

# COMPUTATIONAL MODELING OF INDUCED EMOTION USING GEMS

**Anna Aljanaki**  
Utrecht University  
A.Aljanaki@uu.nl

**Frans Wiering**  
Utrecht University  
F.Wiering@uu.nl

**Remco C. Veltkamp**  
Utrecht University  
R.C.Veltkamp@uu.nl

## ABSTRACT

Most researchers in the automatic music emotion recognition field focus on the two-dimensional valence and arousal model. This model though does not account for the whole diversity of emotions expressible through music. Moreover, in many cases it might be important to model induced (felt) emotion, rather than perceived emotion. In this paper we explore a multidimensional emotional space, the Geneva Emotional Music Scales (GEMS), which addresses these two issues. We collected the data for our study using a game with a purpose. We exploit a comprehensive set of features from several state-of-the-art toolboxes and propose a new set of harmonically motivated features. The performance of these feature sets is compared. Additionally, we use expert human annotations to explore the dependency between musicologically meaningful characteristics of music and emotional categories of GEMS, demonstrating the need for algorithms that can better approximate human perception.

## 1. INTRODUCTION

Most of the effort in automatic music emotion recognition (MER) is invested into modeling two dimensions of musical emotion: valence (positive vs. negative) and arousal (quiet vs. energetic) (V-A) [16]. Regardless of the popularity of V-A, the question of which model of musical emotion is best has not yet been solved. The difficulty is, on one hand, in creating a model that reflects the complexity and subtlety of the emotions that music can demonstrate, while on the other hand providing a linguistically unambiguous framework that is convenient to use to refer to such a complex non-verbal concept as musical emotion. Categorical models, possessing few (usually 4–6, but sometimes as many as 18) [16] classes are oversimplifying the problem, while V-A has been criticized for a lack of discerning capability, for instance in the case of fear and anger. Other pitfalls of V-A model are that it was not created specifically for music, and is especially unsuited to describe induced (felt) emotion, which might be important for some MER tasks, e.g. composing a playlist using emo-

tional query and in any other cases when the music should create a certain emotion in listener. The relationship between induced and perceived emotion is not yet fully understood, but they are surely not equivalent — one may listen to angry music without feeling angry, but instead feel energetic and happy. It was demonstrated that some types of emotions (especially negative ones) are less likely to be induced by music, though music can express them [17].

In this paper we address the problem of modeling induced emotion by using GEMS. GEMS is a domain-specific categorical emotional model, developed by Zentner et al. [17] specifically for music. The model was derived via a three-stage collection and filtering of terms which are relevant to musical emotion, after which the model was verified in a music listening-context. Being based on emotional ontology which comes from listeners, it must be a more convenient tool to retrieve music than, for instance, points on a V-A plane. The full GEMS scale consists of 45 terms, with shorter versions of 25 and 9 terms. We used the 9-term version of GEMS (see Table 1) to collect data using a game with a purpose.

Emotion induced by music depends on many factors, some of which are external to music itself, such as cultural and personal associations, social listening context, the mood of the listener. Naturally, induced emotion is also highly subjective and varies a lot across listeners, depending on their musical taste and personality. In this paper we do not consider all these factors and will only deal with the question to which extent induced emotion can be modeled using acoustic features only. Such a scenario, when no input from the end-user (except for, maybe, genre preferences) is available, is plausible for a real-world application of a MER task. We employ four different feature sets: low-level features related to timbre and energy, extracted using OpenSmile,<sup>1</sup> and a more musically motivated feature set, containing high-level features, related to mode, rhythm, and harmony, from the MIRToolbox,<sup>2</sup> PsySound<sup>3</sup> and SonicAnnotator.<sup>4</sup> We also enhance the performance of the latter by designing new features that describe the harmonic content of music. As induced emotion is a highly subjective phenomenon, the performance of the model will be confounded by the amount of agreement between listeners which provide the ground-truth. As far as audio-based features are not perfect yet, we try to estimate this upper bound for our data by employing human experts, who an-



© Anna Aljanaki, Frans Wiering, Remco C. Veltkamp. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** Anna Aljanaki, Frans Wiering, Remco C. Veltkamp. “COMPUTATIONAL MODELING OF INDUCED EMOTION USING GEMS”, 15th International Society for Music Information Retrieval Conference, 2014.

<sup>1</sup> opensmile.sourceforge.net

<sup>2</sup> jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox

<sup>3</sup> pssystem.wikidot.com

<sup>4</sup> isophonics.net/SonicAnnotator

notate a subset of the data with ten musicological features.

**Contribution.** This paper explores computational approaches to modeling induced musical emotion and estimates the upper boundary for such a task, in case when no personal or contextual factors can be taken into account. It is also suggested that more than two dimensions are necessary to represent musical emotion adequately. New features for harmonic description of music are proposed.

## 2. RELATED WORK

Music emotion recognition is a young, but fast-developing field. Reviewing it in its entirety is out of scope of this paper. For such a review we are referring to [16]. In this section we will briefly summarize the commonly used methods and approaches that are relevant to this paper.

Automatic MER can be formulated both as a regression and classification problem, depending on the underlying emotional model. As such, the whole entirety of machine learning algorithms can be used for MER. In this paper we are employing Support Vector Regression (SVR), as it demonstrated good performance [7, 15] and can learn complex non-linear dependencies from the feature space. Below we describe several MER systems.

In [15], V-A is modeled with acoustic features (spectral contrast, DWCH and other low-level features from Marsyas and PsySound) using SVR, achieving performance of 0.76 for arousal and 0.53 for valence (in terms of Pearson's  $r$  here and further). In [7], five dimensions (basic emotions) were modeled with a set of timbral, rhythmic and tonal features, using SVR. The performance varied from 0.59 to 0.69. In [5], pleasure, arousal and dominance were modeled with AdaBoost.RM using features extracted from audio, MIDI and lyrics. An approach based on audio features only performed worse than multimodal features approach (0.4 for valence, 0.72 for arousal and 0.62 for dominance).

Various chord-based statistical measures have already been employed for different MIR tasks, such as music similarity or genre detection. In [3], chordal features (longest common chord sequence and histogram statistics on chords) were used to find similar songs and to estimate their emotion (in terms of valence) based on chord similarity. In [9], chordal statistics is used for MER, but the duration of chords is not taken into account, which we account for in this paper. Interval-based features, described here, to our knowledge have not been used before.

A computational approach to modeling musical emotion using GEMS has not been adopted before. In [11], GEMS was used to collect data dynamically on 36 musical excerpts. Listener agreement was very good (Cronbach's alpha ranging from 0.84 to 0.98). In [12], GEMS is compared to a three-dimensional (valence-arousal-tension) and categorical (anger, fear, happiness, sadness, tenderness) models. The consistency of responses is compared, and it is found that GEMS categories have both some of the highest (joyful activation, tension) and some of the lowest (wonder, transcendence) agreement. It was also found that GEMS categories are redundant, and valence

and arousal dimensions account for 89% of variance. That experiment, though, was performed on 16 musical excerpts only, and the excerpts were selected using criteria based on V-A model, which might have resulted in bias.

## 3. DATA DESCRIPTION

The dataset that we analyze consists of 400 musical excerpts (44100 Hz, 128 kbps). Each excerpt is 1 minute long (except for 4 classical pieces which were shorter than 1 minute). It is evenly split (100 pieces per genre) by four genres (classical, rock, pop and electronic music). In many studies, musical excerpts are specially selected for their strong emotional content that best fits the chosen emotional model, and only the excerpts that all the annotators agree upon, are left. In our dataset we maintain a good ecological validity by selecting music randomly from a Creative Commons recording label Magnatune, only making sure that the recordings are of good quality.

Based on conclusions from [11, 12], we renamed two GEMS categories by replacing them with one of their sub-categories (*wonder* was replaced with *amazement*, and *transcendence* with *solemnity*). Participants were asked to select no more than three emotional terms from a list of nine. They were instructed to describe how music made them feel, and not what it expressed, and were encouraged to do so in a game context [1]. All the songs were annotated by at least 10 players (mean = 20.8, SD = 14).

The game with a purpose was launched and advertised through social networks. The game,<sup>5</sup> as well as annotations and audio,<sup>6</sup> are accessible online. More than 1700 players have contributed. The game was streaming music for 138 hours in total. A detailed description and analysis of the data can be found in [1] or in a technical report. [2]

We are not interested in modeling irritation from non-preferred music, but rather differences in emotional perception across listeners that come from other factors. We introduce a question to report disliking the music and discard such answers. We also clean the data by computing Fleiss's kappa on all the annotations for every musical excerpt, and discarding the songs with negative kappa (this indicates that the answers are extremely inconsistent (33 songs)). Fleiss's kappa is designed to estimate agreement, when the answers are binary or categorical. We use this very loose criteria, as it is expected to find a lot of disagreement. We retain the remaining 367 songs for analysis.

The game participants were asked to choose several categories from a list, but for the purposes of modeling we translate the annotations into a continuous space by using the following equation:

$$\text{score}_{ij}^1 = \frac{1}{n} \sum_{k=1}^n a_k, \quad (1)$$

where  $\text{score}_{ij}^1$  is an estimated value of emotion  $i$  for song  $j$ ,  $a_k$  is the answer of the  $k$ -th participant on a question whether emotion  $i$  is present in song  $j$  or not (answer is

<sup>5</sup> www.emotify.org

<sup>6</sup> www.projects.science.uu.nl/memotion/emotifydata/

	C1	C2	C3
Amazement	0.01	-0.73	-0.07
Solemnity	-0.07	0.12	0.89
Tenderness	0.75	0.19	-0.22
Nostalgia	0.57	0.46	-0.41
Calmness	0.80	0.22	0.28
Power	-0.80	-0.17	-0.06
Joyful activation	-0.37	-0.74	-0.32
Tension	-0.72	0.20	0.30
Sadness	0.13	0.80	-0.05

Table 1. PCA on the GEMS categories.

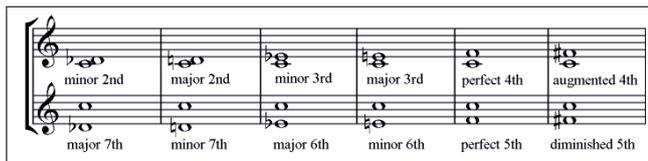


Figure 1. Intervals and their inversions.

either 0 or 1), and  $n$  is the total number of participants, who listened to song  $j$ .

The dimensions that we obtain are not orthogonal: most of them are somewhat correlated. To determine the underlying structure, we perform Principal Components Analysis. According to a Scree test, three underlying dimensions were found in the data, which together explain 69% of variance. Table 1 shows the three-component solution rotated with varimax. The first component, which accounts for 32% of variance, is mostly correlated with calmness vs. power, the second (accounts for 23% of variance) with joyful activation vs. sadness, and the third (accounts for 14% of variance) with solemnity vs. nostalgia. This suggests that the underlying dimensional space of GEMS is three-dimensional. We might suggest that it resembles valence-arousal-triviality model [13].

#### 4. HARMONIC FEATURES

It has been repeatedly shown that valence is more difficult to model than arousal. In this section we describe features, that we added to our dataset to improve prediction of modality in music.

Musical chords, as well as intervals are known to be important for affective perception of music [10], as well as other MIR tasks. Chord and melody based features have been successfully applied to genre recognition of symbolically represented music [8]. We compute statistics on the intervals and chords occurring in the piece.

##### 4.1 Interval Features

We segment audio, using local peaks in the harmonic change detection function (HCDF) [6]. HCDF describes tonal centroid fluctuations. The segments that we obtain are mostly smaller than 1 second and reflect single notes, chords or intervals. Based on the wrapped chromagrams

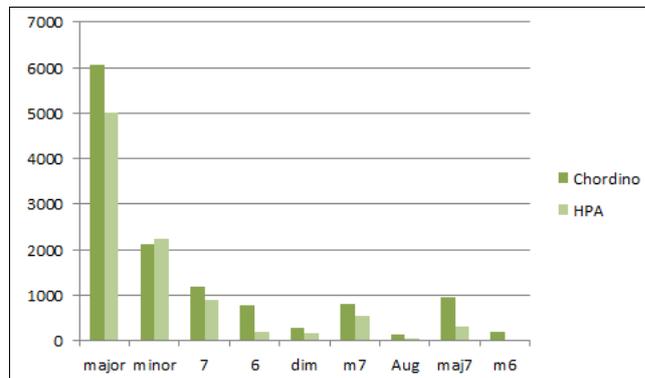


Figure 2. Distribution of chords (Chordino and HPA).

computed from the spectrum of this segments, we select two highest (energy-wise) peaks and compute the interval between them. For each interval, we compute its combined duration, weighted by its loudness (expressed by energy of the bins). Then, we sum up this statistics for intervals and their inversions. Figure 1 illustrates the concept (each bar corresponds to the musical representation of a feature that we obtain). As there are 6 distinct intervals with inversions, we obtain 6 features. We expect that augmented fourths and fifths (tritone) could reflect tension, contrary to perfect fourths and fifths. The proportion of minor thirds and major sixths, as opposed to proportion of major thirds and minor sixths, could reflect the modality. The interval-inversion pairs containing seconds are rather unrestful.

##### 4.2 Chord Features

To extract chord statistics, we used 2 chord extraction tools, HPA<sup>7</sup> (Harmonic Progression Analyzer) and Chordino<sup>8</sup> plugins for Sonic Annotator. The first plugin provides 8 types of chords: major, minor, seventh, major and minor seventh, diminished, sixth and augmented. The second plugin, in addition to these eight types, also provides minor sixth and slash chords (chords for which bass note is different from the tonic, and might as well not belong to the chord). The chords are annotated with their onsets and offsets. After experimentation, only the chords from Chordino were left, because those demonstrated more correlation with the data. We computed the proportion of each type of chord in the dataset, obtaining nine new features. The slash chords were discarded by merging them with their base chord (e.g., Am/F chord is counted as a minor chord). The distribution of chords was uneven, with major chords being in majority (for details see Figure 2). Examining the accuracy of these chord extraction tools was not our goal, but the amount of disagreement between the two tools could give an idea about that (see Figure 2). From our experiments we concluded that weighting the chords by their duration is an important step, which improves the performance of chord histograms.

<sup>7</sup> patterns.enm.bris.ac.uk/hpa-software-package

<sup>8</sup> isophonics.net/npls-chroma

	Tempo	Articulation	Rhythmic complexity	Mode	Intensity	Tonalness	Pitch	Melody	Rhythmic clarity
Amazement	0.50	-0.37	*0.27	**0.24	*0.27				
Solemnity	-0.44	0.39		-0.45					-0.34
Tenderness	-0.48	0.56		0.30	-0.48	*0.29	0.44	0.54	
Nostalgia	-0.47	-0.57			-0.30	*0.28	*0.27	0.50	
Calmness	-0.64	0.48			-0.50		0.36		
Power	0.39	-0.35		*-0.27	0.51		-0.47	-0.43	
Joyful activation	0.76	-0.70	*0.27	**0.24	0.41				0.31
Tension		-0.36		-0.36		-0.47	-0.44	-0.66	
Sadness	-0.45	0.51	-0.38	**0.23	**0.24			*0.27	

**Table 2.** Correlations between manually assessed factors and emotional categories.

## 5. MANUALLY ASSESSED FEATURES

In this section we describe an additional feature set that we composed using human experts, and explain the properties of GEMS categories through perceptual musically motivated factors. Because of huge time load that manual annotation creates we only could annotate part of the data (60 pieces out of 367).

### 5.1 Procedure

Three musicians (26–61 years, over 10 years of formal musical training) annotated 60 pieces (15 pieces from each genre) from the dataset with 10 factors, on a scale from 1 to 10. The meaning of points on the scale was different for each factor (for instance, for tempo 1 would mean ‘very slow’ and 10 would mean ‘very fast’). The list of factors was taken from the study of Wedin [13]: tempo (slow—fast), articulation (staccato—legato), mode (minor—major), intensity (pp—ff), tonalness (atonal—tonal), pitch (bass—treble), melody (unmelodious—melodious), rhythmic clarity (vague—firm). We added rhythmic complexity (simple—complex) to this list, and eliminated style (date of composition) and type (serious—popular) from it.

### 5.2 Analysis

After examining correlations with the data, one of the factors was discarded as non-informative (simple or complex harmony). This factor lacked consistency between annotators as well. Table 2 shows the correlations (Spearman’s  $\rho$ ) between manually assessed factors and emotional categories. We used a non-parametric test, because distribution of emotional categories is not normal, skewed towards smaller values (emotion was more often not present than present). All the correlations are significant with p-value < 0.01, except for the ones marked with asterisk, which are significant with p-value < 0.05. The values that are absent or marked with double asterisks failed to reach statistical significance, but some of them are still listed, because they illustrate important trends which are very probable to reach significance should we have more data.

Many GEMS categories were quite correlated (*tenderness* and *nostalgia*:  $r = 0.5$ , *tenderness* and *calmness*:

$r = 0.52$ , *power* and *joyful activation*:  $r = 0.4$ ). All of these have, however, musical characteristics that allow listeners to differentiate them, as we will see below.

Both *nostalgia* and *tenderness* correlate with slow tempo and legato articulation, but *tenderness* is also correlated with higher pitch, major mode, and legato articulation (as opposed to staccato for *nostalgia*). *Calmness* is characterized by slow tempo, legato articulation and smaller intensity, similarly to *tenderness*. But *tenderness* features a correlation with melodiousness and major mode as well. Both *power* and *joyful activation* are correlated with fast tempo, and intensity, but *power* is correlated with minor mode and *joyful activation* with major mode.

As we would expect, *tension* is strongly correlated with non-melodiousness and atonality, lower pitch and minor mode. *Sadness*, strangely, is much less correlated with mode, but it more characterized by legato articulation, slow tempo and smaller rhythmic complexity.

## 6. EVALUATION

### 6.1 Features

We use four toolboxes for MIR to extract features from audio: MIRToolbox, OpenSmile, PsySound and two VAMP plugins for SonicAnnotator. We also extract harmonic features, described in Section 4. These particular tools are chosen because the features they provide were specially designed for MER. MIRToolbox was conceived as a tool for investigating a relationship between emotion and features in music. OpenSmile combines features from Speech Processing and MIR and demonstrated good performance on cross-domain emotion recognition [14]. We evaluate three following computational and one human-assessed feature sets:

1. **MIRToolbox + PsySound**: 40 features from MIRToolbox (spectral features, HCDF, mode, inharmonicity etc.) and 4 features related to loudness from PsySound (using the loudness model of Chalupper and Fastl).
2. **OpenSmile**: 6552 low-level supra-segmental features (chroma features, MFCCs or energy, and statistical

Feature set	MIRToolbox + PsySound		OpenSmile		MP + Harm		Musicological	
	$r$	RMSE	$r$	RMSE	$r$	RMSE	$r$	RMSE
Amazement	.07 ± .18	.99 ± .16	.19 ± .15	.95 ± .13	.16 ± .15	1.05 ± .11	.35 ± .30	.85 ± .24
Solemnity	.35 ± .14	.80 ± .09	.42 ± .16	.95 ± .13	.43 ± .08	.89 ± .15	.60 ± .24	.84 ± .22
Tenderness	.50 ± .10	.84 ± .10	.52 ± .12	.95 ± .07	.57 ± .12	.85 ± .18	.87 ± .09	.50 ± .19
Nostalgia	.53 ± .16	.82 ± .12	.53 ± .18	.89 ± .07	.45 ± .12	.88 ± .10	.69 ± .24	.69 ± .16
Calmness	.55 ± .14	.83 ± .09	.55 ± .16	.89 ± .07	.60 ± .11	.78 ± .09	.71 ± .17	.70 ± .16
Power	.48 ± .18	.82 ± .13	.56 ± .09	.84 ± .09	.56 ± .11	.80 ± .16	.65 ± .13	.78 ± .26
Joyful activation	.63 ± .08	.77 ± .11	.68 ± .08	.80 ± .08	.66 ± .12	.75 ± .11	.74 ± .28	.58 ± .15
Tension	.38 ± .14	.87 ± .20	.41 ± .19	.94 ± .19	.46 ± .11	.85 ± .13	.58 ± .35	.71 ± .36
Sadness	.41 ± .13	.87 ± .11	.40 ± .18	.96 ± .18	.42 ± .13	.88 ± .12	.39 ± .28	.93 ± .20

**Table 3.** Evaluation of 4 feature sets on the data. Pearson’s  $r$  and RMSE with their standard deviations (across cross-validation rounds) are shown.

functionals applied to them (such as mean, standard deviation, inter-quartile range, skewness, kurtosis etc.).

- MP+Harm:** to evaluate performance of harmonic features, we add them to the first feature set. It doesn’t make sense to evaluate them alone, because they only cover one aspect of music.
- Musicological feature set:** these are 9 factors of music described in section 5.

## 6.2 Learning Algorithm

After trying SVR, Gaussian Processes Regression and linear regression, we chose SVR (the LIBSVM implementation<sup>9</sup>) as a learning algorithm. The best performance was achieved using the RBF kernel, which is defined as follows:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad (2)$$

where  $\gamma$  is a parameter given to SVR. All the parameters,  $C$  (error cost), epsilon (slack of the loss function) and  $\gamma$ , are optimized with grid-search for each feature set (but not for each emotion). To select an optimal set of features, we use recursive feature elimination (RFE). RFE assigns weights to features based on output from a model, and removes attributes until performance is no longer improved.

## 6.3 Evaluation

We evaluate the performances of the four systems using 10-fold cross-validation, splitting the dataset by artist (there are 140 distinct artists per 400 songs). If a song from artist A appears in the training set, there will be no songs from this artist in the test set. Table 3 shows evaluation results. The accuracy of the models differs greatly per category, while all the feature sets demonstrate the same pattern of success and failure (for instance, perform badly on *amazement* and well on *joyful activation*). This reflects the fact that these two categories are very different in their subjectiveness. Figure 3 illustrates the performance of the

systems ( $r$ ) for each of the categories and Cronbach’s alpha (which measures agreement) computed on listener’s answers (see [1] for more details), and shows that they are highly correlated. The low agreement between listeners results in conflicting cues, which limit model performance.

In general, the accuracy is comparable to accuracy achieved for perceived emotion by others [5, 7, 15], though it is somewhat lower. This might be explained by the fact that all the categories contain both arousal and valence components, and induced emotion annotations are less consistent. In [7], *tenderness* was predicted with  $R = 0.67$ , as compared to  $R = 0.57$  for **MP+Harm** system in our case. For *power* and *joyful activation*, the predictions from the best systems (**MP+Harm** and **OpenSmile**) demonstrated 0.56 and 0.68 correlation with the ground truth, while in [5, 15] it was 0.72 and 0.76 for arousal.

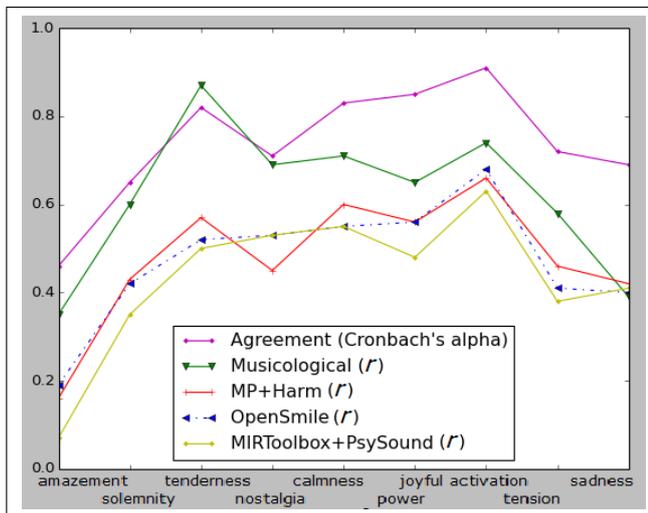
The performance of all the three computational models is comparable, though **MP+Harm** model performs slightly better in general. Adding harmonic features improves average performance from 0.43 to 0.47, and performance of the best system (**MP+Harm**) decreases to 0.35 when answers from people who disliked the music are not discarded. As we were interested in evaluating the new features, we checked which features were considered important by RFE. For *power*, the tritone proportion was important (positively correlated with power), for *sadness*, the proportion of minor chords, for *tenderness*, the proportion of seventh chords (negatively correlates), for *tension*, the proportion of tritones, for *joyful activation*, the proportion of seconds and inversions (positive correlation).

The **musicological** feature set demonstrates the best performance as compared to all the features derived from signal-processing, demonstrating that our ability to model human perception is not yet perfect.

## 7. CONCLUSION

We analyze the performance of audio features on prediction of induced musical emotion. The performance of the best system is somewhat lower than can be achieved for perceived emotion recognition. We conduct PCA and find

<sup>9</sup> www.csie.ntu.edu.tw/~cjlin/libsvm/



**Figure 3.** Comparison of systems' performance with Cronbach's alpha per category.

three dimensions in the GEMS model, which are best explained by axes spanning *calmness—power*, *joyful activation—sadness* and *solemnity—nostalgia*). This finding is supported by other studies in the field [4, 13].

We conclude that it is possible to predict induced musical emotion for some emotional categories, such as *tenderness* and *joyful activation*, but for many others it might not be possible without contextual information. We also show that despite this limitation, there is still room for improvement by developing features that can better approximate human perception of music, which can be pursued in future work on emotion recognition.<sup>10</sup>

## 8. REFERENCES

- [1] A. Aljanaki, D. Bountouridis, J.A. Burgoyne, J. van Balen, F. Wiering, H. Honing, and R. C. Veltkamp: "Designing Games with a Purpose for Data Collection in Music Research. Emotify and Hooked: Two Case Studies", *Proceedings of Games and Learning Alliance Conference*, 2013.
- [2] A. Aljanaki, F. Wiering, and R. C. Veltkamp: "Collecting annotations for induced musical emotion via online game with a purpose Emotify", [www.cs.uu.nl/research/techreps/UU-CS-2014-015.html](http://www.cs.uu.nl/research/techreps/UU-CS-2014-015.html), 2014.
- [3] H.-T. Cheng, Y.-H. Yang, Y.-C. Lin, I.-B. Liao, and H. H. Chen: "Automatic chord recognition for music classification and retrieval", *IEEE International Conference on Multimedia and Expo*, pp. 1505–1508, 2008.
- [4] J. R. J. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth: "The World of Emotions is not Two-Dimensional", *Psychological Science*, Vol. 18, No. 12, pp. 1050–1057, 2007.
- [5] D. Guan, X. Chen, and D. Yang: "Music Emotion Regression Based on Multi-modal Features", *CMMR*, p. 70–77, 2012.
- [6] C. A. Harte, and M. B. Sandler: "Detecting harmonic change in musical audio", *Proceedings of Audio and Music Computing for Multimedia Workshop*, 2006.
- [7] C. Laurier, O. Lartillot, T. Eerola, and P. Toivainen: "Exploring Relationships between Audio Features and Emotion in Music", *Conference of European Society for the Cognitive Sciences of Music*, 2009.
- [8] C. McKay, and I. Fujinaga: "Automatic genre classification using large high-level musical feature sets", *In Int. Conf. on Music Information Retrieval*, pp. 525–530, 2004.
- [9] 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, Special Issue on Scalable Audio-Content Analysis* pp. 735–854, 2010.
- [10] B. Sollberge, R. Rebe, and D. Eckstein: "Musical Chords as Affective Priming Context in a Word-Evaluation Task", *Music Perception: An Interdisciplinary Journal*, Vol. 20, No. 3, pp. 263–282, 2003.
- [11] K. Torres-Eliard, C. Labbe, and D. Grandjean: "Towards a dynamic approach to the study of emotions expressed by music", *Proceedings of the 4th International ICST Conference on Intelligent Technologies for Interactive Entertainment*, pp. 252–259, 2011.
- [12] J. K. Vuoskoski, and T. Eerola: "Domain-specific or not? The applicability of different emotion models in the assessment of music-induced emotions", *Proceedings of the 10th International Conference on Music Perception and Cognition*, pp. 196–199, 2010.
- [13] L. Wedin: "A Multidimensional Study of Perceptual-Emotional Qualities in Music", *Scandinavian Journal of Psychology*, Vol. 13, pp. 241–257, 1972.
- [14] F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro, and K. R. Scherer: "On the Acoustics of Emotion in Audio: What Speech, Music, and Sound have in Common", *Front Psychol*, Vol. 4, p. 292, 2013.
- [15] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. H. Chen: "A Regression Approach to Music Emotion Recognition", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 2, pp. 448–457, 2008.
- [16] Y.-H. Yang, and H. H. Chen: "Machine Recognition of Music Emotion: A Review", *ACM Trans. Intell. Syst. Technol.*, Vol. 3, No. 3, pp. 1–30, 2012.
- [17] M. Zentner, D. Grandjean, and K. R. Scherer: "Emotions evoked by the sound of music: characterization, classification, and measurement", *Emotion*, Vol. 8, No. 4, pp. 494–521, 2008.

<sup>10</sup> This research was supported by COMMIT/.