the Minuet. The main effect of repeats was marginally significant, $F(3,54) = 3.20$, $p = 0.0304$, due to a tendency for longer sixteenth-notes in the second repeat. However, this tendency occurred only in bars 0/8A and 8B/16A, whereas bars 1 and 4 showed the opposite pattern, with longer sixteenth-notes on the first playing. This difference was reflected in a significant bars by repeats interaction, $F(9,162) = 2.63$, $p = 0.0074$.

On the whole, despite all the individual variations, there was a tendency to prolong the sixteenth-notes in the Minuet, as predicted by Clynes' Beethoven pulse. The grand average percentage deviation was 108.1, which is not too far from Clynes' 111. The situation is very different, however, for the Trio and the Coda. In the Trio (bar 23/37), there was an overwhelming trend to shorten the sixteenth-note. Only two pianists (Gieseking and Rosen) played the note with its literal value. The average percentage was 86.4, which strongly contradicts the value of 111 implemented in Clynes' synthesis. Similarly, throughout the Coda (bars 16B–44), except for the last bar, there was a strong tendency to shorten all sixteenth-notes. A few pianists (Gieseking, Ashkenazy, and Perahia) played almost literally, but none showed any lengthening. The average percentage value was 87.6, which again contrasts with Clynes' nominal 111. The first four occurrences tended to be somewhat longer than the second four, which was reflected in a significant difference among the eight values in bars 16B–44, $F(7,126) = 2.73$, $p = 0.0114$. In the final bar (bar 45), some pianists started to lengthen the relative duration of the sixteenth-notes in the first two beats, and the very last occurrence (45c) showed a substantial lengthening in all performances, due to the final ritard. However, the average percentages for the first two occurrences in bar 45, 96.8 and 101.4, are still far below Clynes' nominal 111, while the last occurrence, with an average of 141.2, comes close to Clynes' deliberately extended 147.0.

In summary, these results offer some support for Clynes' specifications during the Minuet, although there were large individual differences, and the most deviant pianists included some of the most renowned Beethoven players (Schnabel, Backhaus, Brendel). The Trio and Coda performances, however, flatly contradict Clynes' Beethoven pulse. It appears that different musical requirements call for radically different temporal microstructures, even for the same composer and within the same piece.

B. Eighth-notes

1. Measurement procedure

Eighth-notes are present throughout the Minuet and form a continuous background movement against which the melody unfolds. However, the onsets of single accompanying eighth-notes were often very difficult to detect in the background noise, and complete measurements did not promise enough information to seem worth the effort. Therefore, eighth-notes were measured only in two selected bars, bars 10 and 12, where they formed part of the melody (cf. Fig. 1). Each of these bars contains three pairs of eighth-notes, with usually four repeats per performance. The relative onset–onset interval durations of each pair were expressed simply as their ratio, with the second onset–onset interval in the numerator.

2. The two-beat Beethoven pulse

For two eighth-notes, the pulse is reduced to a two-beat pattern, [97, 103].[6] Thus the second note in any pair of eighth-notes nested within a quarter-note beat is predicted to be 3% longer than expected, and the ratio between the two onset–onset intervals should be $103/97 = 1.06$.

The three pairs of onset–onset interval durations were measured very carefully in bars 10 and 12 of the computer performance, and their ratios were computed. They were 1.11, 1.12, and 1.17 in bar 10; and 1.08, 1.06, and 1.24 in bar 12. It is evident that, in each bar, the last ratio was considerably larger than prescribed by the pulse specification. The first two ratios in bar 10 also appear too large. The cause of these deviations is unknown; they may represent further personal interventions by Clynes to improve the quality of the computer performance beyond that imparted by the constant pulse pattern.

3. Results

The results for the human performances are easily summarized. The ratio data for the 15 pianists who observed all the repeats were subjected to a repeated-measures analysis of variance with the factors distant repeats (before versus after the Minuet), immediate repeats, bars (bar 10 versus bar 12), and beats (1, 2, 3). There was a single, highly significant effect: the main effect of beats, $F(2,28) = 56.03$, $p < 0.0001$. It was due to an extremely consistent tendency to lengthen the second eighth-note on the third beat, but not on the first and second beats. The three average ratios were: 1.01, 1.01, and 1.18. None of these ratios matches the value of 1.06 predicted by Clynes' specifications; in particular, the first two beats do not show the predicted relative lengthening of the second eighth-note. The substantial lengthening of the last eighth-note in each bar is in agreement with the pattern implemented in the computer performance, but not with the Beethoven pulse.

The consistency of the observed pattern across pianists, including the four not included in the analysis of variance, was striking. Only two pianists did not show the increased onset–onset interval ratio on the third quarter-note beat: Solomon (who played close to mechanical evenness) and Gould (who tended toward slightly positive ratios on all three beats, not unlike the predicted Beethoven pulse). A few pianists (Schnabel, Brendel, Ashkenazy, and Bishop) tended to lengthen the second eighth-note onset–onset interval on the first beat as well, and some (Berman, Gulda, and Ashkenazy) showed some lengthening of the second eighth-note on the second beat, while one (Schnabel) showed some shortening.

On the whole, the eighth-note data show once more that timing relationships are not constant but vary according to local musical requirements. The musical factors that caused lengthening of the last eighth-notes in bars 10 and 12 are phrase-finality (as indicated by the slurs in the score) and, perhaps more importantly, emphasis of the following note (Clarke, 1988, principles 1 and 2).

IV. PERFORMANCE EVALUATION BY LISTENERS

This part of the study had two purposes: first, to get musical listeners' impressions of the "Beethovenian" quality of the performances and to see whether these ratings bear any relationship to the timing measurements, particularly the "Beethoven pulse" factor (factor 3) in the principal components analysis; second, to explore in a very preliminary way the psychological dimensions along which musical performances vary. This part of the study was exploratory in nature, and its methodology did not include all the controls one would want to include in a full-scale study of human performance evaluation. Nevertheless, it yielded some interpretable results.

A. Subjects

Nine subjects participated, all of whom had extensive experience with classical music, played the piano with varying degrees of proficiency, and knew the piece well. They included a senior professor of piano at a major music school; a young professional pianist who frequently concertized in the area; an experienced piano teacher at a community music school; two doctoral students of musicology with a special interest in performance; a distinguished professor of phonetics with a long-standing interest in rhythm; an old lady (the author's mother) with a life-long interest in music performance; and two psychologists with a strong interest in music (the author being one of them). Four of the listeners were of European origin (two Austrians, one German, and one Estonian); the others were American. Four judges were female; five were male.

TABLE VI. The adjective pairs used in performance evaluation.

Beethovenian	un-Beethovenian
Fast	slow
Expressive	inexpressive
Relaxed	tense
Superficial	deep
Cold	warm
Powerful	weak
Serious	playful
Pessimistic	optimistic
Smooth	rough
Spontaneous	deliberate
Consistent	variable
Coherent	incoherent
Sloppy	precise
Excessive	restrained
Rigid	flexible
Effortful	facile
Soft	hard
Realistic	idealistic
Usual	unusual

B. Procedure

Twenty adjective pairs were selected by the author with the intention of capturing dimensions relevant to the judgment of performance variations. They are shown in Table VI.[7] A seven-point rating scale extended between each pair of polar opposites. The most important adjective pair for the present purpose was the first one (Beethovenian/un-Beethovenian), which was explained in the instructions as follows: "This judgment should reflect to what extent the performance 'captures the Beethoven spirit'—that is, whether the general expressive quality of the performance is what the composer may have intended." The second adjective pair concerned the perceived tempo of the performance. The remaining adjective pairs included some deliberately selected because they were thought to correlate with the "Beethovenian" quality; for example, Clynes (1983) characterizes the Beethoven pulse as "restrained." The order of the adjective pairs and the assignment of their members to the ends of the rating scale were fixed in this exploratory study. The lower end of the rating scale represented the more positive connotation for some adjective pairs, the more negative one for others.

Each subject was given a booklet containing 20 identical answer sheets, one for each performance, preceded by detailed instructions that included brief definitions of each adjective pair. The instructions emphasized that the musical, not the sonic quality of each performance was to be judged. They gave the option of entering two separate judgments for Minuet and Trio on each scale; that option was rarely taken, and if it was, the two judgments were averaged in the following data analysis. A few missing judgments were replaced by "4."

All subjects listened at home on their own audio equipment to a cassette tape of the 20 performances, stopping the tape after each performance to enter their judgments. The 20 performances were in the same fixed order for each subject. The first performance was the computer rendition, both to get an immediate reaction to it and to remind the subjects of the music. The order of the human performances was: Gilels, Gulda, Ashkenazy, Arrau, Schnabel, Haskil, Berman, Kempff, Backhaus, Gieseking, Rosen, Frank, Perahia, Bishop, Brendel, Gould, Davidovich, Solomon, and Rubinstein. Performances with missing repeats and/or poor sound quality tended to occur toward the end.[8]

C. Analysis

The raw data constituted a 9 (subjects) $\times$ 20 (performances) $\times$ 20 (adjective pairs) matrix of numerical ratings. Since there was considerable variability between judges' ratings, each adjective scale was first examined as to whether there was any consistency among subjects at all. The criterion was that Kendall's coefficient of concordance be significant ($p < 0.05$). Three adjective scales (serious/playful, effortful/facile, realistic/idealistic) did not meet this criterion and were eliminated. Since two performances, by the computer and by Gould, often elicited the most extreme ratings, the criterion was applied once more with these two performances omitted. Five additional scales (expressive/inexpressive, cold/warm, spontaneous/deliberate, coherent/incoherent, rigid/flexible) were eliminated at this stage. This left 12 scales for further analysis. The highest coefficient of concordance for all 20 performances (0.59, $p < 0.0001$) was shown, not surprisingly, by fast/slow, followed by three scales (smooth/rough, excessive/restrained, Beethovenian/un-Beethovenian) with coefficients of 0.33–0.34 ($p < 0.0001$); the remainder had low but still significant Kendall coefficients. It was gratifying to find that the important Beethovenian/un-Beethovenian scale was used with some consistency by the subjects.

The data were subsequently averaged across the nine subjects' judgments, which resulted in a 20×12 data matrix. The 12×12 intercorrelation matrix was subjected to a principal components analysis with Varimax rotation, to reduce the 12 rating scales to a smaller number of evaluative dimensions.

D. Results

1. Factor structure of rating scales

The analysis yielded four factors with eigenvalues larger than 1; together, they accounted for 88% of the variance in the data. After Varimax rotation, that variance was divided fairly equally among the four factors. The factor loadings of the 12 rating scales (rearranged) are shown in Table VII; loadings smaller than 0.25 have been suppressed for the sake of clarity. The polarities of factors 3 and 4 have been reversed for easier labeling.

The factors can be interpreted without much difficulty. Factor 1 has its highest loadings on "hard," "tense," and "rough," as opposed to soft, relaxed, and smooth. It will be called *force*. Factor 2 loads highly on "excessive," "variable," "unusual," and "sloppy," as opposed to restrained, consistent, usual, and precise. It will be dubbed *individuality*. Factor 3 is characterized by the attributes "deep," "strong," and "Beethovenian," as opposed to superficial, weak, and un-Beethovenian. Interestingly, this factor reveals depth and strength (but not restraint) as the primary correlates of the subjects' idea of "Beethovenian." It will be called *depth*. Finally, factor 4 loads highly on "fast" and "optimistic," as opposed to slow and pessimistic. Clearly, it is a tempo factor, and it is interesting that the listeners associated optimism so strongly with a fast tempo. It will be called *speed*. Note that only depth has a simple relationship to positive/negative evaluation or preference; force, individuality, and speed most likely have a curvilinear relationship, with neither extreme being desirable.

TABLE VII. Sorted rotated factor loadings. Positive loadings represent the second adjective in a pair. Loadings smaller than ± 0.25 have been omitted.

	Factor 1	Factor 2	Factor 3	Factor 4
Soft/hard	0.935			
Relaxed/tense	0.900			
Smooth/rough	0.701	0.463	−0.356	−0.348
Excessive/restrained	−0.349	−0.830		
Consistent/variable		0.814		0.410
Usual/unusual		0.730	−0.372	−0.415
Sloppy/precise	−0.479	−0.701	0.279	
Superficial/deep	−0.268		0.900	
Strong/weak	−0.273		−0.887	
Beet/un-Beethovenian	0.257		−0.771	−0.444
Fast/slow				−0.925
Pessimistic/optimistic				0.842

TABLE VIII. Factor scores of the 20 performances.

	Factor 1	Factor 2	Factor 3	Factor 4
Arrau	−1.385	0.280	−0.657	−2.199
Ashkenazy	−0.283	−0.481	−0.873	0.563
Backhaus	0.511	2.014	−1.024	1.341
Berman	1.719	0.251	0.680	−0.325
Bishop	0.662	−0.725	0.507	1.254
Brendel	1.172	−1.050	0.479	−0.075
Davidovich	−0.559	−0.253	0.657	0.194
Frank	−0.658	0.002	1.332	0.576
Gieseking	−0.103	−0.657	−2.290	1.157
Gilels	−1.318	−0.204	0.305	−1.378
Gould	2.262	1.472	−0.313	−1.727
Gulda	−0.413	−0.654	0.291	1.261
Haskil	−0.418	−0.123	0.506	−0.162
Kempff	−1.082	0.147	−0.783	−0.168
Perahia	0.033	−0.300	1.800	0.245
Rosen	0.348	0.220	0.460	−0.728
Rubinstein	−1.503	1.449	0.517	0.131
Schnabel	0.060	1.677	−0.285	0.943
Solomon	0.523	−1.426	0.479	−0.115
Computer	0.433	−1.639	−1.785	−0.789

2. Factor scores of performances

Let us examine now how the individual performances ranked on these evaluative dimensions, and what characteristics of the performances or of the performers might be responsible for these rankings. The factor scores are shown in Table VIII. On the force factor, the highest scores were exhibited by Gould, Berman, and Brendel; the lowest scores, by Rubinstein, Arrau, Gilels, and Kempff. It is conceivable that this factor was influenced by the relative loudness and sonic quality (e.g., "harshness") of the recordings, which were not controlled in any way; however, the author, having taken notes about these aspects of the recordings, sees no obvious relationship. More likely, some acoustic correlate of the pianists' "touch" was involved, such as their degree of legato playing or amplitude dynamics, which were not measured in the present study. Interestingly, a loose relationship with the artists' age is suggested, older artists tending to have negative scores (i.e., less force). Also, the two women (Davidovich and Haskil) have moderately negative force scores. To the extent that extremes are to be avoided along the force dimension, Perahia and Schnabel obtained the most desirable scores on this factor (close to zero).

The highest scores by far on the individuality factor were shown by Backhaus, Schnabel, Gould, and Rubinstein; the lowest scores, by the computer performance, Solomon, and Brendel, followed by Bishop, Gieseking, and Gulda. Several of these performances with negative scores were relatively deadpan (computer, Solomon, and Gieseking); the others were probably without highly distinctive properties. The four most individual performances, on the other hand, were indeed so: Backhaus introduced striking tempo changes, Schnabel used quirky articulation, Gould was unusually slow and plodding, and Rubinstein used exaggerated expression. Either extreme was avoided most effectively by Frank, Haskil, and Kempff, who scored in the middle range. Given the many different ways in which a performance can be individual, it seems unlikely that a single physical correlate of this dimension could be found.

The depth factor is of special interest here because it represents the listeners' concept of "Beethovenian." The highest score was obtained by Perahia, followed by Frank; these two scores were far ahead of the rest. The lowest scores were shown by Gieseking and the computer performance. Since the latter exhibited Clynes' Beethoven pulse most clearly, it is apparent that depth scores do not have a positive relationship with the presence of such a pulse.

Finally, the speed factor clearly contrasted fast performances (Backhaus, Gulda, and Bishop) with slow ones (Arrau, Gould, and Gilels). Here, indeed, there was a straightforward physical correlate: The correlation coefficient of the factor scores with the computed metronome speeds (Table II, column B) was 0.78 ($p<0.001$), just slightly below the correlation between the average ratings on the fast/slow scale itself and the metronome speeds (0.83).

In an attempt to identify possible correlates of the four evaluative dimensions in the timing patterns of the performances, correlations were computed between the factor scores just discussed (Table VIII) and the factor loadings of the 20 performances in the earlier analysis of the timing data (Table IV). It should be kept in mind that the timing data derived from the Minuet only, while the evaluations were based not only on the complete performances, but also on many other aspects besides timing. Nevertheless, there were several significant correlations. Individuality correlated negatively ($-0.61, p<0.01$) with timing factor 3, which represented the V-shaped pulse; it will be recalled that the computer performance and several other deadpan performances ranked lowest on individuality, which is quite reasonable. Depth correlated positively ($0.62, p<0.01$) with timing factor 1, which represented mainly the marking of major phrase boundaries, but not (-0.14) with timing factor 3 (the Beethoven pulse). Finally, speed correlated positively (0.63, $p<0.01$) with timing factor 2, which represented expressive deviations in bars 1 and 6 as well as a faster tempo for the second half of the Minuet. It also correlated negatively ($-0.45, p<0.05$) with timing factor 1, suggesting that the slower performances tended to emphasize phrase boundaries more, whereas the faster performances tended to focus more on certain other expressive deviations. Force did not correlate significantly with any of the three timing factors.

3. Performance styles

The data were analyzed in yet another way, by transposing the 12×20 matrix of judgments and conducting another principal components analysis. In this case, the correlations were computed across the 12 rating scales and represented the similarities between the rating "profiles" of the 20 performances. (Because there were only 12 scales, the 20×20 correlation matrix was singular and of rank 11.) Five factors (with eigenvalues greater than 1) accounted for 91% of the variance. After Varimax rotation, the first factor accounted for 38% of the variance, with the remainder divided about equally among the other four factors. The factor loadings of the 20 performances are shown in Table IX.

One way of interpreting this structure is that, by means of 12 rating scales (or four evaluative dimensions), the judges were able to distinguish five "performance styles." The first factor seemed to reflect a general performance standard; 10 pianists (Frank, Davidovich, Perahia, Gulda, Bishop, Haskil, Solomon, Brendel, Ashkenazy, and Rubinstein) had high positive loadings, and only one (Gould) had a negative loading. The other four factors seemed to reflect more individual interpretive styles: Factor 2 was represented primarily by Schnabel and Backhaus; factor 3 by Berman, Rosen, Gould, and Brendel; factor 4 by Arrau, Gilels, and Kempff; and factor 5 by Gieseking, the computer, and Ashkenazy. (Note that only two pianists, Brendel and Ashkenazy, have relatively high loadings on more than one factor.) Comparison with Table VIII reveals that factors 2 and 5 reflect high and low individuality, respectively; factors 3 and 4 reflect high and low force, respectively; and factor 1 represents the "middle of the road" performances. Note, however, that all factors were strictly orthogonal. This might be interpreted as indicating that the adjective pairs that defined individuality and force did not constitute true polar opposites but different dimensions. This is quite plausible in the case of individuality, which really represents deviations from the norm in different directions and by different means.

TABLE IX. Sorted rotated factor loadings of the 20 performances on five "performance style" factors that emerged from the principal components analysis of their rating "profiles." Loadings smaller than ±0.25 have been omitted.

	Factor 1	2	3	4	5
Frank	0.944				
Davidovich	0.941				
Perahia	0.932				
Gulda	0.892				0.307
Bishop	0.858				0.279
Haskil	0.815	−0.456			
Solomon	0.789	−0.437			0.296
Brendel	0.642	−0.290	0.623		
Rubinstein	0.598	0.333	−0.283	0.473	−0.358
Schnabel		0.945			
Backhaus		0.880		−0.305	
Berman			0.888		
Rosen	0.397		0.677	0.291	
Gould	−0.627	0.323	0.660		
Arrau				0.918	
Gilels	0.439			0.817	
Kempff	0.398		−0.474	0.616	0.330
Gieseking			−0.308		0.905
Computer		−0.438		0.301	0.773
Ashkenazy	0.610				0.619

V. GENERAL DISCUSSION

The present study addressed three broad issues: (1) the presence or absence of a "Beethoven pulse" in human performances; (2) general characteristics of expressive timing patterns in different expert performances of the same piece; and (3) listeners' evaluation of performances. These topics will be discussed in turn.

A. The search for the Beethoven pulse

One motivation of the present study was to search for a pulselike timing pattern in human performances of a Beethoven piece, similar to that discovered subjectively by Clynes (1983) and implemented in his computer performance of the same piece. Before summarizing the outcome of that search, some limitations and strengths of the study should be pointed out.

One severe limitation is obviously that only a single composition was examined. It could be that the composition chosen is not typical of Beethoven or that its minuetlike character dominated specifically Beethovenian characteristics. In the author's opinion, however, the piece is quite characteristic and not very minuetlike, and its choice was not inappropriate. A more serious problem of restricting the investigation to a single composer is that any timing patterns found, even if they resemble the Beethoven pulse, may not be specific to Beethoven but may be general features of music performance.

A second limitation is that only timing patterns were examined. It could be that accent (amplitude) patterns interact with timing variations, so that considering the timing pattern alone might present a distorted picture. Nevertheless, an examination of the timing component by itself is a defensible methodological strategy, for which there are numerous precedents in the literature. Of course, no conclusions are warranted with regard to the amplitude component of Clynes' Beethoven pulse, which may or may not have been present in the performances examined here.

Third, the present study focused primarily on the "higher level" pulse, at the level of quarter-notes, because of the paucity of sixteenth-notes in the piece. Nevertheless, some information relevant to the "basic," lower level pulse was also obtained.

Balanced against these limitations are a number of strengths of the present research. The study employed a large sample of performances and obtained a large number of measurements from each. The precision of measurement was more than sufficient for its purpose. The artists represent many of the finest interpreters of Beethoven's music in this century. The data were subjected to rigorous statistical analysis. Thus, within the limitations stated above, the search for a Beethoven pulse was fairly exhaustive. Finally, results from an earlier perceptual study (Repp, 1989) indicated that Clynes' Beethoven pulse improved the computer performance of the piece chosen; this provided a valid basis for expecting similar patterns in human performances.

Was a Beethoven pulse found? Certainly, none of the 19 human performances showed a timing pattern closely resembling that of the computer performance. More specifically, none of the human performances showed any relatively constant, pulselike timing pattern; rather, the timing pattern varied from bar to bar according to musical demands. Thus, in general, human performances not only did not show the specific Beethoven pulse "discovered" by Clynes; they did not show *any* constant pulse.

Consistent with this observation is the finding that the computer performance, even though it had been rated favorably by listeners in comparison to a deadpan performance (Repp, 1989), did not fare well in comparison with human performances. It received the lowest rating on the "Beethovenian" scale, even lower than Gould's leaden performance; in terms of the evaluation factors, it scored lowest on individuality and second lowest on depth. Clearly, a pulselike timing pattern is not very favorably received by listeners when the alternatives are musically varied timing patterns.

Despite a number of manual adjustments by Clynes, the computer performance did not exhibit the full richness of a human performance; it was still an artificial performance generated for the purpose of testing the effectiveness of a single, isolated microstructure component. Since there are so many other sources of expressive variation that perturb the surface timing pattern of human performances, the composer's pulse may be "underlying," hidden, or intermittent. The way in which an underlying pulse might combine with other sources of timing variation, or how a listener might perceive a constant pulse through a pattern of surface variability, are issues that Clynes has not discussed explicitly. The present data are not incompatible with the presence of an underlying V-shaped pulse pattern in some of the human performances of the Minuet. The origin and interpretation of that pattern remain uncertain, however. For example, it may just as well derive from structural musical factors than from an underlying pulse: The harmonic and melodic structure of the Minuet tended to weight the first and third quarter-notes in a bar more heavily than the second. Moreover, there may well be a general tendency of performers to shorten the second of three quarter-notes, which is not specific to a particular piece or composer. Indeed, Gabrielsson *et al.* (1983) found lengthening of either the first or the third beat, but never of the second beat, in musicians' performances of folktunes in 3/4 time.

Even if there was some underlying pulse, it apparently contributed little to the impression of a performance on listeners. That impression, to the extent that it derived from the timing pattern, was governed primarily by pianists' temporal marking of major and minor phrase boundaries, and of one prominent melodic excursion (bar 6).

Similar conclusions apply to the basic, lower level pulse, even though there were less data available. The relative timing of sixteenth- and eighth-notes was clearly dependent on the musical context, and no pulselike constancy was evident. In the Trio and the Coda, the observed timing patterns of sixteenth-notes strongly contradicted the ratios specified by Clynes.

The fact that the composer's pulse, as conceived by Clynes, is inherently insensitive to the musical structure of a specific piece seems to preordain a minor role for it, if any, in music performance and evaluation. Since expressive variations based on structural characteristics of the music are large and almost ubiquitous, there is little room for an autonomous pulse to come to the fore. Wolff (1979, p. 15), transmitting Artur Schnabel's views, has made this pertinent comment: "The term 're-creation' has often given rise to the misunderstanding that the interpreter can attempt a revival of the personality of the composer at the moment of creation. The futility of such attitudes is generally acknowledged. We know that all interpretive re-creation depends on the awareness of the structure and objective character of a composition." The presence of a composer's personal pulse in expert performances remains to be demonstrated.

B. Expressive timing in performance

Three general observations can be made from the present data. First, far from being idiosyncratic, the timing patterns of individual artists' performances largely adhere to a common standard. This was demonstrated by the fact that, in the analysis of the Minuet timing patterns, the first two principal components accounted for 71% of the variance. One of these components primarily represents the lengthening at phrase boundaries (a well-known phenomenon; see, e.g., Todd, 1985; Shaffer and Todd, 1987; Clarke, 1988), whereas the other reflects several other types of expressive variation, such as lengthening of salient melodic inflections and tempo changes within and between sections. That the timing variation appears to be governed by two independent dimensions is a finding worth following up in future research. Individual variations consist primarily in the extent to which the structural markers captured by these two dimensions are applied. There seems to be relatively little room for truly idiosyncratic variation in relative timing, at least in the present Minuet, but overall tempo, accent patterns, and articulation offer many additional degrees of freedom to the individual performer.

Second, it is evident that the timing pattern is reproduced with a high degree of precision across repetitions of the same music. This was already noted by Seashore (1938), as well as by others in more recent research (e.g., Gabrielsson, 1987; Palmer, 1989; Shaffer and Todd, 1987). Although some compositions may call for subtle variations between repeats, in the present Minuet there was no evidence of any systematic timing changes, except at the beginnings and ends of sections, where there was either an actual change in the music or the distinction between continuity and finality had to be conveyed.

Third, the timing patterns at all levels are dependent on the musical structure. Thus not only the timing of quarter-notes but also that of eighth- and sixteenth-notes varied substantially with their musical function. For example, sixteenth-notes following dotted eighth-notes were generally prolonged in the Minuet, where they were part of an upbeat, but generally shortened in the Trio, where they fell on the downbeat. The same rhythmic pattern can be performed with very different temporal modulations, according to musical requirements (cf. Gabrielsson *et al.*, 1983).[9]

C. Performance evaluation

The systematic description and evaluation of different performances of the same music have been studied relatively little by psychologists, compared to the considerable literature on listeners' reactions to different compositions. Yet, music critics, jurors at competitions, and discriminating music lovers engage continuously in such judgments that, despite a considerable amount of subjectivity, are by no means totally idiosyncratic. The present results, although they are very preliminary, do suggest that there is some consistency among judges; moreover, part of the variability may well result from lack of skill in using rating scales rather than from genuine judgmental differences.

The individuality factor obtained in the analysis of the rating scales, as well as the "middle-of-the-road" general factor obtained in the analysis of the performances as judged, suggest that musically experienced listeners refer to similar internal performance standards. This common standard is most likely one that includes the basic expressive variations required by the musical structure, without which a performance would be perceived as atypical, impoverished, and unmusical. While there is rarely a single definitive performance of a given piece of music, there can well be a typical or average performance that listeners agree on. It is with respect to their evaluation of deviations from this common standard that listeners show differences of opinion. If the deviations are gross, their reactions may be uniformly negative. (An example is Gould's performance in the present set, although its deviations, apart from its slow tempo, were mostly in aspects other than timing.) However, if the deviations are imaginative, listeners' evaluative judgments may diverge considerably. (Examples in the present set are the performances by Schnabel and Backhaus.) While such performances are stimulating and provide food for discussion, those that come close to the listener's internal standard simply "sound right," and the listener resonates to them. (Perahia and Frank perhaps came closest to that ideal.)

While the individuality dimension obtained in the present analysis implies reference to a conventional standard and thus may be peculiar to the evaluation of cultural artifacts, the other three dimensions—force, depth, and speed—are clearly related to the three traditional dimensions of the semantic differential—potency, evaluation, and activity—which have also been obtained in studies employing varied musical materials (see, e.g., de la Motte-Haber, 1985). This result is very encouraging, as it suggests that effective rating scales for the formal comparison and evaluation of musical performances could be developed.

ACKNOWLEDGMENTS

This research was supported by BRSG Grant RR-05596 to Haskins Laboratories. I am indebted to the record library of the Yale School of Music for providing most of the recordings analyzed here, to Larry Raphael for supplying two additional ones, and to Vin Gulisano for doing the transfers to cassette tape for me. Valuable advice in the course of this research was provided by Manfred Clynes and Robert Crowder, and helpful comments on an earlier version of the

manuscript were obtained from Alf Gabrielsson and Caroline Palmer.

[1]A few words about the artists may be in order for readers not acquainted with them. The sample includes 2 women (Davidovich, Haskil) and 17 men. Nine pianists (Backhaus, Gieseking, Gilels, Gould, Haskil, Kempff, Rubinstein, Schnabel, and Solomon) are dead; the others are still active on the world's concert stages at the time of writing. Schnabel is universally considered an authoritative interpreter of Beethoven's sonatas; Backhaus, Brendel, and Kempff are generally considered Beethoven specialists as well. Other pianists who have played much Beethoven and have recorded the complete sonatas include Arrau, Frank, and Gulda. Since almost every major pianist plays Beethoven sonatas frequently, all of the remaining artists certainly have much experience as Beethoven interpreters, although some are known better from other repertoire. Thus Gould is known primarily for his Bach recordings and often eccentric performances; Haskil is considered a Mozart specialist; Berman is associated with Romantic virtuoso pieces; Gieseking and Rubinstein, although they frequently played Beethoven, are generally not considered ideal interpreters of his music (the former being too fleet, the latter too effusive). The different ages of the artists at the presumed time of recording may also be noted; they range from rather young (Bishop, Gulda, and Perahia) to rather old (Arrau, Backhaus, Kempff, and Rubinstein). Finally, the artists may also be divided into three major national groups: German-Austrian (Backhaus, Brendel, Gieseking, Gulda, Haskil, Kempff, Schnabel, as well as the Chilean-born but German-educated Arrau), Russian (Ashkenazy, Berman, Davidovich, and Gilels), and American-Canadian [Bishop, Frank (of German origin), Gould, Perahia, and Rosen]; this leaves only the British pianist Solomon, and the very cosmopolitan Polish-born Rubinstein. It will be of interest to see whether any of these characteristics are related to expressive timing patterns in music performance.

[2]To guard against variations in speed caused by the multiple transfers of the recordings using a variety of playback equipment, the sound wave corresponding to the initial note of each performance (a B-flat) was subjected to spectral analysis. The FFT spectrum was calculated over a 102.4-ms Hamming window placed roughly over the center of the waveform, and the frequency of the lowest harmonic was determined with a resolution of 4 Hz. Each of the 20 recordings, including the computer performance from Clynes' laboratory, yielded a fundamental frequency of 244 Hz $\pm$ 4 Hz (3 +, 4 −), so that the recording/playback speeds may be considered comparable. The average frequency of the first note, however, was higher than expected. B-flat, being one semitone above the A one octave below the standard A of 440 Hz, should be 5.9% higher than 220 Hz, or at about 233 Hz. Thus it seems that all recordings were played somewhat fast, almost one semitone too high. However, since this difference could not be traced directly to any piece of equipment, no correction was made in the measurements.

[3]It is interesting to compare these metronome values to various opinions about what the tempo *should* be. Urtext editions of Beethoven's piano sonatas (Breitkopf and Härtel, Kalmus, Peters) do not have metronome markings. A perusal of a large number of other editions, however, revealed several in which the editors had inserted their own suggested metronome speeds. For the Minuet, they range from 84 qpm (d'Albert) to 88 qpm (Bülow/Lebert) to 96 qpm (Epstein, Schnabel) to 104–108 qpm (Casella). Schnabel, in addition, gives separate, faster metronome indications for the Trio (108 qpm) and the Coda (100 qpm), whereas Casella indicates that the Trio should be taken slower (96 qpm) than the Minuet. These metronome speeds are quite fast compared with the present performances. Only the slowest marking, by d'Albert, is close to the average speed taken by this group of pianists. A number of pianists, including Schnabel, play the Trio faster than the Minuet, as suggested in the Schnabel edition. However, Schnabel's tempi fall short of his own recommendations; only Backhaus comes close to those.

[4]Actually, the specification was [105, 88, 107], but the deviations were "attenuated" by 50%, according to Clynes' musical judgment in generating the computer performance. Note that the possibility of varying the modulation depth of a pulse defines a pulse family, rather than a single fixed pattern, for a given composer. However, Clynes usually applies attenuation only to the higher-level (slower) pulse.

[5]All 19 performances were included in this analysis. In the four performances that were missing the fourth repeat, the data of the second repeat were duplicated.

[6]The two-beat pulse pattern is obtained from the four-beat values "by adding the duration of tones 1 and 2, and of tones 3 and 4, respectively, to obtain the duration proportions" (Clynes 1983, p. 161). If applied to Clynes' four-beat pulse, this yields [97.5, 103.5] due to the fact that percentage deviations above and below 100 are not perfectly balanced in Clynes' four-beat specification, perhaps due to omission of decimals. Since the percent increase of one eighth-note onset–onset interval in a pair must equal the decrease in the other, the two-beat specification was adjusted to [97, 103]. Clynes' (1983, p. 162) specification of the two-beat Beethoven pulse as [97.5, 100.3] appears to be a mistake.

[7]For one subject (the author's mother), the adjectives were translated into their German equivalents. No particular model was followed in their selection. Although similar adjective pairs have been used in the past to evaluate the expression of different musical compositions (e.g., Hevner, 1936; Bruhn, 1985), at the time the author was not aware of any scales constructed specifically for the evaluation of different performances of one and the same piece. Meanwhile, an article by Senju and Ohgushi (1987) has described such a scale containing 15 adjective pairs (translated from the Japanese).

[8]Since experienced listeners were expected to judge the performances against a well-established internal standard, the fixed order was not considered a serious problem. However, counterbalancing should be employed in future, more extensive research.

[9]Many more specific observations could be made in the present data, which may be of interest to students of music performance. For those interested in pursuing such details, the author will be happy to provide the raw data or graphs of individual performances.

Becking, G. (1928). *Der musikalische Rhythmus als Erkenntnisquelle* (Filser, Augsburg, Germany).

Bengtsson, I., and Gabrielsson, A. (1980). "Methods for analyzing performance of musical rhythm," Scand. J. Psychol. **21**, 257–268.

Bruhn, H. (1985). "Traditionelle Methoden der Musikbeschreibung," in *Musikpsychologie. Ein Handbuch in Schlüsselbegriffen*, edited by H. Bruhn, R. Oerter, and H. Roesing (Urban & Schwarzenberg, Munich, Germany), pp. 494–501.

Clarke, E. F. (1982). "Timing in the performance of Erik Satie's 'Vexations,'" Acta Psychol. **50**, 1–19.

Clarke, E. F. (1988). "Generative principles in music performance," in *Generative Processes in Music*, edited by J. A. Sloboda (Clarendon, Oxford, England), pp. 1–26.

Clynes, M. (1969). "Toward a theory of man: Precision of essentic form in living communication," in *Information Processing in the Nervous System*, edited by K. Leibovic and J. Eccles (Springer, New York), pp. 177–206.

Clynes, M. (1983). "Expressive microstructure in music, linked to living qualities," in *Studies of Music Performance*, edited by J. Sundberg (Royal Academy of Music, Stockholm, Sweden), pp. 76–181.

Clynes, M. (1986). "Generative principles of musical thought: Integration of microstructure with structure," Commun. Cognit. AI **3**, 185–223.

Clynes, M. (1987a). "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 Academy of Music, Stockholm, Sweden), pp. 201–234.

Clynes, M. (1987b). *Computerized System for Imparting an Expressive Microstructure to Succession of Notes in a Musical Score* (United States Patent No. 4,704,682).

de la Motte-Haber, H. (1985). *Handbuch der Musikpsychologie.* (Laaber-Verlag, Laaber, Federal Republic of Germany).

Gabrielsson, A. (1974). "Performance of rhythm patterns," Scand. J. Psychol. **15**, 63–72.

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 Academy of Music, Stockholm, Sweden), pp. 81–104.

Gabrielsson, A., Bengtsson, I., and Gabrielsson, B. (1983). "Performance of musical rhythm in 3/4 and 6/8 meter," Scand. J. Psychol. **24**, 193–213.

Hartmann, A. (1932). "Untersuchungen über metrisches Verhalten in musikalischen Interpretationsvarianten," Archiv für die gesamte Psychologie **84**, 103–192.

Hevner, K. (1936). "Experimental studies of the elements of expression in music," Am. J. Psychol. **48**, 246–268.

Kaiser, J. (1975). *Beethovens 32 Klaviersonaten und ihre Interpreten* (S. Fischer, Frankfurt, Germany).

Lindblom, B. (1978). "Final lengthening in speech and music," in *Nordic*

Prosody, edited by E. Gårding, G. Bruce, and R. Bannert (Department of Linguistics, Lund University, Sweden), pp. 85–100.

Palmer, C. (**1989**). "Mapping musical thought to musical performance," J. Exp. Psychol.: HPP **15**, 331–346.

Povel, D. J. (**1977**). "Temporal structure of performed music: Some preliminary observations," Acta Psychol. **41**, 309–320.

Repp, B. H. (**1989**). "Expressive microstructure in music: A preliminary perceptual assessment of four composers' 'pulses,'" Music Percept. **6**, 243–273.

Repp, B. H. (**1990**). "Further perceptual evaluations of 'composers' pulses' in computer performances of classical piano music," Music Percept. **8**, (in press).

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

Senju, M., and Ohgushi, K. (**1987**). "How are the player's ideas conveyed to the audience?," Music Percept. **4**, 311–324.

Shaffer, L. H. (**1981**). "Performances of Chopin, Bach, and Bartók: Studies in motor programming," Cognit. Psychol. **13**, 326–376.

Shaffer, L. H. (**1984**). "Timing in solo and duet piano performances," Quart. J. Exp. Psychol. **36A**, 577–595.

Shaffer, L. H., Clarke, E. F., and Todd, N. P. (**1985**). "Metre and rhythm in piano playing," Cognition **20**, 61–77.

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 Academy of Music, Stockholm, Sweden), pp. 139–152.

Sloboda, J. A. (**1983**). "The communication of musical metre in piano performance," Q. J. Exp. Psychol. **35A**, 377–396.

Sloboda, J. A. (**1985**). "Expressive skill in two pianists: Metrical communication in real and simulated performances," Canadian J. Psychol. **39**, 273–293.

Sundberg, J. (**1988**). "Computer synthesis of music performance," in *Generative Processes in Music*, edited by J. A. Sloboda (Clarendon, Oxford, England), pp. 52–69.

Thompson, W. F. (**1989**). "Composer-specific aspects of musical performance: An evaluation of Clynes's theory of pulse for performances of Mozart and Beethoven," Music Percept. **7**, 15–42.

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

Wolff, K. (**1979**). *Schnabel's Interpretation of Piano Music* (Faber Music, London).