

CHARACTERIZATION OF MOVIE GENRE BASED ON MUSIC SCORE

Aida Austin, Elliot Moore II, Udit Gupta

Georgia Institute of Technology
School of Electrical and Computer Engineering
210 Technology Circle, Savannah, GA, 31407

Parag Chordia

Georgia Institute of Technology
Department of Music
840 McMillan St, Atlanta GA 30332

ABSTRACT

While it is clear that the full emotional effect of a movie scene is carried through the successful interpretation of audio and visual information, music still carries a significant impact for interpretation of the director's intent and style. The intent of this study was to provide a preliminary understanding on a new database for the impact of timbral and select rhythm features in characterizing the differences among movie genres based on their film scores. For this study, a database of film scores from 98 movies was collected containing instrumental (non-vocal) music from 25 romance, 25 drama, 23 horror, and 25 action movies. Both pair-wise genre classification and classification with all four genres was performed using support vector machines (SVM) in a ten-fold cross-validation test. The results of the study support the notion that high intensity movies (i.e., Action and Horror) have musical cues that are measurably different from the musical scores for movies with more measured expressions of emotion (i.e., Drama and Romance).

Index Terms— Music Genre Classification, Movie Genre Classification, Music Emotion Analysis, Music Information Retrieval

1. INTRODUCTION

Movies are designed to evoke emotion through the careful manipulation of audio and visual imagery. On its own, music carries an inherent sense of mood whose interpretation is generally left to the listener. However, in film scores, music is designed to target specific moods that enhance the intent of the director for a particular set of scenes. For example, a horror scene will likely carry music with dark overtones intended to convey a sense of suspense or fear while a romantic scene may utilize sweeping movements and shifts in a light tone that carries a sense of longing or love. A natural approach of research in analyzing film content has been to utilize the combination of audio-visual cues for analysis [1]. In some cases, audio (e.g., high energy profiles, gunfire, explosions, etc.) has

been used as a key to detect certain types of emotional events [2] in visual scenes.

While it is clear that the full effect of an emotional scene is carried through the successful interpretation of audio and visual information, music still carries a significant impact for interpretation of the director's intent and style. From a computational standpoint, it is desirable to find objective measures that can be used to automatically determine mood and intent for user-oriented applications related to music retrieval [3, 4], creation [5, 6], and general emotion analysis [7-12].

A significant challenge for every computational approach is the determination of a domain of features that carry the necessary information for analysis and modeling. In [13], Tzanetakis and Cook defined timbral, rhythm, and pitch feature sets for music genre classification. Some timbral features utilized included spectral centroid, spectral rolloff, spectral flux, mel frequency cepstral coefficients (MFCC), and time domain zero crossings. The rhythm and pitch features used were extracted from beat and pitch histograms. From these feature sets, a classification accuracy of 61% was achieved for ten musical genres (e.g., jazz, country, etc.). In [14], long-term modulation spectral analysis of spectral and cepstral features was applied to music from the ISMIR2004 GENRE and GTZAN music databases. GTZAN, the same database that was compiled in [13], contains 10 genres, each consisting of 100 tracks of 22050 Hz, mono, 30 second, 16 bit audio. The ISMIR2004 GENRE database consists of a 729 audio track training set and a 700 track testing set. These two sets each consist of six genres (e.g., classical, electronic, jazz/blues, etc.). The GTZAN database resulted in a classification accuracy of 90.6%, while the ISMIR2004 GENRE achieved an accuracy of 86.83%.

In [4], it is suggested the most meaningful characterization of music is by *genre*, *emotion*, *style*, and *similarity*. For this preliminary study, concepts of *genre* and *emotion* are combined and represented by overall descriptors for the type of movie genres under consideration. Four movie genres are considered in this study including: Action (characterized by emotions that are high intensity and invoke excitement), Horror (characterized by emotions instilling a sense of anxiety and fear), Romance (characterized by emotions instilling a sense of longing and

love), and Drama (characterized by a wide variety of emotions from sadness to joy). The purpose of this study is to characterize the measurable traits of the musical scores utilized in these four movie types in an effort to determine what feature categories may carry the most information in distinguishing between them in a broad sense.

2. DATABASE

For this study, a database of film scores from 98 movies was collected. The database contains instrumental (non-vocal) music from 25 romance, 25 drama, 23 horror, and 25 action movies. The primary genre labeling for each film is based on information from the Internet Movie Database (IMDB) (www.imdb.com). Table 1 shows the breakdown of the content of the music tracks. The MP3 files for each music track were purchased, stored, and converted to uncompressed mono WAV files at a sampling rate of 16 kHz with 16-bit samples. Overall, 1728 audio tracks were stored and the middle 60 seconds of each track was isolated for feature extraction and analysis.

Table 1: Music Database Content

Genre	No. of Movies	No. of Tracks
Romance	25	441
Drama	25	458
Horror	23	406
Action	25	423

3. FEATURE EXTRACTION

The use of timbral and rhythm features has proven effective in MIR studies on music genre classification. In this work, these two groups of features are applied for genre classification based on film score feature analysis.

The bulk of the feature extraction was accomplished using JAudio [15], which was designed with the intent of providing a standardized set of features for MIR. All JAudio extraction is performed using windows of length 512 samples with no overlap. JAudio was used to extract a base set of 16 distinct timbral categories along with their 1st and 2nd order statistics resulting in a feature matrix of 216 features.

The timbral features extracted with JAudio are the standard features used in music genre classification and emotion detection, as in [7], [13], and [16]. Timbre is used to define the quality of a sound. It is the way to account for the difference in tone between, for example, a flute and a clarinet. In Table 2, a list of the 16 timbral categories used is shown.

These features are organized into five categories as shown in Table 2. These categories represent subsets of timbral features (Spectral Non-LPC/Non-MFCC, Non-Spectral, MFCC, and LPC) and rhythm features.

Table 2: Features Extracted

Domain	Category	No. of Features
(Timbral) Spectral Non-LPC /Non-MFCC		
Timbral	Spectral Centroid	4
Timbral	Spectral Rolloff Point	4
Timbral	Spectral Flux	4
Timbral	Compactness	4
Timbral	Spectral Variability	4
Timbral	Strongest Frequency via FFT max	4
Timbral	Strongest Frequency via Spectral Centroid	4
Timbral	Method of Moments on Spectrogram	20
Timbral	Area Method of Moments on Spectrogram	40
Timbral	Peak Based Spectral Smoothness	4
(Timbral) Non-Spectral		
Timbral	Root Mean Square of Frames	4
Timbral	Fraction of Low Energy Windows	4
Timbral	Relative Difference Function	4
Timbral	Zero Crossings	8
(Timbral) MFCC		
Timbral	MFCC (first 13 coefficients)	52
(Timbral) LPC		
Timbral	LPC (first 10 coefficients)	40
Rhythm		
Rhythm	Tempo	3
Rhythm	Strongest Beat	6
Rhythm	Beat Sum	6
Rhythm	Pulse Clarity	3

The four rhythm features used are extracted using JAudio, the MIRtoolbox [17], and an autocorrelation based approach from [18]. The JAudio features are calculated using a beat histogram, as in [13]. From this histogram, beat sum, strongest beat, and strength of strongest beat are calculated. Pulse clarity is calculated using the MIRtoolbox. This toolbox contains a set of MATLAB based functions for MIR analysis. Pulse clarity is a feature used to determine how well the beat of the music can be detected [19]. Tempo is extracted using the autocorrelation based approach from [18]. The mode of the tempo vector calculated over 60s was accepted as the tempo feature.

The final feature vector for each of the 1728 audio tracks contained 222 features.

Table 3: Confusion Matrices Using Different Feature Sets (D=Drama; H=Horror; A=Action; R=Romance)

	A	D	R	H
A	63.8%	14.2%	8.0%	13.9%
D	18.4%	55.3%	24.6%	9.9%
R	8.7%	28.8%	58.9%	7.8%
H	21.7%	16.5%	14.4%	43.3%

(Timbral) All

	A	D	R	H
A	58.6%	20.8%	9.5%	11.1%
D	25.5%	40.9%	32.6%	9.2%
R	11.6%	29.3%	53.9%	9.5%
H	29.1%	19.6%	25.1%	22.2%

(Timbral) LPC

	A	D	R	H
A	53.2%	18.9%	11.6%	16.3%
D	26.7%	47.3%	24.3%	9.9%
R	11.8%	30.0%	56.3%	6.1%
H	28.6%	15.8%	18.0%	33.6%

(Timbral) MFCC

	A	D	R	H
A	49.6%	24.6%	10.4%	15.4%
D	15.8%	53.0%	27.2%	12.3%
R	5.4%	38.3%	55.6%	5.0%
H	22.5%	21.0%	18.0%	34.5%

(Timbral) Spectral (non-lpc/non-mfcc)

	A	D	R	H
A	40.7%	35.9%	15.4%	8.0%
D	22.7%	57.2%	21.5%	6.9%
R	14.9%	41.4%	41.6%	6.4%
H	15.8%	39.0%	23.6%	17.5%

(Timbral) Non-spectral

	A	D	R	H
A	35.5%	14.9%	27.0%	22.7%
D	13.2%	29.8%	45.9%	19.4%
R	11.3%	26.5%	44.0%	22.5%
H	11.1%	13.9%	30.5%	40.4%

Rhythm

	A	D	R	H
A	60.0%	15.1%	9.0%	15.8%
D	15.1%	58.2%	24.3%	10.6%
R	9.5%	27.2%	58.6%	9.0%
H	20.3%	18.0%	14.2%	43.5%

All Features

4. RESULTS

The intent of this study was to provide a preliminary understanding on a new database for the impact of timbral and select rhythm features in characterizing the differences among movie genres based on their film scores. Several classification tests were conducted with support vector machines (SVM) and 10-fold cross-validation using the WEKA [20] toolbox. Table 3 shows confusion matrices generated by comparing all four genres utilizing different sets of Timbral and Rhythm features. Overall, the Rhythm features performed the worst among the classification tasks while Timbral features provided better genre characterization. Within the timbral domain, LPC, MFCC, and Non-Spectral features provided the highest genre accuracies for Action (58.6%), Romance (56.3%), and Drama (57.2%), respectively. The Horror genre did not perform as well as the other genres in terms of classification being most often confused with Anger and Drama for timbral features and with Romance for rhythm features. Combining all of the features did not yield a substantial improvement over the use of Timbral features alone, although this is not surprising given the number of timbral versus rhythm features used in the study at this time.

Table 4 investigates pair-wise comparisons of the genres using all of the features from the study (Table 1) in an effort to see which genres are most distinguishable from one another one-on-one. The table shows the highest accuracy for distinguishing Romance vs. Action music scores (81.13%) and the lowest accuracy for distinguishing Drama vs. Romance (65.74%). Additionally, the classification of Drama vs. Action (77.75%) and Horror vs. Romance (77.57%) achieved comparably high accuracies. Table 4 supports the notion that high intensity movies (i.e.,

Action and Horror) have musical cues that are measurably different from the musical scores for movies with more measured expressions of emotion (i.e., Drama and Romance).

Table 4: Results of Classification (D=Drama; H=Horror; A=Action; R=Romance)

Genres	Instances	Correct %	Chance %
D vs. H	864	73.72	53
D vs. R	899	65.74	50.9
D vs. A	881	77.75	52
H vs. R	847	77.57	52.1
H vs. A	829	70.57	51.0
R vs. A	864	81.13	51.0
All	1728	53.94	26.5

5. CONCLUSION

This study presents a preliminary examination on a new database of music collected from film scores in four genres (Action, Romance, Horror, and Drama) utilizing timbral and a select set of rhythm features. Initial results suggest that the music from Action genres is the most clearly distinguishable (particularly from Drama and Romance) with Drama and Romance being the least distinct. For the purposes of conducting a preliminary analysis, all of the music tracks within a single film genre were broadly labeled by the movie genre. However, it is clear that such a labeling scheme is likely too broad as several tracks within a specific genre may exhibit characteristics of music from another genre (e.g., an Action sequence of music in a Drama movie or vice versa). Future work will involve a closer examination of each track to determine the most appropriate groupings of the data and should serve to improve

classification accuracy. Additionally, while rhythm features were considered on a small scale, this study was dominated by timbral features. Future work will seek to broaden the domain of music features to include more rhythm measures and other features such as chroma.

6. REFERENCES

- [1] A. Hanjalic, "Extracting moods from pictures and sounds: towards truly personalized TV," *Signal Processing Magazine, IEEE*, vol. 23, pp. 90-100, 2006.
- [2] M. Xu, L. Chia, and J. Jin, "Affective content analysis in comedy and horror videos by audio emotional event detection," in *IEEE International Conference on Multimedia and Expo*, 2005.
- [3] L. Tao and M. Ogihara, "Content-based music similarity search and emotion detection," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on, 2004*, pp. V-705-8 vol.5.
- [4] L. Tao and M. Ogihara, "Toward intelligent music information retrieval," *Multimedia, IEEE Transactions on*, vol. 8, pp. 564-574, 2006.
- [5] J. Phalip, M. Morphet, and E. Edmonds, "Alleviating communication challenges in film scoring: an interaction design approach," in *Proceedings, CHISIG Computer Human Interaction: Design: activities, artifacts and environments Adelaide, Australia: ACM*, 2007, pp. 9-16.
- [6] <http://www.sonycreativesoftware.com/Cinescore>
- [7] T. Li and M. Ogihara, "Detecting Emotion in Music," in *Proceedings of the International Symposium on Music Information Retrieval*, 2003, pp. 239-240.
- [8] J. Berg and J. Wingstedt, "Relations between selected musical parameters and expressed emotions - extending the potential of computer entertainment," in *Proceedings, ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, 2005, pp. 164-171.
- [9] L. Lie, D. Liu, and Z. Hong-Jiang, "Automatic mood detection and tracking of music audio signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 5-18, 2006.
- [10] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, "Multi-Label Classification of Music into Emotions," in *International Conference on Music Information Retrieval Philadelphia, USA*, 2008, pp. 325-330.
- [11] F. Yazhong, Z. Yueting, and P. Yunhe, "Music information retrieval by detecting mood via computational media aesthetics," in *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*, 2003, pp. 235-241.
- [12] F. Kuo, M. Chiang, M. Shan, and S. Lee, "Emotion-based music recommendation by association discovery from film music," in *Proceedings, ACM International Conference on Multimedia, Hilton, Singapore*, 2005, pp. 507-510.
- [13] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *Speech and Audio Processing, IEEE Transactions on*, vol. 10, pp. 293-302, 2002.
- [14] L. Chang-Hsing, S. Jau-Ling, Y. Kun-Ming, and L. Hwai-San, "Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features," *Multimedia, IEEE Transactions on*, vol. 11, pp. 670-682, 2009.
- [15] D. McEnnis, C. McKay, I. Fujinaga, and P. Depalle, "jAudio: A Feature Extraction Library," in *Proceedings of the International Conference on Music Information Retrieval*, 2005, pp. 600-603.
- [16] C. N. Silla, A. L. Koerich, and C. Kaestner, "Feature Selection in Automatic Music Genre Classification," in *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on*, 2008, pp. 39-44.
- [17] O. Lartillot and P. Taivaine, "MIR in Matlab (II): A Toolbox for Musical Feature Extraction from Audio.," in *International Conference on Music Retrieval Vienna*, 2007.
- [18] <http://labrosa.ee.columbia.edu/>
- [19] O. Lartillot, T. Eerola, P. Toivainen, and J. Fornari, "Multi-feature Modeling of Pulse Clarity: Design, Validation and Optimization," in *International Conference on Music Information Retrieval Philadelphia, USA*, 2008, pp. 521-526.
- [20] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd Edition ed. San Francisco: Morgan Kaufmann, 2005.