Bill Manaris,[*] Juan Romero,[†] Penousal Machado,[‡] Dwight Krehbiel,[§] Timothy Hirzel,[*] Walter Pharr,[*] and Robert B. Davis[¶]
[*]Computer Science Department, College of Charleston
66 George Street, Charleston, SC 29424 USA
{manaris, hirzel, pharr}@cs.cofc.edu
[†]Creative Computer Group, RNASA Lab
Faculty of Computer Science, University of A Coruña, Spain
jj@udc.es
[‡]Centre for Informatics and Systems Department of Informatic Engineering Polo II-University of Coimbra
3030 Coimbra, Portugal
machado@dei.uc.pt
[§]Psychology Department, Bethel College
North Newton, KS 67117 USA
krehbiel@bethelks.edu
[¶]Department of Mathematics and Statistics, Miami University
Hamilton, OH 45011, USA
davisrb@muohio.edu

Zipf's Law, Music Classification, and Aesthetics

The connection between aesthetics and numbers dates back to pre-Socratic times. Pythagoras, Plato, and Aristotle worked on quantitative expressions of proportion and beauty such as the golden ratio. Pythagoreans, for instance, quantified "harmonious" musical intervals in terms of proportions (ratios) of the first few whole numbers: a unison is 1:1, octave is 2:1, perfect fifth is 3:2, perfect fourth is 4:3, and so on (Miranda 2001, p. 6). The Pythagorean scale was refined over centuries to produce well-tempered and equal-tempered scales (Livio 2002, pp. 29, 186).

Galen, summarizing Polyclitus, wrote, "Beauty does not consist in the elements, but in the harmonious proportion of the parts." Vitruvius stated, "Proportion consists in taking a fixed nodule, in each case, both for the parts of a building and for the whole." He then defined proportion as "the appropriate harmony arising out of the details of the work itself; the correspondence of each given detail among the separate details to the form of the design as a whole." This school of thought crystallized into a universal theory of aesthetics based on "unity in variety" (Eco 1986, p. 29).

Computer Music Journal, 29:1, pp. 55–69, Spring 2005

Some musicologists dissect the aesthetic experience in terms of separable, discrete sounds. Others attempt to group stimuli into patterns and study their hierarchical organization and proportions (May 1996; Nettheim 1997). Leonard Meyer states that emotional states in music (sad, angry, happy, etc.) are delineated by statistical parameters such as dynamic level, register, speed, and continuity (2001, p. 342).

Building on earlier work by Vilfredo Pareto, Alfred Lotka, and Frank Benford (among others), George Kingsley Zipf refined a statistical technique known as *Zipf's Law* for capturing the scaling properties of human and natural phenomena (Zipf 1949; Mandelbrot 1977, pp. 344–345).

We present results from a study applying Zipf's Law to music. We have created a large set of metrics based on Zipf's Law that measure the proportion or distribution of various parameters in music, such as pitch, duration, melodic intervals, and harmonic consonance. We applied these metrics to a large corpus of MIDI-encoded pieces. We used the generated data to perform statistical analyses and train artificial neural networks (ANNs) to perform various classification tasks. These tasks include author attribution, style identification, and "pleasantness" prediction. Results from the author attribution and style identification ANN experiments have appeared in Machado et al. (2003, 2004) and Manaris et al. (2003), and these results are summarized in this article. Results from the "pleasantness" prediction ANN experiment are new and therefore discussed in detail. Collectively, these results suggest that metrics based on Zipf's Law may capture essential aspects of proportion in music as it relates to music aesthetics.

Zipf's Law

Zipf's Law reflects the scaling properties of many phenomena in human ecology, including natural language and music (Zipf 1949; Voss and Clarke 1975). Informally, it describes phenomena where small events are quite frequent and large events are rare. Once a phenomenon has been selected for study, we can examine the contribution of each event to the whole and rank it according to its "importance" or "prevalence" (see linkage.rockefeller.edu/wli/zipf). For example, we may rank unique words in a book by their frequency of occurrence, visits to a Web site by how many of them originated from the same Internet address, and so on.

In its most succinct form, Zipf's Law is expressed in terms of the frequency of occurrence (i.e., count or quantity) of events, as follows:

$$F \sim r^{-a} \tag{1}$$

where F is the frequency of occurrence of an event within a phenomenon, r is its statistical rank (position in an ordered list), and a is close to 1. In the book example above, the most frequent word would be rank 1, the second most frequent word would be rank 2, and so on. This means that the frequency of occurrence of a word is inversely proportional to its rank. For example, if the first ranked word appears 6,000 times, the second ranked word would appear approximately 3,000 times (1/2), the third ranked word approximately 2,000 times (1/3), and so on.

Another formulation of Zipf's Law is

$$P(f) \sim 1/f^n \tag{2}$$

where $P(f)$ denotes the probability of an event of rank f, and n is close to 1. In physics, Zipf's Law is a special case of a power law. When n is 1 (Zipf's ideal), the phenomenon is called $1/f$ noise or *pink noise.*

Zipf distributions (e.g., $1/f$ noise) have been discovered in a wide range of human and naturally occurring phenomena including city sizes, incomes, subroutine calls, earthquake magnitudes, thickness of sediment depositions, extinctions of species, traffic jams, and visits to Web sites (Schroeder 1991; Bak 1996; Adamic and Huberman et al. 2000; see also linkage.rockefeller.edu/wli/zipf).

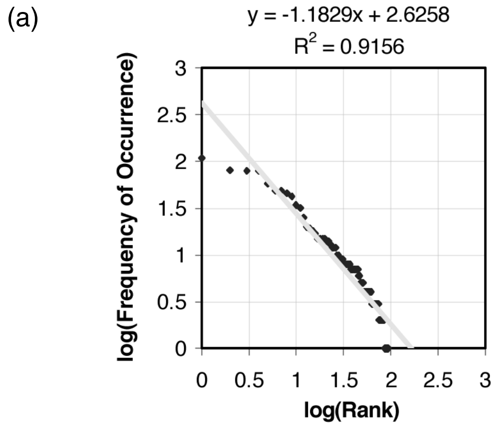
In the case of music, we can study the "importance" or "prevalence" of pitch events, duration events, melodic interval events, and so on. For instance, consider Chopin's *Revolutionary Etude.* To determine if its melodic intervals follow Zipf's Law, we count the different melodic intervals in the piece, e.g., 89 half steps up, 88 half steps down, 80 unisons, 61 whole steps up, and so on. Then we plot these counts against their statistical rank on a log-log scale. This plot is known as the *rank-frequency* distribution.

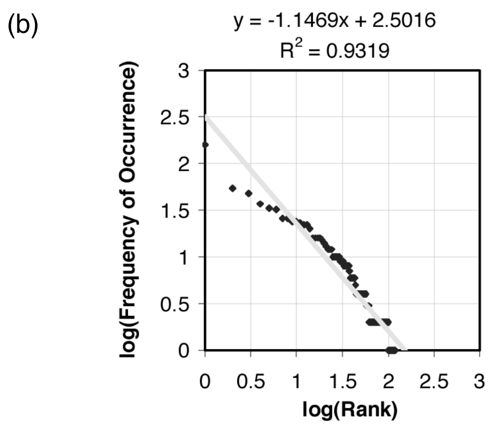
In general, the slope of the distribution may range from 0 to $-\infty$, with -1 denoting Zipf's ideal. In other words, this slope corresponds to the exponent n in Equation 2. The R^2 value can range from 0 to 1, with 1 denoting a straight line. This value gives the proportion of y-variability of data points with respect to the trend line.

Figure 1a shows the rank-frequency distribution of melodic intervals for Chopin's *Revolutionary Etude.* Melodic intervals in this piece approximate a Zipfian distribution with slope of -1.1829 and R^2 of 0.9156. Figure 1b shows the rank-frequency distribution of chromatic-tone distance for Bach's *Air on the G String.* The *chromatic-tone distance* is the time interval between consecutive repetitions of chromatic tones. In this piece, the chromatic-tone distance approximates a Zipfian distribution with slope of -1.1469 and R^2 of 0.9319. It should be noted that the less pronounced fit at the tails (high and low ranking events) is quite common in Zipf plots of naturally occurring phenomena.

In many cases, the statistical rank of an event is inversely related to the event's size. Informally, smaller events tend to occur more frequently, whereas larger events tend to occur less frequently. For instance, the statistical rank of chromatic-tone distance in

Figure 1. (a) Rank-frequency distribution of melodic intervals for Chopin's Revolutionary Etude, Op. 10 No. 12 in C minor; *(b) rank-frequency distribution of chromatic-tone distance for Bach's* Orchestral Suite No. 3 in D, movement no. 2, Air on the G String, *BWV 1068.*





Mozart's *Bassoon Concerto in B-flat Major* is inversely related to the length of these time intervals (Zipf 1949, p. 337). In other words, plotting the counts of various distances against the actual distances (from smaller to larger) produces a near-Zipfian line.

Using size instead of rank on the *x*-axis generates a *size-frequency* distribution. This is an alternative formulation of Zipf's Law that has found application in architecture and urban studies (Salingaros and West 1999). This formulation is also used in the *box-counting* technique for calculating the fractal dimension of phenomena (Schroeder 1991, p. 214).

Zipf's Law has been criticized on the grounds that $1/f$ noise can be generated from random statistical processes (Li 1992, 1998; Wolfram 2002, p. 1014). However, when studied in depth, one realizes that Zipf's Law captures the scaling properties of a phenomenon (Mandelbrot 1977, p. 345; Ferrer Cancho and Solé 2003). In particular, Benoit Mandelbrot, an early critic, was inspired by Zipf's Law and went on to develop the field of fractals. He states:

> Natural scientists recognize in "Zipf's Laws" the counterparts of the scaling laws which physics and astronomy accept with no extraordinary emotion—when evidence points out their validity. Therefore physicists would find it hard to imagine the fierceness of the opposition when Zipf—and Pareto before him—followed the same procedure, with the same outcome, in the social sciences. (Mandelbrot 1977, pp. 403–404)

Zipf-Mandelbrot Law

Mandelbrot generalized Zipf's Law as follows:

$$P(f) \sim 1/(1 + b)f^{(1 + c)} \tag{3}$$

where b and c are arbitrary real constants. This is known as the *Zipf-Mandelbrot Law.* It accounts for natural phenomena whose scaling properties are not necessarily Zipfian.

Zipf's Law in Music

Zipf himself reported several examples of $1/f$ distributions in music. His examples were processed manually, because computers were not yet available. Zipf's corpus consisted of Mozart's *Bassoon Concerto in B-flat;* Chopin's *Etude in F minor, Op. 25, No. 2;* Irving Berlin's *Doing What Comes Naturally;* and Jerome Kern's *Who.* This study focused on melodic intervals and the distance between repetitions of notes (Zipf 1949, pp. 336–337).

Richard Voss and John Clarke (1975, 1978) conducted a large-scale study of music from classical, jazz, blues, and rock radio stations recorded continuously over 24 hours. They measured several parameters, including output voltage of an audio amplifier, loudness fluctuations of music, and pitch fluctuations of music. They discovered that pitch and loudness fluctuations in music follow Zipf's distribution. Additionally, Voss and Clarke developed a computer program to generate music using three different ran-

Table 1. Our Current Set of 20 Simple Metrics Based On Zipf's Law

Metric	*Description*
Pitch	Rank-frequency distribution of the 128 MIDI pitches
Chromatic tone	Rank-frequency distribution of the 12 chromatic tones
Duration	Rank-frequency distribution of note durations (absolute duration in seconds)
Pitch duration	Rank-frequency distribution of pitch durations
Chromatic-tone duration	Rank-frequency distribution of chromatic tone durations
Pitch distance	Rank-frequency distribution of length of time intervals between note (pitch) repetitions
Chromatic-tone distance	Rank-frequency distribution of length of time intervals between note (chromatic tone) repetitions
Harmonic interval	Rank-frequency distribution of harmonic intervals within chord
Harmonic consonance	Rank-frequency distribution of harmonic intervals within chord based on music-theoretic consonance
Melodic interval	Rank-frequency distribution of melodic intervals within voice
Harmonic-melodic interval	Rank-frequency distribution of harmonic and melodic intervals
Harmonic bigrams	Rank-frequency distribution of adjacent harmonic interval pairs
Melodic bigrams	Rank-frequency distribution of adjacent melodic interval pairs
Melodic trigrams	Rank-frequency distribution of adjacent melodic interval triplets
Higher-order intervals	Rank-frequency distribution of higher orders of melodic intervals; first-order metric captures change between melodic intervals; second-order metric captures change between first-order intervals, and so on up to sixth order

dom number generators: a *white-noise* ($1/f^0$) source, a pink-noise ($1/f$) source, and a *brown-noise* ($1/f^2$) source. They used independent random-number generators to control the duration (half, quarter, eighth) and pitch (various standard scales) of successive notes. Remarkably, the music obtained through the pink-noise generators was much more pleasing to most listeners. In particular, the white-noise generators produced music that was "too random," whereas the brown-noise generators produced music that was "too correlated." They noted, "Indeed the sophistication of this '$1/f$ music' (which was 'just right') 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" (1975, p. 318) .

John Elliot and Eric Atwell (2000) failed to find Zipf distributions in notes extracted from audio signals. However, they used a small corpus of music pieces and were looking only for ideal Zipf distributions. On the other hand, Kenneth Hsu and Andrew Hsu (1991) found $1/f$ distributions in frequency intervals of Bach and Mozart compositions. Finally, Damián Zanette found Zipf distributions in notes extracted from MIDI-encoded music. Moreover, he used these distributions to demonstrate that as music progresses, it creates a meaningful context similar to the one found in human languages (see http://xxx.arxiv.org/abs/cs.CL/0406015).

Zipf Metrics for Music

Currently, we have a set of 40 metrics based on Zipf's Law. They are separated into two categories: simple metrics and fractal metrics.

Simple Metrics

Simple metrics measure the proportion of a particular parameter, such as pitch, globally. Table 1 shows the complete set of simple metrics we currently employ (Manaris et al. 2002). Obviously, there are many other possibilities, including size of movements, volume, timbre, tempo, and dynamics.

For instance, the *harmonic consonance* metric operates on a histogram of harmonic intervals within each chord in a piece. It counts the number of occurrences of each interval modulo multiples of the octave, and it plots them against their consonance ranking. In essence, this metric measures the proportion of harmonic consonance, or statistical balance between consonance and dissonance in a piece. We use a traditional music-theoretic ranking of harmonic consonance: unison is rank 1, P5 is rank 2, P4 is rank 3, M3 is rank 4, M6 is rank 5, m3 is rank 6, m6 is rank 7, M2 is rank 8, m7 is rank 9, M7 is rank 10, m2 is rank 11, and the tritone is rank 12.

Simple Zipf metrics are useful feature extractors. However, they have an important limitation. They examine a music piece as a whole, ignoring potentially significant contextual details. For instance, the pitch distribution of Bach's *Air on the G String* has a slope of –1.078. Sorting this piece's notes in increasing order of pitch would produce an unpleasant musical artifact. This artifact exhibits the same pitch distribution as the original piece. Thus, simple metrics could be easily fooled in the context of, say computer-aided music composition, where such metrics could be used for fitness evaluation. However, in the context of analyzing culturally sanctioned music, this limitation is not significant. This is because culturally sanctioned music tends to be well-balanced at different levels of granularity. That is, the balance exhibited at the global level is usually similar to the balance exhibited at the local level, down to a small level of granularity, as will be explained shortly.

Fractal Metrics

Fractal metrics handle the potential limitation of simple metrics in the context of music composition. Each simple metric has a corresponding fractal metric (Manaris et al. 2003). Whereas a simple metric calculates the Zipf distribution of a particular attribute at a global level, the corresponding fractal metric calculates the self-similarity of this distribution. That is, the fractal metric captures how many subdivisions of the piece exhibit this distribution at many levels of granularity.

For instance, to calculate the fractal dimension of pitch distribution, we recursively apply the simple pitch metric to the piece's half subdivisions, quarter subdivisions, etc., down to the level of single measures. At each level of granularity, we count how many of the subdivisions approximate the global distribution. We then plot these counts against the length of the subdivision, producing a size-frequency distribution. The slope of the trend line is the fractal dimension, *D,* of pitch distribution for this piece. This allows us to identify anomalous pieces that, although balanced at the global level, may be quite unbalanced at a local level. This method is similar to the box-counting technique for calculating the fractal dimension of images (Schroeder 1991, p. 214).

Taylor et al. (1999) used the box-counting technique to authenticate and date paintings by Jackson Pollock. Using a size-frequency plot, they calculated the fractal dimension, *D,* of Pollock's paintings. In particular, they discovered two different slopes: one attributed to Pollock's dripping process, and the other attributed to his motions around the canvas. Also, they were able to track how Pollock refined his dripping technique: the slope decreased through the years, from approximately –1 in 1943 to –1.72 in 1952.

Experimental Studies: Zipf-Mandelbrot Distributions in MIDI-Encoded Music

Inspired by the work of Zipf (1949) and Voss and Clarke (1975, 1978), we conducted two studies to explore Zipf-Mandelbrot distributions in MIDI-encoded music. The first study used a 28-piece corpus from Bach, Beethoven, Chopin, Debussy, Handel, Mendelssohn, Schönberg, and Rodgers and Hart. It also included seven pieces from a white-noise generator as a control group (Manaris et al. 2002). The second study used a 196-piece corpus from various genres, including Baroque, Classical, Romantic, Modern, Jazz, Rock, Pop, and Punk Rock. It also included 24 control pieces from DNA strings, white noise, and pink noise (Manaris et al. 2003).

Methodology

Zipf (1949, pp. 336–337) worked with composition data, i.e., printed scores, whereas Voss and Clarke (1975, 1978) studied performance data, i.e., audio

Table 2. Average Results Across Metrics for Various Genres From a Corpus of 220 Pieces

Genre	*Slope*	R^2	*Slope Std. Dev.*	R^2 *Std. Dev.*
Baroque	–1.1784	0.8114	0.2688	0.0679
Classical	–1.2639	0.8357	0.1915	0.0526
Early Romantic	–1.3299	0.8215	0.2006	0.0551
Romantic	–1.2107	0.8168	0.2951	0.0609
Late Romantic	–1.1892	0.8443	0.2613	0.0667
Post Romantic	–1.2387	0.8295	0.1577	0.0550
Modern Romantic	–1.3528	0.8594	0.0818	0.0294
Twelve-Tone	–0.8193	0.7887	0.2461	0.0964
Jazz	–1.0510	0.7864	0.2119	0.0796
Rock	–1.2780	0.8168	0.2967	0.0844
Pop	–1.2689	0.8194	0.2441	0.0645
Punk Rock	–1.5288	0.8356	0.5719	0.0954
DNA	–0.7126	0.7158	0.2657	0.1617
Random (Pink)	–0.8714	0.8264	0.3077	0.0852
Random (White)	–0.4430	0.6297	0.2036	0.1184

recorded from radio stations. Our corpus consisted mostly of MIDI-encoded performances from the Classical Music Archives (available online at www.classicalarchives.com).

We identified a large number of parameters of music that could possibly exhibit Zipf-Mandelbrot distributions. These attributes included pitch, duration, melodic intervals, and harmonic intervals, among others. Table 1 shows a representative subset of these metrics.

Results

Most pieces in our corpora exhibited near-Zipfian distributions across a wide variety of metrics. In the first study, classical and jazz pieces averaged near-Zipfian distributions and strong linear relations across all metrics, whereas random pieces did not. Specifically, the across-metrics average slope for music pieces was –1.2653. The corresponding R^2 value, 0.8088, indicated a strong average linear relation. The corresponding results for control pieces were –0.4763 and 0.6345, respectively.

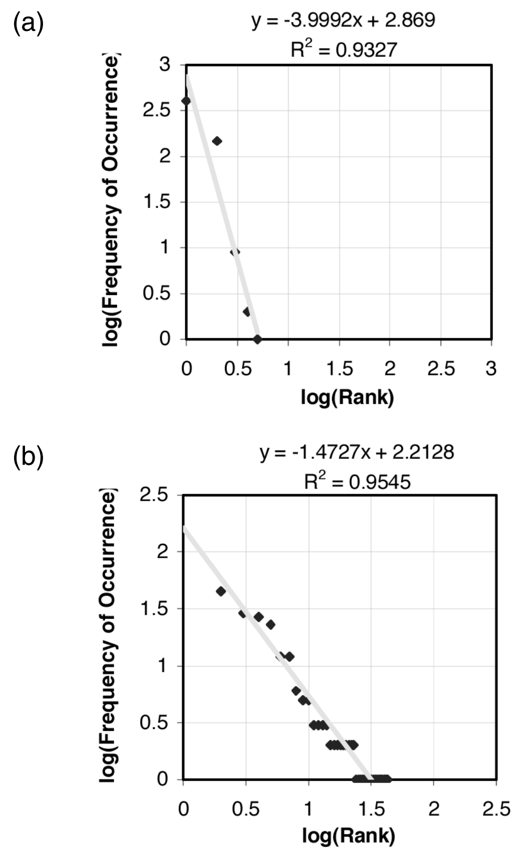
Table 2 shows average results from the second study. In particular, the 196 music pieces exhibited an overall average slope of –1.2023 with standard deviation of 0.2521. The average R^2 is 0.8233 with a standard deviation of 0.0673. The 24 pieces in the control group exhibited an average slope of –0.6757 with standard deviation 0.2590. The average R^2 is 0.7240 with a standard deviation of 0.1218. This suggests that some music styles could possibly be distinguished from other styles and from non-musical data through a collection of Zipf metrics.

Music as a Hierarchical Dynamic System

Mandelbrot observed that Zipf-Mandelbrot distributions in economic systems are "stable" in that, even when such systems are perturbed, their slopes tend to remain between 0 and –2. Systems with slopes less than –2, when perturbed, exhibit chaotic behavior (1977, p. 344). The same stability has also been observed in simulations of sand piles and various other natural phenomena. Phenomena exhibiting this tendency are called *self-organized criticalities* (Bak et al. 1987; Maslov et al. 1999). This tendency characterizes a complex system that has come to rest. Because the system has lost energy, it is bound to stay in this restful state, hence the "stability" of these states.

Mandelbrot states that, because these stable distributions are very widespread, they are noticed and published, whereas chaotic distributions tend not to be noticed (1977, p. 344). Accordingly, in physics, all

Figure 2. (a) Rank-frequency distribution of note durations from the score of Bach's Two-Part Invention No. 13 in A minor, *BWV 784; (b) the more "natural" distribution from the interpretation of this piece as performed by harpsichordist John Sankey.*



distributions with slope less than –2 are collectively called *black noise,* as opposed to brown noise (slope of –2), pink noise (slope of –1, i.e., Zipf's ideal), and white noise (slope of 0). (See Schroeder 1991, p. 122.)

The tendency of music to exhibit rank-frequency distribution slopes between 0 and –2, as observed in our experiments with hundreds of MIDI-encoded music pieces, suggests that perhaps composing music could be viewed as a process of stabilizing a hierarchical system of pitches, durations, intervals, measures, movements, etc. In this view, a completed piece of music resembles a dynamic system that has come to rest.

For a piece of music to resemble black noise, it must be rather monotonous. In the extreme case, this corresponds to a slope of negative infinity (–∞), i.e., a vertical line. Other than the obvious "minimalist" exceptions, such as John Cage's *4′33″,* most performed music tends to have some variability across different parameters such as pitch, duration, melodic intervals, etc. Figure 2a 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.

Music Classification and Zipf's Law

There are numerous studies on music classification, such as Aucouturier and Pachet (2003), Pampalk et al. (2004), and Tzanetakis et al. (2001). However, we have found no references to Zipf's Law in this context.

Zipf's Law has been used successfully for classification in other domains. For instance, as mentioned earlier, it has been used to authenticate and date paintings by Jackson Pollock (Taylor et al. 1999). It has also been used to differentiate among immune systems of normal, irradiated chimeric, and athymic mice (Burgos and Moreno-Tovar 1996). Zipf's Law has been used to distinguish healthy from non-healthy heartbeats in humans (see arxiv.org/abs/physics/0110075). Finally, it has been used to distinguish cancerous human tissue from normal tissue using microarray gene data (Li and Yang 2002).

Experimental Studies

We performed several studies to explore the applicability of Zipf's Law to music classification. The studies reported in this section focused on author attribution and style identification.

Author Attribution

In terms of author attribution, we conducted five experiments: Bach vs. Beethoven, Chopin vs. De-

Table 3. Author Attribution Experiment with Five Composers from Various Genres

SCARLATTI VS. PURCELL VS. BACH VS. CHOPIN VS. DEBUSSY							
				Test Set		*MSE*	
Train Patterns (%)	*Test Patterns (%)*	*Architecture*	*Cycles*	*Errors*	*Success Rate (%)*	*Train*	*Test*
652 (86%)	106 (14%)	81-6-5	10000	6	94.4	0.00005	0.07000
			4000	6	94.4	0.00325	0.10905
		81-12-5	10000	6	94.4	0.00313	0.11006
			4000	5	95.3	0.00321	0.10201
541 (71%)	217 (29%)	81-6-5	10000	11	95	0.00386	0.09076
			4000	11	95	0.00199	0.10651
		81-12-5	10000	14	93.6	0.00194	0.14195
			4000	11	95	0.00388	0.09459

MSE = Mean-Square Error.

bussy, Bach vs. four other composers, and Scarlatti vs. Purcell vs. Bach vs. Chopin vs. Debussy (Machado et al. 2003, 2004).

We compiled several corpora whose size ranged across experiments from 132 to 758 music pieces. Our data consisted of MIDI-encoded performances, the majority of which came from the online Classical Music Archives. We applied Zipf metrics to extract various features for each piece. The number of features per piece varied across experiments, ranging from 30 to 81. This collection of feature vectors was used to train an artificial neural network. Our training methodology is similar to one used by Miranda et al. (2003); in particular, we separated feature vectors into two data sets. The first set was used for training, and the second set was used to test the ANN's ability to classify new data. We experimented with various architectures and training regimens using the Stuttgart Neural Network Simulator (see www-ra.informatik.uni-tuebingen.de/SNNS).

Table 3 summarizes the ANN architectures used and results obtained from the Scarlatti vs. Purcell vs. Bach vs. Chopin vs. Debussy experiment. The success rate across the five-author attribution experiments ranged from 93.6 to 95 percent. This suggests that Zipf metrics are useful for author attribution (Machado et al. 2003, 2004).

The analysis of the errors made by the ANN indicates that Bach was the most recognizable composer. The most challenging composer to recognize was Debussy. His works were often misclassified as scores of Chopin.

Style Identification

We have also performed statistical analyses of the data summarized in Table 2 to explore the potential for style identification (Manaris et al. 2003). We have discovered several interesting patterns.

For instance, our corpus included 15 pieces by Schönberg, Berg, and Webern written in the twelve-tone style. They exhibit an average chromatic-tone slope of –0.3168 with a standard deviation of 0.1801. The corresponding average for classical pieces was –1.0576 with a standard deviation of 0.5009, whereas for white-noise pieces it was 0.0949 and 0.0161, respectively. Clearly, by definition, the chromatic-tone metric alone is sufficient for identifying twelve-tone music. Also, DNA and white noise were easily identifiable through pitch distribution alone. Finally, all genres commonly referred to as classical music exhibited significant overlap in all of the metrics; this included Baroque, Classical, and Romantic pieces. This is consistent with average human competence in discriminating between these musical styles.

Subsequent analyses of variance (ANOVA) revealed significant differences among some genres. For instance, twelve-tone music and DNA were identifiable through harmonic-interval distribution alone. Similarly to author attribution, we expect that a combination of metrics will be sufficient for style identification. To validate this hypothesis, we are currently conducting a large ANN-based style-identification study.

Aesthetics and Zipf's Law

Arnheim (1971) proposes that art is our answer to entropy and the Second Law of Thermodynamics. As entropy increases, so do disorganization, randomness, and chaos. In Arnheim's view, artists subconsciously tend to produce art that creates a balance between chaos and monotony. According to Schroeder, this agrees with George Birkhoff's Theory of Aesthetic Value:

> [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 $1/f^2$, nor like an unpredictable white noise, with a frequency distribution of $1/f^0$. (Schroeder 1991, p. 109)

As mentioned earlier, in the case of music, Voss and Clarke (1975, 1978) have shown that classical, rock, jazz, and blues music exhibits $1/f$ power spectra. Also, in our study, 196 pieces from various genres exhibited an average Zipf-Mandelbrot distribution of approximately $1/f^{1.2}$ across various music attributes (Manaris et al. 2003).

In the visual domain, Spehar et al. (2003) have shown that humans show an aesthetic preference for images exhibiting a Zipf-Mandelbrot distribution between $1/f^{1.3}$ and $1/f^{1.5}$. Finally, Mario Livio (2002, pp. 219–220) has demonstrated a connection between a Zipf-Mandelbrot distribution of $1/f^{1.4}$ and the golden ratio (0.61803 . . .).

Zipf's Law and Human Physiology

Boethius believed that musical consonance "pleases the listener because the body is subject to the same laws that govern music, and these same proportions are to be found in the cosmos itself. Microcosm and macrocosm are tied by the same knot, simultaneously mathematical and aesthetic" (Eco 1986, p. 31).

One connection between near-$1/f$ distributions in music and human perception is the physiology of the human ear. The basilar membrane in the inner ear analyzes acoustic frequencies and, through the acoustic nerve, reports sounds to the brain. Interestingly, $1/f$ sounds stimulate this membrane in just the right way to produce a constant-density stimulation of the acoustic nerve endings (Schroeder 1991, p. 122). This corroborates Voss and Clarke's finding that $1/f$ music sounds "just right" to human subjects, as opposed to $1/f^0$ music, which sounds "too random," and $1/f^2$ music, which sounds "too monotonous" (Voss and Clarke 1978).

Functional magnetic resonance imaging (fMRI) and other measurements are providing additional evidence of $1/f$ activity in the human brain (Zhang and Sejnowski 2000; see also arxiv.org/PS_cache/cond-mat/pdf/0208/0208415.pdf). According to Carl Anderson, to perceive the world and generate adaptive behaviors, the brain self-organizes via spontaneous $1/f$ clusters or bursts of activity at various levels. These levels include protein chain fluctuations, ion channel currents, synaptic processes, and behaviors of neural ensembles. In particular, "[e]mpirical fMRI observations further support the association of fractal fluctuations in the temporal lobes, brainstem, and cerebellum during the expression of emotional memory, spontaneous fluctuations of thought and meditative practice" (Anderson 2000, p. 193).

This supports Zipf's proposition that composers may subconsciously incorporate $1/f$ distributions into their compositions because they sound right to them and because their audiences like them (1949, p. 337). If this is the case, then in certain styles such as twelve-tone and aleatoric music, composers may subconsciously avoid such distributions for artistic reasons.

Experimental Study: "Pleasantness" Prediction

We conducted an ANN experiment to explore the possible connection between aesthetics and Zipf-

Table 4. Twelve Pieces Used for Music Pleasantness Classification Study

Composer	Piece	Duration	Human Rating Average	Human Rating Std. Dev.
Beethoven	*Sonata No. 20 in G, Opus 49. No. 2*	1'00"	72.84	11.83
Debussy	*Arabesque No. 1 in E (Deux Arabesques)*	1'34"	78.30	17.96
Mozart	*Clarinet Concerto in A,* K.622 (first movement)	1'30"	67.97	12.43
Schubert	*Fantasia in C minor, Op. 15*	1'58"	68.17	13.67
Tchaikovsky	*Symphony 6 in B minor, Op. 36,* second movement	1'23"	68.59	13.52
Vivaldi	*Double Violin Concerto in A minor, F. 1, No. 177*	1'46"	63.12	15.93
Bartók	*Suite, Op. 14*	1'09"	42.46	14.58
Berg	*Wozzeck* (transcribed for piano)	1'38"	35.75	15.79
Messiaen	*Apparation de l'Eglise Eternelle*	1'19"	39.75	17.12
Schönberg	*Pierrot Lunaire* (fifth movement)	1'13"	44.00	15.85
Stravinksy	*Rite of Spring,* second movement (transcribed for piano)	1'09"	43.19	15.58
Webern	*Five Songs* (1. "Dies ist ein Lied")	1'26"	39.74	13.04

The first six pieces were rated by subjects as "pleasant" overall; the last six pieces were rated as "unpleasant" overall. (Neutral is 50.)

Mandelbrot distributions at the level of MIDI-encoded music. In this study, we trained an ANN using Zipf-Mandelbrot distributions extracted from a set of pieces, together with human emotional responses to these pieces. Our hypothesis was that the ANN would discover correlations between Zipf-Mandelbrot distributions and human emotional responses and thus be able to predict the "pleasantness" of music on which it had not been trained.

Methodology

We used a corpus of twelve excerpts of music. These were MIDI-encoded performances selected by a member of our team with an extensive music theory background. Our goal was to identify six pieces that an average person might find pleasant and six pieces that an average person might find unpleasant. All excerpts were less than two minutes long to minimize fatigue for the human subjects. Table 4 shows the composer, name, and duration of each excerpt.

We collected emotional responses from 21 subjects for each of the twelve excerpts. These subjects were college students with varied musical backgrounds. The experiment was double-blind in that neither the subjects nor the people conducting the experiment knew which of the pieces were presumed as pleasant or unpleasant.

Subjects were instructed to report their own emotional responses during the music by using the mouse to position an "X" cursor within a two-dimensional space on a computer monitor. The horizontal dimension represented "pleasantness," and the vertical dimension represented "activation" or arousal. The system recorded the subject's cursor coordinates once per second. Positions were recorded on scales of 0–100 with the point (50, 50) representing emotional indifference or neutral reaction. Table 4 shows the average pleasantness rating and standard deviation.

Much psychological evidence indicates that "pleasantness" and "activation" are the fundamental dimensions needed to describe human emotional responses (Barrett and Russell 1999). Following established standards, the emotion labels "excited," "happy," "serene," "calm," "lethargic," "sad," "stressed," and "tense" were placed in a circle around the space to assist the subjects in the task. These labels, in effect, helped the subjects discern the semantics of the selection space. Similar methods for continuous recording of emotional response to music have been used elsewhere (Schubert 2001). It is worth emphasizing that the subjects were not

Table 5. Summary of Results from Twelve-Fold, Cross-Validation ANN Experiment, Listed by Composer of Test Piece

			MSE	
Composer	*Success Rate (%)*	*Cycles*	*Train*	*Test*
Beethoven	100.00	32200	0.008187	0.003962
Debussy	100.00	151000	0.001807	0.086451
Mozart	100.00	222200	0.004430	0.003752
Schubert	100.00	592400	0.001982	0.004851
Tchaikovsky	100.00	121400	0.004268	0.004511
Vivaldi	100.00	431600	0.003870	0.009643
Bartók	100.00	569200	0.001700	0.008536
Berg	80.95	4600	0.015412	0.100619
Messiaen	100.00	35200	0.008392	0.001315
Schönberg	100.00	8000	0.016806	0.015803
Stravinksy	100.00	311200	0.004099	0.002693
Webern	100.00	468600	0.002638	0.013540
Average	98.41	245633	0.006133	0.021306
Std. Dev.	5.27	212697	0.004939	0.032701

reporting what they liked or even what they judged as beautiful. We are not aware of any studies relating how one's musical preferences or formal training might affect one's reporting of pleasantness. However, there is evidence that pleasantness and liking are not the same (Schubert 1996). Also, it has been shown that pleasantness represents a more useful predictor of emotions than liking when using the above selection space in the music domain (Ritossa and Rickard 2004).

For the ANN experiment, we divided each music excerpt into segments. All segments started at 0:00 and extended in increments of four seconds. That is, the first segment extended from 0:00 to 0:04, the second segment from 0:00 to 0:08, the third segment from 0:00 to 0:12, and so on. We applied Zipf metrics to extract 81 features per music increment. Each feature vector was associated with a desired output vector of (1, 0) indicating pleasant and (0, 1) indicating unpleasant. This generated a total of 210 training vectors.

We conducted a twelve-fold, "leave-one-out," cross-validation study. This allowed for twelve possible combinations of eleven pieces to be learned and one piece to be tested. We experimented with various ANN architectures. The best one was a feed-forward ANN with 81 elements in the input layer, 18 in the hidden layer, and two in the output layer. Internally, the ANN was divided into two $81 \times 9 \times 1$ "Siamese-twin" pyramids, both sharing the same input layer. One pyramid was trained to recognize pleasant music, the other unpleasant. Classification was based on the average of the two outputs.

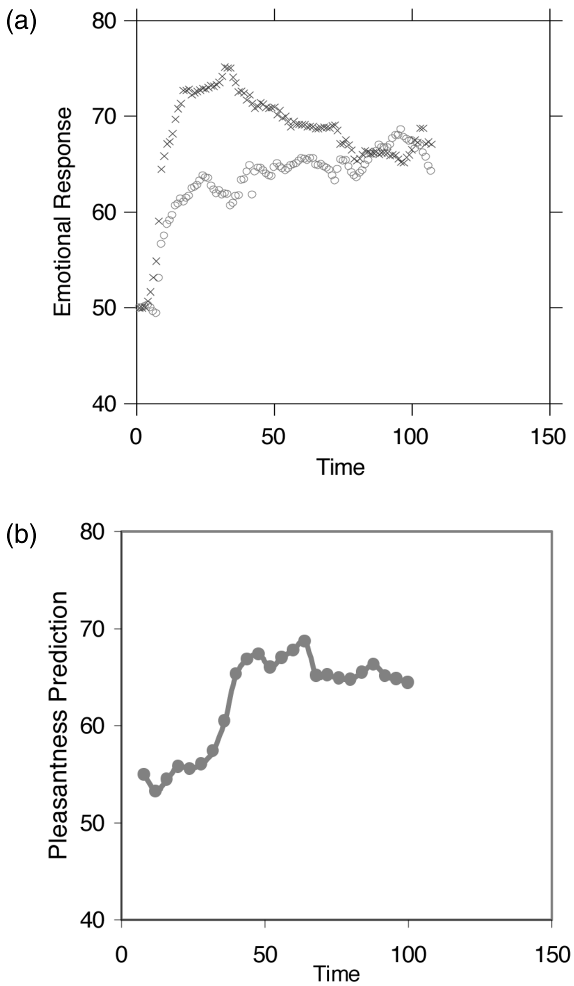
Results

Table 5 shows the results from all 12 experiments. The ANN performed extremely well with an average success rate of 98.41 percent. All pieces were classified with 100 percent accuracy, with one exception: Berg's piece was classified with only 80.95 percent accuracy. The ANN was considered successful if it rated a music excerpt within one standard deviation of the average human rating; this covers 68 percent of the human responses.

There are two possibilities for this "failure" of the ANN. Either our metrics fail to capture essential aspects of Berg's piece, or the other eleven pieces do not contain sufficient information to enable the interpretation of Berg's piece.

Figure 3a displays the average human ratings for Vivaldi's *Double Violin Concerto in A minor*, F. 1., No. 177. Figure 3b shows the pleasantness ratings predicted by the ANN for the same piece. The

Figure 3. *(a) Average pleasantness (o) and activation (x) ratings from 21 human subjects for the first 1 min, 46 sec of Vivaldi's* Double Violin Concerto in A minor, F. 1, No. 177. *A rating of 50 denotes a neutral response. (b) Pleasantness classification by ANN of the same piece having been trained on the other 11 pieces.*



ANN prediction approximates the average human response.

Relevance of Metrics

The analysis of ANN weights associated with each metric gives an indication of its relevance for a particular task. A large median value suggests that, for at least half of the ANNs in the experiment, the metric was useful in performing the particular task. There were 13 metrics that had median ANN weights of at least 7. Table 6 lists these metrics in descending order with respect to the median. It also lists the mean ANN weights, standard deviations, and the ratio of standard deviation and mean.

Among the metrics with the highest medians, two of them stand out: *harmonic consonance* and *chromatic tone.* This is because they have a high mean and relatively small standard deviation, as indicated by the last column of Table 6. It can be argued that these metrics were most consistently relevant for "pleasantness" prediction across all twelve experiments.

As mentioned earlier, *harmonic consonance* captures the statistical proportion of consonance and dissonance in a piece. "Pleasant" pieces in our corpus exhibited similarities in their proportions of harmonic consonance: the slope ranged from –0.8609 (Schubert, 0:08 sec) to –1.8087 (Beethoven, 0:40) with an average of –1.2225 and standard deviation of 0.1802. "Unpleasant" pieces in our corpus also exhibited similarities in their proportions of harmonic consonance; in this case, however, the slope ranged from –0.2284 (Schönberg, 0:24) to –0.9919 (Berg, 0:20) with an average of –0.5343 and standard deviation of 0.1519. Owing to the overlap between the two ranges, the ANN had to rely on additional metrics for disambiguation.

Chromatic tone captures the uniform distribution of pitch, which is characteristic of twelve-tone and aleatoric music. Such music was rated consistently by our subjects as rather "unpleasant." The chromatic tone slope for "unpleasant" pieces ranged from –0.0578 (Webern, 0:48) to –1.4482 (Stravinsky, 0:32), with an average of –0.6307 and standard deviation of 0.3985. On the other hand, the chromatic tone slope for "pleasant" pieces ranged from –0.4491 (Debussy, 0:16) to –1.8848 (Mozart, 0:68), with an average of –1.3844 and standard deviation of 0.3075. The chromatic tone metric was less relevant for classification than harmonic consonance owing to the greater overlap in the ranges of slopes between "pleasant" and "unpleasant" pieces. Other relevant metrics include chromatic-tone distance, pitch duration, harmonic interval, harmonic and melodic interval, harmonic bigrams, and melodic bigrams.

Discussion

These results indicate that, in most cases, the ANN is identifying patterns that are relevant to human aesthetic judgments. This supports the hypothesis

Table 6. Statistical Analysis of ANN Weights for Metrics Used in the "Pleasantness" Prediction ANN Experiment (Ordered by Median)

Metric	*Median*	*Mean*	*Std. Dev.*	*Std. Dev./Mean*
Harmonic-melodic interval (simple slope)	57.43	64.22	45.09	0.70
Harmonic consonance (simple slope)	44.54	44.48	17.13	0.39
Harmonic bigram (simple slope)	37.76	41.37	31.88	0.77
Pitch duration (simple slope)	32.34	32.53	20.01	0.62
Harmonic interval (simple R^2)	23.54	23.25	15.18	0.65
Chromatic-tone distance (simple slope)	21.82	27.69	16.00	0.58
Chromatic tone (simple slope)	19.93	21.83	7.75	0.36
Melodic bigrams (simple slope)	15.74	20.83	17.82	0.86
Duration (simple R^2)	9.38	10.23	7.60	0.74
Harmonic interval (simple slope)	8.39	8.21	5.90	0.72
Fourth high order (fractal slope)	8.16	8.26	4.86	0.59
Melodic interval (fractal slope)	7.81	8.37	3.42	0.41
Harmonic bigram (simple R^2)	7.46	10.28	11.64	1.13

that there may be a connection between aesthetics and Zipf-Mandelbrot distributions at the level of MIDI-encoded music.

It was interesting to note that harmonic consonance approximated a $1/f$ distribution for pieces that were rated as pleasant and a more chaotic $1/f^{0.5}$ distribution for pieces that were rated as unpleasant. Because the emotional responses used in this study were actually psychological self-report measures, this suggests the influence of a higher level of organization. Also, because an emotional measure is involved, this likely reflects some higher-level pattern of intellectual processing that exhibits $1/f$ organization. This processing likely draws upon other, non-auditory information in the brain.

Conclusions

We propose the use of Zipf-based metrics as a basis for author- and style-identification tasks and for the assessment of aesthetic properties of music pieces. The experimental results in author-identification tasks, where an average success rate of more than 94 percent was attained, show that the used set of metrics, and accordingly Zipf's Law, capture meaningful information about the music pieces. Clearly, the success of this approach does not imply that other metrics or approaches are irrelevant.

As noticed by several researchers, culturally sanctioned music tends to exhibit near-ideal Zipf distributions across various parameters. This suggests the possibility that combinations of Zipf-based metrics may represent certain necessary but not sufficient conditions for aesthetically pleasing music. This is supported by our pleasantness study where an ANN succeeds in predicting human aesthetic judgments of unknown pieces with more than 98 percent accuracy.

The set of 40 metrics used in these studies represent only a small subset of possible metrics. The analysis of ANN weights indicates that harmonic consonance and chromatic tone were related to human aesthetic judgments. Based on this analysis and on additional testing, we are trying to determine the most useful metrics overall and to develop additional ones.

It should be emphasized that the metrics proposed in this article offer a particular description of the musical pieces, where traditional musical structures such as motives, tonal structures, etc., are not measured explicitly. Statistical measurements, such as Zipf's Law, tend to focus on general trends and thus can miss significant details. To further explore the capabilities and limitations of our approach, we are developing an evolutionary music generation system in which the proposed classification methodology will be used for fitness assignment.

Once developed, this system will be included in a hybrid society populated by artificial and human agents, allowing us to perform further testing in a dynamic environment.

In closing, our studies show that Zipf's Law, as encapsulated in our metrics, can be used effectively in music classification tasks and aesthetic evaluation. This may have significant implications for music information retrieval and computer-aided music analysis and composition, and may provide insights on the connection among music, nature, and human physiology. We regard these results as preliminary; we hope they will encourage further investigation of Zipf's Law and its potential applications to music classification and aesthetics.

Acknowledgments

This project has been partially supported by an internal grant from the College of Charleston and a donation from the Classical Music Archives. We thank Renée McCauley, Ramona Behravan, and Clay McCauley for their comments. William Daugherty and Marisa Santos helped conduct the ANN experiments. Brian Muller, Christopher Wagner, Dallas Vaughan, Tarsem Purewal, Charles McCormick, and Valerie Sessions helped formulate and implement Zipf metrics. Giovanni Garofalo helped collect human emotional response data for the ANN pleasantness experiment. William Edwards, Jr., Jimmy Wilkinson, and Kenneth Knuth provided early material and inspiration.

References

Adamic, L. A., and B. A. Huberman. 2000. "The Nature of Markets in the World Wide Web." *Quarterly Journal of Electronic Commerce* 1(1):5–12.

Anderson, C. M. 2000. "From Molecules to Mindfulness: How Vertically Convergent Fractal Time Fluctuations Unify Cognition and Emotion." *Consciousness & Emotion* 1:2:193–226.

Arnheim, R. 1971. *Entropy and Art: An Essay on Disorder and Order.* Berkeley: University of California Press.

Aucouturier, J.-J., and F. Pachet. 2003. "Representing Musical Genre: A State of the Art." *Journal of New Music Research* 32(1):83–93.

Bak, P. 1996. *How Nature Works: The Science of Self-Organized Criticality.* New York: Springer-Verlag.

Bak, P., C. Tang, and K. Wiesenfeld. 1987. "Self-Organized Criticality: An Explanation for 1/*f* Noise." *Physical Review Letters* 59:381–384.

Barrett, L. F., and J. A. Russell. 1999. "The Structure of Current Affect: Controversies and Emerging Consensus." *Current Directions in Psychological Science* 8(1):10–14.

Burgos, J. D., and P. Moreno-Tovar. 1996 "Zipf-Scaling Behavior in the Immune System." *Biosystems* 39(3): 227–232.

Eco, U. 1986. *Art and Beauty in the Middle Ages.* H. Bredin, trans. New Haven: Yale University Press.

Elliot, J., and E. Atwell. 2000. "Is Anybody Out There? The Detection of Intelligent and Generic Language-Like Features." *Journal of the British Interplanetary Society* 53(1/2):13–22.

Ferrer Cancho, R., and R. V. Solé. 2003. "Least Effort and the Origins of Scaling in Human Language." *Proceedings of the National Academy of Sciences, U.S.A* 100(3):788–791.

Hsu, K. J., and A. Hsu. 1991. "Self-Similarity of the '1/*f* Noise' Called Music." *Proceedings of the National Academy of Sciences, U.S.A.* 88(8):3507–3509.

Li, W. 1992. "Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution." *IEEE Transactions on Information Theory* 38(6):1842–1845.

Li, W. 1998. "Letter to the Editor." *Complexity* 3(5):9–10.

Li, W., and Y. Yang. 2002. "Zipf's Law in Importance of Genes for Cancer Classification using Microarray Data." *Journal of Theoretical Biology* 219:539–551.

Livio, M. 2002. *The Golden Ratio.* New York: Broadway Books.

Machado, P., et al. 2003. "Power to the Critics—A Framework for the Development of Artificial Critics." *Proceedings of 3rd Workshop on Creative Systems, 18th International Joint Conference on Artificial Intelligence (IJCAI 2003).* Coimbra, Portugal: Center for Informatics and Systems, University of Coimbra, pp. 55–64.

Machado, P., et al. 2004. "Adaptive Critics for Evolutionary Artists." *Proceedings of EvoMUSART2004—2nd European Workshop on Evolutionary Music and Art.* Berlin: Springer-Verlag, pp. 437–446.

Manaris, B., T. Purewal, and C. McCormick. 2002. "Progress Towards Recognizing and Classifying Beautiful Music with Computers: MIDI-Encoded Music and the

Zipf-Mandelbrot Law." *Proceedings of the IEEE SoutheastCon 2002.* New York: Institute of Electrical and Electronics Engineers, pp. 52–57.
Manaris, B., et al. 2003. "Evolutionary Music and the Zipf-Mandelbrot Law: Progress towards Developing Fitness Functions for Pleasant Music." *Proceedings of EvoMUSART2003—1st European Workshop on Evolutionary Music and Art.* Berlin: Springer-Verlag, pp. 522–534.
Mandelbrot, B. B. 1977. *The Fractal Geometry of Nature.* New York: W. H. Freeman.
Maslov, S., C. Tang, and Y.-C. Zhang. 1999. "1/*f* Noise in Bak-Tang-Wiesenfeld Models on Narrow Stripes." *Physical Review Letters* 83(12):2449–2452.
May, M. 1996. "Did Mozart Use the Golden Section?" *American Scientist* 84(2):118.
Meyer, L. B. 2001. "Music and Emotion: Distinctions and Uncertainties." In P. N. Juslin and J. A. Sloboda, eds. *Music and Emotion—Theory and Research.* Oxford: Oxford University Press: 341–360.
Miranda, E. R. 2001. *Composing Music with Computers.* Oxford: Focal Press.
Miranda, E. R., et al. 2003. "On Harnessing the Electroencephalogram for the Musical Braincap." *Computer Music Journal* 27(2):80–102.
Nettheim, N. 1997. "A Bibliography of Statistical Applications in Musicology." *Musicology Australia* 20:94–106.
Pampalk E., S. Dixon, and G. Widmer. 2004. "Exploring Music Collections by Browsing Different Views." *Computer Music Journal* 28(2):49–62.
Ritossa, D. A., and N. S. Rickard. 2004. "The Relative Utility of 'Pleasantness' and 'Liking' Dimensions in Predicting the Emotions Expressed in Music." *Psychology of Music* 32(1):5–22.
Salingaros, N. A., and B. J. West. 1999. "A Universal Rule for the Distribution of Sizes." *Environment and Planning B: Planning and Design* 26:909–923.
Schroeder, M. 1991. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise.* New York: W. H. Freeman.
Schubert, E. 1996. "Enjoyment of Negative Emotions in Music: An Associative Network Explanation." *Psychology of Music* 24(1):18–28.
Schubert, E. 2001. "Continuous Measurement of Self-Report Emotional Response to Music." In P. N. Juslin and J. A. Sloboda, eds. *Music and Emotion—Theory and Research.* Oxford: Oxford University Press, pp. 393–414.
Spehar, B., et al. 2003. "Universal Aesthetic of Fractals." *Computers and Graphics* 27:813–820.
Taylor, R. P., A. P. Micolich, and D. Jonas. 1999. "Fractal Analysis Of Pollock's Drip Paintings." *Nature* 399:422.
Tzanetakis, G., G. Essl, and P. Cook. 2001. "Automatic Musical Genre Classification of Audio Signals." *Proceedings of 2nd Annual International Symposium on Music Information Retrieval.* Bloomington: University of Indiana Press, pp. 205–210.
Voss, R. F., and J. Clarke. 1975. "1/*f* Noise in Music and Speech." *Nature* 258:317–318.
Voss, R. F., and J. Clarke. 1978. "1/*f* Noise in Music: Music from 1/*f* Noise." *Journal of the Acoustical Society of America* 63(1):258–263.
Wolfram, S. 2002. *A New Kind of Science.* Champaign, Illinois: Wolfram Media.
Zhang, K., and T. J. Sejnowski. 2000. "A Universal Scaling Law between Gray Matter and White Matter of Cerebral Cortex." *Proceedings of the National Academy of Sciences, U.S.A* 97(10):5621–5626.
Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort.* New York: Addison-Wesley.