ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

*International Journal of Advanced Research in Computer and Communication Engineering*
**Vol. 4, Issue 1**, *January 2015*

# Music emotion recognition using support vector machines and regression approach

**T.N.CHARANYA[1], R.VIJAYALAKSHMI[2]**

Asst Prof, Dept of Computer Science and Engineering, DMI College of Engg, Chennai, India[1,2]

**Abstract:** Content-based retrieval has emerged in the face of content explosion as a promising approach to information access. In this paper, we focus on the challenging issue of recognizing the emotion content of music signals, or music emotion recognition (MER). Specifically, we formulate MER as a regression problem to predict the arousal and valence values (AV values) of each music sample directly. Associated with the AV values, each music sample becomes a point in the arousal-valence plane, so the users can efficiently retrieve the music sample by specifying a desired point in the emotion plane. Because no categorical taxonomy is used, the regression approach is free of the ambiguity inherent to conventional categorical approaches. To improve the performance, we apply principal component analysis to reduce the correlation between arousal and valence, and RReliefF to select important features. An extensive performance study is conducted to evaluate the accuracy of the regression approach for predicting AV values. The best performance evaluated in terms of the *R2* statistics reaches 58.3% for arousal and 28.1% for valence by employing support vector machine as the regressor. We also apply the regression approach to detect the emotion variation within a music selection and find the prediction accuracy superior to existing works. A group-wise MER scheme is also developed to address the subjectivity issue of emotion perception.

**Keywords:** Music emotion recognition (MER), arousal, Valence, regression, support vector machine.

## 1. INTRODUCTION

Music plays an important role in human's history, even more so in the digital age. Never before has such a large collection of music been created and accessed daily by people. As the amount of content continues to explode, the way music information is organized has to evolve in order to meet the ever increasing demand for easy and effective information access. Music classification and retrieval by emotion is a plausible approach, for it is content-centric and functionally powerful. Emotion recognition from music signal is a challenging task due to the following reasons. First, emotion perception is intrinsically subjective and people can perceive different emotions for the same song. This subjectivity issue makes the performance evaluation of an MER system fundamentally difficult because a common agreement on the classification result is hard to obtain. Second, it is not easy to describe emotion in a universal way because the adjectives used to describe emotions may be ambiguous, and the use of adjectives for the same emotion can vary from person to person. Third, it is still inexplicable how music evokes emotion. What intrinsic element of music, if any, creates a specific emotional response in the listener is still far from well-understood.

To uncover the relationship between music and emotion, many previous works [1]–[8] have categorized emotions into a number of emotion classes and applied the standard pattern recognition procedure to train a classifier. The methods described in [1], [2] adopt the basic emotions such as happy, angry, sad and fear as the emotion classes, whereas the methods described in [3]–[8] recognize the ambiguity of adjectives and define the emotion classes in terms of arousal (how exciting or calming) and valence (how positive or negative). For example, the emotion classes can be divided into the four quadrants in Thayer's arousal-valence emotion plane [12], Fig. 1.

However, even with the emotion plane, the categorical taxonomy of emotion classes is still inherently ambiguous. Each emotion class represents an area in the emotion plane, and the emotion states within each area may vary a lot. For example, the first quadrant of the emotion plan contains emotions such as excited, happy, and pleased, which are different in nature. This ambiguity confuses the subjects in the subjective test and confuses the users when retrieving a music piece according to their emotion states.

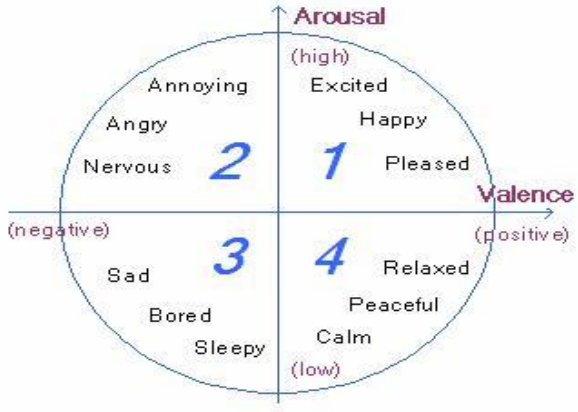


Fig. 1 Thayer's arousal-valence emotion plane.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

*International Journal of Advanced Research in Computer and Communication Engineering*
**Vol. 4, Issue 1**, *January 2015*

An alternative is to view the emotion plane as a continuous space and recognize each point of the plane as an emotion state. In this way, the ambiguity associated with emotion classes or adjectives can be successfully avoided since no categorical classes are needed. This continuous perspective has been adopted by psychologists to model the emotional response of the subjects [13], [14]. In [15], the software 'FEELTRACE' is developed to let subjects track the emotion content of a stimulus (such as speech, music, or video) as they perceive it over time. However, a major issue of the continuous perspective is that arousal and valence are not necessarily independent and can in fact impact each other. Whether the emotion states should be modeled as categories or continua has been a debate in psychology, and either perspective has its pros and cons. For MER, the continuous perspective is considered more appropriate since it resolves the ambiguity issue.

Specifically, with the continuous approach, we first compute the arousal and valence values (AV values) of each music sample and view the music sample as a point in the emotion plane. Then the user can retrieve music by specifying a point in the emotion plane according to his/her emotion state, and the system would return the music pieces whose locations are closest to the specified point. In this way, apparently, the efficiency and accuracy of music retrieval is much improved. The viability of the continuous approach heavily lies in the prediction accuracy of the AV values. Since automatic calculation of the AV values (AV computation) is still at its early stage, and the performance of existing approaches [8]–[11] is unsatisfactory in many aspects (see Section II), a primary task of this paper is to develop an effective method for AV computation.

We propose to formulate MER as a *regression* problem and use regression techniques to directly predict the AV values of music samples from the extracted features. This computational algorithm has sound theoretical basis, allows thorough performance study, and generally exhibits reliable prediction performance. The other main issue, the dependency between arousal and valence, is addressed by reducing the data correlation by principal component analysis [16].

An extensive performance study is conducted to evaluate the prediction accuracy of the proposed regression approach by using different combination of data spaces, feature spaces, and regression algorithms. Support vector regression [18] is found to produce better prediction accuracy than linear regression [17] and AdaBoost.RT [20]. The *R2* statistics [17] reaches 58.3% for

arousal and 28.1% for valence. Because there are no other existing systems viewing MER from a continuous perspective, we apply the regression approach to detect the emotion variation within music selections and find it is superior to the one proposed in [10].

In summary, the primary contributions of the paper include: To our best knowledge, this work represents one of the first attempts that develop an MER system from a continuous perspective and represent each song as a point in the emotion plane. This approach is free of the ambiguity issue of MER.

TABLE I COMPARISON OF WORKS ON MUSIC EMOTION

| Field | Perspective | Description |
|---|---|---|
| **MER [1]–[8]** | categorical | Classifying music selections into several classes based on emotion. |
| **MEVD [8]–[11]** | continuous | Detecting the emotion variation within a music selection. |
| **MER (this work)** | continuous | Representing each music selection as a point in the emotion plane. |

A novel AV computation method based on the regression theory is proposed. Principal component analysis [16] is employed to reduce the data correlation, and RReliefF [22] is utilized for feature selection (Sections 3 and 4).

An extensive performance study is conducted to demonstrate the accuracy and effectiveness of the regression approach for both music emotion recognition and music emotion variation detection (Section 5).

A group-wise MER scheme is proposed to solve the subjectivity issue of MER (Section 6).

## 2. LITERATURE REVIEW

Despite a great deal of effort has been made for MER in recent years [1]–[8], little attention has been paid to view the emotion plane from a continuous perspective. Some exceptions can be found in the music emotion variation detection (MEVD) field [8]–[11], where the emotion content of music is quantified as a time-varying continuous variable, and some statistical methods are developed to predict the emotion variation.

However, detecting the emotion variation is different from representing each song individually as a point in the emotion plane. Our work represents one of the first attempts pioneering this novel perspective. See Table I for a comparison. In the following we give brief review of existing AV computation methods and illuminate the rational of adopting the regression approach rather than these methods.

### *2.1. Arousal and Valence Modeling (AV modeling)*

To detect the emotion variation in video sequences, AV modeling is proposed in [9] to compute the AV values. The arousal and valence models are weighted combinations of some component functions that are computed along the timeline. The resulting arousal and valence curves are combined to form an *affective curve*, making it easy to trace the emotion variation of video content and to identify the segments with high emotional content. The component functions used for arousal are the motion vectors between consecutive video frames, the changes in shot lengths, and the energy of sound. Valence is modeled by the pitch of sound.

## 3. THE REGRESSION APPROACH

Regression theory is a well-studied theory aiming at predicting a real value from observed variables (or features). It has a sound theoretical foundation, allows easy performance analysis and optimization, and generally provides reliable prediction performance [17]. Besides, no temporal information or geometric operation is needed.


DOI 10.17148/IJARCCE.2015.4124

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

*International Journal of Advanced Research in Computer and Communication Engineering*
**Vol. 4, Issue 1**, *January 2015*

Therefore, formulating MER as a regression problem seems to be a promising approach.

Below we first describe how the formulation is made, then present the system description in detail. The performance study of the regression approach is reported in Section V. Given $N$ inputs ($xi$, $yi$), $1 \leq i \leq N$, where $xi$ is a feature vector for the $i$th input sample, and $yi \in$ ( denotes a set of real values) is the real value to be predicted for the $i$th sample, the regression system trains a regression algorithm (regressor) $R(\cdot)$ such that the mean squared error ε is minimized [17]:

$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N}(y_i - R(x_i))^2 ,$$

where $R(xi)$ is the prediction result for the $i$th sample. Since the AV values are viewed upon as real values from the continuous perspective, the regression theory can be well applied to directly predict arousal and valence. To formulate MER as a regression problem, the following considerations are taken into account:

1) Domain of : The Thayer's emotion plane is viewed as a coordinate space spanned by arousal and valence, where each value is confined within [–1, 1].

2) Ground truth: The ground truth is set via a subjective test by averaging the subjects' opinions about the AV values of each music sample (see Section 4.3).

3) Feature extraction: The extracted features need to be relevant to emotion perception for the repressors to be accurate (see Sections 4.2 and 5.3).

4) Regression algorithm: Although regression theory has been well studied and many good regressors are readily available [17], the performance of a regressor is case dependent. A number of regressors should be adopted and compared to find the best one (see Section 4.4).

5) Number of regressors: Since we want to predict both arousal and valence, two regressors are required and are referred to as *RA* and *RV* .

6) Training fashion: As mentioned in Section I, there is a certain degree of dependency between arousal and valence. Therefore, apart from training *RA* and *RV* solely independently, we need to study whether the prediction accuracy is improved if the dependency of the AV values is considered (see Section 5.2).

TABLE II THE ADOPTED FEATURE EXTRACTION ALGORITHMS

| Method | Number of feature | Description |
|---|---|---|
| **PsySound [27]** | (P) 44 | Extracts features including loudness, level, pitch multiplicity, and dissonance based on psychoacoustic models. |
| **Marsyas [29]** | (M) 30 | Extracts timbral texture, rhythmic content and pitch content features. It has been shown useful in music genre classification. |
| **Spectral contrast [3]** | (SC)12 | Represents the relative characteristics of each spectral subband, and reflects the distribution of harmonic components. |
| **DWCH [2]** | (D) 28 | Daubechies wavelets coefficient histogram, which has better ability in representing both local and global information. |
| **Total (ALL)** | 114 | |

TABLE III THE 15 PSYSOUND FEATURES (PSY15) RECOMMENDED IN [8]

| Feature | Description |
|---|---|
| **1 Spectral Centroid** | The centroid of spectral density function. |
| **2 Loudness** | Human perception of sound intensity. |
| **3, 4₁ Sharpness** | A pitch-like (low-high) aspect of timbre. |
| **5 Timbral** | Width The flatness of a loudness function. |
| **6 Volume** | Human perception of the size of sound. |
| **7, 8₁ Spectral Dissonance** | Roughness of all spectrum components. |
| **9,10₁ Tonal Dissonance** | Roughness of just the tonal components. |
| **11 Pure Tonal** | The audibility of the spectral pitches. |
| **12 Complex Tonal** | The audibility of the virtual pitches. |
| **13 Multiplicity** | The number of pitches heard. |
| **14 Tonality** | Major-minor tonality, e.g., A major. |
| **15 Chord** | Musical pitches sounded simultaneously. |

## 4. SYSTEM DESCRIPTION

Our MER system represents each music selection as a point in the emotion plane and provides a friendly user interface for music retrieval and management. The system diagram is shown in Fig. 2, and the details are described below.

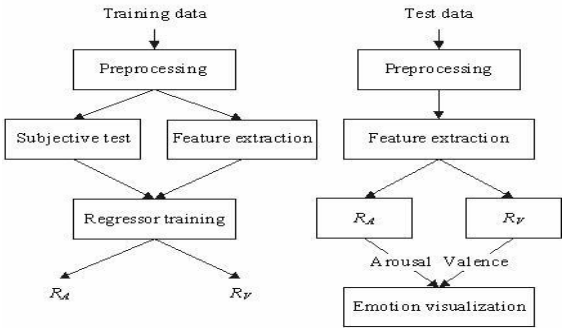


Fig. 2. System diagram of the proposed regression approach. Left: training phase; right: testing phase. *RA* denotes the regressor for arousal, and *RV* denotes the regressor for valence.

### 4.1. Data Collection and Preprocessing

The music database is made up of 195 popular songs selected from a number of Western, Chinese, and Japanese albums [8]. Two criteria are used in the selection: 1) These songs should be distributed uniformly in each quadrant of the emotion plane. 2) Each music sample should express a certain dominant emotion.

Note the genre of our database is popular music of different countries rather than the western classical music, which is commonly adopted in previous works [2]–[5], [10], [11]. Western classical music is often chosen because it is much easier to gain agreement on perceived emotion and thus has less subjectivity issue [3]. However, since the purpose of MER is to facilitate music retrieval and management in everyday music listening, and since it is the popular music that dominates the everyday music-listening, we should not shy away from the subjectivity issue by using only western classical music. More discussions on the subjectivity issue are provided in Section 6.

### *4.2. Feature Extraction*

After preprocessing, we use the spectral contrast algorithm [3], DWCH algorithm [2], and two computer programs PsySound [27] and Marsyas [29] to extract musical features


DOI 10.17148/IJARCCE.2015.4124

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

*International Journal of Advanced Research in Computer and Communication Engineering*
**Vol. 4, Issue 1**, *January 2015*

and construct a 114-dimension feature space, which is referred to as ALL hereafter. The extracted features, which are described in detail below, have been used for MER in pervious works. See Table II for denotations and brief descriptions.

Dissonance measures are related to the perception of short irregularities in a sound; any note in music that does not fall within the prevailing harmony is considered dissonant. Pitch measures are related to the perceived fundamental frequency of a sound. Because of this psychoacoustical foundation, the features extracted by PsySound have been found much related to emotion perception, especially 15 of them [8]. Therefore, we utilize these 15 features to form a second feature space called Psy15. It generates 19 timbral texture features (spectral centroid, spectral rolloff, spectral flux, time domain zero-crossing and MFCC), 6 rhythmic content features (by beat and tempo detection) and 5 pitch content features (by multi-pitch detection). Spectral centroid, spectral rolloff, and spectral flux describe spectral shape properties, zero-crossing measures the noisiness of the signal, and MFCC (Mel-frequency cepstral coefficient) is a non-musical pitch scale commonly used in speech and audio signal processing.

Spectral contrast features capture the relative spectral information in each subband and utilize the spectral peak, spectral valley, and their dynamics as features [3]. The spectral contrast features also roughly reflect the relative distribution of the harmonic and non-harmonic components in the spectrum.

### 4.3. Regressor Training

The 195 (*xi*, *yi*) inputs from feature extraction and subjective test are then used to train the following three regression algorithms: multiple linear regression (MLR) [17], support vector regression (SVR) [18], and AdaBoost.RT (BoostR) [20]. MLR is a standard regression algorithm which assumes a linear relationship between variables and estimates the linear relationship by a least squares estimator. We treat MLR as the baseline approach for its simplicity.

Comparatively, SVR nonlinearly maps input feature vectors to a higher dimensional feature space by the kernel trick, and yields prediction functions that are expanded on a subset of support vectors [18]. As its name indicates, SVR is an extension of the famous support vector classification, which has been found in many cases superior to existing machine learning methods. A number of previous works have adopted support vector classification for MER and reported excellent classification performance [2], [4], [6]. BoostR is another nonlinear regression algorithm in which a number of regression trees are trained iteratively and weighted according to the prediction accuracy. After the iterative process, the prediction result of each regression tree is combined (weighted mean) to form the final hypothesis. The basic underlining concept of the boosting process is based on the observation that finding a number of weak predicting rules is much easier than finding a single, highly accurate one [20].

Boosting algorithms, which are the state-of-the-art methods for face detection [21], have been successfully applied in many machine learning problems.

## 5. PERFORMANCE ESTIMATION AND STUDY

We run a series of experiments to evaluate the performance of the regression approach. Different ground truth data spaces, feature spaces, and regression algorithms are compared in terms of the *R2* statistics, which is a standard way for measuring the goodness of fit for regression models [17],

$$R^2 = 1 - \frac{N\varepsilon}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \quad (5)$$

where $y$ is the mean of the ground truth, and the normalization of the total squared error ($N\varepsilon$) by the energy of the ground truth makes *R2* comparable between experiments. *R2* is often interpreted as the proportion of underlying data variation that is explained by the fitted regression model [32]. An *R2* of 1.0 means the model perfectly fits the data, while a negative *R2* means the model is even worse than simply taking the sample mean. However, the kind of the *R2* statistics that is satisfactory is case-dependent.

We evaluate the performance of regression by the 10–fold cross validation technique [16], in which the whole dataset is randomly divided into 10 parts, 9 of them for training and the remaining one for testing. The above process is repeated 20 times before we compute the average result. *R2* for each data dimension (say, arousal and valence) is computed separately.

### *5.1. Performance Evaluation of Regressor*

We first evaluate the prediction accuracy of different regression algorithms in terms of *R2*. The implementation of SVR is based on the library LIBSVM [19], along with a grid parameter search to find the best parameters. BoostR is implemented in Java language. The threshold $\varphi$ for demarcating correct and incorrect predictions are empirically determined as 0.1, and the number of iterations is 30. MLR can be easily implemented in Matlab.

We run each of the regressor with the same configuration: the data space is AV and the feature space is Psy15. Result shown in the first three rows of Table 4 indicates that the *R2* of SVR reaches 57.0% for arousal and 22.2% for valence, representing the most prominent prediction accuracy among the three, and BoostR exhibits prediction accuracy similar to the baseline method MLR. Consequently, we employ SVR as the regressor in the following experiments.

TABLE IV THE $R_2$ STATISTICS FOR DIFFERENT COMBINATION OF DIFFERENT METHODS, DATA SPACES, AND FEATURE SPACES

| Method | Data Space | Feature Space | $R_2$ statistics | |
|---|---|---|---|---|
| | | | *a* | *v* |
| **MLR** | AV | Psy15 | 56.8% | 10.9% |
| **BoostR** | AV | Psy15 | 55.3% | 11.7% |
| **SVR** | AV | Psy15 | 57.0% | 22.2% |
| **SVR** | PC | $RRF_{18,15}$ | 58.3% | 28.1% |
| **Test-retest** | [1] N/A | N/A | 80.5% | 58.6% |

## 6. CONCLUSION

In this paper, a music selection is quantified as a point in the arousal-valence emotion plane. This continuous view of

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

*International Journal of Advanced Research in Computer and Communication Engineering*
**Vol. 4, Issue 1**, *January 2015*

music emotion makes the proposed MER system free of the inherent ambiguity issue from which conventional categorical approaches suffer. In addition, because there is more freedom in describing a song, the subjectivity issue is alleviated to some extent.

The accuracy of the AV computation determines the viability of the MER system. We formulate MER as a regression problem and adopt the support vector regression for direct estimation of the AV values. Comparing to existing AV computation algorithms, the regression approach has a sound theoretical foundation, exhibits promising prediction accuracy, and needs no temporal information or geometric operations.

Future work will focus on exploiting features about lyrics and singing of a song, evaluating the regression approach on a large-scale database, and realizing the GWMER scheme to further address the subjectivity issue.

## REFERENCES

[1] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," *Proc. ACM SIGIR*, pp. 375–376, 2003

[2] T. Li and M. Ogihara, "Content-based music similarity search and emotion detection," *Proc. Int. Conf. Acoustic, Speech, and Signal Processing*, Toulouse, France, pp. 17–21, 2006.

[3] L. Lu, D. Liu, and H.-J. Zhang, "Automatic mood detection and tracking of music audio signals," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 5–18, 2006.

[4] M.-Y. Wang, N.-Y. Zhang, and H.-C. Zhu, "User-adaptive music emotion recognition," *Proc. Int. Conf. Sig. Processing*, pp. 1352–1355, 2004.

[5] T.-L. Wu and S.-K. Jeng, "Extraction of segments of significant emotional expressions in music," *Proc. Int. Workshop on Computer Music and Audio Technology*, pp. 76–80, 2006.

[6] D. Yang and W. Lee, "Disambiguating music emotion using software agents," *Proc. Int. Conf. Music Information Retrieval*, pp. 52–58, 2004.

[7] V. Carvalho and C. Chao, "Sentiment retrieval in popular music based on sequential learning," *Proc. ACM SIGIR*, 2005.

[8] Y.-H. Yang, C.-C Liu, and H. H. Chen, "Music emotion classification: A fuzzy approach," *Proc. ACM Multimedia*, Santa Barbara, USA, pp. 81–84, 2006.

[9] A. Hanjalic and L.-Q. Xu, "Affective video content representation and modeling," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 143–154, 2005.

[10] M. D. Korhonen, D. A.Clausi, and M. E. Jernigan, "Modeling emotional content of music using system identification," *IEEE Trans. Systems, Man., and Cybernetics*, vol. 36, no. 3, pp. 588– 599, 2006. Data available at: http://www.sauna.org/kiulu/emotion.html. [11] E. Schubert, "Measurement and time series analysis of emotion in music,"

Ph.D. dissertation, School of Music & Music Education, Univ. New South Wales, Sydney, Australia, 1999.

[12] R. E. Thayer, *The Biopsychology of Mood and Arousal*, New York, Oxford University Press, 1989.

[13] P. J. Lang, "The emotion probe, " *American Psychologist*, vol. 50, no. 5, pp. 372–385, 1995.

[14] 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*, vol. 57, no. 3, pp. 493–502, 1989.

[15] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. McMahon, M. Sawey, and M. Schröder, "'FEELTRACE': An instrument for recording perceived emotion in real time," *Proc. Speech and Emotion, ISCA Tutorial and Research Workshop*, Newcastle, U.K., pp. 19–24, 2000.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Recognition*, New York, John Wiley & Sons, Inc., 2000.

[17] A. Sen and M. Srivastava, *Regression Analysis: Theory, Methods, and Applications*, New York, Springer, 1990.

[18] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, 2004.

[19] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001. Available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[20] D.P. Solomatine and D.L. Shrestha, "AdaBoost.RT: A boosting algorithm for regression problems," *Proc. IEEE Int. Joint Conf. Neural Networks*, pp. 1163–1168, 2004.

[21] P. Viola and M. J. Jones, "Robust real-time face detection,"

*Int. Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[22] M. R. Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning Journal*, vol. 53, pp. 23–69, 2003.

[23] L. C. Molina, L. Belanche, and A. Nebot, "Feature selection algorithms: A survey and experimental evaluation," *Proc. IEEE Int. Conf. Data Mining*, pp. 306–313, 2002.

[24] J. M. Keller, M. R. Gray, and J. A. Givens, "Fuzzy k-nearest neighbor algorithm," *IEEE Trans. Systems, Man, and. Cyber*netics, vol. 15, no.4, pp. 580–584, 1985.

[25] D. Tran, M. Wagner, and T. Zheng, "Fuzzy nearest prototype classifier applied to speaker identification," *Proc. ESIT*, 1999.

[26] N. C. Maddage, C. Xu, M. S. Kankanhalli, and X. Shao, "Content-based music structure analysis with applications to music semantics understanding," *Proc. ACM Multimedia*, NY, USA, pp. 112–119, 2004.

[27] D. Cabrera, "PSYSOUND: A computer program for psychoacoustical analysis," *Proc. Australian Acoustic Society Conf.*, pp. 47–54, 1999. Available at: http://www.psysound.org.

[28] J. Ricard, "Towards computational morphological description of sound," Ph. D dissertation, Univ. Pompeu Fabra, Barcelona, 2004.

[29] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech and Audio Processing*, vol. 10, no.5, pp. 293–302, 2002. Available at: http://marsyas.sness.net/.

[30] A. Meng, P. Ahrendt, J. Larsen, and L. K. Hansen, "Temporal feature integration for music genre classification," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1654–1663, 2007.

[31] Musicovery: interactive webRadio, http://www.musicovery.com/ [32] D. C. Montgomery, G. C. Runger, and N. F. Hubele, *Engineering Statistics*, Wiley, 1998.

[33] R. J. Cohen and M. Swerdlik, *Psychological Testing and Measurement: An Introduction to Tests and Measurement*, Mountain View, CA, Mayfield Publishing Company, 1996.

[34] W. Chai and B. Vercoe, "Using user models in music information retrieval systems", *Proc. Int. Symp. Music Information Retrieval*, 2000.

[35] M. Lesaffre, M. Leman, and J.-P. Martens, "A user-oriented approach to music information retrieval," *Content-Based Retrieval*, Dagstuhl SeminarProceedings,2006.