Music Information Retrieval
Based on Social Tags and Emotion

Hsin-Min Wang
whm@iis.sinica.edu.tw

Speech, Language and Music Processing Laboratory
Institute of Information Science
Academia Sinica, Taipei, Taiwan
http://slam.iis.sinica.edu.tw
Institutional divisions

- Mathematics and Physical Sciences
- Life Sciences
- Humanities and Social Sciences
People

- **269 Academicians**
  - 102 are Domestic Academicians and 167 are Overseas Academicians
  - 117 Mathematics and Physical Sciences; 92 in Life Sciences; and 60 in Humanities and Social Sciences

- **850 Principle Investigators**
  - 110 Distinguished Research Fellows; 341 Research Fellows; 245 Associate Research Fellows; 154 Assistant Research Fellows

- **91 Specialists**
  - 18 Senior Research Specialists; 29 Associate Research Specialists; 44 Assistant Research Specialists.

- **1044 Postdoctoral Fellows**

- **>2000 research assistants and students**
Research Groups at IIS

- Bioinformatics Laboratory
- Computation Theory and Algorithms Laboratory
- Computer System Laboratory
- Information Processing and Discovery (iPAD) Laboratory
- Multimedia Technologies Laboratory
- Natural Language and Knowledge Processing Laboratory
- Network System and Service Laboratory
- Programming Languages and Formal Methods Laboratory
People

39 Principle Investigators
- 5 Distinguished Research Fellows; 20 Research Fellows; 11 Associate Research Fellows; 3 Assistant Research Fellows

1 Assistant Research Specialist

41 Postdoctoral Fellows

~200 research assistants and students
Our Research Topics

- **Speech**
  - Speech/Speaker/Language recognition
  - Spoken document retrieval/summarization
  - Speech synthesis/voice conversion

- **Music**
  - Music tag annotation and retrieval
  - Emotion-based music annotation and retrieval
  - Music retrieval using query-by-singing/humming/example
  - Context-aware music and recommendation
Music Information Retrieval (MIR)

- Users need to find the “right” songs for
  - a specific listening context (driving, studying, exercising)
  - a specific mood (sad, happy, angry)
  - a specific event (wedding, party)
  - accompanying a video (home video)

- Current solution
  - Manual browsing or selection
  - Keyword search (artist, title, lyrics)
  - Social recommendation
  - Content-based retrieval (query-by-singing/humming, fingerprinting)
Outline

- Social Tagging-based MIR
  - Automatic music tagging, Tag-based music retrieval, Query interfaces, Real-time music tagging, Music player visualization

- Emotion-based MIR
  - Music emotion recognition (MER), Emotion-based music retrieval, Query interfaces

- Relationship between Tags and Dimensional Emotions of Music

- Automatic Generation of Music Video
Outline

■ Social Tagging-based MI R
  ➢ Automatic music tagging, Tag-based music retrieval, Query interfaces, Real-time music tagging, Music player visualization

■ Emotion-based MI R
  ➢ Music emotion recognition (MER), Emotion-based music retrieval, Query interfaces

■ Relationship between Tags and Dimensional Emotions of Music

■ Automatic Generation of Music Video
Social Tagging

- Social tagging, also known as collaborative tagging, is the process by which many users add metadata in the form of keywords to shared content (Golder and Huberman, 2006)

- In the Web 2.0 era, people tag various kinds of shared content on the web
  - Image – Flickr
  - Music – Last.fm

- These tags form a folksonomy (大眾分類法)
Social Tags for Images

Postcard Beach
It didn't take much editing to get it there. A polarizing filter in late afternoon sun did most of the work, the old fashioned way. Just an excuse for more palm tree photos :)

Comments and favorites

Tags
- palm
- trees
- beach
- sand
- sky
- blue
- rocks
- ocean
- water

Additional info
- Resolution: 1180 x 783, 300 x 600
- License: All Rights Reserved

Institute of Information Science, Academia Sinica, Taiwan
Social Tags for Music

Music tags describe different aspects of a music clip, e.g., genre, mood, instrumentation, users’ preference.

lost.fm

Tags

60s 70s acoustic alternative alternative rock amazing awesome ballad ballads beatles beautiful brilliant british british invasion britpop calm chill chillout classic classic rock classics cool downtempo easy listening english favorite favourites favourite songs favourites good great guitar indie john lennon love male vocalist male vocalists melancholic melancholy mellow moody night oldies paul mccartney perfect piano pop pop rock psychedelic rock rock ballad rolling stones top 500 songs of all time sad singer-songwriter sweet the beatles uk uplifting 1970
Tag Sparsity (Cold-start)

- No tags available for unpopular or new music tracks
- Research into automatic music tag annotation and tag-based music retrieval from an untagged music database
Audio Tag Classification in MIREX

- Music Information Retrieval Evaluation eXchange (MIREX) is an annual evaluation campaign for Music Information Retrieval (MIR) algorithms since 2005

- Audio Tag Classification became an evaluation task in 2008
  - To associate descriptive tags with 10-second audio clips of songs

- Audio Train/ Test Tasks
  - Audio Artist Identification - artist tags
  - Audio Genre Classification - genre related tags
  - Audio Music Mood Classification - mood related tags
  - Audio Classical Composer Identification - composer tags
Music Tag Annotation and Retrieval

Annotating music clips with tags

A Music Clip

Annotate Music Using One Predictor for Each Tag

Scores of Tag Predictors

Female R&B Guitar Metal Bass

Retrieving music clips using a tag query

A Query: Rock

Rank Music Clips based on the Scores of the Rock Predictor

Ranked List for the Query

High Relevance Low Relevance
Issues for MIR Using Social Tags

■ Jointly model the auditory features and music tags of a tagged music corpus (training)

■ Apply the model for content-based music retrieval from an untagged music database by tag-based queries
Web-based Music Tagging Games

- Web-based music tagging games with a purpose (GWAP) for collecting useful and reliable tags

**MajorMiner.org**

- Search
  - music using automatic descriptions
  - Qualities: british, soft, ambient, distortion, solo, acoustic, dark, hard
  - Instruments: guitar, piano, strings, violin, flute

- Browse
  - music with similar descriptions
  - at 1:00 in "Mary" from Patty Griffin’s album *Concerts for a Landmine Free World* [similar] Autotags: female, soft, vocal, soulful
  - at 0:30 in "Longfellow" from The Gentle Giant’s album *Our Little Rocket* [similar] Autotags: ballad, acoustic, soft, love, sad, piano, slow
  - at 1:30 in "Smiles (feel session)" from Spiritualized’s album *Friendly Fire* [similar] Autotags: distortion, instrumental, guitar, metal, noise, loud, rock, heavy, punk
  - at 1:00 in "American Boy (MINR go west remix)" from Estelle’s album *Let’s Get It: The Album* [similar] Autotags: R&B, female, hip-hop, soulful, vocal, duet, acoustic

- Play
  - a game about describing music

**Tag a Tune**

**Institute of Information Science, Academia Sinica, Taiwan**
Tagged Data from MajorMiner.org

- 10-second music audio clips with tag labels
- Multiple tags with different counts

Table 1. Some Examples of Audio Clips with Associated Tags Obtained from the MajorMiner Website

<table>
<thead>
<tr>
<th>Song</th>
<th>Album</th>
<th>Clip Start Time</th>
<th>Artist</th>
<th>Associated Tags (Tag Counts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hi-Fi</td>
<td>Head Music</td>
<td>0:00</td>
<td>Suede</td>
<td>drum (9), electronic (3), beat (2)</td>
</tr>
<tr>
<td>Universal</td>
<td>Talkie Walkie</td>
<td>4:00</td>
<td>Air</td>
<td>synth(7), electronic(4), vocal(5), female(4) voice(2), slow(2), ambient(2), soft(3), r&amp;b (3)</td>
</tr>
<tr>
<td>Traveler</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safe</td>
<td>Travis</td>
<td>1:00</td>
<td>The Invisible Band</td>
<td>guitar(5), male(4), pop(4), vocal(3), acoustic(2)</td>
</tr>
<tr>
<td>Moritat</td>
<td>Saxophone Colossus</td>
<td>0:50</td>
<td>Sonny Rollins</td>
<td>jazz(9), saxophone(12)</td>
</tr>
<tr>
<td>Pacific Heights</td>
<td>Ascension</td>
<td>2:30</td>
<td>Pep Love</td>
<td>male(4), synth(2), hip hop(8), rap (6)</td>
</tr>
<tr>
<td>Trouble</td>
<td>The Chillout</td>
<td>3:40</td>
<td>Coldplay</td>
<td>male(6), pop(3), vocal(5), piano(7)           voice(3), slow(2), soft(2), r&amp;b(2)</td>
</tr>
</tbody>
</table>
Music Tag Co-occurrence (Co-Tag)

- **Music tags include different types of musical information**
  - genre, mood, instrumentation, personal preference, original artists, etc.

- **Co-tag examples**
  - guitar, distortion, and drum commonly result in rock, loud, and metal
  - saxophone and piano usually lead to jazz and soft
  - rap mostly co-occurs with hip-hop
  - synth and beat often give electronic and dance
Tag Popularity (Observed by Tag Counts)

- A tag will enter the common musical vocabulary once it is adopted by a large number of users.
- A set of commonly used music tags can be predefined.
- We propose a content-based music retrieval system that accepts a query comprised of multiple tags with multiple levels of preference (denoted as an MTML query hereafter) to search for music in an untagged music database.
MTML Query Interface – Coloring Tags in Tag Cloud

Online: http://slam.iis.sinica.edu.tw/demo/SoTags/

80s acoustic ambient bass beat british country
dance distortion drum-machine
drums electronic electronica
fast female funk guitar
hip-hop horns house instrumental jazz
keyboard loud male metal noise organ
piano pop punk quiet r&b rap rock
saxophone slow soft solo strings
synth techno trumpet vocal voice
Modeling Auditory Features and Music Tags Through Latent Feature Classes

When tagging a song, people commonly choose one or more un-describable co-tag patterns according to the auditory characteristics of the song.

Define latent feature classes to link the auditory features and the music tag distribution.

Gaussian Mixture Model

Tag-based Aspect Model

A Group of Music Feature Vectors → Latent Feature Classes

Latent Co-tag Pattern $\beta$

Music Tag Distribution

Auditory Music Features
Acoustic GMM Training

Acoustic GMM (Gaussian Mixture Model) for auditory feature encoding

Frame-based
Feature vectors

EM (Expectation-Maximization) Training

Audio signal
Generative Flow of Tag-based Music Aspects over $K$ Latent Feature Classes

Feature Vectors

GMM

$$p(x) = \sum_{k=1}^{K} \pi_k N_k(x \mid \mu_k, \Sigma_k)$$

Frame-level Encoding

Frame Posterior

$$p(z_k \mid x_t) = \frac{\pi_k N_k(x_t \mid \mu_k, \Sigma_k)}{\sum_{h=1}^{K} \pi_h N_h(x_t \mid \mu_h, \Sigma_h)}$$

Song-level Encoding

Song Representation

$$\theta_k \leftarrow p(z_k \mid s) = \frac{1}{T} \sum_{t=1}^{T} p(z_k \mid x_t)$$

Global GMM

GMM Posterior
Generative Flow of Tag-based Music Aspects over $K$ Latent Feature Classes

- Feature Vectors
- Global GMM
- GMM Posterior
- Tag-based Music Aspects (Multinomial)
- $p(w_m \mid s) = \sum_k \theta_k \beta_{km}$
Learning the Tag-based Music Aspect Model by the EM Algorithm (Maximum Likelihood)

**Tag-Level**

\[
p(w_m | s_n) = \sum_{k=1}^{K} p(z_k | s_n) p(w_m | z_k) = \sum_{k=1}^{K} \theta_{nk} \beta_{km}
\]

**Song-Level**

\[
p(w | s_n) = \prod_{m=1}^{M} p(w_m | s_n)^{c(n,m)} = \prod_{m=1}^{M} \left( \sum_{k=1}^{K} \theta_{nk} \beta_{km} \right)^{c(n,m)}
\]

**Corpus-Level**

\[
L = \sum_{n=1}^{N} \sum_{m=1}^{M} c(n,m) \log \sum_{k=1}^{K} \theta_{nk} \beta_{km}
\]

Maximize the objective function \( L \) w.r.t. \( \beta_{km} \) by the EM algorithm.

Tag-based Music Aspects:

- \( \beta_1 \)
- \( \beta_2 \)
- \( \beta_{K-1} \)
- \( \beta_K \)
MTML MIR - System Overview

1. GMM Posterior ($\theta_1 \sim \theta_K$)

2. Co-Tag Distribution

Feature Extraction → Feature Vectors → Feature Indexing

Music Tracks

Indexing Phase

Music Database

Matching based on VSM Cosine Similarity

Tag Queries

rock 0.3,
pop 0.9,
synth 0.7,
female 1.0

Retrieved Music

Retrieval Phase

<table>
<thead>
<tr>
<th>Approach</th>
<th>Indexing</th>
<th>Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold-in</td>
<td>GMM Posterior</td>
<td>Pseudo Song-based Matching</td>
</tr>
<tr>
<td>Auto-tagging</td>
<td>Co-Tag Affinity</td>
<td>Co-Tag Affinity-based Matching</td>
</tr>
</tbody>
</table>

Institute of Information Science, Academia Sinica, Taiwan
Given an untagged music clip in the untagged music database.
Online Estimation of a Pseudo Song by the Tag-based Music Aspects Model

Learned Tag-based Music Aspects

\[ \beta_1 \rightarrow \lambda_1 \]
\[ \beta_2 \rightarrow \lambda_2 \]
\[ \beta_{K-1} \rightarrow \lambda_{K-1} \]
\[ \beta_K \rightarrow \lambda_K \]

Compare the pseudo song with GMM posterior-based songs in the database

MTML Query

GMM Posterior Like Representation

Pseudo Song
### Matching Methods

<table>
<thead>
<tr>
<th>Name</th>
<th>Indexing Vector</th>
<th>Matching Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>System 1</td>
<td>Tag Affinity Distribution</td>
<td>Tag Affinity-based Matching</td>
</tr>
<tr>
<td>System 2</td>
<td>Auditory Posterior Distribution</td>
<td>Pseudo Song-based Matching</td>
</tr>
</tbody>
</table>

![Diagram showing the process of matching methods](image)
Real-Time Music Tagging

- Visualizing audio signals during playback has long been a fundamental function of music players
- Most visual effects are generated by audio signal processing directly and render meaningless or incomprehensible displays

How about visualizing the dynamic tag distribution during music playback?
Synchronization of Visual Display and Music Playback

- Frames of each chunk are summarized by an acoustic GMM posterior representation.
- Prediction of the tag distribution is just a simple linear combination.
From CAL500 to CAL500exp (1/2)

Track-level CAL500

Piano, Violin, Cello, Male-Vocal, and Female-vocal

Segment-level CAL500exp

Violin, and Male-vocal  Piano  Cello  Violin  Piano, and Female-vocal
From CAL500 to CAL500exp (2/2)

1. Segment and Clustering
   We adopt Foote & Cooper’s segmentation algorithm to pre-process every track in CAL500. After that, there are in total 18,664 segments, with each track being partitioned to 37.3 segments on average. Because many segments of a song could be repetitive, we perform k-medoids clustering to obtain the representative segments of each song, leading to on average 6.4 representative segments per track.

2. Segment-level music player
   Segment player with the repeat function

3. Tag categories
   Show a category of tags at a time

4. Tag label checkbox & initialization
   To alleviate the annotation labor, we initialize the tag labels as default by using CAL500 labels and retagging [1] for each segment and ask the subjects to modify the default labels by insertion and deletion in the checkbox.

5. Friendly copy
   The tag labels of some representative segments of a track might be similar. Therefore, we include a ‘copy’ function: for a new segment, subjects can copy the tag labels of a previously done segment and make modification upon them.

6. Addition song information
   If subjects intend to look for more information about the song, they can be redirected to the Last.fm page of this song.

7. Track-level music player
   Global view of the whole track
Music Segmentation Using Self-Similarity Matrix

Foote and Cooper, 2003
Outline

- **Social Tagging-based MIR**
  - Automatic music tagging, Tag-based music retrieval, Query interfaces, Real-time music tagging, Music player visualization

- **Emotion-based MIR**
  - Music emotion recognition (MER), Emotion-based music retrieval, Query interfaces

- **Relationship between Tags and Dimensional Emotions of Music**

- **Automatic Generation of Music Video**
Emotions in MIR

Why is emotion so important in MIR?
- Music is the finest language of emotion
- People use music to convey or modulate emotion
- Music cannot be composed, performed, or listened to without affection involvement
- Smaller semantic gap, compared to genre
- Could be a clue for context-aware music recommendation

Expressed or Perceived
- Indicate the mood conveyed by a music piece
- Indicate how you feel when listening to a music piece
Emotions as Discrete Categories

- 5 mood categories used in the MI REX Audio Mood Classification task:
  - Cluster_1: passionate, rousing, confident, boisterous, rowdy
  - Cluster_2: rollicking, cheerful, fun, sweet, amiable/good natured
  - Cluster_3: literate, poignant, wistful, bittersweet, autumnal, brooding
  - Cluster_4: humorous, silly, campy, quirky, whimsical, witty, wry
  - Cluster_5: aggressive, fiery, tense/anxious, intense, volatile, visceral

- Debate on categorical emotions

- Music emotion recognition is a classification problem
Dimensional Emotion

**Activation–Arousal**
- Energy or neurophysiological stimulation level

**Evaluation–Valence**
- Pleasantness
- Positive and negative affective states

The Circumplex model (Russell 1980)

- Annoying
- Angry
- Nervous
- (negative)

- Excited
- Happy
- Pleased
- (high)

- Sad
- Bored
- Sleepy
- Calm
- (low)
- Relaxed
- Peaceful
- (positive)
Emotion-based MIR Systems

Musicover
Play your mood

Energetic

Dark

Calm

Mufin Player

Mr. Emo developed by Yang et al.

Institute of Information Science, Academia Sinica, Taiwan
The Valence-Arousal Model

- Emotions are considered as numerical values (instead of discrete labels) over some emotion dimensions
- Good visualization and intuitive
- Easy to capture **temporal change** of emotion
Valence-Arousal Annotations

- Emotion is subjective, but aggregation of annotations among users exists.
- Dimensional emotion of a song can be described by a bivariate Gaussian distribution.
- Predict the emotion of a song as a single Gaussian.
Regression for Gaussian Parameters

- The **Gaussian-parameter approach** directly learns five regression models to predict the mean, variance, and covariance of valence and arousal, respectively.
- No joint modeling and estimation of the Gaussian parameters.
The Acoustic Emotion Gaussians Model

- A principled probabilistic/statistical approach
- Represent the acoustic features of a song by a probabilistic histogram vector
- Develop a model to comprehend the relationship between acoustic features and VA space (annotations)

Acoustic GMM posterior representations of songs
Generative Process of VA GMM (1/2)

A corpus with audio signals and associated VA annotations

Audio Signal of Each Clip
Generative Process of VA GMM (2/2)

- The VA space is modeled by a VA GMM
- Each latent feature class (a component in the acoustic GMM) can generate a component VA Gaussian

A Mixture of Gaussians in the VA Space

Audio Signal of Each Clip

Acoustic GMM
Learning the VA GMM by the EM Algorithm (Maximum Likelihood)

Annotation-Level

$$p(e_{ij} \mid u_{ij}, s_i, \theta_i) = \sum_{k=1}^{K} \theta_{ik} \mathcal{N}(e_{ij} \mid \mu_k, \Sigma_k)$$

Corpus-Level

$$p(E \mid X) = \sum_i p(E_i, s_i \mid X) = \sum_i p(s_i \mid X)p(E_i \mid s_i, X)$$
$$= \sum_i p(s_i \mid X) \sum_j p(e_{ij}, u_{ij} \mid s_i, X)$$
$$= \sum_i p(s_i \mid X) \sum_j p(u_{ij} \mid s_i, X)p(e_{ij} \mid u_{ij}, s_i, X)$$
$$= \sum_i \sum_j p(u_{ij} \mid s_i, X) \sum_k \theta_{ik} \mathcal{N}(e_{ij} \mid \mu_k, \Sigma_k).$$

$p(u_{ij}, s_i \mid X)$ is the annotation prior for each annotator $\mu_{ij}$ of $s_i$.

Maximize the objective function w.r.t. $\mu_k, \Sigma_k$ by the EM algorithm.
The Learning of VA GMM on MER60
Music Emotion Recognition (MER)

- Given the acoustic GMM posterior of a test song, predict the emotion as a single VA Gaussian

\[ p(e | \hat{\theta}) = \sum_{k=1}^{K} \hat{\theta}_k \mathcal{N}(e | \mu_k, \Sigma_k) \]

\[ \mathcal{N}(e | \hat{\mu}, \hat{\Sigma}) \]
Finding the Representative Gaussian

- The representative Gaussian has the minimal cumulative distance from all the component VA Gaussians

\[ N(\mathbf{e}|\hat{\mu}, \hat{\Sigma}) = \arg\min_{\{\mu, \Sigma\}} \sum_k \hat{\theta}_k D_{KL}(N(\mathbf{e}|\mu_k, \Sigma_k) \| N(\mathbf{e}|\mu, \Sigma)) \]

- The optimal parameters of the Gaussian are

\[ \hat{\mu} = \sum_{k=1}^{K} \hat{\theta}_k \mu_k, \]

\[ \hat{\Sigma} = \sum_{k=1}^{K} \hat{\theta}_k \left( \Sigma_k + (\mu_k - \hat{\mu})(\mu_k - \hat{\mu})^T \right) \]
## Emotion-Based Music Retrieval

The process of emotion-based music retrieval is illustrated in the diagram above. The main phases are:

### Indexing Phase
- **Feature Extraction**: Extracting features from the music clip.
- **Feature Indexing**: Indexing the extracted features.

### Retrieval Phase
- **Emotion Query**: Querying the music database with an emotion query.
- **Query Processing**: Processing the query against the music database.

### Table: Approach, Indexing, and Matching

<table>
<thead>
<tr>
<th>Approach</th>
<th>Indexing</th>
<th>Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold-In</td>
<td>Acoustic GMM Posterior</td>
<td>Cosine Sim ((K\text{-dim}))</td>
</tr>
<tr>
<td>Emotion Prediction</td>
<td>Predicted VA Gaussian</td>
<td>Gaussian Likelihood</td>
</tr>
</tbody>
</table>
The Fold-In Approach

The query is dominated by the VA Gaussian of $A_2$

$\hat{\lambda} = \arg\max_{\lambda} \log \sum_k \lambda_k \mathcal{N}(\hat{e} | \mu_k, \Sigma_k)$

Using the EM algorithm
Personalization Using MAP Adaptation

The personal annotation can be applied to clips exclusive to the background training set.
Sec-by-Sec Music Emotion Tracking

Top: ground truth

Bottom: predicted
Outline

- **Social Tagging-based MIR**
  - Automatic music tagging, Tag-based music retrieval, Query interfaces, Real-time music tagging, Music player visualization

- **Emotion-based MIR**
  - Music emotion recognition (MER), Emotion-based music retrieval, Query interfaces

- **Relationship between Tags and Dimensional Emotions of Music**

- **Automatic Generation of Music Video**
Relationship between Mood Tags and VA Dimensional Emotion

- Can the mood tags be mapped into the VA space?
Acoustic Tag Bernoullis (ATB)

- Given a tagged music dataset
- Learn ATB that describes the generative process of each song in the dataset from acoustic features to a tag

\[ p(\text{Angry} \mid s) = \sum_k \theta_k \beta_k \]

A Mixture of Bernoullis
Aligned with Acoustic GMM

Acoustic GMM Posterior
Acoustic Emotion Gaussians (AEG)

- Given a VA-annotated music dataset
- Learn AEG that describes the generative process of each song in the dataset from acoustic features to the VA space

![Diagram of AEG Model]

- \( \theta_1 \)
- \( \theta_2 \)
- \( \theta_{K-1} \)
- \( \theta_K \)

Acoustic GMM Posterior

VA GMM Aligned with the Acoustic GMM

Representative VA Gaussian

Arousal

Valence
Mapping a Tag into the VA Space
MIR Interface Integration

- The integration of tag colorizing-based and Mr. Emo interfaces
Outline

- **Social Tagging-based MIR**
  - Automatic music tagging, Tag-based music retrieval, Query interfaces, Real-time music tagging, Music player visualization

- **Emotion-based MIR**
  - Music emotion recognition (MER), Emotion-based music retrieval, Query interfaces

- **Relationship between Tags and Dimensional Emotions of Music**

- **Automatic Generation of Music Video**
Automatic Generation of Music Video

- Google challenge (one of the six Multimedia Grand Challenges at ACM MM 2012)
  - Recommending appropriate music to a user video
  - Recommending suitable video clips to accompany a soundtrack

- Add significant entertainment value to user videos
- Make video sharing a lot more fun for the users

- We extended the Acoustic Emotion Gaussians (AEG) model to an Audiovisual Emotion Gaussians (AVEG) model for the challenge, and won the FIRST PRIZE
  - Utilize the perceived emotion of multimedia content as a bridge to connect music and videos
The Audiovisual Emotion Gaussians Model for Automatic Generation of Music Video

- Construct a novel Audiovisual Emotion Gaussians (AVEG) framework for learning the tripartite relationship among music, video, and emotion.

- Represent music clips and video sequences as stochastic distributions in the 3-D emotion space (3DES) and perform cross-modality matching in the 3DES.
System Diagram of AVEG

Audio
- Sound energy
- Tempo and beat strength
- Rhythm regularity
- Pitch

Video
- Lighting key
- Shot change rate
- Motion intensity
- Color (saturation, color energy)
Demonstration 1

- Video to Audio Retrieval
  - Negative-valence, low-arousal, low-potency

Predicted video emotion

Predicted music emotion
Demonstration 2

Audio to Video Retrieval

- Positive-valence, mid-arousal, mid-potency

Predicted music emotion

Predicted video emotion
Acknowledgements

Special thanks to

Dr. Ju-Chiang Wang
Institute of Information Science
Academia Sinica

Dr. Yi-Hsuan Yang
Research Center for Information Technology Innovation
Academia Sinica
References


More papers are available at http://slam.iis.sinica.edu.tw/paper.htm
Thank You !