Music Information Retrieval and Music Emotion Recognition

Yi-Hsuan Yang Ph.D.
http://www.citi.sinica.edu.tw/pages/yang/
yang@citi.sinica.edu.tw

Music & Audio Computing Lab,
Research Center for IT Innovation,
Academia Sinica
About Me & CITI, AS

• Yi-Hsuan Yang, Ph.D., Assistant Research Fellow
  ➢ Education
  Ph.D. in GICE, National Taiwan University, 2006-2010
  B.S. in EE, National Taiwan University, 2002-2006
  ➢ Research Interests
  Music information retrieval, multimedia applications, and machine learning

• Research Center for IT Innovation, Academia Sinica
  ➢ Music and Audio Computing Lab
    - Since 2011/09
    - Research assistants
    - PhD students
    - Postdocs
  ➢ Industrial collaborations: KKBOX, HTC, iKala
Outline

• What is and why music information retrieval?
• Current projects
• Example project: music and emotion
Digital Music Industry
Proliferation of Mobile Devices

- 1.5 billion handsets were sold in 2011
- 1/3 of them are smart phones
- 6 billion mobile-cellular subscriptions

Mobile behavior related to multimedia

- Took photos
- Played games
- Recorded video
- Social networking
- Listened to music
- Watched video

#Statistics from ITU
Music Information Retrieval

• User need: find the “right” song
  ➢ For a specific listening context (in a car, before sleep)
  ➢ For a specific mood (feeling down, in an anger)
  ➢ For a specific event (wedding, party)
  ➢ For accompanying a video (home video, movie)

• Current solution
  ➢ Manual
  ➢ Keyword search
  ➢ Social recommendation
“Smart” Content-Based Retrieval

Music audio

↓

Music content analysis (e.g., similarity estimation)

↓

Content-based retrieval

Recommendation

- Can’t Get It Out of My Head
  Greatest Hits - Electric Light Orchestra

- The Things We Do For Love
  Greatest Hits 1972-78 - 10cc

- I Wouldn’t Want To Be Like You
  I Robot - Alan Parsons Project

- Rikki Don’t Lose That Number
  A Decade Of Steely Dan - Steely Dan

- Take the Long Way Home
  The Very Best of Supertramp - Supertramp

- Nights on Broadway
  Their Greatest Hits - The Record (Disk 1) - Be...

- Your Mama Don’t Dance
  The Best Of Friends - Loggins & Messina

Query by humming

Search for songs by humming or tapping.
Demos

Pop Danthology 2012 – Mashup of 50+ Pop Songs
Scope of MIR

• **Music signal analysis**
  - Timbre, rhythm, pitch, harmony, tonality
  - Melody transcription, audio-to-score alignment
  - Source separation

• **Content-based music retrieval**
  - Metadata-based
    - Genre, style, and mood analysis
  - Audio-based
    - Query by example / singing / humming / tapping
    - Fingerprinting and digital rights management
    - Recommendation, personalized playlist generation
    - Summarization, structure analysis
Scope of MIR (Cont’)

• By nature inter-disciplinary
Current Projects 1/4: Music Emotion

- Music retrieval and organization by “emotion”
  - Music is created to convey and modulate emotions
  - The most important functions of music are social and psychological (Huron, 2000)
Current Projects 2/4: Listening Context

On-device music feature extraction

Mobile phone sensing

Accelerometer
Ambient light
Compass
Dual cameras
GPS
Gyroscope
Microphone
Proximity
Running apps
Time
Wifi

Getting up
Eating breakfast
Getting dressed
Current Projects 3/4: Singing Voice Separation

• Useful for modeling singing voice timbre, instrument identification and melody transcription
Current Projects 4/4: Musical Timbre

Category Level: brass / woodwind / string / percussion / ...

Instrument Level: violin / viola / cello / trumpet / trombone / ...

Expression Level: vibrato / sordino / pizzicato / sul ponticello ...

Research Center for Information Technology Innovation, Academia Sinica
Focus: Emotion-based Recognition & Retrieval

Activation–Arousal ○ Energy or neurophysiological stimulation level

Evaluation–Valence ○ Pleasantness ○ Positive and negative affective states
Music Retrieval in the Emotion Space

- Automatic computation of music emotion
  - No need of human labeling
  - Scalable
  - Easy to personalize/update

- Emotion-based music retrieval / recommendation
  - Content-based
  - Intuitive
  - Fun
Learning to Predict Music Emotion

- Learn the *mapping* between ground truth and feature using *pattern recognition* algorithms
Audio Feature Analysis

Perceptual features audio:
- time signature / tempo
- key/ mode
- timbre
- pitch
- loudness
- structure
Short-Time Fourier Transform and Spectrogram

- Time domain: energy, rhythm
- Frequency domain: pitch, harmonics, timbre
Timbre

- The perceptual feature that makes two sounds with same **pitch** and **loudness** sound **different**
  - Temporal attack-delay
  - Spectral shape

(a) **Flute**

(b) **Clarinet**
Spectral Timbre Features

- Widely used in all kinds of MIR tasks
- Spectral centroid (brightness)
- Spectral rolloff
  - The freq. which 85% of spectral power is concentrated
- Spectral flux
  - Amount of frame-to-frame spectral amplitude difference (local change)
- Spectral flatness
  - Whether the spectral power is concentrated
- Mel-frequency cepstral coefficient (MFCC)
- Vibrato
Extension 1: Time-varying Prediction

Application to Video content understanding
Extension 2: Affect-Based MV Composition

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

- **Video**
  - Lighting key
  - Shot change rate
  - Motion Intensity
  - Color (saturation, color energy)
Demos

- Music → video
- Video → music

- ACM MM 2012 Multimedia Grand Challenge First Prize
Extension 3: User Mood & Music Emotion

Hey friends, it's been awhile since I last updated but my weekend was really good at the retreat and then Monday was a blaze...sooo tired and yesterday I went and bought my Halloween Costume with Jelinek yay haha im going to be a fairy its basically pink!!! my favorite color..im happy haha..gotta be ready for our Halloween Dance put on by yearbook...yeah u better be there ALL OF U!!! haha well cyas later

Current Mood: 🌟 happy
Current Music: Something Corporate- Down

Well my internet is still not working...soo its been awhile since i updated..im sry but nothing hasnt been life while to write about..but now there is...looks like some things dont always go well and some people know how to make ppl feel like crap..hmm..way to go asshole

Current Mood: 🙁 angry
Current Music: goo goo dolls- black balloon

• In addition to blog writing, users
  ➢ enter an emotion tag (user mood)
  ➢ enter a song title & artist name (music emotion)
Mood-Congruent or Mood-Incongruent
Emotion-Based Music Recommendation

Training data (multimedia signal)

- Feature extraction
- Manual annotation

Emotion value

Model training

Test data

- Feature extraction

Emotion value

Automatic Prediction

Emotion-based recommendation

Personalization

User feedback

Human affect/activity detection (e.g., facial expression, speech intonation)

- Melody
- Timbre
- Dynamics
- Rhythm
- Lyrics

Research Center for Information Technology Innovation, Academia Sinica
Wrap-Up

• Introduction of the field ‘Music information retrieval’
  ➢ Music signal analysis
  ➢ Query by example (humming, similarity)
  ➢ Query by text (genre, emotion)

• Current projects at our lab
  ➢ Context & listening behavior
  ➢ Source separation
  ➢ Modeling musical timbre
  ➢ Music and emotion
    ➢ 2-D visualization
    ➢ Time-varying prediction
    ➢ Emotion-based music video composition
    ➢ Music emotion and user mood; emotion-based recommendation
Int. Society for Society for Music Information Retrieval (ISMIR)

• **General chairs**: Jyh-Shing Roger Jang (NTU) et al.
• **Program chairs**: Yi-Hsuan Yang (Academia Sinica) et al.
• **Music chairs**: Jeff Huang (Kainan University) et al.
• **Call for Music**: ISMIR/WOCMAT 2014 Main Theme – “Oriental Thinking” (Due: June 1, 2014)
MIREX (MIR Evaluation eXchange)

- Audio Classification (Train/Test) Tasks
- Audio K-POP Genre Classification
- Audio K-POP Mood Classification
- Audio Tag Classification
- Audio Music Similarity and Retrieval
- Symbolic Melodic Similarity
- Structural Segmentation
- Audio Tempo Estimation
- Audio Onset Detection
- Audio Beat Tracking
- Audio Key Detection

- Multiple Fundamental Frequency Estimation & Tracking
- Real-time Audio to Score Alignment (a.k.a Score Following)
- Audio Cover Song Identification
- Discovery of Repeated Themes & Sections
- Audio Melody Extraction
- Query by Singing/Humming
- Query by Tapping
- Audio Chord Estimation