Determination of the meter of musical scores by autocorrelation

Judith C. Brown
Physics Department, Wellesley College, Wellesley, Massachusetts 01281 and Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 14 September 1992; revised 3 May 1993; accepted 29 June 1993)

A single voice extracted from the scores of compositions from the Baroque, Classical, Romantic, and Contemporary periods has been studied in order to determine the feasibility of the determination of musical meter by computer. The method of autocorrelation is appropriate for this calculation since it is a measure of the frequency of occurrence of events following an event at time zero. If a greater frequency of events occurs on the downbeat of a measure as predicted by Palmer and Krumhansl ["Mental Representations for Musical Meter," J. Exp. Psychol.: Hum. Percept. Perform. **16**, 728–741 (1990)], then a peak in the autocorrelation function should indicate the time for a single measure. The results of these computations indicate that computer determination of meter from score events is indeed possible. An example is included to show that this method of analysis can be applied to live performance data as well.

PACS numbers: 43.75.St

INTRODUCTION

Any discussion of the temporal properties of musical events involves the use of a number of terms which must be precisely defined. The following quotation from Dowling and Harwood (1986) clarifies the use in this paper of the most important of these terms.

Duration is the psychological correlate of time.

Beat refers to a perceived pulse marking off equal durational units.

Tempo refers to the rate at which beats occur, and *meter* imposes an accent structure on beats (as in "*one*, two, three, *one*, two, three ..."). Meter thus refers to the most basic level of rhythmic organization and does not generally involve durational contrasts.

Rhythm refers to a temporally extended pattern of durational and accentual relationships.

In this terminology then, rhythm is a general way of talking about the time-dependent properties of music. Meter refers specifically to the timing of written music, i.e., music that can be classified in the key signature as (3 4) or (4 4) etc, and the parsing of musical events into units called measures is what is meant by meter determination. An event in this context refers to the onset of a musical note.

The problem of human and/or computer determination of rhythm or meter has been studied by a number of workers (Longuet-Higgins and Lee, 1984; Lee, 1986; Povel and Essens, 1985; Schloss, 1985; Dannenburg and Mont-Reynaud, 1987; Chung, 1989; Katayose *et al.*, 1989; Allen and Dannenberg, 1991; Rosenthal, 1991, 1992; and Desain, 1992). See Rosenthal (1991, 1992) for a discussion of previous work. Some preliminary results in this study were reported by Brown (1991, 1992).

The determination of rhythm/meter is an example of a quantity, which is easily extracted from large quantities of input data by human beings, but which represents considerable difficulties as a computational problem. Fortunately, in contrast to the audio rates required for calculations on fundamental frequency tracking, musical tempo variations occur in time frames measured in seconds, which results in significant data reduction.

Palmer and Krumhansl (1990) analyzed the number of note occurrences at different metrical positions in a measure. They assumed the actual measure as indicated in the score in order to count up these note occurrences. The counts were made on the scores of musical compositions from the Baroque, Classical, Romantic, and Contemporary periods, each of which included examples of different meters. They concluded that the number of note occurrences depends upon the meter and that the highest number is at the position of the beginning of the measure (or measure onset).

A more interesting and challenging question from the point of view of signal processing and machine perception is to determine if a computer is able to pick out directly the measures of a piece of music from the score or from a musical performance of the score. It is this question which will be addressed. With a few additions the same portions of the same scores studied by Palmer and Krumhansl were chosen for our calculations, since these are representative pieces with different meters from the different periods. These compositions are listed in Table I. There were at least 20 measures included in each of the pieces studied. The compositions are listed in Table I.

The method of autocorrelation was chosen for this calculation since it is a measure of the number of events following an event at time zero. If events are correlated from measure to measure with a higher frequency of events occurring with the time separation of the measure, then peaks in the autocorrelation function should indicate the times at which measures begin. This differs from the study of Palmer/Krumhansl in that all time separations for all events are included in the present calculation; whereas they examined the sum of all events falling on a particular beat in the measure.

An advantage of the autocorrelation method is that it

 0001-4966/93/94(4)/1953/5/$6.00

TABLE I. Summary of results arranged by musical period.

Integration time:	Meter	Half of score	Short =3 s	Maximum time
Mary Had a Little Lamb	(4 4)	W	W	W
Bach Invention #13	(4 4)	W	–	W
Bach Suite #3	(4 4)	W	W	W(1/2)
Bach Suite #4	(6 8)	–(m)	W(1/2)	–(m)
Bach Suite #6	(2 4)	W	W(1/2)	W
Brahms Op. 118 No. 2	(3 4)	W	W	W
Brahms Op. 119 No. 2	(3 4)	–	–	–
Brahms Op. 119 No. 3	(6 8)	W	W	W
Brahms Op. 119 No. 4	(2 4)	W	W	W
Mozart Sonata K. 310				
1st Movement	(4 4)	W	–	W
2nd Movement	(3 4)	–	–	–
Mozart Sonata K. 311				
1st Movement	(4 4)	W	W	W
2nd Movement	(2 4)	–(m)	–	W
Mozart Sonata K. 545	(4 4)	W(1/2)	W	W(1/2)
Mozart Sonata K. 576	(6 8)	W(1/2)	W(1/2)	W(1/2)
Shostokovich Prelude 3	(4 4)	W	W	W
Shostokovich Fugue 3	(6 8)	W(1/2)	W(1/2)	W(1/2)
Shostokovich Fugue 6	(3 4)	W	W	W
Shostokovich Fugue 11	(2 4)	W(1/2)	W(1/2)	W
Number of wins out of 19		15	14	16

does not rely upon postulates about mental processes, and examines the "physical properties" of the notated score. It is based upon score (or performance) information only, with no grammar or rule-based heuristics intended to incorporate considerations of human perception of musical structure. Thus we make no claim for an autocorrelation mechanism for human perception. There is, nevertheless, convincing evidence that such a mechanism exists for information processing by bats (Suga, 1990), and, whatever the mechanism, it is probable that humans do have access to the periodicities turned up by autocorrelation.

I. PROCEDURES AND CALCULATIONS

A single melodic line or voice was extracted from the score for analysis. An input file for the analysis was constructed consisting of weighted amplitudes at the positions of onsets of the notes and zeros (0's) elsewhere. The weighting was based on the note duration as indicated in the score; for example a half note would be given an amplitude twice that of a quarter note. This weighting is roughly equivalent to the so-called interonset interval (defined as the time between successive onsets and abbreviated IOI), which is thought to be a cue in rhythm perception by humans (Lee, 1986). The results weighted with the IOI are slightly better overall than those with equally weighted note events obtained in preliminary calculations. An effective sampling rate of 200 per second was chosen because the typical overall resolution of most MIDI (musical interactive digital interface) devices is about 5 ms; thus all of the software written for this study can be used in the next stage of this work for the analysis of live performances

MOZART K. 310 Second Movement IOI Wt

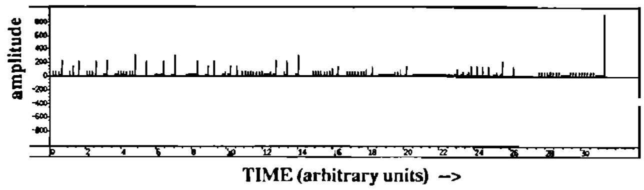

FIG. 1. Score events in Mozart K. 310 Second Movement weighted with the interonset interval (IOI) plotted against time in arbitrary units where the quarter note is given the value of 0.25.

recorded in standard MIDI file format. The resulting input file is illustrated in Fig. 1 for Mozart's Sonata K. 310 Second Movement. This example had the most complex temporal structure of any of the pieces studied.

The melodic line $x[n]$ was then subjected to a short-time autocorrelation calculation defined as

$$A[m]=\sum_{n=0}^{N-1} x[n]x[n+m], \qquad (1)$$

where the average is taken over N samples and m is the autocorrelation time in samples. The adjustable parameters for the calculation are the *upper limit for the autocorrelation time* (maximum value of m above) and the *time over which the autocorrelation is averaged or integration time N.*

For our calculation the upper limit of the autocorrelation time was chosen so as to include several measures, and the integration time was varied to determine its effect on the result. For example, humans are thought to determine the meter rather quickly so a perceptually based meter determination would necessarily have a short integration time.

II. RESULTS FROM SCORES

An example of the autocorrelation results is graphed in Fig. 2 for score events in the Fugue No. 6 by Shostokovich. In this and other figures the peak at autocorrelation

SHOSTOKOVICH FUGUE NO 6 (3 4) IOI Wt

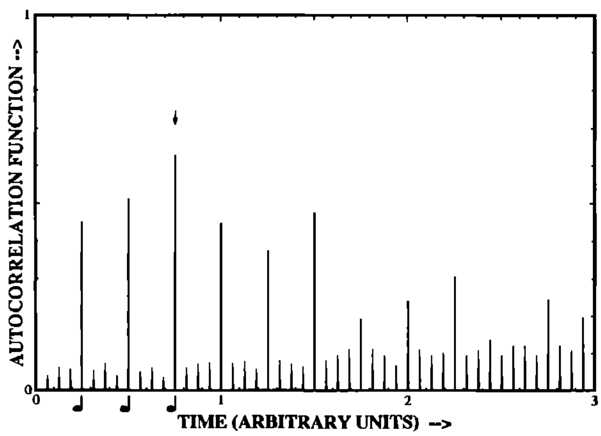

FIG. 2. Autocorrelation function for score events in Shostokovich Fugue No. 6 weighted with the IOI plotted against autocorrelation time in arbitrary units where the quarter note is given the value of 0.25. The arrow indicates the correct position of the measure as notated in the score.

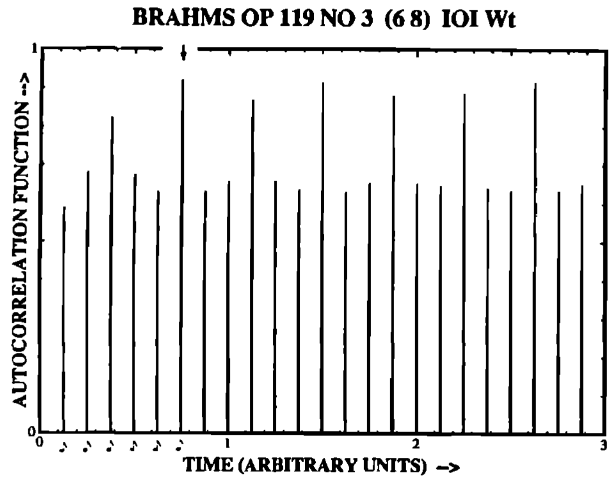

FIG. 3. IOI weighted autocorrelation function for score events in Brahms Op. 119 No. 3. The arrow indicates the correct position of the measure as notated in the score.

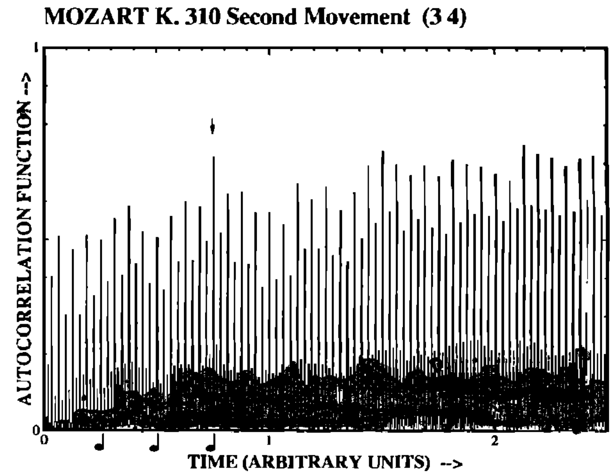

FIG. 4. Narrowed autocorrelation function for score events in Mozart K. 310 Second Movement plotted against autocorrelation time in arbitrary units where the quarter note is given the value of 0.25. The arrow indicates the correct position of the measure as notated in the score.

time 0 represents the total number of events on which the calculation was carried out, and is normalized to 1. The piece is in (3 4) time and the highest peak falls at 0.75 indicating the position of three quarter notes with our arbitrary time choice assigning 0.25 s to the quarter note. We have emphasized this by labeling the horizontal axis with quarter notes in the appropriate positions out to the position of the measure. Within the measure there are three major peaks with a 0.25-s separation which indicates a beat of one quarter note. Thus it is possible in this case to predict not only the measure position but also the key signature of (3 4) correctly.

As second example, the autocorrelation results for Brahms Op 119 No. 3 are shown in Fig. 3 where the key signature (6 8). Again it would be possible to predict the key signature as well as the meter. The measure determination was considered successful if the peak at the correct position is greater than the peaks preceding it and greater than those following it up to the position of the second measure. This criterion was adopted to avoid the problems with peaks at integral multiples of the measure for the following reasons.

First, this method was developed for "tracking" human performances, and here the timing would certainly be known to much better than a factor of two. Second, one can use various heuristics (Brown and Zhang, 1991) such as requiring the winning peak to be more than a certain percentage greater than the peak at half its time value; for a periodic function this criterion would not hold for the peak at the position of two measures. Third, the vast majority of scores are in 2/4, 3/4, or 4/4 so one could make a case for analyzing peaks over the time corresponding to the first five quarter notes in the score limiting the search more than with our method. Finally, we could use the method of narrowed autocorrelation described below to sharpen the first peak.

The IOI weighted autocorrelation function for the second movement of Mozart's Sonata K. 310 did not indicate the meter of the piece successfully by the criteria previously stated. The temporal behavior of the input for this autocorrelation calculation is shown in Fig. 1. This piece was chosen as an example of extremely complex temporal structure. A narrowed autocorrelation calculation (Brown and Puckette, 1989), where terms of the form $f(t)f(t+2\tau)$, $f(t)f(t+3\tau)$, etc. are included, was carried out for this piece. (See Fig. 4.) This resulted in a sharpening of the peak at the position of the measure giving a successful determination of the meter. This could be predicted, and this calculation could be carried out on all pieces for a more definitive determination of the position of the measure. This calculation is more intensive computationally, however, and was not necessary since results with ordinary correlation are quite satisfactory.

The overall results are summarized in Table I. The pieces studied are given in the left column followed by the meter as noted in the score. Results for three different calculations are given in the next three columns. For the calculation of the first column the average of the autocorrelation function was taken over half of the segment studied [$N=1324$ to 3900 in Eq. (1)].

In the last two columns, the integration time was varied with a very short time [$N=600$ in Eq. (1)] in the next to last column and a longer time, chosen to be the maximum possible for each piece, in the last column [$N=1650$ to 7200 in Eq. (1)].

A "W" (for win) in the column indicates success in determining the position of the measure by the criterion described previously. The notation W(1/2) indicates that this criterion was met for the peak at the position of half the measure, for example at the (2 4) position when the score notation indicates (4 4) time. This was considered to be successful, since human listeners make this judgement as well. Also some of these peaks at the position of the half measure were only a percent or so higher than those at the notated position.

Failures to determine the position of the measure or half measure by the above criterion are denoted by a minus

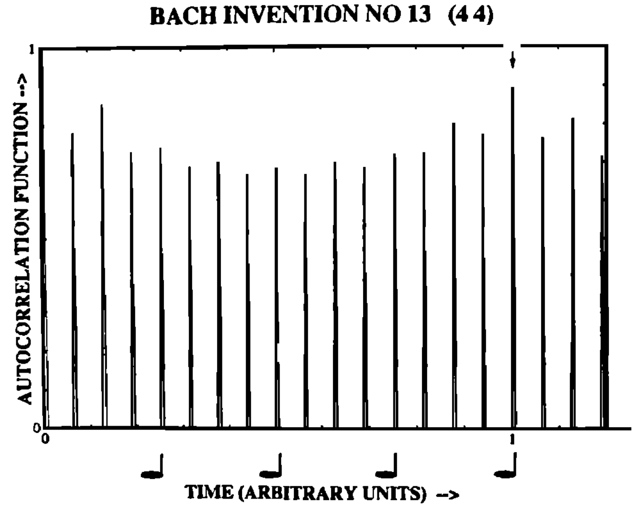

FIG. 5. Autocorrelation function for score events in Bach's Invention No. 13 plotted against autocorrelation time in arbitrary units where the quarter note is given the value of 0.25. The arrow indicates the correct position of the measure as notated in the score.

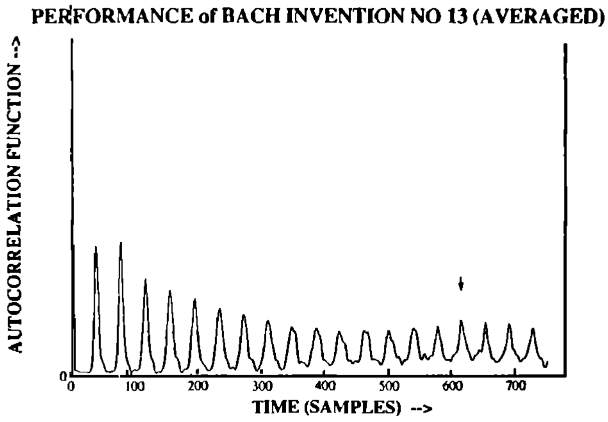

FIG. 6. Autocorrelation function for a performance by Michael Hawley of Bach's Invention No. 13 plotted against autocorrelation time in samples. One second is equal to 200 samples. The performance data was smoothed by spreading out each event over 20 ms of time relaxing the simultaneity condition for the autocorrelation calculation.

sign. The minus sign followed by "(m)" indicates a marginal miss, in many cases peaks of visually indistinguishable height when graphed, but with a slightly lower peak at the position of the measure when the numbers are examined.

Of the 19 pieces studied, the best score of 16 wins was made with the maximum integration time. Other integration times resulted in only one (integration time of half the score) or two (short integration time) more errors. Thus results are largely independent of integration time.

III. RESULTS FROM PERFORMANCE

Much information is available from these autocorrelation calculations in addition to that used for the extraction of the meter of the piece. One of the most interesting possibilities is the comparison of autocorrelation calculations of actual performance data with the calculations which we have done using data from the score. Here one can extract information regarding a performer's timing choices for musical purposes, as well as testing the length and accuracy of human temporal memory.

Comparisons of score and performance data can be seen in Figs. 5 and 6. The autocorrelation function obtained from the score for the upper voice of Bach's Invention No. 13 appears in Fig. 5. The autocorrelation function obtained from performance data in standard MIDI file format taken from a performance by Michael Hawley on a computer monitored Bosendorfer SE concert grand piano appears in Fig. 6. The upper voice was extracted from a normal performance (two hands) for the calculation. The peaks on the first couple of note events are extremely sharp in the performance data indicating near perfect memory for the timing of the preceding note over this short time frame. As the autocorrelation time increases, the performance peaks become broader indicating poorer timing or memory or conceivably a relaxation from a rigid beat for musical purposes. More performance data is needed to distinguish among these mechanisms.

IV. DISCUSSION

Results overall were good and offer promise for the determination of meter for performances by a live performer playing duets with a computer as second performer. Here it is necessary for the computer to be aware of the tempo of the live performer in order to adjust its own tempo to keep in synchrony. For the analysis of performance data there exists the option of weighting the events by their dynamic levels, and it is probable that the inclusion of this amplitude information for performances would make the results even better.

The achieving of slightly better results with IOI weighting is consistent with theories about human parsing of metrical events. Lee (1986) states that listeners attempt to place long notes on strong beats as they listen to a piece and attempt to determine its meter. Our results would indicate that composers write in such a way as to give these cues. Lee also maintains that listeners attempt to establish a metrical analysis as soon as possible; whereas we did not find evidence that these scores were written in such a way as to make it easier to determine the meter from a short segment at the beginning of the piece than from a long segment.

This method holds promise as an excellent tool for the study of memory, motor control, and performance. The method may be particularly valuable in objective tests of musical questions about style and expression.

ACKNOWLEDGMENTS

I am very grateful to David Rosenthal for many hours of interesting discussions and for his extremely valuable suggestions on approaching this problem. Edward Carterette was extremely generous with his time and made many invaluable suggestions which were incorporated into the final version of the manuscript. Finally I would like to thank Caroline Palmer for taking the time to read and comment on the manuscript.

Allen, P., and Dannenberg, R. (1991). "Tracking Musical Beats in Real Time," Proceedings of the 1990 International Computer Music Conference, 140–143.

Brown, J. C., and Puckette, M. S. (1989). "Calculation of a narrowed autocorrelation function," J. Acoust. Soc. Am. **85**, 1595–1601.

Brown, J. C. (1991). "Determination of Musical Meter Using the Method of Autocorrelation," 1991 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, 10.1.

Brown, J. C. (1992). "Determination of musical meter using the method of autocorrelation," J. Acoust. Soc. Am. **91**, 2374–2375.

Brown, J. C., and Zhang, B. (1991). "Musical frequency tracking using the methods of conventional and narrowed autocorrelation," J. Acoust. Soc. Am. **89**, 2346–2354.

Chung, J. (1989). "An Agency for the Perception of Musical Beats," MS thesis in EECS, MIT, Cambridge, MA.

Desain, P. (1992). "A (De)Composable Theory of Rhythm Perception," Music Percept. **9**, 439–454.

Dannenberg, R. B., and Mont-Reynaud, B. (1987). "Following an Improvisation in Real Time," Proc. ICMC 241–247.

Dowling, W. J., and Harwood, D. L. (1986). *Music Cognition* (Academic, New York), p. 185.

Katayose, H., Kato, H., Imai, M., and Inokuchi, S. (1989). "An Approach to an Artificial Music Expert," Proceedings of the International Computer Music Conference, 139–146.

Lee, C. S. (1986). "The Rhythmic Interpretation of Simple Musical Sequences: Towards a Perceptual Model," in *Musical Structure and Cognition*, edited by P. Howell, I. Cross, and R. West (Academic, London), pp. 53–69.

Longuet-Higgins, H., and Lee, C. (1982). "The Perception of Musical Rhythms," Perception **11**, 115–128.

Palmer, C., and Krumhansl, C. (1990). "Mental Representations for Musical Meter," J. Exp. Psychol.: Hum. Percept. Perform. **16**, 728–741.

Povel, D. J., and Essens, P. (1985). "Perception of Temporal Patterns," Music Percept. **2**, 411–440.

Rosenthal, D. (1991) "Emulation of Human Rhythm Perception," Comput. Music J. **16**, 64–76.

Rosenthal, D. (1992) "Machine Rhythm: Computer Emulation of Human Rhythm Perception," MIT Media Lab Ph.D. thesis.

Schloss, A. (1985). "On the Automatic Transcription of Percussive Music—from Acoustic Signal to High-Level Analysis," Ph.D. thesis, Department of Music Report No. STAN-M-27, Stanford University.

Suga, N. (1990). "Cortical Computational Maps for Auditory Imaging," Neural Networks **3**, 3–21.