Deep Learning for Music Analysis and Generation

Introduction

Course

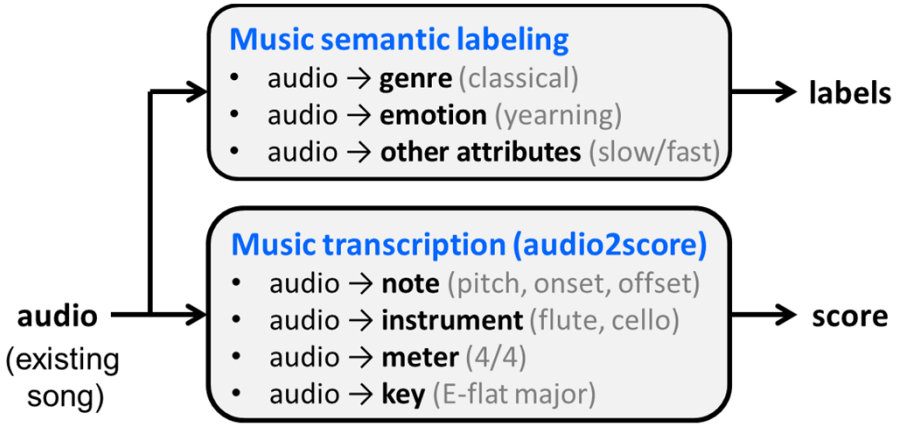
Yi-Hsuan Yang Ph.D.
yhyangtw@ntu.edu.tw

Outline

- **Music & AI**
- The course
- Brief intro to Music Information Retrieval (MIR)

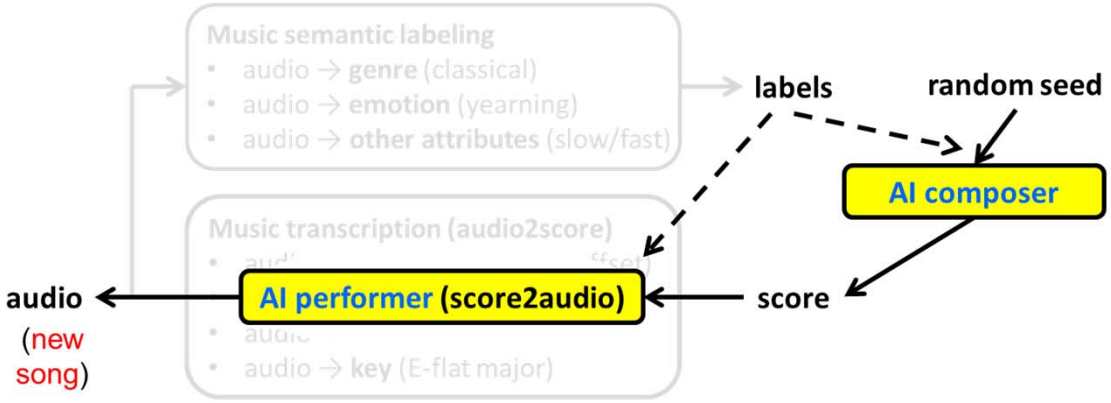
Music AI; or Music Information Research (MIR)

- **Music analysis**



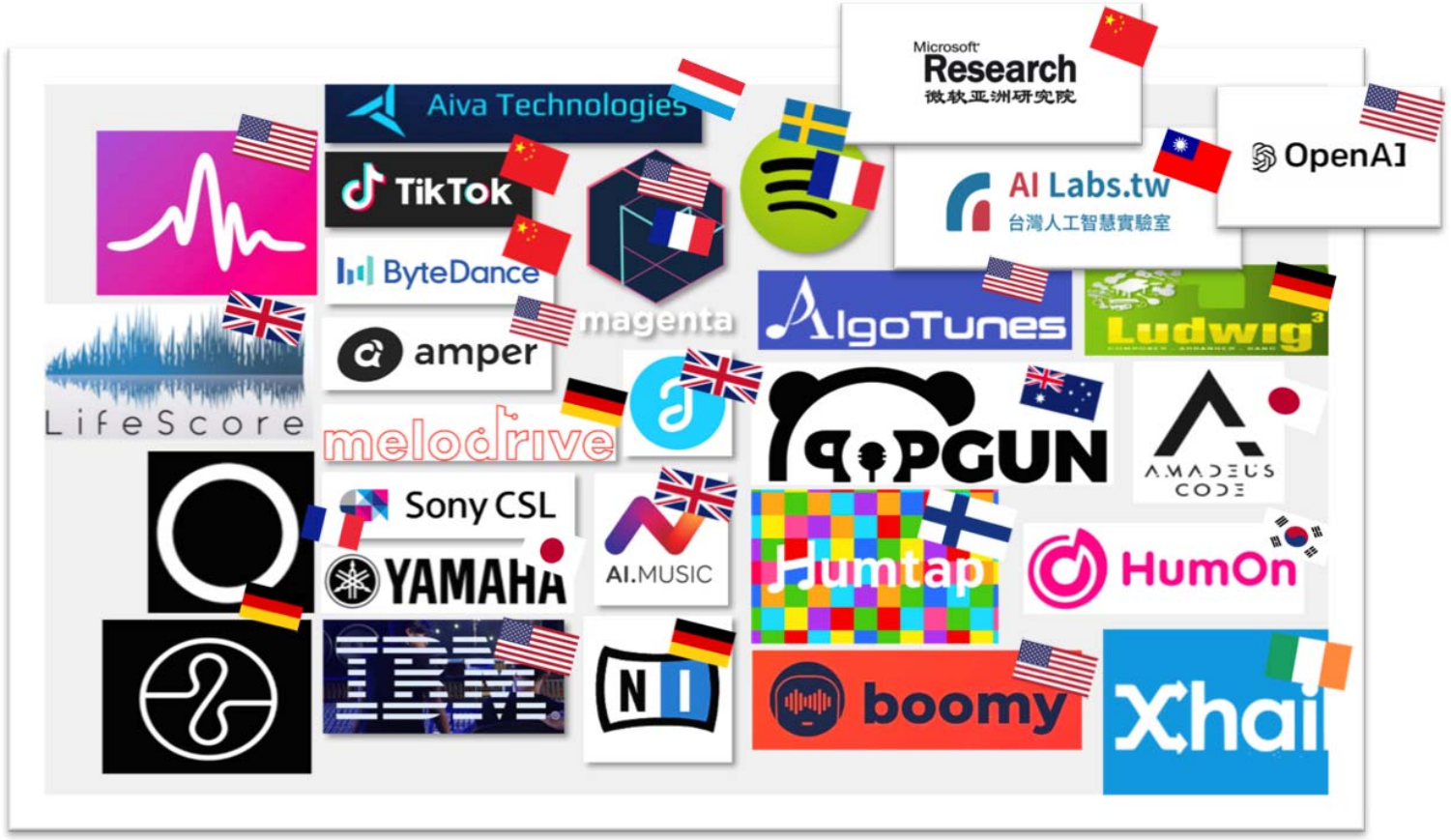
  - music understanding
  - music search
  - music recommendation

- **Music generation**



  - MIDI generation
  - audio generation
  - MIDI-to-audio generation

Global Interest in Music AI



(**Slide from Rujing Huang, Bob L. T. Sturm, and Andre Holzapfel**, "De-centering the West: East Asian Philosophies and the Ethics of Applying Artificial Intelligence to Music," ISMIR 2021)

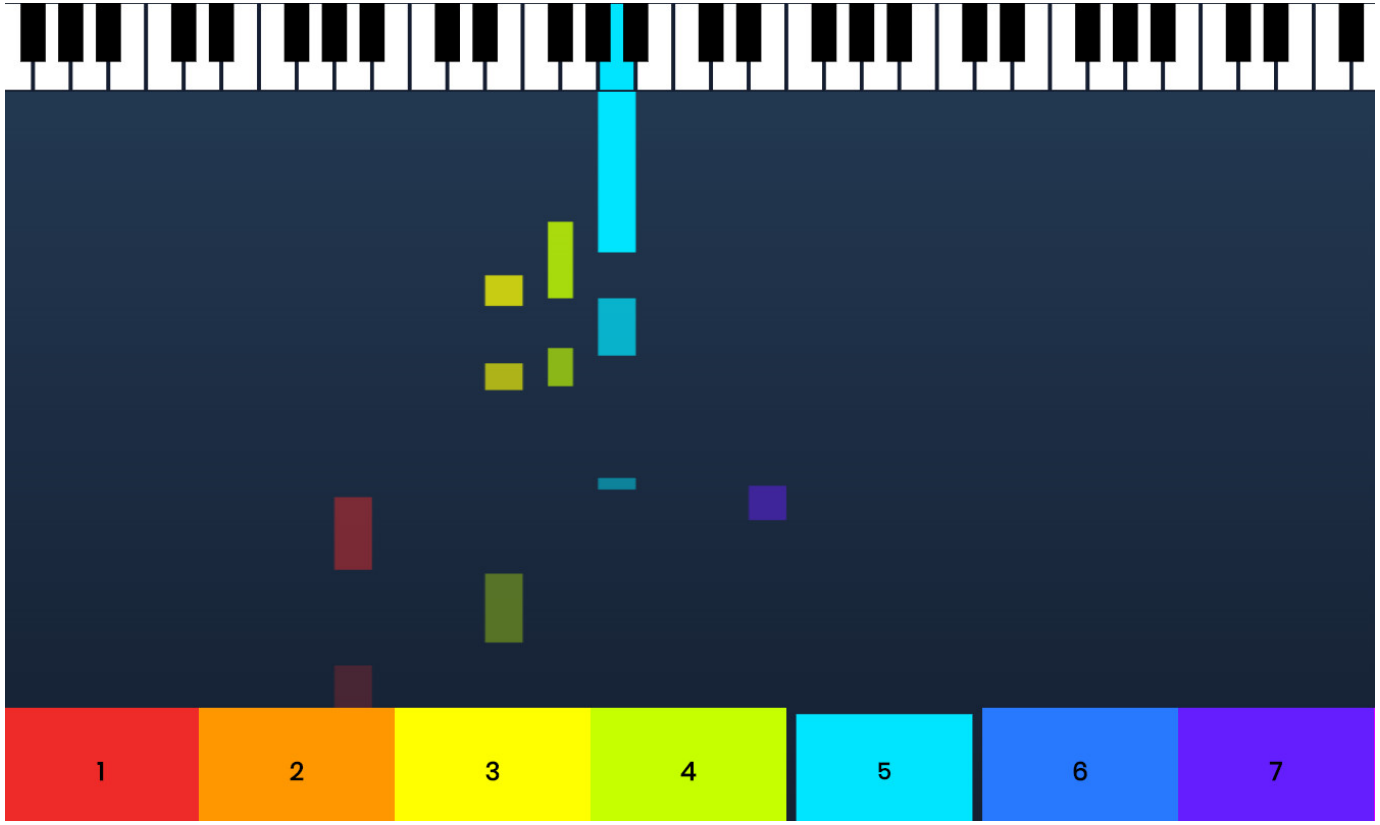
Positive Use Cases

- **Make music easier to play with** (i.e., for common people)
  - "democratization" of music creation
- **Make musicians' life easier** (i.e., for musicians)
  - inspire ideas
  - suggest continuations
  - suggest accompaniments
- **Create copyright free music for videos or games**
- **Music education / learning**
- **Digital archival**

Demo 1: Piano Genie

(Make music easier to play with)

https://magenta.tensorflow.org/pianogenie



Demo 2: Tone Transfer

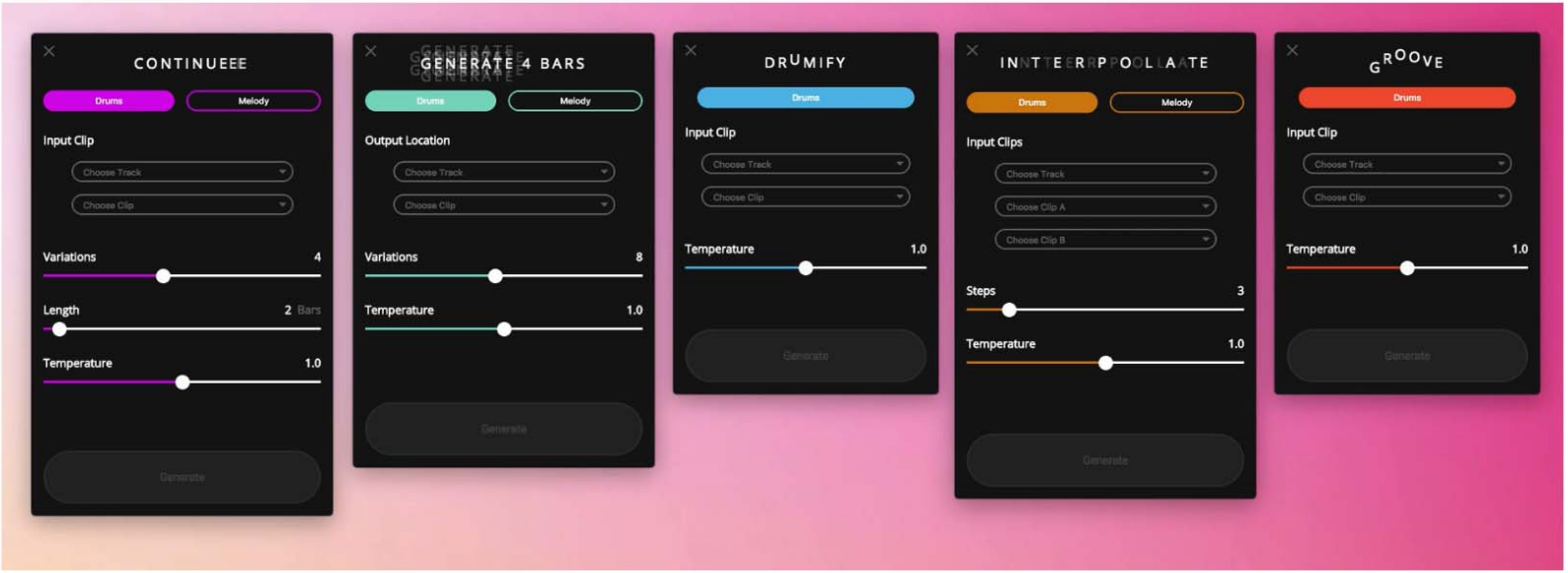
(Make music easier to play with)

https://sites.research.google/tonetransfer

Demo 3: Magenta Studio

(Make musicians' life easier)

https://magenta.tensorflow.org/studio/

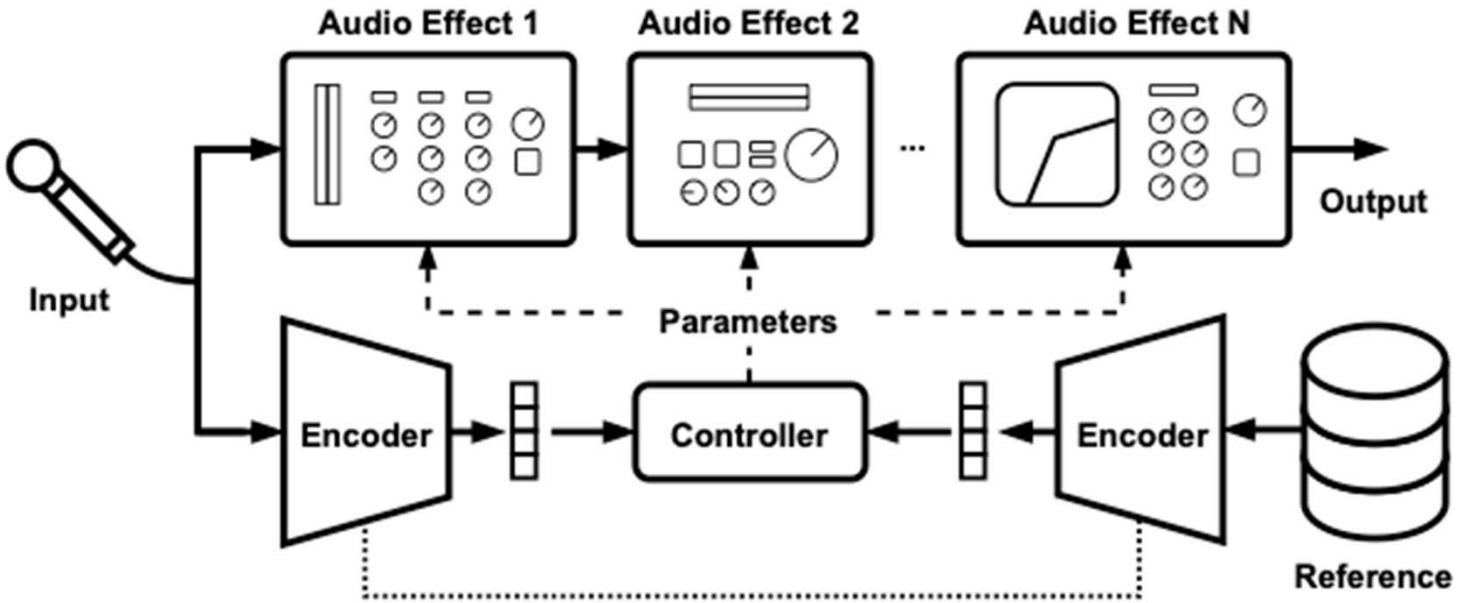


Continue," "Generate 4 bars," "Drumify," "Interpolate," "Groove"

Demo 4: Mixing Style Transfer

(Make musicians' life easier)

https://csteinmetz1.github.io/DeepAFx-ST/

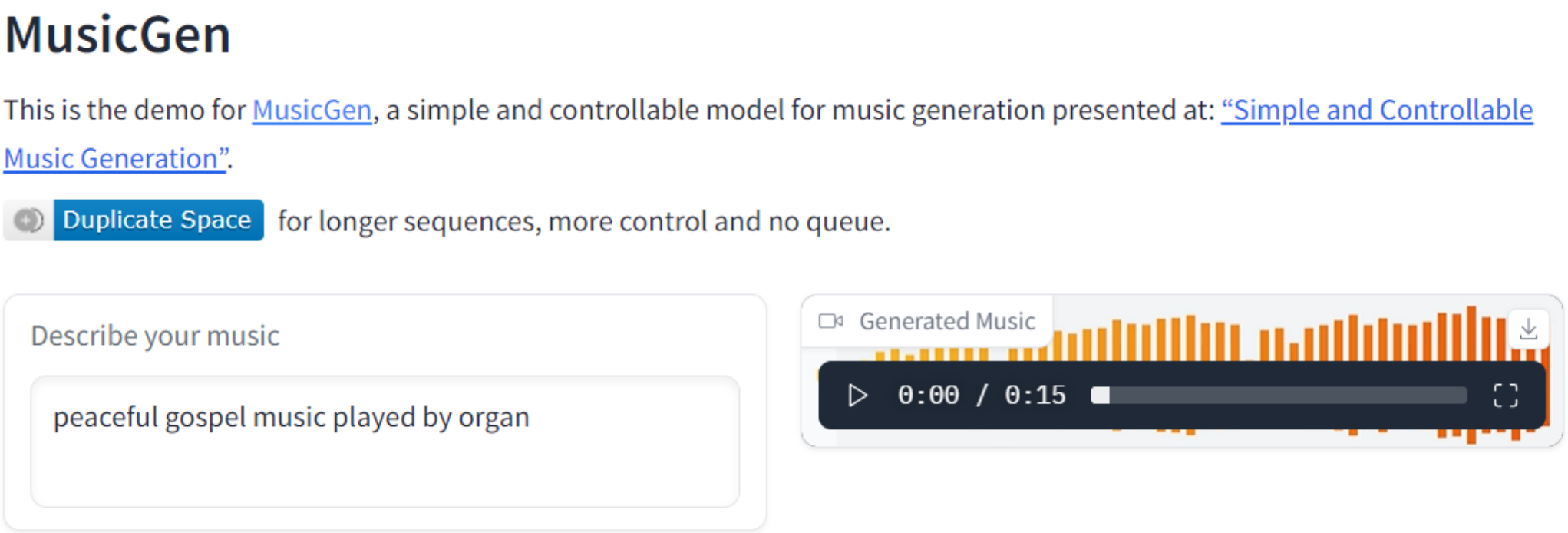


Demo 5: Text-to-Music

(Create copyright free music for videos or games)

https://ai.honu.io/papers/musicgen/

https://huggingface.co/spaces/facebook/MusicGen



Demo 6: Image-to-Sound by Adobe Firefly

https://www.youtube.com/watch?v=30xueN12guw
https://www.adobe.com/tw/sensei/generative-ai/firefly.html

Future Vision: Adobe Firefly for @AdobeVideo

Demo 7: AI MV

https://www.ziaxaza.com/

Demo 8: KaraSinger

https://jerrygood0703.github.io/KaraSinger/

Lyrics:
In this paper we propose
a novel neural network model
called Karaoke singer for a less studied
singing voice synthesis task
named score-free SVS
in which the prosody and melody are
spontaneously decided by machine.

Demo 9: AI Sandee

https://www.youtube.com/watch?v=nWTuZIRU80A

「音樂製作人的工作是無法被取代的」。AI vocal 要怎麼唱，能唱得多好，終究需要專業音樂製作人，以人類的美學和經驗去引導 AI，要如何將 AI 昇華到情感面，終究還是需要製作人的能力，以及對音樂的想像力。

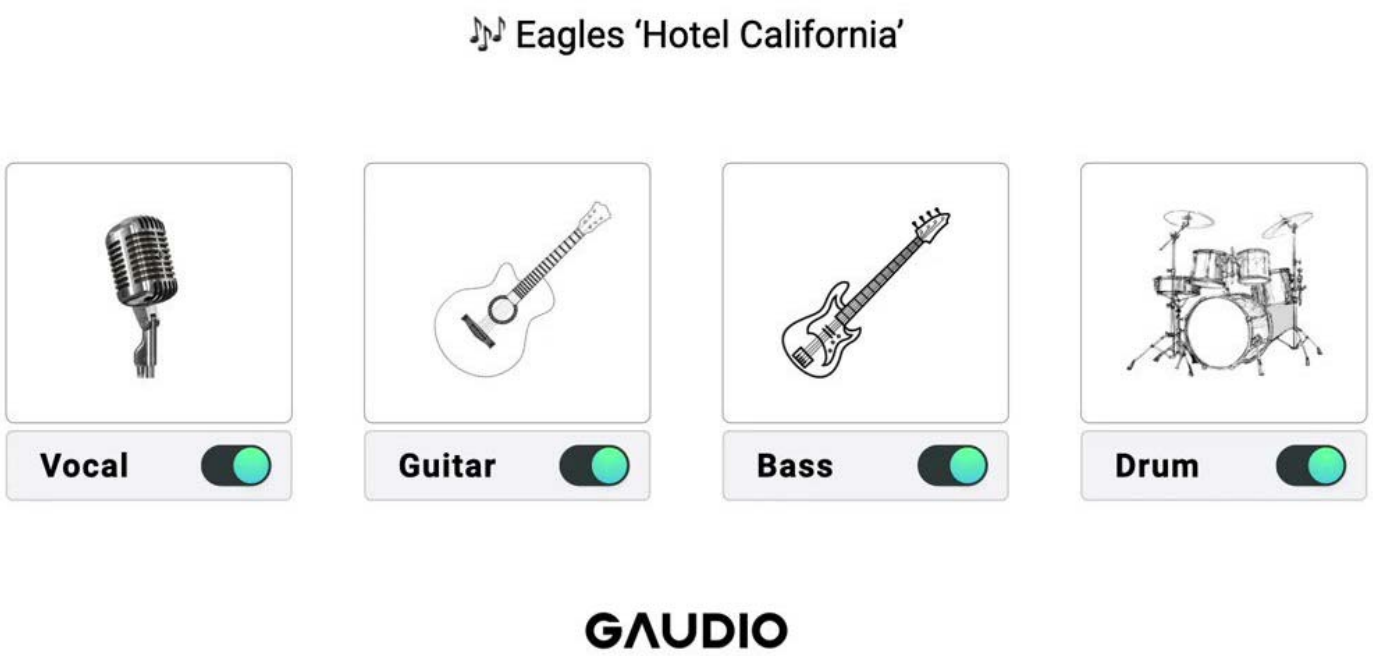
作為一個仍在線上的歌手與製作人，由我親自處理自己的AI vocal，讓這首歌傳達出「創作者、歌者不怕 AI 的挑戰」、「我們擁有自己聲音的控制權」等訊息，同時也是「人類的思考和意志，才是人之所以為人」的巨大宣示。

透過聆聽《教我如何做你的愛人》，試著探討：「若 AI 已經能模擬原唱的一切，那麼原唱歌手的價值會是什麼？」

當 AI 真正學會唱歌之後，就是創作人與歌手，重新理解自身價值的時候了。....by公主

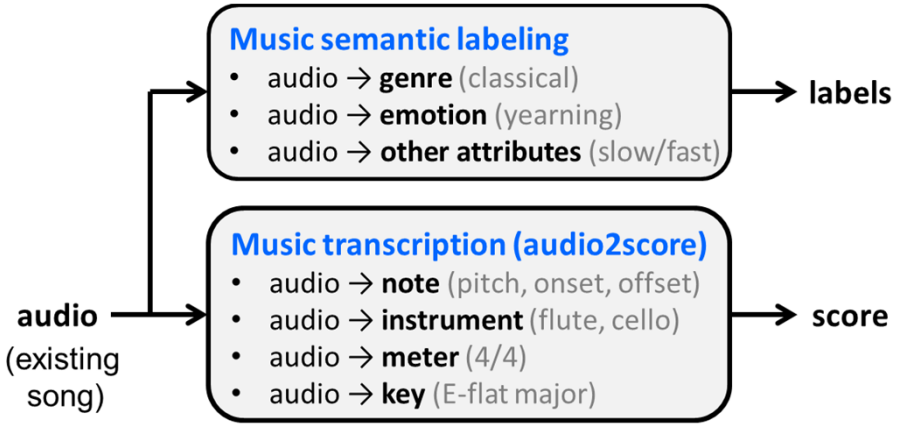
Demo 10: Source Separation

https://www.gaudiolab.com/technology/source-separation



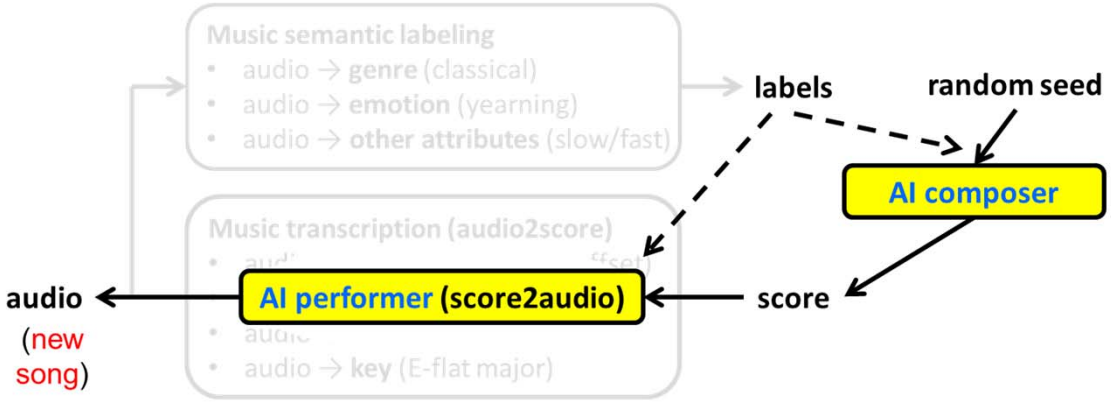
Music AI: {signal processing, machine learning} + music

- **Music analysis**



  - music understanding
  - music search
  - music recommendation

- **Music generation**



  - MIDI generation
  - audio generation
  - MIDI-to-audio generation

Outline

- Music & AI
- **The course**
- Brief intro to Music Information Retrieval (MIR)

This Course: Prerequisites

- **Graduate level** (CommE5070) @ NTU GICE
  - It's ***NOT*** a music course
  - It's an EE/CS graduate-level course working on music data/problems
- Prerequisites
  - Great interest in music
  - Good background in machine learning & deep learning
  - Good coding experience in python and a deep learning framework such as PyTorch
- If you know little about deep learning → **Don't** take this course

This Course: Wills and Won’ts

- Will talk about
  - Domain knowledge in music data representation
  - Domain knowledge in music analysis: timbre, rhythm, pitch
  - Deep learning-based music analysis
  - Deep learning-based audio generation
  - Deep learning-based MIDI sequence generation
- **Won’t** talk about
  - Basics in deep learning
  - Applications in other domains

Lecturer

- Lecturer
  - Yi-Hsuan Yang (楊奕軒)
    - https://affige.github.io/
    - yhyangtw@ntu.edu.tw

- Office hour
  - Thursday 9:30-11:30, or by appointment
  - Office: EE2-337 (電二)

Teaching Assistants

- TA
  - Fischer Yeh (葉軒瑜)
    - fish90510@gmail.com
  - Wei-Jaw Lee (李維釗)
    - weijaw2000@gmail.com

- Office hour
  - Thursday 13:20-14:10, or by appointment
  - Office: BL-505 (博理館)

Location & Time

- Location: EE2-229
- Time: Thursday 7,8,9
  - **7**: 14:20-15:10
  - **8**: 15:20-16:10 (i.e., 10 mins earlier)
  - **9**: 16:20-17:10 (i.e., 10 mins earlier)

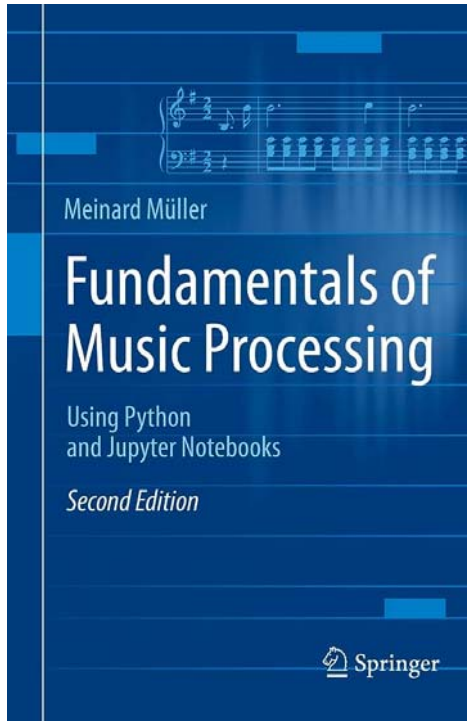
Textbook

(for the music analysis part)

- Reference textbook

Meinard Müller
Fundamentals of Music Processing
Using Python and Jupyter Notebooks

ISBN: 978-3-030-69808-9
Springer, April 2021
https://www.audiolabs-erlangen.de/fau/professor/mueller/bookFMP
https://www.audiolabs-erlangen.de/resources/MIR/FMP/C0/C0.html



- Related book
  - *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*, Wiley
    https://github.com/alexanderlerch/pyACA
    https://github.com/alexanderlerch/ACA-Slides

FMP Notebook

https://www.audiolabs-erlangen.de/resources/MIR/FMP/C0/C0.html

Part	Title	Notions, Techniques & Algorithms	HTML	IPYNB
B	Basics	Basic information on Python, Jupyter notebooks, Anaconda package management system, Python environments, visualizations, and other topics	[html]	[ipynb]
0	Overview	Overview of the notebooks (https://www.audiolabs-erlangen.de/FMP)	[html]	[ipynb]
1	Music Representations	Music notation, MIDI, audio signal, waveform, pitch, loudness, timbre	[html]	[ipynb]
2	Fourier Analysis of Signals	Discrete/analog signal, sinusoid, exponential, Fourier transform, Fourier representation, DFT, FFT, STFT	[html]	[ipynb]
3	Music Synchronization	Chroma feature, dynamic programming, dynamic time warping (DTW), alignment, user interface	[html]	[ipynb]

Part	Title	Notions, Techniques & Algorithms	HTML	IPYNB
4	Music Structure Analysis	Similarity matrix, repetition, thumbnail, homogeneity, novelty, evaluation, precision, recall, F-measure, visualization, scape plot	[html]	[ipynb]
5	Chord Recognition	Harmony, music theory, chords, scales, templates, hidden Markov model (HMM), evaluation	[html]	[ipynb]
6	Tempo and Beat Tracking	Onset, novelty, tempo, tempogram, beat, periodicity, Fourier analysis, autocorrelation	[html]	[ipynb]
7	Content-Based Audio Retrieval	Identification, fingerprint, indexing, inverted list, matching, version, cover song	[html]	[ipynb]
8	Musically Informed Audio Decomposition	Harmonic/percussive separation, signal reconstruction, instantaneous frequency, fundamental frequency (F0), trajectory, nonnegative matrix factorization (NMF)	[html]	[ipynb]

Grading Policy

- Grading policy
  - **Assignments** (60%), 3 times
  - **Final Project** (40%): for teams of 2 or **3** (recommended)

- **Work hard to get high score**
  - I don’t plan to please the students

Syllabus

- W1. Introduction to the course
- W2. Fundamentals & Music representation
- W3. **Analysis** I (timbre): Automatic music classification and representation learning
- (**HW1**: Singer classification)
- W4. **Generation** I: Source separation
- W5. **Generation** II: GAN & Vocoders
- W6. **Generation** III: Synthesis of notes and loops
- (**HW2**: Vocoder & Loop generation)
- W7. **Analysis** II (pitch): Music transcription, Melody extraction, and Chord Recognition
- W8. **Analysis** III (rhythm): Beat/downbeat tracking
- W9. **Generation** IV: Symbolic MIDI generation
- (**HW3**: Pop music transformer)
- W10. **Generation** V: Singing voice generation
- W11. **Generation** VI: Differentiable DSP models and automatic mixing
- **W12**. Proposal of ideas of final projects
- W13. **Generation** VII: Symbolic MIDI generation: Advanced Topics
- W14. **Generation** VIII: Text-to-music generation
- W15. Miscellaneous Topics
- **W16**. Oral presentation of final projects

Assignments

- **Programming** (in python) + **report** (in English)
  - We assume that you have good coding experience in python and a deep learning framework such as PyTorch
  - The assignments can be **quite hard** for deep learning beginners
  - Submit **code** + **model** + **report**
  - NO cheating: Will run *plagiarism detector*

- HW1: **singer classification** (accuracy *leaderboard*)
- HW2: **GAN-based vocoder** (generation quality *leaderboard*)
- HW3: **piano MIDI generation** (generation quality *leaderboard*)

NO Cheating

- Once caught: **failure** of the course

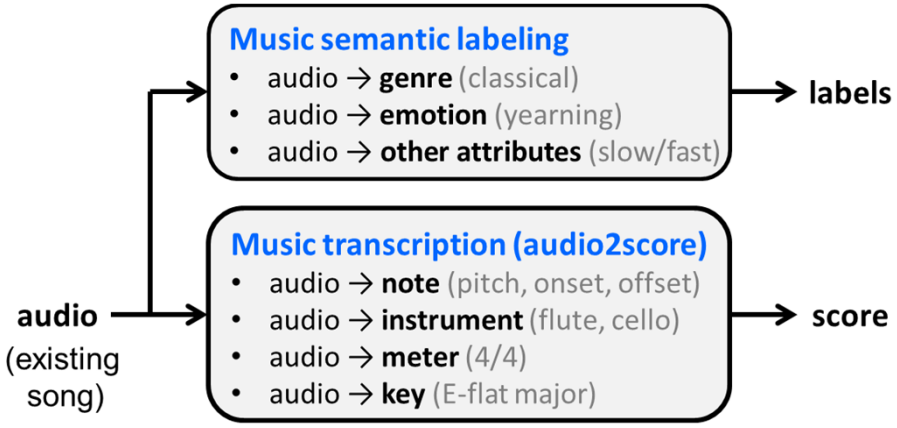
Final Project

- For teams of 2 or **3** (recommended)
- Start earlier & form teams

- Deadline for **team-up**: W10
- **Project pitch**: W12
- **Final presentation**: W16
- Deadline for **final report**: W16+2

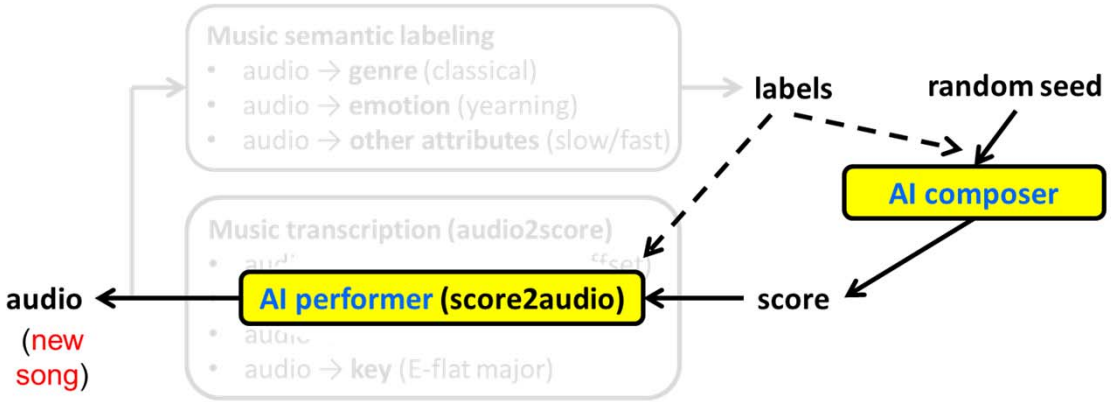
Final Project

- **Music analysis**



  - music semantic labeling
  - music transcription
  - source separation

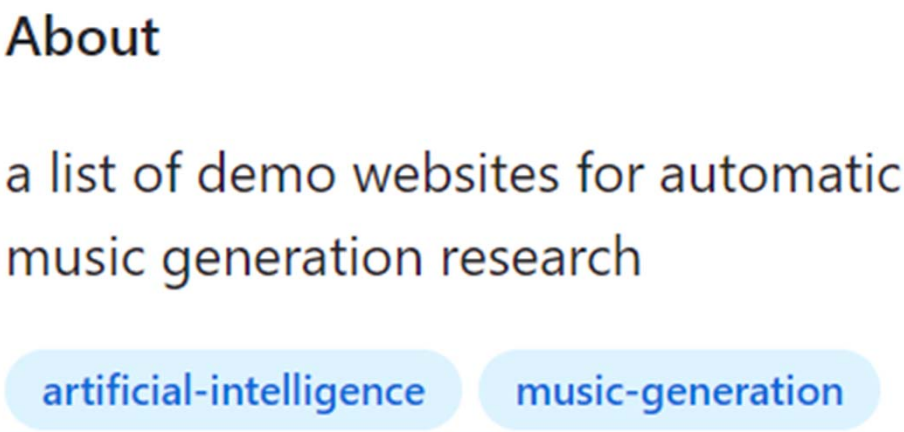
- **Music generation**



  - MIDI generation
  - audio generation
  - MIDI-to-audio generation

genmusic_demo_list

https://github.com/affige/genmusic_demo_list



Resources

- ML/DL
  - http://speech.ee.ntu.edu.tw/~tlkagk/courses.html
  - https://www.csie.ntu.edu.tw/~htlin/course/
  - https://www.csie.ntu.edu.tw/~yvchen/teaching
  - https://courses.cs.washington.edu/courses/cse599i/20au/ (generative models)
- Music information research
  - https://www.audiolabs-erlangen.de/fau/professor/mueller/teaching/2023w_mpa
  - https://musicinformationretrieval.com/
  - https://mac.kaist.ac.kr/~juhan/gct634/index.html
  - http://www.jordipons.me/apps/teaching-materials/
  - https://www.upf.edu/web/smc/audio-signal-processing-for-music-applications

Resources

- Conference proceedings
  - **Int'l Soc. Music Information Retrieval Conf.** (ISMIR)
  - Int'l Conf. Acoustic, Speech, and Signal Processing (ICASSP)
  - ACM MM, ACM ICMR, ACM SIGIR, IEEE ICME
- Transactions
  - IEEE Trans. Audio, Speech and Language Processing (TASLP)
  - IEEE Trans. Multimedia (TMM)
  - IEEE Trans. Signal Processing (TSP)

Course Website

https://cool.ntu.edu.tw/courses/27790

https://affige.github.io/teaching.html

Additional Enrollment

https://forms.gle/LcXaqhgwMcEtVfxm9

- Fill the form before 23:59, **September 9** (Saturday)
  - ML background
  - Music background
  - Ideas for final project
- Will announce the result the next **Tuesday**
  - Will only send a mail to those qualified
  - Will also post the result online at https://affige.github.io/teaching.html

Outline

- Music & AI
- The course
- **Brief intro to Music Information Retrieval (MIR)**