Testing Zipf's Law:
The Mathematics and Aesthetics of Performance

by Jordan Bennett Louis Smith

submitted in partial fulfillment of the requirements
for an AB degree with honors in Music & Physics

Department of Music
Harvard College

March 20, 2006

Table of Contents:

ACKNOWLEDGEMENTS

ABSTRACT

It has been observed by Manaris et al. (2005) that in a particular performance of a particular Bach cantata, while the statistical distribution of note lengths as defined by the score corresponds to a highly repetitive distribution called 'black noise,' the subtle changes in timing in the performance of the cantata led to a more even distribution of note lengths indicative of 'pink noise.' In order to explore how generalizable this effect was, a more substantial data set was collected, including five solo piano pieces of varied genres played four times each by six skilled performers. The note lengths were analyzed in addition to other metrics. Two main observations were made: first, for any particular piece, the distributions for a given metric remained roughly the same across all performances by all performers; and second, that the distributions of some metrics were substantially changed – often in strikingly similar ways – when the performed versions were compared to the scored versions. Contrary to the expectations set up by Manaris et al. that the distributions would seek out more regular, 'Zipfian' properties in performance, it appears that the changes mostly led only to an increase in statistical randomness.

OUTLINE

Section 1 is devoted to a detailed exposition of Zipf's Law, beginning with its conception as a description of word frequency, moving on to describe the significance of the exponent that defines it and the kinds of noise different exponents are related to, and finally relating it all to music. This is followed by a detailed description of Manaris' findings, which forms the motivation for this thesis, and a discussion of the aesthetic and philosophical background relevant to Manaris' discovery. Following this is a review of the research relating Zipf's Law to music.

The first part of section 2 will formally state the hypotheses of the experiment. Section 2.2 will describe the design of the experiment, including the performers and the experimental apparatus, the rationale for the selection of pieces, and finally the basic procedures used in analysis. A descriptive account of the results will follow in section 2.3, including all the results of the statistical tests applied.

Section 3 will then discuss the implications of the results with respect to the hypotheses and philosophical groundings of the experiment. The limitations of the experiment will be acknowledged and new directions for research recommended.

Appended information includes the scores of the pieces used (Appendix A), the MATLAB scripts used to analyze the scores and the many performances (Appendix B), the tables (Appendix C) and finally the figures (Appendix D).

§ 1   AN INTRODUCTION TO ZIPF'S LAW

§ 1.1   ZIPF'S LAW IN WORDS

Zipf's Law was originally formulated by George Kingley Zipf, a professor of linguistics at Harvard at the time, to describe an interesting property of word frequency in language. It is most easily explained using an example. Begin by taking a text of at least a few thousand words, such as William Shakespeare's *Hamlet*. Then create a kind of 'concordance' by counting the number of times each word occurs and ordering them from most frequent to least frequent. For instance, in *Hamlet*, the most frequent word is 'the' (appearing 1142 times), followed by 'and' (964 appearances), and so on.[1] Once the list is complete, assign each word a non-
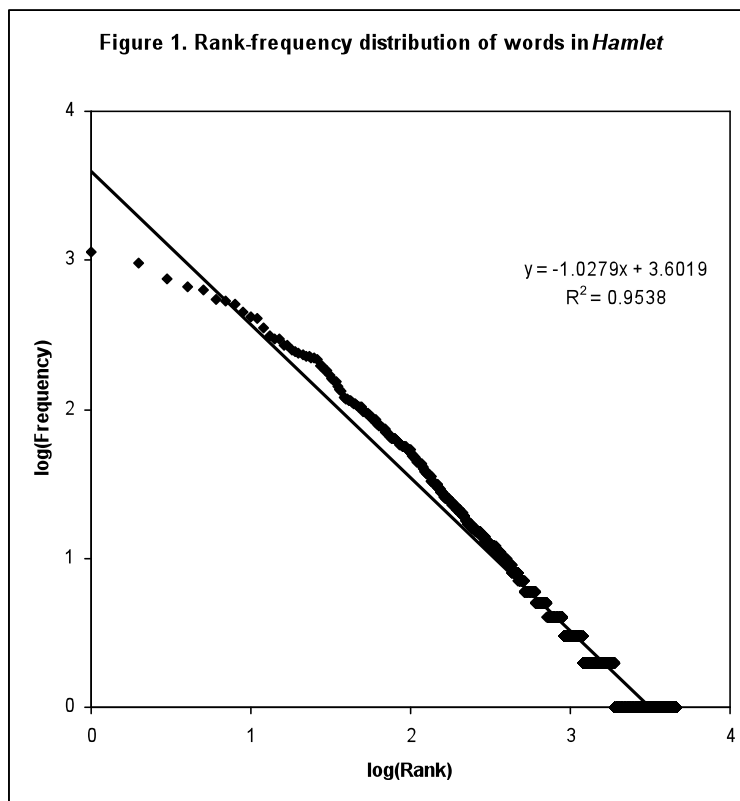
Figure 1. Rank-frequency distribution of words in *Hamlet*

$$y = -1.0279x + 3.6019$$
$$R^2 = 0.9538$$

---

1   With modern computing power and large texts freely available on the internet, investigations of this sort are phenomenally simple. The data on *Hamlet* were derived in seconds using the free software program "Concordance 3.2," developed by R. J. C. Watt, and using a copy of Hamlet from "The Web's first edition of the Complete Works of Shakespeare," hosted by "The Tech," an MIT newspaper. These resources are freely available at <http://www.bluesofts.com/download/121/11353/Concordance.html> and <http://www-tech.mit.edu/Shakespeare/hamlet/index.html>, respectively.

repeating rank: in our case, 'the' receives rank 1, 'and' ranks 2, and 'doth,' 'matter,' and 'Rosencrantz,' which are all tied for 160th place, are assigned ranks 160, 161, and 162 in any order.

Proceeding down the list, a surprising trend emerges where the frequency of each word is inversely proportional to its rank. That is, the nth most common word appears roughly 1/n times as often as the most common word. Clearly, this law is approximate: 'and' did not occur half as often as 'the,' and words only appear with integer frequencies, the 4657th and last word on the list, 'zone,' naturally did not occur 1142/4657 times – it occurred once. Yet, as Figure 1 demonstrates, the 1/n relationship is still quite strong. This relationship, that the frequency of a word varies inversely to its rank, is Zipf's Law.

How exactly does Figure 1 demonstrates this? First, translate Zipf's Law into mathematical form. Zipf's Law states that frequency is inversely proportional to rank. Letting $f(x)$ be the frequency with which the word x occurs, and r_x the rank of word x, we have

$$(1.1) \qquad f(x) = A / r_x{}^a$$

where A is a constant related to the frequency of the most common word, and where a is the number 1. (We will shortly see the advantage of treating a as a variable and manipulating the most general form of the equation.) Taking the natural logarithm of both sides, one obtains the relation

$$(1.2) \qquad \ln f(x) = \ln (A) - a * \ln r_x .$$

Comparing it to the general equation for a straight line,

(1.3) $$y = b + mx \,,$$

we can see that we have obtained a straight line of slope $m = -a = -1$. This is

precisely the relationship found in Figure 1, a logarithm plot of rank vs. frequency

of all 4659 different words found in *Hamlet*. A graph of this type is called a *rank-

frequency distribution* (RFD). A line of best fit estimates the slope of the

distribution as -1.0279, extremely close to the prediction of -1 and indicating that

the set of words was highly Zipfian. One should note that the poorer linear fit at

either end of the distribution is normal, and the tails of the data are often excluded

when the slope is estimated. We can also recognize that the constant A as being

inconsequential: it depends only on the sample size (here, the number of words in

Hamlet) and does not affect the shape of the distribution.

This relationship between the rank and frequency with which words occur

might seem merely serendipitous on its own, but the same power law has surfaced

in statistical studies in many other domains. Among other phenomena, Zipf's Law

has been found to describe the sizes of cities, the magnitude of earthquakes, the

number of species per genus, the frequency of web traffic, and still more.[2]

However, Zipf's Law remains experimental: although it is frequently

discovered to describe various data sets, explanations are more controversial. For

instance, although the Zipfian nature of large bodies of text has been known for at

least 50 years, no widely satisfactory explanation has been proposed.[3] This paper

---

2 References are abundant but also outside the scope of this paper. Wentian Li maintains a
   comprehensive index of Zipf-related studies at <http://www.nslij-genetics.org/wli/zipf/>.
3 Zipf himself attempted to explain it in a theory of 'least effort,' in which humans have carefully
   balanced the redundancy and efficiency of language.

will thus delicately avoid the question of *why* Zipf's Law applies to music, instead focusing on the questions of whether and how it applies to music. On the other hand, the philosophical implications of the law's application to music remains an important consideration for this study.

§ 1.2   THE ZIPF EXPONENT AND NOISE

So far we have kept *a* close to 1, and seen that this corresponds to a rank-frequency distribution of slope -1, but we would like to better understand what property is being described by the parameter *a*. To this end we will divide both sides of equation (1.1) by the number of words in the entire text, producing the new equation

$$(1.4) \qquad\qquad p(x) = B / r_x{}^a .$$

Here r_x and *a* mean the same as before and *B* is another constant equal to *A* divided by the total number of words counted. Likewise, instead of considering the frequency of the word *x,* we consider the probability *p(x)* with which the word *x* occurs, its frequency relative to the size of the text.

What kinds of values can *a* have? Well if we bring *a* all the way up to 0, we will have increased the slope of the distribution to 0, creating a horizontal line. This indicates that every event is equally frequent: in other words, the occurrence of events is on average random – white noise. In language, this distribution would be hard to accomplish on a large scale – imagine trying to write a coherent 300-word essay using 300 unique words – but in music, examples abound: a strict

twelve-tone piece could have every note as frequent as any other.

As before, $a = 1$ corresponds to a slope of -1 in the rank-frequency plot, and is defined as the 'ideal' Zipfian distribution found in language. This kind of probability distribution is sometimes nicknamed pink noise in analogy to white noise. If you produce an audio signal with equal power over all frequencies (white noise), the result is very harsh on human ears.[4] If you dampen the higher frequencies according to Zipf's Law, so that the power of each frequency f falls as $1/f$, the pink noise that results is almost soothing in comparison.

As a increases and the slope of the rank-frequency plot steepens, we eventually produce 'brown' noise (when $a = 2$) or 'black' noise (for $a > 2$). This indicates an increase in the repetitiveness of the events: the most common event (of rank 1) begins to dominate the distribution, since now the 2nd most common event occurs only one fourth (instead of one half) as often as the most common event, and so on. However, the extreme of repetitiveness only occurs when $a = -\infty$, in which case only one event ever occurs. A text possessing such a slope would simply consist of one word repeated ad nauseum, like "Also also also also also...."

We have seen what the literal audio analogy of Zipf's Law entails in terms of noise 'colours,' but how might actual pieces of music reflect these distributions? Suppose we were to calculate the rank-frequency distribution of

---

4   In the following discussion, it is difficult to unambiguously describe what the different colours of noise sound like. If one is not already familiar with the aural qualities of noise described, audio samples of each kind of noise can be found in the Wikipedia article "Colours of noise." <http://en.wikipedia.org/wiki/Colours_of_noise>.

notes, or pitches, in a piece of music: instead of making a concordance of all the words used, we simply count up the number of times each note occurs. If a distribution turned out to have a slope of 0, it would indicate that every note occurred equally often, perhaps the property of a strictly twelve-tone piece. The unlikely slope of minus infinity has perhaps only been achieved in one serious work: John Cage's *4'33"*, for solo piano, in which the performer plays no notes at all (Manaris 2005, 61). Clearly, more conventional music must fall between these two extremes; the next section will explain where.

§ 1.3 MUSICAL ZIPF

In the history of music theory, the idea of music as language and vice versa is a common recurring theme: philosophers from Archytras and Aristoxenus, who "considered grammar to be included under music, and ... taught both" (Quintilian, quoted in Rousseau 324) to Jean-Jacques Rousseau who "thought that music came first and language was a subspecies" (Lidov, 1), philosophers and theorists have imagined metaphoric connections between language and music for thousands of years.

Perhaps influenced by this philosophical tradition, Zipf himself searched for a musical analogue to Zipf's Law. He focused his attention on two musical parameters: melodic intervals (the difference in semitones between successive notes in a melody) and the time distance between repetitions of the same note, which he called pitch distance (Zipf 336). The four pieces of music Zipf analyzed

included two works by Mozart and Chopin, and two jazzier numbers by Irving

Berlin and Jerome Kern. In each, Zipf reported finding 1/f distributions (Zipf

337).

His tests were conducted painstakingly by hand,[5] but since Zipf, the

advent of computers and of convenient MIDI[6] technology has allowed researchers

to analyze greater numbers of works faster and in greater detail. There exists now

a wealth of studies expanding on Zipf's results and supporting the notion that it is

a widespread fact. Whatever the metric – be it melodic intervals, pitch distance,

volume, or any other parameter – finding the slope of its rank-frequency

distribution amounts to the same four simple steps:

1. **Count up every instance of each type of the given parameter.** For
   instance, if the RFD of melodic intervals is desired, tally the number of
   single-semitone steps, the number of two-semitone steps, and so on;
2. **Sort these frequency tallies in decreasing order and assign each a
   strict non-repeating rank.** That is, if the 5-semitone intervals and 8-
   semitone intervals are tied for 4th most frequent, one is assigned rank 4
   and the other rank 5; it makes no difference which;
3. **Plot the logarithm of the frequency against the logarithm of the rank**;

---

5   ...by his three loyal students, of course: Messrs. C. H. Bridge, Jr., Daniel Scarlett, and W. S.
    Wheeling. Wheeling, incidentally, also composed the 1949-50 Hasty Pudding musical "Heart
    of Gold."
6   MIDI = Musical Instrument Digital Interface, is a widespread music encoding technology.
    Rather than encode the entire audio signal of a piece of music, it compactly encodes only the
    pitch, onset time, duration, and velocity (loudness) of each note.

4. **Find the line of best fit and determine its slope**.

The first to follow in Zipf's footsteps were Richard Voss and John Clarke, who in 1975 measured the pitch and loudness fluctuations of classical, jazz, and rock radio station signals over long periods (Manaris 2005, 57). They found a rough *1/f* relationship in the loudness fluctuations of each station as well as in the power spectrum of pitch fluctuations (Voss & Clarke 317-318). In 1990 and 1991, Hsü and Hsü published papers showing that the frequency with which melodic intervals occur in various Bach and Mozart compositions varied inversely to the size of the interval. This higher order relationship, Zipfian in form, suggests a fractal[7] geometry embedded in the music.

More recently, Manaris has led a number of studies exploring the relationship between Zipfian statistics and pleasantness. In his 2003 study, he analyzed a corpus of 200 pieces from many genres (from baroque to romantic, to jazz, to punk rock, to pink noise generators), using up to 20 metrics (mostly variations on pitch, pitch distance, duration, and melodic and harmonic intervals). He found that, over all genres, over all pieces, and over all metrics, the average slope was -1.2023, with a standard deviation of 0.2521 (Manaris 2005, 60). Considering the diversity of genres and metrics included, this is a remarkably restricted range.[8]

---

7   There exists a wealth of papers dealing with fractal distributions in addition to Zipfian ones, but a discussion of these, while related to the present discussion, is beyond the scope of this thesis.

8   Examining how the average slope varies by genre is fascinating: the shallowest slope, -0.8193,

In all of these studies, existing music was analyzed and elements of it found to follow Zipfian distributions. Could Zipf's Law be used to run the experiment the other way? That is, could one turn pink noise into pleasant music? Voss and Clarke attempted just that, developing a computer program that used two independent random number generators (which could produce white noise, pink noise, and brown noise) to control the duration and pitch of successive notes. "Remarkably," writes Manaris, "the music obtained through the pink-noise generators was much more pleasing to most listeners" (2005, 58). Furthermore, the subjects found that white-noise music tended to be "too random" and the brown-noise music "too correlated" – the pink-noise music "was judged by most listeners to be much more pleasing" than either of the others (Voss & Clarke, 318). Voss and Clarke noted that "the sophistication of this '1/f music' … extends far beyond what one might expect from such a simple algorithm, suggesting that a '1/f noise' (perhaps that in nerve membranes?) may have an essential role in the creative process" (V&C 318, quoted in Manaris, 58).

In his research, Manaris has pursued some of the consequences of this idea, proposing and working towards several possible applications of Zipf's Law. In 2003 he imagined an "Evolutionary Music Framework" which would 'evolve' computer-generated music based on its Zipfian fitness. In his 2005 paper, he discussed the value of using Zipfian metrics to automatically distinguish between genres, and tested a composer attribution program based on Zipfian analyses of

---

corresponds to highly random twelve-tone music, while the steepest slope, -1.5288, corresponds to punk rock. Most other genres hover in the -1.2 – -1.3 range, with jazz straying out to -1.0510.

musical scores. Finally, he compared Zipfian analyses to human subjects' ratings of the "pleasantness" of various pieces of music, and used the data to train a computer program to predict fairly accurately a human's real time judgement of the pleasantness of a piece.

To summarize briefly, although we earlier defined the relatively pleasant pink noise in opposition to white noise, these colours of noise have revealed themselves to carry potentially enormous aesthetic significance. Between Zipf, Manaris, and Voss and Clarke, there indeed seems to be some 'ideal' structural quality common to language and music. The intuition that reflects this structure was once expressed by Manfred Schroeder, a founding member of IRCAM:

> [F]or a work of art to be pleasing and interesting, it should neither be too regular and predictable nor pack too many surprises. Translated to mathematical functions, this might be interpreted as meaning that the power spectrum of the function should behave neither like a boring 'brown' noise, with a frequency dependence of $1/f^2$, nor like an unpredictable white noise, white a frequency distribution of $1/f^0$. (109, quoted in Manaris 2005, 63)

§ 1.4   PERFORMING ZIPF'S LAW

However, in his 2005 article Manaris makes note of a striking case of black noise in music. His finding is described briefly enough that it is worth quoting in its entirety:

Figure [2a] [not reproduced here] shows an example of black noise in music. It depicts the rank-frequency distribution of note durations from the MIDI-encoded score of Bach's *Two-Part Invention No. 13 in A minor.* This MIDI rendering has an unnatural, monotonous tempo. The Zipf-Mandelbrot slope of -3.9992 reflects this monotony. Figure [2b] depicts the rank-frequency distribution of note durations for the same piece, as interpreted by harpsichordist John Sankey. The Zipf-Mandelbrot slope of -1.4727 reflects the more 'natural' variability of note durations found in the human performance." (Manaris 2005, 61)

The finding is not described in any greater detail, nor is it mentioned again elsewhere in the article. Indeed, it seems that nowhere else in the literature on Zipf's Law are issues of performance even addressed.[9] Conversely, while articles that provide quantitative or statistical analyses of performance are plentiful (and will be described later), none of them invoke Zipf's Law. Manaris' reference to Zipf's Law's application to performance may constitute the first intersection of these two fields of research (analyses of scores using Zipf's Law, and other mathematical analyses of performances). Given its novelty, then, the bold wording of Manaris' finding may seem rather brazen. That the reduced slope reflects something 'natural' about performance may be a rather attractive

---

9   The reader may correctly point out that at least one study already mentioned, that of Voss & Clarke, involved the analysis of live music via radio signals. However, their study (which only looked at pitch and loudness fluctuations, and could not address timing) was carried out over a global scale that involved averaging over many songs. Furthermore, since specific songs were never identified, it was impossible for them to compare their distributions to those found in the relevant scores.

conclusion, it must be admitted that it has been leapt to on the basis of rather anecdotal evidence.

He never says so explicitly, but embedded in Manaris' account of Sankey's performance is the belief that something other than random chance changed the slope from -4 to -1.47. The notes have been variously lengthened and shortened necessarily, as a result of the 'natural' and unavoidable impreciseness in human performance compared to the score, or perhaps even deliberately, as a reflection of the performer's expressive interpretation of the score. Either way, the implication seems to be that there is something special about the number it was reduced to: the slope did not fall to -3, nor did it plunge as far as -0.2; rather, it found its way to almost within a standard deviation of the grand average of all metrics over 200 pieces (i.e. to within -1.2 ± .25). In other (highly suggestive) words, through a performer's act of interpretation, black noise has been "corrected" to more closely approximate the pink noise "ideal." The -1.47 suggests not just a greater degree of randomness in the note durations, but a greater degree of *musicality*. Sankey's performance somehow avoided introducing both too few deviations from the score (leading potentially to "brown noise") or too many deviations (leading to "white noise"), and instead deviated just enough to resemble Zipf's ideal.

To better understand the implications that Manaris might have meant the reader to draw from the Sankey example, it is illuminating to briefly consider the aesthetic viewpoint that he occupies with respect to music theory. Manaris

situates his work within a larger tradition – one might say the largest tradition – of music theory: that of connecting music to numbers. This tradition, Buzarovski notes, "owes its origins to Pythagoras' discovery of the ratios of the musical intervals" found to be the most harmonious (167): thus the octave can be represented as 2:1, the perfect fifth as 3:2, and the perfect fourth as 4:3. Manaris emphasizes this tradition rather explicitly: he begins his 2005 paper by relating the achievements of the Pythagoreans, who "quantified 'harmonious' musical intervals in terms of proportions," and immediately proceeds to quote a number of ancient philosophers to the end that music is mathematically beautiful: says Polyclitus, "Beauty does not consist in the elements, but in the harmonious proportion of the parts" (quoted in Manaris, 55).

The application of Zipf's Law to pieces of music is thus about much more than simply trying to fit frequency distributions to power laws; rather, the implicit project is to fit them to *simple* power laws, with ideally integer exponents. Hence Manaris found the Sankey example notable not because the slope of the duration metric was changed in performance, but because it came nearer to the theoretical ideal. The possible implications are compelling: if, for instance, a composition's note lengths were so randomly distributed as to have a metric slope of -0.6, would an interpretation of the piece steepen the slope to a value nearer to -1?

Returning to the example of Sankey's performance, and disregarding for the moment this Pythagorean aesthetic framework, one may point out that a change from a highly-ordered to a more random distribution is not all that

surprising. After all, it is part of the design of Western musical notation to use exact beats and note lengths, while variations in timing are prescribed by other instructions. Not only could no human performer, no matter how skilled or intent, hope to replicate the timing as precisely denoted in the score, but expressive instructions and conventions of performance indicate that he shouldn't; random timing fluctuations may be inevitable, but intentional time fluctuations are a part of the composed structure of the piece. Since an increase in randomness and variability indicates a decrease in the slope, the lower slope detected by Manaris is expected.

But despite the tentative confirmation of Manaris' expectation, one finds that his experiment ultimately raises more questions than it answers. If there is some principle at work causing the slopes to become less steep, how systematic is it, and what does it depend on? We have just supposed that both random and planned timing variations – due respectively to performers' imperfect timing and artistic sensibilities – contribute to this flattening slope. Although it seems likely that both contribute to this change, it is possible that one effect dominates over the other. For instance, if unintentional timing changes dominated over the player's intentional timing changes, then if a musician were to play a song multiple times one would expect the duration metric to change differently each time: perhaps if Sankey played the Bach invention again he would be as likely to get -0.7 as -2.1. There are a number of other factors that change could depend on. These include:

The score's duration metric value: Given two different Bach inventions

with duration metrics -4 and -3, would a performance by Sankey change both to -1.47, or would the changes depend on the original value? Or, perhaps there exists something particular in the distinct design of each piece not captured by the duration metric that governs how it changes in performance.

The piece's genre: Manaris also reported broad differences between genres in terms of each one's average slope over all metrics (2005). If, instead of another Bach piece, Sankey were to perform a Beethoven sonata whose duration metric was also around -4, would he change it exactly as he did the Bach piece (to roughly -1.47), or might it converge instead on another slope?

The performer's style: If we were to analyze a different musician's performance of the same Bach invention that Sankey played, would we find the duration metric changed in the same way? Or, perhaps there is something in whatever characterizes a performer's "style" of interpretation that partly determines how duration metrics are changed.

None of these four possible influences (random effects, duration metric, genre, and performer) are mutually exclusive: in fact, it seems likely that all of them contribute to the change in slope to some degree. One could imagine still more factors – the performer's skill level, the tempo or key signature of the piece, or countless others – but due to the limitations of this experiment, we will restrict our discussion and consideration to only these four possible factors.

There is one last detail to consider. While we have restricted our speculation so far to the note duration metric, a number of other metrics stand to

be affected by the changes incurred by performance. The pitch distance metric

(the time distance between successive instances of the same pitch), may seem

more rigidly prescribed by the score than the note duration metric, and thus the

changes in performance more slight, but it still seems apt to be influenced by the

factors mentioned above. And although Manaris' study did not include any

measures of dynamic variations in pieces, Voss & Clarke's 1975 study found that

loudness fluctuations followed Zipf's Law (see §1.3). Certainly the dynamics

prescribed in the score are as subject to change as the timing and are worth

investigating. This experiment will thus examine these metrics (pitch distance and

dynamics) in addition to the pitch duration metric.

Clearly, an extended study of Zipf's Law in performance has the potential

to offer a wealth of undiscovered knowledge. As a first foray into this domain,

this thesis will attempt to test in experiment the influence of each of the above

factors (duration metric, genre, and performer) on changes in note duration, pitch

distance, and dynamic incurred in performance. The following subsection will

develop hypotheses for this investigation, and the design of the experiment is

described in §2.

§ 1.5   QUANTITATIVE PERFORMANCES

It has already been stated that although nowhere else in the literature on

Zipf's Law is performance mentioned, there nevertheless exists a great number of

quantitative studies on performance timing. The results of these, while they could

not conclusively settle the questions raised in the previous subsection, are nevertheless crucial in forming expectations of how the metrics may be affected in performance. This section will discuss the relevant literature in order to formulate some hypotheses.

One should note that in most of the literature on quantitative analyses of musical performance, although many of the same kinds of techniques Manaris employed are used, the focus of the investigation is on local timing *deviations*, rather than the properties of global timing distributions. This is the case for each of the studies considered in this section, including Shaffer's 1984 study, which focuses on beat length and beat asynchrony as a function of score position, and Windsor and Clarke's comparison between the output of a performance algorithm and actual recorded performances.

An interest in local timing differences is reflected in a tendency to regard such analysis as either a quantitative substantiation or a mere assistant to theoretical analysis. For instance, in Cook's 1987 study comparing two performances of a Bach prelude by Helmut Walcha and Glenn Gould, his observation that "Walcha's and Gould's [timing] profiles for a given section are generally quite different" leads to a discussion of "how ... to interpret such lengthenings and shortenings of individual notes" (262). Windsor notes that a number of "quantitative analyses of the timing of piano performances clearly identify the ways in which the metrical structure of a piece is 'expressed' through systematic patterns of timing" (129). In this study, we will not be explicitly

concerned with the details of the structural information in the piece, but instead with a broad statistical view of the structure, as expressed in the rank-frequency distributions of the events. If, as both researchers and theorists seem to believe, "the source of such musical expression lies within the structure of the music that is played" (Windsor 129) then it seems likely that the metrical slopes found in performances will not vary arbitrarily. Rather, those metrics explicitly tied to the structure of the piece, such as pitch distance and harmonic interval, will perhaps depend in a predictable way on the original values in the score.

Guerino Mazzola is a researcher who has produced a number of articles using statistical approaches to analyzing musical structure and performance, including at least one that speaks to the timing differences incurred by performers when compared to the score. In his 1999 paper, for example, he used a computer model to extract different kinds of structural information from a score. This information was then used to create a number of 'weight functions' describing the theoretical importance of each note. He was then able to use these functions to evaluate and compare a number of professional performances. He sorted the 28 analyzed performances into 4 broadly different categories reflecting different kinds of timing decisions, categories defined by systematic timing deviations. For instance, where one category of performances tended to "emphasize the 4-measure periodicity of the harmonic structure," those in another "[had] a global curve with a peak around the 15th measure as a dominant feature" (72). If the rank-frequency distribution of note durations has any ability to reflect these kinds

of differences, we may expect significant differences between various performers. However, because Zipf's Law tends to average information over the entire piece, it could also be that much of this kind of detailed information will instead be averaged out.

Although there are very few studies that compare data from multiple performances, the consensus seems to be that timing differences between multiple performances by the same performer are very slight. In a 1985 paper by Shaffer et al., for instance, three performances of a Satie piano piece were recorded by a single pianist, and the subsequent analysis showed quite convincingly "that note timing was quite precise: very similar timing profiles were produced in the three performances and in the repeats of the theme within a performance" (72). A similar study by the same principal author even found those same similarities in two performances recorded a year apart (Shaffer 1984, 581). In our study, then, we will expect not to discover any significant differences between successive performances by the same person.

§ 2   METHODOLOGY AND RESULTS

§ 2.1   OVERVIEW

The experiment described in this paper endeavoured to explore the many possible dependencies that changes in duration metric caused by performance could exhibit. To test the possibility that this change in slope depends strongly on the initial conditions – that is, that the change will somehow be proportional or related to the original slope of the score – I have selected five pieces with different metrical slopes to investigate. To test the possibility that style or interpretation could dominate the change in slope, I have obtained six different skilled musicians to play these pieces. To test the possibility that random timing fluctuations are great enough to dominate all other factors, each performer was recorded playing the piece four separate times. Finally, although Manaris noted no observation of any changes in the other metrical slopes – for instance, pitch distance, harmonic interval, and velocity – this experiment presents a natural opportunity to investigate changes in each of these too, since all the data will have been collected anyway.

My hypotheses are:

1. based on Manaris and a Pythagorean philosophy, that for all metrics subject to change in performance (i.e. all metrics other than pitch and pitch-class), the slope found in performance will be nearer to -1 than the slope in the score, even when the score's metrics lie between 0 and -1;

2. based on the many studies led by Shaffer, that differences between successive performances by the same player will be extremely small;

3. based on Mazzola, that for all timing-based metrics, the slopes for different performers will tend to fall into one or more well-defined categories.

§ 2.2 EXPERIMENTAL DESIGN

PERFORMERS

The six performers who acted as participants were all highly-skilled pianists: with the exception of one, a tutor of music at Harvard with at least 45 years of experience, all were college students who had studied piano for at least 13 years. The participants were also all told well in advance of the experiment which pieces they would be required to play, giving them ample time to prepare or even mark up a personal copy of the scores, as they wished. Despite the generally high skill level and sufficient preparation time, the subjects nevertheless proved a diverse sample of aptitude and level of preparation: some subjects breezed through the pieces quickly and with dexterity, while others stumbled over various passages. Each subject had been informed of the topic of the experiment only in the broadest sense (for instance, that it was about 'measuring global mathematical or statistical properties' of musical interpretation), but were not availed of the hypothesis or information on Zipf's Law. Participants were

instructed to keep their interpretations of the score "consistent," but beyond this were given few guidelines: they were encouraged to play the pieces at whatever tempo and in whatever order they liked.

MUSICAL SELECTIONS

While it would have been desirable to have the subjects play many pieces many times each, in order to minimize the fatigue of the performers and to ensure that a maximum of subjects would be available for the study, only five solo piano pieces were selected. These were also edited down for length so that each piece would take roughly 90 seconds to play. The scores for the five pieces are reproduced (in edited form) in Appendix A. The pieces are: *Sonate (Pathétique)*, Op. 13 No. 8, Ludwig van Beethoven (mm. 1-10); *The Entertainer*, Scott Joplin (mm. 1-38, no repeats, second endings); *Sechs Kleine Klavierstücke*, Op. 19, No. 6, Arnold Schoenberg (mm. 1-9); *1ˢᵗ Gymnopédie*, Erik Satie (mm. 1-39); *Minuet in G, BWV Anh. 114 No. 4*, Johann Sebastian Bach[10] (mm. 1-32, no repeats).

These five selections were selected on the basis of both their great familiarity (which minimized the chance that subjects would not have known the pieces beforehand) and their rich diversity: among them, five different composers are represented as well as five different genres (respective to the list above: classical, jazz, atonal, late romantic, and baroque). The pieces also represent a variety of difficulties (compare the Bach and Beethoven pieces) and expressive

---

10 This piece, although it appeared in the 1725 Notebook for Magdalena Bach, has actually
    recently been attributed to Christian Petzold, although for simplicity I will continue to refer to
    the composer as Bach.

qualities (compare the pensive Satie with the lively Joplin). Furthermore, since the primary motivation of this experiment was to investigate changes in the note duration metric, the selections were chosen so that the note-duration RFDs of their edited-down versions represented a range of slopes (from roughly -0.71 in the Schoenberg piece to -2.89 in the Joplin).

APPARATUS

Participants played on a Yamaha digital piano situated on the top floor of the Loeb Music Library of Harvard. The output signal, containing all MIDI information, was passed through a Midisport 2x2 MIDI interface, which converted the signal to a USB cable connected to a laptop. The incoming signal was recorded there using the program Cakewalk Pro Audio 9.03, and the final timing resolution was 5 milliseconds. The separate performances were then individually saved as MIDI files, ready to be input into MATLAB.

To quickly address the timing resolution, it should be noted that 5-millisecond accuracy was more than sufficient for our needs. In fact, when analyzed, the time information was always rounded to the nearest 20 millisecond interval in light of E. F. Clarke's findings that listeners' sensitivity to timing differences is generally limited to differences on the order of "30 to 50 milliseconds" (2002, *Musical Performance: A Guide to Understanding,* 192). In fact, the results of his 1989 study showed that "listeners are able to perceive as little as 20 ms lengthening in the context of notes lasting between 100 and 400

ms" (*The perception of expressive timing in music*, 2). Making distinctions

between note lengths differing by less than 20 ms threatened to lead to a greater

perceived randomness in the RFDs, which would lead to an underestimate of the

slope's steepness, so the durations were rounded up to the threshold of audible

detectability.

ANALYTIC PROCEDURE

Each of the 120 MIDI files (6 performers, playing 5 pieces, 4 times each)

that comprised my collected data was individually converted into a "note matrix"

using the MidiToolbox, a free compilation of MATLAB commands developed by

two faculty members of the University of Jyväskylä in Finland (Eerola 2004). In a

note matrix, each row represents a separate note event, listing its onset time and

duration in seconds, its velocity[11] (in arbitrary units from 0 to 127, where 64

denotes 'average'), and its MIDI pitch (measured in semitones and ranging from 0

to 127, with middle C = 60). These pieces of information were sufficient to find

the rank frequency distributions of each metric (the metrics, as well as brief

descriptions of each, are summarized in Table 1, found in Appendix C). A

separate function was written for each metric (these are reproduced in Appendix

B), although each function operates in essentially the same way, following the

four-step procedure outlined in §1.3. Each function takes a note matrix as input

and gives three output values: the slope of the rank-frequency distribution, its

---

11 That is, the volume or loudness of the event. It's called velocity because it relates to how fast
    the piano key is struck.

corresponding R^2 value, and the standard estimate of error for the slope, which were found using the formula

(3.1)   standard error $= \sigma_y * (1 - r^2)^{1/2} * (N / N\text{-}2)^{1/2}$

where σ_y is the standard deviation of the set of y values, r is the correlation coefficient, and N is the size of the sample (the number of y values used).[12]

§ 2.3   RESULTS AND ANALYSIS

The raw data – that is, the rank-frequency distributions discovered for each performance, are too numerous (120 performances with 8 metrics makes 960 in all) to present entirely, but are all very similar in appearance. Two typical examples are shown in Figures 2a-b: the RFD of the note durations in one performance of Beethoven's *Pathétique*, and the RFD of note velocities in a performance of Joplin's *The Entertainer*. The information extracted from each RFD is presented with it: the equation for the line of best fit, the R^2 value, and the standard estimate of the slope error determined according to Equation 3.1. For all analyses made, the average R^2 value was 0.8012, indicating good fits overall, while the average estimate of error was 0.3154.

The first observation to make was whether, as expected, the slope of the metrics remained roughly constant across different performances by the same performer. Presented in Figure 3 are the slopes of the RFD for all 8 metrics in four performances of Bach's Minuet by a single performer. The graph clearly

---

12 How to find the standard deviation and correlation coefficient are topics covered in any good
     statistics textbook. The statistics resource used to advise this analysis was Moore and McCabe.

suggests that variation between performances is minimal, especially with respect to the error estimates already placed on them. The homogeneity of the results in Figure 3 is typical of all performers and all pieces. Thus, for the following calculations, the slopes were averaged over the four performances.

Next, we wished to determine how large variations were across different performers. Figures 4a-e plot each performer's average slope (for each metric) against each other, along with standard estimates of error calculated using the following equation, which finds the appropriate error ε_{av} for a set of n data points with errors ε_i:

$$(3.2) \qquad \varepsilon_{av} = sqrt[\ (\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2)\ /\ n]$$

Although for most pieces and most metrics the slopes once again remain quite constant between performances, there appeared to be significant variation in some cases. To help determine for which metrics and for which pieces there were significant differences among performers, a one-way analysis of variance (ANOVA) was performed on each ($F(5,18) = 6.2$, p=0.05). The output of these tests – an "F-statistic" that, if greater than a critical value of 6.2, indicated a significant difference – are presented in Table 2. Evidently significant differences were absent for most metrics and in most pieces, although there are notable exceptions, which will be discussed in the following section.

For the time being, one may judge the slopes to be similar enough overall between performers to justify averaging them together so they can be compared to the scores. In Figures 5a-h, the metric slopes have been further averaged across all

performers (once again recalculating errors according to Equation (3.2)), and are now compared with the metric slopes of the original scores.[13] By plotting all the information together for each metric, a number of suggestive trends emerge: the pitch and pitch-class metrics, information that should only change in a performance containing errors, indeed changed very little between score and performance; the pitch distance and pitch-class distance metric slopes for each piece were slightly flattened in performance by similar amounts each time; both kinds of harmonic interval metric slopes seemed to be flattened in performance, but less consistently than pitch distances; and finally, the duration and velocity metric slopes were both significantly flattened in performance. in sporadic ways that seemed not to depend on their original value. The average change in slopes between score and performance is listed for each metric in Table 3, while the average metric found in the performances for each metric is plotted, along with standard deviations, in Table 4.

Indeed, although the original values of the duration and velocity metrics vary between the five pieces, the changes between score and performance seem to be quite similar. To compare the changes in metric slope between the different songs, the data in Figures 4a-e have been replotted in Figures 6a-h, which graph

---

13 The reader may note that no error bars were included for the score slopes. This is because the estimate of the standard error of the slope (as determined by Equation (3.1)) includes an ($N / N-2$) term that rapidly increases as the number of data points is reduced. For instance, if there were only two different note lengths in an entire piece, the $N/N-2$ term would diverge to infinity. However, this is because linear regression is typically performed on a set of data that represents only a sample of a larger population – the assumption being that more points could be found if the study were expanded. In this case, no such other points exist – the entire population, in a sense, has been sampled. The standard estimates of error, whose meaning is not even clearly useful for the purposes of this analysis, were thus greatly overestimated, and hence are omitted.

individually for each metric the reduction in slope for all pieces. The trends suggested in these figures are enough to inspire searching for some rudimentary models to predict the change in metric slope. To conclude the results section I will propose a model for each metric and describe the calculations for them; discussion of the models will be reserved for the following section.

The performance slopes for pitch and pitch class distance, as well as stacked and rooted harmonic intervals, seem to have linear dependences on the score slopes, while the performance slope for duration and velocity seem to remain constant regardless of the score's slope. To test these observations, linear regressions were fitted on each type of metric (that is, both pitch distances were treated together, as well as both stacked and rooted intervals). The results are summarized in Table 5, where the coefficient of dependence and the intercept represent the values m and b respectively in the fitted linear equation $y = mx + b$.

The linear regression produced somewhat poor linear fits for pitch distance and harmonic interval. The relatively low values of R^2 indicate a poor correlation between metric slopes in performance and score, and the large estimates of error do not give much confidence in the estimate of the correlation coefficient.

On the other hand, our observations seem to be supported for the duration and velocity metrics. The even lower R^2 values indicate very little correlation between the performance and score slopes for both metrics; a slope of 0, indicating no dependence at all, does not lie outside the standard estimate of error

for either coefficient of dependence. It seems that in performance the velocity metric slopes tend to converge roughly on a single value, regardless of the original slopes found in the score. The same seems to nearly be true for the duration metric slopes, except that a very small linear dependence remains.

§ 3.4   SUMMARY

Across different performances given by the same performer, the slopes of the rank-frequency distribution seem to stay roughly constant for all metrics. Across different performers playing the same piece of music, the metric slopes seem to vary only a little, although significant differences are sometimes detected. Across different pieces of music, the way in which the metric slopes in performance differ from the slopes found in the score seem to be systematic according to metric. Based on these results, a number of models are proposed to relate the slope in performance s_p to the metric slope found in the score s_s:

(3.3)   Pitch distance / Pitch class distance: $s_p = 0.46s_s + 0.10$

(3.4)   Stacked interval / Rooted interval: $s_p = 0.55s_s + 0.26$

(3.5)   Duration: $s_p = 0.16s_s + 0.45$

(3.6)   Velocity: $s_p = 0.00s_s + 0.66$

§ 3 DISCUSSION

Due to the small sample sizes and simplicity of the statistical tests, nearly all of the conclusions drawn from the data have come with serious caveats. Nevertheless, the broadest expectations of the hypothesis were well supported by the data. The expectations, caveats, and potential explanations are all discussed in this section.

§ 3.1 PERFORMANCE AND PERFORMER DIFFERENCES

First, it was observed that the different performances given by any single performer tended in every case and with every metric to produce the same RFD slopes. This is neither a controversial nor an unexpected observation: besides being presaged in the studies conducted by Shaffer and described in section 3, it is supported by the consensus in the literature, which states that a performance constitutes an analysis and 'expression' of its underlying structure. It should be noted here that because of the instructions given to the subjects to keep their multiple performances of the same piece "consistent," these performances in fact represent four versions of the same interpretation. It seems likely that the small differences between these versions are attributable to either the random timing fluctuations posited before or the accidental and isolated errors made by the performers. Of course, there is also an outside chance that they are simply due to errors in the linear regression analysis, but in any case it was seen that differences between multiple performers were greater than differences between multiple

performances.

Second, it was found that overall, the slopes of the RFDs produced by different performers were very similar, although there were undoubtedly some significant differences detected. Attempts to explain this on the basis of outliers or sampling error are probably unfounded, since over 20% of the ANOVA tests found significant differences, including the duration metrics of 4 of the 5 pieces. However, within those metrics exhibiting significant differences, it was rarely the case that a large range of slopes were found: rather, it was usually that five of the performers had more similar values, leaving the sixth an odd one out. This situation is depicted in the pitch-class metric for the Beethoven piece (Figure 4a), in which all slopes were found to be roughly the same except for performer #2. The same seems to be the case for the the stacked intervals of the Joplin (Figure 4b), the pitch class distance and velocity metrics of the Satie (Figure 4d), and the duration metric of the Bach (Figure 4e). Furthermore, even where the ANOVA did not discover significant differences, configurations where the metric slopes clumped together into groups of four and two were readily detected, usually in the interval metrics: see, for instance, the stacked interval metric of the Satie.

Although there are too few performers to compare to be certain, these results do recall Mazzola's findings that the timing and loudness profiles of many performers could be sorted into distinct "clusters." From this perspective, the significantly differing metrics could be seen as reflecting distinctly different categories of interpretation. This could also suggest a convenient explanation for

the remarkable variety of slopes found in the duration and velocity metrics for the Schoenberg piece. Certainly the Schoenberg piece was the least well known (all but performer #5 were learning it for the first time), and was the only representative from the atonal genre, with which the subjects undoubtedly had the least experience; perhaps, lacking the same musical understanding that guided their performances of the other more familiar pieces, the subjects produced very dissimilar interpretations that were reflected in the more continuous range of metric slopes.

If indeed differences in interpretation are being reflected in slopes that stray from the mean, one might expect such differences to be observable in more than one isolated metric. Once again the data prove frustratingly inconclusive with regard to this possibility, although some notable anecdotal evidence can be found. For instance, performer #2 tended to produce steeper slopes than the rest when playing the Beethoven selection, and performer #5, the only one to have encountered the Schoenberg piece before, tended to produce slopes distinct from the others. If systematic differences in interpretation could be detected across multiple pieces, one might consider this an indication of style. This is perhaps reflected in the tendency for performer #5 to produce duration metric slopes shallower than the others.

To treat these conjectures on performer differences as conclusions would be unwise, given the limited and high-error data. Thus the source of metric differences between performers remains unaccounted for. Nonetheless, the results

suggest at least one interesting possibility that would be readily testable. To test the hypothesis that different metric slopes are the result of different interpretations, a study similar to this one, except where performers with identifiably different approaches to the same piece of music are solicited, could be carried out to try to establish a correspondence between interpretation differences and metric differences. Part of the difficulty in relating these results to previous ones is that very different properties have been measured: in the present study, global statistical differences in timing; in Shaffer's and Mazzola's studies, trends in local deviations in timing. If a study such as the one proposed were to collect both global and local timing information, a correspondence between the two could perhaps be established.

Finally, it remains a possibility that differences in metric slope reflect not stylistic differences but differences in skill. Although all the performers in this study were skilled and had been trained in classical piano for at least 13 years, they nevertheless constituted a range of skill: some play sporadically in chamber groups, others currently take regular lessons, and at least one is a professional concert pianist. These skill differences were not consistently reflected anywhere in the data, but perhaps if a more dramatic gradient of skill level were obtained – say, by recruiting one pianist just learning to play, another with 10 years of lessons, and a professional concert pianist – a more interesting investigation could be conducted into this possibility.

§ 4.2 ZIPF'S LAW TESTED

Regarding Manaris' discovery of the flattening of the duration metric slope when a particular piece was performed, the data overwhelmingly suggest that this trend is completely generalizable across most pieces, genres, composers, and performers. As reflected in Figure 6g, the average duration metric slope was flattened in performance for each piece, even if only slightly, corresponding to more random distributions of note lengths.

However, as stated in section 3.1, the hypothesis predicted not that slopes would all get flatter, but that slopes would all converge roughly towards -1. The data strongly contradict this hypothesis. Only one slope, the -0.71 obtained for the Schoenberg piece, was originally set above -1, and it was found to increase still more to -0.34. Moreover, rather than approach -1, the slopes for three of the other four pieces shifted even further: only the duration metric of the Satie piece remained steeper that -0.9.

These trends in the duration metric were mirrored in all of the metrics other than pitch and pitch-class: for instance, the pitch distance slopes, three of which were as steep or steeper than -1 in the score, were all flattened to slopes between 0 and -1. Contradicting the established results of Zipf and Manaris, the rank-frequency distribution slope of -1 does not seem to represent a common, much less ideal, distribution for performed music. To recall another of Manaris' findings, he found that when the slopes of a number of metrics were averaged over a number of scores of different genres, an overall slope of -1.20 was obtained

with standard deviation 0.25. In the present study, when the metric slopes from all performances by all performers of pieces of different genres were all averaged together, an overall slope of -0.76 was obtained with standard deviation 0.27. Thus with similar accuracy, two strikingly different results have been obtained.

This result is both exciting and confusing: while unearthing a new and unexpected difference between performance and score, the hypotheses that might have explained it have been largely, if inconclusively refuted. If the metric slopes of the score are flattened in performance, what governs this change? The similarity of the slopes obtained in multiple performances led earlier to the suggestion that random timing fluctuations did have any great on the slope change between performance and score. Next, although noting some significant differences in the slopes obtained by different performers, differences between performers overall were found to be small. The irrelevance of performer differences is suggested even more by the sharp differences between score and performance for the duration and velocity metric slopes. While the largest inter-performer differences in duration metric were on the order of ~0.2 (see Figures 4a-e), the average change between score and performer was 2.34 (see Table 3). This was similarly the case for the velocity metric. If not random fluctuations or differences in interpretation, what, then, might account for these systematic differences in slope?

As a modest first step towards answering this question, simple linear regression was employed to attempt to model the data. The resulting fits,

summarized in Table 5, were very poor, with modest R^2 values and large estimates error. The changes in the pitch and pitch-class distance metrics were pooled, given their similar means, and likewise the changes in the harmonic interval metrics. The lines found each had slopes of roughly -0.5, with estimate errors of roughly 0.2. Thus, although the exact nature of the dependency is far from certain, it seems probable that the slopes found in performance relate strongly to the slopes found in the score.

On the other hand, the models produced for duration and velocity seem to suggest that the performance slopes exhibit very little dependence on the score slopes. For velocity, the tiny coefficient of dependence suggests that this metric remains completely independent from the velocity metric of the score. This is a compelling result: it suggests that whatever the dynamic variations prescribed in the score, whether highly specific as in the Schoenberg or consisting only of two dynamic markings in the Satie, the dynamic variations expressed in performance will be statistically the same.

The same could be true of note duration, since a coefficient of dependence of 0 is well within the standard estimate of error of the calculated value, but the data seem to indicate a very slight dependence on the information contained in the score.

CONCLUSION

Contrary to the expectations set up by Manaris' 2005 study, it was found that the slopes of the rank-frequency distributions of all timing- and velocity-sensitive metrics did not converge towards the -1 ideal observed for the metric slopes of scores. On the contrary, slopes tended to converge on values between -0.5 and -0.9, which indicate more random rank-frequency distributions than expected. The possibility that these differences are due to random timing fluctuations incurred in human performance or by differences in interpretation between performers were not supported by the data; indeed, differences between multiple performances and performers turned out to be rather small compared to the great differences between performance and score slopes.

Although the implications of this violation in Zipf's Law in performances remains unexplained, a handful of rudimentary linear models were fitted to the current data with suggestive results. A larger scale which did not use up resources searching for inter-performer and inter-performance differences could easily determine more accurate and reliable models worthy of theoretical consideration.

BIBLIOGRAPHY

Works Cited

Bent, Ian D., and Anthony Pople. "Analysis," *Grove Music Online.* Ed. L. Macy
(Accessed 12 March 2006), <http://www.grovemusic.com>.

Buzarovski, Dimitrije. "**Generative Ideas in the Aesthetics of Music.**"
International Review of the Aesthetics and Sociology of Music. 17.2
(Dec. 1986): 163-184.

Clarke, Eric. F. "The perception of expressive timing in music." *Psychological
Research.* 51 (1989): 2-9.

Clarke, Eric. F. "Listening to Performance." *Musical Performance: A Guide to
Understanding.* Ed. J. Rink. Cambridge: Cambridge UP, 2002. 185-196.

Cook, Nicholas. "Structure and Performance Timing in Bach's C Major
Prelude (WTC I): An Empirical Study." *Music Analysis.* 6.3 (Oct. 1987):
257-272.

Eerola, T. & P. Toiviainen. (2004). *MIDI Toolbox: MATLAB Tools for Music
Research.* University of Jyväskylä: Kopijyvä, Jyväskylä, Finland.
Available at http://www.jyu.fi/musica/miditoolbox.

Hsü, Kenneth J., and Andreas J. Hsü. "Fractal geometry of music." *Proceedings
of the National Academy of Science of the USA.* 87.3 (Feb. 1990): 938-
941.

Hsü, Kenneth J., and Andrew Hsü. "Self-Similarity Of The '1/f Noise' Called
Music." *Proceedings of the National Academy of Science of the USA.* 88.8
(Apr. 1991): 3507-3509.

Lidov, David. *Is Language a Music? Writings on Musical Form and
Signification.* Bloomington: Indiana UP, 2005.

Manaris, Bill, et al. "Evolutionary Music and the Zipf-Mandelbrot Law:
Developing Fitness Functions for Pleasant Music." *Lecture Notes in
Computer Science.* Vol. 2611 (2003) 522-534.

Manaris, Bill, et al. "Zipf's Law, Music Classification, and Aesthetics."
Computer Music Journal. 29:1 (Spring 2005): 55-69.

Mazzola, Guerino, & Jan Beran. "Analyzing Musical Structure and Performance – A Statistical Approach." *Statistical Science* 14.1 (1999): 47-79.

Moore, David S., and George P. McCabe. *Introduction to the Practice of Statistics.* 5th ed. New York: W.H. Freeman and Company, 2006.

Repp, Bruno H. "The Detectability of Local Deviations from a Typical Expressive Timing Pattern." *Music Perception* 15.3 (Spring 1998): 265-289.

Rousseau, Jean-Jacques. "Essai sur l'origine des langues." *Musical Aesthetics: A Historical Reader*. Ed. Lippman, Edward. Vol. 1. New York: Pendragon Press, 1986.

Shaffer, L. Henry. "Performances of Chopin, Bach, and Bartok: Studies in Motor Programming." *Cognitive Psychology*. 13 (1981): 326-376.

Shaffer, L. Henry. "Timing in Solo and Duet Piano Performances." *The Quarterly Journal of Experimental Psychology*. 36A (1984): 577-595.

Shaffer, L. Henry, Eric F. Clarke, and Neil P. Todd. "Metre and rhythm in piano playing." *Cognition.* 20.1 (June 1985): 61-77.

Voss, R. F., and J. Clarke. "'1/f noise' in music and speech." *Nature.* 258 (1975): 317-318.

Windsor, W. Luke, and Eric F. Clarke. "Expressive Timing and Dynamics in Real and Artificial Musical Performances: Using an Algorithm as an Analytical Tool." *Music Perception* 15.2 (Winter 1997): 127-152

Zanette, Damien. "Zipf's law and the creation of musical context." Submitted 7 June 2004, accessed 14 October 2005. Available at http://arxiv.org/abs/cs.CL/0406015

Zipf, G. K. *Human Behaviour and the Principle of Least Effort.* Cambridge, MA: Addison-Wesley, 1949.

Additional readings

Clarke, Eric F. "Understanding the psychology of performance." Rink 59-72.

Clarke, Eric. F. "Listening to Performance." Rink 185-196.

Clarke, Eric F., and Nicholas Cook, eds. *Empirical Musicology: Aims, Methods, Prospects.* Oxford: Oxford UP, 2004.

Deutsch, Diana, ed. *The Psychology of Music.* Orlando: Academic Press, Inc., 1982.

Gabrielsson, Alf, ed. *Action and Perception in Rhythm and Music.* Stockholm: Musikaliska akademien, 1987.

Mazzola, Guerino, Thomas Noll, and Emilio Lluis-Puebla, eds. *Perspectives in Mathematical and Computational Music Theory*. Electronic Publishing Osnabrück, 2004.

Moles, Abraham. *Information Theory and Esthetic Perception*. London: U of Illinois Press, 1966.

Rink, John. "Analysis and (or?) performance." Rink 35-58.

Rink, John, ed. *Musical Performance: A Guide to Understanding.* Cambridge: Cambridge University Press, 2002.

Parncutt, Richard, and Gary E. McPherson, eds. *The Science & Psychology of Music Performance*. Oxford: Oxford UP, 2002.

Seashore, Carl E. *Psychology of Music*. New York: Dover Publications, Inc., 1938.

Sloboda, John A., ed. *Generative Processes in Music. The Psychology of Performance, Improvisation, and Composition.* Oxford: Clarendon Press, 1988.

Appendix A: Scores

Five pieces were used for the study described in this thesis. They are listed below and reproduced in this appendix.

Sonate (Pathétique), Op. 13 No. 8, Ludwig van Beethoven (mm. 1-10).

The Entertainer, Scott Joplin (mm. 1-38, no repeats, second endings).

Sechs Kleine Klavierstücke, Op. 19, No. 6, Arnold Schoenberg (mm. 1-9).

1st Gymnopédie, Erik Satie (mm. 1-39).

Minuet in G, BWV Anh. 114 No. 4, Johann Sebastian Bach (mm. 1-32, no repeats). [The composer is actually Christian Petzold; see footnote 10.]

Appendix B: MATLAB Functions

    1   PRFD: pitch
    2   PCRFD: pitch class
    3   PDRFD: pitch distance
    4   PCRFD: pitch class distance
    5   IRFD: interval (stacked and rooted)
    6   DRFD: duration
    7   VRFD: velocity

```
% 1
function [slope rsq se] = prfd(nmat)

% Function: Given a note matrix as input, finds and plots the
rank-frequency distribution for all 128 MIDI pitches, and return
the equation for the line of best fit (slope), along with r
squared value (rsq) and standard estimate of error (se).

% Step one: count up the notes
v4 = nmat(:,4);                             % isolate the fourth column = pitches
v4 = sort(v4);                              % sorts the pitches in ascending order
len = length(v4);

y = [1];                  % y will eventually become a histogram of note tallies
j = 1;                           % j will keep track of which note we're counting
for i=2:len,                                          % for each successive note,
    if v4(i) == v4(i-1),                     % if it's the same as the last one,
        y(j) = y(j) + 1;                              % add it to the current bin
    else y = [y; 1];                        % otherwise, start a new bin
        j=j+1;                                        % and go to the next bin
    end
end

% Step two: assign a rank to each tally
y = log(sort(y, 'descend'));

leny = length(y);                         % This italicized segment was sometimes
for i=1:leny-1,                           % used to remove an excess number of
    if y(leny+1-i) == 0,                  % zeroes from the y data, which would
        y = y(1:leny-i);                          % skew data if left in.
    end
end
                                          %  No data were removed when
                                          %  analyzing scores.
x = 1:length(y);                          % x will contain the ranks of each y datum
x = log(x)';
ly=length(y);
```

```
% Steps three and four: plot the data and find the regression
plot(x, y,'k.');                                        % plots the RFD
slope = polyfit(x, y, 1);                    % finds the line of best fit
rsq = regstats(x, y, 'linear', 'rsquare');
rsq = rsq.rsquare;
se = std(y) * (1 - rsq)^.5 * (ly/(ly-2))^.5;

end
```

% 2
function [slope rsq se] = pcrfd(nmat)

```
% Function: Given a note matrix as input, finds and plots the
rank-frequency distribution for all 12 pitch-classes (i.e.
treating notes as octave-equivalent), and return the equation for
the line of best fit (slope), along with r squared value (rsq)
and standard estimate of error (se).
```

% Step one: count up the notes
```
v4 = nmat(:,4);                           % isolate the fourth column = pitches
v4 = mod(v4, 12) + 1;               % convert pitches to pitch class classes,
                                           % setting C = 1, C# = 2, etc.
v4 = sort(v4);                       % sorts the pitches in ascending order
len = length(v4);

y = [1];              % y will eventually become a histogram of note tallies
j = 1;                     % j will keep track of which note we're counting
for i=2:len,                                    % for each successive note,
    if v4(i) == v4(i-1),               % if it's the same as the last one,
        y(j) = y(j) + 1;                        % add it to the current bin
    else y = [y; 1];                        % otherwise, start a new bin
        j=j+1;                                  % and go to the next bin
    end
end
```

% Step two: assign a rank to each tally
```
y = sort(y, 'descend');                          % sort in descending order
ly = length(y);
y = y(3:length(y));                      % The first 3 data points were omitted
                                  % from each distribution to improve the linear fit.
                                        % No data were removed when
                                            % analyzing the scores.

y = log(y);
x = log([1:length(y)]');
ly = length(y);
```

% Steps three and four: plot the data and find the regression
```
plot(x, y);                                        % plots the RFD
slope = polyfit(x, y, 1);                       % finds line of best fit
rsq = regstats(x, y, 'linear', 'rsquare');
rsq = rsq.rsquare;
se = std(y) * (1 - rsq)^.5 * (ly/(ly-2))^.5 ;
end
```

```
% 3
function [slope rsq se] = pdrfd(nmat)

% Function: Given a note matrix as input, finds and plots the
rank-frequency distribution for pitch distance, the time distance
between successive instances of the same MIDI pitch. Returns the
equation for the line of best fit (slope), along with r squared
value (rsq) and standard estimate of error (se).
% Note that pitch distances are rounded to the nearest pth of a
second where
p = 50

% Step one: count up the notes
v46 = nmat(:,[4 6]);                        % isolate the fourth and sixth columns
                                             % = pitch and onset time in seconds
v46 = sortrows(v46);                         % arrange rows according to pitch
len=length(v46);

for i=1:(len-1),                             % turns onset times into time distances
     v46(i,2) = v46(i+1,2) - v46(i,2);
end

v6=[];                                       % eliminate the time distances bridging
for i=1:len-1,                                      % different pitches
     if v46(i,1) == v46(i+1,1),
          v6 = [v6; v46(i,2)];
     end
end

v6 = round(v6*p)/p;                          % round time distances to nearest 1/p secs.
v6 = sort(v6);
len = length(v6);

y = [1];                                       % tally up the number of each
ly = 1;                                      % pitch ditsance into the vector y
for i=2:len,
     if v6(i) == v6(i-1), y(ly)=y(ly)+1;
     else y = [y; 1];                                             ly = ly + 1;
     end;
end

% Step two: assign a rank to each tally
y = sort(y, 'descend');                                  % sort in descending order
y = log(y);

leny = length(y);                            % This italicized segment was sometimes
for i=1:leny-1,                              % used to remove an excess number of
     if y(leny+1-i) == 0,                    % zeroes from the y data, which would
          y = y(1:leny-i);                            % skew data if left in.
     end
end
y = y(2:,length(y)-1);                       % Alternatively, where no excess zeroes
                                             % were found, it improved results to remove
                                             % the first and last data points.
```

```matlab
                                     %  No data were removed when
                                     %  analyzing scores.
ly = length(y);
x = log([1:ly]');

% Steps three and four: plot the data and find the regression
plot(x, y, 'k.');                                    % plots the RFD
slope = polyfit(x, y, 1);                     % finds line of best fit
rsq = regstats(x, y, 'linear', 'rsquare');
rsq = rsq.rsquare;
se = std(y) * (1 - rsq)^.5 * (ly/(ly-2))^.5 ;


% 4
function [slope rsq se] = pcdrfd(nmat)

% Function: Given a note matrix as input, finds and plots the
rank-frequency distribution for pitch-class distance, the time
distance between successive instances of the same octave-
equivalent pitch. Returns the equation for the line of best fit
(slope), along with r squared value (rsq) and standard estimate
of error (se).
% Note that pitch distances are rounded to the nearest pth of a
second where
p = 50

% Step one: count up the notes
v46 = nmat(:,[4 6]);                    % isolate the fourth and sixth columns
                                        % = pitch and onset time in seconds
v46(:,1) = mod( v46(:,1), 12) + 1;           % convert pitches to pitch
                                        % classes, setting C = 1, C# = 2, etc.
v46 = sortrows(v46);                     % arrange rows according to pitch
len=length(v46);

for i=1:(len-1),                        % turns onset times into time distances
        v46(i,2) = v46(i+1,2) - v46(i,2);
end

v6=[];                                  % eliminate the time distances bridging
for i=1:len-1,                                      % different pitches
        if v46(i,1) == v46(i+1,1),
                v6 = [v6; v46(i,2)];
        end
end

v6 = round(v6*p)/p;                 % round time distances to nearest 1/p secs.
v6 = sort(v6);
len = length(v6);

y = [1];                                        % tally up the number of each
ly = 1;                                         % pitch ditsance into the vector y
for i=2:len,
        if v6(i) == v6(i-1), y(ly)=y(ly)+1;
```

```
        else y = [y; 1];                                              ly = ly + 1;
        end;
end


% Step two: assign a rank to each tally
y = sort(y, 'descend');                        % sort in descending order
y = log(y);


leny = length(y);                       % This italicized segment was sometimes
for i=1:leny-1,                           % used to remove an excess number of
        if y(leny+1-i) == 0,              % zeroes from the y data, which would
                y = y(1:leny-i);                     % skew data if left in.
        end
end
y = y(2:,length(y)-1);                     % Alternatively, where no excess zeroes
                                         % were found, it improved results to remove
                                                % the first and last data points.


                                                  %  No data were removed when
                                                     % analyzing scores.
ly = length(y);
x = log([1:ly]');                                          % set up the x data


% Steps three and four: plot the data and find the regression
plot(x, y, 'k.');                                              % plots the RFD
slope = polyfit(x, y, 1);                             % finds line of best fit
rsq = regstats(x, y, 'linear', 'rsquare');
rsq = rsq.rsquare;
se = std(y) * (1 - rsq)^.5 * (ly/(ly-2))^.5 ;


% 5
function [slopes rsqs ses sloper rsqr ser] =
irfd(nmat)

% Function: Given a note matrix as input, finds and plots the
rank-frequency distribution for both stacked harmonic intervals
(each pitch in a chord is related to the next lower note) and
rooted harmonic intervals (each pitch in a chord is related to
the lowest root of the chord). Returns the equations for both
lines of best fit (slopes and sloper), along with r squared
values (rsqs, rsqr) and standard estimate of error (ses, ser). In
both cases harmonic intervals are treated as octave equivalent
(so that a minor 9th and a minor 2nd count as equals).
% Note that the intervals are recounted for every time-slice they
occur in where the time-slices are 1 pth of a second, where
p = 50

% create a vector to count the number of each kind of interval
ys = 0*[1:12];
yr = 0*[1:12];
```

```matlab
% Step one: count up the intervals
v = nmat(:,[4 6 7]);                      % isolate the fourth (pitch), sixth
len = length(v);                          % (onset) and seventh (duration) columns,
v = sortrows(v, [2 1 3]);                 % sort the pitches in chronological
v = round(v*p)/p;                         % order, and round times off to the
                                          % nearest pth of a second

w=[];          % this little loop gets rid of rows with
for i=1:len,                              % notes of zero length, which sometimes are
      if v(i,3) ~= 0,                     % accidentally obtained in performance
            w = [w; v(i,:)];
      end
end
v=w;
len = length(v);

v1 = v(:,2)+v(:,3);                       % add duration to onset times
v1 = max(v1);                             % find the latest instant at which
                                          % notes are sounding
v1 = v1*p;  % count how many time steps are in the piece

% What the next little subroutine does is, for every note event
in the piece (for each row i), it finds out how many timesteps it
sounds during (isteps), and then adds 'isteps'-many notes of the
same pitch v(i,1) starting at each successive timestep beginning
at the onset time v(i,2).

w=[];
for i=1:len,
      isteps = v(i,3)*p;
      for j=1:isteps,
            w = [w; v(i,1), v(i,2)*p + j];
      end
end

w = sortrows(w, [2 1]);
lenw = length(w);

% At this point, separate calculations are required for the
stacked and rooted RFDs. We will do stacked first, and save a
copy of w as ww for the rooted intervals later.
ww = w;

for i=1:lenw-1,                           % tallies up the different intervals
      if w(i,2) == w(i+1,2),
            a = mod(w(i+1,1) - w(i,1), 12) + 1;
            ys(a) = ys(a) + 1;
      end
end
```

```matlab
% Step two: assign a rank to each tally
ys = sort(ys, 'descend');                        % sort the data in descending order
for i=1:11, % remove any zeroes
     if ys(13-i) == 0,
          ys = ys(1:12-i);
     end
end
ys = log( ys );

s = ys(2:,length(ys)-1);                  % There were never any excess zeroes
                                          % for the stacked intervals, so only
                                          % the first and last data points were
                                          % removed to improve the linear fit.
                                          %  No data were removed when
                                          %  analyzing scores.

ly = length(ys);
x = log( 1:ly );

% Steps three and four: plot the data and find the regression
plot(x, ys, 'k.');                                      % plots the RFD
slopes = polyfit(x, ys, 1);                    % find line of best fit
rsqs = regstats(x, ys, 'linear', 'rsquare');
ys=ys';
rsqs = rsqs.rsquare;
ses = std(ys) * (1 - rsqs)^.5 * (ly/(ly-2))^.5 ;


% Now we go back to the rooted interval part:
% Step one (continued): count up the intervals

rooti = 1;  % sets an initial root index
for i=1:lenw-1,                            % tallies the different intervals,
     if ww(i,2) == ww(i+1,2),                        % relative to the root
          a = mod(ww(i+1,1) - ww(rooti,1), 12) + 1;
          yr(a) = yr(a) + 1;
          else rooti = ww(i+1,1);
     end
end

% Step two: assign a rank to each tally
yr = sort(yr, 'descend');                          % sort in descending order
for i=1:11, % take out zeroes
     if yr(13-i) == 0,
          yr = yr(1:12-i);
     end
end
yr = log(yr);

                              %  No data were removed for rooted intervals when
                              %  analyzing either scores or performances!


leny = length(yr);
x = log( 1:leny );
```

```
% Steps three and four: plot the data and find the regression
plot(x, yr, 'k.');                                      % plots the RFD
sloper = polyfit(x, yr, 1);                      % find line of best fit
rsqr = regstats(x, yr, 'linear', 'rsquare');
% plot(x, yr);
yr=yr';
rsqr = rsqr.rsquare;
ly = length(yr);
ser = std(yr) * (1 - rsqr)^.5 * (ly/(ly-2))^.5 ;

end


% 6
function [slope rsq se] = drfd(nmat)

% Function: Given a note matrix as input, finds and plots the
rank-frequency distribution for note durations. Returns the
equation for the line of best fit (slope), along with r squared
value (rsq) and standard estimate of error (se).
% Note that durations are rounded to the nearest pth of a second
where
p = 50

% Step one (continued): count up the intervals
v7 = nmat(:,7);                        % isolate the seventh column = durations
v7 = sort(v7, 'descend');                       % sort out the pitches
len = length(v7);
v7 = round(v7*p)/p;                      % round the durations to the nearest
                                                    % pth of a second

y = [1];                               % y will tally the various note lengths
ly = 1;                                % ly is the length index of this vector
for i=2:len,
    if v7(i) == v7(i-1), y(ly)=y(ly)+1;
                                % add one to the current vector if it's the
    else y = [y; 1];                        % otherwise move onto the next entry
    ly = ly + 1;                    % reflect the increase the length index
    end;
end

% Step two: assign a rank to each tally
y = sort(y, 'descend');
y = log(y);

leny = length(y);                      % This italicized segment was sometimes
for i=1:leny-1,                        % used to remove an excess number of
    if y(leny+1-i) == 0,               % zeroes from the y data, which would
        y = y(1:leny-i);               % skew data if left in.
    end
end
y = y(2:,length(y)-1);                 % Alternatively, where no excess zeroes
                                       % were found, it improved results to remove
                                       % the first and last data points.
```

```
                                                         %  No data were removed when
                                                         % analyzing scores.
ly = length(y);
x = log([1:ly]');

% Steps three and four: plot the data and find the regression
plot(dx, dy,'k.');                                          % plots the RFD
slope = polyfit(x, y, 1);                          % find line of best fit
rsq = regstats(x, y, 'linear', 'rsquare');
rsq = rsq.rsquare;
ly=length(y);
se = std(y) * (1 - rsq)^.5 * (ly/(ly-2))^.5 ;

end
```

% 7
function [slope rsq se] = vrfd(nmat)

% Function: Given a note matrix as input, finds and plots the
rank-frequency distribution for note velocities. Returns the
equation for the line of best fit (slope), along with r squared
value (rsq) and standard estimate of error (se).

% Step one (continued): count up the intervals

```
v5 = nmat(:,5);                          % isolate the fifth column = velocities
v5 = sort(v5, 'descend');                              % sort out velocites
len = length(v5);

y = [1];                      % y will tally up the frequency of each velocity
ly = 1;                       % ly is the length index of this vector
for i=2:len,
    if v5(i) == v5(i-1), y(ly) = y(ly) + 1;
                                      % add one to the current vector if it's the
                                      % same value as the previous entry
        else y = [y; 1];              % otherwise move onto the next entry
        ly = ly + 1;                  % reflect the increase the length index
        end;
end
```

% Step two: assign a rank to each tally
```
y = sort(y, 'descend');
y = log(y);
```

leny = length(y);   % This italicized segment was sometimes
for i=1:leny-1,   % used to remove an excess number of
*   if y(leny+1-i) == 0,   % zeroes from the y data, which would*
*   y = y(1:leny-i);   % skew data if left in.*
*   end*
end
y = y(2:,length(y)-1);   % Alternatively, where no excess zeroes
                                        % were found, it improved results to remove

```matlab
                                                      % the first and last data points.

                                                      %  No data were removed when
                                                      % analyzing scores.
ly = length(y);
x = log([1:ly]');

% Steps three and four: plot the data and find the regression
plot(x, y,'k.');                                      % plots the RFD
slope = polyfit(x, y, 1);                             % find line of best fit
rsq = regstats(x, y, 'linear', 'rsquare');
rsq = rsq.rsquare;
se = std(y) * (1 - rsq)^.5 * (ly/(ly-2))^.5 ;

end
```

Appendix C: Tables

Table 1. Summary of analyzed metrics

Metric	Function name	Description
Pitch	prfd	RFD of the 128 MIDI pitches
Pitch-class	pcrfd	RFD of the 12 pitch classes
Pitch distance	pdrfd	RFD of the time distances between pitch repetitions
Pitch-class distance	pcdrfd	RFD of the time distances between pitch-class repetitions
Interval (stacked)	irfd	RFD of chord intervals, relative to next-lower note
Interval (rooted)	"	RFD of chord intervals, relative to lowest note of chord
Duration	drfd	RFD of note durations
Velocity	vrfd	RFD of note velocities (volume)

Table 2. Results of ANOVA tests for differences between performers. $F(5, 18) = 6.2$, $p=0.05$. Values above critical value of 6.2 (boldfaced) indicate significant differences.

Piece	Pitch	Pitch Class	Pitch Distance	PC Distance	Interval (Stacked)	Interval (Rooted)	Duration	Velocity
Beethoven	4.84	2.11	5.01	**21.80**	1.45	1.71	**12.81**	5.29
Joplin	0.66	1.83	3.25	5.80	**9.22**	4.73	3.70	3.73
Schoenberg	1.27	3.91	3.36	4.20	3.34	3.29	**12.02**	**20.31**
Satie	2.14	2.10	3.95	**6.31**	2.60	2.07	**6.60**	**12.02**
Bach	1.65	2.86	3.78	2.13	5.60	1.18	**9.70**	4.57

Table 3. Average change in metric slope between score and performance.

Metric	Average change in slope	Standard deviation of slope changes
Pitch distance	0.45	0.19
Pitch-class distance	0.56	0.37
Stacked interval	0.35	0.23
Rooted interval	0.18	0.12
Duration	2.34	1.09
Velocity	3.19	1.88

Table 4. Average over all performances for metric slope with standard deviation

Metric	Average slope	Standard deviation of slopes
Pitch	0.79	0.20
Pitch-class	0.88	0.29
Pitch distance	0.56	0.23
Pitch-class distance	0.65	0.31
Stacked interval	0.96	0.20
Rooted interval	0.85	0.23
Duration	0.71	0.26
Velocity	0.65	0.06
All metrics averaged together:	0.76	0.27

Table 5. Results of linear regression analysis to determine four models of metric slope change in performance.

Metric	Coefficient of dependence	Intercept	R^2	Estimate of error
Pitch and pitch class distance	0.46	0.10	0.38	0.25
Stacked and rooted intervals	0.55	0.26	0.56	0.20
Duration	0.16	0.45	0.30	0.25
Velocity	-0.003	0.66	0.003	0.06

Appendix D: Figures

Figure 2. Rank-frequency distribution of note duration for one performance of Beethoven`s Pathétique Sonata.

$$y = -1.0315x + 4.8204$$
$$R^2 = 0.8960$$
$$error = 0.3288$$
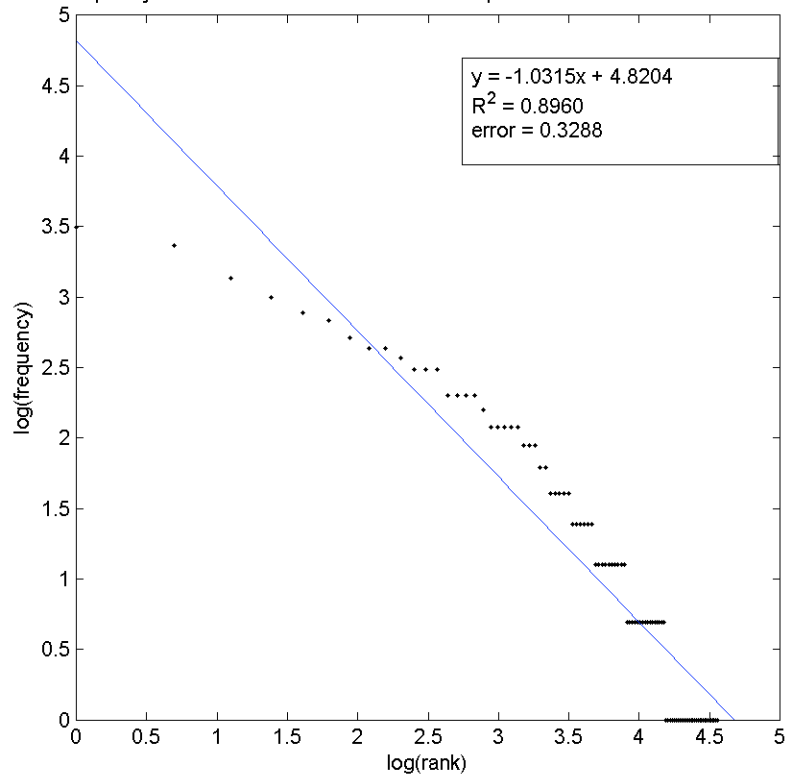
Figure 2b. Rank-frequency distribution of note velocity for one performance of Joplin`s `The Entertainer`.

$$y = -1.0315x + 4.8204$$
$$R^2 = 0.8960$$
$$error = 0.3288$$

g 4a, Beethoven, Pathetique. Comparison between metrical analysis of the score and average metrical analysis for each perfc

The 8 metrics: Pitch, Pitch Class, Pitch Distance, Pitch Class Distance, Interval (Stacked), Interval (Rooted), Duration, Veloc
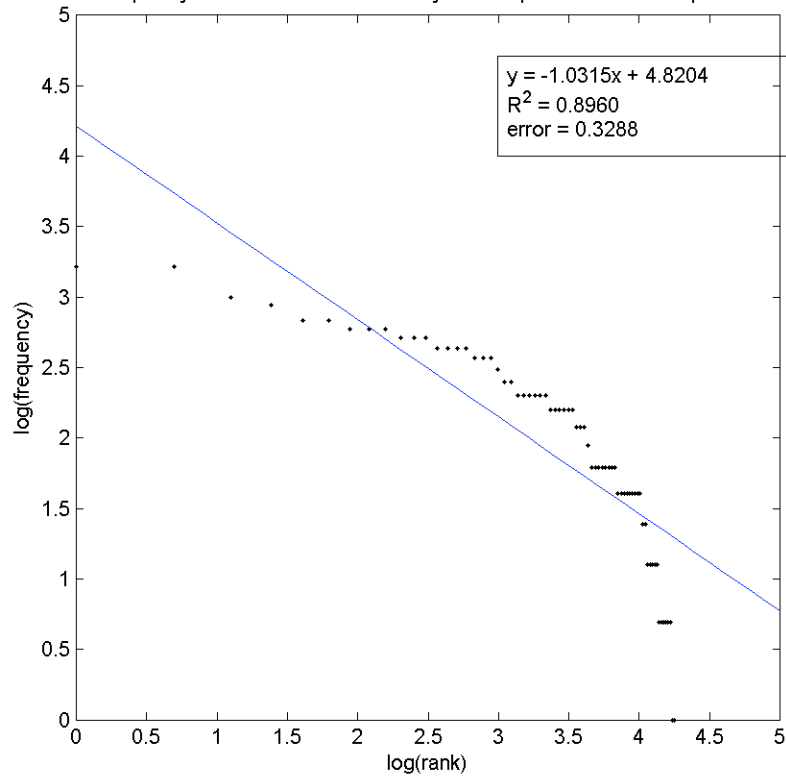
ig 4b, Joplin, The Entertainer. Comparison between metrical analysis of the score and average metrical analysis for each perfo
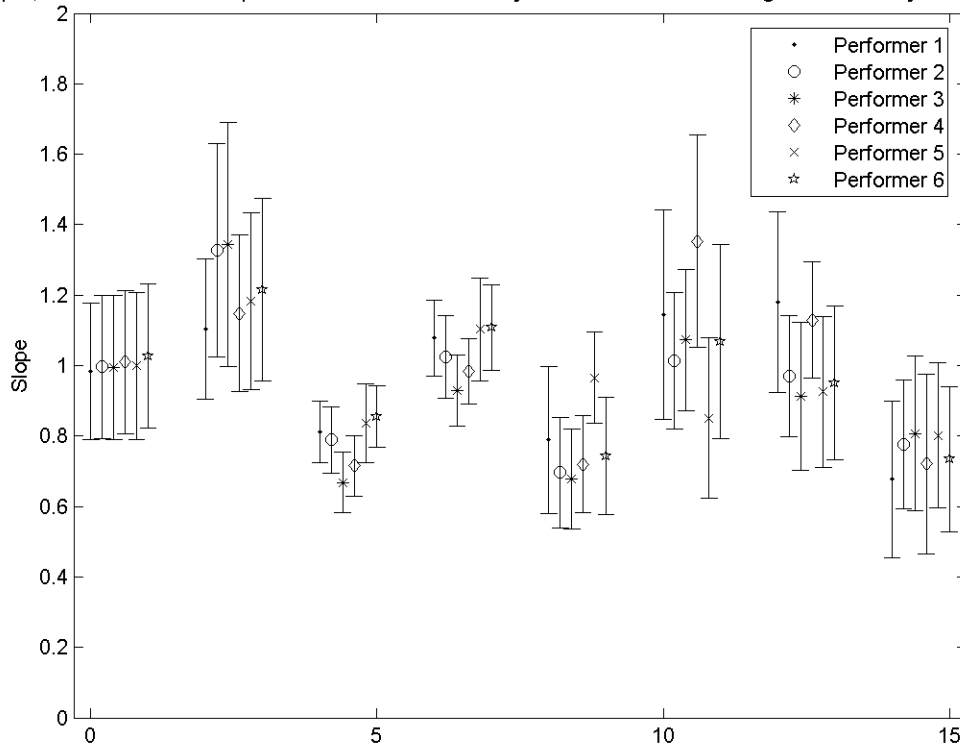
The 8 metrics: Pitch, Pitch Class, Pitch Distance, Pitch Class Distance, Interval (Stacked), Interval (Rooted), Duration, Veloc

choenberg, Short Piece Piece no. 6. Comparison between metrical analysis of the score and average metrical analysis for eac
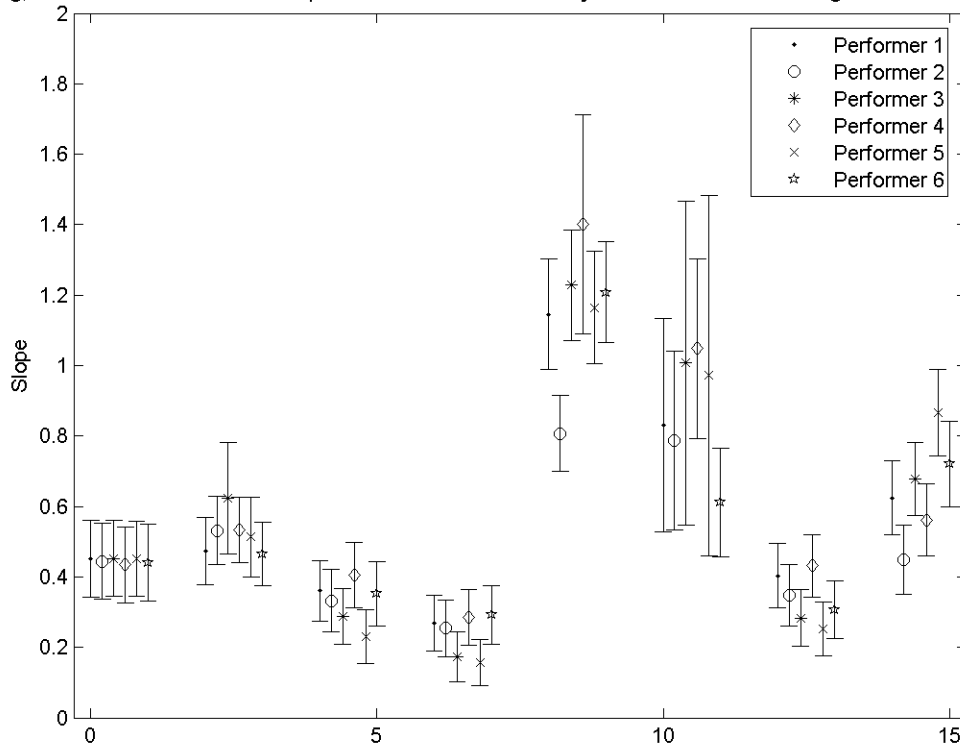
The 8 metrics: Pitch, Pitch Class, Pitch Distance, Pitch Class Distance, Interval (Stacked), Interval (Rooted), Duration, Veloc

The 8 metrics: Pitch, Pitch Class, Pitch Distance, Pitch Class Distance, Interval (Stacked), Interval (Rooted), Duration, Veloc
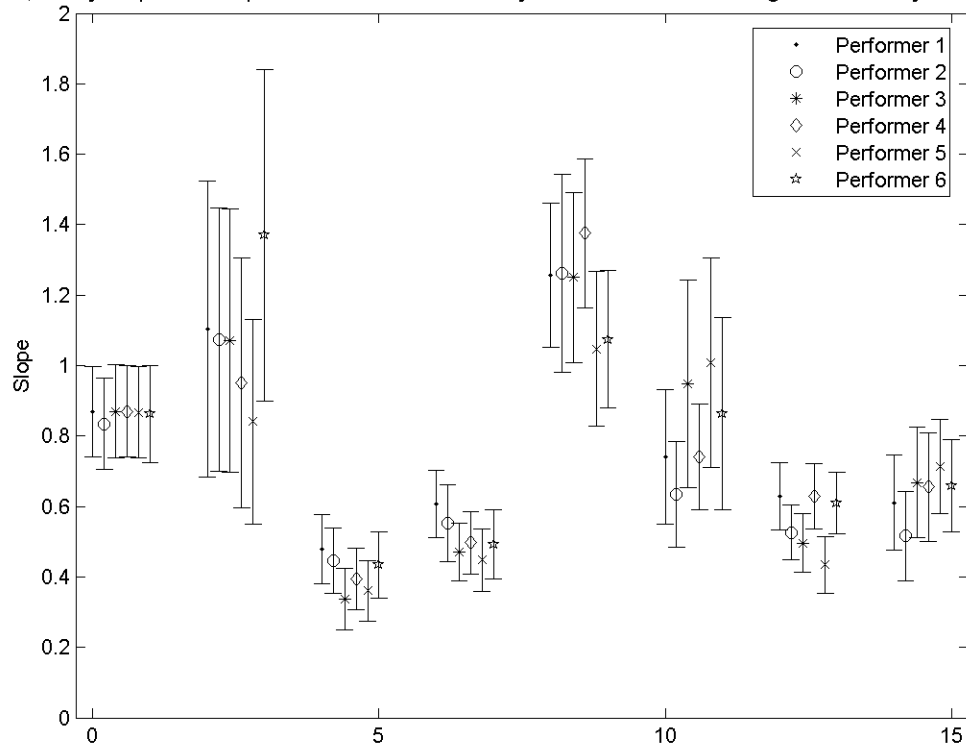
Fig 4e, Bach, Minuet in G. Comparison between metrical analysis of the score and average metrical analysis for each perform

The 8 metrics: Pitch, Pitch Class, Pitch Distance, Pitch Class Distance, Interval (Stacked), Interval (Rooted), Duration, Veloc
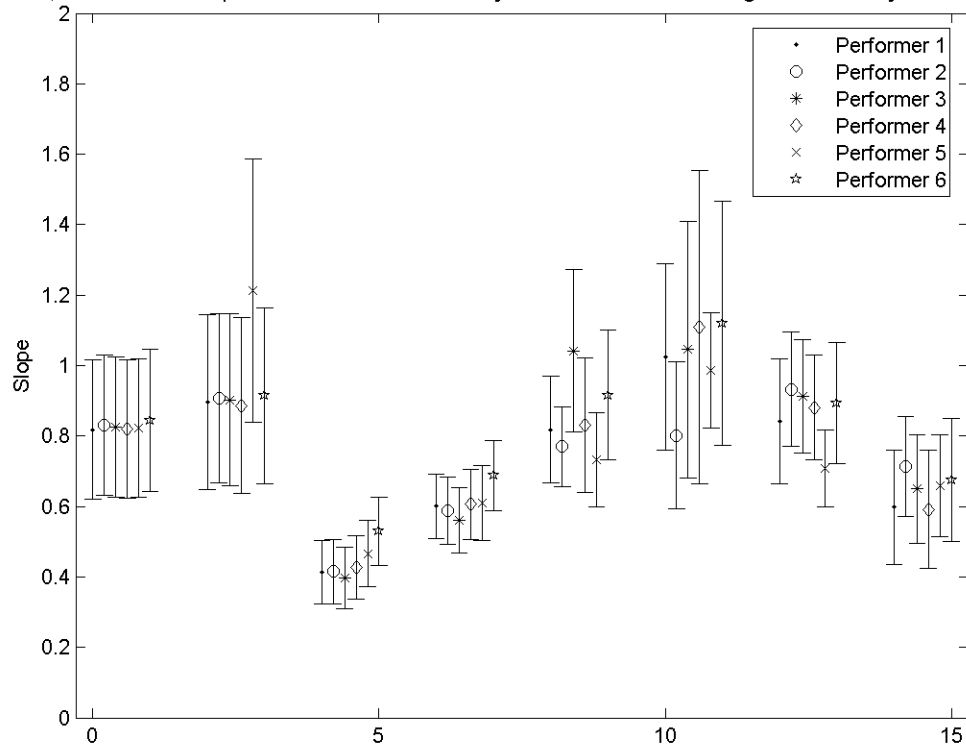
Figure 6a. Pitch. Difference in average metric slope between score and performance.
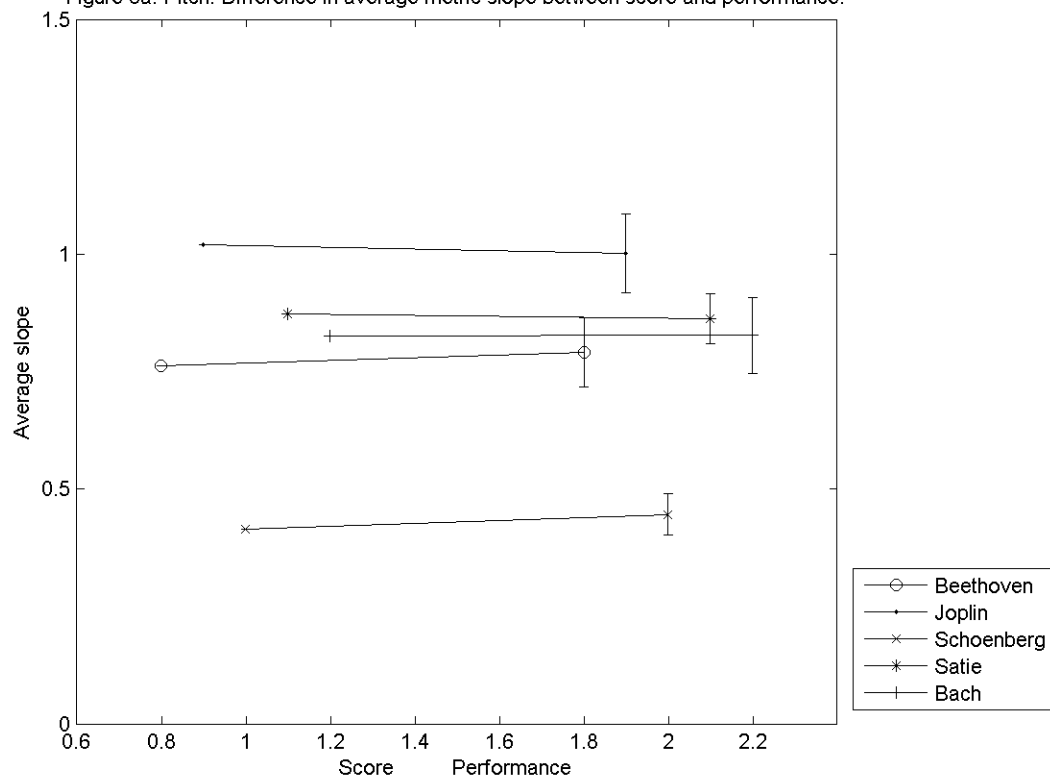
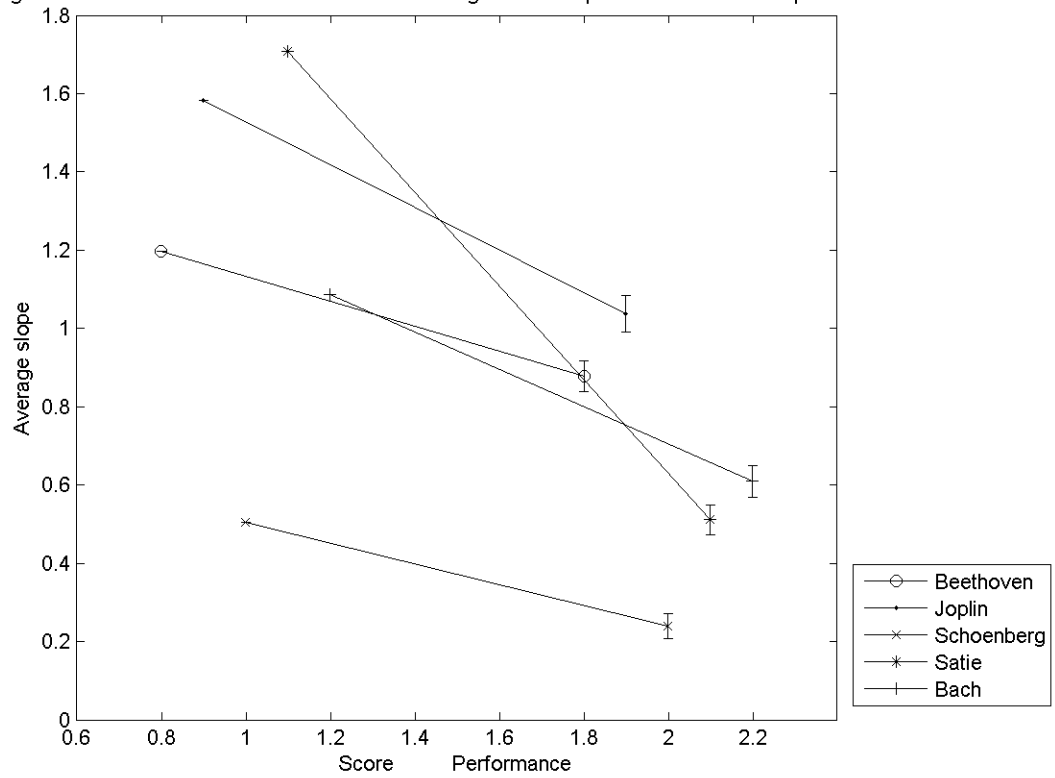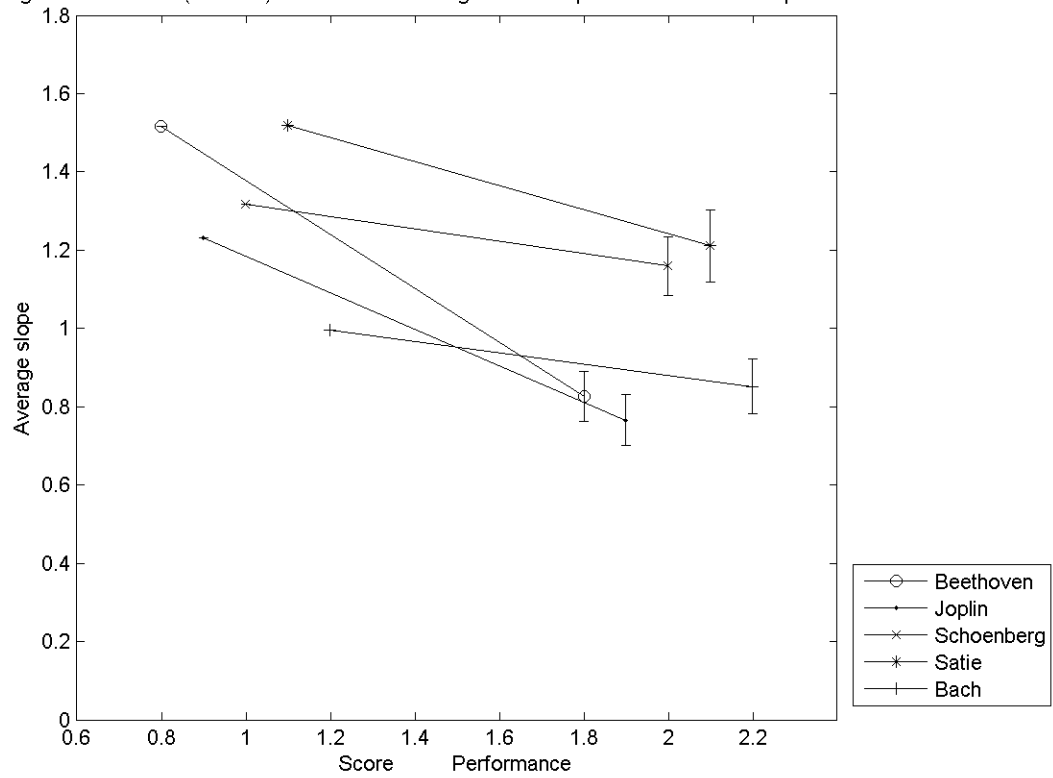Figure 6b. Pitch Class. Difference in average metric slope between score and performance.
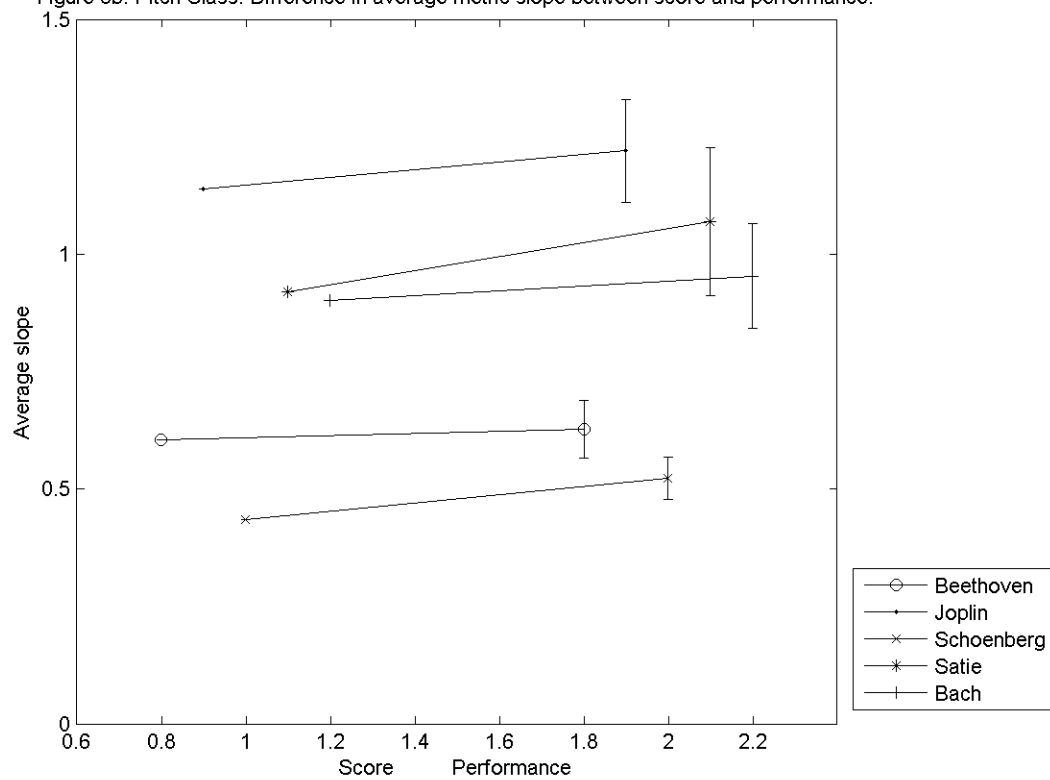
Figure 6c. Pitch Distance. Difference in average metric slope between score and performance.
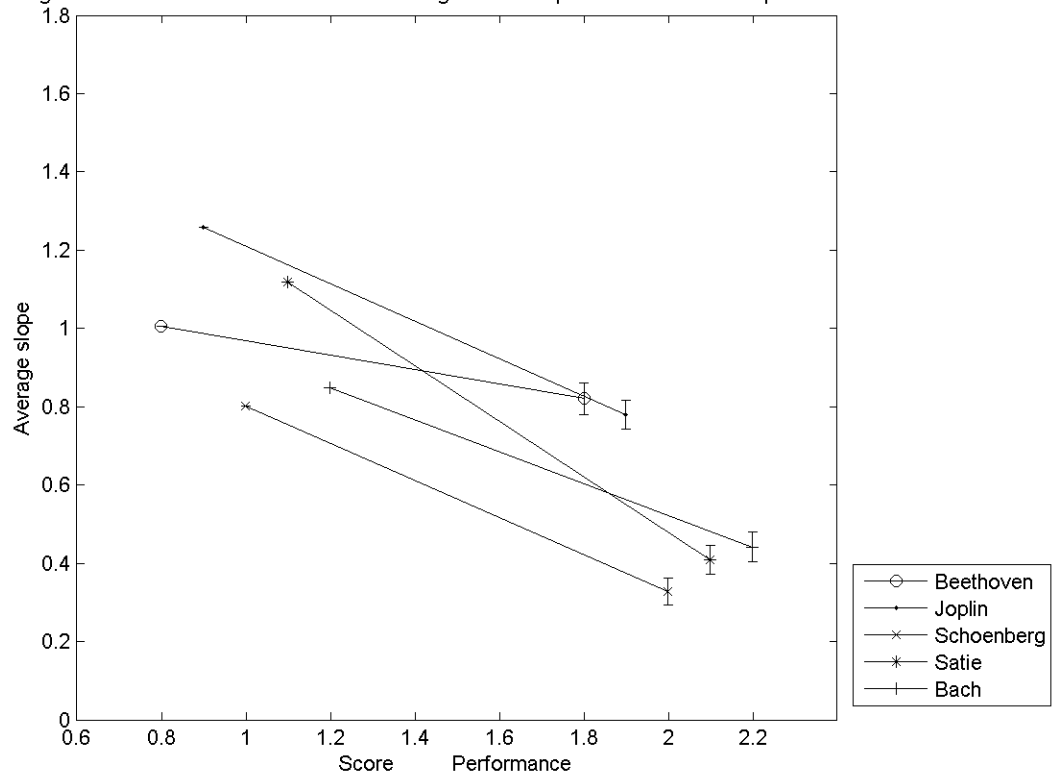
Figure 6d. Pitch Class Distance. Difference in average metric slope between score and performance.

Figure 6e. Intervals (Stacked). Difference in average metric slope between score and performance.

Figure 6f. Intervals (Rooted). Difference in average metric slope between score and performance.
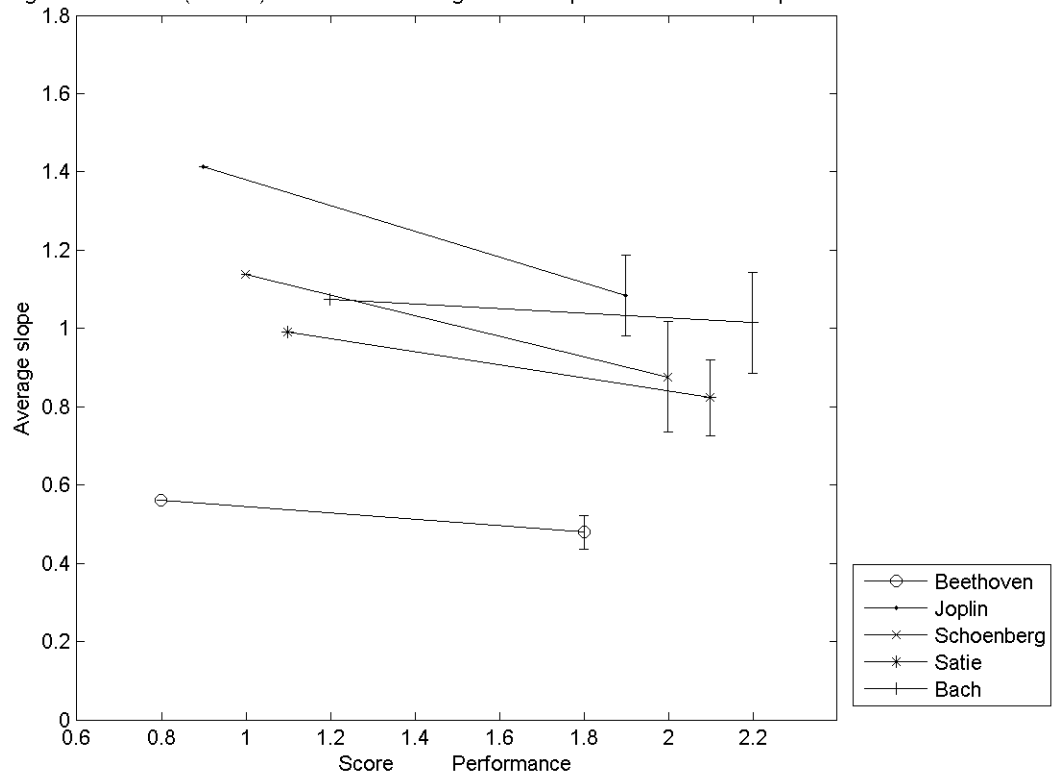
Figure 6g. Note Duration. Difference in average metric slope between score and performance.
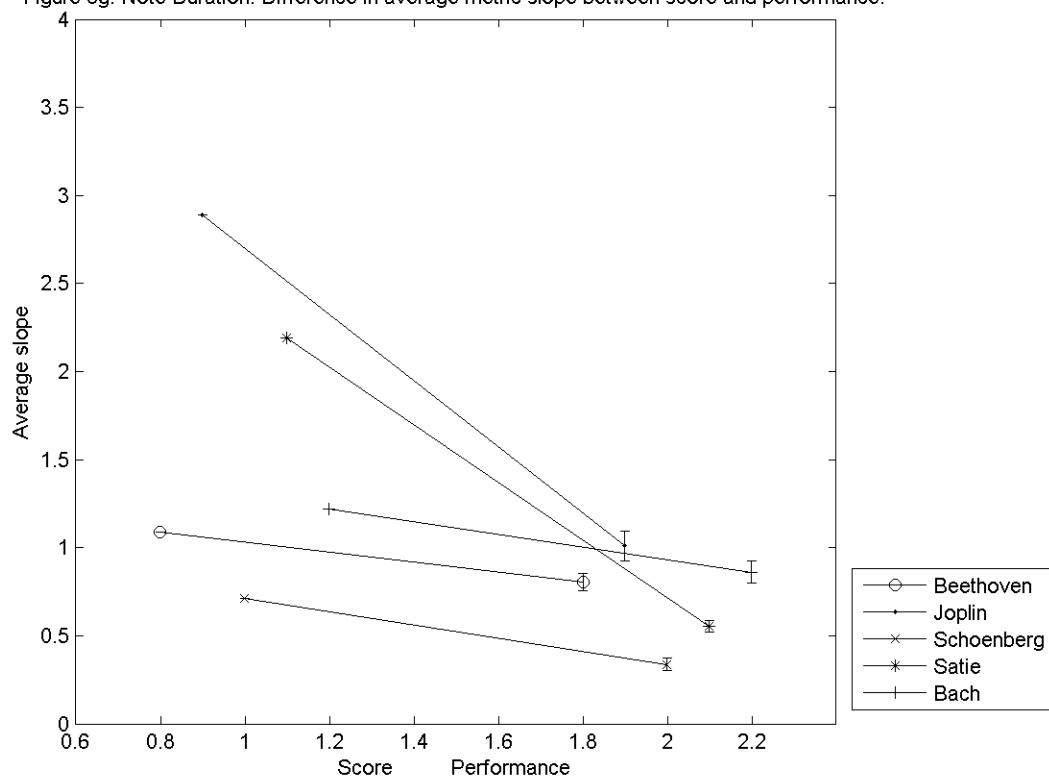

Figure 6h. Note Velocity. Difference in average metric slope between score and performance.