Empirical Musicology: An Attempt to Decipher the Mysteries of Music Performance

A case study on Bruno Repp’s Diversity and commonality in music performance: An analysis of timing microstructure in Schumann’s “Träumerei”

Maria Beatriz Ramos

Main Supervisor: Dr. Henkjan Honing
Second Supervisor: Prof. Dr. Rokus de Groot

Final Essay
MA Degree
Musicology department
Universiteit van Amsterdam
August 2005
1.0 Introduction

Why has Glenn Gould’s Goldberg Variations become the milestone it became in the history of piano performance? Why does Furtwängler occupy a high rank in the hierarchy of the numerous conductors of the 20th century? Why do you like his playing and do I prefer hers? In short, how do we formulate in words, more specifically, how do we scientifically substantiate the reasons and factors which cause our hearts and ears to prefer one performance over another?

As the development and broadening of the field of Musicology has and is taking place, the sole focus of the analyses of the score, which for so long had been the nerve center for music theorists, has left its headquarters and welcomed a new horizon. In the last two decades an important shift has occurred from music as an art (or art object) to music as a process in which the performer, the listener and music as sound play a central role. Even though it seems obvious that a score is nothing without its performance, the measurement of musical performances, which is of great interest to studies in musicology, music psychology and music performance practice, have for a long time not been considered the main (attainable) issue. One of the reasons for this was the simple fact that musicologists found their hands tied without methods and devices able to efficiently process the various performance data. With the advancement of computer techniques and technology as a whole, significant developments took place, which endowed musicologists with valuable tools, without which the search for the understanding of musical performance from a musicological point of view could not be efficiently initiated. These technical developments brought forth instruments, amongst others the invention of MIDI, which make the process of capturing performance data easier and in a way, closer to apprehending what for so long was left untouched, reasoned by some to be simply the “magical” and “mysterious” aspect of music.

As long as it can be remembered, the tools used to distinguish and describe a performance or performer have been mostly based on mere adjectives; “An elegant player”, “a brilliant performance”, “a noble interpretation”. Although volumes have been written about different performers and their characteristics, these discussions rarely go beyond mere descriptions, and the vocabulary used (such as the adjectives quoted above) are not specifically linked to particular performance properties. The issue of intersubjectivity can reason that among musicians, avid music lovers, etc., these terms have gained a more defined and concrete significance. They became shared meanings constructed by this particular musical community in their interaction through the years and have been used as an everyday resource to interpret their meanings. If a music critic describes a performer as “noble”, even though it can remain a rather abstract term for the layman, the so called “in-crowd” may automatically associate

---

1 Høning (2004a: 241)
2 Musical Instrument Digital Interface; commercial standard for the exchange of information between electronic instruments and computers (See glossary).
3 Intersubjectivity refers to the shared meanings constructed by people in their interactions with each other and used as an everyday resource to interpret the meaning of elements of social and cultural life.
4 The musical world consists of a large variety of musical communities and groups. Each one has its own vocabulary. Musicians, music lovers, music critics, amongst others, of the classical ‘standard’ repertoire make up the particular musical community referred to above.
“noble” to a number of factors related to the intersubjectively established interpretation of the term. It is possible to do research on such usage by means of e.g. semantic differential methods. Still, the vocabulary used remains not specifically linked to particular performance properties. Music criticism is an art rather than a science, and the critic’s impressions, accurate as they may be, are filtered through an idiosyncratic web of personal experiences, expectations, preferences, and semantic associations. Disagreements within music criticism are very common.

Wilhelm Furtwängler was one of the great conductors of the 20th century, who died in 1954, and who left an important mark in the history of performance. Even though his contribution to the musical world can hardly be disputed, during the late twentieth century, there were two figures that could not disagree more with each other in regard to Furtwängler’s conducting qualities. They were Paul Henry Lang, who was a record critic for High Fidelity magazine in the late 1960s, and Peter Pirie, a musicologist and author of many books, including one on Furtwängler entitled “Furtwängler and the art of conducting” (1980).

According to Lang, Furtwängler was a ‘dyed-in-the-wool romantic, favoring arbitrary and highly subjective procedures in tempo, dynamics and phrasing’. Additionally, he wrote of the most famous of all Furtwängler’s interpretations, the 1951 recording of Beethoven’s Ninth Symphony, that it is a ‘stereo “reconstruction also playable on mono”, but it does not make any difference how you play it. This recording, with the Bayreauth Festspielhaus forces, is something of a treasured unique, though I can’t see the reason for the admiration’. At best, Lang concluded that Furtwängler’s recordings of Beethoven were ‘for collectors and historians, not for enjoyment’.

On the other hand, Peter Pirie, the “ultimate” Furtwängler connoisseur, could not disagree more with Lang’s characterization of Furtwängler’s conducting. For Pirie, the way Furtwängler performed Beethoven was anything but arbitrary; on the contrary, says Pirie (referring specifically to the first movement of the Ninth Symphony), ‘his interpretation analyzed the structure’. Pirie describes the first movement of Beethoven’s Ninth Symphony as the movement, ‘in which Beethoven left behind the classical allegro with its fixed tempo in favor of a flexible declamation’. What Pirie calls ‘flexible declamation’ lies at the heart of the Furtwängler style.

In an important article on performing the Ninth Symphony, Richard Taruskin characterized this style as follows: ‘his tempo are eternally in flux, accelerating and decelerating in great waves’. Lang thought – or affected to think – that these great waves were consequence of Furtwängler’s incompetence as a conductor; speaking this time of the Seventh Symphony, he refers to ‘Furtwängler’s inability to keep to a steady tempo…This is not always the result of deliberate intention; it is due rather to a certain lack of the sort of orchestral discipline we expect from our conductors’. It can be concluded that the serious disagreement between Lang and Pirie focus on Furtwängler’s modifications of tempo. Were they arbitrary and uncontrolled, as Lang maintained, or a response to structure, as Pirie asserts? Who is right and who is wrong? These types of situations and disagreements in music.

---

5 Repp (1992: 2547)
6 Lang (1978: 17)
7 Ibid. 24.
8 Lang (1978: 17)
9 Pirie (1980: 49)
10 Ibid.
11 Taruskin (1989: 244)
12 Lang (1978: 22)
criticism are a constant. Assertion is of course not enough. It remains subjective if
not grounded. What was needed was evidence that could help solve this dispute and clear this
impasse. Nicholas Cook, one of the important figures in the field of empirical musicology,
provided it.

Cook (1995) explores Furtwängler’s conducted performances of Beethoven’s Ninth
Symphony. In his study, Cook focuses on tempo in these performances. He assembles bar-by-
bar tempo information for the first movement of two live performances (1951; with Chor und
Orchester der Festpiele Bayreuth/ EMI CDH 7 69801 2, and 1953; with Wiener
Philharmoniker and Wiener Singakademie) available on CD (1951: EMI CDH 7 69801 2 / 1953:
DG 435 325-2). The technique he used to acquire the tempo data involved playing the CD in the
CD-ROM drive of a computer, and tapping the space bar of the computer keyboard in
synchrony with the onset of each bar. This resulted in detailed graphics where the timing
patterns of both of the recordings were illustrated. See Figure below:

![FIG.1. Timing patterns of Furtwängler’s 1951 and 1953 live recordings of Beethoven’s Ninth Symphony (first movement).](image)

By illustrating both 1951 and 1953 timing patterns in the same graphic, one can easily
compare the two. The picture is striking. The two live performances, recorded two years apart,
were almost identical in tempo fluctuation, with slight variations at times, most likely
reflecting a change in interpretation choices between the years. Such graphic clearly discredits
Lang’s criticism. One may not agree with Furtwängler’s interpretation and the various tempo
fluctuations characteristic of it. Yet, saying that these fluctuations resulted from his inability to
keep a steady tempo is a clear mistake. Cook’s graphics and their similarity prove that
Furtwängler knew exactly what he was doing and based his tempi in well-considered choices
of interpretation. The impasse could now be resolved.

Empirical musicology emerged from ‘a desire to ground theories on empirical
observation and to construct theories on the basis of the analysis and interpretation of such
observations’14. To the traditional music scholar, it must look for ‘all the world like science is

---

13 Fragment adapted from Cook (1995: 113)
14 Honing (2004a: 2)
muscling-in on musicology’. According to David Huron, eminent figure in the field of empirical musicology, ‘the rise of empiricism has nothing to do with “science”. It arises from within music scholarship, and is motivated by the desire to learn as much as possible from the information available to us -- including the additional information that might be assembled with a little effort.’

A way of describing a field of enquiry is by listing some of the questions its practitioners hope to answer. The following are an example of some of the questions that help to shape the field of empirical musicology.

- Why does rubato exist -- why isn’t music played strictly according to the notated timing?
- ‘What does the visual information contained in music performance offer the observer?’
- Do children experience music the same way adults do?
- Why do our musical preferences sometimes change over time?

At the moment, musicologists find themselves just at the beginning of exploring this yet infant strand of musicological research. They face various challenges. Much of the empirical research taking place in the field of music performance at the moment can only look at one factor of performing at a time (e.g. timing), and not integrate and look at several factors simultaneously, thus not being able to explain the relationship between the different factors, to name just one of the many problems. Furthermore, skeptic views and several question marks posed to the idea of we ever being able to formalize and objectively take a look at the art of music performance and all that occurs within and around it are a constant. Thoughts such as “music [is] too spiritual a thing to be measured” head much of the criticism. The belief is widespread that performance differences cannot be characterized objectively. The truth is though that so far, one cannot say sufficient research has been done, which provide evidence to justify this belief. In the contrary, several significant results have already been achieved through empirical research (e.g. Cook, 1995), which have led to genuine insights about music and musicality.

Bruno Repp’s Diversity and commonality in music performance: An analysis of timing microstructure in Schumann’s “Träumerei” (1992) is one of the landmarks in the young history of empirical musicology that played an important part in defining the field. In this study, Repp attempted to characterize the temporal commonalities and differences in the timing patterns among distinguished pianists’ interpretations of Robert Schumann’s “Träumerei” through the objective analysis of 28 performances of this work. Repp obtained the performance data necessary for his study by deriving the timing profiles from existing sound recordings of the 28 performances (the methods through which Repp obtained these data will later be discussed in

---

15 Huron (1999)
16 David Huron is head of the Cognitive and Systematic Musicology Laboratory at Ohio State University and author of significant books and articles on the topic of empirical musicology.
17 Huron (1999)
19 Wing (1940: 343)
20 Repp (1992: 2547)
21 Clarke & Cook (2004: 5; 78)
detail). By extracting the timing profile of every performance, Repp was able to objectively look at what all 28 pianists did in each of their specific timing pattern. This allowed Repp to later objectify the diversities and commonalities he was able to substantiate among these performances.

Most music lovers would agree that Arthur Rubinstein and Vladimir Horowitz were two very different pianists, whose performance of similar repertoire, are instantly distinguishable. But what is it, really, that makes them so different and individual? And do they have anything in common at all? How far can empirical musicology help us answer these questions?

The results obtained by Repp in his study are of significant value to the debate of how much one can possibly learn from objective performance analysis and its contribution to the musicological discourse. My goal in this essay is to introduce empirical musicology through Repp’s landmark, question and substantiate the value of empirical research for both musicologists and musicians, and attempt to forecast what the future may hold for this rising field of research.

---

22 Repp (1992: 2547)
1.1 Overview

This paper is organized as follows. The next chapter (Chapter 2) provides a quick overview of the role occupied by performance through history. What it meant to be a performer two hundred years ago in relation to what is expected from performing artists nowadays. How have technical and social developments influenced the performing arts and effected public’s expectations. Chapter 3 presents an overview of the most significant empirical performance studies so far, which have delineated the, yet young, history of empirical musicology. This overview culminates into Repp’s 1992 landmark. Chapters 4, 5 and 6 provide a detailed description of Repp’s 1992 research. An in-depth analysis of “Träumerei” is presented, including a thorough description of its historical background and a full-fledged structural analysis. Additionally, the analytical methods, which Repp applied during his research, are examined in detail, leading up to an elaborate discussion of two important results derived from Repp’s research. In order to fully understand the way Repp reached these results, each result and the path leading up to each of it, will be explicated step by step. Elaborate statistical analytical methods will be descriptively broken down, so that the reader can fully absorb every facet of the process and its results.

Chapter 8 holds the conclusion of this essay. Thoughts and reflections on the future of empirical musicology are presented, supported by a review of the results and discussions brought to light throughout the essay. Additionally, an interlude reflects notes on a meeting with Repp in July of 2005. Thirteen years after his 1992 study, I question him on what his projects have been since; what he forecasts for the future of empirical musicology and what he believes its contribution is to the musicological discourse, amongst others.
2.0 The art of performance

Classical music in our age revolves around the phenomenon of performance. One may wonder if this has not always been the case, still, in this case, performance is referred to in its 21st century definition, which differs greatly from its earlier connotations. In contrast to the history of performance as a whole, what we expect from a concert experience is something quite different in comparison to earlier centuries. Whilst well into the eighteenth century it was very common for performers to improvise, within certain norms, from the notated score, nowadays the supremacy of the score demands allegiance and preciseness from the performer. Additionally, premieres of works are no longer considered the main events of concert life. Instead, the public’s thirst for what has become known as ‘standard repertoire’ offered over and over again, gave rise to performers as grand personalities and superstar icons. The focus has shifted from the composer and work, to the musician and performance.

Young musicians in search of career opportunities participate in competitions where their playing (sometimes of the very same piece) is judged by a panel. Along with the concert hall, the public also encounters a vast offer in recording material. By entering any record store, one can run into several recordings of the same piece by different performers. To go even a bit further, one can even find two different performances of the same work by the very same artist (Think of Glenn Gould’s 1955 and 1981’s recordings of Bach’s Goldberg Variations). For some of the more popular works, the Schwann catalogue lists dozens of performances of the same work by various performers; if deleted records, available in libraries and private collections are counted, they may run into the hundreds or even thousands.

More than at any earlier period in musical history, the contemporary scene in serious music is dominated by the performer. As Repp points out, this is not only the result of clever marketing by record companies and the promotion of superstars, but also in great part this results from the ability of music lovers to distinguish and appreciate different performances. The more developed listeners are able to distinguish certain characteristics in a performer, which lead them to search out this performer’s concerts and recordings. As earlier mentioned, when describing why a particular performer catches one’s attention, most of the listeners rely in using abstract terms, such as adjectives, to describe the qualities of a certain performer. It is possible to conduct research on the intersubjectivity of the terminology used, still, the vocabulary remains not specifically linked to particular performances properties.

All topics related to above find their source in the differences between performers, and how our preference goes to the one, which in our eyes, brings about the most successful difference, as it were. Still, there are also strong commonalities, which we can find between interpretations of the same work. The source of these commonalities can be traced back to music education. Music teachers, even though each in their own particular way, try to transmit to

---

23 The Schwann catalogue consists of ‘Schwann Artist’ and ‘Schwann Opus’. ‘Schwann Artist’ is an annually published index of classical recordings by performer (under the subject headings: instrumentalist, vocalist, ensemble, and miscellaneous). Schwann Artist also includes an index of conductors, listing the ensemble which they conduct. Main entries include artist/group, composers and pieces, other artists on recording, recording date, label and catalog number, and SPARS code (three-letter code [A=Analog, D=Digital] which explains the recording, editing, and mastering processes used). Schwann Opus is published quarterly and lists currently available classical recordings on CD and cassette. Entries are indexed by composer and then alphabetically by piece.

24 Repp (1992: 2546)

25 See Lipman (1990)
their pupils unwritten rules of a performance tradition that goes back to 19th century central Europe, where most of the standard repertoire originated.\textsuperscript{26} What exists are certain norms of musical performance, which the performer is expected to pay attention to. These norms may change through the influence of factors such as time, group and location. In relation to the performance tradition linked to the standard repertoire mostly originated in 19\textsuperscript{th} century central Europe, despite various changes in performance practices during the last 200 years (most of them of narrowly technical nature), there are generally accepted norms of musical performance, according to which the artist’s actions are largely subordinated to the musical structure.\textsuperscript{27}

The task of conveying the musical structure to the listener is the primary task of the performer; as Repp expresses, ‘the expression of the musical structure, so it can be grasped and appreciated by the listener, and make an impression on him or her’.\textsuperscript{28} The performer has an array of techniques to which he can resort in order to succeed in achieving this goal. The variances we find between performers are due to the distinct manner each of them chooses to apply these techniques. There are but a few performers, which succeed in capturing the audience in the present marketplace with its fierce competition. Even though one may think there is already an abundance of ‘acceptable’ performances offered to the public at the current market, there is an even greater abundance of performances which are labeled unacceptable and which, unless they have the mark of inspired iconoclasm (as do some of the performances by the late Glenn Gould),\textsuperscript{29} collapse behind the iron curtain of the fierce competition, never entering the “stage”.

One can gather from the facts above that there are thus two basic aspects which constitute a musical performance, namely a normative aspect (i.e., commonality) that represents what is expected from a competent performer and is largely shared by different artists, and an individual aspect (i.e., diversity) that differentiates performers. The normative aspects emerge from, amongst others, the unwritten rules of performance tradition linked to the standard repertoire and what was discussed earlier. According to Repp, the individual aspect may be conceived of as deviations from a single ideal norm; more profitably, however, it may be thought of as individual settings of free parameters in the definition of the normative behavior.\textsuperscript{30} What happens is that variety emerges from the different interpretations of the set norm, without deviating so far as to breaking that norm. In short, different performers can agree that in a work a certain section of the musical structure should be highlighted, but the way in which each of them chooses to highlight it might be completely different. An example of a relatively free parameter is tempo. Two performances of the same work can be equally successful, while at the same time, differing highly in tempi. As Repp suggests, one may opt for a slow tempo and large tempo changes, while the other may prefer a faster tempo and smaller deviations from the rhythmic beat.

It becomes clear that there is no question in regard to the powerful influence the so-called ‘unwritten rules of performance tradition’ have on performers, and at the same time, how each in their own way chose to interpret these rules. Still, what has yet to be developed is

\textsuperscript{26} Repp (1992:2546)
\textsuperscript{27} Ibid.
\textsuperscript{28} Ibid. 2547. See also Lussy (1882); Riemann (1884); Stein (1962).
\textsuperscript{29} Ibid. 2546.
\textsuperscript{30} Repp (1992: 2547)
a vocabulary, which can objectively capture these diversities. The question is: what could these methods be, and what will they give us?

Particularly among musicians, objective performance analysis is often viewed with considerable skepticism. “It is sometimes unfortunate”, says Wing (1939), “that the attitude of the musician to his art is like that of the mother to her child. Both are inclined to resent any attempt on the part of an investigator to apply anything savouring of mathematics or the foot to their protégé”. 31

“...the belief is widespread that performance differences cannot be characterized objectively”. 32

Repp advocates that such a statement, though, cannot be made unless a real attempt in the direction of objective analysis has been made and has failed. This has clearly for Repp not yet been the case. He conveys that one may indeed wonder if objective analysis may ever be able to capture every facet of a performance. Yet, this should not be the vital question. The important question to be asked is how much can one learn, how much information can be extracted from objective analysis and how far can one, through this discoveries, unravel the mysteries of musical performance.

According to Repp, even a partial characterization in terms of verifiable and replicable observations can make a valuable contribution to our understanding of performance commonalities and differences. There are more ways in clarifying distinctions between two performances then through the use of rather abstract adjectives, which are not based in any kind of objective ground. As Repp mentions, it is all too easy to keep blaming the concept of ‘artistic personality’ in order to give answers to the diversities among performances. There is work to be done in another direction, namely the one, which may lead us to be able to find more objective grounds on which to ground our perceptions and relate our musical preferences to.

31 Chalmers (1977: 32)
32 Ibid.
3.0 A brief history of empirical performance analysis

Even though performance has always been a vital part of the musical experience, systematic studies of performance can be only traced back to the turn of the twentieth century. Only once methods had been developed to record either the sounds of performance, or the actions of instruments, was any kind of detailed study possible – and so the piano roll, record, magnetic tape, and computer have all played their part at different stages in the short history of empirical studies of performance.33

We can trace the beginning of the rise in empirical performance analysis to Seashore and his collaborators in the early 1930’s. At the University of Iowa, they developed an extensive research program in music performance and pioneered the use of acoustic analysis techniques to derive “performance scores” that show the exact variations of pitch, timing, and intensity produced by an artist on some instrument. Their work represents the first landmark in the field of empirical research on performance. In his book “Psychology of Music” [1967 (1938)], Seashore reports much of what took place during that time. A considerable amount of data was collected in Seashore’s laboratory, but their analyses remained rudimentary and focused primarily on technical aspects such as pitch accuracy and vibrato in singing and string playing, and chord synchrony on the piano.34 There was no focus in the capturing of commonalities and differences between different performances. One of the reasons for this was the fact that American psychology at that time pointed towards more of a behavioristic approach. Additionally, the lack of advanced statistical methods and psychological performance models at that time help clarify the path Seashore chose to take.

Still, Seashore did make some conclusions, which already then, evoked the thoughts Repp would later present in his study:

“...there is a common stock of principles which competent artists tend to observe;...We should not, of course, assume that there is only one way of phrasing a given selection, but, even with such freedom, two artists will reveal many common principles of artistic deviation. Furthermore, insofar as there are consistent differences in their phrasing, these differences may reveal elements of musical individuality”.35

In Germany, during the same time, Hartmann (1932) carried on similar research. He compared timing patterns of two famous pianists playing a movement of a Beethoven sonata. With the information gathered, he later described the difference between the two performers through detailed numerical representations, but the sheer small scale of this research, limits the generality of Hartmann’s outcomes.

Between 1940 and 1970, roughly, not much took place in the area of performance research. The so-called ‘modern’ period of performance research was born with Povel (1977), and Bengston and Gabrielsson (1977) and has continuously grown since. Most of the

33 Clarke & Cook (2004: 5; 77)
34 Repp (1992: 2547)
35 Seashore (1947: 77)
researches taking place at the moment, have taken a case study approach. Not many choose to focus on individual differences (diversities), but instead, on the occurrences of certain principles (commonalities). Povel’s research looked at expressive timing in several Bach harpsichord performances, whilst Bengtsson and Gabrielsson analyzed Swedish folk tunes. In the 1970s, Shaffer designed a piano/computer interface to analyze timing, coordination, expression, and the cognitive representation of complex movements. Shaffer (1981) produced the first substantial paper to report results obtained from direct computer monitoring of the piano.

The arrival of new technologies, most notably that of MIDI, and the personal computer, were instrumental to the considerable increase in the number of empirically oriented investigations into music. A former student of Shaffer, Eric Clarke, went on to make significant contributions to the field. During the 1980s, Clarke conducted case studies of piano performance of Satie’s music and wrote at length on how musical structures are expressed in timing variations.

Sundberg, Fryden, and Askenfelt (1983) is the first published attempt to produce an artificial model of performance expression using a collection of separate rules that relate to local features of the music. Neil Todd, another former student of Shaffer, published in 1985 a subsequent attempt to achieve the same goal, only this time, by using a single rule, which was applied recursively. Todd’s computational model of timing at the phrase level has been revised and extended to dynamics in his most recent work (Todd, 1992a, b).


The analysis and study of different performances of the same work, the same method applied in Repp (1992), was already applied in research by Gabrielsson (1987) and Palmer (1989), but in smaller scale. Gabrielsson (1987) analyzed five pianists’ performances of the first eight measures of Mozart’s Piano Sonata in A major, K. 331. Amongst others, Gabrielsson measured the different timing, intensities, and articulation between the five performances.

Palmer (1989) is worth describing in more detail. Her goal was to gain better understanding of the relationship between a performer’s musical intention and his musical realization. She designed two scientific experiments. Her first experiment, like Gabrielsson’s (1987), also involved the performance of the first eight measures of Mozart’s Piano Sonata in A major, K. 331. Palmer’s, unlike Gabrielsson’s research, involved the same excerpt being played in a “musical” fashion and an “unmusical” (mechanical) fashion by six pianists. The data were recorded on an IBM personal computer attached to a weighted 88-key MIDI keyboard. The pianists were instructed to play in a musical fashion and then in an unmusical fashion. Palmer examined the performances in terms of three rules: chord asynchrony, rubato, and note overlap. Palmer’s work showed first that unmusical

---

36 Musical Instrument Digital Interface; commercial standard for the exchange of information between electronic instruments and computers (See glossary).
37 See Clarke (1999); Gabrielsson (1999).
38 Palmer defines chord asynchrony as “the difference between note onsets that are notated in the musical score as synchronous.” (Palmer, 1989) Rubato is usually perceived as a tempo change
performances contained less of these three elements than musical performances. Furthermore, she found that the musical performances of experts featured more of these rules than the musical performances of students.

In the second experiment, eight performers played the first sixteen measures of a Brahms Intermezzo and then notated their interpretation on an unedited musical score of the work. In her study, Palmer illustrates how one of the performers marked phrase boundaries and circled the melodic line that was emphasized. It is this specificity on the performer's part that then permits the experimenter to examine a relationship between intention (the marks on the score) and performance (the MIDI data in the computer). From this study Palmer discovered that the notated melody is played earlier than the other elements appearing simultaneously. She also found that the line notated by the performer features the greatest amount of note overlap and that the notated phrase boundaries feature the largest amount of rubato. One of her conclusions after conducting both studies was that ‘pianists share a common set of expressive timing methods for translating musical intentions into sounded performance’, an echo, as it were, of Seashore’s (1938) earlier presented statement and now substantiated by systematically collected empirical evidence!

Gabrielsson’s and Palmer’s studies, even though very significant, still used relatively small samples. An “X” number of performances by the same pianist are needed (repeated performances), instead of only one, in order to be able to separate chance, error or unintentional patterns from those that are ‘under control’. Additionally, the number of performers involved in their experiments was also small. Repp (1990a) is the first paper to look at a larger body of performance data, using commercial recordings and extracting performance data from recorded sound. The study focused on nineteen complete performances by distinguished pianists of a Beethoven sonata movement. The analysis was limited to the timing patterns at the level of quarter-note beats and also had as goal a search for Clynes’ (1983) elusive “Beethoven pulse” within the timing patterns. As in Bengstsson and Gabrielsson (1980), Repp applied principal components factor analysis to these timing data, so that he could determine how many independent timing patterns were instantiated. Two factors emerged, the first representing primarily phrase-final lengthening, while the second factor captured other types of expressive variation. Repp also obtained evaluations by listeners of the different performances, and was able to find some relationships between the measured timing patterns and the listener’s judgments. Since this first paper, Repp has gone on to investigate larger collections of performances, in some cases analyzing over 100 recorded performances of the same work.

In the empirical study that will be analyzed in this essay, Repp chose to apply the same general approach as in Repp (1990). This time however, without the aim of testing Clynes’ theory of composer’s pulse any longer. Its main purpose was to assemble a large sample of performances of a particular composition by outstanding artists, and to analyze

---

39 Palmer (1989: 345)
40 See page 10.
41 Clarke & Cook (2004: 5; 79)
42 See glossary.
43 Repp (1992: 2548)
the timing patterns in detail, using various statistical methods. Through the use of these methods, Repp hoped to be able to distinguish commonalities from differences between the various performances. The common patterns would reveal how most pianists transmit musical structure and expression through timing variations, and it was of interest to determine whether there would be a single common factor or several. A few of the performances observed were from legendary pianists whose individuality is the trademark of their mastery. Therefore, the delineation of individual differences was also of interest to the research.

For this study, Repp chose Robert Schumann’s “Träumerei”, which is one of the most well known pieces from the romantic piano repertoire. According to Repp, “Träumerei” was specifically selected due to its highly expressive character, which permits much freedom in performance parameters, and hence much room for individual difference in interpretation. The abundance of available recordings of this work positively contributed to Repp’s research.

44 Ibid.
4.0 “Träumerei”

To be able to fully understand Repp’s results, a thorough analysis of “Träumerei” and its score have to be digested by the reader. This section will devote itself to dissecting the work, from its historical background to a full-fledged structural analysis.

4.1 Historical background and structural analysis

“Träumerei” (“Dreaming”), is the seventh of thirteen short pieces, which make up Robert Schumann’s (1810-1856) “Kinderszenen” (“Scenes from Childhood”), op. 15. A masterpiece in its genre, this suite was written by Schumann in 1838, when he was secretly engaged to Clara Wieck. Schumann had written around 30 pieces for Clara at that time and selected thirteen of those to form this suite. Even though the title refers to childhood, these pieces are far from intended for children. “Kinderszenen”, Schumann’s aphorisms of innocence, take a very mature and ‘lived’ performer to properly bring them to life. Alfred Brendel, one of the most distinguished pianists of our time, wrote an essay on this work whose title evokes the same feelings mentioned earlier: “Testing the Grown-Up Player: Schumann’s Kinderszenen”. “Kinderszenen” is as it were, an adult’s miniature collection of his childhood’s memories. Each piece in the cycle has a title.

“Träumerei” occupies a central position in the “Kinderszenen” suite, not only by its location but also by its duration and structural function. It serves as a resting and turning point in the cycle, which shows so many intricate thematic connections that it may be considered a set of free variations. Its key signature, meter and motivic content also single it out as the hub of the suite. Still, it is also often played by itself. Horowitz often performed it as an encore and several popular versions and arrangements of “Träumerei” have been made. Repp, in his paper, even goes so far as to suggest that “Träumerei” might be considered the most popular Romantic piano work. Its score is shown in Figure 2.

---

45 Repp (1992: 2548)
46 See Reti (1951), Traub (1981)
FIG. 2. The piano score of “Träumerei”, created with MusicProse software following the Clara Schumann (Breitkopf and Härtel) edition (with some deviations in minor details due to software limitations). The layout of the score on the page is intended to highlight the structure of the music.\textsuperscript{48}

\textsuperscript{48} Repp (1992: 2549)
In order to gain full understanding of the terminology Repp uses in his study, a melodic/rhythmic structure analysis of “Träumerei” is vital. In Figure 3, the melodic/rhythmic structure (or grouping structure) is systematically described. In order to understand Figure 3 properly, Repp’s detailed description follows.

“The piece is composed of three 8-bar periods (A, B, A’), the first of which is obligatorily repeated. Each period is subdivided into two 4-bar phrases, which are represented by staff systems in Fig. 2 and by large rectangular boxes in Fig. 3. (Actually the beginning and end of each phrase extend slightly beyond the four bars, overlapping with the preceding and following phrases, respectively). There are two phrase types, a and b, each of which recurs three times with slight variations (indicated by subscripts in Fig 3). Phrase a1 (period A) is repeated literally in period A’. Phrases b2 and b3, which constitute period B, are structurally identical but differ in key, harmony, and some other details.

FIG. 3. Schematic representation of the melodic/rhythmic structure of “Träumerei”.

“The melodic events are divided horizontally (roughly, along the dimension of relative

---

49 See glossary; Lerdahl & Jackendoff (1983)
50 Repp (1992: 2549)
pitch) among four registers or voices (S=soprano, A=alto, T=tenor, B=bass). Vertically (along the dimension of metrical distance), the events within phrases are grouped into melodic gestures, which are represented by filled boxes in F. 3. Characteristically, they extend across bar lines (vertical solid and dashed lines in Fig.3). A melodic gesture (MG) is a unit composed of at least two and rarely more than seven successive tones. It is defined here to begin with the onset of its first tone and to end with the onset of its last one. (In Fig. 3, each MG box extends one eighth-note space beyond the metrical onset of the last tone). Blank spaces represent time spans devoid of MGs; they may contain single tones, sustained tones, or rests. Multivoices chords having some gestural quality are represented by vertical ellipses in Fig.3. Arrows indicate continuity of a MG across a line break or with a subsequent melodic event. Thus MG1 and MG6 continue from the end of one line to the beginning of the next, and MG3b and MG5b in the soprano voice are closely linked. When an arrow points into a blank space, it points to the onset of a single-tone event that coheres with the MG.

As you can see below, the layout of Figure 3 corresponds to that of Figure 2.

“MGs are divided into primary (p) and secondary (s) ones. The latter are usually
shorter and accompany primary MGs; they are represented by boxes filled in lighter shades. An exception to this classification is MG2i, a delayed imitation of MG2 divided between the tenor and alto voices (bars 10 and 14). In the following timing analyses, we will essentially be concerned only with the primary MGs, which represent the leading voice(s) in the polyphonic quartet. As can be seen in Fig. 3, in phrases of type a the primary MGs (including the final MG6f in bar 24) are all in the soprano, except for MG6a, which is in the bass and overlaps both MG5a and MG1. In type b phrases, during their second half, the primary MGs cascade down through the four voices, overlapping each other.  

Having a thorough analytical background of the melodic and rhythmic structure of “Träumerei” is significant to Repp’s study since from these data, amongst others, the possible relationship can be analyzed between the timing structure of the performers and the musical structure of the work.

4.2 The Performances

Repp selected 28 performances by 24 distinguished pianists to be analyzed during his study. Table I shows the list of performances and recording references. Cortot and Horowitz are the only two artists who are represented by three different recordings each. From the 28 recordings, 17 are on long-playing records, 3 on cassettes and 8 on CDs (including a transfer of a 1929 recording by Fanny Davies from a very scratchy original; Fanny Davies was a one-time student of Clara Schumann and for self-evident reasons, of great interest to Repp’s research). 23 performances come from complete performances of “Kinderszenen”, while five are of “Träumerei” only. Two of those five performances are from live concerts (encores), while the others are studio recordings. In Table I, actual or estimated recording dates can be found.

Table I. Artists and their recordings used by Repp. Abbreviations: CD = compact disc; C = cassette; <

---

31 Repp (1992: 2548-2549)
The 24 artists can be grouped according to gender (6 female, 18 male), country of origin (5 from Russia; 4 each from Austria and France), 3 from England; 2 each from Germany and Brazil; one each from Czechoslovakia, Argentina, and Chile; one unknown), and approximate age at the time of recording (about equal numbers of young, middle-aged, and old). Besides some of the most renowned pianists of this century, the 24 artists include also some less well known. At least twelve of these pianists are no longer alive.

---

52 Repp (1992: 2550)
5.0 Analysis Methods

In order to obtain performance data from existing sounds recordings to be able to delineate the timing profiles of the 28 performances listed in Table I, Repp measured the intertone onset intervals (IOIs) of each performance manually, using a waveform editor (SoundEdit 16) to display the waveform at a suitable level of temporal resolution. The digitized waveform was displayed on the screen of a computer terminal. A cursor was placed at clearly recognizable note onsets, and used to label those points in the waveform file. A permanent “label” was attached to that point in the waveform file. The differences between the time points of successive labels yielded the IOIs, which were noted down to the nearest milliseconds. After some practice, it took about 2 hours for Repp measure one complete performance.54

Figure 4 illustrates how the data acquired from the IOI measurements becomes vivid in graphical form. Raw numerical data from each performance is transformed into graphic profiles. Through this adaptation of the data, it becomes much easier to compare the different performances to one another.

![Figure 4](image)

**FIG.4.** The measured tempo flow (tempo rubato) in a number of performances of “Träumerei”. The graph shows the different possible ways of slowing down at the end of the composition; a straight horizontal line would illustrate a metronomic performance.55

6.0 Significant results

---

53 The interonset interval of a note, usually taken to be a note’s primary rhythmic property, is defined as the time from the start of any note to the start of the next note in the same part. From these data one could probably not derive the precise rhythmic property of a note, but instead, what exactly is represented is the performer’s interpretation of a given rhythmic pattern, which we all know, usually varies from a metronomic rendition. (Clarke & Cook, 2004: 5; 80)

54 Repp (1992: 2550)

55 Honing (2004b: 9)
From the data Repp acquired from the IOI measurements of all the 28 performances by distinguished pianists of Schumann’s “Träumerei”, he was able to compare each performance in great detail with regard to their timing structure. Not only did he objectively look at the more global properties such as overall tempo, he also subjected the data to a variety of statistical analyses on large and also more detailed sections of the work. Every minutia in the timing pattern was addressed during the analytical process.

Repp was able to distinguish commonalities from differences between the timing structures of the various performances. From the abundance of data that derived from Repp’s study, the differences among the performances are at first glance more striking than the commonalities. The affluence in the differences accounted provides ample material to support the view that every artist’s performance is, in some sense, unique and unlike any other artist’s, even if just one physical dimension (timing) is considered.56 Referring to the analyzed recordings by Horowitz and Cortot (three performances each, see table 1), Repp reports that ‘even the same artist’s performances on different occasions, while demonstrably similar, are sufficiently different to be considered distinct and individual events’.57 With such information, grasping a common ground between these 28 very unique performances seems impossible. Still, commonalities did come to the front, which reflected apparent constraints on performance linked to musical convention (unwritten rules of tradition associated to the standard repertoire). What exist are generally accepted norms of musical performance (i.e. constraints on performance), which the performer is expected to pay attention to. According to Repp, the variations among performances seem to take place within these constraints. If all 28 performances analyzed, products of established and successful pianists, seem to respect a certain norm, the question arises: How should the boundaries musicians are confronted with be characterized? Are they purely conventional, (i.e. arbitrary) or do they represent more general laws of motor behavior and perception that music performance must conform to in order to sound “natural”?58

This section will highlight the one commonality Repp substantiated among the performances. Additionally, the explicit variations in the ‘Overall Tempo’ in the 28 performances will serve as an example of one of the diversities Repp was able to objectively illustrate in his research.

56 Repp (1992: 2564)
57 Ibid.
58 Ibid.
6.1 **An example of Commonality:**

**Timing Patterns and the Hierarchical Grouping Structure**

One of Repp’s most significant results was to reveal a number of underlying regularities in all these performances, so-called norms that performers seem to observe. Even though each performance was unique, all 28 performances reflected the musical structure in their timing. Besides dynamics and articulation, pianists use timing to communicate their interpretation of the musical structure of a piece. Through his study, Repp was able to substantiate that structural hierarchy seems to influence the timing structure, especially with regard to final phrase lengthenings (i.e. ritardandi). Just as in speech, performers slow down at the end of a sentence or phrase in order to announce its end.\(^{59}\)

### 6.1.1 Todd (1985)

“Boundaries in the hierarchical grouping structure are generally marked by ritardandi whose extent is roughly proportional to the ‘depth’ of the boundary.” \(^{60}\)

This principle, which is the base of the timing constraint Repp refers to in his study, was already theorized and implemented as a computer algorithm by Todd’s (1985) model of expressive timing at the phrase level. It represents the general principle that there is a slowing of the tempo at major structural boundaries, in proportion to the importance of the boundaries [cq phrases; Palmer (1989)] Performers seem to ritard the most at the end of a piece; other substantial ritardandi take place at end of major sections (in “Träumerei”, the 4-bar periods); and smaller ritardandi occur at the ends of individual phrases and gestures.

Furthermore, Repp noticed that variability within the timing structures of the various performances increased at lower levels of the structural hierarchy. Every performer seemed to observe the major ritardandi at the end of the major sections, even though in different degrees. On the other hand, in lower levels of the structural hierarchy, variability increased among the timing structure, which presumably reflected the relaxation of the performance constraint imposed by the musical structure at lower levels of the hierarchy.

In order to understand the relationship of the phrases and depth of the boundaries in the hierarchical grouping structure of “Träumerei”, a phrase structure tree is a helpful visualization (see Figure 5).

---

\(^{59}\) Honing (2004c)

\(^{60}\) Todd (1985)
There are three eight-bars periods (A, B, A') in “Träumerei”. Each period can be divided into two 4-bar phrases. The ‘depth’ of the boundaries relate to the location of a certain phrase or period in the structure. The second phrase of each period has a heavier hierarchical weight then the first phrase. The hierarchical weight of each period is also higher the later the period is located. Period A' is therefore hierarchically heavier in weight then period B, and so on. On a scale of 1-6 (1 = heaviest hierarchical weight, 6= lightest hierarchical weight), the number in the right side of the phrase-structure tree shows the level of depth of each phrase in the grouping structure.\(^{61}\)

**Grand average timing pattern**

In order to analyze the relation between the timing structure and the musical structure, Repp had to first obtain a *grand average timing pattern*, by computing the geometric mean of corresponding IOIs across all 28 performances. Even though the *grand average timing profile* is not necessarily an optimal performance timing pattern, according to Repp, it succeeds in capturing features that many individual performances have in common, despite masking random and idiosyncratic timing deviations that differ from performance to performance.

In Figure 6, the grand average timing profile of all six phrases of “Träumerei” is illustrated. According to their relation and similarity, on the left-hand panel, the timing patterns for the three type-a phrases are shown, and on the right-hand panel, the timing profiles of the three type-b phrases. The abscissa represents the “score time” in eight-note steps. The bar numbers on the abscissa though, refer to bars 0-8 only; for the other (later) phrases an appropriate constant must be added.\(^{62}\) The higher the score of the interonset interval in milliseconds, the slower the tempo.

---

\(^{61}\) Results from analysis.
\(^{62}\) Repp (1992:2553)
FIG. 6. Grand average IOIs (geometric means across all 28 performances). The two panels show the timing profiles for phrases of type a and b, respectively. Primary melodic gestures (see Fig. 2) are indicated by brackets.\(^{63}\)

\(^{63}\) Ibid.
When focusing in the last bars of each phrase (MG5a, MG6f, MG6a, MG5b, MG6b), the shape of the timing profile clearly represent the principle that slowing of the tempo occurs at major structural boundaries, in proportion to the importance of the boundaries. In the left-hand panel of Fig.6, we can see that the most extreme ritardando (prescribed in the score) occurs at the end of the piece (bars 21-24/see arrow). The second most pronounced slowing down (also prescribed) occurs at the end of the second period (bar 16-left hand panel/arrow). Next come the end of the first period (bar 8-left panel) and the end of phrase b1 (bar 12-left panel), which show about the same amount of ritardando. The end of the phrase a1 (bars 4 and 20-right panel) shows the smallest, but still quite noticeable slowing down.
Another method applied by Repp to research commonalities among the different performances was the statistical technique of principal components analysis (PCA). This method decomposes the data matrix (N performances by M IOIs) into a number of independent components or factors, each of which resembles a timing profile (“factor scores”). The original data can be approximated by a weighted sum of these factors: the weights, which differ for each performance, are called “factor loadings” and represent the correlations between the performance timing profiles and the factor score profiles. The higher the similarity between the factor score profiles and the performance timing profiles, the higher the factor loadings. VAF (the percentage of variance accounted for) expresses the degree of approximation between the factor score profiles and the performance timing profiles. The degree of approximation is dependent on the number of factors that are considered significant. The first factor always stands for the largest amount of variance accounted for (highest degree of approximation) and represents a kind of central tendency or most common pattern between the performances. Additional factor loadings account for increasingly less variance (lower degree of approximation), thus representing patterns shared by fewer performances.

PCA thus, can reveal whether there is more than one shared timing pattern represented in the sample of 28 performances. If there is essentially only one way of executing a piece, a single factor should therefore be able to explain most of the variance in the data. In the case of several dramatically different timing patterns, various factors should emerge.

A PCA conducted by Repp on the data of the 28 performances in their entirety, yielded a single factor, which indicated that all pianists observed the major ritardandi. This result confirmed once again the presence of a structurally-related timing constraint under which performers seem to work.

---

64 See glossary. (PCA)
65 Repp (1992: 2554)
66 Ibid.
6.2 **An example of Diversity:**

**Overall tempo**

“What is a metronome for?” exclaimed Mendelssohn. “It is a most useless instrument. A musician who cannot guess the time of a piece at first sight is a blockhead.”

I could easily have answered that there were a good many blockheads, but I held my peace.

[…] One day he asked me to show him the score of the *King Lear* overture, which I had just finished writing at Nice. First he read it over slowly and attentively, and then, just as he was about to play it on the piano (which he did with matchless ability), “Give me the exact time,”, said he.

“Why should I? Did you not say yesterday that any musician who could not guess the time of a piece at first sight was a blockhead?”

I refer to this rather amusing dialogue from Hector Berlioz’ *Memoirs* between Berlioz and Mendelssohn, on the matter of tempo, in order to substantiate my choice to highlight the aspect of ‘overall tempo’ as an example of diversity between the performances; considering the large spectrum of choices. Even though Mendelssohn’s remark on metronome and tempo, is quickly discredited by the speaker himself during this dialogue. The hypothesis that a work possibly dictates its tempo is not entirely alien to the musicological discourse (Think of Clynes’ provocative theory of a “composer’s pulse” in performance timing).

The overall tempo each artists chooses to use is considered to be one of the relatively free parameters in one’s interpretations of a work. It accounts for a significant factor in which performances vary and greatly serves as an example of one of the explicit differences illustrated in Repp’s study. According to Repp, even though there may be an “ideal tempo”(e.g. tempo markings prescribed by composer) for a piece of music, this remains an abstraction. The belief that there is in fact a range of acceptable tempi, and individuals may differ considerably in what they consider the ideal tempo is illustrated by the performance sample examined by Repp. A wide range of tempi is represented, that according to Repp, nearly all seem acceptable to a musical listener. Even though Repp relies on subjective judgment in this aspect, the fact that these performances derive from established artists in the musical world, contributes to the justification of such assumption.

---

67 Berlioz [1932(1870): 278]
68 Repp (1992: 2565)
Repp achieved tentative estimates of the global tempi in the different performances by computing the first quartiles (the 25% point) of the individual eighth-note IOI distributions, multiplying these millisecond values by 2, and dividing them into 60 000. Repp motivates his choice for the first quartiles due to the fact that that expressive lengthening of IOIs is both more frequent and more pronounced than shortening.

Another reason was that by considering the first quartiles, the tempi for two pianists, namely BRE and DAV, came closely to, respectively, Brendel’s (1981) statement of his preferred tempo and Clara Schumann’s (DAV’s teachers) recommended tempo. The tempo range extended from 48 to 79 quarter-notes per minute, as shown in Figure 5.

Repp concluded that apart from the fact that the three fastest performances are all old recordings, there does not seem to be any systematic relationship between tempo and the time at which the recording was made, nor with pianists’ gender, age at the time of recording, or country of origin.

---

Repp (1992: 2551)
7.0 Discussion, interlude and conclusion

The main purpose behind Repp (1992) was to assemble a large sample of performances of a particular composition by outstanding artists, and to analyze the timing patterns in details, using various statistical methods. Through the use of these methods, Repp hoped to be able to distinguish commonalities from differences between the various performances. The common patterns would reveal how most pianists transmit musical structure and expression through timing variations, and it was of interest to determine whether there would be a single or several common factors between the performances. As illustrated in this essay, Repp succeeded in achieving his goal.

Through empirical analysis, the timing profiles of each performance left its abstract existence in sound and became concrete through graphics. These graphics allowed a detailed scrutiny of all the data comprised in each timing profile, which permitted the differences and commonalities found, to be objectively illustrated. Prior to Repp’s research, the commonality-diversity distinction may already have seemed an obvious enough concept. Namely, the existence of two basic aspects of music performance: a normative aspect (i.e. commonality) that represents what is expected from a competent performer and is largely shared by different artists, and an individual aspect (i.e. diversity) that differentiates performers. However, there was little tangible evidence to substantiate it. Only through empirical analysis, could Repp come to reveal, amongst others, a number of underlying regularities in all the 28 analyzed performances and go on to conclude that timing patterns are strongly related to the hierarchical grouping structure of a composition. In this way, one of the so-called generally accepted norms of musical performance, which performers are expected to pay attention to, could be determined and illustrated.

Repp’s research provides enough evidence capable to refute the belief that performances differences cannot be characterized objectively. There are many physical properties of musical performance that, without any question, can be measured objectively. Empirical studies such as Repp (1992) and Cook (1995) permit a kind of discussion to take place, that would be hard or impossible to sustain without the data that empirical methods of research provide.

As a musician myself, being able to study and observe the striking individualities of pianists such as Horowitz and Rubinstein fascinated me. Through the empirical data assembled by Repp, their ‘artistry’ became more tangible. I was not unfamiliar to the concept of final-phrase lengthening. Slowing down at the end of a phrase to announce its ending was one of the many ‘unwritten rules’ that I was passed on by my many teachers. When getting acquainted with Repp’s results, at first, it seemed nothing more then obvious that such ‘rule’ would be highlighted. However, the fact that the 28 performances showed that in the performance process, the extent of the ritardandi were proportional to the boundaries in the hierarchical grouping structure they marked, was something that I simply as musician, could not objectify.
Will these results drastically influence how musicians work? No. It is my belief that tradition, without any doubt, will continue to be the main source of guidance; the unwritten rules of tradition will never stop carrying its role. Empirical musicology will not assume the position of teacher or reference point for performers. Instead, it will serve as reference.

Can we now, through these results, answer what makes the performance of a pianist like Horowitz so different and individual? Or why we prefer one performer over another?

My answer to both is: ‘not entirely’. There are many parameters that contribute to the subjective impression of a performance. Besides timing, other factors such as timbre, articulation, pedaling, etc. influence the listener’s impression. Repp succeeded in analyzing one of these factors. We can now only better understand what is so different and individual in Horowitz’ timing. The others factors that influence the listener’s impression have partially been or have yet to be tackled in future performance research. An example of such an aspect is timbre. It is no mystery for scientists, and for musicians, that when playing the piano, the moment you hit the key, the velocity in which the key is struck defines the sound quality. The faster the velocity in which a pianist hits a key, the louder it sounds. The piano is a percussive instrument by definition. Still, one of the most important features, which plays a significant role in the listener’s perception of a performance, is the so-called “touché” of an artist. Even though according to the theory and definition of the instrument this aspect should easily be able to be analyzed, the variations of fast and soft keystrokes, that supposedly lead to the so far mysterious “touché” equation, has yet to be decoded.

In regard to finding more objective grounds on which to relate our musical preferences to, not only empirical research on music is necessary. Besides the differences in performances and performers, each listener has a different background which directly influences his or hers preferences. Psychological and sociological empirical investigation is also needed in order to be able to study musical preferences. Repp (1992) did not assemble this type of data. However, in a study conducted earlier, Repp (1990a) obtained evaluations by listeners of different performances, and was able to find some relationships between the measured timing patterns and the listener’s judgments.

Therefore, confirming the success of this combined method of research.

Some of the existent criticism towards empirical performance analysis is rooted in the belief that it is unrealistic to think that empirical research will ever be able to capture all the innumerable factors involved in performance practice. In a discussion with Jacques Boogaart, assistant professor at the University of Amsterdam, he expressed the opinion that ‘even if in theory one should be able to analyze all the aspects involved in performance practice, to work out this theory into practice is something different’. Boogaart highlighted his opinion by giving the following example:

“Choose one piece of music. No sounding music yet, only the score. Then invite ten performers to each play the piece. The results of these performances amount to thousands of differences among these performances, all of which may play a role in the listener’s response. If you then let 100 persons listen to it, the numbers of impressions you will receive are innumerable. The background of each listener

---

Repp (1990a: 622)
plays a vital role during his listening experience, and each listener has a different background.” He goes on to say, that “I am sure that if one wanted, in theory one should be able to count how many drops of water the ocean contains, but the practice is another story.”

Whether empirical analysis will ever be able to capture a performance exhaustively and unravel all of its mysteries may indeed remain in doubt; however, this should not be seen as something negative. There are many physical properties of musical performance that, without any question, can be measured objectively, which when analyzed, lead to genuine insights about music and musicality. Repp (1992) and Cook (1995) are examples of some of the research that already succeeded in convincingly substantiating some of them.

Another source of criticism originates from those who see empiricism in music as a ‘dirty word’. When introducing the topic of empirical research in music to friends and colleagues layman to the subject, several at first utter their frustration by asking why we should want to formalize music? By looking at such graphics as the ones presented earlier in this essay, some shake their heads in disapproval. Some musicians among them, in particular, see it as the dismantling and dissecting of their beloved art.

In a recent e-mail I received from Repp, he refers to a young pianist (a graduate student at the Yale School of Music) who he had once invited to participate in one of his studies. Repp writes that he received a message from this student, in which he explained he would rather not participate in the study. “The message basically said that he thought performance was something mysterious and unexplainable, and he wanted it to stay that way.” Chalmers (1977) writes that ‘for many, science and art are antithetical. To apply one in the study of the other seems absurd, if not blasphemous’. Wing (1940) reports that his earlier work was strongly criticized by practising musicians on the grounds that music was said to be too spiritual a thing to be measured statistically. As the pianist who did not want to collaborate in Repp’s experiment, many fear if music’s “secret’ is unraveled, its magic will be automatically lost. And why would we want to lose what is music’s most precious gift?

It is my belief that knowing more does not mean hearing less. Music’s magic cannot be lost, especially not by means of enriching our knowledge of its wonders. Those who believe objective performance analysis will collaborate with the ‘killing’ of music through its formalization, miss in my opinion, the beauty of empirical musicology. Since the emergence of empirical musicological research, particularly since the late 1970s, a more systematic description of a whole variety of phenomena that was previously either unknown, or only known in broad outline, has become available. The rise in performance studies as a research area has, in addition, brought a focus on different performance traditions, the nature of performance interpretations and its relationship to analysis, the legacy of historical recordings (now dating back 100 years) and what it can tell us about changes in performance styles. We

---

71 This discussion took place in Amsterdam in August of 2005.
72 During the process of writing this essay, besides meeting Repp in person during a conference in Belgium, he greatly contributed insight on various topics through e-mail correspondence. This particular e-mail I refer to was received on August 19, 2005.
73 Chalmers (1977: 32)
74 Wing (1940: 343)
75 Clarke & Cook (2004: 5; 77)
are now more able to objectively study how historical events may have influenced the changes in performance style. In April of 2004, the AHRC Research Centre for the History and Analysis of Recorded Music (CHARM) was established. A partnership of Royal Holloway (University of London) with King’s College (London) and the University of Sheffield, CHARM’s aim is to promote the study of music as performance through a specific focus on recordings. One of CHARM’s projects at the moment examines the historical circumstances that have shaped sound recordings as commercial and technological artifacts, in order to provide a contextual dimension to the study of performance style.\footnote{The AHRC Research Centre for the History and Analysis of Recorded Music, http://www.charm.rhul.ac.uk}

In July of this year, the tenth edition of the Rhythm Perception and Production workshop (RPPW) took place in Bilzen, Belgium. This workshop is an opportunity for experts in Rhythm Perception and Production to discuss their ongoing experimental and theoretical work with expert colleagues. The workshop has an interdisciplinary character, with a focus on cognitive and experimental psychology and music research, but also welcoming contributions from other domains (e.g. linguistics, computer science, neurosciences,...).\footnote{RPPW site; http://www.rppw.org/rppw2005.html} Repp would be present during RPPW 2005 and at the beginning of July, I found myself going down to Belgium to meet the \textit{man} behind the \textit{study}.

After attending a very interesting session of presentations in the morning, I met Repp for lunch at the busy RPPW cafeteria, where the circa 70 participants and colleagues present, carried on their animated discussions and conversations over a glass of wine. Seating at the furthest end of one of the long tables, I eagerly awaited Repp’s arrival. Wearing a t-shirt that referred to a similar conference he attended years earlier in Japan and glasses from which the degree could be easily guessed, Repp set across from me at the table and kindly welcomed me with a smile. After briefly introducing myself, giving a short description of the topic of this essay and what had driven me to focus on this subject, I started my quest in having my many questions answered.

What has he done since 1992? What has he considered the most important results brought forth from empirical research after 1992? What are for him the most significant researches taking place in the field of empirical musicology at the moment? How does he see the future of empirical musicology?

As many of the answers that would come to surprise me, his first one at first shattered the expectations that had subconsciously evolved in my mind in anticipation to this meeting. Unconsciously, I had no doubt that when asking him what he had been doing since 1992, he would overwhelm me by naming different studies he would have conducted exponent from the similar quest of his 1992 research; namely the better understanding of musical performance choices and interpretations through objective analysis. His answer could not be more different. Since \textit{Diversity and Commonality in Music Performance: An Analysis of Timing Microstructure in Schumann’s “Träumerei”}, he had barely conducted research related to the topic. Instead, he

\textit{Interlude}
had directed his focus to other areas, such as the research topic he presented at RPPW 2005; Does an Auditory Distractor Sequence Affect Self-paced Tapping?

One of the few studies that did somewhat come close to the topic of performance, but not with the same purpose as his 1992 study, was an research article Repp co-wrote with Günther Knoblich entitled Perceiving Action Identity: How Pianists Recognize Their Own Performances. The article was published in 2004 in the journal “Psychology Science”. Here follows a short description by Repp and Knoblich of their research topic and the results they obtained during their study.

“Abstract – Can skilled performers, such as artists or athletes, recognize the products of their own actions? We recorded 12 pianists playing 12 mostly unfamiliar musical excerpts, half of them on a silent keyboard. Several months later, we played these performances back and asked the pianists to use a 5-point scale to rate whether they thought they were the person playing each excerpt (1=no, 5=yes). They gave their own performances significantly higher ratings than any other pianist’s performances. In two later follow-up tests, we presented edited performances from which differences in tempo, overall dynamic (i.e., intensity) level, and dynamic nuances had been removed. The pianists’ ratings did not change significantly, which suggests that the remaining information (expressive timing and articulation) was sufficient for self-recognition. Absence of sound during recording had no significant effect. These results are best explained by the hypothesis that an observer’s action system is most strongly activated during perception of self-produced actions”.

Following his description of this study, the time seemed to have come to ask him what he thought empirical musicology could contribute to practicing musicians. Repp agreed with my belief that tradition, without any doubt, would continue to be the main source of guidance for performers. Empirical musicology will only serve as reference. For the listeners, Repp expressed that getting acquainted with such results, as the ones obtained in his study, do not deduct, by the least amount, any of the ‘magical’ effect a particular performance of “Träumerei” might have had prior to such analysis. Knowing more does not mean hearing less.

In regard to the future of empirical musicology and the search for the unraveling of music performance’s mystery, he numbered several factors of performance, which empirical musicology has yet to find a way of capturing and analyzing, including the aspect of sound quality (timbre) that I mentioned earlier. When asked how he thinks one may succeed in analyzing this aspect, Repp simply answered: “I don’t know”.

In the train on my way back to Amsterdam, I let all Repp had said slowly sink into my mind. The youth of the field of empirical musicology once again became clear to me.

**

78 Knoblich & Repp (2004; 604-609)

79 Knoblich & Repp (2004; 604)
Among all prospects for the future of empirical research in performance, and its evident potential in leading to more genuine insights on music and musicality, there is one field of research within empirical musicology, which I am skeptical about; namely research on expressive performance synthesis. Despite decades of effort and some moderate progress, especially by the Stockholm group (Sundberg, Friberg, and colleagues) nobody can yet make a computer play a piece at the same artistic level as a fine artist. The saddest example is the work of Mazzola\(^{80}\), whose mathematically complex and sophisticated performance system generates nothing worth listening to—at least, as far as I have heard it.

It is my belief that the wealth of empirical musicology does not lie in the attempt to create a formula that describes the essence of a successful performance or an equation that will be able to reduce music’s magic to numbers and graphics in order to illustrate its essence. Such is, as Mazzola’s results suggest, an unrealistic goal and in my opinion, even undesired.

One of the significant requirements for success in future empirical research lies in the questions that will be asked; the quality of these questions. Research begins when we ask questions about the world. In the case of music, it is clear that there is a multitude of worthwhile questions that can be posed. The power of good and thought-provoking questions is that they rightly challenge scholars to do their best to assemble evidence that might help produce informed (albeit limited and provisional) answers. What is vital, is that the aim behind the posing of each question, is clear and focused in every single one of them. It is important to avoid ‘unfocused’ research, so that valuable research time and money is not wasted.

With regard to empirical investigation on the subject of musical preferences, in the development of empirical methods for future investigation researchers should always keep the following in mind; namely the relationship between the data that empirical performance research can capture from a performance and what the listener can capture in his or her own. In addition, empirical musicology alone won’t be able to answer all of the questions. Empirical evaluations of the listeners’ psychological and sociological backgrounds have to be incorporated in the research process, in order to come to conclusive results and answers. I advocate for an empirical field of musical research, that is, as Henkjan Honing (Honing 2004) describes ‘based on empirical observation and rigorous method, but at the same time is also aware of, and accounts for, the social and cultural context in which music functions’.\(^{81}\)

---


\(^{81}\) Honing (2004a: 242)
Afterword

As a musician and musicologist myself, I was ‘brought up’ in the traditional musical educational system, where the analysis of the score and its historical context predominate. It was during a ‘music analysis’ class in the fall of 2004 that I for the first time learned about empirical musicology: this new area of musicological research. The prospect of a field, which took upon itself unraveling the wonders of musical performance, marveled me. Going through the existing literature on the subject, I kept being surprised by the achievements this young field had already obtained. The more I read, the more my thirst increased in learning all that presently existed on the subject. Alongside my enthusiasm, nonetheless, skeptic feelings also found their way. Thoughts, such as music being simply too subjective to be fully analyzed in an objective manner, kept coming to my mind. The results and achievements I came along during my readings were indeed significant, but at the same time, still small in number and scale to irrefutably succeed in discrediting my doubts. On the other hand, I had only but begun to emerge myself on the topic, and knew I had not yet enough knowledge to base any conclusion on. It became clear to me that there was no better subject to embrace myself with during my MA essay project, then my quest in properly being able to value and assess empirical musicology, and its worth and contribution to the musicological discourse.

In June of 2005, my intense research began on the topic, resulting in this essay. It was a very exciting and challenging journey, that enriched my knowledge both as a musician and musicologist. My gratitude goes to Henkjan Honing, for his mentorship during three very exciting months; Rokus de Groot, for many inspiring and stimulating discussions and comments, and for being so generous with his time; Jacques Boogaart for his contribution; and Bruno Repp, who generously contributed to this essay by making some of his valuable time available to meet me and respond to my many e-mails. In addition, I would like to thank my family and all of my friends, who supported me through this endeavor. In special, my gratitude goes to Alexa Gevers and Anna Borleffs; my very dear friends.

Amsterdam, August 31, 2005.
Glossary

G.1 MIDI

MIDI, is an industry-standard protocol that defines each note precisely and concisely, allowing electronic musical instruments and computers to talk to each other. The MIDI standard was first proposed by Dave Smith in 1981 in a paper to the Audio Engineering Society and the MIDI Specification 1.0 was published in August 1983. With MIDI, researchers were now able to record and store on computer all keyboard events that happened during a performance. A dilemma in the beginning years of MIDI was the fact that the keyboards in which it was applied, poorly resembled a real instrument. Touch and sound qualities were so far away from the real instrument, that serious research was not yet able to take place. However, developments soon occurred. In the mid 1980s Yamaha produced first a synthesizer keyboard with more realistically weighted keys and then a MIDI grand piano – a standard acoustic grand piano fitted with a photoelectric cell system, which picked up the movements of the piano’s keys, hammers and pedals, and translated them into MIDI signals.

MIDI technique has been applied to several other instruments. There are MIDI drum pads, wind controllers, guitar and other string controllers, as well as pitch-to-MIDI conversion systems designed to take input from a microphone and to convert the signal into MIDI information. In Penfold (1990) and Manning (1993) one can find general introductions to MIDI. For now, I will limit myself to describing the features in MIDI, which apply to keyboard performance.

Eric Clarke gives a detailed description of MIDI’s function in capturing performance data in his article “Empirical Methods in the Study of Performance” (2004). Clarke names six features of keyboard performance captured by MIDI, namely:

1) The identity of any key that is struck
2) The time at which a note starts
3) The time at which a note stops
4) The velocity of the key press or of the piano hammer as it hits the string (which is directly correlated with the loudness of the sound produced)
5) The time at which either the sostenuto or soft pedal is depressed
6) The time at which either the sostenuto or soft pedal is released

The data that is gathered from the six items listed above is easily able to be processed and stored in any computer which runs a sequencer of some kind, such as Vision, Cubase, and Logic. Another option is also a more sophisticated graphical programming environment called Max. Most of the data described above is available directly after the capturing of a performance. Besides these, two other pieces of data are also available, but less directly. These two are the interonset interval (IOI) and the articulation.
The interonset interval of a note, usually taken to be a note’s primary rhythmic property, is defined as the time from the start of any note to the start of the next note in the same part. From these data one could probably not derive the precise rhythmic property of a note, but instead, what exactly is represented is the performer’s interpretation of a given rhythmic pattern, which we all know, usually varies from a metronomic rendition. In multi-layered musical textures, identifying the part to which the IOI belongs to can be very tricky and may rely in pragmatic decision, rather than objective choices. In the case of a note-against-note texture (such as a two-part invention), the definition of each part is clear enough.

Articulation, the other piece of ‘hidden’ data, is determined by the relationship between the sounding and silent parts of a note’s time span. For example, a staccato note is one in which the sounding portion of a note’s IOI is shorter than the total IOI, and a legato note is one in which the sounding portions is either the same, or greater than the total IOI - in this last case overlapping with the following note in the same part. (Clarke & Cook, 2004: 5; 79-84)

G.2 Principal Components Analysis (PCA) / Factor Analysis

Statistical technique used to trace down the contributive value of certain components. The purpose of factor analysis is to discover simple patterns in the pattern of relationships among variables. In particular, it seeks to discover if the observed variables can be explained largely or entirely in terms of a much smaller number of variables called factors.

The main applications of factor analytic techniques are: (1) to reduce the number of variables and (2) to detect structure in the relationships between variables, that is, to classify variables. Therefore, factor analysis is applied as a data reduction or structure detection method (the term factor analysis was first introduced by Thurstone, 1931).

Many statistical methods are used to study the relation between independent and dependent variables. Factor analysis is different; it is used to study the patterns of relationship among many dependent variables, with the goal of discovering something about the nature of the independent variables that affect them, even though those independent variables were not measured directly. Thus answers obtained by factor analysis are necessarily more hypothetical and tentative than is true when independent variables are observed directly. The inferred independent variables are the so-called factors. (Rummel, 1970).

G.3 Grouping Structure (Lerdahl and Jackendoff - 1983)

Grouping structure expresses a hierarchical segmentation of the piece into units such as motives, phrases, sections, etc (p. 8). In A Generative theory of Tonal Music (1983), Lerdahl and Jackendoff define it as the most basic component of musical understanding. It is presented graphically as several levels of arcs below the musical staff. Due to the demand for strict hierarchy, in A Generative theory of Tonal Music, Lerdahl and Jackendoff treated all music as
essentially homophonic; that is, they assumed that a single grouping analysis suffices for all voices of a piece. (Lerdahl and Jackendoff, 1983)

G.4 Todd (1985)

In musical performance, as in speech, slight decelerations and pauses frequently occur at the ends of phrases. This phenomenon, called phrase-final lengthening, forms the point of departure for Todd’s study entitled “A Model of Expressive Timing in Tonal Music”. Todd incorporates Lerdahl and Jackendoff’s (1983) technique of time-span reduction (TSR) into his own mathematical model for generating a preferred duration structure for performances of a tonal composition. In Todd’s algorithm, the degree of phrase-final lengthening increases at higher levels of the TSR. For example, a greater deceleration should occur at the end of an eight-bar parallel period than at the half cadence of the period’s antecedent phrase. (Todd, 1985)
References:


Instrumental Music (London: Novello).