Automatic Mood Detection from Acoustic Music Data

Dan Liu, Lie Lu, Hong-Jiang Zhang

Contact: dliu@Cogsci.ucsd.edu

Outline

- Introduction
- Feature Extraction
- Mood Detection
- Mood Tracking
- Experiment
- Conclusion

1. Introduction

- Music Mood
  - A semantic metadata to archive music from database
  - Objective or Subjective
    - Depend on many factors such as culture, education, experience…
    - Consistent within a given cultural context

1. Introduction

- Relevant works
  - Concentrate on MIDI or symbolic representations
  - Use various mood descriptors

1. Introduction

- Music Mood Taxonomy
  - Hevner's adjective checklist (1935)

- Descriptors are ambiguity
- Difficult for computational modeling

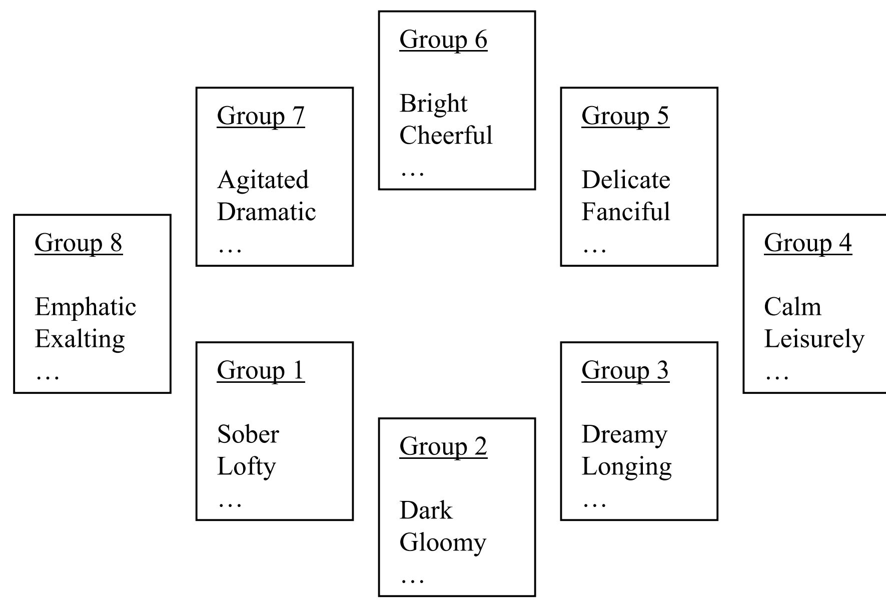

Figure 1: Hevner's adjective checklist

1. Introduction

- Music Mood Taxonomy
  - Thayer's two-dimensional model (1990)
    - Descriptors are explicit and discriminatable
    - Easier for computational modeling

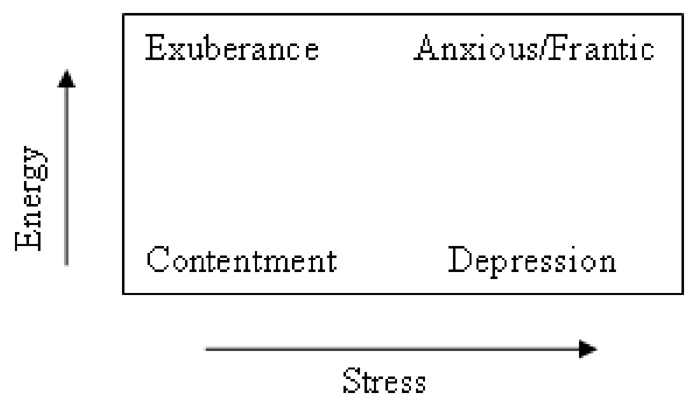

Figure 2: Thayer's model of mood

2. Feature Extraction

- Relevant Music Features
  - Intensity Features
  - Timbre Features
  - Rhythm Features
  - Mode Features (not available)

2. Feature Extraction

- Timbre Features
  - Spectral Shape Features

    (centroid, bandwidth, roll off, spectral flux)
  - Spectral Contrast Features
    - Sub-band Peak
    - Sub-band Valley
    - Sub-band Average

2. Feature Extraction

- Intensity Features
  - Sub-band Intensity: root mean-square (RMS) in each sub-band
  - Total Intensity: sum of sub-band Intensity

2. Feature Extraction

- Rhythm Features

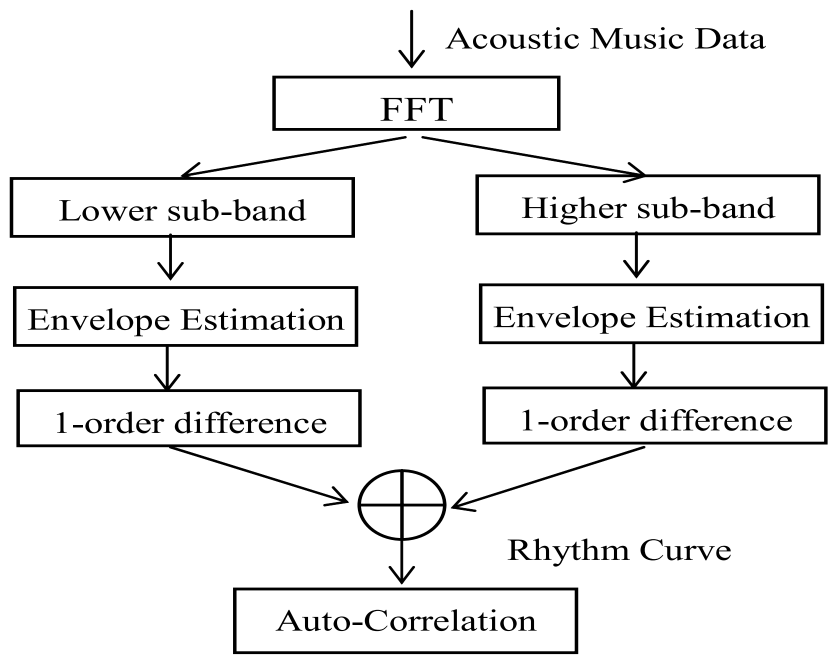

Figure 3: Rhythm features extraction

2. Feature Extraction

- Rhythm Features
  - Average Strength: average strength of bass instrumental onsets.
  - Average Correlation Peak: average of the maximum three peaks in the auto-correlation curve.
  - Average Tempo: the common divisor of the peaks of the auto-correlation curve.

3. Mood Detection

- Hierarchical Framework

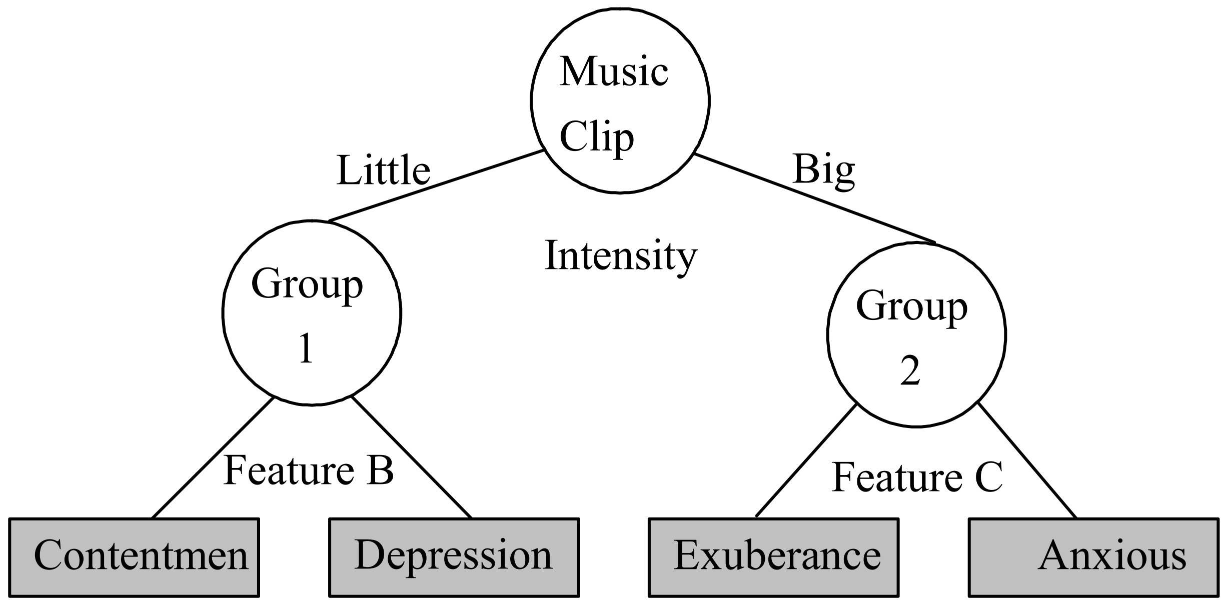

Figure 4: The hierarchical mood detection framework

3. Mood Detection

- Hierarchical Process

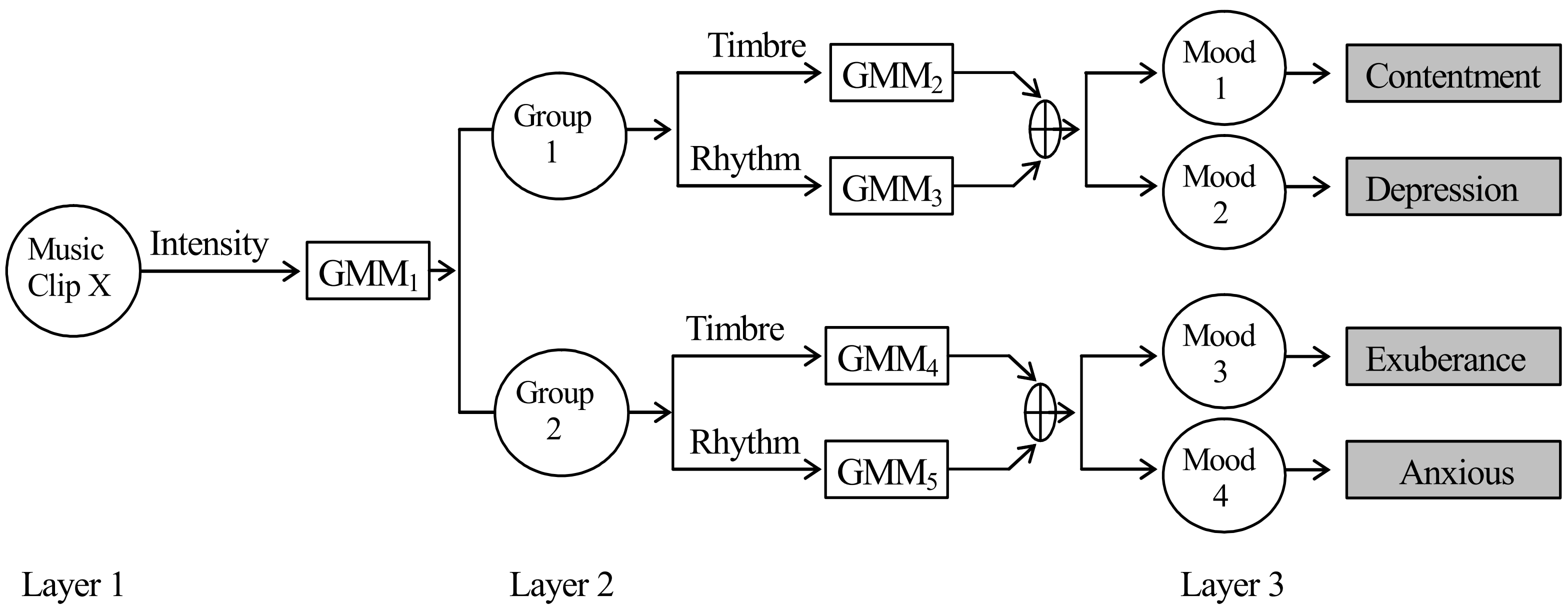

Figure 5: The hierarchical Mood detection process

3. Mood Detection

Step1. Group Classification

$$\frac{P(G_1 \mid I)}{P(G_2 \mid I)} \begin{cases} \geq 1, & Select \quad G_1 \\ < 1, & Select \quad G_2 \end{cases}$$

Step 2. Mood classification in each group

$$P(M_j | G_1, T, R) = \lambda_1 \times P(M_j | T) + (1 - \lambda_1) \times P(M_j | R) \quad j = 1,2$$

$$P(M_j | G_2, T, R) = \lambda_2 \times P(M_j | T) + (1 - \lambda_2) \times P(M_j | R) \quad j = 3,4$$

4. Mood Tracking

- Why we need to track the mood
  - Mood is changeable in music
- How to track the changeable mood
  - Segmentation based on music features (timbre and intensity)
  - Mood detection in each segment

4. Mood Tracking

- Segmentation Procedure
  1) Compute the distance between two contiguous windows based on timbre and intensity features

$$D = \frac{1}{2} tr[(C_i - C_j)(C_j^{-1} - C_i^{-1})]$$

  2) Compute confidence of being a boundary

$$Conf_I = \frac{1}{A_I}\exp(\frac{D_I - \mu_I}{\sigma_I}), \quad Conf_T = \frac{1}{A_T}\exp(\frac{D_T - \mu_T}{\sigma_T})$$

$$Conf = \alpha \times Conf_I + (1 - \alpha) \times Conf_T$$

4. Mood Tracking

- Segmentation Procedure

3) Detect potential boundaries

$$I.\quad Conf(i, i+1) > Conf(i+1, i+2)$$

$$II.\quad Conf(i, i+1) > Conf(i-1, i)$$

$$III.\quad Conf(i, i+1) > Th_i$$

$$Th_i = \alpha \times \frac{1}{2 \times N} \sum_{n=-N}^{N} Conf(i-n-1, i-n)$$

4) Refine potential boundaries

5. Experiment

- Mood Detection on Music Clips
  - Database
    - 250 pieces of music, mainly in the classical period and romantic period
    - 200 representative music clips of 20 seconds long for each of the four mood clusters
  - Experiment
    - Cross-validation evaluation with 25% used for testing and 75% for training.
    - Iterated with different random partitions and the results are averaged

5. Experiment

- Experiment results on hierarchical framework

(1) Optimal average accuracy achieved when

$$\lambda_1(\text{weighting of Timbre in Group } 1) = 0.8$$
$$\lambda_2(\text{weighting of Timbre in Group } 1) = 0.4$$

Timbre features are more important to classify Contentment and Depression in Group 1, and rhythm features are more important to discriminate Exuberance and Anxious in Group 2.

5. Experiment

– Experiment result on hierarchical framework

(2)Only 1.6% music in Group 1 (Contentment and Depression) is classified into Group 2 (Exuberance and Anxious), while only 0.4% music in Group 2 is classified into Group 1

This result confirms the good performance of intensity features in discriminating the two groups of mood clusters.

5. Experiment

- Comparison of hierarchical framework and non-hierarchical framework (See Results)
  - Overall classification accuracy for hierarchical framework is up to 86.3%, about 5.7% better than the non-hierarchical framework, and its standard deviation decreases from 10.7% to 5.2%.
  - Classification accuracies for all of the four clusters are improved by using hierarchical framework, especially for Exuberance (85.5% improved from 64.7%).

  Hierarchical framework has a better performance than its non-hierarchical counterpart, by using the most efficient features for different mood clusters.

5. Experiment

- Mood Tracking
  - Haydn's "Serenade" : constantly Contentment
  - Second movement of Beethoven's "Symphony No. 3": mainly Depression
  - Tchaikovsky's "1812 Overture": changeable

Mood tracking performance based on segmentation is better than that of detecting mood every 20 seconds.

6. Conclusion

- Thayer's model of mood is adopted for mood taxonomy
- Intensity, timbre and rhythm feature sets are extracted directly from acoustic data.
- A hierarchical framework is developed to detect the mood in a music clip.
- A segmentation scheme is presented to track the mood in a whole piece of music.

Thank You !

10/27/2003

Table 1: Mood detection confusion matrix based on hierarchical framework

	Contentment	Depression	Exuberance	Anxious
Contentment	**76.6±7.6**	21.8±7.2	0.5±0.8	1.2±1.2
Depression	4.0±3.5	**94.5±3.4**	0±0	1.5±2.5
Exuberance	0±0	0.8±1.3	**85.5±3.2**	13.7±4.8
Anxious	0±0	0±0	11.5±6.7	**88.5±6.7**

Table 2: Mood detection confusion matrix based on non-hierarchical framework

	Contentment	Depression	Exuberance	Anxious
Contentment	**75.0±11.8**	25.0±11.8	0±0	0±0
Depression	5.8±2.6	**94.2±2.6**	0±0	0±0
Exuberance	1.5±2.6	0.7±1.3	**64.7±20.5**	33.0±18.3
Anxious	0±0	0±0	11.5±6.7	**88.3±7.9**

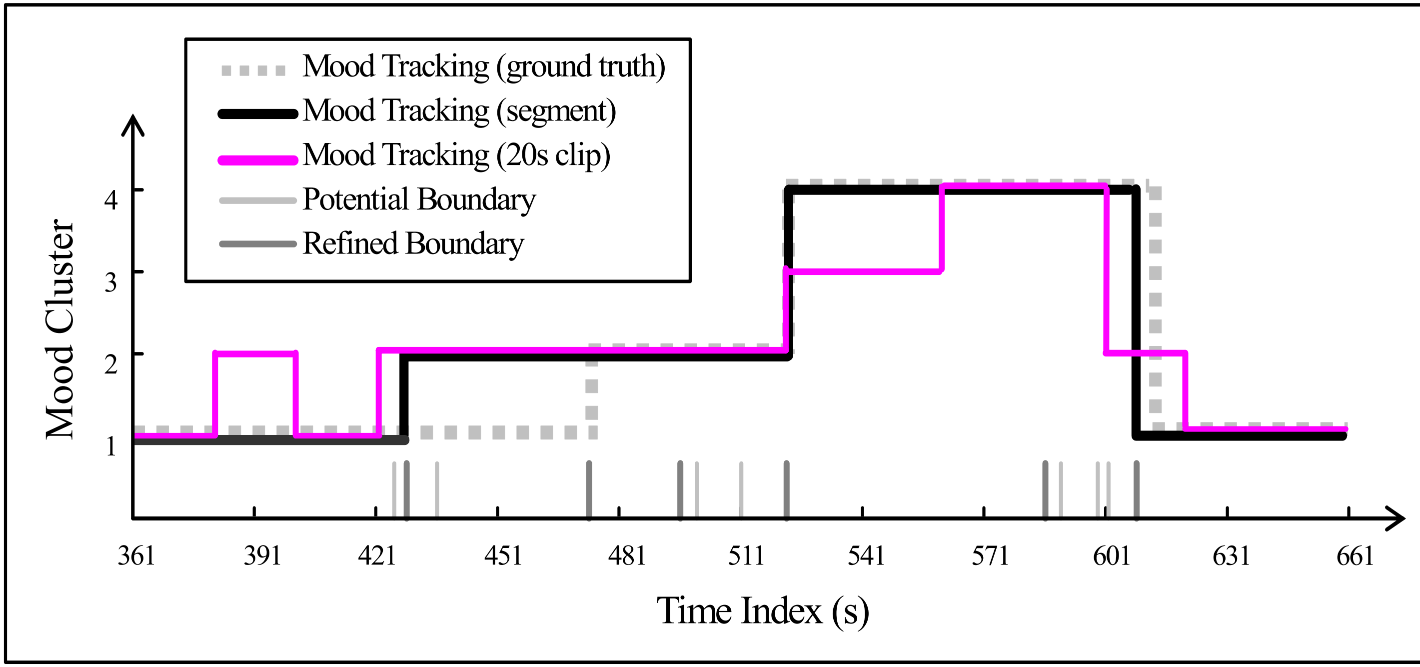

Figure 6: Mood tracking results on a part of “1812 Overture” (from 361s – 661s)