Artificial Intelligence and Creativity in Music

Jérémy Cohen
CNRS, CREATIS,
Université Claude Bernard,
Lyon, France



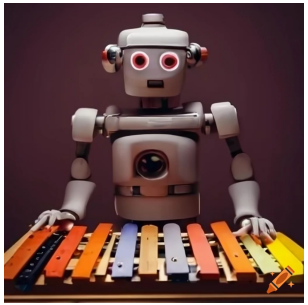
Generative AI

- Generative AI is state-of-the-art for text and image problems.
  - ChatGPT, Le Chat, ...
  - Sora, (among many...), Crayon (lightweight)

Example (2024):
music and AI







- What about music ?
  - 2024: Examples from MusicGen by (Meta AI)

Generative AI for music is evolving fast !

- 2025: Industry has taken over ! Example: Suno



Generate commercial-sounding music from:
- A text prompt describing the music
- An audio recording (voice, instruments)
- Some lyrics

Really good on the vocal part (writes lyrics, realistic sounding voices).

Latest evolutions: give the "artist" more control.

Gen AI: Boosting creativity vs stealing jobs

+ Lowers the technical barrier to start making music. More people can enjoy the creative process.
+ Generate new ideas, AI and artists can work together.
+ ?

- Issue with copyright infringement (many audio AI companies are under lawsuits by major labels).
- Producers replaced by AI ?
- ?

Note: I am not an expert on AI ethics.

1. There are many AI tools beside generative AI.

2. Artist have already started to make music with AI.

A competition on AI music production

Since 2020 in Netherlands exists an international contest on music production with AI !



A competition on AI music production

Winner 2022

Artist: **Yaboi Hanoi**
Title: อสุระเทวะชุมนุม - Enter Demons & Gods (“Asura Deva Choom Noom”)

Audio and explanations: [link]



Winner 2024

Artist: **Onda Corta**
Title: Sudamérica

Audio and explanations: [link]



Song contest: what use of AI ?

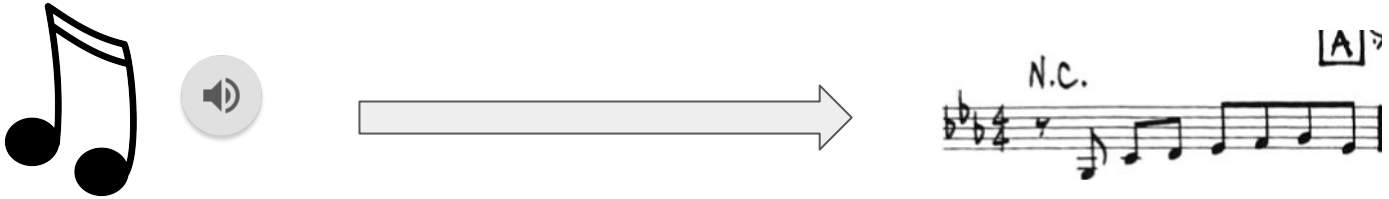
- Generative AI
  - Lyrics
  - Text to singing voice [ElevenLabs]
  - Drum patterns [demo]
  - Arranging/ideas

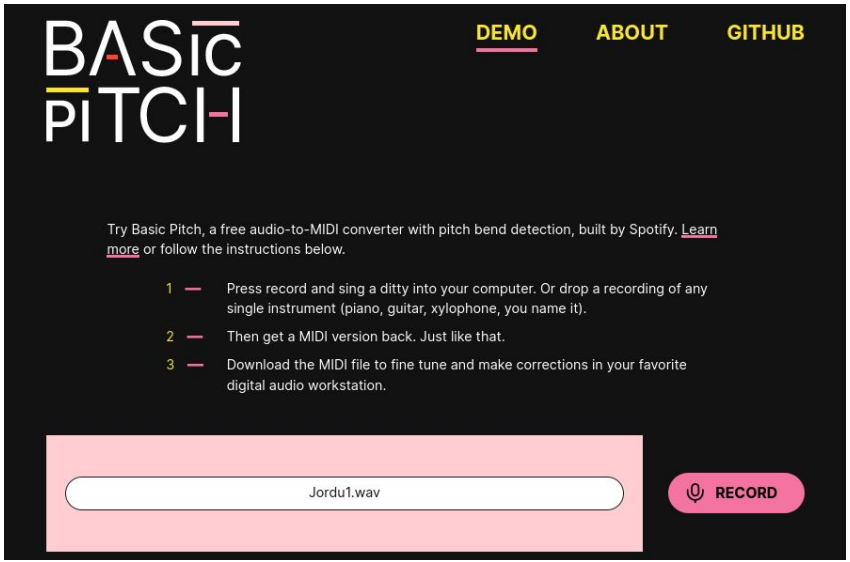
- Specific tools
  - **Pitch detection**
  - **Tone Transfer**
  - Stem separation [Moises]
  - Mixing/mastering
  - Text to texture
  - Singing voice improvement
  - Synth patch generation

A subset of tools used

- ChatGPT
- Claude
- Gemini
- AIDrummachine
- ElevenLabs
- Moises
- Mofi
- ToneTransfer
- BasicPitch
- Magenta VST
- Musenet
- AIVA
- This patch does not exist
- ...

Automatic pitch detection ?







Tone Transfer ??

INPUT

Acapella singing

Birds chirping

Carnatic singing

Cello performing

Pots and pans clanging

Synthesizer riffing

+ Add your own



TRANSFORMATION

Concert

None

Flute

Saxophone

Trumpet

Violin

The roadmap: explaining two AI tools

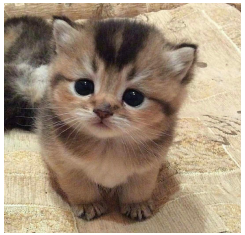
1. Machine learning concepts for music information retrieval
2. Automatic pitch detection: Basic Pitch
3. Tone Transfer: DDSP

AI for the beginner

AI in science has several ambiguous meaning. We rather talk about **machine learning**.

Machine Learning: use a computer to analyse data.

Supervised Learning



Cat

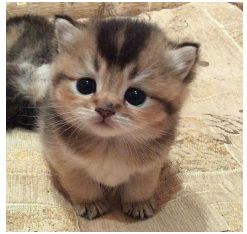


Dog



??

Unsupervised Learning



??



??



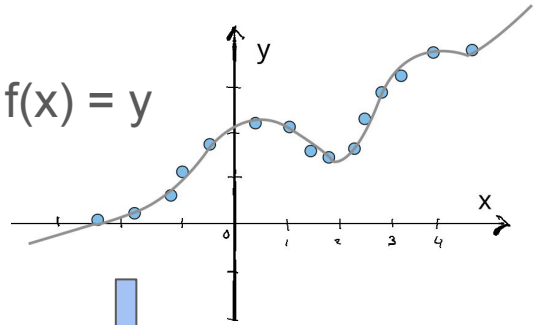
??

3 pillars of supervised learning

1. Training Data

2. Model

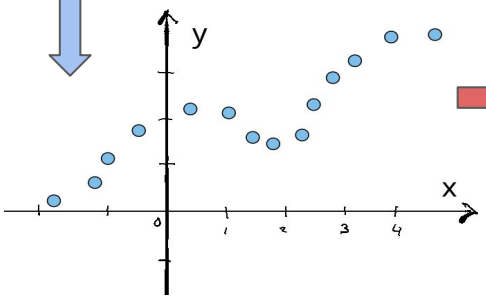
3. Training

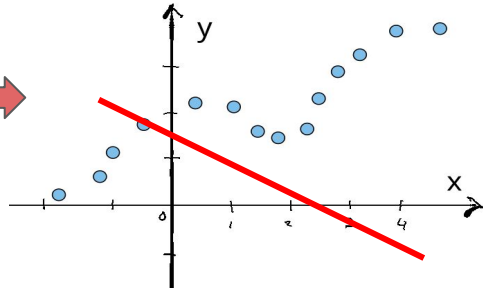


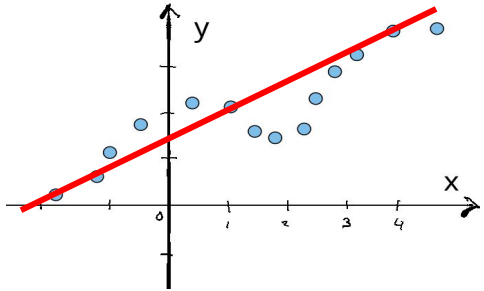
Exemple: linear regression
Learn a function from data samples

model(x; a, b) = a.x+b

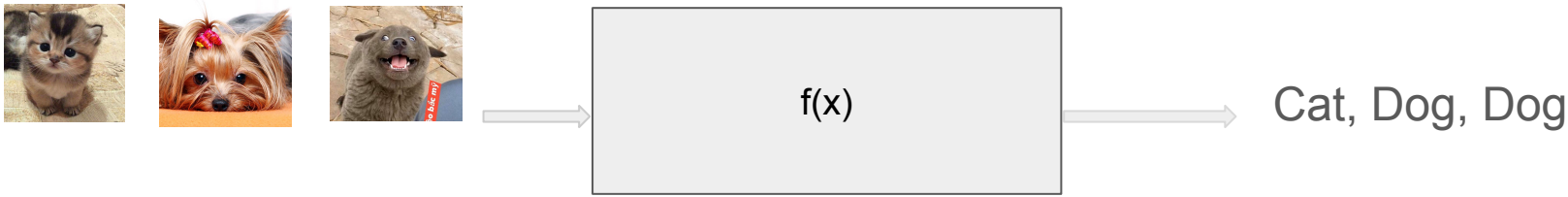
model(x; a', b') = a'.x+b' ≅ f(x)



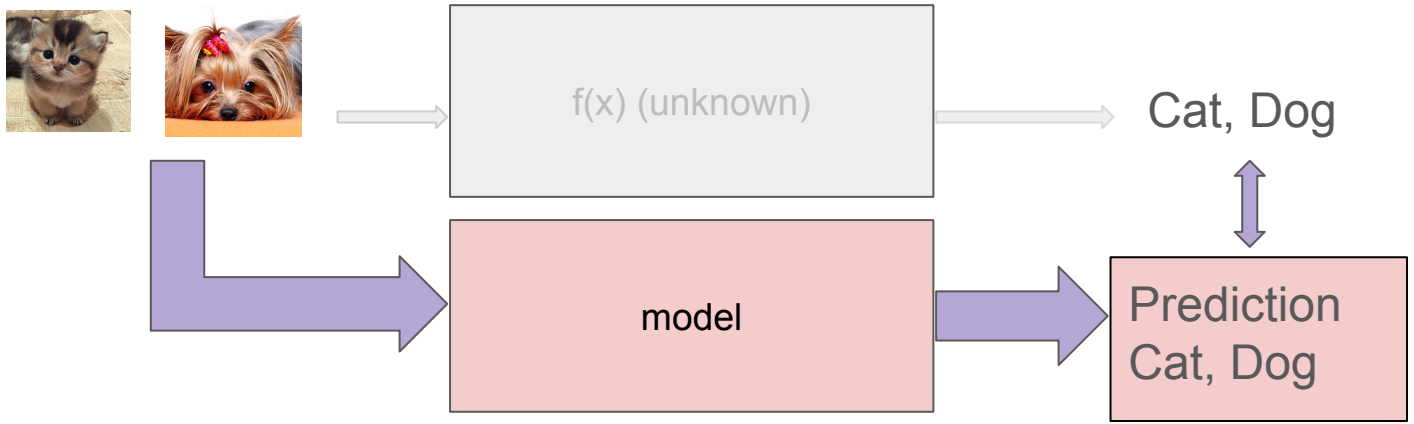




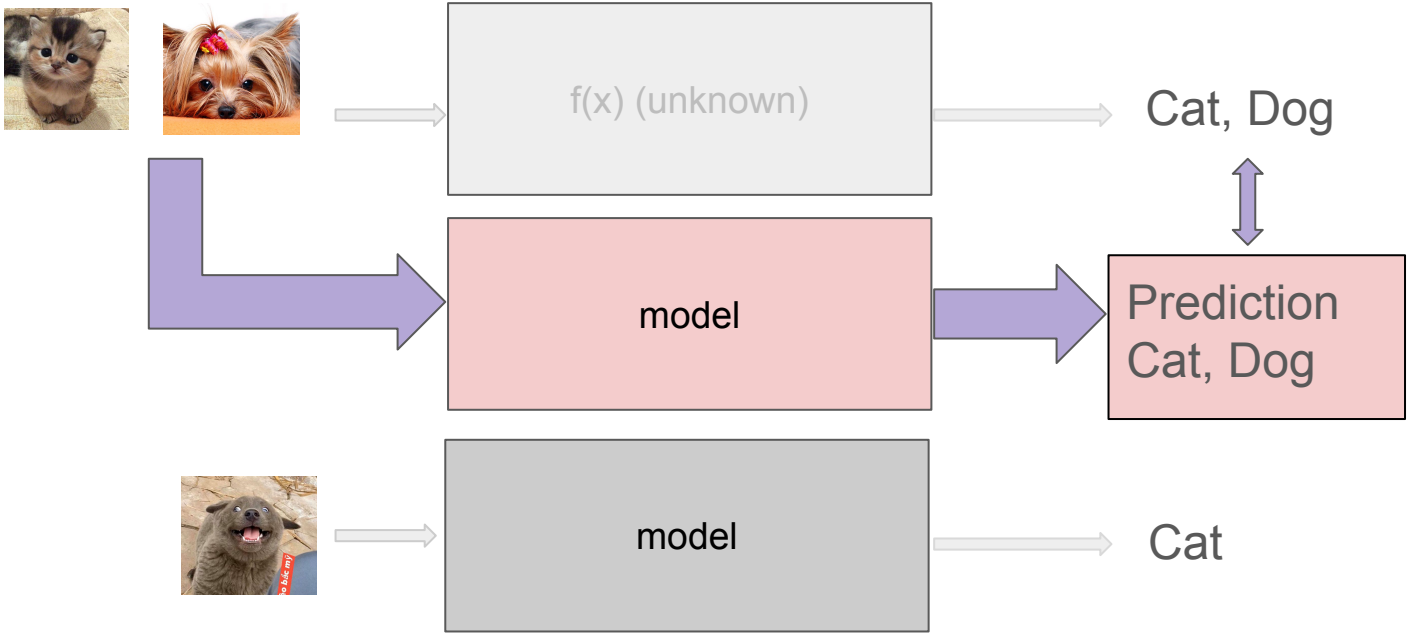
Machine Learning: classification



Machine Learning: classification

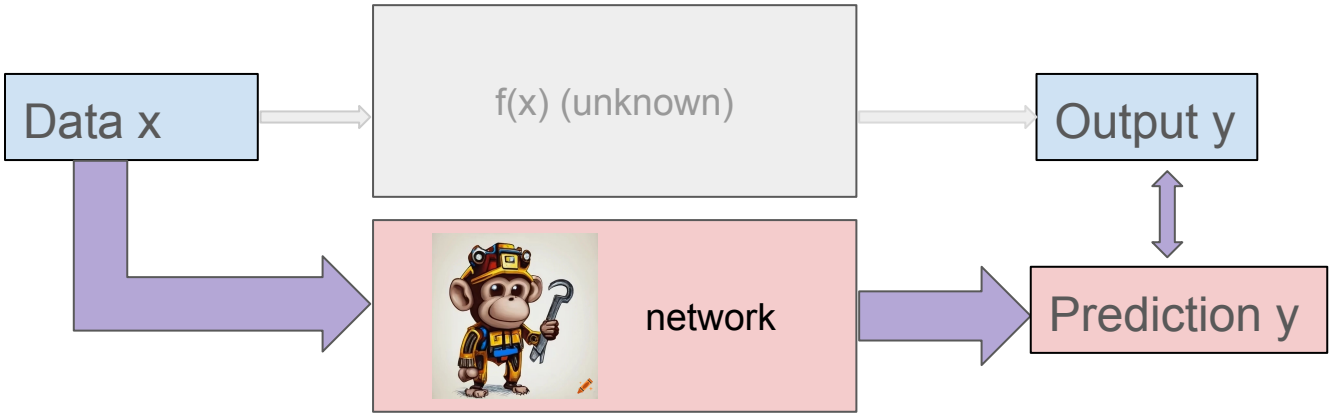


Machine Learning: classification



Neural Networks, a multifunction tool

Neural Network = Generic non-linear model

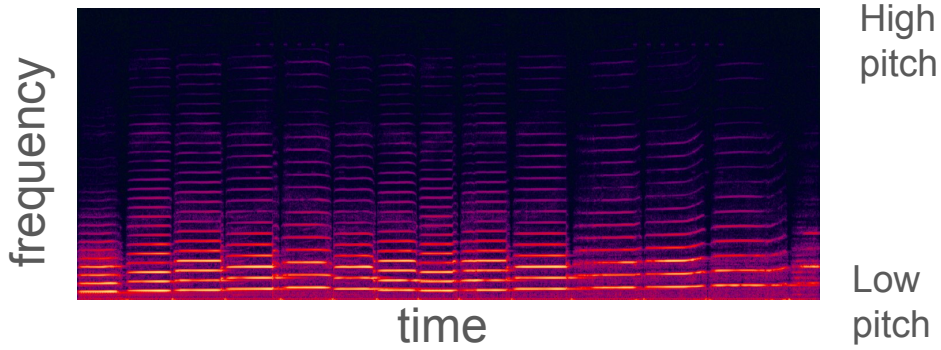


Once trained,  can approximately reproduce function f.

Data in music

- What is "x":
  - Temporal audio signal
  - Spectrogram: frequencies heard at each time instant

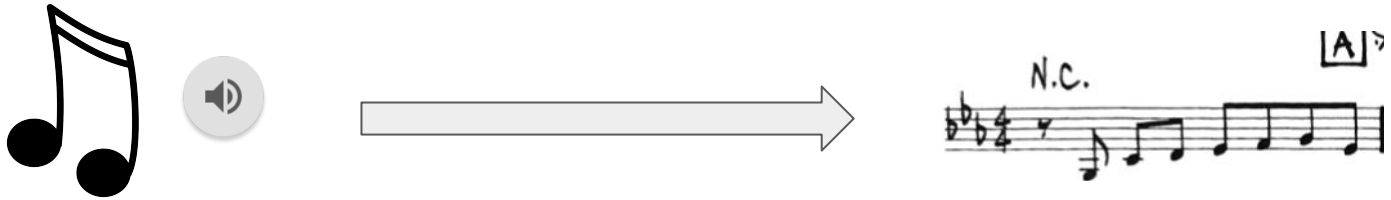
[Demo]

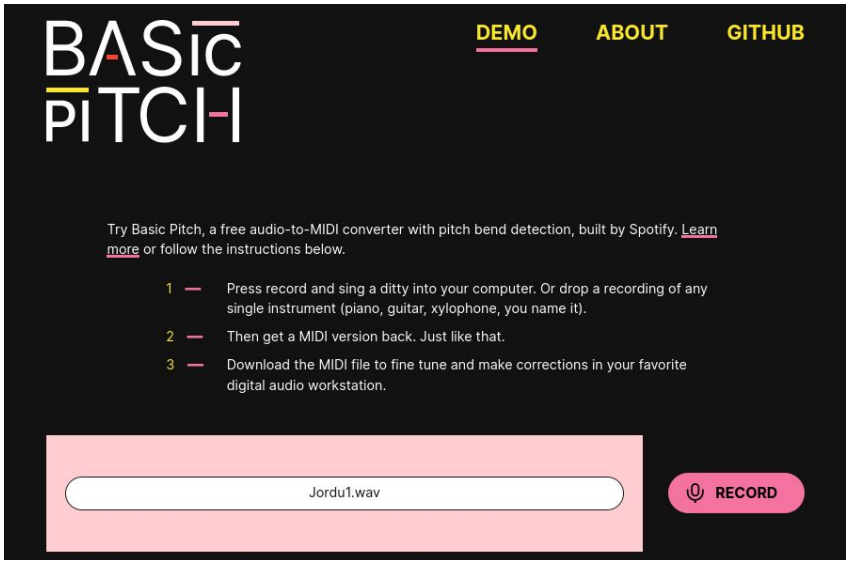


- What is "y":
  - Played notes (end to end pitch detection)
  - A spectrogram or an audio signal
  - …

Let's dive into the actual AI tools

Automatic pitch detection (reminder)







Pitch detection: challenges

- Multi pitch instruments (guitar, piano)
- Handle the various timber and recording conditions (reverb)
- Pitch bend (singing voice, wind instruments)
- Detecting when the note is played on/off (onset, offset)

Dozen of scientific articles on that topic, and as many algorithms.

The (probably) most famous: Yin algorithm (no AI)
(2002, Alain de Cheveigné, IRCAM, and Hideki Kawahara, Wakayama University)

Basic Pitch



Published at the ICASSP 2022 conference

"A Lightweight Instrument-Agnostic Model for Polyphonic Note Transcription and Multipitch Estimation", Rachel M. Bittner et. al.

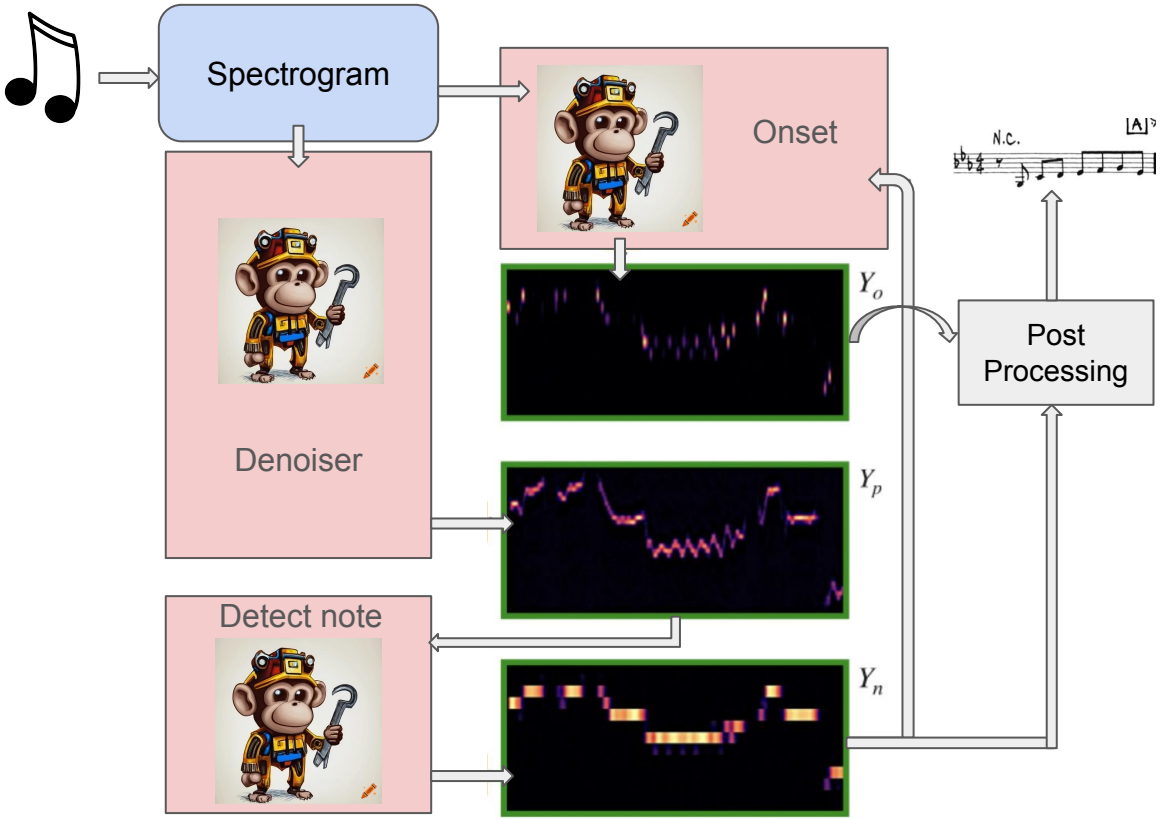
Spotify®

Main idea: decompose the problem into three complementary tasks

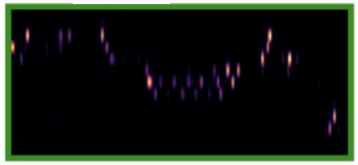
1. Detect onsets (attacks only) → (Yo)
2. Detect if a note is being played → (Yn)
3. Finely denoise the spectrogram → (Yp)

and a twist on the spectrogram (HCQT).

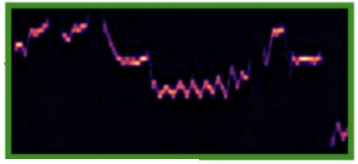
Basic Pitch: details



Basic Pitch: Postprocessing



Y_o



Y_p



Y_n

Note Segmentation

How easily a note should be split into two.

Split Notes — Merge Notes

0.50

Model Confidence Threshold

The model confidence required to create a note.

More Notes — Fewer Notes

0.30

Minimum Pitch

The minimum allowed pitch, in Hz.

Lower notes — Higher notes

0

Maximum Pitch

The maximum allowed pitch, in Hz.

Lower notes — Higher notes

3000

Minimum Note Length

The minimum length required to emit a note, in milliseconds.

Short Notes — Long Notes

11

MIDI File Tempo

The tempo used to export the MIDI file.

120

Basic Pitch: use cases ?

- **To re-synthesized sounds**
  The fundamental frequency F0 can trigger another instrument !

- For musicien work:
  - Hear training
  - Voice training

- … other ideas ?

There are many other tools (Melodyne in particular), but this one is free and we know all the science behind. Also some AI tools (Samplab, AIMIDI).

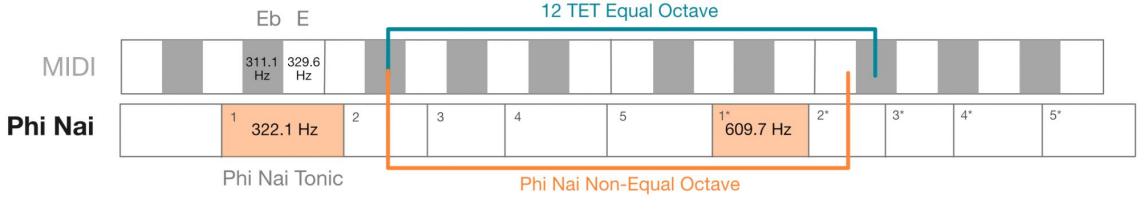
Not yet one big AI company that has completely cracked this yet !!

AI usage by Yaboi Hanoi

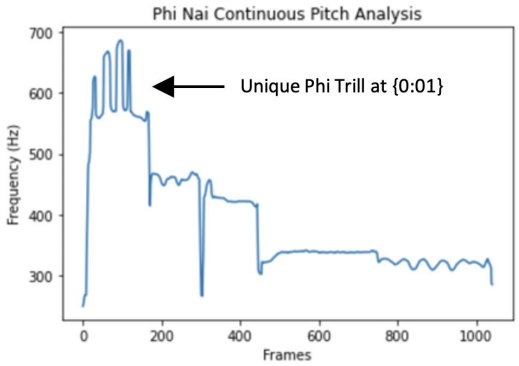


AI usage by Yaboi Hanoi





F0 tracking



Tone Transfer







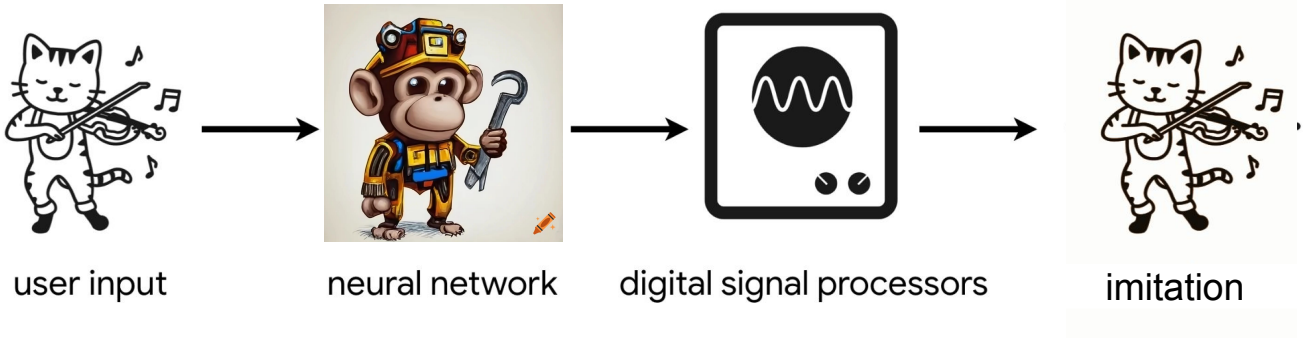
Before ToneTransfer, DDSP

Published in conference ICLR 2020

"DDSP: Differentiable Digital Signal Processing",
Jesse Engel et. al.

magenta

Main idea: Have a neural network play a synthesizer to reproduce audio !



ToneTransfer: training





Source Audio

Pitch tracking (CREPE)

F0

Power (RMS)



F0 (n=1)

Harmonic Distribution (n=60)

Amplitude (n=1)

Frequency Response (n=65)

Harmonic Synth

Filtered Noise Synth

Reverb

Synth Audio

Multi-Scale Spectrogram Loss

ToneTransfer: training





Source Audio

Pitch tracking (CREPE)

F0

Power (RMS)





Reverb

Synth Audio

Multi-Scale Spectrogram Loss

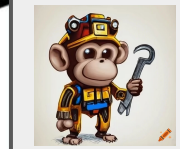
ToneTransfer: usage



Pitch tracking (CREPE)

F0

Power (RMS)





Reverb

Synth Audio



Music and AI, many tools for creativity

BASIC PITCH

ddsp

TONE TRANSFER

Moises,
Stem separation

magenta

Magenta Studio (VST)



This patch does not exist

SUNO

and many others

Eleven Labs

Some final thoughts

- Music AI used to be mainly driven by research groups, now by industry

    + Tools are getting significantly better
    - We have no technical details on how these tools work (but we know the general idea)

- Some research topics are still out of reach for AI startups (transcription !)

- Most existing tools stem from a few research group, in particular Google's Magenta !

- Generative AI is mainly focusing on **Western commercial music** with its caveats (simple harmonies, hyper-compressed sound, equal temperament, piano/guitar/bass/drums…)

→ follow [Yaboi Hanoi] for music/talks on Thai AI.

Thank you for your attention !

Any question?

Find these slides on my personal webpage: https://jeremy-e-cohen.jimdofree.com/