

## On Determining the Basic Tempo of an Expressive Music Performance

BRUNO H. REPP

Haskins Laboratories, 270 Crown Street, New Haven,  
CT 06511-6696, USA

In an expressive music performance, the local tempo varies continuously and often asymmetrically around an implied (nominally constant) basic tempo. This preliminary study explored how pianists organise the expressive timing structure around an intended basic beat rate, how listeners judge the basic tempo of such a modulated performance, and what objectively measurable property of the performances the intended and/or judged tempi might correspond to. Two pianists played Robert Schumann's *Träumerei* three times at each of three intended tempi cued by a metronome. Tempo judgments (metronome settings) for the initial 8 bars of each performance were subsequently obtained from listeners who were pianists themselves. The judged tempi were generally slower than the intended tempi, which was attributed to a tendency of the performers to play slower than intended, especially at the faster tempi. The timing microstructure of each performance was quantified in terms of the frequency distribution of (raw or transformed) beat inter-onset intervals (IOIs). The judged tempi were generally close to the reciprocal of the mean of this distribution (transformations had little effect on the mean tempo), which thus seems to be the parameter that best corresponds to the perceived beat rate of an expressively modulated performance, at least when there are no extreme *ritardandi*.

### Introduction

The tempo at which a piece of music is to be performed is often indicated in the composer's score by a metronome (M.M.) number, such as "M.M. = 60", meaning 60 beats per minute or a beat duration of one second. Such an instruction is easy to follow when the music in question has a steady rhythm; if necessary, the obedient performer can practice with the metronome ticking, aligning beats with ticks. Similarly, it is easy to determine the tempo of such a performance by finding the metronome rate that aligns itself with the beats, or by counting the number of beats in one minute of music.

When the beat is relatively slow and the music is highly expressive, however, the tempo is not so easily implemented or determined. Expression calls for considerable deviations from rigidity in timing, and these deviations are more often lengthenings than shortenings of beat durations, because *ritardandi* have the important function of marking structural boundaries at several levels (Todd, 1985; Repp, 1992). As a result, a count of the number of beats per minute may underestimate the tempo. Nor is it possible to align a metronome with the beats: Since an expressive performance calls for a continuous modulation of the tempo, it is virtually impossible to find a stretch of music during which the tempo is constant, and the occurrence of *ritardandi* destroys the synchrony between metronome and music. What, then, is the tempo of such a performance? And how does a performer implement an intended tempo *con espressione*?

It might be argued that such performances do not have a basic tempo. This objection can be dismissed, however, because even highly expressive pieces are commonly preceded by metronome indications in the score. If composers and editors think such music has a tempo, there must be a principled (if intuitive) way of following their tempo prescriptions. Conversely, musicologists and music critics are often interested in how the tempo of a performance compares with the metronome number in the score. Thus, there should be a way of determining the underlying beat rate (M.M. number) of even a highly expressive performance. The problem of identifying the underlying or basic tempo of a performance is also relevant to methods of "quantisation" in automatic rhythm detection and computer transcription of music (see, e.g. Desain and Honing, 1989, 1992), to theories of human rhythm perception (e.g. Jones and Boltz, 1989; Desain, 1992), and to performance modelling and synthesis (e.g. Todd, 1985).

That this is a non-trivial problem which apparently has not been addressed directly in the literature became evident to the author during a recent analysis of the expressive timing patterns in 28 performances of Robert Schumann's well-known piano piece, *Träumerei* (Repp, 1992). Most editions of the score contain either of two tempo prescriptions, one (M.M. = 100) being attributed to the composer and the other (M.M. = 80) to his wife, the pianist Clara Schumann. Questions of authenticity aside, both tempi seem unusually fast to contemporary ears (cf. Brendel, 1981, 1990; but see also Csipák and Kapp, 1981). Most, probably all, of the 28 performances examined by Repp (1992) were slower than M.M. = 80. But what exactly *were* their tempi?

Informal clues to the intended tempi of two of these performances were available. In an article written at about the time his recording was made, Alfred Brendel (1981) mentions that his own preferred tempo for *Träumerei* is M.M. = 69 (a statement repeated in Brendel, 1990). Another recording was by Fannie Davies, who had been a pupil of Clara Schumann. Although her performance was recorded much later in her career (in 1928), it was the fastest in the set, which suggested that she may have intended to adhere to her teacher's recommendation of M.M. = 80. By a curious coincidence, the reciprocals of both of these informal tempo estimates happened to correspond to the 16th percentile of the total beat inter-onset interval (IOI) distribution of each pianist's performance. This percentile was unexpectedly low, however, suggesting that both informal tempo estimates may have been too high. (Fast tempi imply short IOIs.)

Nevertheless, these informal observations suggested a hypothesis: that the tempo of an expressive performance might be best characterised by an M.M. number corresponding to some fixed point along the IOI distribution, perhaps the median (50th percentile) or some point below it, or the mode (most frequent value). Alternatively, while the arithmetic mean of a skewed IOI distribution underestimates the basic tempo, the possibility remains that the mean of a *transformed* IOI distribution comes close to the "real" tempo. Reasonable choices of transformations are logarithms (the antilogarithm of whose mean is the geometric mean of the original IOI distribution) and reciprocal values, which (when IOIs are expressed in fractions of a minute) represent estimates of *local tempo*. A logarithmic transformation could be

justified on the basis that it takes into account Weber's law, which holds approximately for duration discrimination above 300 ms (e.g. Drake and Botte, 1993). In fact, Wagner (1974) used the geometric mean to estimate tempo, based on this consideration. The reciprocal transformation seems reasonable because it represents tempo directly. Both transformations have the effect of reducing the asymmetry of the IOI distribution. As to the location of a given tempo estimate on the cumulative IOI distribution, it should be noted that it depends only on the order of IOI values and therefore is unaffected by any monotonic transformation (though the reciprocal transformation reverses the order of values).

The following exploratory experiment investigated the relationships among (1) the tempo intended by the performer, (2) the actual timing microstructure of the performance, and (3) the tempo judged by musically trained listeners. The specific question of interest was whether the underlying tempo of a performance can be expressed in terms of an invariant statistical parameter of the (original or transformed) IOI distribution.

FIG. 1

The initial 8 bars of Schumann's *Träumerei*, with the phrase-final chord extended through the second half of bar 8.

The music was again Schumann's *Träumerei*. The study focused on the initial section of the music, which is shown in Figure 1. This section begins with the upbeat in "bar 0" and ends with the chord on the third beat of bar 8, a total of 32 quarter-note beats. There were three reasons for restricting attention to this section: (1) The performers, who intended to follow a tempo cued by a metronome, naturally remembered the tempo best at the beginning of the performance and were expected to be most accurate there. In fact, there is evidence from a more detailed analysis of these and other performances of *Träumerei* that the tempo usually slows down later in the piece, presumably

for expressive reasons (Repp, 1992, 1994). (2) Performance excerpts of the same length were used in the tempo judgement task described below, to limit the duration of the test. (3) The initial section does not contain any extreme *ritardandi*, such as typically occur at the end of the middle section (bar 16) and at the end of the piece (bars 23–24). It seems likely that players and listeners would not include such extreme tempo changes in their mental estimates of the basic tempo. It seemed that the first 8 bars contain sufficient local tempo variation to address the question posed here in a meaningful, if preliminary, way. (For plots of the “timing profiles” of the performances analysed here, see Repp, 1994.)

## PERFORMANCE ANALYSIS

### Methods

**Pianists.** Two pianists participated. One was a professional musician (LPH) in her mid-thirties; the other was a serious amateur (BHR, the author) in his late forties. Both were thoroughly familiar with Schumann’s *Träumerei* and had played it many times previously.

**Equipment.** The instrument was a Roland RD250S digital piano connected to a microcomputer that registered performance data in MIDI format (onset time, offset time, and key velocity), with a temporal resolution of 5 ms. “Piano 1” sound was used, and a simple on/off sustain pedal switch (DP-2) was taped to the floor. The sound output was monitored over earphones by the performer. A brand new Franz LM-FB-4 electric metronome stood nearby on a table.

**Procedure.** Each pianist performed the complete piece 9 times from the score, three times at each of three intended tempi. At the beginning of the recording session, she/he warmed up on the keyboard and then played the piece once at her/his preferred tempo while being recorded. The beginning of this MIDI recording was subsequently played back, and the pianist set the metronome to the beat frequency that best corresponded to the tempo of the performance (as in the tempo judgement task described below). The settings chosen by LPH and BHR were M.M. = 63 and 66 (“medium tempo”), respectively. The recording of this initial performance was discarded.

The desired tempo for each subsequent performance was indicated by the metronome, which was left running at the desired speed for a while and was turned off just before each performance started. “Slow” and “fast” metronome settings were chosen by the author to surround the medium tempi: M.M. = 54 and 72, respectively, for LPH, and M.M. = 56 and 76, respectively, for BHR. These tempi were intended to be within the range of aesthetic acceptability for *Träumerei* (cf. Repp, 1992) and thus did not force the pianists to play in an unnatural manner. LPH played in the order slow-fast-medium (repeated twice), whereas BHR played in the order medium-slow-fast (repeated twice). The performances were free of hesitations and major technical accidents, and were judged by the author to be appropriately expressive renditions of the score. (See Repp, 1994, for a more detailed analysis of their expressive microstructure.)

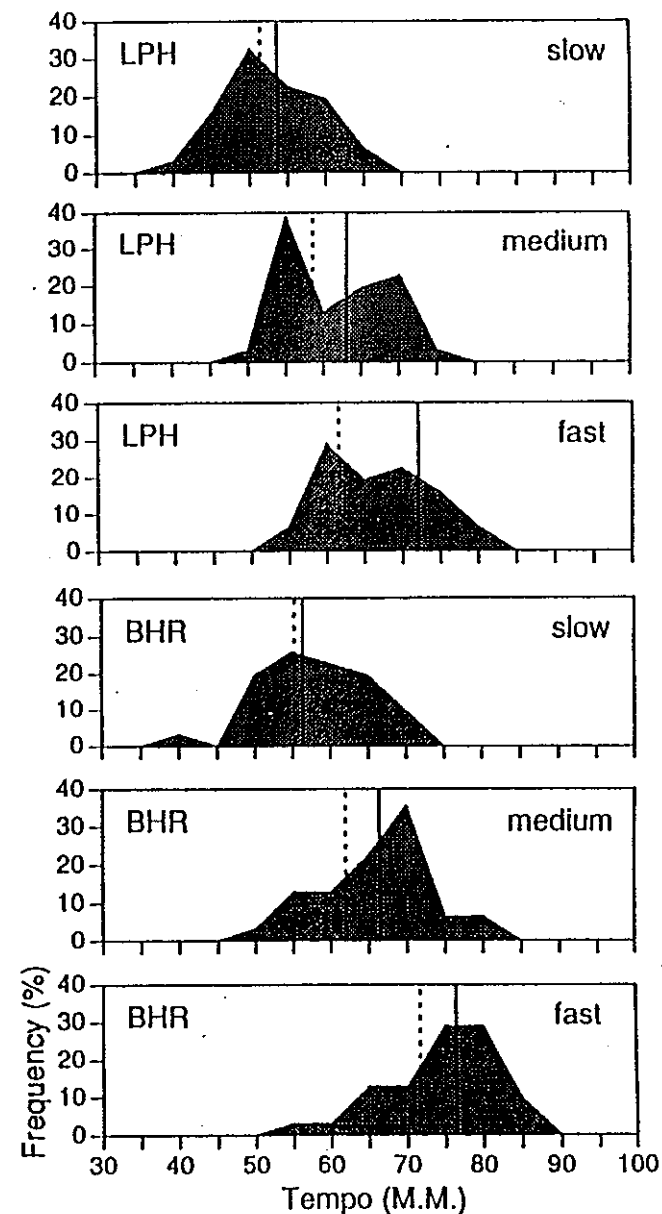


Fig. 2

Frequency distributions of local tempi at each of three intended tempi for the two pianists, based on IOIs averaged over the three performances at each intended tempo. The bin width is 5, and frequency values are plotted at the upper limits of bins. Solid vertical lines indicate intended tempi, dotted lines judged tempi.

**Analysis.** To determine beat (quarter-note) onset times, the tone with the highest pitch in any cluster of nominally simultaneous tones falling on a beat was picked. Inter-onset intervals (in milliseconds) were subsequently computed from this reduced "MIDI score". Missing beat onsets (of which there were 4; see Figure 1) were interpolated by subdividing longer IOIs into smaller intervals of equal duration. Average performances for each intended tempo were obtained by linearly averaging the corresponding IOIs of each pianist's three individual performances. The raw individual and average IOIs were also transformed into logarithms and into local tempo estimates (beats per minute,  $M.M. = 60000/IOI$ ). Means were calculated and, for raw and logarithmic IOIs, transformed into tempo estimates ( $M.M. = 60000/\text{mean}$  and  $M.M. = 60000/e^{\text{mean}}$ , respectively).

## Results

Figure 2 shows the local tempo distributions of the two pianists' average performances at each of the three intended tempi. Local tempi varied by as much as 30 beats per minute. As expected, the distributions shifted towards faster tempi as the intended tempo increased. The solid vertical line indicates the intended tempo. Its location does not generally coincide with the mode of the distribution.

Figure 3a shows the percentiles of the cumulative local tempo distributions of the 18 individual performances that correspond to the intended tempi. It is evident that there is no fixed point along these distributions that characterises the intended tempo. For each pianist the percentile increases as the intended tempo increases. Also, there is a large difference between the two pianists, with LPH showing higher percentiles than BHR, and there is considerable variability within each tempo category.

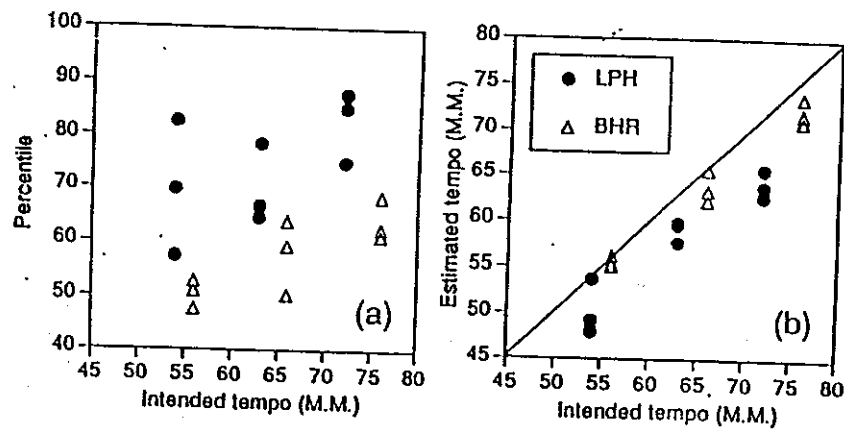


FIG. 3

(a) Percentiles of the cumulative local tempo distributions corresponding to intended tempi, for all individual performances. (b) The relationship between intended tempo and the mean of the local tempo distribution. The diagonal line indicates equality.

Figure 3b compares the intended tempo with the mean of the local tempo distribution for each of the 18 performances. It is evident that the mean consistently underestimates the intended tempo, more so for LPH than for BHR, and more so at fast than at slow intended tempi. (Analogous plots using tempo estimates based on the arithmetic or geometric mean of the original IOI distribution show similar but slightly larger differences.) The figure also suggests that LPH played relatively slower than BHR, even when the differences in intended tempi are taken into account.

The intended tempi thus do not correspond to any invariant parameter of the IOI distribution. However, it is possible that the two pianists did not implement the intended tempi accurately. In particular, they may have played slower than intended, and more so at fast than at slow tempi. Perceptual estimates of the tempi of these performances may shed light on this issue.

## PERCEPTUAL JUDGEMENTS

### Methods

**Listeners.** The listeners were nine skilled pianists, most of them graduate students of piano performance at the Yale School of Music, who were paid for their services. They were tested individually in a 1-hour session that began with a different task using the same materials (Repp, 1994).

**Procedure.** The initial 8-bar sections of the 18 performances were excerpted and stored in separate MIDI sequence files. The final chord was extended to provide a pleasing conclusion (*cf.* Figure 1). The author, who conducted the experimental session, called up these files in a different sequence for each listener, according to a counterbalanced schedule. Each sequence was constructed so that performances by the two pianists alternated, performances in the same tempo category did not follow each other, and each block of 6 included one performance of each pianist in each intended tempo category. The listener sat in front of a computer keyboard, wore earphones connected to the digital piano, and manipulated the metronome on the table next to the keyboard. He/she was shown how to start, stop, and restart MIDI playback by pressing certain keys. The task was to find the beat frequency that best approximated the tempo of each performance by adjusting the metronome and reporting the M.M. number, which was recorded by the author. The strategy for accomplishing the task was left up to the listener. He/she could take as much time as necessary, start and stop MIDI playback at will, use the metronome in sound (click and flash) or silent (flash only) mode, and adjust it while the music was on or off. Most listeners repeatedly alternated between listening and adjustment periods and took about 1 minute per judgement. The resolution of the metronome in the region of interest was 2 steps (beats per minute) up to  $M.M. = 60$ , 3 steps up to  $M.M. = 72$ , and 4 steps above that.

### Results

The tempo judgements were averaged over listeners. Their average standard deviation was 4 metronome steps, so that the average standard error was about 1.3 steps. The judged tempi are shown as a function of intended tempi in

Figure 4. It is evident that most performances were judged to have slower tempi than the pianists had intended, particularly when the intended tempo was fast. In fact, the pattern of these data is quite similar to that in Figure 3b. Although, in principle, the discrepancies between intended and judged tempo could represent systematic errors of judgement (*i.e.* tempo underestimation), it seems more plausible that they reflect performance deviations. After all, there were nine judges but only one pianist per performance.

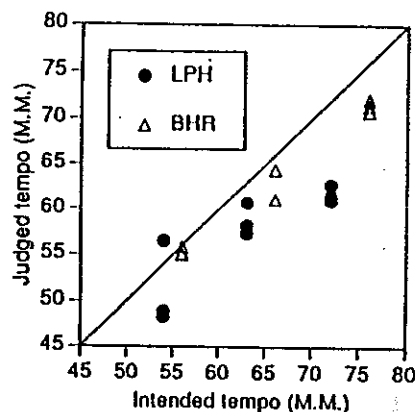


FIG. 4

Average judged tempo as a function of intended tempo for individual performances. The diagonal line indicates equality.

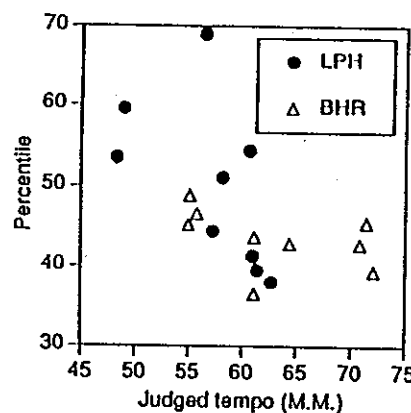


FIG. 5

Percentiles of the cumulative local tempo distributions corresponding to judged tempi, for all individual performances.

Assuming, then, that the listeners' average judgements represent reasonably accurate estimates of the basic tempi of these performances, we may ask now whether *they* correspond to a particular parameter of the tempo distributions. Figure 5 shows the percentiles of the tempo distributions that correspond to the judged tempi. It can be seen that there is still no constancy, especially not for pianist LPH. A glance back at Figure 2 also shows that the judged tempi (vertical dotted lines) do not generally coincide with the modes of the tempo distributions.

Figure 6 compares the judged tempi with the mean tempi, calculated either from the raw IOIs (Figure 6a) or from the local tempo estimates (Figure 6b); results based on the logarithmic IOIs are extremely similar to those in Figure 6b. Here there is a very satisfactory match. There were only small differences among the three types of objective tempo estimates, due to the absence of extreme asymmetries in the IOI distributions. The estimates in Figure 6a are slightly lower than those in Figure 6b, but the match with judged tempi is comparably good. Thus, these data do not favour a particular transformation of the IOI values. Rather, they suggest that any estimate of average tempo is a reasonable approximation of the basic (perceived) tempo of an expressively modulated performance. One may anticipate, however, that the mean of the local tempo distribution (Figure 6b) or the reciprocal geometric mean IOI will be preferable for more strongly skewed IOI distributions.

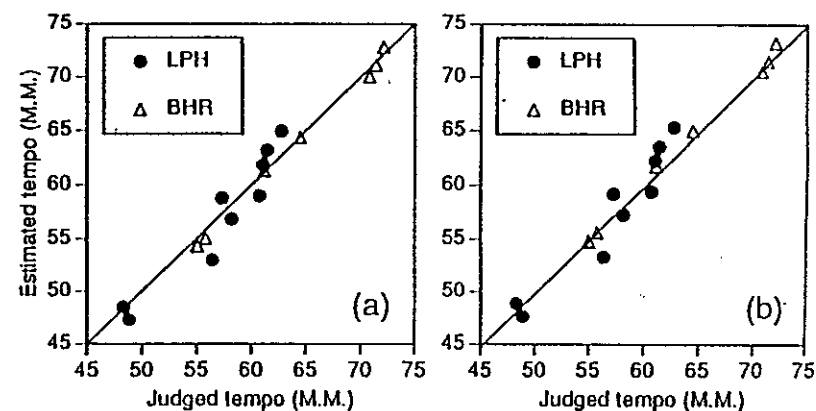


FIG. 6

(a) Judged tempo versus mean tempo based on raw IOIs. (b) Judged tempo versus mean tempo based on local tempo estimates. The diagonal line indicates equality.

### Discussion

This study addressed a question that apparently has not been asked previously in the psychological or musicological literature, though it seems related to the problem of "quantisation" in computer music applications (see Desain and Honing, 1989, 1992). Quantisation algorithms attempt to recover a rigid

metrical structure from an expressively modulated performance. Simple algorithms assume a constant tempo (a metric grid) and consequently make many errors when the timing modulations are large and/or asymmetric, as they often are. More sophisticated algorithms attempt to track tempo changes. In doing so, they negate (or at least do not address) the idea of a single basic tempo. Thus the goal of quantisation is somewhat different from the question pursued in the present study.

In studies of music performance, measured IOIs have often been plotted as percentage deviations from a horizontal baseline (see, e.g. Gabrielsson, 1987; Palmer, 1989). This tradition goes back to Seashore's (1938, p. 9) famous dictum that musical expression is "deviation from the regular". "The regular", in the case of timing, is a mechanically exact rendition of the underlying beat of the expressively modulated performance. In the performance studies referred to, the baseline was placed at the mean IOI. Somewhat unexpectedly, the present data seem to vindicate this procedure. However, it must be remembered that the excerpt investigated here did not contain extreme *ritardandi*, which would have pulled down the estimate of average tempo. Therefore, the author still prefers to avoid a baseline and to plot original IOIs on a log scale, which conveys relative as well as absolute magnitudes (see Repp, 1992, 1994).

The absence of extreme asymmetries in the present IOI distributions raises the question of whether the present findings would generalise to musical excerpts containing severe *ritardandi*. On one hand, there surely would be a larger difference between arithmetic, geometric, and local tempo means, probably favouring the latter. On the other hand, it seems implausible that a listener would include such extreme slow-downs in his or her estimate of basic tempo. Presumably there is a limit to what a listener is willing to accept as being played at the same basic tempo; beyond this limit, the tempo is simply perceived as changing or different. A more precise determination of this limit must await further research.

Some ambiguity remains in the present data, for as long as the basic tempo is not known, it is impossible to separate performance inaccuracy from systematic biases in tempo judgement. Although it seems rather clear that the present performers played slower than intended, there is no guarantee that the listeners judged the tempo accurately. For example, it is possible that the pianists, rather than playing too slow across the board, drifted towards their preferred (medium) tempo. If so, then the listeners must have consistently underestimated the basic tempo. This in turn might account for the unexpectedly close match of the judged tempi with the tempo estimates derived from the arithmetic mean IOI (Figure 6a). A tendency of musicians to underestimate tempi has been reported previously by Madsen (1979). This issue also requires further research.

Repp (1992: Table 3) based his tentative tempo estimates for 28 performances of *Träumerei* on the first quartiles (25th percentiles) of the distributions of eighth-note (half-beat) IOIs for the whole music. (The Brendel and Davies data, originally analysed in this format, had suggested percentiles near the first quartile.) It seems now that this measure probably overestimated the tempo, though the inclusion of the later sections of the piece, with their somewhat slower tempo and large *ritardandi*, may have reduced the error. Revised

estimates representing the mean local tempo during the first rendition of the initial 8-bar section, based on quarter-note (beat) IOIs, are indeed much slower than those reported in Repp (1992). Contrary to his own stated preference of M.M. = 69, Brendel's tempo is only M.M. = 58, and Davies's tempo, at M.M. = 71 still the fastest in the set, falls short of Clara Schumann's recommended M.M. = 80. Thus these initial clues towards defining the basic tempo appear to have been misleading: According to the present data, the basic tempo does not correspond to a particular percentile or the mode of the IOI distribution but rather to the mean of the (transformed) distribution. However, this conclusion will have to be tested further with more extensive sets of data.

#### Acknowledgements

This research was made possible through the generosity of Haskins Laboratories (Carol A. Fowler, President). Additional support came from NIH Grants RR05596 and MH51230. I am grateful to LPH for her patient participation in this study, and to Caroline Palmer, Henry Shaffer, and an anonymous reviewer for helpful comments on earlier versions of the manuscript.

#### References

- Brendel, A. (1981). Der Interpret muß erwachsen sein: Zu Schumann's *Kinderszenen*. *Musica*, 35, 429-433.
- Brendel, A. (1990). Testing the grown-up player: Schumann's *Kinderszenen*. In: A. Brendel, *Music sounded out* (pp. 142-153). New York: Noonday Press. [Revised and translated version of Brendel, 1981.]
- Csipák, K. and Kapp, R. (1981). *Träumerei*. *Musica*, 35, 438-443.
- Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception*, 9, 439-454.
- Desain, P. and Honing, H. (1989). The quantisation of musical time: a connectionist approach. *Computer Music Journal*, 13, 56-66.
- Desain, P. and Honing, H. (1992). *Music, mind and machine*. Amsterdam: Thesis Publishers.
- Drake, C. and Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception and Psychophysics*, 54, 277-286.
- Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K. 331): A comparison of five performances. In: A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Swedish Academy of Music (Publication No. 55).
- Jones, M. R. and Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96, 459-491.
- Madsen, C. K. (1979). Modulated beat discrimination among musicians and non-musicians. *Journal of Research in Music Education*, 27, 57-67.
- Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331-346.
- Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's *Träumerei*. *Journal of the Acoustical Society of America*, 92, 2546-2568.
- Repp, B. H. (1994). Relational invariance of expressive microstructure across changes in global tempo: An exploratory study using Schumann's *Träumerei*. *Psychological Research*, in press.
- Seashore, C. E. (1938). *Psychology of music*. New York: McGraw-Hill. (Reprinted in 1967 by Dover Publications.)
- Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception*, 3, 33-58.
- Wagner, C. (1974). Experimentelle Untersuchungen über das Tempo. *Osterreichische Musikzeitschrift*, 29, 589-604.