PAT 498/598 (Winter 2025)

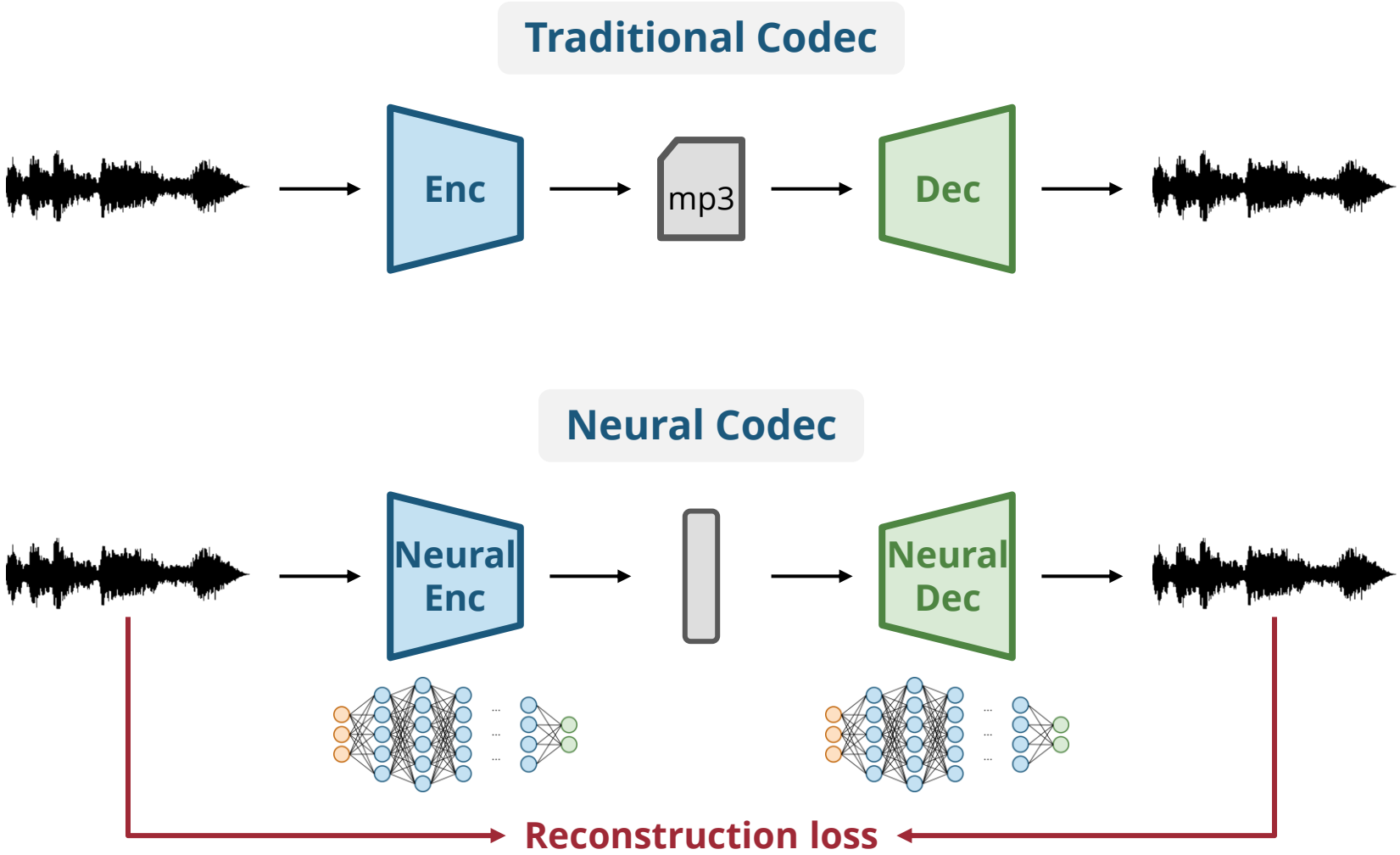
Music & AI

Lecture 18: Music Search & Recommendation

Instructor: Hao-Wen Dong

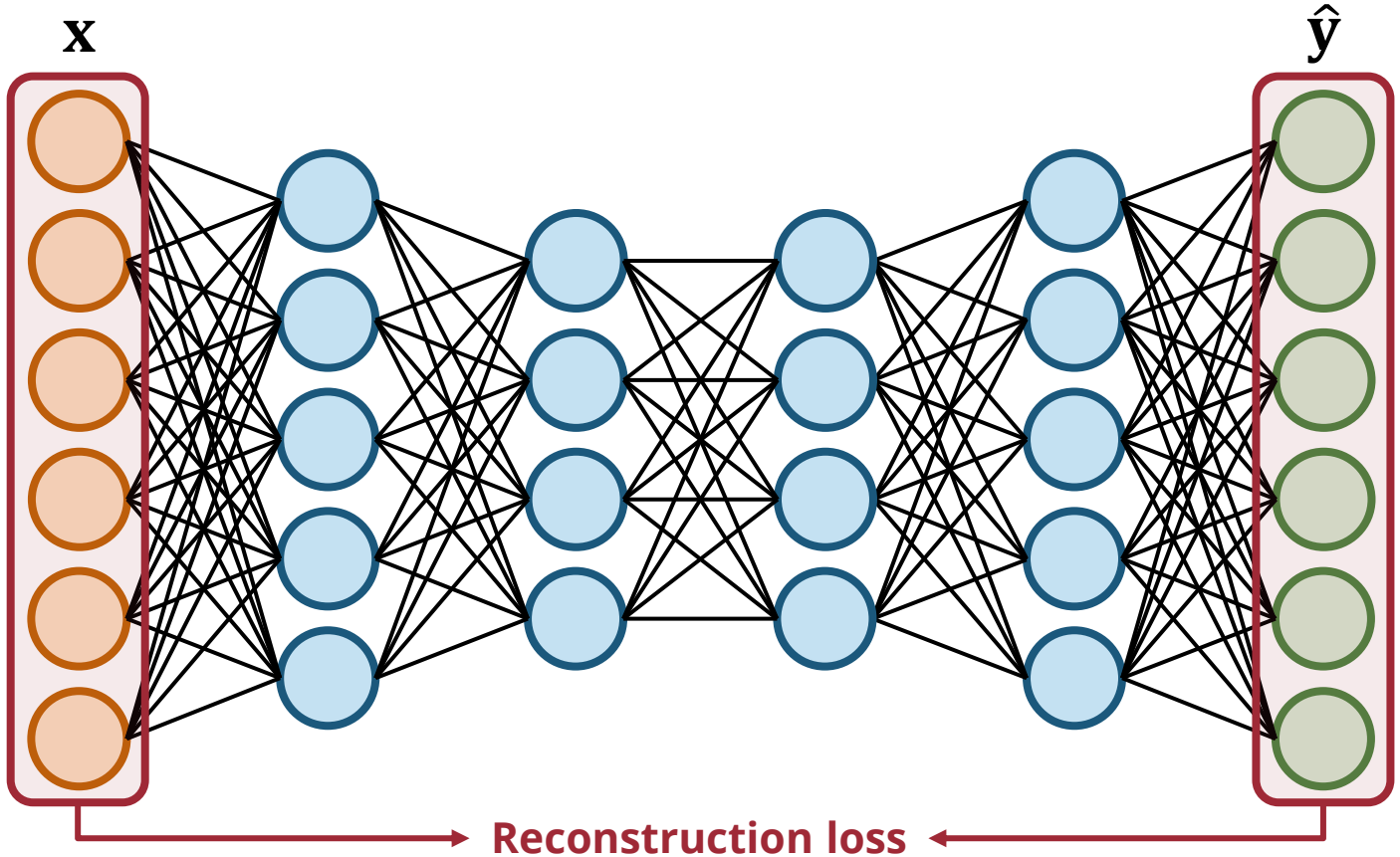
SCHOOL OF MUSIC, THEATRE & DANCE
PERFORMING ARTS TECHNOLOGY
UNIVERSITY OF MICHIGAN

(Recap) Neural Codec

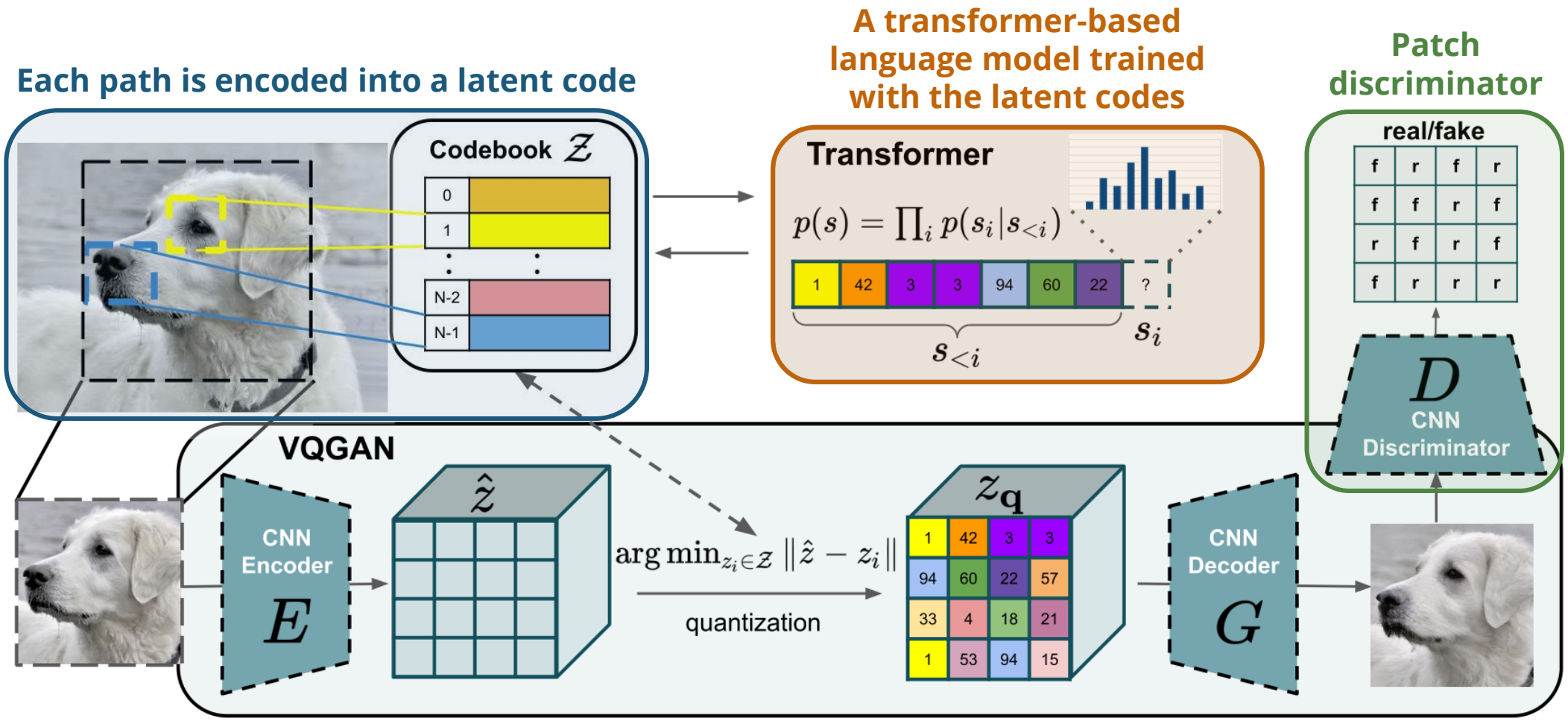


(Recap) Autoencoders

- A neural network where the **input and output are the same**



(Recap) VQGAN

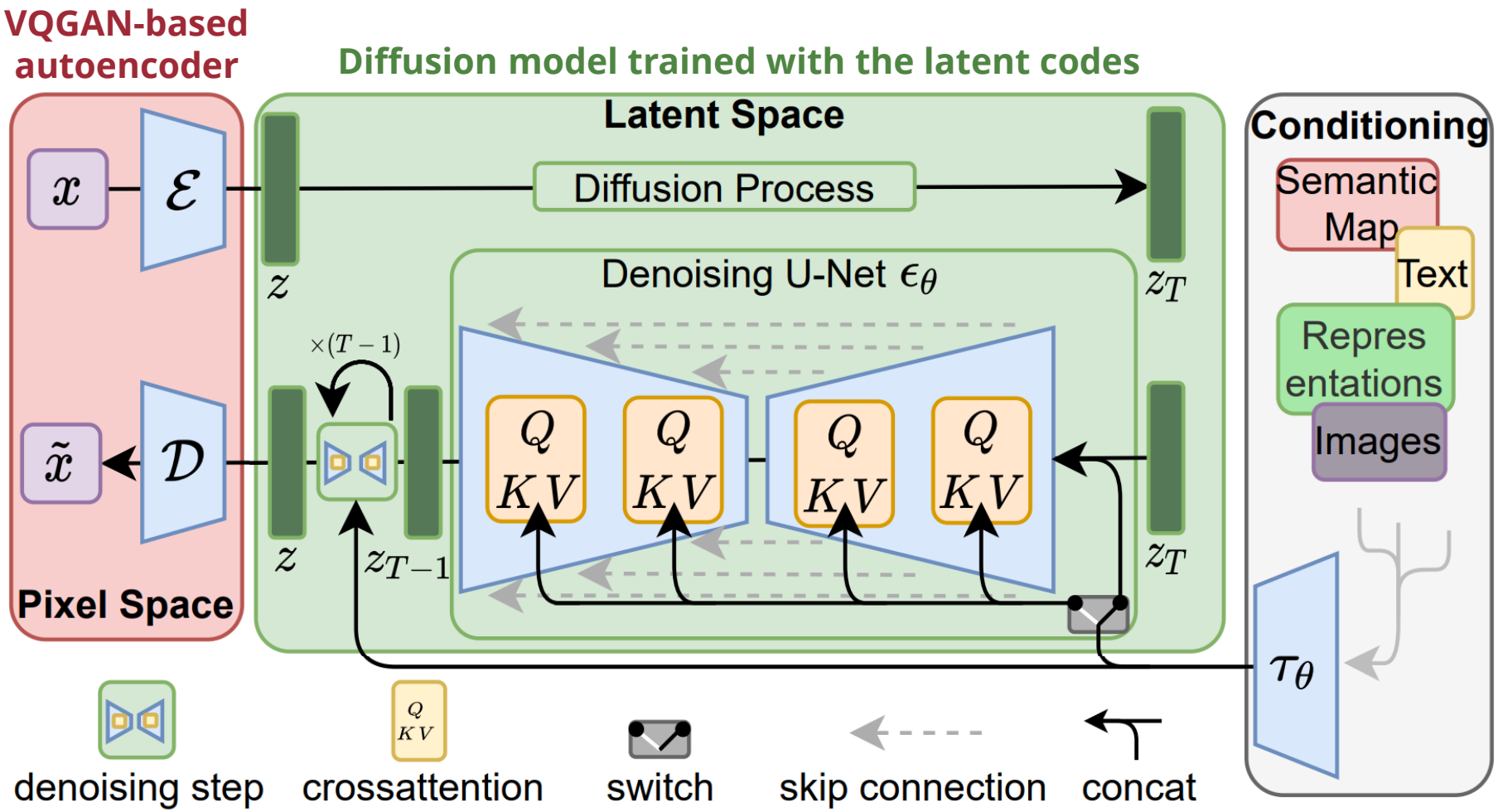


(Source: Esser et al., 2021)

A VQGAN is a VQVAE equipped with adversarial loss

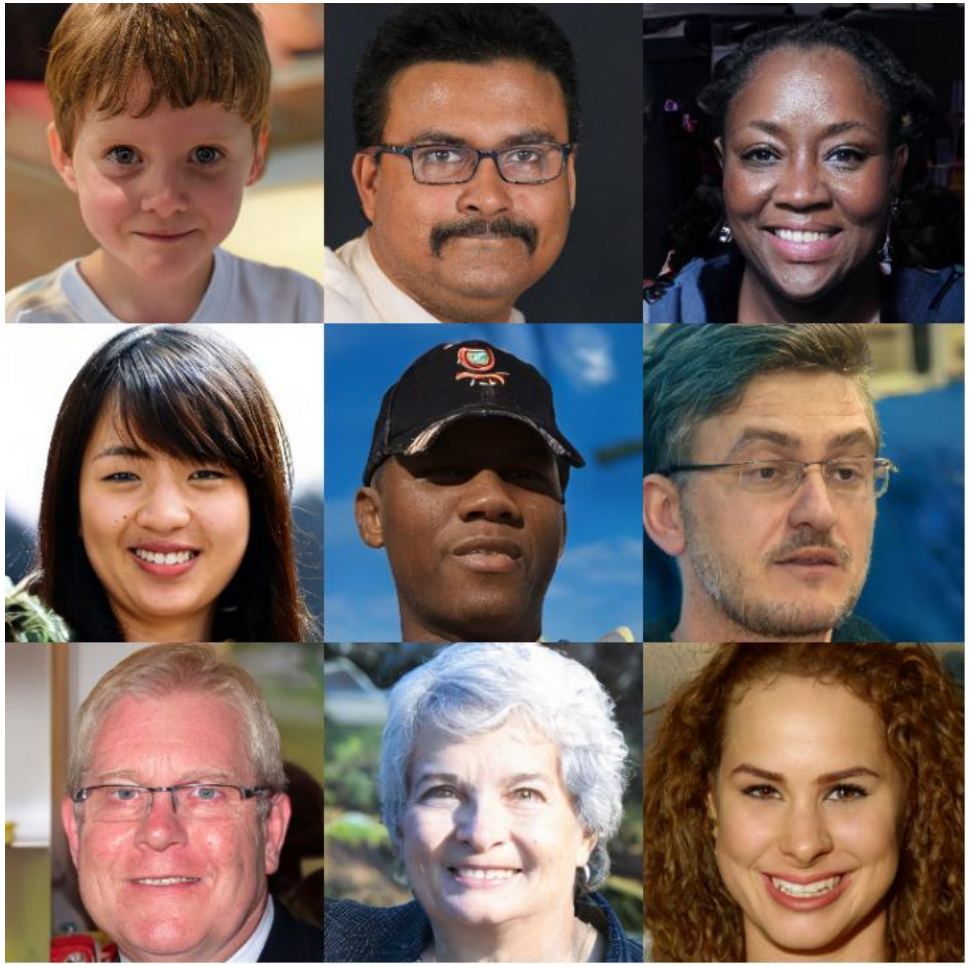
Patrick Esser, Robin Rombach, and Björn Ommer, "Taming Transformers for High-Resolution Image Synthesis," *CVPR*, 2021.

(Recap) Latent Diffusion Models (LDMs)



(Source: Rombach et al., 2022)

(Recap) LDMs: Examples

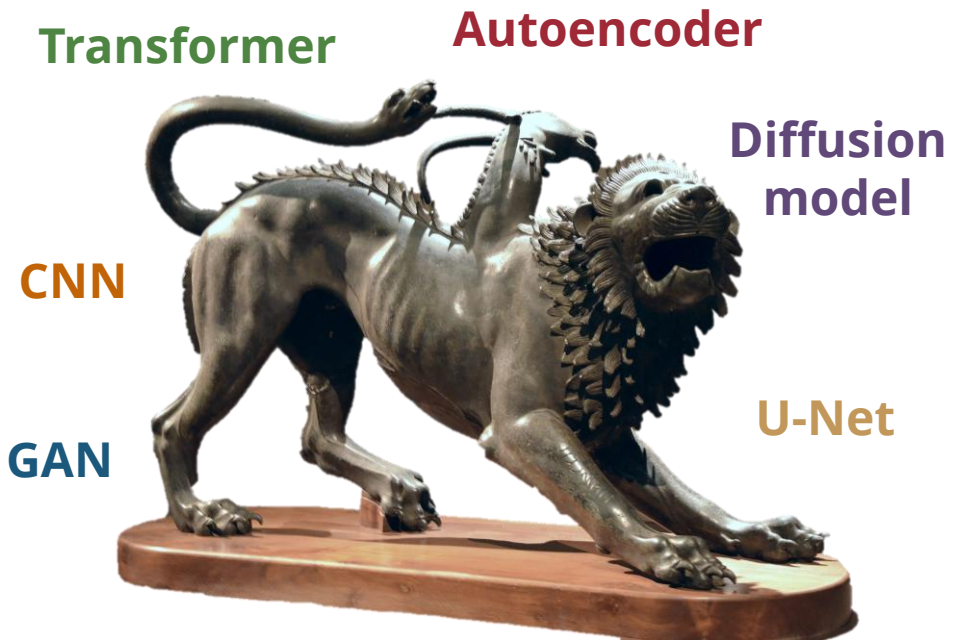


(Source: Rombach et al., 2022)

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," *CVPR*, 2022.

(Recap) Latent Diffusion Model is a Chimera

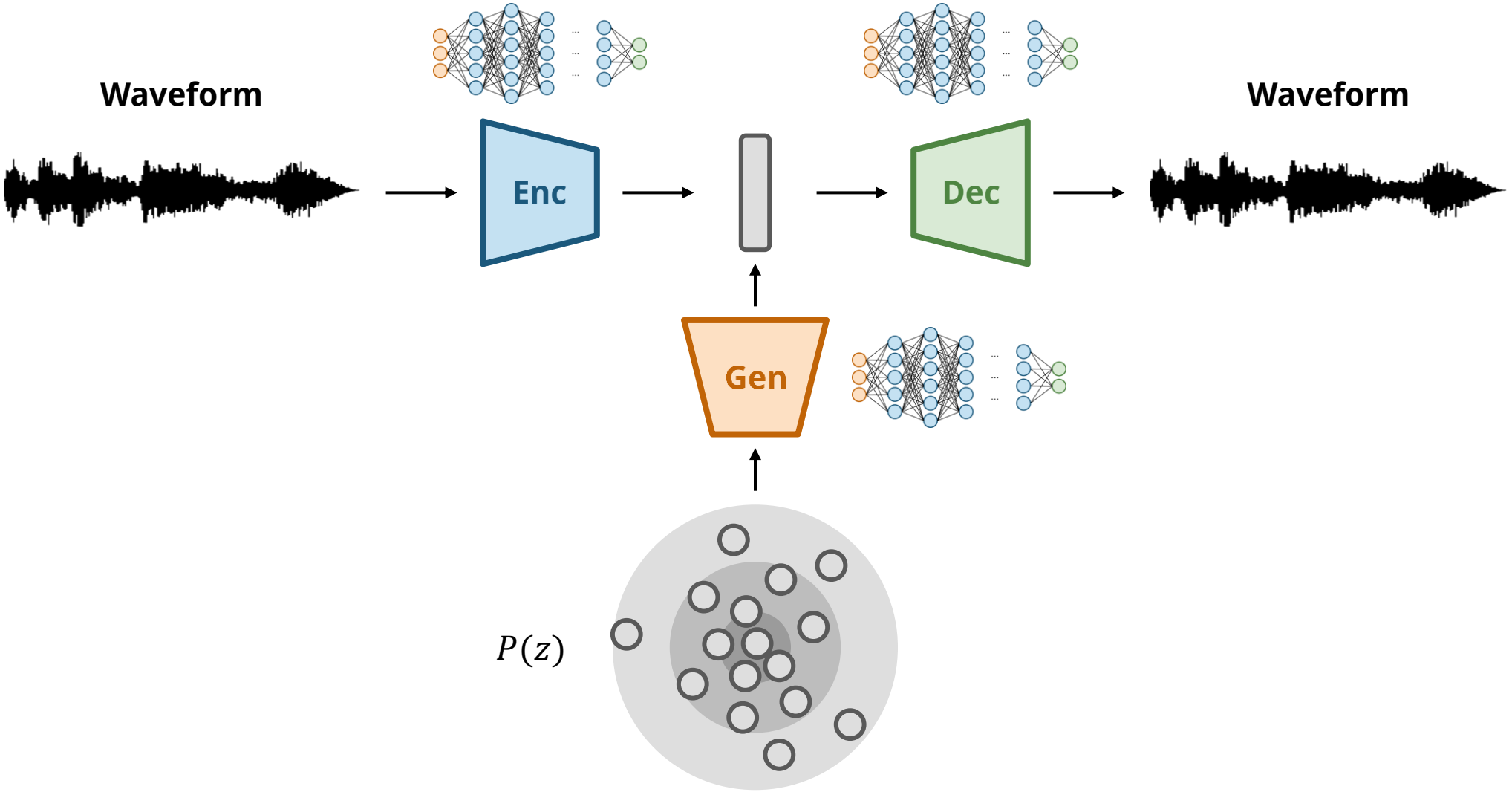
- **A neural codec**
  - An CNN-based autoencoder
  - Trained with a GAN-like adversarial loss
- **Diffusion model in the latent space**
  - A denoising U-Net
- **A conditioning module**
  - Transformer-like cross-attention mechanism



(Source: Raddato via worldhistory.org)

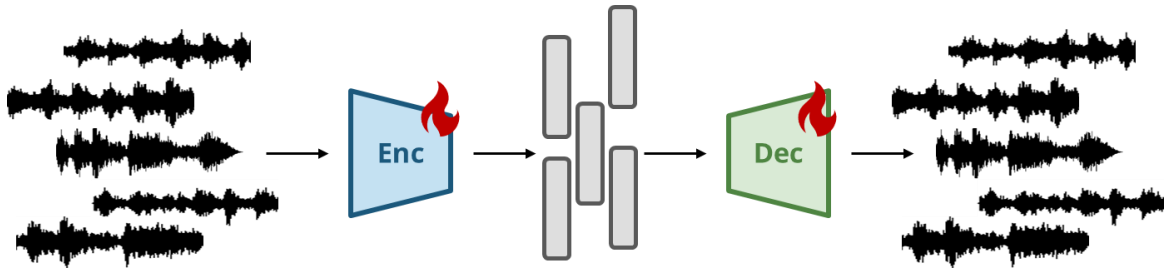
worldhistory.org/image/5462/chimera-of-arezzo-florence/

(Recap) Latent-based Audio Synthesis

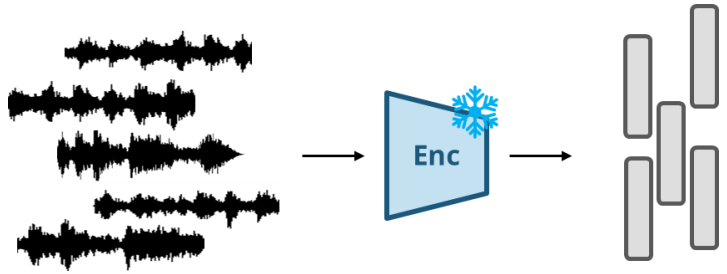


(Recap) Pipeline

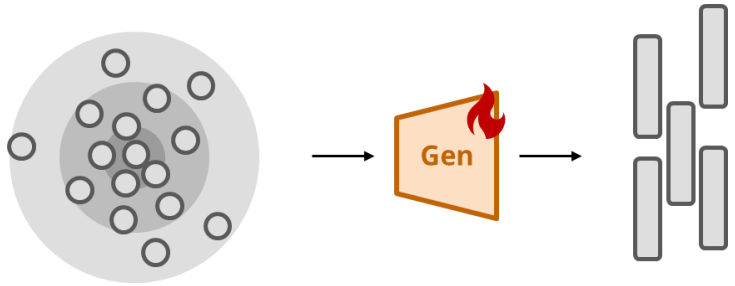
Step 1: Train an Autoencoder



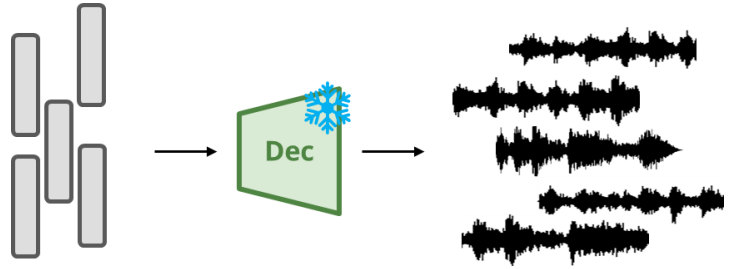
Step 2: Compute the Latent Vectors



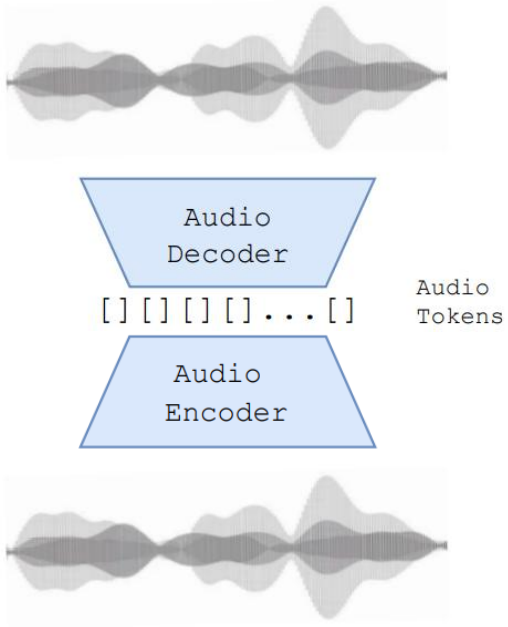
Step 3: Train a Latent Generative Model



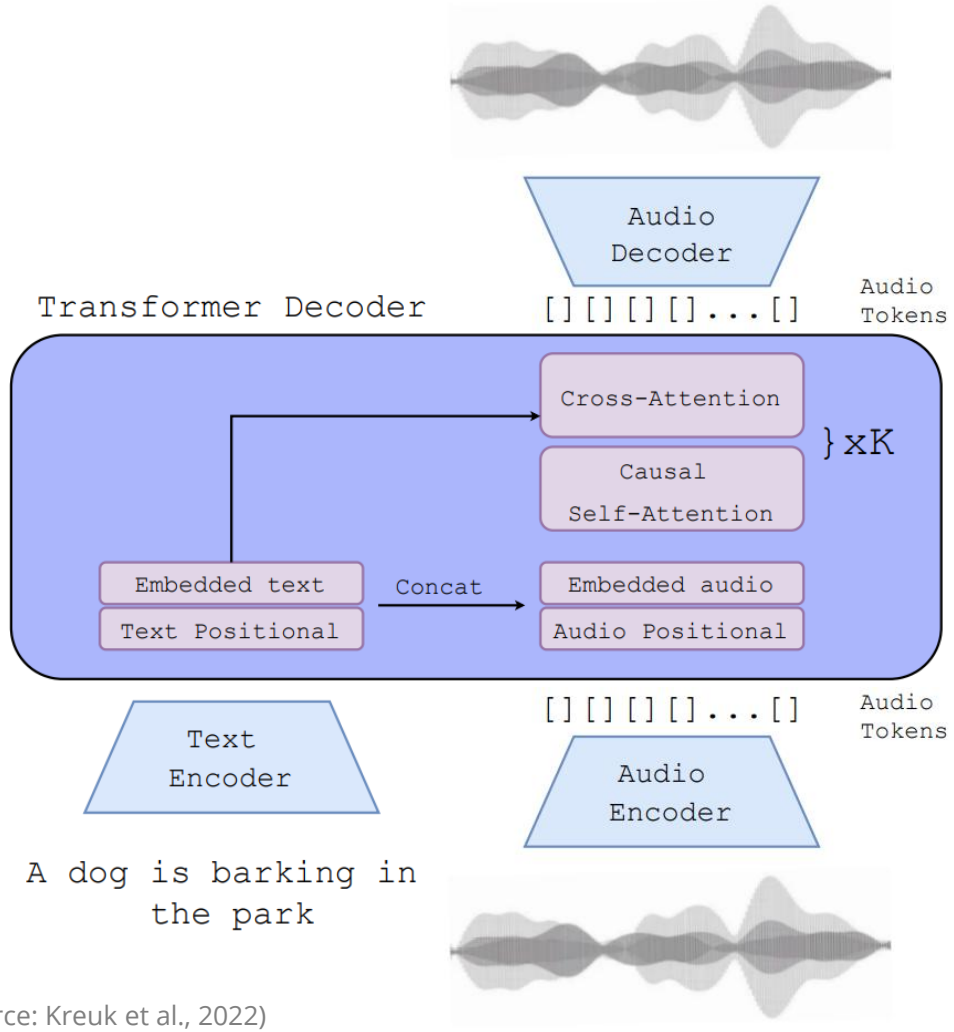
Step 4: Decode the Latent Vectors



(Recap) AudioGen (Kreuk et al., 2023)



4k hours
(speech, music, sound effects)



(Source: Kreuk et al., 2022)

Felix Kreuk, Gabriel Synnaeve, Adam Polyak, Uriel Singer, Alexandre Défossez, Jade Copet, Devi Parikh, Yaniv Taigman, and Yossi Adi, "AudioGen: Textually Guided Audio Generation," *ICLR*, 2023.

(Recap) MusicGen (Copet et al., 2023)

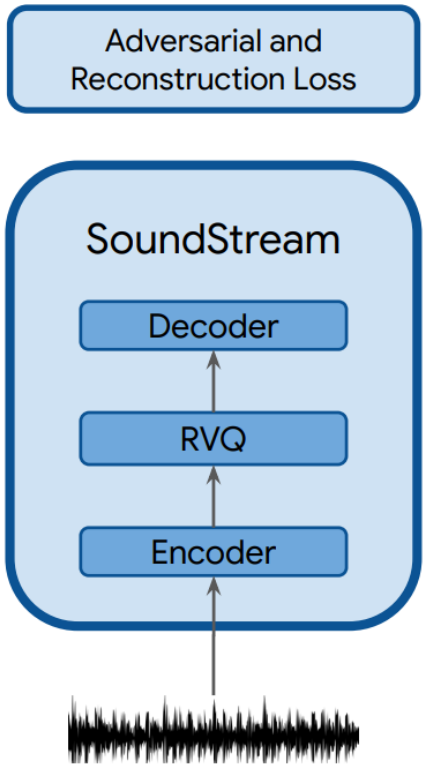
- AudioGen for Music
- Use EnCodec (Défossez et al., 2022) as the autoencoder
  - instead of SoundStream for AudioGen (Kreuk et al., 2023)
- **20k hours** of licensed music
  - Internal dataset 10k High-quality (private)
  - ShutterStock 25k Instrument-only
  - Pond5 365k Instrument-only

ai.honu.io/papers/musicgen/

Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez, "Simple and Controllable Music Generation," *NeurIPS*, 2023.

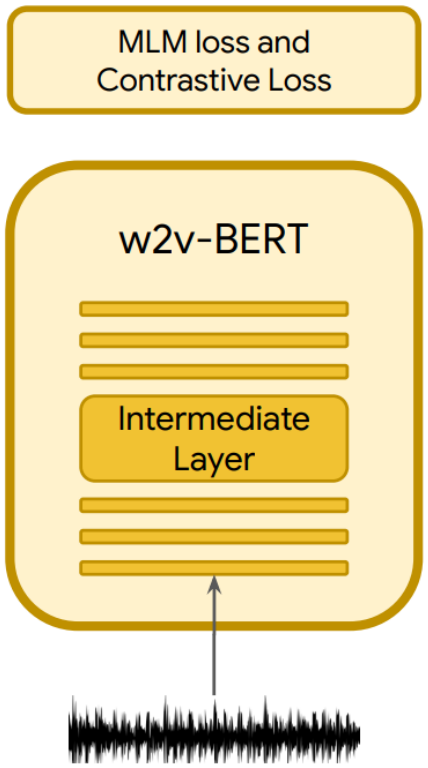
(Recap) MusicLM (Agostinelli et al., 2023)

Audio autoencoder

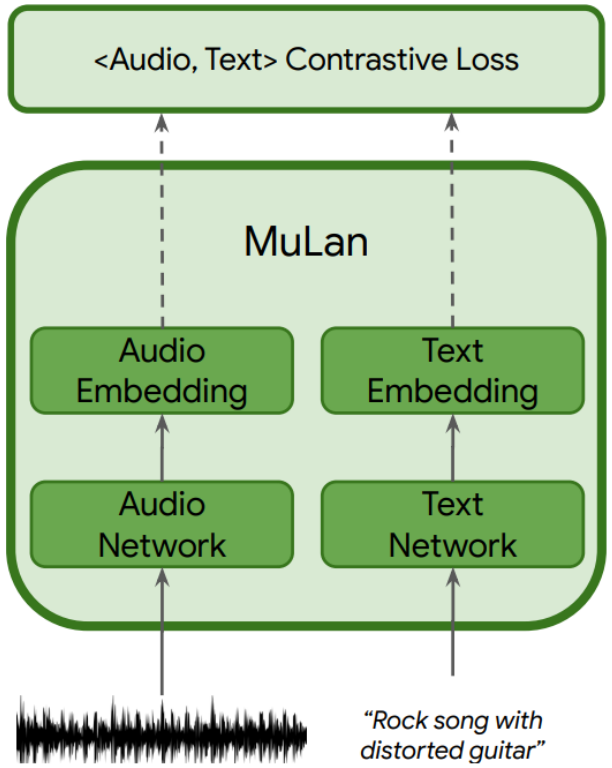


106k songs, 8.2k hours

Semantic representation



Text-music correspondence

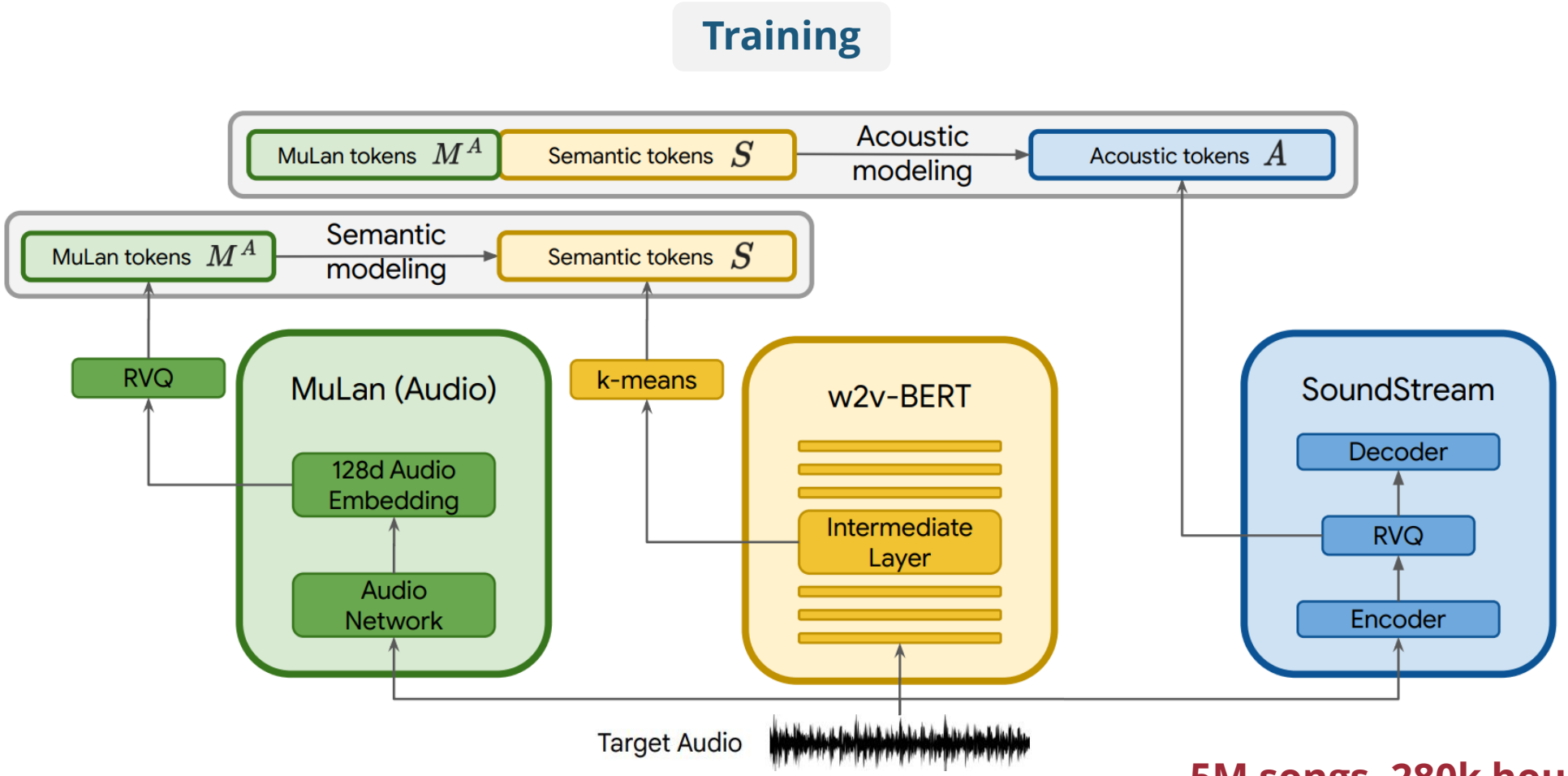


44M 30-sec clips, 370k hours

(Source: Agostinelli et al., 2022)

Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank, "MusicLM: Generating Music From Text," *arXiv preprint arXiv:2301.11325*, 2023.

(Recap) MusicLM (Agostinelli et al., 2023)

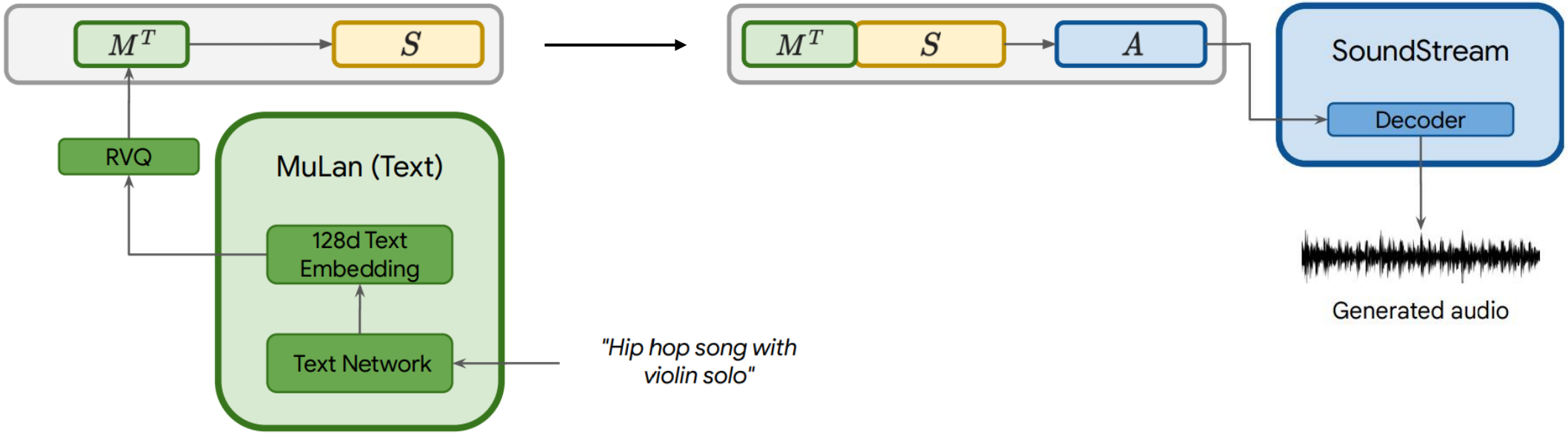


(Source: Agostinelli et al., 2022)

Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank, "MusicLM: Generating Music From Text," *arXiv preprint arXiv:2301.11325*, 2023.

(Recap) MusicLM (Agostinelli et al., 2023)

Inference

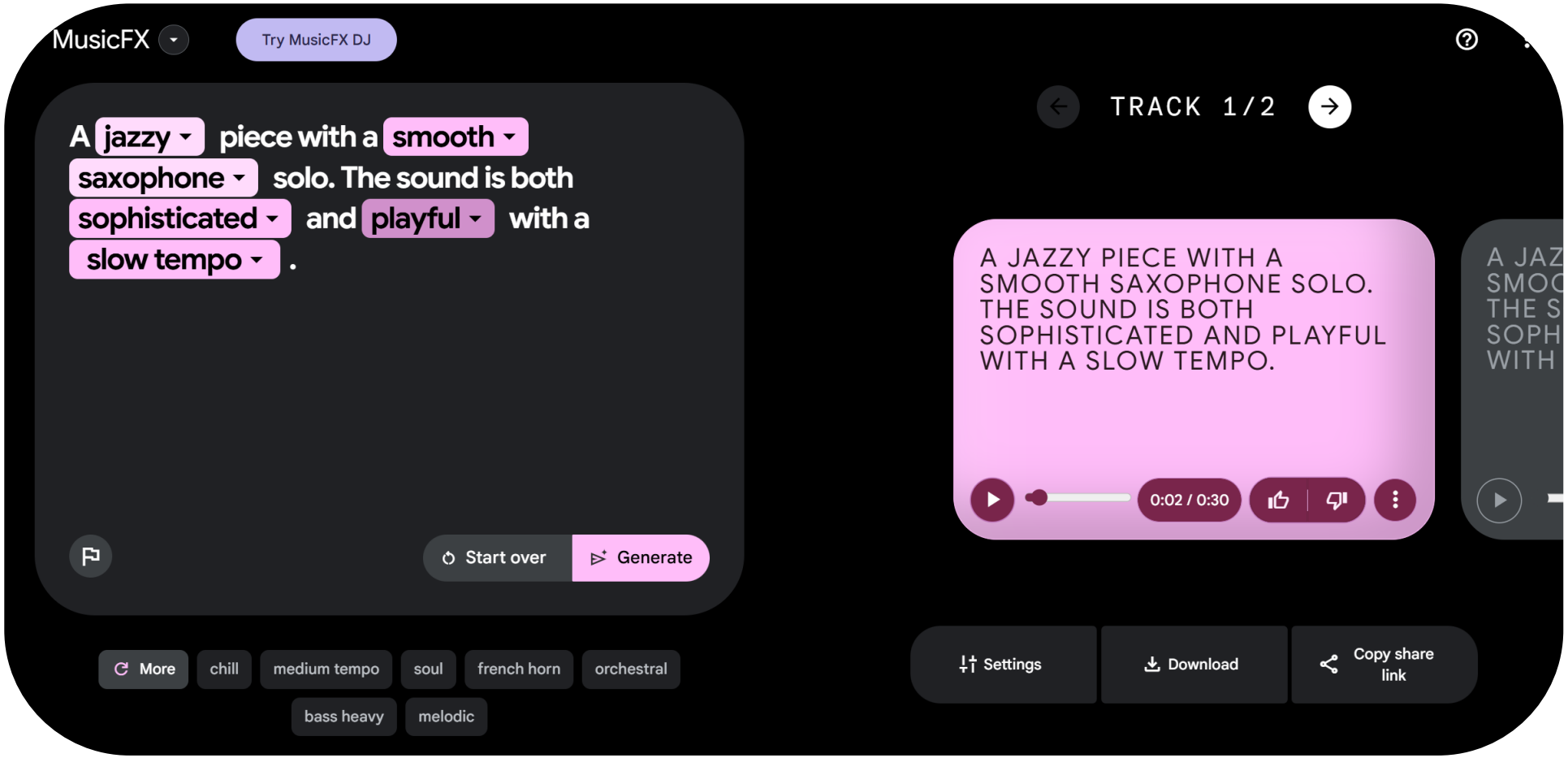


(Source: Agostinelli et al., 2022)

google-research.github.io/seanet/musiclm/examples/

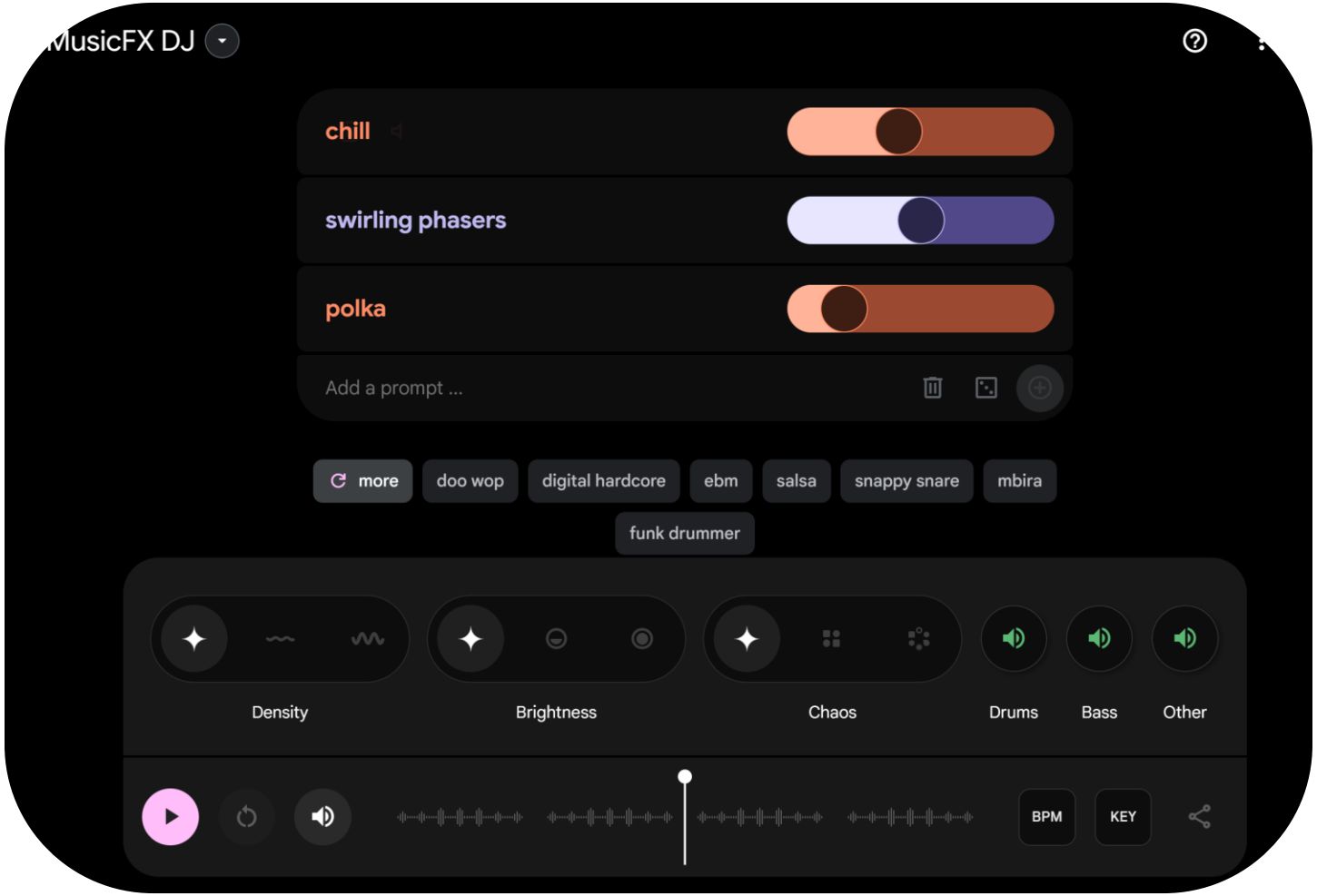
Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank, "MusicLM: Generating Music From Text," *arXiv preprint arXiv:2301.11325*, 2023.

(Recap) Music FX (2024)



aitestkitchen.withgoogle.com/tools/music-fx

(Recap) Music FX DJ (2024)

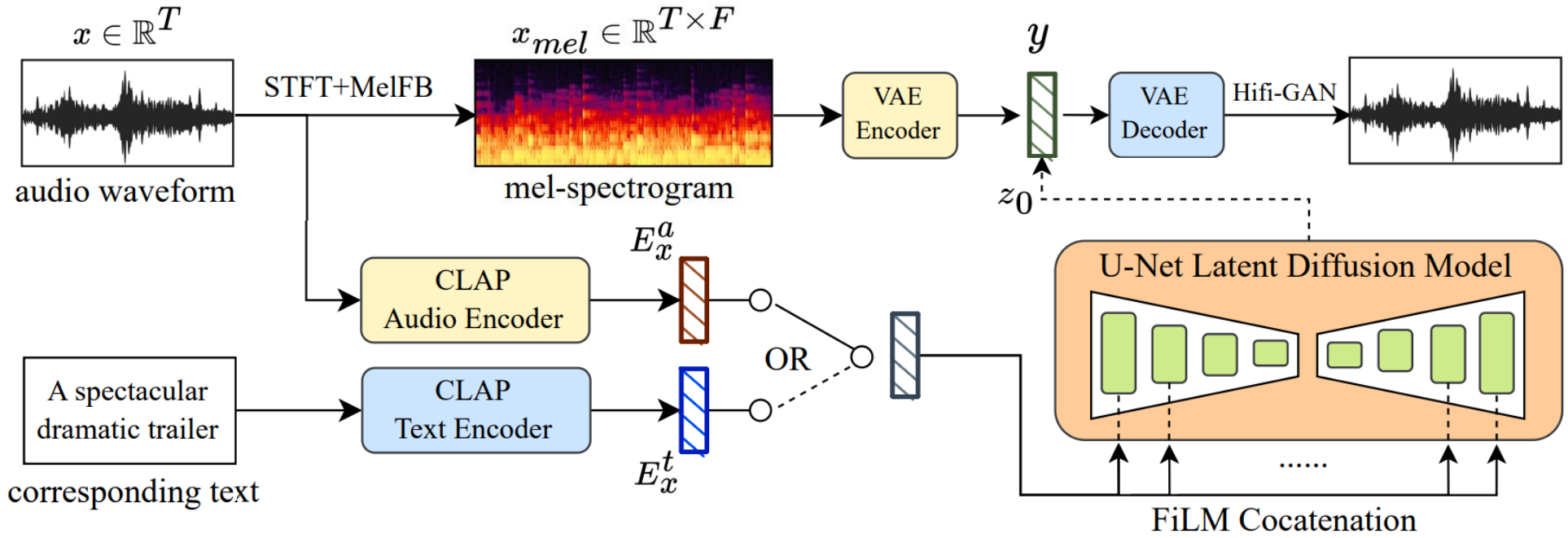


aitestkitchen.withgoogle.com/tools/music-fx-dj

(Recap) Music FX DJ (2024)

youtube.com/live/IUQW5LgBZvQ

(Recap) Example: MusicLDM (Chen et al., 2023)

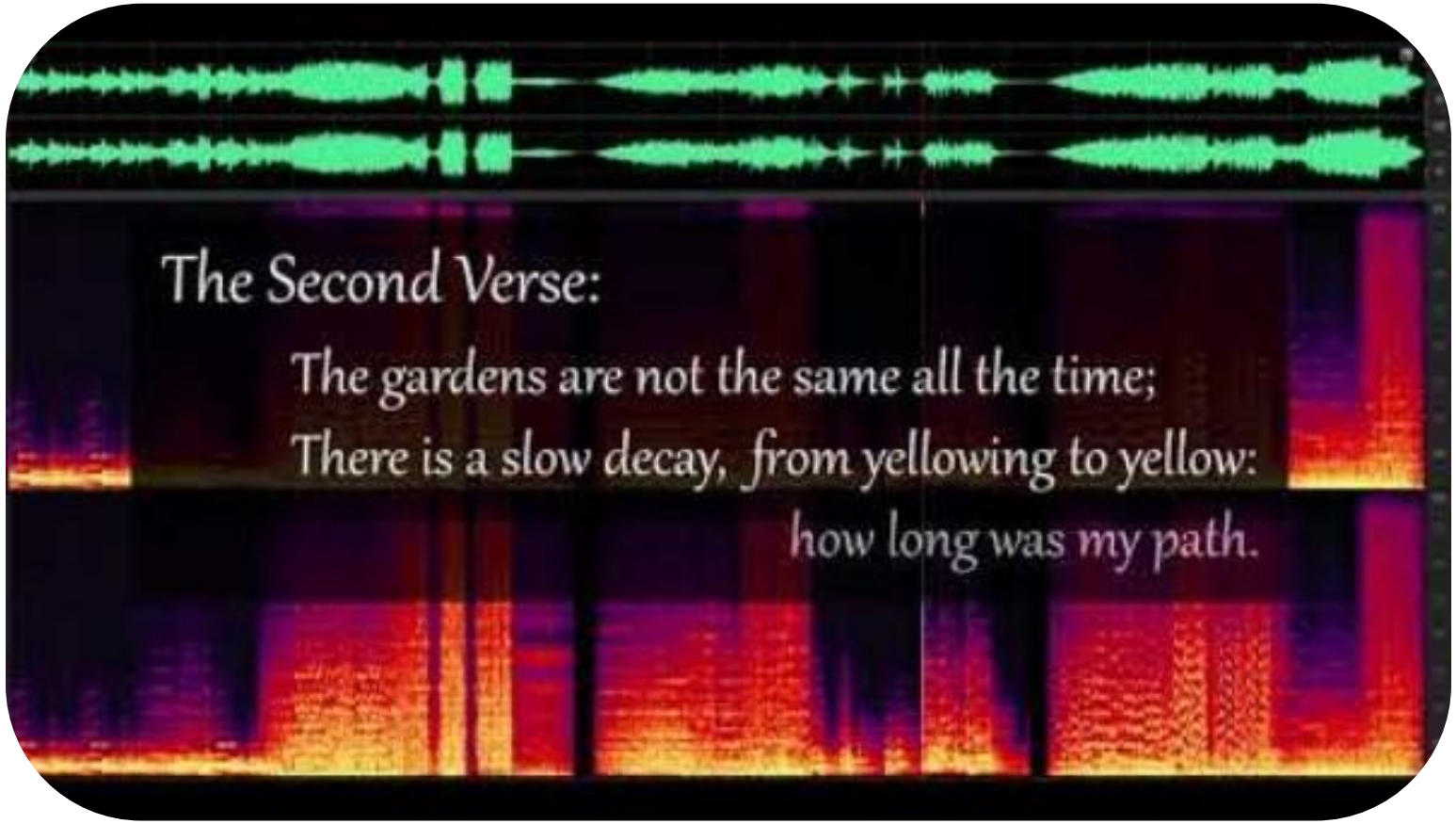


(Source: Ke et al., 2023)

musicldm.github.io

Ke Chen, Yusong Wu, Haohe Liu, Marianna Nezhurina, Taylor Berg-Kirkpatrick, and Shlomo Dubnov, "MusicLDM: Enhancing Novelty in Text-to-Music Generation Using Beat-Synchronous Mixup Strategies," *ICASSP*, 2024.

(Recap) Example: MusicLDM (Chen et al., 2023)



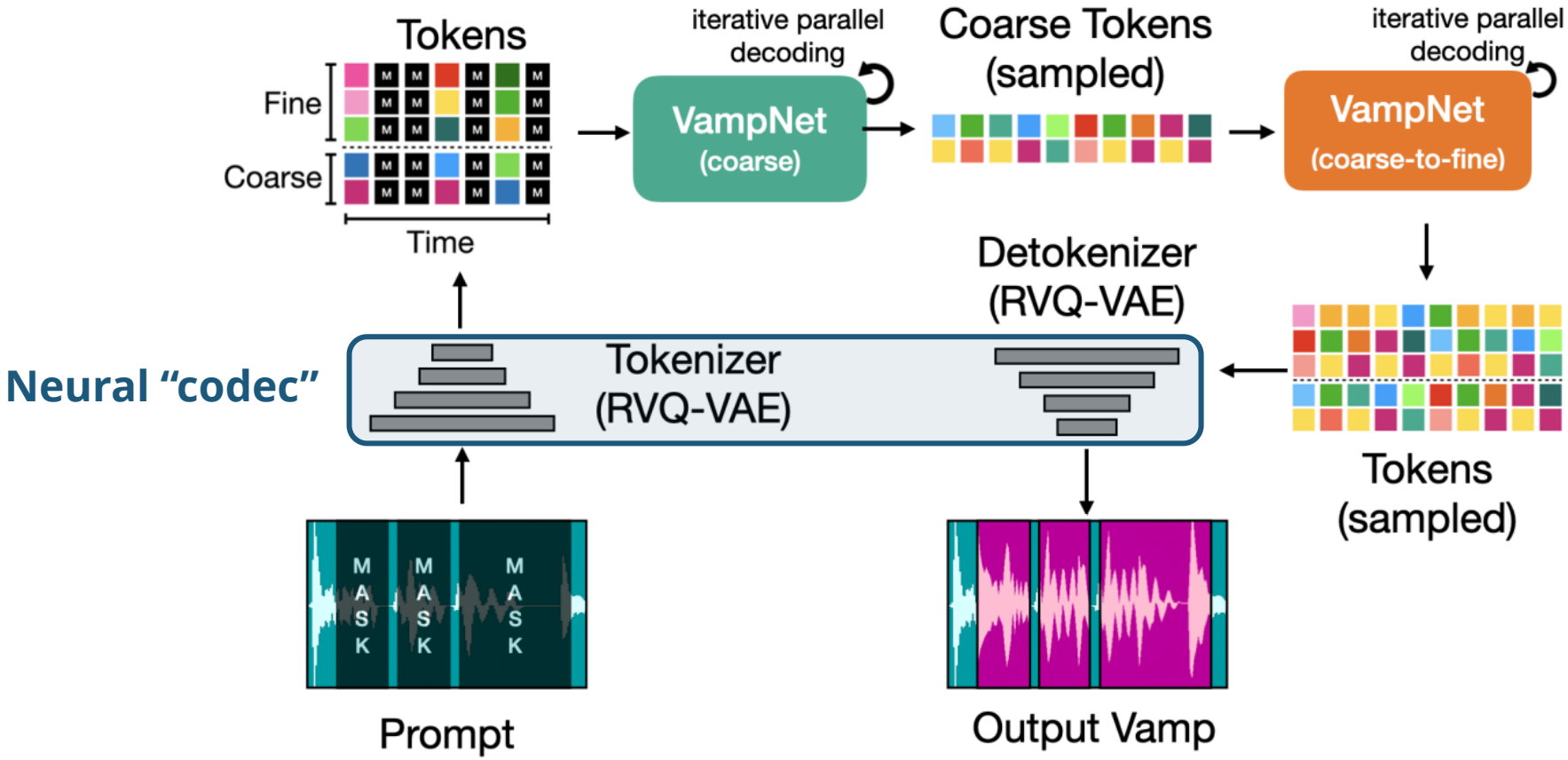
youtu.be/DALv7ea6cv0

(Recap) unloop (Garcia et al., 2023)

youtu.be/yzBI8Vcjd2s

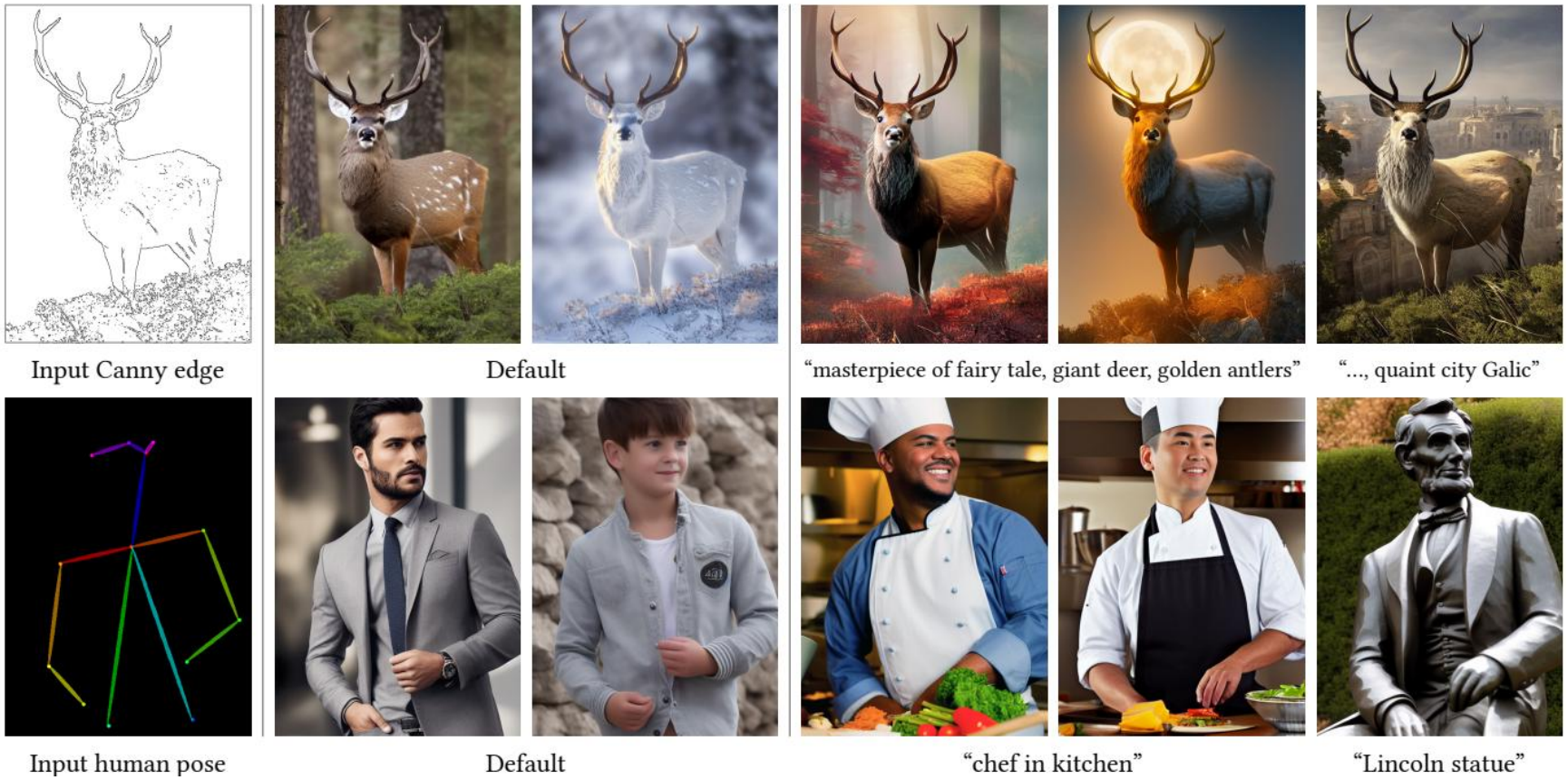
github.com/hugofloresgarcia/unloop

(Recap) VampNet (Garcia et al., 2023)



(Source: Garcia et al., 2023)

(Recap) ControlNet (Zhang et al., 2023)

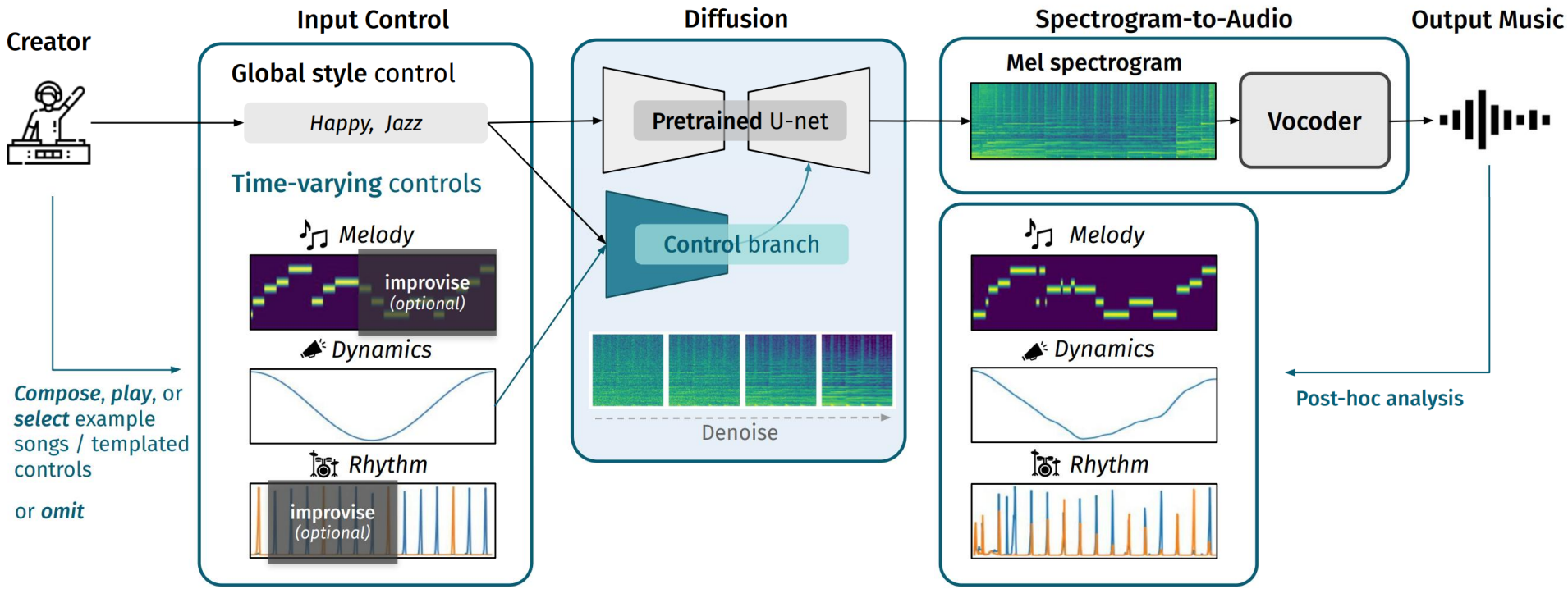


(Source: Zhang et al., 2023)

Can we add controls to a trained text-to-image diffusion model?

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala, "Adding Conditional Control to Text-to-Image Diffusion Models," *ICCV*, 2023.

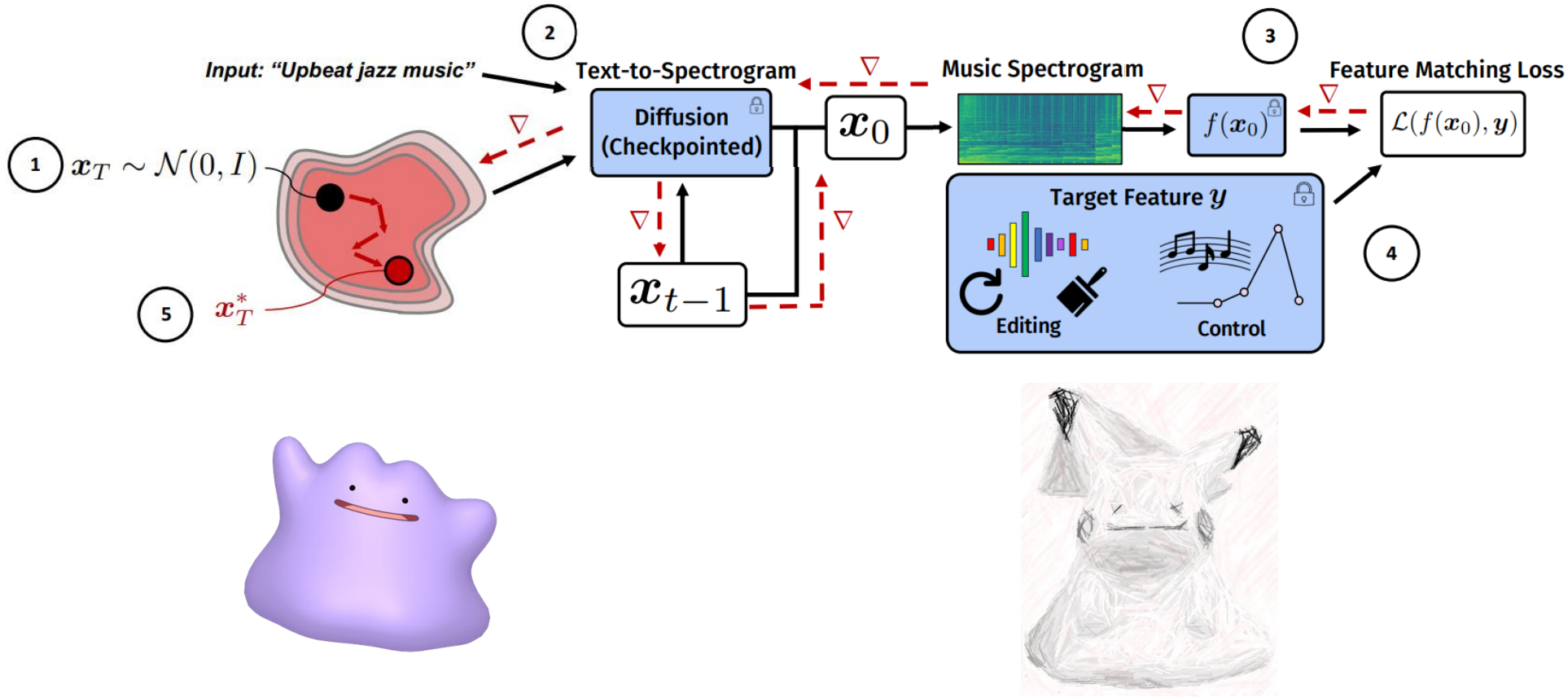
(Recap) Music ControlNet (Wu et al., 2024)



(Source: Wu et al., 2024)

Shih-Lun Wu, Chris Donahue, Shinji Watanabe, and Nicholas J. Bryan, "Music ControlNet: Multiple Time-varying Controls for Music Generation," *TASLP*, 2024.

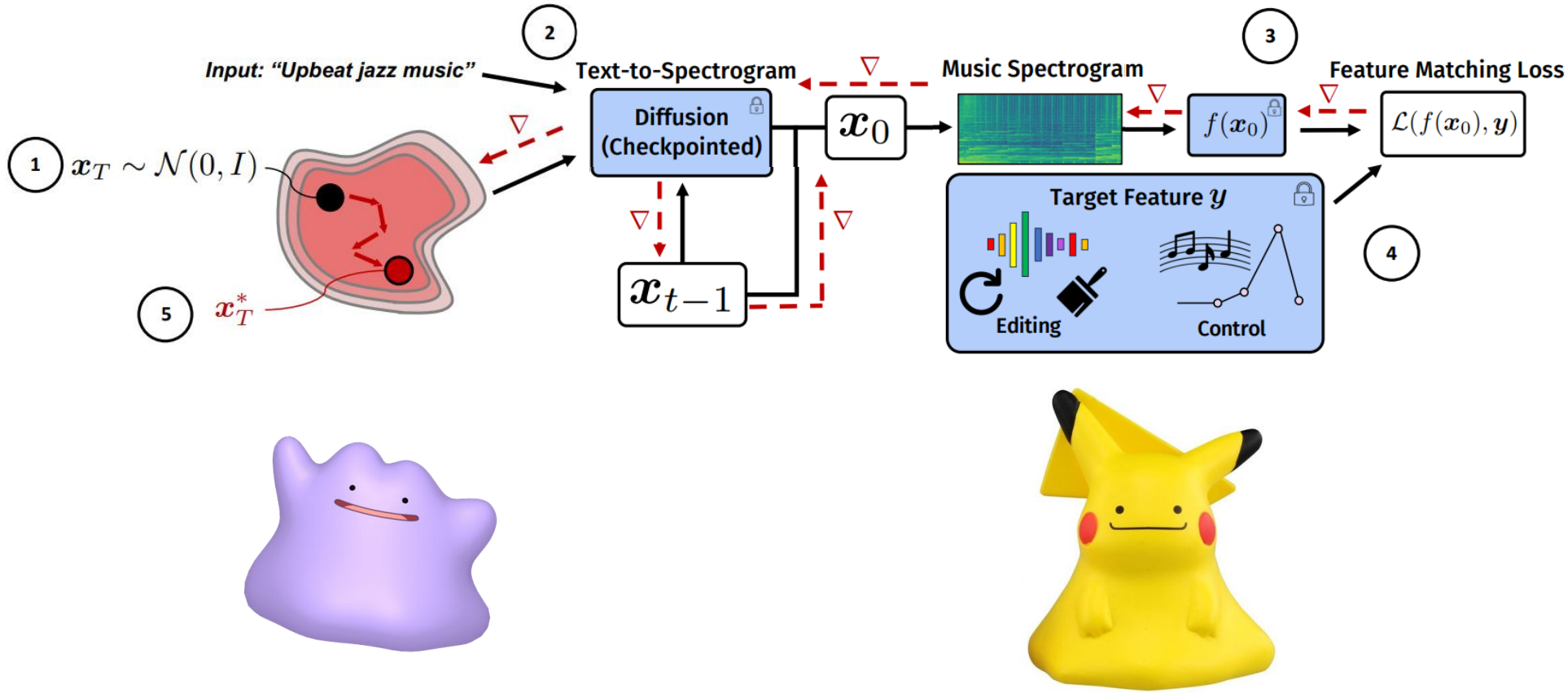
(Recap) DITTO (Novack et al., 2024)



(Source: Novack et al., 2024)

Zachary Novack, Julian McAuley, Taylor Berg-Kirkpatrick, and Nicholas J. Bryan, "DITTO: Diffusion Inference-Time T-Optimization for Music Generation," *ICML*, 2024.

(Recap) DITTO (Novack et al., 2024)

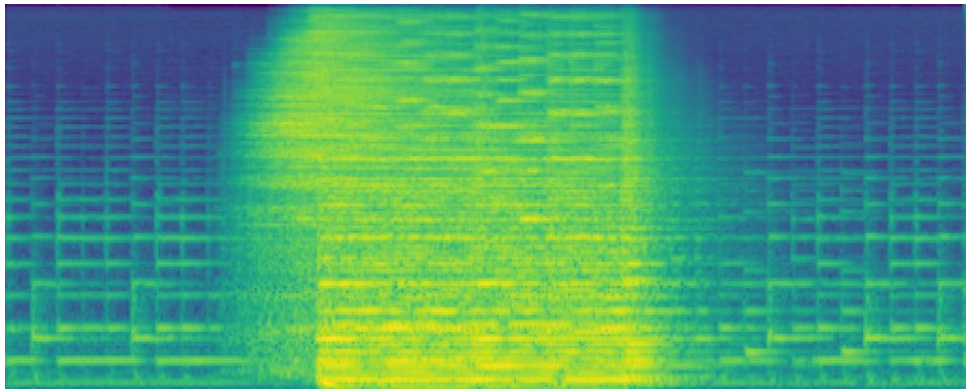


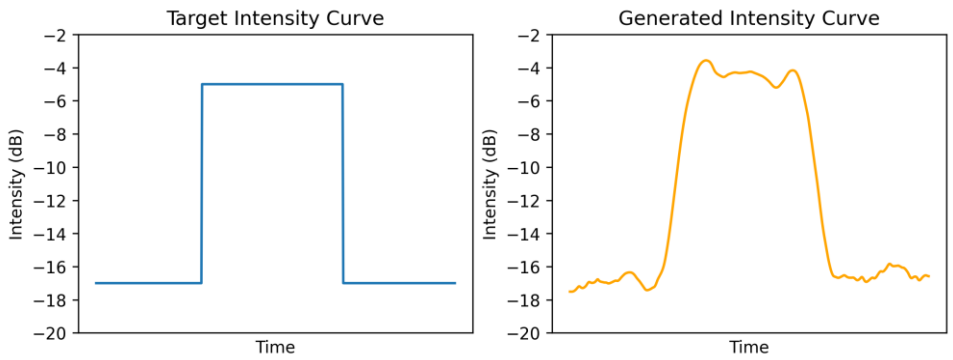
(Source: Novack et al., 2024)

Zachary Novack, Julian McAuley, Taylor Berg-Kirkpatrick, and Nicholas J. Bryan, "DITTO: Diffusion Inference-Time T-Optimization for Music Generation," *ICML*, 2024.

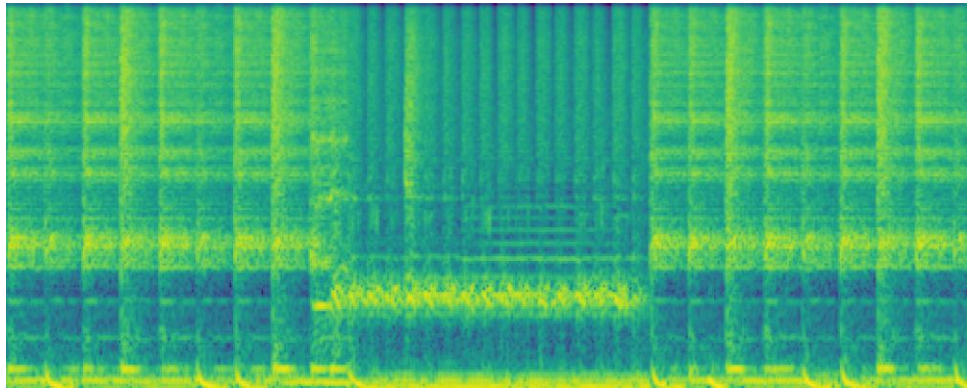
(Recap) DITTO (Novack et al., 2024)

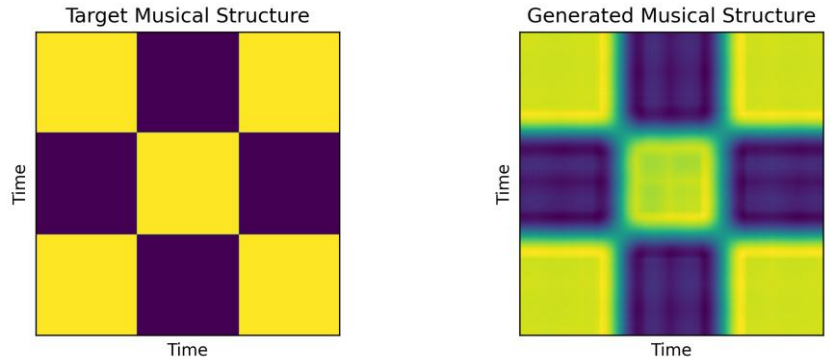
Intensity control





Structure control





(Source: Novack et al., 2024)

Zachary Novack, Julian McAuley, Taylor Berg-Kirkpatrick, and Nicholas J. Bryan, "DITTO: Diffusion Inference-Time T-Optimization for Music Generation," *ICML*, 2024.

Music Fingerprinting

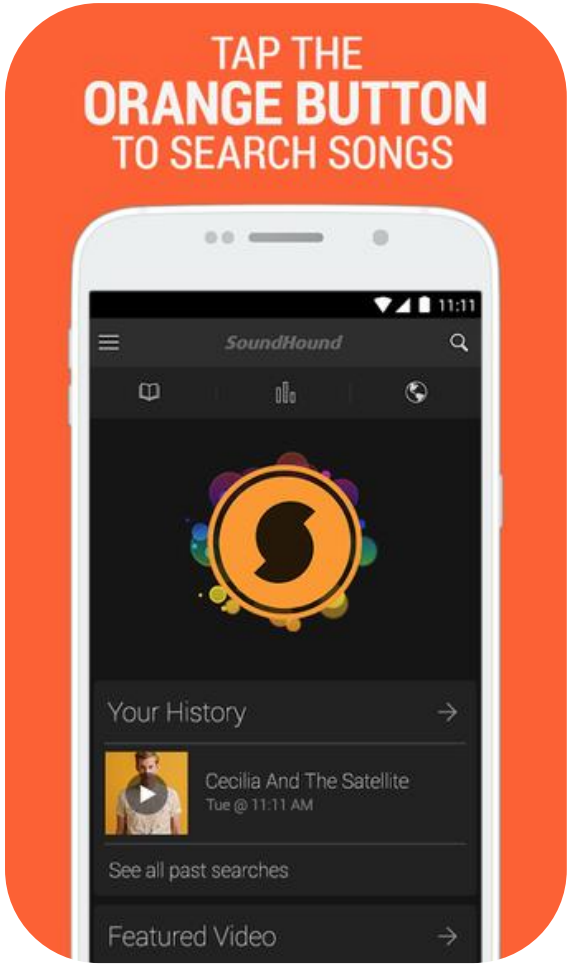
Shazam & Siri

(Source: Shazam User Guide)

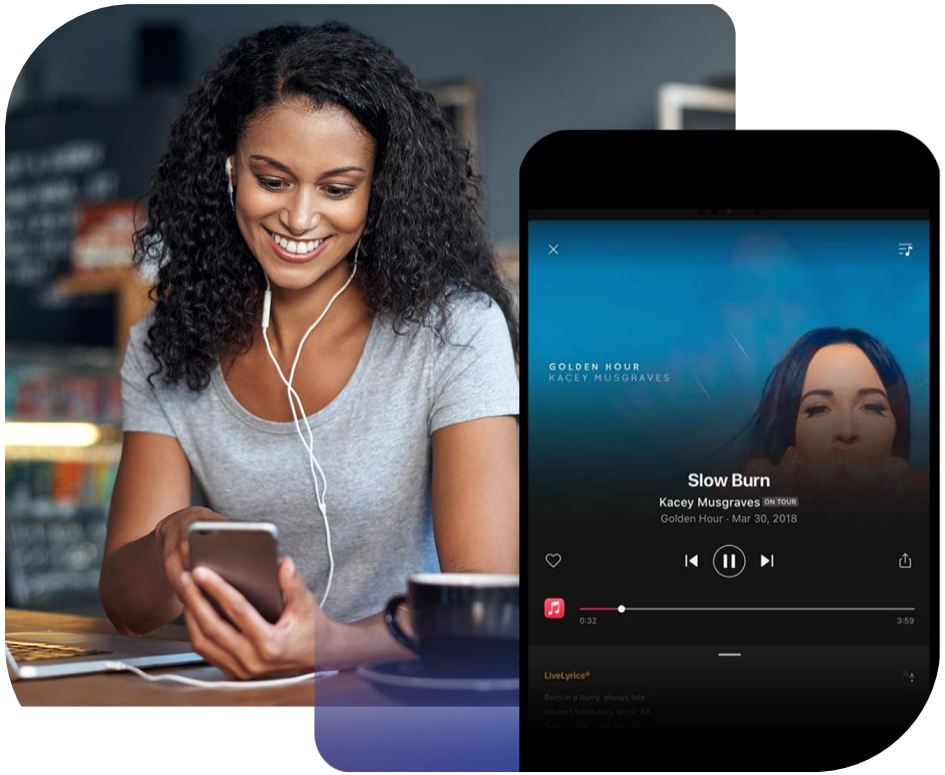
(Source: OSXDaily)

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

SoundHound



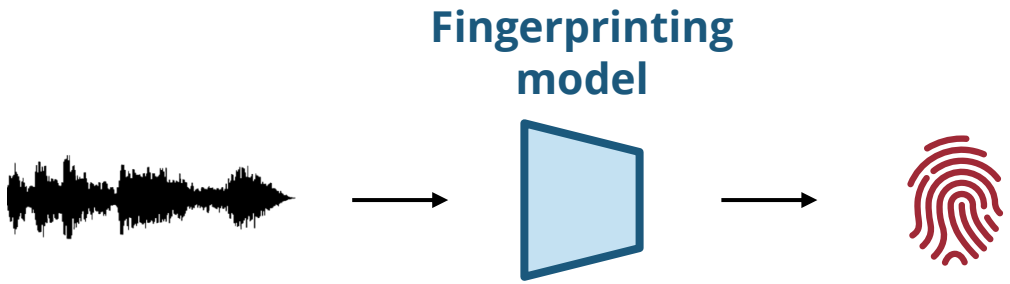
(Source: CNet)



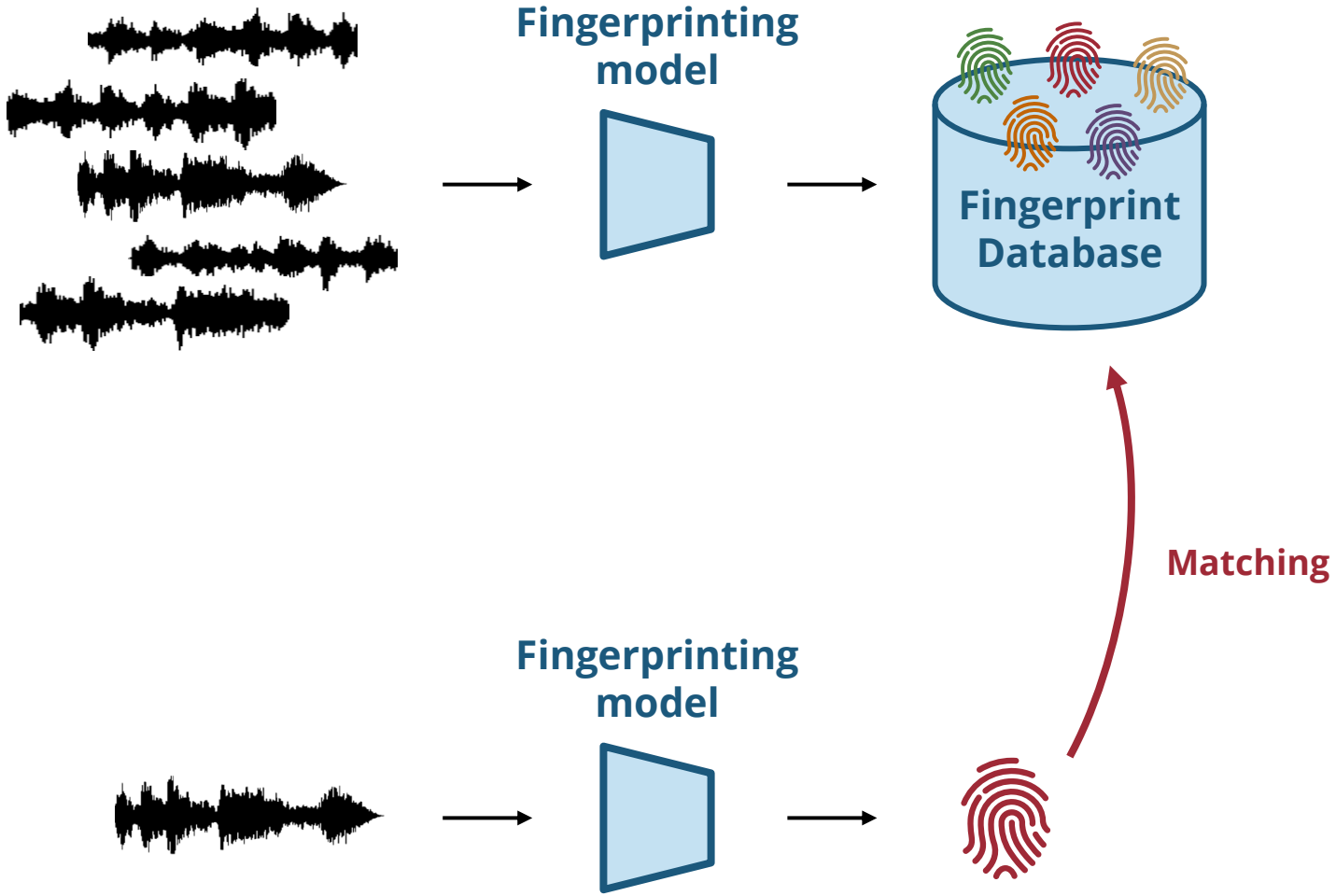
(Source: SoundHound)

download.cnet.com/soundhound-music-discovery-hands-free-player/3000-2141_4-77341748.html
soundhound.com/soundhound/

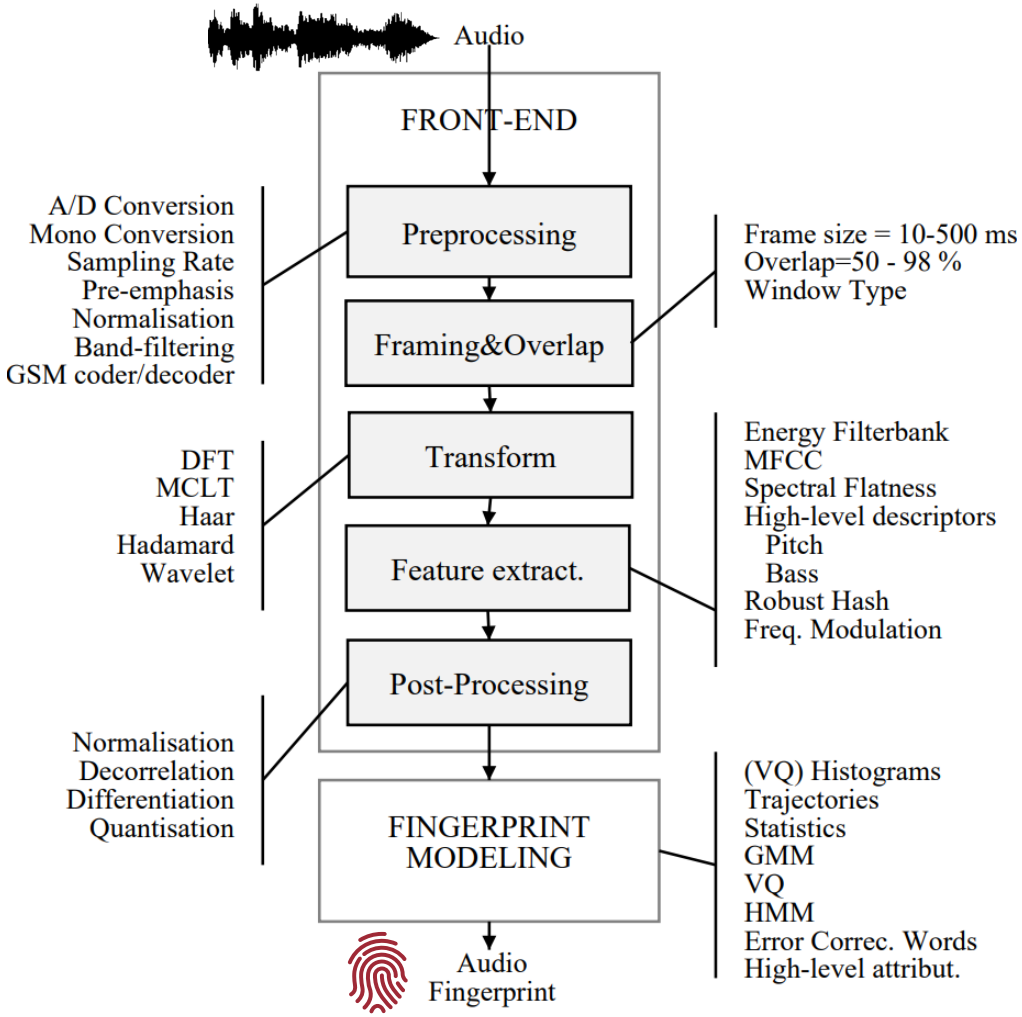
Audio Fingerprinting



Audio Fingerprinting for Audio Identification



Audio Fingerprinting

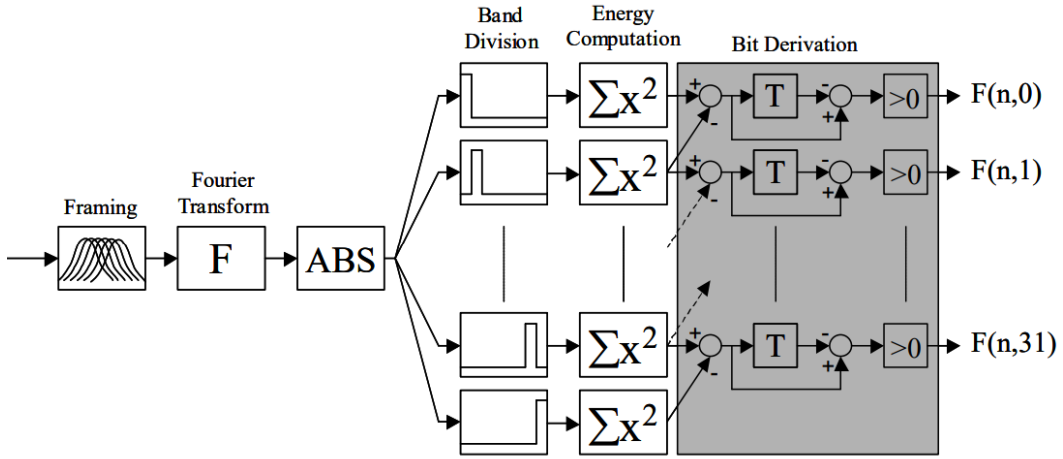


(Source: Cano et al., 2002)

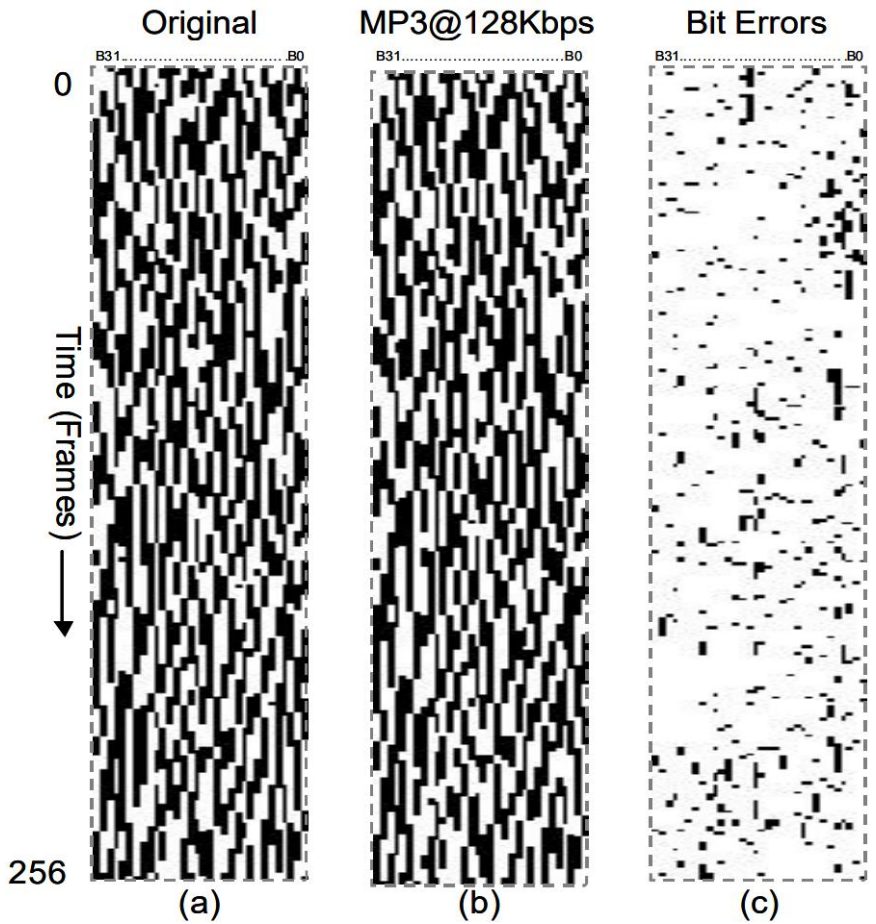
Pedro Cano, Eloi Batlle, Ton Kalker, and Jaap Haitsm, "A Review of Algorithms for Audio Fingerprinting," *MMSP*, 2002.

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

Fingerprint extraction



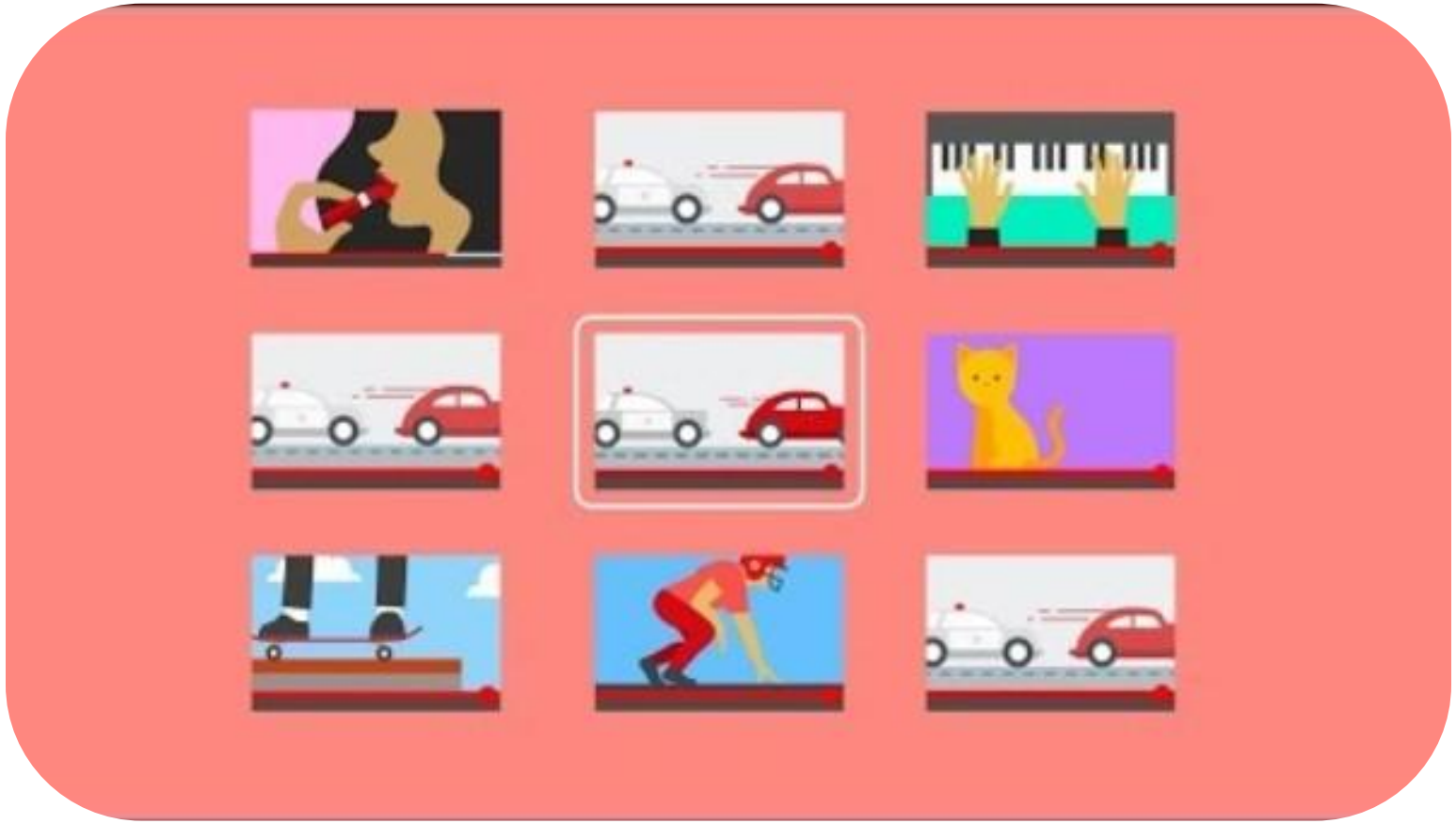
Example fingerprint blocks



(Source: Haitsma et al., 2002)

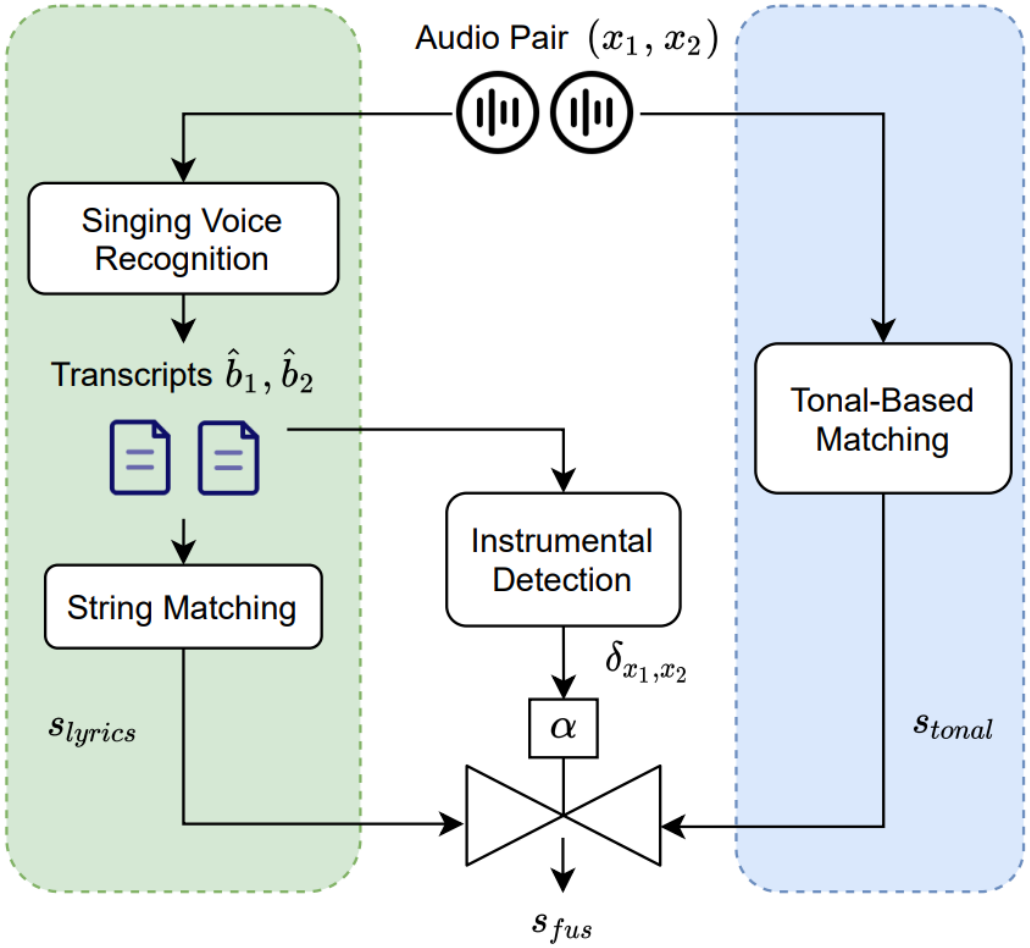
Jaap Haitsma and Ton Kalker, "A Highly Robust Audio Fingerprinting System," *ISMIR*, 2002.

YouTube’s Content ID



youtu.be/9g2U12SsRns

Cover Song Identification

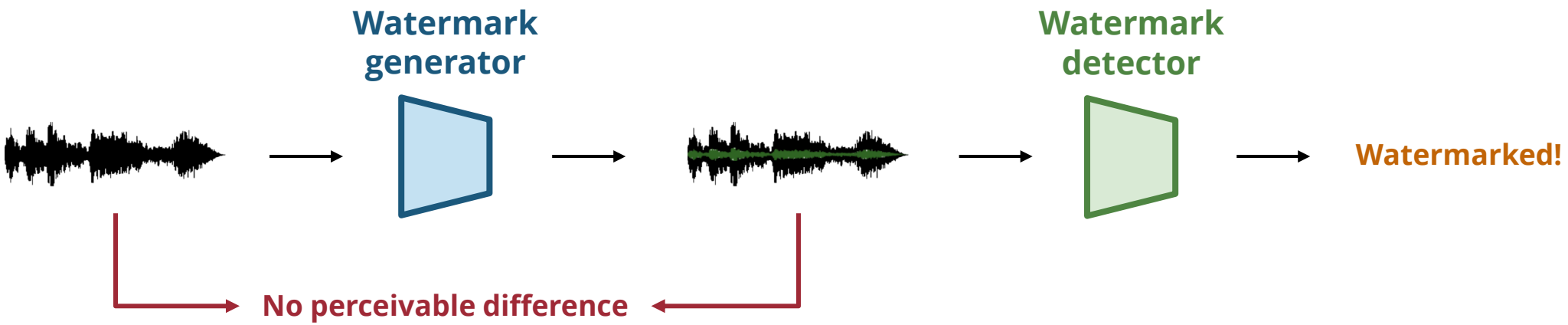


(Source: Vaglio et al., 2021)

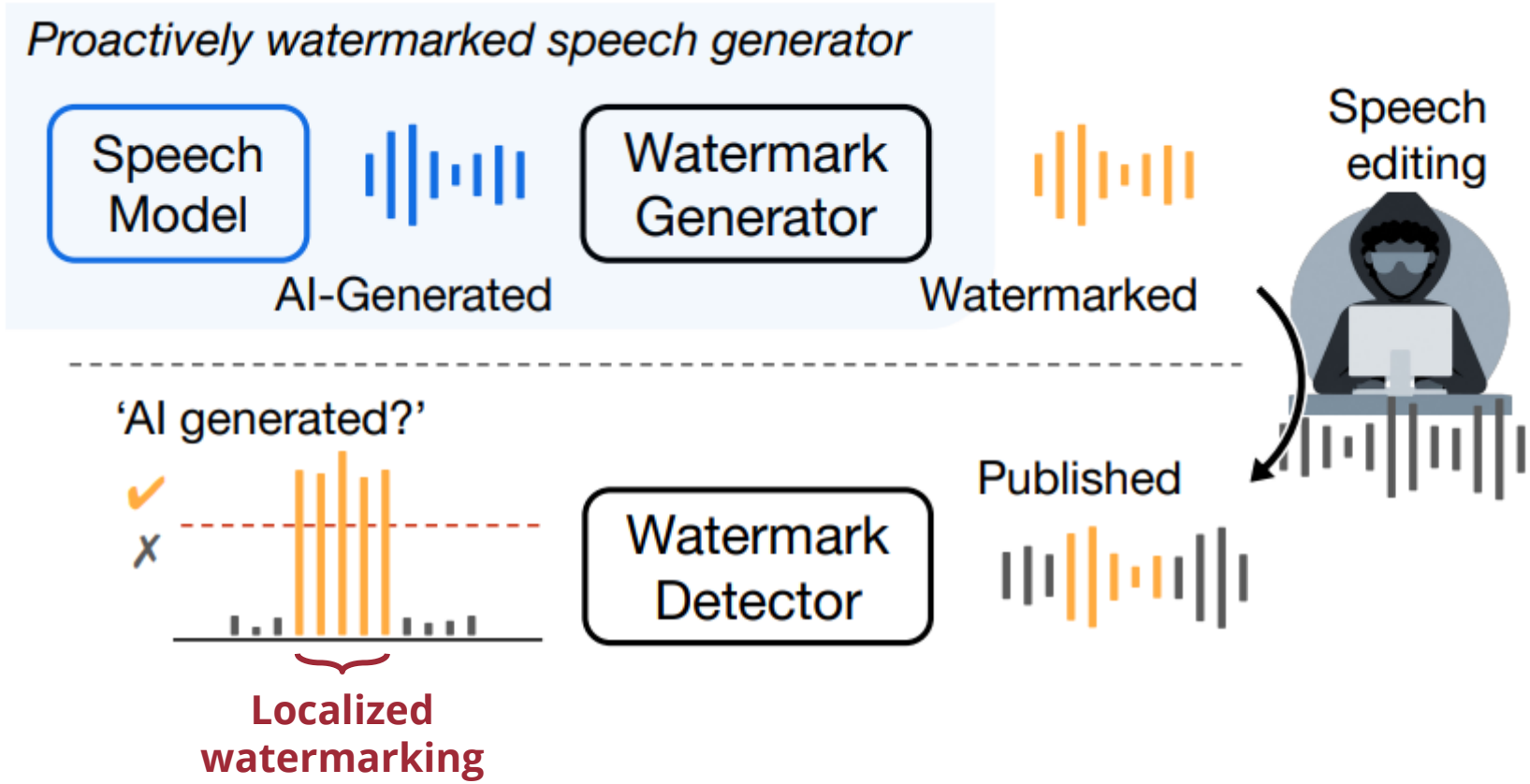
Andrea Vaglio, Romain Hennequin, Manuel Moussallam, and Gaël Richard, "The Words Remain the Same: Cover Detection with Lyrics Transcription," *ISMIR*, 2021.

Music Watermarking

Audio Watermarking



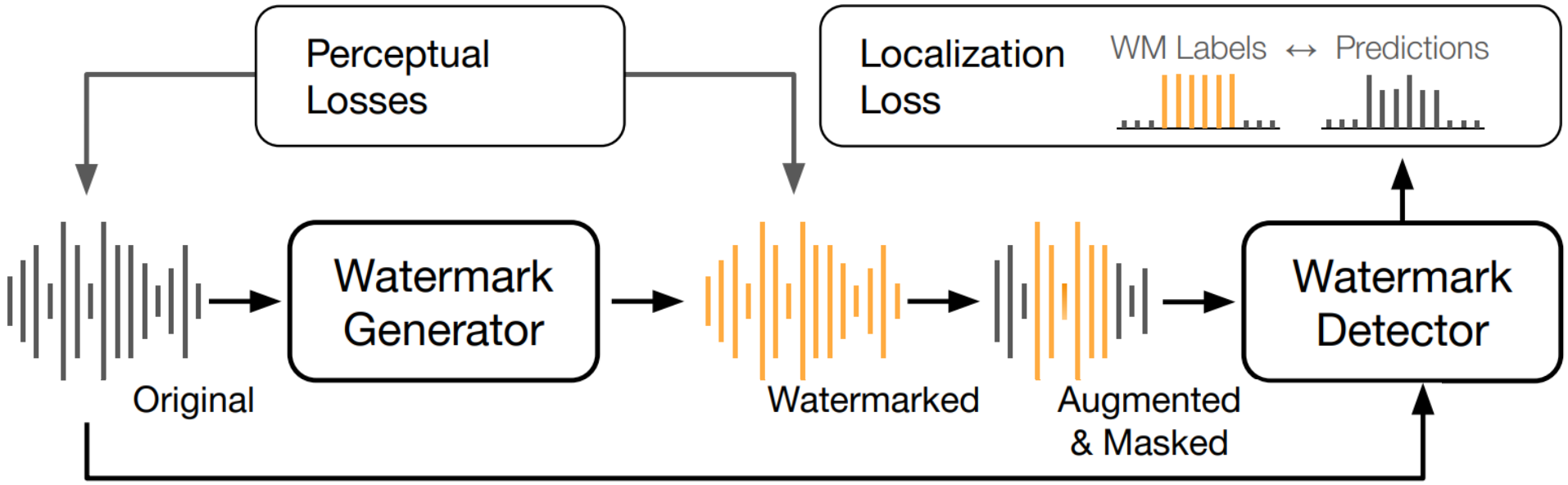
Audio Watermarking Against Generated Audio



(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.

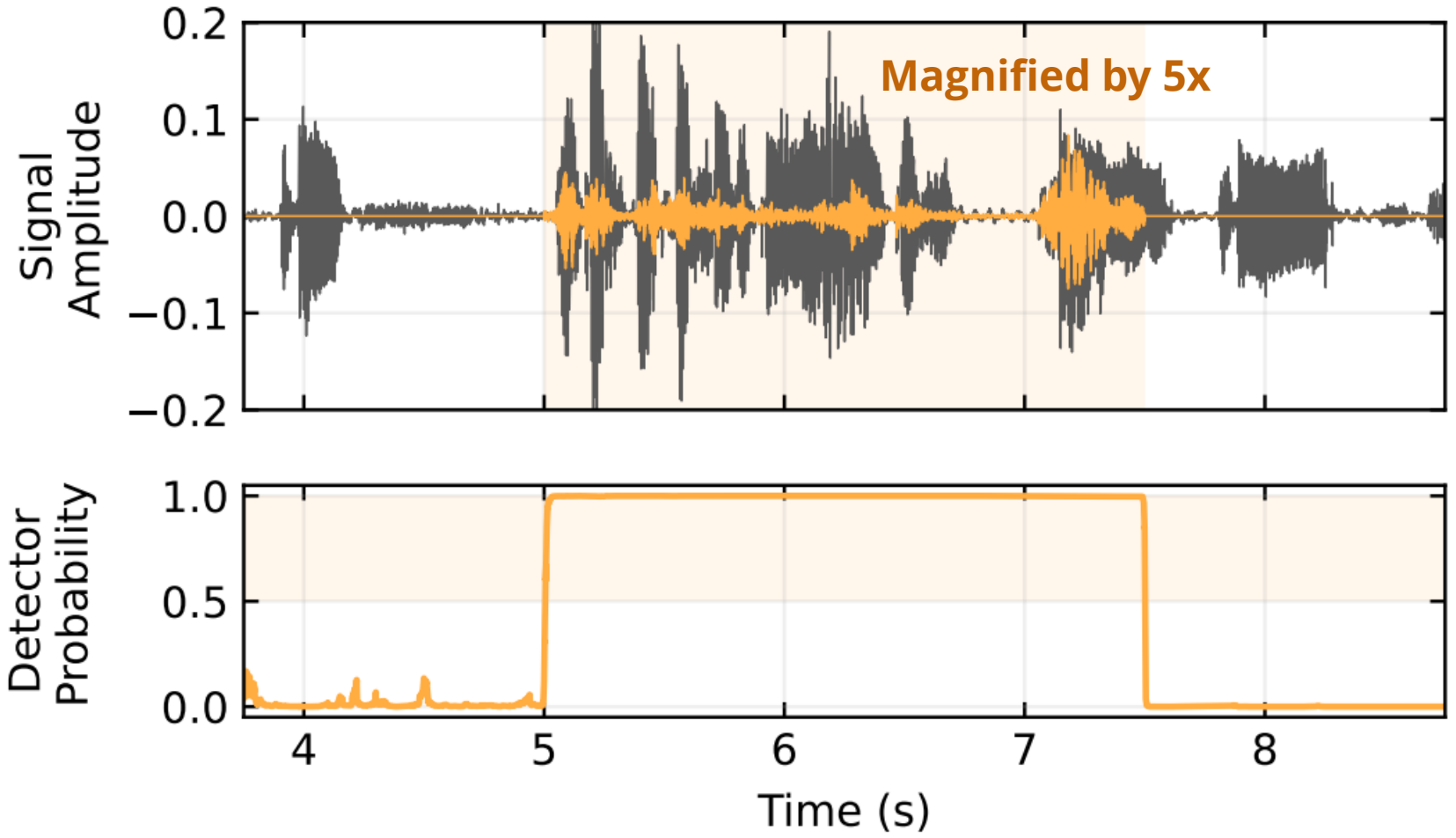
AudioSeal (Roman et al., 2024)



(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.

AudioSeal (Roman et al., 2024)

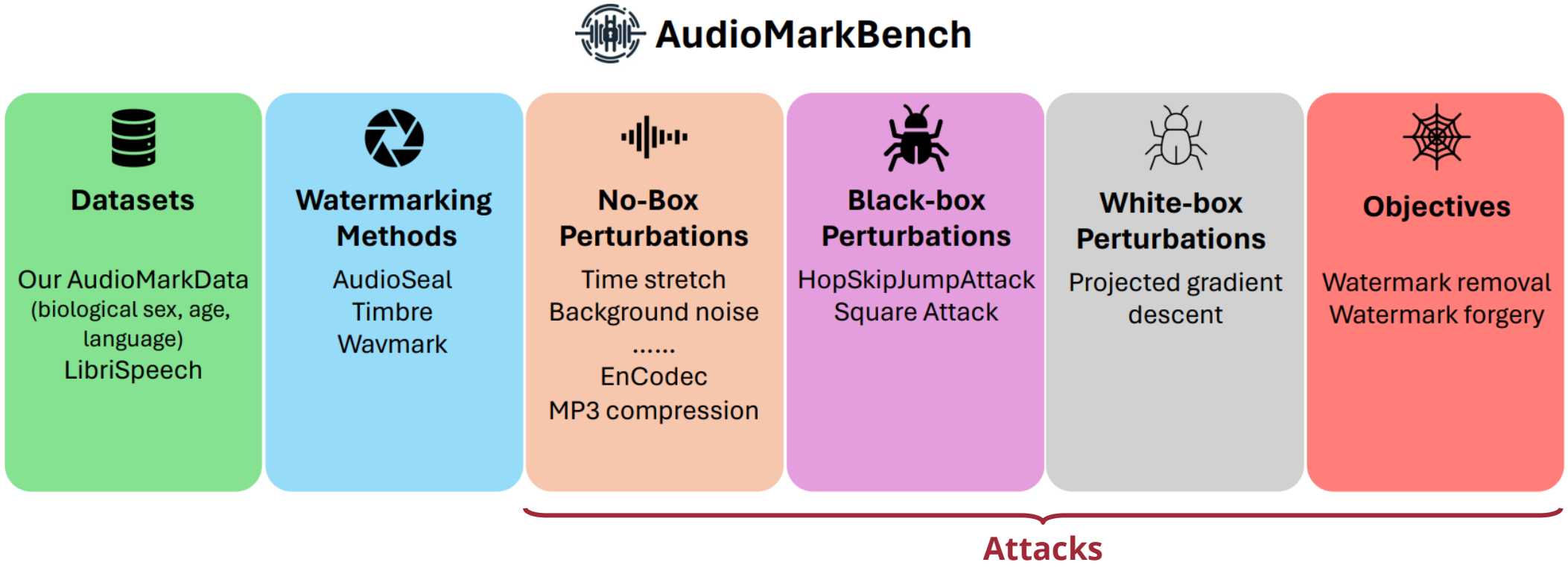


Original **Watermarked**

(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.

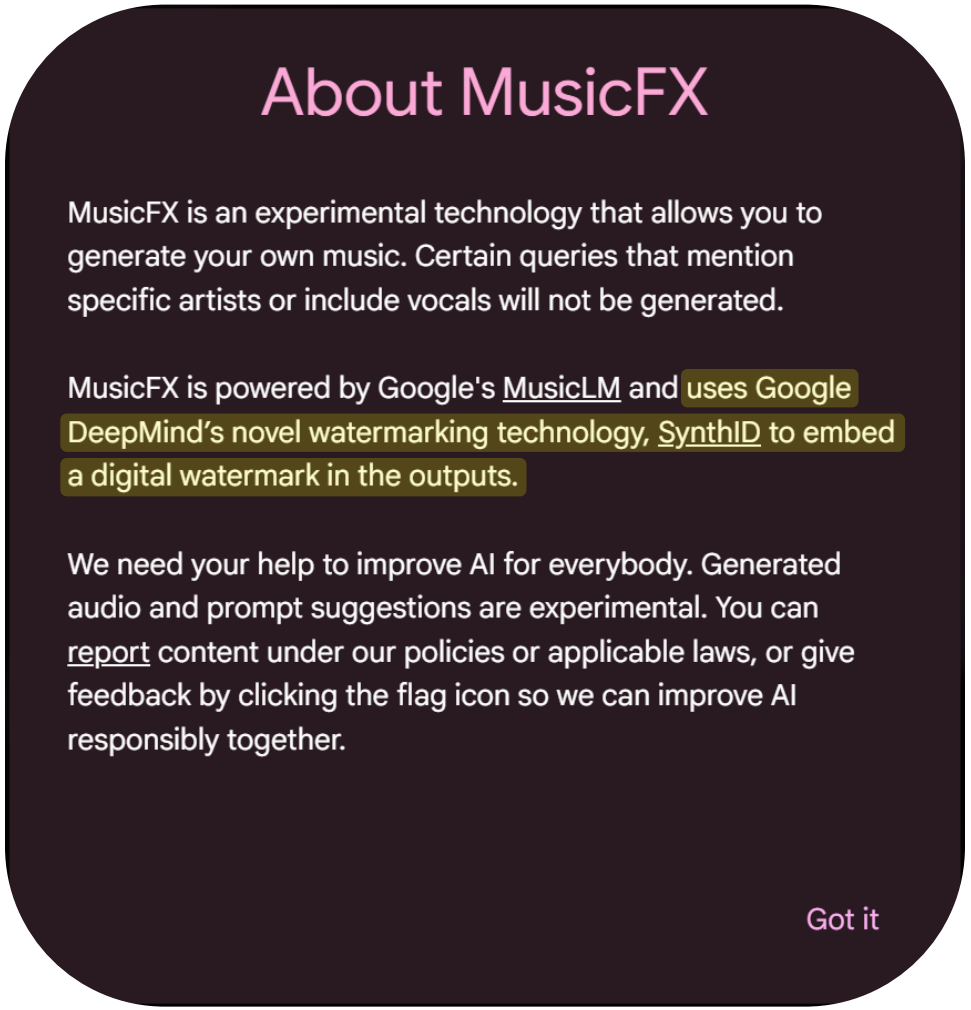
AudioMarkBench (Liu et al., 2024)



(Source: Liu et al., 2024)

Hongbin Liu, Moyang Guo, Zhengyuan Jiang, Lun Wang, and Neil Zhenqiang Gong, "AudioMarkBench: Benchmarking Robustness of Audio Watermarking," *NeurIPS Datasets and Benchmarks Track*, 2024.

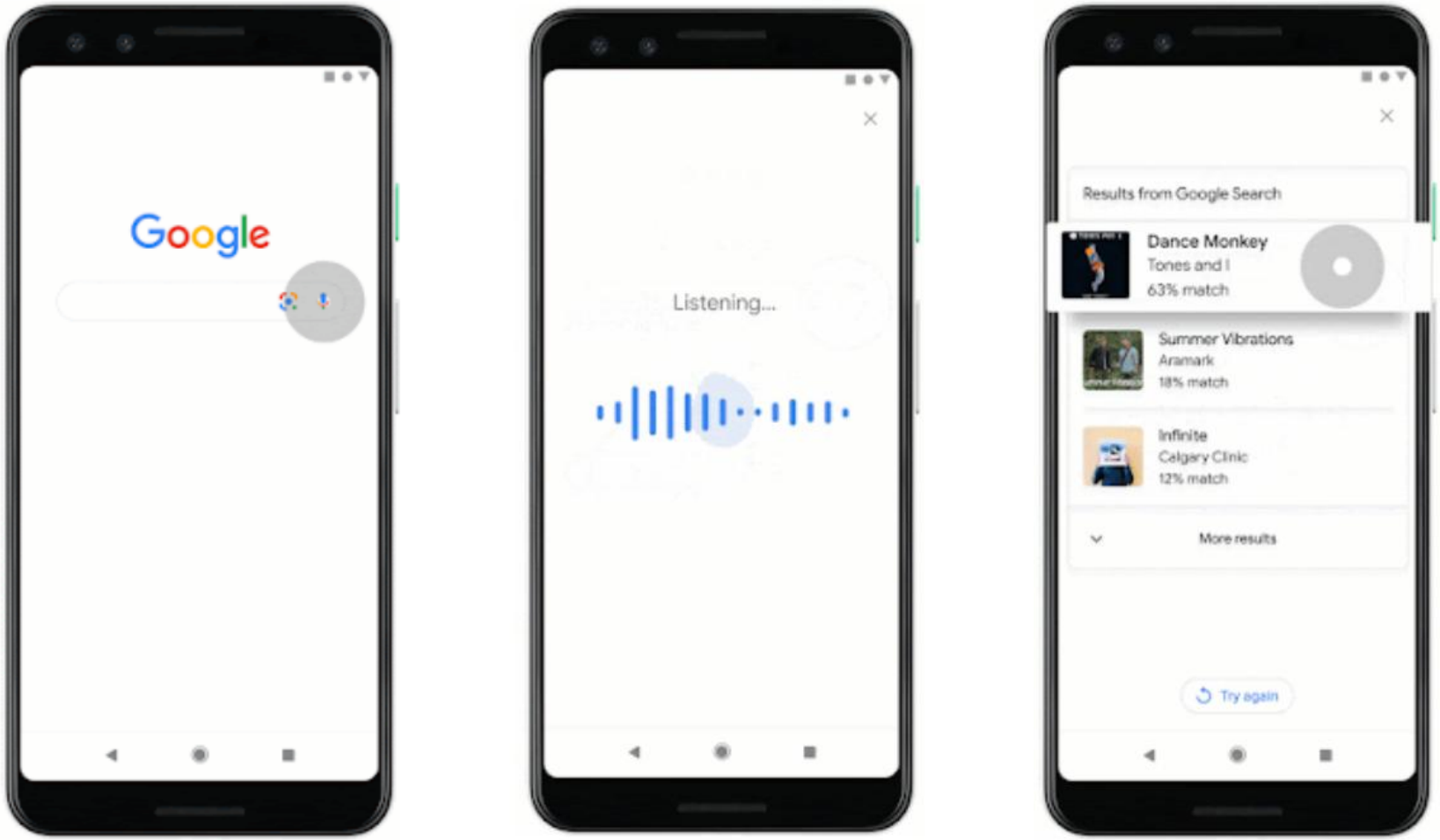
MusicFX's SynthID (Google)



labs.google/fx/tools/music-fx
deepmind.google/technologies/synthid/

Query by Humming

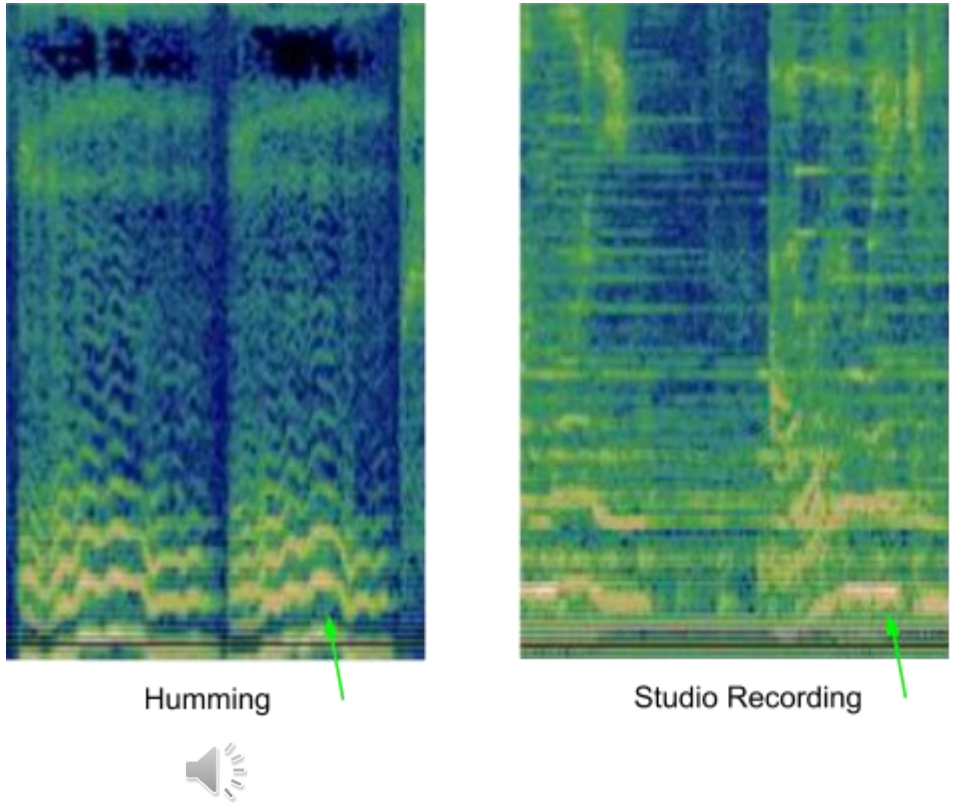
Hum to Search (Google)



(Source: Google Research Blog)

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

Hum to Search (Google)

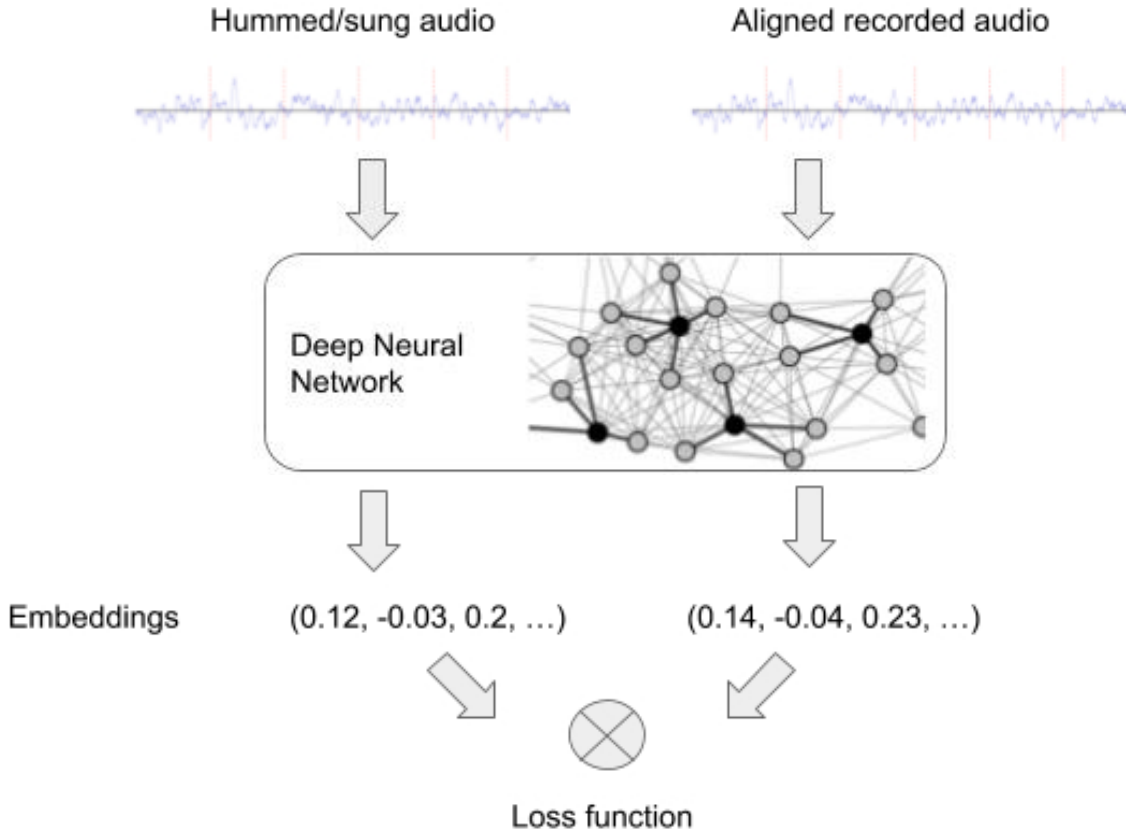


(Source: Google Research Blog)

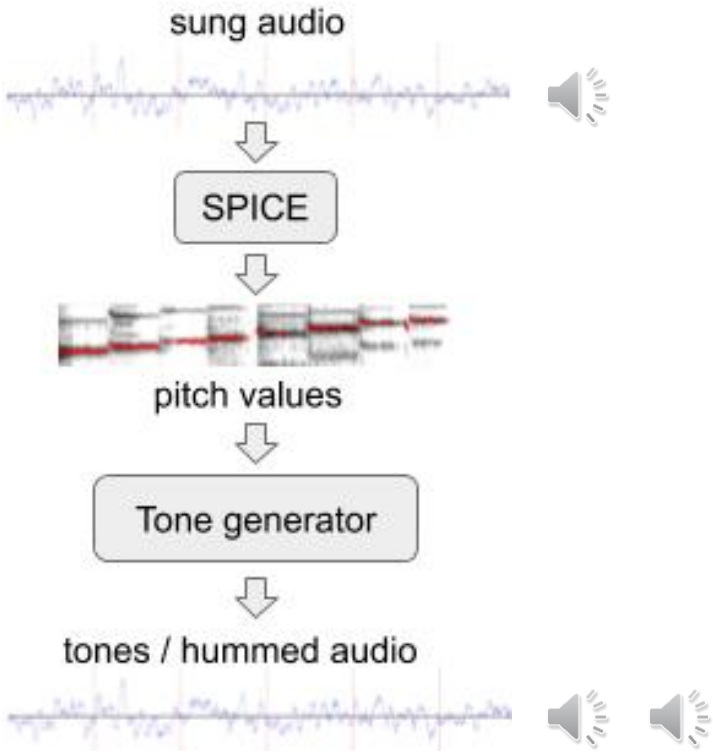
research.google/blog/the-machine-learning-behind-hum-to-search/

Hum to Search (Google)

Audio encoder



Data augmentation

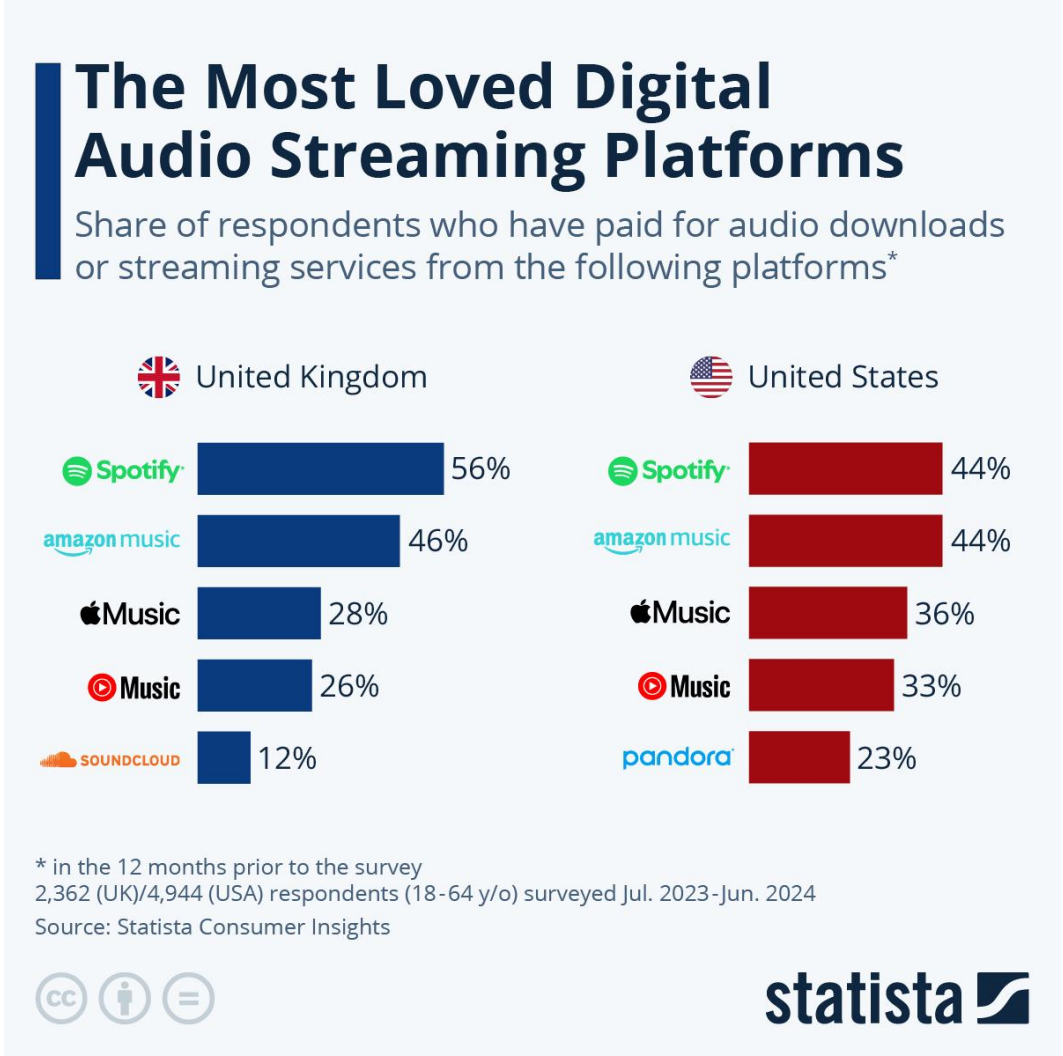


(Source: Google Research Blog)

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

Music Recommendation

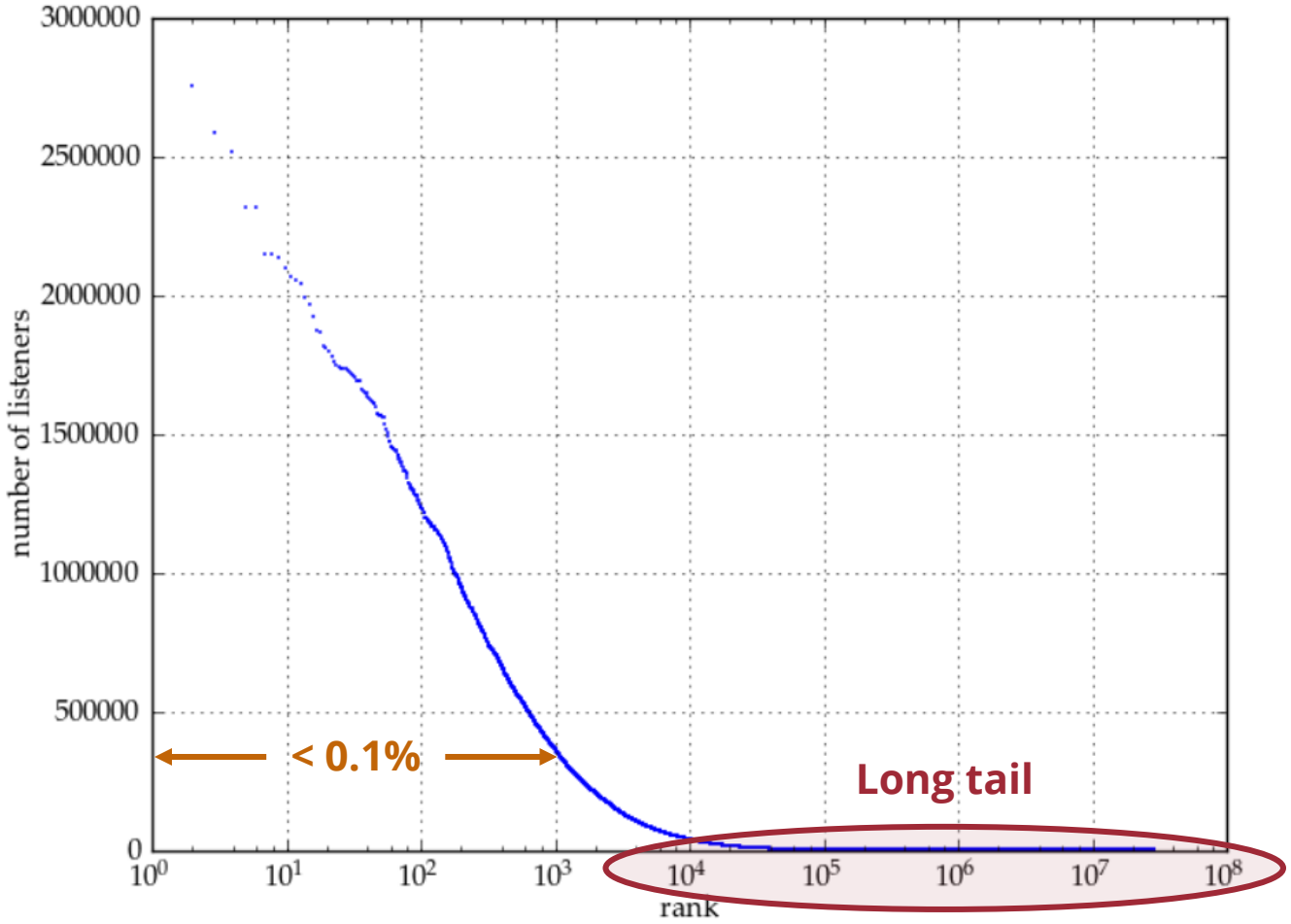
Music Recommendation



Anna Fleck, "The Most Loved Digital Audio Streaming Platforms," *Statista*, September 30, 2024.

Music Recommendation

The Long Tail Problem

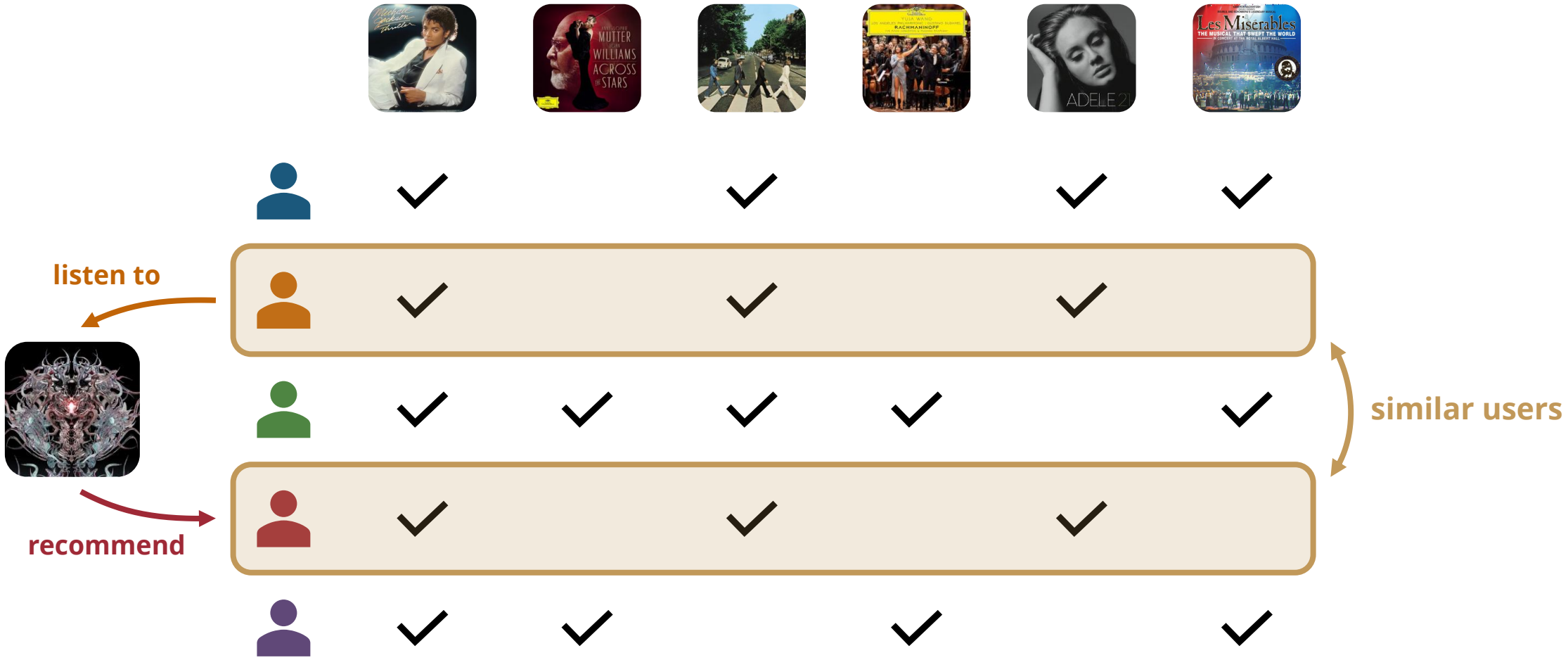


(Source: Levy & Bosteels, 2010)

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

Collaborative Filtering

User-based Collaborative Filtering



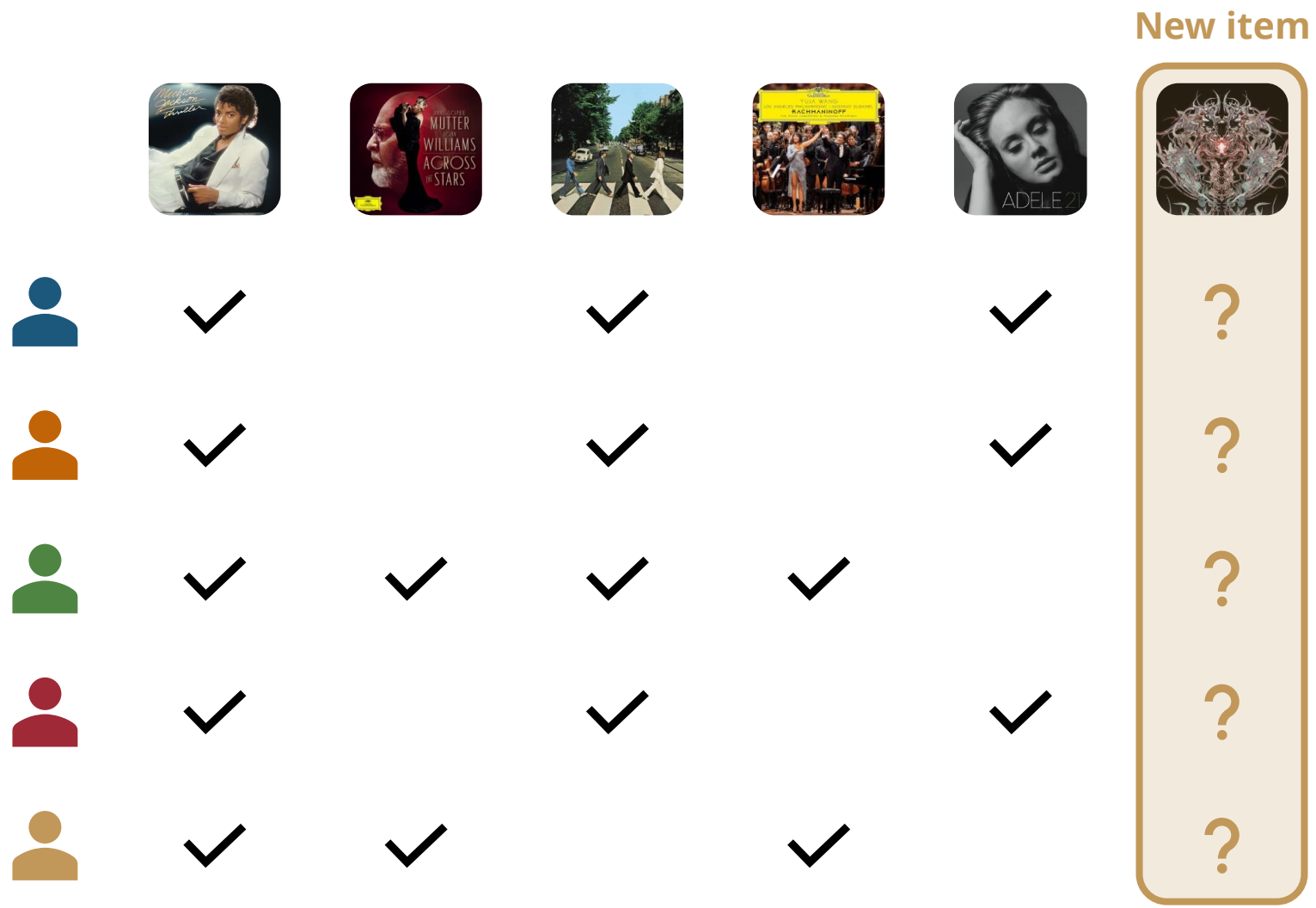
Collaborative Filtering

Item-based Collaborative Filtering

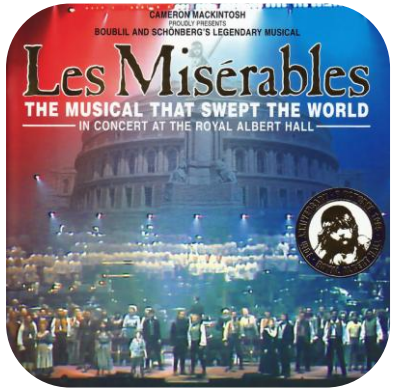
Item-based Collaborative Filtering

Item-based Collaborative Filtering

Cold Star Problem



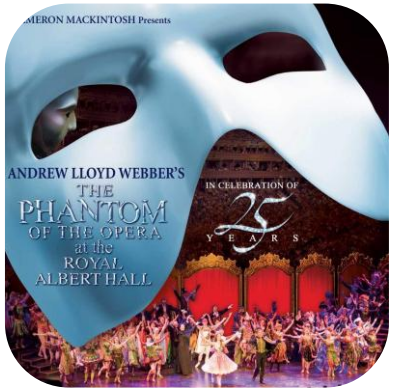
Content-based Filtering



Musical

Live concert version

Concert at Royal Albert Hall



Musical

Live concert version

Concert at Royal Albert Hall

similar songs

Cold Star Problem

User Profile Modelling

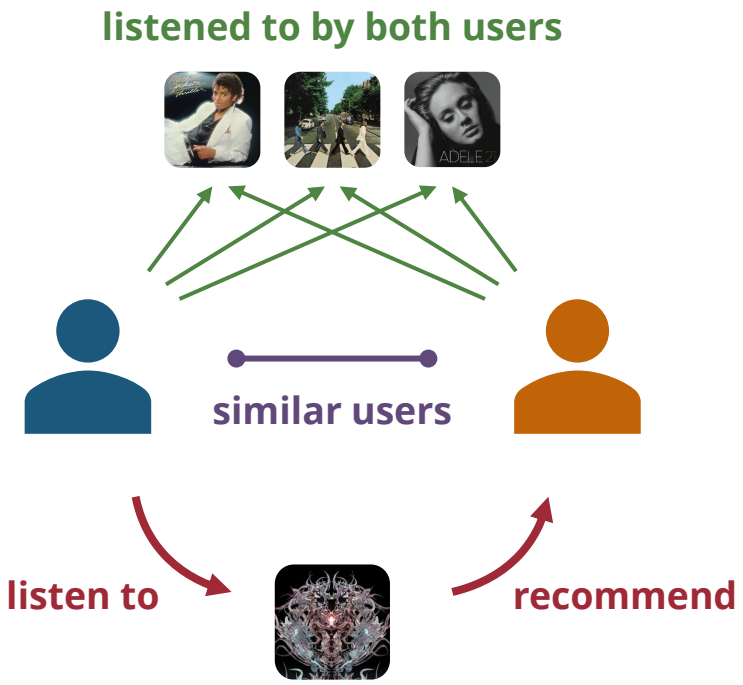
Data type	Example
Demographic	Age, marital status, gender etc.
Geographic	Location, city, country etc.
Psychographic	*Stable*: interests, lifestyle, personality etc. *Fluid*: mood, attitude, opinions etc.

(Source: Song et al., 2012)

Yading Song, Simon Dixon, and Marcus Pearce, "A Survey of Music Recommendation Systems and Future Perspectives," CMMR, 2012.

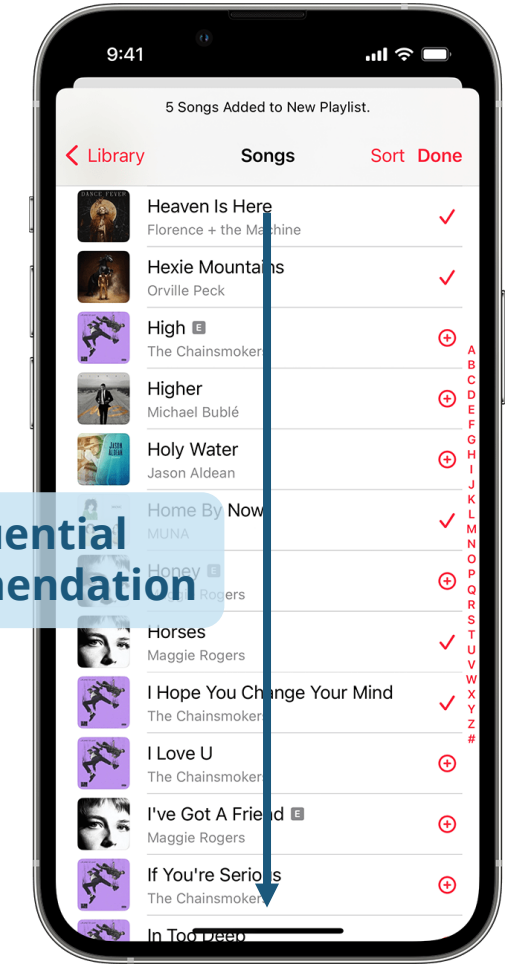
Collaborative Filtering vs Content-based Filtering

Collaborative filtering



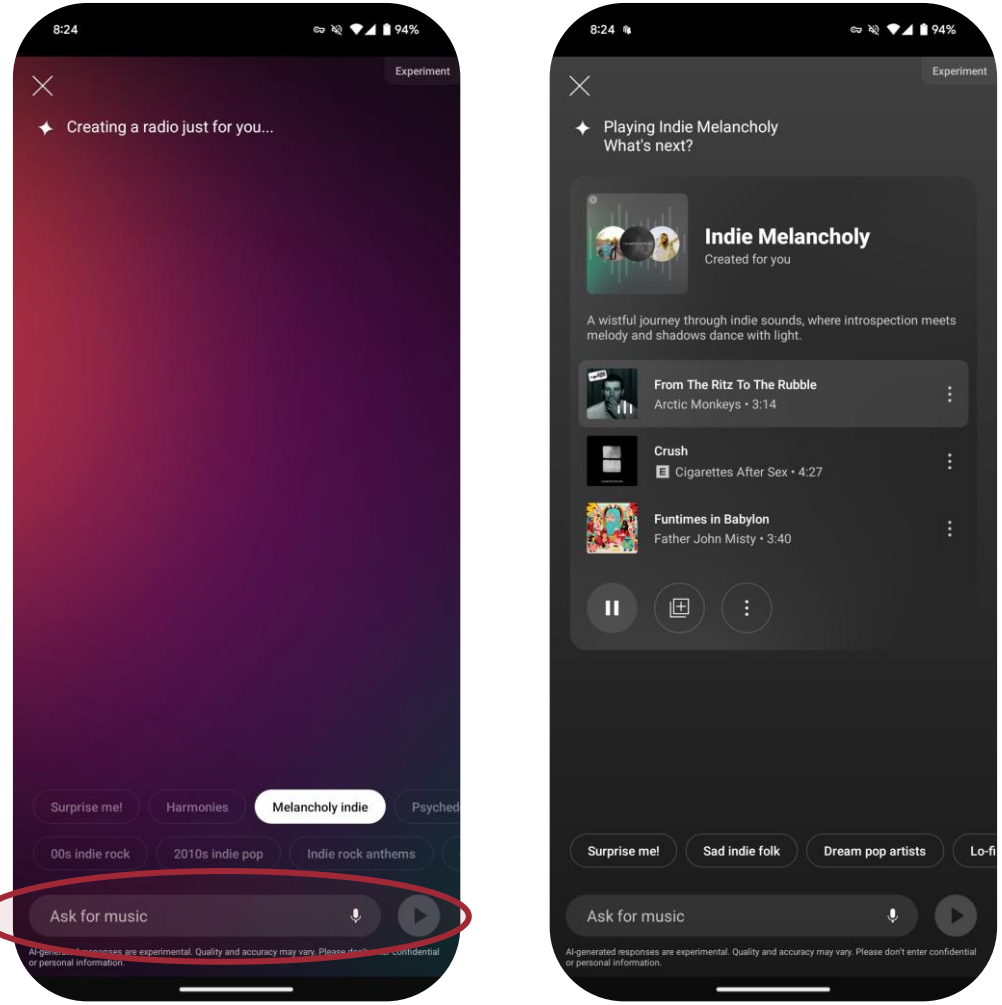
Content-based filtering

Music Playlist Generation



Sequential recommendation

Ask Music (YouTube Music)

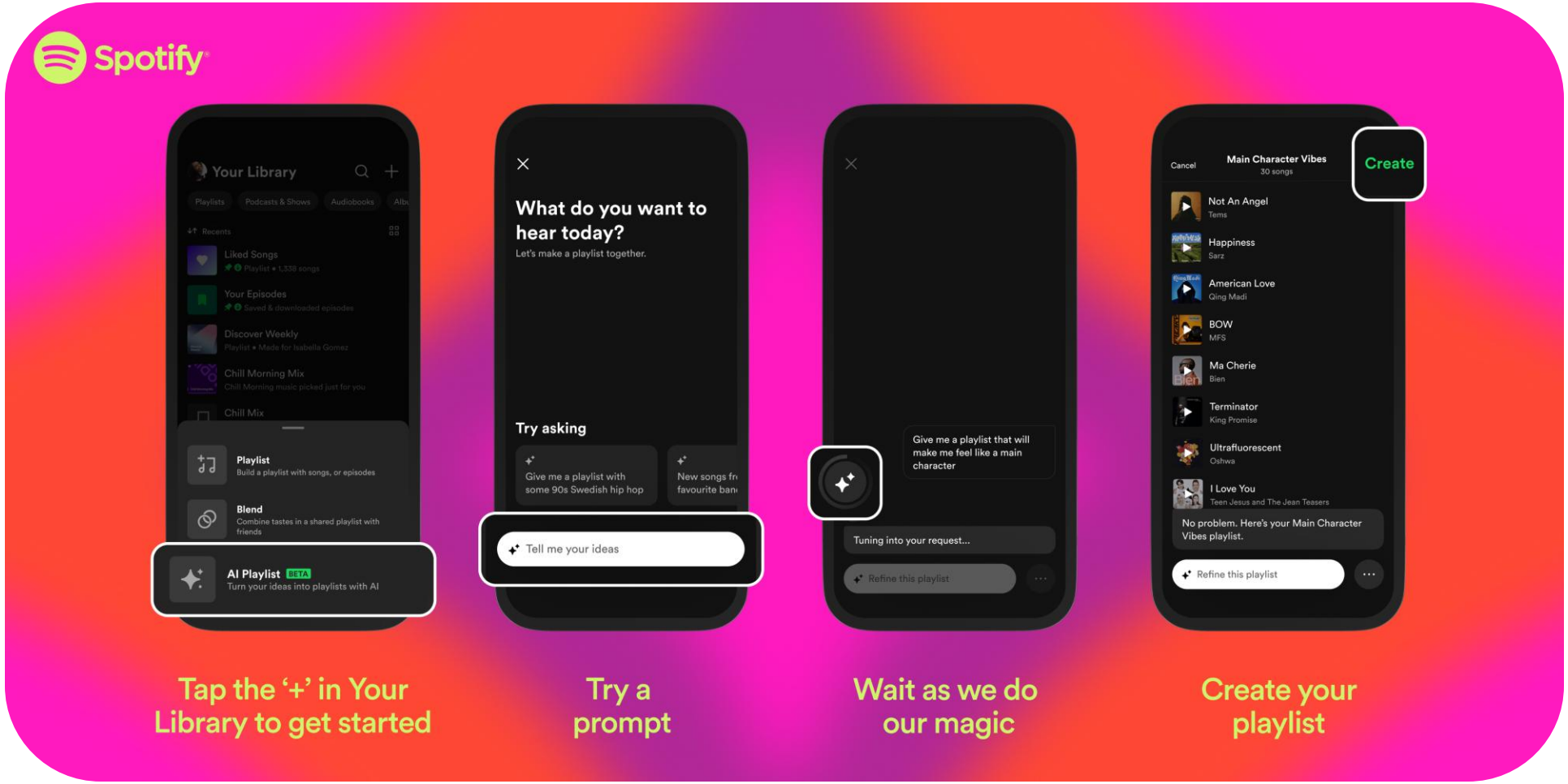


Text prompt

(Source: Android Police)

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

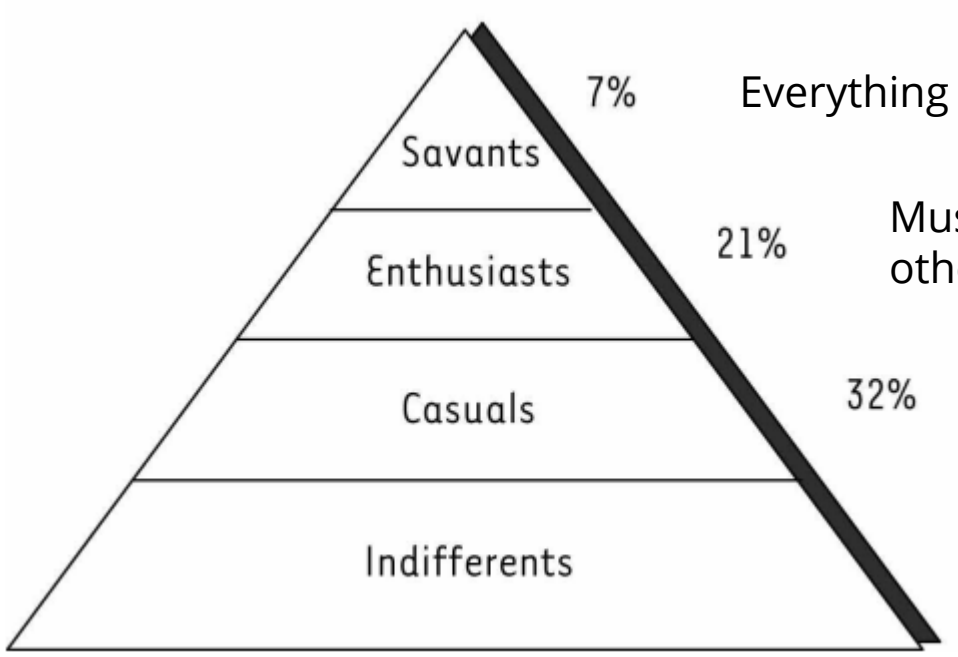
AI Playlist (Spotify)



(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/

User Listening Experience Modelling



Everything in life seems to be tied up with music

Music is a key part of life but is also balanced by other interests

Music plays a welcome role, but other things are far more important

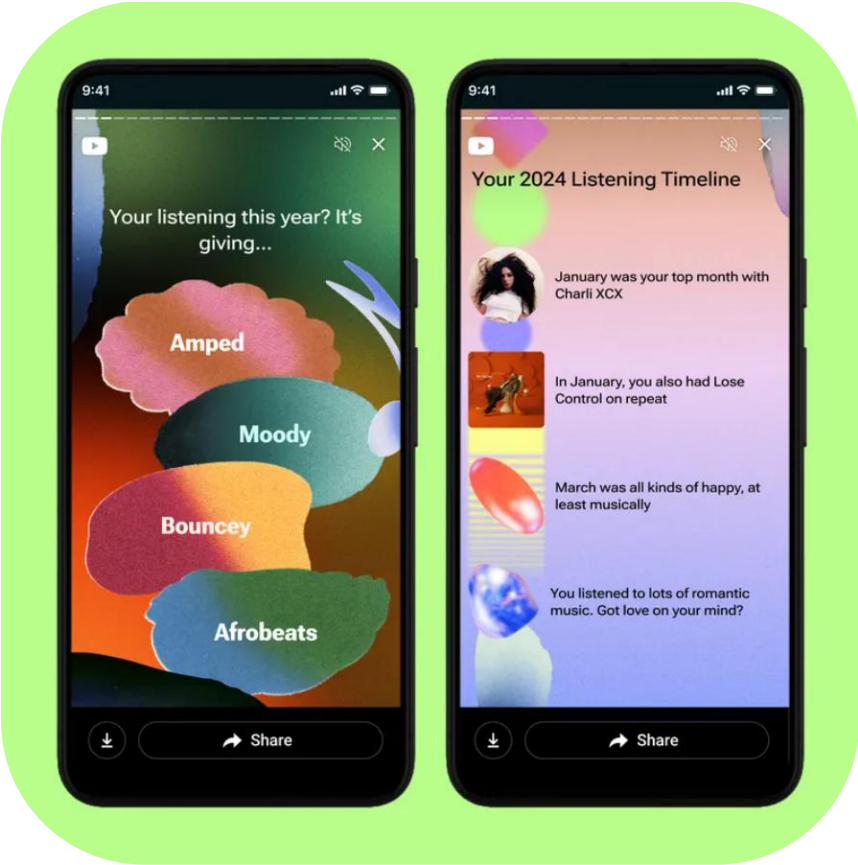
Would not lose much sleep if music ceased to exist

(Source: Jennings, 2007)

David Jennings, "Net, Blogs and Rock 'n' Roll: How Digital Discovery Works and What it Means for Consumers," Boston, MA: Nicholas Brealey Publishing, 2007.

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/