EVALUATION OF MUSICAL FEATURES FOR EMOTION CLASSIFICATION

Yading Song, Simon Dixon, Marcus Pearce
Centre for Digital Music, Queen Mary University of London
{yading.song, simon.dixon, marcus.pearce}@eecs.qmul.ac.uk

ABSTRACT

Because music conveys and evokes feelings, a wealth of research has been performed on music emotion recognition. Previous research has shown that musical mood is linked to features based on rhythm, timbre, spectrum and lyrics. For example, sad music correlates with slow tempo, while happy music is generally faster. However, only limited success has been obtained in learning automatic classifiers of emotion in music. In this paper, we collect a ground truth data set of 2904 songs that have been tagged with one of the four words "happy", "sad", "angry" and "relaxed", on the Last.FM web site. An excerpt of the audio is then retrieved from 7Digital.com, and various sets of audio features are extracted using standard algorithms. Two classifiers are trained using support vector machines with the polynomial and radial basis function kernels, and these are tested with 10-fold cross validation. Our results show that spectral features outperform those based on rhythm, dynamics, and, to a lesser extent, harmony. We also find that the polynomial kernel gives better results than the radial basis function, and that the fusion of different feature sets does not always lead to improved classification.

1. INTRODUCTION

In the past ten years, music emotion recognition has attracted increasing attention in the field of music information retrieval (MIR) [16]. Music not only conveys emotion, but can also modulate a listener's mood [8]. People report that their primary motivation for listening to music is its emotional effect [19] and the emotional component of music has been recognised as most strongly associated with music expressivity [15].

Recommender systems for managing a large personal music collections typically use collaborative filtering [28] (historical ratings) and metadata- and content-based filtering [3] (artist, genre, acoustic features similarity). Emotion can be easily incorporated into such systems to subjectively organise and search for music. Musicovery [1], for example, has successfully used a dimensional model of emotion within its recommendation system.

Although music emotion has been widely studied in psychology, signal processing, neuroscience, musicology and machine learning, our understanding is still at an early stage. There are three common issues: 1. collection of ground truth data; 2. choice of emotion model; 3. relationships between emotion and individual acoustic features [13].

Since 2007, the annual Music Information Retrieval Evaluation eXchange (MIREX) [2] has organised an evaluation campaign for MIR algorithms to facilitate finding solutions to the problems of audio music classification. In previous studies, significant research has been carried out on emotion recognition including regressor training: using multiple linear regression [6] and Support Vector Machines (SVM) [23,37], feature selection [35,36], the use of lyrics [13] and advanced research including mood classification on television theme tunes [30], analysis with electroencephalogram (EEG) [18], music expression [32] and the relationship with genre and artist [12]. Other relevant work on classification suggests that feature generation can outperform approaches based on standard features in some contexts [33].

In this paper, we aim to better explain and explore the relationship between musical features and emotion. We examine the following parameters: first, we compare four perceptual dimensions of musical features: *dynamics*, *spectrum*, *rhythm*, and *harmony*; second, we evaluate an SVM associated with two kernels: polynomial and radial basis functions; third, for each feature we compare the mean and standard deviation feature value. The results are trained and tested using semantic data retrieved from last.fm [3] and audio data from 7digital [4].

This paper is structured as follows. In section 2, three psychological models are discussed. Section 3 explains the dataset collection we use in training and testing. The procedure is described in section 4, which includes data preprocessing (see section 4.1), feature extraction (see section 4.2) and classification (see section 4.3). Section 5 explains four experiments. Finally, section 6 concludes the paper and presents directions for future work.

[1] http://musicovery.com/

[2] http://www.music-ir.org/mirex/wiki/MIREX_HOME
[3] http://www.last.fm/
[4] http://www.7digital.com/

2. PSYCHOLOGICAL EMOTION MODELS

One of the difficulties in representing emotion is to distinguish music-induced emotion from perceived emotion because the two are not always aligned [5]. Different psychological models of emotion have been compared in a study of perceived emotion [7].

Most music related studies are based on two popular approaches: categorical [10] and dimensional [34] models of emotion. The categorical approach describes emotions with a limited number of innate and universal categories such as happiness, sadness, anger and fear. The dimensional model considers all affective terms arising from independent neurophysiological systems: valence (negative to positive) and arousal (calm to exciting). Recently a more sophisticated model of music-induced emotion - the Geneva Emotion Music Scale (GEMS) model - consisting of 9 dimensions, has been proposed [42]. Our results and analysis are based on the categorical model since we make our data collection through human-annotated social tags which are categorical in nature.

3. GROUND-TRUTH DATA COLLECTION

As discussed above, due to the lack of ground truth data, most researchers compile their own databases [41]. Manual annotation is one of the most common ways to do this. However, it is expensive in terms of financial cost and human labour. Moreover, terms used may differ between individuals. Different emotions may be described using the same term by different people which would result in poor prediction [38]. However, with the emergence of music discovery and recommendation websites such as last.fm which support social tags for music, we can access rich human-annotated information. Compared with the traditional approach of web mining which gives noisy results, social tagging provides highly relevant information for music information retrieval (MIR) and has become an important source of human-generated contextual knowledge [11]. Levy [24] has also shown that social tags give a high quality source of ground truth data and can be effective in capturing music similarity [40].

The five mood clusters proposed by MIREX [14] (such as rollicking, literate, and poignant) are not popular in social tags. Therefore, we use four basic emotion classes: *happy*, *angry*, *sad* and *relaxed*, considering these four emotions are widely accepted across different cultures and cover the four quadrants of the 2-dimensional model of emotion [22]. These four basic emotions are used as seeds to retrieve the top 30 tags from last.fm. We then obtain a list of songs labelled with the retrieved tags. Table1 and table 2 show an example of the retrieved results.

Given the retrieved titles and the names of the singers, we use a public API to get preview files. The results cover different types of pop music, meaning that we avoid particular artist and genre effects [17]. Since the purpose of this step is to find ground truth data, issues such as cold start, noise, hacking, and bias are not relevant [4, 20].

Most datasets on music emotion recognition are quite small (less than 1000 items), which indicates that 2904 songs (see table 3) for four emotions retrieved by social tags is a good size for the current experiments. The dataset will be made available [5], to encourage other researchers to reproduce the results for research and evaluation.

Happy	**Angry**	**Sad**	**Relax**
happy	angry	sad	relax
happy hardcore	angry music	sad songs	relax trance
makes me happy	angry metal	*happysad*	relax music
happy music	angry pop music	sad song	jazz relax
happysad	angry rock	sad & beautiful	only relax

Table 1. Top 5 tags returned by last.fm

Singer	**Title**
Noah And The Whale	5 Years Time
Jason Mraz	I'm Yours
Rusted Root	Send Me On My Way
Royksopp	Happy Up Here
Karen O and the Kids	All Is Love

Table 2. Top songs returned with tags from the "happy" category.

Emotion	**Number of Songs**
Happy	753
Angry	639
Sad	763
Relaxed	749
Overall	2904

Table 3. Summary of ground truth data collection

4. PROCEDURES

The experimental procedure consists of four stages: data collection, data preprocessing, feature extraction, and classification, as shown in figure 1.

4.1 Data Preprocessing

As shown in Table 1, there is some noise in the data such as confusing tags and repeated songs. We manually remove data with the tag *happysad* which existed in both the happy and sad classes and delete the repeated songs, to make sure every song will only exist once in a single class. Moreover, we convert our dataset to standard wav format (22,050 Hz sampling rate, 16 bit precision and mono channel). The song excerpts are either 30 seconds or 60 seconds, representing the most salient part of the song [27], therefore there is no need to truncate. At the end, we normalise the excerpts by dividing by the highest amplitude to mitigate the *production effect* of different recording levels.

4.2 Feature Extraction

As suggested in the work of Saari and Eerola [35], two different types of feature (mean and standard deviation) with

[5] The dataset can be found at *https://code.soundsoftware.ac.uk/projects-/emotion-recognition*

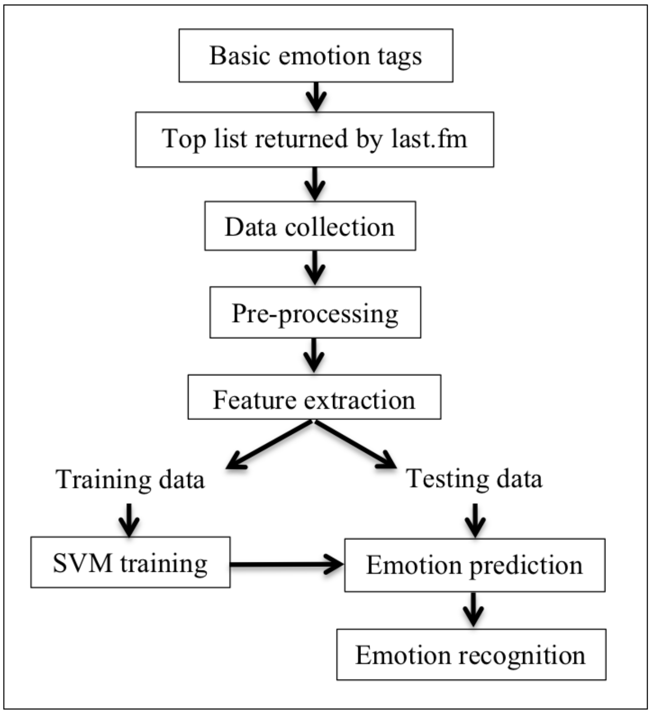

Figure 1. Procedure

a total of 55 features were extracted using the MIR toolbox [6] [21] (shown in table 4). The features are categorized into the following four perceptual dimensions of music listening: *dynamics*, *rhythm*, *spectral*, and *harmony*.

4.3 Classification

The majority of music classification tasks [9] (genre classification [25, 39], artist identification [29], and instrument recognition [31]) have used k-nearest neighbour (K-NN) [26] and support vector machines (SVM) [2]. In the case of audio input features, the SVM has been shown to perform best [1].

In this paper, therefore, we choose support vector machines as our classifier, using the implementation of the sequential minimal optimisation algorithm in the Weka data mining toolkit [7]. SVMs are trained using polynomial and radial basis function (RBF) kernels. We set the cost factor $C = 1.0$, and leave other parameters unchanged. An internal 10-fold cross validation is applied. To better understand and compare features in four perceptual dimensions, our experiments are divided into four tasks.

Experiment 1: we compare the performance of the two kernels (polynomial and RBF) using various features.

Experiment 2: four classes (perceptual dimensions) of features are tested separately, and we compare the results to find a dominant class.

Experiment 3: two types of feature descriptor, mean and standard deviation, are calculated. The purpose is to compare values for further feature selection and dimensionality reduction.

[6] Version 1.3.3: https://www.jyu.fi/music/coe/materials/mirtoolbox
[7] http://www.cs.waikato.ac.nz/ml/weka/

Dimen.	No.	Features	Acronyms
Dynamics	1-2	RMS energy	RMSm, RMSstd
	3-4	Slope	Ss, Sstd
	5-6	Attack	As, Astd
	7	Low energy	LEm
Rhythm	1-2	Tempo	Ts, Tstd
	3-4	Fluctuation peak (pos, mag)	FPm, FMm
	5	Fluctuation centroid	FCm
Spec.	1-2	Spectrum centroid	SCm, SCstd
	3-4	Brightness	BRm, BRstd
	5-6	Spread	SPm, SPstd
	7-8	Skewness	SKm, SKstd
	9-10	Kurtosis	Km, Kstd
	11-12	Rolloff95	R95s, R95std
	13-14	Rolloff85	R85s, R85std
	15-16	Spectral Entrophy	SEm, SEstd
	17-18	Flatness	Fm, Fstd
	19-20	Roughness	Rm, Rstd
	21-22	Irregularity	IRm. IRstd
	23-24	Zero crossing rate	ZCRm, ZCRstd
	25-26	Spectral flux	SPm, SPstd
	27-28	MFCC	MFm, MFstd
	29-30	DMFCC	DMFm, DMFstd
	31-32	DDMFCC	DDm, DDstd
Harmony	1-2	Chromagram peak	CPm, CPstd
	3-4	Chromagram centroid	CCm, CCstd
	5-6	Key clarity	KCm, KCstd
	7-8	Key mode	KMm, KMstd
	9-10	HCDF	Hm, Hstd

Table 4. The feature set used in this work; m = **mean**, std = **standard deviation**.

Experiment 4: different combinations of feature classes (e.g., spectral with dynamics) are evaluated in order to determine the best-performing model.

5. RESULTS

5.1 Experiment 1

In experiment 1, SVMs trained with two different kernels are compared. Previous studies [23] have found in the case of audio input that the SVM performs better than other classifiers (Logistic Regression, Random Forest, GMM, K-NN and Decision Trees). To our knowledge, no work has been reported explicitly comparing different kernels for SVMs. In emotion recognition, the radial basis function kernel is a common choice because of its robustness and accuracy in other similar recognition tasks [1].

	Polynomial		RBF		
Feature Class	Accuracy	Time	Accuracy	Time	No.
Dynamics	37.2	0.44	26.3	32.5	7
Rhythm	37.5	0.44	34.5	23.2	5
Harmony	47.5	0.41	36.6	27.4	10
Spectral	**51.9**	0.40	48.1	14.3	32

Table 5. Experiment 1 results: time = model building time, No. = number of features in each class

The results in table 5 show however that regardless of the features used, the polynomial kernel always achieved the higher accuracy. Moreover, the model construction times for each kernel are dramatically different. The average construction time for the polynomial kernel is 0.4 seconds, while the average time for the RBF kernel is 24.2

seconds, around 60 times more than the polynomial kernel. The following experiments also show similar results. This shows that polynomial kernel outperforms RBF in the task of emotion recognition at least for the parameter values used here.

5.2 Experiment 2

In experiment 2, we compare the emotion prediction results for the following perceptual dimensions: *dynamics*, *rhythm*, *harmony*, and *spectral*. Results are shown in figure 2). Dynamics and rhythm features yield similar results, with harmony features providing better results, but the spectral class with 32 features achieves the highest accuracy of 51.9%. This experiment provides a baseline model, and further exploration of multiple dimensions is performed in experiment 4.

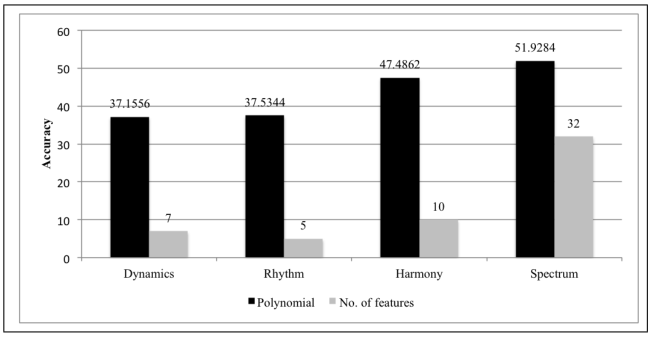

Figure 2. Comparison of classification results for the four classes of features.

5.3 Experiment 3

In this experiment, we evaluate different types of feature descriptors, mean value and standard deviation for each feature across all feature classes, for predicting the emotion in music. The results in table 6 show that the use of both mean and standard deviation values gives the best results in each case. However, the processing time increased, so choosing the optimal descriptor for each feature is highly desirable. For example, choosing only the mean value in the harmony class, we lose 2% of accuracy but increase the speed while the choice of standard deviation results in around 10% accuracy loss. As the number of features increases, the difference between using mean and standard deviation will be reduced. However, more experiments are needed to explain why the mean in harmony and spectral features, and standard deviation values of dynamics and rhythm features have higher accuracy scores.

5.4 Experiment 4

In order to choose the best model, the final experiment fuses different perceptual features. As presented in table 7, optimal accuracy is not produced by the combination of all features. Instead, the use of spectral, rhythm and harmony (but not dynamic) features produces the highest accuracy.

Features Class	Polynomial	No. features
Dynamics all	**37.2**	7
Dynamics mean	29.7	3
Dynamics std	*33.8*	3
Rhythm all	37.5	5
Rhythm mean	28.7	1
Rhythm std	**34.2**	1
Harmony all	**47.5**	10
Harmony mean	*45.3*	5
Harmony std	38.3	5
Spectral all	**51.9**	32
Spectral mean	*49.6*	16
Spectral std	47.5	16
Spec+Dyn all	**52.3**	39
Spec+Dyn mean	*50.5*	19
Spec+Dyn std	48.7	19
Spec+Rhy all	**52.3**	37
Spec+Rhy mean	*49.8*	17
Spec+Rhy std	47.8	17
Spec+Har all	**53.3**	42
Spec+Har mean	*51.3*	21
Spec+Har std	50.3	21
Har+Rhy all	**49.1**	15
Har+Rhy mean	*45.6*	6
Har+Rhy std	41.2	6
Har+Dyn all	**48.8**	17
Har+Dyn mean	*46.9*	8
Har+Dyn std	42.4	8
Rhy+Dyn all	**41.7**	12
Rhy+Dyn mean	32.0	4
Rhy+Dyn std	*38.8*	4

Table 6. Comparison of mean and standard deviation (std) features.

Features	Accuracy	No. features
Spec+Dyn	52.3	39
Spec+Rhy	52.3	37
Spec+Har	53.3	42
Har+Rhy	49.1	15
Har+Dyn	48.8	17
Rhy+Dyn	41.7	12
Spec+Dyn+Rhy	52.4	44
Spec+Dyn+Har	53.8	49
Spec+Rhy+Har	**54.0**	47
Dyn+Rhy+Har	49.7	22
All Features	53.6	54

Table 7. Classification results for combinations of feature sets.

6. CONCLUSION AND FUTURE WORK

In this paper, we collected ground truth data on the emotion associated with 2904 pop songs from last.fm tags. Audio features were extracted and grouped into four perceptual dimensions for training and validation. Four experiments were conducted to predict emotion labels. The results suggest that, instead of the conventional approach using SVMs trained with a RBF kernel, a polynomial kernel yields higher accuracy. Since no single dominant features have been found in emotion recognition, we explored the performance of different perceptual classes of feature for predicting emotion in music. Experiment 3 found that dimensionality reduction can be achieved through removing either mean or standard deviation values, halving the number of features used, with, in some cases, only 2% accuracy loss. The last experiment found that inclusion of dynamics features with the other classes actually impaired the performance of the classifier while the combination of spectral, rhythmic and harmonic features yielded optimal performance.

In future work, we will expand this research both in depth and breadth, to find features and classes of features which best represent emotion in music. We will examine higher-level dimensions such as temporal evolution features, as well as investigating the use of auditory models. Using the datasets retrieved from Last.fm, we will compare the practicability of social tags with other human-annotated datasets in emotion recognition. Through these studies of subjective emotion, we will develop methods for incorporating other empirical psychological data in a subjective music recommender system.

7. ACKNOWLEDGEMENTS

We acknowledge the support of the Queen Mary University of London Postgraduate Research Fund (QMPGRF) and the China Scholarship Council. We would like to thank the reviewers and Emmanouil Benetos for their advice and comments.

8. REFERENCES

[1] K. Bischoff, C. S. Firan, R. Paiu, W. Nejdl, C. Laurier, and M. Sordo. Music Mood and Theme Classification - A Hybrid Approach. In *10th International Society for Music Information Retrieval Conference*, number Ismir, pages 657–662, 2009.

[2] E. Boser, N. Vapnik, and I. M. Guyon. Training Algorithm Margin for Optimal Classifiers. In *ACM Conference on Computational Learning Theory*, pages 144–152, 1992.

[3] P. Cano, M. Koppenberger, and N. Wack. Content-based Music Audio Recommendation. In *Proceedings of the 13th annual ACM international conference on Multimedia*, number ACM, pages 211–212, 2005.

[4] O. Celma. Foafing the Music : Bridging the Semantic Gap in Music Recommendation. In *The Semantic Web-ISWC*, 2006.

[5] T. Eerola. Are the Emotions Expressed in Music Genre-specific? An Audio-based Evaluation of Datasets Spanning Classical , Film , Pop and Mixed Genres. *Journal of New Music Research*, 40(March 2012):349–366, 2011.

[6] T. Eerola, O. Lartillot, and P. Toiviainen. Prediction of Multdimensional Emotional Ratings in Music from Audio Using Multivariate Regression Models. In *10th International Society for Music Information Retrieval Conference*, number Ismir, pages 621–626, 2009.

[7] T. Eerola and J.K. Vuoskoski. A Comparison of the Discrete and Dimensional Models of Emotion in Music. *Psychology of Music*, 39(1):18–49, August 2010.

[8] Y. Feng and Y. Zhuang. Popular Music Retrieval by Detecting Mood. In *International Society for Music Information Retrieval Conference*, volume 2, pages 375–376, 2003.

[9] Z. Fu, G. Lu, K.M. Ting, and D. Zhang. A Survey of Audio-based Music Classification and Annotation. *IEEE Transactions on Multimedia*, 13(2):303–319, 2011.

[10] K. Hevner. Experimental studies of the elements of expression in music. *The American Journal of Psychology*, 48:246–268, 1936.

[11] X. Hu, M. Bay, and J.S. Downie. Creating a Simplified Music Mood Classification Grouth-truth Set. In *International Conference on Music Information Retrieval*, pages 3–4, 2007.

[12] X. Hu and J.S. Downie. Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata. In *8th International Conference on Music Information Retrieval*, 2007.

[13] X. Hu, J.S. Downie, and A.F. Ehmann. Lyric Text Mining in Music Mood Classification. In *10th International Society for Music Information Retrieval Conference*, number Ismir, pages 411–416, 2009.

[14] X. Hu, J.S. Downie, C. Laurier, and M. Bay. The 2007 MIREX Audio Mood Classification Task: Lesson Learned. In *International Society for Music Information Retrieval Conference*, pages 462–467, 2008.

[15] P. N. Juslin, J. Karlsson, E. Lindström, A. Friberg, and E. Schoonderwaldt. Play it Again with Feeling: Computer Feedback in Musical Communication of Emotions. *Journal of experimental psychology. Applied*, 12(2):79–95, June 2006.

[16] Y.E. Kim, E.M. Schmidt, R. Migneco, B.G. Morton, P. Richardson, J. Scott, J.A. Speck, and D. Turnbull. Music Emotion Recognition: A State of the Art Review. In *11th International Society for Music Information Retrieval Conference*, number Ismir, pages 255–266, 2010.

[17] Y.E. Kim, D.S. Williamson, and S. Pilli. Towards Quantifying the Album Effect in Artist Identification. In *International Society for Music Information Retrieval Conference*, 2006.

[18] S. Koelstra, C. Muhl, and M. Soleymani. Deap: A Database for Emotion Analysis Using Physiological Signals. *IEEE Trans. on Affective Computing*, pages 1–15, 2011.

[19] C.L. Krumhansl. Music : A Link Between Cognition and Emotion. *American Psychological Society*, 11(2):45–50, 2002.

[20] P. Lamere. Social Tagging and Music Information Retrieval. *Journal of New Music Research*, 37(2):101–114, June 2008.

[21] O. Lartillot and P. Toiviainen. MIR in Matlab (II): A Toolbox for Musical Feature Extraction from Audio. In *International Conference on Music Information Retrieval*, number Ii, pages 237–244, 2007.

[22] C. Laurier and J. Grivolla. Multimodal Music Mood Classification Using Audio and Lyrics. In *Int. Conf. Machine Learning and Applications*, pages 1–6, 2008.

[23] C. Laurier, P. Herrera, M. Mandel, and D. Ellis. Audio Music Mood Classification Using Support Vector Machine. In *MIREX task on Audio Mood Classification*, pages 2–4, 2007.

[24] M. Levy. A Semantic Space for Music Derived from Social Tags. In *Austrian Compuer Society*, volume 1, page 12. Citeseer, 2007.

[25] B. Lines, E. Tsunoo, G. Tzanetakis, and N. Ono. Beyond Timbral Statistics : Improving Music Classification Using Percussive. *IEEE Transactions on Audio, Speech and Language Processing*, 19(4):1003–1014, 2011.

[26] T. M. Cover and P. E. Hart. Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[27] K.F. MacDorman, S. Ough, and C. Ho. Automatic Emotion Prediction of Song Excerpts: Index Construction, Algorithm Design, and Empirical Comparison. *Journal of New Music Research*, 36(4):281–299, December 2007.

[28] T. Magno and C. Sable. A Comparison of Signal of Signal-based Music Recommendation to Genre Labels, Collaborative Filtering, Musicological Analysis, Human Recommendation and Random Baseline. In *Proceedings of the 9th International Conference of Music Information Retrieval*, pages 161–166, 2008.

[29] M. Mandel. Song-level Features and Support Vector Machines for Music Classification. In *Proc. International Conference on Music Information Retrieval*, 2005.

[30] M. Mann, T.J. Cox, and F.F. Li. Music Mood Classification of Television Theme Tunes. In *12th International Society for Music Information Retrieval Conference*, number Ismir, pages 735–740, 2011.

[31] J. Marques and P.J. Moreno. A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines, 1999.

[32] L. Mion and G.D. Poli. Score-Independent Audio Features for Description of Music Expression. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):458–466, 2008.

[33] F. Pachet and P. Roy. Analytical Features: A Knowledge-Based Approach to Audio Feature Generation. *EURASIP Journal on Audio, Speech, and Music Processing*, 2009(2):1–23, 2009.

[34] J.A. Russell, A. Weiss, and G.A. Mendelsohn. Affect Grid: A Single-item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology*, 57(3):493–502, 1989.

[35] P. Saari, T. Eerola, and O. Lartillot. Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6):1802–1812, 2011.

[36] E.M. Schmidt, D. Turnbull, and Y.E. Kim. Feature Selection for Content-Based, Time-Varying Musical Emotion Regression Categories and Subject Descriptors. In *Multimedia Information Retrieval*, pages 267–273, 2010.

[37] B. Schuller, J. Dorfner, and G. Rigoll. Determination of Nonprototypical Valence and Arousal in Popular Music: Features and Performances. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010:1–19, 2010.

[38] D. Turnbull, L. Barrington, and G. Lanckriet. Five Approaches to Collecting Tags for Music. In *Proceedings of the 9th International Conference of Music Information Retrieval*, pages 225–230, 2008.

[39] G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[40] D. Wang, T. Li, and M. Ogihara. Tags Better Than Audio Features? The Effect of Joint use of Tags and Audio Content Features for Artistic Style Clutering. In *11th International Society on Music Information Retrieval Conference*, number ISMIR, pages 57–62, 2010.

[41] D. Yang and W.S. Lee. Disambiguating Music Emotion Using Software Agents. In *Proceedings of the 5th International Conference on Music Information Retrieval*, pages 52–58, 2004.

[42] M. Zentner, D. Grandjean, and K.R. Scherer. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion (Washington, D.C.)*, 8(4):494–521, August 2008.