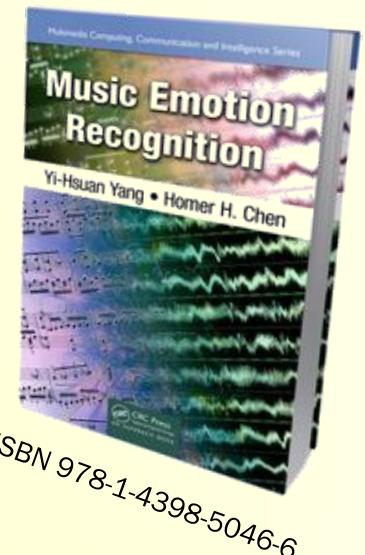


# Music Emotion Recognition

---

Homer H. Chen  
National Taiwan University

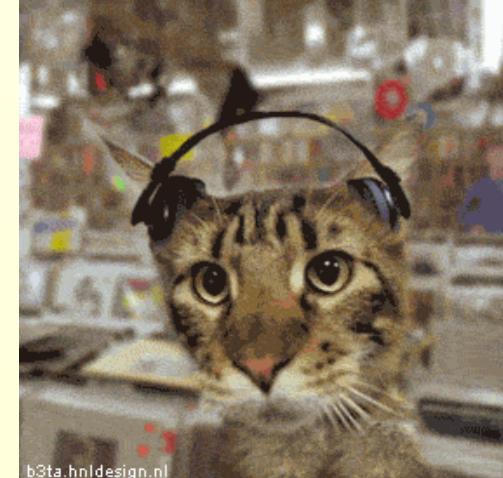
[homer@cc.ee.ntu.edu.tw](mailto:homer@cc.ee.ntu.edu.tw)



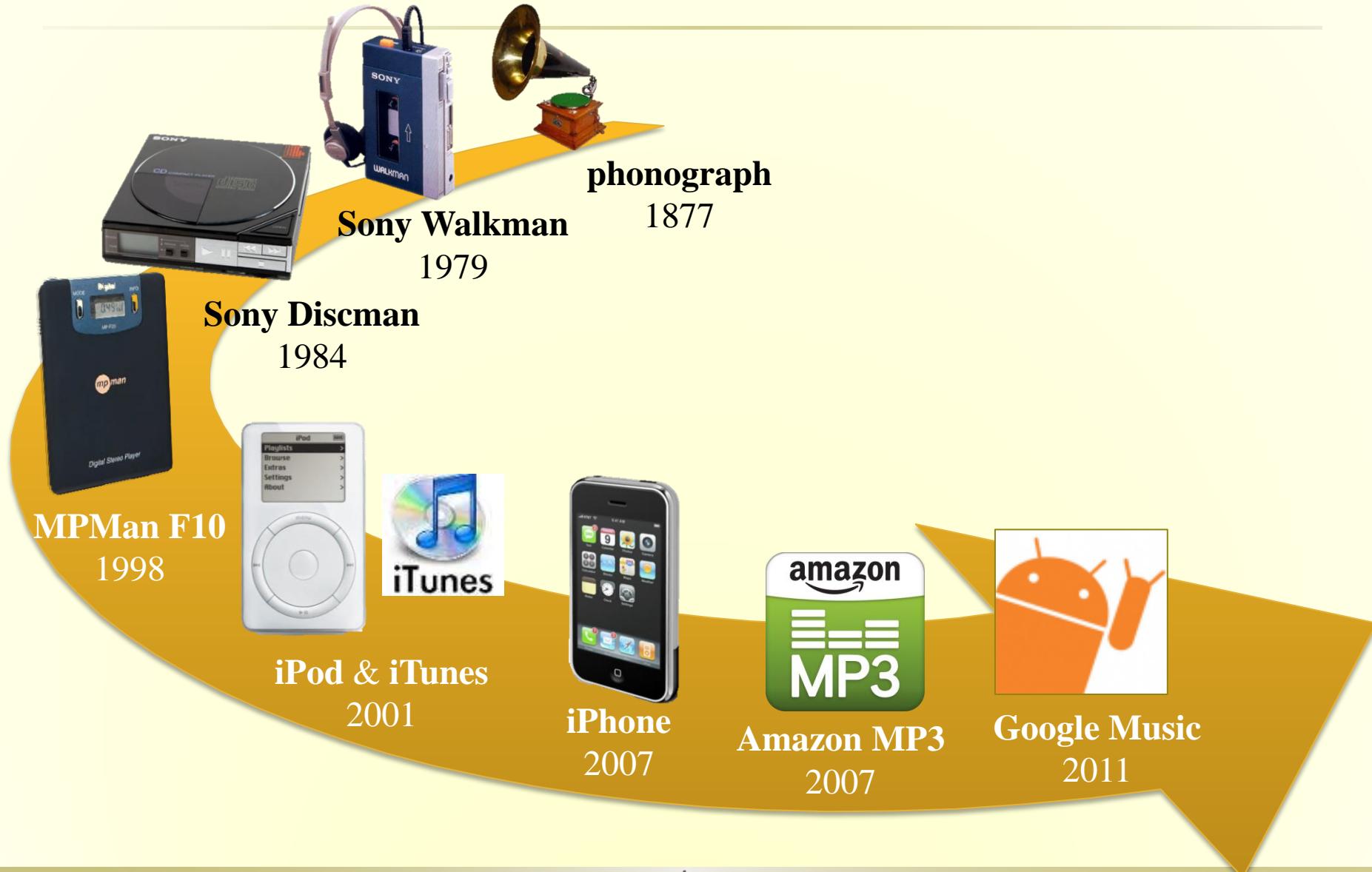
# Music Listening

---

Anytime, anywhere, anyone ...

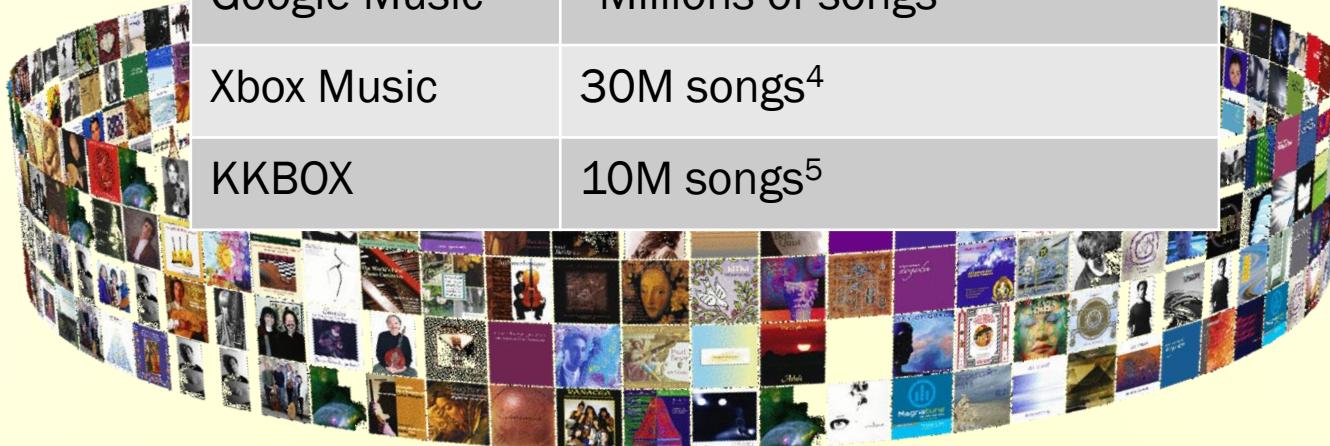


# Evolution of Music Playback



# Explosive Growing of Digital Music

Music provider	Statistics
iTunes store	26M songs <sup>1</sup> 25B downloads (Feb. 2013)
Amazon MP3	25M songs <sup>2</sup>
Google Music	“Millions of songs” <sup>3</sup>
Xbox Music	30M songs <sup>4</sup>
KKBOX	10M songs <sup>5</sup>



<sup>1</sup>Retrieved Sep. 2012, <http://www.apple.com/pr/library/2012/09/12Apple-Unveils-New-iTunes.html>

<sup>2</sup>Retrieved Jun. 2013, <http://www.amazon.com/MP3-Music-Download/b?ie=UTF8&node=163856011>

Source :

<sup>3</sup><https://play.google.com/about/music/>

<sup>4</sup>Retrieved Jun. 2013, <http://www.microsoft.com/en-us/news/Press/2012/Oct12/10-14XboxMusicPR.aspx>

<sup>5</sup>Retrieved Jun. 2013, <http://www.ithome.com.tw/itadm/article.php?c=80653&s=1> <http://tw.kkbox.com>

<http://www.apple.com>

# Music & Emotion

- Music expresses emotions

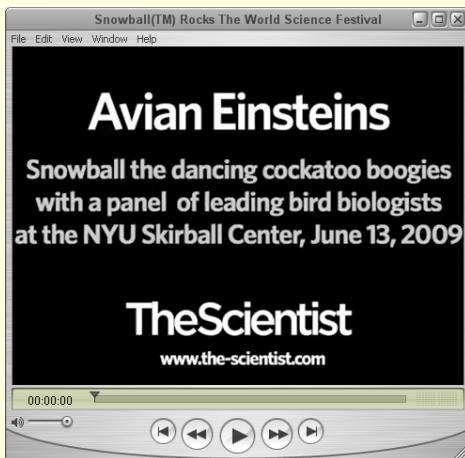


# Music & Emotion

- Music induces emotions



# Power of Music



# Music Emotion Description

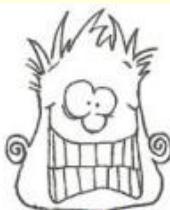
- **Categorical Approach:** Divide music into groups and describe each group by an adjective



Exhausted



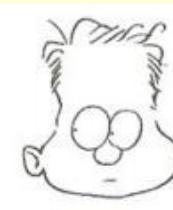
Confused



Ecstatic



Guilty



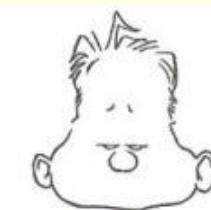
Suspicious



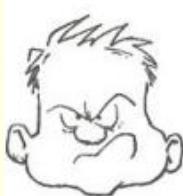
Enraged



Ashamed



Cautious



Angry



Hysterical



Frustrated



Sad



Satisfied



Overwhelmed



Cheerful



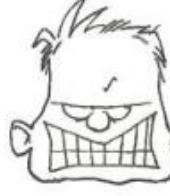
Lonely



Embarrassed



Happy



Mischiefous



Disgusted



Frightened



Boring



Surprised



Anxious

# Traditional Music Information Retrieval (MIR)

- Text based
  - Artist name
  - Song name
  - Album title
  - Lyrics
  - Genre

- KKBOX



- Last.fm

在 Last.fm 上寻找音乐

搜索音乐库 :

- Pandora

Create a New Station

PANDORA®

Enter an artist, track or composer to create a station

# All Music Guide

## Mariah Carey



Photo by Markus Klinko & Indrani

**Picture Browser**  
[< Previous](#) [Next >](#)

**Born**  
Mar 27, 1970 in Huntington, NY

**Years Active**  
1910 20 30 40 50 60 70 80 90  
2000

**Genres** **Styles**  
**R&B** · Dance-Pop  
· Adult Contemporary  
· Pop  
· Adult Contemporary  
R&B

**Biography** by Jason Ankeny

The best-selling female performer of the 1990s, Mariah Carey rose to superstardom on the strength of her stunning five-octave voice. An elastic talent who moved easily from glossy ballads to hip-hop-inspired dance-pop, she earned frequent comparison to rivals Whitney Houston and Celine Dion, but did them both one better by composing all of her own material. Born in Long Island, NY, on March 27, 1970, Carey moved to New York City at the age of 17 -- just one day after graduating high school -- to pursue a music career; there she befriended keyboardist Ben Margulies, with whom she began writing songs. Her big break came as a backing vocalist on a studio session with ... [» Read more](#)



**Moods**

- Confident
- Party/Celebratory
- Sensual
- Sexy
- Carefree
- Exuberant
- Playful
- Refined/Mannered
- Stylish
- Amiable/Good-Natured
- Bright
- Energetic
- Fun
- Laid-Back/Mellow
- Reflective
- Romantic
- Sentimental
- Slick
- Sparkling
- Sweet

**Instruments**

- Vocals

**Other Entries**

- Classical Music Entry
- Movie Entry

**AMG Artist ID**  
P 62404

Watch music videos by this artist!

<http://www.allmusic.com>

17

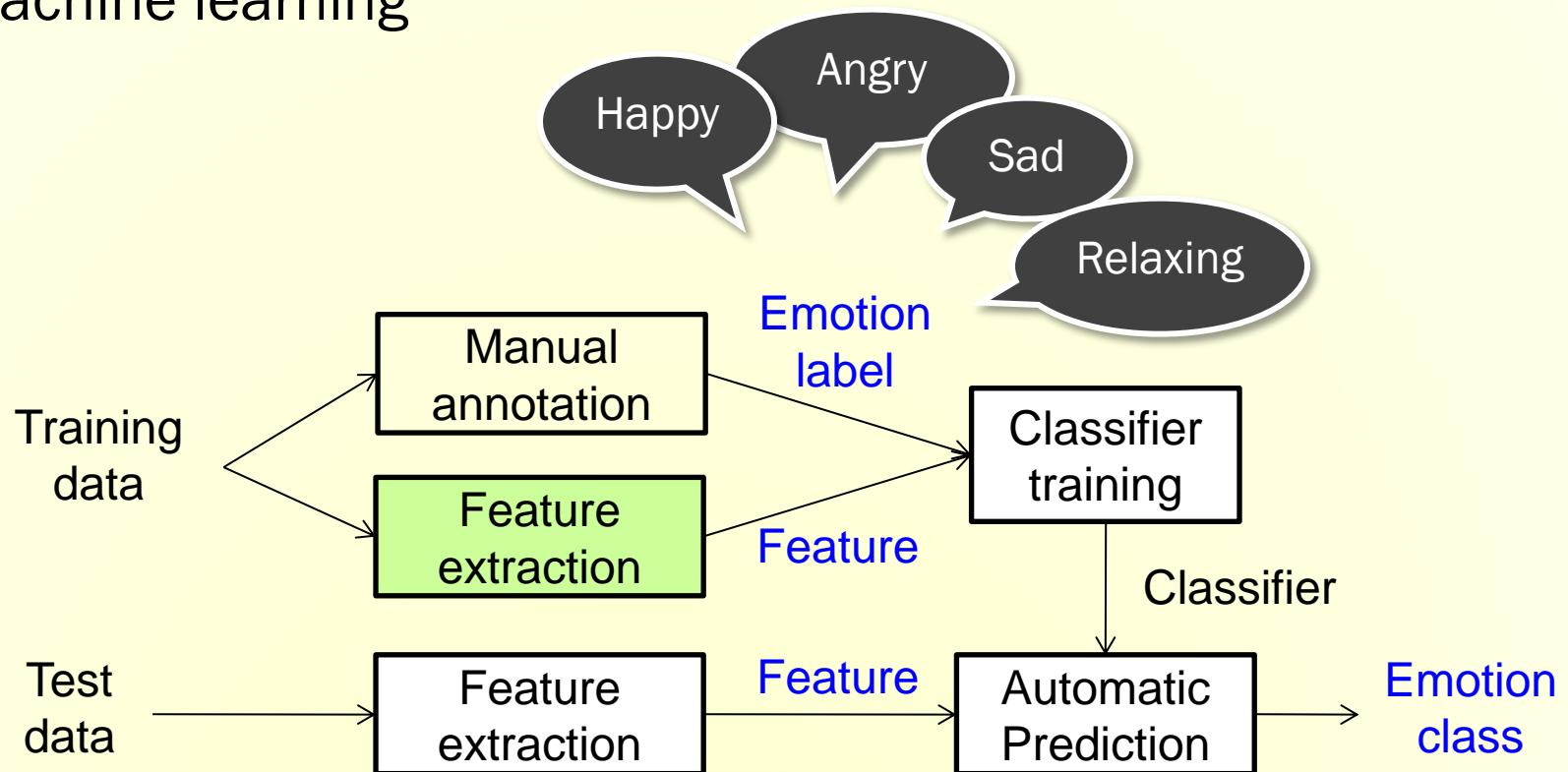
# Gracenote Media Manager

The screenshot shows the Gracenote Media Manager interface. On the left is a navigation tree under 'Audio TreeView' with categories like 'Track Artists', 'Albums', 'Genre', etc. The main area displays a search results grid with columns: Name, Track Artist, Album, Title, Mood, and Tempo. A red box highlights the 'Mood' column. The results include tracks from artists like The Association, Astrud Gilberto, and B.B. King.

Name	Track Artist	Album	Title	Mood	Tempo
(All)				Idealistic / Stirring	Medium Fast
Track Artists	The Association	The Association's Greatest Hits	Everything That Touches You	Energetic Groovy / Bitters...	Medium Fast
Album Artists	The Association	The Association's Greatest Hits	Never My Love	Tender	Medium
Albums	The Association	The Association's Greatest Hits	Cherish	Intimate / Nostalgic / Bitte...	Medium
Classical	Astor Piazzolla	Sur	Tristeza/Separation	Sweet	Medium
Years	Astrid Hadad & Los Tar...	Ay!	iAy Qué Dolor Vivir!	Sultry / Swank	Medium
Genre	Astrud Gilberto	Verve Jazz Masters 9	The Girl From Ipanema [Live]	Tender	Medium Slow
Origin	Astrud Gilberto	Verve Jazz Masters 9	A Certain Sadness	Sophisticated / Lush / Ro...	Medium Slow
Era	Astrud Gilberto	Verve Jazz Masters 9	The Shadow Of Your Smile	Hard Positive Excitement	Fast
Artist Type	The Ataris	Blue Skies, Broken Hearts...N...	San Dimas High School Foot...	Hard Positive Excitement	Fast
Mood	Audioslave	Audioslave	Cochise	Heavy Brooding	Medium Fast
Peaceful	Audra	The Arbitrary Width Of Shad...	There Are No Snakes In Hea...	Frenetic Euphoric Bliss	Fast
Romantic	Aurora Feat. Naimee C...	Euro Dance Hits	Ordinary World	Dark Groovy / Savvy	Medium Fast
Sentimental	Average White Band	Funkgasm	Pick Up The Pieces	Hard Dark Excitement	Fast
Tender	Avril Lavigne	The Best Damn Thing -Specia...	Girlfriend	Loud Strength & Glory	Medium
Yearning	Avril Lavigne	The Best Damn Thing -Specia...	When You're Gone	Loud Strength & Glory	Medium Fast
Easygoing	Avril Lavigne	The Best Damn Thing	Keep Holding On	Creepy / Eerie / Ominous	Medium
Sensual	Aynur Dogan	Crossing The Bridge: The Sou...	Ehmedo	Energetic Groovy / Bitters...	Fast
Somber	Aztec Camera	High Land, Hard Rain	Walk Out To Winter	Gritty / Earthy / Soulful	Medium Slow
Melancholy	B.B. King	King Of Blues	When My Heart Beats Like A...	Cheerful / Playful	Medium Fast
Blue	B.J. Thomas	The Very Best Of Burt Bachar...	Raindrops Keep Fallin' On My...	Relaxed Groove	Medium Slow
Defiant	B.T. Express	Disco Classics	Do It ('Til You're Satisfied)	Edgy / Sexy	Medium Fast
Cool	B5	Radio Disney: Party Jams	Let's Groove	Heavy Groove	Medium Fast
Fiery	The B-52's	The B-52's	52 Girls	Arrogant / Attitude / Defi...	Medium
Sophisticated	Baby Bash Feat. T-Pain	Cyclone	Cyclone	Energetic Alienation / Mela...	Medium Fast
Urgent	Babyshambles	The Blinding E.P.	The Blinding	Confident / Celebratory	Medium Fast
	Bachman-Turner Overd...	BTO's Greatest	You Ain't Seen Nothing Yet		

# Music Emotion Classification (MEC)

Machine learning



# Subjects Annotation

---

- Ask human subjects to annotate music emotion



- Happy
- Sad
- Angry
- Relaxed

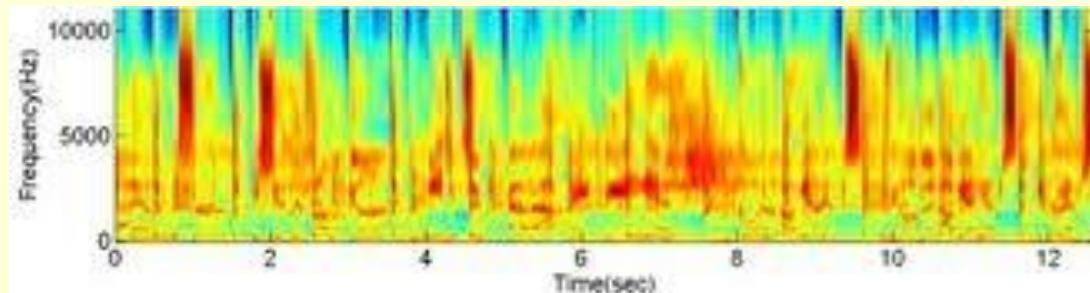


- Happy
- Sad
- Angry
- Relaxed

# Music Features

---

- Spectral
    - Spectral centroid, spectral rolloff, spectral flux, MFCCs
  - Temporal
    - Zero-crossing rate, temporal centroid, log-attack time
  - Melody/harmony
    - Pitch class profile, key clarity, harmonic change, musical mode
  - Rhythmic
    - Beat histogram, average tempo (BPM)
- :



# Spectral Features

- **Spectral centroid**
  - average frequency of the signal weighted by magnitude
- **Spectral roll-off**
  - how much of the frequencies are concentrated below a given threshold
- **Spectral flux**
  - how much the frequency varies over time
- **Spectral flatness**
  - Whether the spectral power is concentrated

$$\sum_{n=1}^{R_t} M_t[n] = 0.5 * \sum_{n=1}^N M_t[n].$$

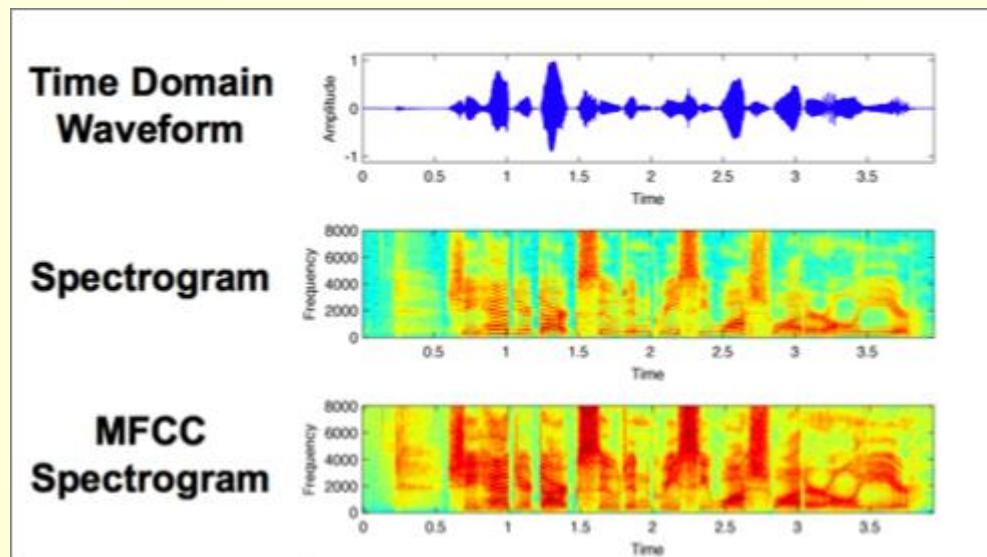
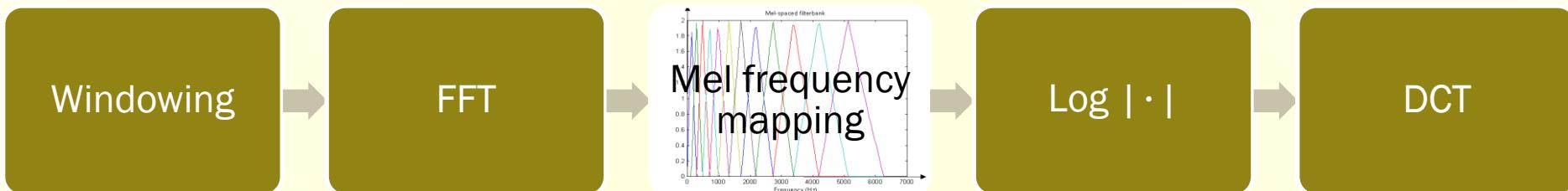
$$\sum_{n=1}^{R_t} M_t[n] = 0.85 * \sum_{n=1}^N M_t[n].$$

$$F_t = \sum_{n=1}^N (N_t[n] - N_{t-1}[n])^2$$

$$\frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\left(\frac{\sum_{n=0}^{N-1} x(n)}{N}\right)}$$

# Spectral Features (Cont'd)

- MFCCs



# Loudness

- Root-mean-square energy (RMS)

- $\sqrt{\text{mean}(A^2)}$

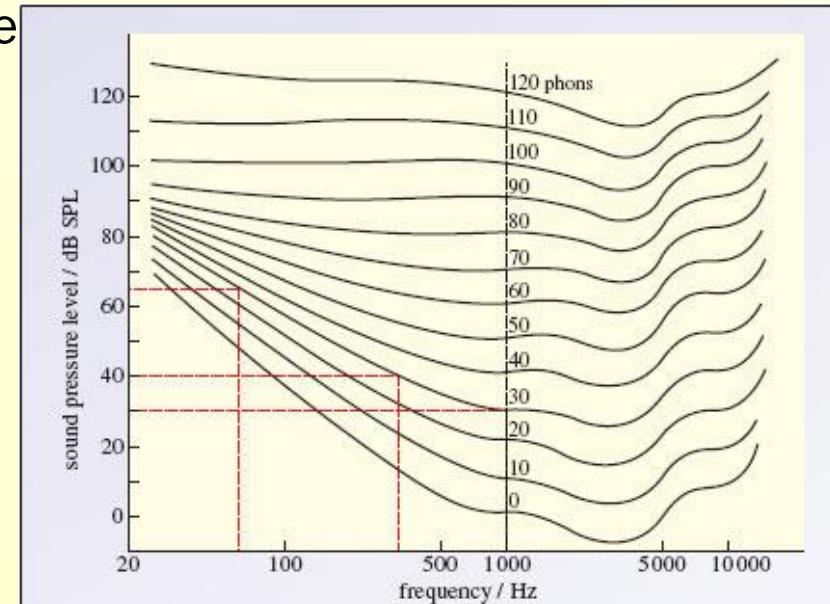
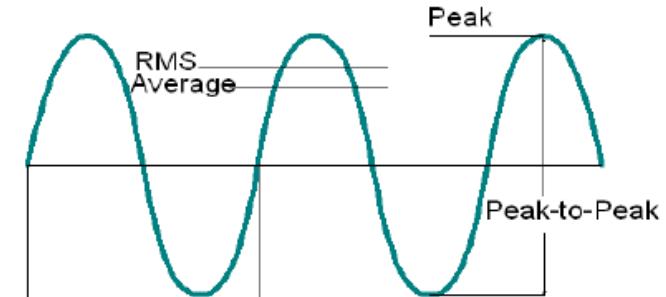
- Classifying exciting/relaxing music

- Low-energy feature

- Percentage of “texture window” (1 sec) with RMS value under average
- Example: Vocal music with silence

- Intensity vs. loudness

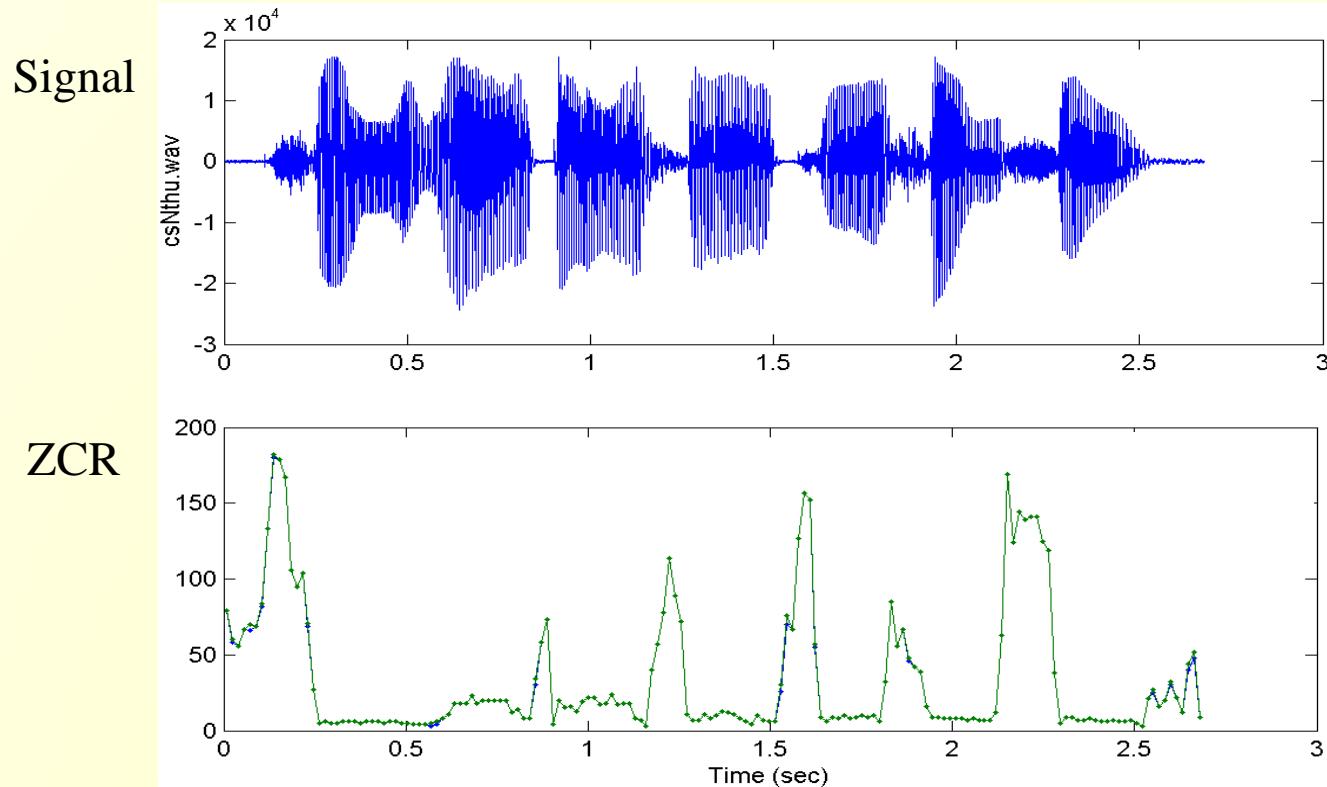
- Physical / psychological
- Sound pressure level (db SPL) / phone



# Zero Crossing Rate

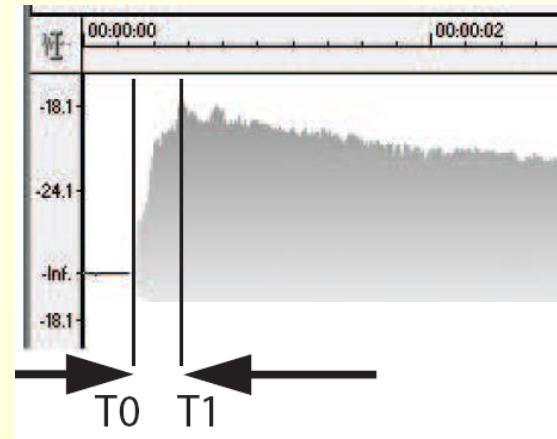
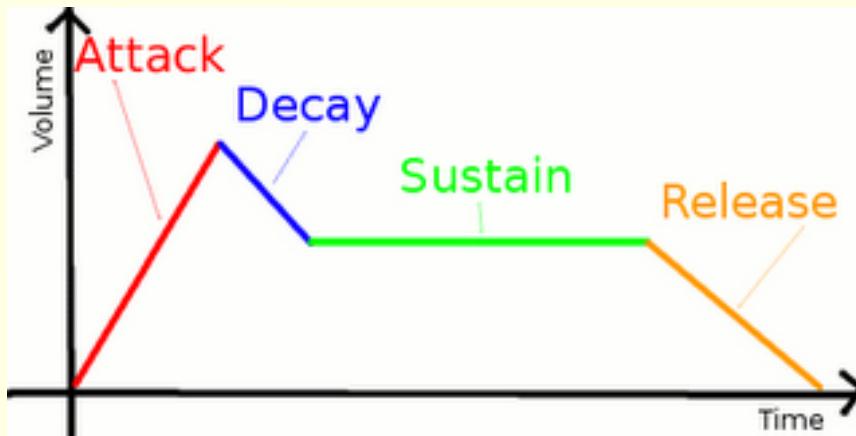
- Number of time domain crossings within a frame

$$Z_t = \frac{1}{2} \sum_{n=1}^N |sign(x[n]) - sign(x[n-1])|$$



# ADSR

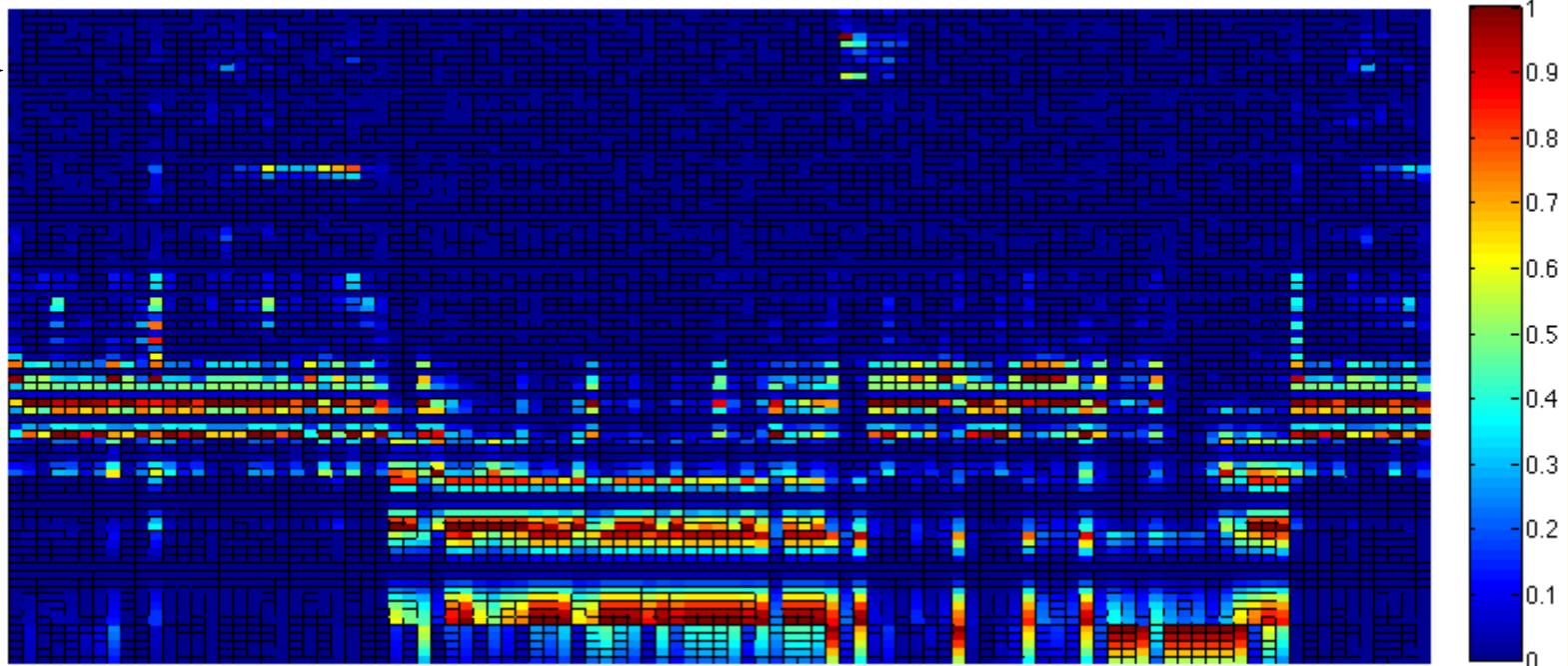
- Attack, Decay, Sustain, Release



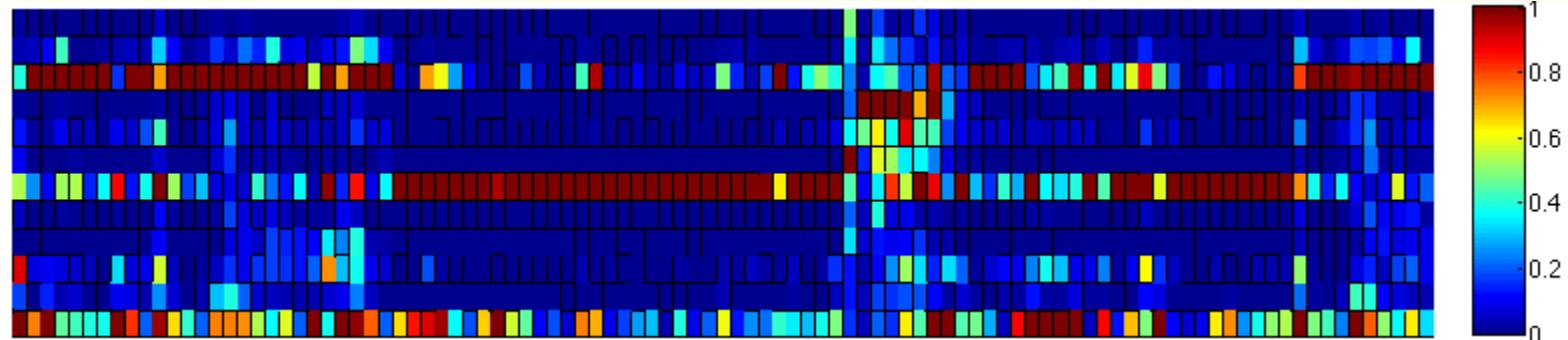
- Log attack time  $LAT = \log_{10}(T1 - T0)$ 
  - the time it takes to reach the maximum amplitude of a signal from a minimum threshold time

# Pitch Class Profile

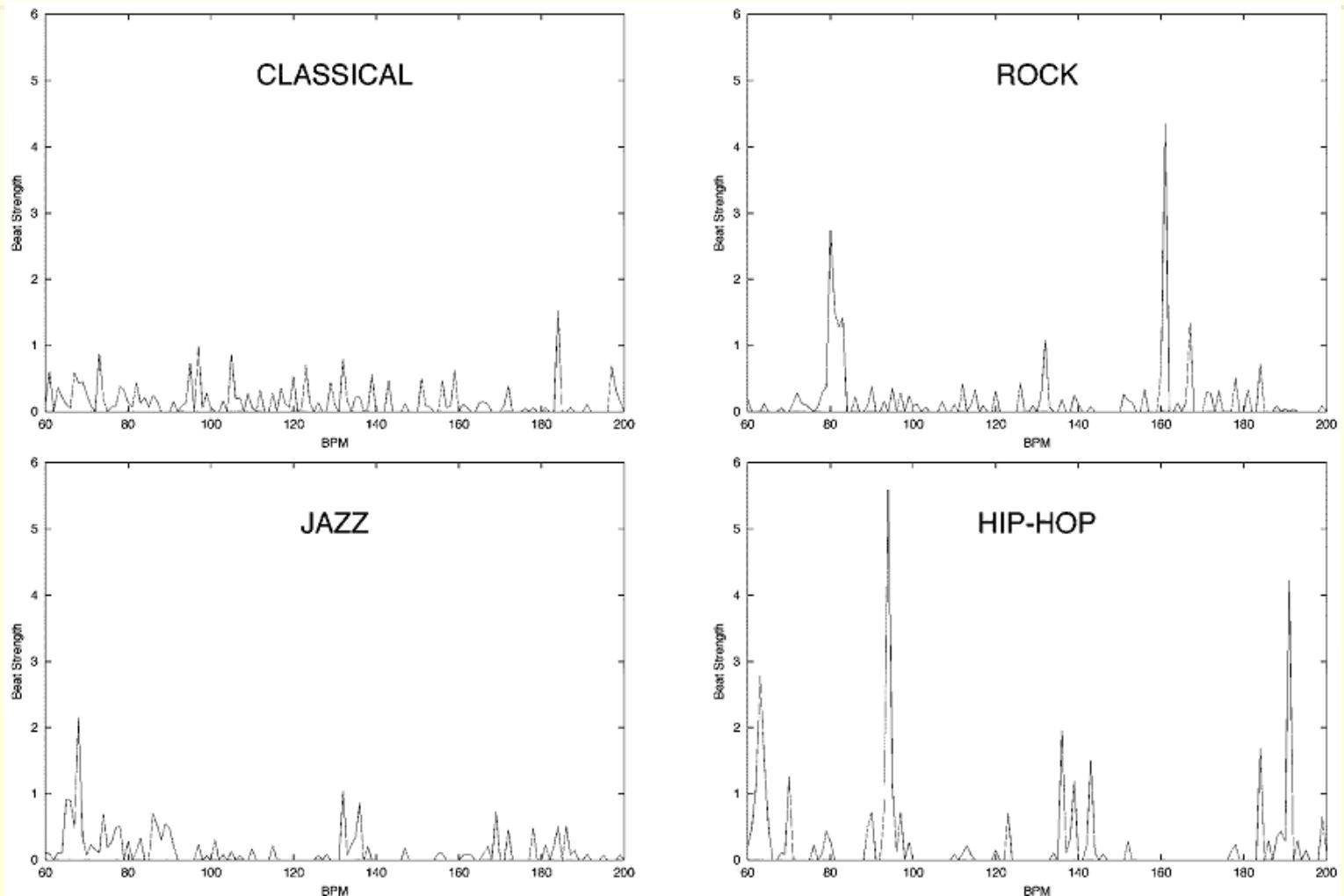
Unfolded



Folded



# Beat Histogram



G. Tzanetakis and P. Cook "Musical Genre Classification of Audio Signals", *IEEE Transactions on Speech and Audio Processing*, 10(5), July 2002

# Problems of Categorical Approach

---

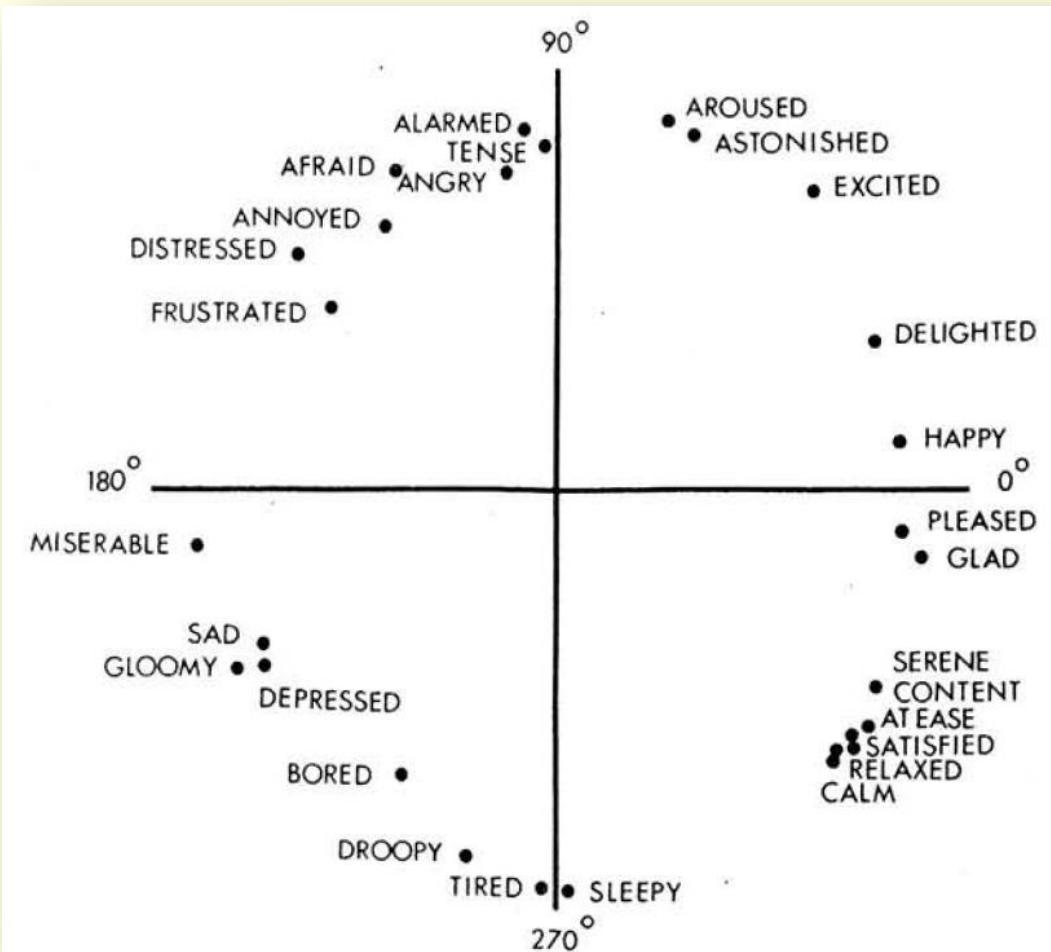
- Ambiguity
  - Happy = joyous = cheerful = delighted?
  - Sad = sorrowful = depressed = gloomy?
- Granularity
  - Small number of emotion classes
    - Insufficient to describe the richness of human perception
  - Large number of emotion classes
    - Difficult to obtain ground truth values

abandoned, abashed, abused, aching, admiring, adoring, adrift, affectionate, affronted, afraid, aggravated, aglow, ailing, alarmed, alienated, alienated, alone, ambivalent, anguished, annoyed, annoyed, antagonistic, anxious, apart, apologetic, appalled, appreciative, apprehensive, ardent, ashamed, attached, attentive, awful, awkward...

# Dimensional Approach

## Arousal

- Activation, activity
- Energy and stimulation level

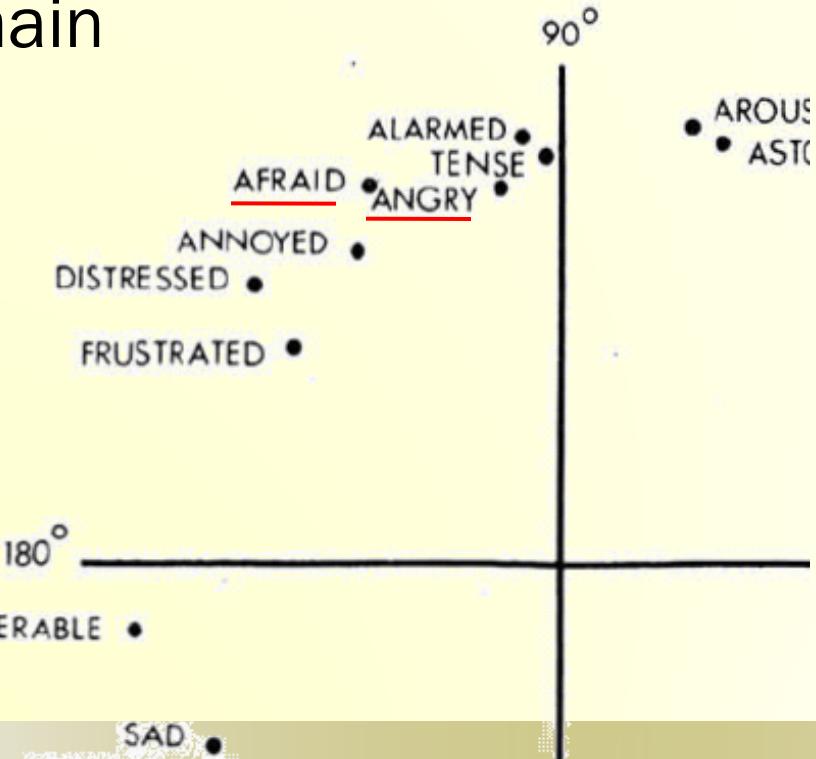


## Valence

- Pleasantness
- Positive and negative affective states

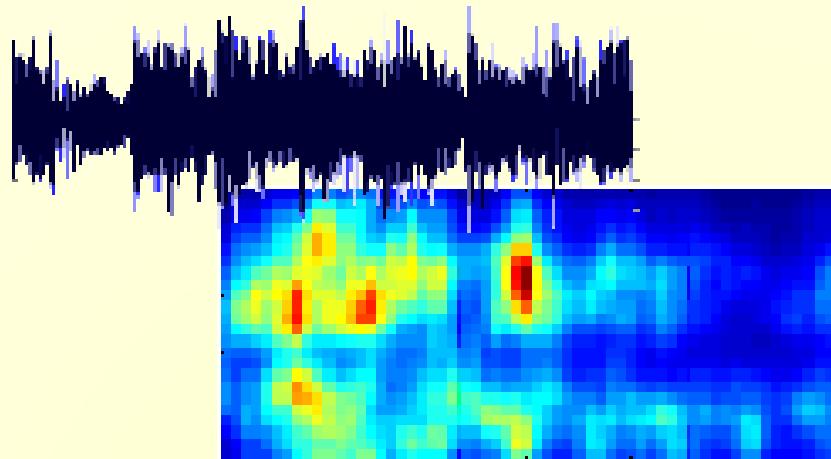
# Dimensional Approach

- Strength
  - No need to consider which and how many emotions
  - Generalize MER from categorical domain to real-valued domain
  - Provide a simple means for 2D user interface
- Weakness
  - Blurs important psychological distinctions
    - Afraid, angry



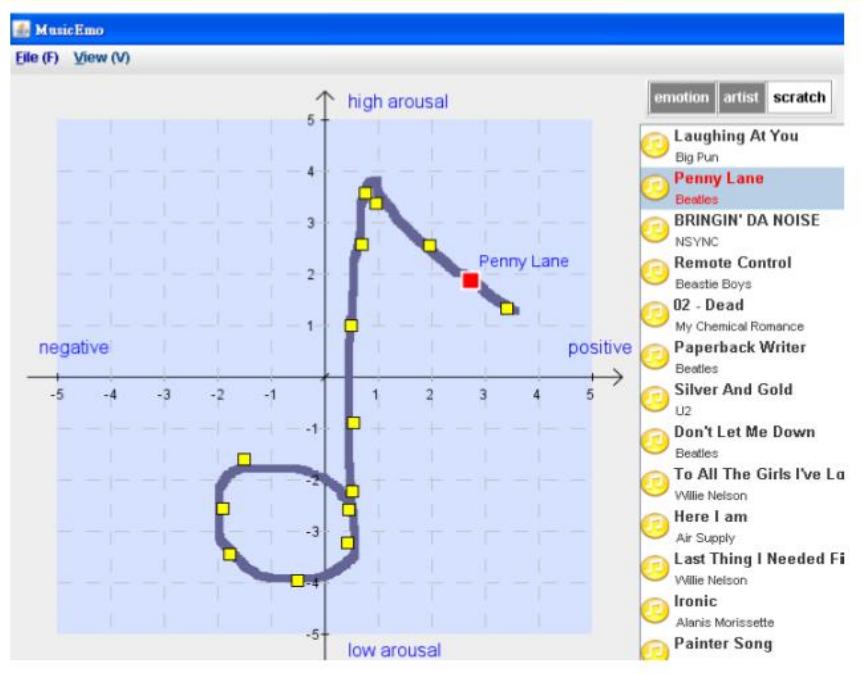
# Arousal and Valence Features

- Arousal
  - Pitch: high/low
  - Tempo: fast/slow
  - Timbre: bright/soft
  - :
- Valence
  - Harmony: consonant/dissonant
  - Mode: major/minor
  - Tonality: tonal/atonal
  - :



# Mr. Emo

- Developed by our lab at NTU
- Each music piece is a point in the emotion space
- A great app for smart phones



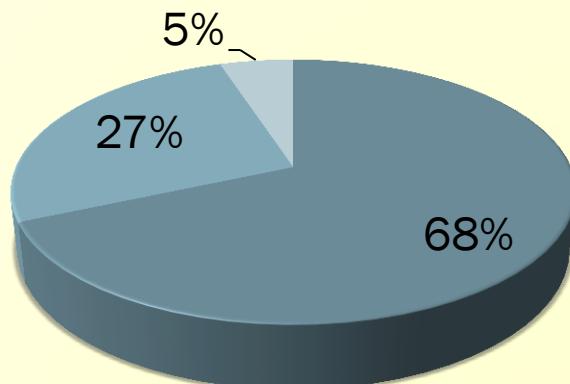
demo.ink

[Live Demo](#)

YouTube  
<http://www.youtube.com/watch?v=ra55xO20UHU>

# Retrieving Music by Emotion

- Complementary to traditional MIR method
- Survey: Would you like to retrieve music by emotion?
  - 615 subjects (mostly college students):



- Strong desire: 421 (68%)
- Moderate interest: 163 (26%)
- No interest: 31 (5%)

# Musicover

**Musicover**

Search an artist

[Log In](#) [Sign Up](#) [U.S.A.](#) ▾

Mood Dance Artist radio New Widget Lab

Favorite Songs Favorite Artists **Hits** Discovery

Energetic

Dark

Positive

Calm

50s 60s 70s **80s** 90s 00s  
1958

Roy Orbison

LEGEND BOB MARLEY

The Sun Years

Fool's Hall of Fame

Roy Orbison

Could you be loved

Bob Marley

Blue moon

Nat King Cole

Desafinado

Eliane Elias

Layla

Eric Clapton

Bitter sweet

The Verve

Tesla girls

O.M.D

Ziggy Marley

Billie Holiday

Stan Getz & Joao

Live 2004

Sugar Daddy

Joe Cocker

Live 1980

Monty Alexander

share my mood with this song [f](#) [t](#)

All Genres

<input checked="" type="checkbox"/> Rock	<input checked="" type="checkbox"/> Reggae	<input checked="" type="checkbox"/> Jazz
<input checked="" type="checkbox"/> Pop	<input checked="" type="checkbox"/> Soul	<input checked="" type="checkbox"/> Blues
<input checked="" type="checkbox"/> Soundtrack	<input checked="" type="checkbox"/> Gospel	<input checked="" type="checkbox"/> Vocal Pop
<input checked="" type="checkbox"/> Latino	<input checked="" type="checkbox"/> Disco	<input checked="" type="checkbox"/> World
<input checked="" type="checkbox"/> Rap	<input checked="" type="checkbox"/> Funk	<input checked="" type="checkbox"/> Classical
<input checked="" type="checkbox"/> R & B	<input checked="" type="checkbox"/> Electro	<input checked="" type="checkbox"/> Metal

<http://musicover.com/>

# Moodagent

- Automatically profiles music based on emotion, mood, genre, style, tempo, beat, vocals, instruments and production features
- Create mood-based playlists by setting the mood-sliders or choosing a seed track
- Need to profile a song on PC if online database does not have the entry



<http://www.moodagent.com/>

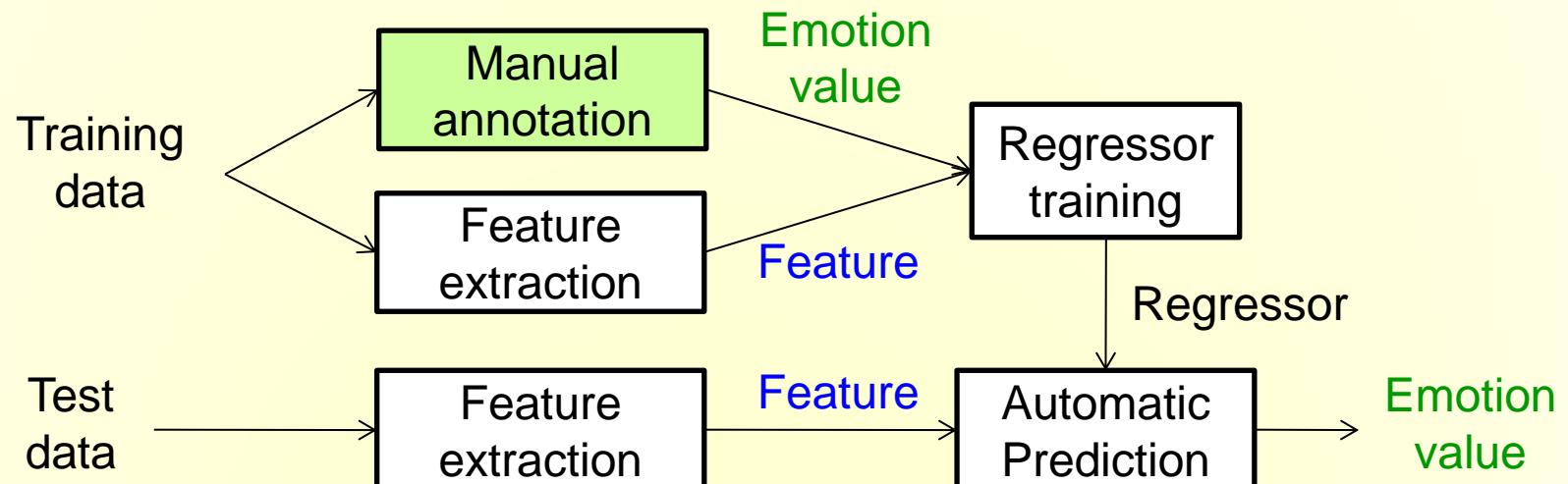
# Obtain Music Emotion Rating

- Subject annotation
  - Rates the VA values of each song

- Ordinal rating scale

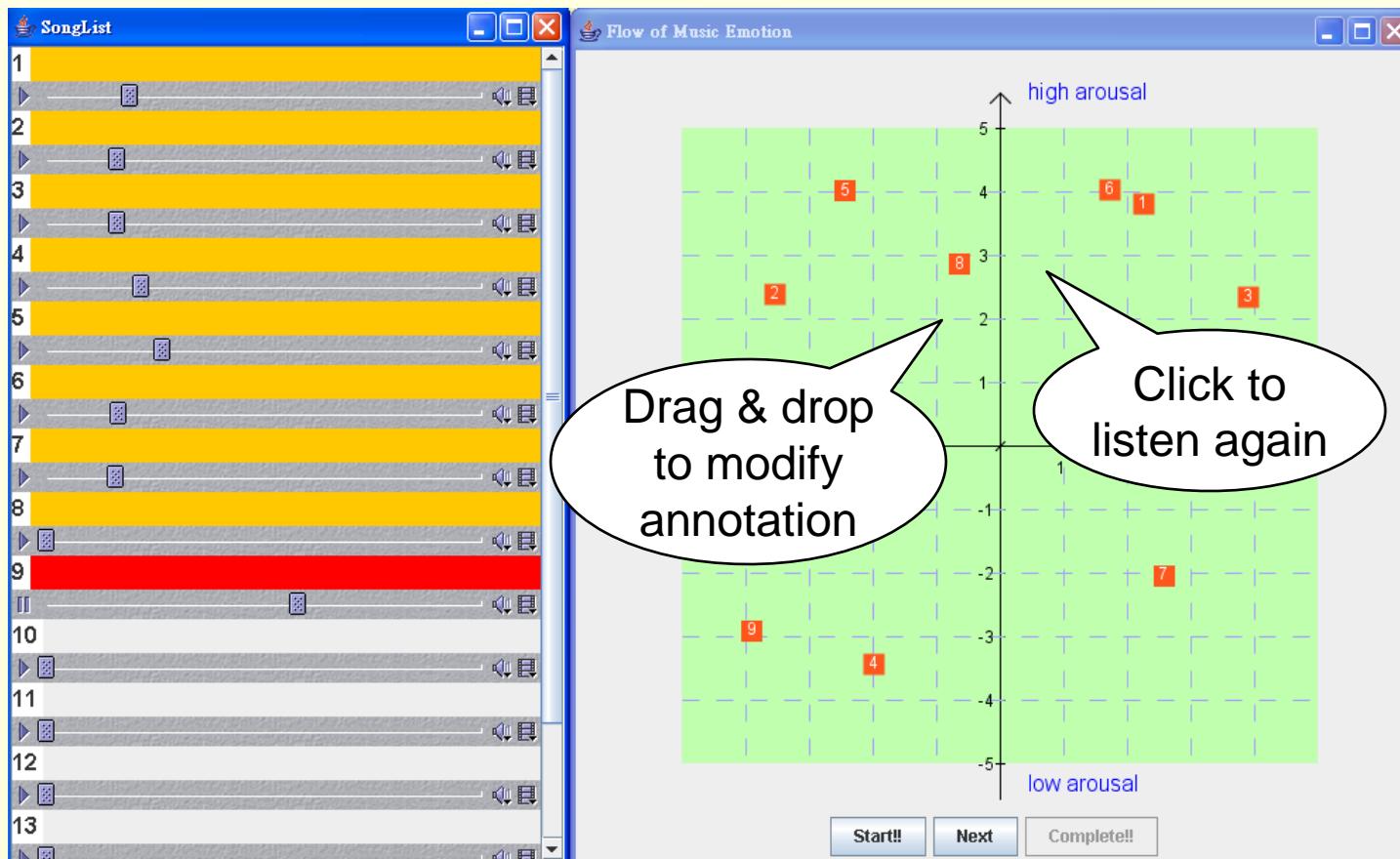


- Scroll bar



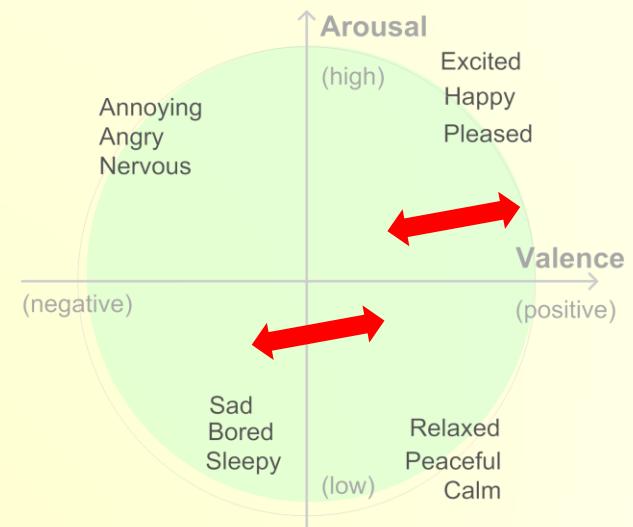
# AnnoEmo: GUI for Emotion Rating

- Easy to differentiate



# Difficulty of Emotion Annotation

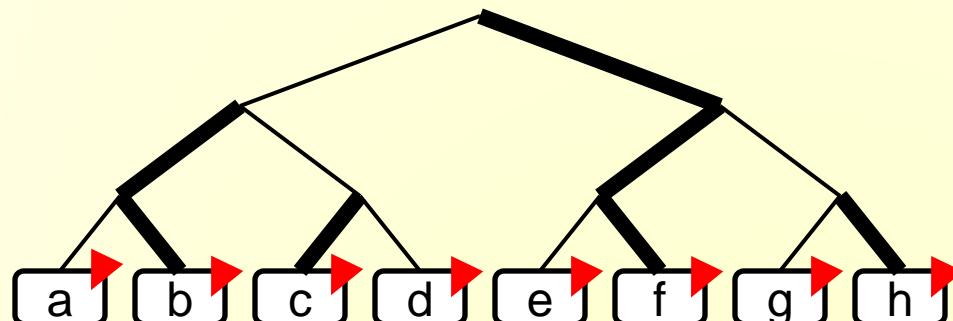
- The cognitive load is high
- Difficult to ensure the rating scale is used consistently
  - Is the distance between 0.6 and 0.8 equals to the distance between -0.1 and 0.1?
  - Does 0.7 mean the same for two subjects?



# Ranking-Based Emotion Annotation

- Emotion tournament

- Requires only  $n-1$  pairwise comparisons
- The global ordering can later be approximated by a greedy algorithm



	a	b	c	d	e	f	g	h
a								
b								
c								
d								
e								
f								
g								
h								

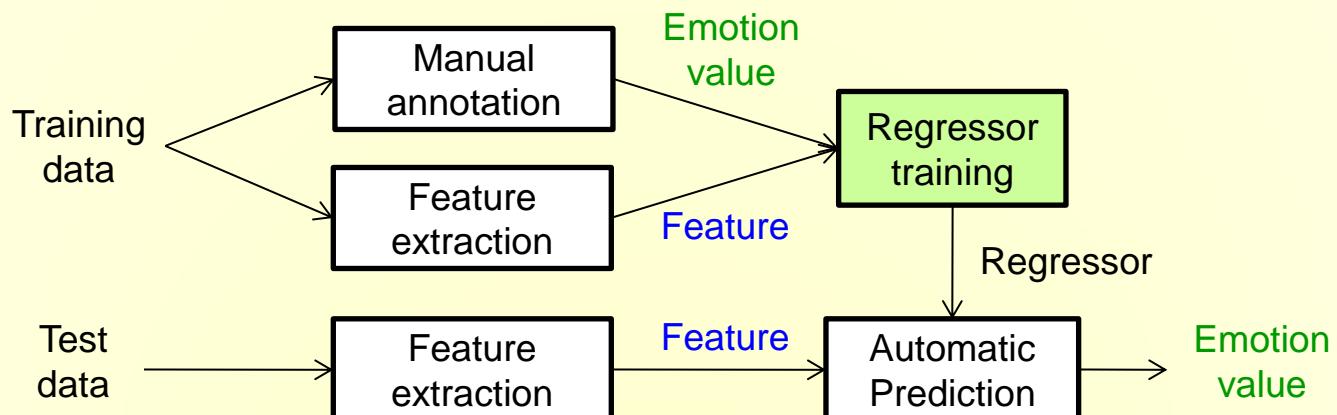
0  
3  
1  
0  
0  
7  
0  
1

Which songs is more positive?

$f > b > c = h > a = d = e = g$

# Regression

- Predict the VA values
    - Trains a regression model (regressor)  $f(\cdot)$  that minimizes the mean squared error (MSE)
    - One for valence; one for arousal
- $y_i$ : numerical emotion value  
▪  $x_i$ : feature (input)  
▪  $f(x_i)$  : prediction result (output)  
e.g. linear regression  
 $f(x_i) = w^T x_i + b$



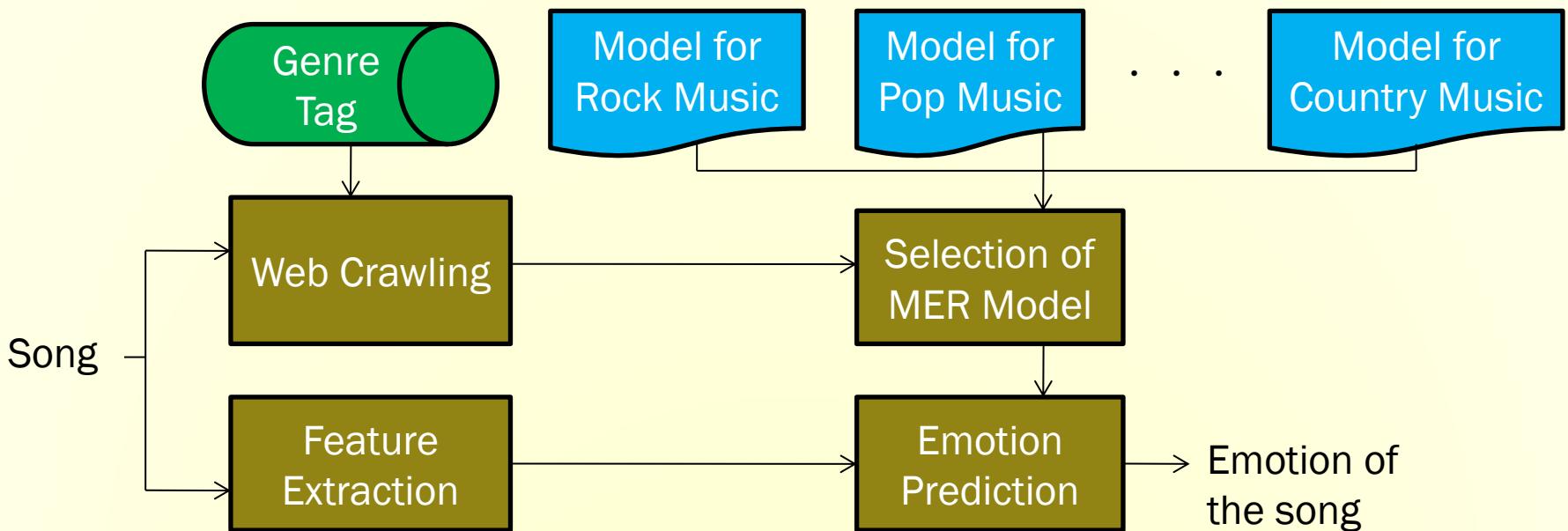
# Improving Valence Recognition by Lyrics

- Lyrics
- 張惠妹 - 人質
  - Without lyrics  Original
  - Lyrics - 在我心上用力的開一槍  
讓一切歸零 在這聲巨響  
如果愛是說什麼都不能放  
我不掙扎
  - Without lyrics – neutral
  - With lyrics – sad
- Improves accuracy of valence by 19.9%



# Improving MER by Genre Tags

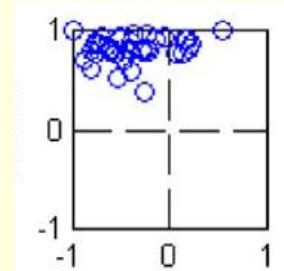
- Genre information makes music retrieval more effective



- MER accuracy increases 13.0%

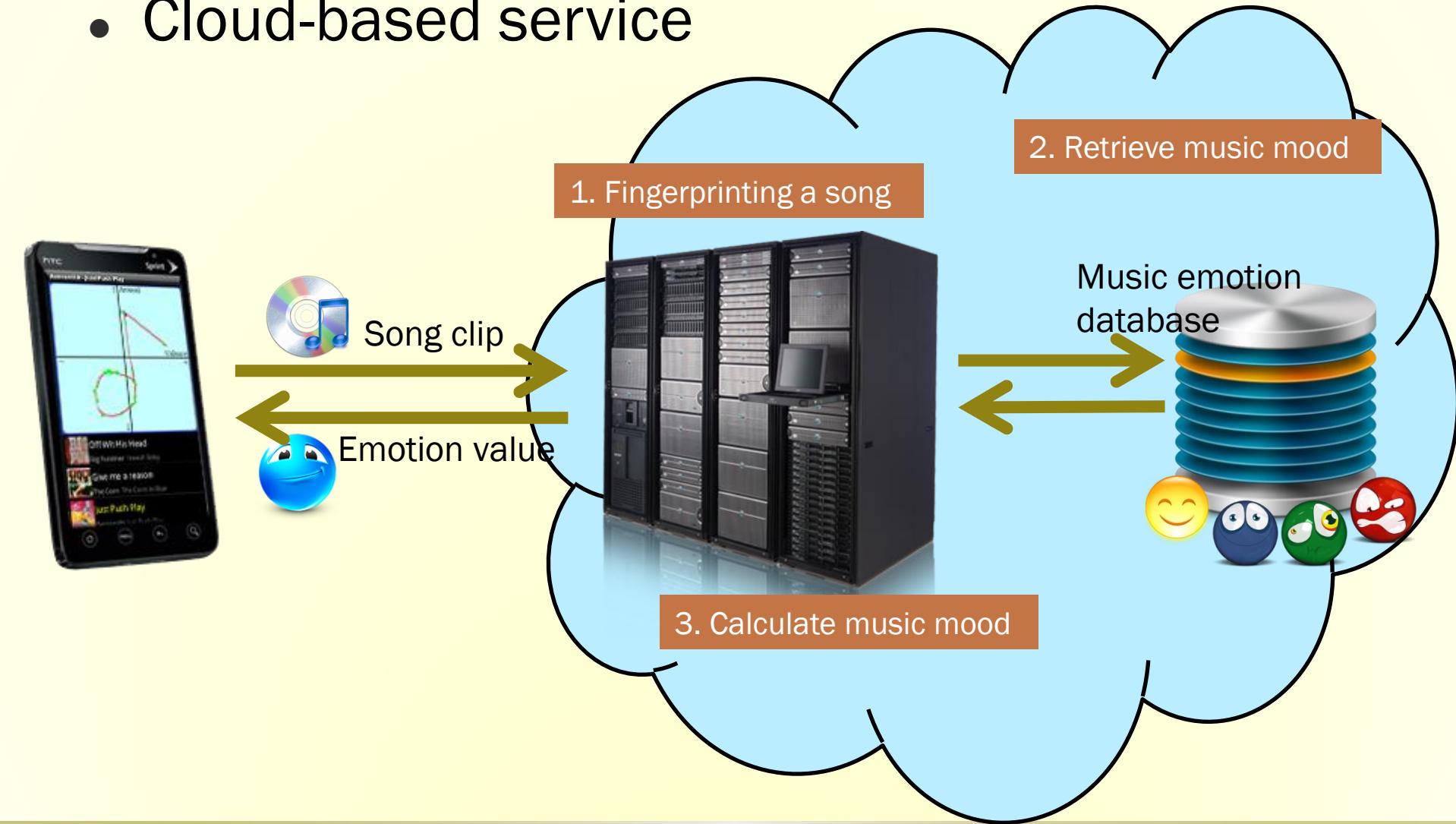
# Personalized MER System

- People feel the same song differently
- A general model that fits everyone is almost impossible
  - Smells Like Teen Spirit by Nirvana 
  - Most people annotate it as negative valence
  - Rock music fans annotate it as positive valence
- Our approach
  - Build personalized model by user feedback
  - Choose an MER model according to personal information (gender, age, music preference, etc.)



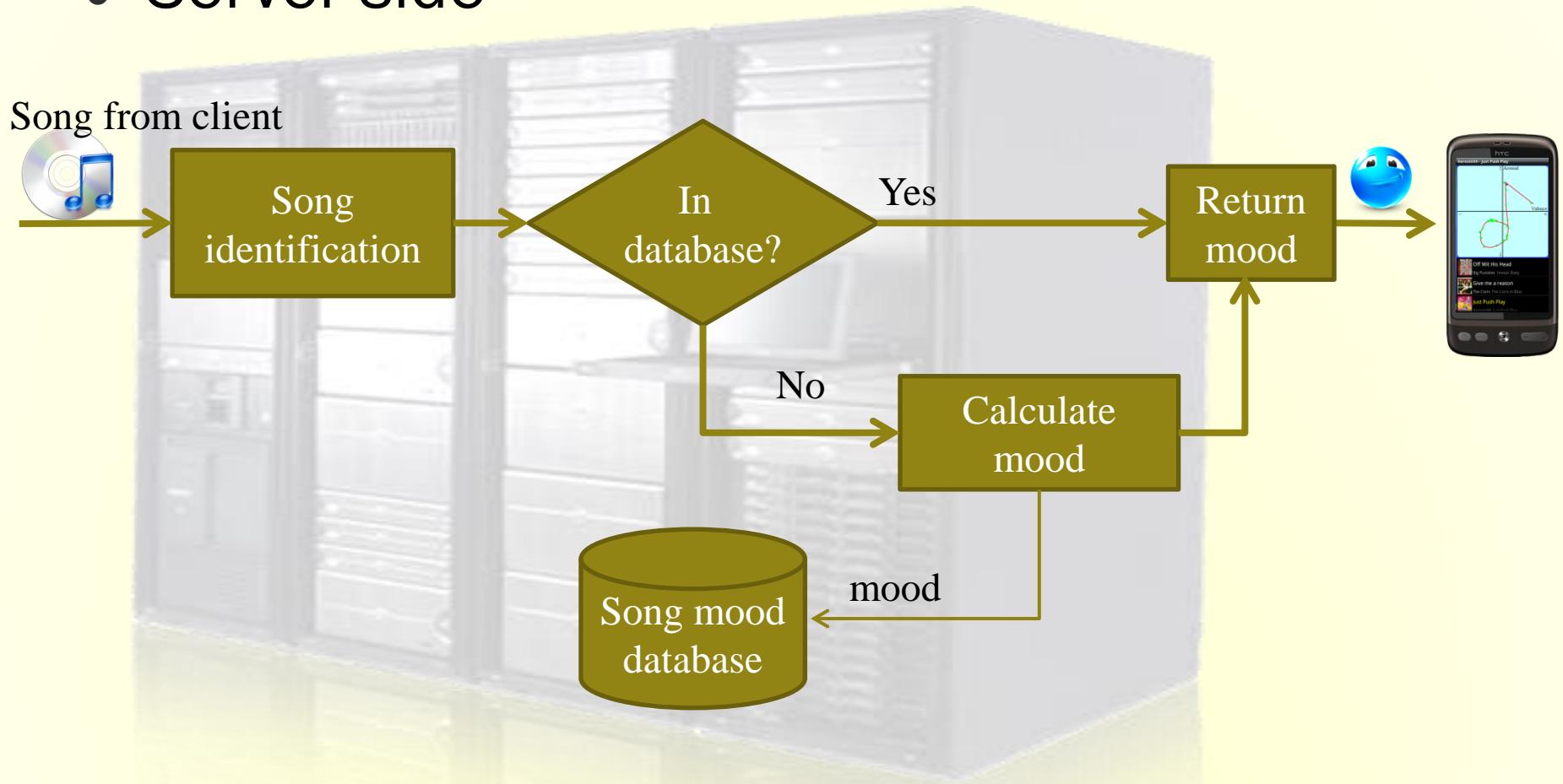
# Emotion-Based Music Retrieval System

- Cloud-based service

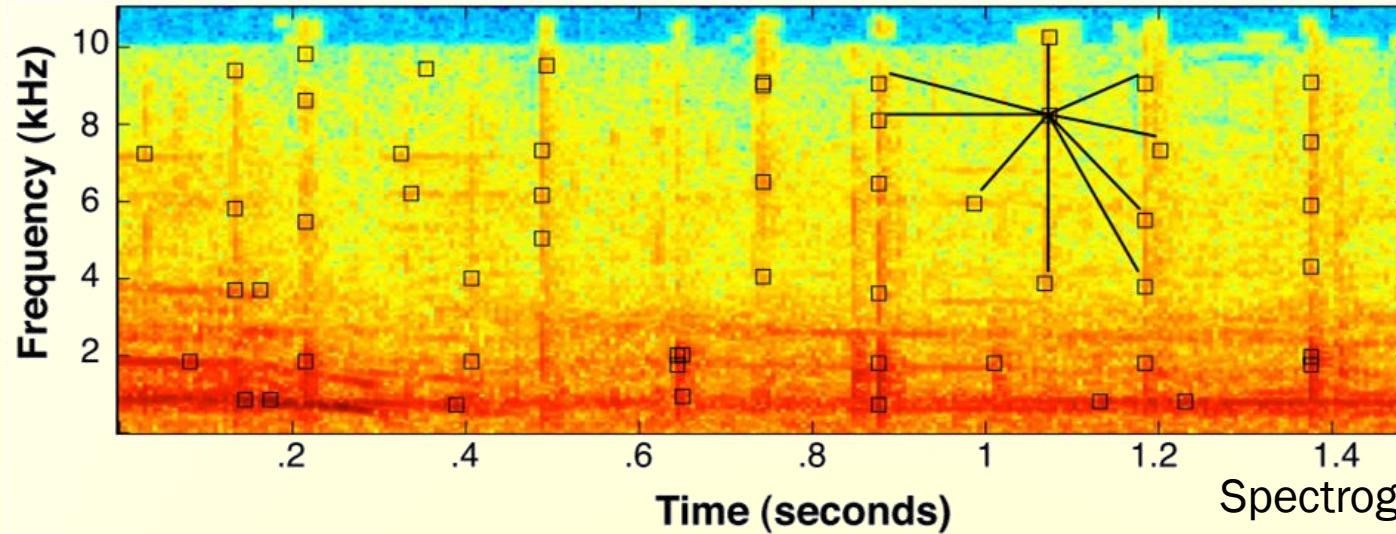


# Emotion-Based Music Retrieval System

- Server side



# Audio Fingerprinting



Spectrogram of a song  
clip with peak intensities  
marked

Frequency (Hz)	Time (seconds)
823.44	1.054
1892.31	1.321
712.84	1.703
...	...
819.71	9.943

Example fingerprint

# Emotion-Based Music Retrieval System

- Client side
  - Automatic song profiling
  - Retrieve songs by emotion coordinates
  - Represent playlist by emotion trajectory
  - Show the mood distribution of all songs of an artist
- Ported to Android phone and iPhone



# Acknowledgement



Yi-Hsuan Yang



Sighter Liu



Ming-Yen Su



Yu-Ching Lin



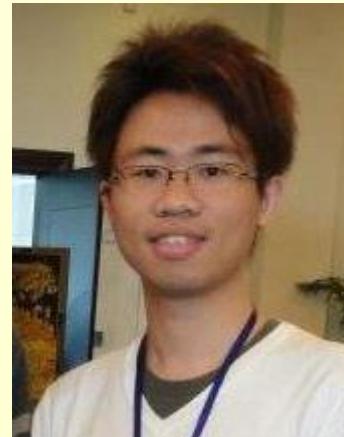
Cheng-Te Lee



Ya-Fan Su



Ann Lee



Keng-Sheng Lin



Cheng-Ya Sha



Pei-Chun Chen



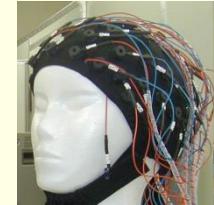
Heng-Tze Cheng

# Extensions

---



Recognize users' emotion by EEG  
for music recommendation



EEG



Music therapy



NeuroSky



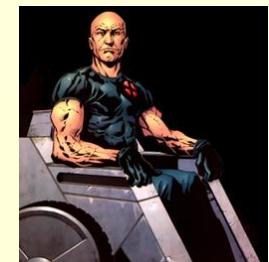
Music accompaniment



Incidental music selection



Karaoke system

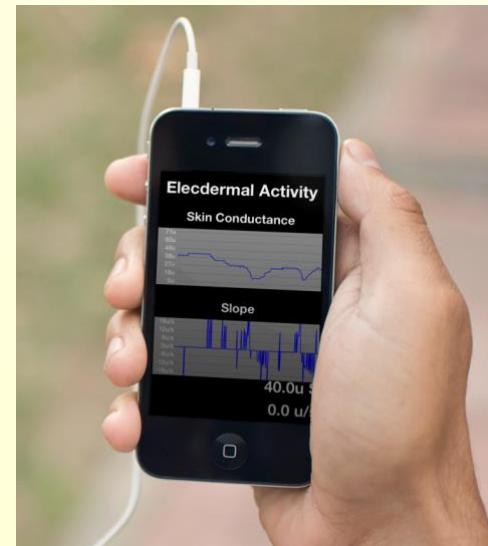


Prof. Charles Xavier

# Electrodermal Sensor



Affectiva



NTU

# Automatic Video Highlight Extraction

- Traditional system only considers low-level visual features
  - Motion, color
- Use the emotion of incidental music to improve accuracy

Buzzer Beat  
34 sec



The flower shop without Roses  
22 sec



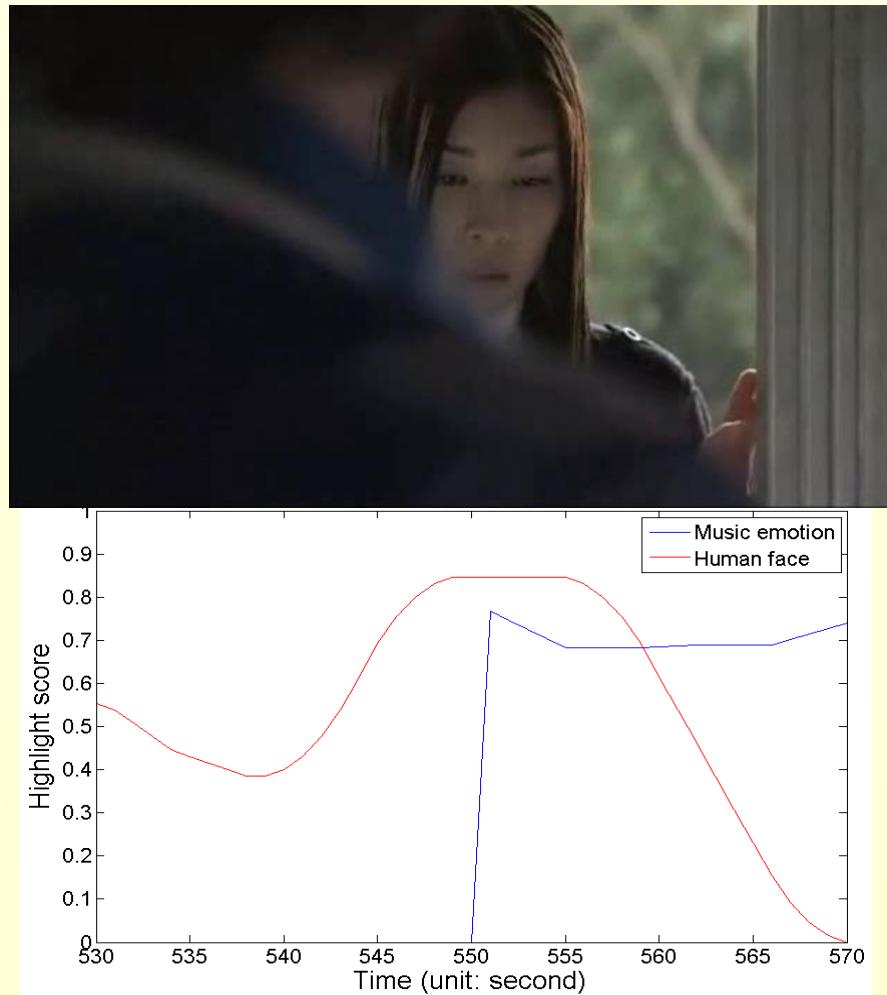
Last Friends  
24 sec



K.-S. Lin, A. Lee, Y.H. Yang, and H. H. Chen, "Automatic highlights extraction for drama video using music emotion and human face features," in *Proc. IEEE Workshop on Multimedia Signal Processing*, Nov. 2011.

# Romantic Music

- Human face and music



# Automatic Transcription

- Classical music
  - No lyrics
  - Arouses emotion by melody
- Automatic Transcription of Piano Music
  - Extracts melody information

Song	Prelude and Fugue No.2 in C Minor	Sonata no. 8 Pathetique in C minor, 3 <sup>rd</sup> movement	Moments Musicaux No.4	Sonata K.333 in Bb Major, 1 <sup>st</sup> Movement
Composer	Bach	Beethoven	Schubert	Mozart
Original				
Result				

C.-D. Lee, Y.-H. Yang, and H. H. Chen, “Multipitch estimation of piano music by exemplar-based sparse Representation,” *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 608-618, Jun. 2012.

# Singing Voice Timbre Classification

---

- Using singing voice Timbre to classify music
- Build a new data set for this task
- Empirically validate that
  - Using vocal segment detection and singing voice separation improves the classification accuracy
  - Voice features are remarkably effective
- Applications: singing voice timbre as a high-level feature

C.Y. Sha, Y.-H. Yang, Y.-C. Lin and H. H. Chen, "Singing voice timbre classification of Chinese popular music," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013.

# The KKTIM data set

- KKBOX Timbre
  - 387 Chinese songs (91 singers)
  - 272 Japanese songs (66 singers)
- 6 singing timbre classes
- Multi-label, per-song instead of per-singer

Chinese	#Song
低沉 (Deep)	74
沙啞 (Gravelly)	57
渾厚 (Powerful)	70
甜美 (Sweet)	54
空靈 (Ethereal)	63
高亢 (High-pitched)	81

Japanese	#Song
轉音多、鼻音 (Run riffs/Nasal)	39
沙啞 (Gravelly)	50
渾厚 (Powerful)	50
活力偶像、甜美 (Sweet)	50
乾淨、明亮 (Bright)	40
高亢 (High-pitched)	50

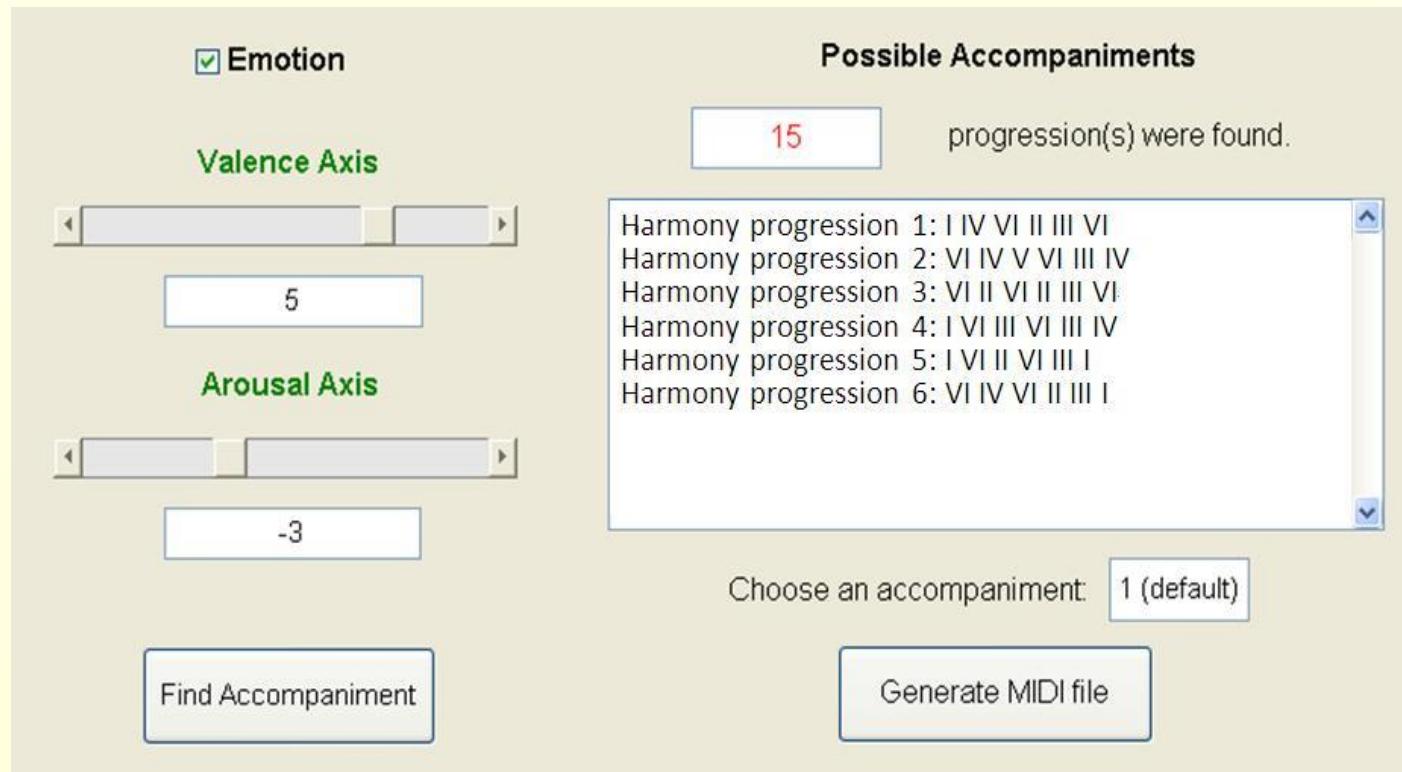
# Demo

- Chinese songs
- Probability output of six classifiers

Song ID	低沉 Deep	沙啞 Gravelly	渾厚 Powerful	甜美 Sweet	空靈 Ethereal	高亢 High-pitched	標記 Ground truth
384	0.72	0.49	0.51	0.10	0.56	0.58	1,0,0,0,0,0
117	0.43	0.62	0.67	0.28	0.13	0.21	0,1,0,0,0,0
631	0.56	0.52	0.60	0.30	0.05	0.40	0,0,1,0,0,0
632	0.25	0.50	0.32	0.91	0.20	0.45	0,0,0,1,0,0
443	0.30	0.47	0.45	0.38	0.84	0.45	0,0,0,0,1,0
371	0.67	0.54	0.50	0.31	0.11	0.48	0,0,0,0,0,1
636	0.14	0.54	0.47	0.77	0.83	0.63	0,1,1,0,0,0

Probability of belonging to the class

# Emotional Accompaniment Generation

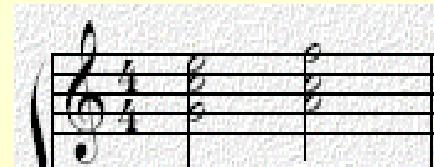
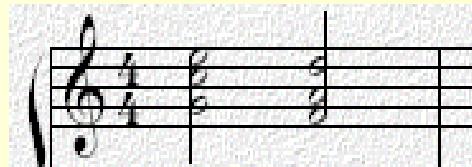


P.-C Chen, K.-S. Lin, and H. H. Chen, "Emotional accompaniment generation system based on harmonic progression," *IEEE Trans. Multimedia*, v. 15, no. 7, pp. 1-11, Nov. 2013

# Valence and Harmonic Progression

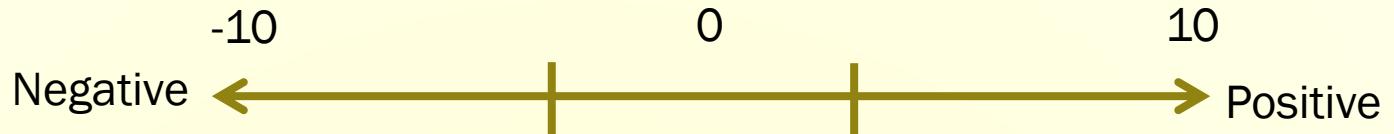
---

- Main valence-affecting features
  - ✓ Mode: major key-happy; minor key-sad
  - ✓ Chord: focus is on the consonance of a single chord
- Music is a delicate temporal art with emphasis on the flow of chords =>  
**Harmonic progression**



# Putting Everything Together

Valence values



Melody



-10

0

10

Negative

Positive



Arousal values



Block Chord



-10

0

10

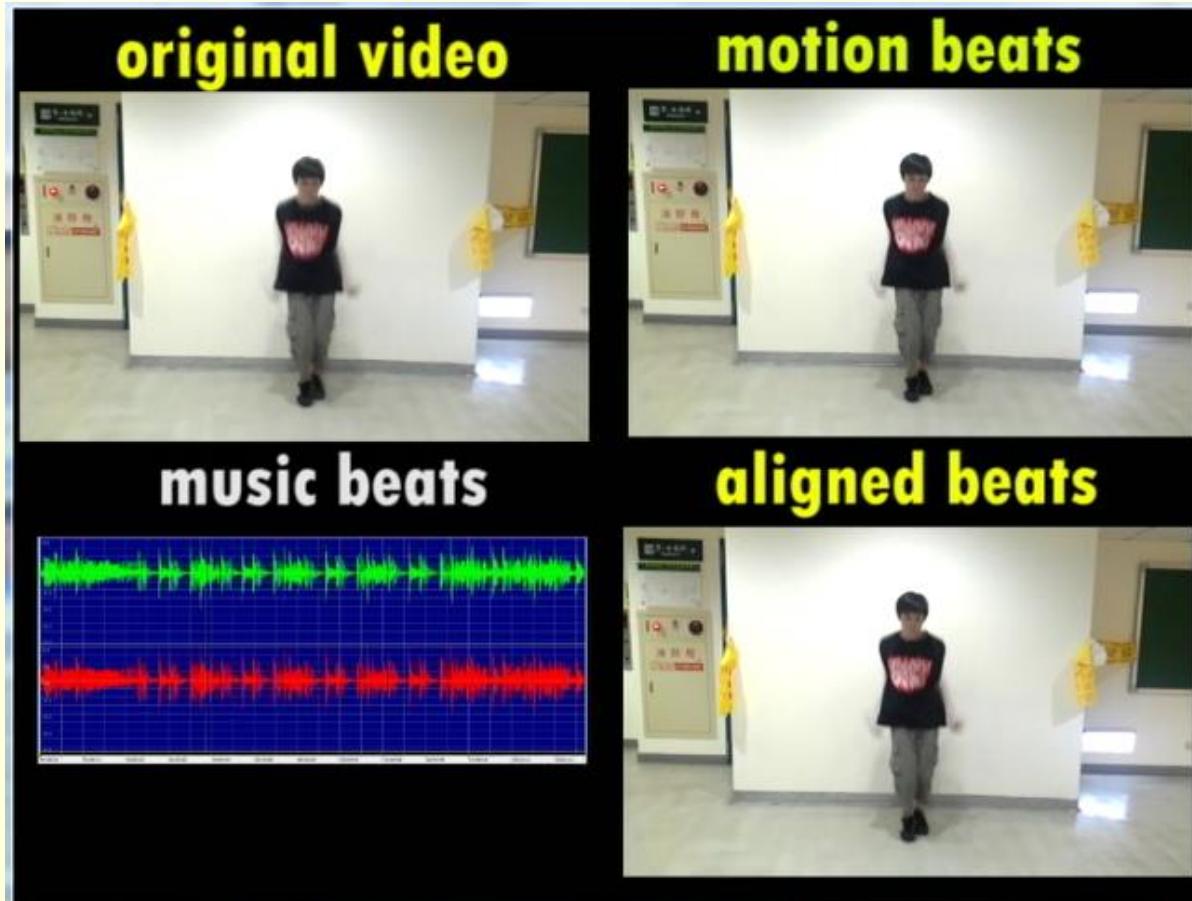
Calming

Exciting

Arpeggio



# Alignment Evaluation of Motion Beats



C. Ho, W.-T. Tsai, K.-S. Lin, and H. H. Chen, “Extraction and alignment and evaluation of motion beats for street dance,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013.