Emotion Based Categorization of Music Using Low Level Features and Agglomerative Clustering

Rajib Sarkar, Saikat Dutta, Aneek Roy and Sanjoy Kumar Saha
Computer Science & Engg. Dept., Jadavpur University

Research Goals

- Emotion perception based acoustic feature selection.
- End goal is to categorize music excerpts according to their emotional properties.

Related Work

- Saari, Eerola and Lartillot (2011):
  - Wrapper selection method to select suitable features from a wide range of acoustic features.
  - Classifier: Naive Bayes, k-NN and SVM.
- Gomez & Caceres (2017):
  - Used features like spectral centroid, spectral roll-off and MFCC.
  - Classifier: k-Nearest Neighbors (kNN).

Feature Selection

- The way musical accents are patterned through time leads listeners to anticipate the emotional essence.
- Feature set is designed considering the relation between emotional response and musical structure.
- Features Considered:
  - RMS Energy, ZCR, Linear prediction cepstral coefficients (LPCC), Spectral Features (Flux, Rolloff, Flatness Measure and Crest Factor).

Classifier

- K-Means clustering:
  - K is taken as the number of emotional groups.
  - Position of cluster-centroids are updated by averaging the points present in respective clusters.
- Agglomerative clustering:
  - a bottom-up approach to hierarchical clustering.
  - Initially assumed that each data points belong to a separate cluster.
  - In each iteration two closest clusters are merged.
  - When number of clusters reaches the number of emotional category the process stops.

Challenges

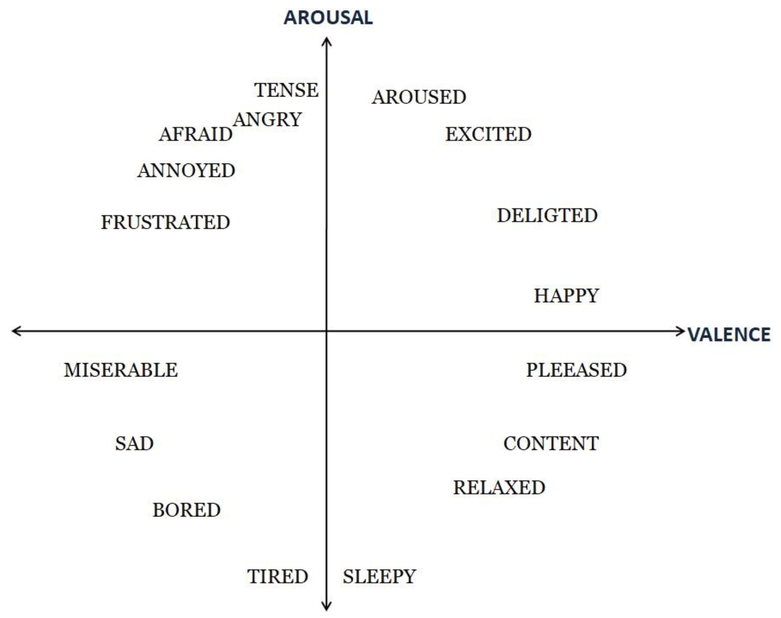

Fig 1: Thayer's 2D Emotion Plane

- Emotion is subjective.
- The way of experiencing emotional feelings is the most difficult to describe or measure.

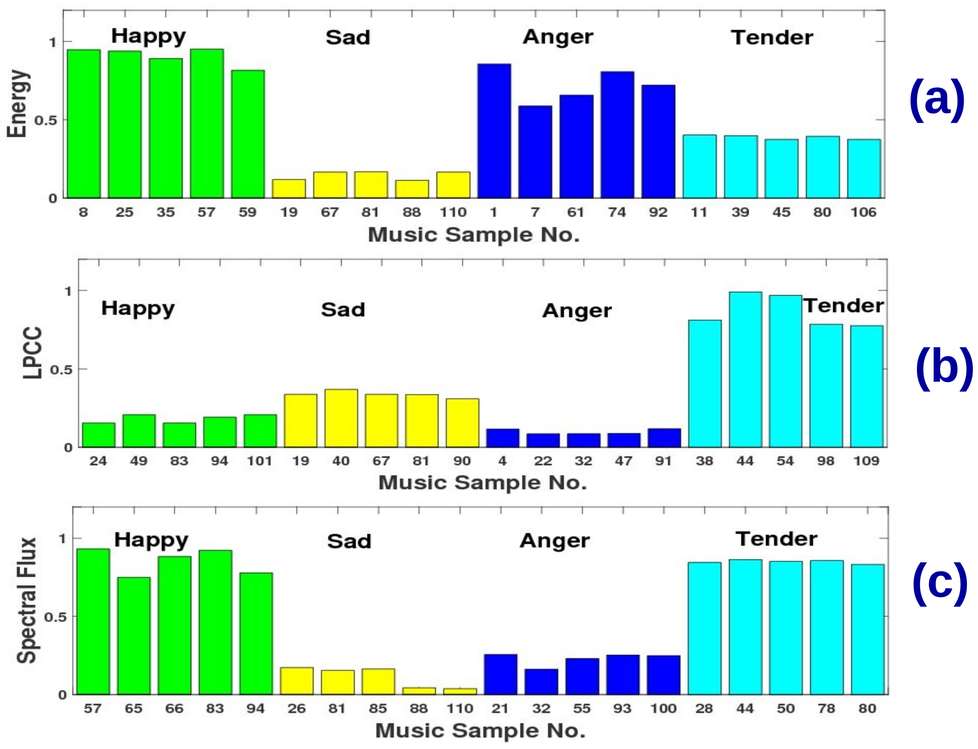

Fig 2: Plot for (a) Energy, (b) LPCC and (c) Spectral Flux for different category of emotion.

Method	Features Used	Accuracy
K-means	Set A – Energy, Spectral Rolloff, ZCR	59.09%
Agglomerative Clustering	Set B – Spectral Features (Crest Factor, Flatness, Flux, Rolloff), LPCC	**63.63%**
SVM	Set A	51.43%
SVM	Set B	58.08%
k-NN BE Saari et al. (2011)*	Mode Majorness, Key clarity, dynamical, rhythmical, structural	56.50%
SVM BE Saari et al. (2011)*	Dynamical, rhythmical, pitch, structural, timbral, Harmony (Wrapper selection)	54.30%

*Saari, P., Eerola, T., Lartillot, O.: Generalizability and simplicity as criteria in feature selection: Application to mood classification in music. IEEE Transactions on Audio, Speech, and Language Processing 19(6) (2011) 1802–1812

NCVPRIPG 2017, IIT Mandi, 16 - 19 December, 2017.