Appears in Proceedings of 8th IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2012), Piraeus-Athens, Greece, pp. 436-440, July 20, 2012.

Fractal Dimensions of Music and Automatic Playlist Generation

Similarity Search via MP3 Song Uploads

Dana Hughes and Bill Manaris
Department of Computer Science
College of Charleston
Charleston, SC, United States
dthughes@g.cofc.edu, manaris@cs.cofc.edu

Abstract— **We present an automated approach to music search and playlist generation based on fractal dimensions of music. We compute 372 power-law metrics per song capturing statistical proportions of musical material. Using attribute selection and principal component analysis, we have reduced these metrics to approximately 45 independent features. These have been shown to capture important aspects of music aesthetics and similarity. Through an audio-to-MIDI transcription process, users may upload MP3 songs as search queries, in real time. This new development enables construction of music recommendation systems, which may work with previously unknown music. Unlike Pandora, last.fm, and Genius, such systems will analyze the actual music (potentially like the human ear), as opposed to harvesting information from humans (e.g., websites, user preferences, or musicologist recommendations). This approach combines time-frequency and spectral processing, information retrieval and audio analysis, and music classification. We present two on-line demos, using corpora from Magnatune and 7digital.**

Keywords: Fractal geometry of music; Zipf's law; power laws; music similarity; audio-to-MIDI transcription; playlist generation

I. Introduction

We report the latest results from a decade-long project investigating fractals in music, and their relationship to human aesthetics. This project explores Zipf's Law (and related power laws) in music data mining and information retrieval [1, 2]; and in computer-aided music analysis, composition, and performance [3, 4]. One project outcome is an experimental music search engine, called *Armonique*, which automatically identifies aesthetic similarities in musical content (www.armonique.org).

In this paper, we explore how fractal dimensions of musical material can be used to automatically generate music recommendation lists in real-time, with audio-to-MIDI transcription and analysis of MP3 songs, as a front-end to the Armonique system. This is similar to a Google-like engine, which accepts actual music as search queries. This system improves on music fingerprinting (e.g., Shazam), in that it not only can identify uploaded songs, but it can also recommend similar songs, ordered by musical and timbral resemblance.

Our approach utilizes hundreds of power-law metrics, which calculate the fractal dimension of music-theoretic, timbral, and other musical attributes of pieces. These fractal dimensions include pitch, duration, repeated-pitch distance, repeated-duration distance, melodic interval, harmonic interval, and many others [1-4]. The advantage of using fractal dimensions to calculate music similarity is that we analyze actual musical content (as opposed to harvesting user preferences in social or collaborative filtering systems, such as iTunes Genius). This approach has emerged from studying the relationship between fractal dimensions of music and human aesthetics. Various earlier experiments have explored this relationship using artificial neural networks (ANNs), e.g., composer identification: 93.6% - 95.3% accuracy; style classification: 71.5% - 96.6% accuracy; pleasantness prediction: 90.7% accuracy [1-3]. Moreover, psychological experiments comparing Armonique's judgments to emotional and physiological responses of human subjects have validated the aesthetic similarity of retrieved pieces from a human listener perspective [1-3].

II. Background

A. *Automatic Music Playlist Generation*

Automatic music recommendation and playlist generation has received much attention recently in music information retrieval (e.g., [5-9]). For instance, playlists can be generated to reflect a particular mood or activity, or be used as a tool to discover previously unknown music. Playlist generation systems are content-based or metadata/context-based [5]. Content-based systems extract features from audio or MIDI, while metadata-based systems rely on individual songs being tagged with information such as artist, style, genre, etc. Popular systems like Pandora, last.fm, and iTunes Genius utilize metadata generated through collaborative filtering (e.g., user ratings), and/or musicologists. Below is the state-of-the-art in content-based automatic playlist generation:

Turnbull, et al. have developed a content-based autotagging system, which generates semantic data for a given song [6]. This system automatically annotates music with coarse textual descriptors, such as "tender", "hip-hop", and "exercising", using a corpus of 1305 tracks and a vocabulary of 348 words.

This work has been funded in part by NSF grants IIS-0736480, IIS-0849499 and IIS-1049554.

Results show that listeners preferred iTunes Genius when the query song was known, but the autotagging system performed equally well or better with unknown music. Given its nature, this system generates coarsely-related playlists (e.g., "exercising"), compared to our approach that focuses on finer melodic and timbral similarity.

Listeners seem to be very subjective in their evaluation of automatically generated playlists [8]. Maillet et al., propose an automatic playlist generator, which allows users to customize the recommendation model based on a tag cloud [9]. This hybrid system extracts 180 attributes from four audio features (autocorrelation coefficients, Mel-frequency cepstral coefficients, "danceability" and long-term loudness level) for each song. Using existing radio-station playlists, the system trains artificial neural networks (ANNs) to take in results produced by the four coarse audio features and remove unlikely candidates. Finally, a metadata-based system is used to fine-tune song selection based on the tag cloud provided by the user. This system utilizes both content and context data, making it more complicated to use, as the user must seed it with a tag cloud. Results are still inconclusive, but, according to Maillet et al., this is a first step towards a more effective hybrid system.

Our approach is strictly content-based, utilizing the fractal dimension of hundreds of musical and timbral attributes (as seen in Section III).

B. *Audio Transcription*

Musical timbre, musical pitch, and other audio characteristics (e.g., prosody of vocalizations) are all represented by frequency and amplitude changes in recorded audio signals. In order to calculate our fractal-dimension metrics, the audio signal has to be converted to a higher-level representation. This representation should capture timbral characteristics (e.g., frequencies) and musical notes.

Automated transcription of musical signals is still an open research area. The goal of most transcription techniques in the literature is to generate an accurate representation of the musical structure (i.e., notes) in an audio file (e.g., [10-14]). Typically, these techniques extract notes from a time-frequency representation (e.g., short-time Fourier Transform or Constant Q Transform) of an audio signal using a trained model (e.g., neural network, hidden Markov model, etc.).

Our approach utilizes a highly tunable adaptation of the Constant Q Transform algorithm (as seen below).

III. APPROACH

Our content-based approach to playlist generation employs hundreds of power-law metrics, which measure the fractal dimensions of music-theoretic, timbral, and other musical attributes of pieces. As earlier studies show (e.g., [1-3]), these metrics, collectively, act as a statistical signature for authorship-attribution, style-identification, and pleasantness-prediction tasks. Our hypothesis is that these metrics (by capturing the interplay between chaos and monotony) are related to how the human auditory system processes musical information. Evaluation results appear to support this hypothesis.

Using attribute selection and principal component analysis, we reduce these metrics to approximately 45 independent features. These are shown to capture important aspects of music aesthetics and similarity.

We automatically compute these measurements on various corpora. These vectors are stored on an on-line database. Then, through an audio-to-MIDI transcription process (discussed below), users may upload MP3 songs as search queries, in real time. Similarity results (i.e., playlist recommendations) are generated using a simple Euclidean distance calculation. We have also developed a binary-search-like algorithm (to be reported elsewhere) which reduces search from $O(n)$ to $O(logn)$, where n is the number of songs in the database. This makes it possible to use this system with very large music corpora.

A. *Power-Law Metrics And Feature Extraction*

Currently, we calculate the *rank-frequency* (or *size-frequency*) distribution of a particular attribute (e.g., pitch, duration, etc.), although other techniques have been explored. The rank-frequency distribution is:

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

where $P(f)$ denotes the probability of an event of rank f, and a and b are real constants.

When a and b approximate 1, this is known as *Zipf's law*, after George K. Zipf, the Harvard linguist who first observed the ubiquity of this distribution in natural phenomena and human ecology. Actually, Zipf was the first to observe this distribution in music (see [15], p. 336). Zipf's law is also known as *1/f* or pink-noise distribution. In general, equation (2) is known as a *power law*.

Numerous studies demonstrate that music exhibits power laws across timbre, melodic, and rhythmic attributes (e.g., [15-18]). In particular, Voss and Clarke demonstrate that music audio properties (e.g. loudness and pitch fluctuation) exhibit power-law relationships [16]. Hsü and Hsü discuss *1/f* distributions in musical pitch [17]; and Levitin, et al. explore *1/f* distributions in musical rhythm [18].

We compute 372 power-law metrics per song capturing various fractal dimensions of music-theoretic, timbral, and other musical attributes of pieces. These musical attributes include pitch, chromatic tone, duration, pitch and duration combined, distance between repeated notes, distance between repeated durations, melodic and harmonic intervals, melodic and harmonic consonance, melodic and harmonic bigrams, chords, and rests (a total of 300 MIDI metrics). The timbral attributes capture various fractal dimensions of frequency components (a total of 72 audio metrics).

Each metric calculates the *rank-frequency* distribution of the attribute in question [15], and returns two values:

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

Since these measurements relate to the entropy of the provided signal, the 372 metrics we currently utilize tend to be highly correlated. For example, a musical piece with chaotic pitch (e.g., *12-tone* music) tends to be chaotic across other attributes as well (e.g., melodic intervals, harmonic intervals, etc.). Using attribute selection and principal component analysis, we reduce these metrics to approximately 45 independent features. These are shown to capture important aspects of music aesthetics and similarity (see Section IV).

B. Transcription Algorithm

Through an audio-to-MIDI transcription process, users may upload MP3 songs as search queries, in real time. For this, we have implemented our own version of the constant Q transform (CQT) algorithm [14]. The main advantage of the CQT algorithm is that it is very efficient, as it focuses on frequencies that correspond to MIDI notes (as opposed to all possible frequencies in the signal). Since our approach utilizes MIDI notes as an intermediate representation, this simplification is quite appropriate. Our algorithm allows training of various parameters via genetic algorithms to improve its accuracy. (This is on-going work to be reported elsewhere.)

Initially, the transcription algorithm adjusts the audio volume by segmenting the signal into windows of fixed duration. The total energy of each window is normalized, which helps maintain a constant audio volume of the signal. This simplifies the calculation of threshold values. We currently use a window size of 2 seconds to normalize the volume.

Then the CQT algorithm converts the audio signal to a spectrogram. This bins the audio signal into frequencies according to individual MIDI notes. For efficiency, we currently constrain frequency analysis to MIDI notes between 32 and 108.

We then refine the spectrogram by removing noise and other spurious data. The variable-length windows for each frequency are converted to a fixed-length window of 9.375 *ms* by interpolating the calculated values of each bin. Potential harmonics of lower frequencies are eliminated within each window by using the following heuristic formula:

$$X(f_k) < H^k \, X(f_0) \tag{1}$$

where k is the harmonic index, and H is a parameter (between 0 and 1) adjusting harmonic removal sensitivity (for the experiments below, 0.55 was used).

The final stage involves extracting MIDI notes from the spectrogram. Within each window, the peaks of the spectrogram are determined; then any non-zero, non-peak values are eliminated. The remaining values are converted to MIDI notes.

For each note, we capture the start time, end time and average volume. We keep a maximum of 10 notes per window. We eliminate notes shorter than 65.625 *ms* in duration or have a MIDI velocity (volume) less than 25 (both adjustable). We save all remaining notes in a MIDI file for metric calculation.

IV. Evaluation

We performed multiple classification experiments using artificial neural networks (ANNs), attribute selection, and principal components analysis (PCA). ANN Experiments performed 10-fold cross-validation on the evaluation corpora using Weka's multilayer perceptron model. These experiments were designed to explore the information content of our metrics.

A. 264-Piece, 6-Genre Corpus (Personal Dataset)

We generated a small, yet distinct corpus of well-known musical pieces. This corpus consists of 6 genres (264 pieces): *Classical* (56 pieces by Bach, Beethoven, and Mozart); *Jazz* (51 pieces by Charles Mingus, Charlie Parker, John Coltrane, Miles Davis, Ornette Coleman, and Thelonious Monk), *Nouveau Flamenco* (47 pieces by Ottmar Liebert, Oscar Lopez, Jesse Cook, Sergio Lara, and others); *Greek Rempetika* (41 pieces by the Athenian Company and Markos Vamvakaris); *70s Rock* (35 pieces by Deep Purple, The Doors, Led Zeppelin, Rainbow, and Ten Years After); and *House-Techno* (34 pieces by DJ Dado, Eiffel 65, DJ Keoki, and others).

Using the first 120 seconds of each song, we calculated the 372 power-law metrics. This duration was chosen because it minimizes fatigue for human listeners, yet still captures enough of a piece for its statistical signature to emerge.

We performed Attribute Selection using the Weka toolkit's Chi-squared method to evaluate each individual metric. This sorted our metrics by their informedness (or contribution) relative to the classification task. By iteratively removing the 10 least-informed metrics, and repeating the classification, we reduced the amount of metrics required to maximize the genre classification results to 250. Surprisingly, this subset was more accurate (90.15%) than the original 372 metrics (83.33%). We believe this is due to the heuristic (vs. exhaustive) nature of Weka's attribute selection algorithm.

Finally, using Weka's Principle Components Analysis (PCA), we further reduced the set of necessary features to 46 principal components (linearly uncorrelated variables). So, in summary, 46 features achieve 90.15% classification accuracy. Fig. 1 displays the confusion matrix.

Predicted

a	b	c	d	e	f		
52	0	1	3	0	0	a = classical	Actual
0	42	0	2	2	1	b = flamenco	
0	0	41	0	0	0	c = greek	
4	1	1	40	5	0	d = jazz	
2	0	1	1	30	1	e = rock	
0	0	0	1	0	33	f = techno	

Figure 1. Confusion matrix for 264-piece, 6-genre classification.

The pieces in this corpus are copyrighted and, thus, are not included in the Armonique music database.

B. 2400-Piece, 8-Genre Corpus (Million-Song Dataset)

Using metadata provided with the Million Song Dataset [19], we automatically generated an 8-genre corpus (2400 pieces total, 300 pieces each): *Ambient*, *Blues*, *Classic-Rock*, *Classical*, *Country*, *Hip-hop*, *Jazz*, and *Techno*.

Given that *genre overlap* is a common problem with standard benchmarks (e.g., the ISMIR 2004 genre dataset), we attempted to maximize genre distinction by selecting songs whose genre term (as provided in the metadata):

- had a very high frequency of occurrence (90-100%), and
- included none of the other genre terms (0% frequency of occurrence).

For example, a *Blues* song was selected if the term "blues" was associated more than 90% with this song; and none of the other terms (e.g., "classic-rock", "classical", "country", "hip-hop", "jazz", and "techno") were associated with this song.

For each song, we downloaded the free preview clip from 7digital.com. These MP3 clips usually span 30 seconds (or more), are contiguous, but may come from anywhere in the song.

Using the first 30 seconds of each preview, we calculated the 372 power-law metrics. The ANNs were trained using all 372 (300 MIDI metrics and 72 audio) metrics.

In an 8-genre classification experiment consisting of all 2400 songs (300 per genre), the ANN achieved a success rate of 51.3% (which is high, compared to 12.5% for random selection). The average ROC area value was 82.7%. (The ROC area statistic captures how often, given one positive and one negative example, the ANN will pick correctly.) The ROC area values for each of the genres were as follows: *Ambient* 75.4% - *Blues* 67.8% - *Classical* 96.1% - *Classic-Rock* 77.1% - *Country* 85.1% - *Hip-hop* 88.6% - *Jazz* 81.3% - and *Techno* 89.8%.

Three genres (*Ambient*, *Blues*, and *Classic-Rock*) have a high degree of misclassifications. Upon listening to a few 7digital preview clips, it became clear that some genres (e.g., *Ambient*) are more ambiguous, in this corpus, than other genres (e.g., *Classical*). This is also captured by the ANN confusion matrix (see Fig. 2).

			Predicted					
A	b	c	d	e	f	g	h	
128	13	42	11	27	18	34	27	a = ambient
47	53	24	31	23	36	74	12	b = blues
24	0	259	3	0	0	14	0	c = classical
59	27	15	55	53	33	41	17	d = rock
44	11	13	25	158	14	28	7	e = country
10	5	2	9	7	213	13	41	f = hip hop
26	24	46	14	13	16	155	6	g = jazz
21	2	4	5	6	41	11	210	h = techno

Figure 2. Confusion matrix for 2400-song, 8-genre classification.

We also ran a classification experiment consisting of 1500 songs across five genres, namely *Classical*, *Country*, *Hip-hop*, *Jazz*, and *Techno* (i.e., excluding the three most ambiguous genres). The ANN achieved a success rate of 78.5% (compared to 20.0% for random selection). The average ROC area value was 93.6%. The ROC area values for each of the genres were as follows: *Classical* 97.9% - *Country* 94.2% - *Hip-hop* 92.9% - *Jazz* 89.8% - and *Techno* 93.6%.

Using Attribute Selection and PCA, we reduced the set of necessary features to 104 (vs. 46 in the previous experiment).

C. Playlist Generation - 14K Songs (Magnatune)

Using the PCA results from these two experiments, we created two versions of the Armonique music search engine. Both allow for music searches with MP3 song uploads.

The first demonstration of the automatic playlist generation uses the Magnatune.com corpus. At the time of this writing, it contains 13,829 (~14K) songs. Uploading an MP3 file, as a search query, transcribes the first 120 seconds, computes 372 power-law metrics, and extracts 46 PCA features (as per Section A above). This takes about one minute (including transcription, metric extraction, and results generation). The on-line demo is available at:

http://armonique.org/magnatune

D. Playlist Generation - 10K Songs (7digital)

The second demonstration of the automatic playlist generation uses a subset of the 7digital.com corpus. As with the Magnatune corpus, this subsut is independent of the corpus used to calculate the PCA formula. At the time of this writing, this contains 9221 (~10K) songs. Uploading an MP3 file, as a search query, transcribes the first 30 seconds, computes 372 power-law metrics, and extracts 104 PCA features (as per Section B above). This takes about 30 secs. The on-line demo is available at:

http://armonique.org/7digital

E. Discussion and Results

We invite the reader to search for music similarity within the two corpora above (Magnatune and 7digital). Also, we invite the reader to search by uploading arbitrary MP3 files. However, since our server is not ready yet for high-volume traffic (and since transcription is currently done server-side), we ask the reader to be cognizant of our CPU-cycle limitations. Accordingly, we ask to not publicize these URLs (for now).

Although both demos utilize multi-genre corpora of at least 10K pieces each, clearly, the playlists generated by the Magnatune demo are more genre-cohesive and of much higher quality (in terms of similarity) compared to the 7digital demo. Why?

The Magnatune demo uses 120 seconds of musical material; also, excerpts come from the beginning of a piece. On the other hand, the 7digital demo uses 30 seconds of musical material; excerpts come from an arbitrary place in a piece. We believe this inconsistency in the positioning of excerpts (as seen in 7digital preview clips) is interfering with the quality of the fractal signatures computed for each piece (i.e., the fractal dimensions are not representative and/or comparable). The 7digital preview clips are most likely selected based on some notion of activity (i.e., a high-energy contiguous excerpt in a piece). This acts as a normalizer - and as such, it misses differences in musical structure development starting at the beginning of each piece. This is how human listeners normally compare pieces. Also, the smaller size of 7digital preview clips (30 secs) interferes with capturing fractal dimensions that are representative of a piece. Small excerpts initially appear chaotic; when enough material and musical structure has unfolded, metrics begin to differentiate and converge to power laws.

This conclusion is supported by the ability to reduce the required features in the first experiment from 372 down to 46 (see Section A above), as opposed to 104 in the second experiment (see Section B). In summary, it appears the following parameters are very important in generating high-quality similarity results using fractal dimensions of music (as captured by the presented approach):

- size of musical excerpts (30 secs vs. 120 secs of audio) - the latter is better;
- starting point of musical excerpts (anywhere vs. beginning of the piece) - again, the latter is better.

V. Conclusions and Future Work

We have presented an automated approach to music search and playlist generation based on fractal dimensions of music. This approach utilizes an efficient audio-to-MIDI transcription algorithm, which allows for users to perform real-time, on-line searches by uploading arbitrary MP3 songs.

The generated playlists resemble those of Pandora, last.fm, and iTunes Genius except that they are generated automatically from the fractal dimensions (i.e., similarities in statistical proportions) of musical pieces, as opposed to using human musicologists, capturing user preferences, or harvesting text about music from the web. This allows even for new, undiscovered music to be used as input to (or output from) the system. It also allows for exploring cross-genre similarities that may be missed by human musicologists.

This approach enables development of music recommendation systems and other practical applications involving searching, archiving, classification and metadata extraction of audio content. We are planning to evaluate Armonique's similarity recommendations with human subjects, using the experimental methodology in [1].

On a more theoretical level, these results contribute to the on-going investigation of fractal geometry of music, and of power laws and their applicability to classification of arbitrary sounds, such as music, human and animal vocalizations, and other audio recordings.

Acknowledgments

This work has been funded in part by the US National Science Foundation (NSF grants IIS-0736480, IIS-0849499 and IIS-1049554). Thomas Zalonis has contributed to the development of the transcription algorithm. We also received generous donations of music corpora from the Classical Music Archives (www.classicalarchives.com) and the Music Library of Greece (www.mmb.org.gr). The following have contributed to development of power-law metrics: Thomas Zalonis, Brys Sepulveda, Patrick Roos, Dwight Krehbiel, Luca Pellicoro, Timothy Hirzel, Penousal Machado, and Juan Romero.

References

[1] B. Manaris, P. Roos, D. Krehbiel, T. Zalonis, and J.R. Armstrong, "Zipf's law, power laws and music aesthetics," in T. Li, M. Ogihara, G. Tzanetakis (eds.), *Music Data Mining*, pp. 169-216, CRC Press - Taylor & Francis, 2011.

[2] B. Manaris, D. Krehbiel, P. Roos, T. Zalonis, "Armonique: Experiments in content-based similarity retrieval using power-law melodic and timbre metrics," *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, Philadelphia, PA, pp. 343-348, Sep. 2008.

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

[4] B. Manaris, D. Hughes, and Y. Vassilandonakis, "Monterey Mirror: Combining Markov models, genetic algorithms, and power laws," *Proceedings of 1st Workshop in Evolutionary Music, 2011 IEEE Congress on Evolutionary Computation (CEC 2011)*, New Orleans, LA, pp. 33-40, Jun. 2011.

[5] L. Barrington, R. Oda and G. Lanckriet, "Smarter than genius? Human evaluation of music recommender systems," *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pp. 351-356, Kobe, Japan, Oct. 2009.

[6] D. Turnbull, L. Barrington, D. Torres and G. Lanckriet, "Semantic annotation and retrieval of music and sound effects," *IEEE Transactions on Audio, Speech and Language Processing*, 16(2), pp. 467-476, Feb. 2008.

[7] K. Yoshii and M. Goto, "Continuous pLSI and smoothing techniques for hybrid music recommendation," *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pp. 339-344, Kobe, Japan, Oct. 2009.

[8] J.H. Lee, "How similar is too similar? Exploring users' perceptions of similarity in playlist evaluation," *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami, FL, pp. 109-114, Oct. 2011.

[9] F. Maillet, D. Eck, G. Desjardins and P. Lamere, "Steerable playlist generation by learning song similarity from radio station playlists," *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pp. 345-350, Kobe, Japan, October 26-30, 2009.

[10] A.P. Klapuri: "Automatic music transcription as we know it today," *Journal of New Music Research*, 33(3), pp. 269-282, Sep. 2004.

[11] A. Dessein, A. Cont and G. Lemaitre, "Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence," *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, Utrecht, Netherlands, pp. 489-494, Aug. 2010.

[12] B. Niedermayer, "Non-negative matrix division for the automatic transcription of polyphonic music," *Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, Philadelphia, PA, pp. 544-549, Sep. 2008.

[13] G. Grindlay and D. Ellis, "Multi-voice polyphonic music transcription using eigeninstruments," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, pp. 53-56, Oct. 2009.

[14] J.C. Brown and M.S. Puckette, "An efficient algorithm for the calculation of a constant Q transform," *Journal of the Acoustical Society of America* 92(5), pp. 2698-2701, Nov. 1992.

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

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

[17] K.J. Hsü and A.J. Hsü, "Fractal geometry of music", *Proceedings of the National Academy of Sciences (PNAS)*, 87(3), pp. 938–941, Feb. 1990.

[18] D.J. Levitin, P. Chordia, and V. Menon, "Musical rhythm spectra from Bach to Joplin obey a 1/f power law", *Proceedings of the National Academy of Sciences (PNAS)*, 109(10), pp. 3716-3720, Mar. 2012.

[19] T. Bertin-Mahieux, D.P.W. Ellis, B. Whitman, and P. Lamere, "The Million Song Dataset", *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami, FL, pp. 591-596, Oct. 2011