Deep Learning for Music Production

Enhanced Symbolic Music Generation with Beat Shaper

Taylor Peer
Co-Founder of Beat Shaper

Background

Current

Beat Shaper
Co-Founder & ML Engineer

Bern University of Applied Sciences
Lecturer: Machine Learning & Data Engineering

Past

Studies in Software Engineering
Master's Thesis: Comparing Neural Network Architectures for Drum Pattern Generation

Cortical.io
Director of Data Science

Simple Generative Music Application

AI-909: Intelligent Drum Machine

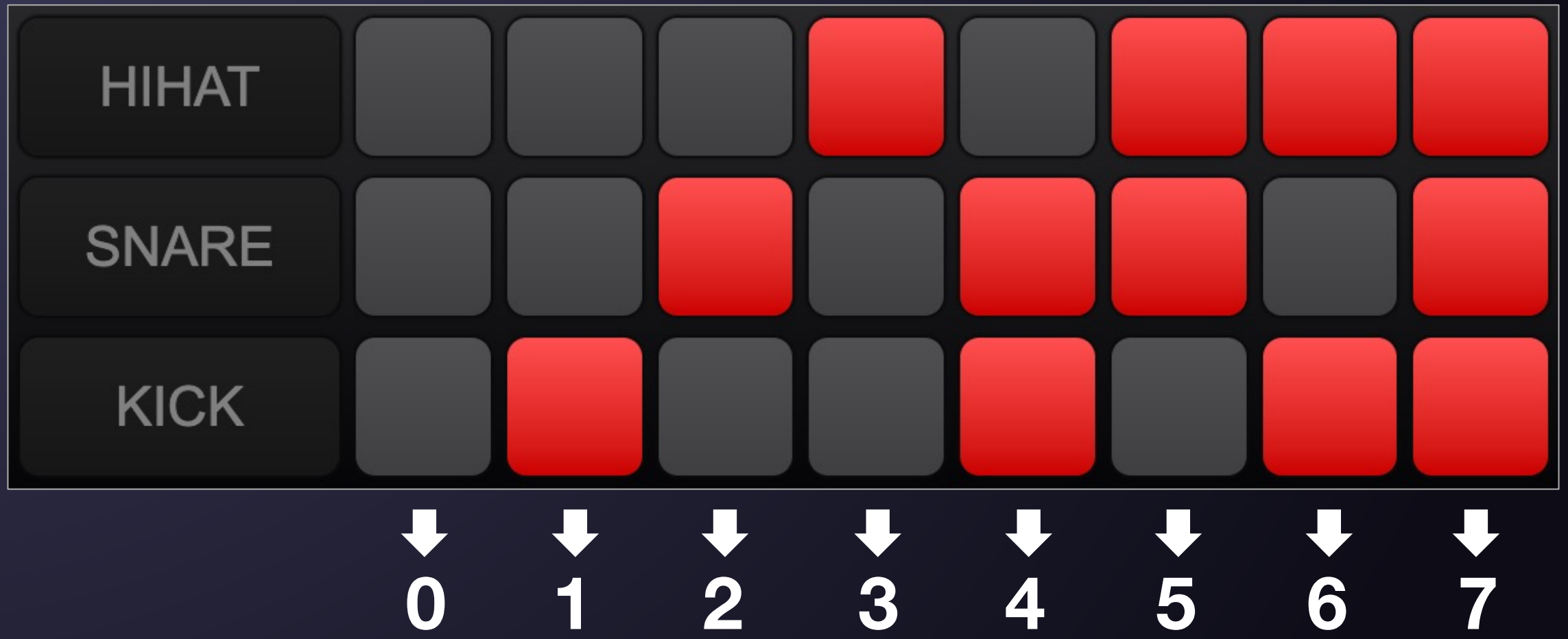
- Hackathon project:

  Numenta Challenge
  San Francisco
  November 2015
- Built in 2 days
- 16-step drum patterns
- Supports three drums

GitHub: https://github.com/TaylorPeer/AI-909

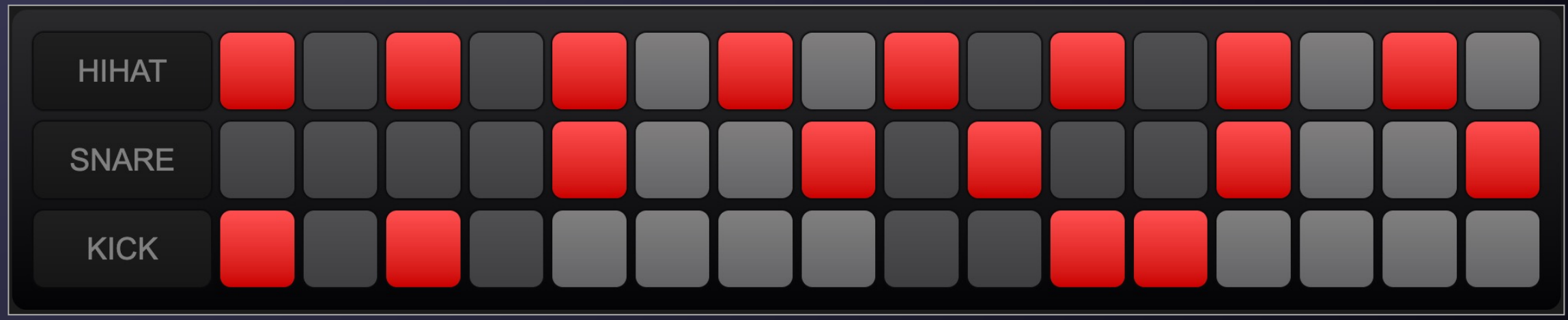
AI-909 Data Encoding

8 Possible States Represented by Integer Values



Training

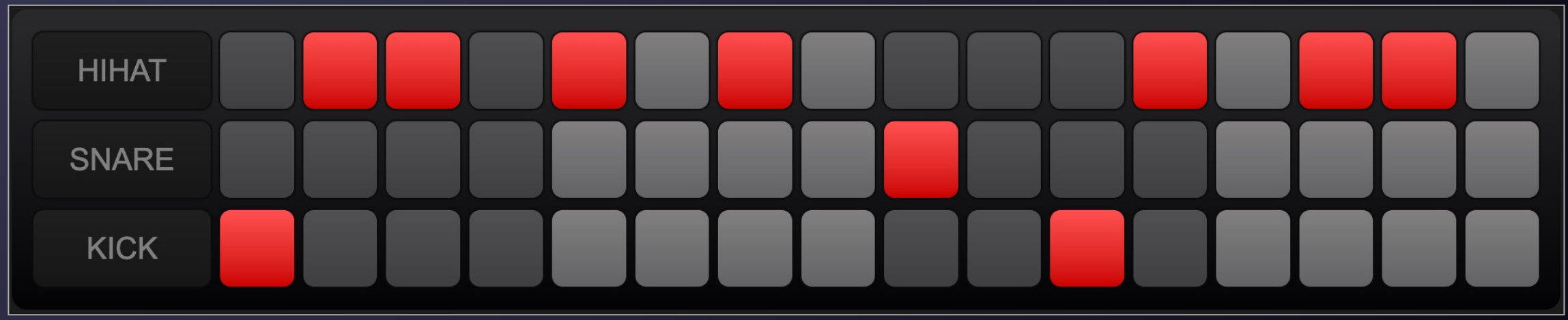
Learning Sequences of Integer Values



6 0 6 0 5 0 3 2 3 2 6 1 5 0 3 2

Training

Learning Sequences of Integer Values



6 0 6 0 5 0 3 2 3 2 6 1 5 0 3 2

→ 1 3 3 0 3 0 3 0 2 0 1 3 0 3 3 0

Training

Learning Sequences of Integer Values

	6	0	6	0	5	0	3	2	3	2	6	1	5	0	3	2
	1	3	3	0	3	0	3	0	2	0	1	3	0	3	3	0
→	6	3	6	3	5	3	6	2	3	5	6	5	5	1	3	5

Training

Learning Sequences of Integer Values

6	0	6	0	5	0	3	2	3	2	6	1	5	0	3	2
1	3	3	0	3	0	3	0	2	0	1	3	0	3	3	0
6	3	6	3	5	3	6	2	3	5	6	5	5	1	3	5

Generation

Apply Trained Model: Stepwise Autoregressive Prediction

1 0 3 0 1 1 2 0 3 0 3 0 1 1 2 0

Generative Music Workflow

Input Data

Example music sequences

Encoding

Tokenization & vectorization

Learning

Transformer, RNN, Diffusion, CNN, …

Generation

Output new sequences

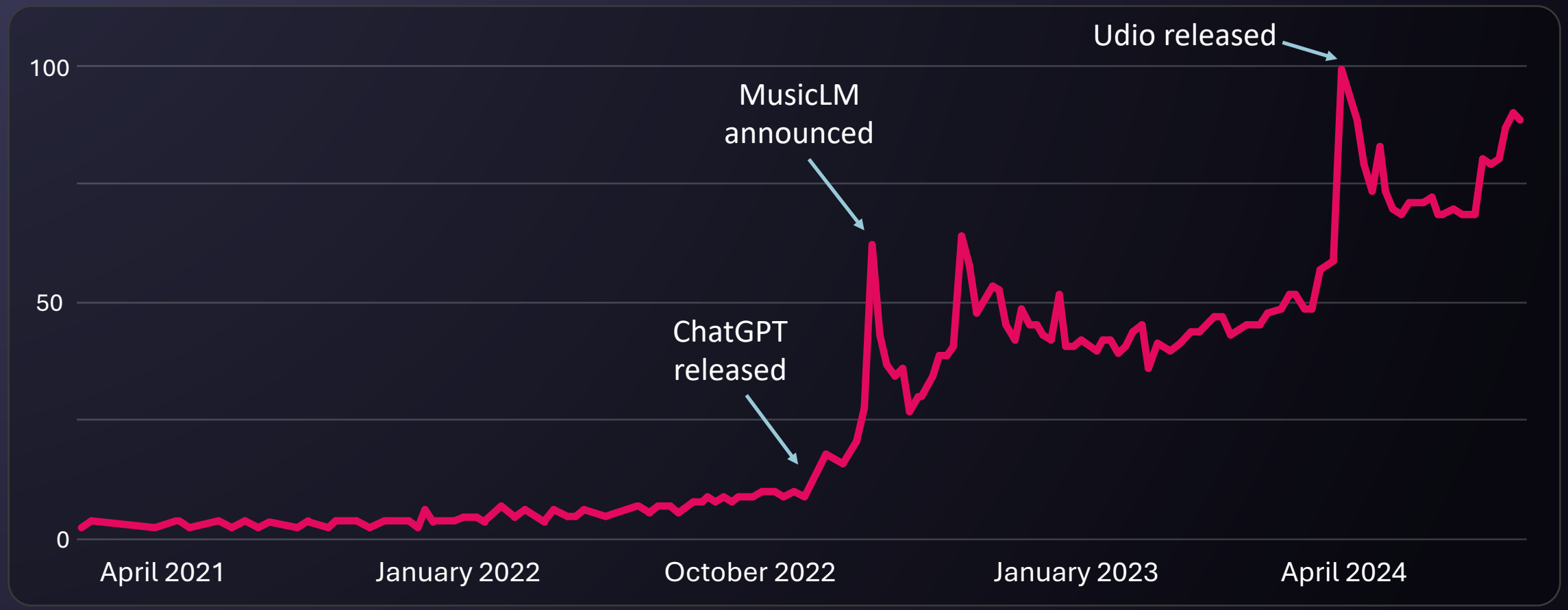
Decoding

Render back to target format

Main outstanding challenges:

- User controllability
- Size, availability, and legality of datasets
- Consistency & quality
- Long-term structure

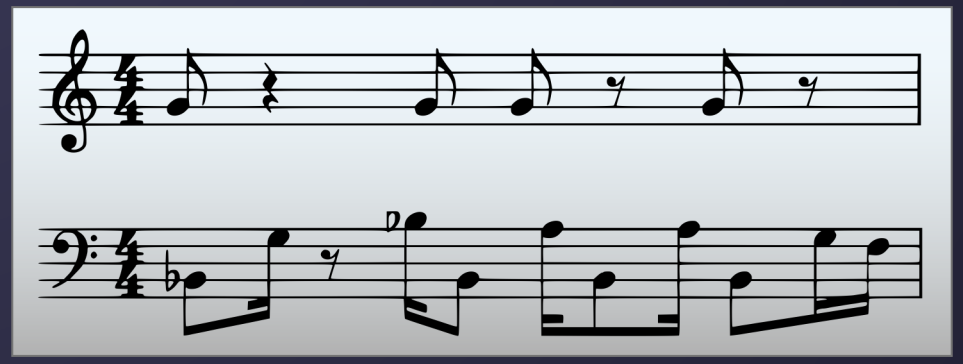
Interest in AI-Generated Music



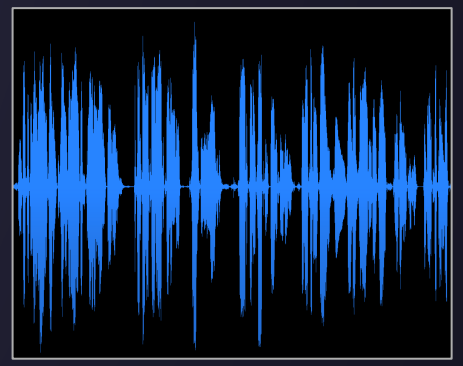
Relative Google search popularity for the query “AI music”

Source: trends.google.com/trends/explore?date=today%205-y&q=ai%20music&hl=en

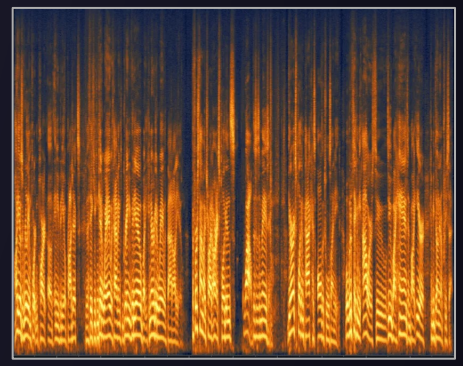
Symbolic Music vs. Audio



Music Notation



Audio Waveform



Audio Spectrogram

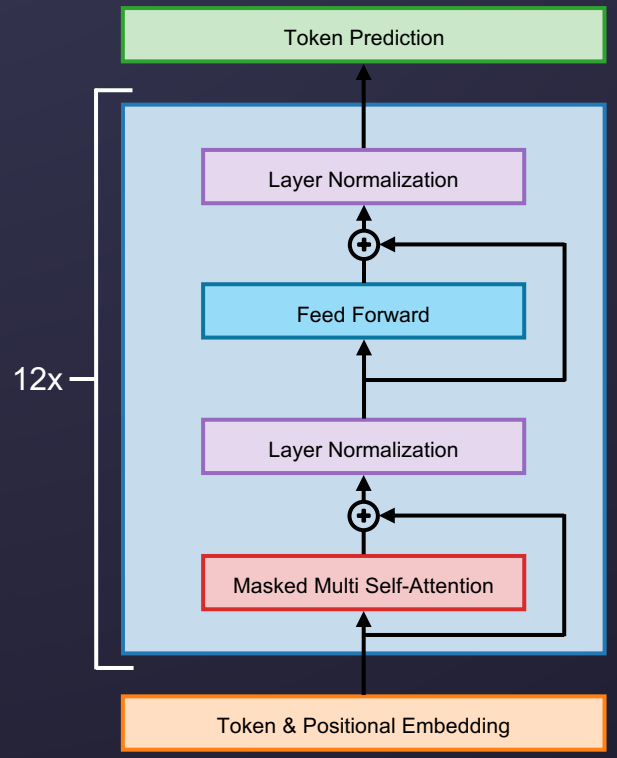
- Musical instructions
- To be played back by a human or device
- 1 minute of multi-instrument music notation is 20 kB
- Output is easily editable

- Digital audio
- Can be played back directly
- 1 minute of stereo 16-bit 44.1 kHz WAV is 10 MB (10,000 kB)
- Limited post-generation editing options

Transformers

MuseNet Architecture & Data Representation

GPT-2 Style Decoder



Plaintext Representation with Byte Pair Encoding

```
bach piano_strings start tempo90 piano:v72:G1 piano:v72:G2
piano:v72:B4 piano:v72:D4 violin:v80:G4 piano:v72:G4
piano:v72:B5 piano:v72:D5 wait:12 piano:v0:B5 wait:5
piano:v72:D5 wait:12 piano:v0:D5 wait:4 piano:v0:G1
piano:v0:G2 piano:v0:B4 piano:v0:D4 violin:v0:G4 piano:v0:G4
wait:1 piano:v72:G5 wait:12 piano:v0:G5 ...
```

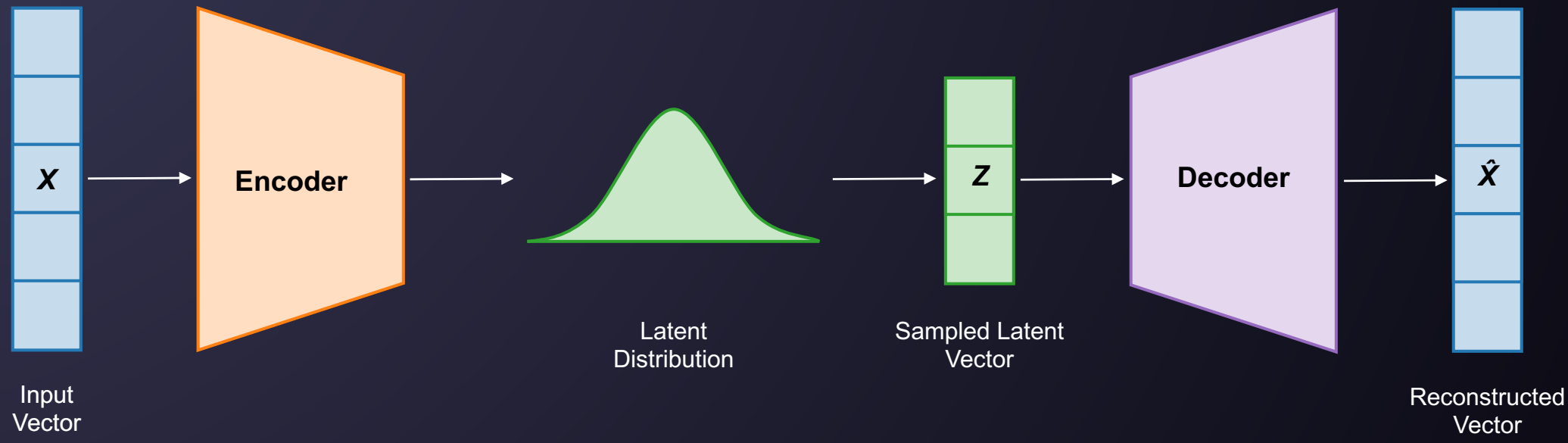
- Conditioning tokens: `bach`, `piano_strings`
- Metadata: `start`, `tempo90`
- Note & performance tokens: `piano:v72:G1`, `violin:v80:G4`

`instrument:volume:note+octave`

OpenAI (2019). *MuseNet: A generative model for music*.

Variational Autoencoders (VAEs)

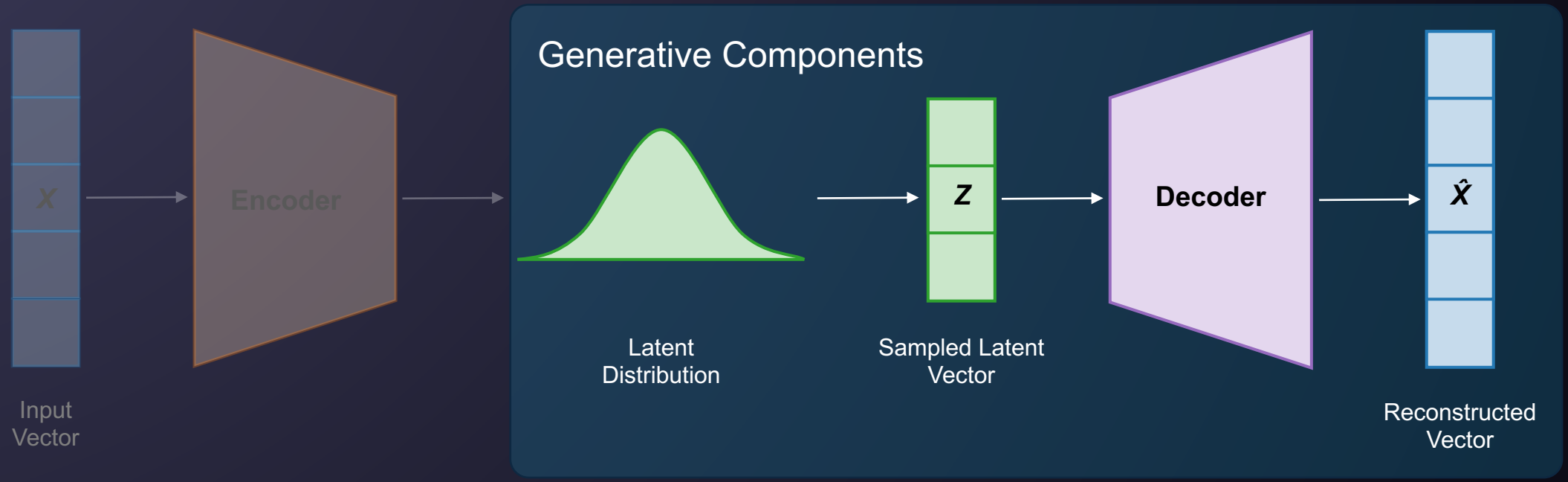
Generation Based on Probabilistic Sampling



1. Network is first trained to reconstruct its input

Variational Autoencoders (VAEs)

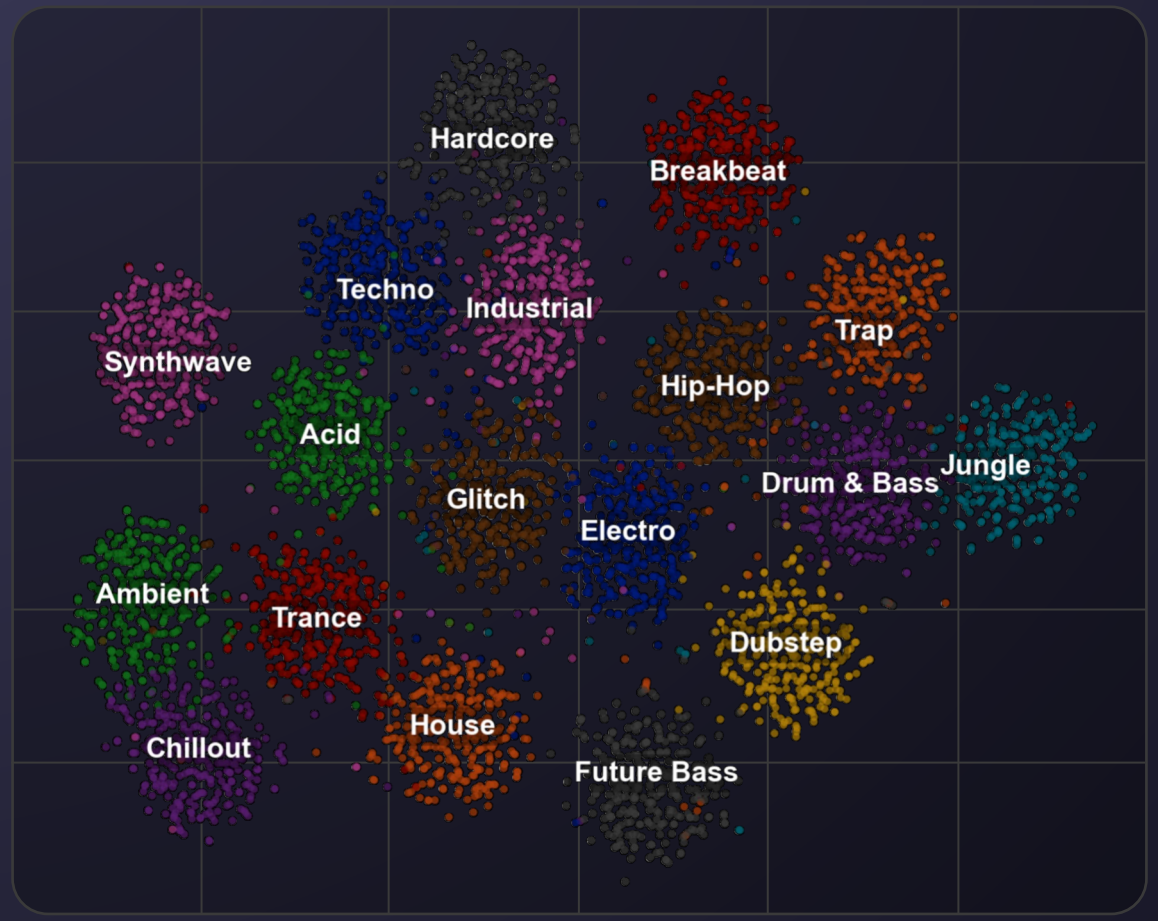
Generation Based on Probabilistic Sampling



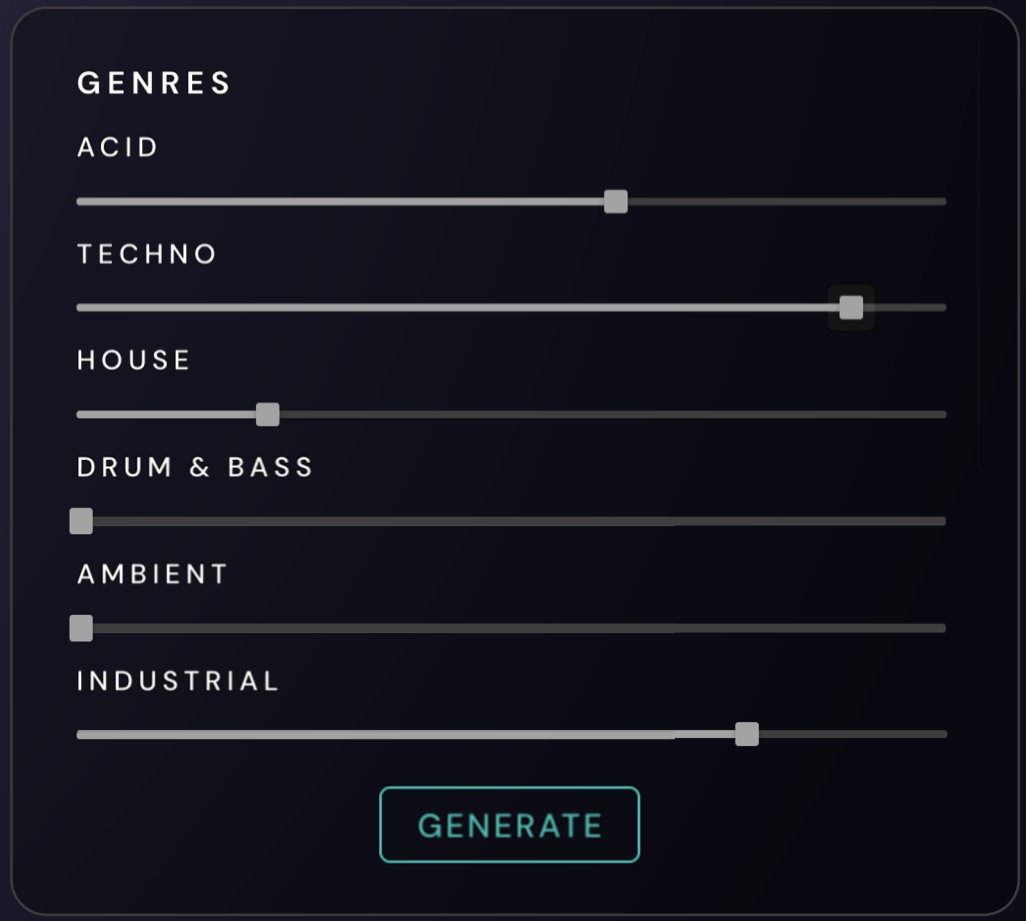
1. Network is first trained to reconstruct its input
2. After training, sampled points from the learned latent space can be decoded to generate new outputs

Variational Autoencoders (VAEs)

Controllable Generation via Attribute Vectors



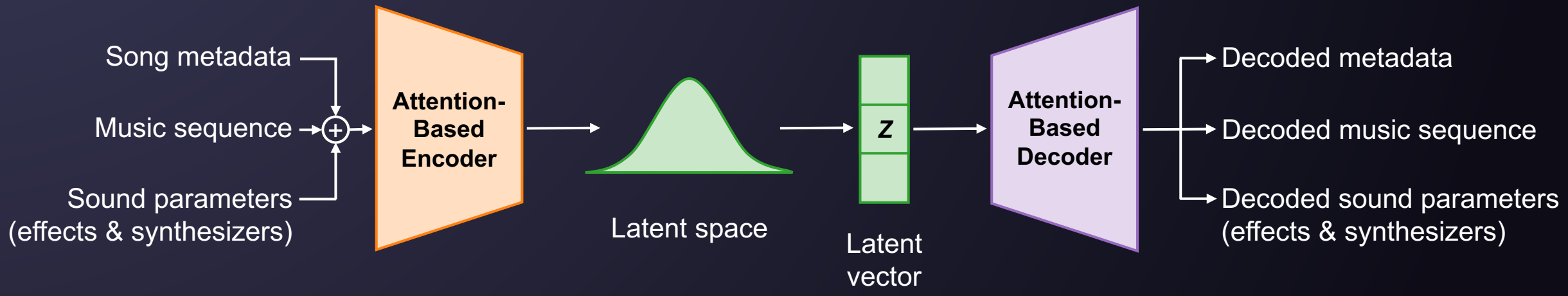
2D T-SNE Visualization of the VAE's Latent Encodings



User Control over VAE Decoding

Beat Shaper

Variational Transformer Architecture



Evaluation

Benchmarks in Related Fields

Language Modeling on Penn Treebank: Best Model Perplexity Scores over Time

Source: paperswithcode.com/sota/language-modelling-on-penn-treebank-word

Evaluation

Measuring Performance of Generative Music Models

Model-specific metrics, e.g.:

- Loss function values

Novelty measurements, e.g.:

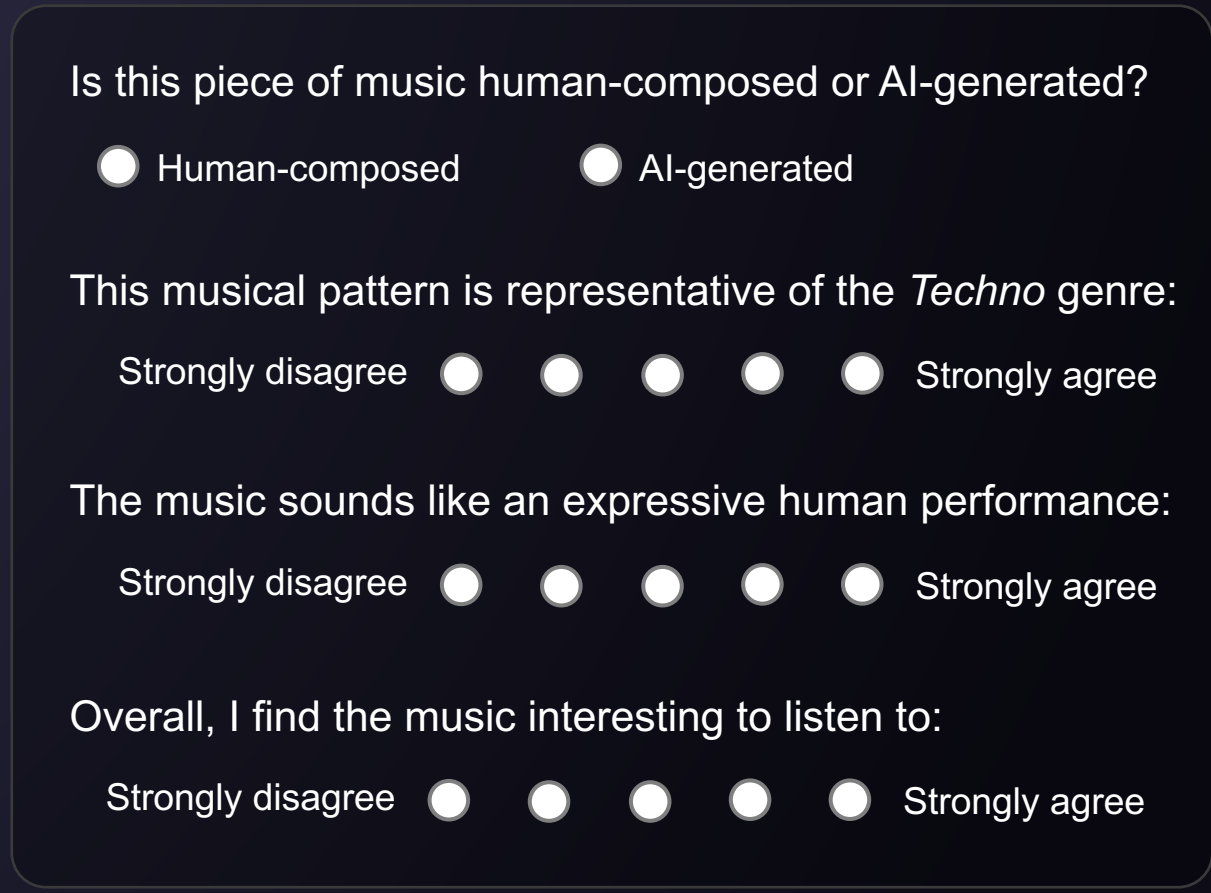
- Similarity to the training data

Musical properties, e.g.:

- Distribution of generated note pitches

Ratings by Human Evaluators, e.g.:

- Preference tests
- Turing tests
- Likert ratings



Example Subjective Listening Survey

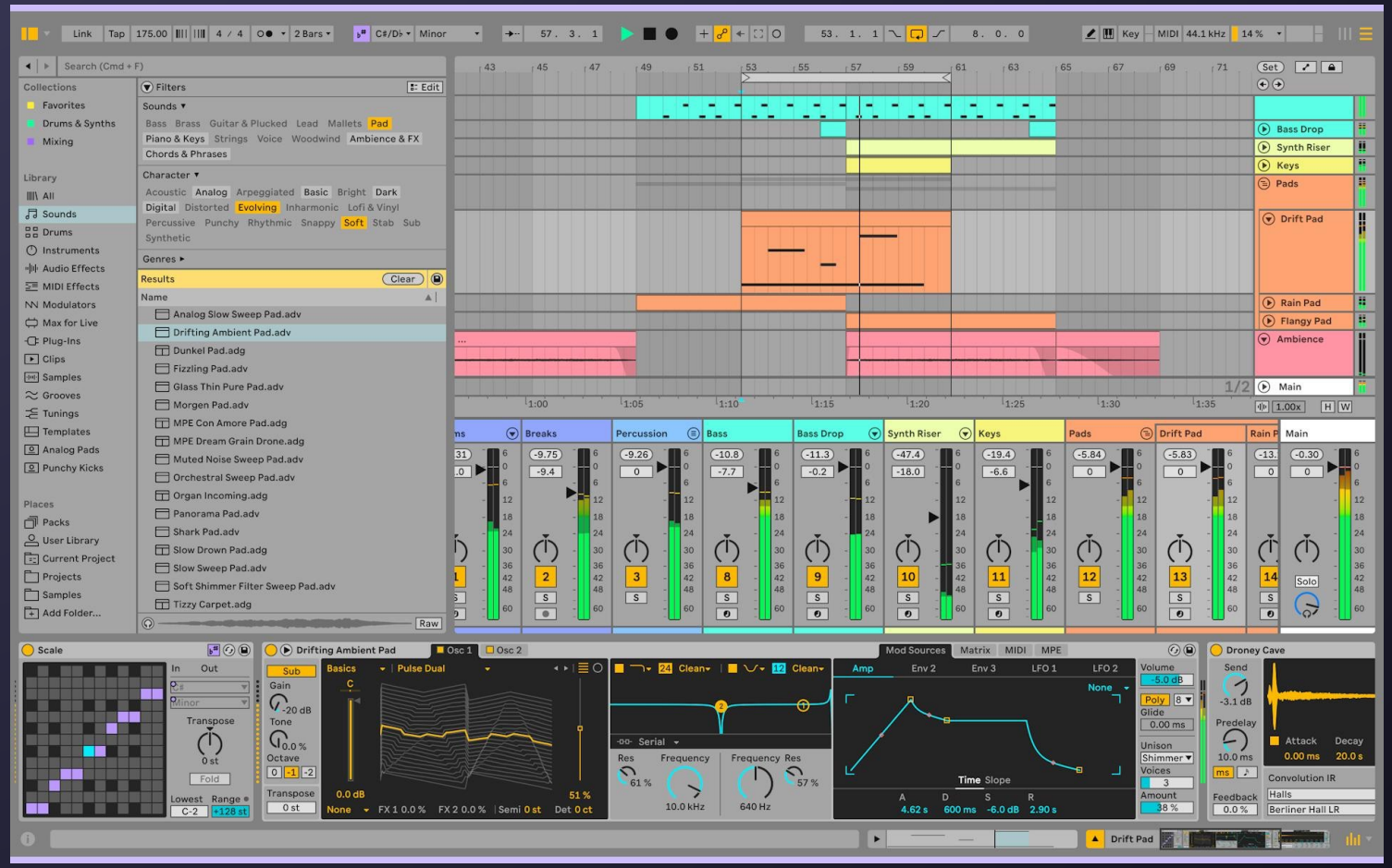
Evaluation

Methods from Selected Publications

Model	Metrics	Listening Tests	Dataset
Museformer	Perplexity, similarity error	10-point Likert scale (10 participants)	LMD
Symbolic Diffusion	Self-similarity, FD, MMD	None	LMD
MusicBERT	Downstream tasks	None	MMD
Compound Word Transformer	Longest common subsequence	5-point Likert scale (18 participants)	Pop Piano
Pop Music Transformer	Beat and downbeat STD	Preference test (76 participants)	Pop Piano
Jazz Transformer	Pitch class entropy	5-point Likert scale (59 participants)	WJazzD
Music Transformer	Negative log-likelihood	Preference test (unknown participant count)	JSB
LakhNES	Perplexity	Turing & preference tests (unknown participants)	LMD
GrooVAE	Timing MAE & MSE	Turing & preference tests (unknown participants)	GMD
MusicVAE	Reconstruction accuracy	Preference test (unknown participant count)	Web MIDI

Real-World Application

Simplifying Music Production with Deep Learning



Ableton Live Digital Audio Workstation

BEAT SHAPER
augmenting human creativity

SIGN OUT

Generate new sequences with natural language

GENERATE

SEQUENCE SETTINGS

TEMPO

COMPLEXITY

VIBE

DARK

UPBEAT

AMBIENT

ENERGETIC

GENRES

ACID

TECHNO

HOUSE

Note	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
G																																
F#																																
F																																
E																																
D#																																
D																																
C#																																
C																																
B																																
A#																																
A																																

OSC A

DETUNE WIDTH FMOD AMOD

NOISE A

Type

White

Pink

Brown

LEVEL

SUB OSC

LEVEL

LFO

DEPTH RATE

FILTER

Type

Low Pass

High Pass

Notch

RES FREQ Q

ADSR

ATTACK DECAY SUSTAIN RELEASE

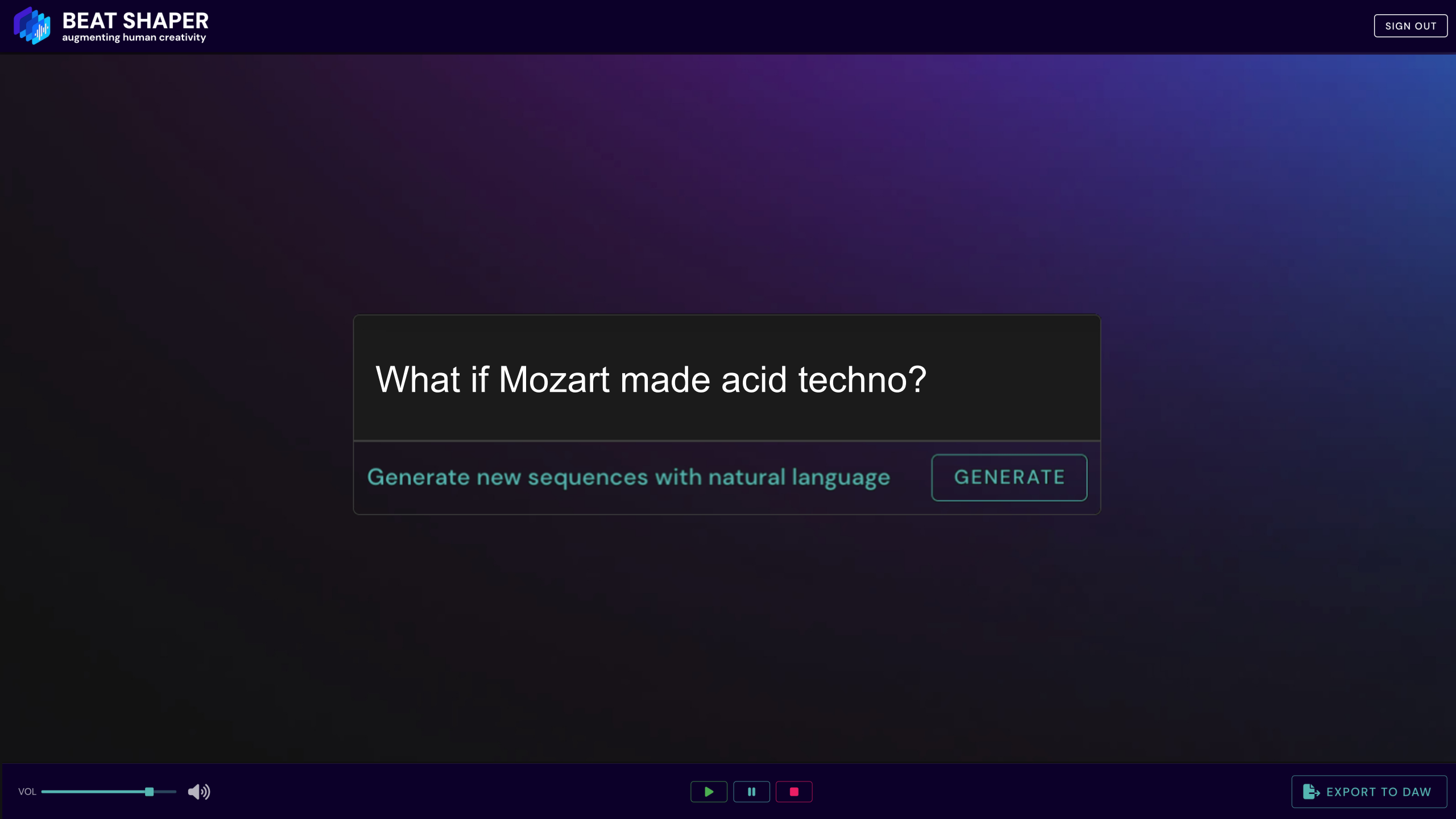
MIXER

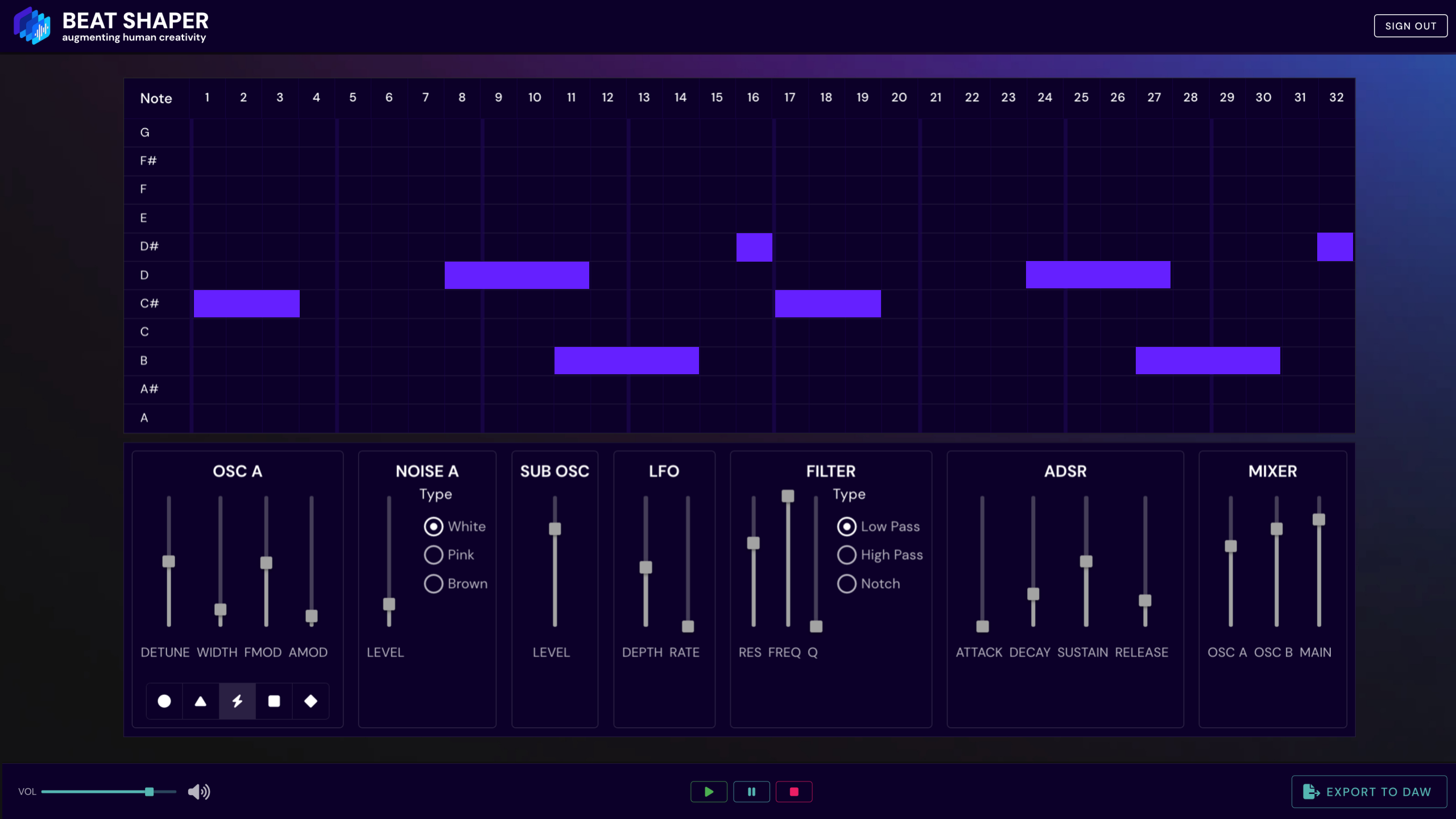
OSC A OSC B MAIN

Note	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
High Hat																																
Open High Hat																																
Snare																																
Kick																																

VOL

EXPORT TO DAW


BEAT SHAPER
augmenting human creativity
SIGN OUT
What if Mozart made acid techno?
Generate new sequences with natural language
GENERATE
VOL
EXPORT TO DAW


BEAT SHAPER
augmenting human creativity
SIGN OUT
Note
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
G
F#
F
E
D#
D
C#
C
B
A#
A
OSC A
DETUNE WIDTH FMOD AMOD
NOISE A
Type
White
Pink
Brown
LEVEL
SUB OSC
LEVEL
LFO
DEPTH RATE
FILTER
Type
Low Pass
High Pass
Notch
RES FREQ Q
ADSR
ATTACK DECAY SUSTAIN RELEASE
MIXER
OSC A OSC B MAIN
VOL
EXPORT TO DAW

SEQUENCE SETTINGS

TEMPO

COMPLEXITY

VIBE

DARK

UPBEAT

AMBIENT

ENERGETIC

GENRES

ACID

TECHNO

HOUSE

BEAT SHAPER
augmenting human creativity

SIGN OUT

What if Mozart made acid techno?

Generate new sequences with natural language

GENERATE

SEQUENCE SETTINGS

TEMPO

COMPLEXITY

VIBE

DARK

UPBEAT

AMBIENT

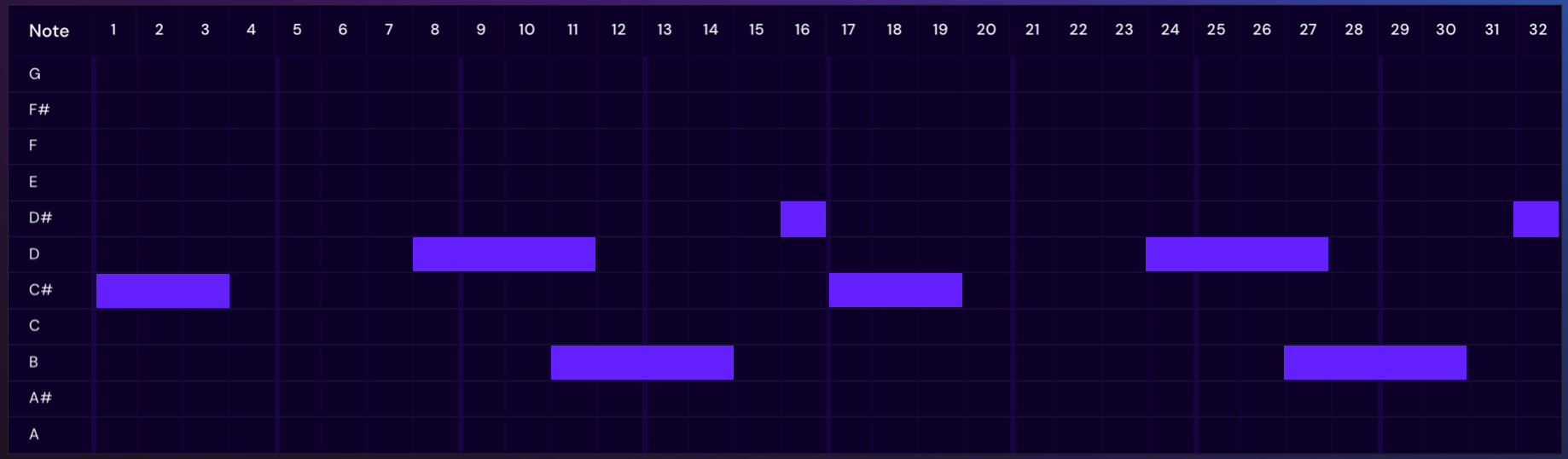
ENERGETIC

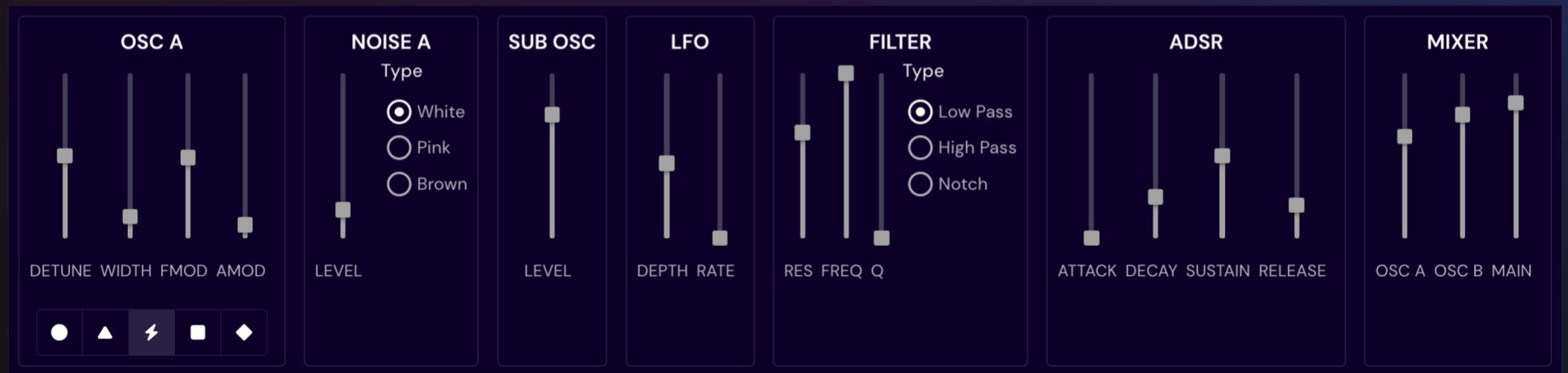
GENRES

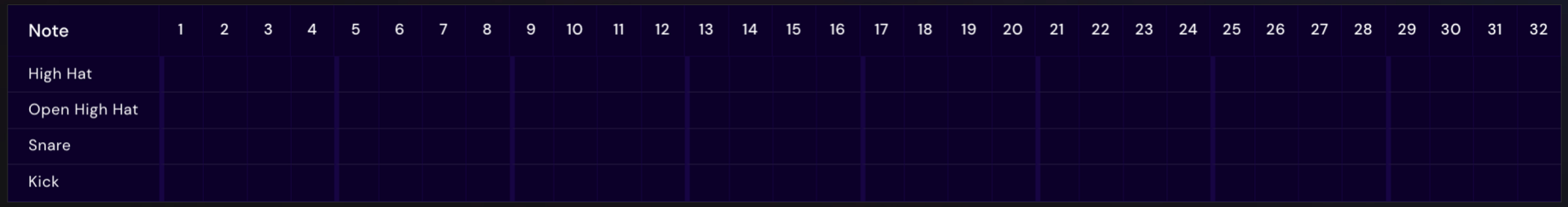
ACID

TECHNO

HOUSE







VOL

EXPORT TO DAW

BEAT SHAPER

beatshaper.ai

taylor@beatshaper.ai
Vienna, Austria