A Corpus-Based Hybrid Approach to Music Analysis and Composition

Bill Manaris[1], Patrick Roos[2], Penousal Machado[3], Dwight Krehbiel[4],
Luca Pellicoro[5], and Juan Romero[6]

[1,2,5] Computer Science Department, College of Charleston, 66 George Street, Charleston, SC 29424, USA,
{manaris, patrick.roos, luca.pellicoro}@cs.cofc.edu
[3] CISUC, Department of Informatics Engineering, University of Coimbra, 3030 Coimbra, Portugal, machado@dei.uc.pt
[4] Psychology Department, Bethel College, North Newton KS, 67117, USA, krehbiel@bethelks.edu
[6] Creative Computer Group - RNASA Lab - Faculty of Computer Science, University of Coruña, Spain, jj@udc.es

Abstract

We present a corpus-based hybrid approach to music analysis and composition, which incorporates statistical, connectionist, and evolutionary components. Our framework employs artificial *music critics*, which may be trained on large music corpora, and then pass aesthetic judgment on music artifacts. Music artifacts are generated by an evolutionary *music composer*, which utilizes music critics as fitness functions. To evaluate this approach we conducted three experiments. First, using music features based on Zipf's law, we trained artificial neural networks to predict the popularity of 992 musical pieces with 87.85% accuracy. Then, assuming that popularity correlates with aesthetics, we incorporated such neural networks into a genetic-programming system, called NEvMuse. NEvMuse autonomously "composed" novel variations of J.S. Bach's *Invention #13 in A minor* (BWV 784), variations which many listeners found to be aesthetically pleasing. Finally, we compared aesthetic judgments from an artificial music critic with emotional responses from 23 human subjects. Significant correlations were found. We provide evaluation results and samples of generated music. These results have implications for music information retrieval and computer-aided music composition.

Introduction

Music composition is one of the most celebrated activities of the human mind across time and cultures. According to Minsky and Laske (1992), due to its unique characteristics as an intelligent activity, it poses significant challenges to existing AI approaches, with respect to (a) formalizing music knowledge, and (b) generating music artifacts.

In this paper, we present a corpus-based hybrid approach to music composition, which incorporates statistical, connectionist, and evolutionary components. We model music composition as a process of iterative refinement, where music artifacts are generated, evaluated against certain aesthetic criteria, and then refined to improve their aesthetic value. In terms of formalizing music knowledge, we employ *artificial music critics* – intelligent agents that may be trained on large music corpora, and then pass

aesthetic judgment on music artifacts. In terms of generating music action, we employ an evolutionary *music composer* – an intelligent agent, which generates music through genetic programming, utilizing artificial music critics as fitness functions.

Artificial Art Critics

The process of music composition depends highly on the ability to perform aesthetic judgments, to be inspired by the works of other composers, and to act as a critic of one's own work. However, most music generation systems developed in the past few years neglect the role of the listener/evaluator in the music composition process (e.g. see survey by Wiggins et al., 1999). We believe that modeling the aesthetic judgment part of the human composer is an important, if not necessary, step in the creation of a "successful" artificial composer.

Artificial Art Critics (AACs) are intelligent agents capable of classifying/evaluating human- or computer-generated artifacts using as learning base a set of taxonomized examples (Romero et al. 2003; Machado et al., 2003). In particular, the AAC architecture incorporates a *feature extractor* and an *evaluator* module (see figure 1). The *feature extractor* is responsible for the perception of music artifacts, generating as output a set of measurements that reflect relevant characteristics. These measurements serve as input for the *evaluator*, which assesses the artwork according to a specific criterion or aesthetics.

In this paper, we explore the development of AACs specific to music evaluation. We focus on two main aspects, the use of metrics based on Zipf's law for the development of AACs, and the use of these AACs for fitness assignment in an evolutionary composition system.

Evolutionary Music Composition

The main difficulty in the application of evolutionary computing (EC) techniques to music tasks involves choices for (a) an appropriate representation and (b) an appropriate fitness assignment scheme. It can be argued that music (at least conventional music) has a hierarchic structure; thus developing representations that capture and take advantage of this structure may be an important step in the development of an effective EC music system. Typically, genetic algorithm (GA) approaches use a linear
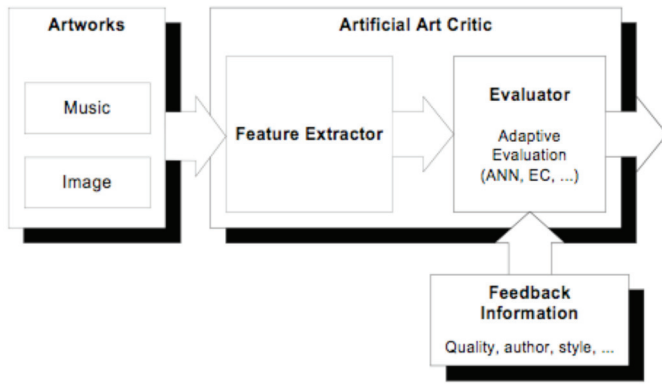
Figure 1. Overview of the AAC architecture.

representation, while genetic programming (GP) approaches prefer tree-based representations. According to Papadopoulos and Wiggins (1999) the hierarchical nature of GP representations make them more suited for musical tasks.

The use of a robust, adaptive and flexible method such as genetic programming, together with a mechanism of internal evaluation based on examples of "good" artifacts, may facilitate the autonomous generation of new music themes similar to (a) a particular composer's style (e.g., J.S. Bach), (b) a particular musical genre (e.g., Jazz), (c) an individual's eclectic aesthetic preferences (as identified by the chosen "good" examples), and possibly (d) various combinations of the above.

The proposed framework is evaluated through three experiments. The first experiment evaluates the ability of artificial critics to classify music based on aesthetics. The second experiment focuses on evolutionary music composition utilizing such critics. The third experiment compares aesthetic judgments of an artificial critic with those of human listeners.

Related Work

Horner and Goldberg (1991) applied a GA to perform thematic bridging, in what became the first work exploring the use of an EC approach in a music-related task. Since then a vast number of papers on the subject have been published (for a thorough survey see Todd and Werner, 1998; Miranda and Biles, 2007). Today, EC music comprises a wide array of tasks, including composition, harmonization, sound synthesis, and improvisation.

Fitness assignment plays an important role in any EC system; musical tasks are not an exception. There are essentially five different approaches to fitness assignment:

- *interactive evaluation* – fitness values are provided by humans (e.g. Horowitz, 1994);
- *similarity based* – fitness depends on "proximity" to a specific sound or music piece (e.g. Horner and Goldberg, 1991);

- *hardwired fitness functions*, typically based on music theory (e.g. Phon-Amnuaisuk and Wiggins, 1999);
- *machine learning approaches*, such as neural networks (e.g. Gibson and Byrne, 1991); and
- *co-evolutionary approaches* (e.g. Todd and Werner, 1998).

The combination of several of the above methods has also been explored (e.g. Spector and Alpern, 1994; 1995).

We are interested in EC composition systems that employ machine-learning methods to supply fitness. Among related systems, the work of Johanson and Poli (1998) is similar to our approach. They employ a GP system, where the function set consists of operations on sets of notes (e.g. play_twice), and the terminal set consists of individual notes and chords. Initially, small tunes are evolved by interactive evolution. These are then used to train an artificial neural network (ANN) with shared weights, which is able to handle variable length inputs. Once trained, the ANN is used to assign fitness to new individuals. Our approach is different with respect to (a) the ANN input values (notes vs. extracted features); (b) the topology of the ANN; and (c) the algorithm used to build the training corpus.

In Machado et al. (2007) we present a similar approach in the visual domain. In this case, the AAC is trained to distinguish between external images (e.g., paintings) and images created by evolutionary artists. The iterative refinement of the AAC forces the GP engine to explore new paths, leading to a stylistic change. The inclusion of a fixed set of external images provides an "aesthetic referential" promoting the relation between evolved imagery and conventional aesthetics.

Music Aesthetics and Power Laws

Our approach builds on earlier research, which suggests that power laws provide a promising statistical model for music aesthetics.

According to Salingaros and West (1999), most pleasing designs in human artifacts obey a power law. "The relative multiplicity p of a given design element, i.e., the relative number of times it repeats (frequency), is determined by a characteristic scale size x as roughly $px^m = C$, where C is related to the overall size of the structure, and the index m is specific to the structure."

A logarithmic plot of p versus x has a slope of m, where $-1 \leq m \leq -2$. Exceptions to this rule correspond to "incoherent, alien structures" (ibid, p. 909).

In many cases, *statistical rank* may be used instead of *size*. This variation is known as Zipf's law, after the Harvard linguist, George Kingsley Zipf, who studied it extensively in natural and social phenomena (Zipf, 1949). Figure 2 shows the *rank-frequency* distribution of melodic intervals in Chopin's "Revolutionary Etude", which approximates Zipf's law.

Voss and Clarke (1978) have shown that classical, rock, jazz, and blues music exhibit a power law with slope approximately -1. They generated music artifacts
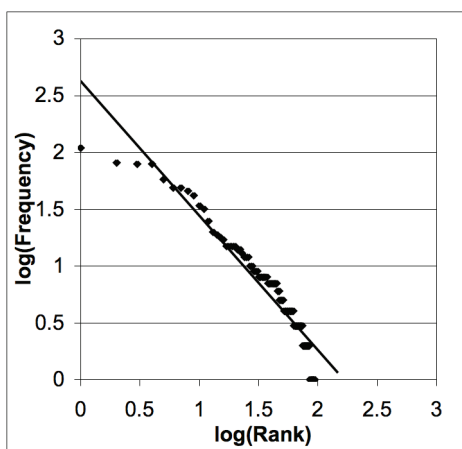
Figure 2. Distribution of melodic intervals for Chopin's "Revolutionary Etude", Opus 10 No. 12 in C minor. Slope is –1.18, R^2 is 0.92.

exhibiting power law distributions with m ranging from 0 (*white noise*), to –1 (*pink noise*), to –2 (*brown noise*). Pink-noise music was much more pleasing to most listeners, whereas white-noise music sounded "too random", and brown-noise music "too correlated".

Manaris et al. (2003) showed that 196 "socially-sanctioned" (popular) music pieces exhibit power laws with m near –1 across various music attributes, such as pitch, duration, and melodic intervals.

Power laws have been applied to music classification, in terms of composer attribution, style identification, and pleasantness prediction, as follows:

Composer Attribution. Machado et al. (2003, 2004) trained ANNs to classify music pieces between various combinations of composers including Bach, Beethoven, Chopin, Debussy, Purcell, and Scarlatti. Features were extracted from these pieces using power-law metrics. Corpora ranged across experiments from 132 to 758 MIDI-encoded music pieces. Success rates ranged from 93.6% to 95% across experiments.

Style Identification. In similar experiments, we have trained ANNs to classify music pieces from different styles. Our corpus consisted of Baroque (161 pieces), Classical (153 pieces), Country (152 pieces), Impressionist (145 pieces), Jazz (155 pieces), Modern (143 pieces), Renaissance (153 pieces), Rock (403 pieces), Romantic (101 pieces). ANNs achieved success rates ranging from 71.52% to 96.66% (under publication).

Pleasantness Prediction. Manaris et al. (2005) conducted an ANN experiment to explore correlations between human-reported pleasantness and metrics based on power laws. Features were extracted from 210 excerpts of music, and then human responses to these pieces were recorded. The combined data was used to train ANNs. Using a 12-fold cross-validation study, the ANNs achieved an average success rate of 97.22% in predicting (within one standard deviation) human emotional responses to those pieces.

Feature Extraction

Similarly to the above experiments, we employ music metrics based on power laws to extract relevant features from music artifacts.

Each metric measures the entropy of a particular music-theoretic or other attribute of music pieces. For example, in the case of *melodic intervals*, a metric counts each occurrence of an interval in the piece, e.g., 168 half steps, 86 unisons, 53 whole steps, and so on. Then it calculates the slope and R^2 values of the logarithmic *rank-frequency* distribution (see figure 2). In general, the slope may range from 0 to –∞, with 0 corresponding to high entropy and –∞ to zero entropy. The R^2 value may range from 0 to 1, with 1 denoting a straight line; this captures the proportion of y-variability of data points with respect to the trendline. Our metrics are categorized as follows:

Regular Metrics. These capture the entropy of a regular attribute or event (an 'event' is anything countable, e.g., a melodic interval). We currently employ 14 regular metrics related to pitch, duration, harmonic intervals, melodic intervals, harmonic consonance, bigrams, chords, and rests.

Higher-Order Metrics. These capture the entropy of the *difference* between two consecutive regular events. Similarly to the notion of derivative in mathematics, for each regular metric one may construct an arbitrary number of higher-order metrics (e.g., the difference of two events, the difference of two differences, etc.).

Local Variability Metrics. These capture the entropy of the difference of an event from the local average. In other words, local variability, $d[i]$, for the ith event is

$$d[i] = abs(tNN[i] - average(tNN, i)) / average(tNN, i)$$

where tNN is the list of events, *abs* is the absolute value, and *average(tNN, i)* is the local average of the last, say, 5 events (Kalda et al., 2001). One local variability metric is provided for each of the above metrics.

It should be noted that these metrics implicitly capture significant aspects of musical hierarchy. Similarly to Schenkerian analysis, music events (e.g., pitch, duration, etc.) are recursively reduced to higher-order ones, capturing long-range structure in pieces. Consequently, pieces without hierarchical structure have significantly different measurements than pieces with structure.

A Simple Music Critic Experiment

Since artificial music critics are integral to the success of the proposed approach, we decided to evaluate their effectiveness. However, assessing *aesthetic judgment* is similar to assessing *intelligence*; there is no objective way to do so, other than perhaps a variant of the Turing Test. So, assuming a correlation between *popularity* and *aesthetics*, one could post music pieces on a website and collect download statistics over a long period of time. Another possibility is to ask human subjects to evaluate the aesthetics of music artifacts, and then compare these

judgments with those of music critics. This section explores the first approach. The second approach is explored later in this paper.

For this experiment, we used the Classical Music Archive corpus, which consists of 14,695 MIDI pieces (http://www.classicalarchives.com). We also obtained download logs for one month (November 2003), which contains a total of 1,034,355 downloads. Using these data, we identified the 305 most popular pieces.

A piece was considered *popular* if it had a minimum of 250 downloads for the month. For example, the five most popular pieces were:
- Beethoven's *Bagatelle No. 25 in A minor, "Fur Elise"* (9965 requests);
- J.S. Bach's *Jesu, Joy of Man's Desiring, BWV147* (8677 requests);
- Vivaldi's *'Spring' Concerto, RV.269, "The Seasons", 1. Allegro* (6382 requests);
- Mozart's *Divertimento in D, K.136, 1. Allegro* (6190 requests); and
- Mozart's *Sonata in A, K.331 (with Rondo alla Turca)* (6017 requests).

Using the same download statistics, we also identified 617 unpopular pieces. To ensure a clear separation between the two sets (and thus control for other variables, such as physical placement of links to music pieces within the website), we selected pieces with only 20 or 21 downloads for the month. This separated the two sets (popular and unpopular) by several thousand pieces. For example, five unpopular pieces were (all at 20 requests):
- Marchetto Cara's, *Due frottole a quattro voci, 1. Crudel, fugi se sai*;
- Niels Gade, *String Quartet in D, Op. 63, 3. Andante, poco lento*;
- Ernst Haberbier, *Studi-Poetici, Op. 56, No. 17, Romanza*;
- George Frideric Handel, *Tamerlano, HWV18 – Tamerlano's aria "A dispetto d'un volto ingrato"*; and
- Igor Stravinsky, *Oedipus Rex, Caedit nos pestis – Liberi, vos liberado*.

ANN Classification Tests

Several ANN classification tests were conducted between popular and unpopular pieces. The metrics described earlier were used to extract features (slope and R^2 values) for each music piece.

In the first classification test, we trained an ANN with 225 features extracted per piece. We carried out a 10-fold, cross-validation experiment using a feed-forward ANN trained via backpropagation. The ANN trained for 500 epochs using values of 0.2 for momentum and 0.3 for learning rate. The ANN architecture involved 225 elements in the input layer and 2 elements in the output layer. The hidden layer contained *(input nodes + output nodes)/2* nodes.

For control purposes, we ran a second experiment identical to the first, using randomly assigned classes for each music piece.

Finally, we ran a third experiment with the same setup as the first, but using only the 79 most relevant attributes. These were the attributes most highly correlated with a class, and least correlated with one another.

Results and Discussion

In the first test, the ANN achieved a success rate of 87.85% (correctly classified 810 of 922 instances). The ANN in the control test achieved a success rate of 49.68% (458 of 922 instances). This result suggests that the high success rate of the first ANN is due mainly to the effectiveness of the metrics. In the third classification test, using the 79 most relevant features, the ANN achieved a success rate of 86.11%.

Clearly, the prominent issue in using artificial music critics is finding appropriate corpora to train them. It is quite easy to find popular (socially sanctioned) music, but much harder to find truly unpopular (bad) music, since, by definition, the latter does not get publicized or archived.[1] Even without access to truly bad "music", this experiment demonstrates the potential for developing artificial music critics that may be trained on large music corpora.

A Simple Music Composer Experiment

The second experiment evaluated the effectiveness of an evolutionary music composer incorporating artificial music critics for fitness assignment.

We implemented a genetic-programming system, called NEvMuse (Neuro-Evolutionary Music environment). This is an autonomous genetic programming system, which evolves music pieces using a fitness mechanism based on examples of desirable pieces. Assuming a correlation between popularity and aesthetics, NEvMuse utilizes ANNs, trained on various music corpora, as fitness functions. The input to NEvMuse consists of:

- a harmonic outline of the piece to be generated (MIDI);
- a set of "melodic genes" to be used as raw material (MIDI); and
- a music critic.

The harmonic outline provides a harmonic and temporal template to be filled in. The melodic genes may be a few notes (e.g., a scale, a solo, etc.), or a complete piece; these may be broken up into individual notes or phrases of random lengths.

The system proceeds by creating random arrangements of the melodic genes, evaluating them using the music critic, and recombining them using standard genetic operators. This process continues until a fitness threshold is reached.

---

[1] Even the unpopular music in the music critic experiment is not truly bad, as it has to be somewhat aesthetically pleasing to some listeners for it to have been performed, published, and archived.

```
(+ (+ (retro (+ (+ n(16,18) n(21,22)) (+
n(37,38) (+ (+ (+ (retro (+ (+ (+ n(12,14)
n(11,13)) n(12,14)) n(13,16))) n(0,2))
n(95,98)) ...
```

Figure 3. S-expression (LISP) tree genotype sample. ('+' stands for concatenation, '*retro*' for retrograde, and '*n(x,y)*' for "melodic gene" notes *x* through *y*).

Figure 4. Score phenotype of sample genotype shown in figure 3 (excerpt).

Genotype Representation

NEvMuse represents the genotype of an individual as a symbolic expression (LISP) tree (see figure 3). This tree is comprised of a set of operators (nonterminal nodes), which are applied to a set of MIDI phrases (terminal nodes). The phenotype is a MIDI file (see figure 4).

The genotype operators model traditional music composition devices. These include superimposing two phrases (polyphony); concatenating two phrases (sequence); retrograding a phrase; inverting a phrase; transposing a phrase; augmenting a phrase; and diminishing a phrase.

The system uses two standard genetic operators to evolve genotypes: (a) swap-tree crossover (with a variable number of crossover points) and (b) random subtree mutation (replacing a subtree by a randomly generated one). The selection scheme used is roulette wheel. Setup parameters include population count (e.g., 500), fitness threshold (e.g., 0.99), max generations (e.g., 1000), elite percentage (e.g., 15%), crossover rate (e.g., 0.5), crossover points (e.g., 2), mutation rate (e.g., 0.8), and parameters to dynamically adjust genotype tree depth.

Music Generation Tests

Several music generation tests were conducted exploring different possibilities. To reduce the number of variables, we instructed NEvMuse to create variations of a single piece, namely J.S. Bach's *Invention #13 in A minor* (BWV 784). Thus, we evaluated five music critics:

(A) Popular vs. Unpopular: Fitness was determined by a static ANN trained to recognize popular vs. unpopular music (see previous experiment).

(B) Actual vs. Random Music: Fitness was determined by a static ANN trained with actual music ("popular" corpus) vs. random music (generated off-line by NEvMuse via random fitness assignment).

(C) Actual vs. Artificial Music: Fitness was determined by a dynamic ANN that was trained during evolution. Initially, the ANN was trained with an actual vs. random music corpus (same as B). The ANN was then retrained at the end of each generation; the training corpus was "bootstrapped" by adding the latest population into the random corpus. Evolution stopped when the ANN training error became too large, i.e., the ANN could not differentiate between actual and generated music.

(D) Mean Square Error (MSE): Fitness was determined by calculating the MSE between a genotype's features (slope and R^2 values) and the features of a target piece. In other words, high fitness was assigned to genotypes with statistical proportions similar to the target piece.

(E) Random: Fitness was determined by a random number generator (for control purposes).

For melodic genes, we explored three choices:

(1) Original Notes: Melodic genes were all notes in the original piece.

(2) Minor Scale: Melodic genes were half notes, quarter notes, 8th notes, and 16th notes in the A minor scale.

(3) 12-Tone Scale: Melodic genes were half notes, quarter notes, 8th notes, and 16th notes in the chromatic scale.

Below, we use the notation *x.n* to refer to a system configuration incorporating music critic *x* (where *x* may be A, B, C, D, or E), and melodic gene choice *n* (where *n* may be 1, 2, or 3).

Results and Discussion

NEvMuse autonomously "composed" many variations of BWV 784, which a wide variety of listeners have informally judged as aesthetically pleasing.

It should be noted that, ultimately, this experiment is a *performance test* of our power-law based metrics. Any imperfections in the generated music correspond to deficiencies in how the metrics model music aesthetics. Thus, the generated music samples provide a *sonification* of these deficiencies; they are invaluable in refining the metrics (see **http://www.cs.cofc.edu/~manaris/nevmuse**).

In terms of aesthetics, configurations D.1, C.1, B.1, and A.1 performed well, probably in this order. Surprisingly, even configuration E.1 sometimes produced interesting pieces. (Again, "1" means original notes.) We believe this is because the melodic genes effectively implement a probabilistic scheme: repeated notes (e.g., tonic, 5th, minor 3rd) in the original piece have more chances of appearing in genotypes.

In terms of effectiveness, critic C is the only critic that produced relatively interesting results with gene types 2 and 3. We suspect, critic A was less effective because the ANN was trained to classify between two types of actual music, whereas NEvMuse's early populations do not resemble actual music; critic B was less effective because the ANN is static; critic D was less effective because it rewarded statistical similarity with a single piece (as opposed to many pieces by the ANN-based critics). This suggests that the ANN "bootstrapping" approach is very promising for EC composition.

An Aesthetic Judgment Experiment

An experiment was conducted to compare the aesthetic judgment of an MSE-based artificial music critic to that of 23 human subjects recruited from undergraduate psychology classes (6 male, 17 female; age 18-22; 0-14 years of private music lessons). Both artificial and human participants rated J.S. Bach's *Invention #13 in A minor* (BWV 784) and 17 variations generated by NEvMuse (see http://www.cs.cofc.edu/~manaris/nevmuse). Two of these variations were created to be "unpleasant", for comparison.

Barrett and Russell (1999) describe *pleasantness* and *activation* "as basic and universal dimensions of affect". Our human participants provided continuous ratings of pleasantness and activation, while listening to the music. This was done by moving a computer cursor on a two-dimensional space with emotion labels around the periphery, e.g., "happy", "serene", "calm", "lethargic", "sad", "stressed", and "tense" (Barrett and Russell, 1999; Schubert, 2001). Subjects were carefully instructed to report their own feelings rather than their judgments of composer or performer intent.

The artificial critic provided aesthetic judgments by calculating the similarity between each variation and BWV 784, using the MSE approach (see previous section). In particular, the 15 "pleasant" variations were assigned high aesthetic values (i.e., low MSE ranging from 0.0086 to 0.0644); whereas the two "unpleasant" variations were assigned the lowest aesthetic values (i.e., highest MSEs of 0.1515 and 0.1685, respectively).

Results and Discussion

Results were analyzed using hierarchical linear modeling (HLM 6.0, Scientific Software International), with variations over time as level-1 variables, and participant characteristics as level-2 variables.

The interaction of time and MSE was highly predictive of pleasantness ($p < 0.001$), as well as of activation ($p < 0.001$). These interactions reflect the fact that the changes in ratings over time were different for the original and the variations, especially the two "unpleasant" ones (see figure 5). Additional significant predictors in the activation model were the separate variables of time (activation decreasing over time, $p < 0.001$) and MSE (increasing as MSE increased, $p = 0.034$). Pearson correlations, with MSE
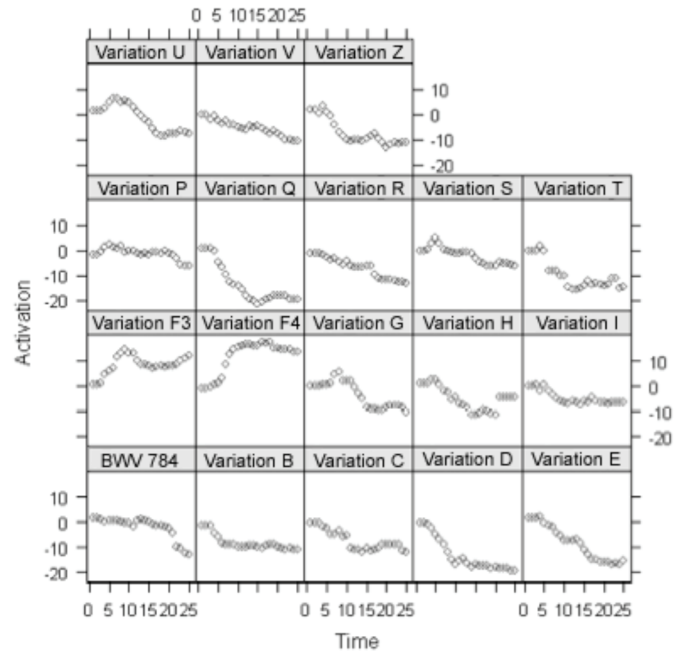
Figure 5. Plots of mean self-reported activation (n = 23) over time, recorded during J.S. Bach's *Invention #13 in A minor* (BWV 784) and 17 variations. Note the two "unpleasant" variations (F3 and F4).

calculated on data averaged over time and over participants, were –0.620 for pleasantness and 0.747 for activation. Thus, the aesthetic judgment of the artificial music critic was a strong predictor of both pleasantness and activation ratings of human listeners; this relationship emerged in spite of large differences between participants, which were highly significant in HLM models. This further confirms the aesthetic relevance of the considered power-law metrics.

Conclusion

We have described a corpus-based approach to music analysis and composition involving (a) music critics utilizing power-law metrics, which may be trained on large music corpora; and (b) an evolutionary music composer that utilizes such music critics as fitness functions.

This approach has obvious implications for intelligent music retrieval tasks, such as identifying music similar to a set of favorite songs. One possibility is a music search engine based on aesthetic similarity. For example, see a demo at http://www.cs.cofc.edu/~manaris/music-search.

Finally, the use of a robust, adaptive and flexible method such as genetic programming, together with a mechanism of internal evaluation based on examples of "good" artifacts, supports the generation of new music themes. Tools based on this framework could be utilized by human composers as cognitive prostheses to help generate new ideas, to overcome "writer's block", and to explore compositional spaces.

References

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

Gibson, P. M., and Byrne, J. A. (1991), "Neurogen, Musical Composition Using Genetic Algorithms and Cooperating Neural Networks", *Second International Conference on Artificial Neural Networks*, 309-313.

Horner, A., and Goldberg, D. E. (1991), "Genetic Algorithms and Computer-Assisted Music Composition", *International Computer Music Conference* (ICMC-91), Montréal, Québec, Canada, 479-482.

Horowitz, D. (1994), "Generating Rhythms with Genetic Algorithms", *International Computer Music Conference* (ICMC'94), Aarhus, Denmark, 142-143.

Johanson, B., and Poli, R. (1998), "GP-Music: An Interactive Genetic Programming System for Music Generation with Automated Fitness Raters", *Proceedings of Third Annual Genetic Programming Conference*, Madison, WI, 181-186.

Kalda, J., Sakki, M., Vainu, M., and Laan, M. (2001), "Zipf's Law in Human Heartbeat Dynamics", http://arxiv.org/abs/physics/0110075.

Machado, P., Romero, J., Manaris, B., Santos, A., and Cardoso, A. (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), Acapulco, Mexico, 55-64.

Machado, P., Romero, J., Santos, M.L., Cardoso, A., and Manaris, B. (2004), "Adaptive Critics for Evolutionary Artists", *Applications of Evolutionary Computing*, LNCS 3005, Springer-Verlag, 437-446.

Machado, P., Romero, J., Manaris, B. (2007), "Experiments in Computational Aesthetics – An Iterative Approach to Stylistic Change in Evolutionary Art", *The Art of Artificial Evolution*, Springer-Verlag (to appear).

Manaris, B., Vaughan, D., Wagner, C., Romero, J. and Davis, R.B. (2003), "Evolutionary Music and the Zipf–Mandelbrot Law – Progress towards Developing Fitness Functions for Pleasant Music", *Applications of Evolutionary Computing*, LNCS 2611, Springer-Verlag, 522-534.

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): 55-69.

Minksy, M., and Laske, O. (1992), "Forward: A Conversation with Marvin Minksy", *Understanding Music with AI: Perspectives on Music Cognition*, AAAI Press.

Miranda, E.R. and Biles, A. (2007), *Evolutionary Computer Music*. Springer-Verlag.

Papadoulos, G., and Wiggins, G. (1999), "AI Methods for Algorithmic Composition: A Survey, a Critical View and Future Prospects", *Proceedings of AISB'99 Symposium on Musical Creativity*, Edinburgh, UK, 110-117.

Phon-Amnuaisuk, S., and Wiggins, G.A. (1999), "The Four-Part Harmonisation Problem: A Comparison between Genetic Algorithms and a Rule-Based System", *Proceedings of AISB'99 Symposium on Musical Creativity*, Edinburgh, UK, 28-34.

Romero J., Machado P., Santos A., Cardoso A., (2003), "On the Development of Critics in Evolutionary Computation Artists", *Applications of Evolutionary Computing*, LNCS 2611, Springer-Verlag, 559-56.

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

Spector, L., and Alpern, A. (1994), "Criticism, Culture, and the Automatic Generation of Artworks", *Proceedings of Twelfth National Conference on Artificial Intelligence,* Seattle, WA, 3-8.

Spector, L., and Alpern, A. (1995), "Induction and Recapitulation of Deep Musical Structure", *Proceedings of Workshop on Artificial Intelligence and Music, 14th International Joint Conference on Artificial Intelligence* (IJCAI 1995), Montréal, Québec, Canada, 41-48.

Schubert, E. (2001). "Continuous Measurement of Self-Report Emotional Response to Music". In P. N. Juslin & J. A. Sloboda (Eds.), *Music and Emotion: Theory and Research*, 391-414, Oxford University Press.

Todd, P. M., and Werner, G.M. (1998), "Frankensteinian Methods for Evolutionary Music Composition", *Musical Network*, 313-339, MIT Press/Bradford Books.

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): 258–263.

Wiggins, G.A., Papadopoulos, G., Phon-Amnuaisuk, S., and Tuson, A. (1999), "Evolutionary Methods for Musical Composition", *International Journal of Computing Anticipatory Systems*, 1(1), 1999.

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