Deep Learning for Music Analysis and Generation

Automatic Mixing
(audio → audio)

Yen-Tung Yeh

r12942179@ntu.edu.tw

Outline

- Introduction
- Background
- Mixing-related Research
- Resources

Introduction

Audo Production

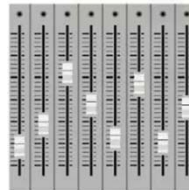## STAGES OF AUDIO ENGINEERING

TRACKING

RECORDING OF DIFFERENT INSTRUMENTS, SOUNDS AND VOCALS

EDITING

EDITING AND CORRECTING THE RECORDINGS

MIXING

ART OF PROCESS OF COMBINING ALL THE INSTRUMENTS AND MIXING THE SOUNDS

MASTERING

PROCESS OF MAKING ALL THE SONGS SOUND COHERENT

image from

Mixing

- **Audio mixing** is the process by which multiple sounds are combined into one or more audio channels by compositions of audio effect (Gain, Panning, EQ, Reverb, Compression, etc.)

Commercial Software (e.g., iZotope Neutron/Ozone)

AUDIO REPAIR

RX 10

Industry-standard audio repair tool used on movies and TV shows to restore damaged, noisy audio to pristine condition.

MIXING

Neutron 4

Mix smarter and faster while staying in your flow. Eight professional plug-ins combine to create your modern and intelligent mixing experience.

MASTERING

Ozone 11

Harness the power of Ozone 11, the industry-standard mastering suite. Featuring new processing like Clarity, Stem Focus, and Transient/Sustain for professional sound with ease and precision.

VOCAL PRODUCTION

Nectar 4

Get your vocals to sit in the mix with the most sophisticated set of tools for vocal production.

Importance of Mixing

- Demo: https://sh-lee97.github.io/grafx-prune/

Music without mixing

Music with mixing

Although these come from the same source, one is clearly much better for listening

Multitrack Mixing is Hard

- High demand of mixing technique
    - Requires good listening skills
    - Requires creativity
    - Requires understanding of music (instruments, genre)
    - Requires skills of handling large number of input tracks
- Time consuming

Can we let machine intelligently help us to mix audio?

Background

Problem Formulation of Automatic Mixing

Create a final "good" mixture by given multiple raw tracks
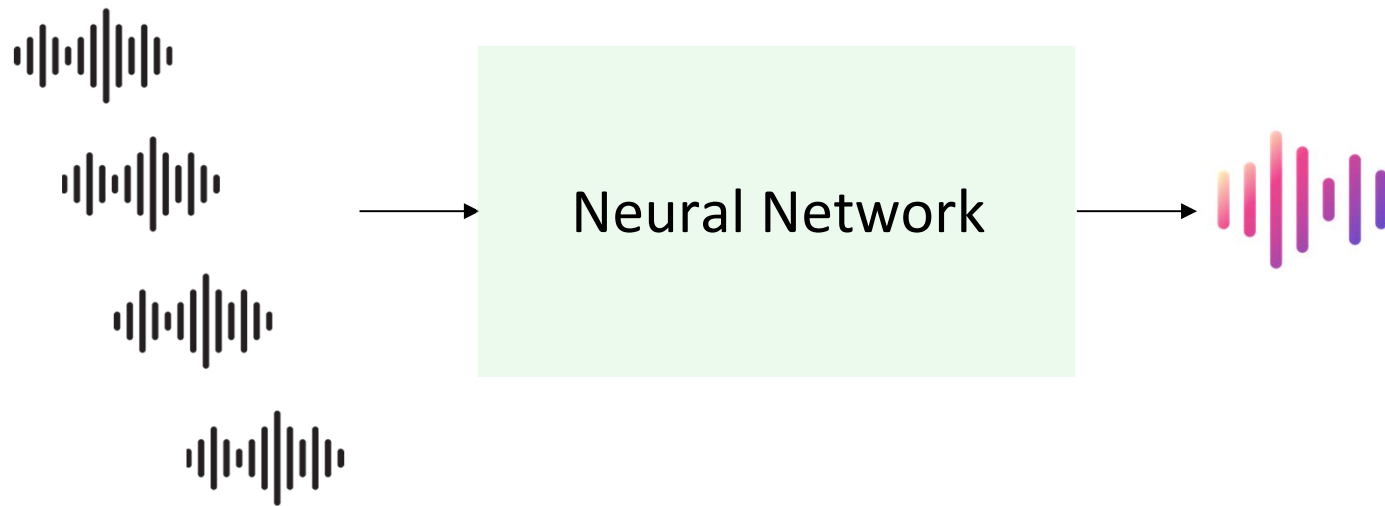
Audio Transformation

- Audio to audio
- Audio transformation

How do we **model the audio transformation** or **generate the audio transformation**?

- Direct Transformation
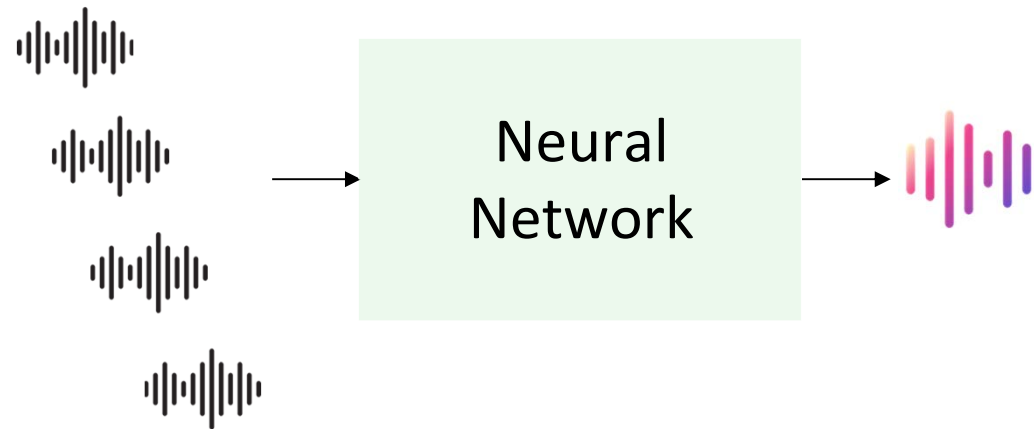- Parameter Estimation

Direct Transformation

- Neural Network will learn the transformation in black-box manner
  - Pros:
    - Potential for creativity
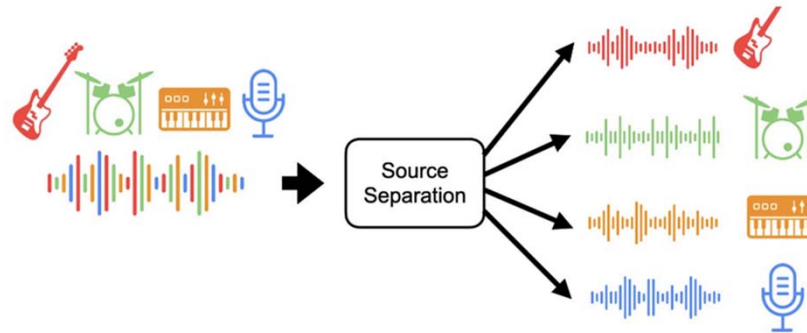   - Easy to implement
  - Cons:
    - No interpretability
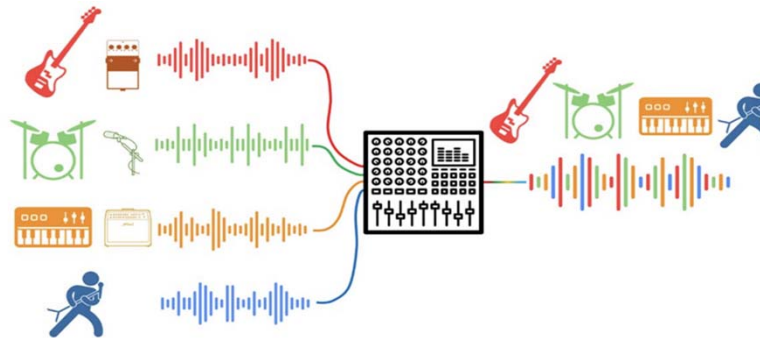   - No controllability
    - Data hungry

Mixing and Source Separation
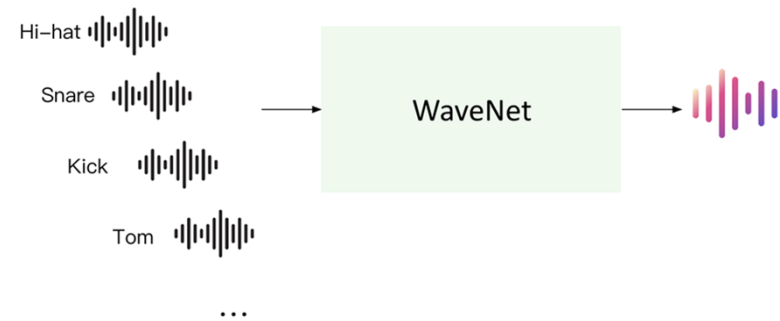
- Source separation

- Mixing

Mixing (Inverse Source Separation)

- Using Wave-U-Net for Drum Mixing
  - The first work to formulate mixing problem as inverse ss
  - Pros:
    - High quality
    - Directly learn audio transformation
  - Cons:
    - Fixed input tracks
    - Fixed order
    - Only drums

Martinez Ramirez, M., Daniel Stoller, and David Moffat. "A deep learning approach to intelligent drum mixing with the wave-u-net." Audio Engineering Society, 2021.

Recap: Audio Plugin and Mixing

- Mixing engineer often use "analog device" or "audio plugin" for mixing

Audio Effect (Modeling)

- Neural Proxy
  - Using neural network to emulate specific audio effects
  - Pros:
    - Exact behavior of specific device
  - Cons:
    - Emulating all possible effects is time consuming
- Differentiable Digital Signal Processing (DDSP)
  - Using generic algorithm for specific audio effects
  - Pros:
    - Easy and general
  - Cons:
    - Lack of "personality"

Example
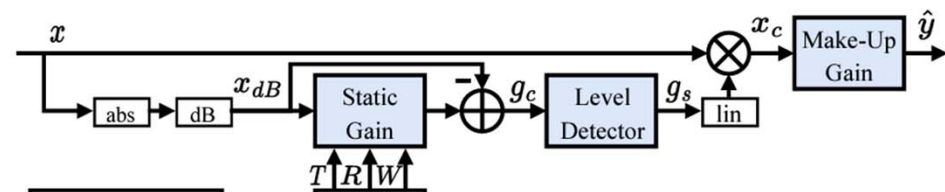
Parameter Estimation

- Neural Network will estimate the parameter of DDSP Audio Effect Processors
  - Pros:
    - Interpretable
    - Controllability
    - High-quality
    - Require less data
  - Cons
   - Limited creativity
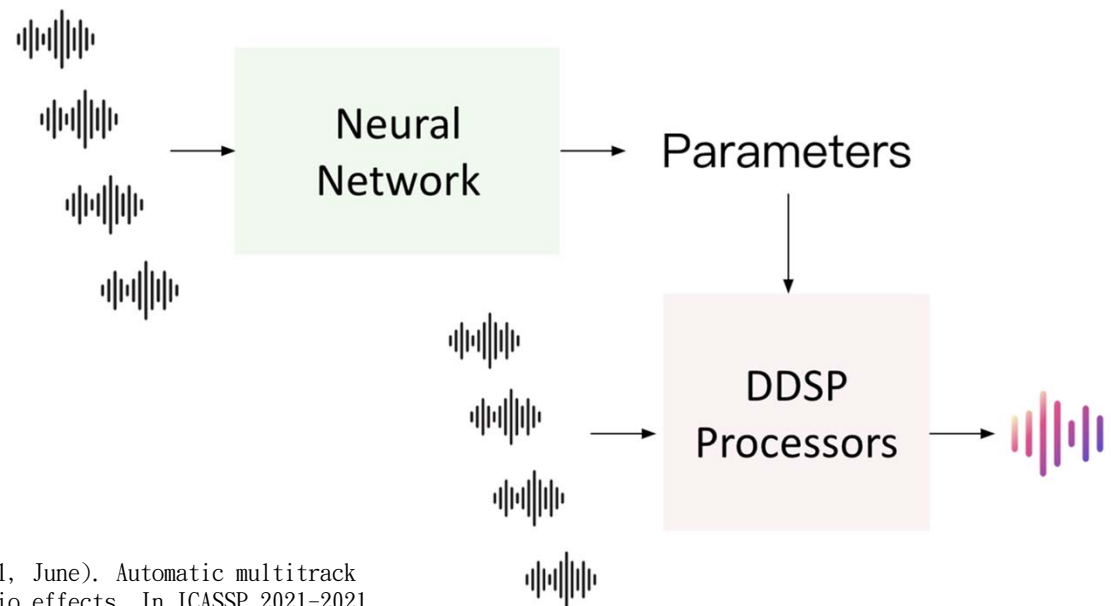
Steinmetz, C. J., Pons, J., Pascual, S., & Serra, J. (2021, June). Automatic multitrack mixing with a differentiable mixing console of neural audio effects. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 71-75). IEEE.

Neural Network (Mixing Console)

- In the real world, mixing engineers often used mixing console to control audio effects

Neural Network =
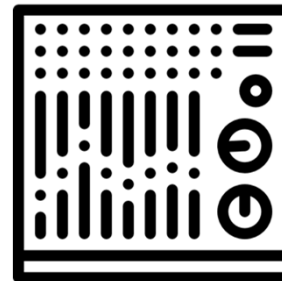
DDSP Processors

- In the real world, the plugins used by mixing engineers are implemented based on DSP
    - For example:
    - Parametric EQ is usually implemented by cascade of low shelf filter, peak filter, and high shelf filter

DDSP
Processors

=

Dataset

from https://github.com/csteinmetz1/automix-toolkit

Dataset	Size(Hrs)	no. of Songs	no. of Instrument Category	no. of tracks	Type	Usage Permissions	Other info	Remarks
MedleyDB	7.2	122	82	1-26	Multitrack, Wav	Open	44.1KHz, 16 bit, stereo	-
ENST Drums	1.25	-	1	8	Drums, Wav/AVI	Limited	44.1KHz, 16 bit, stereo	Drums only dataset
Cambridge Multitrack	>3	>50	>5	5-70	Multitrack, Wav	open	44.1KHz, 16/24 bit, Stereo	Not time alligned, recordings for all the songs are not uniform

Loss

- Audio domain loss
  - Time domain
    - L1 Loss
  - Freq domain
    - MS-STFT Loss
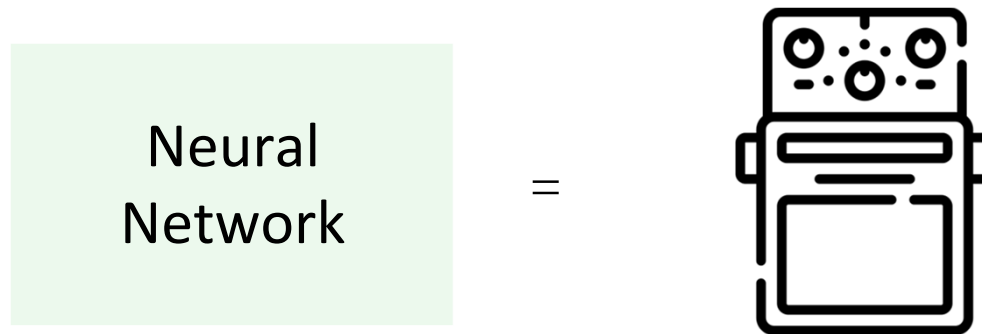- Parameter Loss

Evaluation Metric

- Objective:
  - Reconstruction Loss
  - Audio Feature
  - FAD
- Subjective
  - Listening test

Usually, listening test is the best evaluation method for mixing
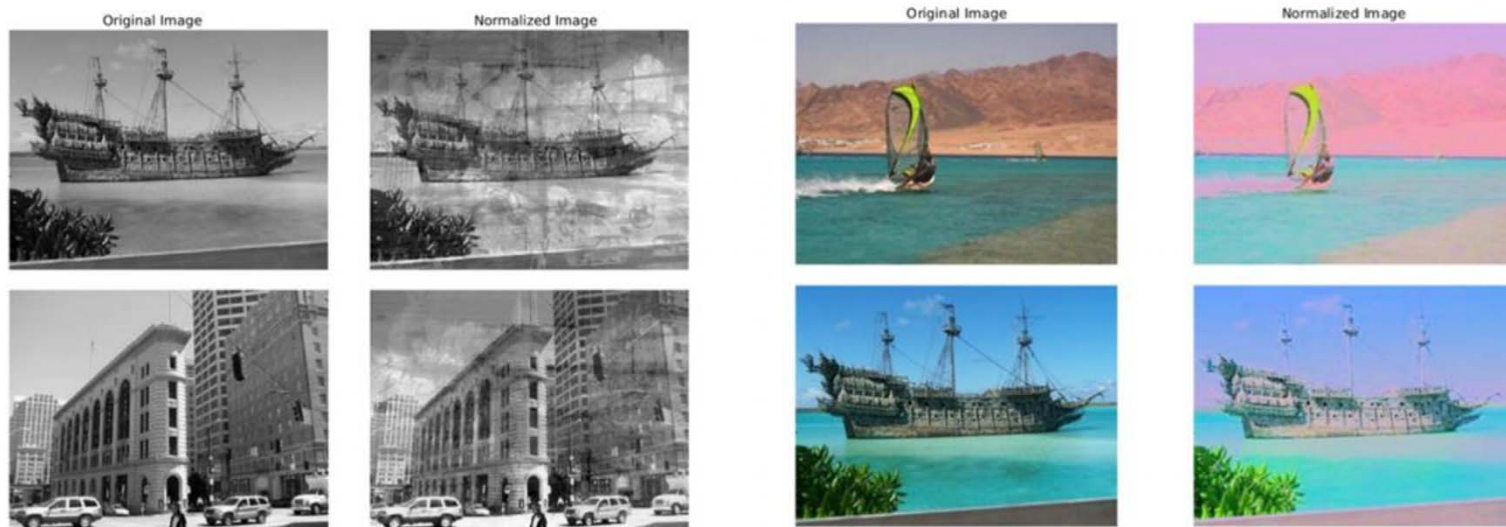
Mixing-related Research

Audio Effect Modeling

- Given target audio effects (analog devices), using neural network to emulate its behavior

Neural Network =

Ref: Audio Effect Research

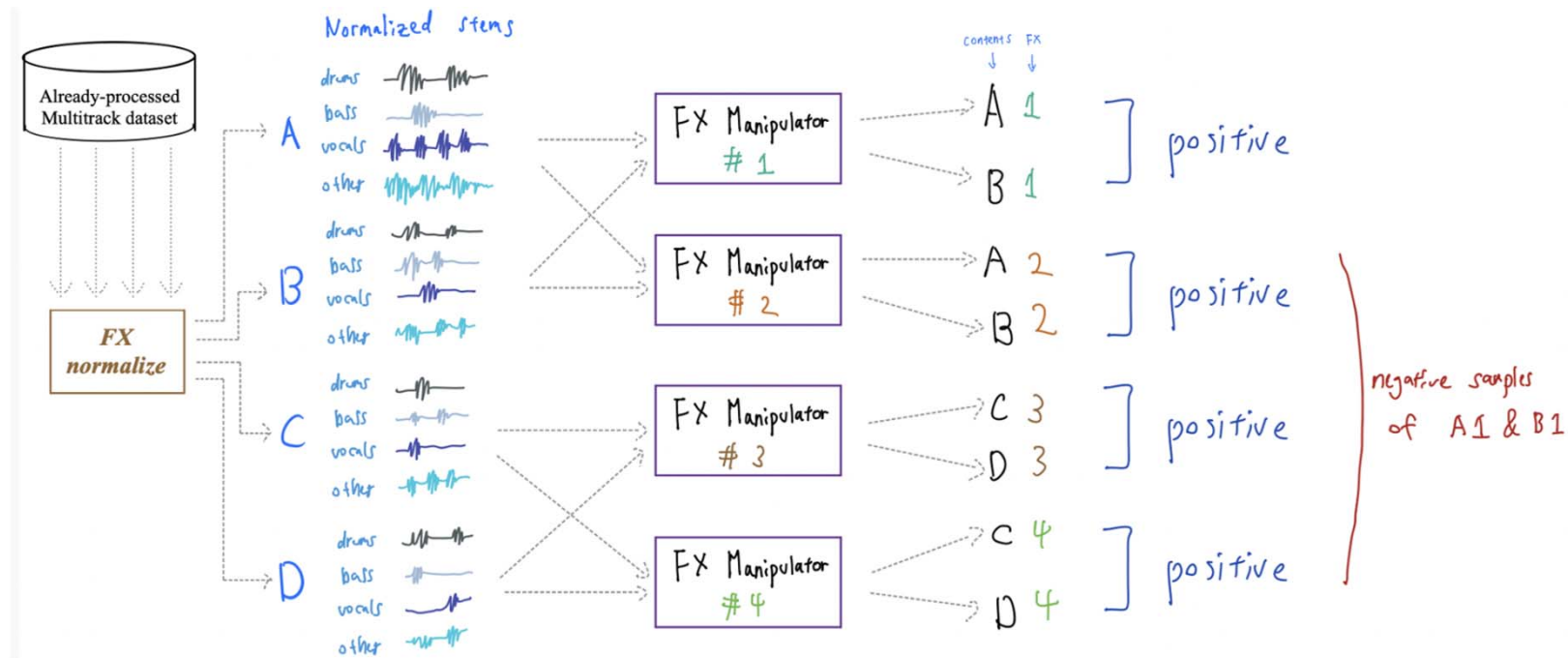# Audio Effect Normalization

- Data normalization for Audio Effect

Martínez-Ramírez, M. A., Liao, W. H., Fabbro, G., Uhlich, S., Nagashima, C., & Mitsufuji, Y. (2022). Automatic music mixing with deep learning and out-of-domain data. arXiv preprint arXiv:2208.11428.
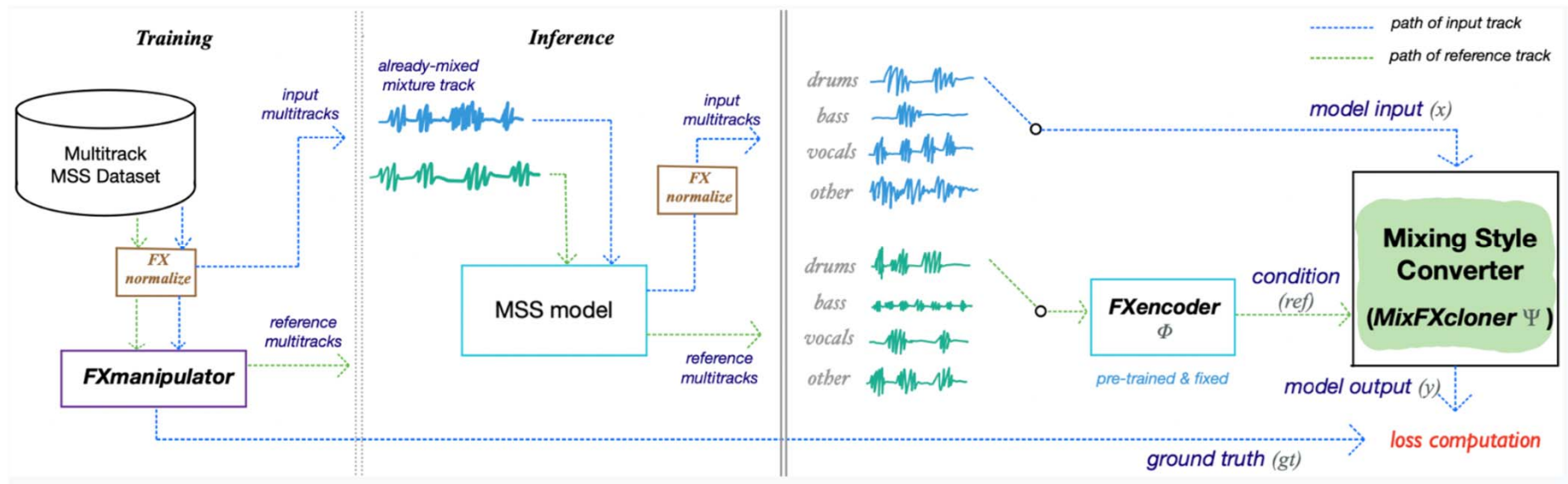
Audio Effect Representation Learning

- Using contrastive learning for effect disentanglement

Koo, J., Martinez-Ramirez, M. A., Liao, W. H., Uhlich, S., Lee, K., & Mitsufuji, Y. (2023, June). Music mixing style transfer: A contrastive learning approach to disentangle audio effects. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.

Mixing Style Transfer

- We want the "mixing style" of song A apply to song B

Koo, J., Martinez-Ramirez, M. A., Liao, W. H., Uhlich, S., Lee, K., & Mitsufuji, Y. (2023, June). Music mixing style transfer: A contrastive learning approach to disentangle audio effects. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.
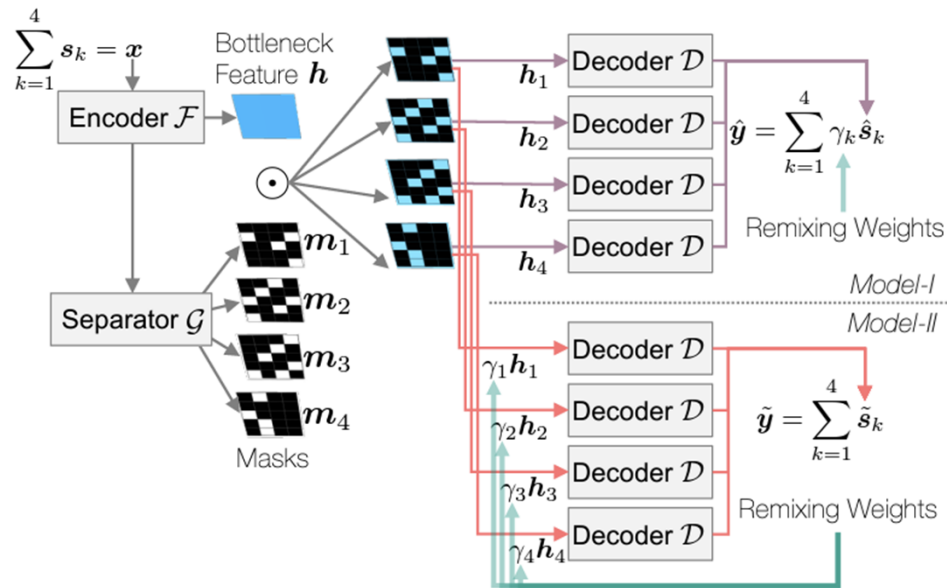
Reverse Engineering of Mixing

- Given raw input tracks and mixture, can we reverse engineer the "mixing graph"?
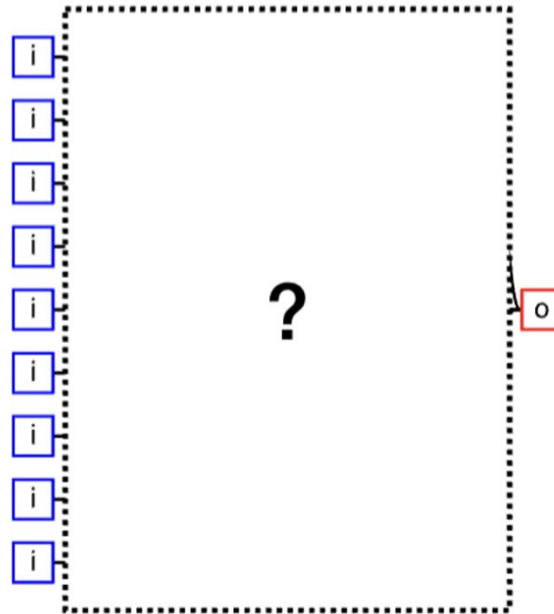
Lee, S., Martínez-Ramírez, M. A., Liao, W. H., Uhlich, S., Fabbro, G., Lee, K., & Mitsufuji, Y. (2024). Searching For Music Mixing Graphs: A Pruning Approach. *arXiv preprint arXiv:2406.01049*.
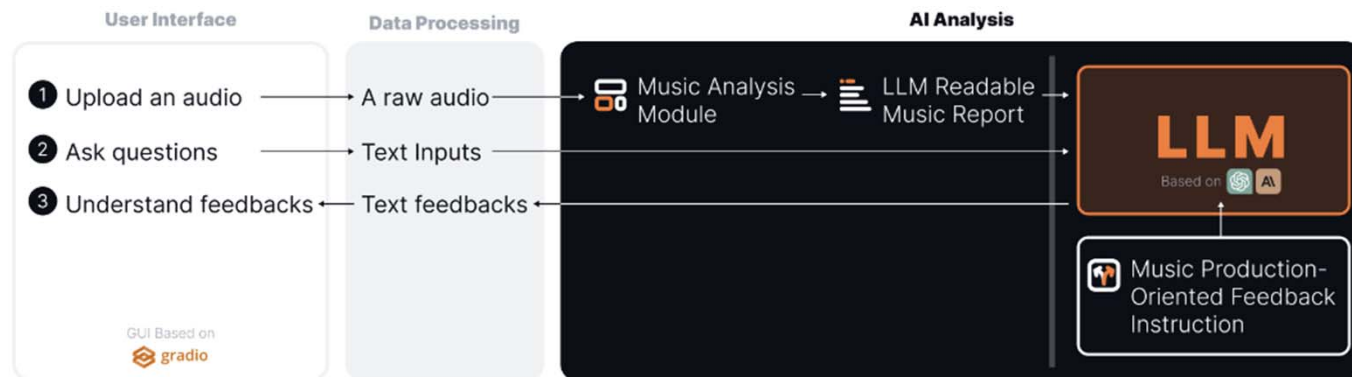
Remixing

- Given mixture, we want to apply other cool mixing style

Yang, H., Firodiya, S., Bryan, N. J., & Kim, M. (2022, May). Don't separate, learn to remix: End-to-end neural remixing with joint optimization. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 116-120). IEEE.

Mixing Agent

- Leveraging LLM to act as a mixing agent, giving you advice of mixing

Yi-Lin Jiang, Chia-Ho Hsiung, Yen-Tung Yeh, Lu-Rong Chen, Bo-Yu Chen. "AI TrackMate: Finally, Someone Who Will Give Your Music More Than Just "Sounds Great!" " . accepted at Neuralips 2024 Creativity AI Track.

Difficulties in Mixing

- No reliable objective metric
- Audio transformation is hard to learn for multitrack input
- Mixing is generation task (one-to-many)

Conclusion

- Mixing is a highly creative task and quite important to music.
- Many open questions and challenges in this field

Resources

Toolkit:

- [Dasp-pytorch](#) (DDSP Fx processors in Pytorch)
- [Grafx](#) (DDSP Fx processors for Mixing Reverse Engineering)
- [PyNeuralFx](#) (Framework for Audio Effect Modeling)
- [Automix-toolkit](#) (Toolkit for Automatic Mixing)

Tutorials & Materials:

- [Deep Learning for Automatic Mixing](#)
- [Automatic Mixing Research](#)
- [Audio Effect Research](#)
- [ISMIR22 Mixing Tutorial](#)
- DAFx24 Mixing Tutorial (if interested, please contact me)