從古典音樂看機器聽覺的若干難題

清大電機系 劉奕汶

Nov. 17, 2015
Presented as a special topic at
EE6641 Analysis and Synthesis of Audio Signals

個人簡歷

- 1979-80: 山葉兒童音樂班
- 1981, 84: 獲北市立農國小鋼琴比賽第一名
- 1987: 獲北市明德國中鋼琴比賽第一名
  - （Schubert 即興曲 op.143 no. 3）
  - 擊敗眾多女生
  - 1988：獲邀校慶演出（Chopin 幻想即興曲）
- 1990: 北市建國中學班際合唱比賽伴奏
- 1994-95: 台大電機之夜幫男聲合唱／marimba 同學伴奏
  - 獲邀於廣播電台演出
- 1996: 台大交響樂團畢業公演幫小提琴同學伴奏
  - （Brahms: Sonata for violin and piano No. 1）
- 1996: 海軍航輪兵通四校合唱比賽幫別的營的同梯伴奏
  - 獲榮譽假12小時
  - 再於中秋節晚會擔任孔鏘
- 2000-02: 史丹福台灣同學會piano bar/ 幫同學伴奏
- 2005: 台大合唱團灣區公演
- 1999- 2013: Church choirs

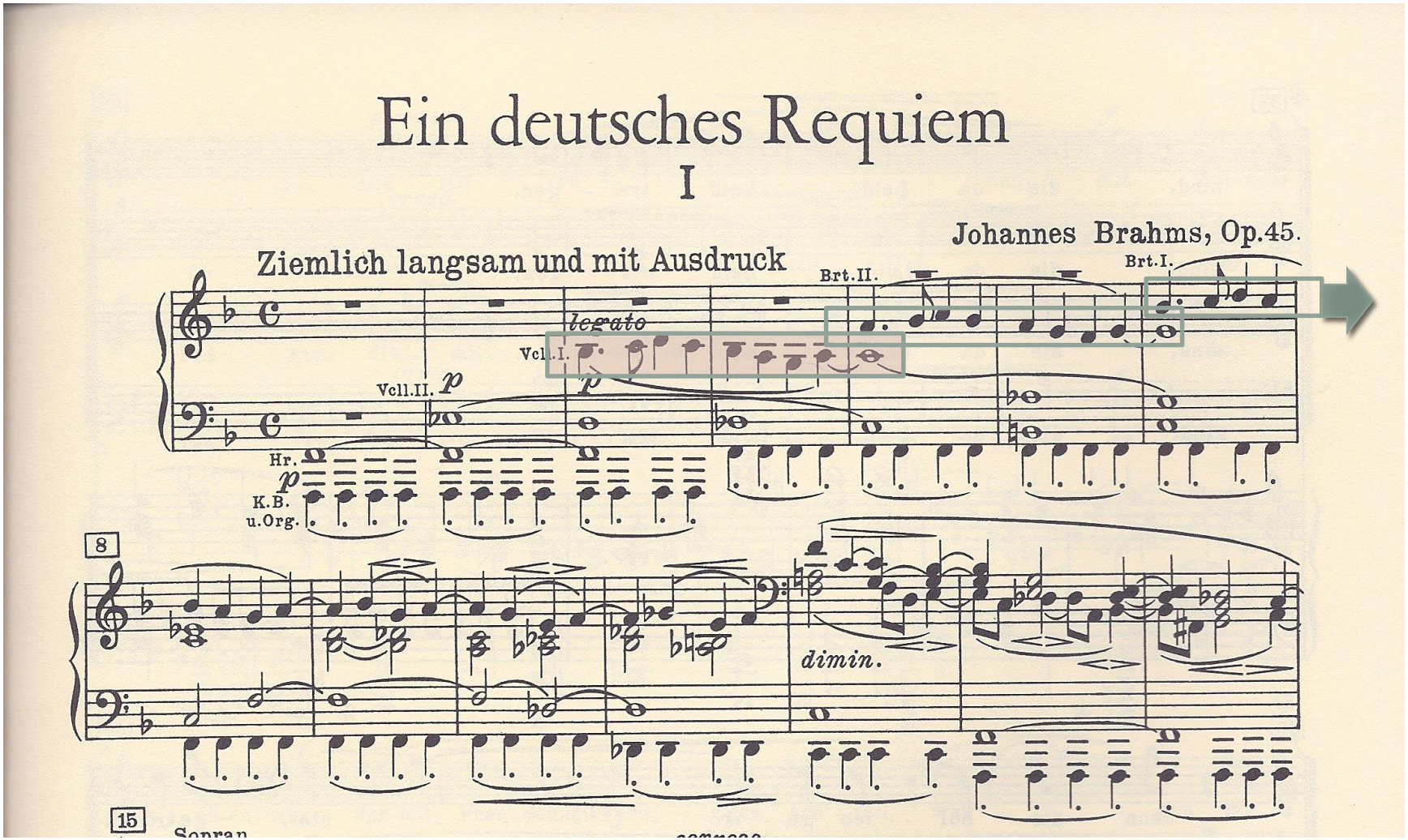
DISCLAIMER:

This presentation generally lacks academic rigor.
View with discretion.

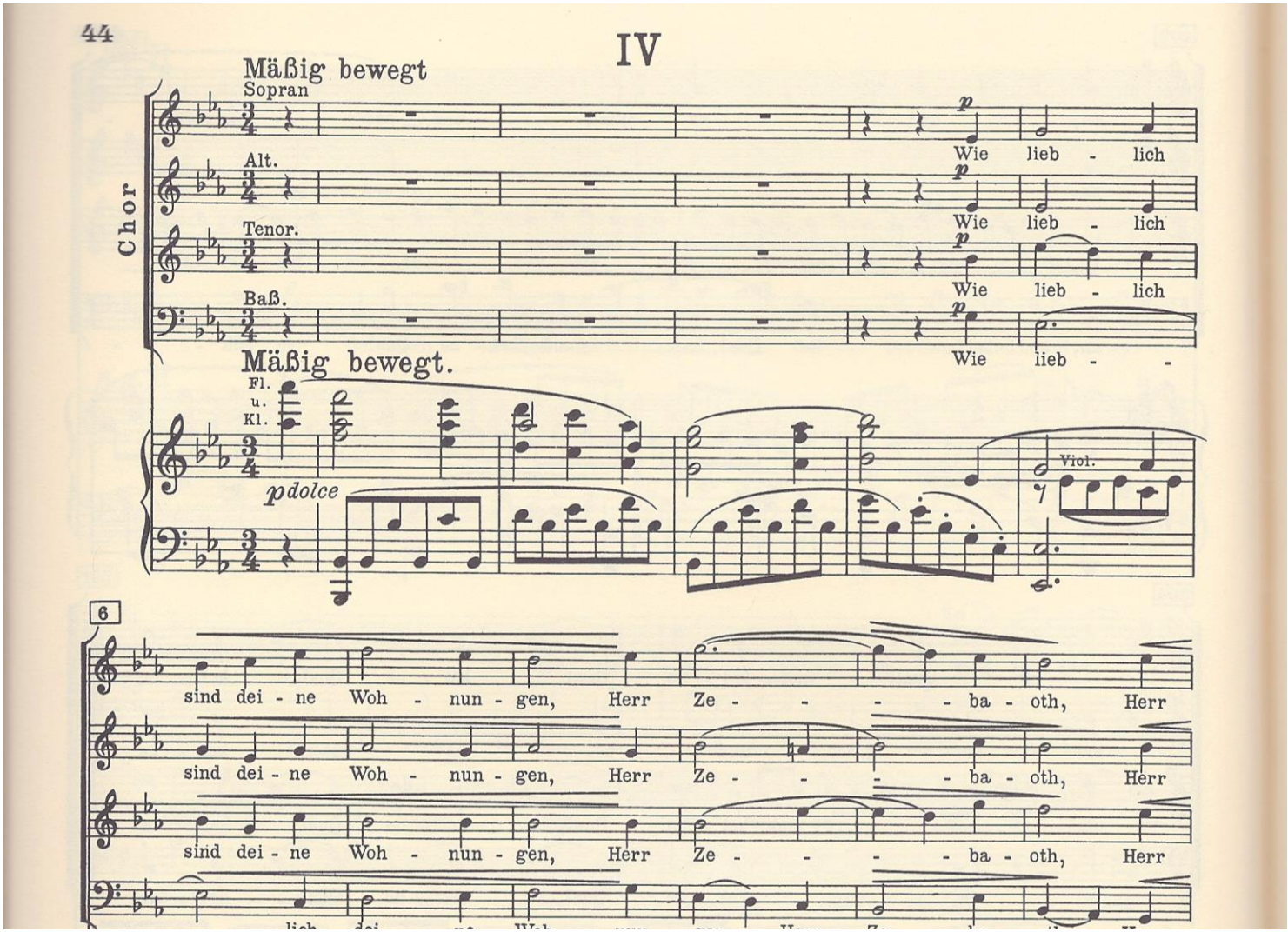
(西洋)古典音樂聽什麼?

- Physics/signals:
  - Pitch 音高, harmony和聲 (frequency)
  - Rhythm節奏, beats 拍子, tempo速度 (time)
  - Dynamics強弱, timbre音色
    - (time-frequency representation)
- Syntax: phrasing樂句, structure結構
- Semantics: style風格, interpretation詮釋

譜例


Ein deutsches Requiem
I
Johannes Brahms, Op.45.
Ziemlich langsam und mit Ausdruck
Brt.II.
Brt.I.
legato
Vcl.I.
Vcll.II.
p
Hr.
K.B.
u.Org.
8
dimin.
15
Sopran

譜例 (ii)


44
IV
Mäßig bewegt
Sopran
Alt.
Tenor.
Baß.
Chor
Wie lieb - lich
Wie lieb - lich
Wie lieb - lich
Wie lieb -
Mäßig bewegt.
Fl. u. Kl.
pdolce
Viol.
6
sind dei - ne Woh - nun - gen, Herr Ze - - - - ba - oth, Herr
sind dei - ne Woh - nun - gen, Herr Ze - - - - ba - oth, Herr
sind dei - ne Woh - nun - gen, Herr Ze - - - - ba - oth, Herr

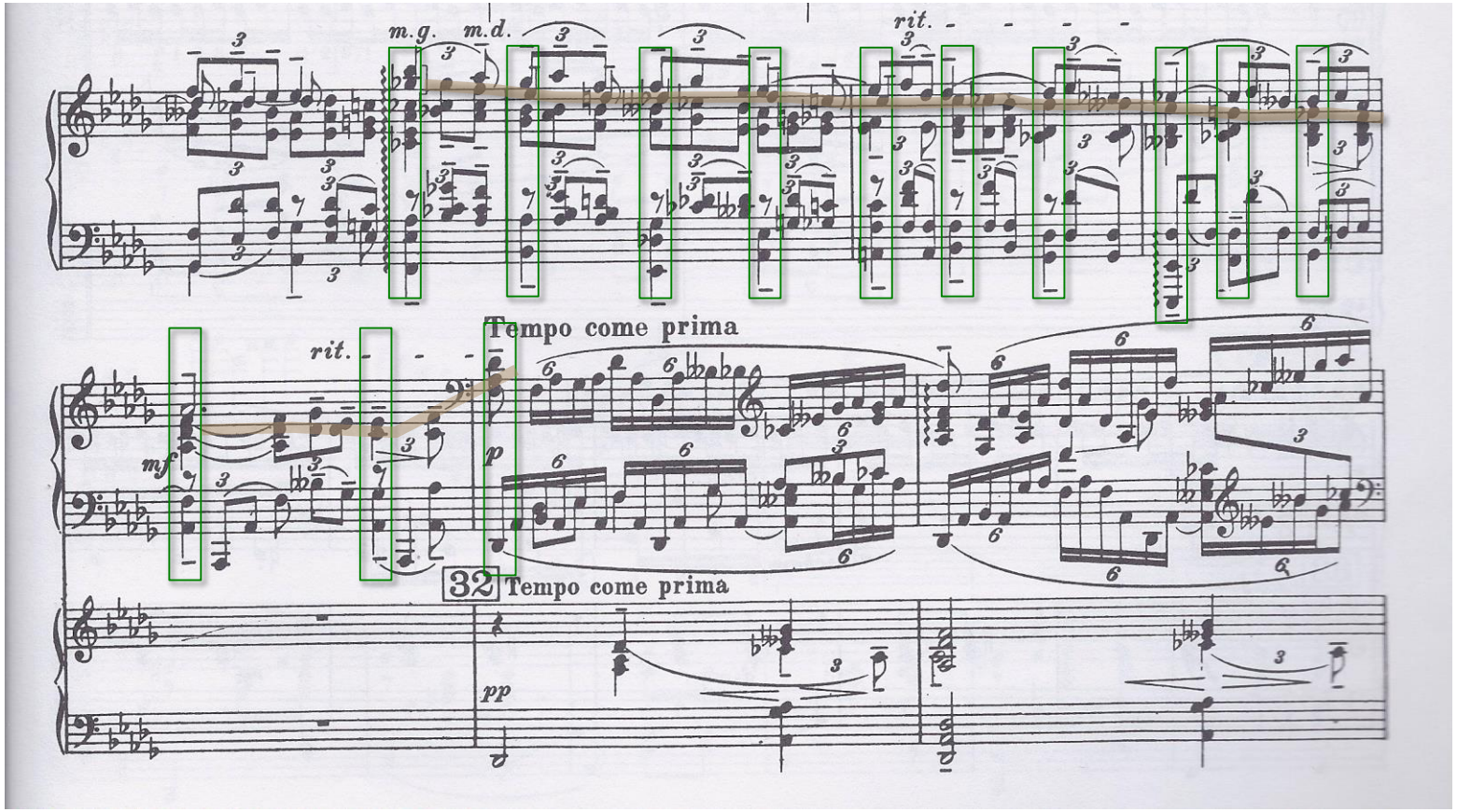
音樂交織於 time-frequency 二度空間中

- Time-varying spectrum = “chord progression” (和絃演進)
- Concurrent melodies = “counterpoint” (對位)

**Listeners can choose to focus on either aspect*

**Performers memorizes music via recognition of structures*

A. Weissenberg (1983)
Bulgarian NSO



Rachmaninoff, Piano Concerto No. 3, 2nd movement

CHALLENGE:

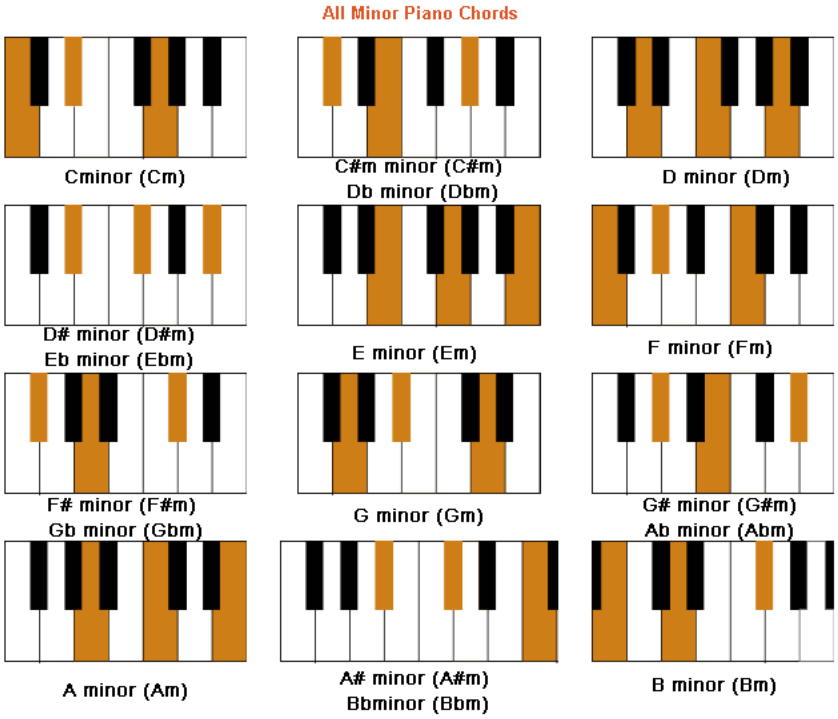
Can machine dig out *interesting* things?

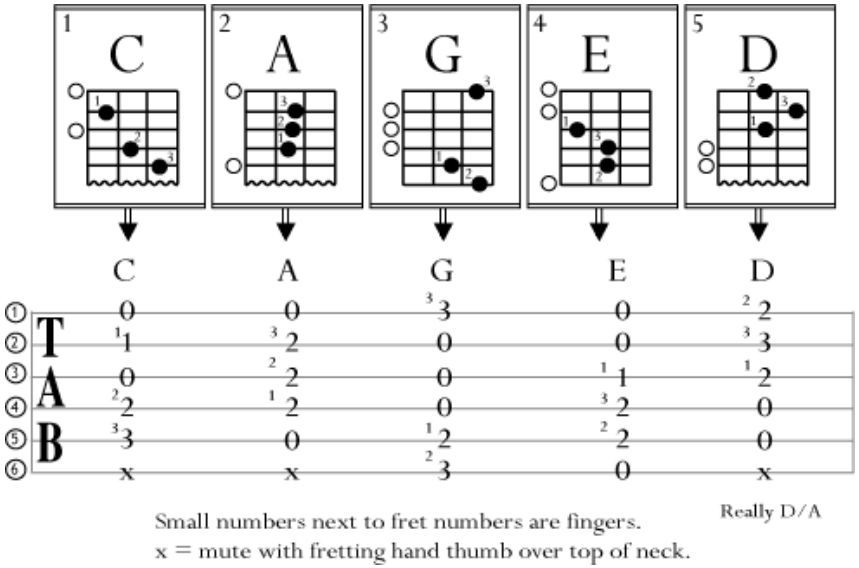
Today's Outline

- 聆聽的理性
  - 和聲的辨認
  - 節奏的辨認
  - 自動轉譜
  - 音樂風格之辨認
- 聆聽的感性：談音樂中的情緒

和聲 / 和絃的辨認

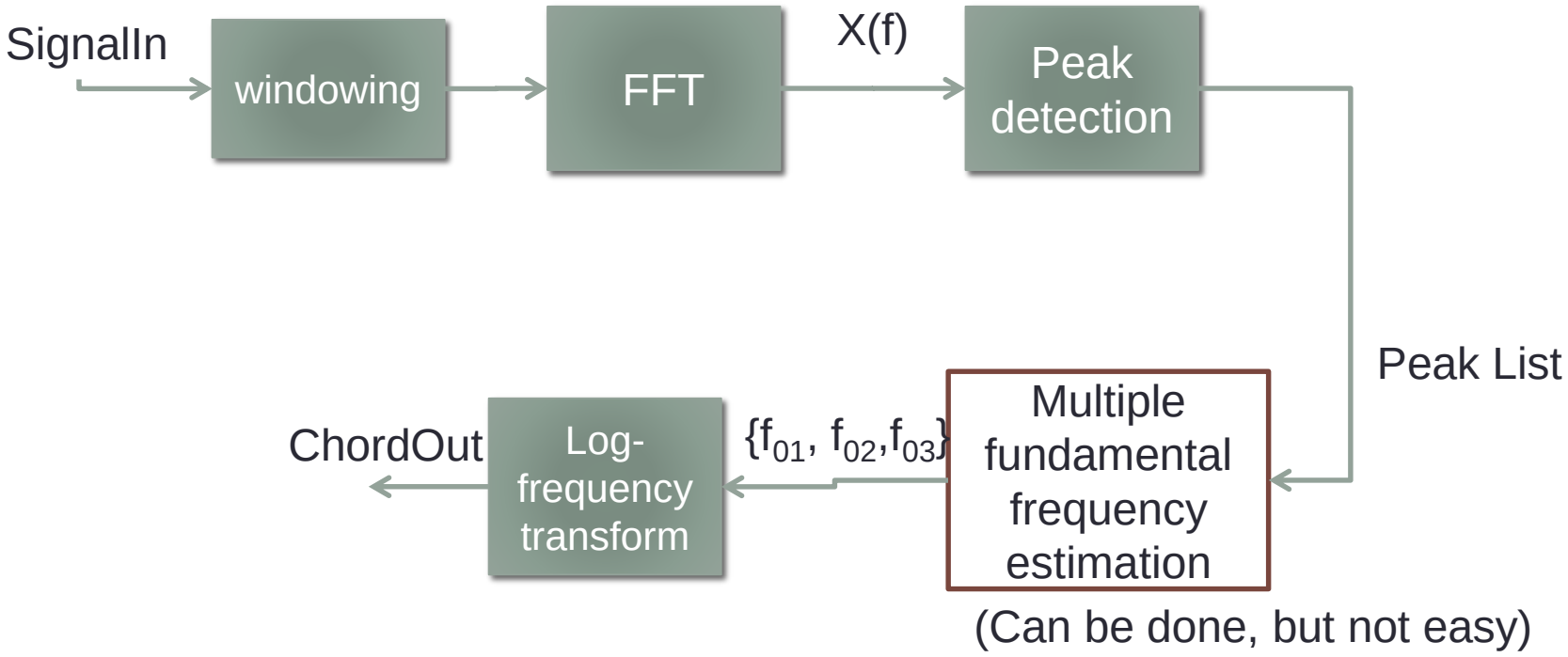
- 和絃的定義：三個音(triad) 以上 **一起**出現





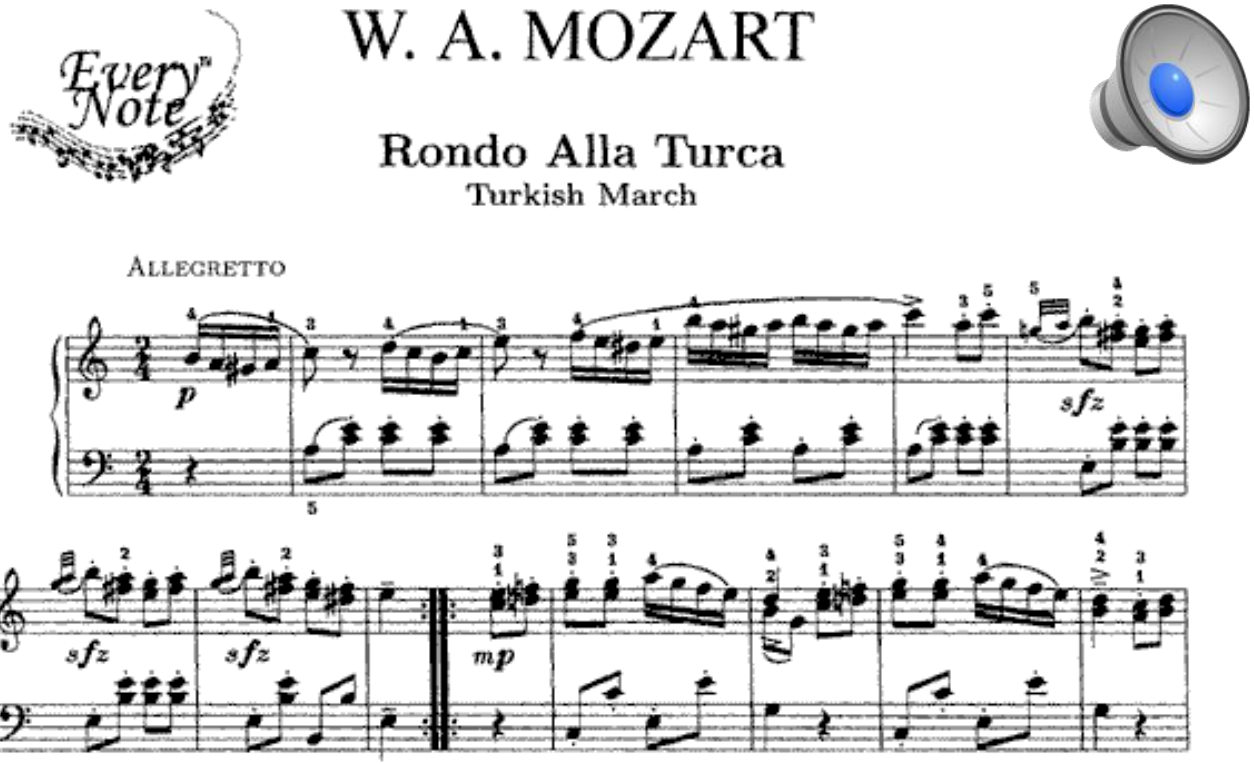
http://www.personal.psu.edu/users/m/j/mjp5109/Chords.html

A prescription for machine chord extraction based on spectral analysis

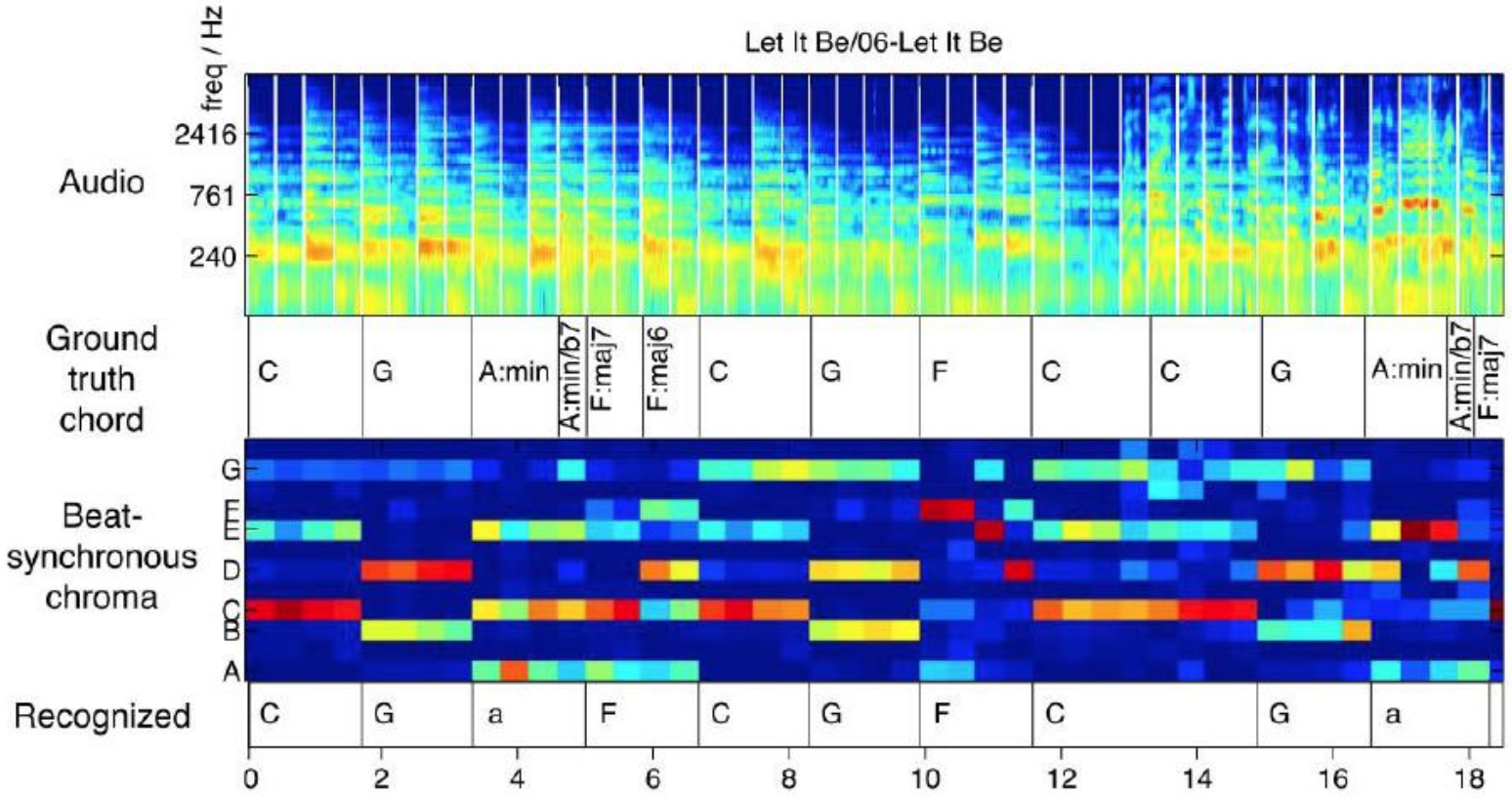


Problems of merely using spectral analysis

- 分散和絃：屬於同一個和絃的音不見得同時出現
- 音樂的旋律不見得每個音都落在和絃上



Chord recognition beyond spectral estimation:
Feature extraction by energy collapsing across octaves



洪暐桓、蔡鈺群 (交大電信) 徐培霖 (清大資工)

EE6641 Analysis and Synthesis of Audio Signals 期末報告

[2012 MIREX competition]

CHALLENGE:
和絃不難，轉位難。

Today's Outline

- Automatic listening
  - 和聲的辨認
  - **節奏的辨認**
  - 自動轉譜
  - 音樂風格之辨認
- 音樂與情緒

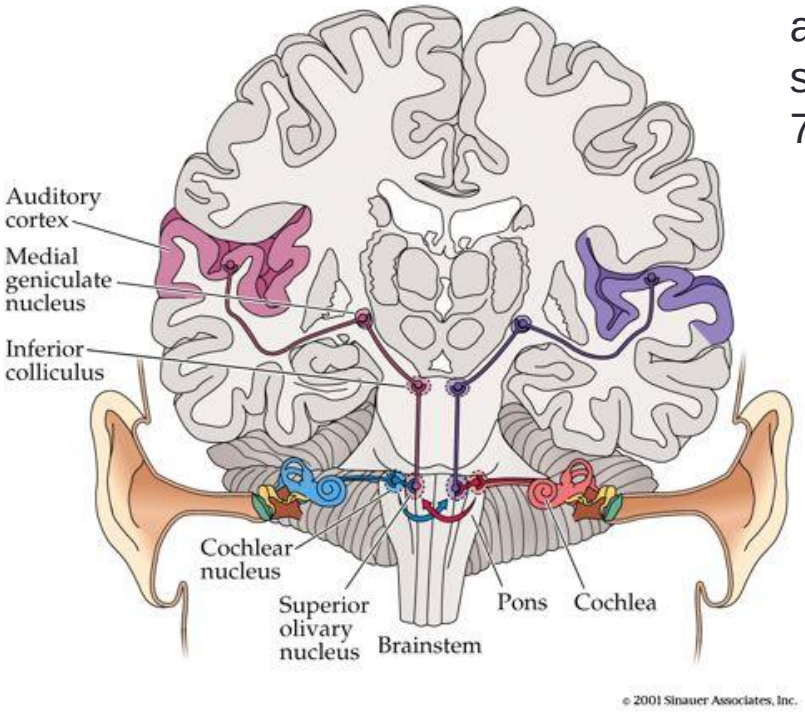
節奏的通則

- 2/4 : | 強弱 | 強弱 | 強弱 | …
- 3/4 : | 強 弱 弱 | 強 弱 弱 | …
- 4/4 : | 強弱 ？ ？ | …
- 6/8 : | 強弱弱 ？ ？ ？ |
- 其他

凡規則必有例外

Can **you** feel the beats and tap along?

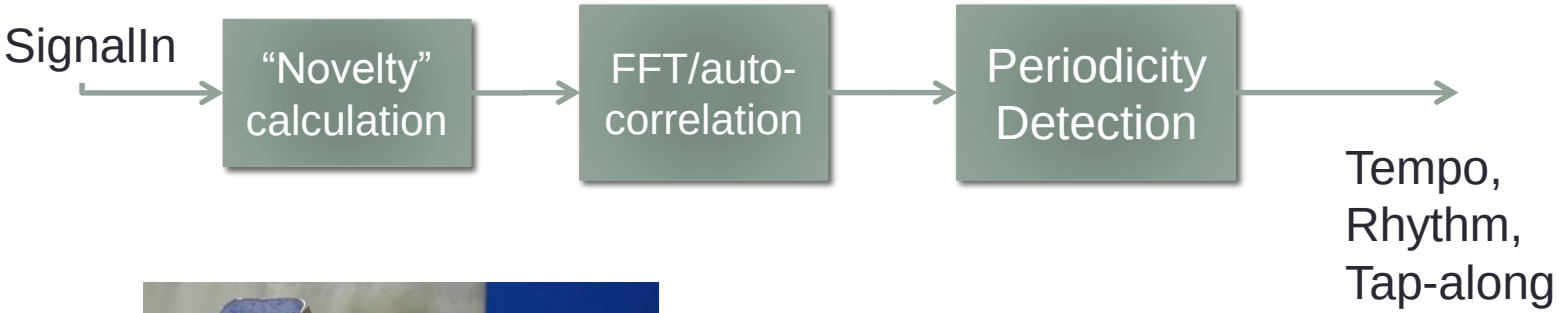
Why is it so natural to follow the beats?

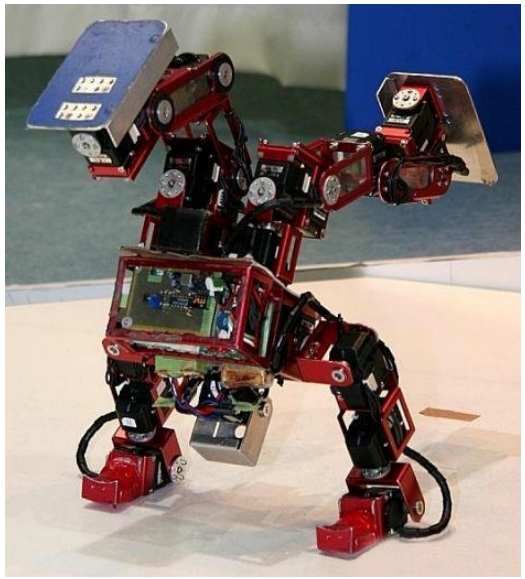


RS Snyder, A Stowell (1944). "Receiving areas of the tactile, auditory, and visual systems in the cerebellum," J Neurophysiol. 7:331-357.

CM Huang, G Liu, R Huang (1982). "Projection from the cochlear nucleus to the cerebellum," *Brain Res* 244:1-8.

A prescription for computer rhythm recognition/automatic tap-along





http://www.dailymail.co.uk/sciencetech/article-1123882/Breakdancing-robots-set-storm-classrooms-future.html

http://inc.ucsd.edu/~poizner/motioncapture.html

Challenges in computer rhythm recognition/automatic tap-along

- Syncopation (切分音):
  when strong beat does not align with measure lines
  - First beat can even be silent.
    - E.g., “[ ] You better run you’d better do what you can”

- Half/Doubling ambiguity

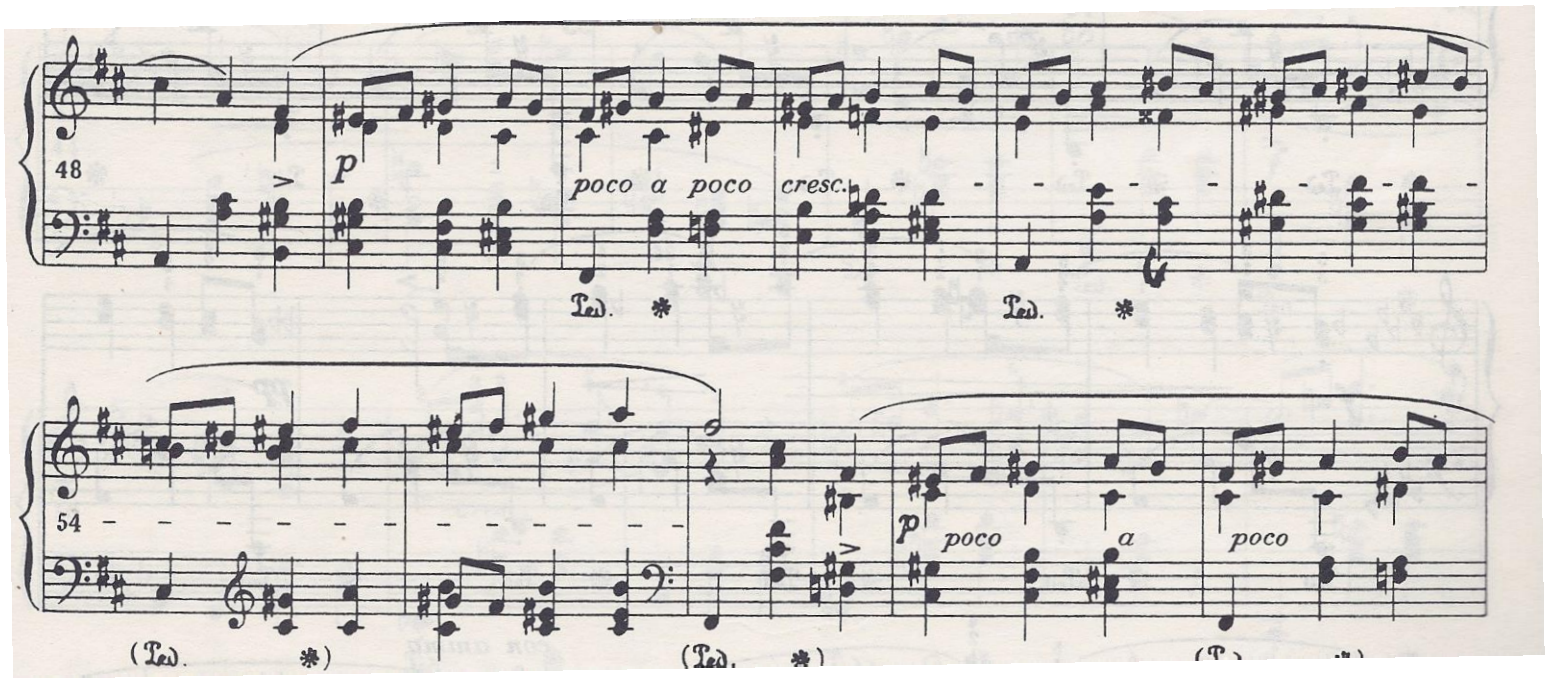
- *Rubato* (彈性速度):
  expression via change of speed

Can computer tap along?

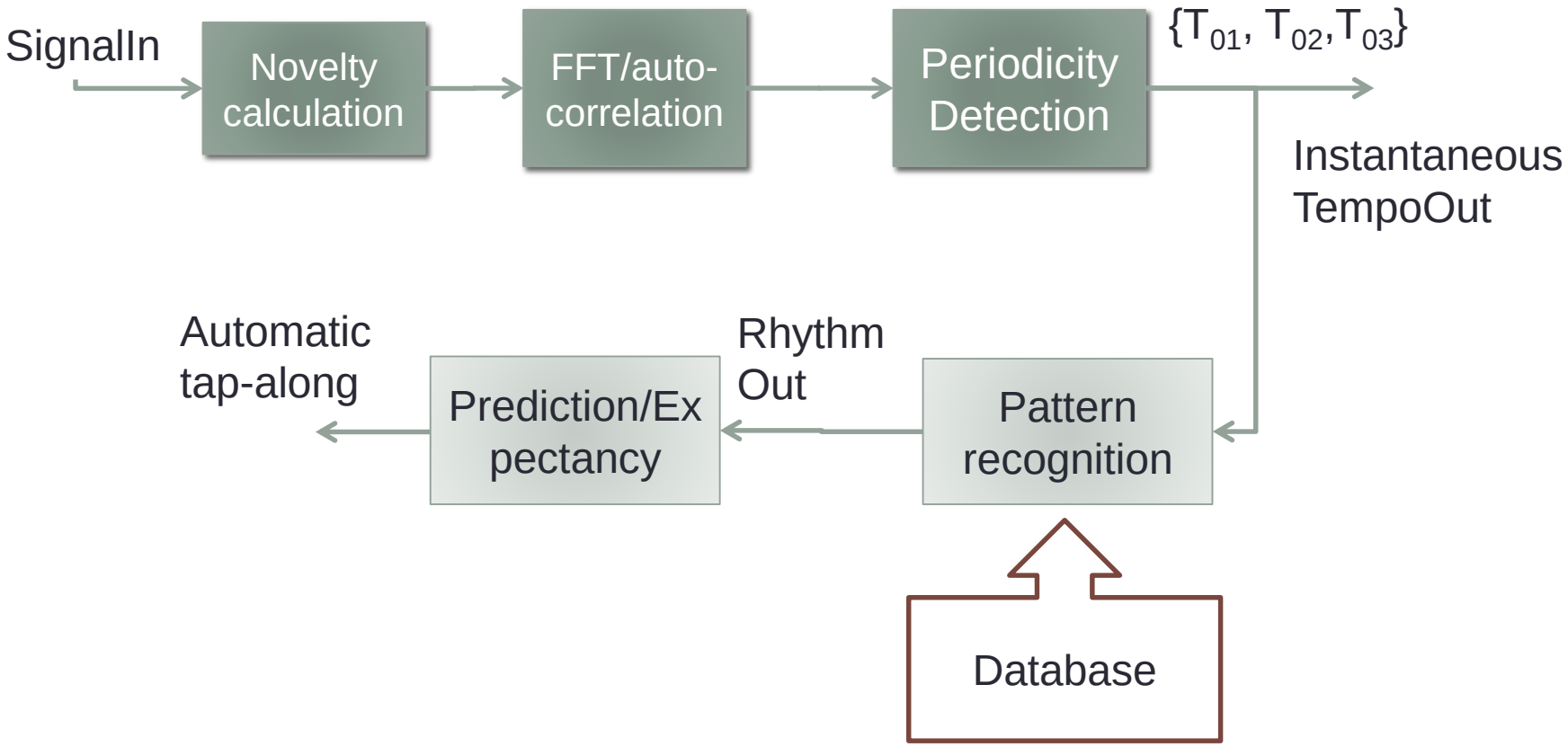
- This excerpt was included in a contest
  - International Conf. Music Info. Retrieval

Performed by Garrick Ohlsson

Performed by A. Rubinstein



A revised prescription for computer rhythm recognition (節奏辨識)/ automatic tap-along (自動跟拍)



Automatic tap-along:

simple idea, many applications

Sonation's Cadenza™

Today's Outline

- Automatic listening
  - 和聲的辨認
  - 節奏的辨認
  - **自動轉譜**
  - 音樂風格之辨認

- 音樂與情緒

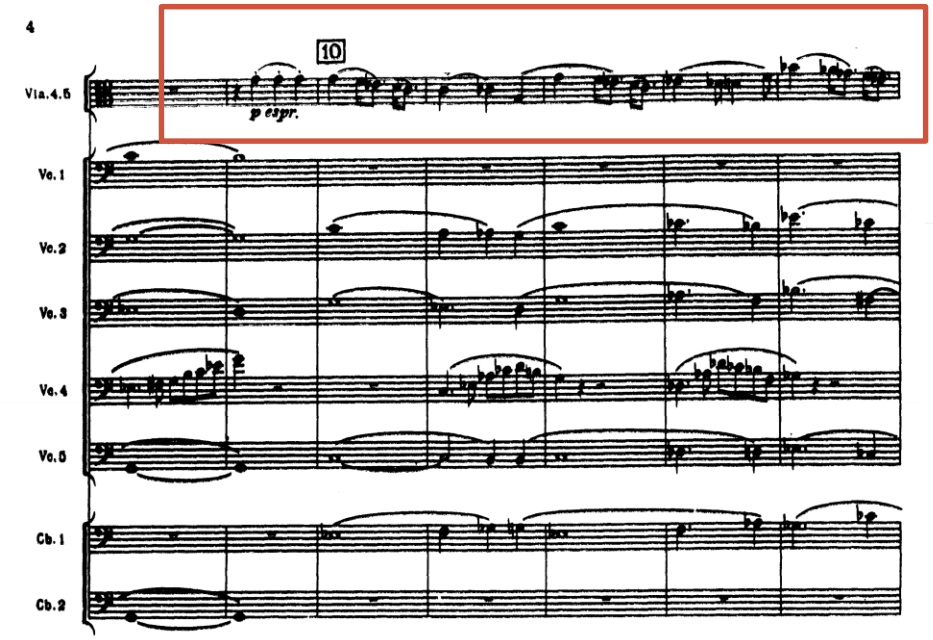
自動轉譜的需克服的細節

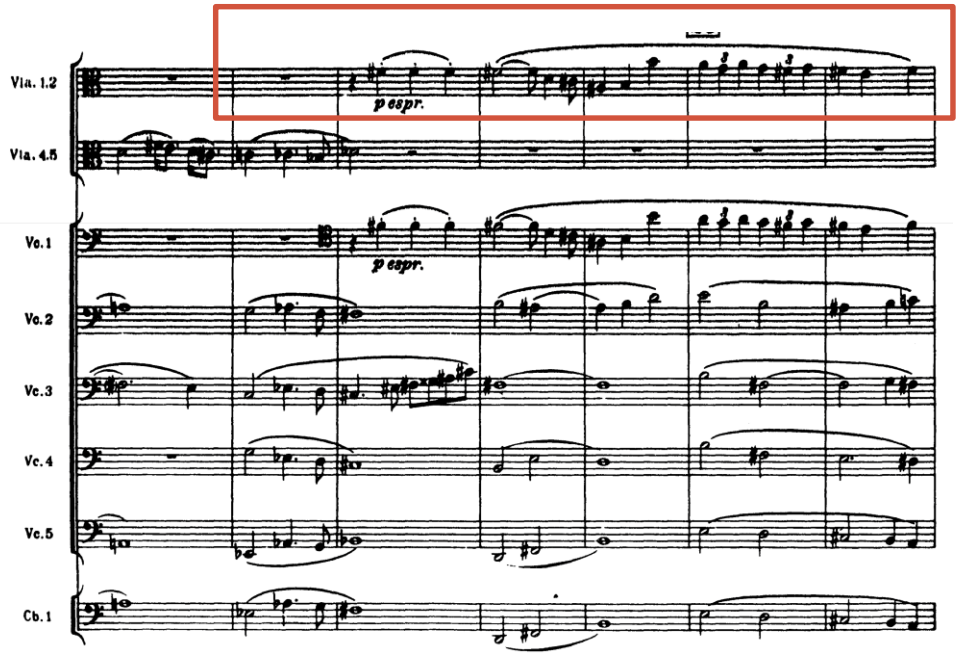
- Multiple pitch estimation
- Quantization in time
- **Notation: musical informatics**
- Beyond listening: Does music have to be heard?

Metamorphosen, by R. Strauss (1945)

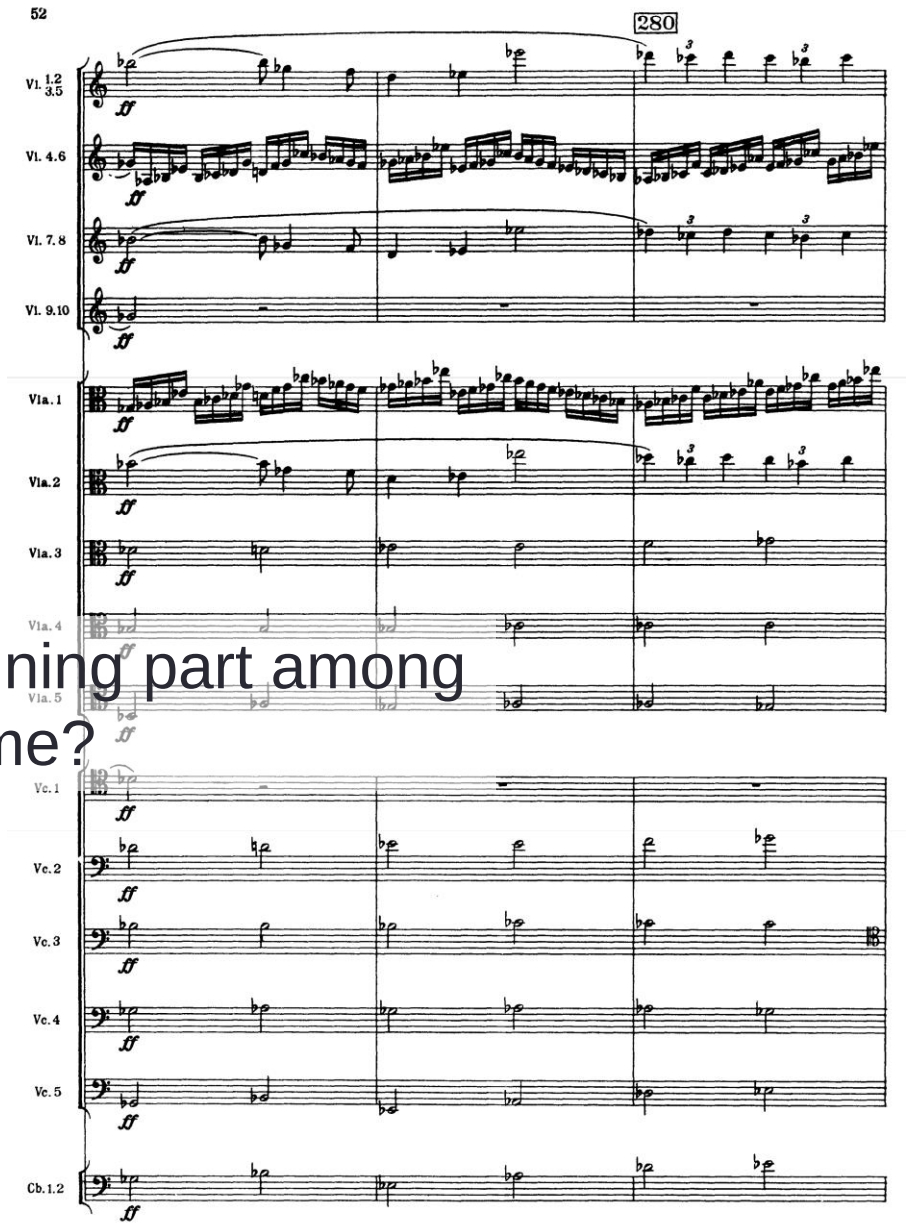
For 23 solo strings

- 10 violins
- 5 violas
- 5 cellos
- 3 double basses

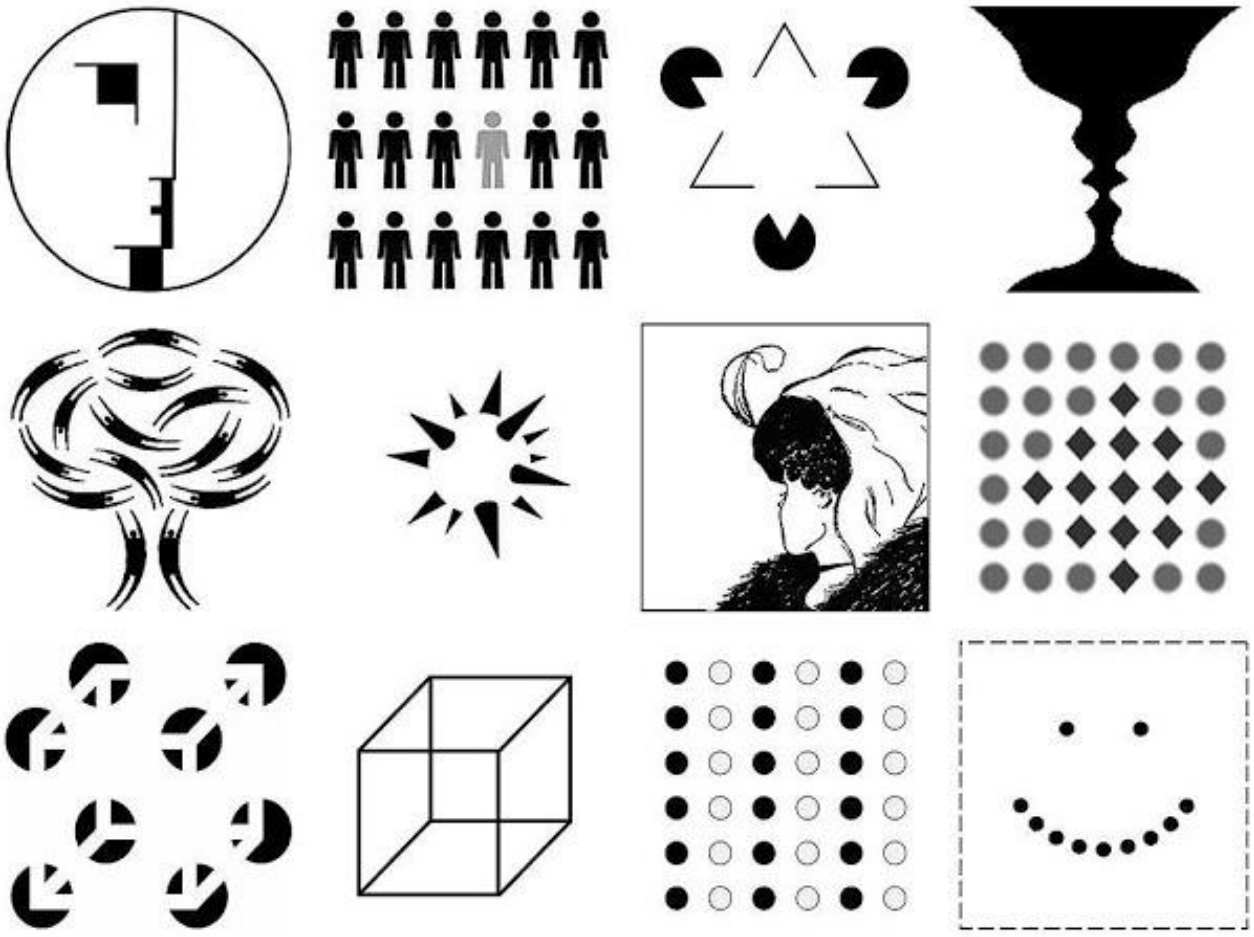




Did you hear a fast-running part among slow-varying main theme?



假作真時真亦假 無為有處有還無



CHALLENGE:

When a machine listens, can it utilize *gestalt psychology* (or, does it need to)?

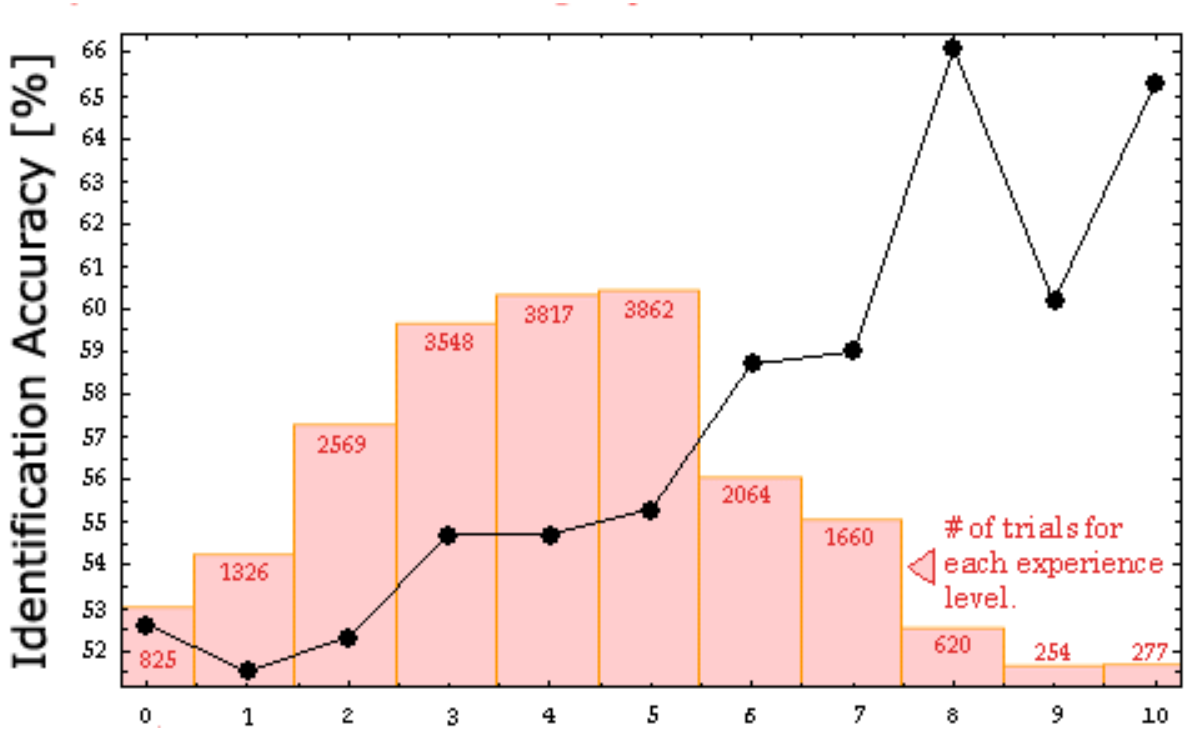
Today's Outline

- Automatic listening
  - 和聲的辨認
  - 節奏的辨認
  - 音樂風格之辨認
    - 類別
    - 作曲家
- 音樂與情緒

A few comments on genre (category) classification

- Task: to tell if a piece of music belongs to pop, jazz, rock, classical, etc.
  - **Relatively easy for humans, hard for computers.**
  - I'd rather hire college students to do it.
- I think the key is rhythm and harmony, not timbre.
  - E.g., jazz vs. classical
  - Many "bark against the wrong tree".
- We can criticize: what for?
  - 音樂難道不能跨界

作曲家的風格：What's different between Mozart (1756-1791) and Haydn (1732-1809)?



Online quiz (2002-2012)

Database = MIDI files for all of their string quartets, total 200+ movements.

visit qq.themefinder.org

Computer recognition of Mozart vs. Haydn's quartet based on note transition probability and Kullback-Leibler divergence

$$D(P^{(1)}||P^{(2)}) = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}^{(1)} \log_2 \frac{P_{i,j}^{(1)}}{P_{i,j}^{(2)}}$$

$$2^{D(P^{\mathrm{U}}||P^{\mathrm{A}})-D(P^{\mathrm{U}}||P^{\mathrm{B}})} = \left(\prod_{t=1}^{L-1} \frac{P^{\mathrm{B}}_{U_t,U_{t+1}}}{P^{\mathrm{A}}_{U_t,U_{t+1}}}\right)^{1/L}$$

$$= \left(\frac{P^{\mathrm{B}}_{U_1}}{P^{\mathrm{A}}_{U_1}} \prod_{t=1}^{L-1} \frac{P^{\mathrm{B}}_{U_t,U_{t+1}}/P^{\mathrm{B}}_{U_t}}{P^{\mathrm{A}}_{U_t,U_{t+1}}/P^{\mathrm{A}}_{U_t}}\right)^{1/L}$$

$$= \left(\frac{p(\mathrm{U}|\mathrm{B})}{p(\mathrm{U}|\mathrm{A})}\right)^{1/L}$$

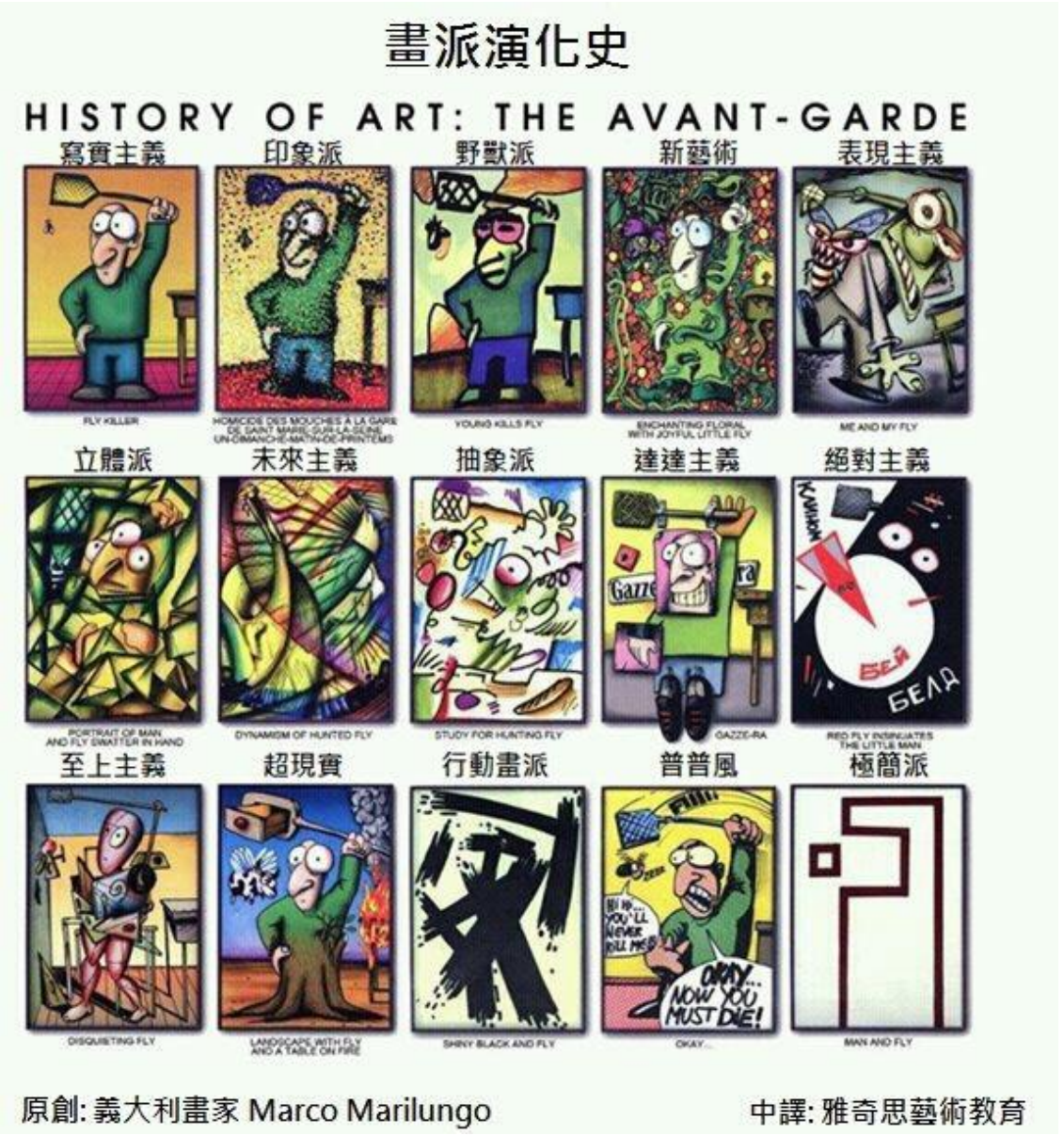
Table 1: Computer identification performance

Part	Mozart	Haydn
Violin I	68.0%	64.2%
Violin II	58.0%	64.2%
Viola	61.0%	53.8%
Cello	57.0%	52.8%

Yi-Wen Liu, Final report for Music254/CS377: *musical informatics* (2002, unpublished) https://ccrma.stanford.edu/~jacobliu/254report.pdf

Discussion: 到底作曲家的風格所在何處？

蕭邦 = 惆悵、憂鬱？
布拉姆斯 = 悶騷？
貝多芬 = 光明正大？
舒曼 = 狂放不羈？
莫札特 = 純真？



Today's Outline

- Automatic listening
  - 和聲的辨認
  - 節奏的辨認
  - 音樂風格之辨認
- 音樂與情緒

楊奕軒：Reduction of emotion onto 2D Cartesian space (2011)

IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, NO. 4, MAY 2011

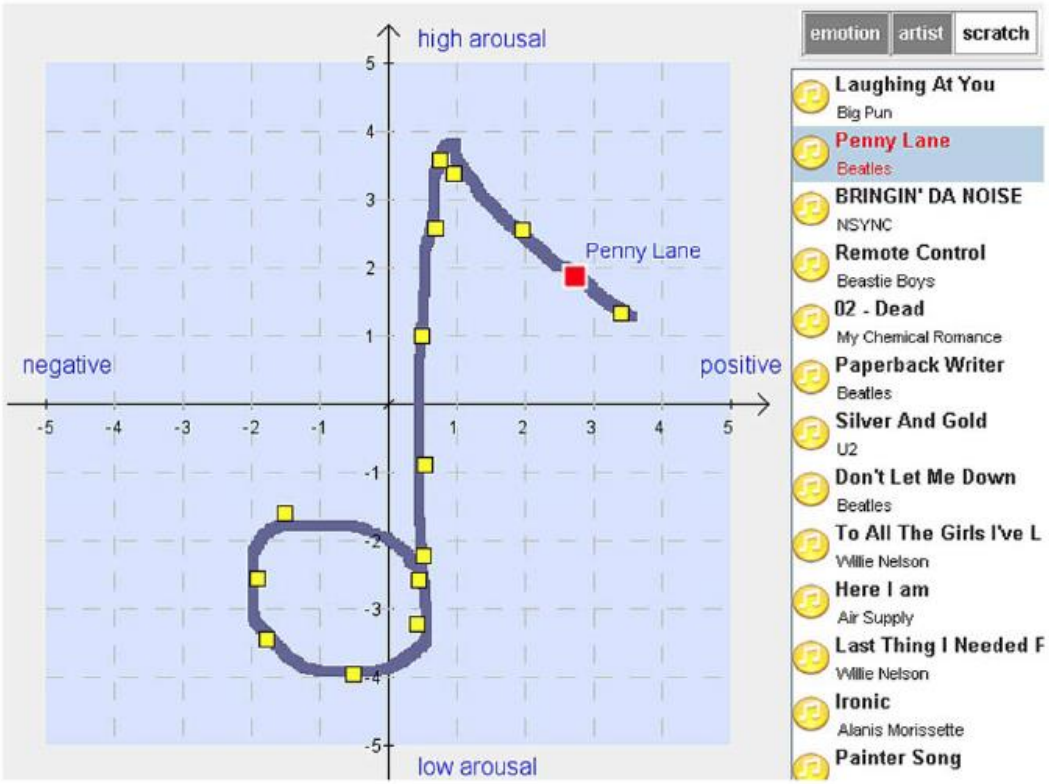


Fig. 1. Associated with the valence and arousal values, each song is represented as a point in the 2-D emotion space [17], where a user can specify points or draw trajectories to retrieve songs of certain emotions [19]–[23].

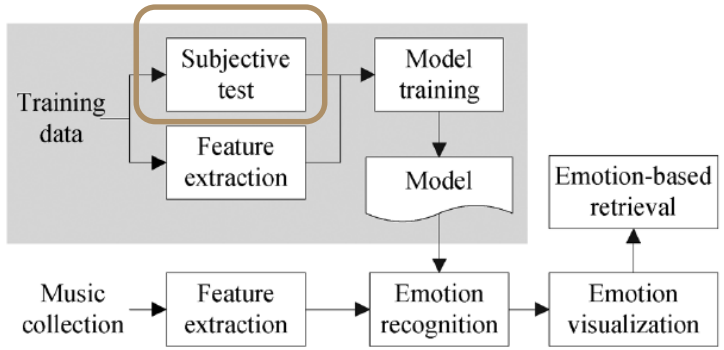
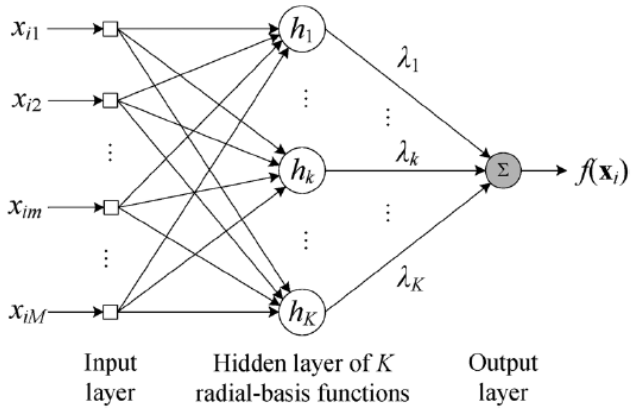


Fig. 5. Schematic diagram of the music emotion recognition system.



Research direction:
從音樂分析與文字之自然語言處理來探討音樂之情緒成份

- 流行樂：歌詞之比對

- 古典樂：表情符號之比對
  - animato, cantabile, con brio, con fuoco, dolce, furioso, leggiero, perdendosi, tranquillo,...
  - Mit guten Humor, wild und lustig

CHALLENGE:

How do we model the **progression** of emotion/feelings in music?

研究音樂之情緒起伏、最大的困難: 難以界定 ground truth

Can ground truth be inferred via physiological measurements?

- EEG/MEG?
- fMRI?
- HRV?
- Breath rates?
- Skin conductance?

How about using a *joystick*

古典音樂中的「**表情記號**」正好提供了很好的素材

- *Animato* animated; lively
- *Cantabile* in a singing style
- *Con Brio* with vigor/spirit
- *Con Fuoco* with energy/ passion
- *Dolce* sweetly
- *Doloroso* sorrowfully
- *Espressivo* expressively
- *Furioso* furious
- *Grandioso* with grandeur
- *Grazioso* gracefully
- *Leggiero* lightly
- *Maestoso* majestically
- *Morendo* dying away
- *Perdendosi* dying away
- *Pesante* heavy
- *Scherzando* humorously
- *Semplice* simple
- *Tranquillo* tranquil

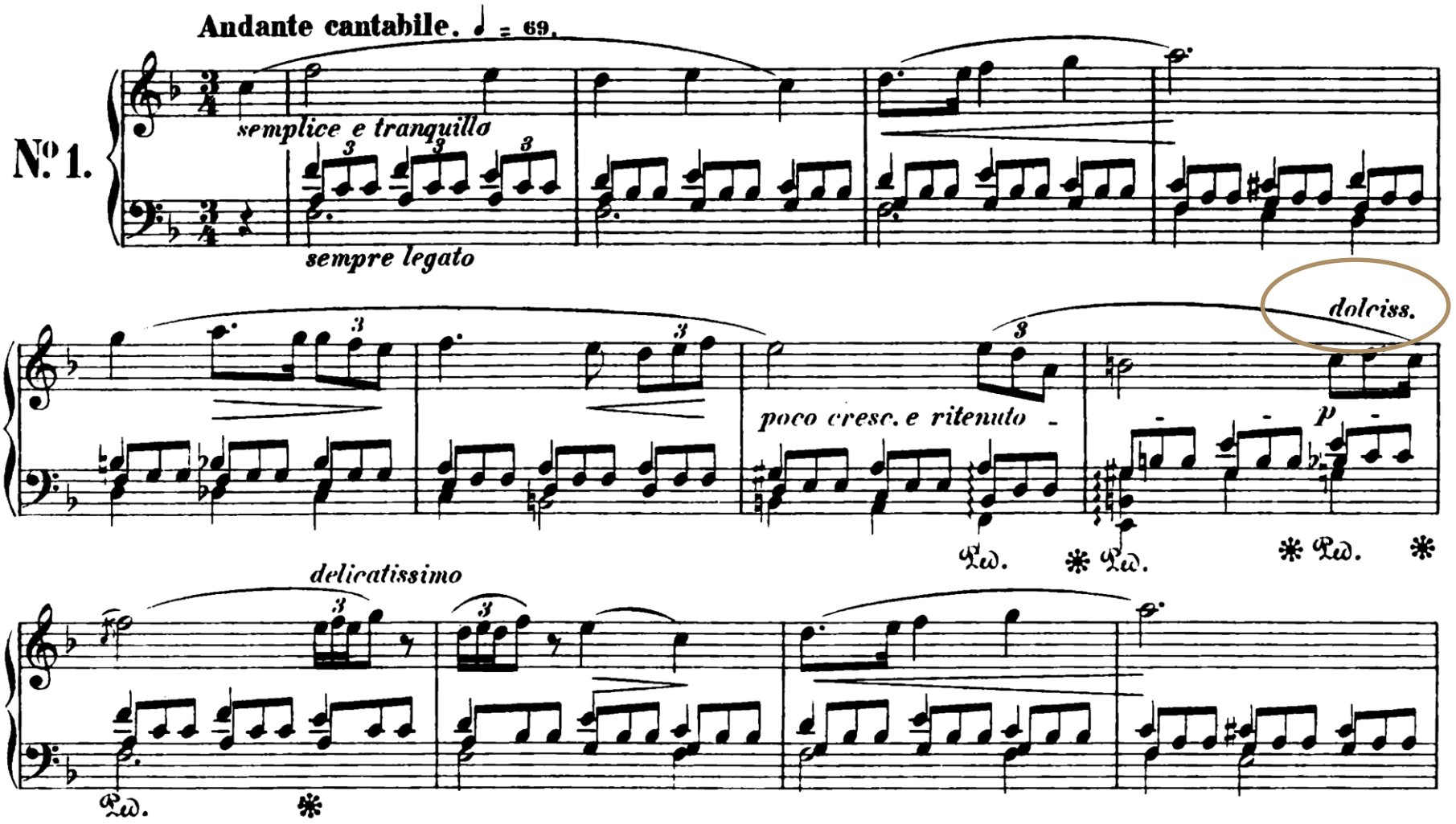
The *Dolcissimo* project
尋找音樂中的甜蜜時刻

- Materials: Rubinstein plays Chopin's 19 Nocturnes

- Dolcissimo/Dolce occurs in 21 measures

Op. 15.

Ferdinand Hiller gewidmet.



- Tempo analysis: 彭玉淮(口琴)、洪偵量(小號)
  - 手工標記所有小節起迄時間
  - 半自動拍點偵測

  => Dolce 出現前，常常有減慢，然後返回原來速度
- Timbre analysis: 陳瑀妮(鋼琴)、楊雅涵(鋼琴)
  - 運用time-reversed novelty function 偵測踏板之深淺度

  => 一般音量分析：太大聲的就排除甜蜜的可能
- Harmonic analysis: 陳政宇(鋼琴)、林洋安(鋼琴)
  - 和絃分析, 因為 dolce 常出現在回到主題樂句時
  - 自動轉譜之嘗試






Statistical analysis on single features shows high probability of false positives

	novelty平均	
dolc\判斷	Y	N
Y	18	2
N	159	180

	novelty標準	
dolc\判斷	Y	N
Y	18	2
N	137	212

	振幅	
dolc\判斷	Y	N
Y	18	3
N	168	170

	相對標準差	
dolc\判斷	Y	N
Y	8	10
N	77	235

	相對平均	
dolc\判斷	Y	N
Y	10	9
N	80	241

How can we compute emotion in music?

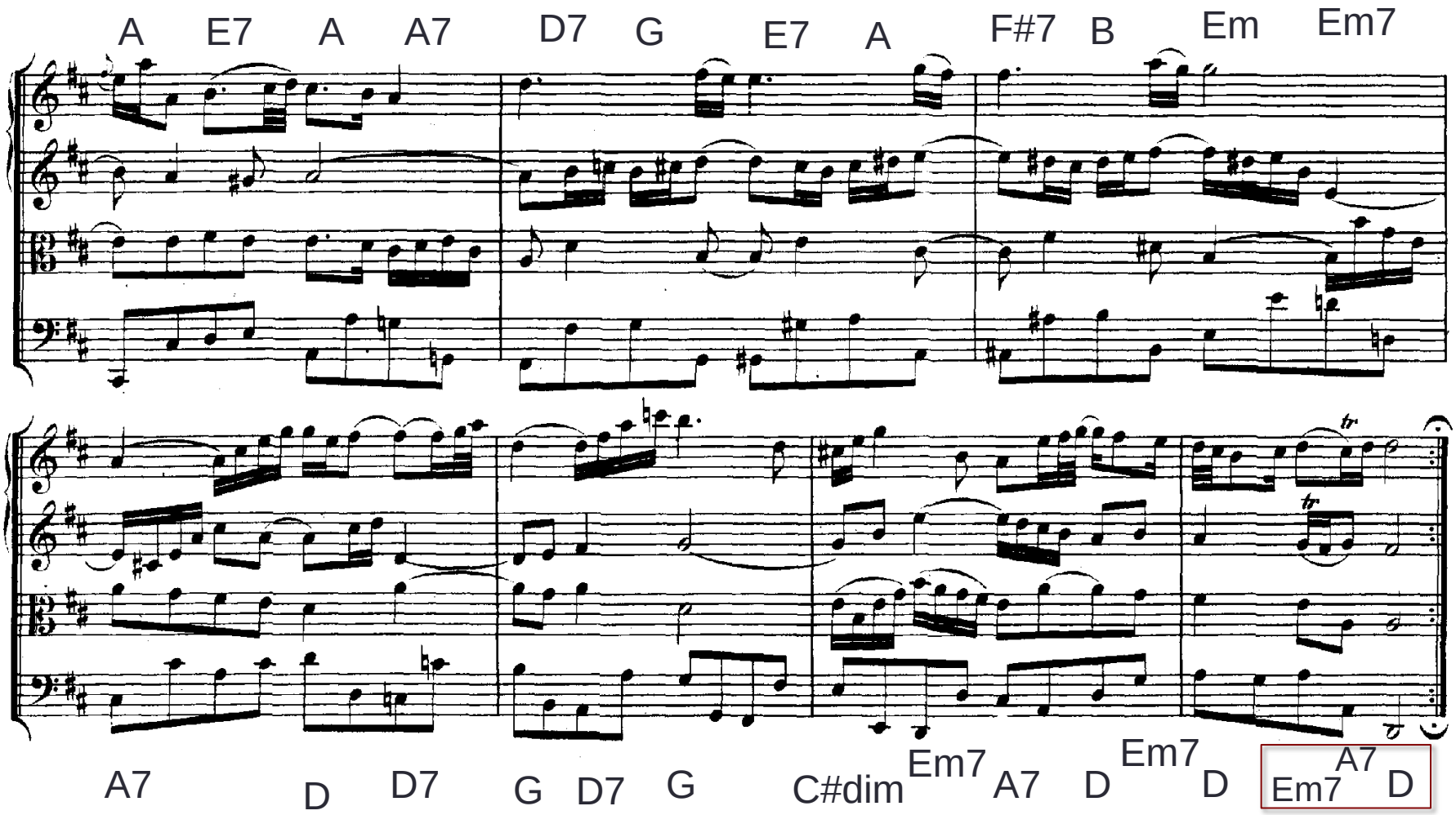
My intended approach

- To quantify tension, surprise and expectancy
  - Of a sequence of features
  - Possibly borrowing “entropy” from information theory

$$H = -\sum_i p_i \log p_i$$

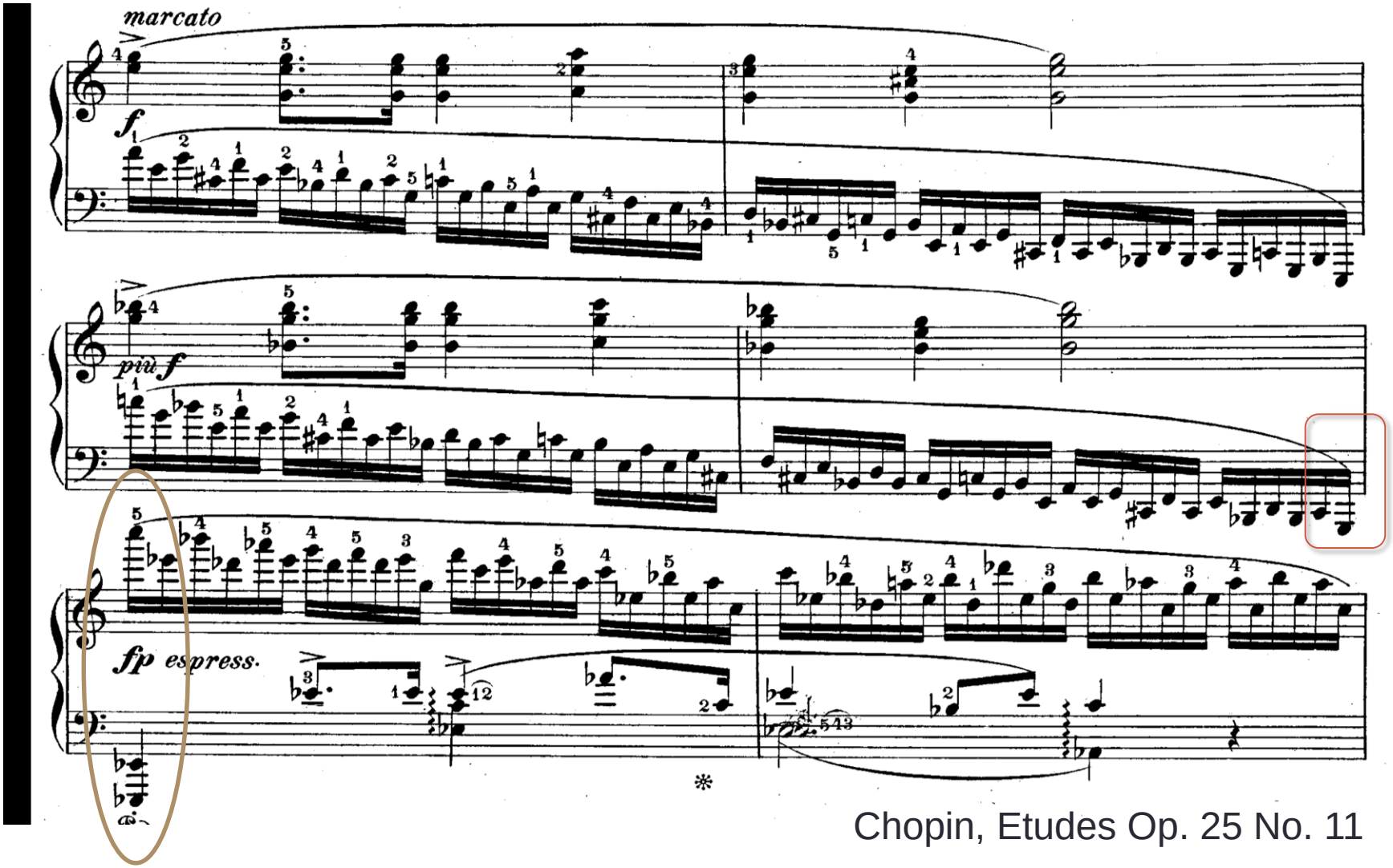
- Then, relate them to the dynamic change of feelings

An example of tension and release



J.S. Bach: “Air in G” in Orchestral Suite No. 3, BWV 1068

An example of release by *surprise*



Chopin, Etudes Op. 25 No. 11

Good music should balance between **predictability** and **surprise**

Final remarks:

Why is it difficult for a machine to **appreciate music?**

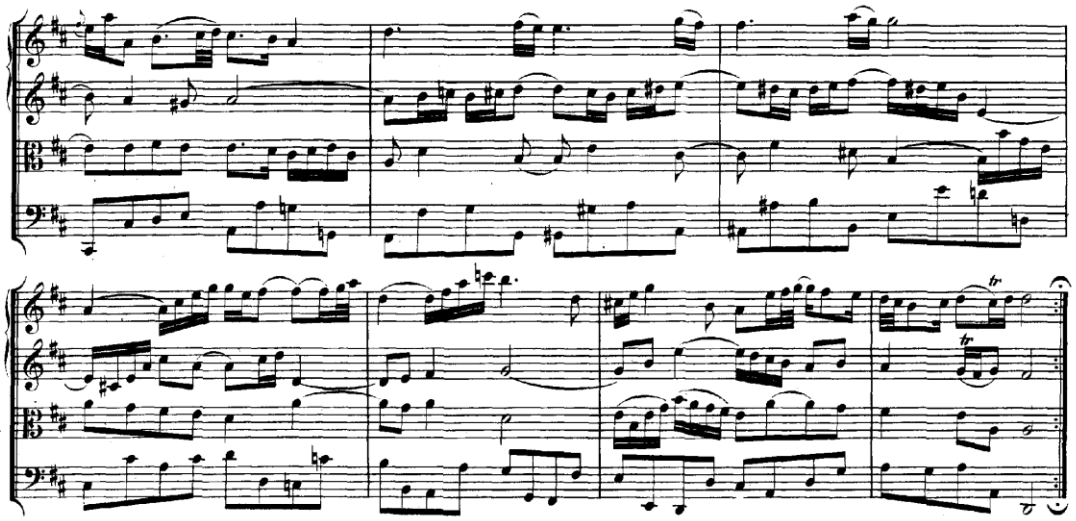
=> Music listening is an active process:
We enjoy expecting/waiting for what's next.

- Even if the music is totally familiar
  - Quite peculiar that we hardly get tired of our favorite music

  - In contrast: do we enjoy
    - listening to the same joke again and again?
    - Watching the same sport event again and again?

- Listening to music is addictive (Dan Levitin, *This is Your Brain on Music*)
  - V Menon and DJ Levitin (2005) "The rewards of music listening: Response and physiological connectivity of the mesolimbic system," NeuroImage 28:175-184.

Thank you!



ADDITIONAL SLIDES

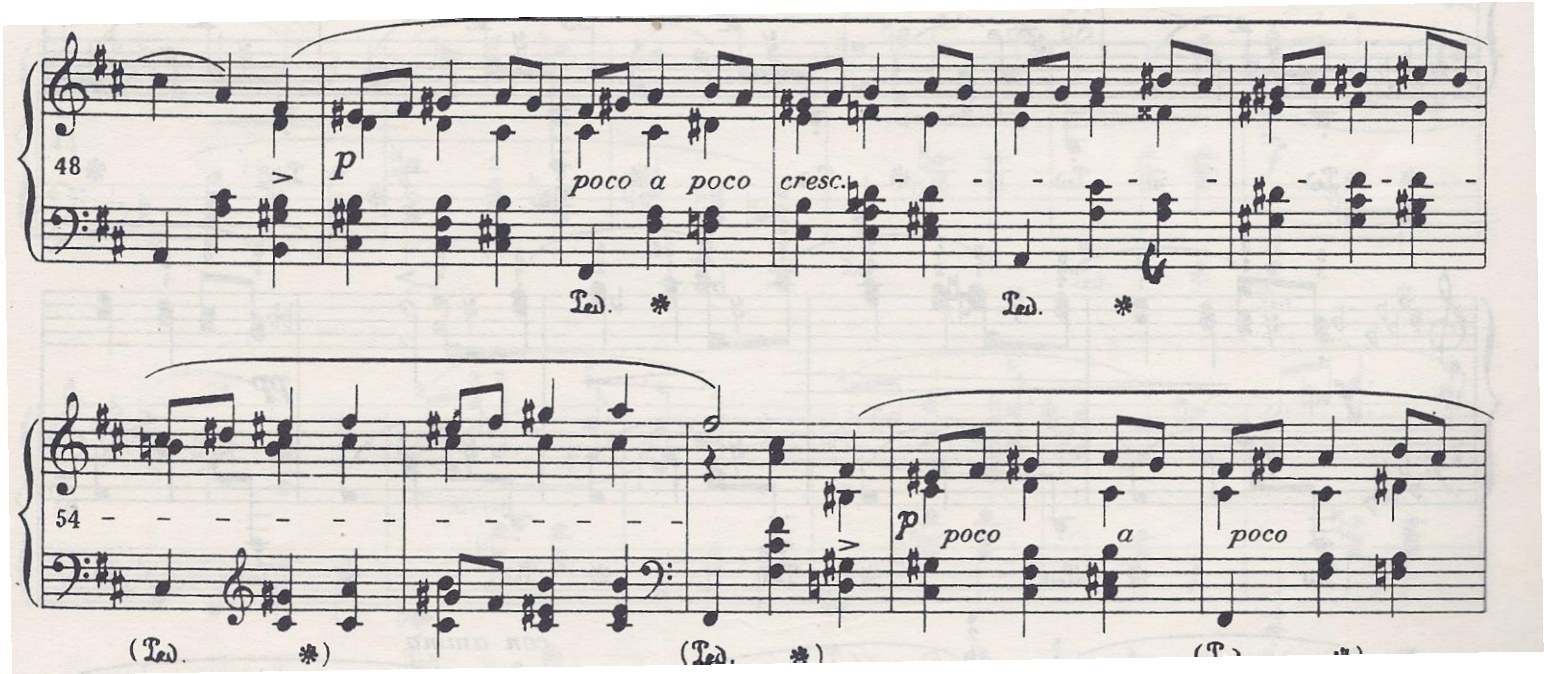
http://tinyurl.com/nthu-demo-mazurka

Petrushka by Stravinsky (1911)

What is the key signature?
(幾拍子)

曲目的風格：以蕭邦之 *Mazurkas* 為例

- 東歐之民族舞蹈
  - 蕭邦的鄉愁
  - 大賽之特別獎項
  - 每一拍都可以是重音



馬厝卡是蕭邦最內在的精神，是他最本能、最個人、心靈最深處的作品。而且蕭邦把節奏和表情完全融合為一，這也就是為何基本上斯拉夫人比較容易理解蕭邦，因為這裡面有民族和語言的天性。

摘自焦元溥《聽見蕭邦》傅聰訪談錄

Mazurkas embedded in Chopin's other works

Concerto No. 1, 1st mov. in E minor

Sonata No. 2, 2nd mov. in E-flat minor

Polonaise No. 5 in F-sharp minor

Andante Spianato and Grand Polonaise Brillante

Waltz No. 10 in B minor