Computer-based approaches to music research

By: Joe George [1] & Lior Shamir [1]

[1]Dept. Computer Science, Lawrence Technological University, Southfield, MI 48075 USA

The digital information era has had a substantial impact on the way music is consumed and produced, and in a relatively short time has revolutionized the music industry and culture. For instance, music stores have almost disappeared, and have been replaced by digital shopping malls, digital music repositories, and social media networks. The mass digitization of music now offers opportunities to study music in a quantitative fashion, leading to a new sub-field in the digital humanities, enabled by advanced signal and audio processing algorithms that can analyze the music in its digital format. Here we present discovery-driven research methods that can be used to study music by using the mathematical representation of the audio. The methodology allows automatic classification of musical pieces by genres, artists, geographical locations, musical instruments, and more, and the quantitative analysis provides an objective view of sound and musical styles.

Index Terms: Music, Music Research, Digital Humanities, Machine Learning, Machine Perception

I. Introduction

Although Still in its infancy, Digital Humanities carries with it the potential to shed light on many interesting problems pertaining to a variety of fields. Online music repositories have grown exponentially since the birth of the digital music era (the end of the previous century) [1] [4]. The physical record store, the records themselves, as well as other forms of media sold at such outlets (I.E. Magazines) that used to be at the forefront of music consumption culture are all but obsolete in the tangible sense. Now, Online stores and software, present in many digital products such as personal computers, cell phones and smart televisions (I.E. Amazon, I-Tunes, Google Play) offer more music and other digital forms of media than ever before. With so many choices available to the consumer, the need for algorithms and tools that can efficiently classify new media in to its respective domain or genre will prove valuable to both infrastructure development and overall usability. Organizing and managing a dataset such that it can be used to generate accurate suggestions based off of the media's content; rather than its rating, suggestions, or the purchasing patterns of other users will allow users new discovery methods to find new media [4]. The accelerated growth of these digital catalogues and repositories will continue to present a unique challenge for both the end user and online retailer alike. Such tools could be used in a manner that would allay inherent human error as well as biases. A study aiming to provide a taxonomy of whale species based solely on analysis of the whale's call/sound. Such calls are considered to be quite complex, proving difficult to classify [2]. The aforementioned experiment demonstrates that it is possible to use computers to provide quantitative analysis where in the past qualative examination would have been employed. Such tools and methodology may be used for research in a number of topics pertaining to digital humanities. For example to shed light on topics such as an artist(s) influence, for example did the rolling stones influence the Beatles, or vice versa? Or perhaps how original in the past such topics were restricted to the realm of conversation and debate, held more or less as one's subjective opinion.

In recent years, computing algorithms and software have been developed allowing human beings the opportunity to delve deep in to problems long thought to be philosophical in nature rather than scientific. Artificial intelligence engines have proven useful in many situations which time consuming and complex calculations are required. Automated classification methods save lives in the medical industry [7] and provide invaluable processing analytical tools for astronomers [8]. The quantitative analysis of music now becomes possible as the elements of bias, confusion, among other types of human error are reduced, allowing a more objective approach to this type of research. In the pre-information era such analysis would have been a monotonous task, employing the use of people rather than technology and sacrificing accuracy for time. Take into account simply listening to the audio for a hundred songs. This task may take several hours where a computer with modest specifications can complete the process in seconds.

Furthermore, the introduction of a confusion matrix allows a machine learning based approach where songs that sound similar to one another are interpreted as such, and weighted in a method allowing the system to differentiate between tracks with no human interaction or definition of perceived genre. The development of such tools will allow the practical ability to solve problems in the realm of human perception and interpretation, which cannot be measured on a finite scale or measured in a domain. Other possible uses include a rating system for the RIAA and MPAA, plagiarism detection, navigating large repositories and databases, biometric security. One could go as far as determining the originality, or influence or a particular artist. The age old question of what constitutes plagiarism aside from inspiration is one that will continue to be debated. Implementations of algorithms like the one described in the following sections will hopefully be able to distinguish one another quantitatively. Establishing a means with which to define boundaries and detect similarities based on computer analysis and mathematical expression.

II. Methods and Result

A study investigating how to classify species of whale based on audio analysis of its call used a similar method and was able to successfully classify each of the whale species [2]. Although more complex than whale calls, music and audio signals have several ways to be described, including harmonicity, pitch, timbre, key, and more [3]. Such a complicated task such as the classification of digital audio requires the exclusion of the subjective nature of the human perception and the use of complex artificial intelligence and Related computing algorithms.

The following is one such method, it utilizes open source code and software made available to any and all who request it, and can be downloaded and executed on several operating systems including Microsoft Windows and various Linux flavors. Our proposed classification method initiates with normalizing a dataset of audio files in MP3 format (FLAC files were originally used however the classification accuracy did not increase given the relatively large file size) to a uniform length such that SOX [10] the software providing the spectrogram (Figure 1) can first sample the audio at any given interval then down mixed from stereo MP3 to Mono MP3. The spectrograms were then organized preserving the original folder structure where each genre represented has an associated folder.

Contained in each folder are songs from multiple artists which correspond to their respective genre. The artists and music chosen were in accordance with critics in the industry as well as audio engineers alike. The data contained within the spectrogram is used to classify the sample quantitatively based on numerous two-dimensional descriptor variables that represent the characteristics of the audio sample [5]. The frequency response of the overall song may be able to be broken down and analyzed on a per instrument basis. Below are the spectrograms and respective instruments for the song Sergeant Pepper's lonely hearts club band by the Beatles, along with brief summaries of what is expressed within them as no scale or key is generated along with the spectrograms.
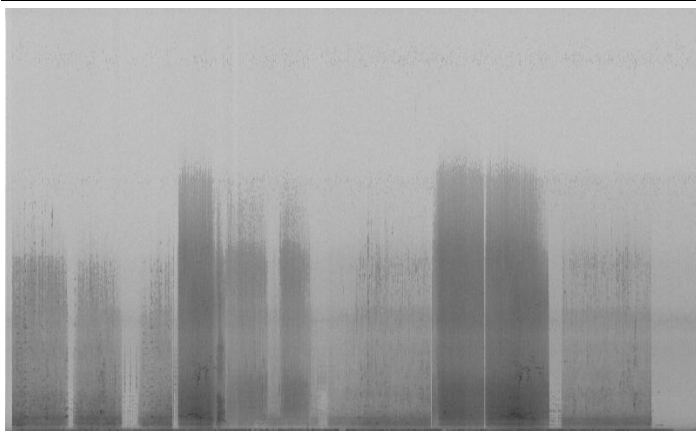

Track 1: In this track all we have is the audience's applause and background noise used as an effect throughout the entire song, Darker shades of grey represent an increase in
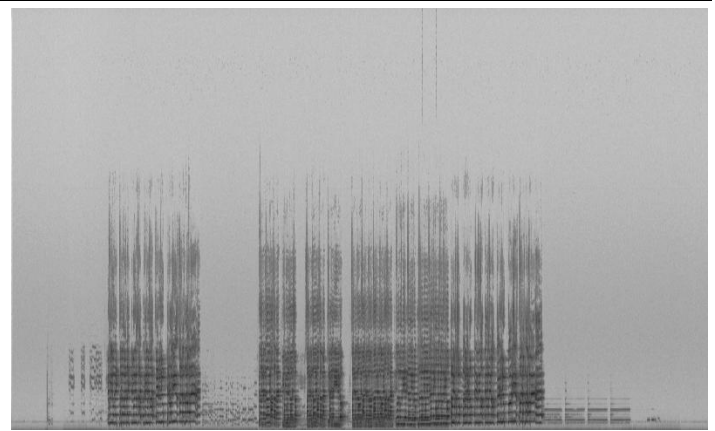
Track 2: This track is comprised of the drum, bass, some but very little backing guitar. The bass is well defined towards the bottom of the spectrogram while The higher frequencies represent some of the drums, and cymbals
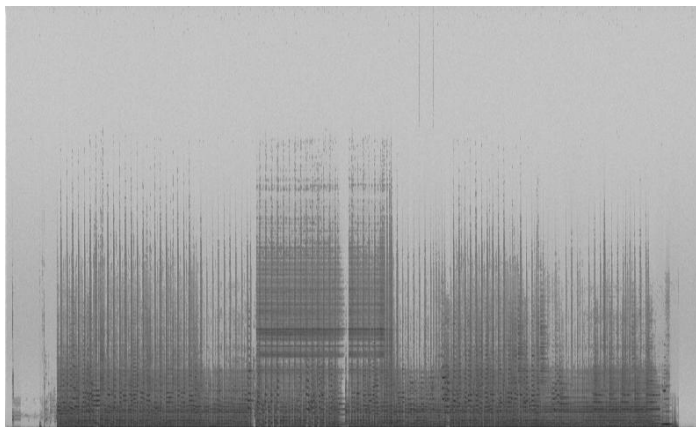
Track 3: a vocal track, things like vibrato as well as layered harmonies which are comprised of two or multiple tracks combined.
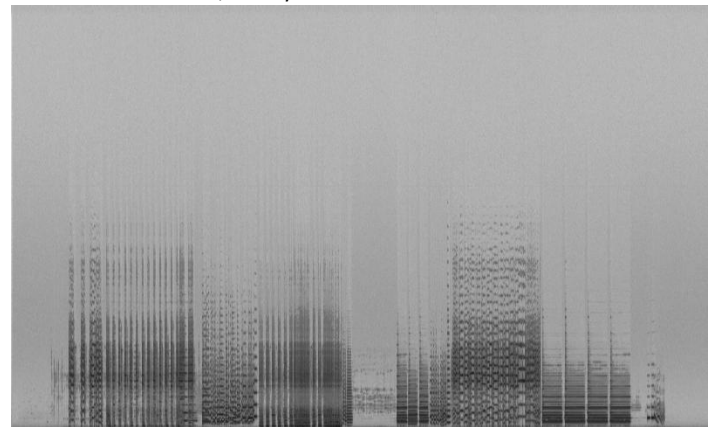
Track 4: The guitar track is comprised of thin slices followed by a brief pause before the subsequent strum or picking of a note. Unlike the longer vertical grey "slices" depicted

These descriptors are not equally informative and consequently are weighted according to their informativeness. Using PHYLIP, software used for visualizing evolutionary progression of species [6], a tree like graph referred to as a phenology is

utilized that depicts the classified data set organized by genre. In an experiment which aimed to differentiate and classify musical content based on genre showed that a total of seven genres including: Rap, Rock, Country, Classical, Rockabilly, and

Modern Alternative, containing 47 samples per genre were able to be classified in a manner that for the most part is in agreement with critics in the music industry. Figure 2 shows the

	Alt	Classical	Country	Jazz	Rap	Rock	Swing
Alt	16	2	3	4	8	34	3
Classical	0	70	0	0	0	0	0
Country	1	4	48	2	1	5	9
Jazz	3	0	5	44	5	5	8
Rap	3	0	0	0	55	12	0
Rock	9	4	5	1	0	41	10
Swing	0	7	4	4	2	2	51

output of the classification experiment using the spectrograms generated using 30 second mono samples. This is only one output of several as the experiment was executed multiple times analyzing different parts of the spectrogram with each iteration. Song descriptor values are then weighted and the sample is compared to the confusion matrix to determine its relative genre. A confusion matrix, similar to a human beings own though process which is not always sure where a particular song may lie in terms of its genre allows a computer processing the data to determine the likelihood that a given song belongs to its respective genre, taking into account its relationship to the other genres (I.E. Eminem songs can be associated with Snoop Dog, who shares the same genre and similar musical style).

This is especially true with new music due to the blurring of the lines so to speak within genres and as sub genres continue to become ever present. The results depicted in figure one show that such a method is able to correctly classify a given a song to its genre showing a clear differentiation between nodes indicating the computer was able to train, learn, and classify musical styles automatically with no human interjection.
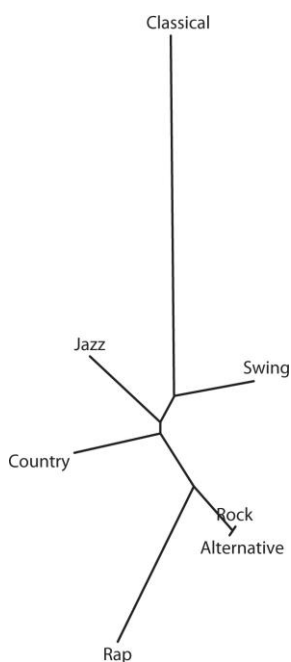
It is interesting note that the algorithm placed rap far from classical identifying rock and roll and alternative rock were closely related although clearly indicating that alternative is derived from rock as this holds true among music critics [9].

III. Conclusion

The outlined methodology will open new doors in the field of digital humanities, as there are a number of uses for algorithms such as the one described. It should be noted that the above algorithm may be applied to other forms of digital media. More experimentation will prove valuable in the years to come as the need for complex tasks, such as music classification continue to emerge, and as more areas of study invoke the use of artificial intelligence and related computer algorithms. The experiment described in section two, demonstrates the outlined method is accurate and more importantly that the conceptual principals that lie at its core will prove useful to an array of research topics. The continued research and development of classification methods will prove valuable as digital media will soon secure its role as the dominant force in the printing, music, and movie industries.

IV. Acknowledgements

V. References

[1] Martin J. Halvey and Mark T. Keane. 2007. Exploring social dynamics in online media sharing. In *Proceedings of the 16th international conference on World Wide Web* (WWW '07). ACM, New York, NY, USA, 1273-1274. DOI=10.1145/1242572.1242804 http://doi.acm.org/10.1145/1242572.1242804

[2] Shamir, L., Yerby, C., Simpson, R., von Benda-Beckmann, A., Tyack, P., Samarra, F., Miller, P., Wallin , J. "Classification of large acoustic datasets using machine learning and crowdsourcing - application to whale call", *Journal of the Acoustical Society of America*, In Press.

[3] Wold, E., T. Blum, D. Kaislar, and J. Wheaton. "Content-based Classification, Search, and Retrieval of Audio." *IEEE Multimedia* 3.3 (1996): 27-36. Print.

[4] Liu, M., C. Wan, and L. Wang. "Content-based Audio Classification and Retrieval Using a Fuzzy Logic System: Towards Multimedia Search Engines." *Soft Computing - A Fusion of Foundations, Methodologies and Applications* 6.5 (2002): 357-64. Print.

[5] Altes, Richard A. "Detection, Estimation, and Classification with Spectrograms." *The Journal of the Acoustical Society of America* 67.4 (1980): 1232. Print.

[6] J. Felsenstein. 2004. PHYLIP Phylogeny Inference Package, Version 36.

[7] Statnikov, A. "A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis." *Bioinformatics* 21.5 (2004): 631-43. Print.

[8] Zhang, Y., and Y. Zhao. "Automated Clustering Algorithms for Classification Of astronomical Objects." *Astronomy and Astrophysics* 422.3 (2004): 1113-121. Print.

[9] McDonaldd, Chris. "Exploring Modal Subversions in Alternative Music." *Popular Music* 19.03 (2000): Print.

[10] SourceForge. Sox: Sound Exchange. 2012. [Online] Available: http://sox.sourceforge