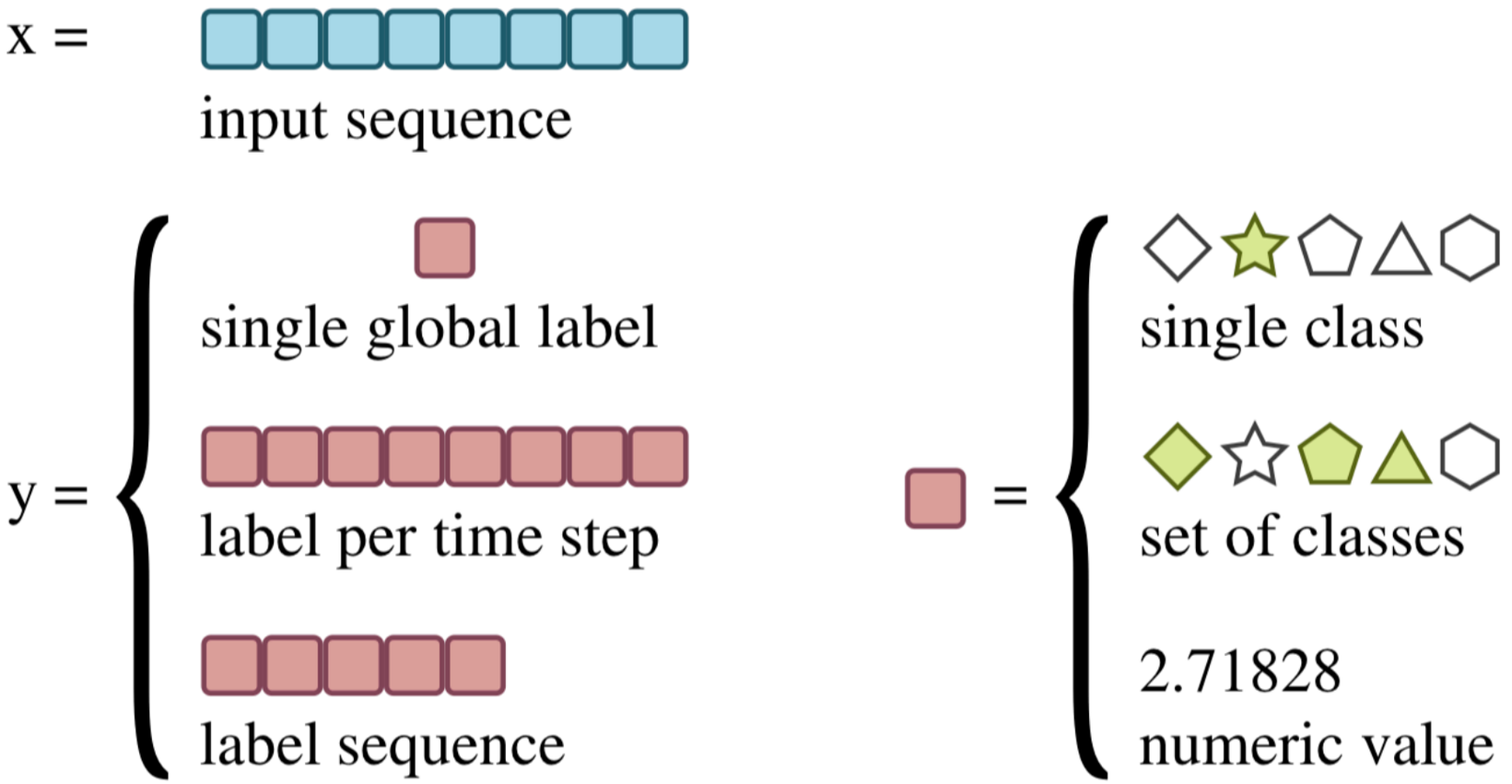
Deep Learning for Audio Signal Processing

Choose your task



Source: https://arxiv.org/pdf/1905.00078.pdf

Motivation

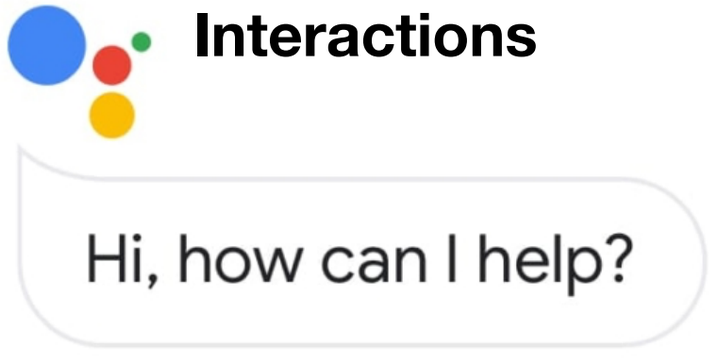
Source: https://magenta.tensorflow.org/2016/07/15/lookback-rnn-attention-rnn

Art & Fun

1. Music Generation
2. Merging Sounds
3. Sound Interpolation

Text-to-Speech

1. Voice Assistants
2. Screen Readers
3. Translation
4. More Natural Computer-Human Interactions



Source: https://podfm.ru/v-google-assistant-dobavlena-opciya-vozobnovleniya-podkasta/

Source: https://towardsdatascience.com/auto-transcribe-google-speech-api-time-offsets-in-python-7c24c4db3720

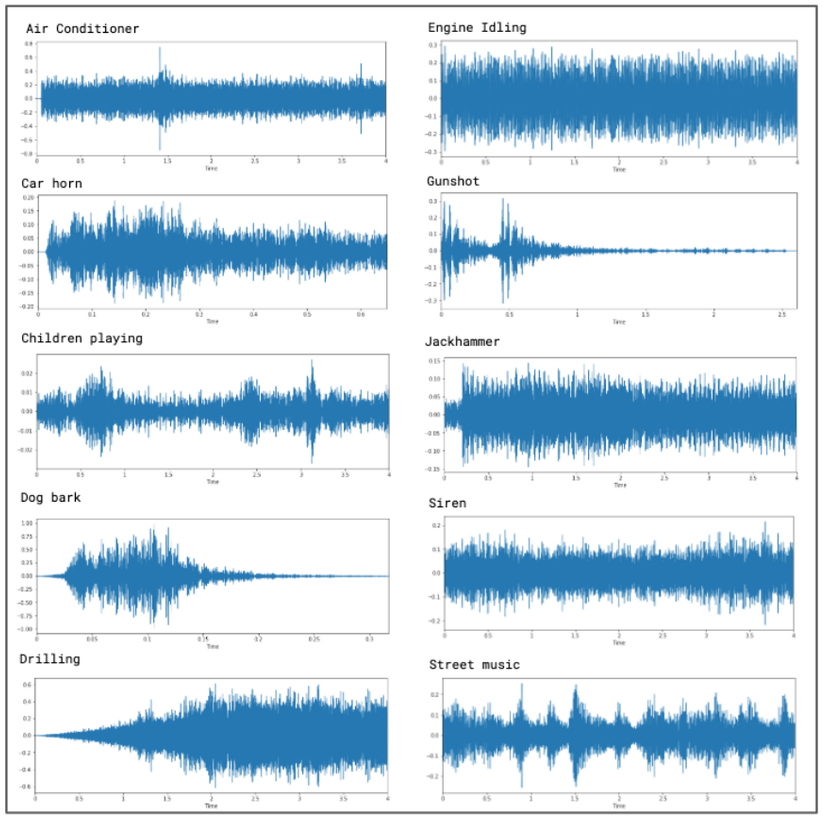
Speech Recognition

1. Natural Language Interfaces
2. Hands-free Control

Motivation

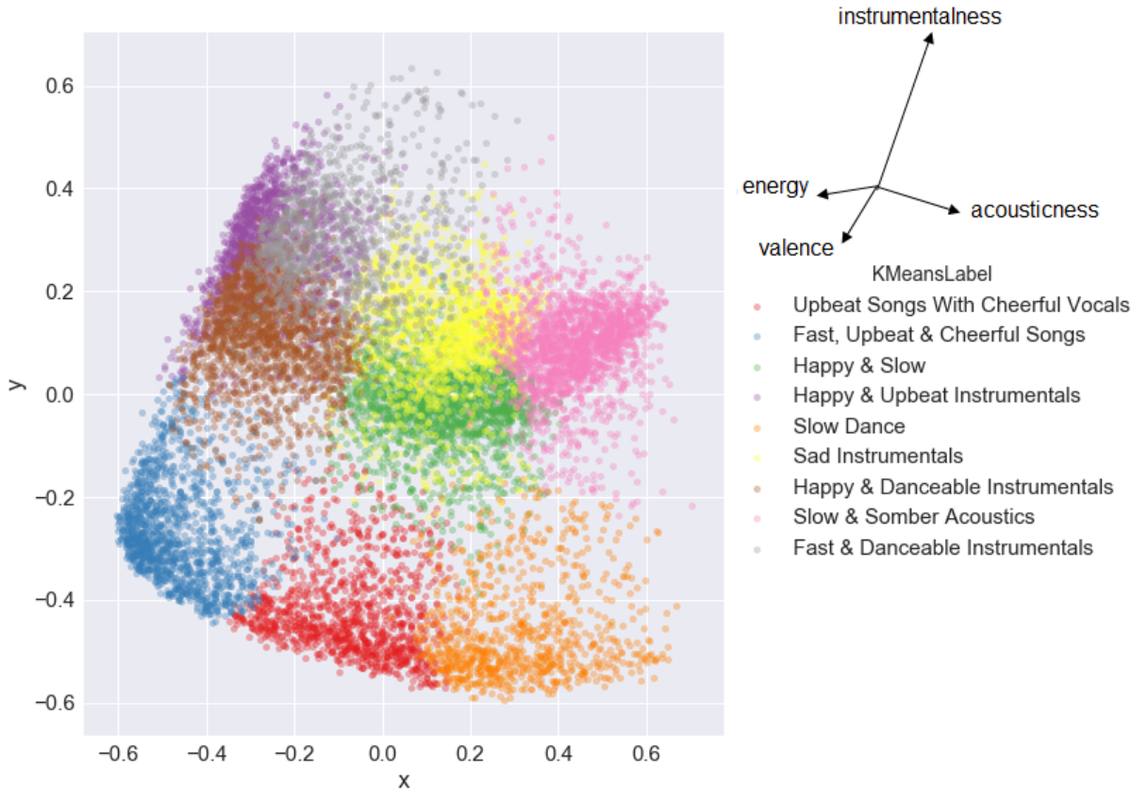
(Less Fancy)

Classification



Source: https://medium.com/@mikesmales/sound-classification-using-deep-learning-8bc2aa1990b7

Clustering

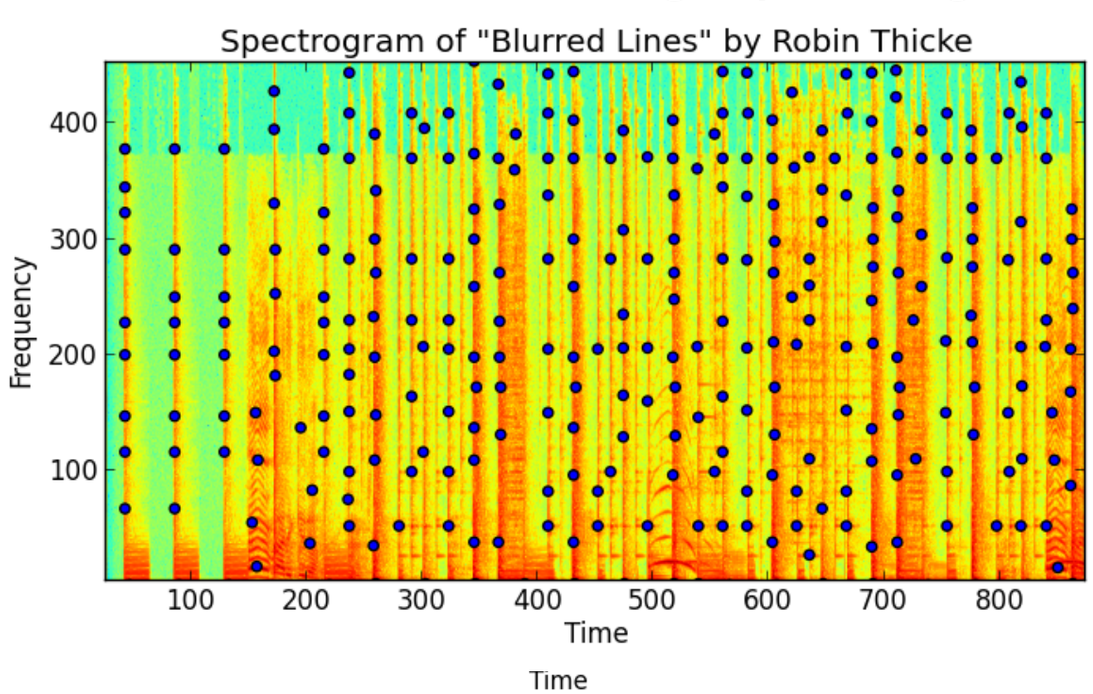


Source: https://medium.com/latinxinai/discovering-descriptive-music-genres-using-k-means-clustering-d19bdea5e443

Real world problems

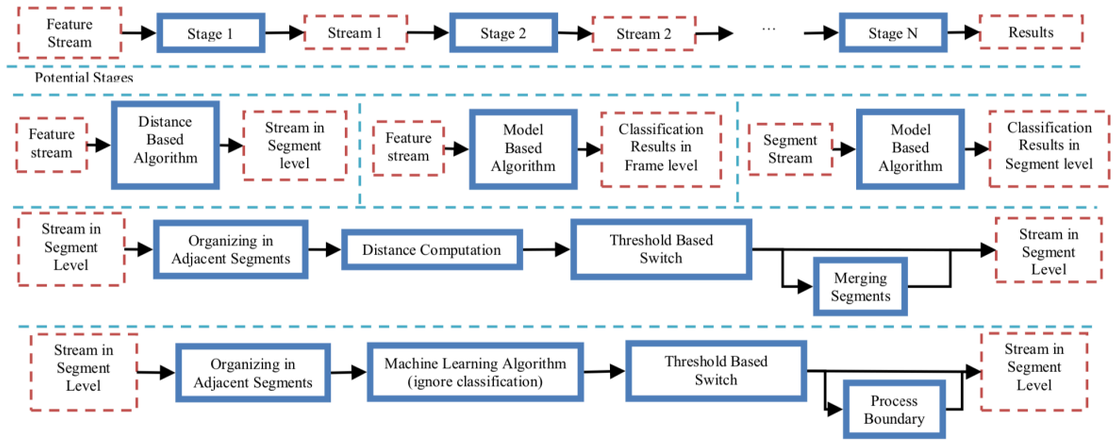
1. Audio Classification
Music Tagging

2. Audio Fingerprinting



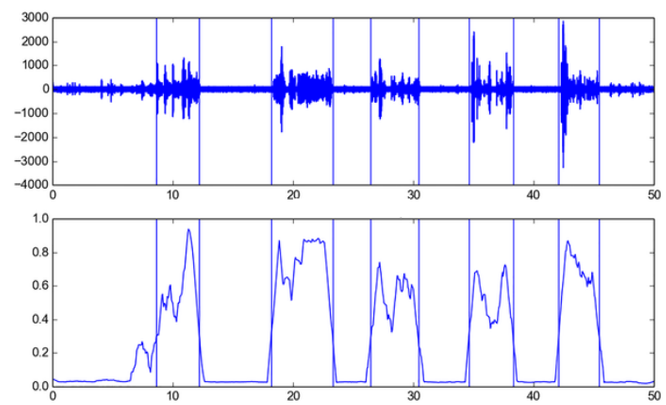
Source: https://willdrevo.com/fingerprinting-and-audio-recognition-with-python/

3. Audio Segmentation
Audio Source Separation
Beat Tracking



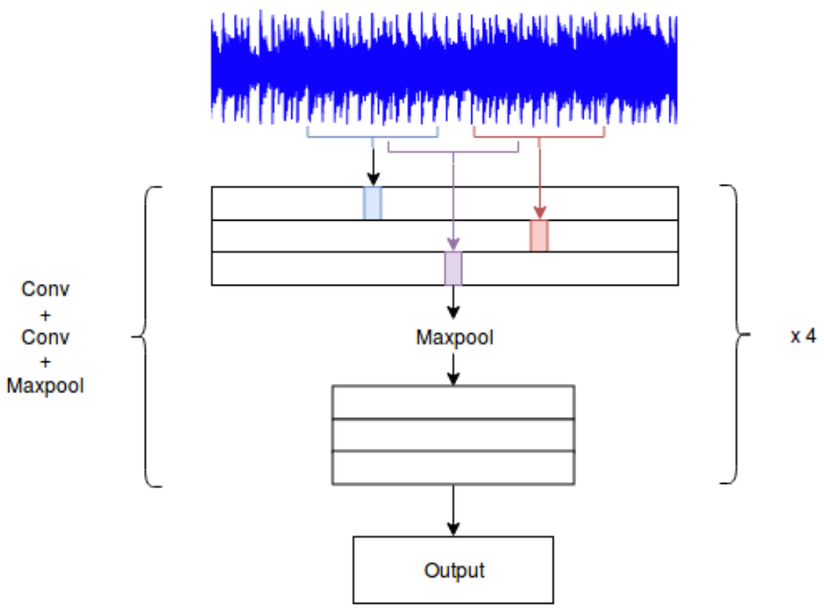
A typical usage scenario involves:

- loading an audio file
- computing a time-frequency transform to obtain a spectrogram, and
- using some of the source separation algorithm (such as non-negative matrix factorization) to obtain a time-frequency mask



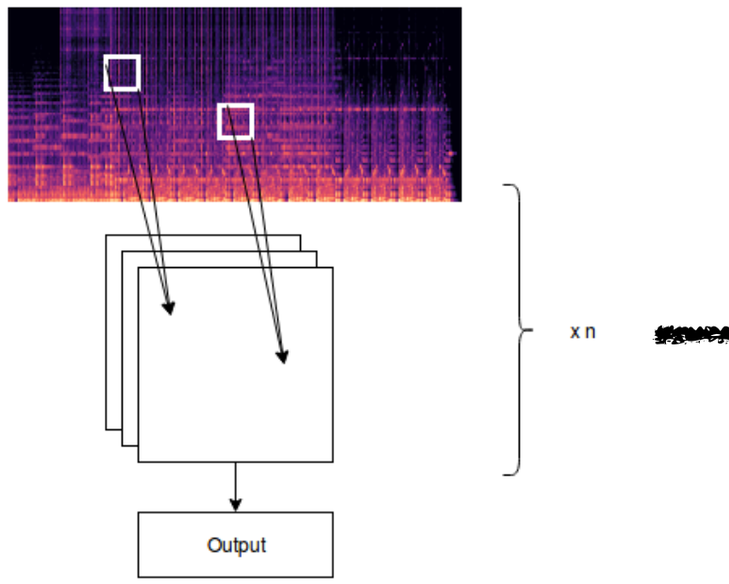
Audio Classification

Raw audio wave
1D convolutions



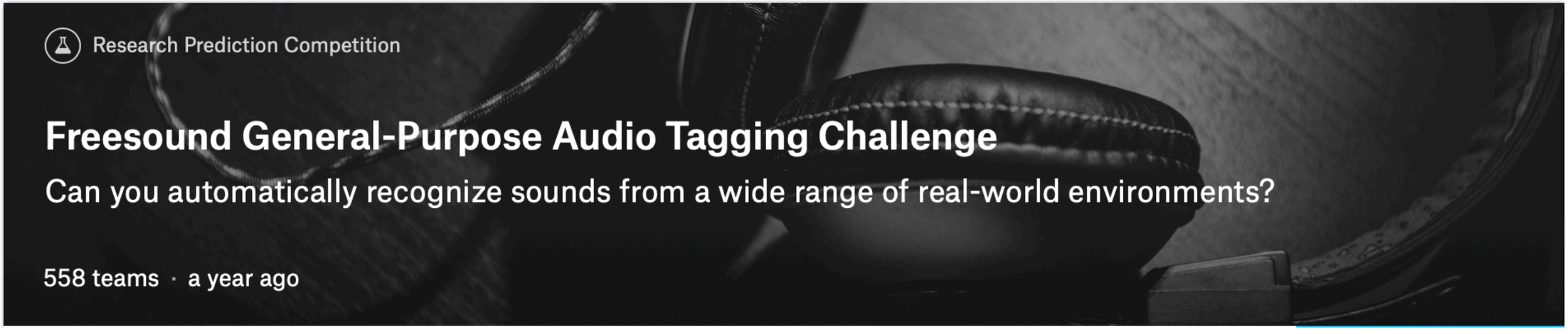
Source: https://medium.com/@CVxTz/audio-classification-a-convolutional-neural-network-approach-b0a4fce8f6c

Log-Mel spectrogram

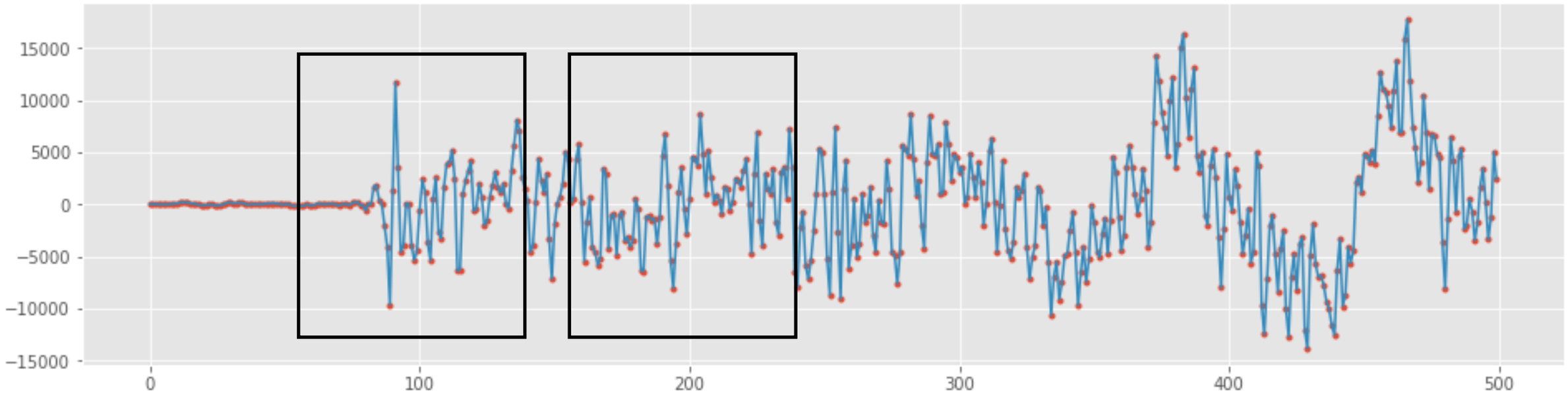


*Source: https://miro.medium.com/max/1044/1*N9cxrH0nMlohogWal3UcfQ.png*

Stack More Models



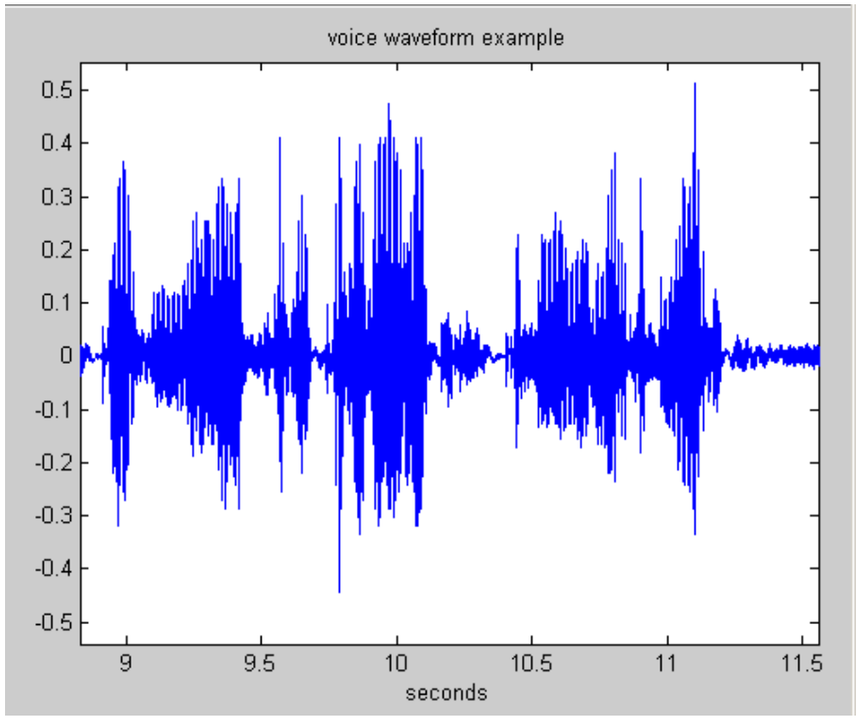
Spectograms & MFCC

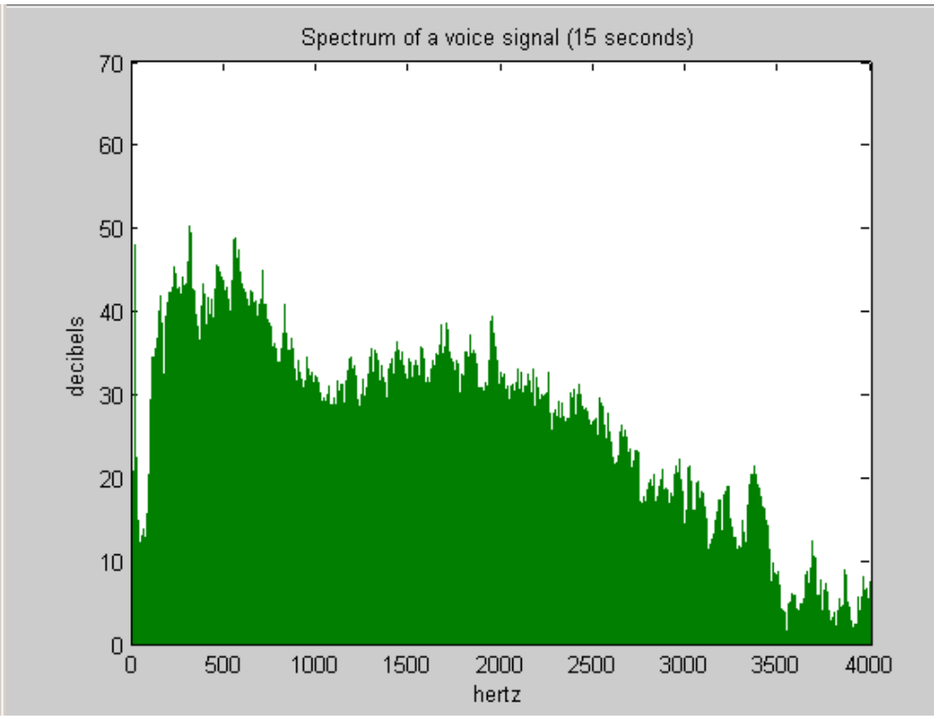


Source: https://www.kaggle.com/fizzbuzz/beginner-s-guide-to-audio-data

FFT **FFT**

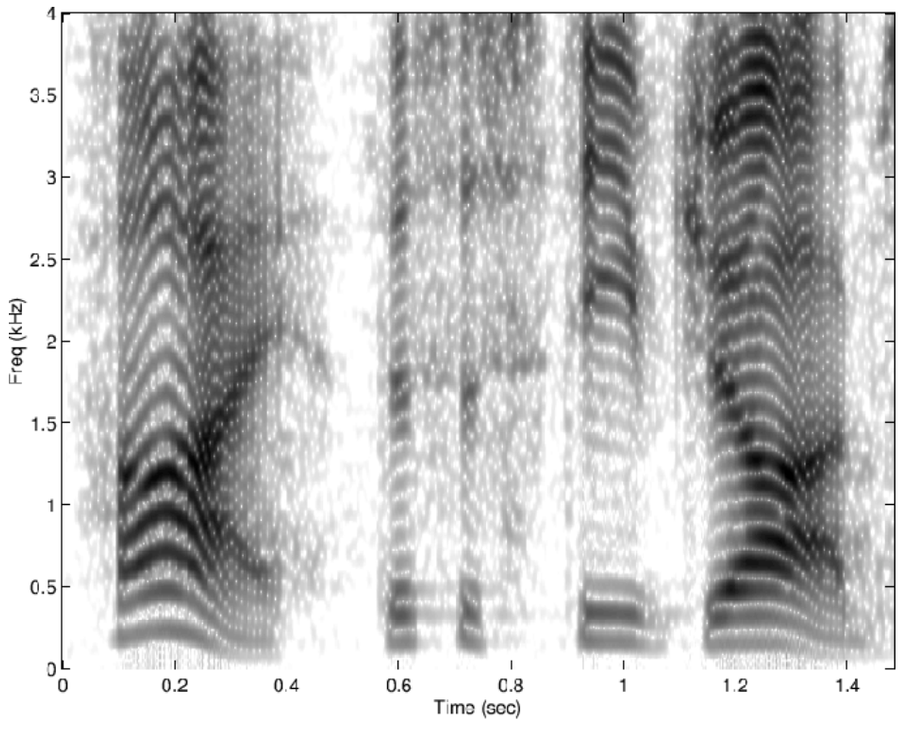
Spectrum



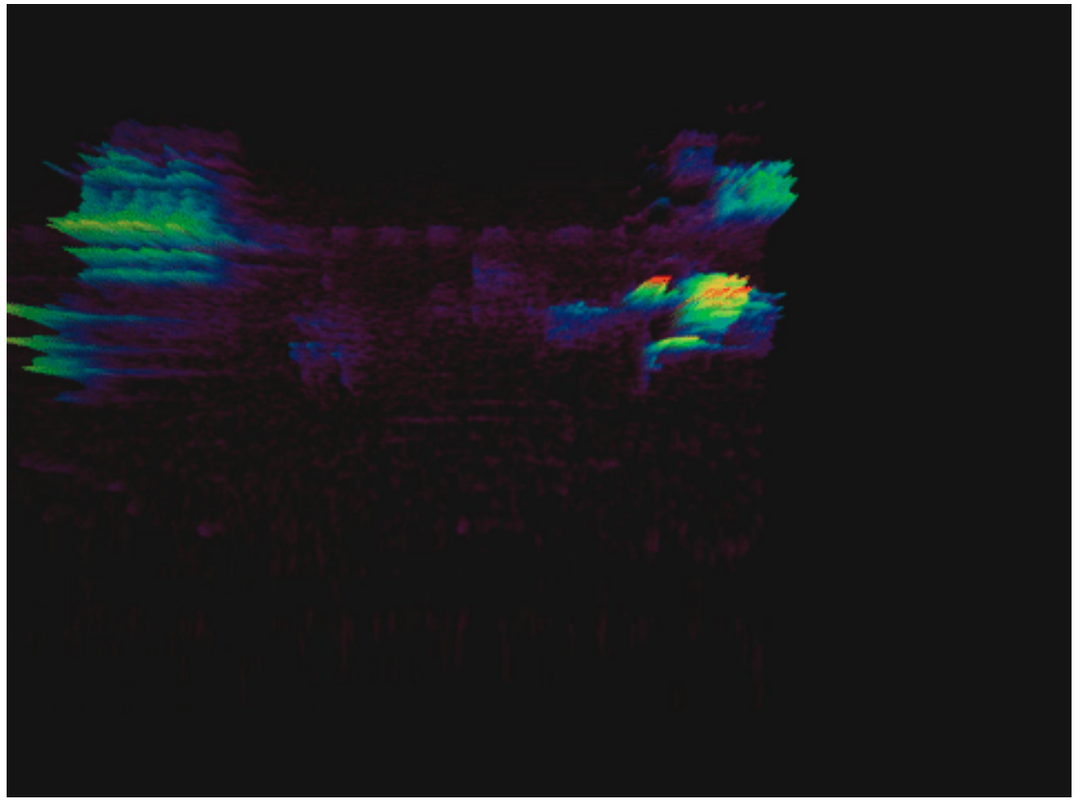


Source: https://habr.com/ru/post/226143/

MAP spectral amplitude to a grey level (0- 255) value.
0 represents black and 255 represents white.
Higher the amplitude, darker the corresponding region.



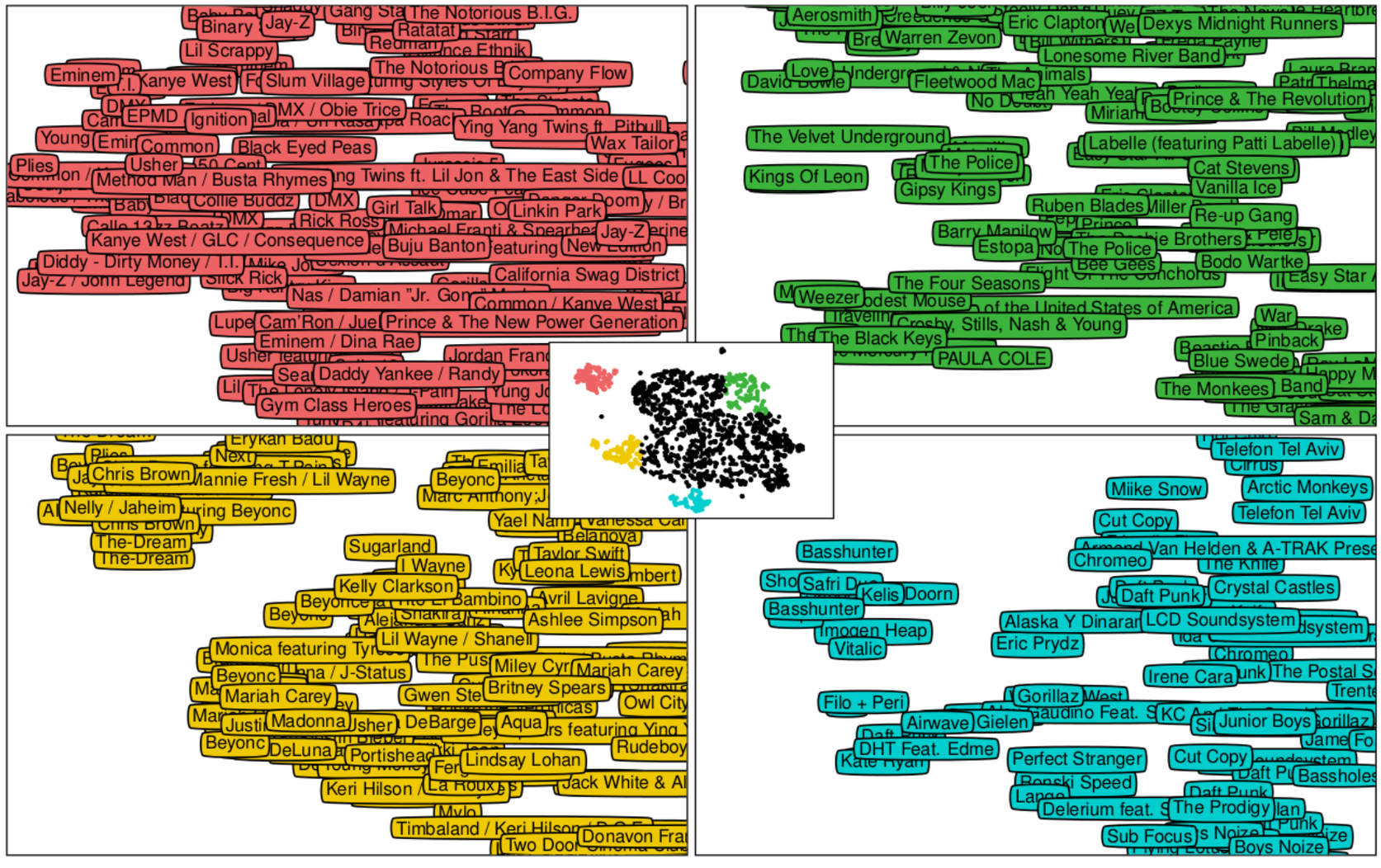
Source: https://www.dsprelated.com/freebooks/sasp/Spectrogram_Speech.html



Source: https://xakep.ru/wp-content/uploads/2019/09/236259/bird.gif

Real world problems

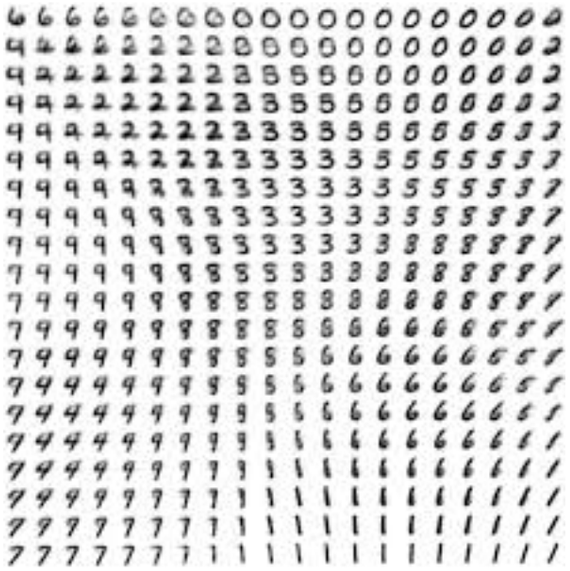
7. Music Recommendation
Collaborative filtering
Content-based recommendation



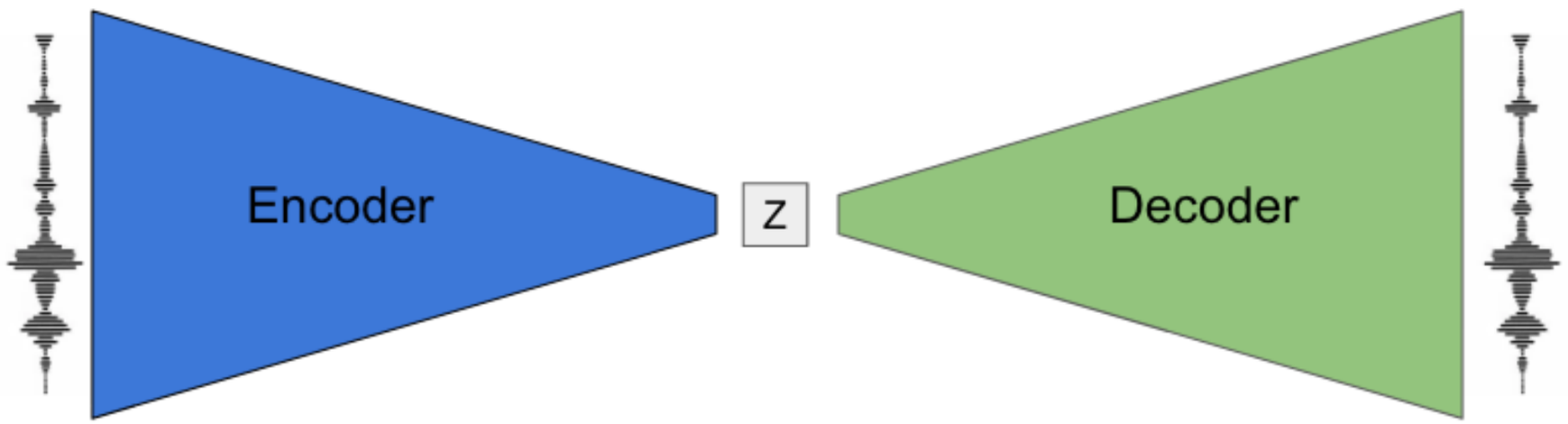
Source: http://benanne.github.io/2014/08/05/spotify-cnns.html

Working with music

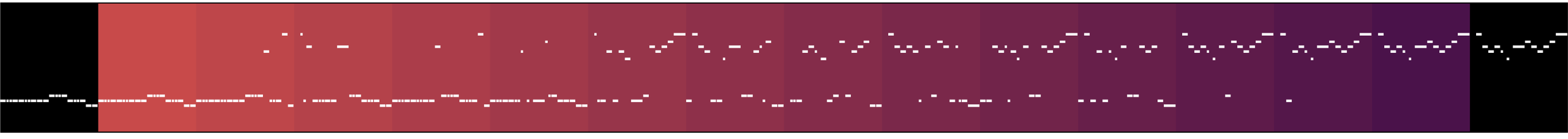
MusicVAE



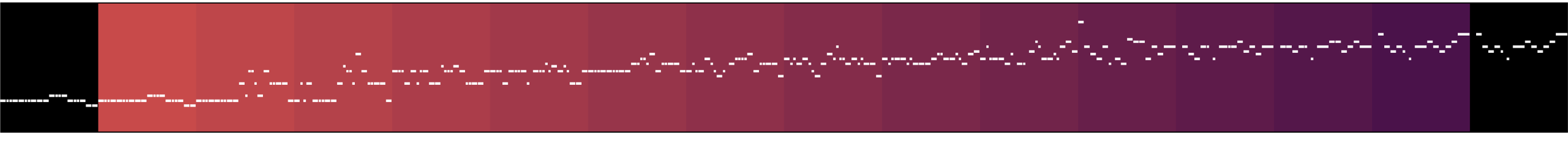
Source: https://www.researchgate.net/figure/VAE-learned-manifold-of-the-MNIST-dataset-A-VAE-with-a-two-dimensional-latent-space-was_fig11_332547811



Source: https://magenta.tensorflow.org/music-vae

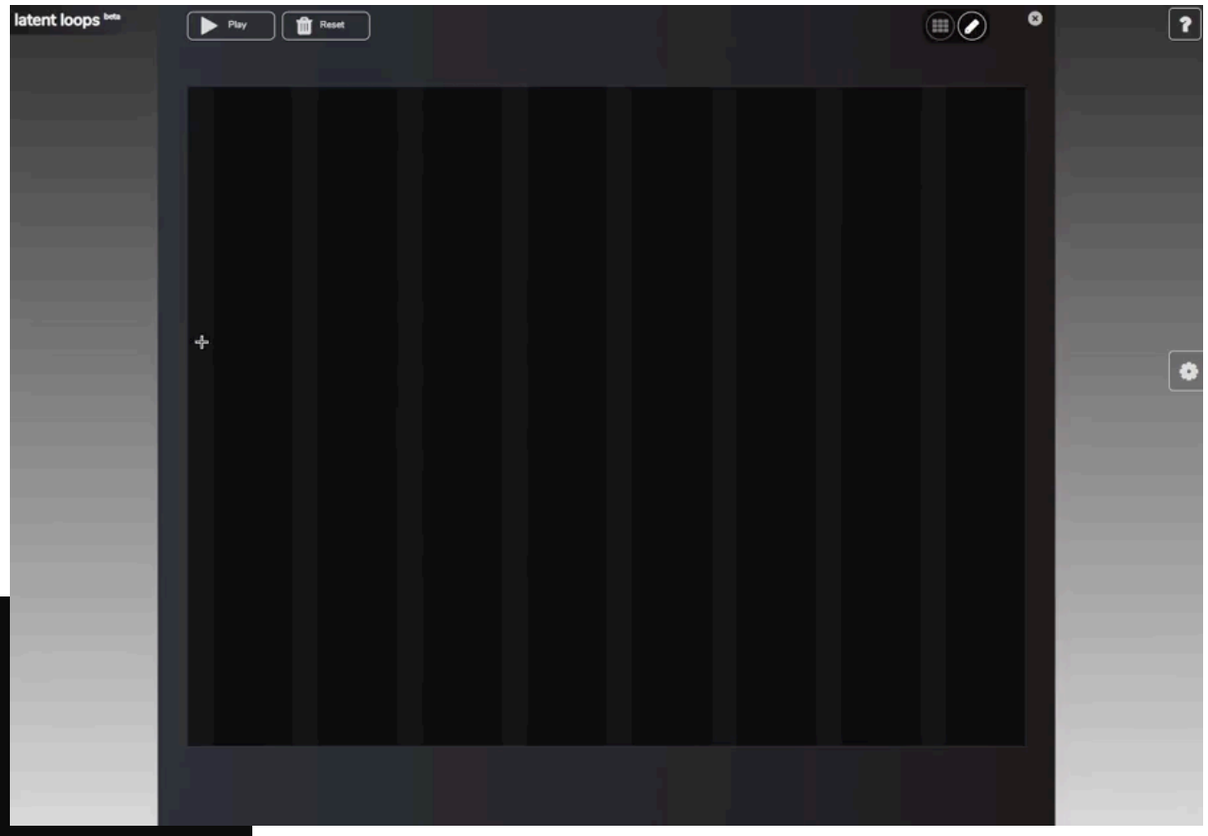


NOT VAE

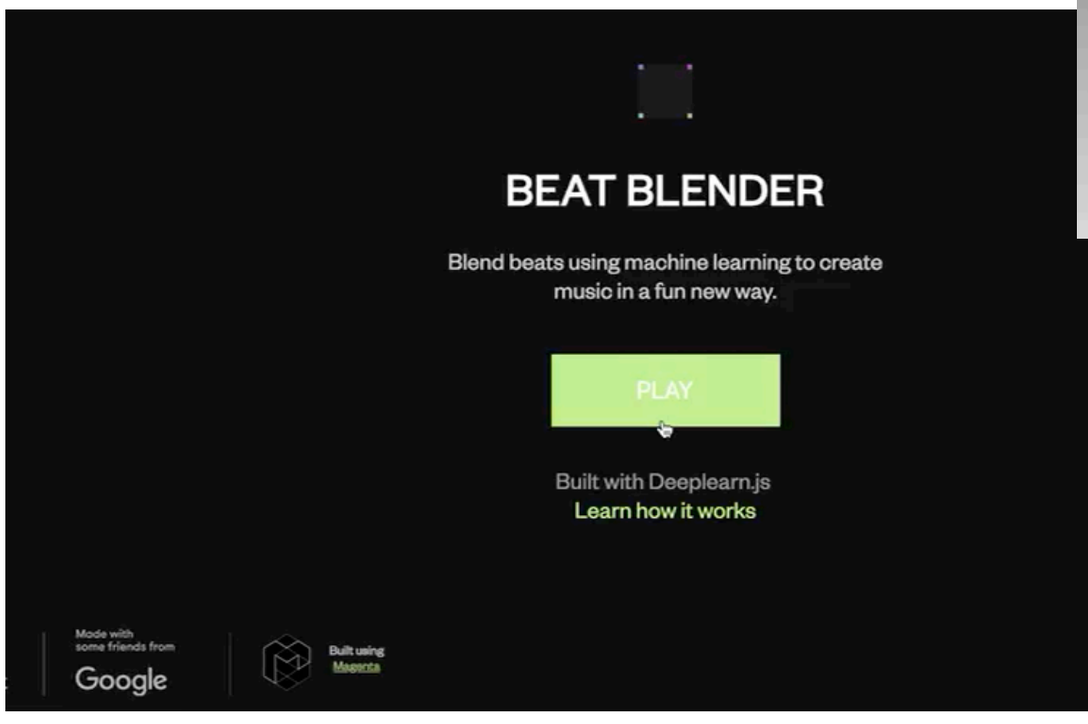


VAE

Tools for musicians



Latent Loops



Beat Blender

NSynth

moving to the actual sound…