Music Perception
Winter 1992, Vol. 10, No. 2, 221–242

821

A Constraint on the Expressive Timing of a Melodic Gesture: Evidence from Performance and Aesthetic Judgment

BRUNO H. REPP
Haskins Laboratories

Discussions of music performance often stress diversity and artistic freedom, yet there is general agreement that interpretation is not arbitrary and that there are standards that performances can be judged by. However, there have been few objective demonstrations of any extant constraints on music performance and judgment, particularly at the level of expressive microstructure. This study illustrates such a constraint in one specific case: the expressive timing of a melodic gesture that occurs repeatedly in Robert Schumann's famous piano piece, "Träumerei." Tone onset timing measurements in 28 recorded performances by famous pianists suggest that the most common "temporal shape" of this (nominally isochronous) musical gesture is parabolic and that individual variations can be described largely by varying a single degree of freedom of the parabolic timing function. The aesthetic validity of this apparent constraint on local performance timing was investigated in a perceptual experiment. Listeners judged a variety of timing patterns (original parabolic, shifted parabolic, and nonparabolic) imposed on the same melodic gesture, produced on an electronic piano under control of a Musical Instrument Digital Interface (MIDI). The original parabolic patterns received the highest ratings from musically trained listeners. (Musically untrained listeners were unable to give consistent judgments.) The results support the hypothesis that there are classes of optimal temporal shapes for melodic gestures in music performance and that musically acculturated listeners know and expect these shapes. Being classes of shapes, they represent *flexible constraints* within which artistic freedom and individual preference can manifest themselves.

Introduction

Much has been written about music performance, with the emphasis generally being on the diversity among interpretations by different artists and in different historic periods. Yet, in each period (and quite likely across periods) there have also been generally accepted performance standards, which were reflected in music education, performance practice, and music

Requests for reprints may be sent to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511–6695.

criticism. The nature of these standards has been discussed in a number of treatises (most notably Lussy, 1882), but rarely in objective and quantitative terms. This is particularly true with regard to the expressive microstructure of performance—all those variations that are not easily captured in music notation but that are essential to the communicative function of interpretation. Musicians are usually only dimly aware of these variations, which they control intuitively rather than deliberately. Similarly, musical listeners perceive the structure and expression conveyed by these variations without being aware of the microstructure as such. It has been up to experimental psychologists to discover and measure these variations objectively (e.g., Gabrielsson, Bengtsson, and Gabrielsson, 1983; Palmer, 1989; Repp, 1990; Seashore, 1938/67; Shaffer, 1980).

Even though a number of studies of expressive microstructure have been published, they have rarely provided evidence of constraints on performance parameters. The principal reason is that they usually were based on very small samples of performances, so no statements could be made about the generality of particular microstructural patterns. Hypotheses about the generality of such patterns, as instantiated for example in the performance rules of Friberg (1991) or in the hierarchical timing model of Todd (1985), remain to be validated on large performance data bases. Moreover, studies of music performance have rarely combined measurements with formal perceptual evaluations to confirm the aesthetic validity of the hypothesized or measured patterns.

The work of Johan Sundberg and his colleagues is a significant exception (see Sundberg, Friberg, & Frydén, 1991). A study by Sundberg and Verrillo (1980) had a purpose very similar to that of the present research. These authors were concerned with the temporal shape of the ritardando, the gradual slowing of tempo commonly observed in performance at the ends of most compositions. They asked whether there was an optimal time course for this slowing down that performers observed and listeners expected. They selected 24 recordings of rhythmically uniform music, mostly by J. S. Bach, and measured the onset intervals between successive tones, whose reciprocals they then plotted as local tempo decreasing over time. Sundberg and Verrillo found that the average function resulting from these measurements could be described in terms of two linear segments, the second steeper in slope than the first. They also conducted a perceptual test in which musically experienced listeners were presented with excerpts that exhibited various forms of ritardando, some corresponding to the observed average function and others having deviant temporal shapes of various kinds. The listeners tended to prefer the ritardandos corresponding to the original performances.

In a later discussion of the same data, Kronman and Sundberg (1987) abandoned the bilinear model and instead fitted the average data points with a single curve (a square-root function), which they claimed was similar to that observed when other rhythmic motor activities, such as locomotion, come to a smooth halt. (Specific references to relevant literature were not given.) This function thus may represent a rather general constraint on the optimal shape of the musical ritardando.

Although these studies exhibit some methodological weaknesses[1] and therefore can only be regarded as preliminary, they nevertheless set a good precedent for the kind of approach to be taken in investigations of performance constraints.

The present study concerns possible constraints on the temporal shape of an expressive melodic gesture. (The performance-oriented term "melodic gesture" is used here to refer to a brief sequence of melody tones that is executed as a single expressive unit. The equivalent term "rhythmic group" is often used in the musicological literature.) By a constraint is meant a restriction on the performance patterns that occur in expert interpretations and that are judged acceptable by musically experienced listeners. Melodic gestures occur throughout Western music in most styles, and they come in a large variety of forms. It seems unlikely that all these forms are subject to any single performance constraint. The nature of these constraints may vary as a function of many factors, including tempo, metric and harmonic structure, style, and so on. Rather than searching for a universal constraint, the present study focused on the timing pattern of one particular melodic gesture. If it could be demonstrated that this pattern is subject to a significant constraint in performance and perceptual judgment, this would at least provide an existence proof of such constraints on expressive microstructure. Moreover, by focusing on a specific case, the constraint can be characterized rather precisely. Questions about its origin and generality may then form the basis for future research.

The melodic gesture under investigation occurs in Robert Schumann's famous piano piece, "Träumerei" (No. 7 of "Kinderszenen," op. 15), whose score is shown in Figure 1. The melodic gesture moves from bar 1 into bar 2 (see Figure 3). In its notated form, it consists of five eighth notes ascending in pitch and a final longer note that repeats the pitch of the preceding eighth note. The gesture recurs six times (eight times, if the obligatory repeat of the first eight bars is counted) during the piece, with some variations in key and interval structure. These recurrences are aligned vertically in Figure 1. The gesture is of central importance to the expressive quality of a performance of "Träumerei" and can be assumed to be given close attention by both performers and listeners.

1. These weaknesses include preselection of performances with "typical" ritardandos, averaging across heterogeneous musical materials, a rather unbalanced and poorly described design in the perception experiment, and great variability in listeners' judgments.

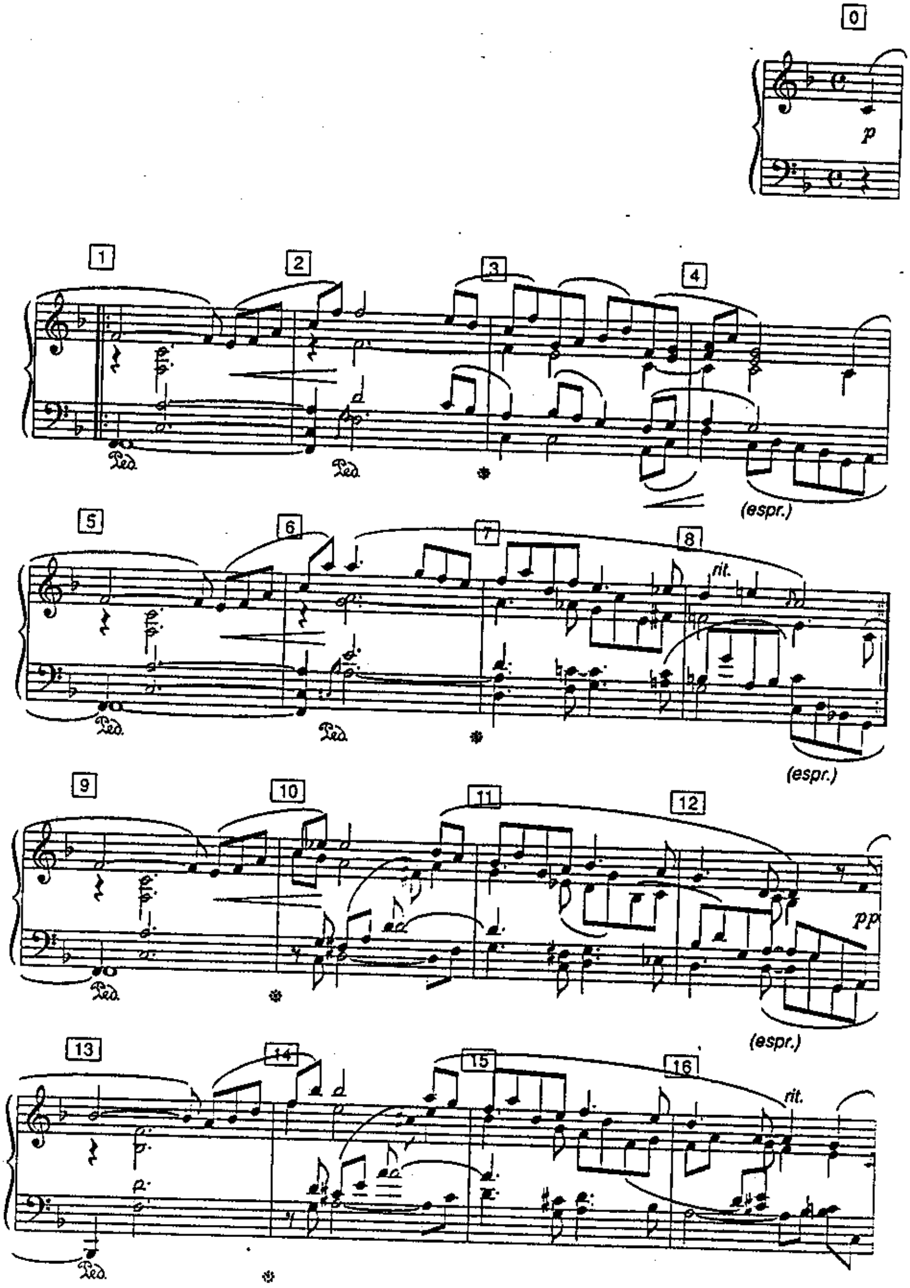

Fig. 1. Piano score of Schumann's "Träumerei," arranged on the page so parallel structures are vertically aligned. The score was created after the Clara Schumann edition (Breitkopf & Härtel) using MusicProse software; minor deviations from the original are due to software limitations.

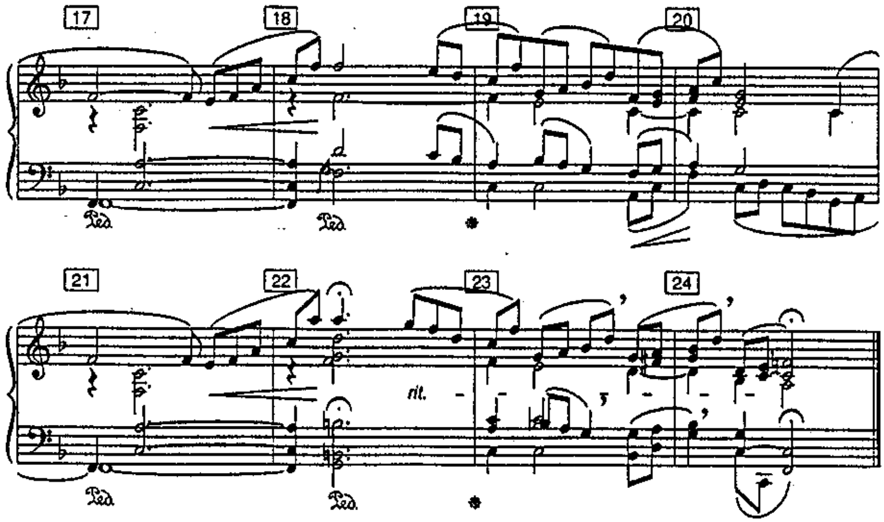

Fig. 1 *Continued*

The onsets of the tones corresponding to the six melody notes define five interonset intervals (IOIs), which would be equally long if the music were performed mechanically (e.g., by a computer). In fact, they are never equal in a human performance; pianists always give an expressive temporal shape to this crucial part of the melody. This temporal shape can be visualized as the pattern of observed IOI durations, plotted as connected points equidistant along the x-axis ("score time"). How many such patterns are there? In principle, the melodic gesture can be performed with any temporal pattern whatsoever.[2] However, the hypothesis pursued here is that only certain patterns actually occur in expert performances and are found acceptable by listeners.

One characteristic of this class of patterns can be predicted on the basis of the general principle of final lengthening (e.g., Lindblom, 1978; Todd, 1985): A slowing down of tempo is often observed at the ends of action units such as phonological phrases in speech or melodic gestures in music, particularly when they coincide with the end of a larger structural unit, such as a clause (subphrase) or phrase. Therefore, the timing patterns to be investigated can be expected to show some lengthening of the last IOI(s). An independent reason for lengthening of the last IOI might be the

2. That is, as long as the pattern does not lead musically literate listeners to conclude that a rhythmic *mistake* has been made. In other words, the performance timing pattern must be compatible with the notated temporal pattern. There is, of course, a grey area here in which faithfulness to the score may be a matter of opinion.

occurrence of two grace notes (essentially a written-out arpeggio) in the left hand during that interval (see Figure 1). However, these grace notes occur only in bars 2, 6, and 18, not in bars 10, 14, and 22. The pattern of execution of these two sets of variants may differ. Another relevant phenomenon is the possible lengthening of accented tones. In the score, the fourth note of the melodic gesture follows a bar line and thus in theory carries a strong metrical accent (downbeat). Based on the notated music, therefore, a lengthening of the fourth intertone interval might be predicted. Musical intuition suggests, however, that this theoretical accent is suspended in performance, and that the accented tone of the melodic gesture is in fact the final one. Whether this is in fact so is an empirical issue to be addressed later. In principle, nothing can prevent a pianist from placing an overt accent on the fourth tone.

In the remainder of this article, a summary of performance measurements is followed by the detailed report of a perceptual experiment. The measurements derive from a comprehensive analysis of timing microstructure in performances of Schumann's "Träumerei"; for details, the reader is referred to Repp (1992).

Performance Measurements

Tone-onset timing measurements were obtained from the digitized waveforms of 28 different performances of "Träumerei," taken from commercial recordings (LP, CD, or cassette) by 24 pianists. Two famous pianists (Alfred Cortot and Vladimir Horowitz) were represented with three different recordings each. The measurements were averaged over the obligatory repeat of bars 1–8 (observed by all but two pianists in the sample) before further analysis. Thus there were data for six instances of the melodic gesture of interest in each of the 28 performances, a total of 168.

Initially, the geometric mean durations of the five IOIs for each of the six instances of the gesture were computed across the 28 performances. These durations (in milliseconds) were plotted as a function of score time (i.e., at equal abscissa intervals), and their pattern was examined as to whether it could be fit by some simple function. These data are shown in Figure 2. It is evident that the timing pattern of each instance was fit well by a smooth curvilinear function, in fact a parabola (quadratic curve). Overall, pianists tended to speed up somewhat in the initial part of the melodic gesture and to slow down at the end. This slowing down was especially pronounced in the last instance of the melodic gesture (bars 21 and 22), where the score indicates a fermata (hold) on the last note. It was least pronounced in the two instances in the middle section of the piece

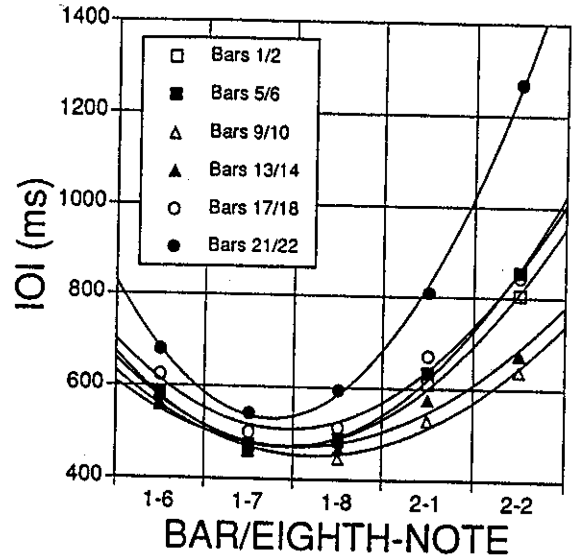

Fig. 2. Timing patterns of six instances of the same melodic gesture in "Träumerei." The data points are the geometric average durations of 28 performances (Repp, 1992), with quadratic functions fitted to them. The abscissa labels refer to bars 1 and 2.

(bars 9 and 10 and bars 13 and 14). All instances, however, were described well by quadratic functions that differed mainly in curvature.

Subsequently, all 168 individual timing patterns were plotted and examined in the same way. It was found that 87% of them could be described rather well by quadratic functions of varying elevation (i.e., average tempo) and curvature (i.e., degree of tempo modulation). All but two of the exceptions followed a single pattern: a relative shortening of the last IOI.[3] This pattern, whose main representative was the French pianist Alfred Cortot in his three performances, suggests a different structural interpretation of the melodic gesture: a division into two subgestures and/or an intention to place an accent on the fourth tone. Three other pianists showed this pattern intermittently; Cortot himself consistently avoided it in bars 21 and 22, where he showed the standard parabolic timing curve.

Further analysis of the coefficients of the quadratic polynomials ($y = a + bx + cx^2$) fit to 87% of all instances revealed some strong relationships among the constant (a), linear (b), and quadratic (c) terms of these functions. The latter two, in particular, were highly correlated. A substantial correlation was found between the quadratic and constant

3. The remaining two exceptions, both occurring in the performance of Brazilian pianist Cristina Ortiz, exhibited a relative lengthening of the third IOI instead (i.e., a W-shaped pattern).

terms as well. Linear regressions among the coefficients made it possible to predict the linear and constant terms from the quadratic term and thus to generate a single family of parabolas by varying the quadratic term alone. (This family is shown in the upper left-hand panel of Figure 4.) It captures a substantial amount of the variance in the data, with deviations occurring mainly in the constant term (i.e., elevation along the ordinate, corresponding to variations in overall tempo), which is irrelevant to the temporal shaping of the melodic gesture.

These quadratic curves represent a strong constraint on the timing pattern of the melodic gesture studied here. Apparently, the large majority of expert pianists achieve a parabolic timing function by controlling a single degree of freedom. No pianist lengthened the second IOI, say, or shortened the first, or showed any pattern (other than the type favored by Cortot) that deviated substantially from a parabolic trajectory (but see footnote 2). Even the Cortot pattern followed a parabolic curve through the first four IOIs. To the author, however, these performances sound mannered. This subjective impression, in conjunction with the overall predominance of parabolic timing patterns, suggested that the more typical parabolic patterns might also be preferred by other musically experienced listeners. This hypothesis was tested in the following perceptual experiment.

Perceptual Experiment

The purpose of this experiment was to demonstrate that listeners' aesthetic preferences converge on the timing patterns that characterize the majority of expert performances. To that end, subjects were presented with the melodic gesture of interest, executed with a variety of timing patterns, each of which was to be rated for acceptability on a 10-point scale. The timing patterns included parabolic and "hybrid" (nonparabolic) shapes. Among the former, some belonged to the family of functions observed in actual performances, whereas others deviated in the location of the minimum. It was expected that listeners would prefer the "normal" over the deviant parabolic shapes. In addition, these functions varied in curvature. As listeners might also exhibit a preference for a particular curvature (degree of tempo modulation) within each class of temporal shapes, some deviant shapes might actually be preferred over some normal shapes. However, for a given degree of curvature, the normal shapes were expected to be preferred most. The hybrid shapes were generated from two normal parabolic patterns of different curvature by interchanging their IOIs in all possible ways. It was expected that listeners' judgments would reflect the hybrids' degree of approximation to a parabolic shape.

The responses were also expected to yield information about what deviations from the normal shapes are more readily tolerated than others. In fact, one of these deviant shapes resembled the Cortot type of pattern.

The role of listeners' musical experience was a rather crucial issue. If the parabolic constraint uncovered in the performance timing measurements reflects a general principle of physical motion (i.e., an optimal pattern of acceleration-deceleration), then even listeners without much musical experience might show a preference for it. The alternative possibility is that listeners need to be attuned to temporal patterns in classical music performance to show reliable preferences in this task. To investigate this issue, subjects both with and without musical experience were tested. A second question, concerning subjects with musical experience, was whether their judgments would be based on general knowledge of performance principles in classical music, or on specific knowledge of "Träumerei" and its performance. This question was not addressed rigorously, but some relevant information was obtained.

METHODS

Subjects

Twenty-six subjects participated. The majority of them had responded to an advertisement in the Yale campus newspaper; others were recruited personally by the author and included some friends and family members who served without pay. Twelve subjects had little musical education; most of them did not play any instrument, while some had studied an instrument for a short time. Fourteen subjects were musically experienced; they included 11 pianists, two violinists, and one flutist, ranging in skill from advanced amateur to professional level.

Stimuli

The stimuli were generated on a Roland RD250S digital piano under MIDI control. Temporal resolution was 5 msec. Each stimulus consisted of the excerpt shown in Figure 3 (from bars 1 and 2 of "Träumerei"), played with one of 45 different timing patterns.

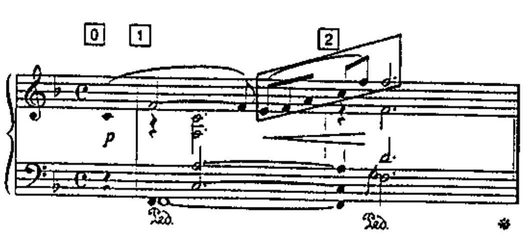

Fig. 3. The musical excerpt used in the experiment (from bars 1 and 2 of "Träumerei," with slightly modified final notes). The melodic gesture of interest is boxed in.

The timing pattern was applied only to the melodic gesture of interest, which comprised five IOIs; the timing of the preceding context, comprising three longer IOIs, was constant at values representing the geometric means of the 28 expert performances measured by Repp (1992): 1065, 1380, and 1825 msec, respectively. The timing of the left-hand grace-note tones during the last IOI of the critical gesture was such that the first tone started after one-third of the IOI had elapsed and ended with the onset of the second tone, which started after one-third of the remaining interval had elapsed. (This timing pattern was fairly common in the 28 performances examined.) To make room for the grace-note tones, the preceding chord, the tied-over quarter notes of the preceding chord in bar 2 were realized as tied-over eighth notes. Sustaining pedal was added as indicated in the score. The tones had a fixed expressive intensity pattern similar to that of one of the expert performances.

The timing patterns of the critical melodic gesture are illustrated in Figure 4. The upper left-hand panel shows the five "normal" patterns, which followed parabolic functions of varying curvature. Each parabola was generated by the equation, IOI (msec) = $C + Lx + Qx^2$, where x stands for the ordinal numbers of the IOIs (1,... 5). The quadratic term (Q) of the polynomial equation was set at values of 20, 40, 60, 80, and 100, which span the range of most empirically observed timing functions. The linear (L) and constant (C) terms of the parabolas were derived according to the empirically determined regression equations, $L = 35 - 5.5Q$ and $C = 388 + 7.8Q$ (see Repp, 1992). The resulting stimuli were named Q20,...Q100.

The lower panels in Figure 4 illustrate two sets of deviant parabolic curves. Each set varied in Q along the same values as the normal set, but the constant and linear terms differed. In the "left-shifted" set, C was decreased by 300 and L was increased by 100, whereas, in the "right-shifted" set, C was increased by 300 and L was decreased by 100. In each case, the change in one parameter was arbitrary, but the change in the other parameter was chosen so as to keep the average IOI duration equal to that of the normal condition with the same Q. The stimuli in the left-shifted set (Q20L,... Q100L) started faster and ended slower than the normal stimuli; the opposite was true for the stimuli in the right-shifted set (Q20R,... Q100R).

The remaining 30 timing patterns were generated as illustrated in the upper right-hand panel of Figure 4. The heavy lines in the figure illustrate the Q20 and Q100 timing patterns, represented here as polygons rather than as smooth curves. Thirty hybrid patterns were generated by interchanging IOI durations from those two patterns. With two possible values for each of five IOIs, there are 32 possible patterns, two of which are the original ones. The original patterns were coded arbitrarily as H00000 (=Q20) and H11111 (=Q100), and hybrid patterns were coded as H10000, H11000, etc. Clearly, some of these hybrids (e.g., H00100, H11011) were very similar to the originals, whereas others were more dissimilar. Although some of them were clearly nonparabolic (e.g., H01010), others might by fit by a left-shifted or right-shifted parabola (H00011 and H11100, respectively). In contrast to the left- and right-shifted parabolic patterns, however, all individual IOIs in the hybrid patterns were within the normal range. One hybrid pattern, H11110, was not unlike the Cortot pattern described earlier.

The stimuli were recorded electronically from the audio output jack of the digital piano onto high-quality cassette tape. Six examples were recorded at the beginning of the tape, the first three with isochronous timing of the melodic gesture (i.e., with constant IOIs of 500 msec), and the second three being stimuli H01101, Q80R, and Q40. These examples were followed by three different randomized sequences of the 45 stimuli. Interstimulus intervals were 5 sec, with an additional 5 sec after each group of 15, and another 5 sec between blocks.

Procedure

Subjects received a dubbed copy of the master cassette, accompanied by detailed printed instructions, an answer sheet, and a questionnaire about their musical experience. They

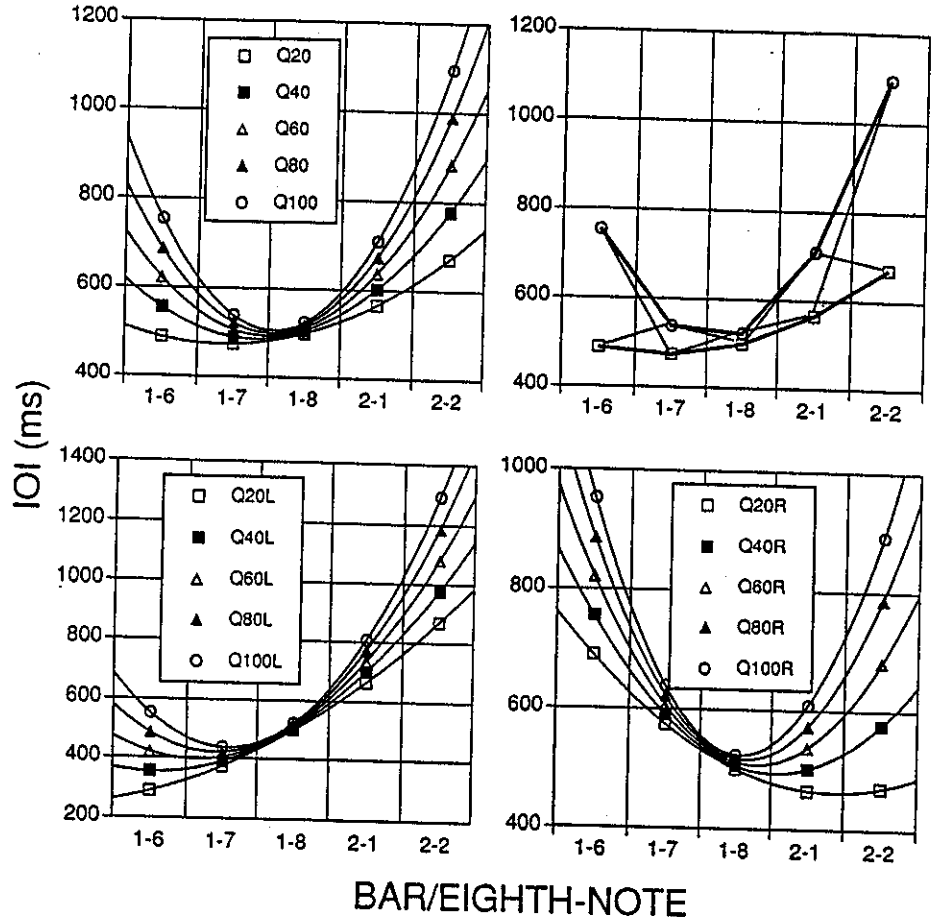

Fig. 4. Timing patterns of the experimental stimuli. Upper left-hand panel: normal parabolic patterns. Lower left-hand panel: left-shifted parabolic patterns. Lower right-hand panel: right-shifted parabolic patterns. Upper right-hand panel: hybrid patterns.

listened on their home audio equipment and returned the completed materials. (Control over sound quality and playback level was not crucial in this study.)

The instructions displayed the score of the excerpt (cf. Figure 3) and included the following crucial sections:

> Each time the excerpt will be played with a slightly different timing pattern of the notes. Your task is to judge the aesthetic appeal of each timing pattern. Clearly, there are no right or wrong responses here; I want to find out what sounds good to *you*. . . .

(After the first set of examples had been introduced:)

> In the following three examples, the eighth-notes vary in duration, as they would in a human performance. Each of the three examples has a different timing pattern, and they may not (in fact, should not) sound equally good to you. Clearly, there

> are some timing patterns that are preferable to others. . . . In the following test, you will indicate [your] preference by giving a numerical rating between 1 and 10 to each excerpt you hear, where 10 is the best possible rating and 1 is the worst. . . . However, don't use these [ratings] in an absolute sense, but try to adjust to the diversity of timing patterns you hear and use the whole scale; that is, give ratings of 9 or 10 to the best patterns you hear in the course of this experiment, and ratings of 1 or 2 to the worst, regardless of how you might judge these patterns in an absolute sense. Avoid giving too many ratings in the middle range; try to use the extremes as well.

Nearly all subjects in fact used the whole range of rating categories.

Results and Discussion

CONSISTENCY OF JUDGMENTS

The first question to ask was whether the subjects were able to perform the task—that is, give reliable judgments. The reliability of their ratings could be determined by correlating the ratings across the three blocks of stimuli. Since the first block served to familiarize subjects with the stimuli, the correlation between the second and third blocks was expected to be higher than that between the first block and either of the other two. However, although this was true for some individual subjects, there was no such overall tendency in the data, and the three interblock correlations were therefore averaged for each subject.

All 14 musically experienced subjects exhibited significant average correlations, ranging from 0.31 ($p<.05$) to 0.79 ($p<.0001$). Of the 12 musically inexperienced subjects, however, only two showed a significant average correlation (0.49 and 0.51, respectively, both $p<.001$); for the rest, the correlations ranged from −0.01 to 0.18. This is a very striking difference. Most musically untrained subjects apparently did not possess a stable criterion by which to judge the stimuli.

A second criterion that separated the two groups of subjects was their response to timing pattern Q20L. As was evident to the author during stimulus generation, this pattern (as well as Q40L) sounded really ridiculous, in contrast to the other patterns, which seemed at least moderately acceptable. Indeed, all musically experienced subjects assigned their lowest ratings to Q20L, with average ratings ranging from 1.0 to 2.0. Ten of the 12 musically inexperienced subjects, however, gave this stimulus average ratings between 5.0 and 9.3 (!). The remaining two subjects gave average ratings of 2.7 and 3.0, respectively; however, they were not the two individuals who showed significant reliability of judgments.

Because of this striking dichotomy in judgmental criteria and consistency between the two subject groups, further analysis was restricted to the data of the 14 musically experienced subjects. Their responses to the parabolic and hybrid patterns were analyzed separately.

PARABOLIC PATTERNS

The parabolic patterns constituted a 3 (Type) by 5 (Curvature) design. The subjects' ratings were averaged over the three blocks and subjected to a two-way repeated-measures ANOVA. The average ratings are plotted in Figure 5.

As is evident from the figure, the prediction that the normal parabolic curves would receive the highest ratings was confirmed. The main effect of Type was highly significant [$F(2,26) = 65.60$, $p < .0001$]. There was also a significant main effect of Curvature [$F(4,52) = 5.50$, $p < .001$], although it was irrelevant in view of a strong two-way interaction [$F(8,104) = 20.95$, $p < .0001$]. This interaction was evidently due to the very different effect of Curvature for left-shifted parabolas than for normal and right-shifted ones.

The latter two stimulus types were analyzed in a separate ANOVA. There were significant effects of Type [$F(1,13) = 40.24$, $p < .0001$] and of Curvature [$F(4,52) = 11.26$, $p < .0001$], but a nonsignificant interaction [$F(4,52) = 1.44$, $p = .24$]. Normal parabolas were rated more highly than right-shifted ones at all degrees of curvature, and for both types the most preferred curvature was Q = 40. By contrast, left-shifted parabolas were judged extremely unfavorably at low curvatures (as noted earlier) and more favorably at high curvatures. At Q = 100, left-shifted functions were almost as acceptable as normal ones, and more acceptable than right-shifted ones. This indicates that the subjects were particularly averse to hearing a short first IOI; for stimuli Q60L to Q100L, the reduction of the starting tempo apparently compensated for the exaggeration of the final slowdown.

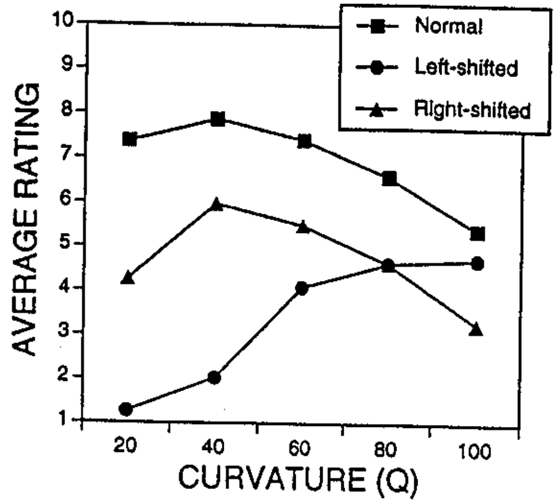

Fig. 5. Average ratings given to the parabolic patterns.

HYBRID PATTERNS

The 30 hybrid patterns, together with the parent patterns Q20 and Q100, formed a 2 × 2 × 2 × 2 × 2 design: Each of the five IOIs could either have a short duration (from Q20) or a long duration (from Q100). These five positions will be referred to by the letters A, B, C, D, and E in the following. A five-way repeated-measures ANOVA was conducted on the subjects' ratings. Significant main effects in this analysis would indicate that the listeners preferred a shorter or longer IOI duration in particular positions. Such effects were more likely in the positions where Q20 and Q100 differed most, namely, E and A (cf. Figure 4). Of greater interest were any interactions among the five position factors, which would indicate that the relationships among several IOIs mattered. The average ratings are shown in Table 1.

Only one of the five main effects reached significance, that of position A [$F(1,13) = 8.10$, $p < .02$]: Listeners preferred the shorter IOI in that position (see Table 1, bottom row).[4] The main effect for the last position, E, was nonsignificant [$F(1,13) = 1.06$, $p = .32$], even though the change in duration was larger (424 msec vs. 264 msec). This is interesting in view of the "Cortot pattern" mentioned earlier, in which the last IOI is abnormally shortened; apparently, the present listeners were not very consistent in their responses to different degrees of final lengthening.[5]

TABLE 1
Average Ratings for the Hybrid Stimuli

Code	Rating	Code	Rating	Code	Rating	Code	Rating
H00000	7.4	H01000	6.5	H10000	5.4	H11000	6.9
H00001	7.6	H01001	6.5	H10001	4.9	H11001	5.1
H00010	6.8	H01010	6.1	H10010	5.2	H11010	5.6
H00011	6.0	H01011	6.1	H10011	4.6	H11011	5.3
H00100	7.1	H01100	6.7	H10100	5.7	H11100	5.5
H00101	7.0	H01101	6.1	H10101	5.5	H11101	6.0
H00110	6.8	H01110	6.6	H10110	4.7	H11110	5.7
H00111	6.8	H01111	6.8	H10111	4.2	H11111	5.4
H00...	6.9	H01...	6.4	H10...	5.0	H11...	5.7
H0....	6.7			H1....	5.4		

4. This seems to contradict the earlier observation that subjects disliked a short first IOI in left-shifted patterns. Note, however, that the first IOI of stimulus Q20 corresponded to that of stimulus Q80L (cf. Figure 4). Thus, while listeners disliked abnormally short first IOIs, within the normal range they preferred short over long first IOIs.

5. It should not be inferred that the subjects were unable to detect the difference in duration of the final IOI (424 msec). The present task was one of aesthetic, not sensory, discrimination.

Several interactions were significant, however, which indicated that listeners did not judge IOI durations individually. The three largest interactions were AB [$F(1,13) = 16.59$, $p < .002$], ABCD [$F(1,13) = 16.15$, $p < .002$], and CE [$F(1,13) = 12.00$, $p < .005$]. Four additional interactions, ACD, BCD, BD, and BDE, were significant at $p < .05$, and three further interactions, ACDE, ABCE, and BCDE, were nearly significant ($p < .06$). It is perhaps noteworthy that the only two positions that were never involved together in a significant interaction are A and E. The beginning and the end of the timing pattern thus seemed to be judged independently.

These interactions indicate that it is the pattern of IOIs that mattered, not individual IOI durations. The AB interaction, for example, shows that a shorter second IOI (B) was preferred when the first IOI (A) was short, but a longer B was preferred when A was long (see Table 1, penultimate row). The BD and CE interactions show a similar pattern of preferred positive covariation between two positions. Now consider a more complex interaction, ABCD, which subsumes two other significant interactions, ACD and BCD. It can be viewed as four CD interactions, one for each of the four combinations of A and B values. Three of these four two-way interactions exhibit the positive covariation described earlier, but one (that for long A and short B) exhibits negative covariation. That is, in that specific condition listeners preferred a long C when D was short, and a short C when D was long. The reason for this complex interaction is not obvious, but it is remarkable that position C, which had a duration difference of only 24 msec, was so strongly involved in it. Listeners' sensitivity to small deviations in that position mirrors the restricted range of observed IOI durations.

Another prediction can be examined in the hybrid pattern data. The parent patterns, Q20 and Q100, were parabolas of the normal type. The hybrid patterns approached parabolas in various degrees. Therefore, none of them should have been rated higher than the parent patterns, whereas quite a few should have been rated lower. However, the difference in average ratings between Q20 and Q100 of about two points (cf. Figure 5) must be taken into account. The revised prediction, therefore, is that no hybrid pattern should have been judged more acceptable than Q20, but some should have been judged less acceptable than Q100.

The first part of the prediction was confirmed: Only one hybrid stimulus, H00001 (i.e., Q20 with a lengthened final IOI), received a higher average rating than Q20 (7.6 vs. 7.4), but this difference was certainly not significant. The second part of the prediction also was supported: Seven hybrid stimuli received lower average ratings than Q100 (5.4). The lowest rated stimulus, H10111 (4.2), corresponded to Q100 with a shortened second IOI. In fact, Table 1 shows that all stimuli of the type H10...

received low ratings whose range (4.2 to 5.7) did not overlap at all with the ratings of H00... and H01... stimuli (range: 6.1–7.6); H11... stimuli were in between (range: 5.1–6.9). This again confirms the relative importance of the first IOI in relation to the second. Clearly, listeners did not like a relatively long first IOI. This is easily explained by the tonal structure: The first tone of the melodic gesture, E, is a half-step below the tonic (i.e., a "dissonant lower neighbor") and, moreover, in a metrically weak position, which calls for a quick resolution.

It might also be asked whether the ratings of the hybrid stimuli reflected the degree to which they approached a parabolic timing curve. As the results for parabolic patterns show, however, what matters is not so much the parabolic shape itself as its parameters. That deviations from a parabolic shape can be tolerated is illustrated by the ratings for hybrid stimulus H11110 (5.7), which were slightly higher than those for the perfectly parabolic stimulus H11111, alias Q100 (5.4). H11110 resembles the Cortot pattern, and Cortot may have taken advantage of listeners' tolerance for variations in the final IOI. The resulting timing shape is only moderately acceptable, however, which matches the author's impression from listening to Cortot's recordings (whatever other qualities they may have).

General Discussion

The present results are limited in a number of ways, which will be discussed next. Within these limitations, however, they provide a clear indication of a constraint on performance timing that is shared by expert performers and musically acculturated listeners. While it may be perfectly obvious to some theorists that such constraints must exist, their objective demonstration and characterization has rarely been undertaken before. The local constraint examined here is flexible enough to permit a large variety of concrete timing patterns; yet there is reason to believe that, in a specific musical context, a single pattern may be optimal. Because of the contextual timing variation inherent in different performances, the evidence for optimality comes from the perceptual data alone. For the specific musical excerpt presented here, the timing shape labeled Q40 seemed to be best, on the average.

It is necessary to discuss now what possible generality this finding may have. Three major issues concern individual differences in preferences and experience, the specific stimulus conditions of the experiment, and the specific musical excerpt selected.

Individual differences among listeners did exist, of course, as they do in nearly all psychological studies. However, the high levels of significance of some of the effects obtained suggest considerable agreement. More extensive replications of judgments per subject would be needed to interpret individual differences. A few observations are offered here: All subjects but one gave some of their highest ratings to parabolic patterns of the normal type; the exception was a professional pianist who gave her highest ratings to stimuli H11110 (the Cortot-like pattern), H11010, and Q40R, which all shared an initial accelerando but had a reduced ritardando at the end. Among the normal parabolic patterns, most subjects' preference fell on patterns with lower curvature (Q20, Q40, or Q60), although two subjects, both accomplished pianists, preferred those with higher curvature. One subject, interestingly the youngest in the group (an 11-year-old girl who studies the piano), did not differentiate much among the different degrees of curvature, though she clearly preferred the normal patterns over the left- and right-shifted ones. How the internal standard by which such patterns are judged are acquired in the course of music education is of course a very interesting question for future research.

That musical experience is a sine qua non for reliable performance in the experimental task was demonstrated convincingly here. The precise nature of the necessary experience is less clear, however. The subject sample did not include individuals who cannot play an instrument but listen extensively to classical music; the musically experienced subjects were all instrumentalists of various degrees of proficiency. The several professional pianists in the group, who surely had the most extensive musical education, actually were not the most reliable judges. It is entirely possible that professional musicians' criteria are less fixed than those of amateurs and ordinary music lovers, because constant interaction with other musicians as students, ensemble players, and teachers may encourage tolerance of a large variety of interpretative nuances.

It seems unlikely that specific knowledge of "Träumerei" and exposure to performances of this music in the past played a significant role in subjects' judgments. Most subjects were very familiar with the piece, but some were not. One subject, a flutist, indicated that she did not know it at all; two others, who are string players, and the 11-year-old pianist indicated they were "fairly" familiar with it. Yet, these subjects gave reliable judgments consistent with those of the other musicians. The more important argument is one of plausibility: Although some surface characteristics of previously heard performances may well be part of the memories of familiar pieces of music, performance rules must be stored in a more abstract form also, so as to be applicable to music never heard before. Musically experienced listeners surely can judge the performance quality of novel music in a familiar style, just as performers can sightread new music with good expression, again provided the style is familiar (as it is in the case of any piece from the Romantic period).

Subject variables thus do not seem to impose a serious limitation on the generality of the present results. Stimulus variables are more of a problem. At least three factors that may influence subjects' judgments were kept constant in the experiment. One is the melodic and harmonic content of the musical excerpt. As pointed out in the section on performance measurements, the melodic gesture under examination occurs six times in "Träumerei," and only two of these instances are exactly identical. As Figure 2 showed, the average timing curve of the excerpt used in the rating task (which occurs in bars 1 and 2 and 17 and 18) has only moderate curvature, comparable to the Q40 stimulus. A lower curvature was typical of the variants in the middle section of the piece, while high curvatures were mainly associated with the last instance preceding the *fermata*. Thus the listeners indeed preferred the curvature appropriate to the excerpt offered, but they may well prefer a different curvature for other variants. However, the preference for normal parabolic shapes should hold across all variants.

A second factor is the timing (and the implied tempo) of the context in which the critical melodic gesture was presented. The IOIs of the preceding musical events were set somewhat arbitrarily at the geometric means of the performance sample. It is possible, even likely, that a different choice of IOIs for the context would have influenced subjects' preferences. For example, if the IOIs had been longer (implying a slower tempo), listeners may have opted for a more curved or elevated timing function. This would be interesting to test in future experiments. As it was, however, listeners were presented with an average contextual timing pattern, and they preferred a curvature that also corresponded to the average, which seems appropriate. Their general preference for normal parabolas should be independent of variation in contextual timing.

A third factor is the intensity microstructure of the melodic gesture, which also was held fixed. It was derived from an individual performance, and its contour may not have been close to the average.[6] It did constitute a *crescendo,* however, as marked in the score. Very little is known at this time about the perceptual interdependence, if any, of timing and intensity microstructure. It is conceivable that a different intensity contour would change subjects' curvature preferences. Again, however, there is no reason to believe that the subjects would prefer atypical timing patterns, as long as the intensity microstructure stayed within the normal range of variation.

A final consideration is the selection of timing functions presented in the experiment. Clearly, many possible shapes were not included, mainly because they were expected to sound terrible and might have offended musical listeners' sensibilities. This is not a serious omission. On the other hand, it is conceivable that there are timing curves superior to Q40 in this

6. The methods of deriving and transferring the intensity values will not be defended here, as they are still in need of validation. Suffice it to say that the dynamic variation sounded appropriate to the author. An analysis of the intensity microstructure of the entire sample of 28 "Träumerei" performances remains to be conducted.

particular context. The left- and right-shifted parabolas constituted fairly gross deviations, and there are other functions closer to the normal ones that, in a sufficiently sensitive perceptual test, might prove even more highly acceptable. It must also be noted that implicit tempo (which is difficult to quantify in a temporally modulated performance) was confounded with curvature to some extent, Q100 having a slower tempo than Q20. Listeners' overall preference for Q40 may have constituted a preference for (contextually appropriate) tempo as much as for curvature. This would have to be sorted out by varying the constant and quadratic parameters of the timing curves independently.

In summary, consideration of various stimulus-related factors suggests that listeners' preference for a particular curvature of the timing function may well be context-dependent; however, their general preference for normal parabolic shapes most likely is not. It should also be remembered that the normal family of parabolic shapes was derived from a set of performances that varied widely in the performance parameters (tempo, contextual timing, intensity microstructure) whose possible role in perception was just considered. The generality of the parabolic constraint across this performance variation should have a parallel in perceptual preference across similar variation.

This leads to the broader question concerning the generality of the parabolic constraint to other kinds of melodic gestures and musical styles. One obvious limitation is that the constraint can meaningfully apply only to melodic gestures that have at least four IOIs. The more IOIs, the stronger the constraint may manifest itself. Repp (1992) examined the timing patterns of three other melodic gestures in "Träumerei," each comprising four IOIs near the end of a phrase; they, too, seemed to follow the constraint, but somewhat less consistently than the five-IOI gesture examined here. Gestures with less than four IOIs, of course, cannot violate the parabolic constraint; they are simply irrelevant to it.

Another limitation is that the gestures may need to have a ritardando in them. This was true for all the instances examined by Repp (1992). Moreover, the present results are in strong agreement with the performance and perception results of Sundberg and Verrillo (1980), who focused on final ritardandos in Baroque music. The parabolic constraint thus may characterize ritardandos at all levels of the grouping structure, and quite possibly across different musical styles. It may indeed represent a "natural" way of changing tempo, including both accelerando and ritardando, although the evidence for accelerando is limited to the initial part of the melodic gesture examined here.

These tempo changes, moreover, must be uninterrupted. This perhaps constitutes the most serious limitation of the parabolic constraint. It can apply only to gestures that are rhythmically uniform and do not contain tones that receive special emphasis for harmonic or melodic reasons. If so, it characterizes only a small minority of the melodic gestures in a musical piece, although they may be the most salient ones, which mark the ends of major sections. This minority, however, turns into a majority if all short melodic gestures in which the constraint applies trivially are included. It is noteworthy that Todd (1992), in the process of extending his coarse-grained model of expressive timing (Todd, 1985) to detailed local timing patterns, has been assuming a linear velocity function of tempo change for melodic gestures ("segments") of any length, apparently with good success. A linear velocity change is equivalent to a quadratic timing function for the raw IOIs during a unidirectional tempo change. Previously, Todd (1985, 1989) presented data suggesting that the global timing shapes of whole phrases can be modeled by a family of parabolic functions. His current, somewhat modified conception promises to constitute a valid basis of a general performance model.

The parabolic functions used in Repp (1992) and in the present study were empirically derived and may eventually have to give way to similar but theoretically motivated functions such as proposed by Todd (1992), provided that they fit the data equally well. The extramusical origin of constraints on performance timing is still a matter of speculation, but it is likely to lie in aspects of physical movement that have invaded musical performance and ultimately account for the frequent allusions to "musical motion" in the musicological literature. Although musical motion is often attributed to tonal sequences without explicitly appealing to performance, it seems likely that music needs to be *set into motion* by a performer, real or imaginary. Once the physical movement has entered the music, it will in turn be able to "move" a listener, provided it has the properties that the sensitive listener is attuned to. The kinds of melodic gestures that are most "moving" in a good performance are probably those that give the timing constraint a chance to emerge clearly and impress itself on the listener.[7,8]

References

Friberg, A. Generative rules for music performance: A formal description of a rule system. *Computer Music Journal,* 1991, *15,* 56–71.

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

---

7. This research was made possible by the generosity of Haskins Laboratories (Michael Studdert-Kennedy, president). Additional support came from NIH BRSG Grant RR-05596 to the Laboratories. I am grateful to Pat Shove for many stimulating discussions.

8. A short version of this paper was presented at the Second International Conference on Music Perception and Cognition in Los Angeles, February 1992.

Kronman, U., & Sundberg, J. Is the musical ritard an allusion to physical motion? In A. Gabrielsson (ed.), *Action and perception in rhythm and music.* Stockholm: Royal Swedish Academy of Music, 1987, Pub. No. 55, pp. 57–68.

Lindblom, B. Final lengthening in speech and music. In E. Gårding, G. Bruce, & R. Bannert (eds.), *Nordic prosody.* Lund, Sweden: Lund University, 1978, pp. 85–101.

Lussy, M. *Musical expression: Accents, nuances, and tempo in vocal and instrumental music.* London: Novello, 1882.

Palmer, C. Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance,* 1989, *15,* 331–346.

Repp, B. H. Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America,* 1990, *88,* 622–641.

Repp, B. H. Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America,* 1992, *92* (in press).

Seashore, C. E. *Psychology of music.* New York: McGraw-Hill/Dover, 1967. (Original work published 1938)

Shaffer, L. H. Performances of Chopin, Bach, and Bartók: Studies in motor programming. *Cognitive Psychology,* 1981, *13,* 326–376.

Sundberg, J., Friberg, A., & Frydén, L. Threshold and preference quantities of rules for music performance. *Music Perception,* 1991, *9,* 71–91.

Sundberg, J., & Verrillo, V. On the anatomy of the retard: A study of timing in music. *Journal of the Acoustical Society of America,* 1980, *68,* 772–779.

Todd, N. A model of expressive timing in tonal music. *Music Perception,* 1985, *3,* 33–58.

Todd, N. A computational model of rubato. *Contemporary Music Review,* 1989, *3,* 69–88.

Todd, N. The kinematics of musical expression. Manuscript submitted for publication, 1992.