CS4347

Sound and Music Computing

L3: Introduction to Music Representation, Analysis and Transcription

Wang Ye

wangye@comp.nus.edu.sg

Office: AS6-04-08

1

Topics to Cover (*selective approach*)

Part A: The Core

- Introduction
- Review of DFT, Audio Representation, and Machine Learning
- Music Representation, Analysis and Transcription
- Automatic Music Transcription (AMT)
- Automatic Speech Recognition (ASR)
- Generative Models for Text-to-Speech (TTS) & Singing Voice Synthesis (SVS)

Midterm break

Part B: The Breadth

- Speech Assessment
- Real-time Audio Programming in Practice
- Automatic Music Generation
- Singing voice processing
- Music production audio effects
- Synthesis of sound & music – a DSP approach
- Project presentations/demo

2

Today's topics

1) Recap what we have learnt in the past 2 weeks
2) Music representations – music notation demystified
3) Music analysis (e.g. transcription)

Is music just a form of entertainment?

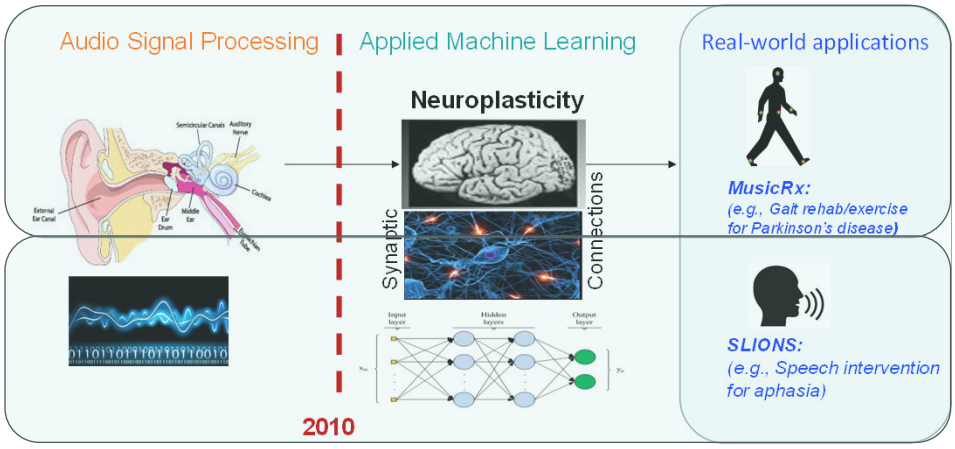
My two key research questions in the past decade
motivated by NMT (RAS & MIT)

1) How to enable users to discover their preferred music that satisfies clinical requirements for gait rehabilitation and exercise via music **search**, **recommendation** and **generation**?

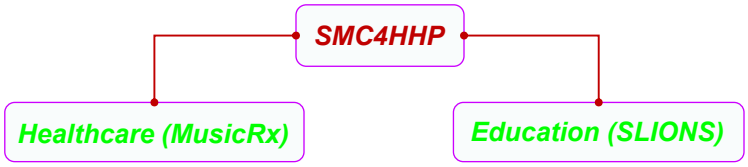
2) How to leverage on the relationship between speech and singing to build applications for speech intervention?

NMT: Neurologic Music Therapy
RAS: Rhythmic Auditory Stimulation
MIT: Melodic Intonation Therapy

Programmatic Research Agenda:
Sound and Music Computing for Human Health and Potential (SMC4HHP)



SMC4HHP's relevance to the local community: Human Health and Potential (HHP)



My Keynote Speech @ ISMIR2014 :
Sound and Music Computing for Exercise and Rehabilitation

Li, Z., Xiang, Q., Hockman, J., Yang, J., Yi, Y., Fujinaga, I., & Wang, Y. (2010, October). **A music search engine for therapeutic gait training**, in Proceedings of the 18th ACM international conference on Multimedia (pp. 627-630). ACM.

Ng, W.F., Zhou, Y., Tan, P., and Wang, Y., "**Using the MOGCLASS in group music therapy with individuals with muscular dystrophy: A pilot study**," Music and Medicine., vol. 4, no. 4, pp. 199-204, 2012.

iRACE: Music-enhanced exercise and motor rehabilitation (ACM MM2014)

Validating an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) for Parkinson's Disease

Shenggao Zhu[1], Robert J Ellis[2], Gottfried Schlaug[3], Yee Sien Ng[4], and Ye Wang[1,2]
[1]NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore
[2]School of Computing, National University of Singapore, Singapore
[3]Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, USA
[4]Department of Rehabilitation Medicine, Singapore General Hospital, Singapore
shenggaozhu@nus.edu.sg; {ellis, wangye}@comp.nus.edu.sg;
gschlaug@bidmc.harvard.edu; ng.yee.sien@sgh.com.sg

ABSTRACT

Movement disorders such as Parkinson's disease (PD) will affect a rapidly growing segment of the population as society continues to age. Rhythmic Auditory Cueing (RAC) is a well-supported evidence-based intervention for the treatment of gait impairments in PD. RAC interventions have not been widely adopted, however, due to limitations in access to personnel, technological, and financial resources. To help "scale up" RAC for wider distribution, we have developed an iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) mobile application to deliver RAC and assess motor performance in PD patients. The touchscreen of the mobile device

1. INTRODUCTION

A 2007 review of published prevalence studies projects that the number of individuals over the age of 50 with Parkinson's disease (PD) living in 15 of the world's most populous nations will double between 2005 (roughly 4.4 million) and 2030 (roughly 9.0 million) [11]. Although administration of carbidopa/levodopa remains the "gold-standard" treatment for motor impairments in PD [32], gait parameters such as cadence, stride length, and velocity remain significantly reduced in PD patients relative to age-matched healthy controls, even when patients are tested during the "peak" effect of medication [25]. Together, these concerns motivate the

Music Magic!

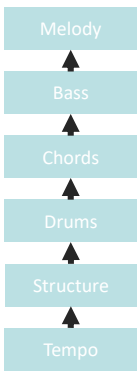
iRACE

An iOS-based Rhythmic Auditory Cueing Evaluation (iRACE) for Parkinson's Disease

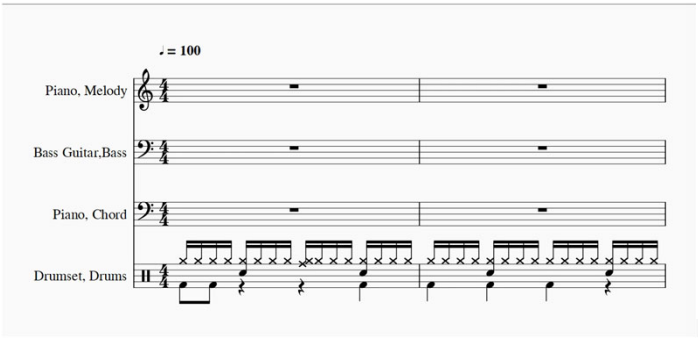
9

9

MusicRx-C: from search engine, recommender systems to automatic music generation to explore music which satisfies both ***clinical requirements*** and ***user preference***.



Decomposition of music



10

10

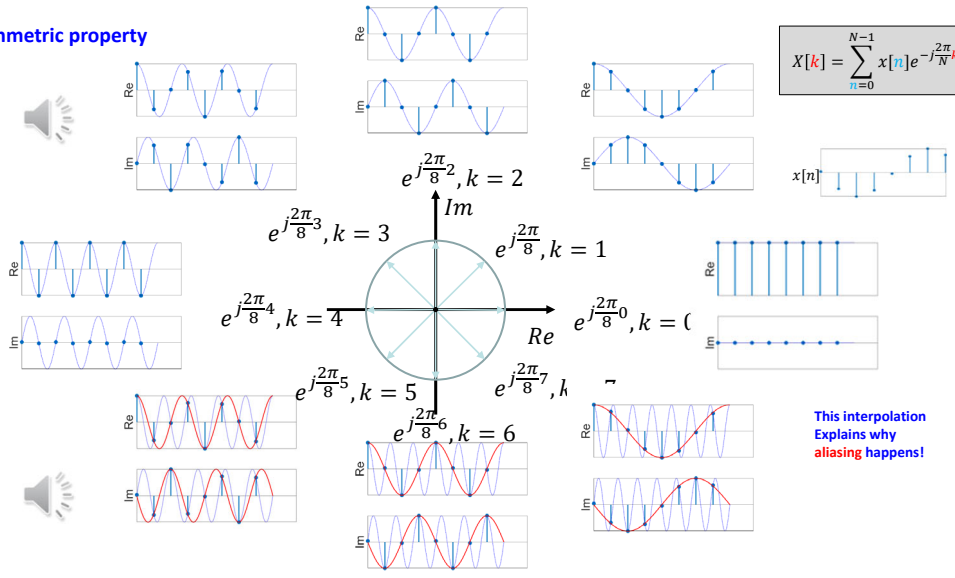
2024

CocoLyricist:

an AI enabled co-creation songwriting system for post-stroke brain training

Symmetric property

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}$$

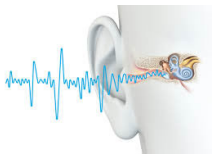


This interpolation Explains why aliasing happens!

Today's topics

1) Recap what we have learnt in the past 2 weeks
2) Music representations – music notation demystified
3) Music analysis (e.g. transcription)

CS4347/CS5647 is about Bridging and Connecting the Dots



Analog

Digital

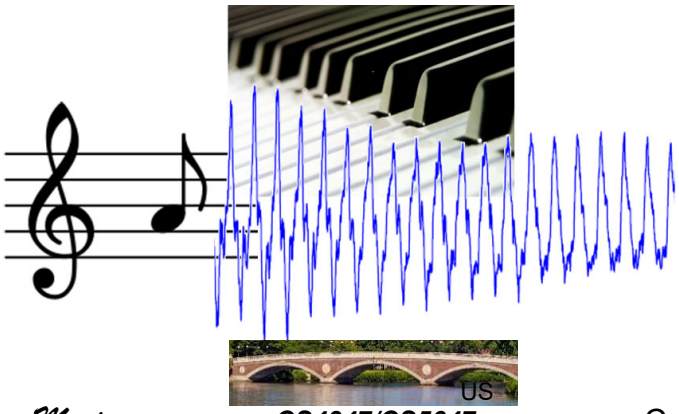
Nyquist Sampling theorem

Discrete Fourier Transform (DFT)

Time

Frequency

Music Representations



Music — ***CS4347/CS5647*** — Computing

ASR (CS5241) AMT (CS4347)

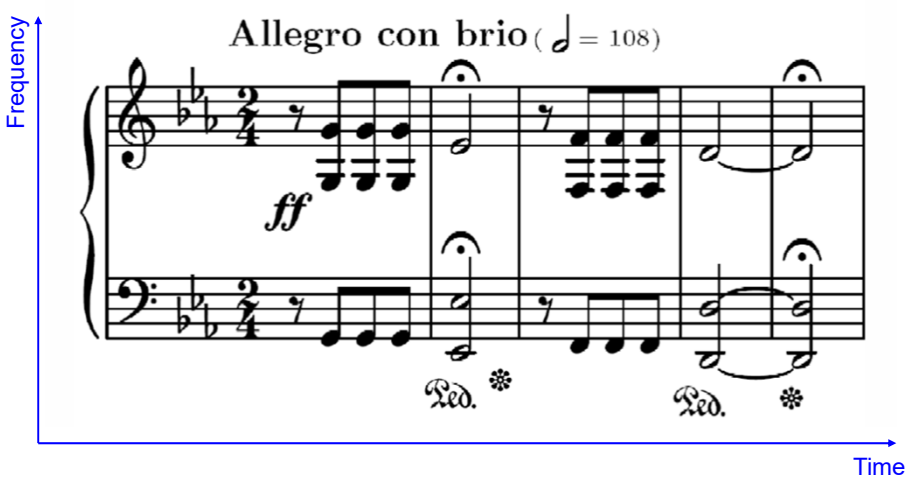
What is music?

The Concise Oxford Dictionary defines music as "the art of combining vocal or instrumental sounds (or both) to produce beauty of form, harmony, and expression of emotion"

Why do we listen to music?

15

Sheet Music Representations



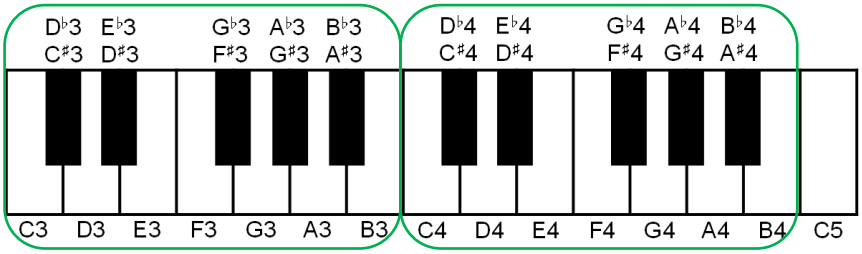
Textual description of music – the language of musicians!

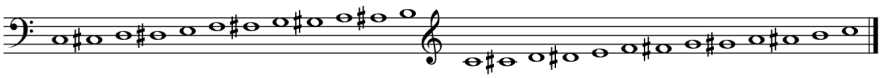
Let's try to demystify it with the spectrogram representation!

16

Sheet Music Representations

Twelve-tone equal-tempered scale in western music

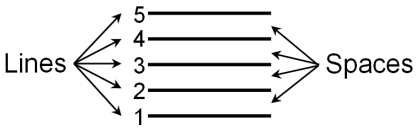


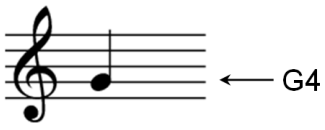


Compare symbols in the sheet music to words and phonemes in speech!

17

Sheet Music Representations



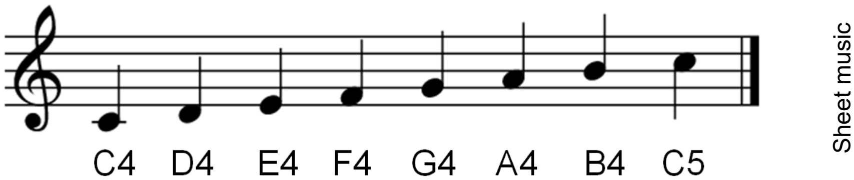


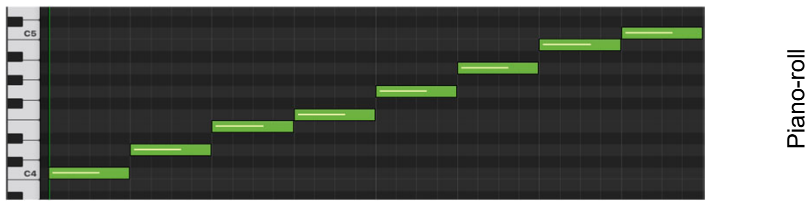
Compare symbols in the sheet music to words and phonemes in speech!

18

Symbolic Music Representations (*pitch*)

C-major scale starting with C4 and ending with C5



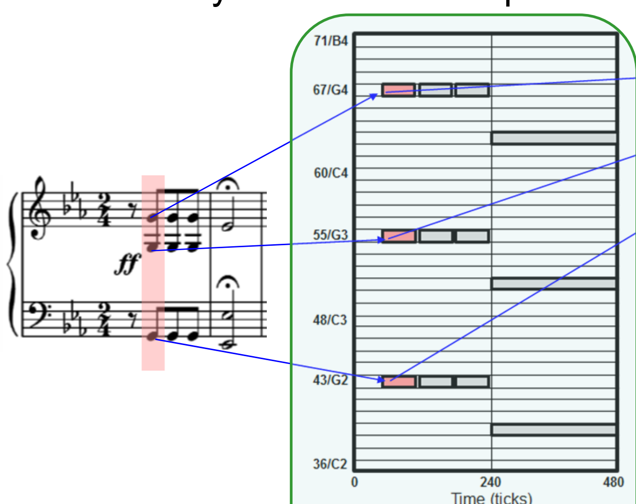


Pitch:

- Defined as the frequency of a sine wave that is matched to the perceived sound by a human listener

19

Symbolic Music Representations



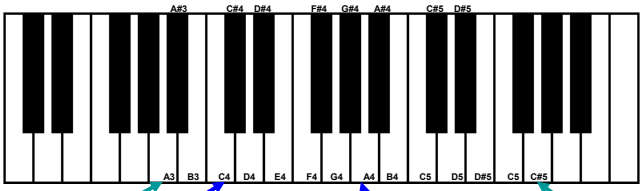
Time (Ticks)	Message	Channel	Note Number	Velocity
60	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	63	100
0	NOTE ON	2	51	100
0	NOTE ON	2	39	100
240	NOTE OFF	1	63	0
0	NOTE OFF	2	51	0
0	NOTE OFF	2	39	0

Sheet music **Piano-roll** **MIDI**

20

Mapping Pitch to Notes, to Keyboard Position

- Keyboard players usually take bearings from “Middle C” (C4)
- Orchestral Players tune to the A above “Middle C” (A4)



NOTE	A3	A#3	B3	C4	C#4	D4	D#4	E4	F4	F#4	G4	G#4	A4	A#4	B4	C5	C#5	D5	D#5	C5	C#5	...	...
PITCH	220	233	247	262	277	294	311	330	349	370	392	415	440	466	493	523	554	587	622	659	698	...	...
MIDI No	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	...	...

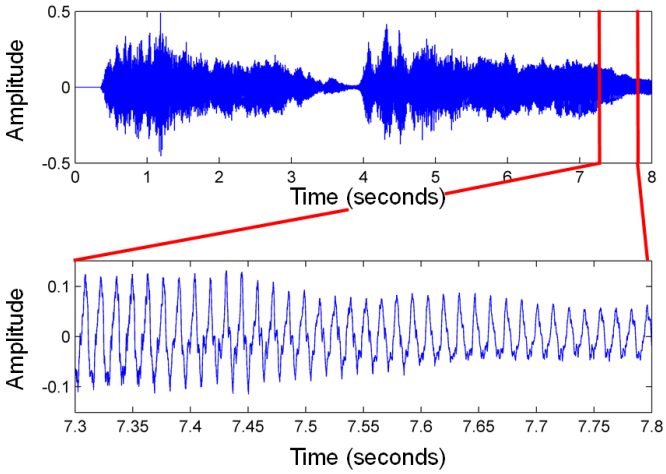
Note Name (With ANSI index), Perceived Pitch (rounded to nearest Hertz) & MIDI No.

NOTE	A	A#	B	C	C#	D	D#	E	F	F#	G	G#	A	A#	B	C	C#	D	D#	C	C#	...	...

Note Name

Live demo from Dr. Torin Hopkins

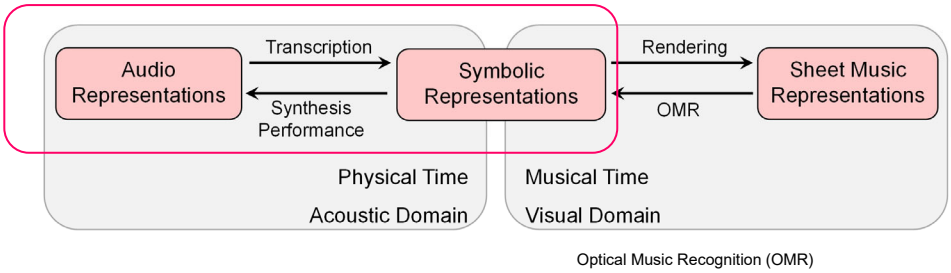
Audio Representations



23

Connecting the Dots

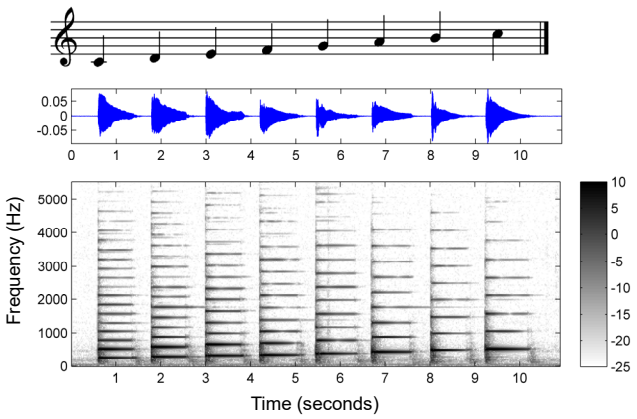
MIDI and Piano-roll



Optical Music Recognition (OMR)

24

Music Representations



Sheet music

Waveform

Spectrogram

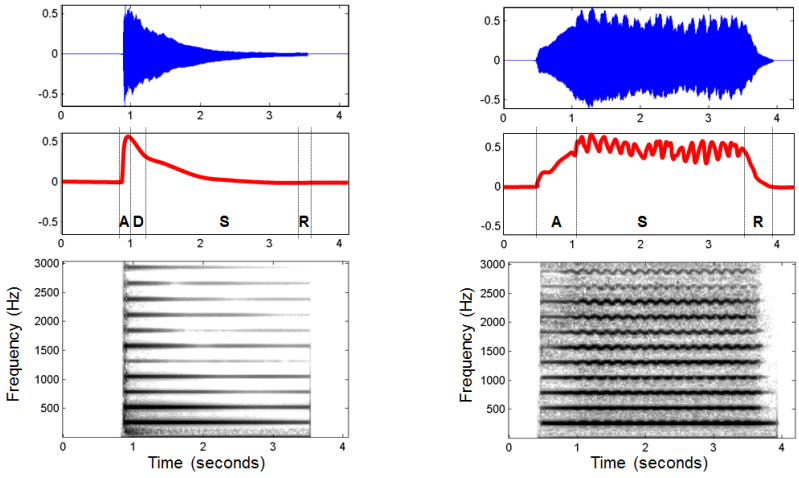
How do you convert the TF representation to a piano roll?

Meinard Müller, Fundamentals of Music Processing
ISBN: 978-3-319-21944-8, Springer, 2015

25

Music Representations

3 basic parameters for an AMT: ***pitch, onset and offset***



Waveform

Contour

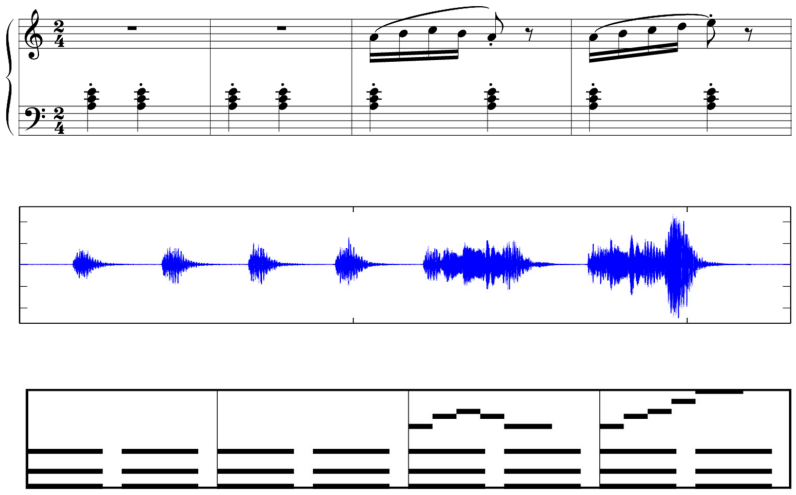
Spectrogram

Meinard Müller, Fundamentals of Music Processing
ISBN: 978-3-319-21944-8, Springer, 2015

$V_{F \cdot T} \approx W_{F \cdot k} \cdot H_{k \cdot T}$

26

Music Representations



Sheet music

Waveform

Piano-roll

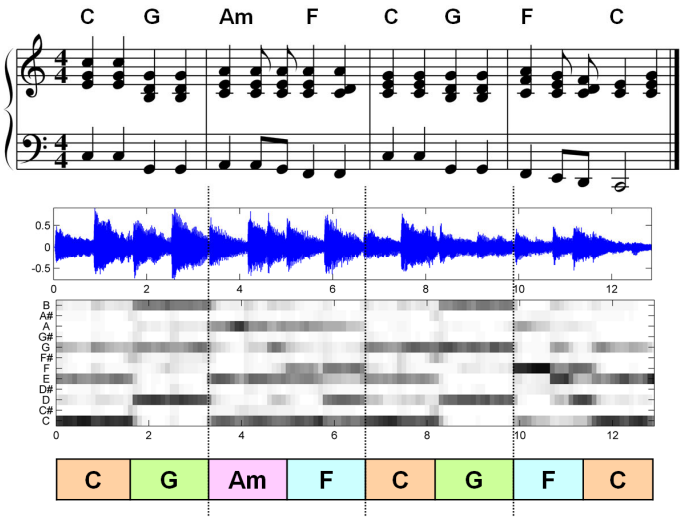
Meinard Müller, Fundamentals of Music Processing
ISBN: 978-3-319-21944-8, Springer, 2015

Non-negative matrix factorization (NMF)

$V_{F \cdot T} \approx W_{F \cdot k} \cdot H_{k \cdot T}$

27

Music Representations



Sheet music

Waveform

Chroma

Chord recognition

Meinard Müller, Fundamentals of Music Processing
ISBN: 978-3-319-21944-8, Springer, 2015

28

Music Representations (MIDI format)

MIDI (**M**usical **I**nstrument **D**igital **I**nterface) is a technical **standard** that describes 3 components:

- Protocol
- Digital Interface
- Connectors

29

What is the purpose of MIDI?

- In 1970 each digital synthesizer manufacturer was creating its own system to link the input device (usually keyboard) to the sounds
- In 1983 the full MIDI 1.0 detailed specification was released

MIDI is a standardized way of communication between synthesizers, computers, samplers and other musical equipment.

30

What is MIDI data?

MIDI does **NOT** carry sound; it carries **event messages**.

What type of events ?

- The start of a note
- The end of a note
- Changing the parameters of the synthesizer
- Other controls and configuration messages

31

What is a MIDI file?

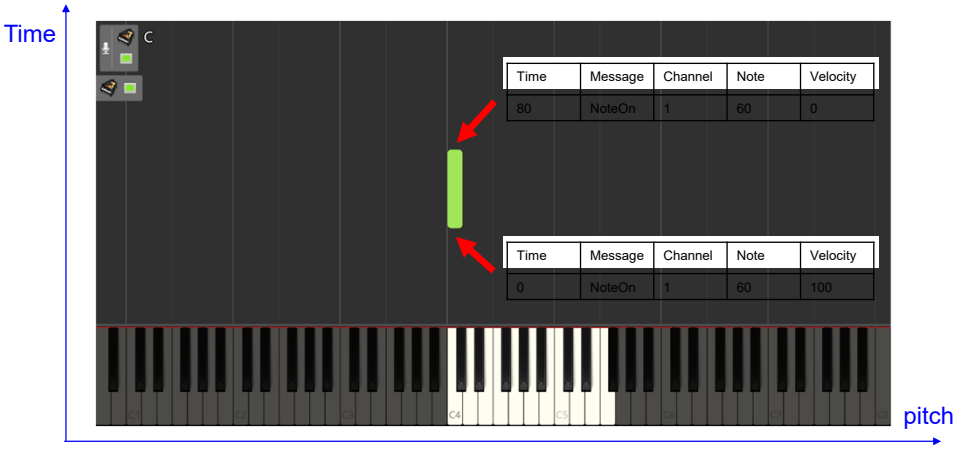
A **MIDI** recording, or **MIDI** file can be thought of as being an enhanced score:

- Exact dynamics (quantified between 0-127)
- Exact timings
- Other information (eg: sustain pedal)

A **MIDI** file does **NOT** contain the music. It contains just "instructions" to play the music, thus the file size of a MIDI recording is greatly reduced.

32

A closer look into MIDI: note messages as an example

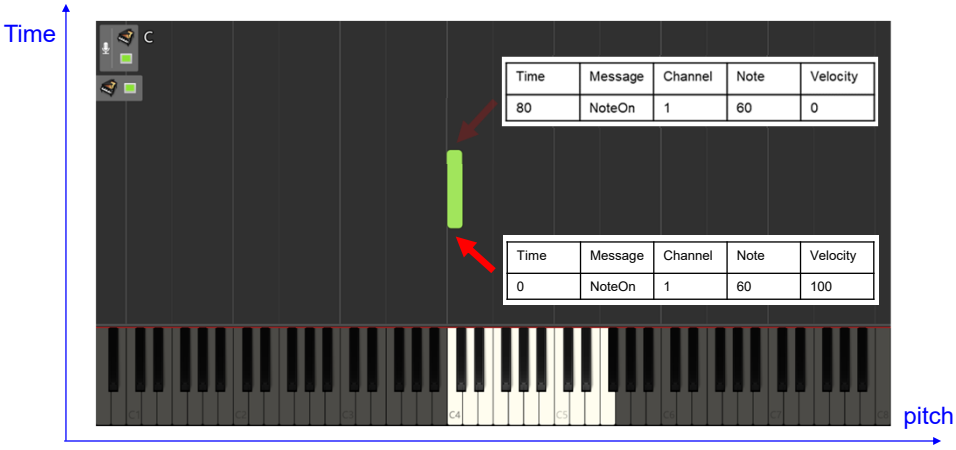


A note is represented as two MIDI messages

One for the onset and one for the offset

33

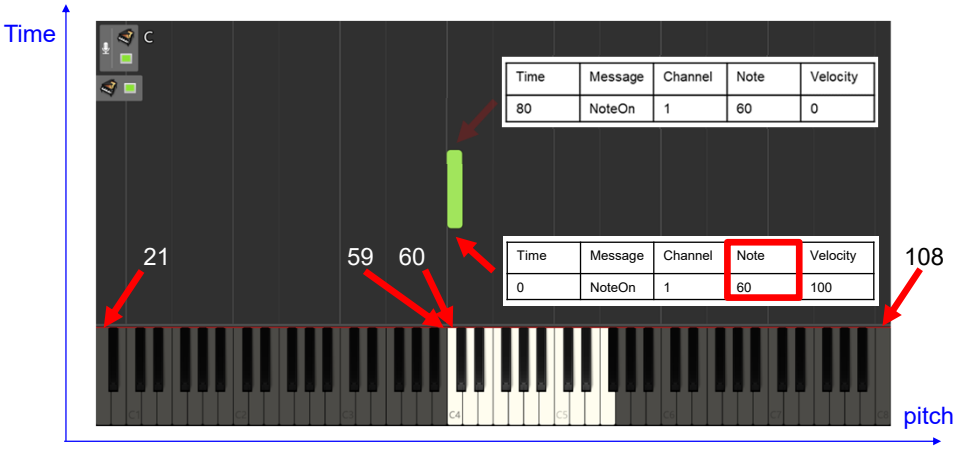
A closer look into MIDI: note messages as an example



A note on message with positive velocity marks the start of the note

34

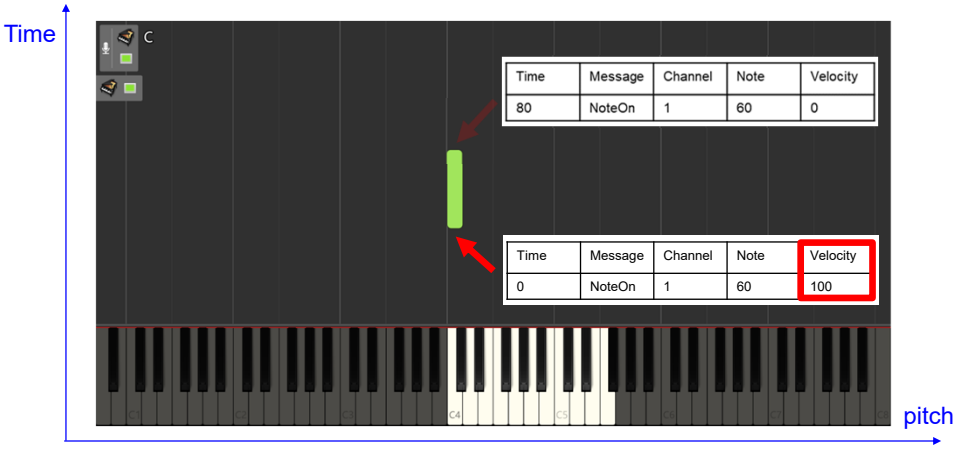
A closer look into MIDI: note messages as an example



The note attribute determines the pitch of the note
For example, 60 is C4, the middle C
59 is B3, 21 is A0, and 108 is C8
MIDI can express a wider range of notes than the piano!

35

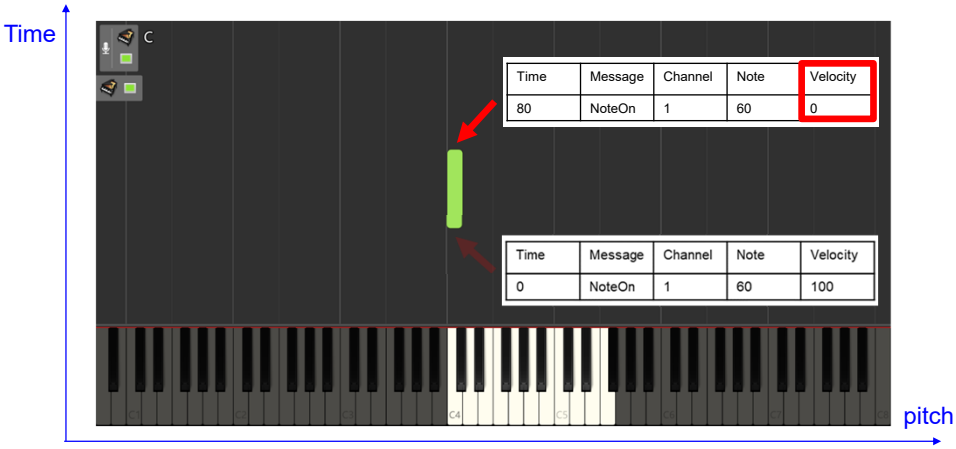
A closer look into MIDI: note messages as an example



The velocity attribute determines the dynamic or loudness of the note
The higher the velocity, the higher the loudness

36

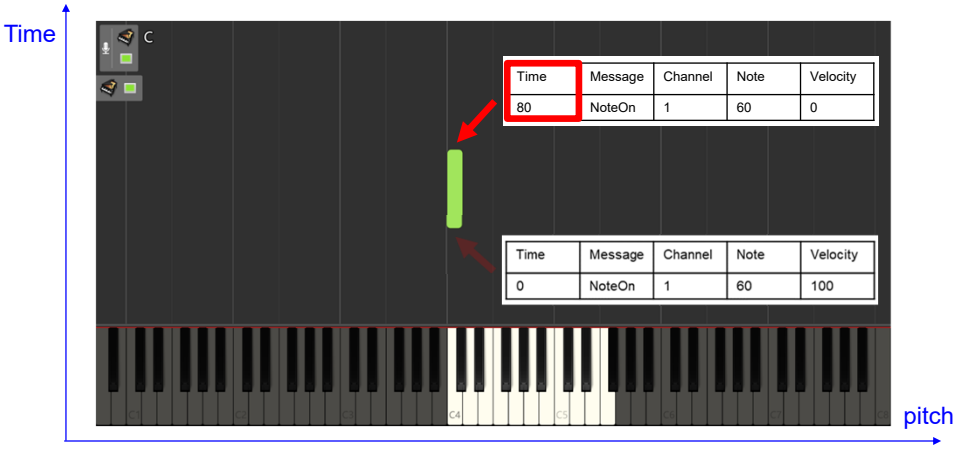
A closer look into MIDI: note messages as an example



A note on message with 0 velocity (or a note off message) marks the end of the note

37

A closer look into MIDI: note messages as an example



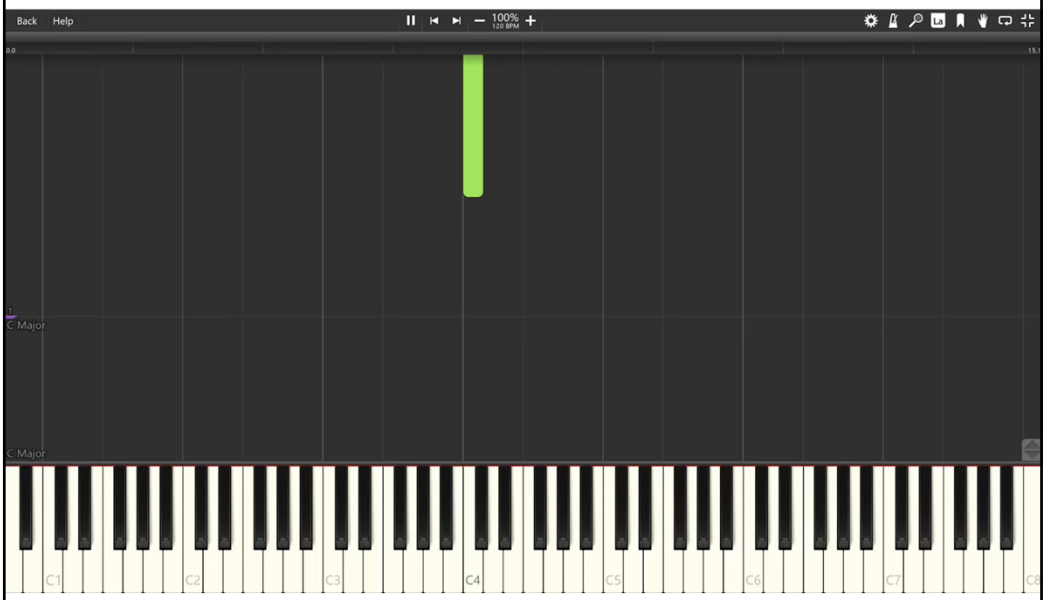
Time is the delta time after the previous message, measured in ticks

The length of a tick is determined by the *tick per beat* and *beat per second* **meta messages**

Hence, the time of a tick in seconds can be calculated after two conversions

38

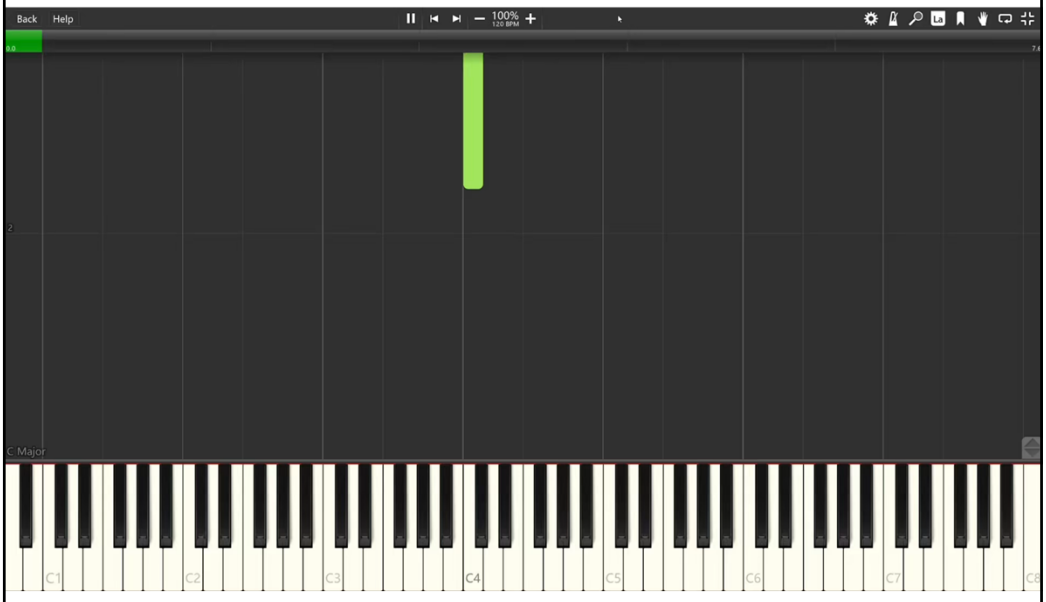
MIDI Notes (60-64)



39

39

MIDI Message: Velocity



40

40

How can we listen to a MIDI file ?

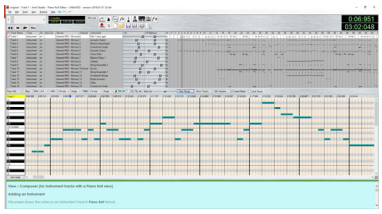
To listen to a MIDI file we need one or more pieces of software or hardware that are able to play the MIDI “instructions”:

- Synthesizers
- Samplers
- Virtual Instruments

Which one is from an acoustic piano?

Pros and Cons of MIDI

- Easy to edit
  - Change notes
  - Change tempo
  - Change instruments
  - Add tracks
  - …
- Free software: GarageBand, Anvil Studio, …
- Python libraries: mido, pretty-midi, …



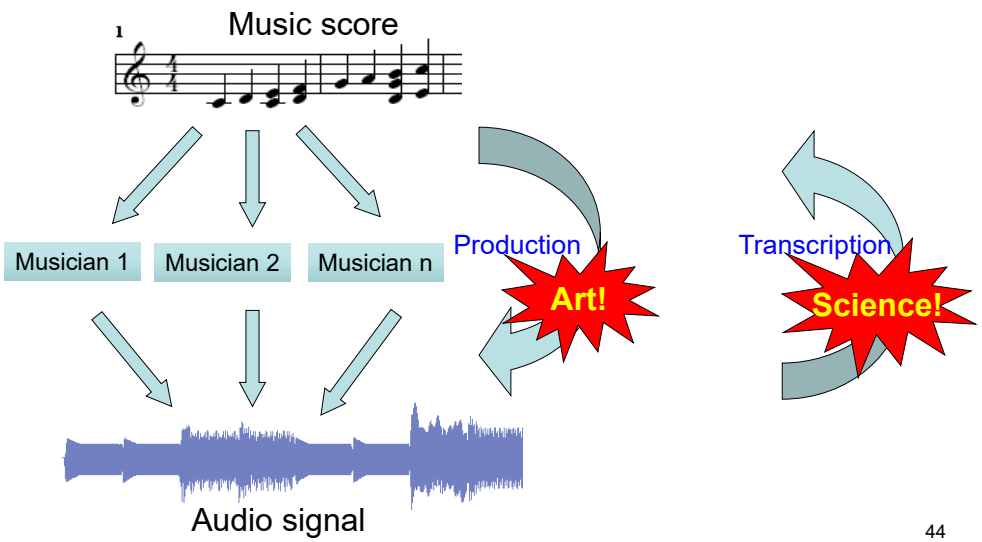
- Inconsistent sound effects
- No vocals
- Less realistic

Today's topics

1) Recall what we have learnt last week
2) Music representations – music notation demystified
3) Music analysis (e.g. transcription)

43

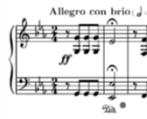
Music Production and Transcription



44

Piano music transcription (examples)

Input audio



Mozart Sonata K. 331, 3rd movement

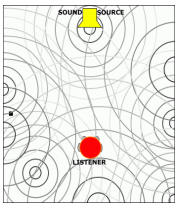
Synthesized audio using transcribed MIDI

Curtis Hawthorne, Erich Elsen, Jialin Song, Adam Roberts, Ian Simon, Colin Raffel, Jesse Engel, Sageev Oore, Douglas Eck, Proceedings of the 19th International Society for Music Information Retrieval Conference, ISMIR 2018, Paris, France, 2018

45

Acoustics

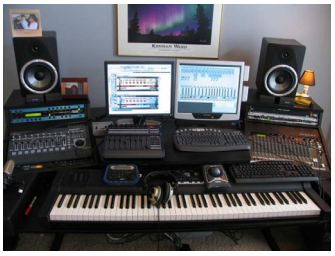
- Reverberation (echo) is caused by reflected waves
  - exaggerated in enclosed rooms
  - amount depends on material, room size



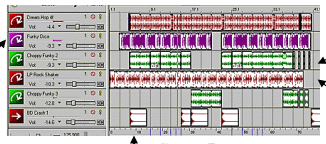
- Recording studios typically have little reverb
  - Acoustically treated walls absorb / diffuse sounds.

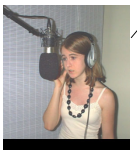
46

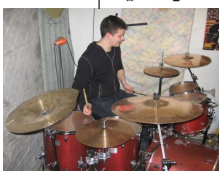
Recording

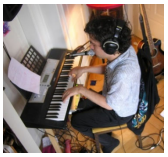


- Studio Engineer









47

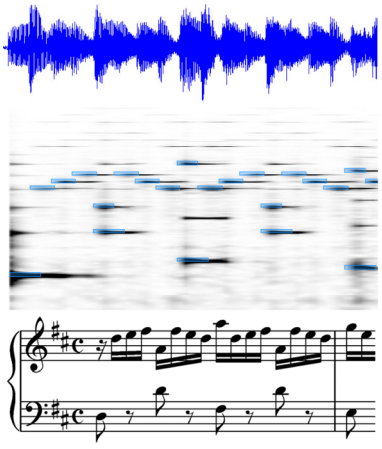
Mixing

- Balance the relative volume, frequency, and dynamical content of a number of sound sources
  - Drums
  - Bass
  - Guitar 1
  - Guitar 2
  - Vocals
  - End product
- Usually, digital WAVE files in 24-bit, 44.1kHz or higher

48

Automatic Music Transcription (AMT)

Automatic music transcription (AMT): the process of converting an acoustic musical signal into some form of music notation (e.g. staff notation, MIDI file, piano-roll,…)



Fundamental (and open) problem in music information research

Applications:

- Search/annotation of musical information
- Interactive music systems
- Music education
- Music production
- Digital/computational musicology

In comparison to ASR, the progress of AMT is quite limited. Why?

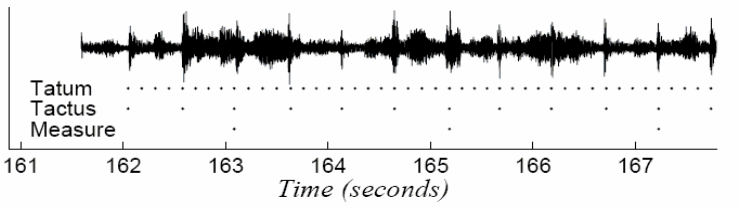
Music Transcription: What to transcribe?

Musical Terminology	Matching Physical property
Onset/Offset	Time measured in samples
Pitch	Fundamental Frequency F0
Loudness	Logarithmic scale (simplifies)
Timbre/Tone Color	Difficult, no single physical attribute describes it completely

- Vibrato and other expressive features
- Lyrics

Temporal Structure

- Goal is to **automatically find musically relevant time information** (e.g. start of single notes) in the recorded signal



51

Temporal Structure

Examples:

Percussive sources:

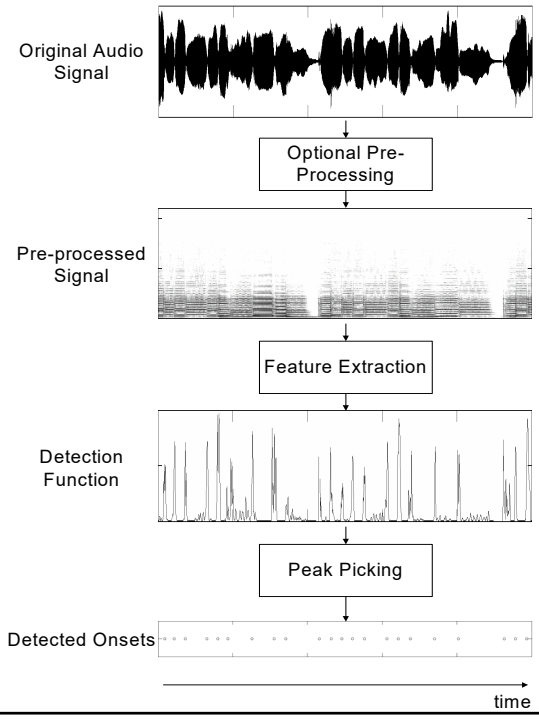
Drums, Cymbals; rather simple to find by looking at signal energy

Harmonic sources:

Singing voice, violin, trumpet; much harder to derive temporal structure from harmonic changes in signal

Conclusion:
We need more sophisticated methods for harmonic signals

52



53

53

Temporal Structure


Source: http://www.cs.tut.fi/~klap/iiro/meter/index.html

54

54

Pitch and harmonic structure

Fundamental frequency:

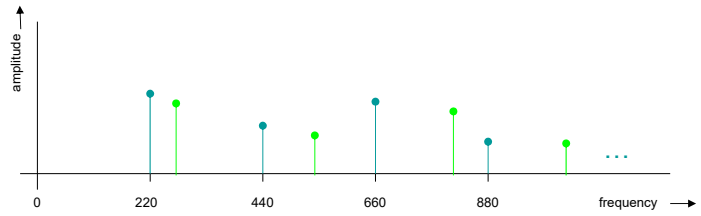
- Is the corresponding physical term
- Only defined for periodic or nearly periodic musical signals

Commonly used pitch detection algorithms include YIN and pYIN.

Pitch and harmonic structure

A challenging problem: multi-pitch detection (polyphonic sound)

Example: A3 (220 Hz) and C4 (262 Hz)



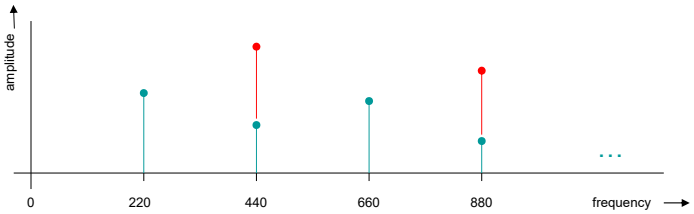
Relation to audio source separation task!

Demystify the cocktail party effect!

Pitch and harmonic structure

A challenging problem: multi-pitch detection (polyphonic sound)

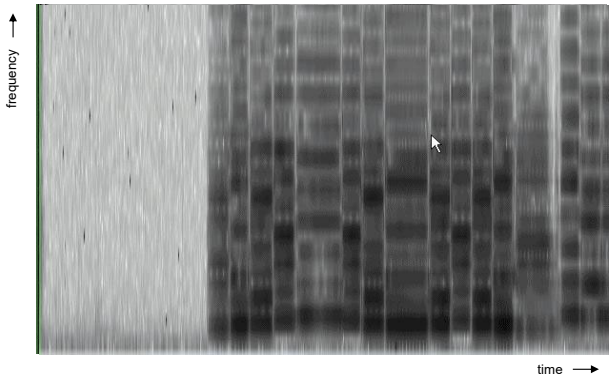
Problem: A3 (220 Hz) and A4 (440 Hz)
The harmonics overlap and such it is very difficult to separate the two tones



57

Pitch and harmonic structure

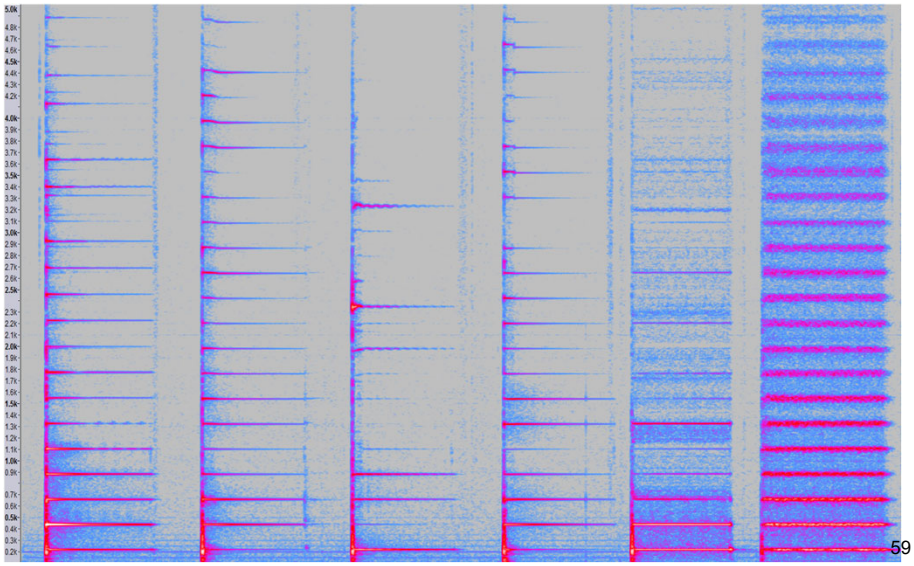
One major problem of the FFT is the Heisenberg-uncertainty



58

Timbre

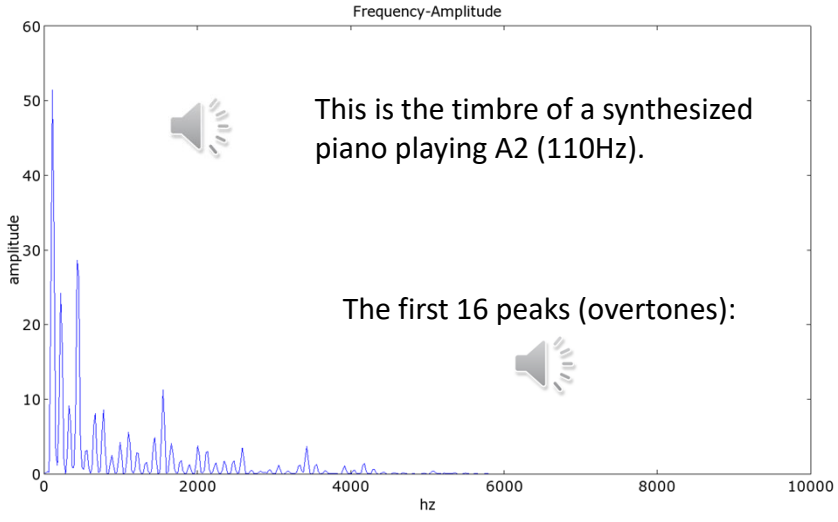
piano harpsichord vibraphone guitar organ string orch.



59

59

One Piano, Many Timbres



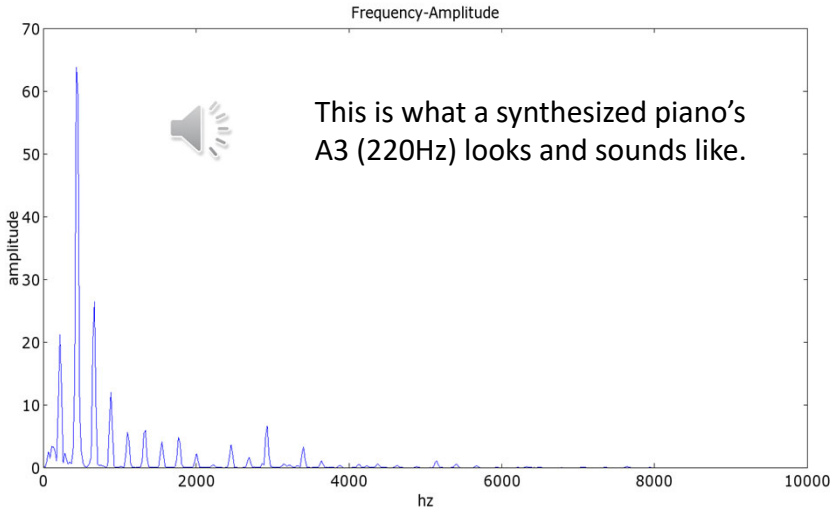
This is the timbre of a synthesized piano playing A2 (110Hz).

The first 16 peaks (overtones):

60

60

One Piano, Many Timbres

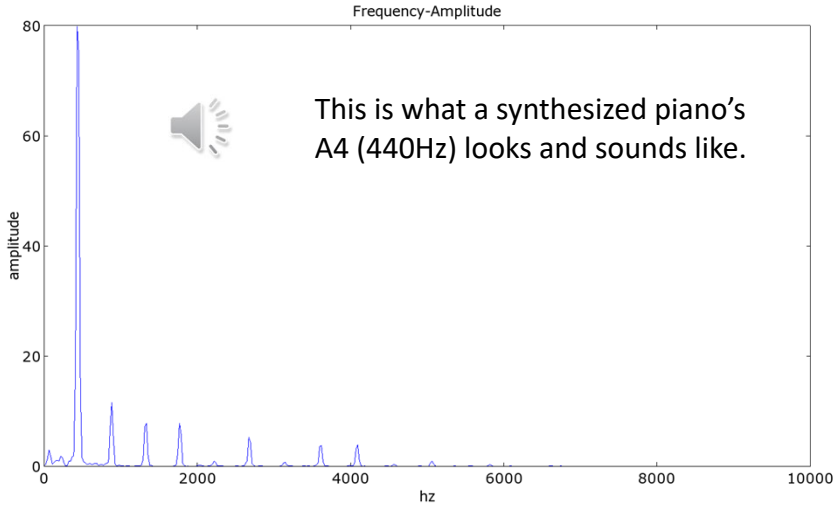


This is what a synthesized piano's A3 (220Hz) looks and sounds like.

61

61

One Piano, Many Timbres

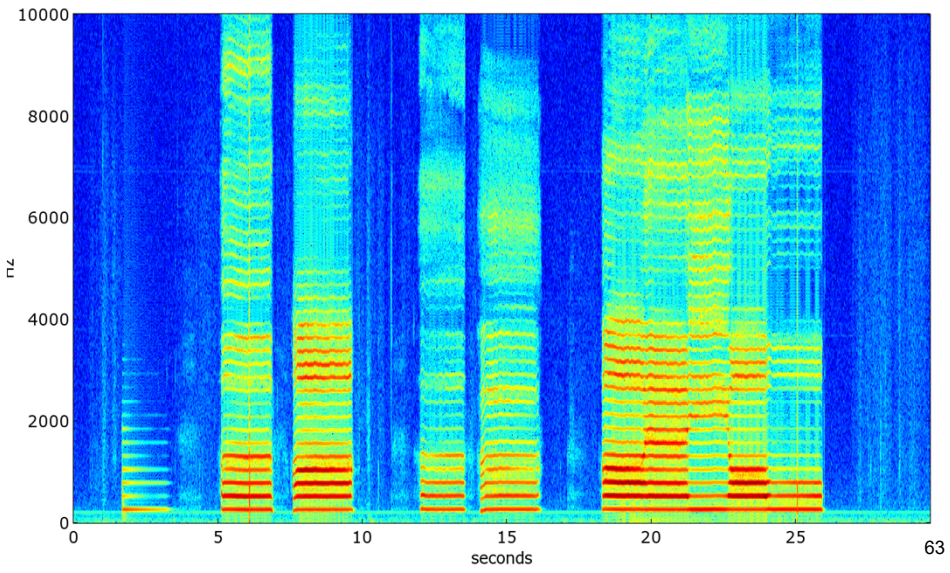


This is what a synthesized piano's A4 (440Hz) looks and sounds like.

62

62

Speech: Voices & Vowels



63

63

Summary of Today's Lecture

1) Recap what we have learnt in the past 2 weeks
2) Music representations – music notation demystified
3) Music analysis (e.g. transcription)

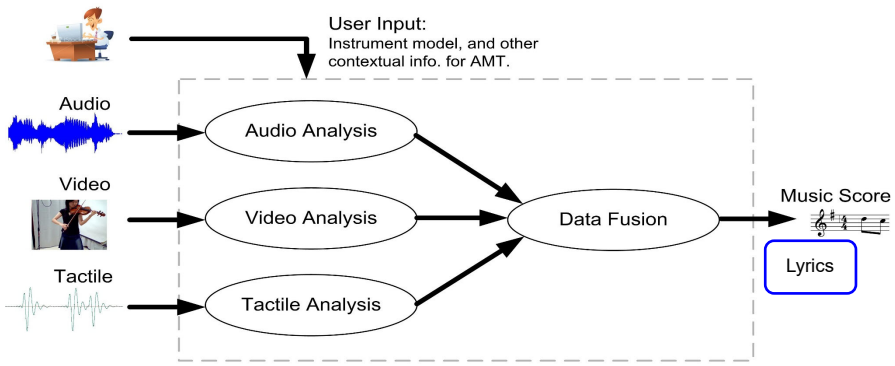
A few words about the group project (we will release the requirements in the 3rd week).

64

64

Multimedia fusion to improve transcription

This has been a default group project since 2022.



Wang & Zhang, "Application-Specific Music Transcription for Tutoring", IEEE Multimedia, Vo.15, No.3, July-September 2008

Gu, Ou, Ong, & Wang, "MM-ALT: A Multimodal Automatic Lyric Transcription System", ACM Multimedia 2022

https://smcnus.comp.nus.edu.sg/archive/pdf/2022_ACM_MM_MM-ALT.pdf

65

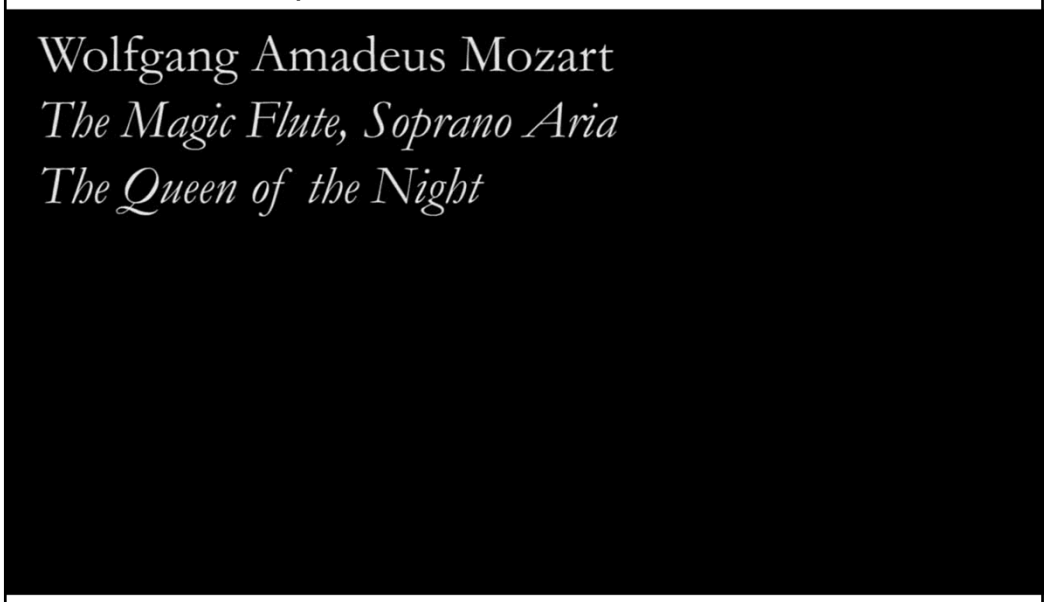
65

Synchronization of Audio and Lyrics: Visualization

66

66

Expressive visualization of music



67