Music Information Retrieval with Python

PyStok, 29.03.2023

Mateusz Modrzejewski @mamodrzejewski

Hi! I’m Mateusz 👋


POLITECHNIKA
1899
1901


www.fotobykaras.pl

LIVE IN CHINA 2019

MM3

Bartek Łuczkiewicz - bass
巴尔托什•乌齐凯维奇 - 贝斯

Michał Milczarek - guitar
米哈乌•米勒查莱克 - 吉他

Mateusz Modrzejewski - drums
马特乌什•莫德热耶夫斯基 - 鼓

28.03 Guangzhou, European Jazz Festival, Xinghai Concert Hall
29.03 Guangzhou, European Jazz Festival, JZ Club
30.03 Zhuhai, European Jazz Festival, Golden Jazz Academy
31.03 Guangzhou, European Jazz Festival, Music Factory
01.04 Zhuhai, Golden Jazz Academy
03.04 Zhongshan, Friends Bar
04.04 Shenzhen, Lavo
06.04 Guiyang, Starlight Theatre
07.04 Zunyi, Zunyi Grand Theatre
08.04 Zunyi, Music Academy
10.04 Chongqing, Nuts Live House
13.04 Chengdu, He Duoling Art Galery

PLATEAUX
COMBINING MIND, BODY & SOUL

波兰驻华大使馆 文化处

ADAM MICKIEWICZ INSTITUTE

波兰共和国 驻成都总领事馆

波兰共和国 驻广州总领事馆

19.10 Shenzhen, Oct-Loft Festival
2017年10月19日 深圳 OCT-LOFT国际爵士音乐节

20.10 Guangzhou, tba
2017年10月20日 广州 即将公布

21.10 Zhongshan, tba
2017年10月21日 中山 即将公布

22.10 Zhuhai, tba
2017年10月21日 珠海 即将公布

23.10 Guangzhou, T:Union
2017年10月23日 广州 TU凸空间

25.10 Chengdu, University of Electronics Science and Technology of China
2017年10月25日 成都 电子科技大学

25.10 Chengdu, University of Electronics Science and Technology of China
2017年10月26日 成都 电子科技大学

27.10 Chengdu, Yanjiyou Bookstore
2017年10月27日 成都 言几又书店

28.10 Chengdu, White Night Bar
2017年10月28日 成都 白夜

JAZZ ON TOUR NA PO LSKU

Live! in China 2017
september / october
爵色波兰2017中国巡演

Michał Milczarek 米哈乌・米勒查莱克 - guitar 吉他
Bartosz Łuczkiewicz 巴尔托什・乌齐凯维奇 - bass 贝斯
Mateusz Modrzejewski 马特乌什・莫德热耶夫斯基 — drums 鼓

Michał MILCZAREK Trio

米哈乌・米勒查莱克三重奏

Jazz Phonographic Debut of the Year in Poland 2014
2014年波兰“爵士留声机首秀”金奖

CHINA-CEEC CULTURAL SEASON

CJA 中国音协爵士乐学会 CHINESE MUSICIANS ASSOCIATION JAZZ ASSOCIATION

波兰共和国 驻广州总领事馆

ADAM MICKIEWICZ INSTITUTE

波兰驻华大使馆 文化处

波兰共和国 驻成都总领事馆


TYSKIE
by Karas
www.fotobykaras.pl


THE
DUMPLINGS
TRASA PRZYKRO MI
8.11 OBORNIKI
9.11 WAŁBRZYCH
10.11 WROCŁAW
14.11 OSTRÓW WLKP.
15.11 ŁÓDŹ
16.11 KATOWICE SOLD OUT
22.11 MYSŁOWICE SOLD OUT
23.11 KRAKÓW SOLD OUT
24.11 CHEŁM
7.12 LUBLIN SOLD OUT
8.12 BIELSKO-BIAŁA SOLD OUT
12.12 WROCŁAW SOLD OUT
13.12 ZIELONA GÓRA SOLD OUT
14.12 POZNAŃ SOLD OUT
15.12 SZCZECIN SOLD OUT
18.12 POZNAŃ SOLD OUT
19.12 BYDGOSZCZ SOLD OUT
20.12 WARSZAWA SOLD OUT
21.12 BIAŁYSTOK SOLD OUT
22.12 GDAŃSK SOLD OUT
27.12 DĄBROWA GÓRNICZA
28.12 WOLBROM
30.12 WARSZAWA SOLD OUT
DO ZOBACZENIA PO PRZERWIE


PATROL
LECH

MIR

Music Information Retrieval

Music Information Retrieval

Processing music in useful ways

Music Information Retrieval

Processing music in useful ways

Generating new musical content

Music Information Retrieval

Processing music in useful ways

Generating new musical content

Enhancing our perception of music

Music

Music

Genre

Music

Genre

Mood

Music

Genre

Mood

Instrumentation

Music

Genre

Mood

Instrumentation

Timbre

Music

Genre

Mood

Instrumentation

Timbre

Structure

Music

Genre

Mood

Instrumentation

Timbre

Structure

Harmony

Music

Genre

Mood

Instrumentation

Timbre

Structure

Harmony

Rhythm

Music

Genre

Mood

Instrumentation

Timbre

Structure

Harmony

Rhythm

Deep, complex form of expression and communication

Interacting with music


OUTPUT EXP.PEDAL INPUT
TREBLE POWER BASS
FREQUENCY BANDPASS GAIN Q
Tube Q
electro-harmonix
MADE IN USA
AMPLIFIER INPUT
VOLUME TONE SUSTAIN
BIG MUFF
electro-harmonix
MADE IN NYC, USA
VOX
OUTPUT SENSITIVITY
MXR
dyna comp
MODE
DISTORTION FILTER VOLUME
RAT
DRIVE TONE LEVEL
Maxon
OD-9
Overdrive
DEPTH SHAPE RATE
Stereo
Pulsar
VARIABLE SHAPE TREMOLO
electro-harmonix
MADE IN NYC, USA
OUTPUT INPUT
LPF MODES
STATUS
DRY OUTPUT
SUB OCTAVE
+1 OCTAVE
+1 OCTAVE DETUNED
+2 OCTAVES
+2 OCTAVES DETUNED
LP FILTER
electro-harmonix
MADE IN USA
POG
POLYPHONIC OCTAVE GENERATOR

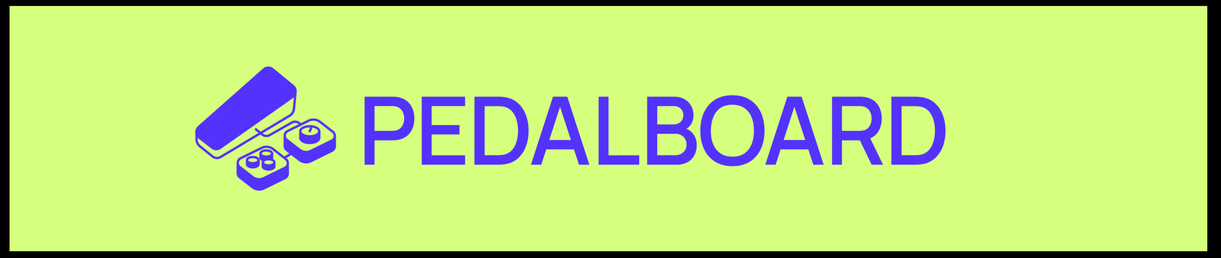
Pedalboard

https://github.com/spotify/pedalboard

Audio effects library by Spotify

Emulates popular musical effects (reverb, delay, distortion…)

Used for data augmentation and working with audio



```python
from pedalboard import Pedalboard, Chorus, Reverb
from pedalboard.io import AudioFile

# Read in a whole audio file:
with AudioFile('some-file.wav') as f:
  audio = f.read(f.frames)
  samplerate = f.samplerate

# Make a Pedalboard object, containing multiple plugins:
board = Pedalboard([Chorus(), Reverb(room_size=0.25)])

# Run the audio through this pedalboard!
effected = board(audio, samplerate)

# Write the audio back as a wav file:
with AudioFile('processed-output.wav', 'w', samplerate, effected.shape[0]) as f:
  f.write(effected)
```

Pyo

https://github.com/belangeo/pyo

Audio synthesis engine

Low-level digital signal processing

Actually used in live performances!

Matthieu Amiguet

Les Chemins De Traverse

Python and Pyo used for:

- **effects control**
- **implementing loopers**
- **implementing control scripts for shows**

Other cool tools

Audiomentations
audio effects
https://github.com/iver56/audiomentations

PyDub
audio editing
https://github.com/jiaaro/pydub

Soundfile
audio editing
https://github.com/bastibe/python-soundfile

APIs

Spotify API

Deezer API

APIs

Spotify API

Features: mood, danceability, energy, acousticness…

Musical structure: segments, beats, structure

Deezer API

Metadata, artist and album data

Recommendations

Datasets

ISMIR datasets

https://ismir.net/resources/datasets

status	dataset	metadata	contents	with audio
☠	200DrumMachines	audio samples	7371 one-shots	yes
✅	ACM_MIRUM	tempo	1410 excerpts (60s)	yes
✅	AcousticBrainz-Genre	15-31 genres with 265-745 subgenres	audio features for about 2000000 songs	no
☠	ADC2004	predominant pitch	20 excerpts	yes
✅	Acoustic Event Dataset	28 event classes	5223 audio snippets	yes
✅	AIST Dance Video Database	street dance videos	13,940 videos for 60 pieces	yes
✅	Amg1608	valence & arousal	1608 excerpts (30s)	no
✅	AMT-pilot	structure by multiple annotators	8 songs	yes

Datasets

mirdata

dataset utilities

https://github.com/mir-dataset-loaders/mirdata

Librosa

https://github.com/librosa

Music processing

Visualization

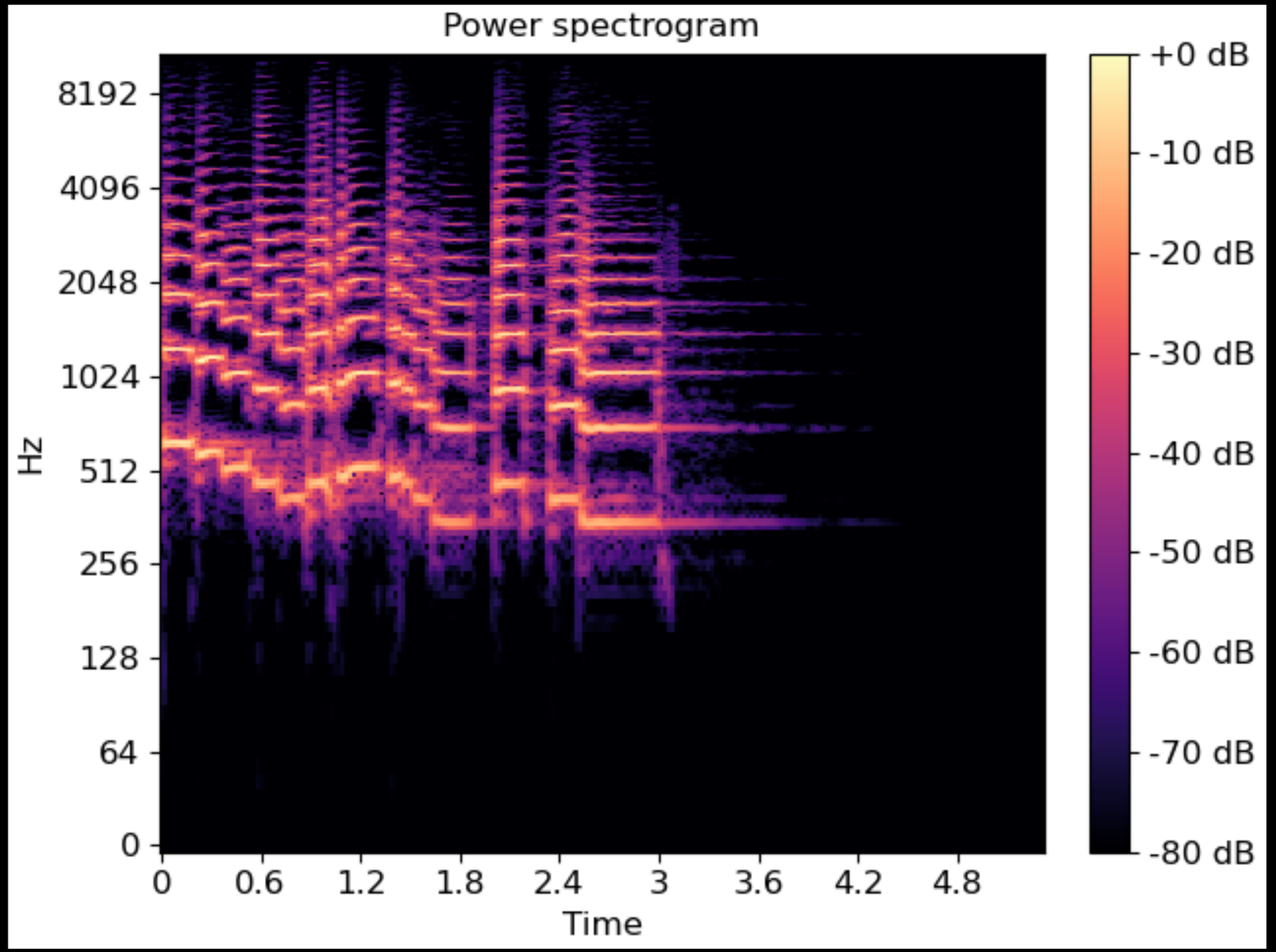
Feature extraction

McFee, Brian, et al. "*librosa: Audio and music signal analysis in Python.*"
Proceedings of the 14th python in science conference. Vol. 8. 2015.

```python
import librosa

# Extract harmonic and percussive parts
y, sr = librosa.load(librosa.ex('choice'))
y_harmonic, y_percussive = librosa.effects.hpss(y)

# Create a spectrogram
y, sr = librosa.load(librosa.ex('trumpet'))
S = np.abs(librosa.stft(y))
```

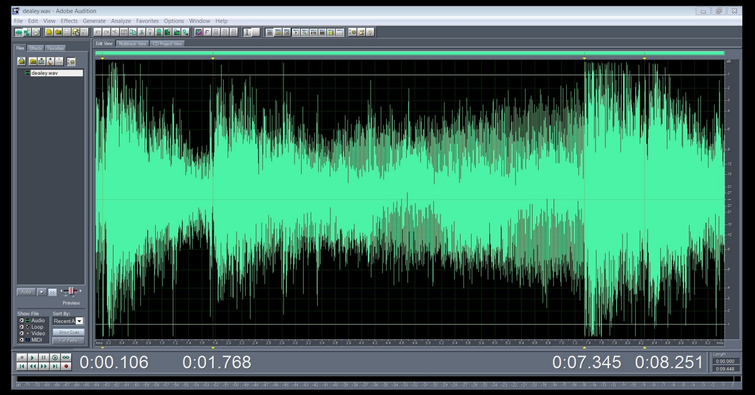

Power spectrogram
8192
4096
2048
1024
512
256
128
64
0
Hz
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Music representations

Music representations

MUSIC

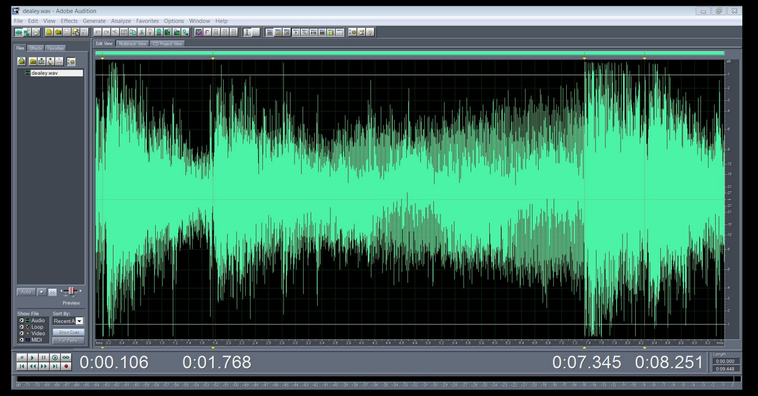
AUDIO
(.wav, .flac, .mp3…)



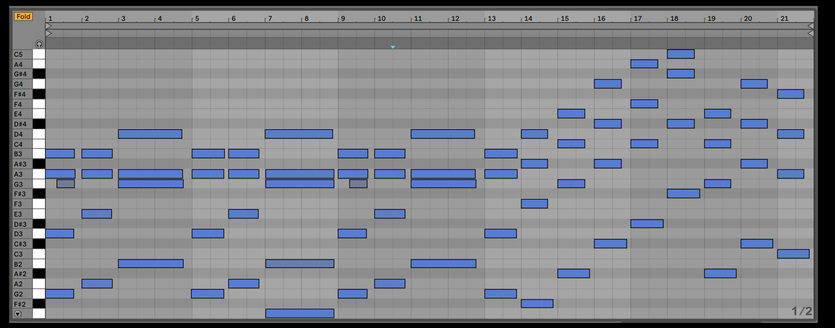
Music representations

MUSIC

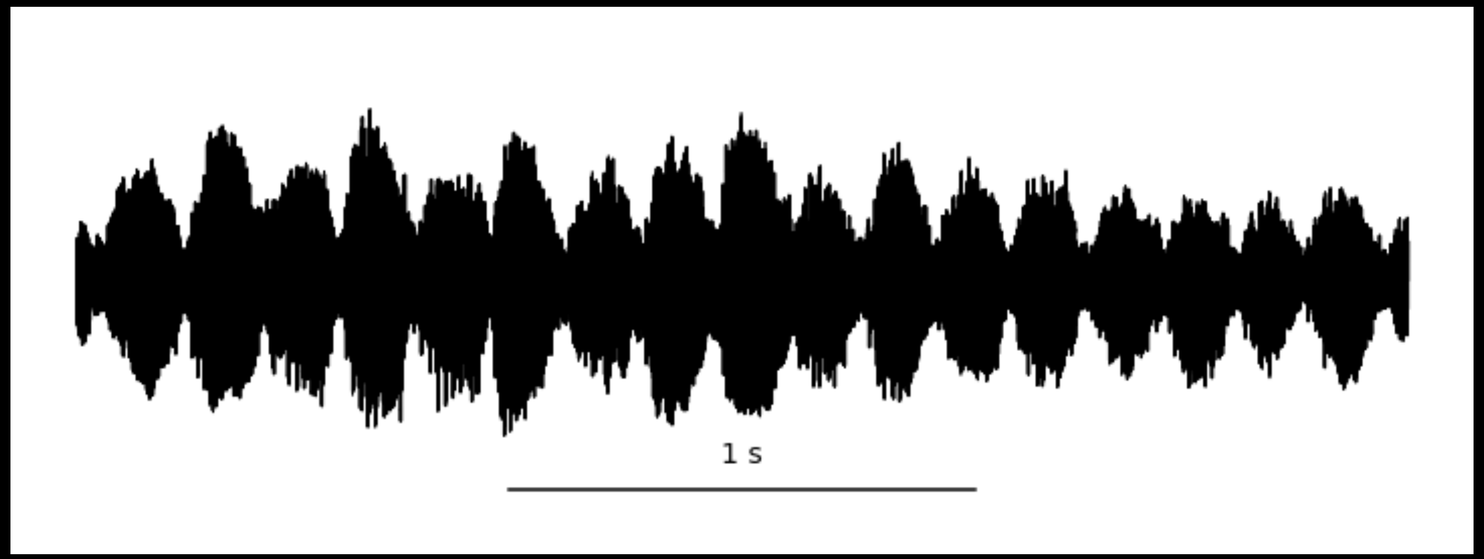
AUDIO
(.wav, .flac, .mp3…)



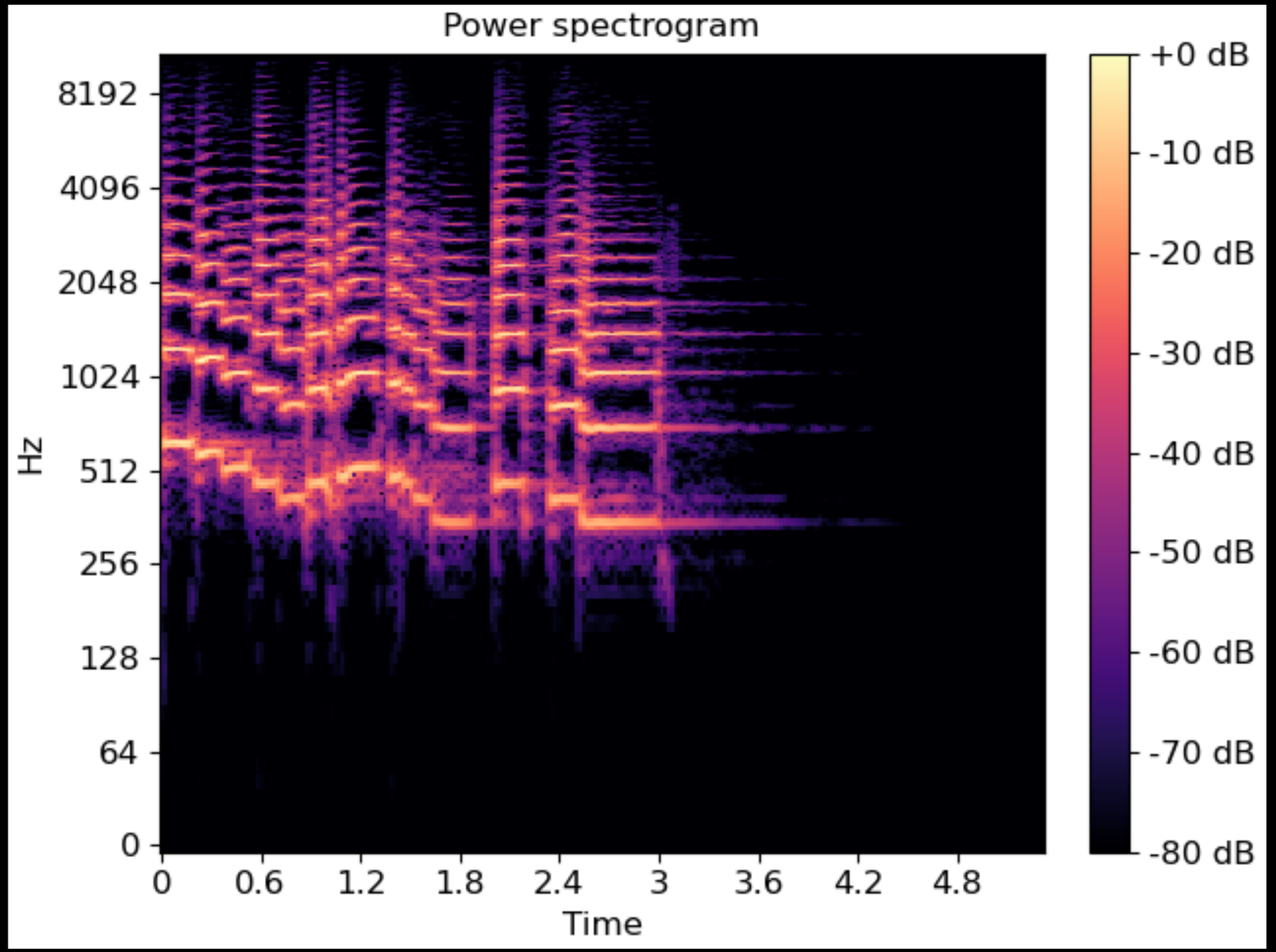
SYMBOLIC
(.mid, .abc, .mxl…)



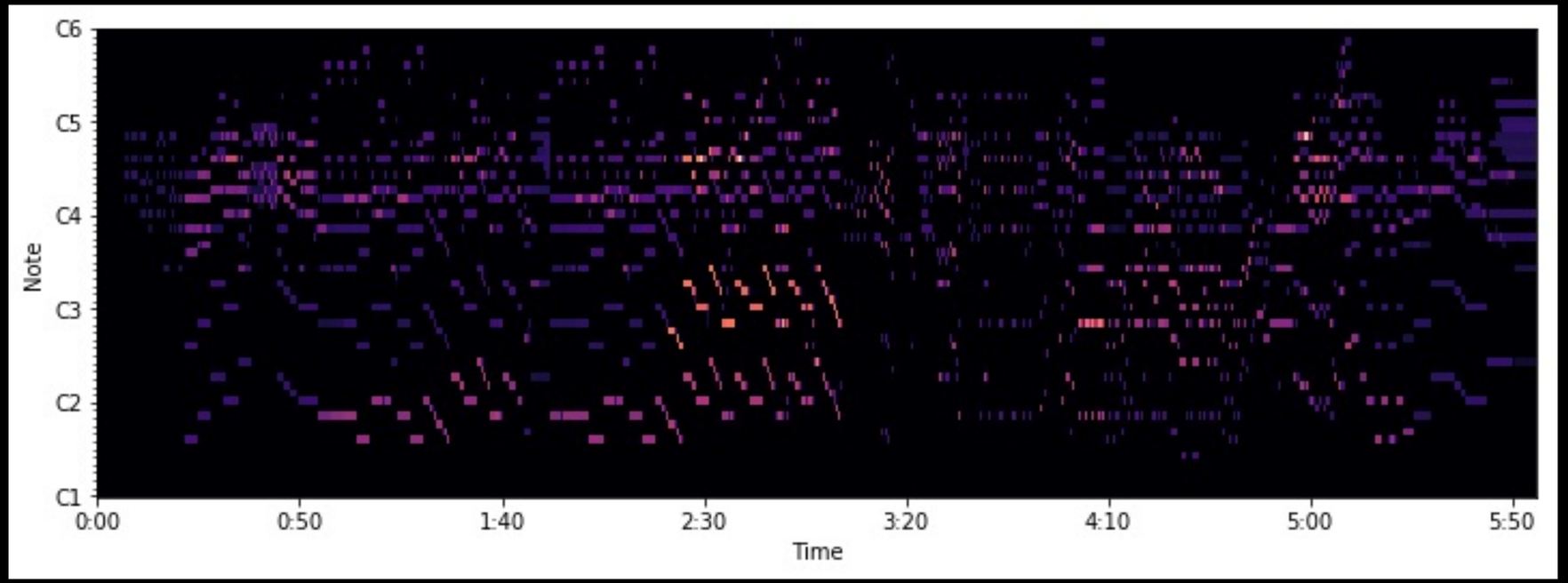
Audio vs symbolic



Jan Van Balen, „What a WAV file looks like”, https://jvbalen.github.io/notes/waveform.html


Power spectrogram
8192
4096
2048
1024
512
256
128
64
0
Hz
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

MIDI



MIDI

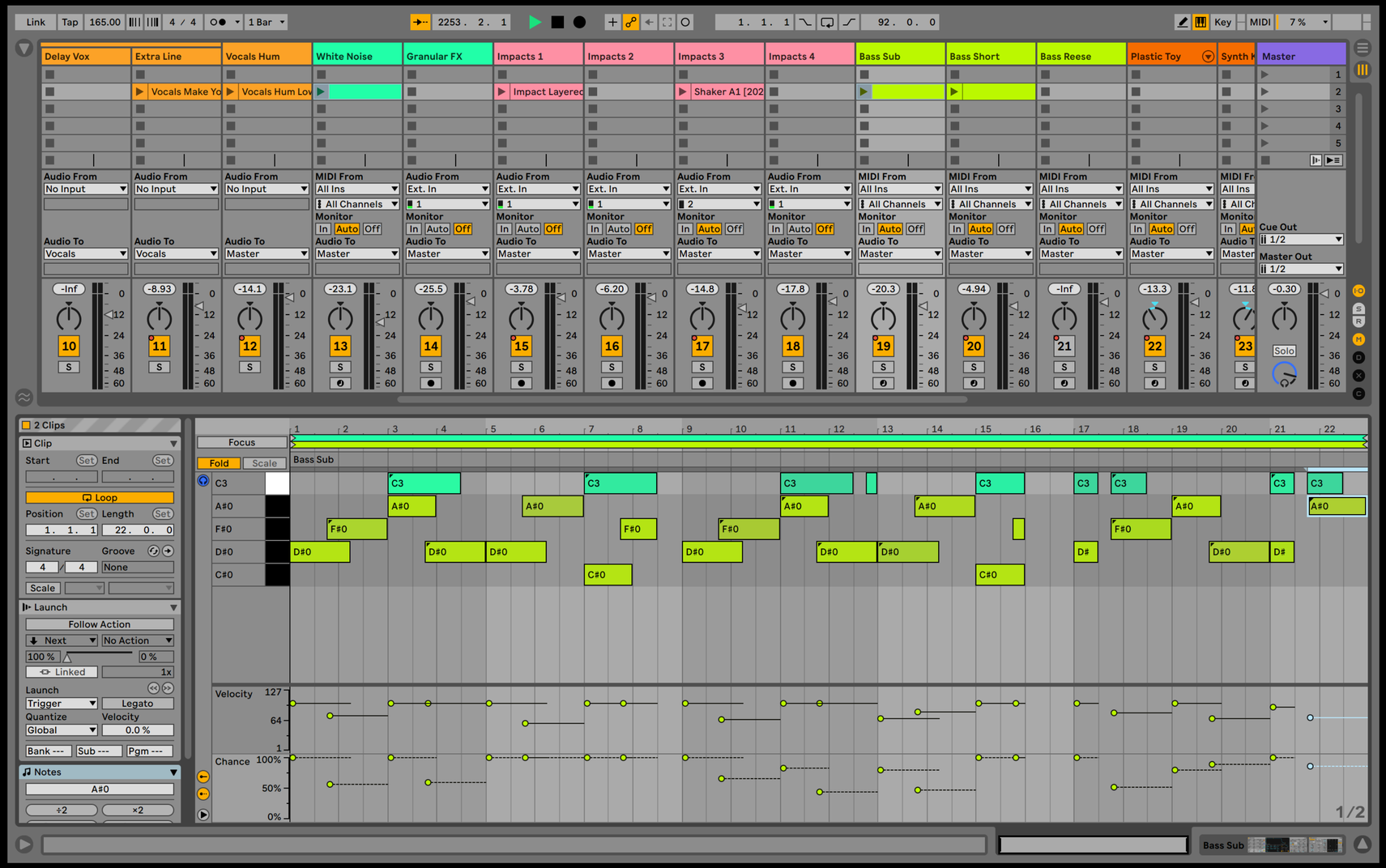
Protocol invented in early 80’s

Communicates synthesizers and computers

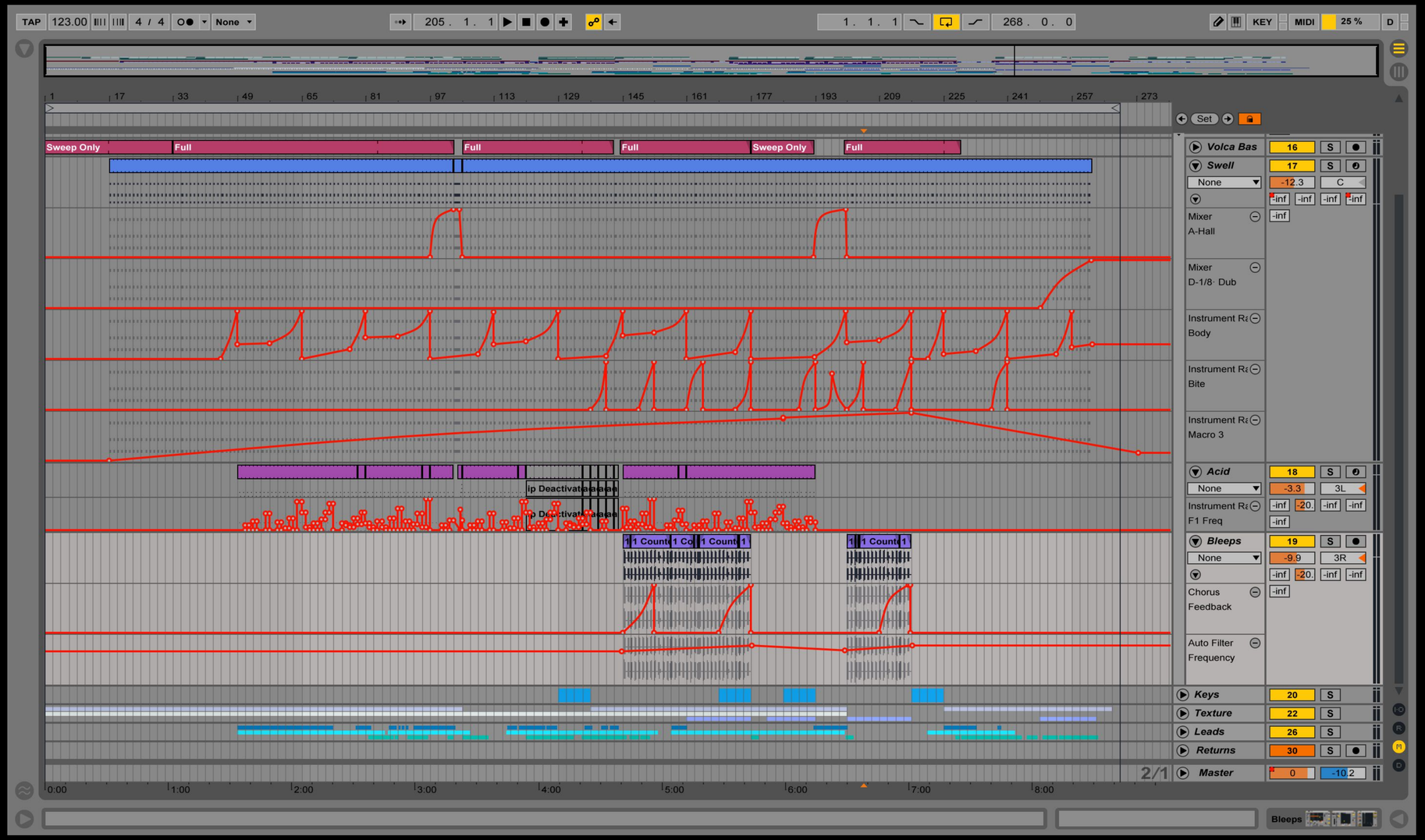
Stores note information and multiple control values

Relevant up to this day

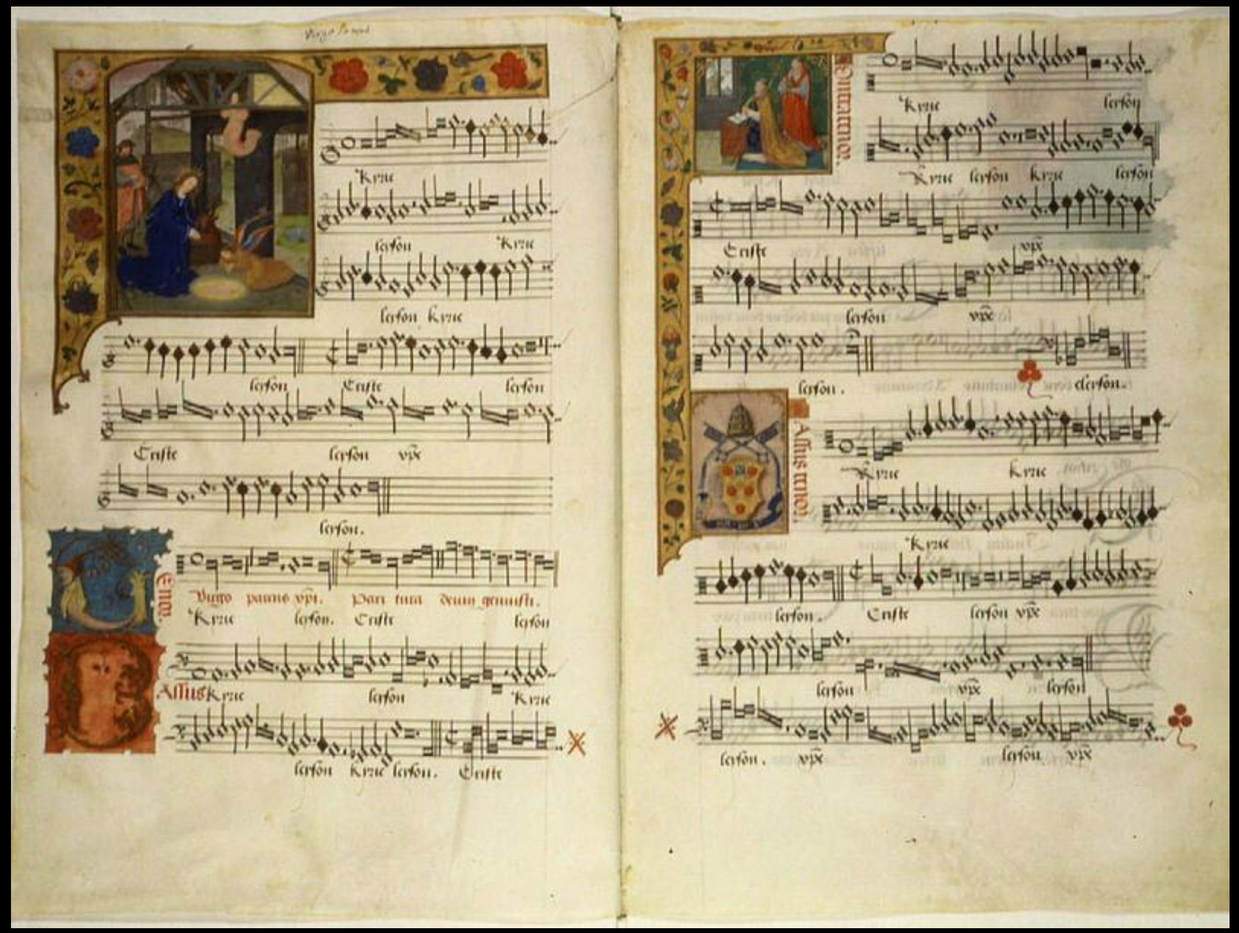
MIDI

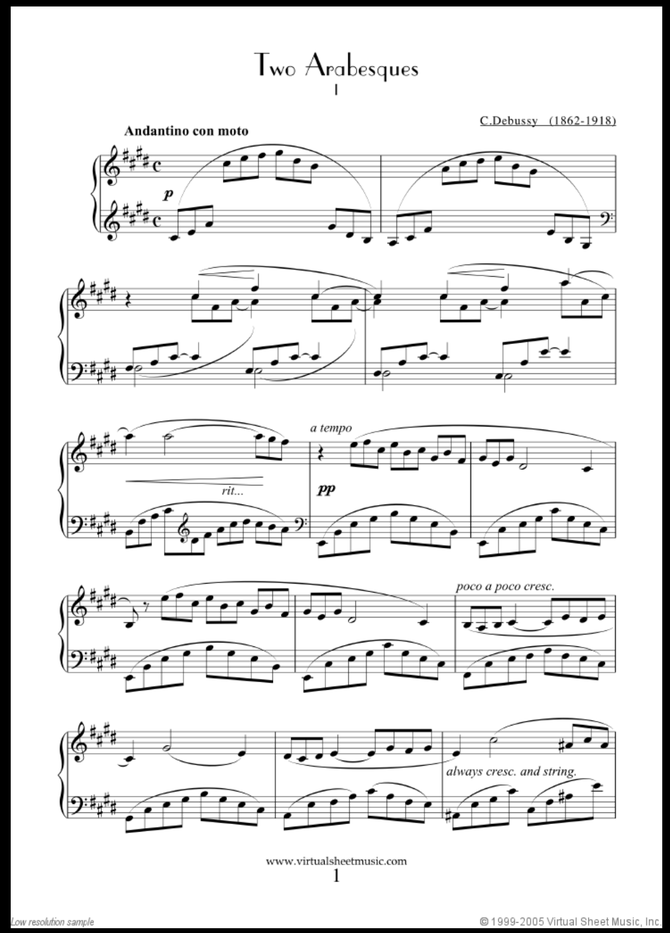

Link
Tap
165.00
4 / 4
1 Bar
2253 . 2 . 1
1 . 1 . 1
92 . 0 . 0
Key
MIDI
7 %
Delay Vox
Extra Line
Vocals Hum
White Noise
Granular FX
Impacts 1
Impacts 2
Impacts 3
Impacts 4
Bass Sub
Bass Short
Bass Reese
Plastic Toy
Master
Vocals Make Yo
Vocals Hum Lov
Impact Layerec
Shaker A1 [202
Audio From
No Input
MIDI From
All Ins
All Channels
Monitor
In Auto Off
Audio To
Vocals
Master
Cue Out
1/2
Master Out
1/2
-Inf
-8.93
-14.1
-23.1
-25.5
-3.78
-6.20
-14.8
-17.8
-20.3
-4.94
-Inf
-13.3
-11.8
-0.30
Solo
2 Clips
Clip
Start
End
Loop
Position
Length
1. 1. 1
22. 0. 0
Signature
Groove
4
4
None
Scale
Launch
Follow Action
Next
No Action
100 %
0 %
Linked
1x
Launch
Trigger
Legato
Quantize
Velocity
Global
0.0 %
Bank ---
Sub ---
Pgm ---
Notes
A#0
÷2
×2
Focus
Fold
Scale
Bass Sub
C3
A#0
F#0
D#0
C#0
D#
Velocity
127
64
1
Chance
100%
50%
0%
1/2
Bass Sub

MIDI - effects and parameters



The original symbolic format 😎





Two Arabesques

I

C.Debussy (1862-1918)

Andantino con moto

a tempo

rit...

poco a poco cresc.

always cresc. and string.

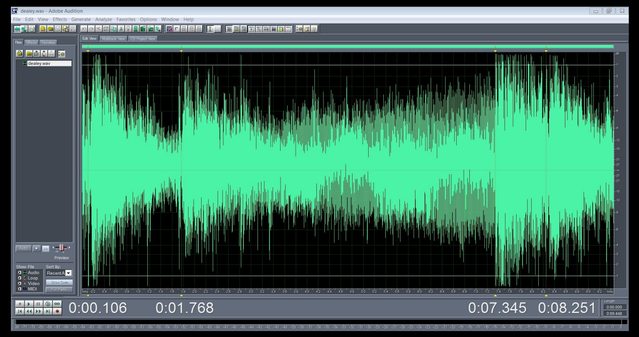
www.virtualsheetmusic.com

1

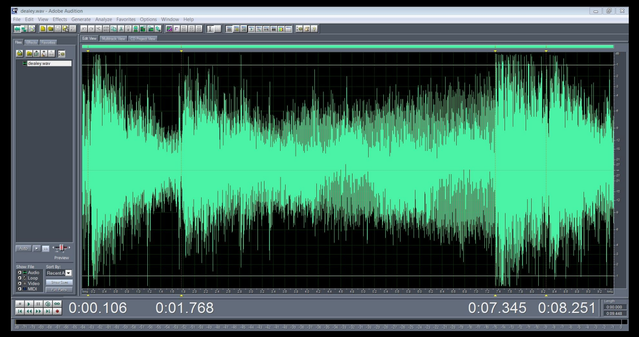
Low resolution sample

Learning from music

Music tagging

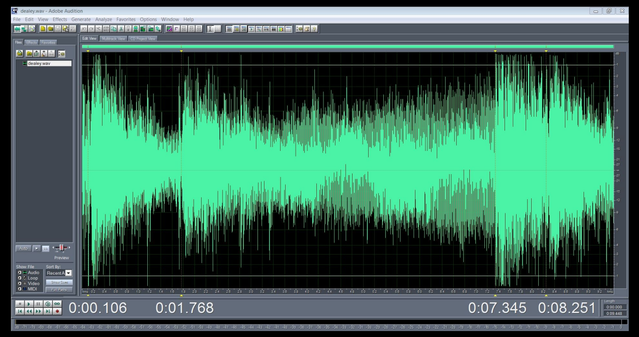


Music tagging



MODEL

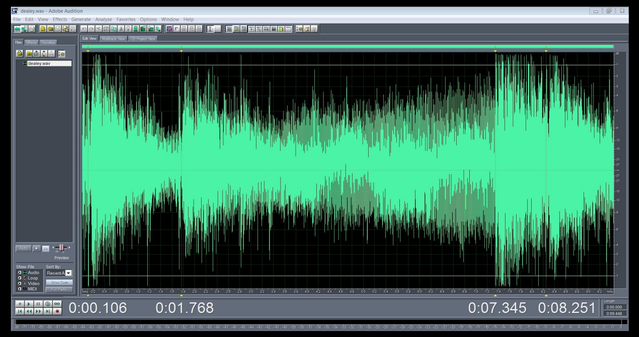
Music tagging



MODEL

Rock

Music tagging



MODEL

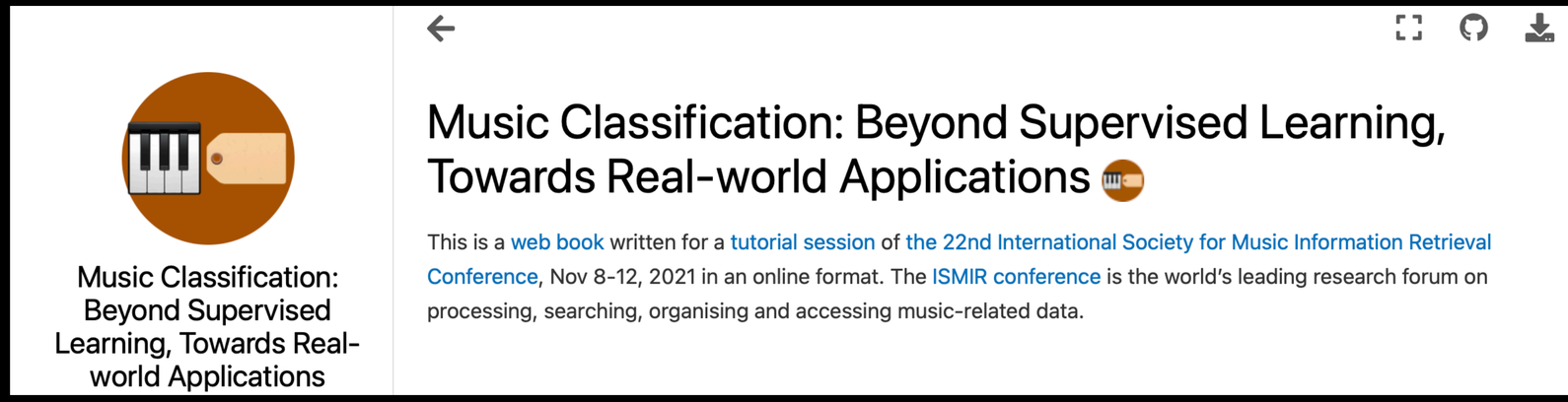
- Rock
- Pop
- Female singer
- Guitar
- Upbeat
- Happy
- Acoustic
- 90’s

Music tagging

Music Classification: Beyond Supervised Learning, Towards Real-world Applications

M. Won, J.Spijkervet, K. Choi, 2021

https://music-classification.github.io/



Music tagging

State of the art music tagging models in Pytorch!

M.Won, 2021

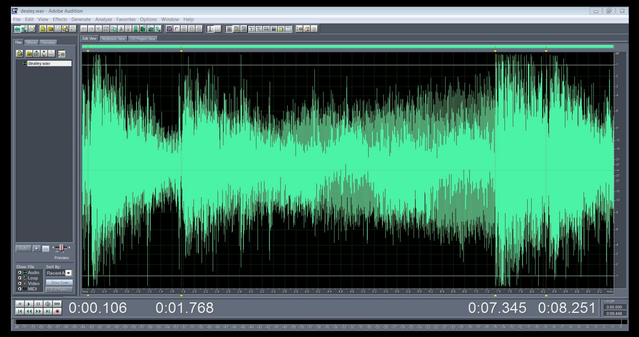
https://github.com/minzwon/sota-music-tagging-models

Available Models

- **FCN** : Automatic Tagging using Deep Convolutional Neural Networks, Choi et al., 2016 [arxiv]
- **Musicnn** : End-to-end Learning for Music Audio Tagging at Scale, Pons et al., 2018 [arxiv]
- **Sample-level CNN** : Sample-level Deep Convolutional Neural Networks for Music Auto-tagging Using Raw Waveforms, Lee et al., 2017 [arxiv]
- **Sample-level CNN + Squeeze-and-excitation** : Sample-level CNN Architectures for Music Auto-tagging Using Raw Waveforms, Kim et al., 2018 [arxiv]
- **CRNN** : Convolutional Recurrent Neural Networks for Music Classification, Choi et al., 2016 [arxiv]
- **Self-attention** : Toward Interpretable Music Tagging with Self-Attention, Won et al., 2019 [arxiv]
- **Harmonic CNN** : Data-Driven Harmonic Filters for Audio Representation Learning, Won et al., 2020 [pdf]
- **Short-chunk CNN** : Prevalent 3x3 CNN. So-called *vgg*-ish model with a small receptieve field.
- **Short-chunk CNN + Residual** : Short-chunk CNN with residual connections.

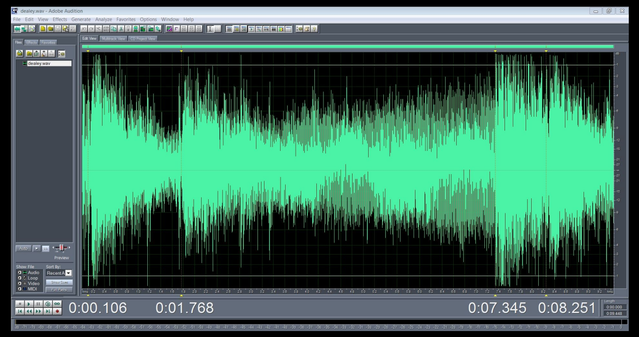
Source separation

Source separation

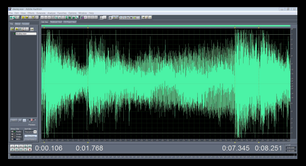


MODEL

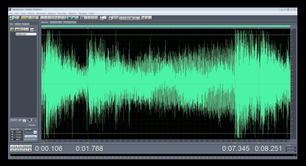
Source separation



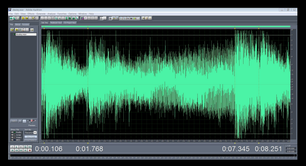
MODEL



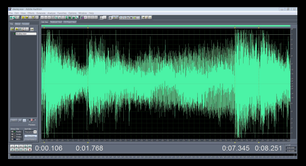
VOCALS



DRUMS



BASS



OTHER

R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam and J. P. Bello, "*MedleyDB: A Multitrack Dataset for Annotation-Intensive MIR Research*", in 15th International Society for Music Information Retrieval Conference, Taipei, Taiwan, Oct. 2014.

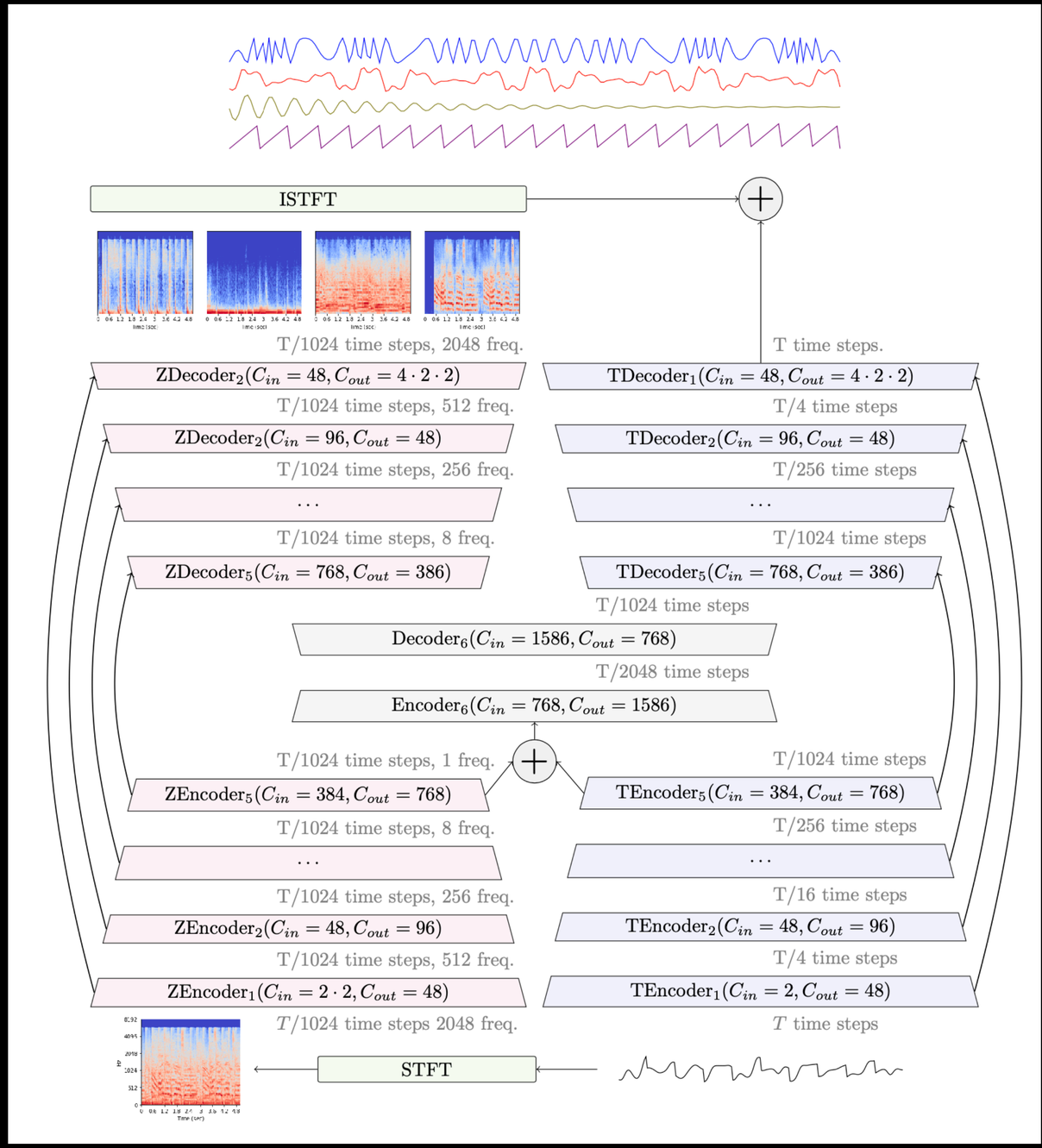
Rafii, Zafar and Liutkus, Antoine and Stöter, Fabian-Robert and Mimilakis, Stylianos Ioannis and Bittner, Rachel, *„The MuseDB corpus for music separation”*, 2017

Source separation

Demucs

Facebook, 2020

https://github.com/facebookresearch/demucs



Défossez, Alexandre. "*Hybrid spectrogram and waveform source separation.*" Proceedings of the ISMIR 2021 Workshop on Music Source Separation, 2021

Source separation

Spleeter

Deezer, 2020

https://github.com/deezer/spleeter

```python
from spleeter.separator import Separator
from spleeter.audio.adapter import AudioAdapter

audio_loader = AudioAdapter.default()
sample_rate = 44100
waveform, _ = audio_loader.load('/path/to/audio/file',
                                sample_rate=sample_rate)

# Separator config
separator = Separator('spleeter:2stems')

# Perform the separation:
prediction = separator.separate(waveform)
```

Hennequin, Romain, et al. "*Spleeter: a fast and efficient music source separation tool with pre-trained models.*" Journal of Open Source Software 5.50 (2020): 2154.

Source separation

Spleeter

Deezer, 2020

https://github.com/deezer/spleeter

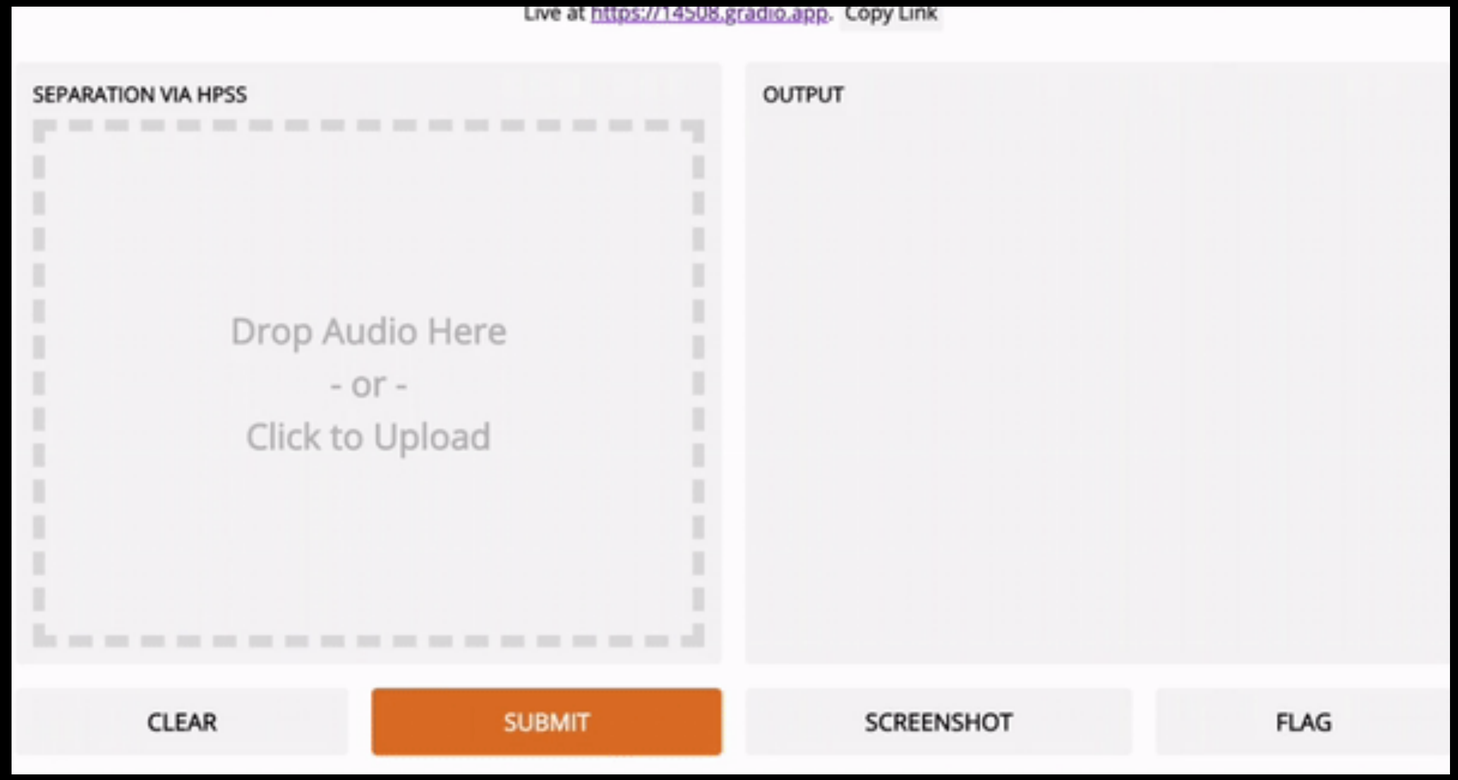
Izotope - Music Rebalance

Source separation

nussl

E.Manilow et al., 2018

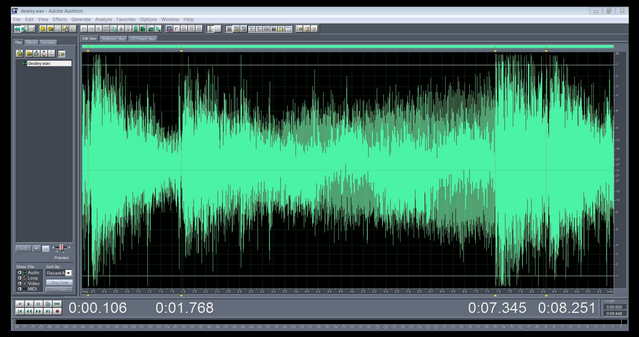
https://github.com/nussl/nussl



Manilow, Ethan, Prem Seetharaman, and Bryan Pardo. "The Northwestern University Source Separation Library." *ISMIR*. 2018.

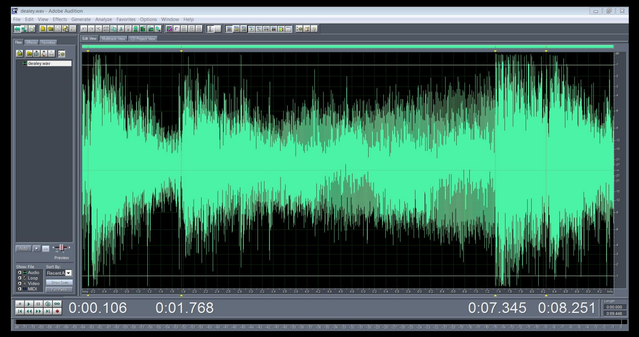
Transcription

Transcription

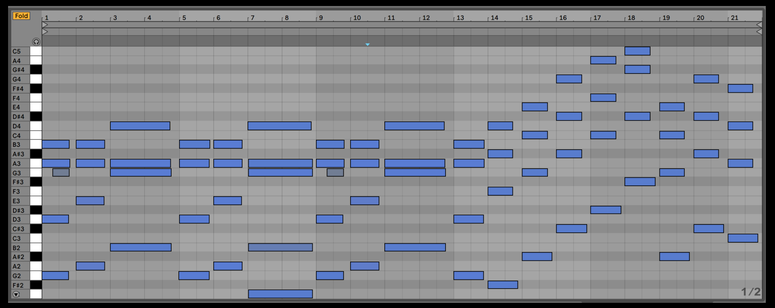


MODEL

Transcription



MODEL



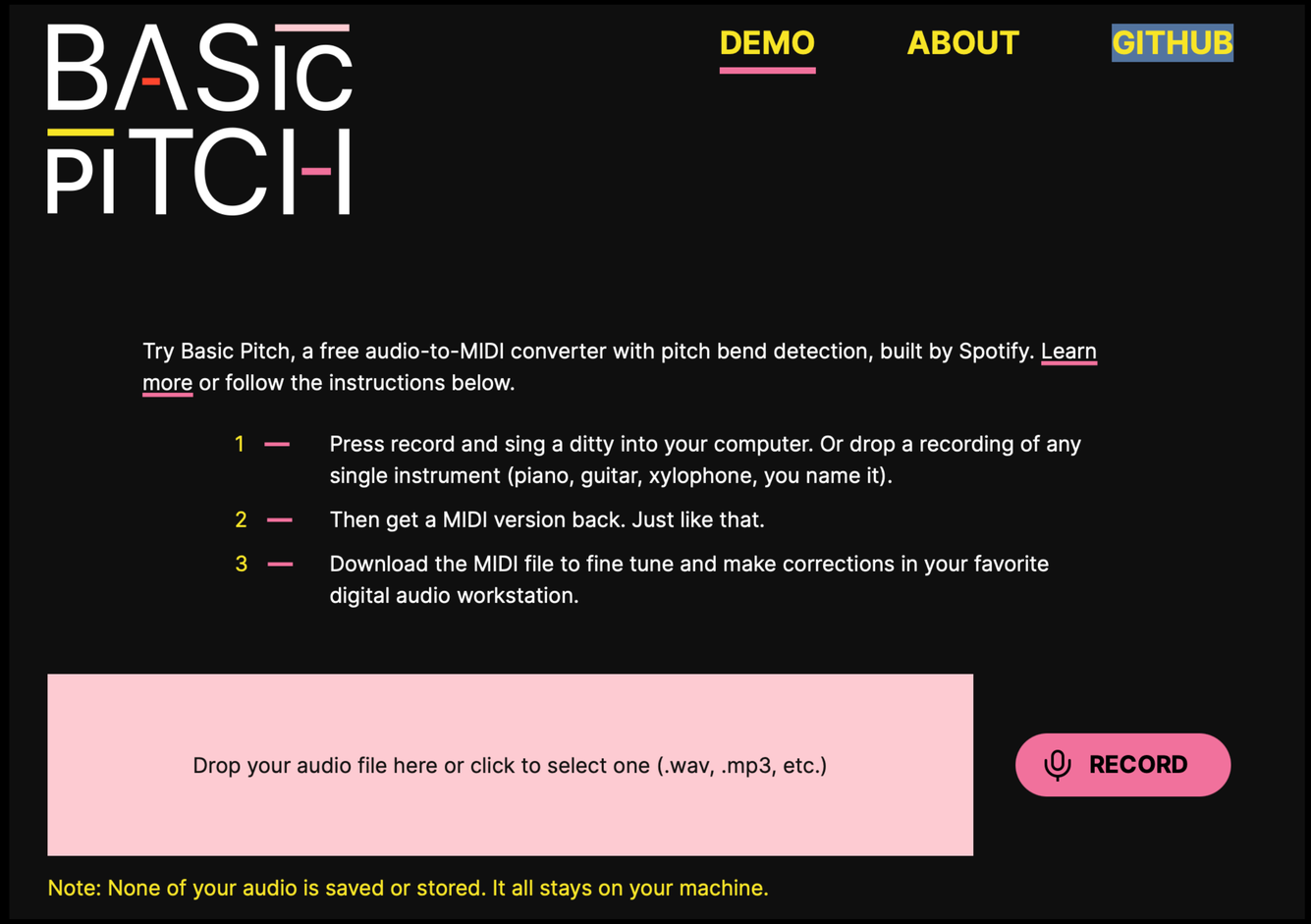
MIDI
(or other symbolic format)

Transcription

Basic Pitch

Spotify, 2022

https://github.com/spotify/basic-pitch



Bittner, Rachel M., et al. "A Lightweight Instrument-Agnostic Model for Polyphonic Note Transcription and Multipitch Estimation." *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022.

Music recommendations

Music recommendations

Complex topic with massive business and cultural impact

Tagging, retrieval, ranking, engineering, A/B testing…

Music recommendations

Complex topic with massive business and cultural impact

Tagging, retrieval, ranking, engineering, A/B testing…

Python libraries: Spotlight, Surprise, Implicit, LightFM, pandas, TensorFlow, PyTorch

Other tasks

Beat tracking
Pitch tracking
Chord detection
Tempo estimation
Structure analysis
Transcription
Artist similarity
Clustering
Music training enhancement

Generating music

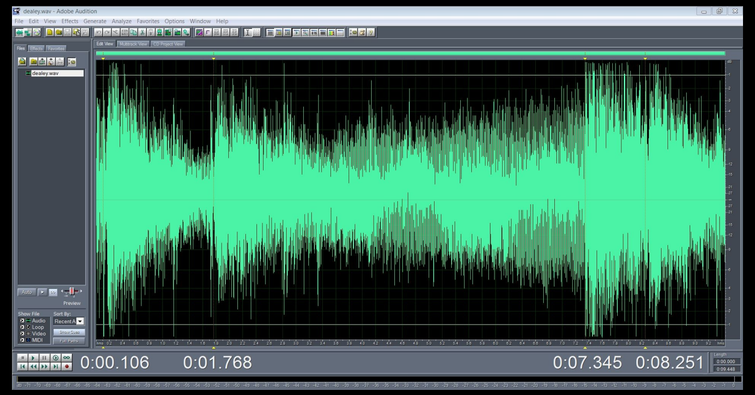
Generating music

Neural audio synthesis

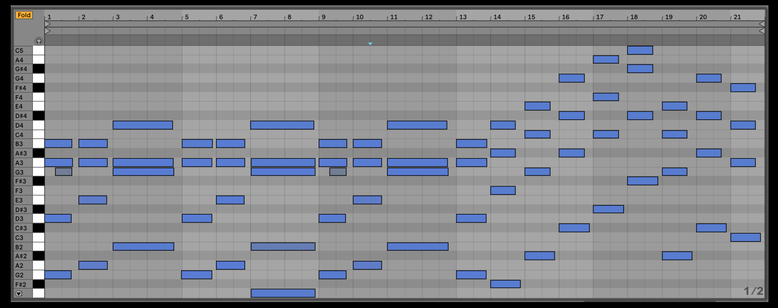
Symbolic music generation

Generating music

Neural audio synthesis



Symbolic music generation



Symbolic music generation

Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016).
Music transcription modelling and composition using deep learning.
arXiv preprint arXiv:1604.08723.

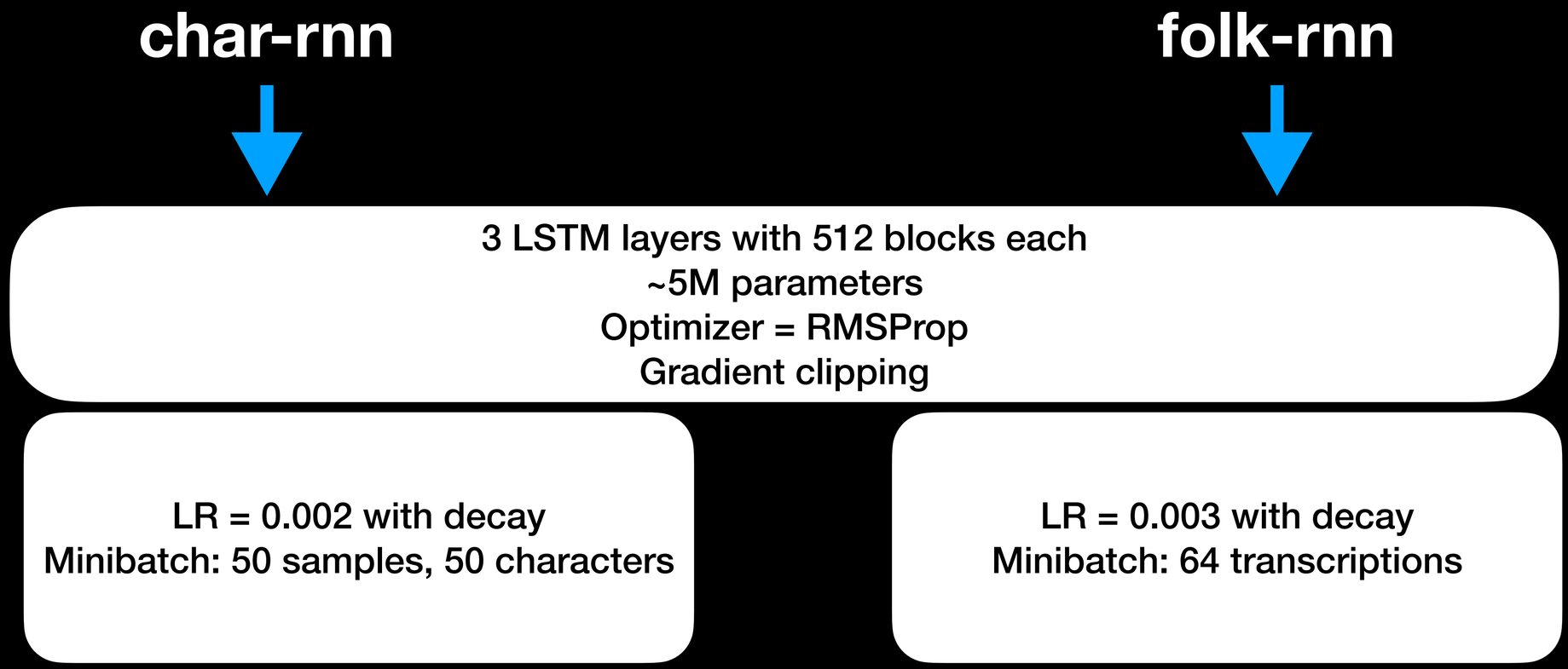
Symbolic music generation

char-rnn

folk-rnn

Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016). *Music transcription modelling and composition using deep learning.* arXiv preprint arXiv:1604.08723.

Symbolic music generation



Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016). *Music transcription modelling and composition using deep learning.* arXiv preprint arXiv:1604.08723.

Symbolic music generation

char-rnn

```
T: A Cup Of Tea
M: 4/4
L: 1/8
K: Amix
|:eA (3AAA g2 fg|eA (3AAA BGGf|eA (3AAA g2 fg|1afge d2 gf:|2afge d2 cd||
|:eaag efgf|eaag edBd|eaag efge|afge dgfg:|

T: A Cup Of Tea
M: 4/4
L: 1/8
K: Ador
eAAa ~g2fg|eA~A2 BGBd|eA~A2 ~g2fg|1af (3gfe dG~G2:|2af (3gfe d2^cd||
eaag efgf|eaag ed (3Bcd|eaag efgb|af (3gfe d2^cd:|
```

folk-rnn

```
<s> M:4/4 K:Cmix |: g c (3 c c c b 2 a b | g c (3 c c c d B B a | g c (3
c c c b 2 a b |1 c' a b g f 2 b a :| |2 c' a b g f 2 e f |: g c' c' b g
a b a | g c' c' b g f d f | g c' c' b g a b g | c' a b g f b a b :| <\s>
<s> M:4/4 K:Cdor g c c c' b 2 a b | g c c 2 d B d f | g c c 2 b 2 a b |1
c' a (3 b a g f B B 2 :| |2 c' a (3 b a g f 2 =e f | g c' c' b g a b a | g
c' c' b g f (3 d e f | g c' c' b g a b d' | c' a (3 b a g f 2 =e f :| <\s>
```

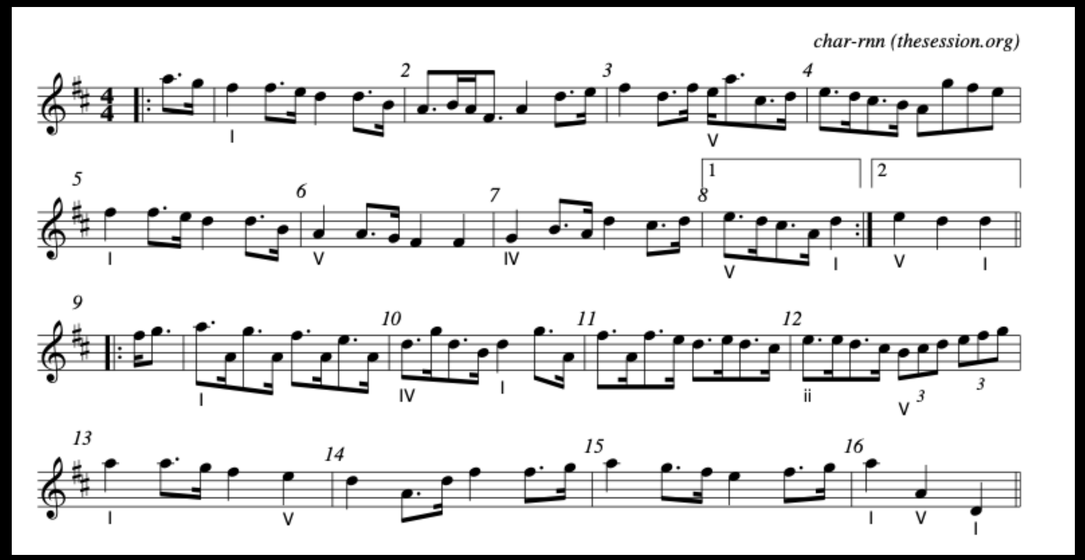
Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016). *Music transcription modelling and composition using deep learning.* arXiv preprint arXiv:1604.08723.

Symbolic music generation

Musical analysis:

- Typical structure of traditional Irish music: 8 bar „tune” and 8 bar „turn”
- Repetition and variation
- Harmony implicit in melody
- Checked against dataset for originality
- „No glaring mistakes”

```
T: Mal's Copporim, The
M: 4/4
L: 1/8
K: Dmaj
|: a>g | f2 f>e d2 d>B | A>BA<F A2 d>e | f2 d>f e<ac>d | e>dc>B Agfe |
f2 f>e d2 d>B | A2 A>G F2 F2 | G2 B>A d2 c>d |[1 e>dc>A d2 :|[2 e2 d2 d2 ||
|: f<g | a>Ag>A f>Ae>A | d>gd>B d2 g>A | f>Af>e d>ed>c | e>ed>c (3Bcd (3efg |
a2 a>g f2 e2 | d2 A>d f2 f>g | a2 g>f e2 f>g | a2 A2 D2 ||
```



Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016). *Music transcription modelling and composition using deep learning.* arXiv preprint arXiv:1604.08723.

Symbolic music generation

Google Magenta
https://magenta.tensorflow.org/

Music Transformer

Music VAE

MelodyRNN

Bach Doodle

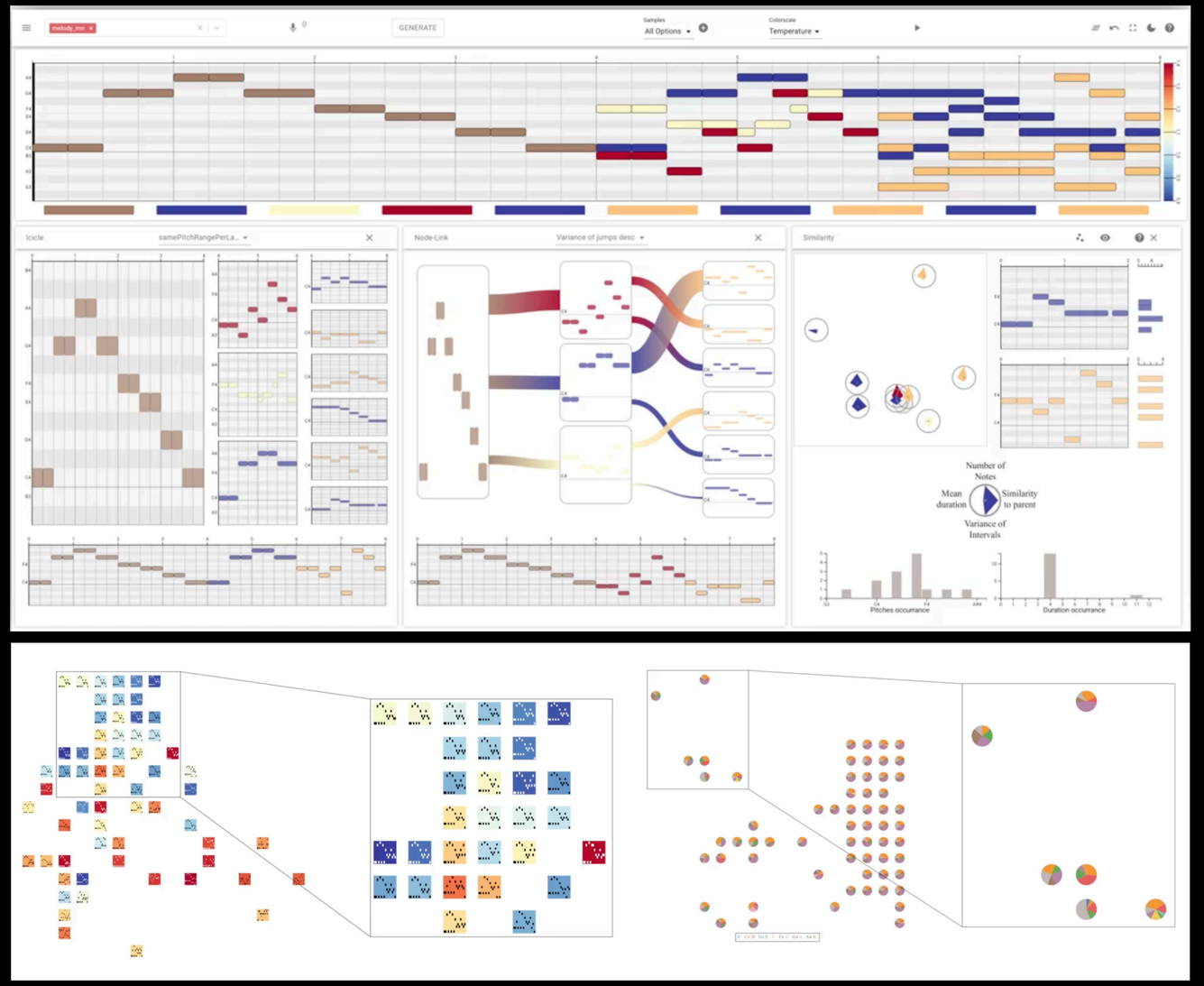
VST plug-ins!

More new tools!

Visualization for AI-Assisted Composing.

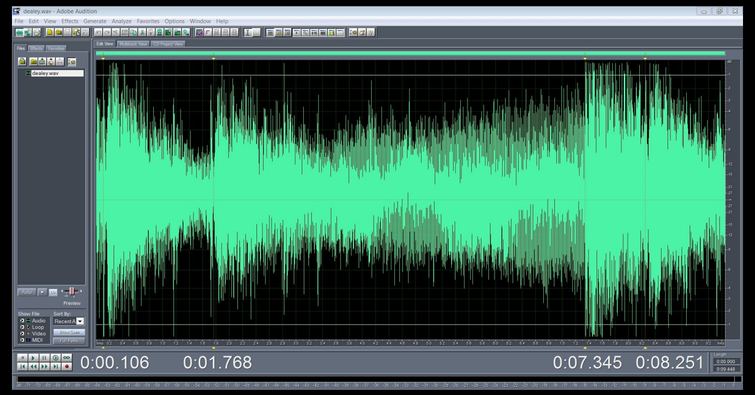
Rau, Simeon, et al.

ISMIR 2022

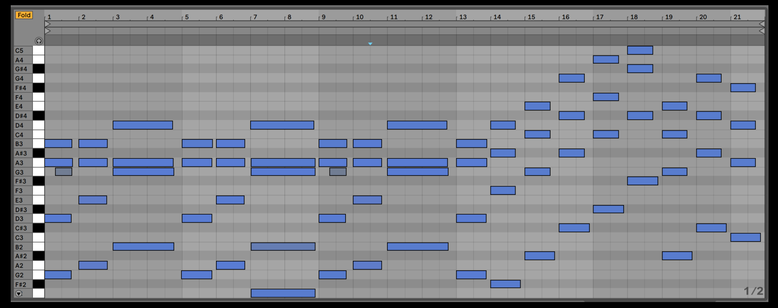


Generating music

Neural audio synthesis



Symbolic music generation



Neural audio synthesis

PixelRNN, Wavenet

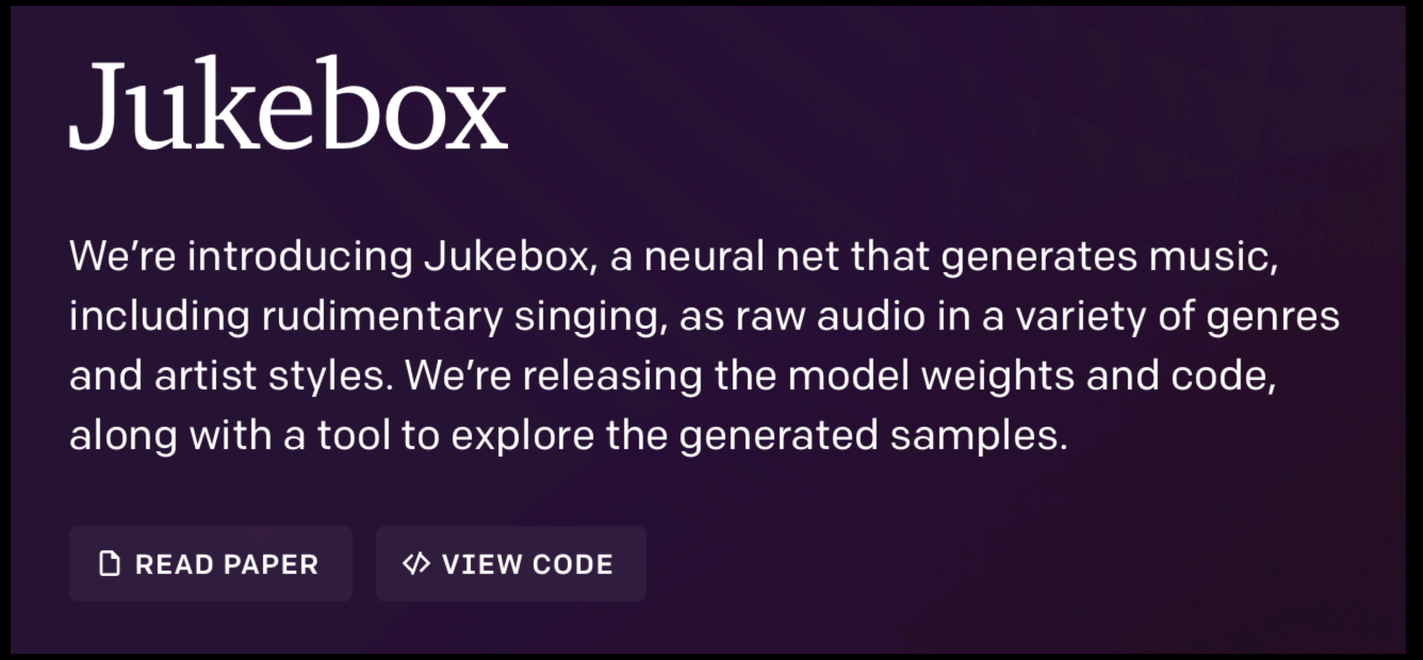
A. van den Oord et al.,
Google / DeepMind, 2016

DDSP

J.Engel et al., Google
Magenta 2020

Neural audio synthesis

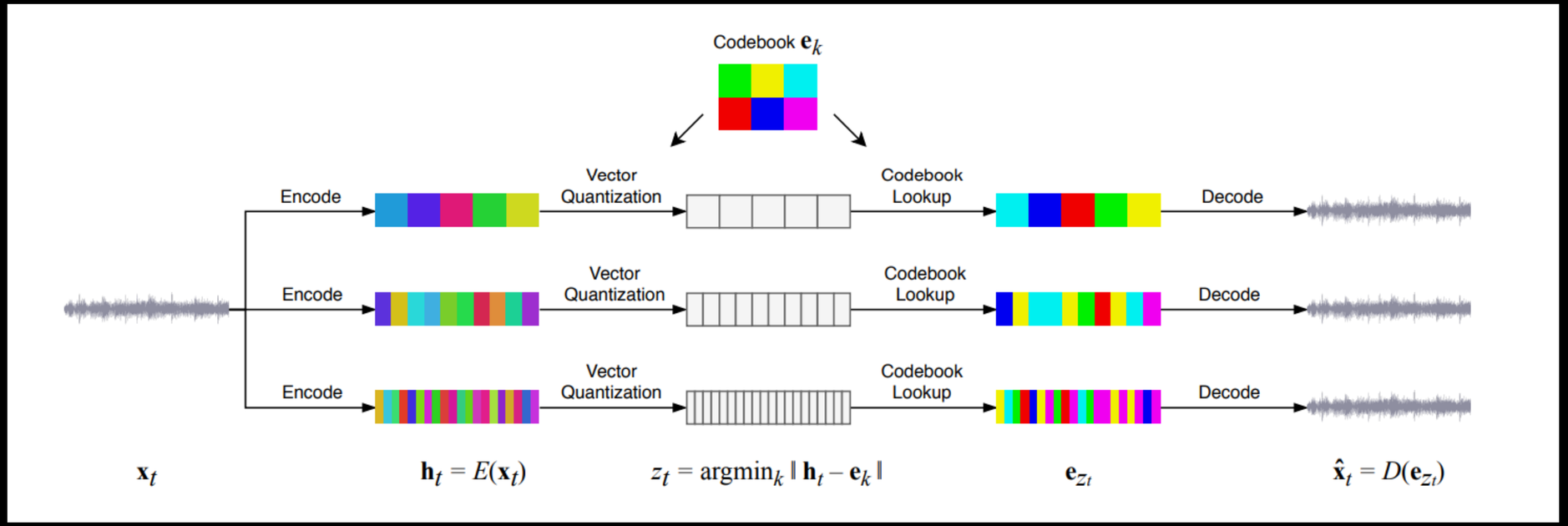
Jukebox
OpenAI, 2020



Neural audio synthesis

Jukebox

OpenAI, 2020



$$\mathcal{L} = \mathcal{L}_{\text{recons}} + \mathcal{L}_{\text{codebook}} + \beta\mathcal{L}_{\text{commit}} \quad (1)$$

$$\mathcal{L}_{\text{recons}} = \frac{1}{T}\sum_t \|\mathbf{x}_t - D(\mathbf{e}_{z_t})\|_2^2 \quad (2)$$

$$\mathcal{L}_{\text{codebook}} = \frac{1}{S}\sum_s \|\text{sg}\,[\mathbf{h}_s] - \mathbf{e}_{z_s}\|_2^2 \quad (3)$$

$$\mathcal{L}_{\text{commit}} = \frac{1}{S}\sum_s \|\mathbf{h}_s - \text{sg}\,[\mathbf{e}_{z_s}]\|_2^2 \quad (4)$$

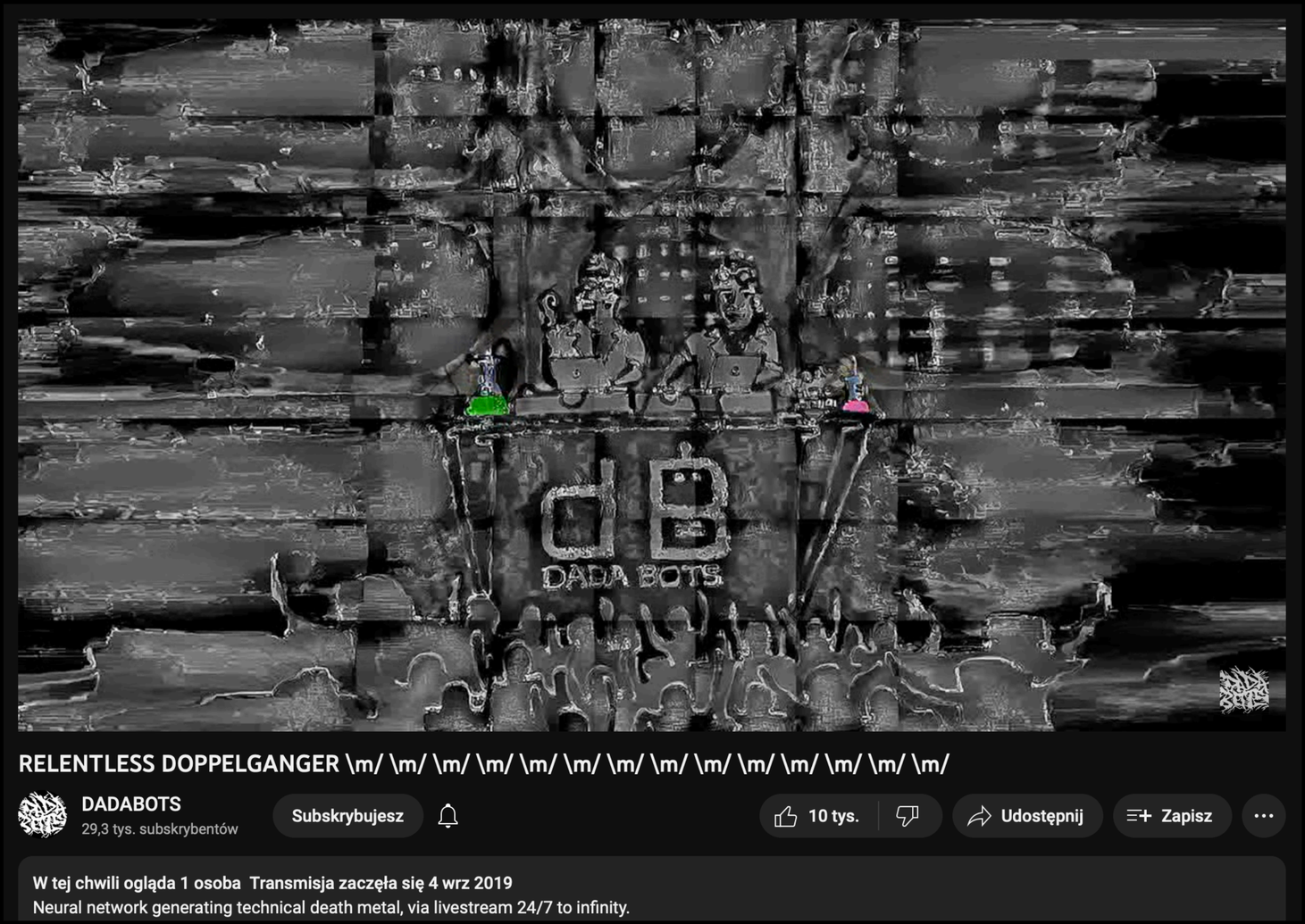
Neural audio synthesis

Dadabots

Neural audio synthesis

Dadabots

Relentless Doppelganger 😂



Neural audio synthesis

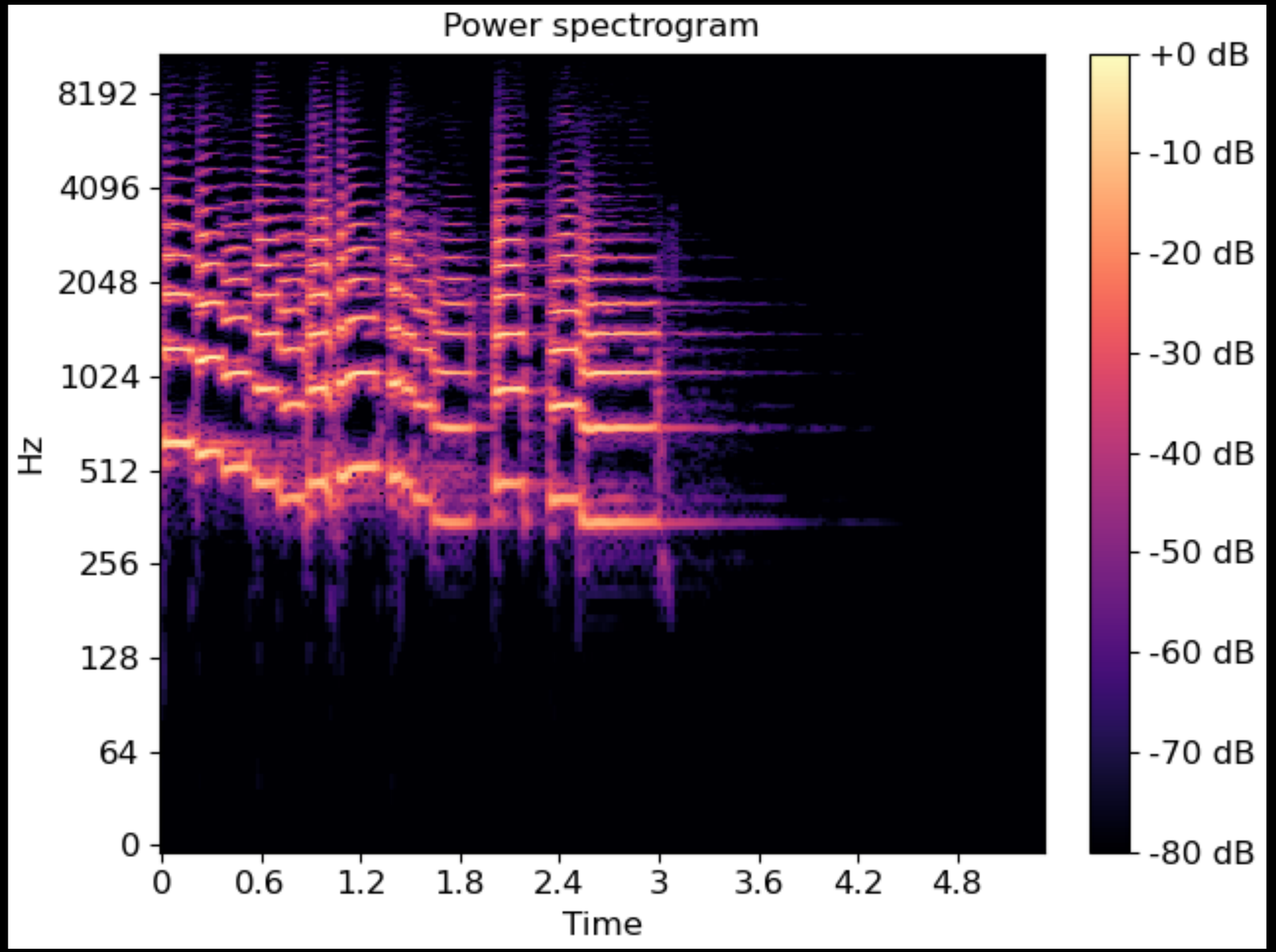
Dadabots

Relentless Doppelganger 😂

Nuns in a Moshpit 😂😂😂

We generated nuns singing “hail satan”, death reeeees, breakbeats, GuitarPro riffs & more using SampleRNN, Jukebox, Catch-A-Waveform, qubit neural synthesis, conceptular synthesis, DadaGP, and raw audio diffusion. Then assembled the pieces together in Ableton.

Dadabots


Power spectrogram
8192
4096
2048
1024
512
256
128
64
0
Hz
0
0.6
1.2
1.8
2.4
3
3.6
4.2
4.8
Time
+0 dB
-10 dB
-20 dB
-30 dB
-40 dB
-50 dB
-60 dB
-70 dB
-80 dB

Riffusion

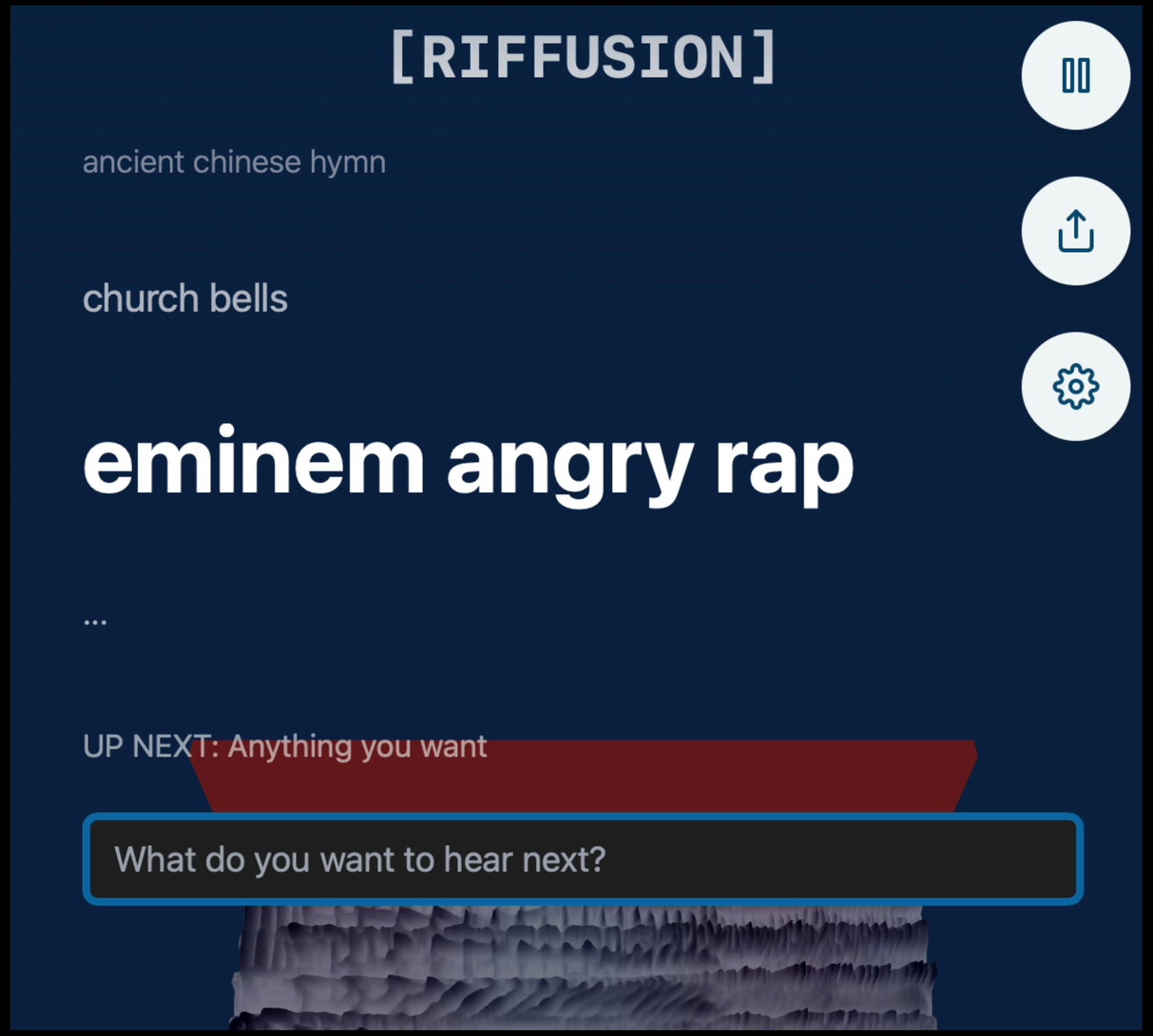
You've heard of Stable Diffusion, the open-source AI model that generates images from text?

photograph of an astronaut riding a horse

Well, we fine-tuned the model to generate images of spectrograms, like this:

funk bassline with a jazzy saxophone solo

Riffusion


[RIFFUSION]
ancient chinese hymn
church bells
eminem angry rap
...
UP NEXT: Anything you want
What do you want to hear next?

MusicLM

MusicLM: Generating music from text

A. Agostinelli et al., Google 2023

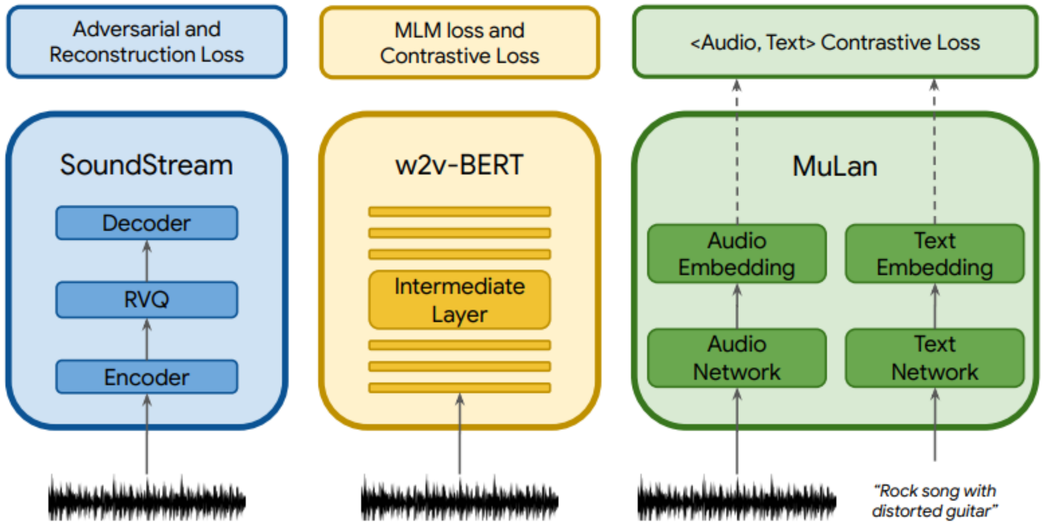


Figure 1. Independent pretraining of the models providing the audio and text representations for MusicLM: SoundStream (Zeghidour et al., 2022), w2v-BERT (Chung et al., 2021), and MuLan (Huang et al., 2022).

Sample synthesis

Sample synthesis

DRUMGAN: SYNTHESIS OF DRUM SOUNDS WITH TIMBRAL FEATURE CONDITIONING USING GENERATIVE ADVERSARIAL NETWORKS

Javier Nistal
Sony CSL
Paris, France

Stefan Lattner
Sony CSL
Paris, France

Gaël Richard
LTCI, Télécom Paris
Institut Polytechnique de Paris, France

CRASH: Raw Audio Score-based Generative Modeling for Controllable High-resolution Drum Sound Synthesis

Simon Rouard*
Sony CSL - CentraleSupélec
simon.rouard@student-cs.fr

Gaëtan Hadjeres*
Sony CSL
gaetan.hadjeres@sony.com

Cultural impact

Summary

Summary

✅ MIR enables new experiences both for listeners and creators

Summary

✅ MIR enables new experiences both for listeners and creators

✅ Autonomous music AI is still immature…

Summary

✅ MIR enables new experiences both for listeners and creators

✅ Autonomous music AI is still immature…

✅ …but it already enhances human creativity in novel ways

Summary

✅ MIR enables new experiences both for listeners and creators

✅ Autonomous music AI is still immature…

✅ …but it already enhances human creativity in novel ways

✅ Python has a rich ecosystem for working with music!

@mamodrzejewski

https://dyspensa.ai/

Thanks! 🙌

Music Information Retrieval with Python

PyStok, 29.03.2023

Mateusz Modrzejewski @mamodrzejewski