Enhanced Polyphonic Music Genre Classification Using High Level Features

Arash Foroughmand Arabi [#1], Guojun Lu [#2]
[#] *Faculty of Information Technology*
Monash University, Melbourne, Australia
[1] arabi.arash@gmail.com
[2] guojun.lu@infotech.monash.edu.au

***Abstract*— The task of classifying the genre of polyphonic music signals is traditionally done using only low level features of the signal. In this paper high level features have been applied to improve the task of music genre classification. The use of statistical chord features and chord progression information in conjunction with low level features are proposed in this paper. The chord progression information is manifested in genre probability descriptors calculated using a pattern matching algorithm. Our proposed method provides an improvement of 12.4% in the classification results over a commonly compared technique.**

***Keywords*— chord progressions, music genre classification, music signal processing, high level features, chord features**

I. INTRODUCTION

In the past 10 years a lot of attention has been paid to automatic content based music indexing and retrieval. Particularly there has been a lot of research in the task of automatic music genre classification. Many researchers have worked on this problem and a lot of novel systems have been developed.

Perhaps the most significant and influential work in automatic music genre classification of polyphonic signals is the work of Tzanetakis and Cook [1].The techniques that they proposed served as the basis of many researches and have been widely cited throughout the literature. According to Google Scholar [1] has been cited over 800 times. Since there is an open source software (Marsyas [1]) implemented based on this technique, it has turned to a kind of benchmark in the subject area of automatic polyphonic music genre classification. The latest implementation of this technique [2] won the MIREX [2] 2008 genre classification task. In this paper we propose improvements over [2] and provide experimental results and analysis.

Our proposed improvement to this technique is based on incorporating high level features. Particularly we focus on chord and chord progression features.

Although there have been many newer techniques than those of [2], because of the following reasons we have selected [2] as the baseline of our system. Firstly, achieving the highest average classification accuracy in the 2008 MIREX genre classification contest, this method is still one of the best performing methods. Secondly as mentioned before an open source software has been implemented by the authors giving us the chance of examining the exact setup that was described in their paper. And finally since this method is used as a benchmark by many researchers in the community it will be a more reliable comparison point both for us and other researchers who want to compare their results with ours.

The remainder of the paper is organized as follows. Review of the most significant related works is presented in Section II and their limitations identified. Then a system is proposed to overcome the limitations of the previous works in Section III. In Section III we also describe the features used in our system in details. We then provide detailed experimental results in Section IV. Finally Section V concludes the paper.

II. RELATED WORK

The basis for the method proposed by [1] is [3]. Although cited numerous times, this paper [3] never existed in print. In 2008 the authors of [3] published a revised version of their paper [4]. They showed that humans are able to detect the genre of the music by listening to very short excerpts of music. From this fact [1] concluded that low level information which define the timbral texture of the sound such as MFCC, Spectral Centroid, Spectral Rolloff, Spectral Flux, and Zero Crossing are enough for classifying music genre because high level features cannot be realized in such short timeframes.

Since then many novel techniques have been proposed to improve the genre classification accuracy. And what they mostly have in common is that they mainly utilize low level features. Here we describe the most significant achievements in this discipline.

Panagakis et al. [5] used a multilinear approach in genre classification. They used tensors in representing spectro-temporal modulation features extracted from the music and used these tensors to classify the songs. Barbedo and Lopes [6] defined set of very wide and deep hierarchical genre taxonomies so a very meticulous comparison can be done between the genres. Their feature set included spectral roll-off, loudness, frequency bandwidth, and spectral flux. Holzapfel and Stylianou [7] proposed the use of Nonnegative matrix factorization (NMF) to describe the

[1] marsyas.sness.net
[2] Music Information Retrieval Evaluation eXchange www.music-ir.org/mirex/2008

978-1-4244-5561-4/09/$26.00 ©2009

timbre of the music and they achieved significant improvements compared to MFCC based models.

To overcome the limitations of mean and standard deviation in summarizing the feature data Meng et al. [8] proposed a multivariate autoregressive feature integration Scheme to integrate temporal features (short-time features, medium-time features, and ling-time features). As a result two new feature sets of DAR (Diagonal Autoregressive) and MAR (Multivariate Autoregressive) were created. Their experiment resulted in significance improvement of classification results compared to the traditional mean/variance method.

Lidy et al. [9] combined audio and symbolic descriptors to classify the songs into genres. In their method they first extracted Rhythm, timbral, and onset features from the songs. Then they converted the songs to MIDI (Musical Instrument Digital Interface) files considering the pitch and duration of the notes, and extracted harmonic, rhythm, note duration, silence duration, pitch, and onset based features from the MIDI files. Lampropoulos et al. [10] separated the musical instrument sources in the signal first and then extract the features from each individual separated source. They used the original feature set of [1].

All the studies outlined above only make use of low level features as deduced from [4]. This may be partly because of the lack of psychoacoustic research in this area. However we have sufficient motivation to experiment with high level features. Here we discuss these reasons.

Chase [11] investigated the ability of Koi (*Cyprinus carpio*) fish to discriminate musical genre. And the results indicate that Koi can discriminate music by their genres. As a result of their study compared to other studies they concluded that animals use both local (low level) features and pattern (high level) features to classify music. However they prefer local features instinctively because features that are instantly recognizable grant a survival advantage.

There has been very limited psychology and psychoacoustic research on how humans perceive and classify music genre. Unlike music emotion classification the discipline of music genre classification really lacks the psychological and musicological basis. The only works known to us in this discipline are [4] and [11].

There has been a few works experimenting with high level features in automatic genre classification observing improvements in the results. This does not mean that [4] is wrong, it means that beside timbral features there are many other factors that affect the music genre and they should not be ignored.

McKay [12] suggested that [4] does not imply that high level musical features should be ignored. The musical form and structure also play an important role in music classification but they are not the basic essential feature used for classifying music genres. They [12] also concluded that based on the works of Tekman and Hortacsu [13], music genres are strongly related to music emotion. There have been many studies in psychoacoustics and music emotion classification which suggest that the high level music features play an important role in music emotion classification [14-16]. So perhaps in the case of genre classification high level features also play a role.

McKay and Fujinaga [17] suggested ways to improve the current automatic genre classification techniques. One of their suggestions was that information from low level, high level and cultural features must be combined. They stated that since currently most genre classification techniques only incorporate low level timbre based features their performance have been limited. Then they argued that timbre only represents a small portion of what humans use for classifying music genre and high level features are central to music performers and composers in performing and composing in different music genres [17].

Perhaps Zhu et al. [18] were one of the few who have truly used high level features for music genre classification. They used features such as the instrument distribution and instrument based notes features to do genre classification and observed significant improvement of 11% in the accuracy when compared to MFCC and energy based features.

In this paper we derive high level features from chords and combine them with low level features for genre classification. When preparing this paper, we noted a paper published recently which also uses high level features based on chords [19]. They extracted chord progression features from MIDI files and used them for genre classification. To classify music genre based on chord progression features they proposed n-grams approach which is similar to our proposed approach.

In our system we propose using a combination of low level and high level features including chord progressions but [19] only concentrates on using chord progression features. Moreover, we extract all the features directly from the music signal (i.e. mp3 and au formats) rather than MIDI music.

III. PROPOSED SYSTEM

To overcome the limitations of the current methods for genre classification of polyphonic music signal we propose using high level features along with low level timbral features. In particular we concentrate on chord and chord progression features. Figure 1 illustrates our proposed system.

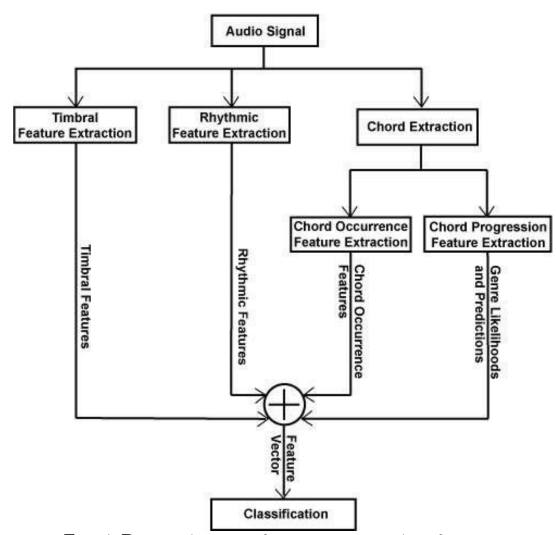

Fig. 1. Proposed system for music genre classification

According to Cambridge dictionary chord is defined as "Three or more musical notes played at the same time". And a chord progression is defined as a number of chords played consecutively.

Our rational and motivation behind using chords come from the works of Paiement et al. [20]. They demonstrated that chord progressions capture high level and sophisticated harmony organizations, which we believe is a good source of information for genre classification. Also Lee [21] used genre information to extract chords of a song more accurately. This shows a relationship between chords and genres and indicates that chord information can be a candidate for increasing the accuracy of genre classification.

In our proposed method we introduce ways to extract useful information from chord and chord progressions and ways to organize these information in a way that can be used in conjunction with other features as an input to machine learning algorithms. We also experiment with rhythmic features, although rhythmic features have been used before [1, 8-10].

For the task of classification we first experiment with 7 different data mining algorithms belonging to different categories (multilayerPerceptron, SMO, RBFNetwork, IBK, Bagging, BayesNet, and Logistics using WEKA) and select the best 2 methods. Beside these 2 algorithms we also perform classification using Kea's SVM[3] which is a specialised algorithm for music classification.

In the following sections we describe the three high level features used in the proposed system: rhythmic features, chord occurrence features and chord progression features.

A. *Rhythmic Features*

Rhythmic features are related to music tempo (beat rate). We extracted beat histogram features using Marsyas. The beat histogram is extracted using an algorithm based on Discrete Wavelet Transform. For more details on this algorithm refer to [1]. To capture the rhythmic content of the song from the beat histogram we calculate the low peak amplitude, low peak BPM (Beats per Minute), high peak amplitude, high peak BPM, high low ratio and 3 sum values representing the sum of energy in 3 different BPM bands. For example if the first sum is higher than the other two it means there is more energy in slower tempos.

B. *Chord Occurrence Features*

For extracting the song chords we have used the chord extracting algorithm of Harte and Sandler [22]. This algorithm is implemented in ChordExtractor which is a part of Clam Annotator[4] software. This algorithm extracts the chords using a Chromagram. Harte and Sandler exploited the psychological theory that it is possible to model the human perception of pitch using a helix [23]. In their initial experiments they could achieve an average of 62.4% accuracy in chord extraction.

The chords are identified using the root which is one of the 12 notes in a chromatic scale and mode (type) which can be Augmented, Diminished, Diminished7, Dominant7, Major, Major7, Minor, Minor7, or MinorMajor7. So in total 108 different chords can be represented.

The number of chords extracted from each song depends on the content of the song (i.e. how many chords are played in that song). Some songs may have very few chords while some other may have a lot chord changes.

For the purpose of classification we construct a 21-dimensional feature vector which contains statistical information on the chords. It consists of 12 dimensions corresponding to the 12 roots and 9 dimensions corresponding to the 9 modes. This vector identifies the number of occurrences of each root and each mode in a song. This is done by simply counting the number of occurrences of roots and modes in a song.

In our initial experiments we have also tested other statistical information such as the statistical mode of the roots, modes, and chords but they did not show any improvements in the results.

C. *Chord Progression Features*

A chord progression is a sequence of chords played in a row. The length and the number of chord progressions in music pieces vary. Since musicians often break or modify the musical clichés to produce original musical pieces, the task of incorporating musical knowledge of chord progressions into genre classification is very hard and necessitates an extremely wide problem space.

To simplify this problem a pattern matching algorithm is proposed to estimate the likelihood of songs belonging to each genre from the chord patterns within the songs. So the number of dimensions of the feature vector will be the same as the number of genres present in the training set. In other words each dimension of the vector represents the likelihood of one genre. This algorithm is inspired by the way that data mining algorithms train and test a model therefore we name it ChordMiner.

In the training phase ChordMiner builds a model for each genre in the dataset. This model consists of the chord patterns that appear in that genre and how often they appear compared to the other chord patterns. ChordMiner trains and tests the model based on a 10 fold cross validation approach. In each of the iterations 90% of the songs in each genre are selected. Then for all the songs in each genre all the patterns of 3 chords and their occurrence frequency is recorded. Then for all the chords where their frequency is larger than 1 a probability descriptor P_g is calculated:

$$P_g(c) = \frac{f_g(c)}{n_g}$$

Where $P_g(c)$ is the probability descriptor of chord pattern c appearing in genre g. And $f_g(c)$ is the number of times that chord pattern c appears in the songs of the genre g. And n_g is the number of unique chord patterns within the songs of the genre g. This process is repeated for all the genres within the dataset.

In the testing stage ChordMiner selects the remaining songs and estimates their likelihood for belonging to different genres.

$$L_g(s) = \sum P_g(c)$$

Where $L_g(s)$ is the likelihood of song s belonging to genre g. And $P_g(c)$ is the probability descriptor of a given chord pattern c for genre g. Note that $P_g(c)$ will be zero if the given chord pattern does not exist in the model of the given genre.

[3] Support vector machine implementation in Marsyas

[4] clam-project.org

In other words for each song we will have one likelihoods per genres and each of these likelihoods are calculated by adding up the genre probability descriptors of the chord patterns present within the song. And the genre probability descriptors of the chord patterns are calculated before by counting the number of times that the chord patterns appear in each genre.

This process is then repeated 9 more times using different chunks of the dataset to complete the 10 fold cross validation. ChordMiner also predicts the genre of the songs by selecting the genre which has the highest likelihood. The estimated likelihood values and the predicted genre are then stored for all the songs so they can later be used in conjunction with other features for classification.

We have experimented with ChordMiner for patterns of 2, 3, 4, and 5 chords and found that the best results are achieved when using permutations of 3 and 4 chords.

IV. EXPERIMENTAL STUDIES

In this section, we study effects of high level features discussed in the previous section on genre classification.

A. Experiment Setup

The base system which we used as a benchmark in our experiment corresponds to those of [2]. We used Marsyas to extract MFCC, Spectral Centroid, Spectral Rolloff, Spectral Flux, and Zero Crossing features. For the details of how these features are defined, extracted, and summarized please refer to [1] and [2]. Then we used Marsyas KEA Support vector Machine (SVM) classifier and two algorithms (SMO and multilayer perceptron) in WEKA[5] to classify the music pieces using the above features. We set the classification results achieved here as the baseline and will compare our results with these.

GATZAN dataset, which is a well known and widely used dataset in the community, is used as the ground truth. This dataset consists of 10 genres and 100 songs per genre which adds up to be 1000 songs. Each song is 30 seconds. The 10 genres in this dataset are blues, classical, country, disco, hip hop, jazz, metal, pop, reggae, and rock.

We have experimented on our proposed method by performing classification on timbral features (T) extracted from GTZAN. These results correspond to those of [2] and provide the baseline for comparison. Then different combinations of beat features (B), chord occurrence features (C), genre likelihoods (L), and genre predictions (P) are added to the feature set. Note that the genre likelihoods and genre predictions are extracted by ChordMiner twice: once using patterns of 3 chords (L3 and P3) and once using patterns of 4 chords (L4 and P4).

In this section of the paper we first present the experimental results of ChordMiner predictions then we move onto the classification results using different feature sets outlined above.

B. ChordMiner Classification Results

Table 1 illustrates the confusion matrix of genre prediction by ChordMiner. This demonstrates the ability of ChordMiner to classify music genre just by using chord progression features and without the aid of sophisticated machine learning algorithms. ChordMiner performs this classification simply by selecting the genre which has the highest likelihood. The average accuracy when using sequence of 3 chords is 46.80% and when using sequence of 4 chords is 54.20%. In both cases the performance is very good for metal but poor in pop and reggae. However a lot of songs have been wrongly classified to metal. The opposite of this phenomenon can be observed in hip hop and jazz were there are only very few false positives classifications.

TABLE 1
Confusion matrix of ChordMiner
a: for sequence of 3 chords, b: for sequence of 4 chords
(bl: Blues, cl: Classical, co: Country, di: Disco, hi: Hip Hop, ja: Jazz, me: Metal, po: Pop, re: Reggae, ro: Rock)

	bl	cl	co	di	hi	ja	me	po	re	ro
bl	41	3	1	14	0	0	29	0	4	8
cl	2	52	14	7	1	1	11	5	1	6
co	1	5	62	5	0	0	18	2	1	6
di	1	3	9	52	0	0	24	0	3	8
hi	7	1	3	14	34	0	27	1	6	7
ja	3	1	4	18	0	39	20	3	4	8
me	1	1	1	2	0	2	89	1	1	2
po	1	4	11	16	0	0	22	17	6	23
re	4	3	8	11	0	0	20	2	17	35
ro	1	1	14	6	0	1	28	1	1	47

(a)

	bl	cl	co	di	hi	ja	me	po	re	ro
bl	45	4	2	9	0	0	31	2	2	5
cl	12	64	7	5	0	1	7	1	0	3
co	1	2	74	4	0	2	11	1	1	4
di	3	1	6	65	0	0	15	1	2	7
hi	8	0	2	10	38	0	23	3	5	11
ja	2	1	2	11	0	58	14	3	6	3
me	1	1	3	2	0	2	81	1	2	7
po	1	2	8	9	0	0	16	27	5	32
re	2	2	4	8	0	0	17	2	32	33
ro	1	1	9	3	0	0	21	1	6	58

(b)

C. Effects of Beat Features on Genre Classification Results

As illustrated in Table 2 although beat features perform very poor when used alone, they do increase the accuracy when used in conjunction with timbral features of the baseline. A maximum of 2.6% increase timbral features is achieved over combining beat and timbral features.

D. Effects of Chord Features on Genre Classification Results

Table 2 also illustrates the effect of statistical chord features. Statistical chord features when used alone perform better than beat features. A maximum accuracy of 42.34% is

5 www.cs.waikato.ac.nz/ml/weka

achieved when using chord features only. However results are improved significantly when the chord features are used in conjunction with the timbral and beat features. A maximum improvement of 5.61% over the baseline is observed when combining chord and timbral features. And a maximum 6.21% improvement over the baseline is observed when chord, beat, and timbral features are used altogether.

E. Effects of Chord progression Features on Genre Classification Results

Chord Progression Features had the most significant effect on the classification as demonstrated in table 2. The genre likelihood features (extracted from chord progressions) perform reasonable when used on their own but they still perform poor compared to timbral features. Note that when only considering genre likelihoods, neural network can classify the genres better than ChordMiner but ChordMiner outperforms SVM (refer to Table 1 and Table 2).

Combining the ChordMiner predictions with timbral features improves the classification by a maximum of 5.3% compared to baseline. But a better maximum performance of 9% is achieved by the combination of genre likelihoods and timbral features. Likelihoods, timbre, beat and chord features combined further improve the performance by achieving a maximum increase of 10.61% in classification accuracy. The maximum improvement of 12.4% can be observed by comparing TBL3 (Combination of Timbral features, Beat features, and genre likelihoods extracted from progressions of 3 chords) and the baseline using the MultilayerPerceptron algorithm.

Another important observation is that although the genre likelihoods extracted from sequences of 4 chords perform better on their own, the likelihoods extracted from sequence of 3 chords perform better when combined with other features

F. Effect of Different Classification Methods

Table 2 indicates the classification results of our experiment using all the features that we have extracted. The classification has been performed using a Multilayer Perceptron implementation of Neural Network, an SMO Sequential Minimal Optimization (SMO) implementation of SVM in WEKA, and another SMV in KEA. The SVM implementation of KEA performs significantly better than the WEKA algorithms. However in WEKA the Neural Network algorithm outperforms the SVM.

In Figure 2 we illustrate and compare the best results we achieved using different feature sets when using Kea's SVM classifier. This classifier has been designed for music classification as a part of Marsyas. As indicated in Figure 2 each of the proposed features increase the classification accuracy. And the maximum improvement is achieved with a combination of timbral features, beat features, and genre likelihoods extracted from progressions of 3 chords. This combination improves the results by 11% when compared to timbral features and by 10.1% when compared to timbral and beat features.

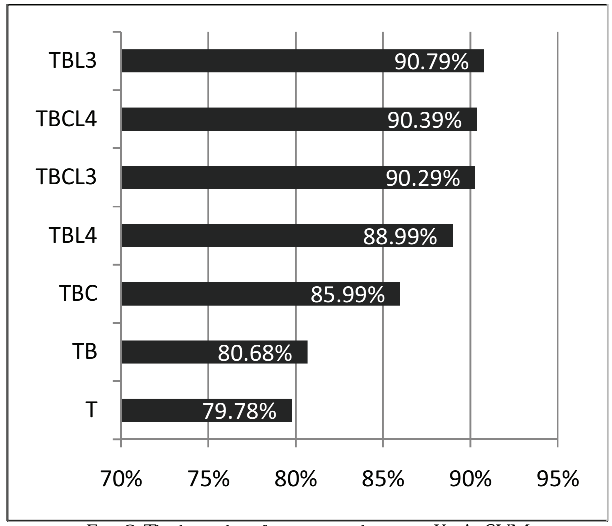

Fig. 2. The best classification results using Kea's SVM

TABLE 2
Classificaiton Results
The first row, T, indicates the baseline. The features are abbreviated to T: Timbral features, B: Beat features, C: Chord occurrence features, L :Genre likelihood features, P: prediction features

	KEA SVM	Multilayer Perceptron	SMO
T	79.78%	71.60%	70.50%
B	24.52%	25.90%	23.80%
TB	80.68%	74.20%	72.30%
C	42.34%	32.80%	36.30%
TC	85.39%	72.50%	71.60%
TBC	85.99%	73%	73%
L3	48.55%	59.80%	46.10%
TL3	86.59%	80.60%	78.30%
TP3	78.68%	75.90%	73.70%
TLP3	82.18%	79.40%	77.60%
TBL3	90.79%	84%	82.80%
TBCL3	90.29%	79.80%	77.10%
L4	43.24%	61.70%	41.40%
TL4	86.89%	80.20%	78.30%
TP4	79.78%	76.90%	75.80%
TLP4	86.89%	81.90%	79.50%
TBL4	88.99%	82.40%	80.20%
TBCL4	90.39%	81.60%	79.70%

VI. CONCLUSION AND FUTURE WORKS

In this paper we have proposed the use of high level features particularly chords and chord progressions and used them in conjunction with low level features. Our results agree with the proposition in [12] that a combination of high and low level features provide the highest classification accuracy.

To capture the chord information we have counted the number of roots and modes of the chords present in songs. And for chord progressions we had developed a pattern matching algorithm called ChordMiner that predicts the likelihood of songs belonging to different genres. This algorithm captures and summarizes the chord progression information to a format that can be used in conjunction of other features as an input to data mining algorithms. In other words ChordMiner translates the chord progressions into genre likelihoods and by doing so the extremely wide, unorganized, and unpredictable problem space is converted to an organized, and narrow problem space so the chord progression information can be coupled with other features for more accurate classification. We achieved a maximum of 61.70% accuracy on a dataset of 10 genres only using chord progressions which is comparable to those of [19] with a maximum accuracy of 64% on a dataset of 9 genres. Our proposed method showed a maximum of 12.4% improvement over the baseline and achieved a maximum accuracy of 90.79%.

The results we achieved confirms the previous works in the sense that low level features are the most important features in genre classification, but it shows that high level features are effective too. We found that high level features do contain extra information which significantly increases the classification accuracy and they should not be ignored.

This signifies the need for more musicology and psychoacoustic research in this area. There is also the need for more research experimenting with other high level features such as conceptual tempo and incorporating musical knowledge of chord progressions and musical instruments to make the algorithms smarter.

ACKNOWLEDGMENT

The authors wish to thank G. Tzanetakis (University of Victoria, Canada) and P. Mcilwain (Monash University School of Music – Conservatorium, Australia) for their useful comments.

REFERENCES

[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing,* vol. 10, pp. 293-302, 2002.

[2] G. Tzanetakis, "MARSYAS Submissions to MIREX 2007," in *MIREX 2007*, 2007.

[3] R. O. Gjerdingen and D. Perrott, "Scanning the dial: An exploration of factors in identification of musical style," in *the annual meeting of the Society for Music Perception and Cognition (SMPC)* Evanston, USA, 1999, unpublished.

[4] R. O. Gjerdingen and D. Perrott, "Scanning the Dial: The Rapid Recognition of Music Genres," *Journal of New Music Research,* vol. 37, p. 93, 2008.

[5] I. Panagakis, E. Benetos, and C. Kotropoulos, "Music genre classification: a multilinear approach," in *Ninth International Conference on Music Information Retrieval (ISMIR 2008)* Philadelphia, USA, 2008.

[6] J. G. A. Barbedo and A. Lopes, " Automatic Genre Classification of Musical Signals," *EURASIP Journal on Advances in Signal Processing 2007,* vol. 2007, 2007.

[7] A. Holzapfel and Y. Stylianou, "Musical Genre Classification Using Nonnegative Matrix Factorization-Based Features," *IEEE Transactions On Audio, Speech, And Language Processing,* vol. 16, pp. 424-434, 2008.

[8] A. Meng, P. Ahrendt, J. Larsen, and L. K. Hansen, "Temporal feature integration for music genre classification," *IEEE Transactions on Audio, Speech and Language Processing,* vol. 15, pp. 1654 - 1664, 2007.

[9] T. Lidy, A. Rauber, A. Pertusa, and J. e. M. Iñesta, "Improving Genre Classification by Combination of Audio and Symbolic Descriptors Using a Transcription System," in *nternational Conference on Music Information Retrieval (ISMIR 2007)*, 2007.

[10] A. S. Lampropoulos, P. S. Lampropoulou, and G. A. Tsihrintzis, "Genre classification enhanced by improved source separation technique," in *International Conference on Music Information Retrieval (ISMIR 2005)* London, UK, 2005.

[11] A. R. Chase, "Music discrimination by carp (Cyprinus carpio)," *Animal Learning & Behavior,* vol. 29, pp. 336 - 353, 2001.

[12] C. McKay, "Automatic Genre Classification of MIDI Recordings," McGill University, 2004.

[13] H. G. Tekman and N. Hortacsu, "Aspects of Stylistic Knowledge: What Are Different Styles Like and Why Do We Listen to Them?," *Psychology of Music,* vol. 30, pp. 28 - 47, 2002.

[14] B. Y. Chua, "Automatic extraction of perceptual features and categorization of music emotional expression from polyphonic music audio signals." vol. Doctor of Philosophy: Monash University, 2007.

[15] A. Gabrielsson and P. N. Juslin, "Emotional Expression in Music Performance: Between the Performer's Intention and the Listener's Experience," *Psychology of Music,* vol. 24, pp. 68-91, 1996.

[16] S. B. Kamenetsky and D. S. Hill, "Effect of Tempo and Dynamics on the Perception of Emotion in Music," *Psychology of Music,* vol. 25, pp. 149-160, 1997.

[17] C. McKay and I. Fujinaga, "Musical genre classification: Is it worth pursuing and how can it be improved?," in *7th International Conference on Music Information Retrieval (ISMIR 2006)* Victoria, Canada 2006.

[18] J. Zhu, X. Xue, and H. Lu, "Musical genre classification by instrumental features," in *Computer Music Conference (ICMC2004)*, 2004, pp. 580-583.

[19] C. Pérez-Sancho, D. Rizo, and J. M. Inesta, "Genre classification using chords and stochastic language models," *Connection Science,* vol. 21, pp. 145 - 159, June 2009.

[20] J. F. Paiement, D. Eck, and S. Bengio, "A Probabilistic Model for Chord Progressions'," in *Sixth International Conference on Music Information Retrieval (ISMIR 2005)* London, UK 2005.

[21] K. Lee, "A System for Automatic Chord Transcription Using Genre-Specific Hidden Markov Models," in *International Workshop on Adaptive Multimedia Retrieval* Paris, France, 2007.

[22] C. Harte and M. Sandler, "Automatic Chord Identifcation using a Quantised Chromagram," in *118th convention of the Audio Engineering Society (AES)* Barcelona, Spain, 2005.

[23] R. N. Shepard, "Circularity in judgment of relative pitch," *The journal of the acoustical society of America,* vol. 35, pp. 2346 - 2353, 1964.