PAT 464/564 (Winter 2026)

Generative AI for Music & Audio Creation

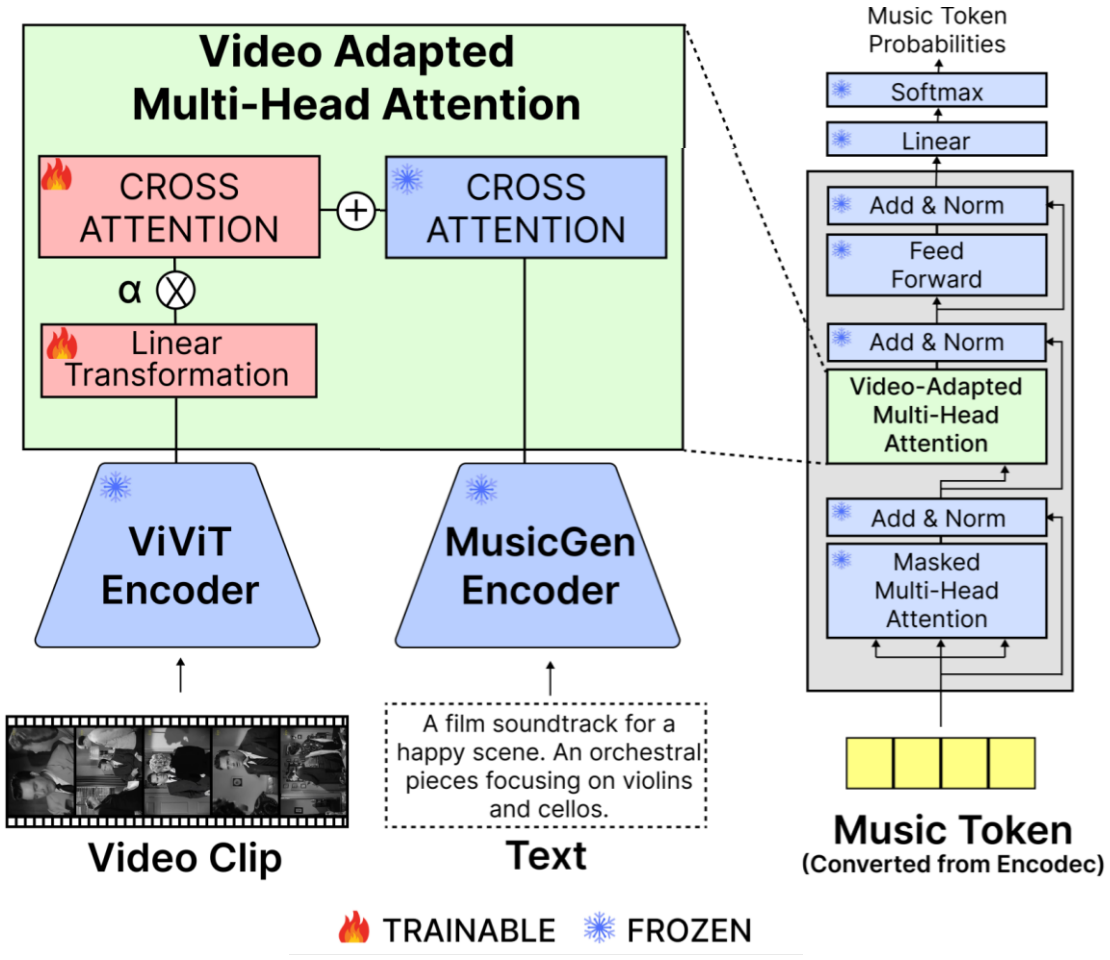
Lecture 16: Multimodal Systems & Music Production

Instructor: Hao-Wen Dong

UNIVERSITY OF MICHIGAN

Multimodal Systems

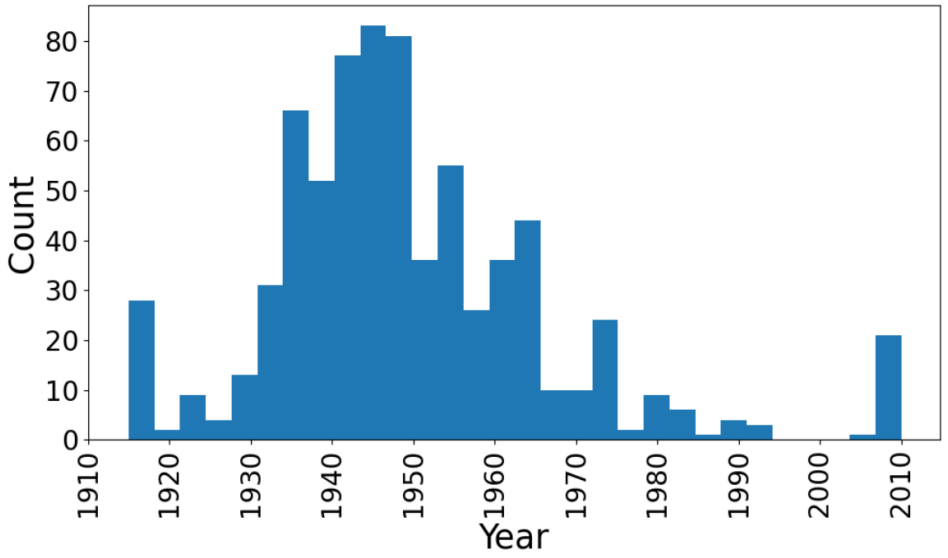
Video-Guided Text-to-Music Generation (Kim et al., 2025)



(Source: Kim et al., 2025)

Open Screen Soundtrack Library (OSSL) (Kim et al., 2025)

- **736 video clips** from **299 films** in **public domain** or **CC-licensed**
- **36.5 hours** in total
- **Mood annotations** as Russell's 4Q (arousal-valence model)



(Source: Kim et al., 2025)

Haven Kim, Zachary Novack, Weihan Xu, Julian McAuley, and Hao-Wen Dong , "Video-Guided Text-to-Music Generation Using Public Domain Movie Collections," *ISMIR*, 2025.

Open Screen Soundtrack Library (OSSL) (Kim et al., 2025)

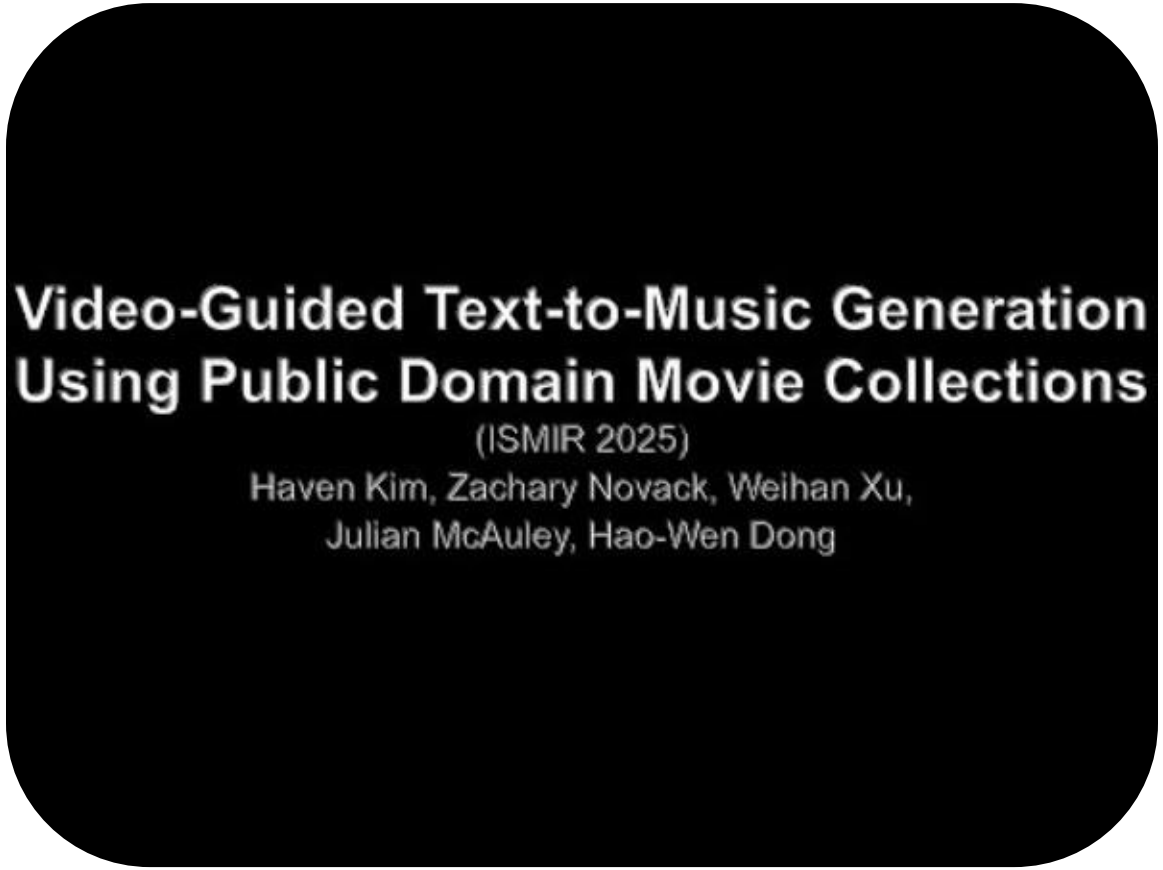





havenpersona.github.io/ossl-v1

Haven Kim, Zachary Novack, Weihan Xu, Julian McAuley, and Hao-Wen Dong , "Video-Guided Text-to-Music Generation Using Public Domain Movie Collections," *ISMIR*, 2025.

Video-Guided Text-to-Music Generation (Kim et al., 2025)



youtu.be/S0BMicbdzmg

Haven Kim, Zachary Novack, Weihan Xu, Julian McAuley, and Hao-Wen Dong , "Video-Guided Text-to-Music Generation Using Public Domain Movie Collections," *ISMIR*, 2025.

Video-Guided Text-to-Music Generation Using Public Domain Movie Collections

(ISMIR 2025)

Haven Kim, Zachary Novack, Weihan Xu,
Julian McAuley, Hao-Wen Dong

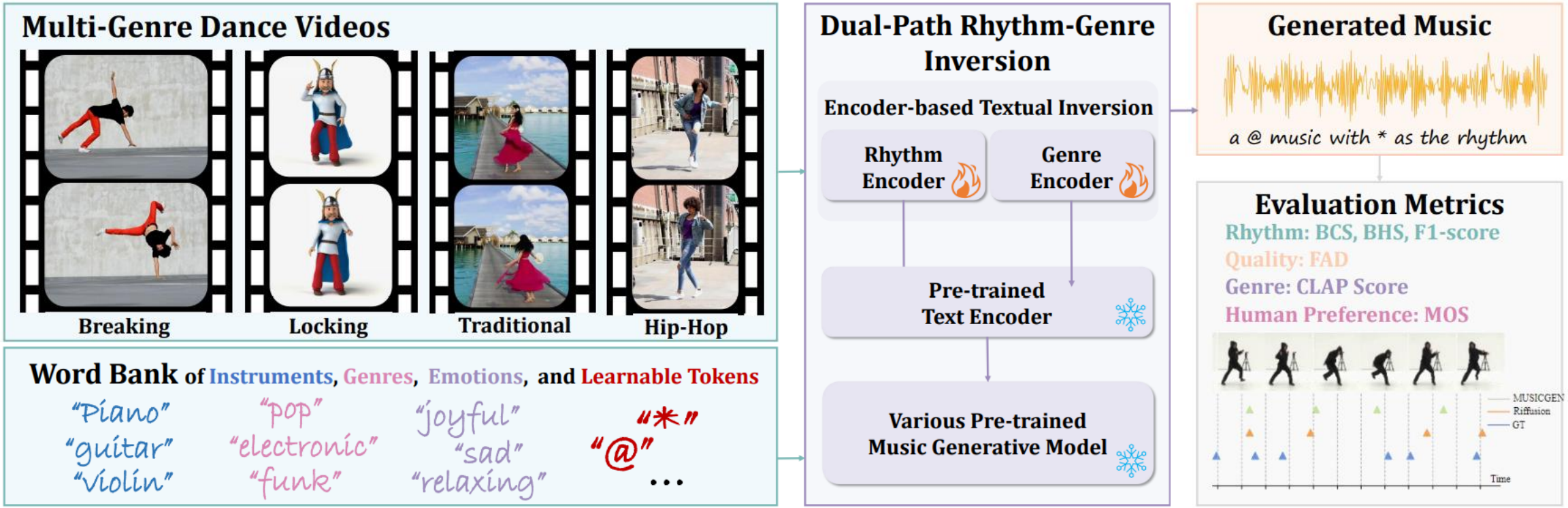
Dance, Music & AI

(Source: Yamaha)

yamaha.com/en/news_release/2018/18013101/

Yamaha Global, "Yamaha Artificial Intelligence (AI) Transforms a Dancer into a Pianist - Short Version," *YouTube*, youtu.be/21injmy1wsU, 2018.

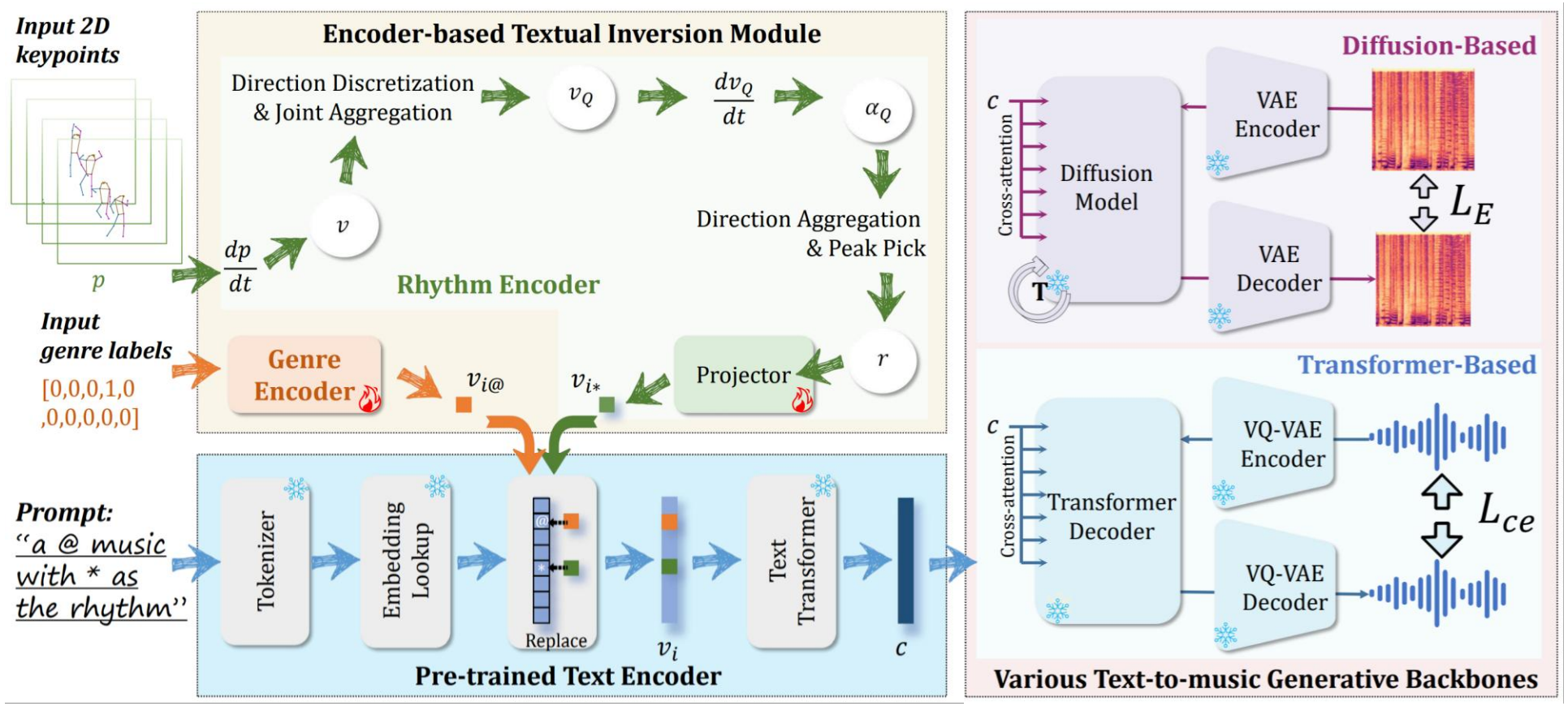
Dance-to-Music Generation (Li et al., 2024)



(Source: Li et al., 2024)

Sifei Li, Weiming Dong, Yuxin Zhang, Fan Tang, Chongyang Ma, Oliver Deussen, Tong-Yee Lee, and Changsheng Xu, "Dance-to-Music Generation with Encoder-based Textual Inversion," *SIGGRAPH ASIA*, 2024.

Dance-to-Music Generation (Li et al., 2024)



(Source: Li et al., 2024)

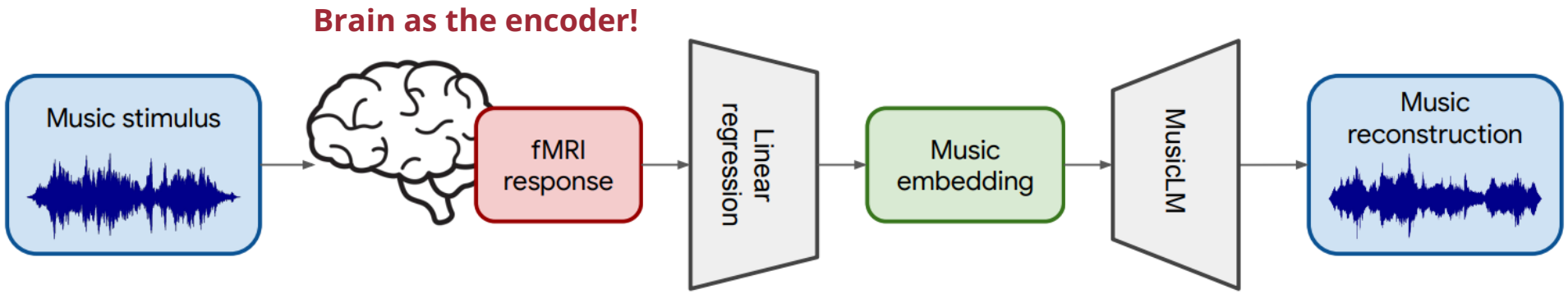
Sifei Li, Weiming Dong, Yuxin Zhang, Fan Tang, Chongyang Ma, Oliver Deussen, Tong-Yee Lee, and Changsheng Xu, "Dance-to-Music Generation with Encoder-based Textual Inversion," *SIGGRAPH ASIA*, 2024.

Dance-to-Music Generation (Li et al., 2024)

youtu.be/y2pG2S5xDLY

Sifei Li, Weiming Dong, Yuxin Zhang, Fan Tang, Chongyang Ma, Oliver Deussen, Tong-Yee Lee, and Changsheng Xu, "Dance-to-Music Generation with Encoder-based Textual Inversion," *SIGGRAPH ASIA*, 2024.

Brain2Music (Denk et al., 2023)

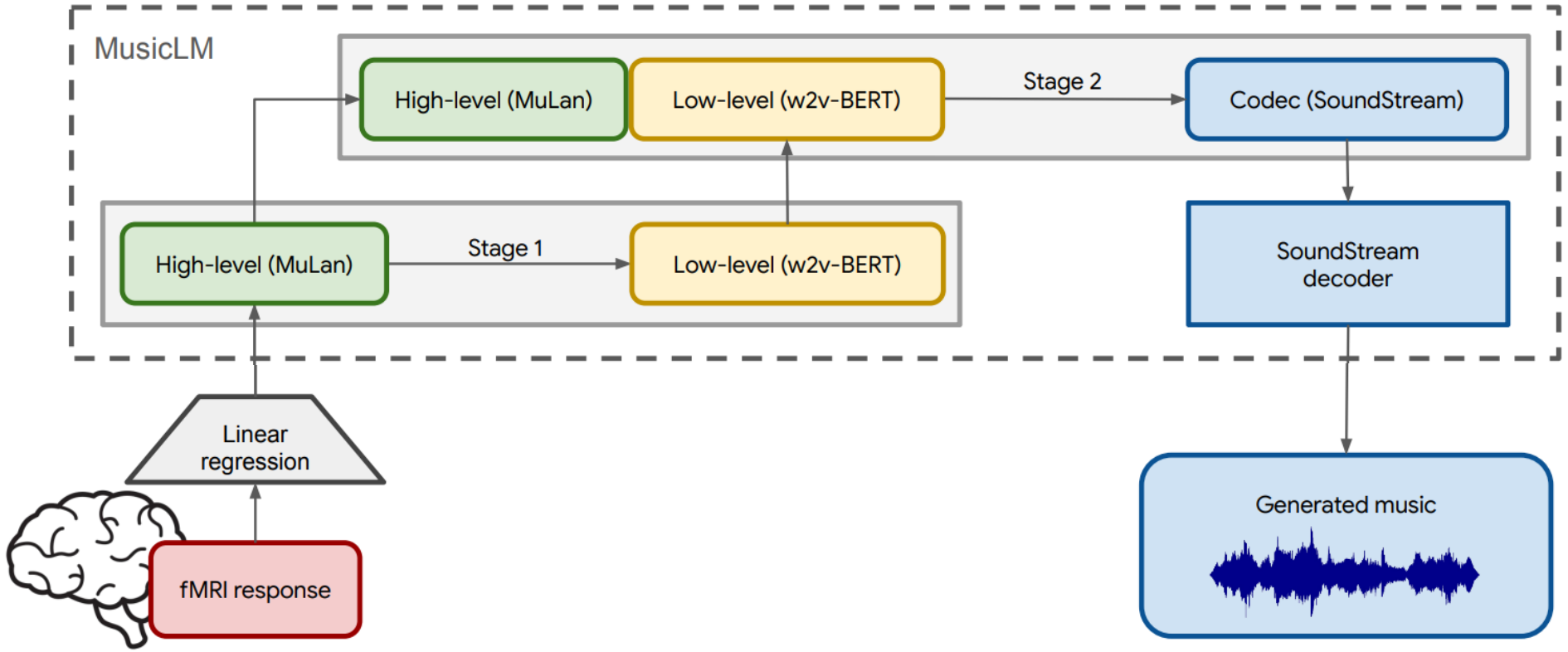


(Source: Denk et al., 2023)

Can we decode human brain-encoded music?

Timo I. Denk, Yu Takagi, Takuya Matsuyama, Andrea Agostinelli, Tomoya Nakai, Christian Frank, and Shinji Nishimoto, "Brain2Music: Reconstructing Music from Human Brain Activity," *arXiv preprint arXiv:2307.11078*, 2023.

Brain2Music (Denk et al., 2023)



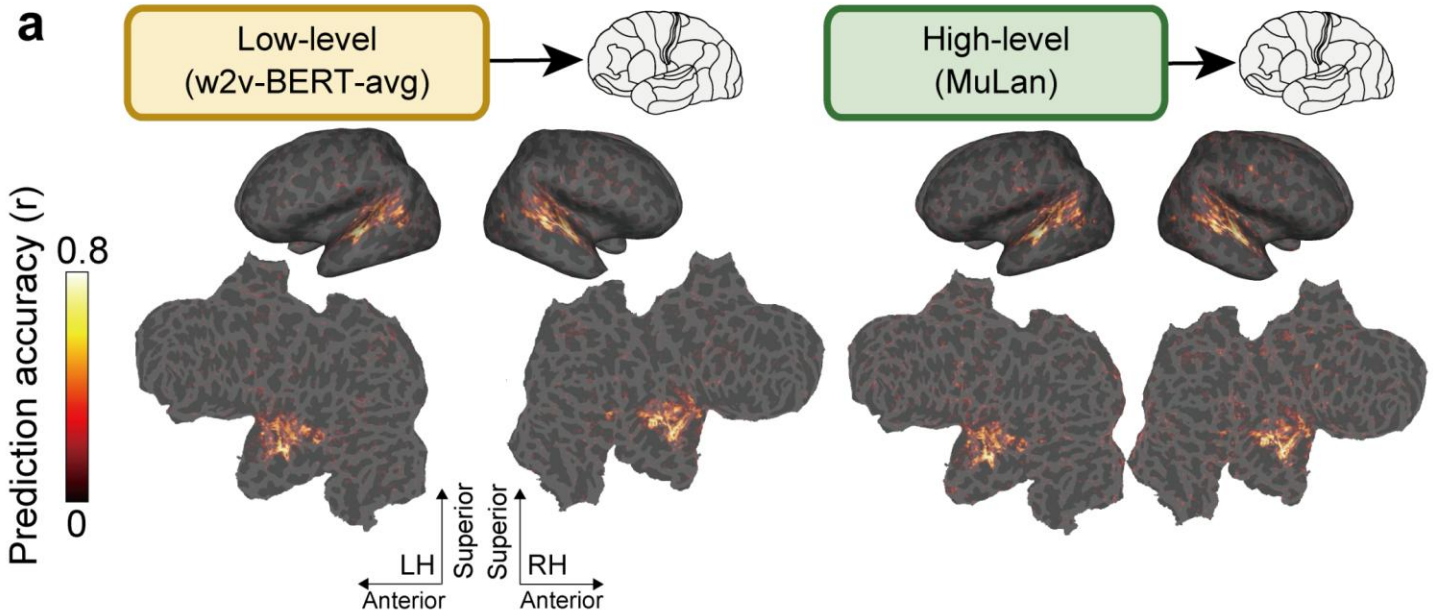
(Source: Denk et al., 2023)

google-research.github.io/seanet/brain2music

Timo I. Denk, Yu Takagi, Takuya Matsuyama, Andrea Agostinelli, Tomoya Nakai, Christian Frank, and Shinji Nishimoto, "Brain2Music: Reconstructing Music from Human Brain Activity," *arXiv preprint arXiv:2307.11078*, 2023.

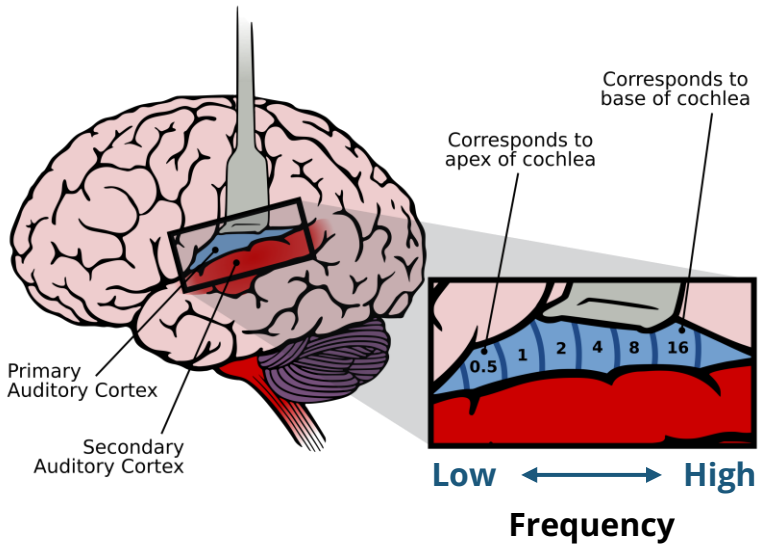
Brain2Music (Denk et al., 2023)

Audio embedding to brain activity prediction



(Source: Denk et al., 2023)

Auditory cortex

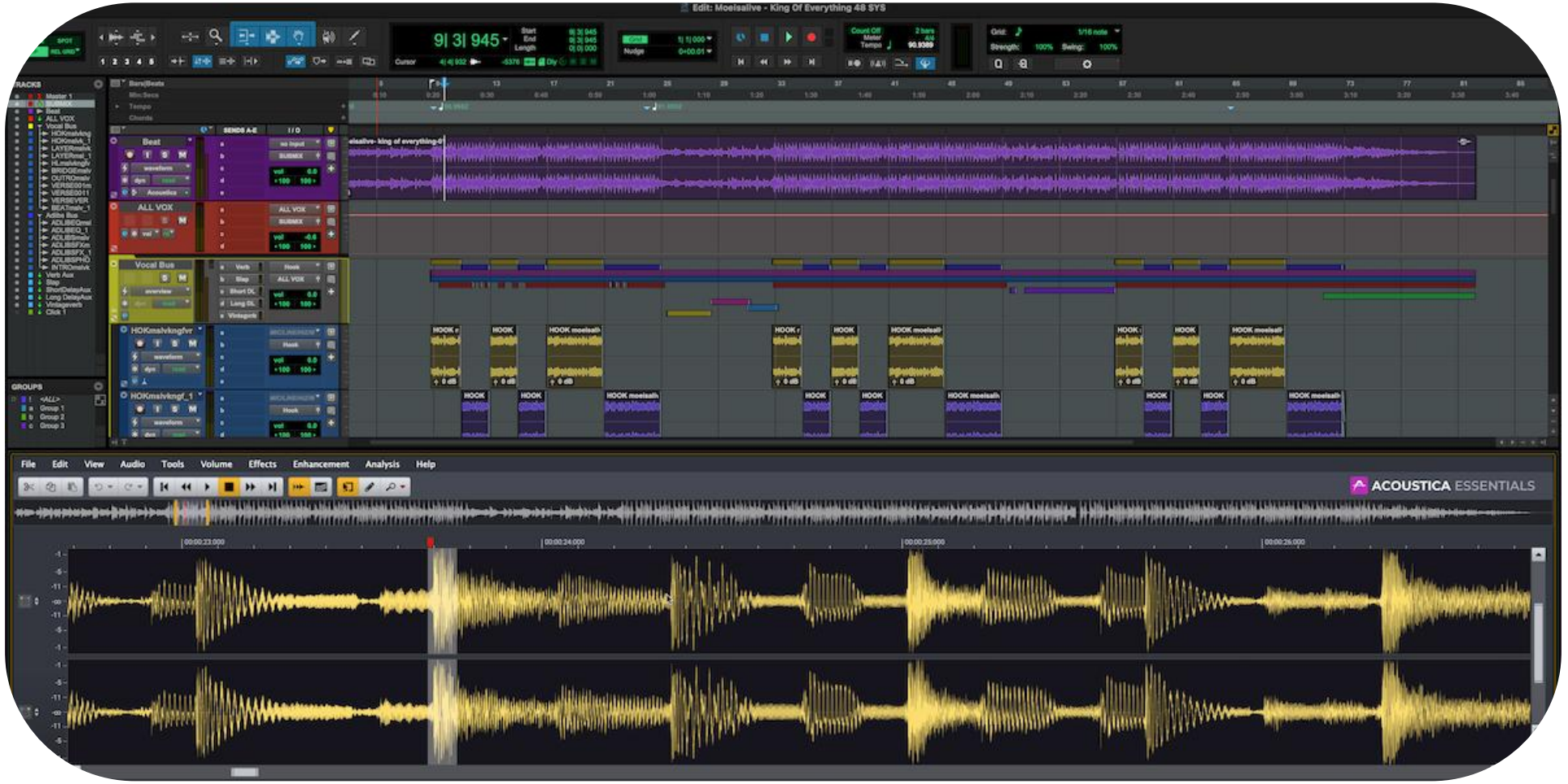


(Source: Wikimedia Commons)

Timo I. Denk, Yu Takagi, Takuya Matsuyama, Andrea Agostinelli, Tomoya Nakai, Christian Frank, and Shinji Nishimoto, "Brain2Music: Reconstructing Music from Human Brain Activity," *arXiv preprint arXiv:2307.11078*, 2023.

Intelligent Music Production

Music Production



(Source: Avid)

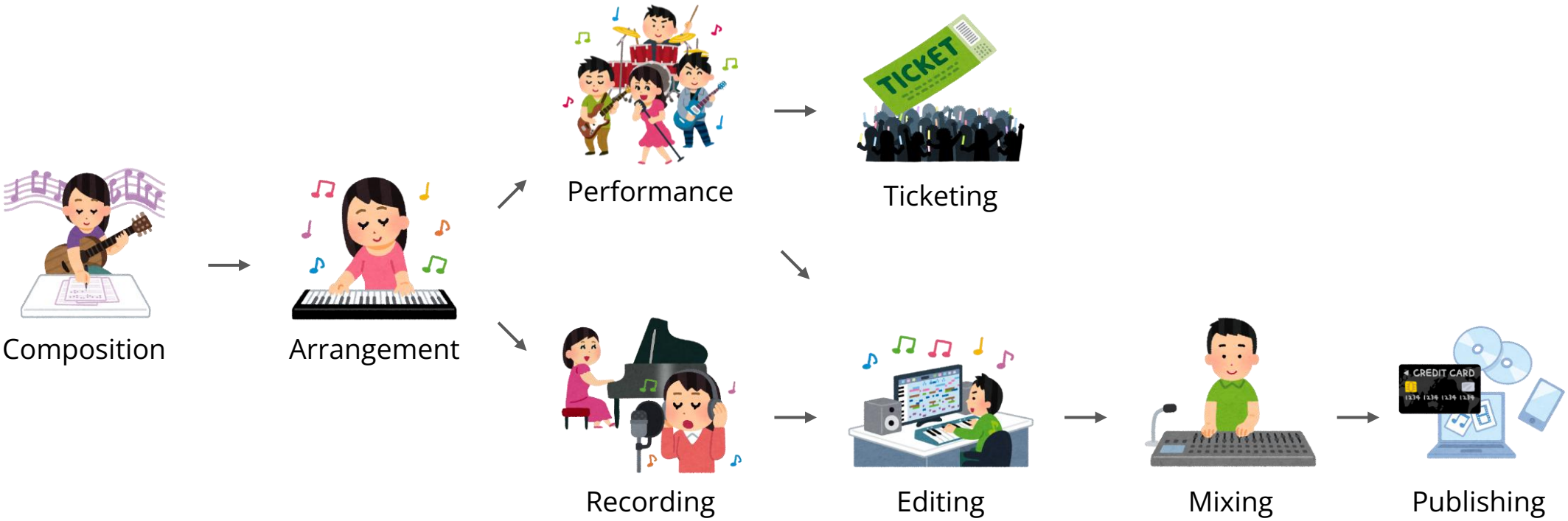
avid.com/pro-tools/whats-new

Music Production

(Source: Avid)

avid.com/pro-tools

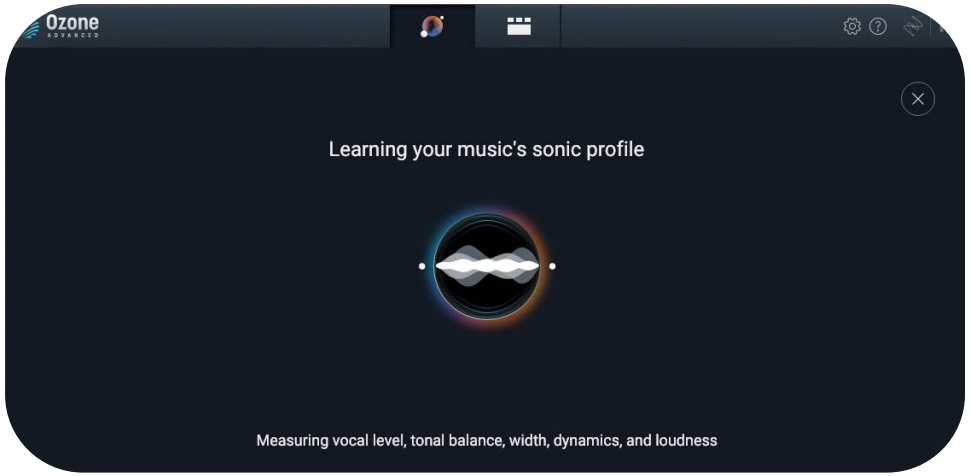
An Overly-Simplified Music Production Workflow



iZotope Neutron's Mix Assistant (2019)

youtu.be/gUD2G9nCc0k?t=458

iZotope Ozone's Master Assistant (2017)



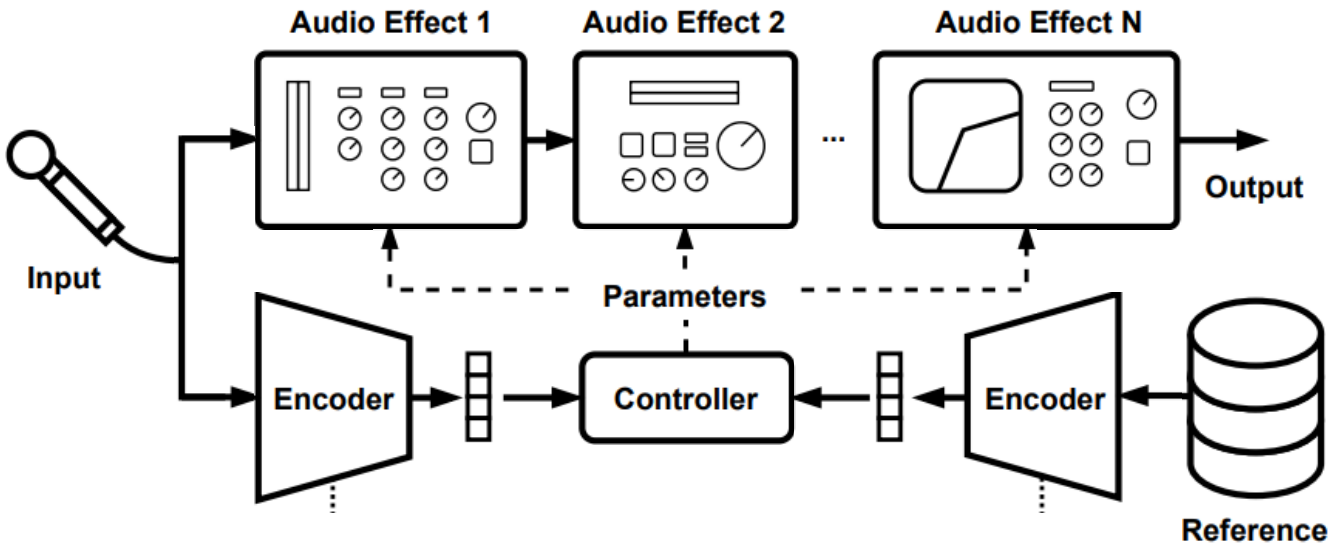
(Source: iZotope)

(Source: iZotope)

izotope.com/en/learn/10-steps-to-a-quick-master-in-ozone.html

Audio Effects Transfer

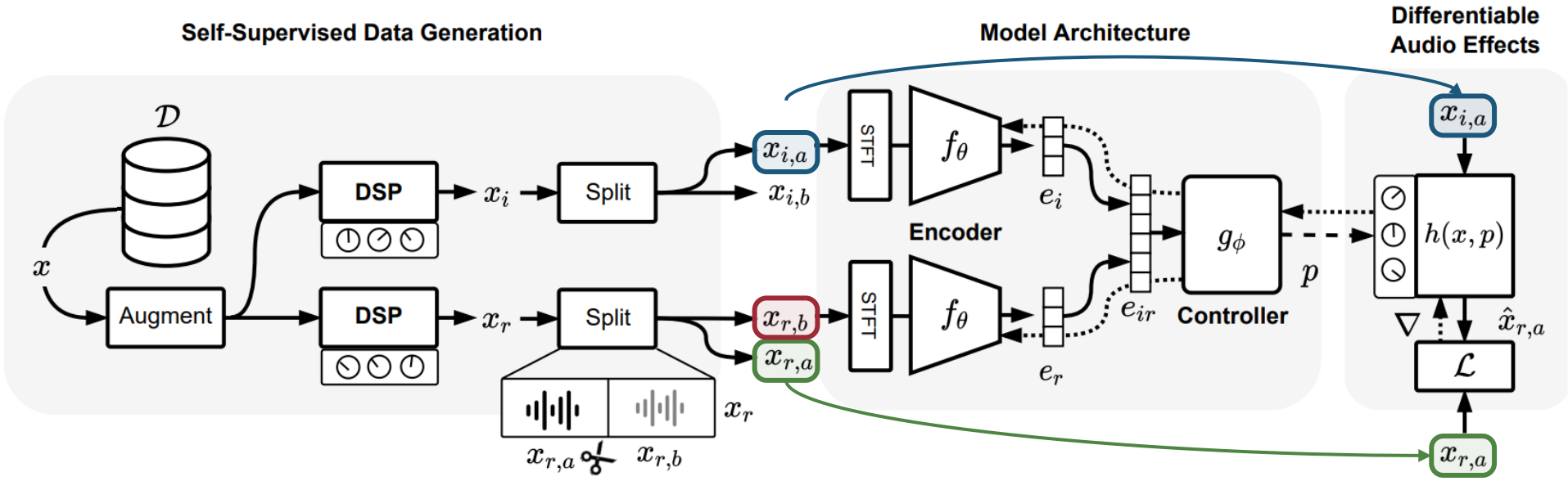
DeepAFx-ST: Effects Transfer (Steinmetz et al., 2022)



(Source: Steinmetz et al., 2022)

Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

DeepAFx-ST: Effects Transfer (Steinmetz et al., 2022)

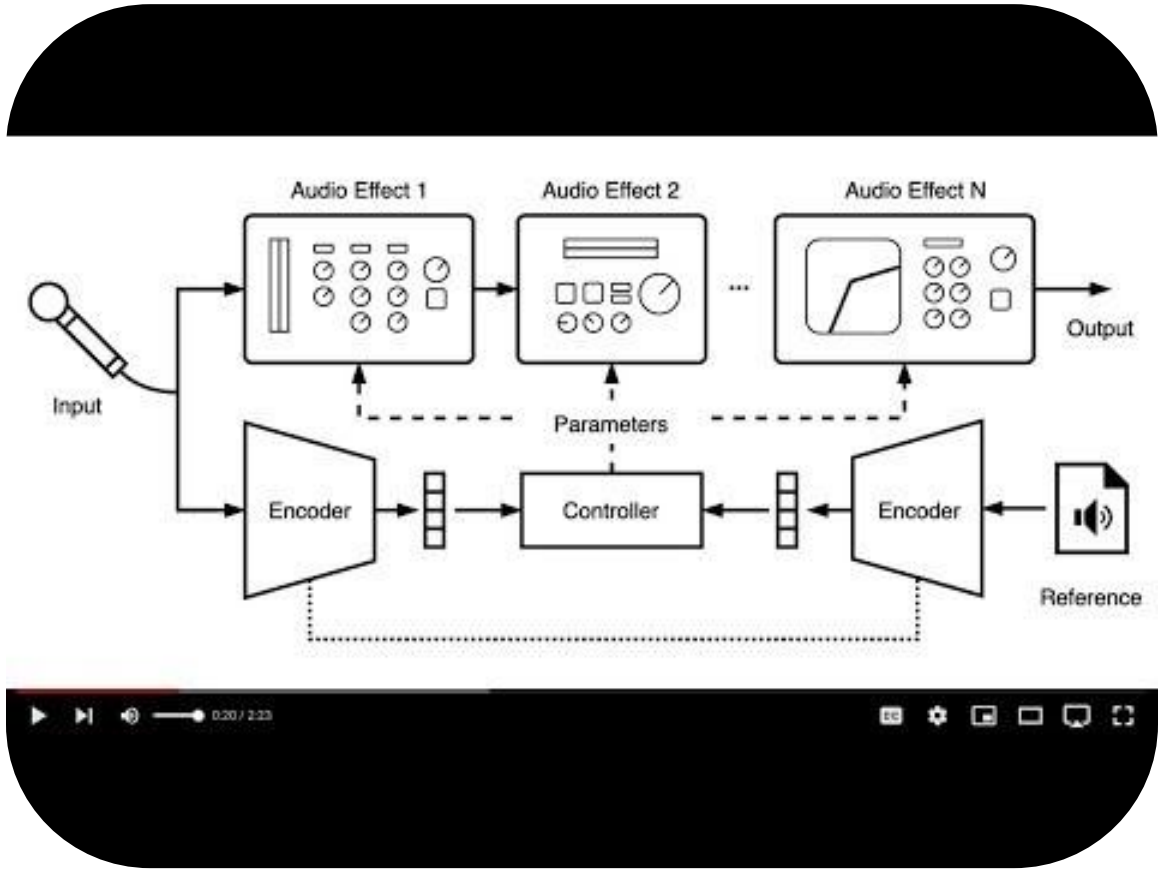


(Source: Steinmetz et al., 2022)

csteinmetz1.github.io/DeepAFx-ST

Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

DeepAFx-ST: Effects Transfer (Steinmetz et al., 2022)

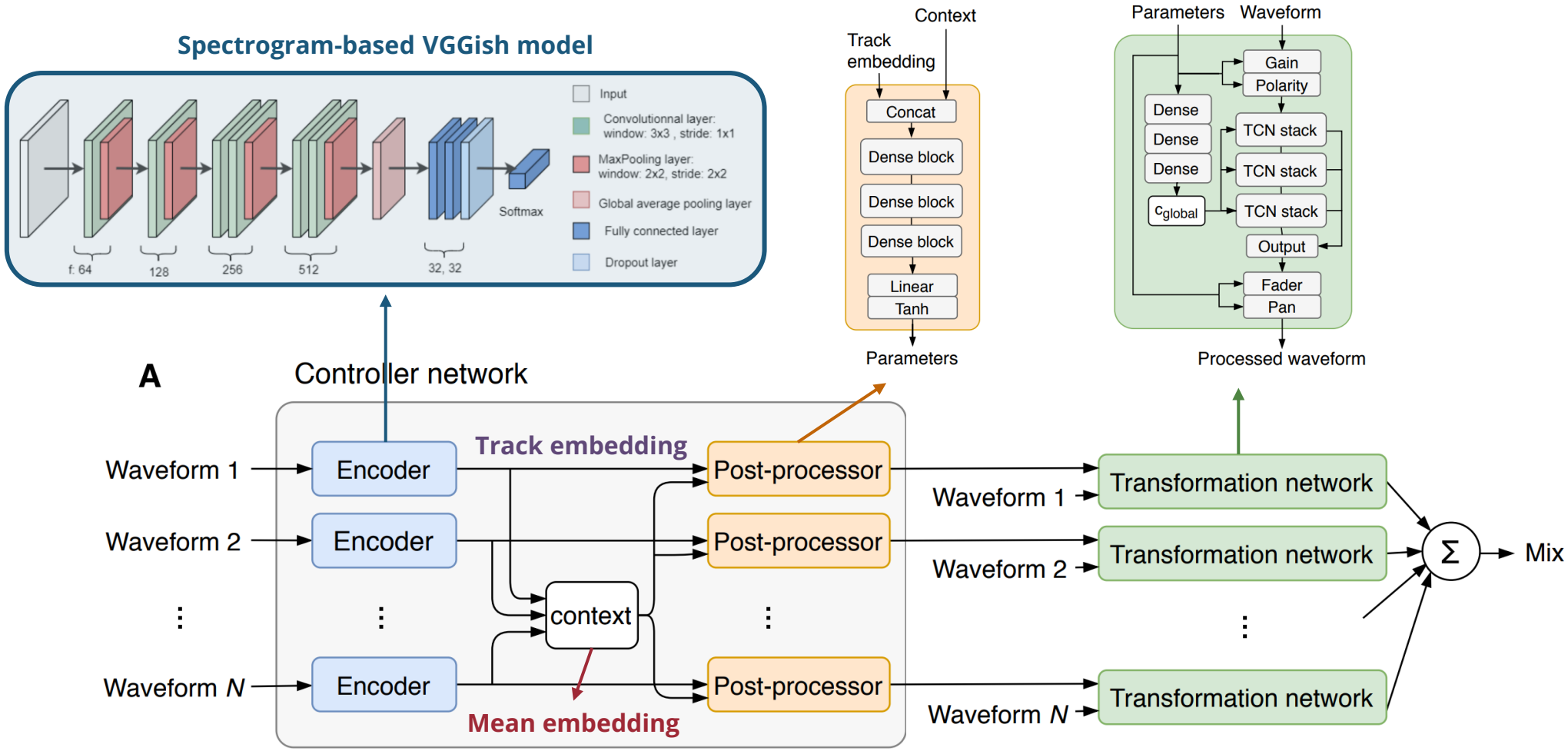


youtu.be/IZp455wiMk4

Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

Automatic Mixing

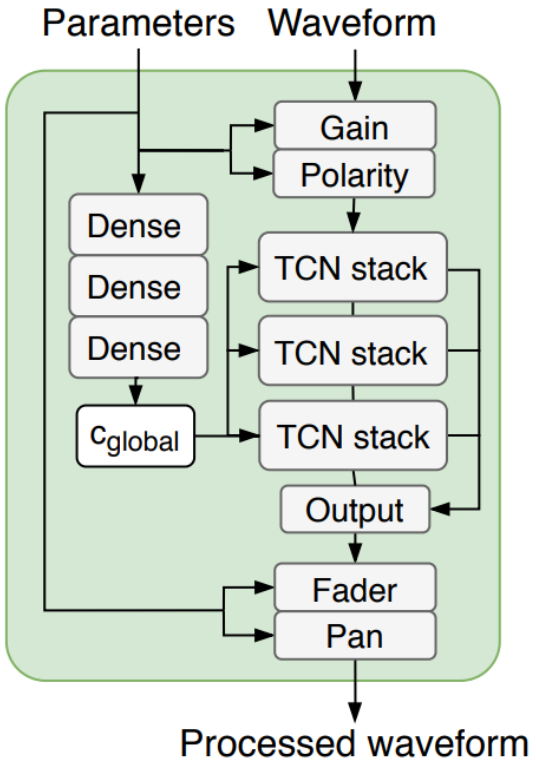
Differentiable Automatic Mixing (Steinmetz et al., 2021)



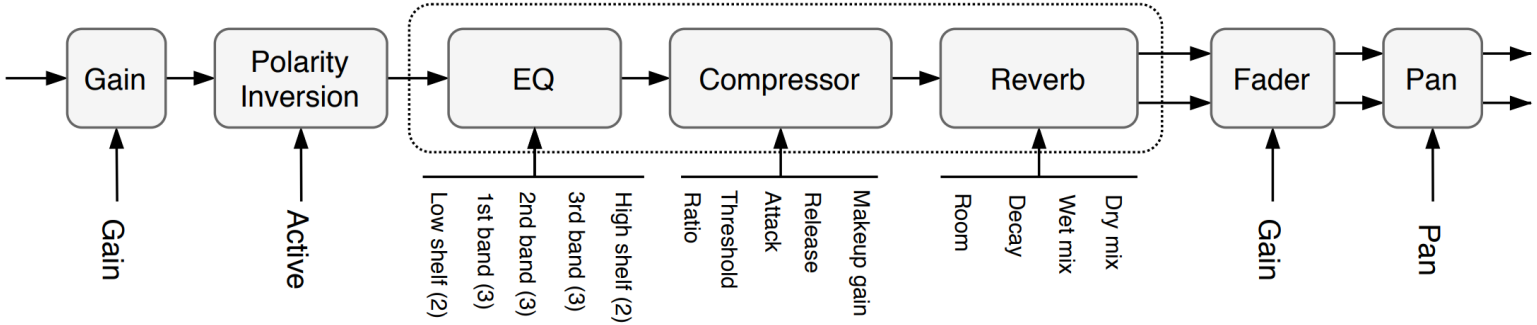
(Source: Steinmetz et al., 2021)

Christian J. Steinmetz, Jordi Pons, Santiago Pascual, and Joan Serrà, "Automatic multitrack mixing with a differentiable mixing console of neural audio effects," *ICASSP*, 2021.

Differentiable Automatic Mixing (Steinmetz et al., 2021)



(Source: Steinmetz et al., 2021)



(Source: Steinmetz et al., 2021)

A differentiable (and thus trainable) mixing console!

github.com/csteinmetz1/pymixconsole

Christian J. Steinmetz, Jordi Pons, Santiago Pascual, and Joan Serrà, "Automatic multitrack mixing with a differentiable mixing console of neural audio effects," *ICASSP*, 2021.

Differentiable Automatic Mixing (Steinmetz et al., 2021)

Transformation Network

Input	Target	Output
		
		

csteinmetz1.github.io/dmc-icassp2021

Christian J. Steinmetz, Jordi Pons, Santiago Pascual, and Joan Serrà, "Automatic multitrack mixing with a differentiable mixing console of neural audio effects," *ICASSP*, 2021.

Differentiable Automatic Mixing (Steinmetz et al., 2021)

Drum mixing

(Same mixing style)

DMC	Mono	Random	Target
			
			

Multitrack mixing

(Diverse mixing style)

DMC	Mono	Random	Target
			
			

csteinmetz1.github.io/dmc-icassp2021

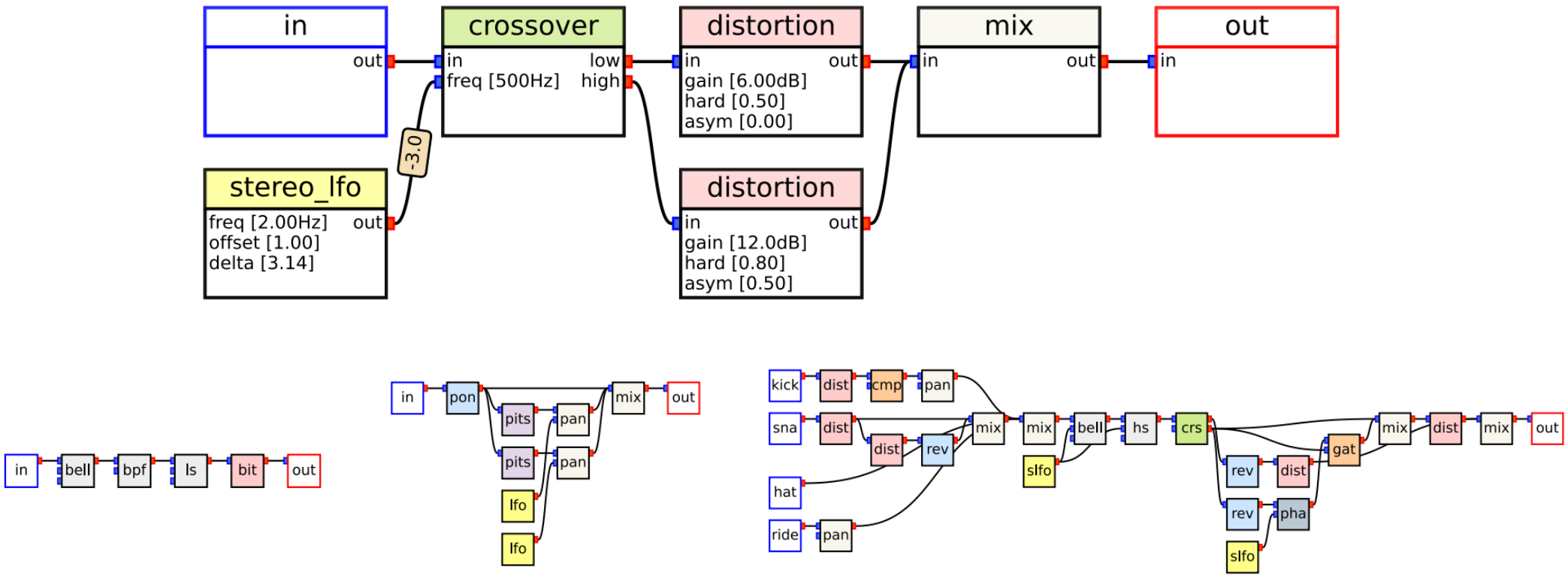
Christian J. Steinmetz, Jordi Pons, Santiago Pascual, and Joan Serrà, "Automatic multitrack mixing with a differentiable mixing console of neural audio effects," *ICASSP*, 2021.

Resources on Automatic Mixing

- Christian J. Steinmetz, Soumya Sai Vanka, Gary Bromham, and Marco A. Martínez Ramírez, "Deep Learning for Automatic Mixing," *Tutorials of ISMIR*, 2022.

Beyond Fixed Processing Graph

Estimating Audio Processing Graph (Lee et al., 2022)



Can we predict the audio processing graph used in a reference recording?

(Source: Lee et al., 2023)

Sungho Lee, Jaehyun Park, Seungryeol Paik, and Kyogu Lee, "Blind Estimation of Audio Processing Graph," *ICASSP*, 2023.

Estimating Audio Processing Graph (Lee et al., 2022)

Supported processors

Processor(s): [inlets, optional*] → [outlets]; [parameters].

Low-order linear filters [15]
- *Second-order low/band/highpass, bandreject, and fourth-order low/band/highpass:* [in, frequency*] → [out]; [frequency, q].
- *Parametric equalizer filters - low/highshelf and bell (peaking filter):* [in, frequency*, gain*] → [out]; [frequency, q, gain].
- *Crossover:* [in, frequency*] → [low, high]; [frequency].
- *Phaser:* [in, mod] → [out]; [frequency, feedback, mix].

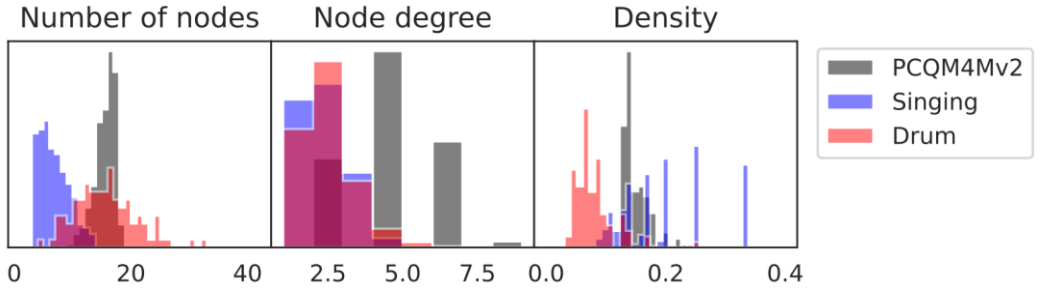
High-order linear filters [16]
- *Chorus/flanger/vibrato:* [in, mod] → [out]; [delay, feedback, mix].
- *Mono and pingpong delay:* [in] → [out]; [delay, feedback, mix, frequency, q, stereo_offset].
- *Reverb (mono and stereo):* [in] → [out]; [size, damping, width, mix].

Nonlinear filters
- *Distortion* [17]: [in] → [out]; [gain, hardness, asymmetry].
- *Bitcrush:* [in] → [out]; [bit].
- *Dynamic range controllers - compressor/noisegate/expander* [18]: [in, sidechain*] → [out]; [threshold, ratio, attack, release, knee].
- *Pitchshift*: [in] → [out]; [semitone].

Utility processors
- *Mix:* [in] → [out]; [].
- *Panning:* [in, pan*] → [out]; [pan].
- *Imager:* [in] → [out]; [width].
- *Mid/side splitter:* [in] → [mid, side]; [].
- *Mid/side merger:* [mid, side] → [out]; [].

Control signal generators
- *Low-frequencyuency oscillator (mono and stereo):* [] → [lfo]; [frequency, phase, stereo_offset].
- *Envelope follower:* [in] → [env]; [attack, release, gain].

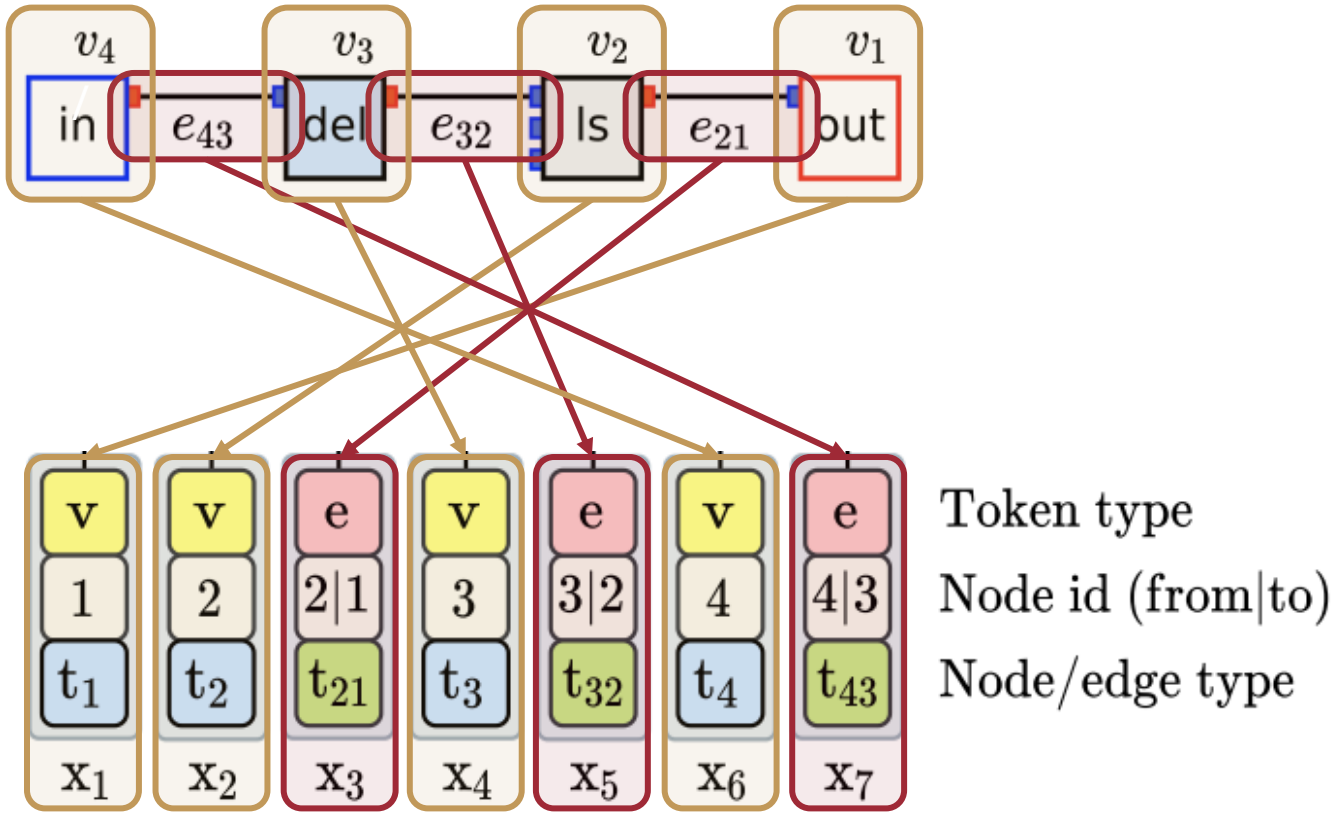
Data statistics



(Source: Lee et al., 2023)

Sungho Lee, Jaehyun Park, Seungryeol Paik, and Kyogu Lee, "Blind Estimation of Audio Processing Graph," *ICASSP*, 2023.

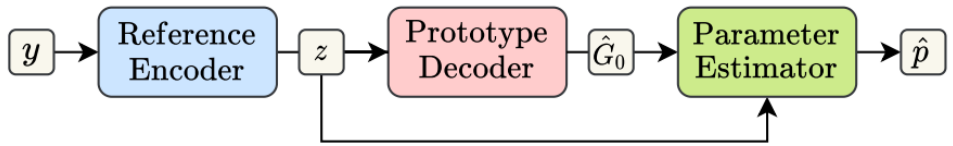
Tokenizing a Graph



Sungho Lee, Jaehyun Park, Seungryeol Paik, and Kyogu Lee, "Blind Estimation of Audio Processing Graph," *ICASSP*, 2023.

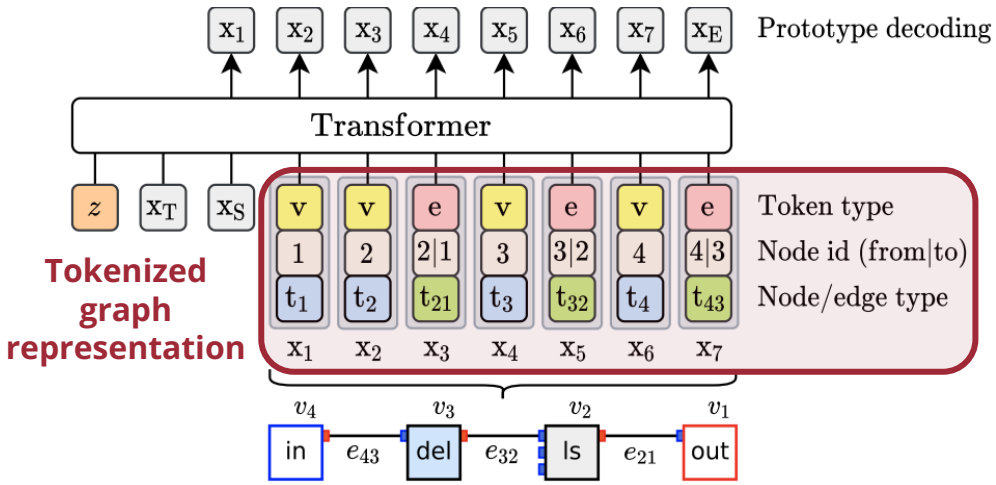
Estimating Audio Processing Graph (Lee et al., 2022)

Blind estimation framework

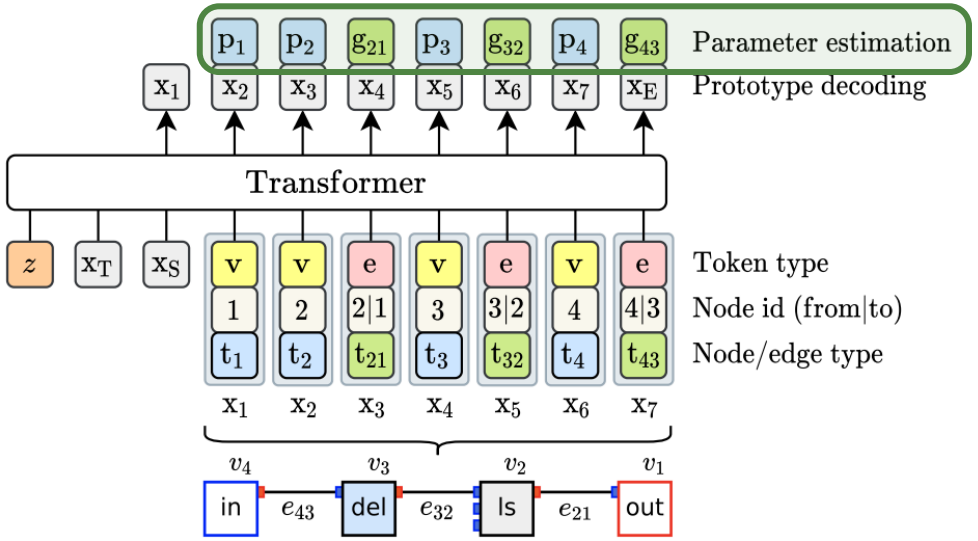


Sources are not given!

Prototype decoder



Parameter estimator



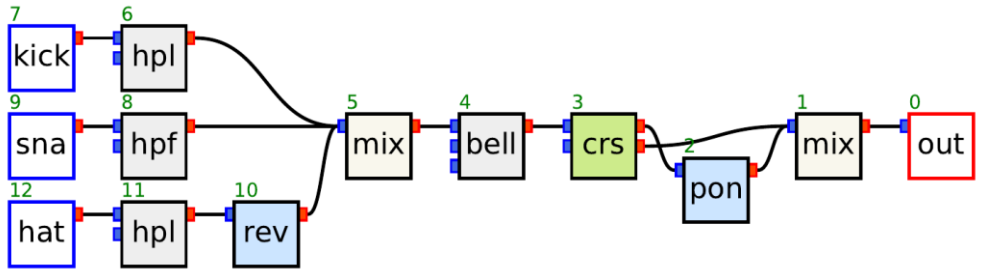
(Source: Lee et al., 2023)

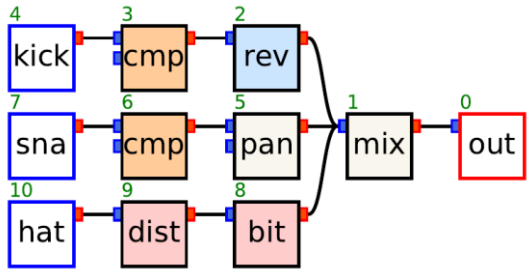
Sungho Lee, Jaehyun Park, Seungryeol Paik, and Kyogu Lee, "Blind Estimation of Audio Processing Graph," *ICASSP*, 2023.

Estimating Audio Processing Graph (Lee et al., 2022)

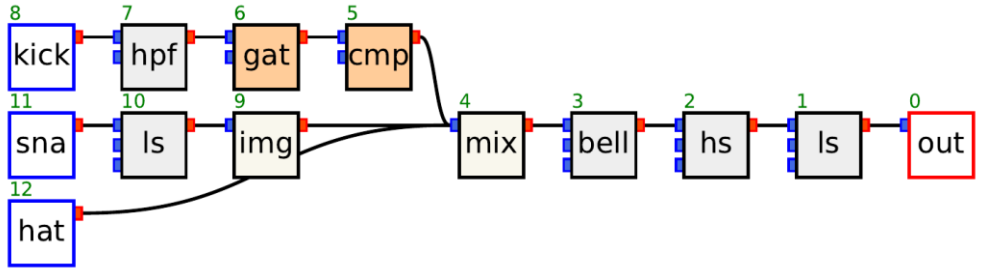
Dry

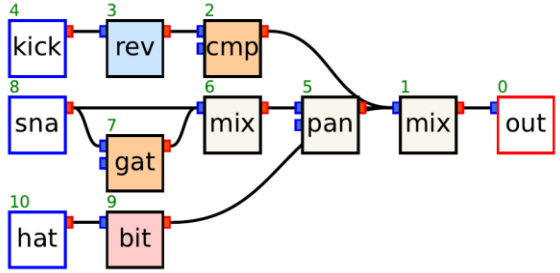
Reference





Estimation



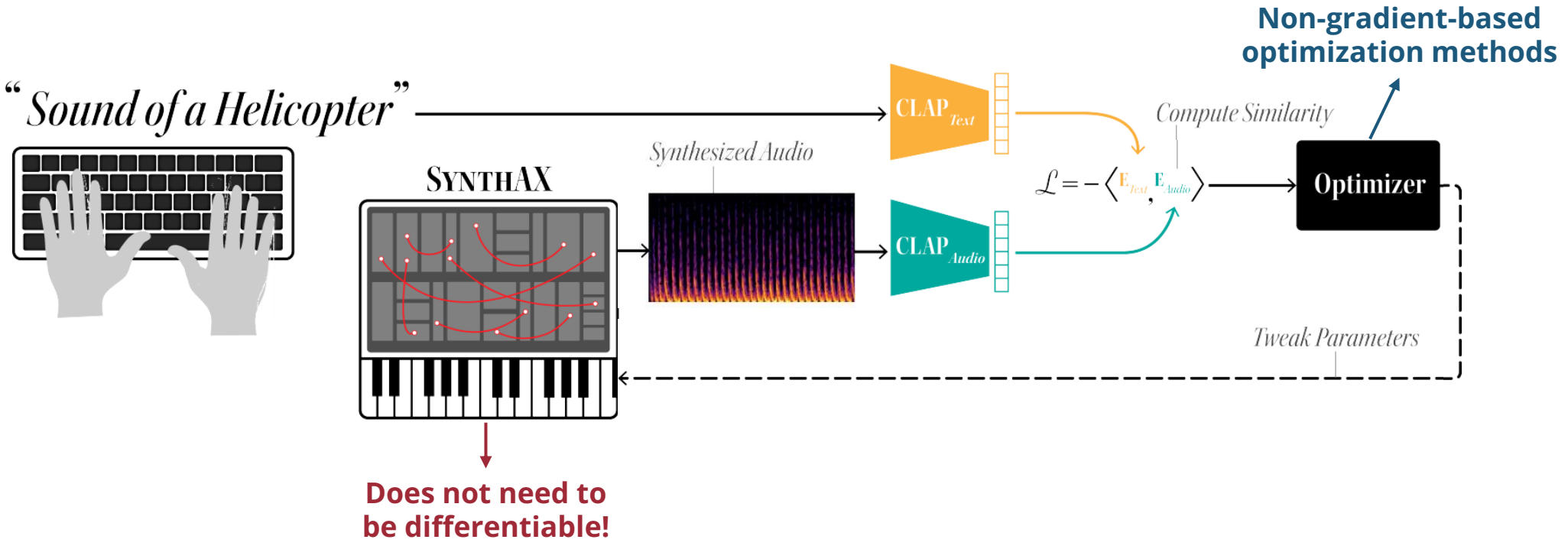


(Source: Lee et al., 2023)

sh-lee97.github.io/apg

Sungho Lee, Jaehyun Park, Seungryeol Paik, and Kyogu Lee, "Blind Estimation of Audio Processing Graph," *ICASSP*, 2023.

CTAG: Synthesizer Programming (Cherep et al., 2024)



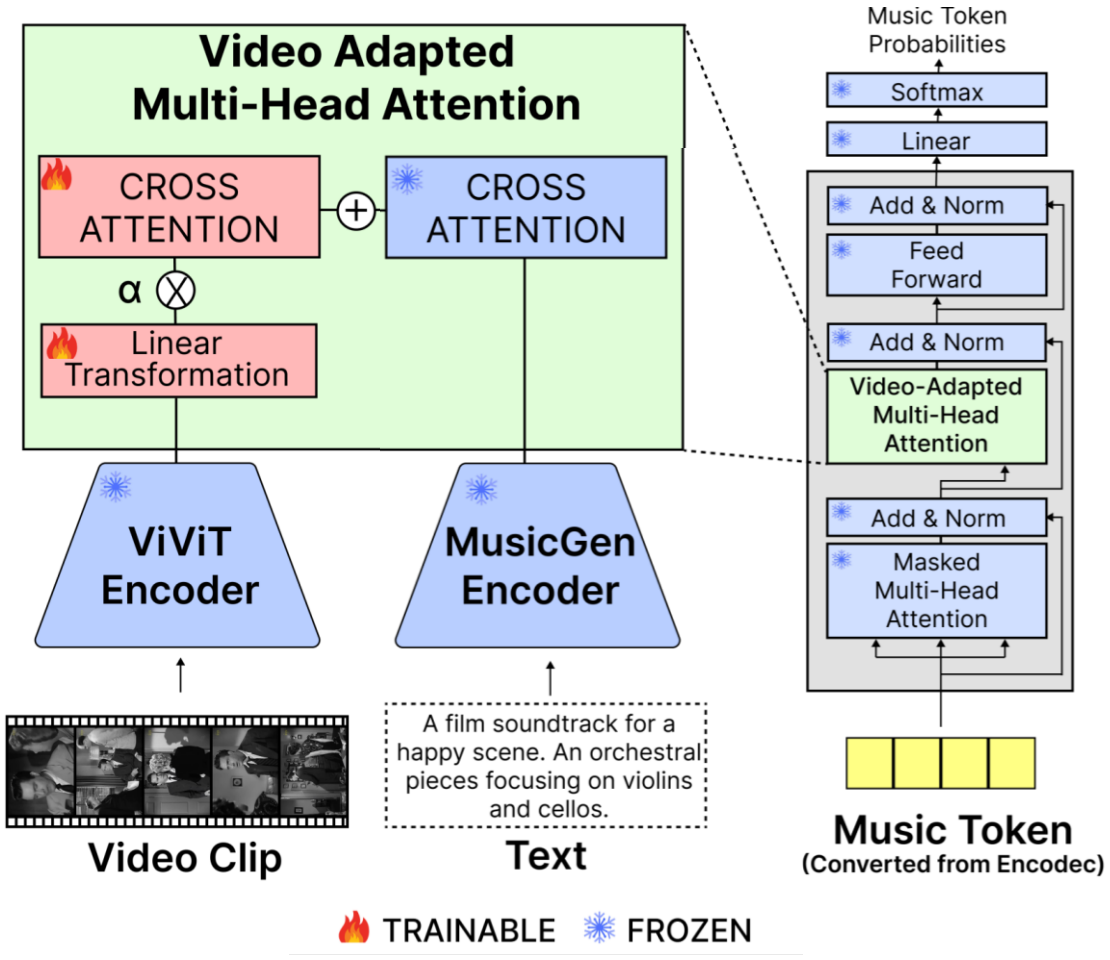
(Source: Cherep et al., 2024)

ctag.media.mit.edu

Manuel Cherep, Nikhil Singh, and Jessica Shand, "Creative Text-to-Audio Generation via Synthesizer Programming," *ICML*, 2024.

Recap

Video-Guided Text-to-Music Generation (Kim et al., 2025)



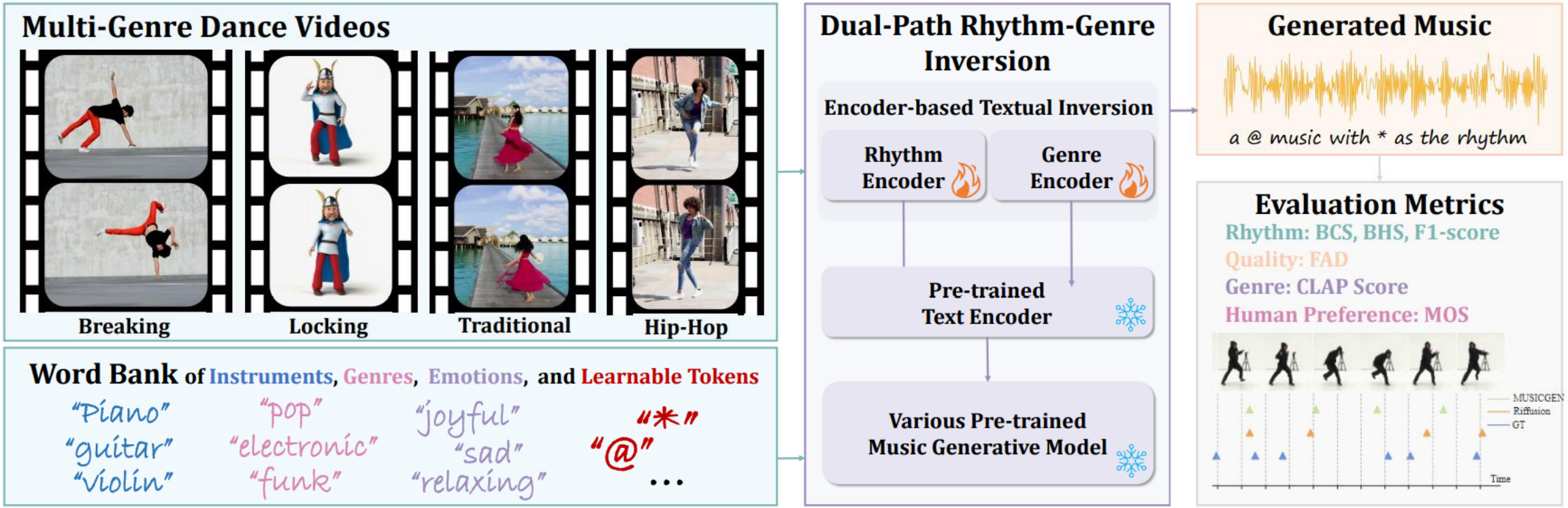
(Source: Kim et al., 2025)

Dance, Music & AI

(Source: Yamaha)

yamaha.com/en/news_release/2018/18013101/
Yamaha Global, "Yamaha Artificial Intelligence (AI) Transforms a Dancer into a Pianist - Short Version," *YouTube*, youtu.be/21injmy1wsU, 2018.

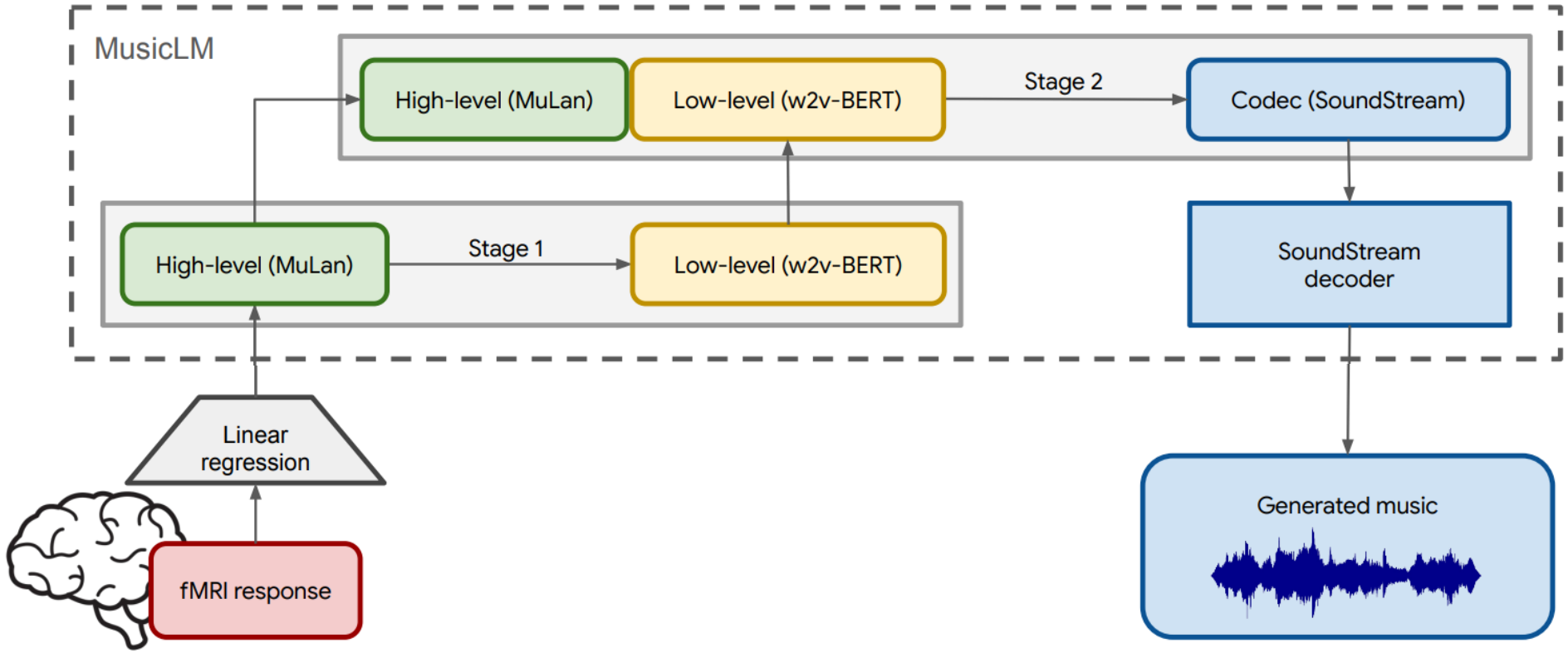
Dance-to-Music Generation (Li et al., 2024)



(Source: Li et al., 2024)

Sifei Li, Weiming Dong, Yuxin Zhang, Fan Tang, Chongyang Ma, Oliver Deussen, Tong-Yee Lee, and Changsheng Xu, "Dance-to-Music Generation with Encoder-based Textual Inversion," *SIGGRAPH ASIA*, 2024.

Brain2Music (Denk et al., 2023)

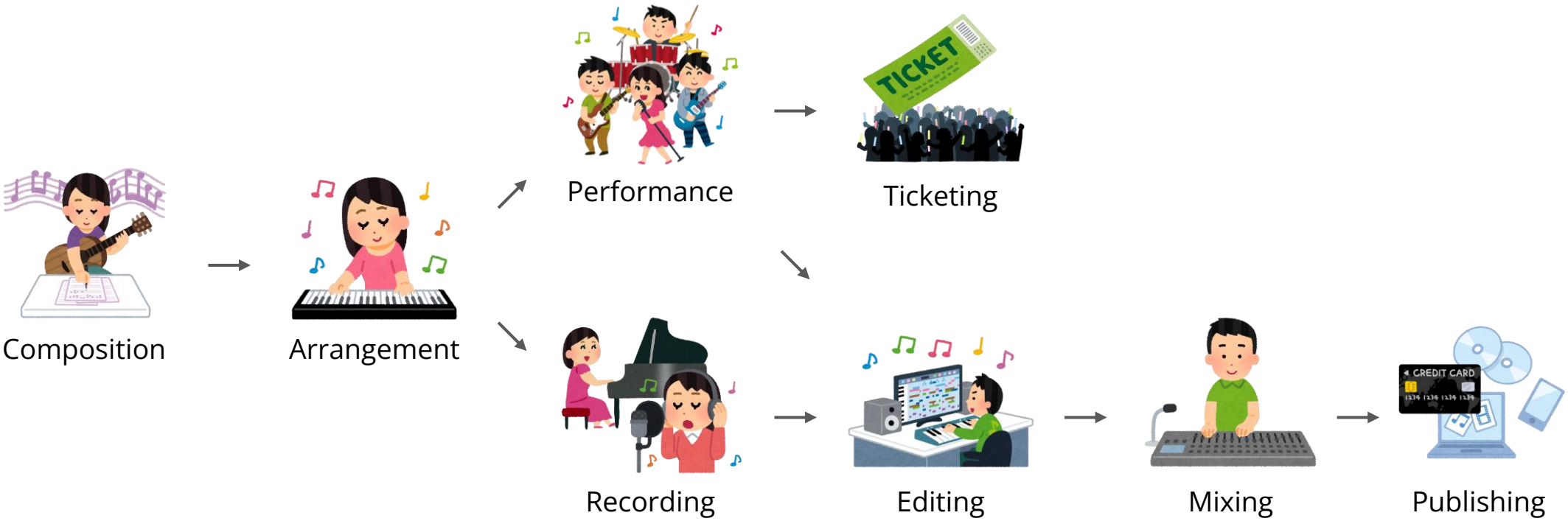


(Source: Denk et al., 2023)

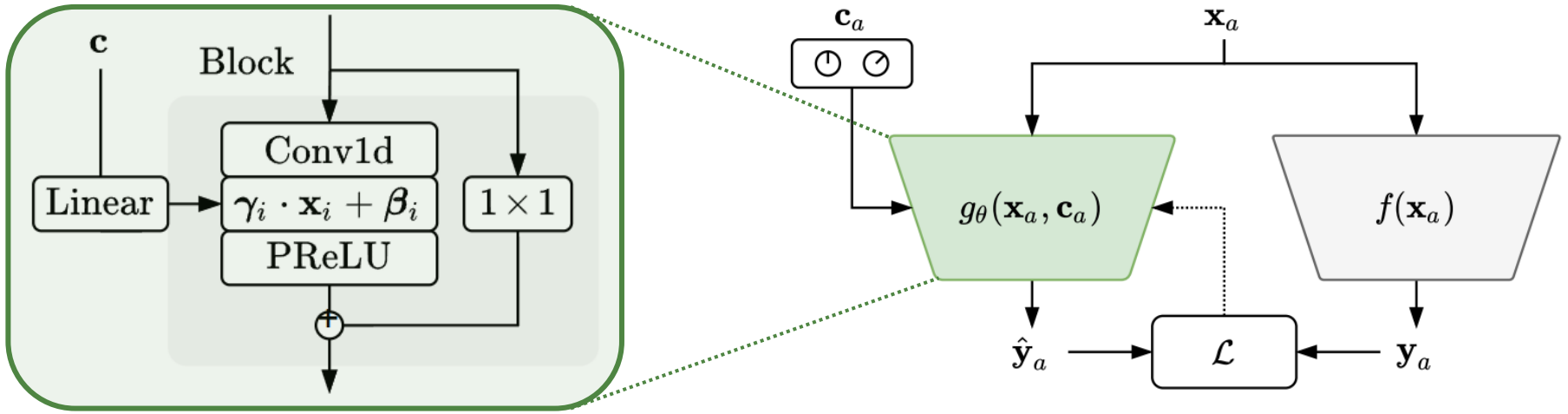
google-research.github.io/seanet/brain2music

Timo I. Denk, Yu Takagi, Takuya Matsuyama, Andrea Agostinelli, Tomoya Nakai, Christian Frank, and Shinji Nishimoto, "Brain2Music: Reconstructing Music from Human Brain Activity," *arXiv preprint arXiv:2307.11078*, 2023.

An Overly-Simplified Music Production Workflow



Neural Audio Effects (Steinmetz et al., 2021)

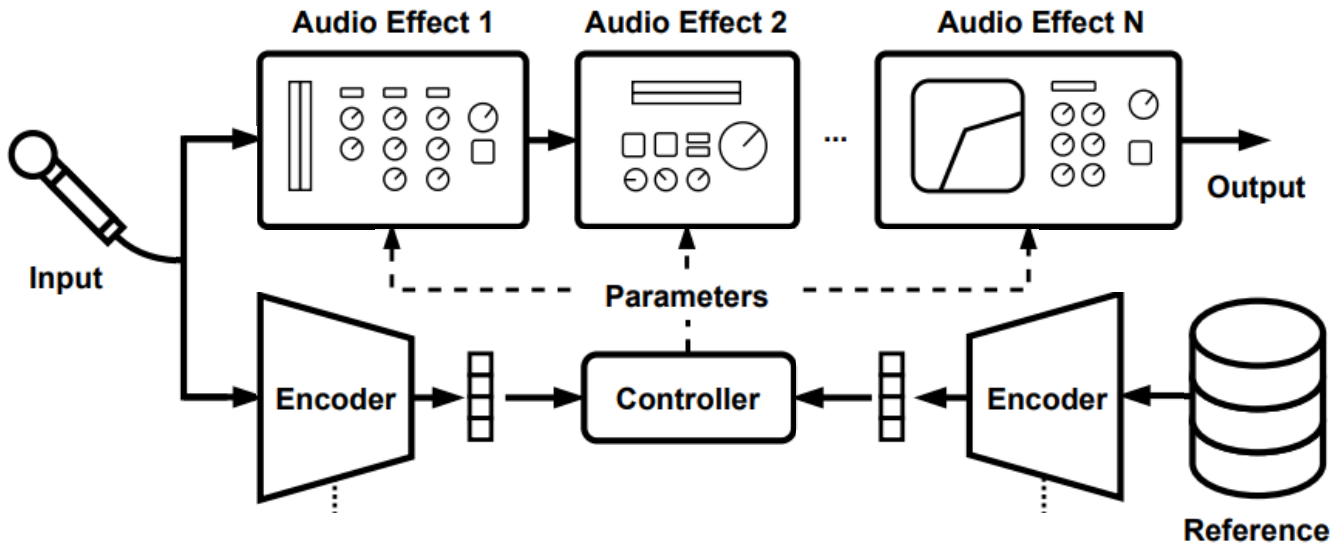


(Source: Steinmetz et al., 2021)

csteinmetz1.github.io/steerable-nafx

Christian J. Steinmetz and Joshua D. Reiss, "Steerable discovery of neural audio effects," *NeurIPS ML4CD Workshop*, 2021.

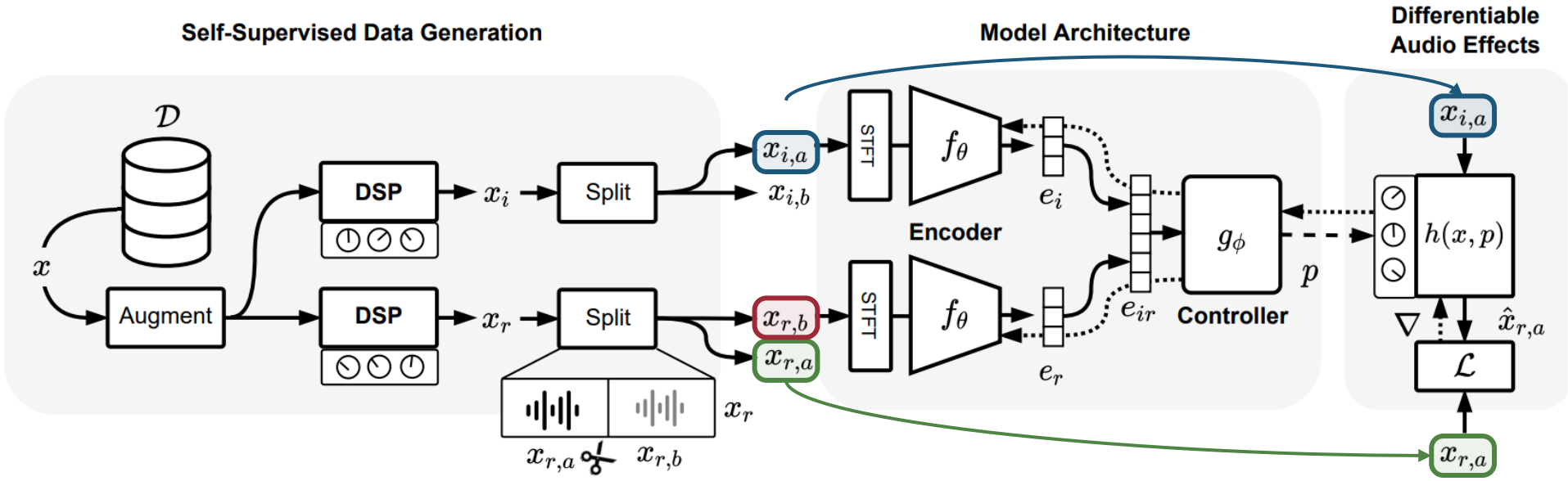
DeepAFx-ST: Effects Transfer (Steinmetz et al., 2022)



(Source: Steinmetz et al., 2022)

Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

DeepAFx-ST: Effects Transfer (Steinmetz et al., 2022)

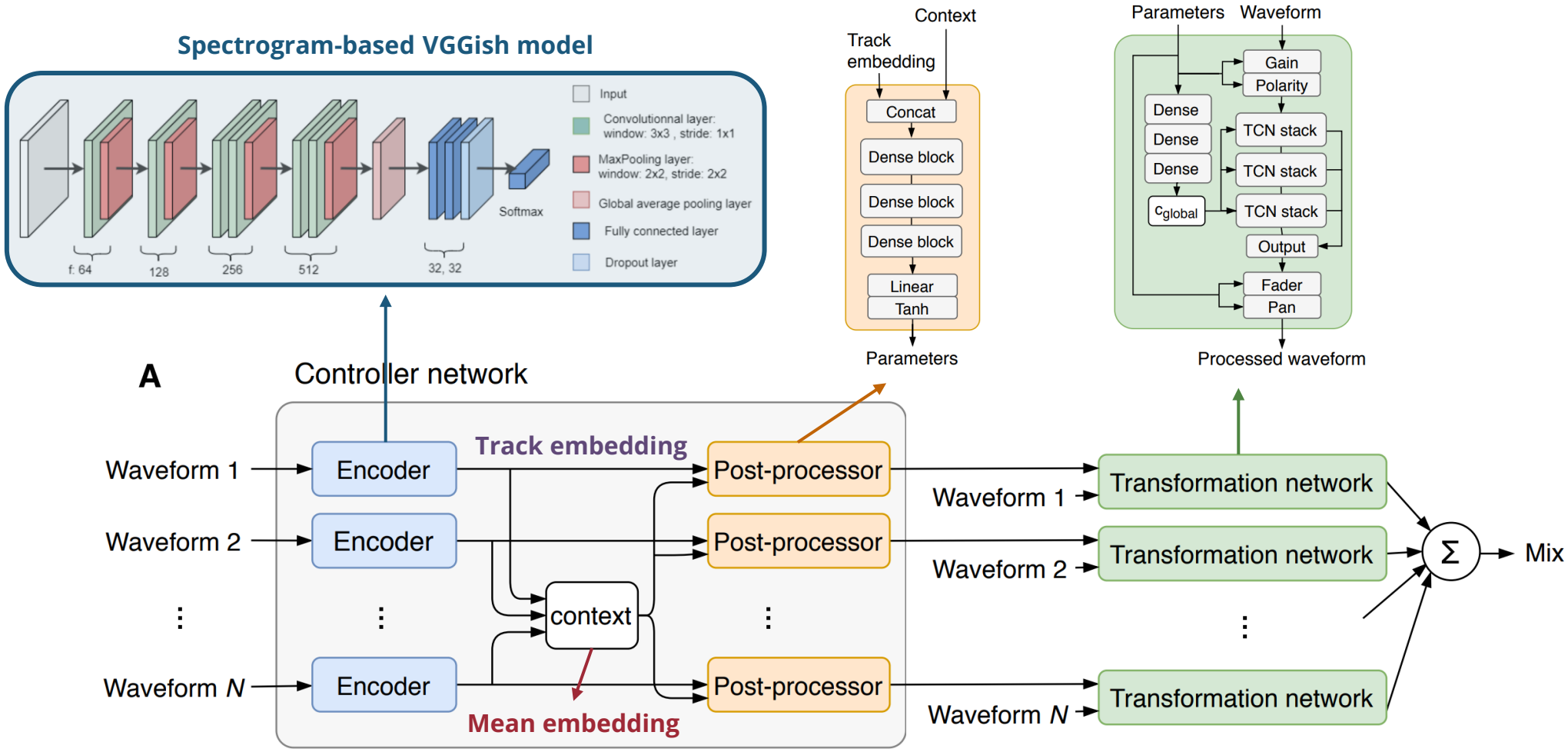


(Source: Steinmetz et al., 2022)

csteinmetz1.github.io/DeepAFx-ST

Christian J. Steinmetz, Nicholas J. Bryan, and Joshua D. Reiss, "Style Transfer of Audio Effects with Differentiable Signal Processing," *JAES*, 2022.

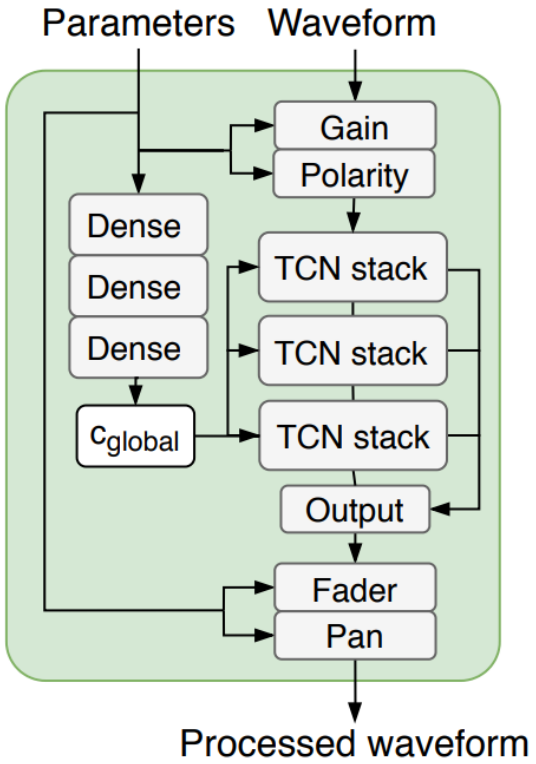
Differentiable Automatic Mixing (Steinmetz et al., 2021)



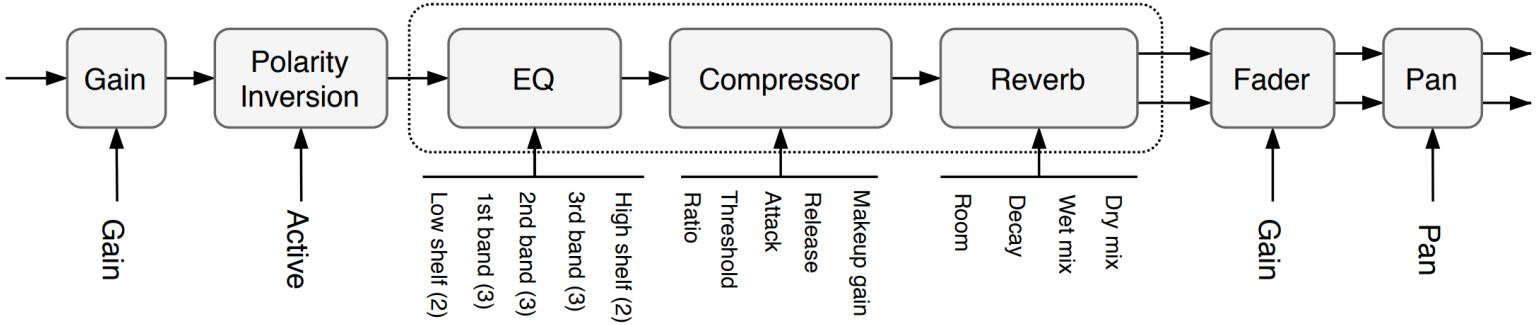
(Source: Steinmetz et al., 2021)

Christian J. Steinmetz, Jordi Pons, Santiago Pascual, and Joan Serrà, "Automatic multitrack mixing with a differentiable mixing console of neural audio effects," *ICASSP*, 2021.

Differentiable Automatic Mixing (Steinmetz et al., 2021)



(Source: Steinmetz et al., 2021)



(Source: Steinmetz et al., 2021)

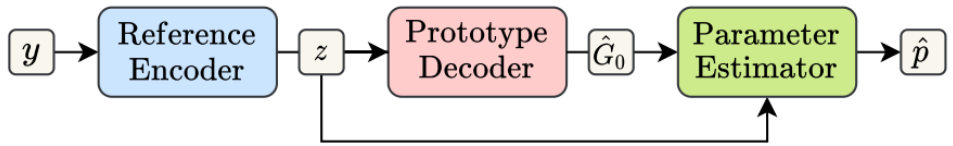
A differentiable (and thus trainable) mixing console!

github.com/csteinmetz1/pymixconsole

Christian J. Steinmetz, Jordi Pons, Santiago Pascual, and Joan Serrà, "Automatic multitrack mixing with a differentiable mixing console of neural audio effects," *ICASSP*, 2021.

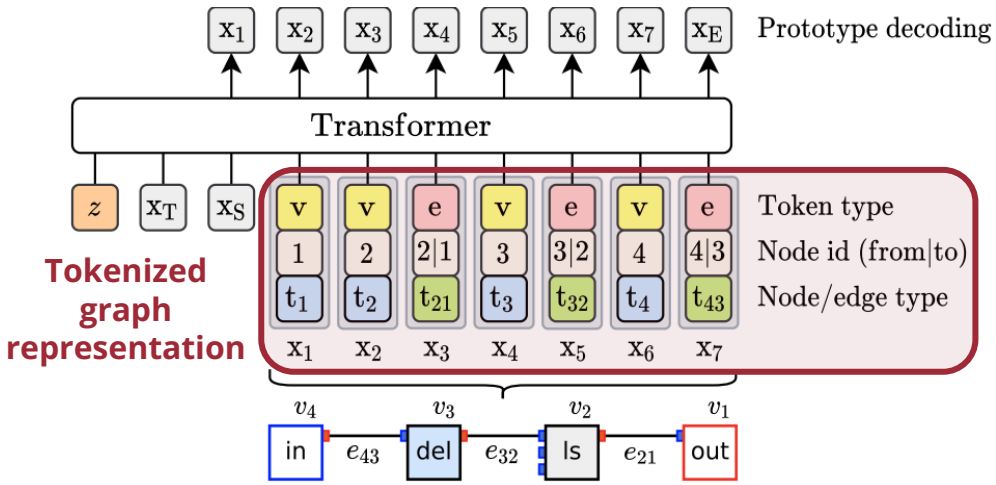
Estimating Audio Processing Graph (Lee et al., 2022)

Blind estimation framework

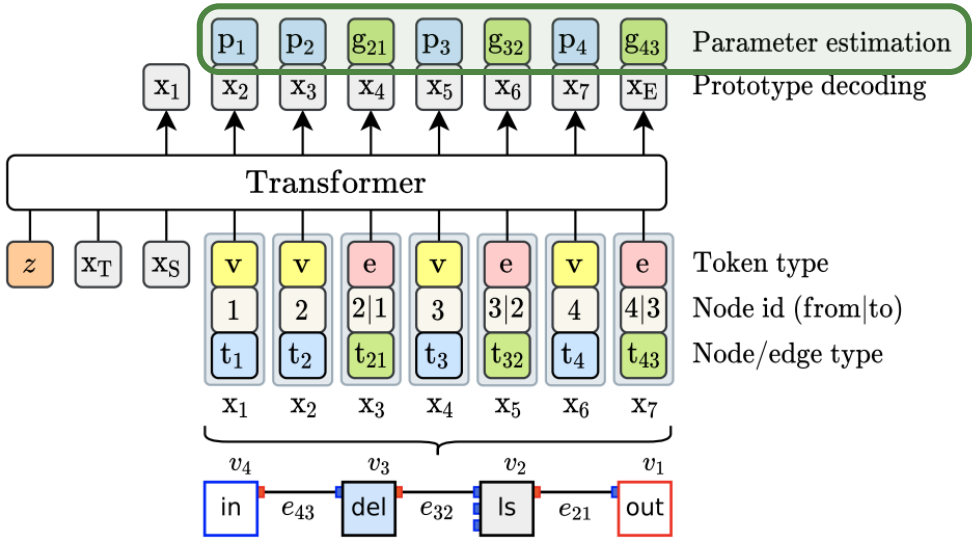


Sources are not given!

Prototype decoder



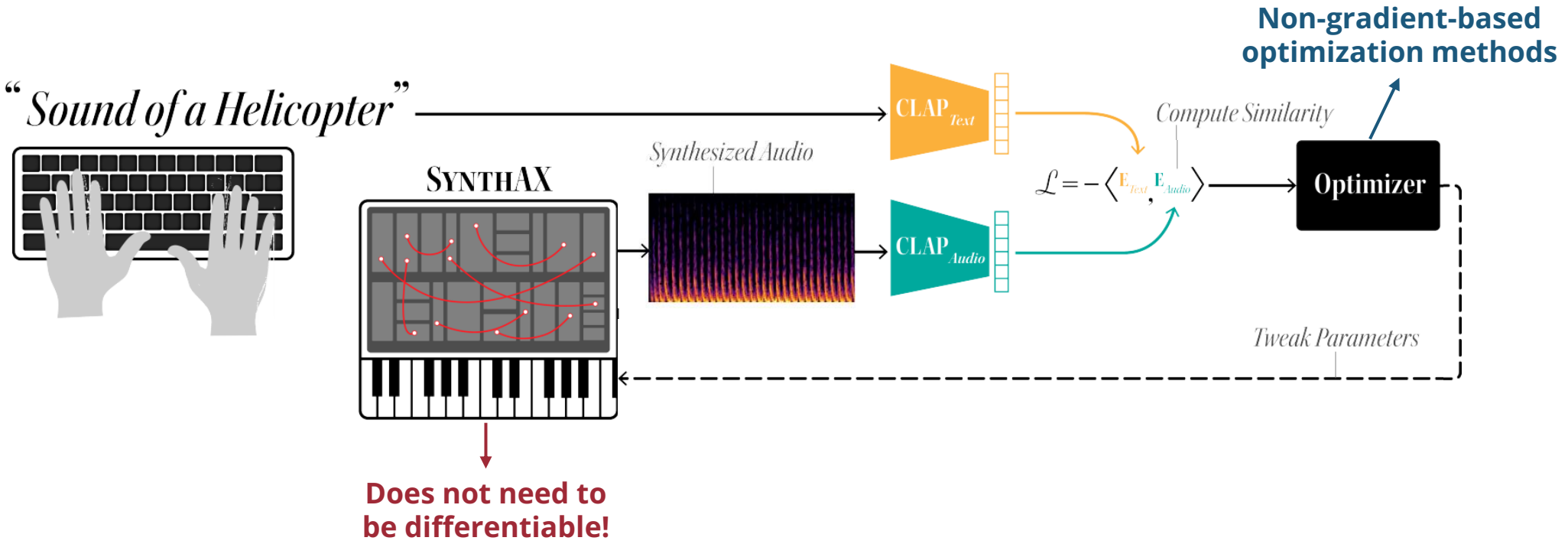
Parameter estimator



(Source: Lee et al., 2023)

Sungho Lee, Jaehyun Park, Seungryeol Paik, and Kyogu Lee, "Blind Estimation of Audio Processing Graph," *ICASSP*, 2023.

CTAG: Synthesizer Programming (Cherep et al., 2024)



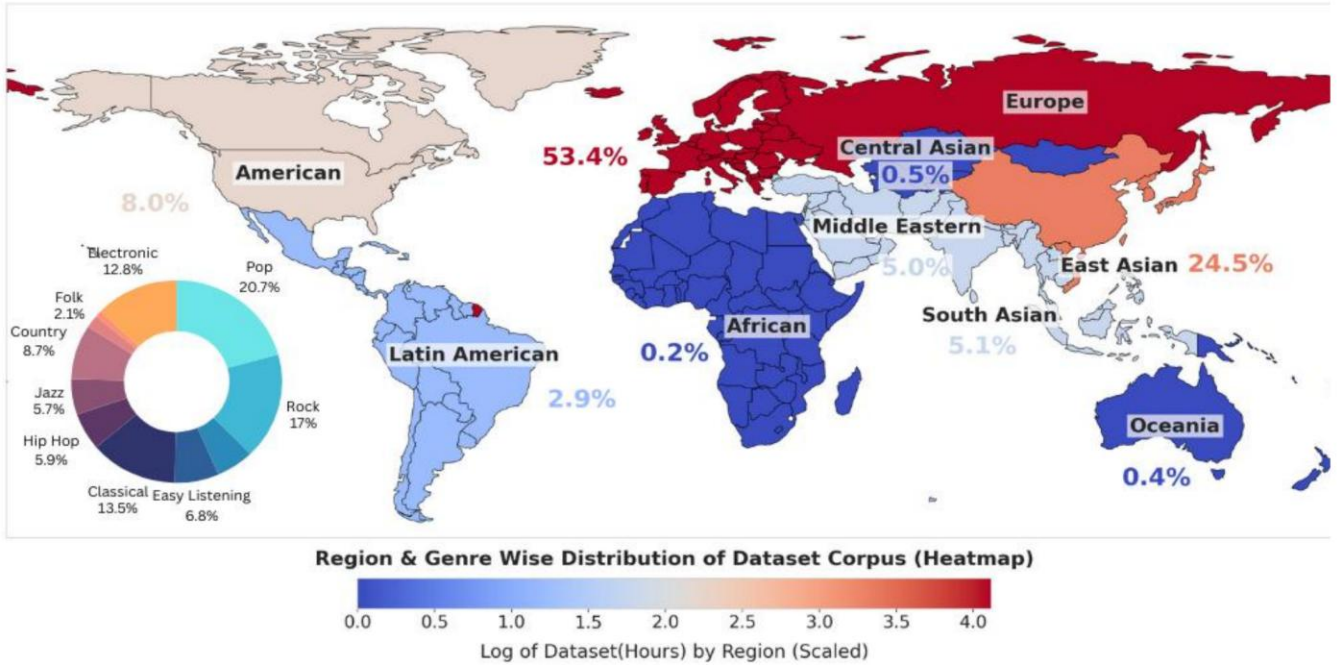
(Source: Cherep et al., 2024)

ctag.media.mit.edu

Manuel Cherep, Nikhil Singh, and Jessica Shand, "Creative Text-to-Audio Generation via Synthesizer Programming," *ICML*, 2024.

Next Lecture

Discussions, Challenges & Opportunities



(Source: Mehta et al., 2024)