PAT 498/598 (Winter 2025)

Music & AI

Lecture 19: Music Production & Editing

Instructor: Hao-Wen Dong

SCHOOL OF MUSIC, THEATRE & DANCE
PERFORMING ARTS TECHNOLOGY
UNIVERSITY OF MICHIGAN

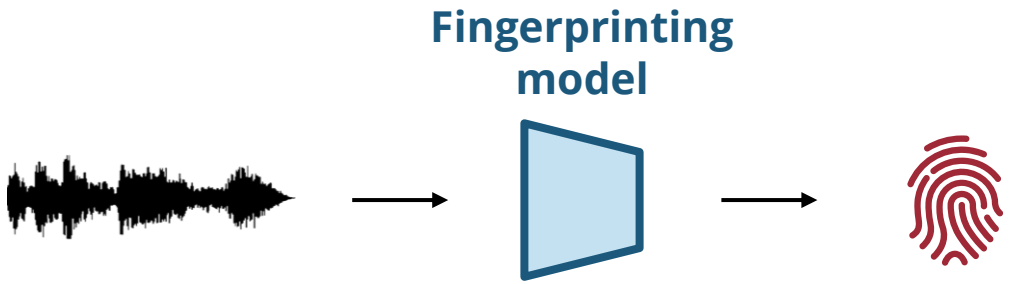
(Recap) Shazam & Siri

(Source: Shazam User Guide)

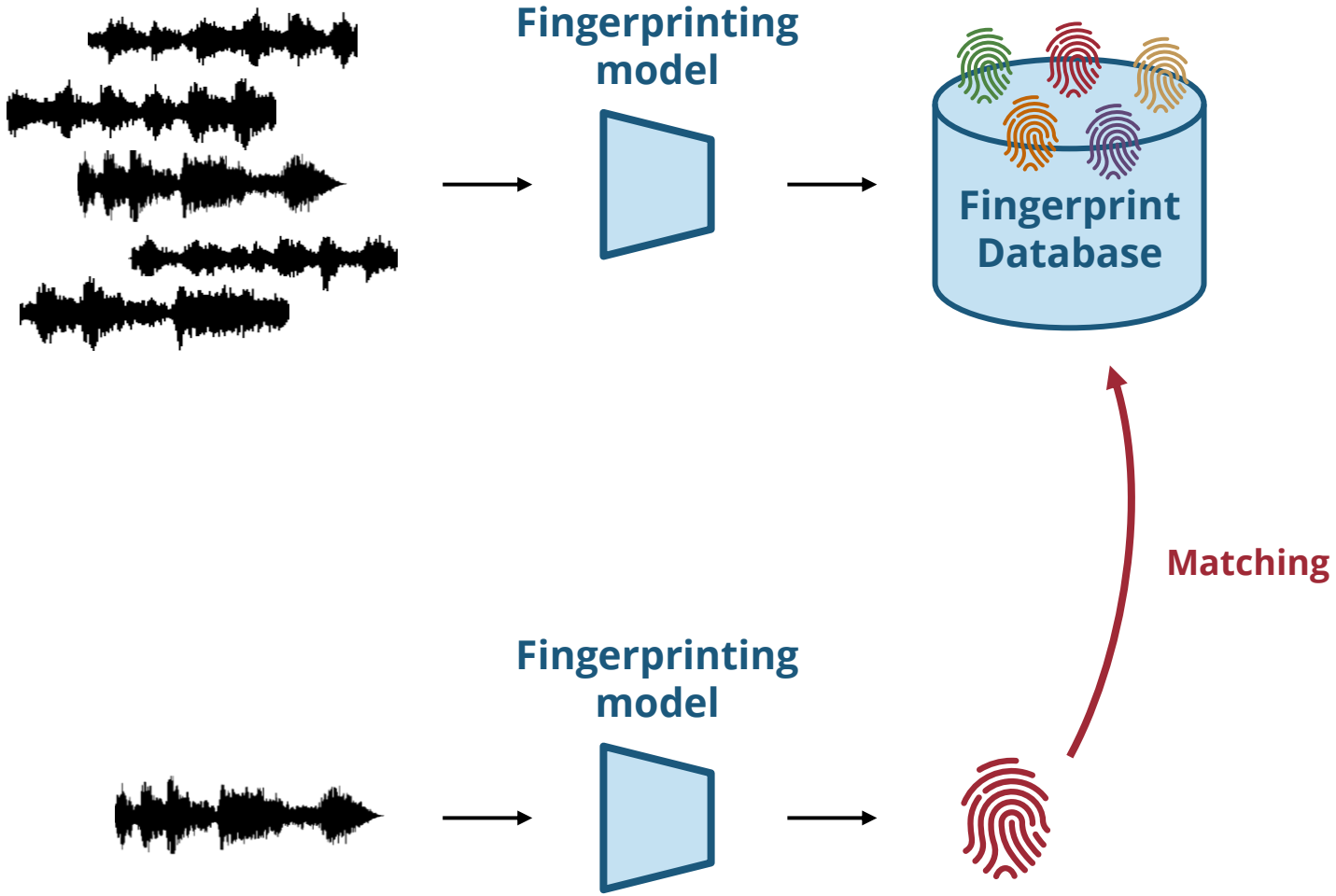
(Source: OSXDaily)

osxdaily.com/2017/09/08/identify-song-playing-mac-siri/
support.apple.com/en-is/guide/shazam/dev9748744b6/web

(Recap) Audio Fingerprinting

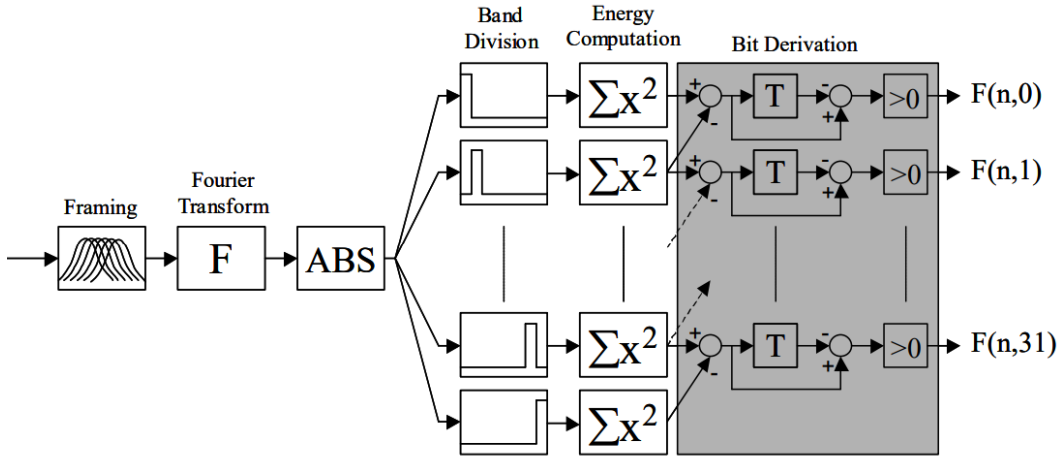


(Recap) Audio Fingerprinting for Audio Identification

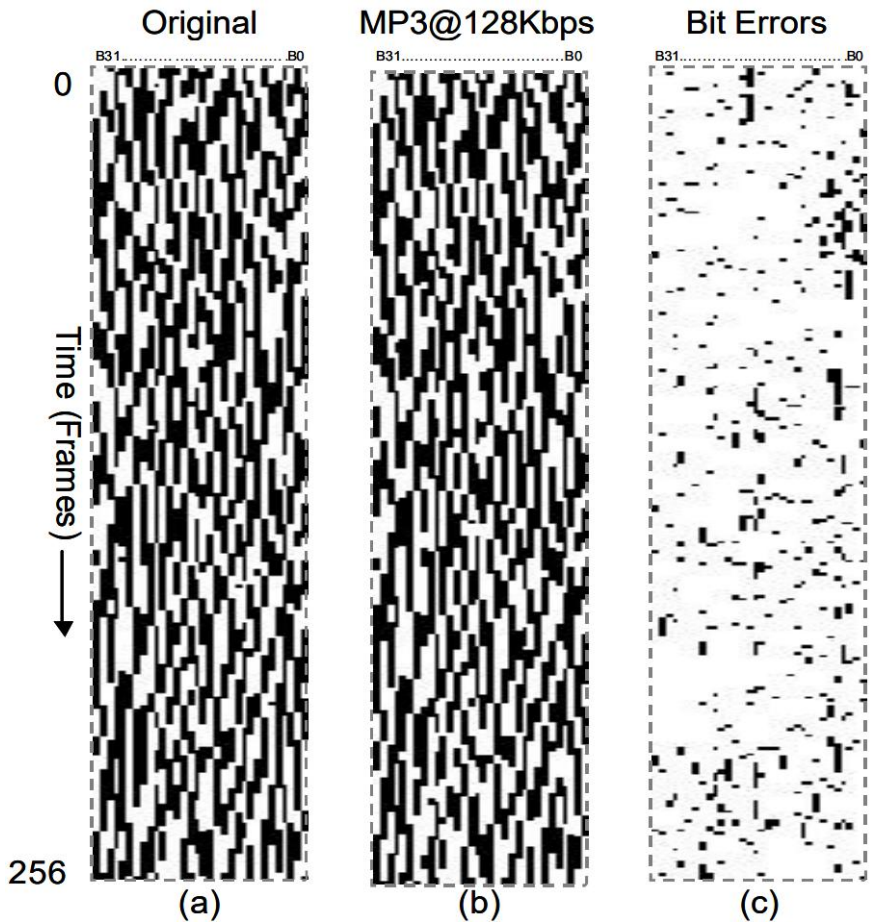


(Recap) Audio Fingerprinting: Examples (Haitsma et al., 2002)

Fingerprint extraction



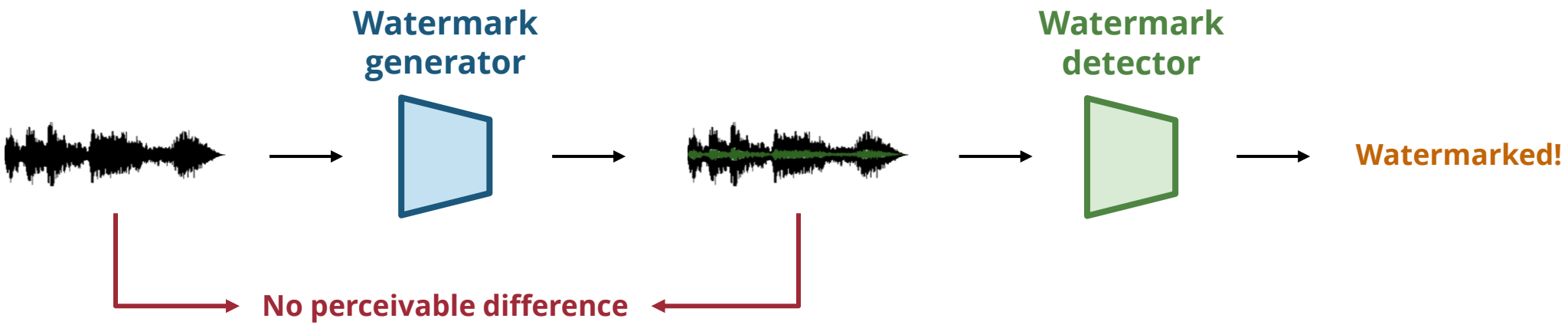
Example fingerprint blocks



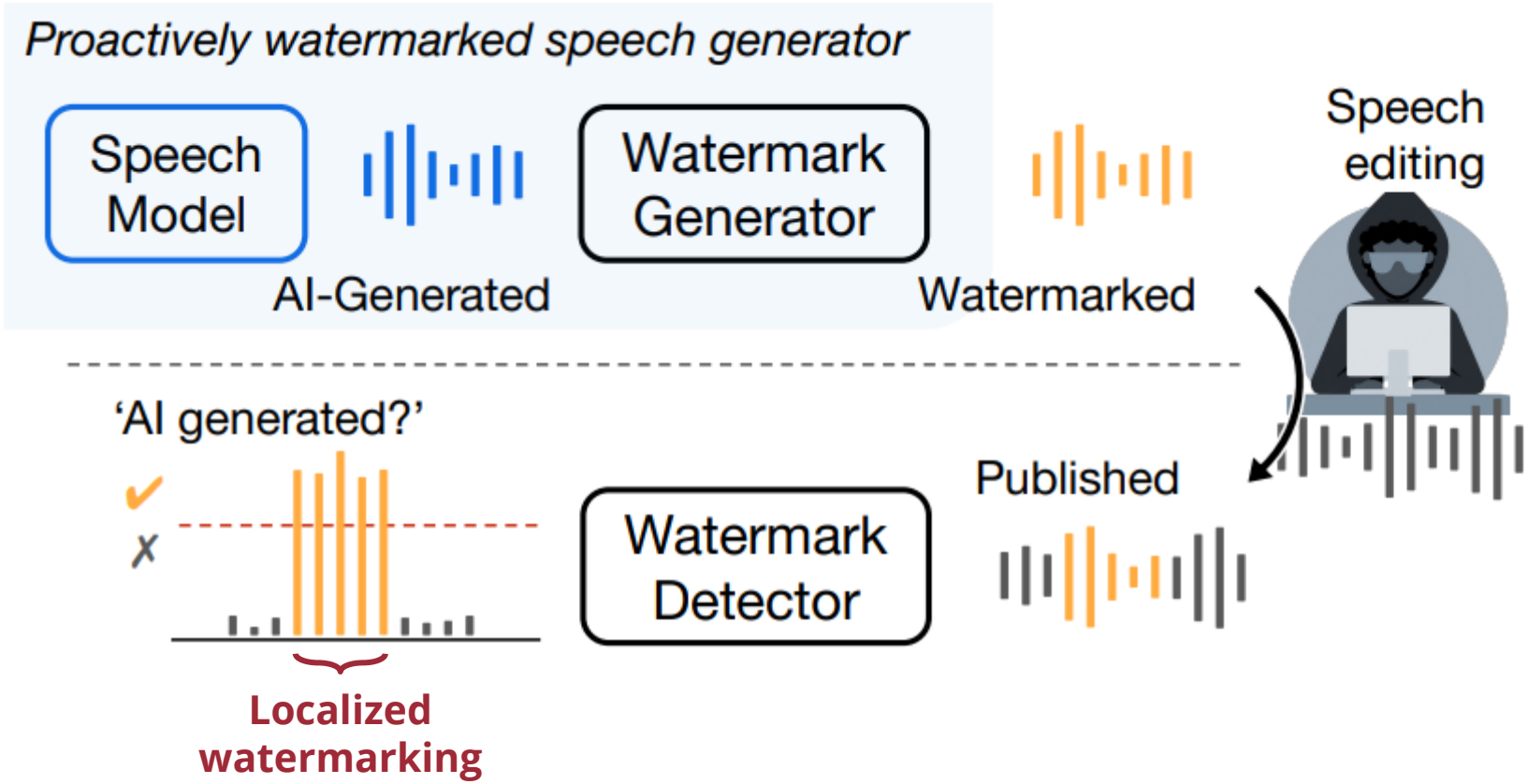
(Source: Haitsma et al., 2002)

Jaap Haitsma and Ton Kalker, "A Highly Robust Audio Fingerprinting System," *ISMIR*, 2002.

(Recap) Audio Watermarking



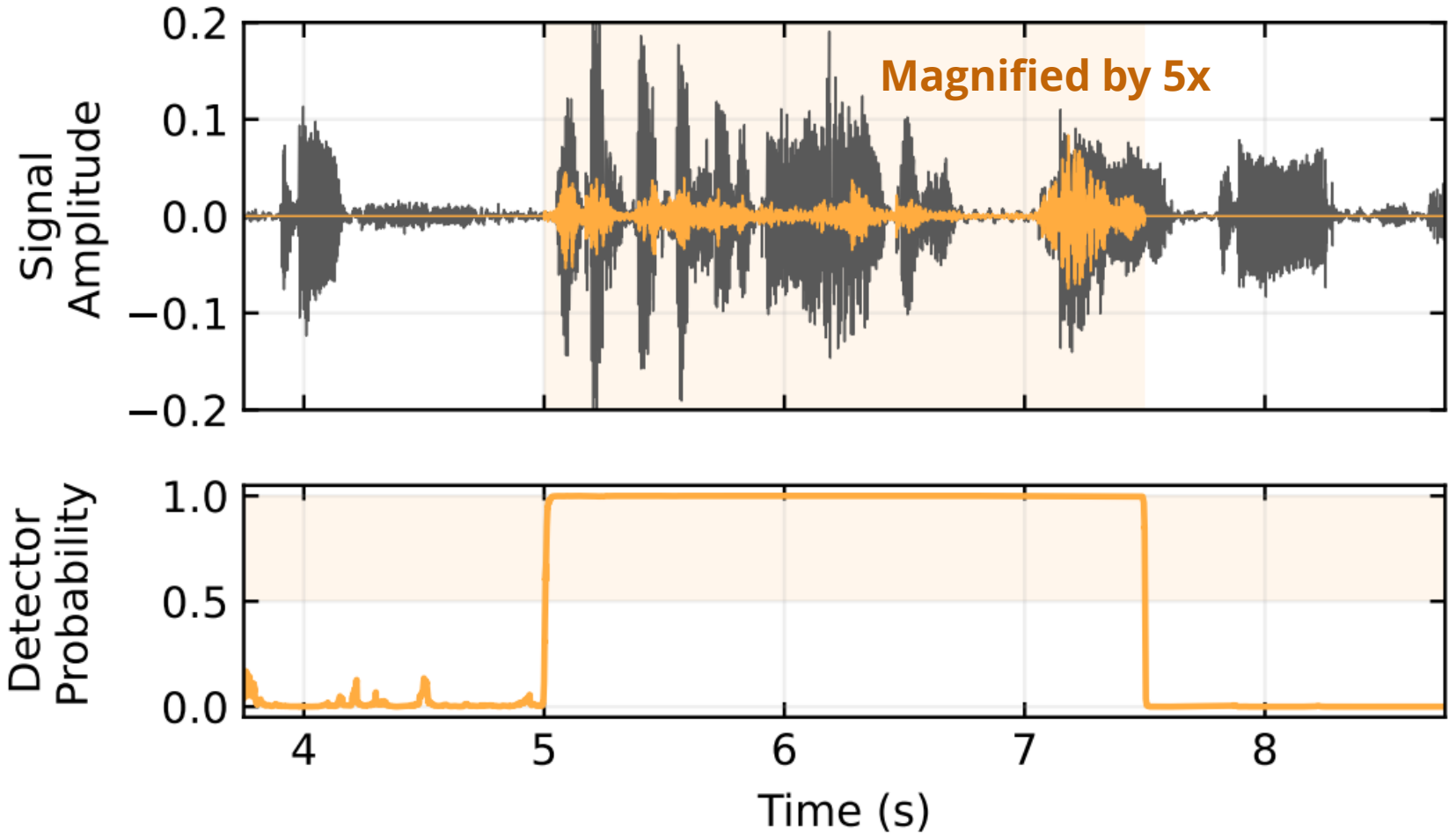
(Recap) Audio Watermarking Against Generated Audio



(Source: Roman et al., 2024)

Robin San Roman, Pierre Fernandez, Hady Elsahar, Alexandre Défossez, Teddy Furon, and Tuan Tran, "Proactive Detection of Voice Cloning with Localized Watermarking," *ICML*, 2024.

(Recap) AudioSeal (Roman et al., 2024)

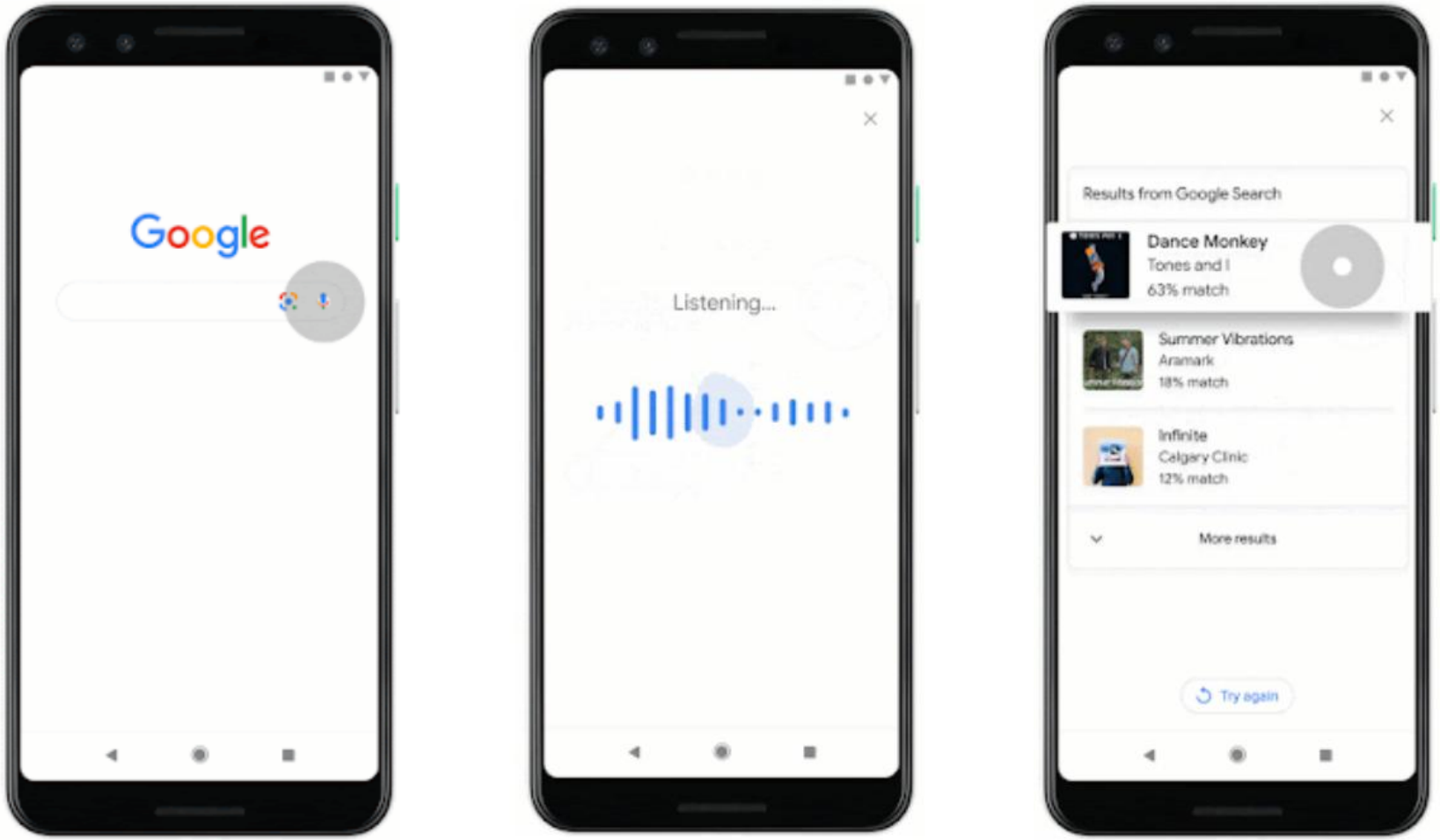


Original **Watermarked**

(Source: Roman et al., 2024)

Robin San Roman, Pierre Fernandez, Hady Elsahar, Alexandre Défossez, Teddy Furon, and Tuan Tran, "Proactive Detection of Voice Cloning with Localized Watermarking," *ICML*, 2024.

(Recap) Hum to Search (Google)

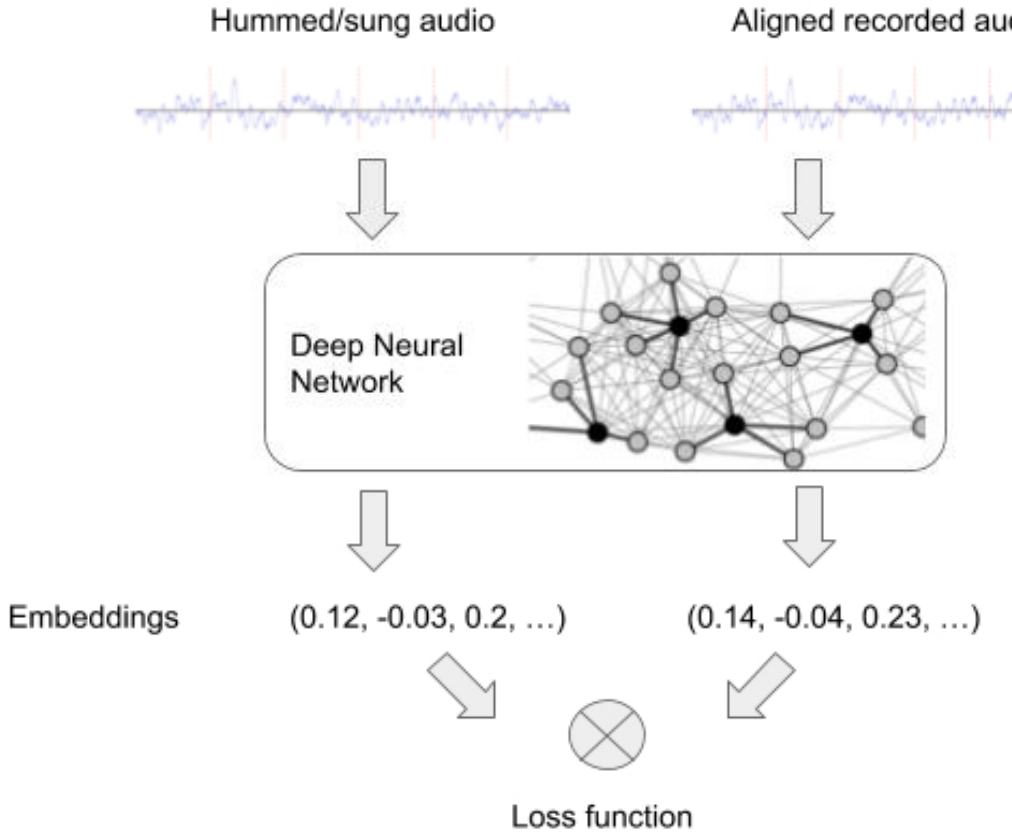


(Source: Google Research Blog)

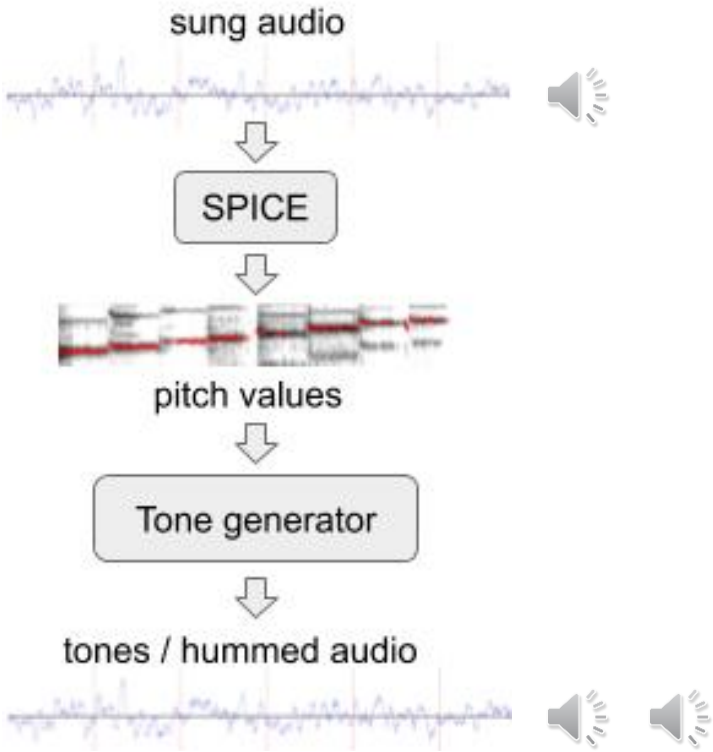
research.google/blog/the-machine-learning-behind-hum-to-search/

(Recap) Hum to Search (Google)

Audio encoder



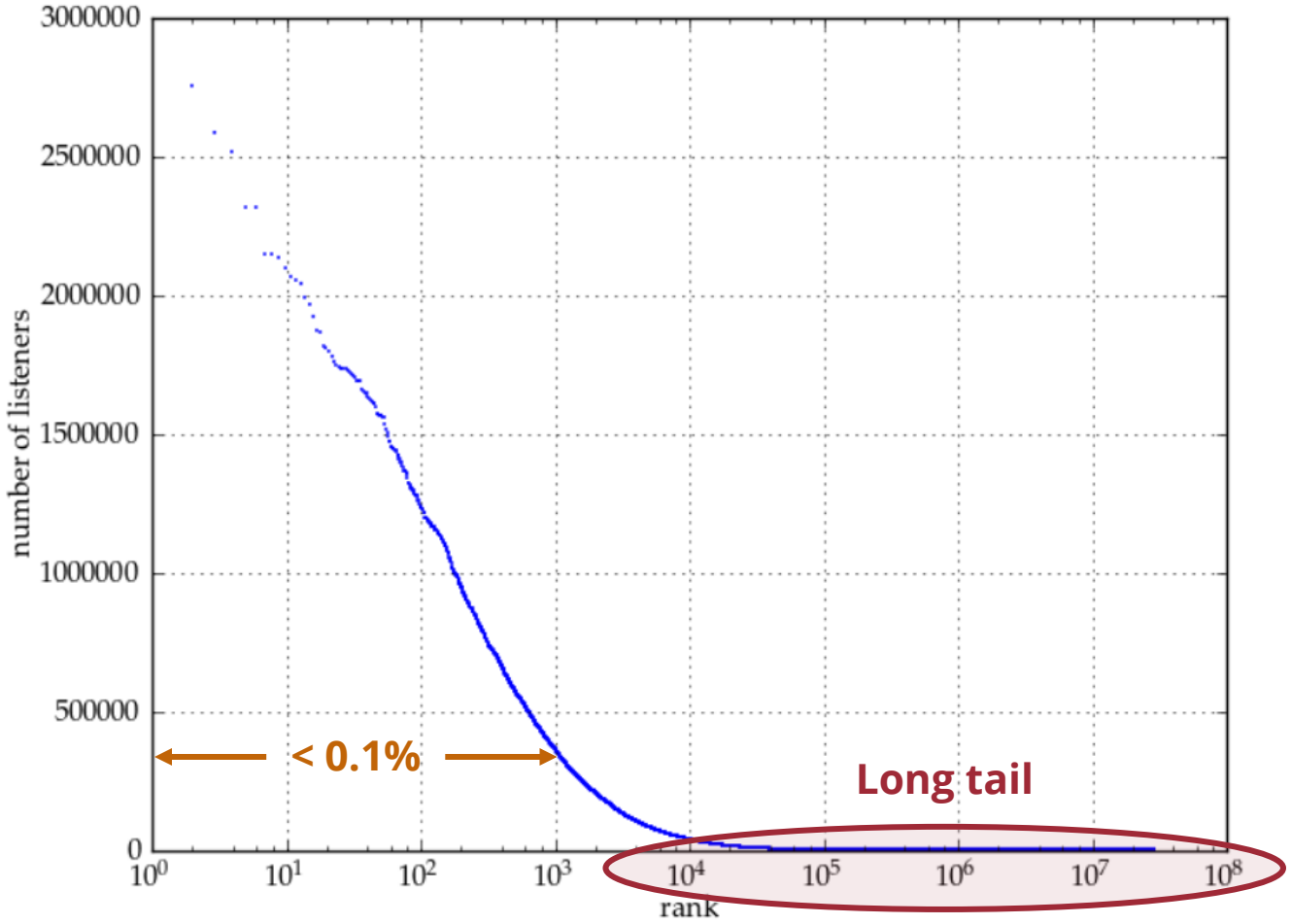
Data augmentation



(Source: Google Research Blog)

research.google/blog/the-machine-learning-behind-hum-to-search/

(Recap) The Long Tail Problem

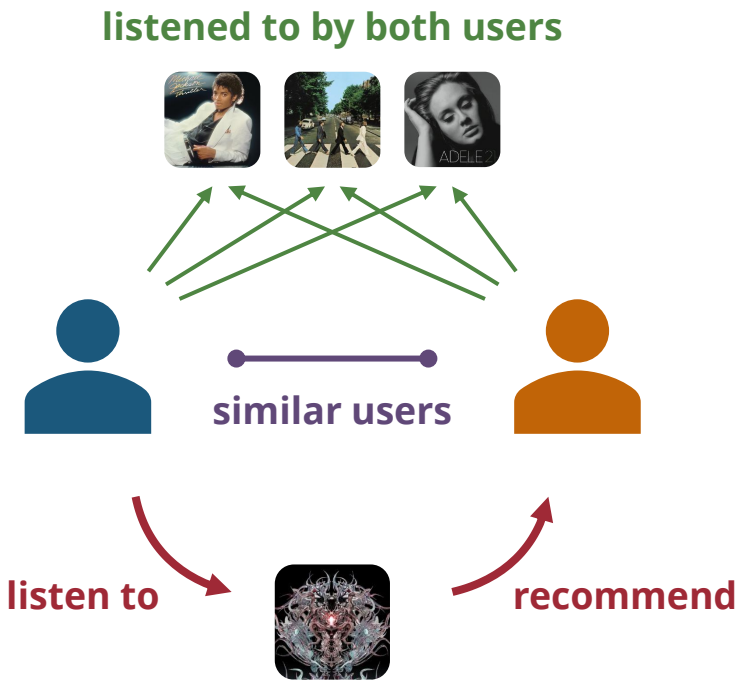


(Source: Levy & Bosteels, 2010)

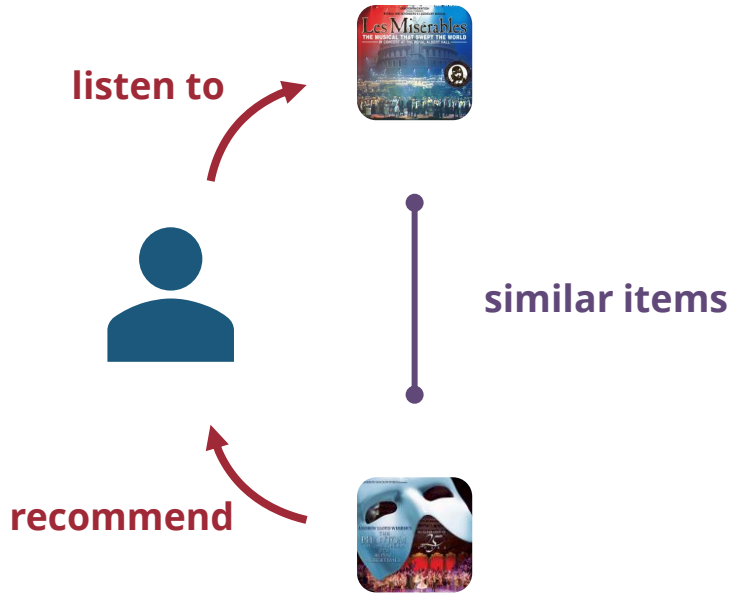
Mark Levy and Klaas Bosteels, "Music Recommendation and the Long Tail," *WOMRAD*, 2010.

(Recap) Collaborative Filtering vs Content-based Filtering

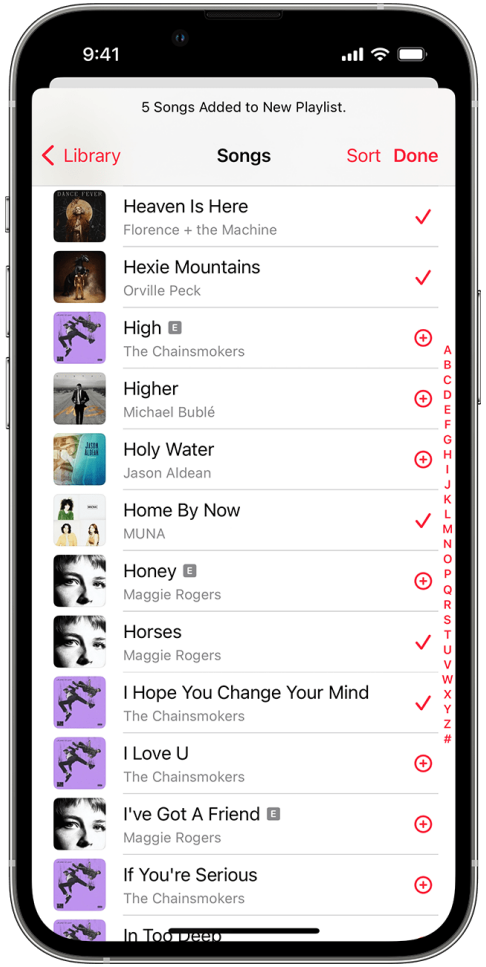
Collaborative filtering



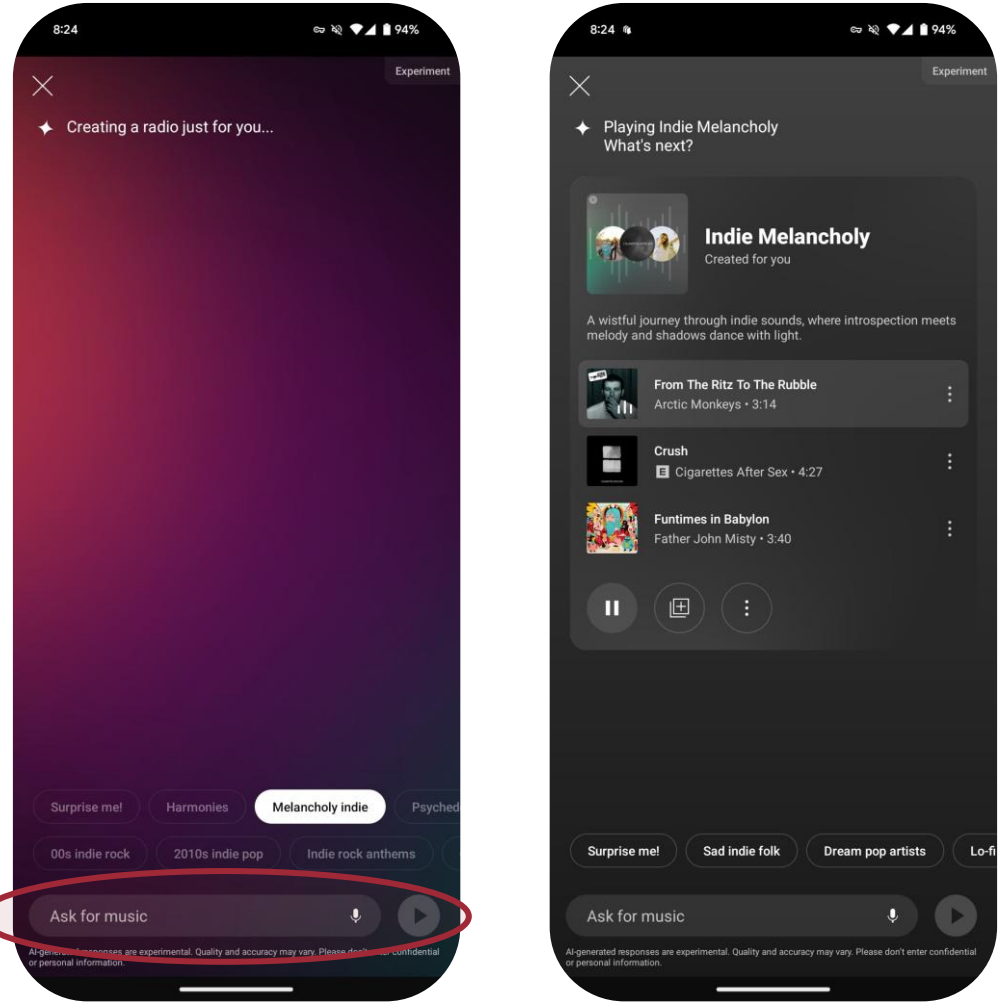
Content-based filtering



(Recap) Music Playlist Generation



(Recap) Ask Music (YouTube Music)

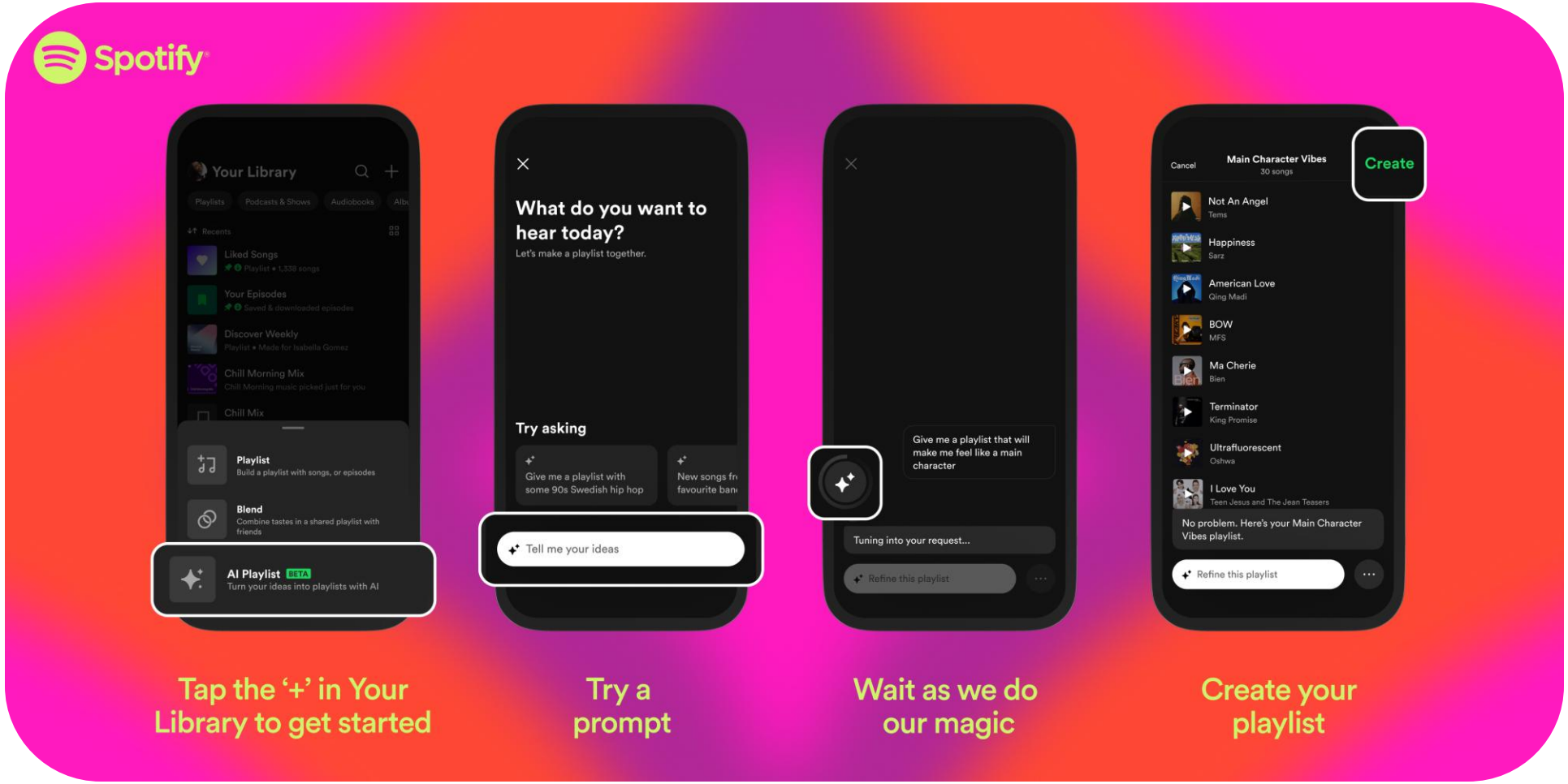


Text prompt

(Source: Android Police)

androidpolice.com/youtube-music-ask-music-ai-playlist-assistant/

(Recap) AI Playlist (Spotify)

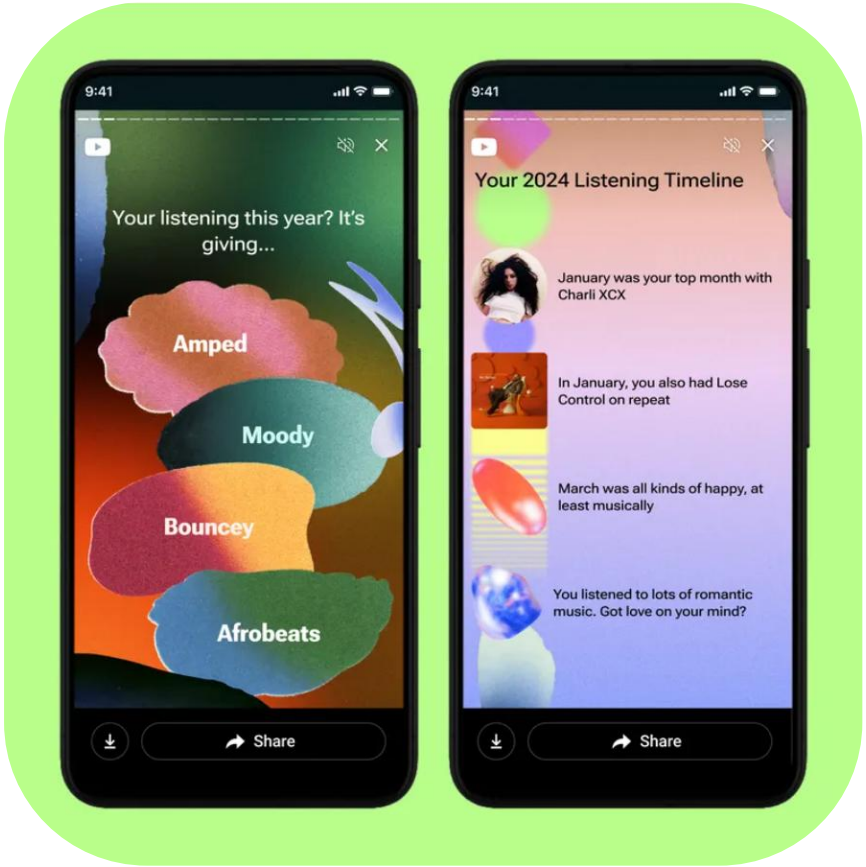


(Source: Spotify)

newsroom.spotify.com/2024-04-07/spotify-premium-users-can-now-turn-any-idea-into-a-personalized-playlist-with-ai-playlist-in-beta/

(Recap) Listening Behavior Analysis

YouTube's Music Recap



(Source: YouTube)

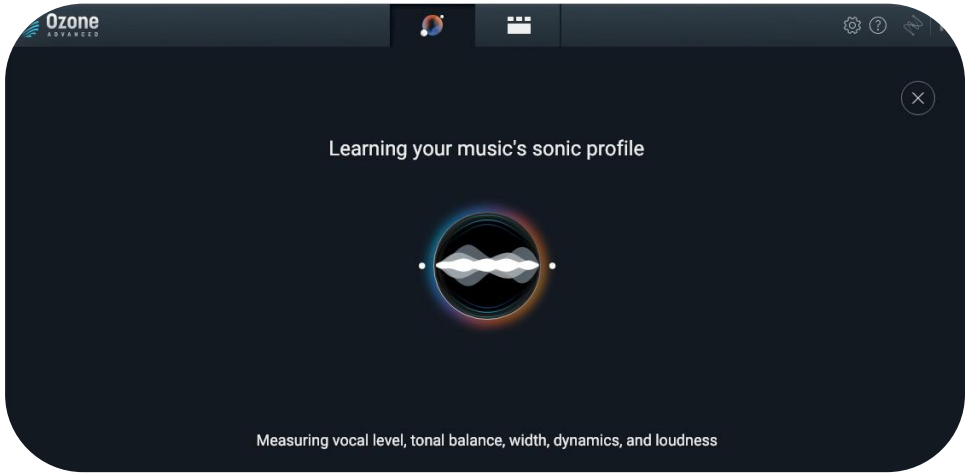
Spotify's Listening Personality

(Source: Spotify)

blog.youtube/news-and-events/2024-music-recap-youtube/
engineering.atspotify.com/2023/01/whats-a-listening-personality/

Auto-mixing

iZotope Ozone's Master Assistant

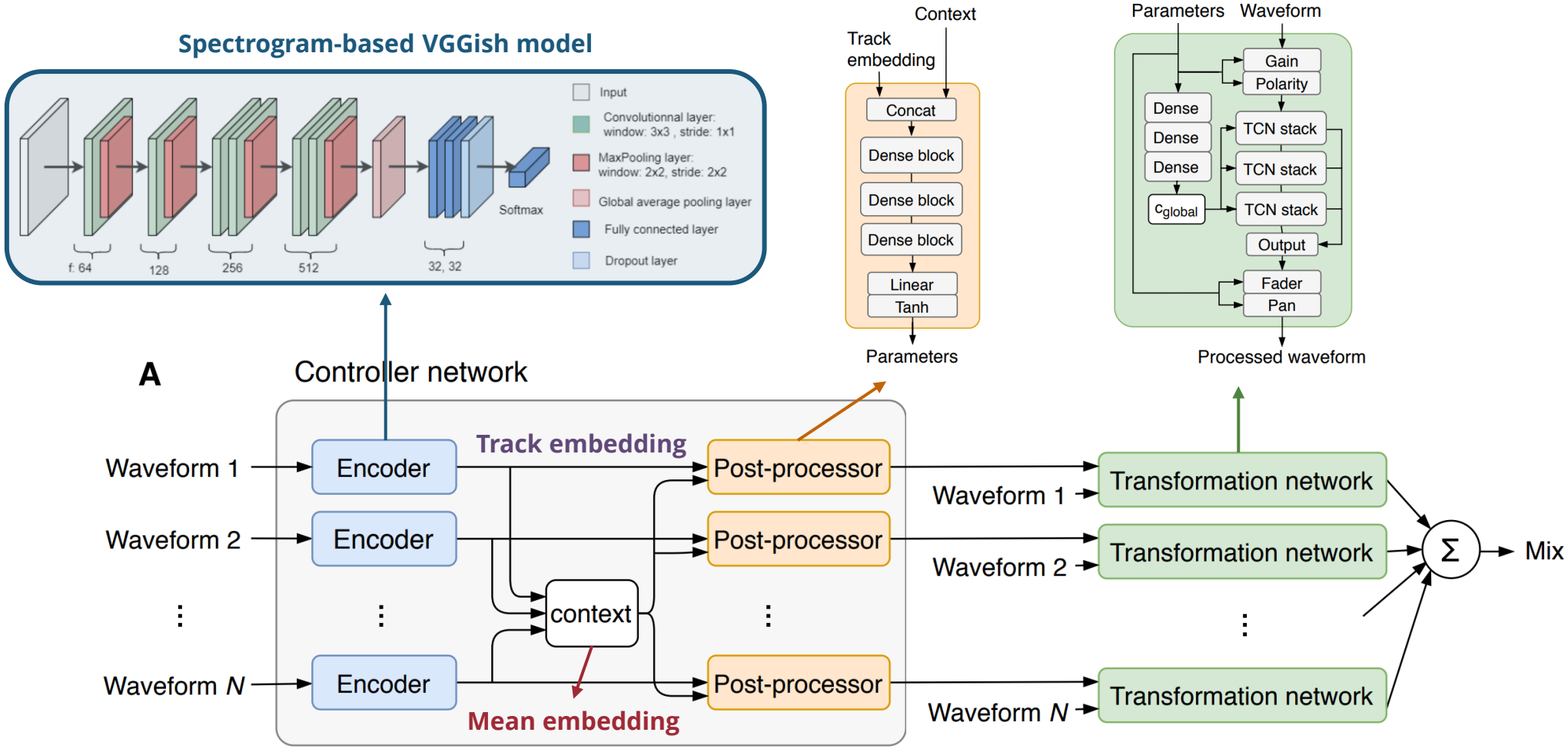


(Source: iZotope)

(Source: iZotope)

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

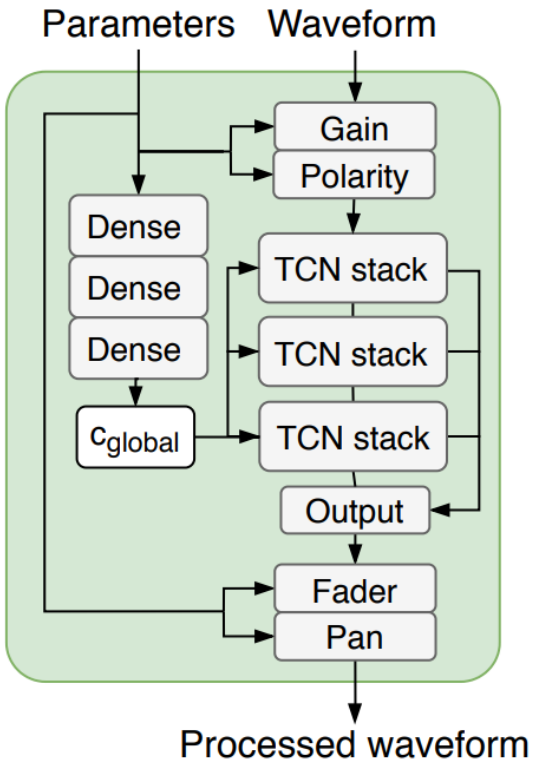
Differentiable Auto-mixing (Steinmetz et al., 2021)



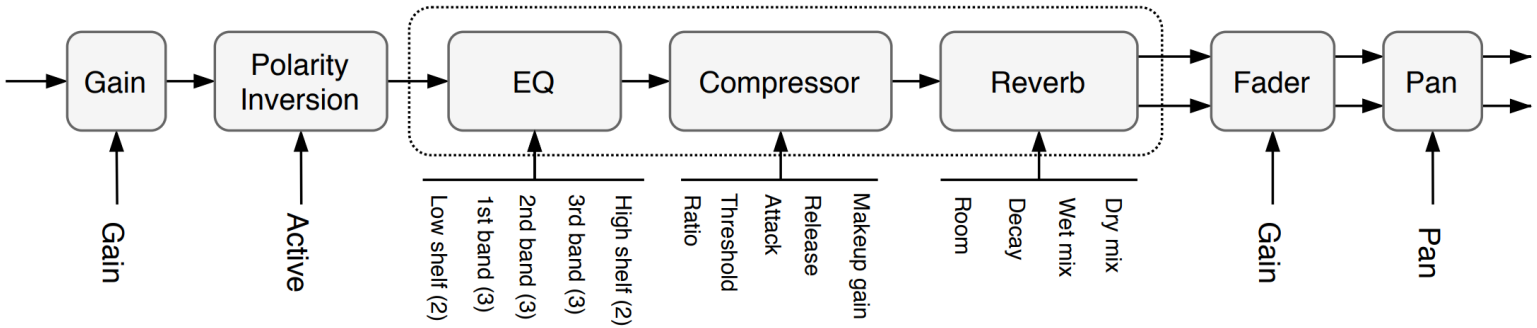
(Source: Steinmetz et al., 2021)

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 Auto-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 Auto-mixing (Steinmetz et al., 2021)

Transformation Network

Input	Target	Output
		
		

csteinmetz1.github.io/dmc-icassp2021

Differentiable Auto-mixing (Steinmetz et al., 2021)

Drum mixing

(Same mixing style)

DMC	Mono	Random	Target
			
			

Multitrack mixing

(Diverse mixing style)

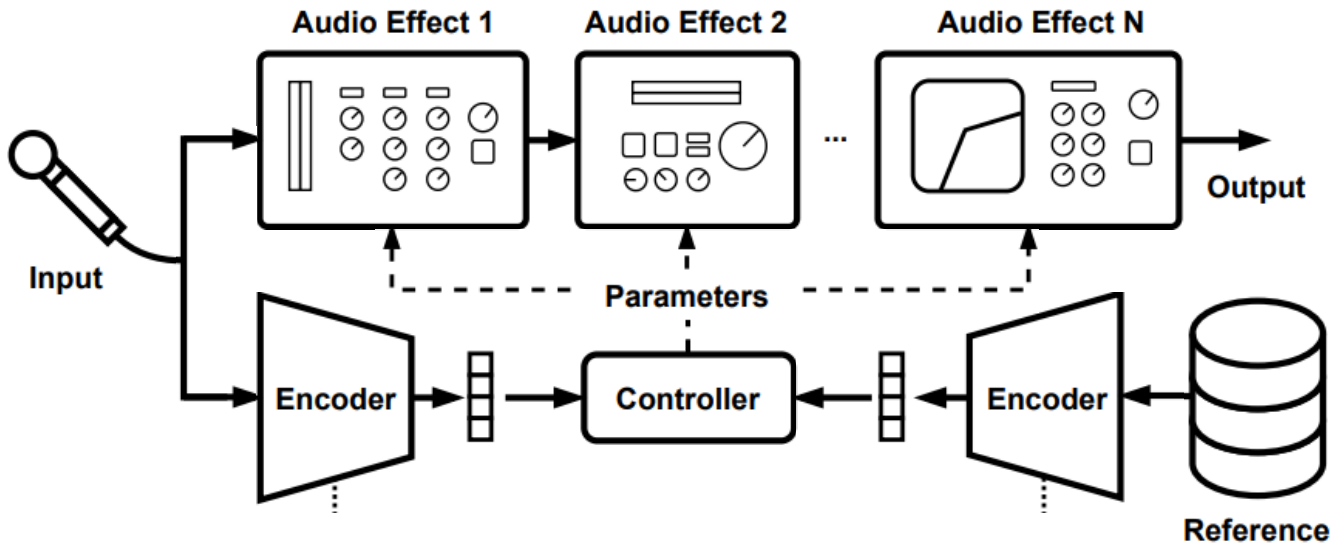
DMC	Mono	Random	Target
			
			

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.

Effects & Mixing Style Transfer

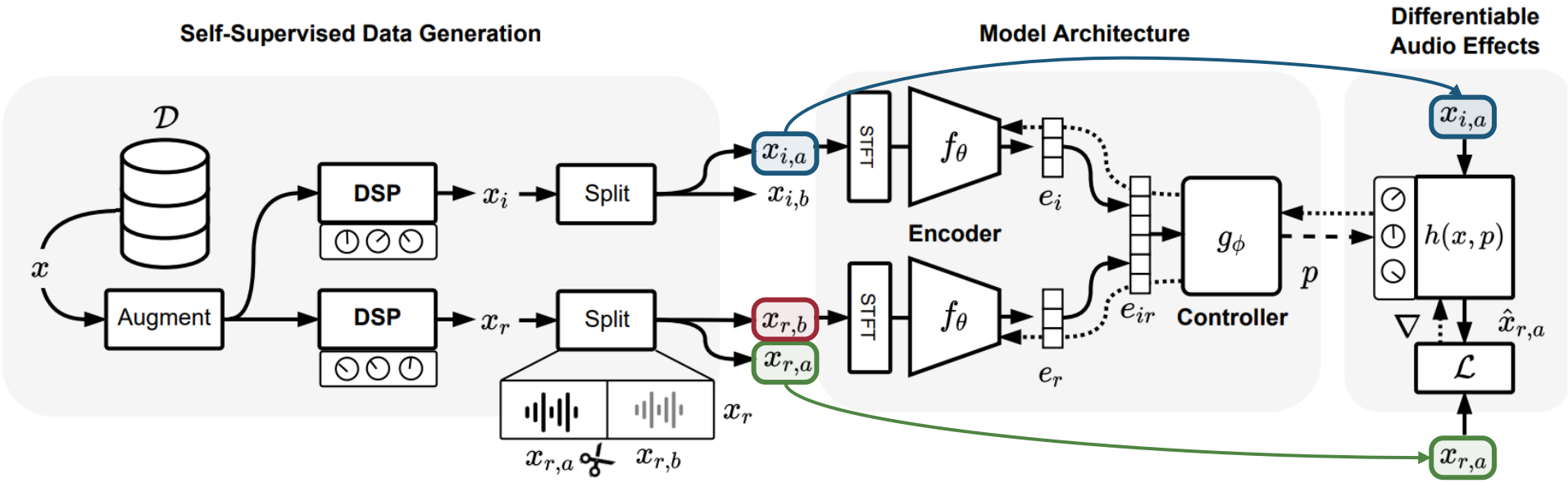
DeepAFx-ST: Effects Transfer (Steinmetz et al., 2022)



(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)

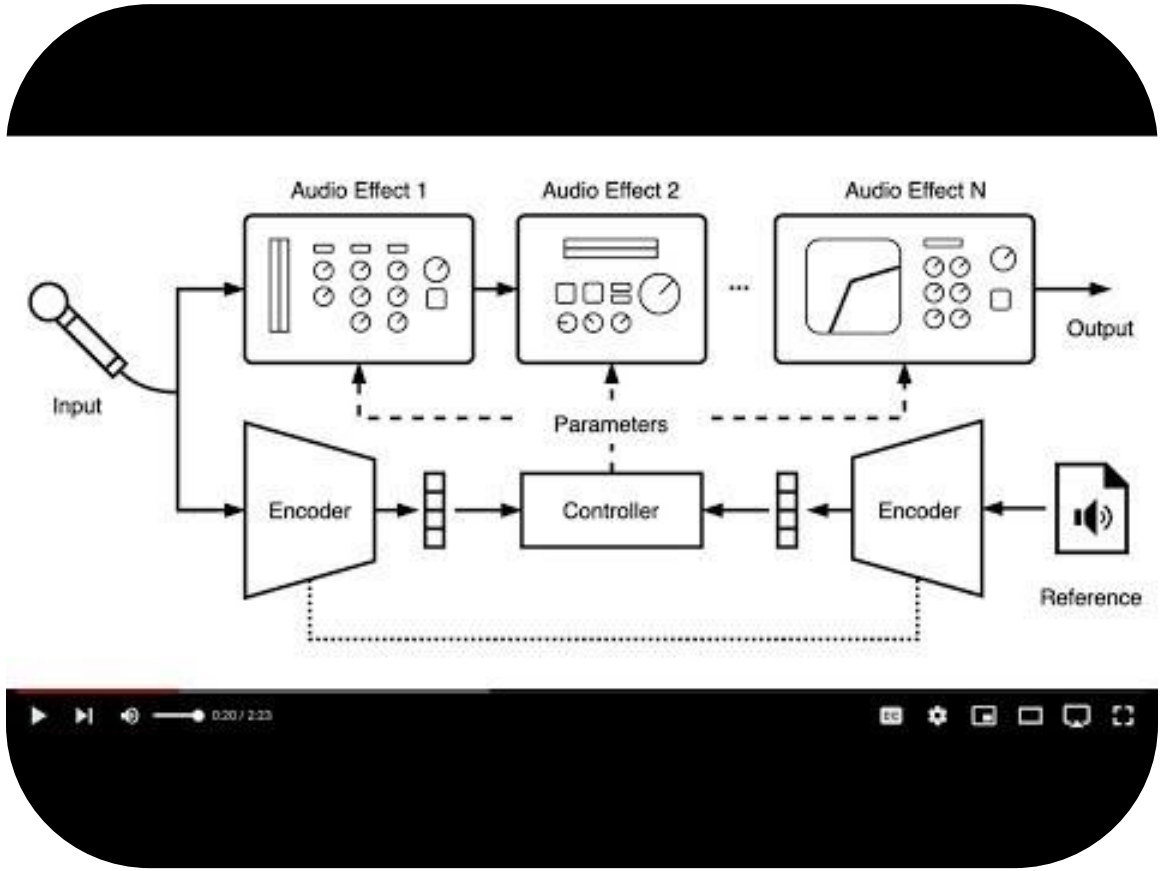


(Source: Steinmetz et al., 2022)

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)

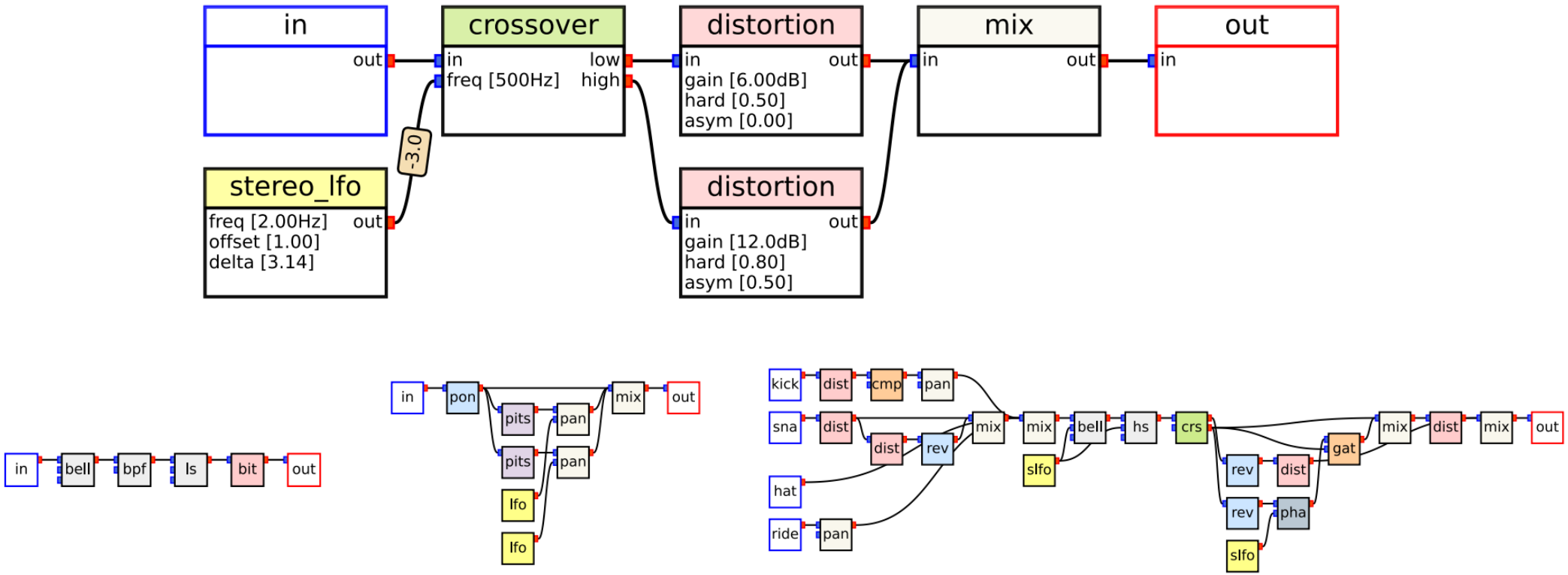


youtu.be/lZp455wiMk4?t=100

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

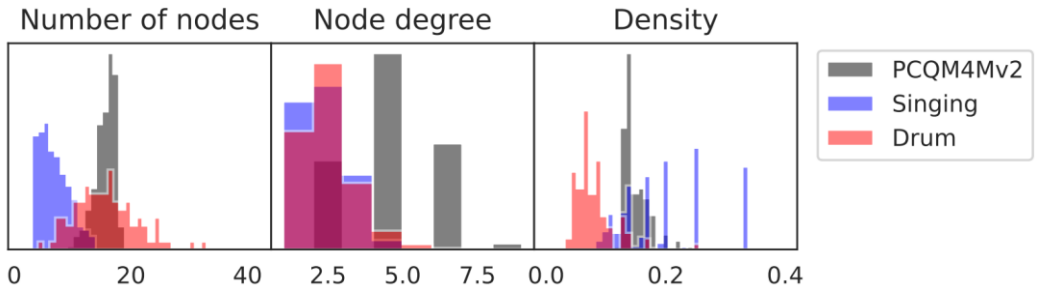
Sungho Lee, Jaehyun Park, Seungryeol Paik, and Kyogu Lee, "Blind Estimation of Audio Processing Graph," *ICASSP*, 2023.

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

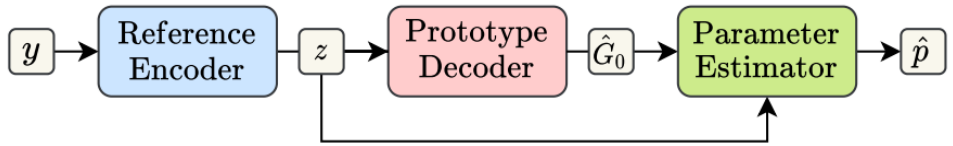


(Source: Lee et al., 2023)

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

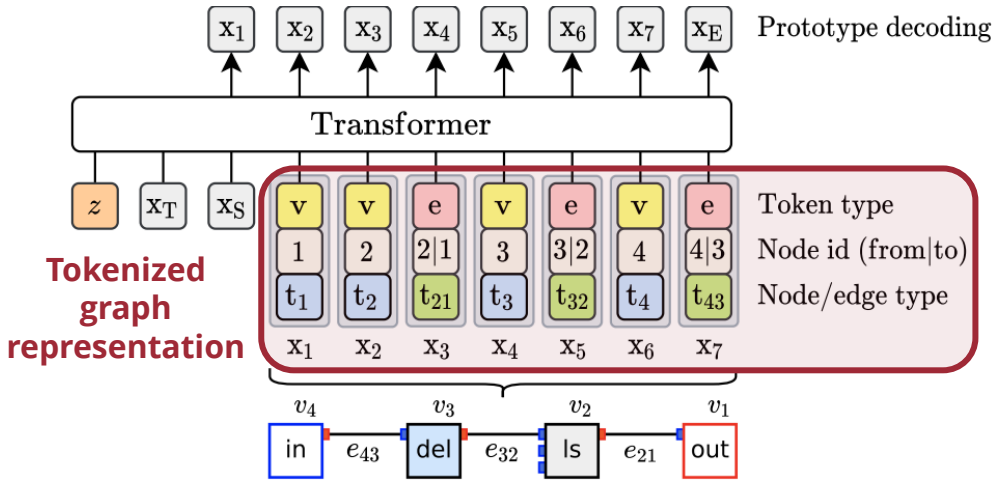
Estimating Audio Processing Graph (Lee et al., 2022)

Blind estimation framework

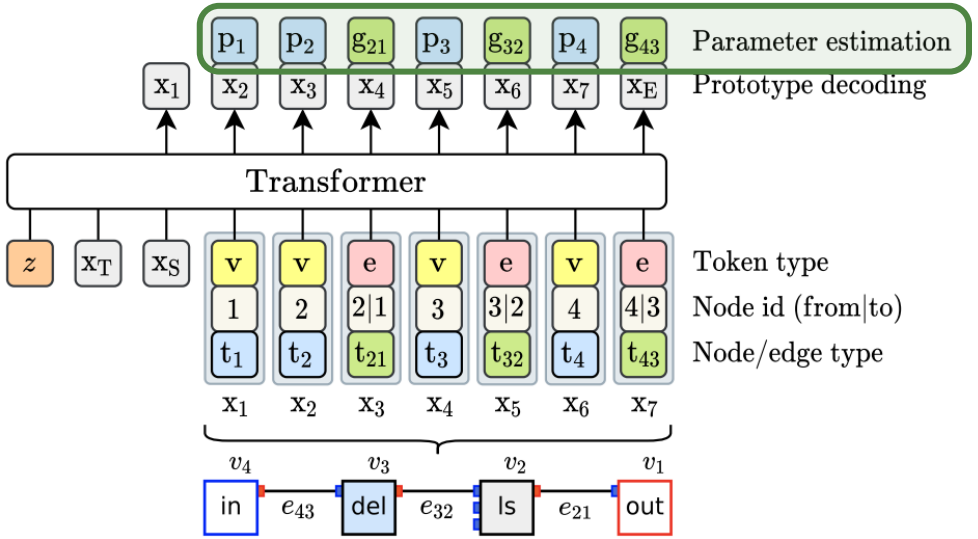


Sources are not given!

Prototype decoder



Parameter estimator

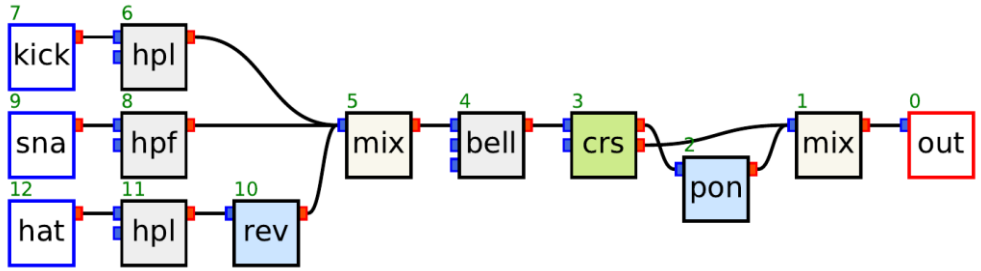


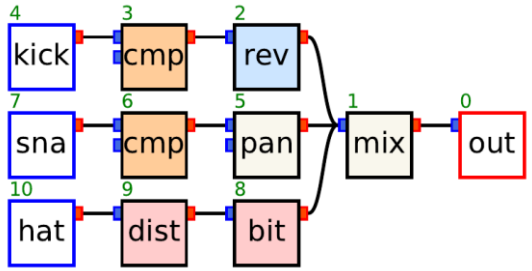
(Source: Lee et al., 2023)

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

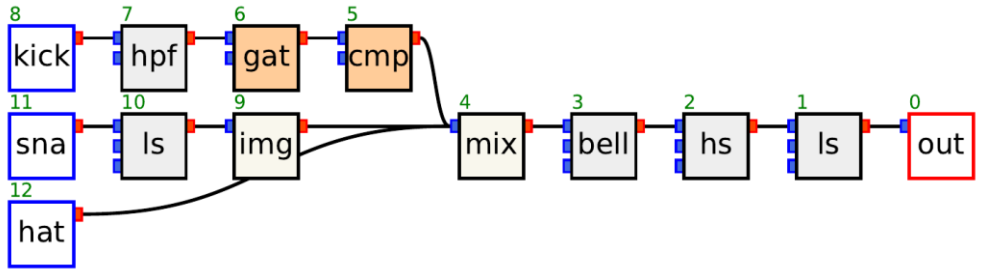
Dry

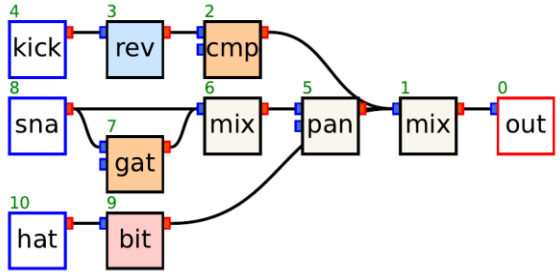
Reference





Estimation



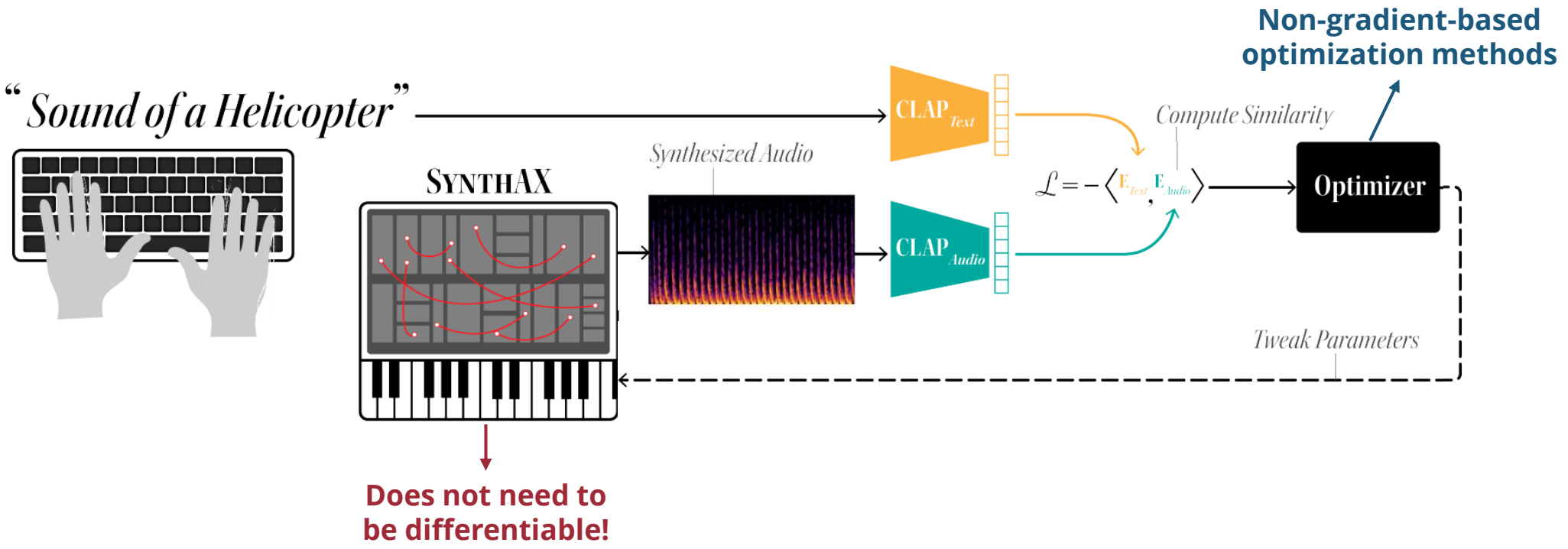


(Source: Lee et al., 2023)

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)



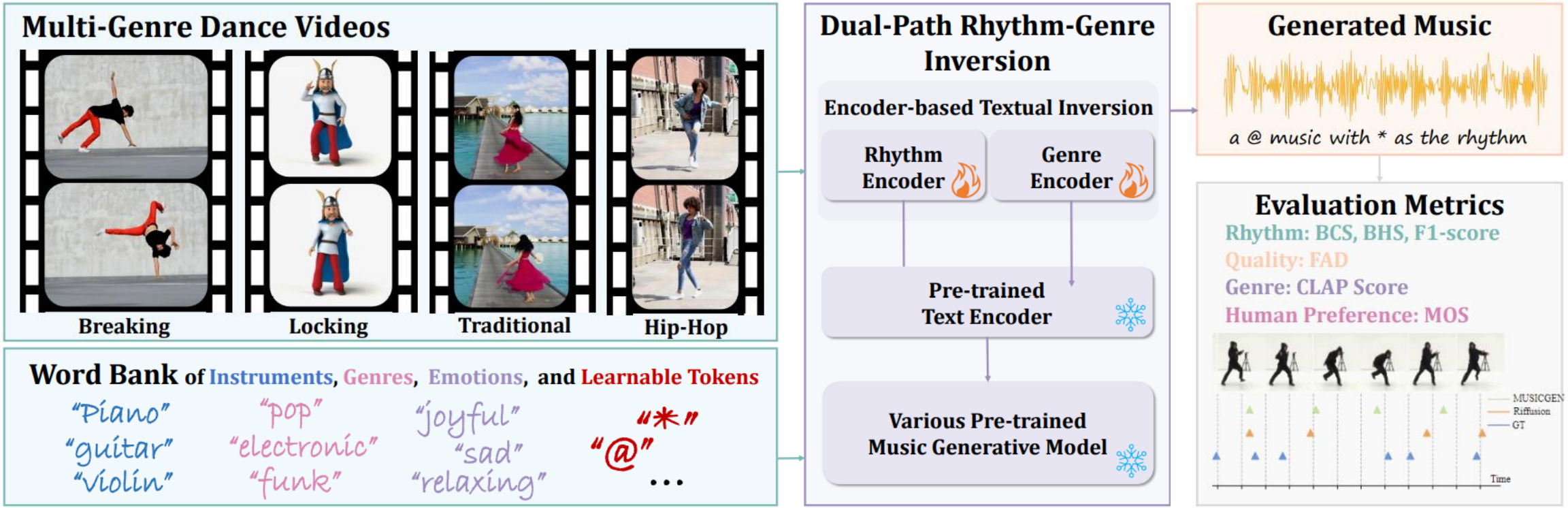
(Source: Cherep et al., 2024)

ctag.media.mit.edu

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

Multimodal Systems

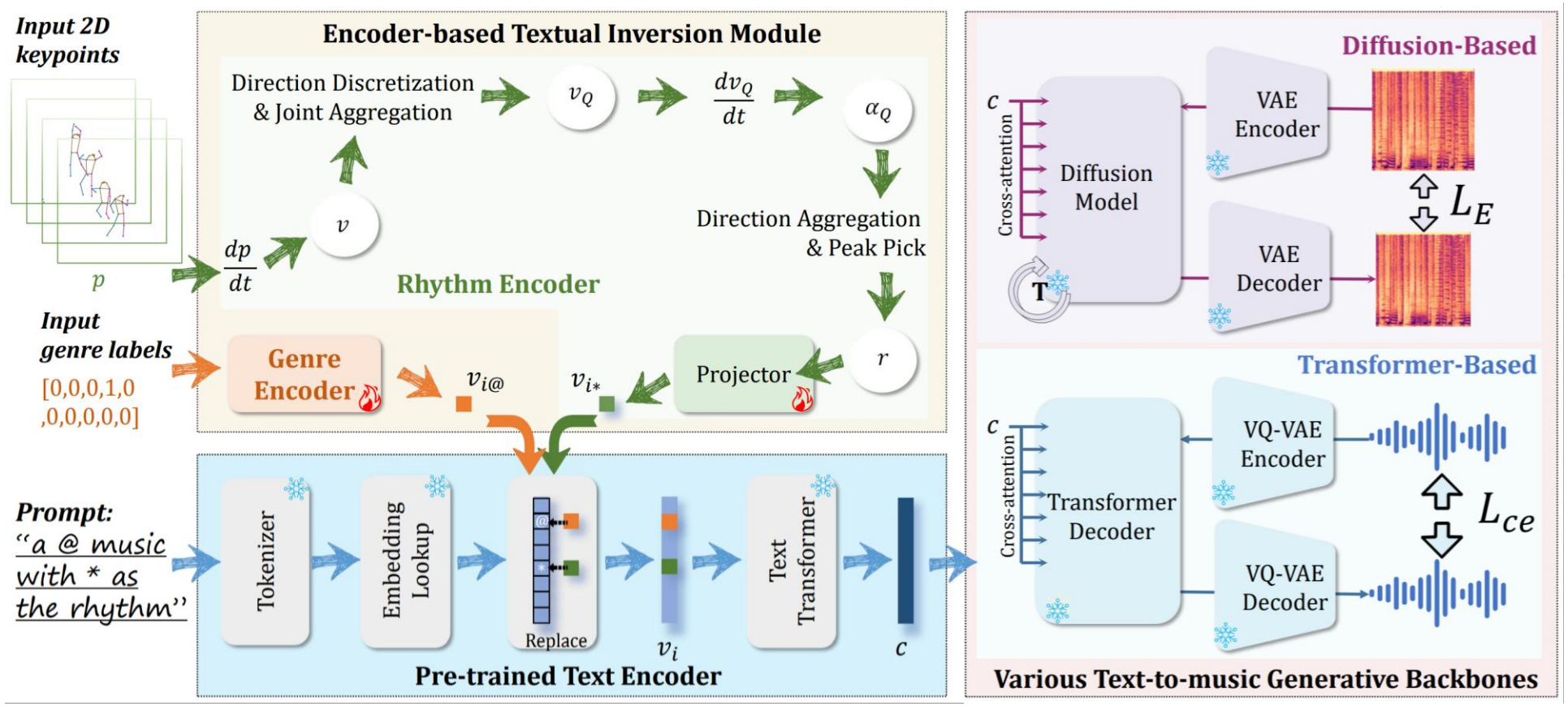
Dance-to-music Generation (Li et al., 2024)



(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)

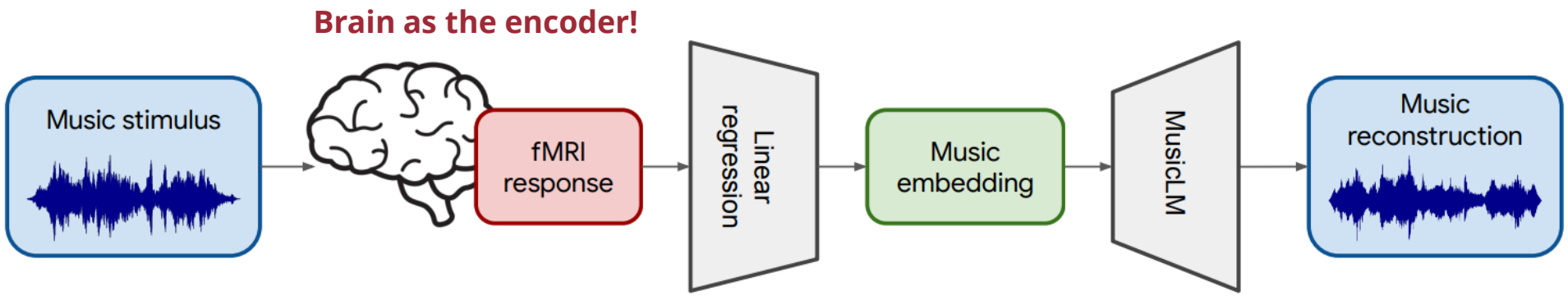
Sifei Li, Weiming Dong, Yuxin Zhang, Fan Tang, Chongyang Ma, Oliver Deussen, Tong-Yee Lee, and Changsheng Xu, "Dance-to-Music Generation with Encoder-based Textual Inversion," *SIGGRAPH ASIA*, 2024.

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)

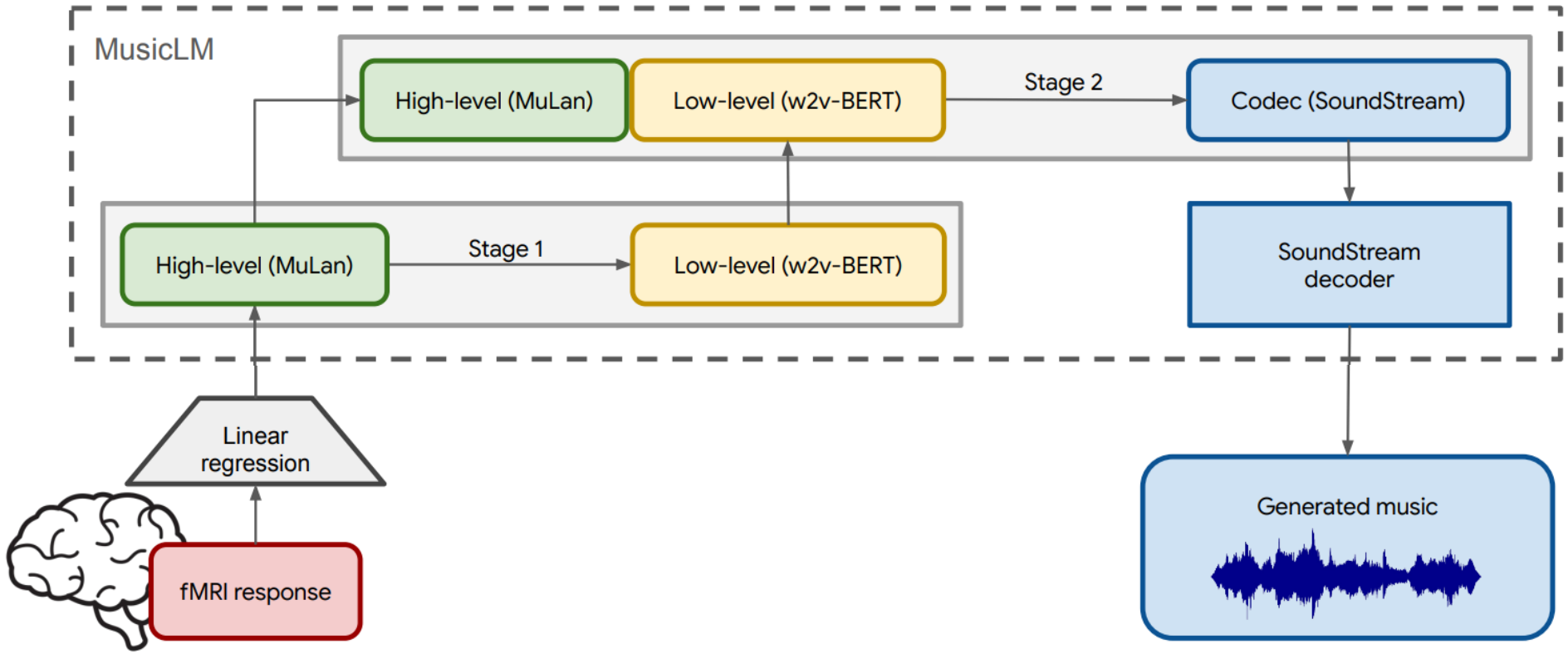


(Source: Denk et al., 2023)

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)



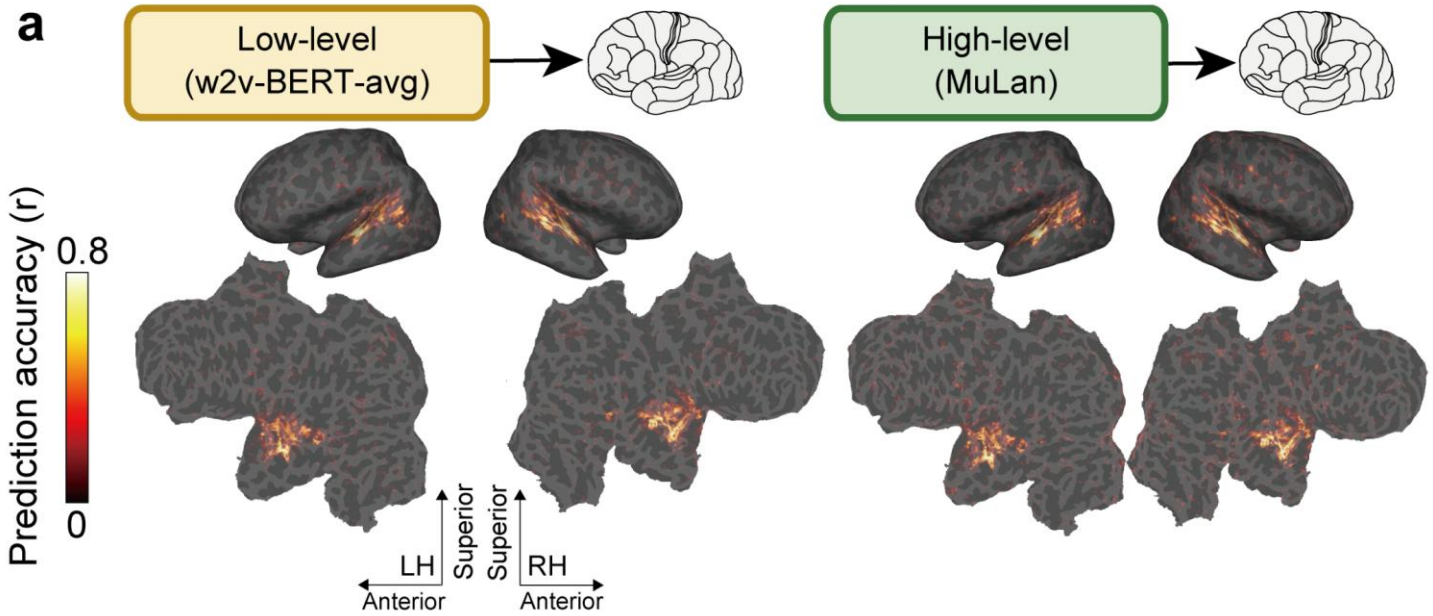
(Source: Denk et al., 2023)

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.

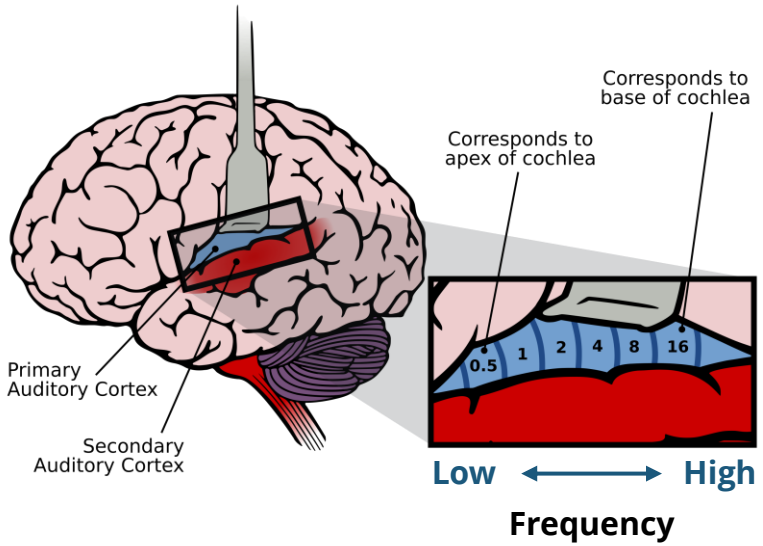
Brain2Music (Denk et al., 2023)

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.