ARMONIQUE: EXPERIMENTS IN CONTENT-BASED SIMILARITY RETRIEVAL USING POWER-LAW MELODIC AND TIMBRE METRICS

Bill Manaris [1], **Dwight Krehbiel** [2], **Patrick Roos** [1], **Thomas Zalonis** [1]

[1] Computer Science Department, College of Charleston, 66 George Street, Charleston, SC 29424, USA
[2] Psychology Department, Bethel College, 300 E. 27th Street, North Newton, KS 67117, USA

ABSTRACT

This paper presents results from an on-going MIR study utilizing hundreds of melodic and timbre features based on power laws for content-based similarity retrieval. These metrics are incorporated into a music search engine prototype, called Armonique. This prototype is used with a corpus of 9153 songs encoded in both MIDI and MP3 to identify pieces similar to and dissimilar from selected songs. The MIDI format is used to extract various power-law features measuring proportions of music-theoretic and other attributes, such as pitch, duration, melodic intervals, and chords. The MP3 format is used to extract power-law features measuring proportions within FFT power spectra related to timbre. Several assessment experiments have been conducted to evaluate the effectiveness of the similarity model. The results suggest that power-law metrics are very promising for content-based music querying and retrieval, as they seem to correlate with aspects of human emotion and aesthetics.

1. INTRODUCTION

We present results from an on-going project in music information retrieval, psychology of music, and computer science. Our research explores power-law metrics for music information retrieval.

Power laws are statistical models of proportions exhibited by various natural and artificial phenomena [13]. They are related to measures of self-similarity and fractal dimension, and as such they are increasingly being used for data mining applications involving real data sets, such as web traffic, economic data, and images [5]. Power laws have been connected with emotion and aesthetics through various studies and experiments [8, 9, 12, 14, 16, 18].

We discuss a music search engine prototype, called *Armonique*, which utilizes power-law metrics to capture both melodic and timbre features of music. In terms of input, the user selects a music piece. The engine searches for pieces similar to the input by comparing power-law proportions through the database of songs. We provide an on-line demo of the system involving a corpus of 9153 pieces for various genres, including baroque, classical, romantic, impressionist, modern, jazz, country, and rock among others. This corpus was originally encoded in

MIDI, which facilitated extraction of melodic features. It was then converted to MP3 for the purpose of extracting timbre features.

In terms of assessment, we conducted an experiment by measuring human emotional and physiological responses to the music chosen by the search engine. Analyses of data indicate that people do indeed respond differently to pieces identified by the search engine as similar to the participant-chosen piece, than to pieces identified by the engine as different. For similar pieces, the participants' emotion while listening to the music is more pleasant, their mood after listening is more pleasant; they report liking these pieces more, and they report them to be more similar to their own chosen piece. These results support the potential of using power-law metrics for music information retrieval.

Section 2 presents relevant background research. Sections 3 and 4 describe our power-law metrics for melodic and timbre features, respectively. Sections 5 and 6 discuss the music search engine prototype, and its evaluation with human subjects. Finally, section 7 presents closing remarks and directions for future research.

2. BACKGROUND

Tzanetakis et al. [15] performed genre classifications using audio signal features. They performed FFT analysis on the signal and calculated various dimensions based on the frequency magnitudes. Also, they extracted rhythm features through wavelet transforms. They reported classification success rates of 62% using six genres (classical, country, disco, hiphop, jazz and rock) and 76% using four classical genres.

Aucouturier and Packet [1] report the most typical audio similarity technique is timbre via spectral analysis using Mel-frequency cepstral coefficients (MFCCs). Their goal was to improve on the overall performance of timbre similarity by varying parameters associated with these techniques (e.g. sample rate, frame size, number of MFCCs used, etc.) They report that there is a "glass ceiling" for timbre similarity that prevents any major improvements in performance via this technique. Subsequent research on this seems to either be a verification of this ceiling or an attempt to pass it using additional similarity dimensions (e.g., [10]).

Lidy et al. [6] discuss a genre classification experiment utilizing both timbre and melodic features. They utilize the typical timbre measures obtained through spectral analysis. They also employ a music transcription system they developed to generate MIDI representations of audio music files. From these files, they extract 37 different features, involving attributes of note pitches, durations and non-diatonic notes. The combined timbre and melodic features are used to conduct genre classification experiments. They report classification accuracies with combined feature sets ranging from 76.8% to 90.4% using standard benchmarks (e.g., ISMIR 2004 audio data).

Cano et al. [4] report on a music recommendation system called *MusicSurfer*. The primary dimensions of similarity used are timbre, tempo and rhythm patterns. Using a corpus of 273,751 songs from 11,257 artists their system achieves an artist identification rate of 24%. On the ISMIR 2004 author identification set they report a success rate of 60%, twice as high as that of the next best systems. MusicSurfer has a comprehensive user interface that allows users to find songs based on artist, genre, and similarity among other characteristics, and allows selection of different types of similarity. However, it does not include melodic features.

2.1. Power Laws and Music Analysis

Our music recommender system prototype employs both melodic and timbre features based on power-laws. Although power laws, fractal dimension, and other self-similarity features have been used extensively in information retrieval (e.g., [5]), to the best of our knowledge, there are no studies of content-based music recommendation systems utilizing power law similarity metrics.

A power law denotes a relationship between two variables where one is proportional to a power of the other. One of the most well-known power laws is *Zipf's law*:

$$P(f) \sim 1 / f^n \qquad (1)$$

where *P(f)* denotes the probability of an event of rank *f*, and *n* is close to 1. Zipf's law is named after George Kingsley Zipf, the Harvard linguist who documented and studied natural and social phenomena exhibiting such relationships [18]. The generalized form is:

$$P(f) \sim a / f^b \qquad (2)$$

where *a* and *b* are real constants. This generalized form is known as the Zipf-Mandelbrot law, after Benoit Mandelbrot.

Numerous empirical studies report that music exhibits power laws across timbre and melodic attributes (e.g., [8, 16, 18]). In particular, Voss and Clarke [16] demonstrate that music audio properties (e.g. loudness and pitch fluctuation) exhibit power law relationships. Using 12 hours worth of radio recordings, they show that power fluctuations in music follow a *1/f* distribution.

Manaris et al. [8] report various classification experiments with melodic features based on power laws. These studies include composer identification with 93.6% to 95% accuracy. They also report an experiment using emotional responses from humans. Using a corpus of 210 music excerpts in a 12-fold cross-validation study, artificial neural networks (ANNs) achieved an average success rate of 97.22% in predicting (within one standard deviation) human emotional responses to those pieces.

3. MELODIC FEATURE EXTRACTION

As mentioned earlier, we employ hundreds of power-law metrics that calculate statistical proportions of music-theoretic and other attributes of pieces.

3.1. Melodic Metrics

We have defined 14 power-law metrics related to proportion of pitch, chromatic tone, duration, distance between repeated notes, distance between repeated durations, melodic and harmonic intervals, melodic and harmonic consonance, melodic and harmonic bigrams, chords, and rests [9]. Each metric calculates the *rank-frequency* distribution of the attribute in question, and returns two values:

- the slope of the trendline, *b* (see equation 2), of the rank-frequency distribution; and
- the strength of the linear relation, r^2.

We also calculate *higher-order* power law metrics. For each regular metric we construct an arbitrary number of higher-order metrics (e.g., the difference of two pitches, the difference of two differences, and so on), an approach similar to the notion of derivative in mathematics.

Finally, we also capture the difference of an attribute value (e.g., note duration) from the local average. Local variability, *locVar[i]*, for the ith value is

$$locVar[i] = abs(vals[i] – avg(vals, i)) / avg(vals, i) \qquad (3)$$

where *vals* is the list of values, *abs* is the absolute value, and *avg(vals, i)* is the local average of the last, say, 5 values. We compute a local variability metric for each of the above metrics (i.e., 14 regular metrics *x* the number of higher-order metrics we decide to include).

Collectively, these metrics measure hierarchical aspects of music. Pieces without hierarchical structure (e.g., aleatory music) have significantly different measurements than pieces with hierarchical structure and long-term dependencies (e.g., fugues).

3.2. Evaluation

Evaluating content-based music features through classification tasks based on objective descriptors, such as artist or genre, is recognized as a simple alternative to listening tests for approximating the value of such features for similarity prediction [7, 11].

We have conducted several genre classification experiments using our set of power-law melodic metrics to extract features from music pieces. Our corpus

consisted of 1236 MIDI-encoded music pieces from the Classical Music Archives (www.classicalarchives.com). These pieces were subdivided into 9 different musical genres (listed here by timeline): Medieval (57 pieces), Renaissance (150 pieces), Baroque (160 pieces), Classical (153 pieces), Romantic (91 pieces), Modern (127 pieces), Jazz (118 pieces), Country (109 pieces), and Rock (271 pieces).

First, we carried out a 10-fold cross validation experiment training an ANN to classify the corpus into the 9 different genres. For this, we used the Multilayer Perceptron implementation of the Weka machine learning environment. Each piece was represented by a vector of 156 features computed through our melodic metrics. The ANN achieved a success rate of 71.52%.

Figure 1 shows the resulting confusion matrix. It is clear that most classification errors occurred between genres adjacent in timeline. For example, most Renaissance pieces misclassified (32/43) were either falsely assigned to the Medieval (7) or Baroque (25) period. Most misclassified Baroque pieces (40/64) were incorrectly classified as Renaissance (23) or Classical (17), and so on. This is not surprising, since there is considerable overlap in style between adjacent genres.

To verify this interpretation, we ran several binary classification experiments. In each experiment, we divided the corpus into two classes: *class 1* consisting of a particular genre (e.g., Baroque), and *class 2* consisting of all other genres non-adjacent in timeline (e.g., all other genres minus Renaissance and Classical). For Jazz, as well as Rock, the other class included all other genres.

Table 1 shows the classification accuracies for all of these experiments. Since the two output classes in each experiment were unbalanced, the ANN accuracy rates should be compared to a majority-class classifier.

4. TIMBRE FEATURE EXTRACTION

We calculate a *base audio metric* employing spectral analysis through FFT. This metric is then expanded through the use of higher-order calculations and variations of window size and sampling rate. Since we are interested in power-law distributions within the human hearing range, assuming CD-quality sampling rate (44,1KHz), we use window sizes up to 1-sec. Interestingly, given our technique, the upper frequencies in this range do not appear to be as important for calculating timbre similarity; the most important frequencies appear to be from 1kHz to 11kHz.

For each of these windows we compute the power spectrum per window and then average the results, across frequencies. We then extract various power-law proportions from this average power spectrum. Again, each power-law proportion is captured as a pair of slope and r^2 values.

```
a   b   c  d  e  f   g   h  i    <-- classified as
92  17  1  2  4  4  23   8  9  |  a = baroque
20 107  0  1  2  5   5   4  9  |  b = classical
 1   1 96  0  0  0   0   9  2  |  c = country
 0   3  1 96  0  2   1  13  2  |  d = jazz
 8   3  0  0 28  0  15   3  0  |  e = medieval
 9   9  1  0  0 78   2  10 18  |  f = modern
25   8  1  0  7  0 107   1  1  |  g = renaissance
 1   2  7 11  0  4   1 242  3  |  h = rock
 9  12  3  3  1 19   1   5 38  |  i = romantic
```

Figure 1. Confusion matrix from 9 genre multi-classification ANN experiment.

Class 1	Class 2	ANN Accuracy	Majority Classifier
Baroque	Non-Adjacent	91.29 %	82.85 %
Classical	Non-Adjacent	92.08 %	84.46 %
Rock	Non-Rock	92.88 %	78.07 %
Jazz	Non-Rock	96.66 %	90.45 %

Table 1. ANN success rates for binary classification experiments.

We have explored various ways to calculate *higher-order* quantities involving both signal amplitudes and power spectrum (i.e., frequency) magnitudes. Again, the idea of a higher-order is similar to the use of derivatives in mathematics where one measures the rate of change of a function at a given point. Through this approach, we have created various derivative metrics, involving raw signal amplitude change, frequency change within a window, and frequency change across windows.

To further explore the proportions present in the audio signal, we vary the window size and the sampling rate. This allows us to get measurements from multiple "views" or different levels of granularity of the signal. For each new combination of window size and sampling rate, we recompute the above metrics, thus getting another pair of slope and r^2 values. Overall, we have defined a total of 234 audio features.

4.1. Evaluation

To evaluate these timbre metrics, we conducted a classification experiment involving a corpus of 1128 MP3 files containing an equal number of classical and non-classical pieces.

We carried out a 10-fold cross-validation, binary ANN classification experiment using a total of 234 audio features. For comparison, each classification was repeated using randomly assigned classes. The ANN achieved a success rate of 95.92%. The control success rate was 47.61%. We are in the process of running additional classification experiments to further evaluate our timbre metrics (e.g., ISMIR 2004 audio data).

5.   A MUSIC SEARCH ENGINE PROTOTYPE

Musicologists consider melody and timbre to be independent/complementary aesthetic dimensions (among others, such as rhythm, tempo, and mode). We have developed a music search-engine prototype, called Armonique, which combines melodic and timbre metrics to calculate sets of similar songs to a song selected by the user. For comparison, we also generate a set of dissimilar songs. A demo of this prototype is available at http://www.armonique.org.[1]

The corpus used for this demo consists of 9153 songs from the Classical Music Archives (CMA), extended with pieces from other genres such as jazz, country, and rock. These pieces originated as MIDI and were converted for the purpose of timbre feature extraction (and playback) to the MP3 format. As far as the search engine is concerned, each music piece is represented as a vector of hundreds of power-law slope and r^2 values derived from our metrics. As input, the engine is presented with a single music piece. The engine searches the corpus for pieces similar to the input, by computing the mean squared error (MSE) of the vectors (relative to the input). The pieces with the lowest MSE are returned as best matches.

We have experimented with various selection algorithms. The selection algorithm used in this demo, first identifies 200 similar songs based on melody, using an MSE calculation across all melodic metrics. Then, from this set, it identifies the 10 most similar songs based on timbre, again, using an MSE calculation across all timbre metrics. Each song is associated with two gauges providing a rough estimate of the similarity across the two dimensions, namely melody and timbre. It is our intention to explore more similarity dimensions, such as rhythm, tempo, and mode (major, minor, etc.).

6.   EVALUATION EXPERIMENT

We conducted an experiment with human subjects to evaluate the effectiveness of the proposed similarity model. This experiment evaluated only the melodic metrics of Armonique, since the timbre metrics had not been fully incorporated at the time.[2]

6.1. Participants

Twenty-one undergraduate students from Bethel College participated in this study. The participants consisted of 11 males and 10 females, with a mean age of 19.52 years and a range of 5 years. They had participated in high school or college music ensembles for a mean of 3.52 years with a standard deviation of 2.21, and had received private music lessons for a mean of 6.07 years with a standard deviation of 4.37.

6.2. Design

A single repeated-measures variable was investigated. This variable consisted of the seven different excerpts, one of them participant-chosen (hereafter referred to as the original piece), three of them computer-selected to be similar to the participant-chosen piece, and three selected to be different (see next section). Of primary interest was the comparison of responses to the original with those to the three similar pieces, of the original to the three different pieces, and of the three similar to the three different pieces.
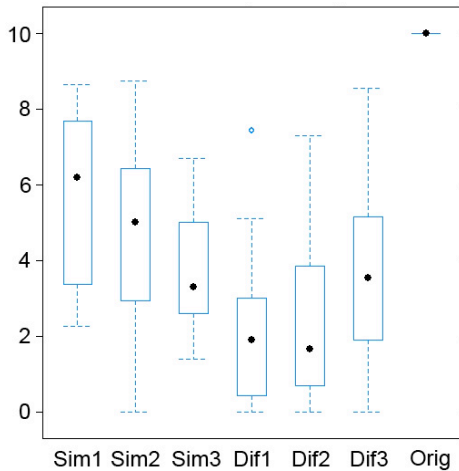
6.3. Data Set

At the time they were recruited, participants were asked to list in order of preference three classical music compositions that they enjoyed. The most preferred of these three compositions that was available in the CMA corpus was chosen for each participant. The search engine was then employed to select three similar and three different pieces. Similarities were based upon the first two minutes of each composition. This corpus is available at http://www.armonique.org/melodic-search.html .

6.4. Procedure

The resulting seven two-minute excerpts (original plus six computer-selected pieces), different for each participant, were employed in a listening session in which participant ratings of the music were obtained during and after the music, and psychophysiological responses were monitored (i.e., skin conductance, heart rate, corrugator supercilii electromygraphic recording, respiration rate, and 32 channels of electroencephalographic recording). The physiological data are not reported here. In some instances the entire piece was shorter than two minutes. A different random order of excerpts was employed for each participant.

Ratings of mood were obtained immediately prior to the session using the Self-Assessment Manikin [3] represented as two sliders with a 1-9 scale, one for pleasantness and one for activation, on the front panel of a LabVIEW (National Instruments, Austin, TX) virtual instrument. Physiological recording baseline periods of one minute were included prior to each excerpt. By means of a separate practice excerpt, participants were carefully instructed to report their feelings during the music by using a mouse to move a cursor on a two-dimensional emotion space [2]. This space was displayed on the front panel of the LabVIEW instrument, which also played the music file and recorded x-y coordinates of the cursor position once per second. After each excerpt participants again rated their mood with the Self-Assessment Manikin, and then rated their liking of the previous piece using another slider on the front panel with a 1-9 scale ranging from "Dislike very much" to "Like very much."

At the conclusion of the listening session, participants rated the similarity of each of the six computer-selected

---

[1] Due to copyright restrictions, some functionality is password-protected.
[2] We are planning a similar assessment experiment involving both melodic and timbre metrics.

Figure 2. Boxplots of similarity ratings across all subjects for the three similar songs recommended by the search engine and the three different songs.

excerpts to the original; these ratings were accomplished on a LabVIEW virtual instrument with six sliders and 0-10 scales ranging from "Not at all" to "Extremely."

6.5. Data Analysis

Data from the ratings during the music were averaged over time, yielding an average pleasantness and activation measure for each excerpt. A procedural error resulted in loss of data for two participants on these measures. Data from each behavioral measure were subjected to planned contrasts based upon a repeated-measures analysis of variance (SYSTAT software, SYSTAT, Inc., San Jose, CA). These contrasts compared the original song with each of the other two categories of music, and each of those categories with each other.

6.6. Results

The results for similarity ratings are shown in Figure 2. A contrast between the three similar and three different pieces indicated that the similar pieces were indeed judged to be more similar than were the different ones ($F(1, 20) = 20.98$, $p < 0.001$). Interestingly, the similar songs recommended by the search engine were ordered by humans the same way as by the search engine.[1]

In terms of the average ratings for pleasantness recorded during the music, the contrast between the similar pieces and the original was significant ($F(1, 18) = 5.85$, $p = 0.026$), while that between the different pieces and the original showed an even more marked difference ($F(1, 18) = 11.17$, $p = 0.004$). The contrast between similar and different pieces approached significance ($F(1,$

$18) = 3.04$, $p = 0.098$). No significant differences were found for these contrasts on the average activation measure.

In terms of the ratings for pleasantness recorded after the music, the contrast between the similar pieces and the original was not significant ($F(1, 20) = 1.21$, $p = 0.285$), while that between the different pieces and the original was significant ($F(1, 20) = 7.64$, $p = 0.012$). The contrast between similar and different pieces again approached significance ($F(1, 20) = 4.16$, $p = 0.055$). No significant differences were found for these contrasts on the activation measure recorded after the music.

Finally, in terms of the ratings for liking recorded after the music, the contrast between the similar pieces and the original showed a clear difference ($F(1, 20) = 20.31$, $p < 0.001$), as did that between the different pieces and the original ($F(1, 20) = 42.09$, $p < 0.001$). The contrast between similar and different pieces was not significant ($p > 0.2$).

7. CONCLUSION AND FUTURE WORK

These assessment results involving human subjects suggest that the model under development captures significant aesthetic similarities in music, evidenced through measurements of human emotional and, perhaps, physiological responses to retrieved music pieces. At the same time they indicate that there remains an important difference in affective responses – greater liking of the piece chosen by the person. Thus, these data provide new insights into the relationship of positive emotion and liking.

It is well documented experimentally that liking increases with exposure to music up to some moderate number of exposures and then decreases as the number of exposures becomes very large [17]. It may be that the pieces chosen by our participants are near that optimum number of exposures whereas the aesthetically similar pieces are insufficiently (or perhaps excessively) familiar to them. Thus, different degrees of familiarity may account for some of the liking differences that were found. It may be desirable in future studies to obtain some independent measure of prior exposure to the different excerpts in order to assess the contribution of this factor.

We plan to conduct additional evaluation experiments involving humans utilizing both melodic and timbre metrics. We also plan to explore different selection algorithms, and give the user more control via the user interface, in terms of selection criteria.

Finally, it is difficult to obtain high-quality MIDI encodings of songs (such as the CMA corpus). However, as Lidy et al. demonstrate, even if the MIDI transcriptions are not perfect, the combined (MIDI + audio metrics) approach may still have more to offer than timbre-based-only (or melodic-only) approaches [6].

This paper presented results from an on-going MIR study utilizing hundreds of melodic and timbre metrics

---

[1] It should be noted that the different songs used in this corpus were not the most dissimilar songs. These selections were used to avoid a cluster of most dissimilar songs, which was the same for most user inputs. However, they also somewhat reduced the distance between similar and dissimilar songs.

based on power laws. Experimental results suggest that power-law metrics are very promising for content-based music querying and retrieval, as they seem to correlate with aspects of human emotion and aesthetics. Power-law feature extraction and classification may lead to innovative technological applications for information retrieval, knowledge discovery, and navigation in digital libraries and the Internet.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] Aucouturier, J.-J. and Pachet, F. (2004). "Improving Timbre Similarity: How High Is the Sky?", *Journal of Negative Results in Speech and Audio Sciences*, 1(1), 2004.

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

[3] Bradley, M.M. and Lang, P.J. (1994). "Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential", *Journal of Behavioral Therapy and Experimental Psychiatry, 25*(1), pp. 49-59.

[4] Cano, P. Koppenberger, M. Wack, N. (2005). "An Industrial-Strength Content-based Music Recommendation System", in *Proceedings of 28th Annual International ACM SIGIR Conference*, Salvador, Brazil, p. 673.

[5] Faloutsos, C. (2003). "Next Generation Data Mining Tools: Power Laws and Self-Similarity for Graphs, Streams and Traditional Data", *Lecture Notes in Computer Science*, 2837, Springer-Verlag, pp. 10-15.

[6] Lidy, T., Rauber, A., Pertusa, A. and Iñesta, J.M. (2007). "Improving Genre Classification by Combination of Audio and Symbolic Descriptors Using a Transcription System", in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, pp. 61-66.

[7] Logan, B. and Salomon, A. (2001). "A Music Similarity Function Based on Signal Analysis", in *Proceedings of the 2001 IEEE International Conference on Multimedia and Expo (ICME '01)*, Tokyo (Japan), pp. 745-748.

[8] Manaris, B., Romero, J., Machado, P., Krehbiel, D., Hirzel, T., Pharr, W., and Davis, R.B. (2005), "Zipf's Law, Music Classification and Aesthetics", *Computer Music Journal*, 29(1), pp. 55-69.

[9] Manaris, B., Roos, P., Machado, P., Krehbiel, D., Pellicoro, L., and Romero, J. (2007). "A Corpus-based Hybrid Approach to Music Analysis and Composition", in *Proceedings of the 22nd Conference on Artificial Intelligence* (AAAI-07), Vancouver, BC, pp. 839-845.

[10] Pampalk, E. Flexer, A., and Widmer, G. (2005). "Improvements of Audio-Based Music Similarity and Genre Classification", in *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK.

[11] Pampalk, E. (2006). "Audio-Based Music Similarity and Retrieval: Combining a Spectral Similarity Model with Information Extracted from Fluctuation Patterns", *3rd Annual Music Information Retrieval eXchange (MIREX'06)*, Victoria, Canada, 2006.

[12] Salingaros, N.A. and West, B.J. (1999). "A Universal Rule for the Distribution of Sizes", *Environment and Planning B: Planning and Design*, 26, pp. 909-923.

[13] Schroeder, M. (1991). *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W. H. Freeman and Company.

[14] Spehar, B., Clifford, C.W.G., Newell, B.R. and Taylor, R.P. (2003). "Universal Aesthetic of Fractals." *Computers & Graphics*, 27, pp. 813-820.

[15] Tzanetakis, G., Essl, G., and Cook, P. (2001). "Automatic Musical Genre Classification of Audio Signals", in *Proceedings of the 2nd International Conference on Music Information Retrieval (ISMIR 2001)*, Bloomington, IN, pp. 205-210.

[16] Voss, R.F. and Clarke, J. (1978). "1/f Noise in Music: Music from 1/f Noise", *Journal of Acoustical Society of America* , 63(1), pp. 258–263.

[17] Walker, E. L. (1973). "Psychological complexity and preference: A hedgehog theory of behavior", in Berlyne, D, E., & Madsen, K. B. (eds.), *Pleasure, reward, preference: Their nature, determinants, and role in behavior*, New York: Academic Press.

[18] Zipf, G.K. (1949), *Human Behavior and the Principle of Least Effort*, Hafner Publishing Company.