MOODetector: Automatic Music Emotion Recognition

Rui Pedro Paiva, PhD

Researcher @ Proyecto Prometeo, Ecuador
Professor @ University of Coimbra, Portugal

Campus Party Quito 3

September, 2013

PROMETEO
Investigación
Formación
Desarrollo

Outline

- **Introduction**
  - Coimbra
  - University of Coimbra
- **CISUC**
- **Music Data Mining / Music Information Retrieval**
- **Music Emotion Recognition**

Introduction

- **Coimbra**
- **University of Coimbra**

Introduction

Coimbra

Introduction

Coimbra

Introduction

Coimbra

Introduction Coimbra

Introduction

UC – Old Campus

Introduction

UC – Old Campus

Introduction

UC – Old Campus

Introduction

UC – Old Campus

CISUC

UC – New Campus

Introduction

UC – New Campus

CISUC

CISUC

- **Centre for Informatics and Systems of the University of Coimbra**
  - https://www.cisuc.uc.pt/

CISUC

- Large Portuguese research center in the fields of **Informatics and Communications**
- Created in **1991**
- About **200 researchers**, including full-time university professors and under-graduate and post-graduate students
- High level of **internationalization**
  - High number of foreign researchers (mostly PhD students from Brazil, India and Europe)
  - High level of cooperation with research centers abroad

CISUC

- Organized into **6 research groups (**covering a substantial segment of **Computer Science** research topics)
  - **Cognitive and Media Systems**
  - **Adaptive Computation**
  - **Software and Systems Engineering**
  - **Communications and Telematics**
  - **Information Systems**
  - **Evolutionary and Complex Systems**
- Increasing activity on **multidisciplinary and emergent research subjects**

Music Data Mining / Music Information Retrieval

- MIR: What and Why?
- Applications

MIR: What and Why?

- **MDM / MIR**
  - **Music Data Mining / Music Information Retrieval**: **interdisciplinary research field** devoted to the study of **information extraction mechanisms** from musical pieces, **retrieval methodologies**, as well as all the processes involved in those tasks in different music representation media.

- **Why MDM / MIR?**
  - MDM / MIR emerges from the necessity to **manage huge collections of digital music** for "preservation, access, research and other uses" [Futrelle and Downie, 2003]

MIR: What and Why?

- **Music and Man**
  - Music expresses “that which cannot be put into words and that which cannot remain silent” (Victor Hugo)
  - “We associate music with the most unique moments of our lives and music is part of our individual and social imaginary” [Paiva, 2006]
    - “By listening to music, emotions and memories, thoughts and reactions, are awakened” [Paiva, 2006]
  - “Life has a soundtrack” [Gomes, 2005] (“Festivais de Verão”, Jornal “Público”)
  - “The history of a people is found in its songs” (George Jellinek)

MIR: What and Why?

- **Music and World economy**
  - **Explosion of the Electronic Music Industry (EMD)**
    - Widespread access to the Internet
    - Bandwidth increasing in domestic and mobile accesses
    - Compact audio formats with near CD quality (mp3, aac)
    - Portable music devices (iPod, mp3 readers)
    - Peer-to-peer networks (Napster, Kazaa, eMule)
    - Online music stores (iTunes, Calabash Music, Sapo Music) → resolution is the song, not the CD
    - Music identification platforms (Shazam, 411-Song, Gracenote MusicID / TrackID)
    - Music recommendation systems (MusicSurfer)

MIR: What and Why?

- **Music and World economy (cont.)**
  - Music industry runs, only in the USA an amount of money in the order of **several billion US dollars per year**.
  - By 2005, Apple iTunes was selling ≈ 1.25 million songs each day [TechWhack, 2005]
    - Until January 2009, over 6 billion songs had been sold in total [TechCrunch, 2009]
  - By 2007, music shows in Portugal sold 30 M€ in tickets [RTP, 2009]

MIR: What and Why?

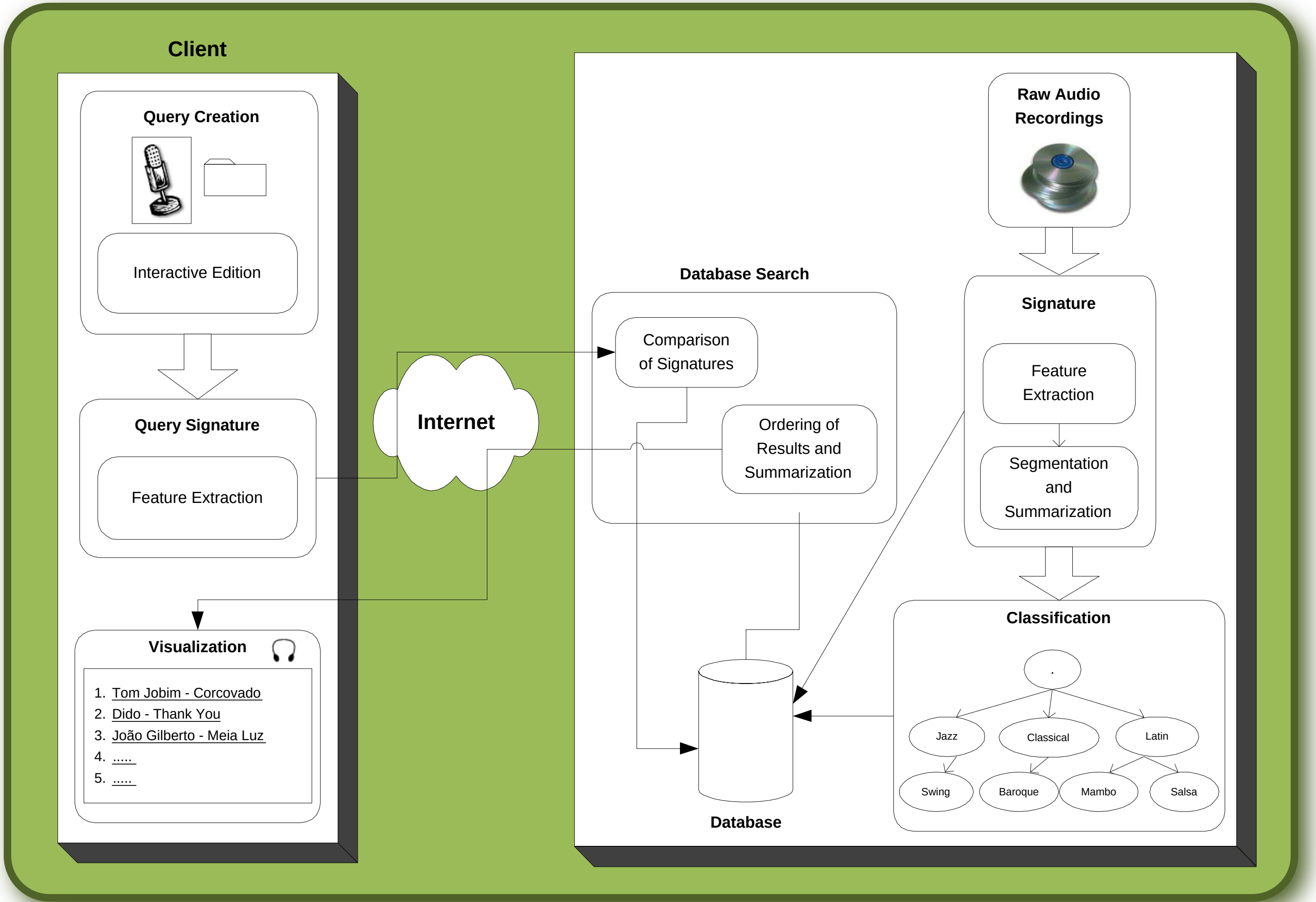
- **Music and World economy (cont.)**
  - **Number and dimension of digital music archives continuously growing**
    - Database size (these days, over 2 million songs)
    - Genres covered
  - Challenges to music providers and music librarians
    - Organization, maintenance, labeling, user interaction
    - Any large music database is only really **useful if users can find what they are looking for in an efficient manner**!

MIR: What and Why?

- **Database Organization and Music Retrieval**
  - **Presently, music databases are manually annotated** → search and retrieval is mostly textual (artist, title, album, genre)
    - Service providers
      - Difficulties regarding manual song labeling: subjective and time-consuming,
    - Customers
      - Difficulties in performing “content-based” queries
        - » “Music’s preeminent functions are social and psychological”, and so “the most useful retrieval indexes are those that facilitate searching in conformity with such social and psychological functions. Typically, such indexes will focus on stylistic, mood, and similarity information” [Huron, 2000].

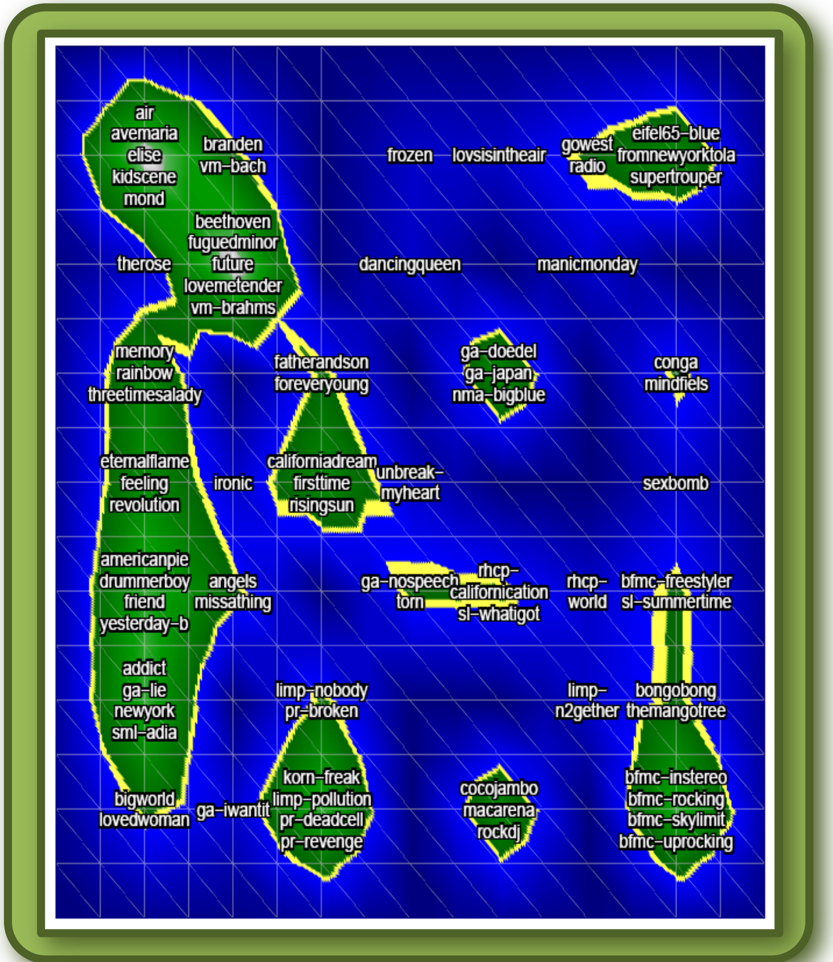
MIR: What and Why?

- **Database Organization and Music Retrieval**



Applications

- **Platforms for EMD**
  - **Similarity-based retrieval tools**
    - **Query-by-example** [Welsh *et al.*, 1999]
      - Music identification (Shazam, trackID, Tunatic) [Wang, 2003; Haitsma and Kalker, 2002; TMN, 2009]
      - Query-by-mood [Panda *et al.*, 2013]
      - Music recommendation [Celma & Lamere, 2007]
      - Islands of music [Pampalk, 2001]
        - » Metaphor of geographic maps: similar genres close together



Applications

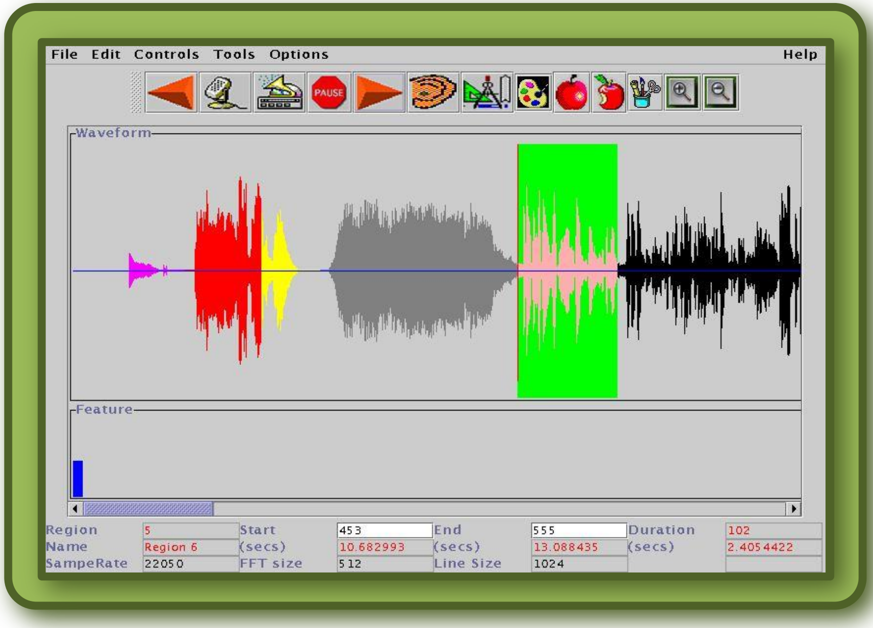
- **Platforms for EMD**
  - Similarity-based retrieval tools
    - **Query-by-melody** (query-by-humming, QBS) [Parker, 2005; Ghias *et al.*, 1995]
    - **Plagiarism detection** [Paiva *et al.*, 2006]
    - **Music web crawlers** [Huron, 2000]
    - **Automatic playlist generation** [Pauws and Wijdeven, 2005; Alghoniemy and Tewfik, 2000]

Applications

- **Music education and training**
  - **Automatic music transcription** [Ryynänen, 2008; Kashino *et al.*, 1995]
    - →Music composition, analysis, performance evaluation, plagiarism detection
- **Digital music libraries**
  - For research issues involving music retrieval, training (learning activities, evaluation, etc.) → Variations [Dunn, 2000]

Applications

- **Audio software**
  - Intelligent audio (music) editors → automatic indexing [Tzanetakis, 2002]



- **Multimedia databases and operating systems [Burad, 2006]**

- **Video indexing and searching**
  - Segmentation based on audio (music) content → detection of scene transitions [Pfeiffer, 1996]

Applications

- **Advertisement and cinema**
  - Tools for mood-based retrieval [Cardoso *et al.* 2011]

- **Sports**
  - Music to induce a certain cardiac frequency [Matesic and Cromartie, 2002]

- **… and so forth…**

Music Emotion Recognition (MER)

- MER: What and Why?
- Applications
- Emotion Models
- MER Data Mining Process
- Other Problems Addressed
- Current Limitations and Open Problems
- Future/Ongoing Work

MER: What and Why?

- **MER**
  - **Music Emotion Recognition**
    - Branch of MIR devoted to the identification of emotions in musical pieces
  - **Emotion vs mood**
    - MIR researchers → use both terms interchangeably
    - Psychologists → clear distinction [Sloboda and Juslin, 2001]
      - Emotion = a short experience in response to an object (e.g., music)
      - Mood = longer experience without specific object connection

MER: What and Why?

- **MER**
  - **Categories of emotions** [Gabrielsson, 2002)]
    - **Expressed emotion**: emotion the performer tries to communicate to the listeners
    - **Perceived emotion**: emotion the listener perceives as being expressed in a song (which may be different than the emotion the performer tried to communicate) → **scope of MIR researchers**
    - **Felt (evoked) emotion**: emotion felt by the listener, in response to the song and performance

MER: What and Why?

- **Why MER?**
  - “**Music’s preeminent functions are social and psychological**”, and so “the most useful retrieval indexes are those that facilitate searching in conformity with such social and psychological functions. Typically, such indexes will focus on stylistic, **mood**, and similarity information” [Huron, 2000].
    - Studies on music information behavior → music mood is an important criterion for music retrieval and organization [Juslin and Laukka, 2004]

MER: What and Why?

- **Why MER?** [Yang and Chen, 2012]
  - **Academia**
    - "**More and more multimedia systems that involve emotion analysis of music** signals have been developed, such as Moodtrack, LyQ, MusicSense, Mood Cloud, Moody and *i*.MTV, just to name a few."
  - **Industry**
    - "**Many music companies**, such as AMG, Gracenote, MoodLogic, Musicovery, Syntonetic, and Sourcetone **use emotion as a cue for music retrieval**."

MER: What and Why?

- **Difficulties [Yang and Chen, 2012]**
  – Emotion perception is by nature **subjective** [Yang and Chen, 2012]
    - People can perceive different emotions for the same song

  – → **Performance evaluation of an MER systems is difficult** [Yang and Chen, 2012]
    - Common agreement on the recognition result is hard to obtain

  – Still **not fully understood how music and emotion are related**
    - Despite several studies on music psychology

Applications

- **Platforms for EMD**
  - Emotion-based retrieval tools
    - Music recommendation, automatic playlist generation, classification, …
- **Game development**
- **Cinema**
- **Advertising**
- **Health**
  - Sports
  - Stress management

Emotion Models

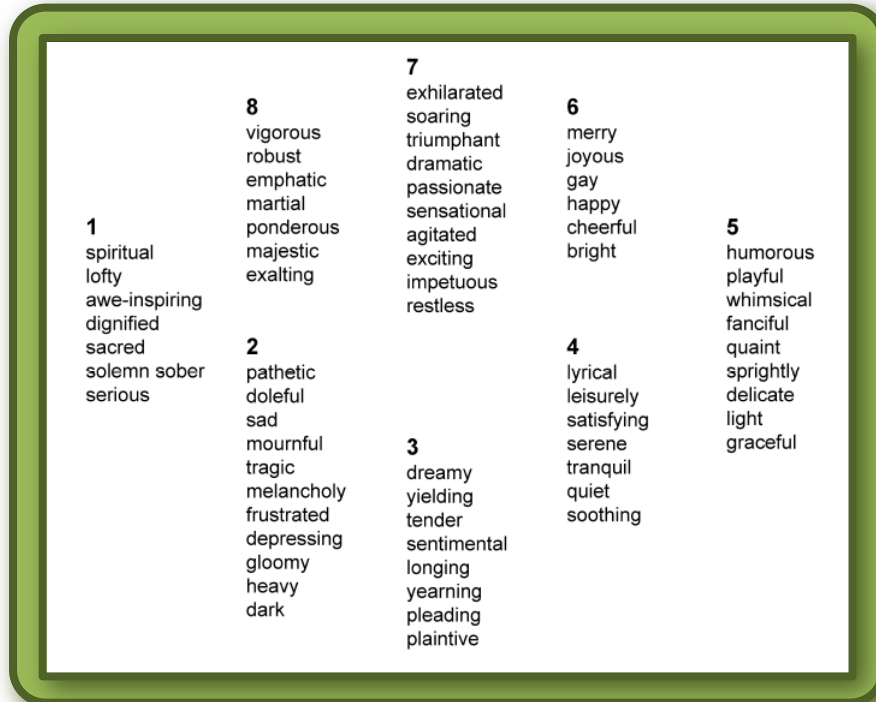
- **Two main conceptualizations of emotion**
  - **Categorical models**
    - Emotions as categories: limited number of **discrete** emotions (adjectives)
  - **Dimensional models**
    - Emotions organized **along axes** (2 or 3)
      - As **discrete** adjectives
      - As **continuous** values

Emotion Models

- **Categorical Models**
  - Main ideas
    - "People experience **emotions as categories** that are distinct from each other" [Yang and Chen, 2012]
    - Existence of **basic** emotions
      - Limited number of **universal and primary emotion classes** (e.g., **happiness, sadness, anger, fear, disgust, surprise** [Ekman, 1992]) from which all other **"secondary" emotion classes** can be derived

Emotion Models

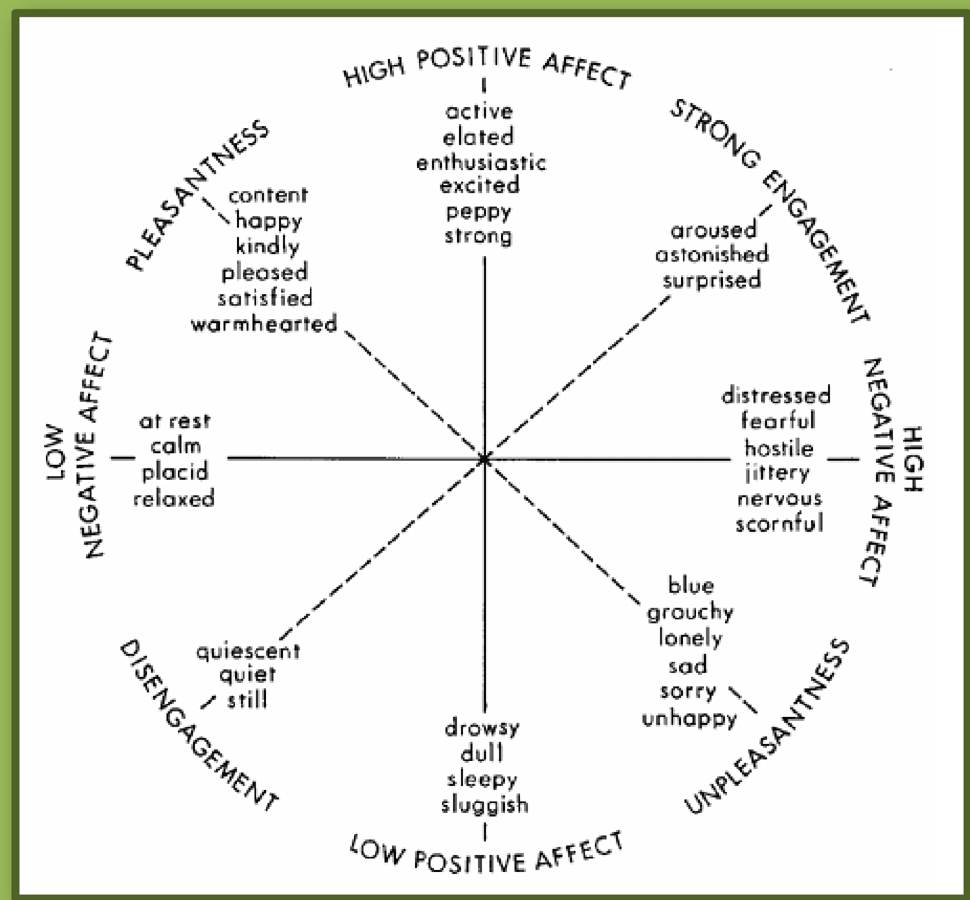
- **Categorical Models**
  - Examples
    - **Hevner's 8 clusters of affective terms** (1935)
    - Regrouped into 10 adjective groups by Farnsworth [Farnsworth, 1954] and into 9 adjective groups by Schubert [Schubert 2003].



© Yang and Chen, 2012

Emotion Models

- **Categorical Models**
  - Examples
    - **Tellegen-Watson-Clark model** (1999)



© Tellegen, Watson and Clark, 1999

Emotion Models

- **Categorical Models**
  - Examples
    - MIREX (2007)
      - 5 clusters, but not supported by psychological models

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Passionate	Rollicking	Literate	Humorous	Aggressive
Rousing	Cheerful	Poignant	Silly	Fiery
Confident	Fun	Wistful	Campy	Tense/anxious
Boisterous	Sweet	Bittersweet	Quirky	Intense
Rowdy	Amiable/good natured	Autumnal	Whimsical	Volatile
		Brooding	Witty	Visceral
			Wry	

Emotion Models

- **Categorical Models**
  - Limitations
    - Limited number of adjectives
    - Larger number may be impractical for psychological studies
    - Adjectives may be ambiguous

Emotion Models

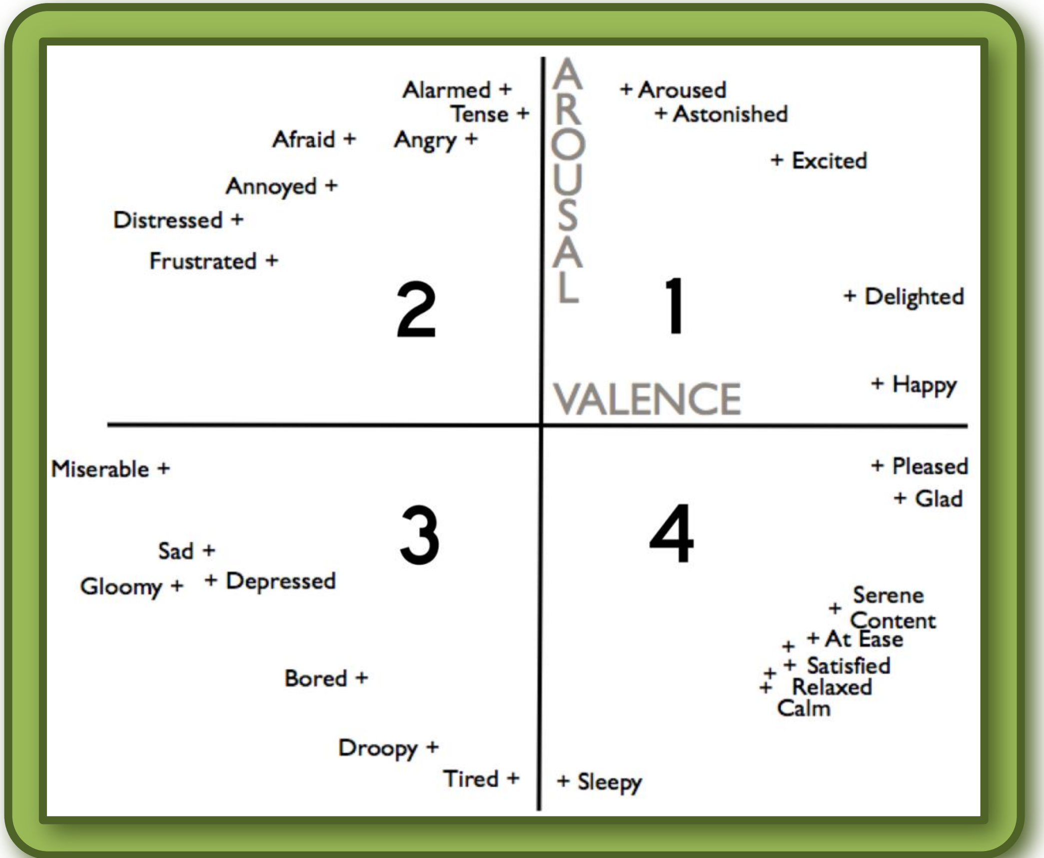
- **Dimensional Models**
  - Main ideas
    - Emotions organized **along axes** (2 or 3)
      - Each emotion is a **located in a multi-dimensional plane**, based on a reduced number of axes (2D or 3D)
      - Argument: correspond to internal human representations of emotions

Emotion Models

- **Dimensional Models**
  - Main ideas
    - **3 main dimensions of emotion** [Yang and chen, 2012]
      - ***valence*** (or pleasantness; positive and negative affective states),
      - ***arousal*** (or activation; energy and stimulation level)
      - ***potency*** (or dominance; a sense of control or freedom to act)
    - **2D used in practice**
      - Valence and arousal regarded as the “core processes” of affect [Yang and Chen, 2002]
      - Simpler to visualize emotions

Emotion Models

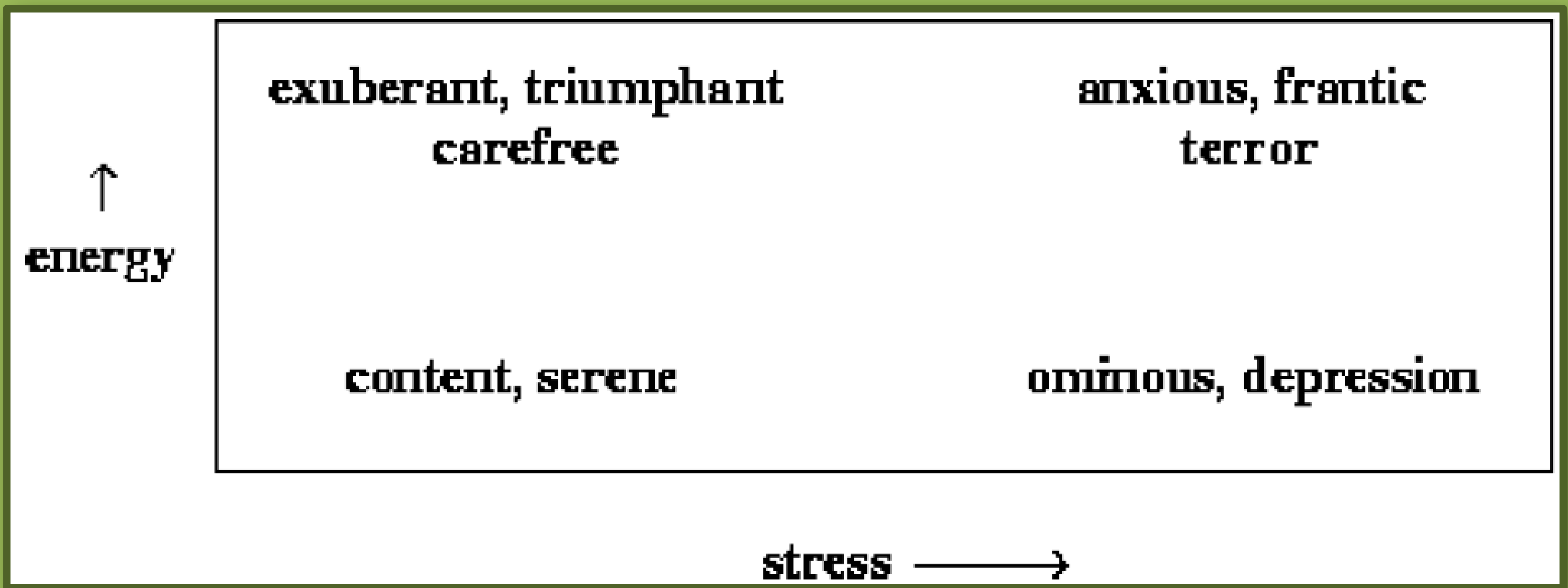
- **Dimensional Models**
  - Examples
    - James Russell's circumplex model [Russell, 1980]



© Laurier, 2011

Emotion Models

- **Dimensional Models**
  - Examples
    - Robert Thayer's model [Thayer, 1989]



© Meyers, 2007

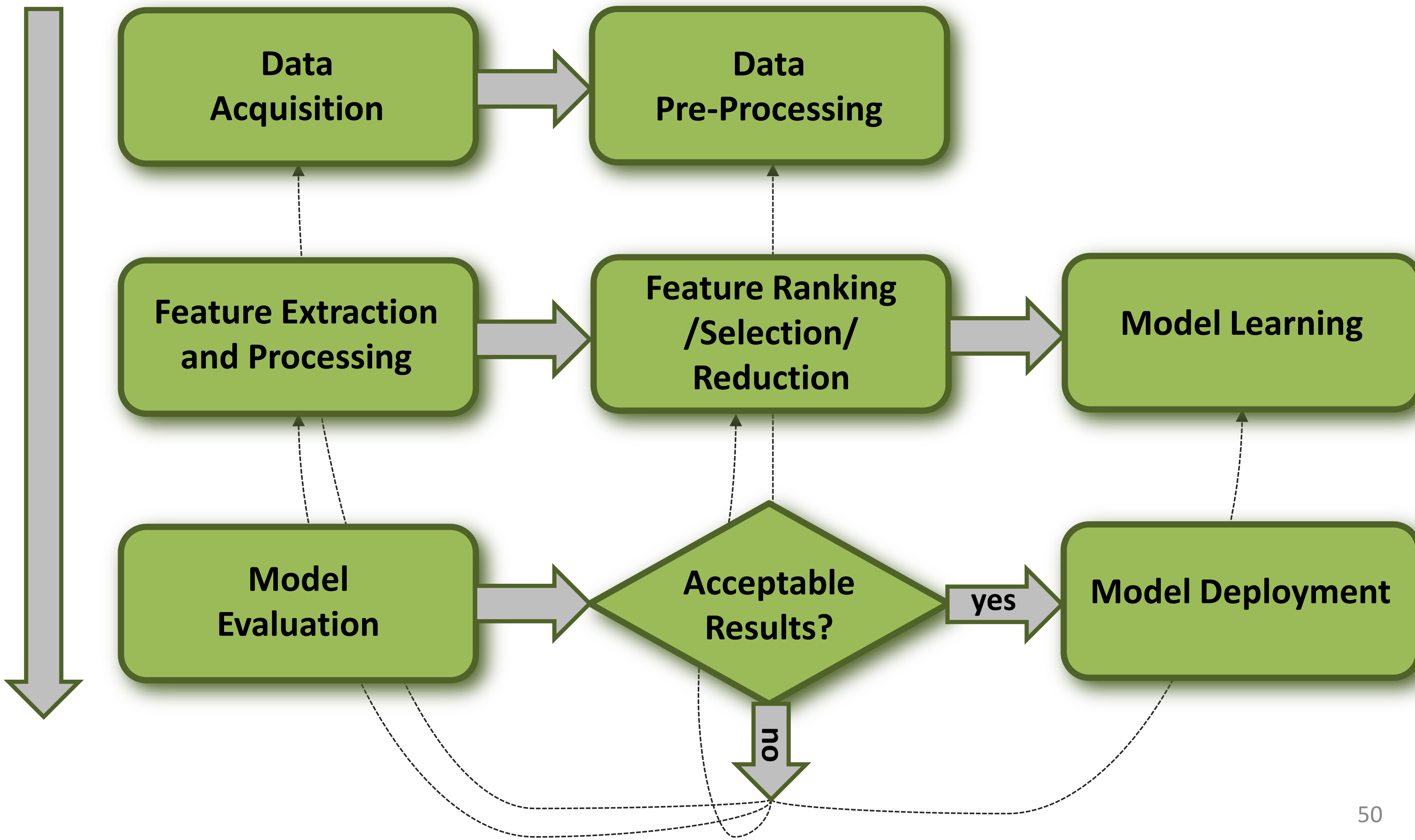
Emotion Models

- **Dimensional Models**
  - Examples
    - **Continuous models**
      - **Emotion plane as a continuous space**
        - » Each point denotes a different emotional state.
        - » Ambiguity related with emotion states is removed

Emotion Models

- **Dimensional Models**
  - Limitations
    - Obscures important aspects of emotion
      - Anger and fear are placed close in the valence-arousal plane
        - » Very different in terms of their implications
      - → **Potency** (dominant–submissive) as the third dimension

MER Data Mining Process



DM Process

Data Acquisition

- **Goals**
  - Get meaningful, **representatives examples** of each concept to capture, **balanced across classes**
    - E.g., songs from different styles, genres, …
  - Get **accurate annotations**
    - Necessary to perform manual data annotation? How many annotators, what profiles, how many songs, song balance, etc.?
      - Can be **tedious**, **subjective** and **error-prone**

There can be no knowledge discovery on bad data!

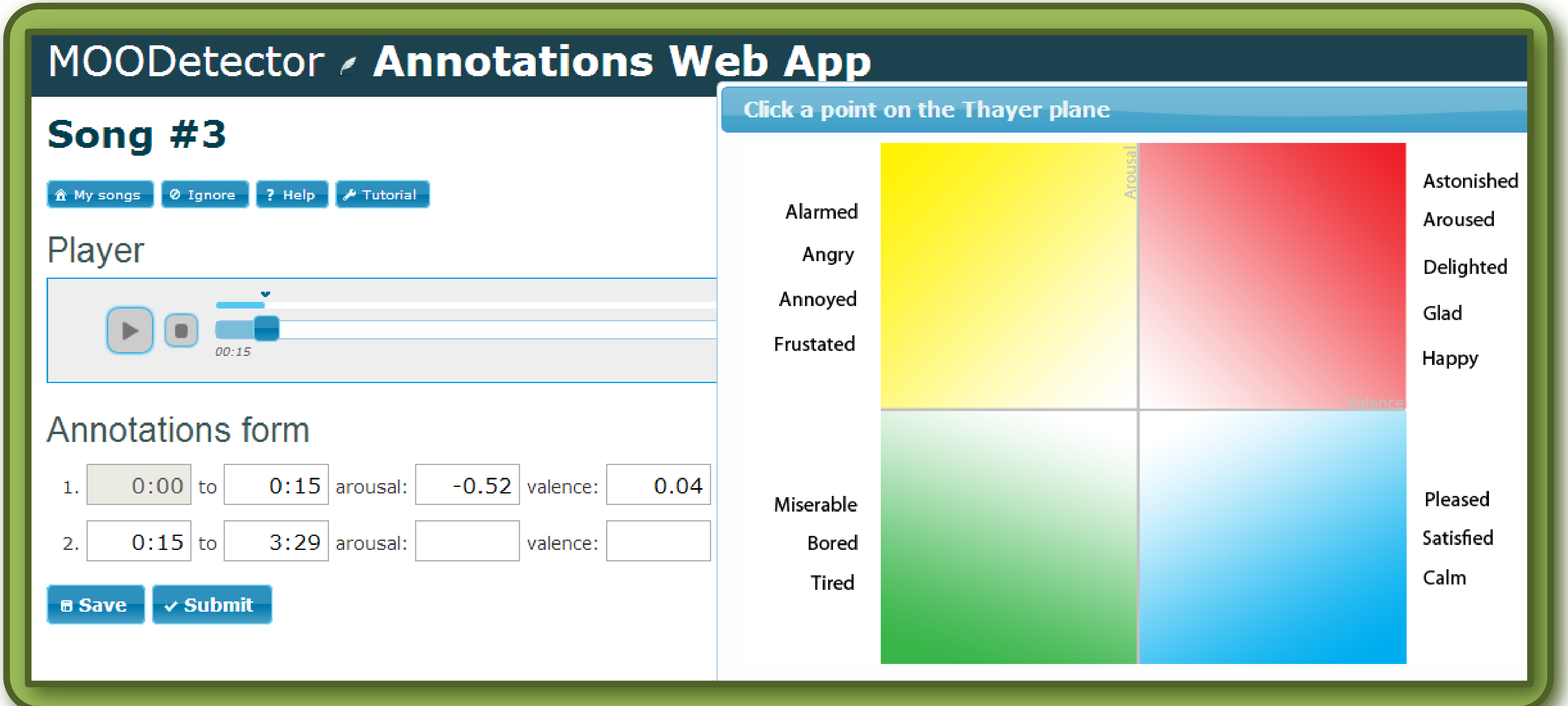
DM Process

- **How?**
  - Selection of an adequate **emotion model**
    - Categorical or dimensional?
    - Which categories?
      - Basic emotions? MIREX?
    - Single label, multi-label or probabilistic classification?

DM Process

Data Acquisition

- **How?**
  - Careful **data annotation protocol**



DM Process

Data Acquisition

- **How?**
  - Careful **data annotation protocol**
    - **Manual annotation process**
      - Use **annotation experts** and/or **music listeners**
        - » Experts: harder to acquire but annotations will likely be more reliable
      - **Distribute the samples across annotators**, guaranteeing that
        - » Each annotator gets a reasonable amount of samples
        - » Each sample is annotated by a sufficient number of people

DM Process

Data Acquisition

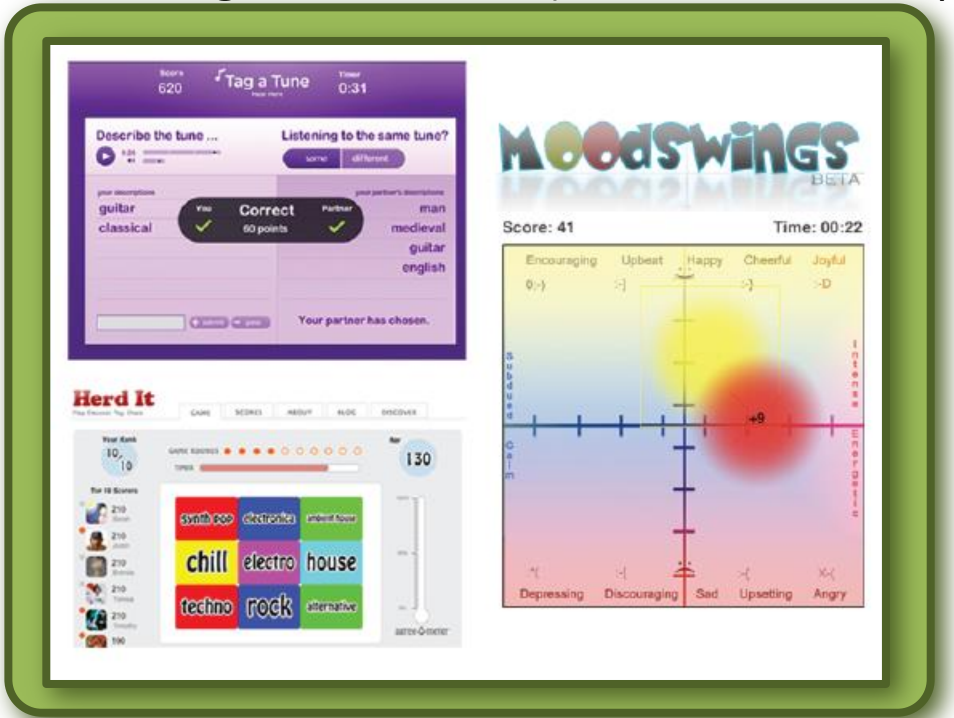
- **How?**
  - Careful **data <u>annotation</u> protocol**
    - **Manual annotation process**
      - **Evaluate sample annotation consistency**
        - » **Remove samples** for which there is **not an acceptable level of agreement**: e.g., too high standard deviation
        - » → Not good representatives of the concept
        - » In the other cases, keep the average, median, etc. of all annotations

DM Process

- **How?**
  - Careful **data annotation protocol**
    - **Manual annotation process**
      - **Evaluate annotator consistency**
        - » Exclude **outlier annotators**
          - Annotators that repeatedly disagree with the majority
        - » Perform a **test-retest reliability study** [Cohen and Swerdlik, 1996]
          - Select a sub-sample of the annotators to repeat the annotations some time later
          - Measure the differences between annotations

DM Process

- **How?**
  - Careful **data <u>annotation</u> protocol**
    - **Manual annotation can be tedious and error-prone**
      - → Annotation games : **GWAP** (Game With A Purpose)



DM Process

- **Goals**
  - **Data preparation** prior to analysis
    - E.g., noise filtering, data cleansing, …
- **How?**
  - **Standardization** of song excerpts
    - Sampling frequency, e.g., 22050 kHz
    - Number of channels, e.g., monoaural
    - Number of quantization bits, e.g., 16

DM Process *Feature Extraction & Processing*

- **Goals**
  - Extract **meaningful**, **discriminative features correlated to emotion** from the information source under analysis (audio, MIDI, lyrics, …)
  - Evaluate the **impact of different features** in each class

DM Process *Feature Extraction & Processing*

- **How?**
  - Based on studies by music psychologists, e.g., [Gabrielsson and Lindström, 2001]
    - **Modes**
      - Major modes related to happiness or solemnity, minor modes associated with sadness or anger [Meyers, 2007].
    - **Harmonies**
      - Simple, consonant, harmonies are usually happy, pleasant or relaxed; complex, dissonant, harmonies relate to emotions such as excitement, tension or sadness, as they create instability in a musical piece [Meyers, 2007].

DM Process *Feature Extraction & Processing*

- **Relevant musical attributes** [Friberg, 2008; Meyers, 2007]
  - Timing
    - Tempo, tempo variation, duration contrast
  - Dynamics
    - Overall level, crescendo/decrescendo, accents
  - Articulation
    - Overall (staccato/legato), variability
  - Timbre
    - Spectral richness, onset velocity, harmonic richness

DM Process *Feature Extraction & Processing*

- **Relevant musical features**
  - Pitch
    - High/low
  - Interval
    - Small/large
  - Melody
    - Range (small/large), direction (up/down)
  - Harmony
    - Consonant/complex-dissonant

DM Process *Feature Extraction & Processing*

- **Relevant musical features**
  - Tonality
    - Chromatic-atonal/key-oriented
  - Rhythm
    - Regular-smooth/firm/flowing-fluent/irregular-rough
  - Mode
    - Major/minor

DM Process *Feature Extraction & Processing*

- **Relevant musical features**
  - Loudness
    - High/low
  - Musical form
    - Complexity, repetition, new ideas, disruption)
  - Vibrato
    - Extent, speed

DM Process *Feature Extraction & Processing*

- **Feature Extraction**
  - Music platforms
    - Audio:
      - Marsyas, MIR Toolbox, PsySound, …
    - MIDI:
      - jSymbolic, MIDI Toolbox, jMusic…
    - Lyrics:
      - jLyrics, SynesSketch, ConceptNet,, …

    - Types of features
      - Key, tonal, timbre, pitch, rhythm

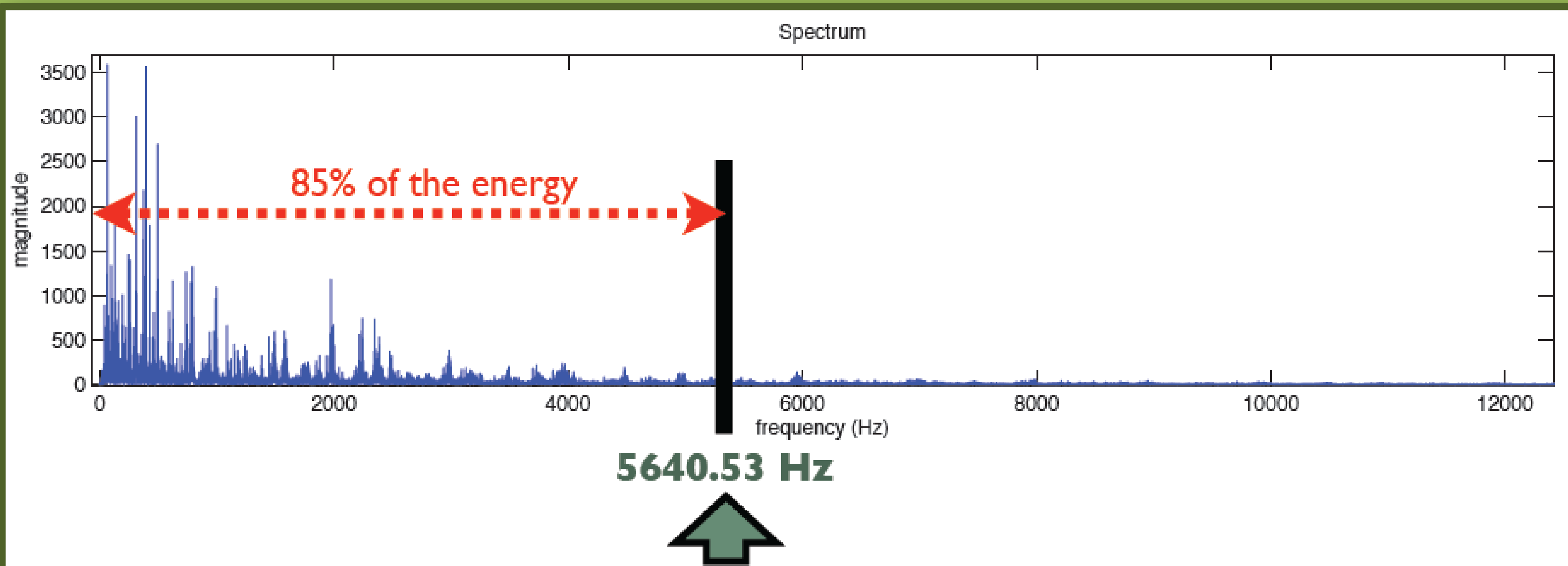
DM Process *Feature Extraction & Processing*

- **Feature Extraction**
  - Music platforms

Framework	# of features	Description
Marsyas	237	Spectral centroid, rolloff, flux, zero cross rate, linear spectral pair, linear prediction cepstral coefficients (LPCCs), spectral flatness measure (SFM), spectral crest factor (SCF), stereo panning spectrum features, MFCCs, chroma, beat histograms and tempo.
MIR Toolbox	177	Among others: root mean square (RMS) energy, rhythmic fluctuation, tempo, attack time and slope, zero crossing rate, rolloff, flux, high frequency energy, Mel frequency cepstral coefficients (MFCCs), roughness, spectral peaks variability (irregularity), inharmonicity, pitch, mode, harmonic change and key.
PsySound	44	Loudness, sharpness, volume, spectral centroid, timbral width, pitch multiplicity, dissonance, tonality and chord, based on psycho acoustic models.

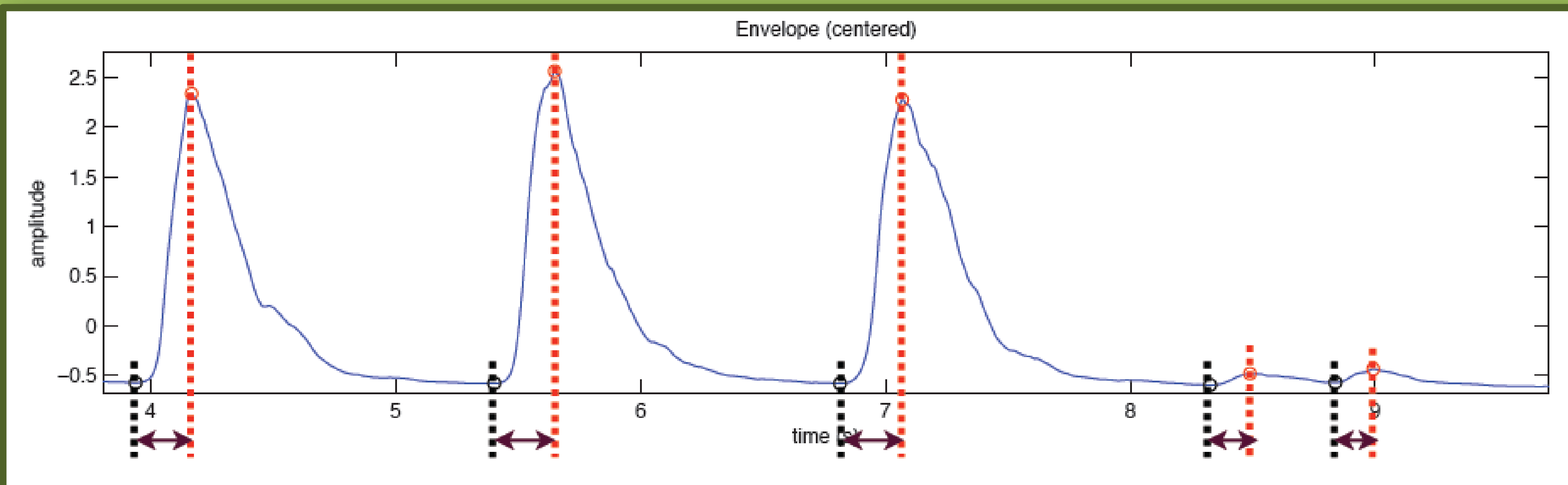
DM Process *Feature Extraction & Processing*

- **Feature Extraction**
  - Music platforms
    - Rolloff: high-frequency energy



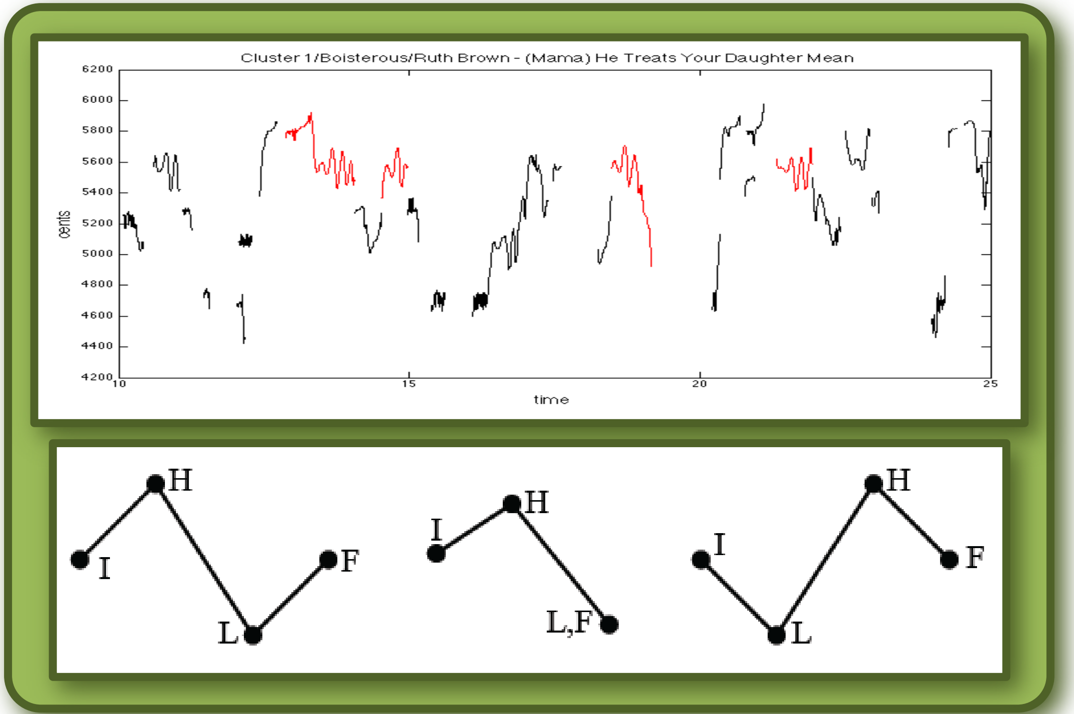
DM Process *Feature Extraction & Processing*

- **Feature Extraction**
  - Music platforms
    - Attack time



DM Process *Feature Extraction & Processing*

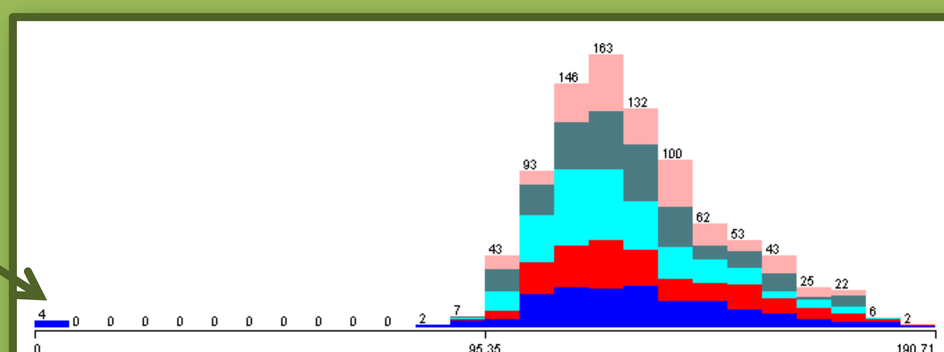
- **Feature Extraction**
  - Programmed by our research team
    - E.g., Melodic features (vibrato, tremolo, dynamics, contour types, ...: 98 features)



DM Process *Feature Extraction & Processing*

- **Feature Processing**
  - **Process features**, if needed
    - **Normalize** feature values
    - **Discretize** feature values
    - Detect and fix/remove **outliers**
      - Errors in feature extraction

Probable outliers:
measurement errors



DM Process *Feature Extraction & Processing*

- **Difficulties**
  - **Few relevant audio features** proposed so far [Friberg, 2008]
    - **Difficult to extract from audio signals** → but easier to extract from symbolic representations (some features are score-based in nature)
  - **Feature inaccuracy**
    - E.g., current algorithms for tempo estimation from audio are not 100% accurate…

DM Process *Feature Ranking/Selection/Reduction*

- **Goals**
  - **Remove redundancies** → eliminate irrelevant or redundant features
    - E.g., Bayesian models assume independence between features → **redundant features decrease accuracy**
  - Perform **dimensionality reduction**
    - Simpler, faster, more accurate and more interpretable models

DM Process *Feature Ranking/Selection/Reduction*

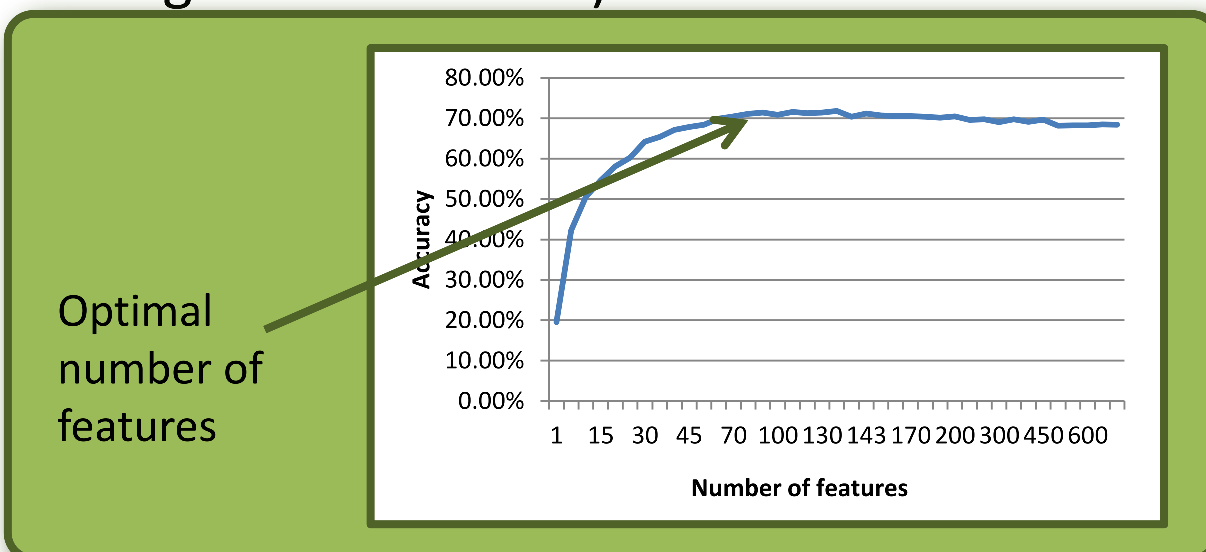
- **Why?**
  - **Improve model performance**
  - **Improve interpretability**
  - **Reduce computational cost**

DM Process *Feature Ranking/Selection/Reduction*

- **How?**
  - Determine the **relative importance** of the extracted features → feature ranking
    - E.g., Relief algorithm, input/output correlation, wrapper schemes, etc.

DM Process *Feature Ranking/Selection/Reduction*

- **How?**
  - Select only the **relevant features**
    - E.g., add one feature at a time according to the ranking, and select the optimum feature set based on the maximum achieved accuracy (see sections on Model Learning and Evaluation)



DM Process *Feature Ranking/Selection/Reduction*

- **How?**
  - **Eliminate redundant** features
    - E.g., find correlations among input features and delete the redundant ones
  - **Map** features to a **less redundant feature space**
    - E.g., using Principal Component Analysis

DM Process *Feature Ranking/Selection/Reduction*

- **Results**
  - Categorical (MIREX)
    - Best audio features: best result s with **11 features** only

Feature Set	Feature Name
SA+MA	1) Vibrato Coverage (VC) (skew) 2) VC (kurt) 3) VC (avg) 4) Vibrato Extent (VE) (avg) 5) VE (kurt) 6) Tonal Centroid 4 (std) 7) Harmonic Change Detection Function (avg), 8) Vibrato Rate (VR) (std), 9) VC (std) 10) VR (avg) 11) VE (std)

DM Process

- **Goals**
  - Tackle the respective learning problem by **creating a good model from data** according to the defined requirements and learning problem
    - **Requirements**
      - **Accuracy**
      - **Interpretability**
      - **…**
    - **Learning problem**
      - **Classification**, **regression**, association, clustering

DM Process

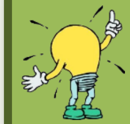
Model Learning

- **How?**
  – Define the **training** and **test sets**
    - Train set: used to learn the model
    - Test set: used to evaluate the model on unseen data

DM Process

Model Learning

- **How?**
  - **Select and compare different models**
    - Performance comparison
      - ***Naïve Bayes*** is often used as baseline algorithm; ***C4.5*** or ***SVMs***, for example, often perform better
      - **Interpretability comparison**
        - » E.g., rules are interpretable, SVMs are black-box



It **has to be shown** empirically from realistic examples that a **particular learning technique is necessarily better than the others**.

When faced with ***N* equivalent techniques**, Occam's razor advises to use the **simplest of them**.

DM Process

- **How?**
  - Perform model **parameter tuning**
    - *Number of neighbors* in ***k*-Nearest Neighbors**
    - *Kernel type*, *complexity*, *epsilon*, *gamma* in **SVMs**
    - *Confidence factor* in **C4.5**
    - …

DM Process *Model Evaluation*

- **Goals**
  - Evaluate **model generalization capability** in a **systematic way**
    - **How the model will perform** on **unseen, realistic, data**,
  - Evaluate **how** one model **compares** to **another**
  - Show that the **learning method** leads to **better performance** than the one achieved **without learning**
    - E.g., chance with 5 classes → 20% accuracy

DM Process

- **How?**
  - Use a **separate test set**
    - Predict the behavior of the model in unseen data
  - Use an adequate **evaluation strategy**
    - E.g., repeated stratified 10-fold cross-validation
  - Use an adequate **evaluation metric**
    - Regression: **RMSE** and **R2** are typically used
    - Classification: **Precision, recall and F-measure** are standard

DM Process

Model Evaluation

- **How?**
  - Check which **kinds of errors** are more prevalent and why
    - What classes/variables show low accuracy? Why?
      - **Valence**: important features might be missing
    - Where is the root of the problem: classifier/regressor, extracted features, method employed for extraction, feature selection approach?
    - Same questions for individual songs

DM Process

Model Evaluation

- **Common Requirements**
  - **Accuracy**
  - **Interpretability**
  - There is often a **trade-off** between accuracy and interpretability
    - E.g., decision tree: trade-off between succinctness (smaller trees) versus classification accuracy
    - E.g., rule induction algorithms might lead to weaker results than an SVM

DM Process

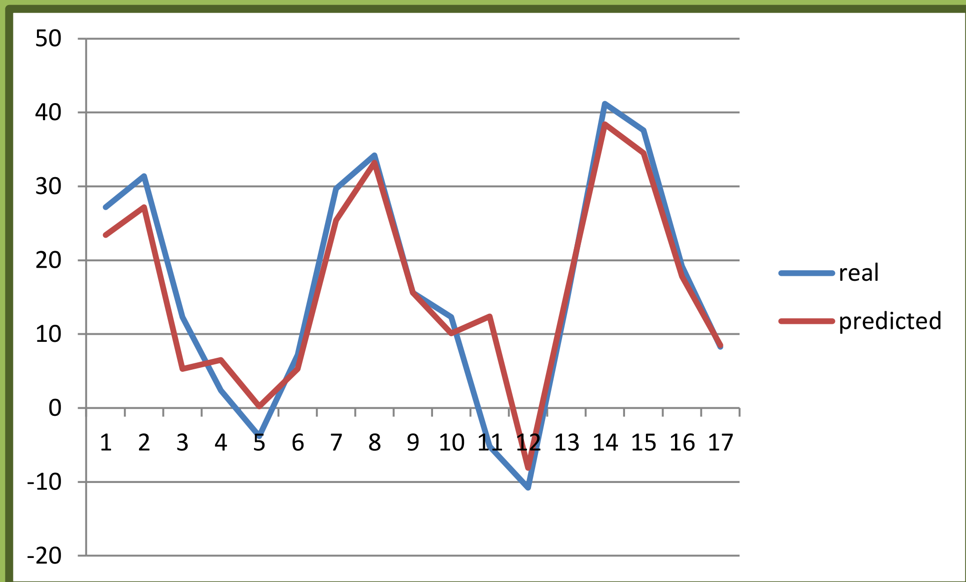
Model Evaluation

- **Performance Metrics**
  - **Regression problems → R2 statistics**

Sample nr.	Real Temp	Predicted Temp
1	27.2	23.4
2	31.4	27.2
3	12.3	15.4
4	2.4	0.1
5	-3.8	0.2
6	7.2	5.3
7	29.7	25.4
8	34.2	33.2
9	15.6	15.6
10	12.3	10.1
11	-5.2	-7.2
12	-10.8	-8.1
13	14.2	15.3
14	41.2	38.4
15	37.6	34.5
16	19.2	17.8
17	8.3	8.5

Example:

Predict temperature for next day at 12:00pm:



DM Process

Model Evaluation

- **Results**
  - Dimensional
    - Yang dataset (**R2** statistics) → **Best world results!**

Classifiers	SA		MA		SA+MA	
	Arousal	Valence	Arousal	Valence	Arousal	Valence
SLR	43.38	-1.22	32.87	-12.73	43.38	-1.22
SLR (fs)	51.87	1.57	32.87	-8.59	51.87	1.57
KNN	55.80	-2.30	28.47	-9.11	54.05	-0.64
KNN (fs)	58.83	7.59	43.84	-3.92	58.26	5.50
SVM	58.03	16.27	39.56	-4.10	58.03	16.27
SVM (fs)	63.17	35.84	42.08	-0.43	**65.69**	**40.56**

DM Process

Model Evaluation

- **Results**
  - Categorical
    - Our MIREX dataset (**F-Measure**)

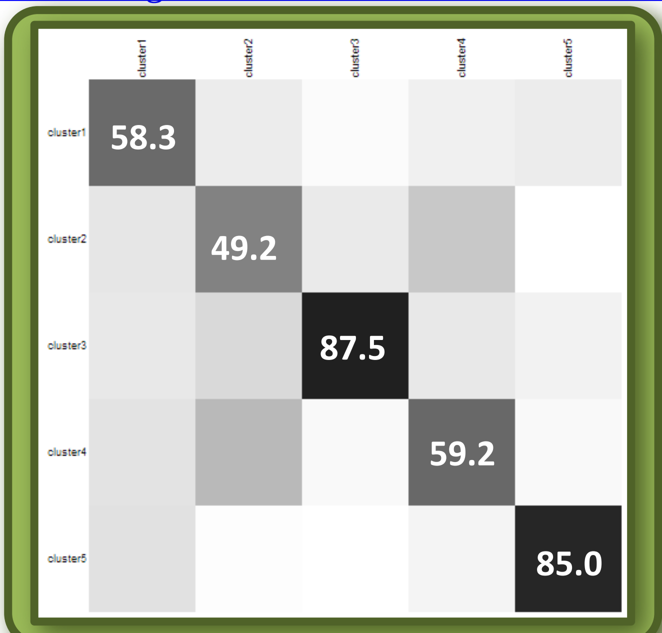
Classifier	SA	MA	SA+MA
NaïveBayes	37.0%	31.4%	38.3%
NaïveBayes*	38.0%	34.4%	44.8%
C4.5	31.4%	53.5%	55.9%
C4.5*	33.4%	56.1%	57.3%
KNN	38.9%	38.6%	41.7%
KNN*	40.8%	56.6%	56.7%
SVM	45.7%	52.8%	52.8%
SVM*	46.3%	59.1%	**64.0%**

DM Process

Model Evaluation

- **Results**
  - Categorical
    - MIREX annual campaign
    - **Best world results!**

http://www.music-ir.org/mirex/results/2012/mirex_2012_poster.pdf



Audio Train/Test: Mood Classification - MIREX08 Dataset

SubID	Participants	Accuracy
PP5	Renato Panda, Rui Pedro Paiva	0.68
WJ2	Ming-Ju Wu, Jyh-Shing Roger Jang	0.68
RW2	Jia-Min Ren, Ming-Ju Wu, Jyh-Shing Roger Jang	0.67
JR1	Jia-Min Ren, Jyh-Shing Roger Jang	0.67
WJ1	Ming-Ru Wu, Jyh-Shing Roger Jang	0.67
RW1	Jia-Min Ren, Ming-Ju Wu, Jyh-Shing Roger Jang	0.67
PP2	Renato Panda, Rui Pedro Paiva	0.66
SSKS1	Klaus Seyerlehner, Markus Schedl, Peter Knees, Reinhard Sonnleitner	0.66
LBLJK1	Shin-Cheol Lim, Karam Byun, Jong-Seol Lee, SEI-JIN JANG, Moo Young Kim	0.65
PP1	Renato Panda, Rui Pedro Paiva	0.65
LBLJK2	Shin-Cheol Lim, Karam Byun, Jong-Seol Lee, SEI-JIN JANG, Moo Young Kim	0.65
PP3	Renato Panda, Rui Pedro Paiva	0.64
YCP1	Deshun Yang, Xiaoou Chen, Liqun Peng	0.64
RW3	Jia-Min Ren, Ming-Ju Wu, Jyh-Shing Roger Jang	0.63
PP4	Renato Panda, Rui Pedro Paiva	0.62
PP6	Renato Panda, Rui Pedro Paiva	0.61
GT1	George Tzanetakis	0.60
PH1	Philippe Hamel	0.57
DM3	Franz de Leon, Kirk Martinez	0.47
DM4	Franz de Leon, Kirk Martinez	0.46

DM Process

Model Evaluation

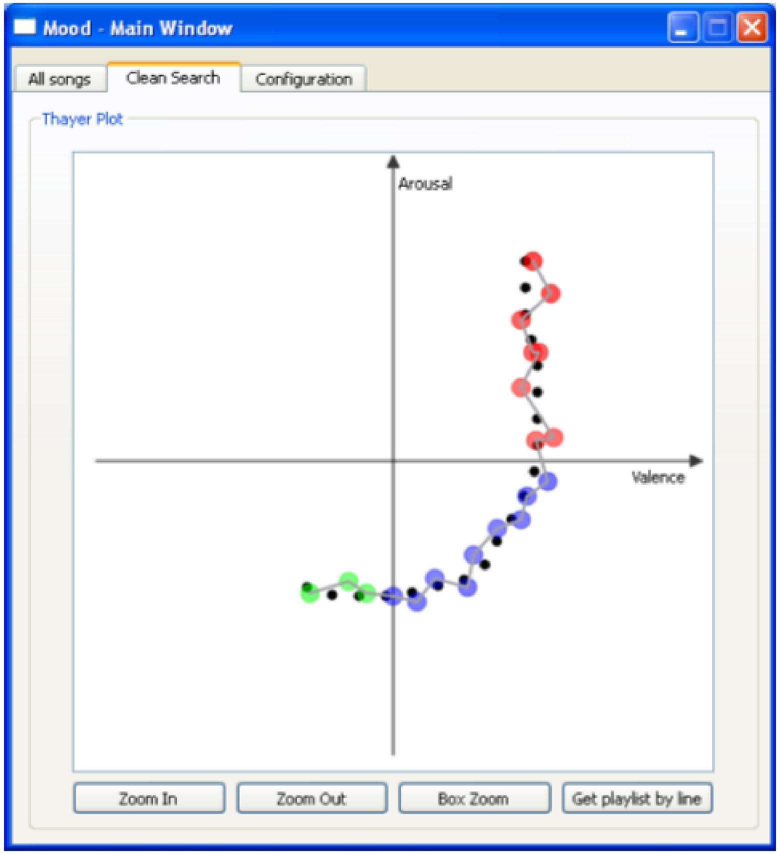
- **Results**
  - Playlist generation

		SA	MA	SA+MA
Top1	All	4.2 %	4.0 %	4.9 %
	FS	6.2 %	3.7 %	5.8 %
Top5	All	21.0 %	17.8 %	21.7 %
	FS	24.8 %	18.4 %	25.4 %
Top20	All	62.7 %	56.6 %	61.0 %
	FS	62.3 %	57.6 %	**64.8** %

DM Process

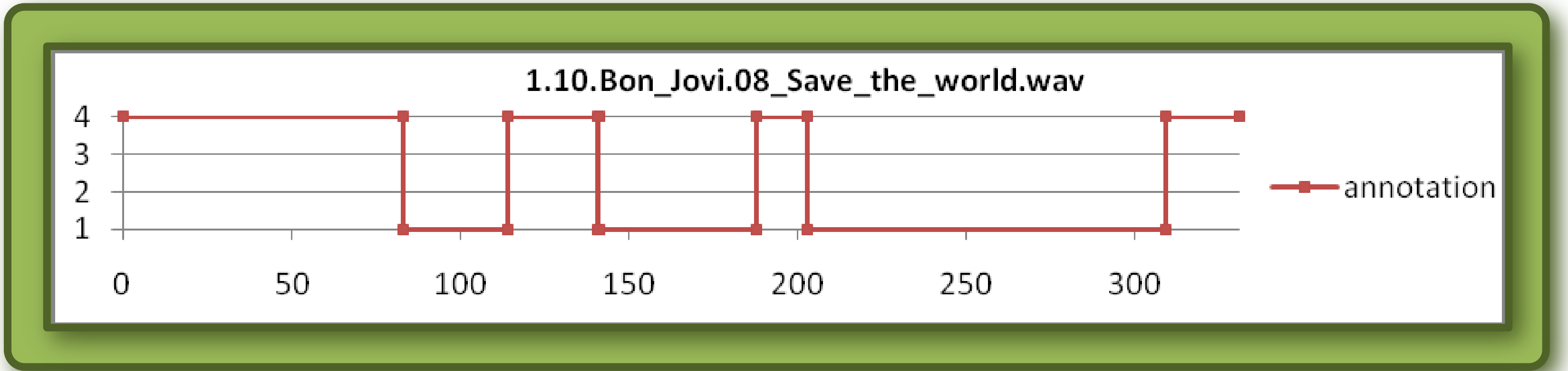
Model Deployment

- **Application development**



Other Problems Addressed

- **Music Emotion Variation Detection (MEVD)**
  - Emotion as a **temporal variable**
    - Emotions may change throughout a song → **MEVD**



Other Problems Addressed

- **Music Emotion Variation Detection (MEVD)**
  - Segment-based classification [Panda and Paiva, 2011]
    - Divide audio signal into small segments
    - Classify each of them as before
  - Ad-hoc strategies (e.g., [Lu *et al.*, 2006])
    - Segmentation based on feature variations
      - Thresholds used and difficult to tune

Other Problems Addressed

- **MEVD**
  - Ground Truth
    - Annotations by 2 subjects: quadrants only
  - Preliminary Results
    - Classification: SVM
    - 4 classes (quadrants)
    - Accuracy: **56%**

Limitations and Open Problems

- **Current Limitations**
  - **Lack of agreement on a usable mood taxonomy**
    - MIREX mood: only 5 moods, with semantic and acoustic overlap [Yang and Chen, 2012]
  - **Lack of sizeable real-world datasets**
    - Dimensional approaches
      - Yang only uses 194 songs
    - Categorical approaches
      - MIREX mood validates using 600 songs [MIREX, 2012]
  - **Accuracy** of current systems is **too low** for most real-world applications
    - MIREX best algorithm ~ 68% accuracy in a 5-class mood problem [MIRExresults, 2012]

Limitations and Open Problems

- **Open problems**
  - **Semantic gap**
    - Novel, semantically-relevant features necessary, able to capture the relevant musical attributes
      - Most important limitation, according to [Friberg, 2008]
    - Multi-modal approaches: combination of different information sources (audio, midi, lyrics)
      - Use MIDI resulting from automatic music transcription
  - **MEVD**
    - Other techniques, e.g., self-similarity techniques [Foote, 1999]
    - Quality datasets

Limitations and Open Problems

- **Open problems**
  - **Multi-label classification** (e.g., [Sanden and Zhang, 2011])
    - Same song belonging to more than one mood category
  - **Knowledge discovery from computational models**
    - E.g., rule induction algorithms, neural-fuzzy approaches [Paiva and Dourado, 2004]

Future/Ongoing Work

- **Multi-modal approaches to MER**
  - Combine audio, MIDI and lyrics
  - Preliminary results: results improve using a multi-modal approach
- **Ground Truth**
  - Basic emotions, larger dataset
  - MEVD A/V dataset
- **Feature Extraction**
  - New features
- **Knowledge Extraction**
  - Rule induction algorithms

Acknowledgements

Acknowledgements

PROMETEO
Investigación
Formación
Desarrollo

Proyecto Prometeo, Ecuador

Campus Party™

Campus Party Quito 3

Universidade de Coimbra, Portugal

About the Author

About the Author

- **More info at http://rppaiva.dei.uc.pt/**

References

- Suggested Initial Literature
- Cited References

Suggested Initial Literature

- **Yang Y.-H. and Chen H.-H. (2011). "Music Emotion Recognition," CRC Press, Boca Raton, Florida, USA, February 2011, 261 pages.**

- Friberg A. (2008). “Digital Audio Emotions - an Overview of Computer Analysis and Synthesis of Emotional Expression in Music”. In Proc. International Conference on Digital Audio Effects - DAFx-08.
- Kim Y. E., Schmidt E. M., Migneco R., Morton B. G., Richardson P., Scott J., Speck J. A. and Turnbull D. (2010). “Music emotion recognition: a state of the art review”. International Society for Music Information Retrieval Conference – ISMIR’2010.
- Lu L., Liu D. and Zhang H.-J. (2006). “Automatic Mood Detection and Tracking of Music Audio Signals”. IEEE Transactions on Audio, Speech and Language Processing, Vol. 14, No. 1, pp. 5-18.
- Yang Y.-H., Lin Y.-C., Su Y.-F. and Chen H. H. (2008). “A Regression Approach to Music Emotion Recognition”. IEEE Transactions on Audio, Speech and Language Processing, Vol. 16, No. 2, pp. 448-457.
- Yang Y.-H. and Chen H.-H. (2012). "Machine recognition of music emotion: a review“, To Appear in ACM Trans. Intelligent Systems and Technology.

Cited References

- Alghoniemy M. and Tewfik A. (2000). “Personalized Music Distribution”, In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP’2000.
- Barlow H. and Morganstern S. (1948). A Dictionary of Musical Themes, New York, Crown.
- Burad A. (2006). “Multimedia Databases”, Seminar Report, Computer Science and Engineering, Indian Institute of Technology.
- Byrd D. (2002). “The History of ISMIR - A Short Happy Tale”, D-Lib Magazine, Vol. 8, No. 1, URL: http://www.dlib.org/dlib/november02/11inbrief.html#BYRD.
- Byrd D. and Crawford T. (2002). “Problems of Music Information Retrieval in the Real World”, Information Systems and Management, Vol. 38, pp. 249-272
- Cardoso L., Panda R. and Paiva R. P. (2011). “MOODetector: A Prototype Software Tool for Mood-based Playlist Generation”. Simpósio de Informática – INForum 2011.
- Celma O. and Lamere P. (2007). “Music Recommendation Tutorial”, International Society for Music Information Retrieval Conference – ISMIR’2007, tutorial.
- Dunn J. W. (2000). “Beyond VARIATIONS: Creating a Digital Music Library”, In International Society for Music Information Retrieval Conference – ISMIR’2000.

Cited References

- Ekman P. (1992). “An argument for basic emotions”, Cognition & Emotion, Vol. 6 (3), 169–200.
- Farnsworth P. R. (1954). “A study of the Hevner adjective list”, J. Aesthetics and Art Criticism 13, 97–103.
- Feng Y., Zhuang Y. and Pan Y. (2003). “Music information retrieval by detecting mood via computational media aesthetics”. In Proc. International Conference on Web Intelligence - IEEE/WIC.
- Friberg A. (2008). “Digital Audio Emotions - an Overview of Computer Analysis and Synthesis of Emotional Expression in Music”. In Proc. International Conference on Digital Audio Effects - DAFx-08.
- Foote J. (1999). “Visualizing music and audio using self-similarity”. In Proc. ACM Multimedia, pp. 77–84, Orlando, Florida, USA.
- Futrelle J. and Downie J. S. (2003). “Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000-2002”, Journal of New Music Research, Vol. 32, No. 2, pp. 121-131.
- Gabrielsson A. (2002). Emotion perceived and emotion felt: Same or different? Musicae Scientiae,123–147. special issue.

Cited References

- Gabrielsson A. and Lindström E. (2001). “The influence of musical structure on emotional expression”. In P. N. Juslin, & J. A. Sloboda (Eds.), Music and emotion: Theory and Research New York: Oxford University Press, 2001, pp. 223-248.
- Ghias A., Logan J., Chamberlin D. and Smith B. C. (1995). “Query by Humming: Musical Information Retrieval in an Audio Database”, In Proc. ACM Multimedia Conference – Multimedia’95.
- Gomes A. R. (2005). “Festivais de Verão: Para Além da Música, Uma Experiência”, Revista XIS – Jornal Público, August 6, 2005, pp. 18-19 (in Portuguese).
- Haitsma J. and Kalker T. (2002). “A highly robust audio-fingerprinting system”, International Society for Music Information Retrieval Conference – ISMIR’2002.
- Hevner K. (1935). “Expression in music: A discussion of experimental studies and theories”, Psychological Review 48, 2, 186–204.
- Huron D. (2000). “Perceptual and Cognitive Applications in Music Information Retrieval”, International Society for Music Information Retrieval Conference – ISMIR’2000.
- Juslin P. N. (1997). “Perceived emotional expression in synthesized performances of a short melody: capturing the listener’s judgment policy,” Music. Sci., Vol. 1, No. 2, pp. 225–256.

Cited References

- Juslin P. N. and Laukka J. (2004). “Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening”. Journal of New Music Research, Vol. 33, No. 3, pp. 217-38, 2004.
- Kashino K., Nakadai K., Kinoshita T. and Tanaka H. (1995). “Organization of Hierarchical Perceptual Sounds: Music Scene Analysis with Autonomous Processing Modules and a Quantitative Information Integration Mechanism”, In Proc. International Joint Conference on Artificial Intelligence – IJCAI’95, pp. 158-164.
- Kassler M. (1966). “Toward Musical Information Retrieval”, Perspectives of New Music, Vol. 4, No. 2, pp. 59-67.
- Kim Y. E., Schmidt E. M., Migneco R., Morton B. G., Richardson P., Scott J., Speck J. A. and Turnbull D. (2010). “Music emotion recognition: a state of the art review”. International Society for Music Information Retrieval Conference – ISMIR’2010.
- Lu L., Liu D. and Zhang H.-J. (2006). “Automatic Mood Detection and Tracking of Music Audio Signals”. IEEE Transactions on Audio, Speech and Language Processing, Vol. 14, No. 1, pp. 5-18.
- Matesic B. C. and Cromartie F. (2002). “Effects music has on lap pace, heart rate, and perceived exertion rate during a 20-minute self-paced run”, The Sports Journal, Vol. 5. No. 1.

Cited References

- Meyers O. C. (2007). “A Mood-Based Classification and Exploration System”. MSc Thesis, School of Architecture and Planning, Massachusetts Institute of Technology.
- MIREX (2012). “MIREX Home”, URL: http://www.music-ir.org/mirex/wiki/MIREX_HOME, available by March 1, 2012.
- MIREXresults (2010). “MIREX 2010 Results”. URL: http://www.music-ir.org/mirex/wiki/2010:MIREX2010_Results, available by March 1, 2012.
- Orio N. (2006). “Music Retrieval: A Tutorial and Review”, Foundations and Trends in Information Retrieval, Vol. 1, No. 1.
- Paiva R. P. (2002). “Content-based Classification and Retrieval of Music: Overview and Research Trends”. Technical Report, Centre of Informatics and Systems of the University of Coimbra, Portugal.
- Paiva R. P. and Dourado A. (2004). “Interpretability and Learning in Neuro-Fuzzy Systems”. Fuzzy Sets and Systems, Vol. 147 (1), pp. 17-38, Elsevier.
- Paiva R. P. (2006). “Melody Detection in Polyphonic Audio”, PhD Thesis, University of Coimbra, Department of Informatics Engineering.
- Pampalk E. (2001). Islands of Music: Analysis, Organization and Visualization of Music Archives, MSc Thesis, Department of Software Technology and Interactive Systems, Vienna University of Technology, Austria.

Cited References

- Panda R. and Paiva R. P. (2011). “Using Support Vector Machines for Automatic Mood Tracking in Audio Music”. Proceedings of the 130th Audio Engineering Society Convention – AES 130, London, UK.
- Parker C. (2005). “Applications of Binary Classification and Adaptive Boosting to the Query-by-Humming Problem”, International Society for Music Information Retrieval Conference – ISMIR’2005.
- Pauws S. and Wijdeven S. (2005). “User Evaluation of a New Interactive Playlist Generation Concept”, International Society for Music Information Retrieval Conference – ISMIR’2005.
- Pfeiffer S., Fischer S. and Effelsberg W. (1996). “Automatic Audio Content Analysis”, In Proc. ACM Multimedia Conference – Multimedia’96, pp. 21-30.
- Ryynänen M. (2008). “Automatic Transcription of Pitch Content in Music and Selected Applications”. PhD Thesis, Tampere University of Technology, Finland.
- RTP (2009). “Cultura: Concertos de música ligeira geraram 30 milhões em 2007”. URL: http://ww1.rtp.pt/noticias/?article=385096&visual=26&tema=5/. Published in January 29, 2009, available by February 2, 2009.
- Russell J. A. (1980). “A circumspect model of affect”, Journal of Psychology and Social Psychology, Vol. 39, No. 6, p. 1161.

Cited References

- Sanden C and Zhang J. Z. (2011). “An Empirical Study of Multi-Label Classifiers for Music Tag Annotation”, International Society for Music Information Retrieval Conference – ISMIR’2003.
- Schubert E. (2003). “Update of the Hevner adjective checklist”, Perceptual and Motor Skills 96, 1117–1122.
- Serra X. (2005). “Bridging the Music Semantic Gap”, talk given at the Information Day on “Audio-Visual Search Technologies” organized by the Commission services of DG Information Society of the EC in Brussels.
- Sloboda J. A. and Juslin P. N. (2001). “Psychological perspectives on music and emotion”. In Music and Emotion: Theory and Research, P. N. Juslin and J. A. Sloboda, Eds. Oxford, University Press, New York.
- Thayer R. E. (1989). “The Biopsychology of Mood and Arousal”, New York, Oxford University Press.
- TechCrunch (2009). “iTunes Sells 6 Billion Songs, And Other Fun Stats From The Philnote”. URL: http://www.techcrunch.com/2009/01/06/itunes-sells-6-billion-songs-and-other-fun-stats-from-the-philnote/, published in January 6, 2009, available by February 2, 2009.

Cited References

- TechWhack (2005). “Apple iTunes Touches an Impressive 250 Million Sales Figure”. URL: http://news.techwhack.com/690/apple-itunes-250-million/, published in January 25, 2005, available by February 2, 2009.
- TMN (2009). “TMN lança inovador serviço de identificação, pesquisa e aquisição de música através do telemóvel”. URL: http://www.tmn.pt/portal/site/tmn/menuitem.db67f528a6dbaa5ac8a71c10a51056a0/?vgnextoid=d072fbb324238110VgnVCM1000005401650aRCRD&vgnextchannel=15b5fa2eb8b37110VgnVCM1000005401650aRCRD&vgnextfmt=default2, published in February 18, 2008, available by February 2, 2009.
- Tzanetakis G. (2002). Manipulation, Analysis and Retrieval Systems for Audio Signals, PhD Thesis, Department of Computer Science, Princeton University, USA.
- Wang A. (2003). “An Industrial-Strength Audio Search Algorithm”, International Society for Music Information Retrieval Conference – ISMIR’2003, presentation.
- Welsh M., Borisov N., Hill J., von Behren R. and Woo A. (1999). Querying Large Collections of Music for Similarity, Technical Report, Computer Science Division, University of California, Berkeley, USA.
- Yang D. and Lee W. (2004). “Disambiguating music emotion using software agents”, International Society for Music Information Retrieval Conference – ISMIR’2004

Cited References

- Yang Y.-H. and Chen H.-H. (2012). "Machine recognition of music emotion: a review“, To Appear in ACM Trans. Intelligent Systems and Technology.
- Yang Y.-H., Lin Y.-C., Su Y.-F and Chen H.-H. (2007), "Music emotion classification: A regression approach“, IEEE Int. Conf. Multimedia and Expo – ICME’2007.