From *Music Information Retrieval* to *Music Emotion Recognition*

Center for Informatics and Systems of the University of Coimbra (CISUC)
Portugal

Rui Pedro Paiva

March 2012

Contents

- MIR: What and Why?
- Applications
- Short MIR Tale
- Representations
- Techniques

MUSIC INFORMATION RETRIEVAL

MIR: What and Why?

- **MIR**
  - **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 MIR?**
  - 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

- **Music industry runs, only in the USA an amount of money in the order of several billion US dollars per year.**
- **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, wma)
  - 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.)**
  - **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]**
  - **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?

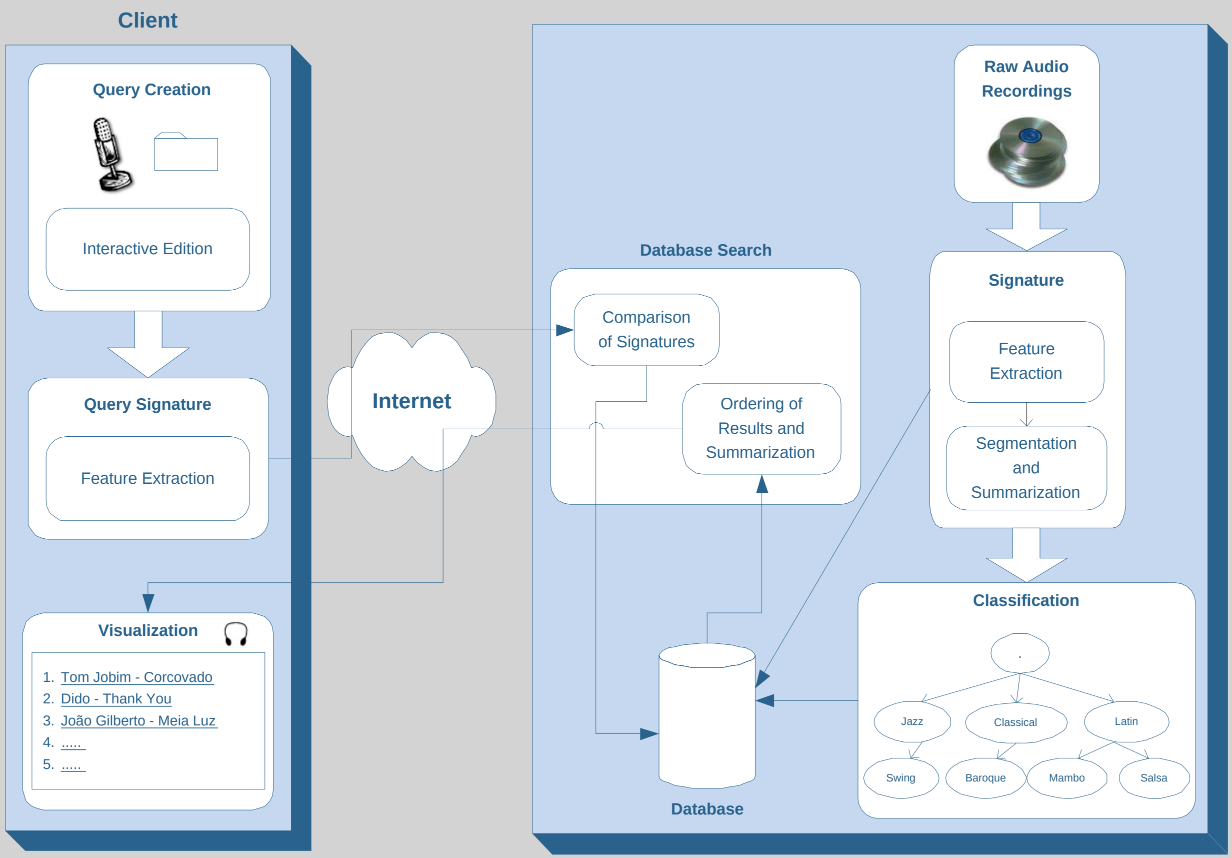
❑ Database Organization and Music Retrieval

- **Presently, 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].

- **→ Music Information Retrieval (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?

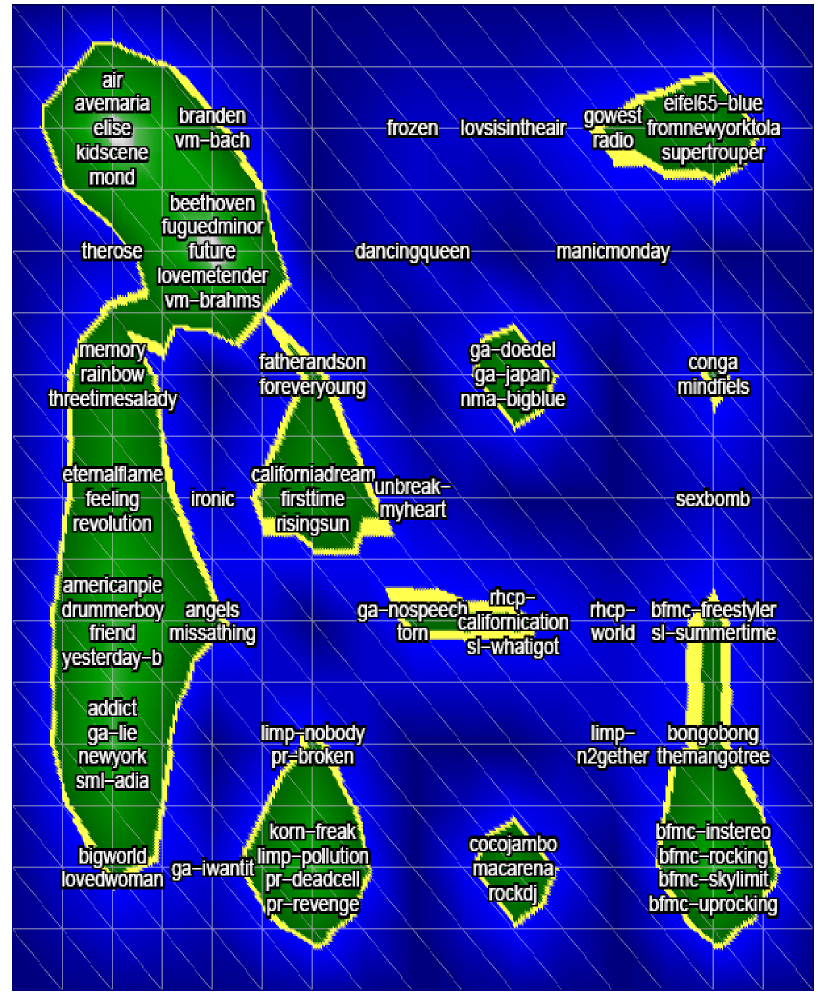
❑ Database Organization and Music Retrieval



© Rui Pedro Paiva, 2002

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 [Cardoso *et al*., 2011]
      - Music recommendation [Celma & Lamere, 2007]
      - Islands of music [Pampalk, 2001]
        - Metaphor of geographic maps: similar genres close together
    - Automatic playlist generation [Pauws and Wijdeven, 2005; Alghoniemy and Tewfik, 2000]



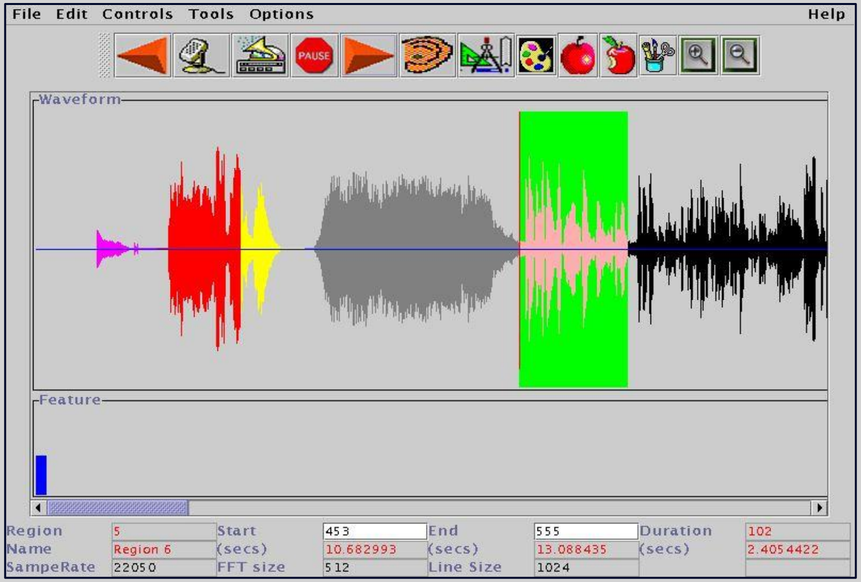
© Elias Pampalk, 2001

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]
- **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]**



© George 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

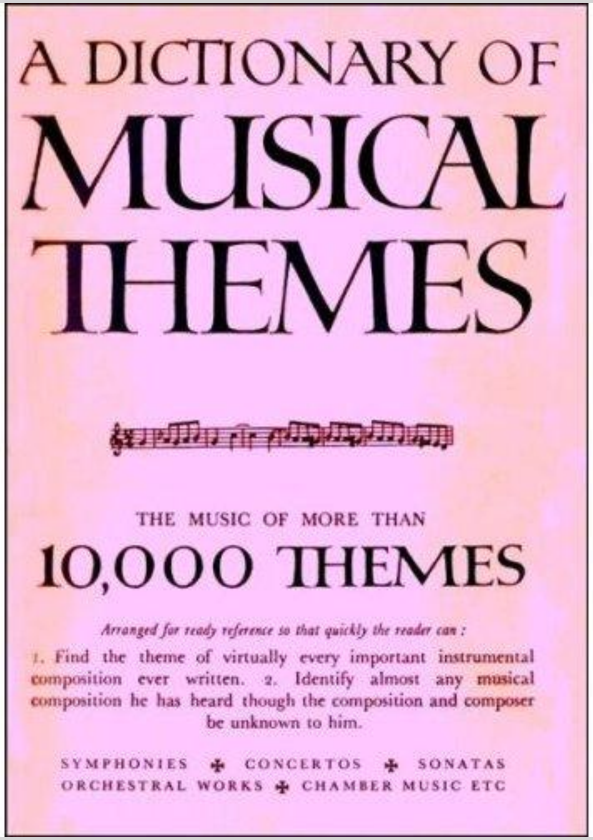
- ❑ **Advertisement and cinema**
  - **Tools for mood-based retrieval [Cardoso *et al.* 2011]**

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

- ❑ **… and so forth…**

Short MIR Tale

- Precursors of computer-based MIR: incipit and theme indexes, e.g., Harold Barlow and Sam Morganstern's dictionary of musical themes [Barlow and Morganstern, 1948]



Short MIR Tale

- 1966: potential of applying automatic information retrieval techniques to music was recognized [Kassler, 1966]
- 1970s and 1980s: automatic music transcription systems
- 1990s: surge of interest, mostly in topics such as query-by-humming (impulse from research on digital libraries)
- 2000: 1st International Symposium on Music Information Retrieval (ISMIR)
  - Now “International Society for Music Information Retrieval Conference”

Short MIR Tale

- 2000-present: active multidisciplinary research field

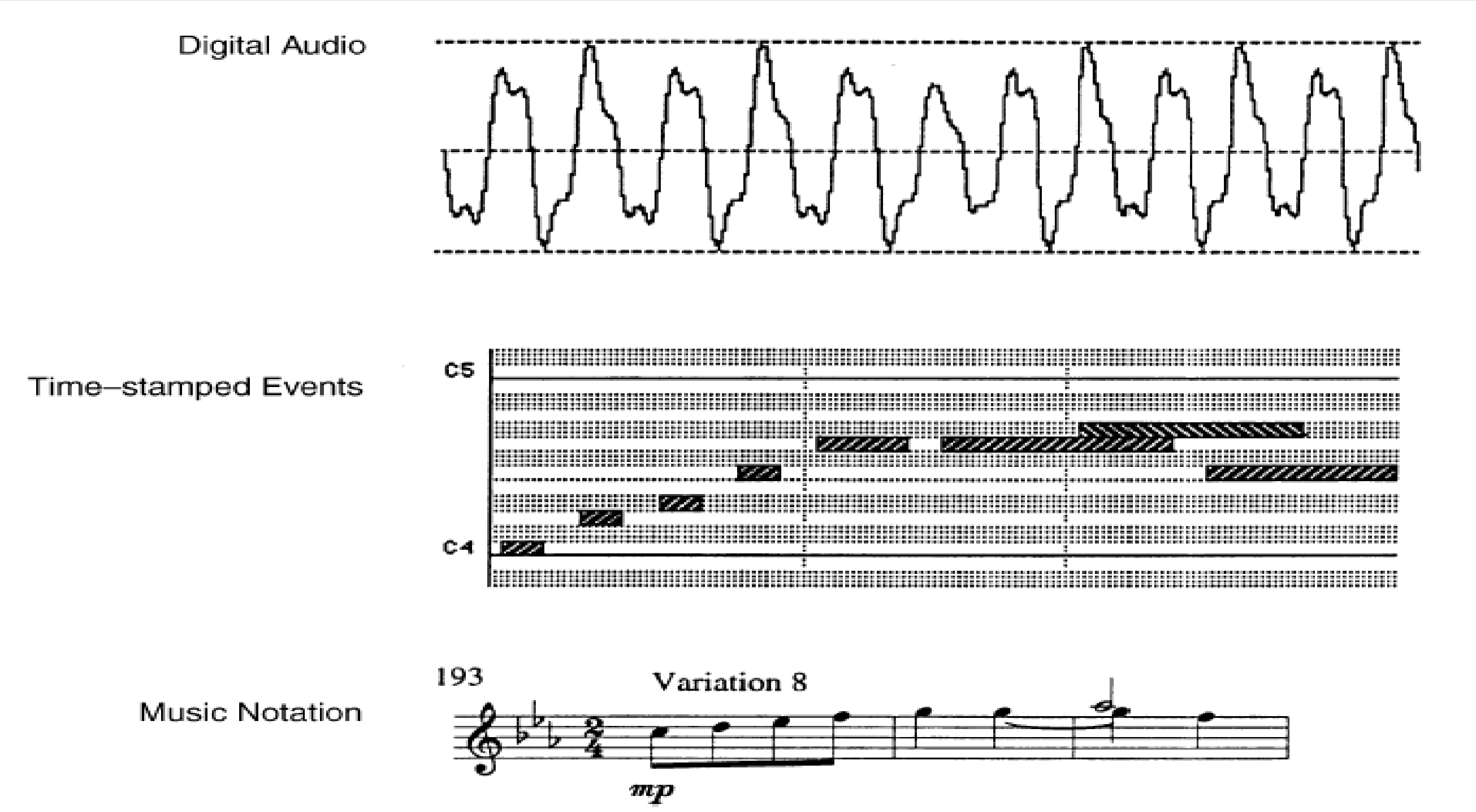
Community	Type of Institution(s)	Typical Research Areas
Computer Science, Information Retrieval	Academic, Commercial	Representation, Indexing, Retrieval, Machine Learning, User Interface Design
Audio Engineering, Digital Signal Processing	Academic, Commercial	Compression, Feature Detection, Pitch Tracking, Machine Learning, Classification, Playlist Generation, Musical Analysis
Musicology, Music Theory	Academic	Representation, Musical Analysis
Library Science	Libraries, Academic	Representation, Metadata, User Studies, Classification, Intellectual Property Rights, User Interface Design
Cognitive Science, Psychology, Philosophy	Academic	Representation, Perception, User Studies, Ontology
Law	Government, Legal Profession, Academic	Intellectual Property Rights

© Joe Futrelle and J. Stephen Downie, 2003

Short MIR Tale

- **2000-present: active multidisciplinary research field**
  - **Content analysis and similarity assessment and retrieval in audio song databases**
    - Metrics of similarity, music identification, music recommendation, audio fingerprinting, music classification and feature extraction, tempo and melody detection, music summarization
  - **Databases systems, indexing, query languages**
  - **Knowledge representation, metadata**
  - **Theme extraction, harmonic and motive analysis, tonality**
  - **Emotion**
  - **Perception and cognition**
  - **Social, legal, ethical and business issues**
  - **User interface design**

Representations



© Donald Byrd and Tim Crawford, 2002

Representations

- **Audio (music content analysis)**
  - **Object**
    - Audio recordings, streaming audio
  - **Topics**
    - Automatic transcription, QBE, classification, recommendation, identification, …
- **Symbolic**
  - **Object**
    - Scores or event-based representations (e.g., MIDI)
  - **Topics**
    - Melodic matching, theme extraction, harmonic analysis, …

Representations

- **Visual**
  - **Object**
    - Printed music
  - **Topics**
    - Optical music recognition
- **Metadata**
  - **Object**
    - Any kind
  - **Topics**
    - Digital libraries, music ontologies (semantic web)

Techniques

- **Idea**
  - **Extract semantic information from low-level data**

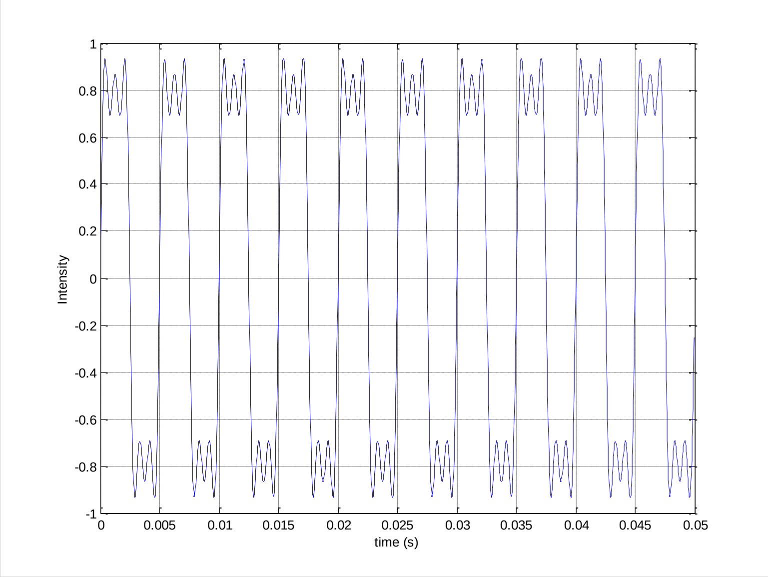
  - **Feature extraction**
    - Physical: F0, intensity, centroid, uniformity, rolloff, flux
    - Perceptual: pitch, loudness, timbre, beat
    - Musicological: notes, melodies, measures, motives, themes
    - Higher-level (semantic) features: emotion, genre, instruments, artist

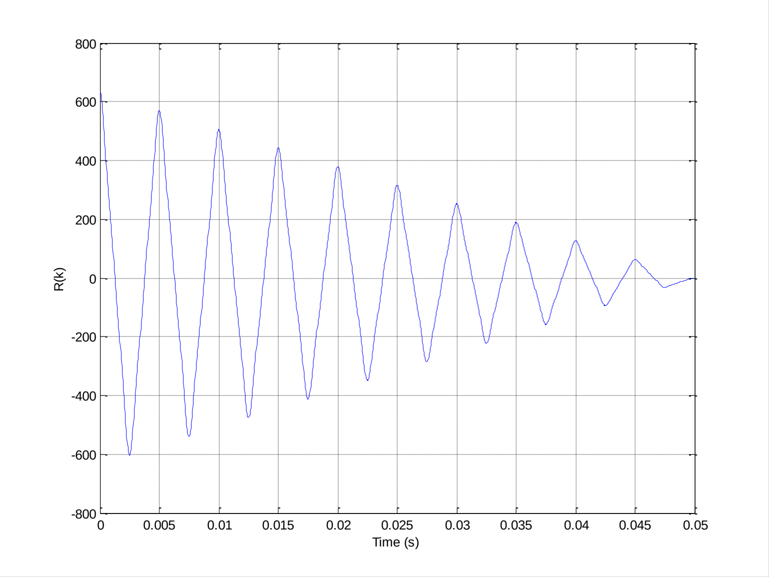
  - **Content and Representation**
    - Audio: which instruments, notes, artist
    - Symbolic: themes, motives
    - Visual: notation, basic units
    - Metadata: genre, artist, MPEG-7 descriptors

Techniques

- **Physical features**
  - **Time domain**
    - Waveform analysis: energy contour, amplitude-based segmentation, auto-correlation, peak detection

$$x(t) = A\sin(\omega_0 t) + \frac{A}{3}\sin(3\omega_0 t) + \frac{A}{5}\sin(5\omega_0 t), \quad A = 1;\ f_0 = 200Hz;\ \omega_0 = 2\pi f_0$$

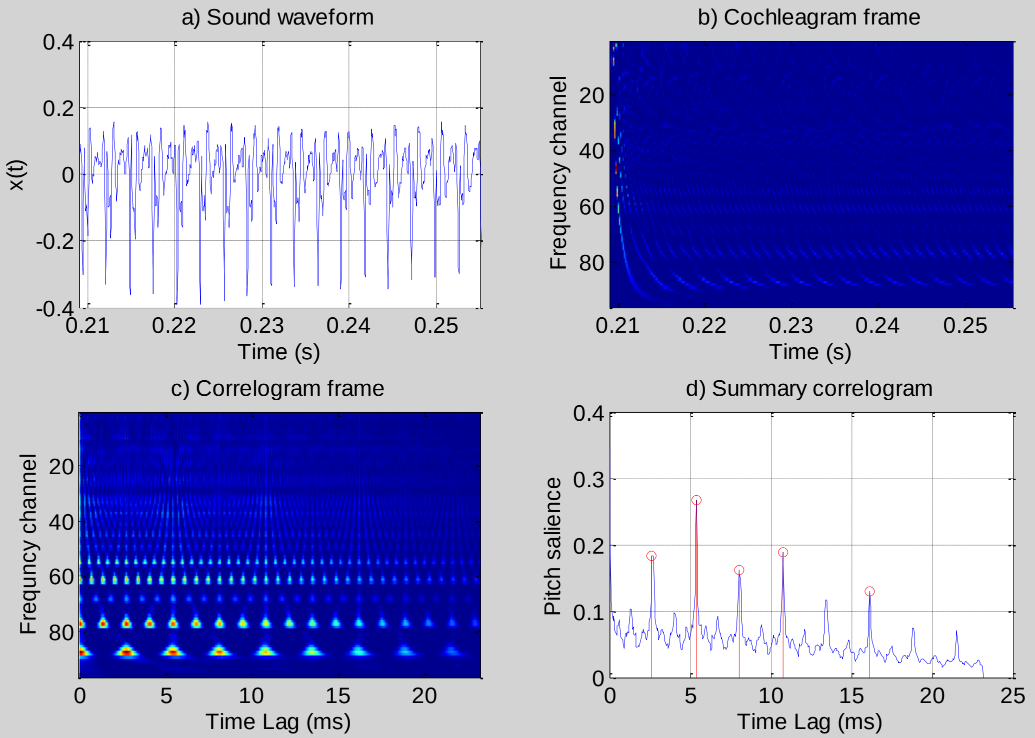


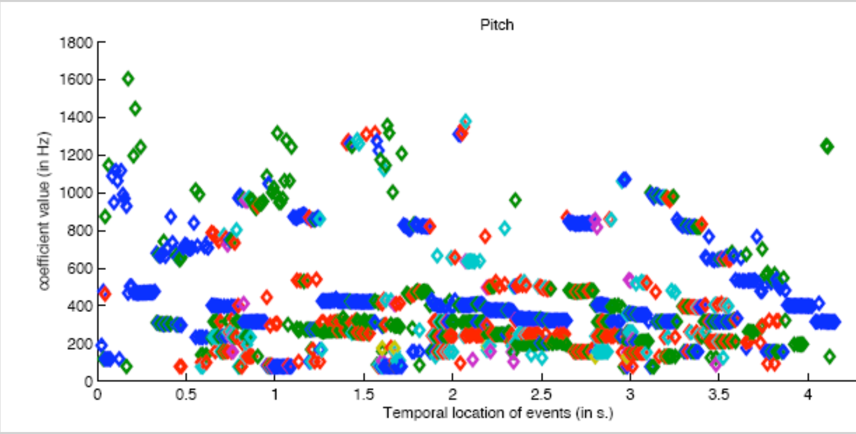


Techniques

❑ Physical features (cont.)

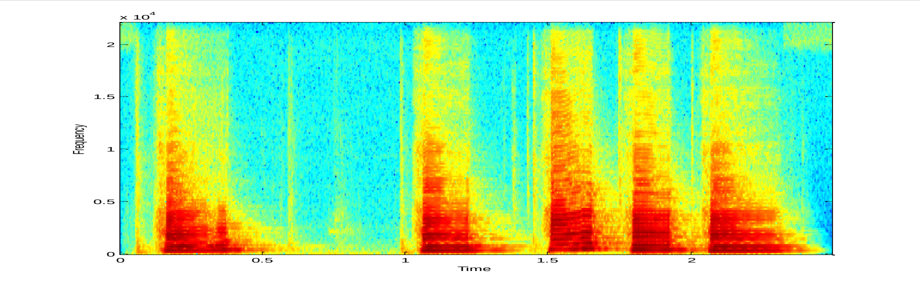
- **Time domain**
  - Auditory model-based F0 detectors, beat detectors (energy-based)

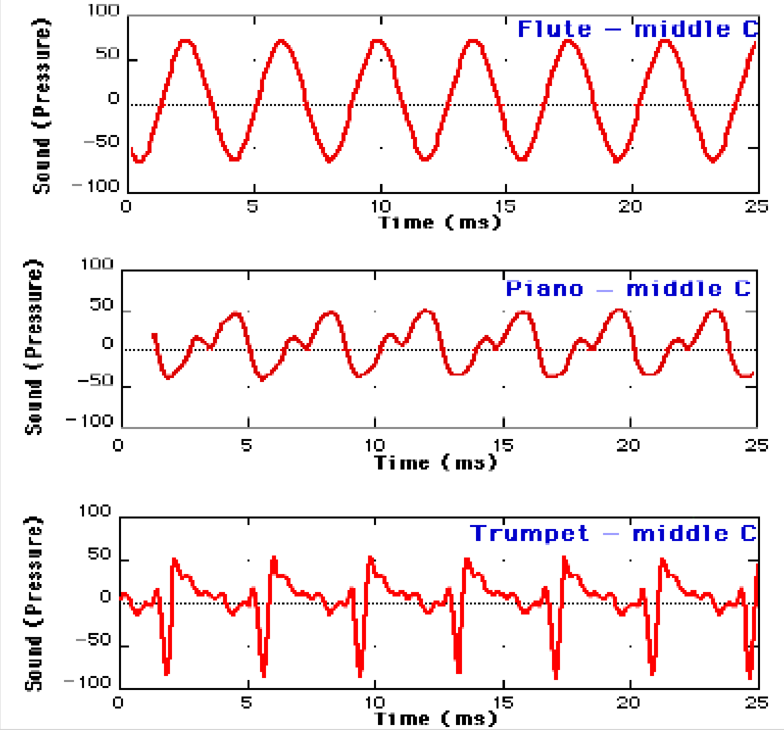


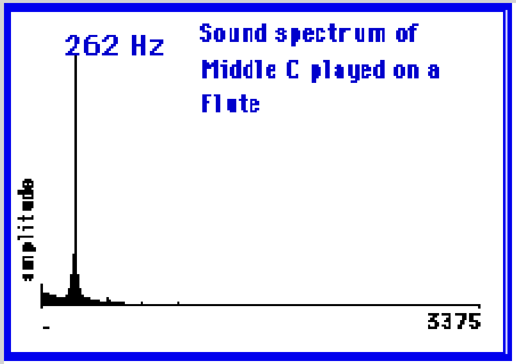


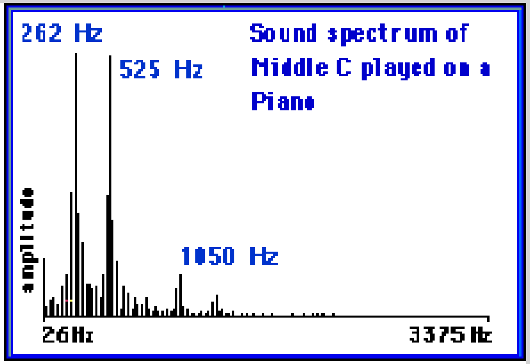
Techniques

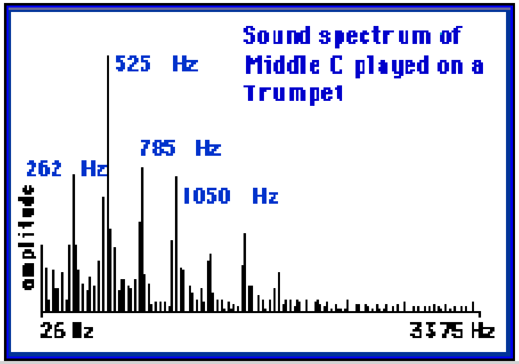
- **Physical features**
  - **Frequency domain**
    - DFT, STFT, spectrogram
    - F0 detectors









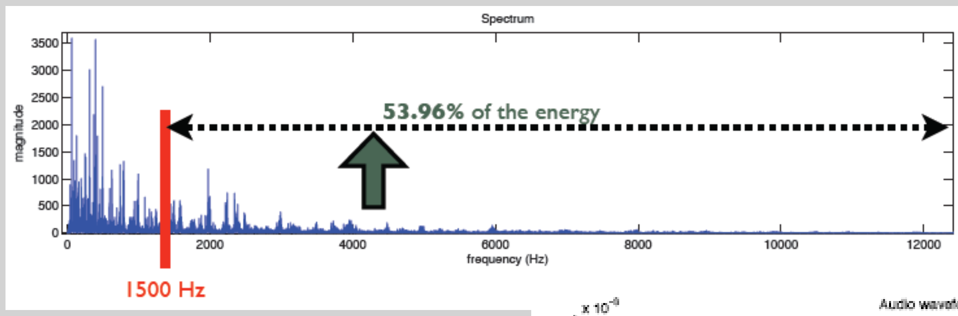


Techniques

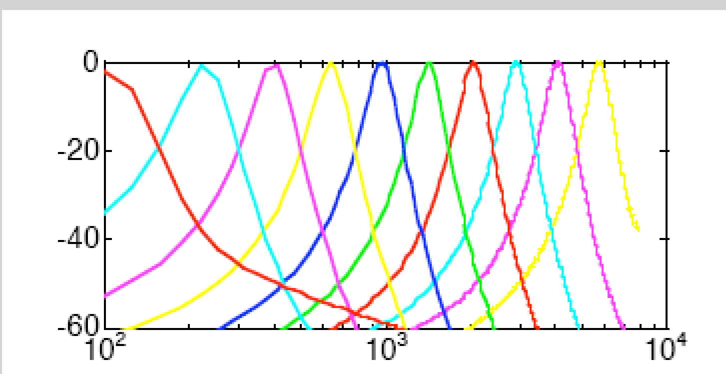
❑ Physical features (cont.)

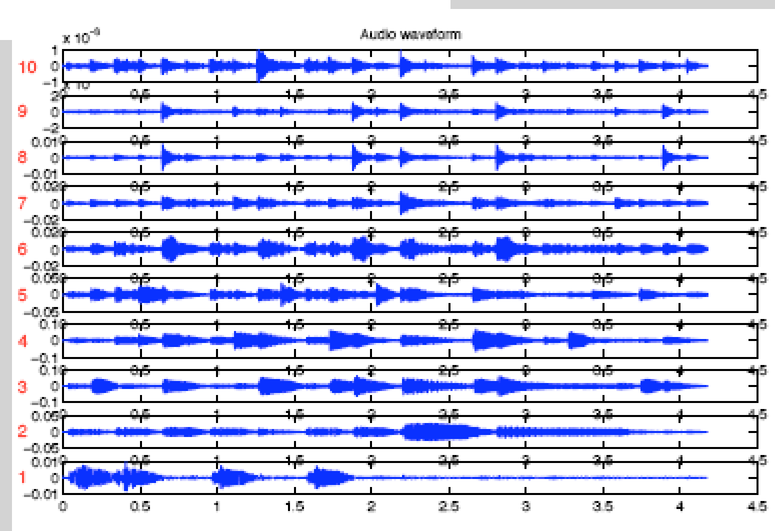
- **Spectral features**
  - MFCCs, centroid, rolloff, flux, harmonicity, high-frequency content, …

© MIR Toolbox, 2008



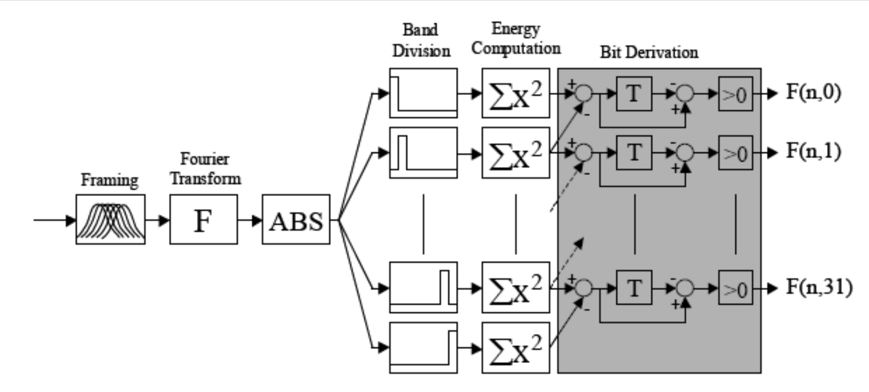
- **Sub-band features**



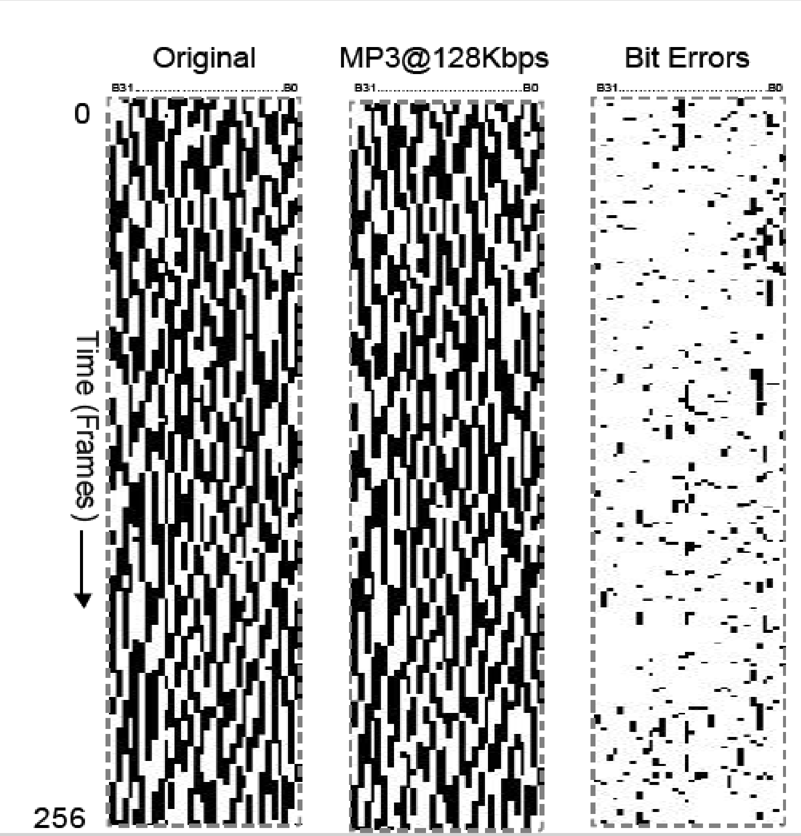


Techniques

- **Physical features (cont.)**
  - **Spectral features + sub-band features (e.g., audio fingerprinting)**



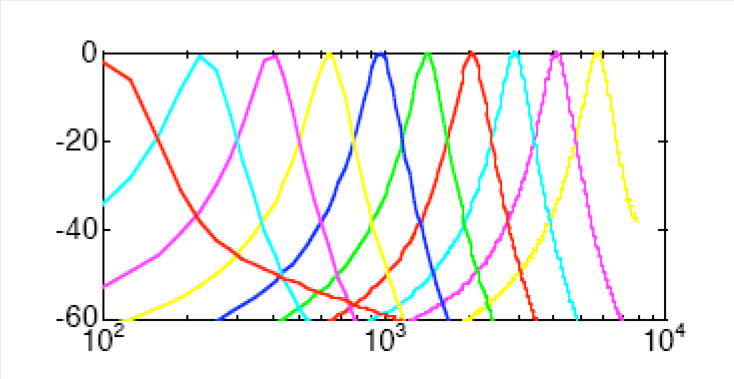
$$F(n,m)=\begin{cases}1 & \text{if } E(n,m)-E(n,m+1)-\big(E(n-1,m)-E(n-1,m+1)\big)>0\\ 0 & \text{if } E(n,m)-E(n,m+1)-\big(E(n-1,m)-E(n-1,m+1)\big)\le 0\end{cases}$$



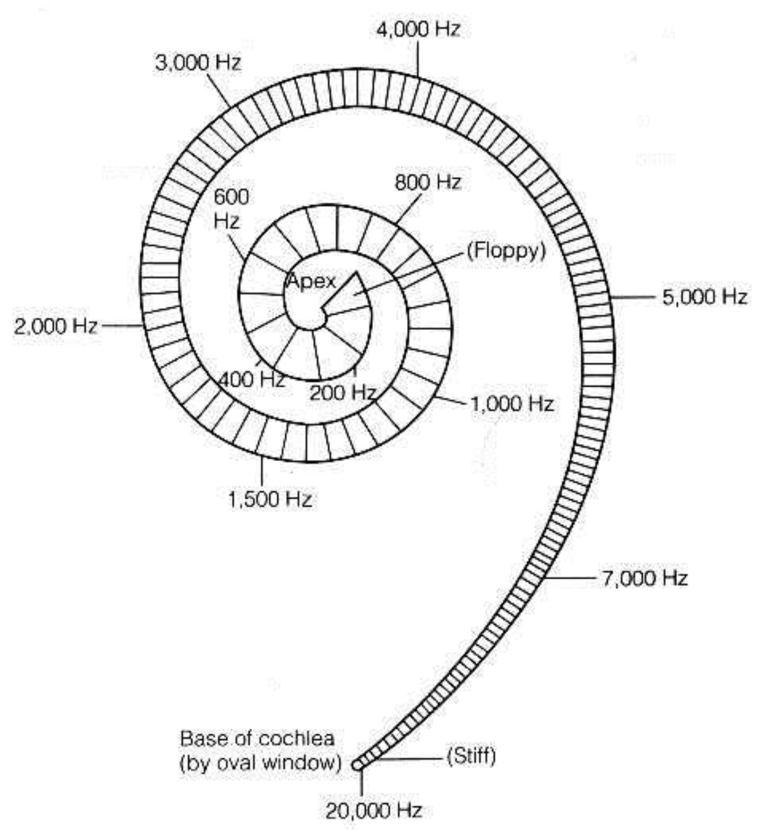
© Haitsma & Kalker, 2002

Techniques

- **Perceptual features**
  - **Pitch**
    - **Frequency**
    - Intensity
    - Context
    - Ear physiology (age)



© MIR Toolbox, 2008

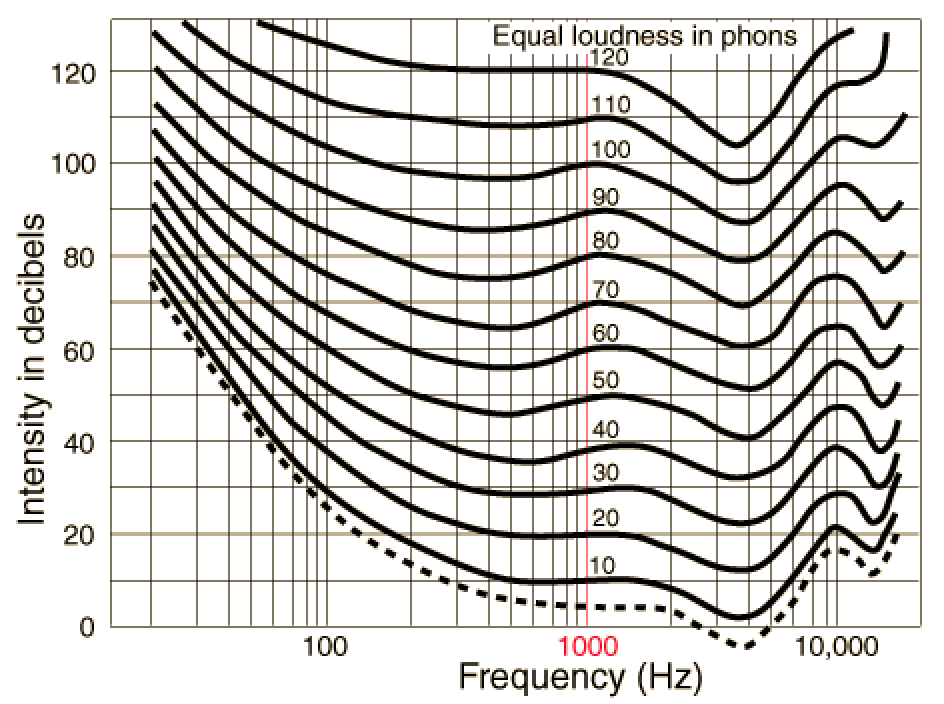


© http://www.ai.rug.nl/~tjeerd/CPSP/docs/cochleaModel.html

Techniques

- **Perceptual features (cont.)**
  - **Loudness**
    - **Intensity**
    - Frequency
    - Context
    - Ear physiology (age)

Fletcher-Munson equal loudness contours

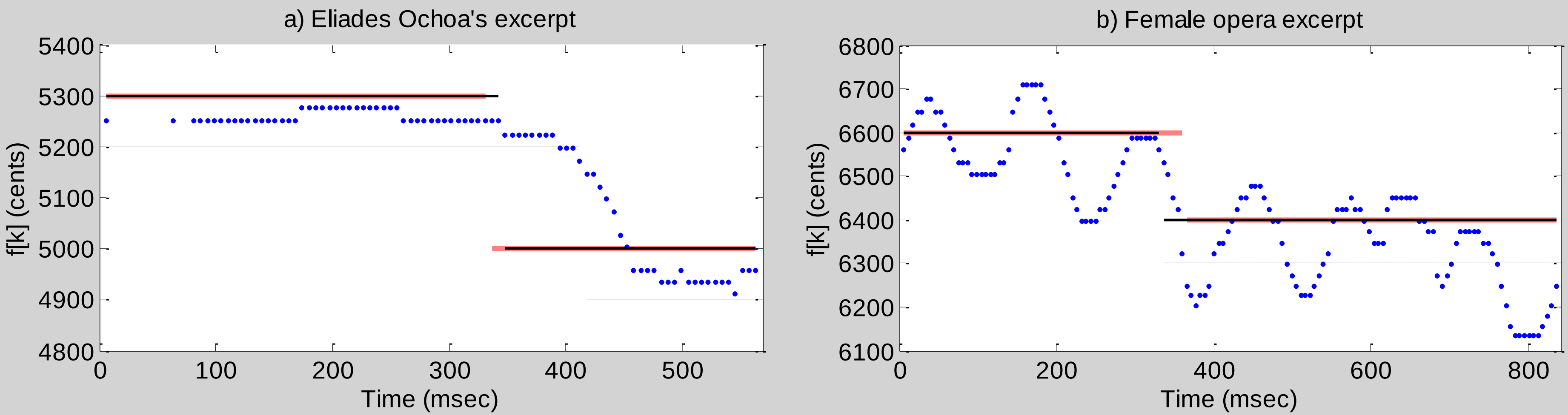


Techniques

- **Perceptual features (cont.)**
  - **Timbre**
    - No physical correlate
    - “what something sounds like”:
      - Spectral content at steady-sate
        - Centroid, rolloff, relative amplitudes of harmonic components, inharmonicty…
      - Signal’s temporal envelope
        - Attack transient
      - Temporal behavior of the harmonics
  - **Melodic contour**
    - UDUEEUUD
  - **Rhythm contour**
    - FSSFEEFS
  - **Beat**

Techniques

- **Musicological features**
  - **Notes from audio**

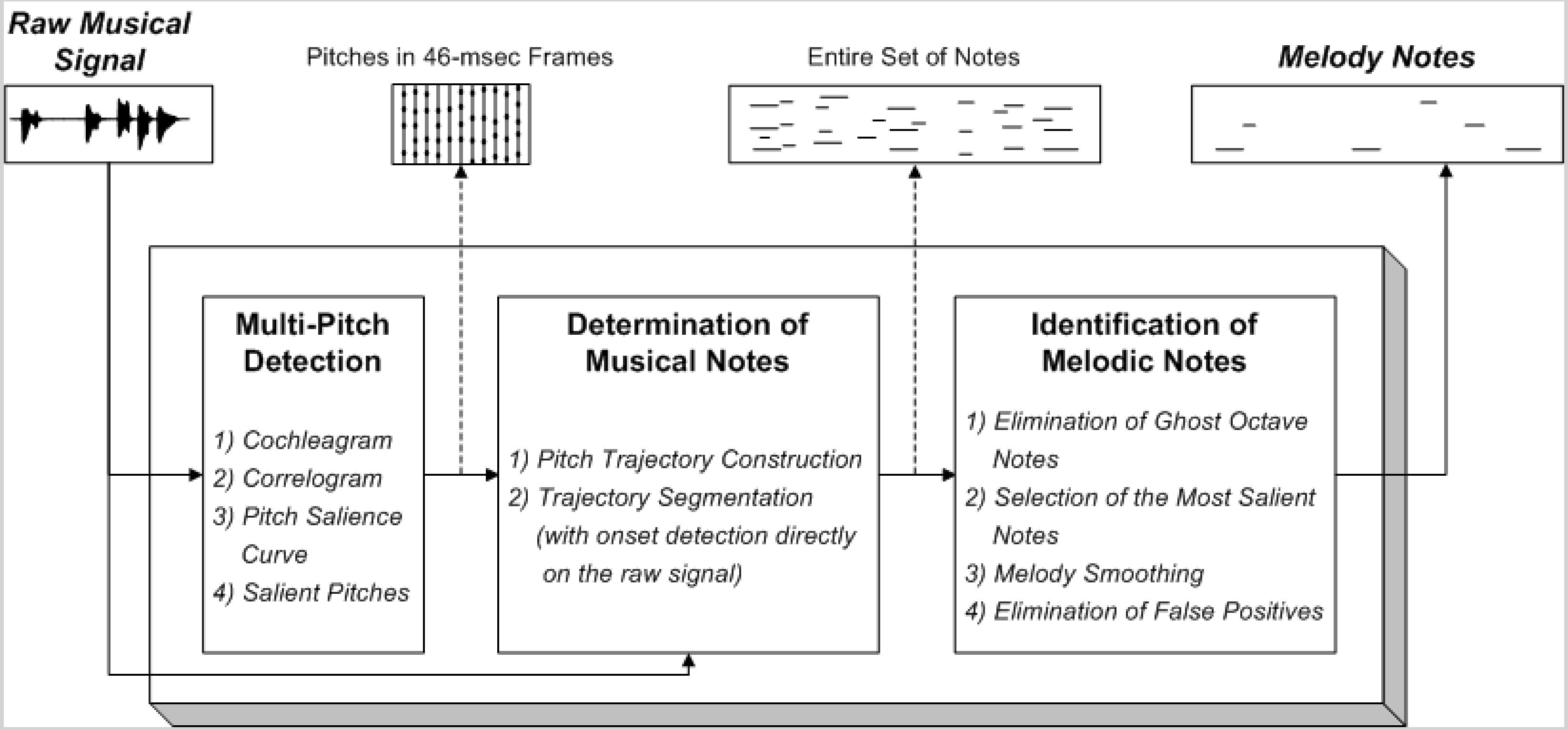


© Rui Pedro Paiva, 2006

  - **Notes from optical music recognition**

Techniques

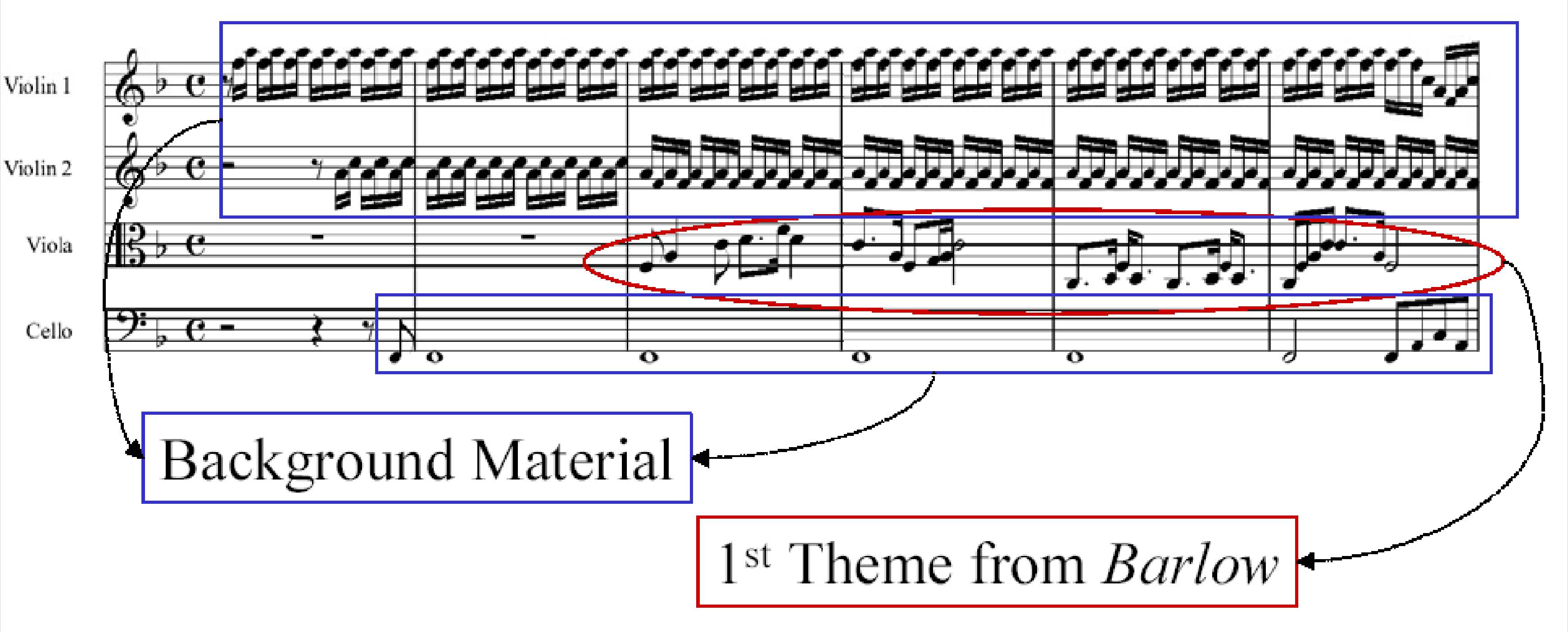
- **Musicological features (cont.)**
  - **Melody**



© Rui Pedro Paiva, 2006

Techniques

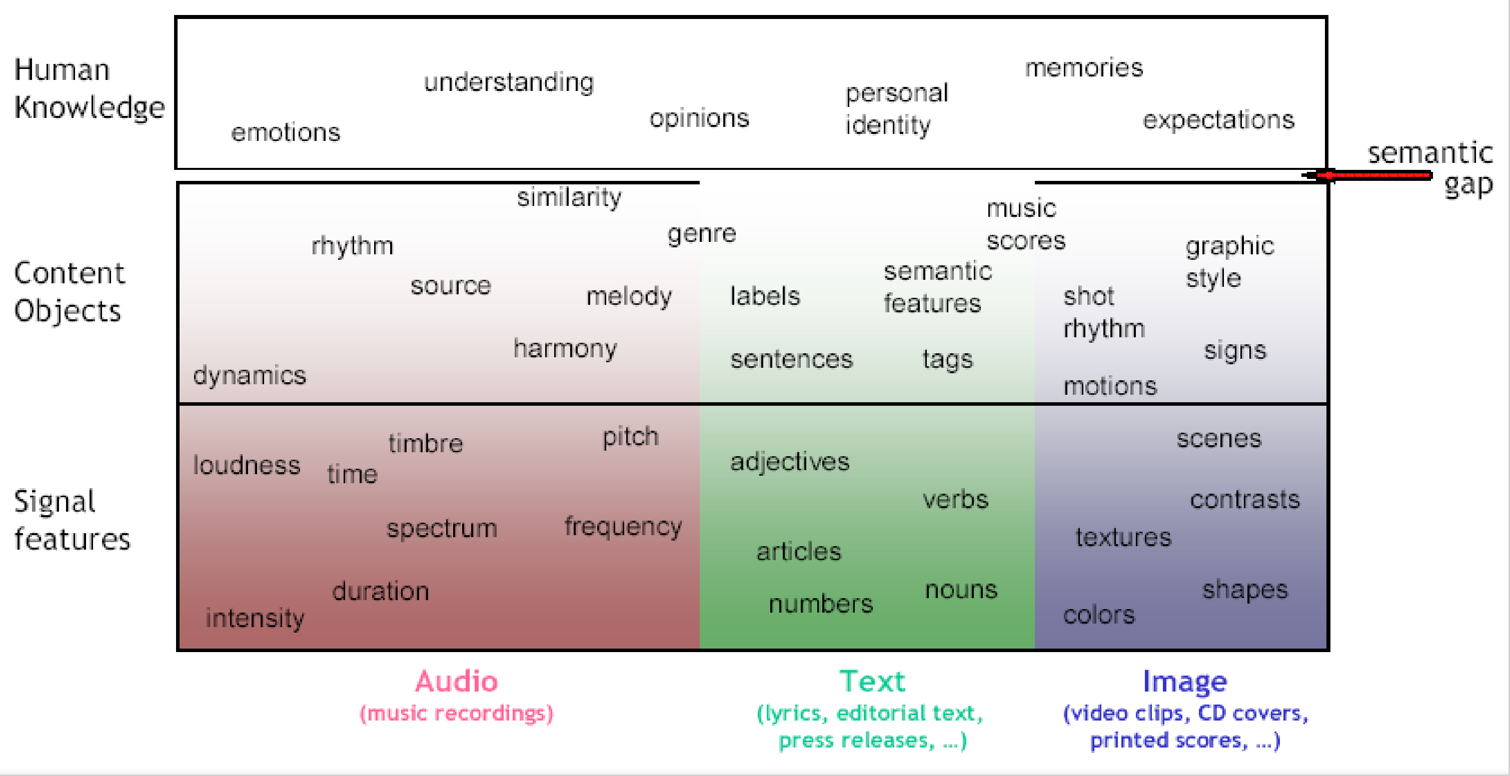
- ❑ **Musicological features (cont.)**
  - **Themes**



© Colin Meek and William Birmingham, 2001

Techniques

- Higher-level features → top-down information flow



© Xavier Serra, 2005

Techniques

- **Higher-level features (cont.)**
  - **Bridge the semantic gap**
  - **Memory, context, expectations**
    - Repetitions, sonic environment, modeling the individual, musicological knowledge
  - **Emotion: *valence* (happy/anxious) and *arousal* (calm/energetic)**
    - Classification approaches resorting to low-level features

- MER: What and Why?
- Applications
- Short MER Tale
- Emotion Models
- Techniques
- MER@CISUC
- Current Limitations and Open Problems

MUSIC EMOTION RECOGNITION (MER)

MER: What and Why?

- **MER**
  - **Music Emotion Recognition: branch of MIR devoted to the identification of emotions/moods 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
  - **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, <u>mood</u>, and similarity information" [Huron, 2000].**
    - Studies on music information behaviour → music mood is an important criterion for music retrieval and organization [Juslin and Laukka, 2004]
  - "In the 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. In the industry, many music companies, such as AMG, Gracenote, MoodLogic, Musicovery, Syntonetic, and Sourcetone use emotion as a cue for music retrieval." [Yang and Chen, 2012]

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**
  - **Mood-based retrieval tools**
    - Music recommendation, automatic playlist generation, classification, …
- ❑ **Game development**
- ❑ **Cinema**
- ❑ **Advertising**
- ❑ **Health**
  - **Sports**
  - **Stress management**

Short MER Tale

- 19th century: initial research on the relations between music and emotion [Gabrielsson and Lindström, 2001]

- 20th century: more active study of music and emotion
  - Relationship between emotions and musical attributes
    - Mode, harmony, tempo, rhythm, dynamics [Meyers, 2007]
  - Mood models
    - E.g., Kate Hevner [Hevner, 1935], James Russell [Russell, 1980]

- 1980-1990s: Initial computational models → mostly emotion synthesis (e.g., [Juslin, 1997])
  - Relationships between emotion and music composition and music expressivity

Short MER Tale

- 1980-1990s: only a few works on emotion analysis
  - Mostly in the symbolic/MIDI domain [Lu *et al*., 2006]

- 2003: first work on mood detection in audio musical signals (by Yazhong Feng [Feng *et al*., 2003])
  - 4 mood categories (happiness, sadness, anger and fear)
  - 2 musical attributes: tempo and articulation
  - Neural network classifier trained using 200 musical pieces
  - Test corpus of 23 pieces, with average precision and recall of 67 and 66%

- 2004: first multi-modal approach combining audio and text (lyrics) [Yang and Lee, 2004]

Short MER Tale

- **2006: first study on <u>MEVD</u> (Music Emotion Variation Detection [Lu *et al*., 2006]**
  - **Temporal variable: emotions may change throughout a song**

- **2007: first mood analysis model based on a <u>dimensional approach</u> [Yang *et al*., 2007]**
  - **Thayer 2D mood model → continuous emotion as function of *arousal* and *valence***

Short MER Tale

- **2007: first <u>MIREX</u> (MIR Evaluation eXchange) track devoted to <u>mood</u> classification [MIREX, 2012]**
  - **Before:**
    - Each approach used a different (and limited) set of features, mood taxonomies, number of classes and test sets.
    - Some approaches constrained the analysis to a particular musical style (e.g., classical music)

- **Present: active research in the field**

Emotion Models

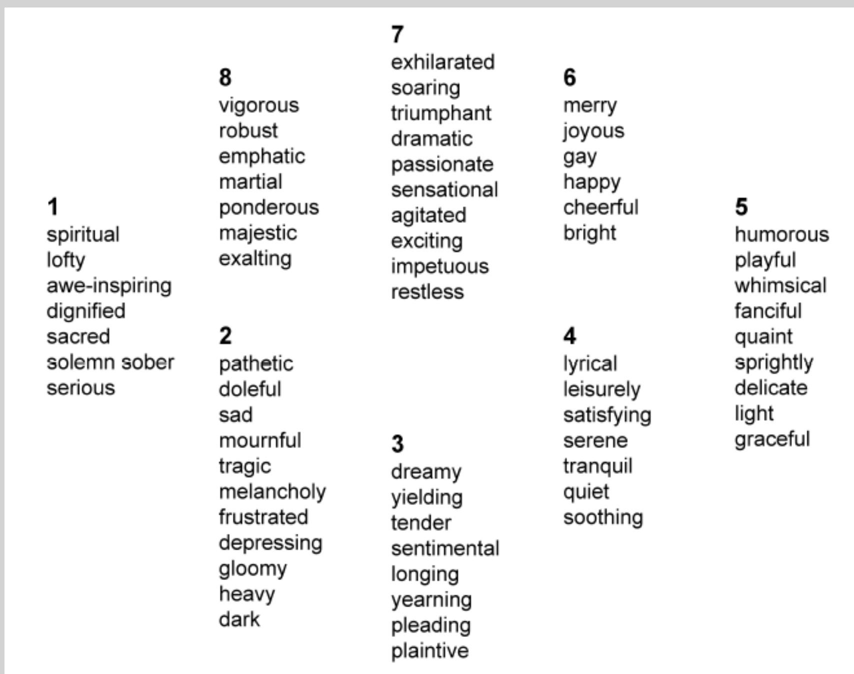
- **Two main conceptualizations of emotion**
  - **Categorical models**
    - Emotions as categories: limited number of **discrete** emotions (adjectives)
  - **Dimensional models**
    - Emotions as **continuous** values, depending on 2 or 3 axes

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) from which all other **“secondary” emotion classes** can be derived

Emotion Models

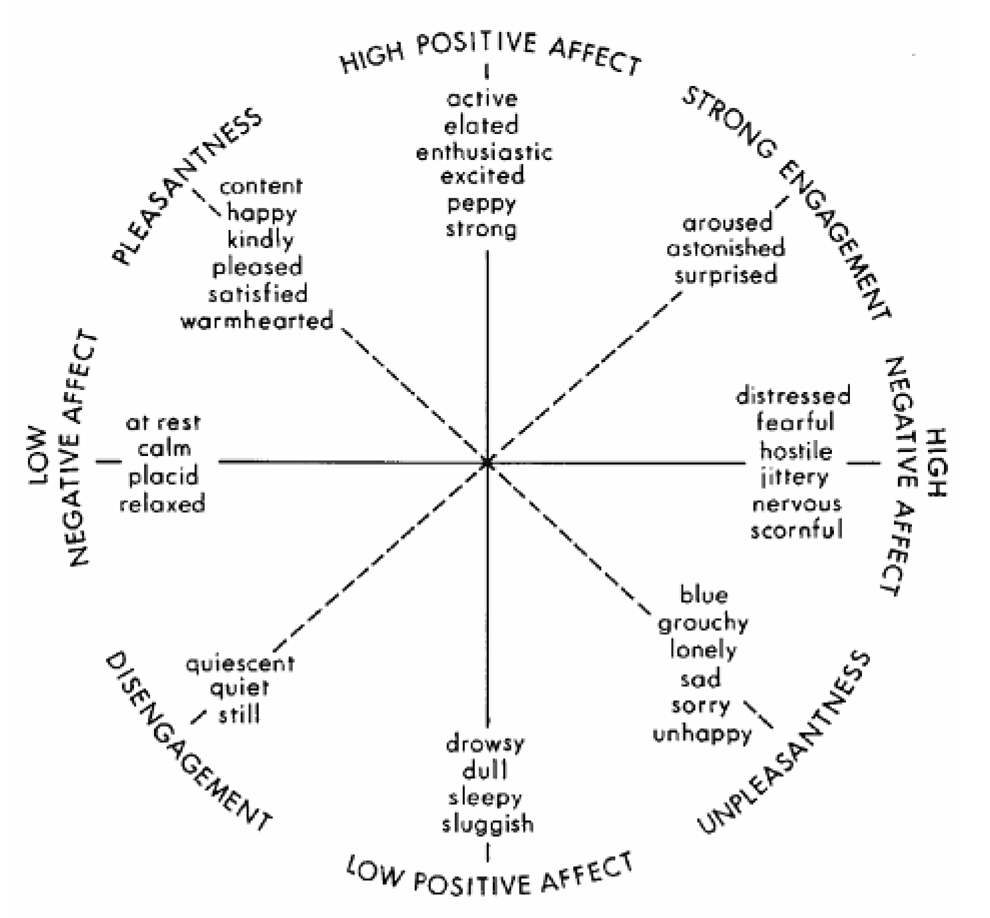
- ❑ **Categorical Models**
  - **Examples**
    - Hevners's 8 clusters of affective terms (1935)
    - Regrouped into 10 adjective groups by Farnsworth [Farnsworth, 1954] and into 9 adjective groups by Schubert [Schubert 2003].



© Yang and Chen, 2012

Emotion Models

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



© Tellegen, Watson and Clark, 1999

Emotion Models

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

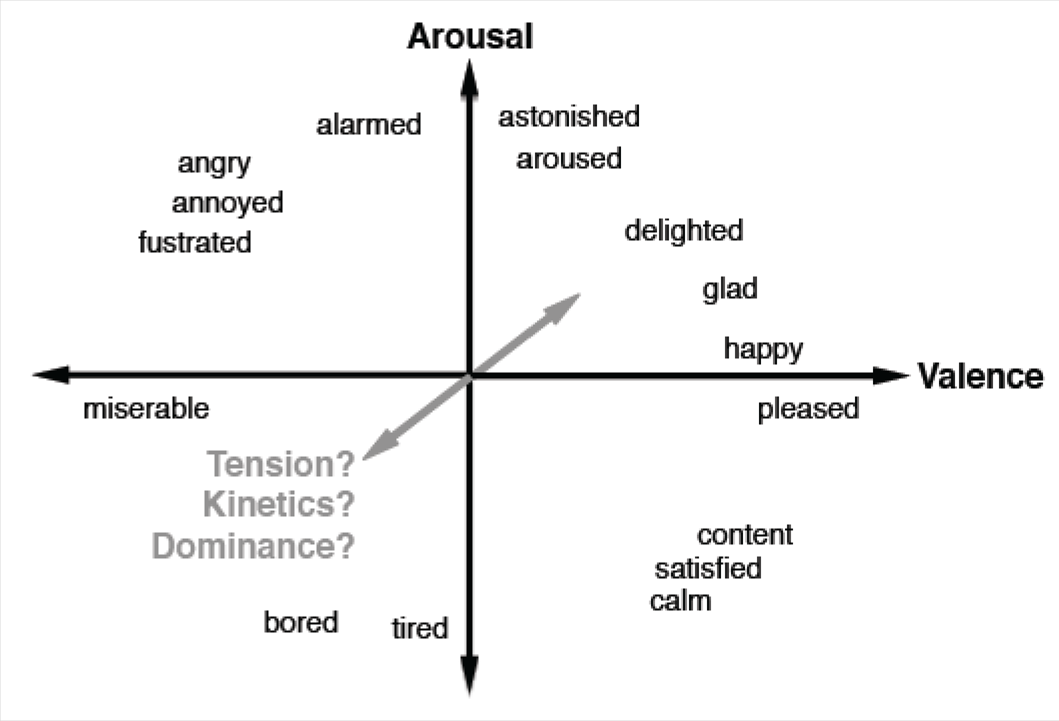
Emotion Models

❑ Dimensional Models

- **Main ideas**
  - Emotions experience as a continuous
    - Each emotion is a **locations in a multi-dimensional plane**, based on a reduced number of axes (2D or 3D)
    - Argument: correspond to internal human representations of emotions
  - 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]

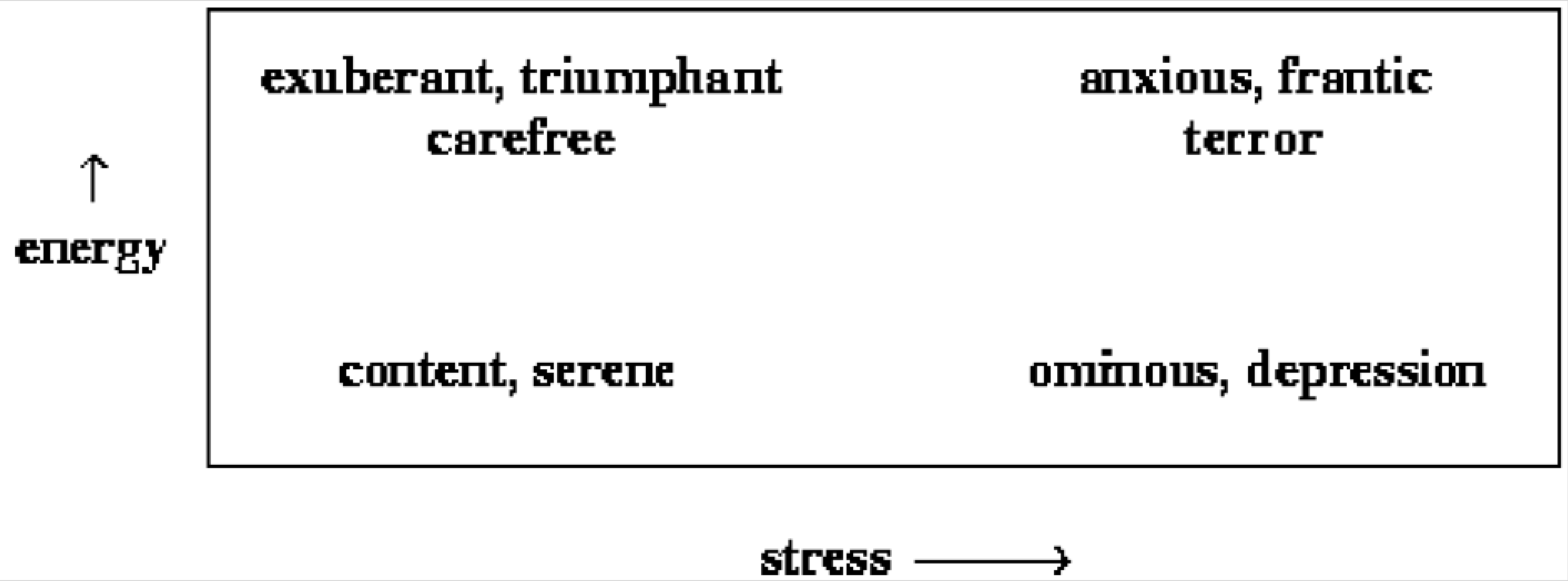


© Kim *et al.*, 2010

Emotion Models

❑ Dimensional Models

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



© Meyers, 2007

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

Techniques

- **Idea**
  - **Extract musical features correlated to emotion from the information source under analysis (audio, MIDI, lyrics, …)**
  - **Temporal variable**
    - Emotions may change throughout a song → MEVD

Techniques

- **Relevant musical attributes**
  - **→ Studies by music psychologists, e.g., [Gabrielsson and Lindström, 2001]**
    - Dynamics, articulation, timbre, pitch, interval, melody, harmony, tonality and rhythm [Friberg, 2008], mode, harmony, tempo, rhythm, dynamics, musical form [Meyers, 2007]
      - 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].
    - 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)

Techniques

- Relevant musical attributes
  - Timing: Tempo, tempo variation, duration contrast
  - Dynamics: overall level, crescendo/decrescendo, accents
  - Articulation: overall (staccato/legato), variability
  - Timbre: Spectral richness, onset velocity, harmonic richness
  - Pitch (high/low)
  - Interval (small/large)
  - Melody: range (small/large), direction (up/down)

Techniques

- Relevant musical attributes
  - Harmony (consonant/complex-dissonant)
  - Tonality (chromatic-atonal/key-oriented)
  - Rhythm (regular-smooth/firm/flowing-fluent/irregular-rough)
  - Mode (major/minor)
  - Loudness (high/low)
  - Musical form (complexity, repetition, new ideas, disruption)

Techniques

- **Overall approach: classical pattern recognition approach**
  - **1) Selection of a mood model**
    - Categorical or dimensional?
    - How many categories / variables? Multi-label or probabilistic classification?
  - **2) Acquisition of training and testing data**
    - Necessary to perform manual data annotation? How many annotators, what profiles, how many songs, song balance, etc.?
  - **3) Feature extraction and selection**
    - Which features? How to extract them from the information source?
    - How to evaluate the impact of different features in each class?

Techniques

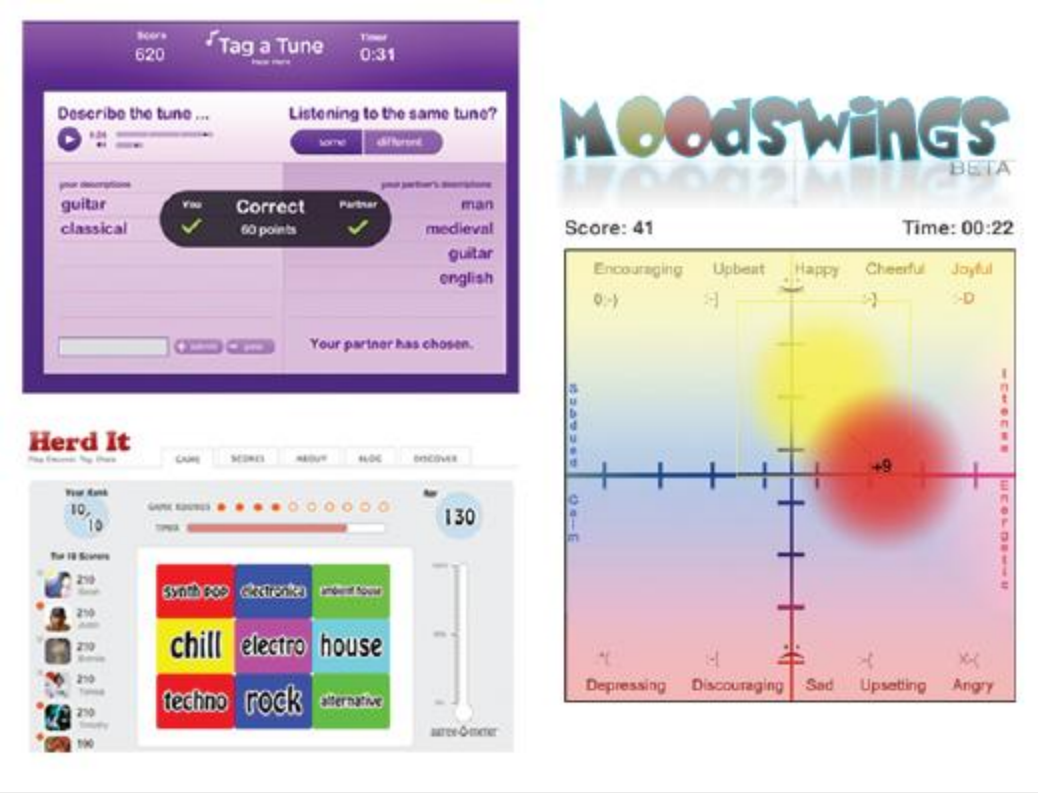
- **Overall approach: classical pattern recognition approach**
  - **4) Selection of a representation model**
    - Black-box, gray-box, white-box?
      - **Black-box** paradigm dominates. **Gray-box** should be researched.
    - SVM, neural networks, GMM, KNN, neural-fuzzy classifiers, rule-base systems?
      - **SVM**s: most accurate results currently
    - NN: how many hidden layers, etc? SVM: which parameters?...
      - SVM parameter optimization: **grid-search**
  - **5) Model training**
    - How to organize the training and testing experience? Cross validation? How many folds?
      - Tipically, **10-fold cross validation**

Techniques

- Overall approach: classical pattern recognition approach
  - **6) Model validation**
    - How to quantitatively evaluate the results? RMSE, SSE, correlation, R2 statistics? Precision, recall and F-measure? Confusion matrix?
      - Regression: **RMSE** and **R2** are tipically used
      - Classification: **Precision, recall and F-measure** are standard
    - What kinds of errors are more prevalent? Why? Where is the root of the problem: classifier/regressor, extracted features, method employed for extraction, feature selection approach?
      - Currently, the bottleneck seems to be the **lack of semantic richness** of current **low-level features** and the **innacuracy of particular feature extraction algorithms** (e.g., tempo, etc.)
    - What classes/variables show low accuracy? Why?
      - **Valence**: important features might be missing
    - Same questions for individual songs.

Techniques

- Acquisition of training and testing data [Kim *et al*., 2010]
  - Subjective tests
  - Annotation games



© Kim *et al*., 2010

Techniques

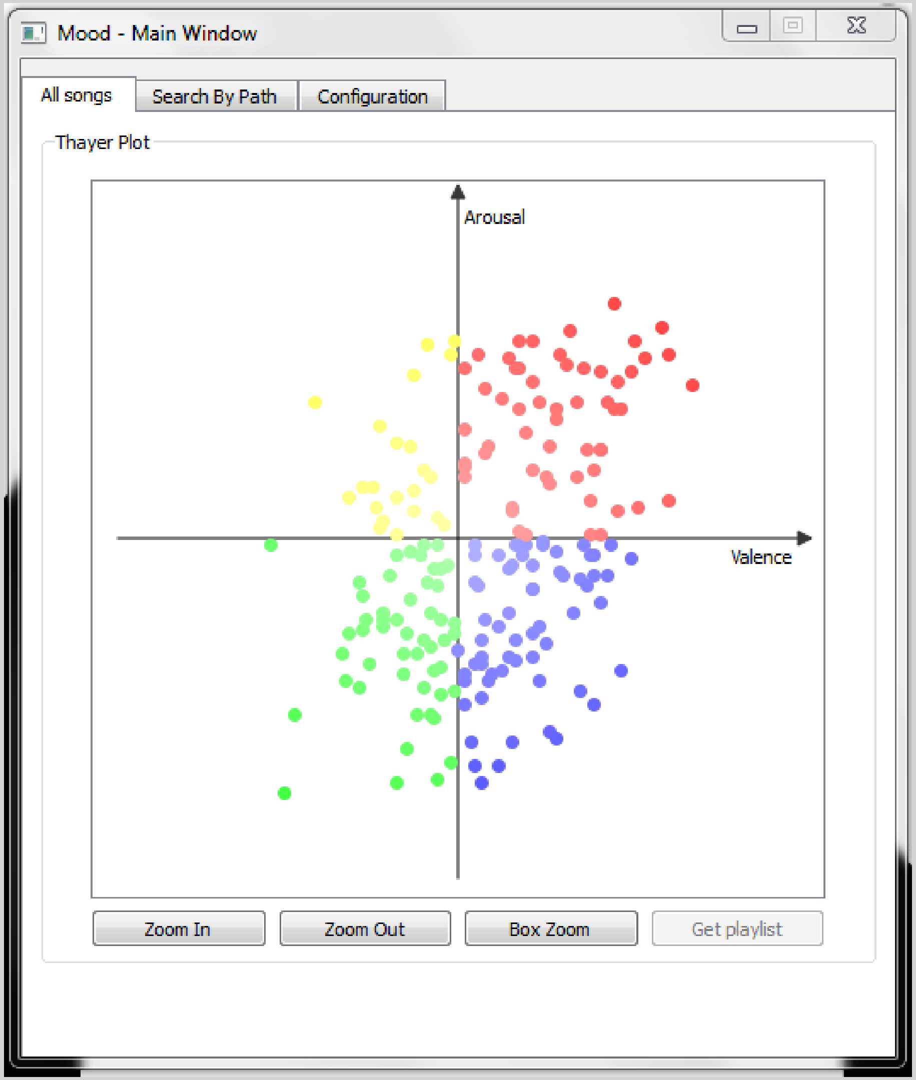
- **Feature Extraction**
  - **Music platforms**
    - Audio: Marsyas, MIR Toolbox, PsySound, …
    - MIDI: jMIR, …
    - Lyrics: ConceptNet, …
  - **Audio: acoustic features developed in other contexts**
    - Music genre classification (e.g., spectral shape features, Mel-frequency cepstral coefficients (MFCC), etc.) and speech recognition
      - Low-level audio descriptors (LLDs): spectral shape features such as centroid, spread, bandwidth, skewness, kurtosis, slope, decrease, rolloff, flux contrast, MFCCs, etc.
  - **Necessary to**
    - Evaluate the relevance of LLDs to mood detection
    - **Program actual mood-related features**

Techniques

- 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

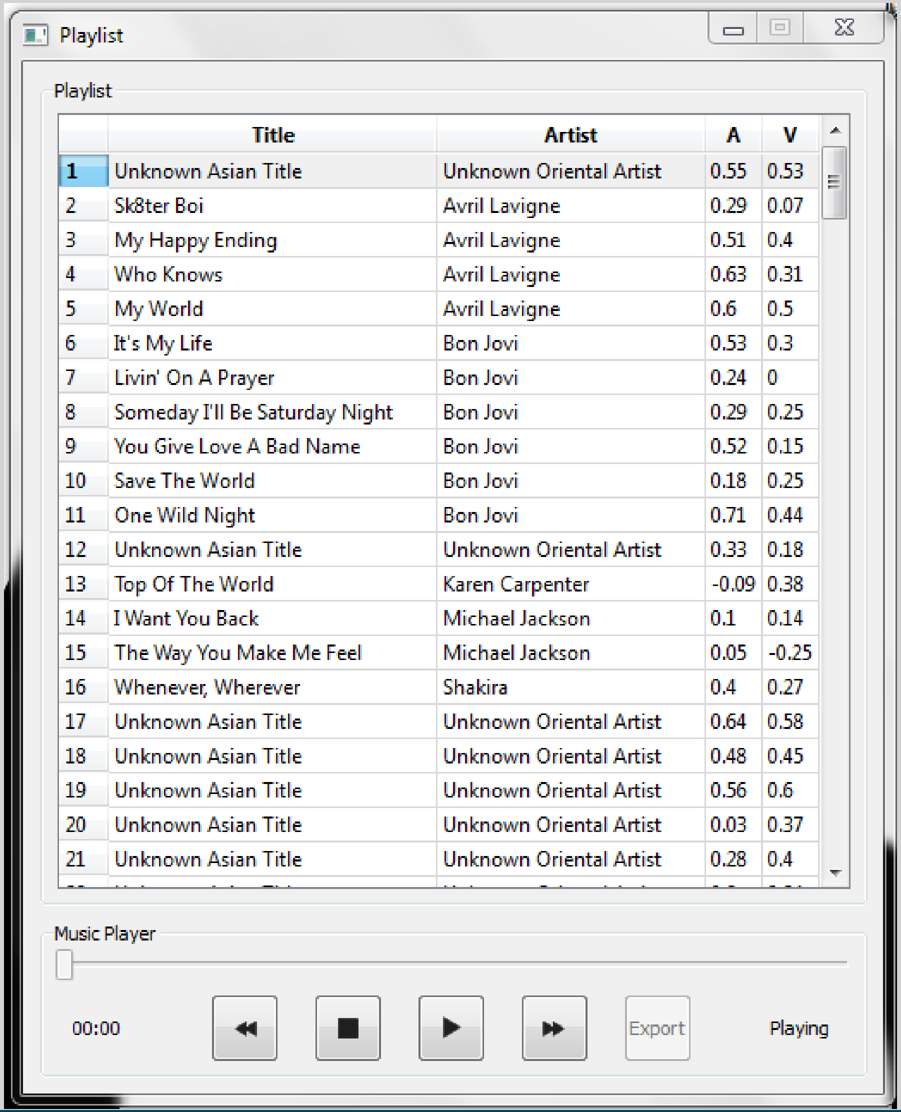
MER@CISUC

- **MOODetector project**
  - **Search**



MER@CISUC

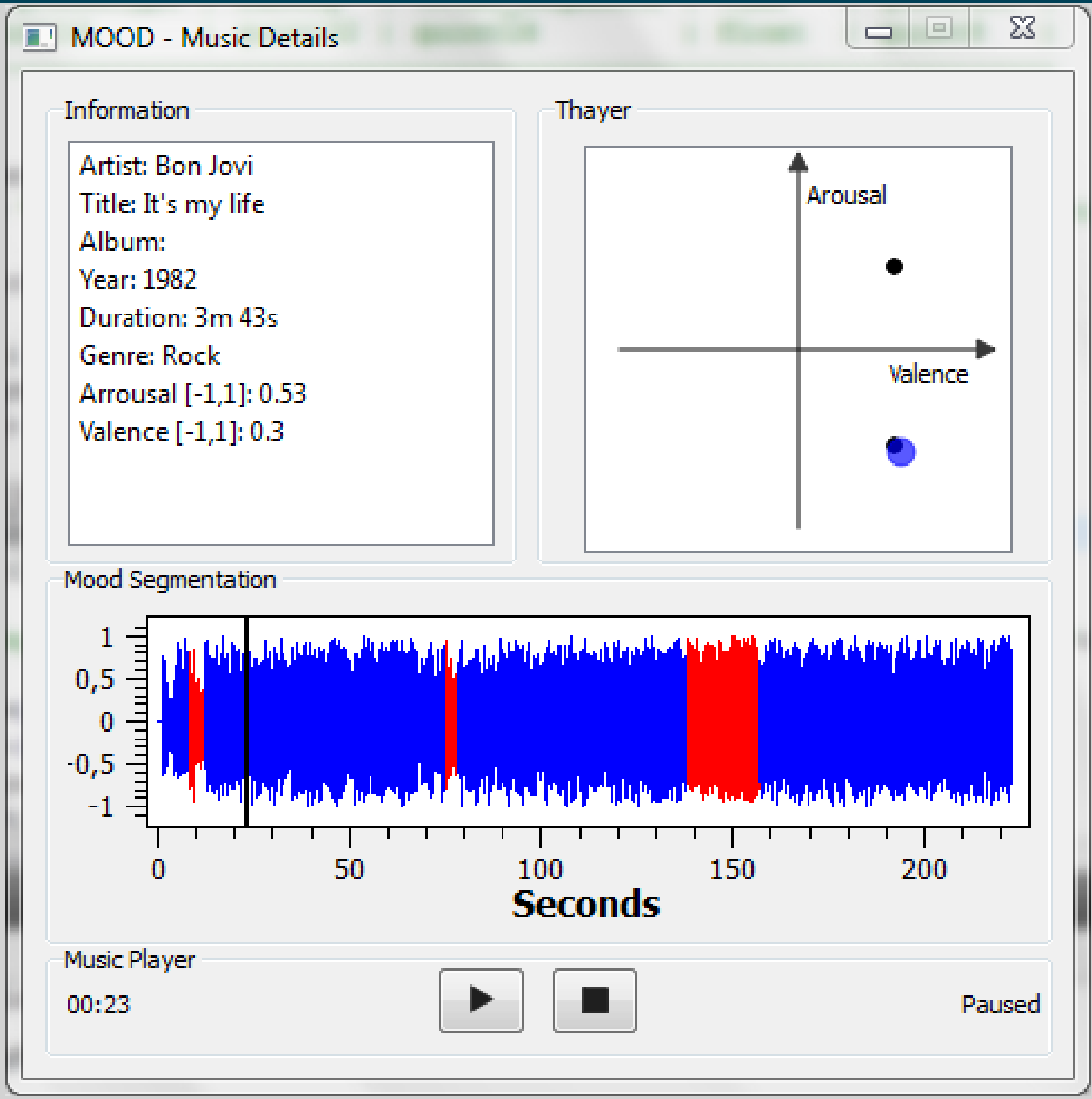
- **MOODetector project**
  - **Generate playlist**



	Title	Artist	A	V
1	Unknown Asian Title	Unknown Oriental Artist	0.55	0.53
2	Sk8ter Boi	Avril Lavigne	0.29	0.07
3	My Happy Ending	Avril Lavigne	0.51	0.4
4	Who Knows	Avril Lavigne	0.63	0.31
5	My World	Avril Lavigne	0.6	0.5
6	It's My Life	Bon Jovi	0.53	0.3
7	Livin' On A Prayer	Bon Jovi	0.24	0
8	Someday I'll Be Saturday Night	Bon Jovi	0.29	0.25
9	You Give Love A Bad Name	Bon Jovi	0.52	0.15
10	Save The World	Bon Jovi	0.18	0.25
11	One Wild Night	Bon Jovi	0.71	0.44
12	Unknown Asian Title	Unknown Oriental Artist	0.33	0.18
13	Top Of The World	Karen Carpenter	-0.09	0.38
14	I Want You Back	Michael Jackson	0.1	0.14
15	The Way You Make Me Feel	Michael Jackson	0.05	-0.25
16	Whenever, Wherever	Shakira	0.4	0.27
17	Unknown Asian Title	Unknown Oriental Artist	0.64	0.58
18	Unknown Asian Title	Unknown Oriental Artist	0.48	0.45
19	Unknown Asian Title	Unknown Oriental Artist	0.56	0.6
20	Unknown Asian Title	Unknown Oriental Artist	0.03	0.37
21	Unknown Asian Title	Unknown Oriental Artist	0.28	0.4

MER@CISUC

- **MOODetector project**
  - **Mood Tracking**



Limitations and Open Problems

❑ Current Limitations

- **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 ~ 65% accuracy in a 5-class mood problem [MIREXresults, 2010]

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
  - **Data sets**
    - Real-world sizeable datasets for categorical, dimensional and MEVD approaches
  - **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
  - **Extraction of knowledge from computational models**
    - E.g., neural-fuzzy approaches [Paiva and Dourado, 2004]

- Suggested Initial Literature
- Cited References

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

- 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.
- 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.

Cited References

- 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.
- 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.

Cited References

- 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.
- Meyers O. C. (2007). “A Mood-Based Classification and Exploration System”. MSc Thesis, School of Architecture and Planning, Massachusetts Institute of Technology.

Cited References

- 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.
- 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.

Cited References

- 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.
- 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.

Cited References

- 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.
- 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.

Cited References

- 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
- 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

- 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.