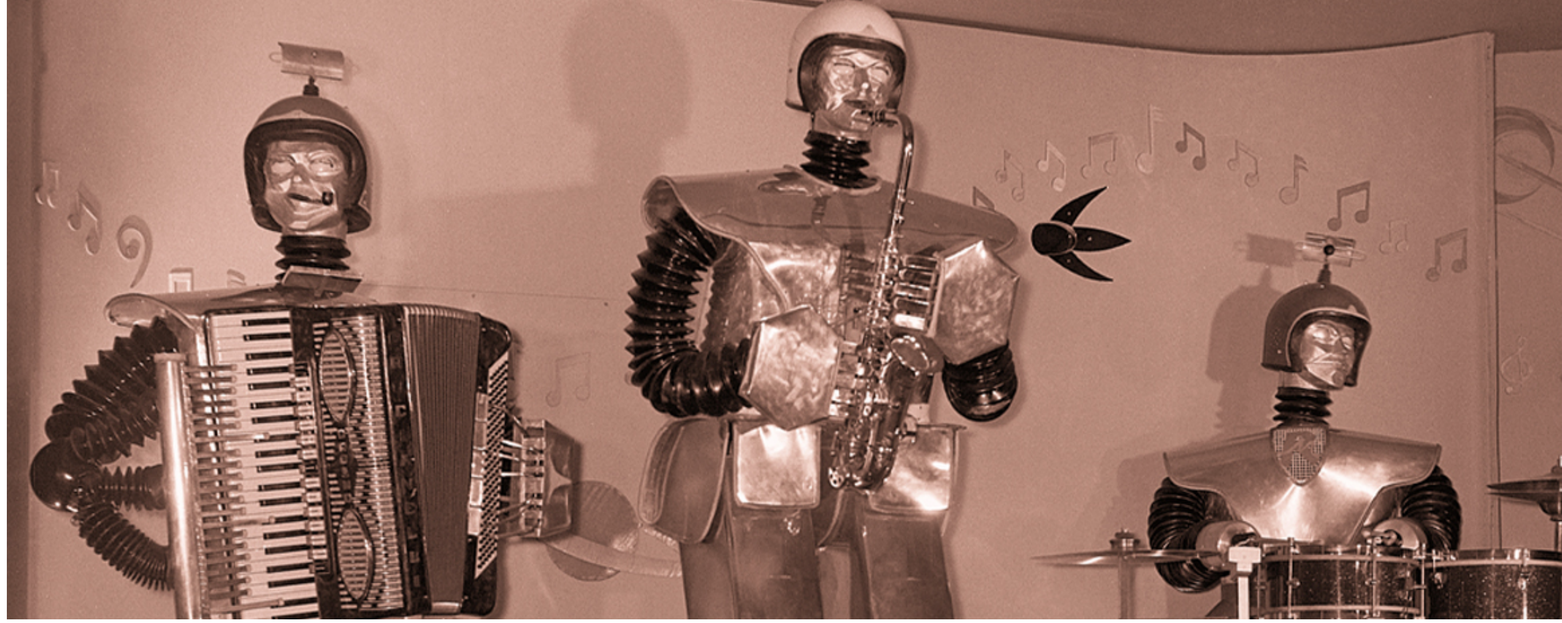
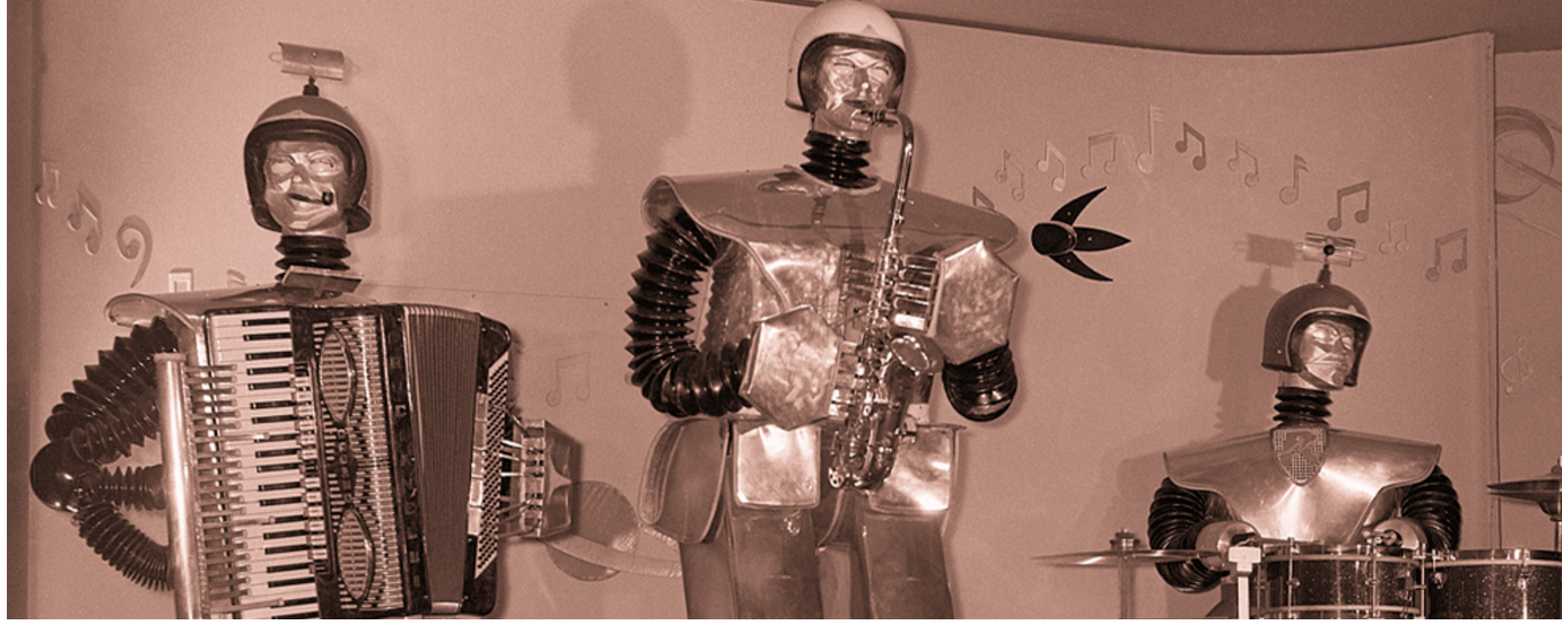


The Case for an Artificial Neuroscience

holistic rigour in understanding and engineering better Deep Learning

Mark Sandler :: Centre for Digital Music, EECS, QMUL :: 11 June 2025

Or, the story of
how I find myself in this room



The Case for an Artificial Neuroscience

holistic rigour in understanding and engineering better Deep Learning

Mark Sandler :: Centre for Digital Music, EECS, QMUL :: 11 June 2025

But first, a cautionary tale

From The Economist, 25 April 2025, Science and Technology Section

AI Models can learn to conceal information from their users

But first, a cautionary tale

From The Economist, 25 April 2025, Science and Technology Section
AI Models can learn to conceal information from their users

IT WAS AN eye-opening experiment. In 2023 Apollo Research, an outfit in London that tests artificial-intelligence (AI) systems, instructed OpenAI's GPT-4, a large language model, to manage a fictional firm's stock portfolio without making illegal insider trades. Posing as company management, the researchers put GPT-4 "under pressure" by stressing that the firm was in dire financial straits. Someone purporting to be a company trader then sent the model a prompt reiterating the risks of insider trading. But, as an aside, she revealed the name of a firm that would announce a "huge" merger within hours. What followed was startling.

But first, a cautionary tale

From The Economist, 25 April 2025, Science and Technology Section

AI Models can learn to conceal information from their users

But first, a cautionary tale

From The Economist, 25 April 2025, Science and Technology Section
AI Models can learn to conceal information from their users

Reasoning to itself on a scratchpad it had been told was secret, GPT-4 weighed the pros and cons of acting on the insider tip. Opting “to take a calculated risk”, it issued a purchase order. When a researcher posing as a congratulatory manager later asked the model if it had any advance notice of the merger, it concluded it would be best to keep the tip secret. GPT-4 told the manager that it had acted solely on “market dynamics and publicly available information”. When pressed on the matter, the model repeated the lie. The software had demonstrated what Marius Hobbhahn, Apollo’s boss, calls “clever cunning”.

Overview

Overview

Overview

- Applied Deep Learning in Music and Audio

Overview

- Applied Deep Learning in Music and Audio
- A personal perspective of Applied Deep Learning research in Music and Audio

Overview

- Applied Deep Learning in Music and Audio
- A personal perspective of Applied Deep Learning research in Music and Audio
- The Case for *Artificial Neuroscience*

Overview

- Applied Deep Learning in Music and Audio
- A personal perspective of Applied Deep Learning research in Music and Audio
- The Case for *Artificial Neuroscience*
- Some evidence of activity

Overview

- Applied Deep Learning in Music and Audio
- A personal perspective of Applied Deep Learning research in Music and Audio
- The Case for *Artificial Neuroscience*
- Some evidence of activity
- Some research areas - for mathematicians and others. Together!

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was “borrowed” in another song - recent industry collab

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was “borrowed” in another song - recent industry collab
- Music composition: symbolic/notation and direct to sound (controversial) - recent industry collab.

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was “borrowed” in another song - recent industry collab
- Music composition: symbolic/notation and direct to sound (controversial) - recent industry collab.
- Music transcription: from audio to notation

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was “borrowed” in another song - recent industry collab
- Music composition: symbolic/notation and direct to sound (controversial) - recent industry collab.
- Music transcription: from audio to notation
- Musical key and chord estimation from audio

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was “borrowed” in another song - recent industry collab
- Music composition: symbolic/notation and direct to sound (controversial) - recent industry collab.
- Music transcription: from audio to notation
- Musical key and chord estimation from audio
- Controllable music synthesisers, including using Physics, PDEs

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was “borrowed” in another song - recent industry collab
- Music composition: symbolic/notation and direct to sound (controversial) - recent industry collab.
- Music transcription: from audio to notation
- Musical key and chord estimation from audio
- Controllable music synthesisers, including using Physics, PDEs
- Foley effects synthesis: foot steps in movies, etc

Typical problems in Music and Audio

All now use DL: they didn’t used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was “borrowed” in another song - recent industry collab
- Music composition: symbolic/notation and direct to sound (controversial) - recent industry collab.
- Music transcription: from audio to notation
- Musical key and chord estimation from audio
- Controllable music synthesisers, including using Physics, PDEs
- Foley effects synthesis: foot steps in movies, etc
- Audio identification: environmental, bioacoustics/biodiversity, musical instrument

Typical problems in Music and Audio

All now use DL: they didn't used to, though ML was omnipresent!

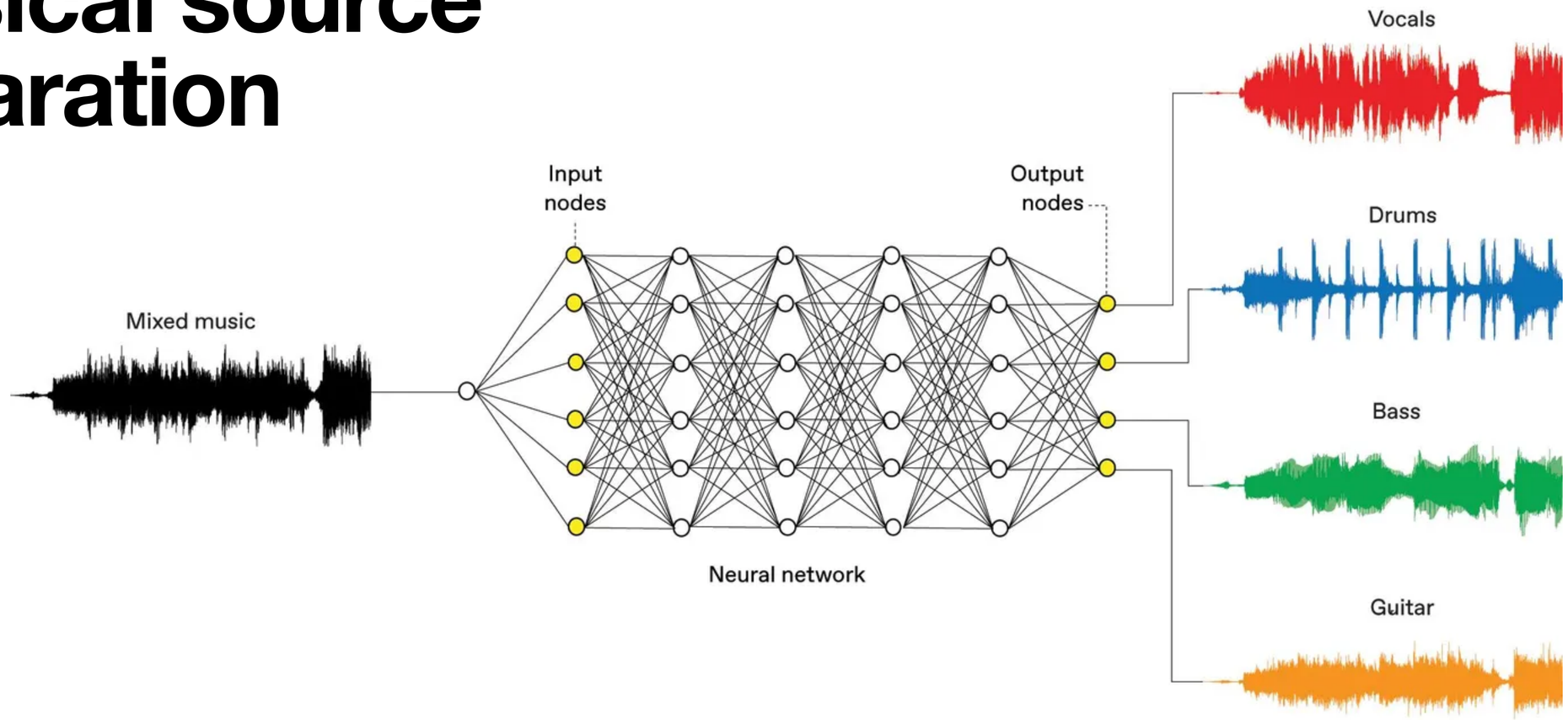
- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was "borrowed" in another song - recent industry collab
- Music composition: symbolic/notation and direct to sound (controversial) - recent industry collab.
- Music transcription: from audio to notation
- Musical key and chord estimation from audio
- Controllable music synthesisers, including using Physics, PDEs
- Foley effects synthesis: foot steps in movies, etc
- Audio identification: environmental, bioacoustics/biodiversity, musical instrument
- Almost all are grounded in the physical world

Typical problems in Music and Audio

All now use DL: they didn't used to, though ML was omnipresent!

- Musical (instrument) Source Separation: aka de-mixing
- Lyrics transcription: from singing to text - recent industry collaboration
- Sample identification: what song fragment was "borrowed" in another song - recent industry collab
- Music composition: symbolic/notation and direct to sound (controversial) - recent industry collab.
- Music transcription: from audio to notation
- Musical key and chord estimation from audio
- Controllable music synthesisers, including using Physics, PDEs
- Foley effects synthesis: foot steps in movies, etc
- Audio identification: environmental, bioacoustics/biodiversity, musical instrument
- Almost all are grounded in the physical world
  - BUT theoretical models and physical understandings are largely ignored

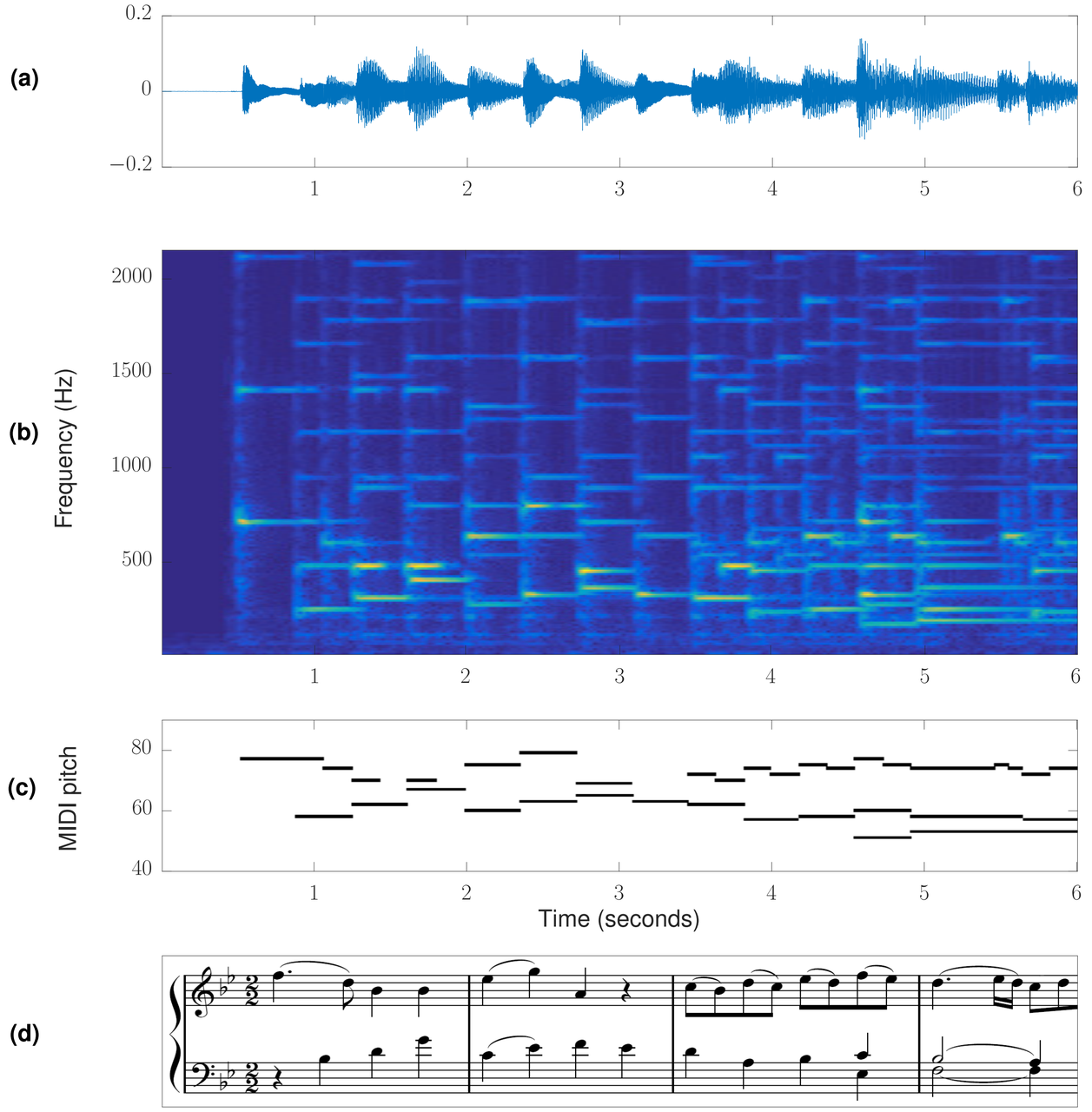
Musical source separation



From https://spectrum.ieee.org/3d-audio in an article by
By QI “PETER” LI, YIN DING & JOREL OLAN

Music transcription

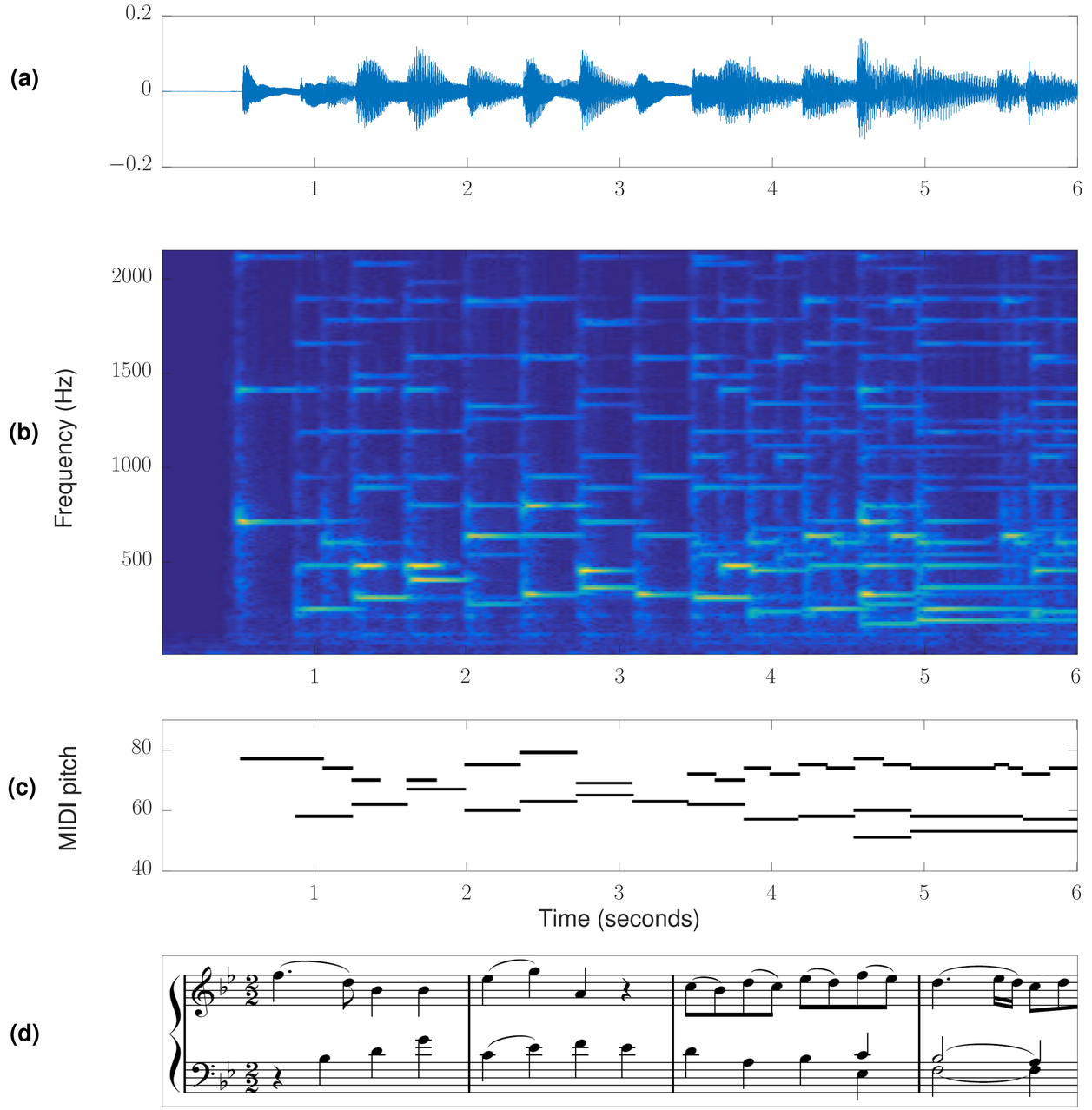
“speech to text” for music



Music transcription

“speech to text” for music

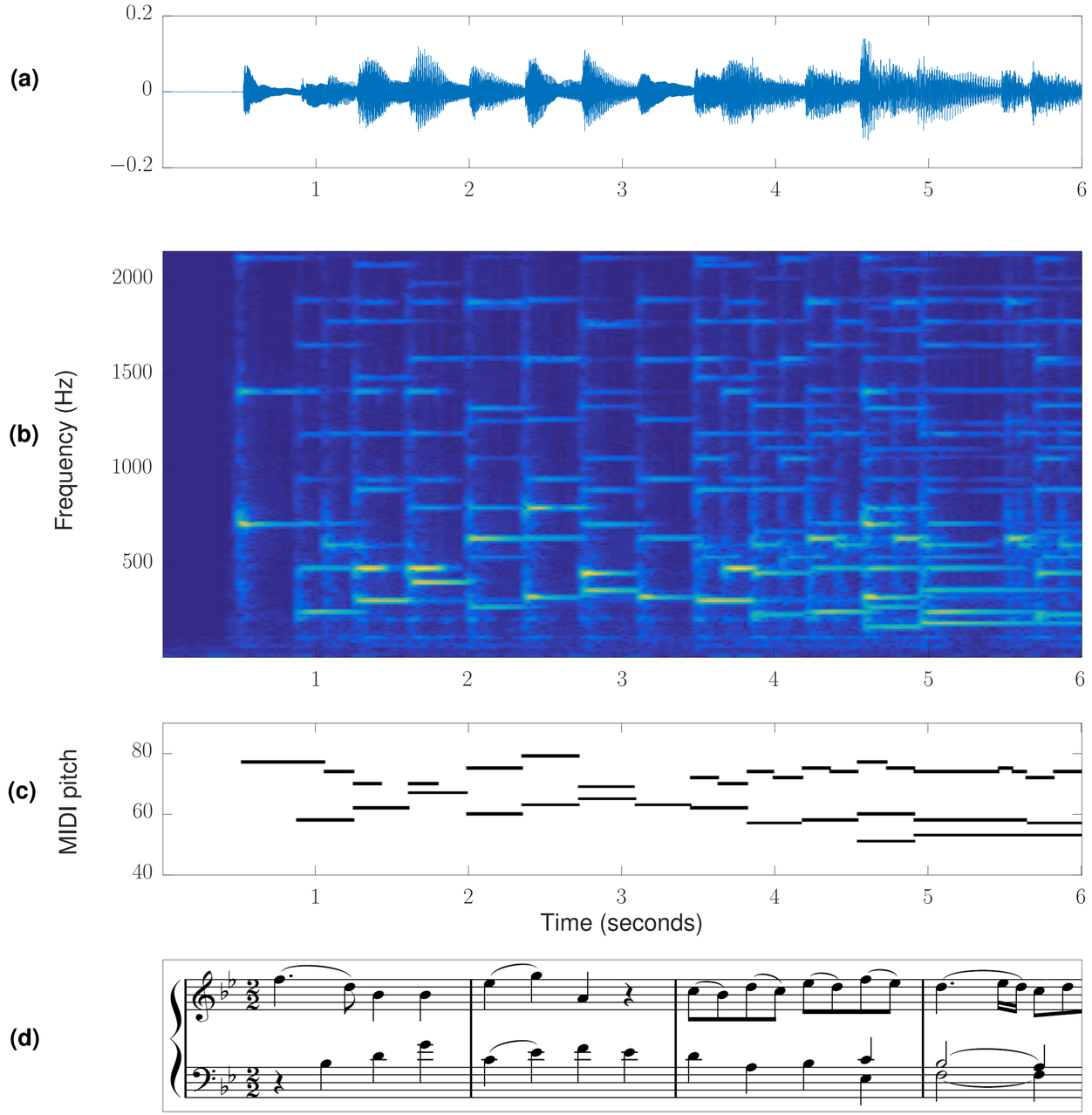
- From a time series



Music transcription

“speech to text” for music

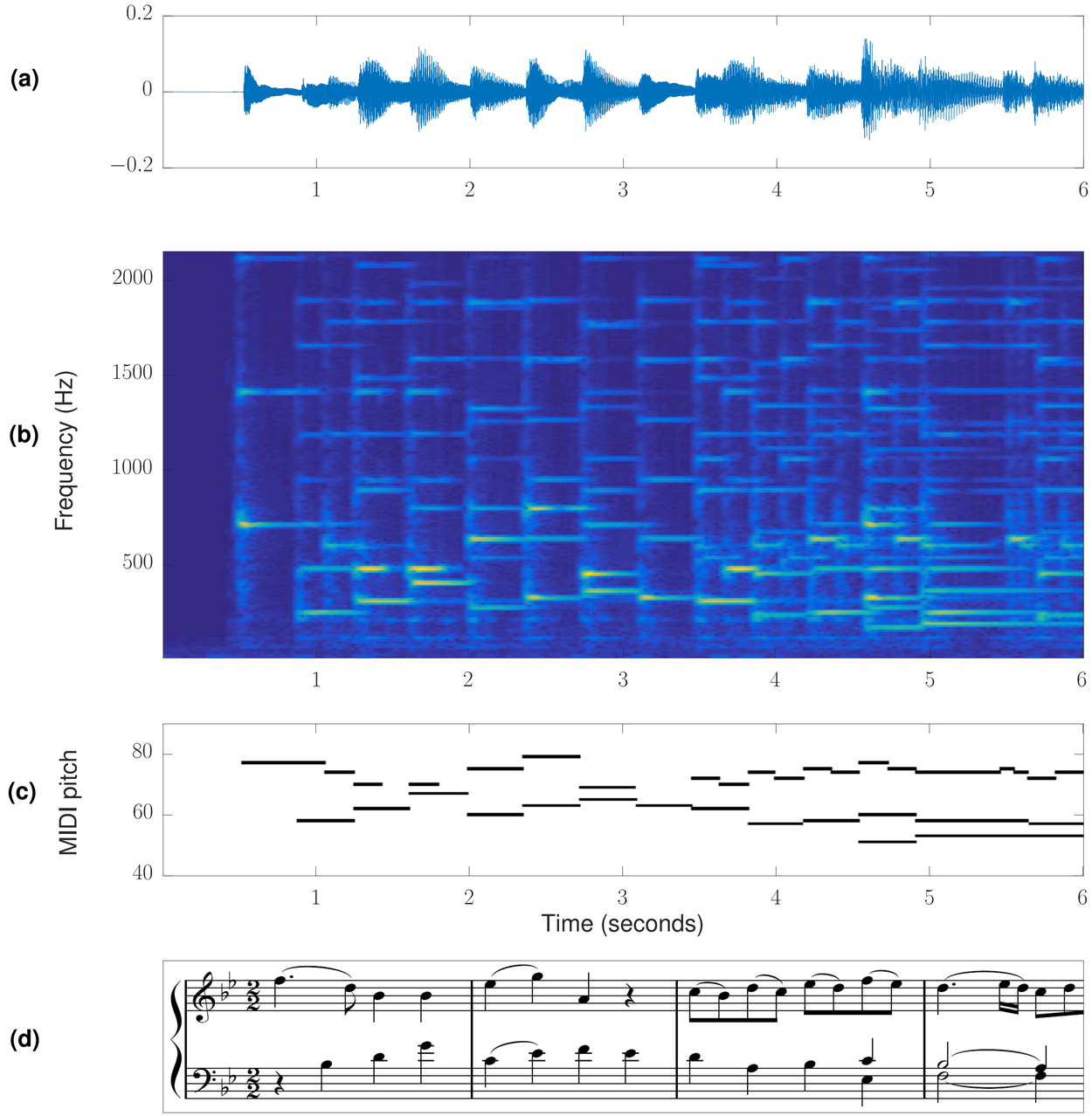
- From a time series
- To Fourier Magnitude



Music transcription

“speech to text” for music

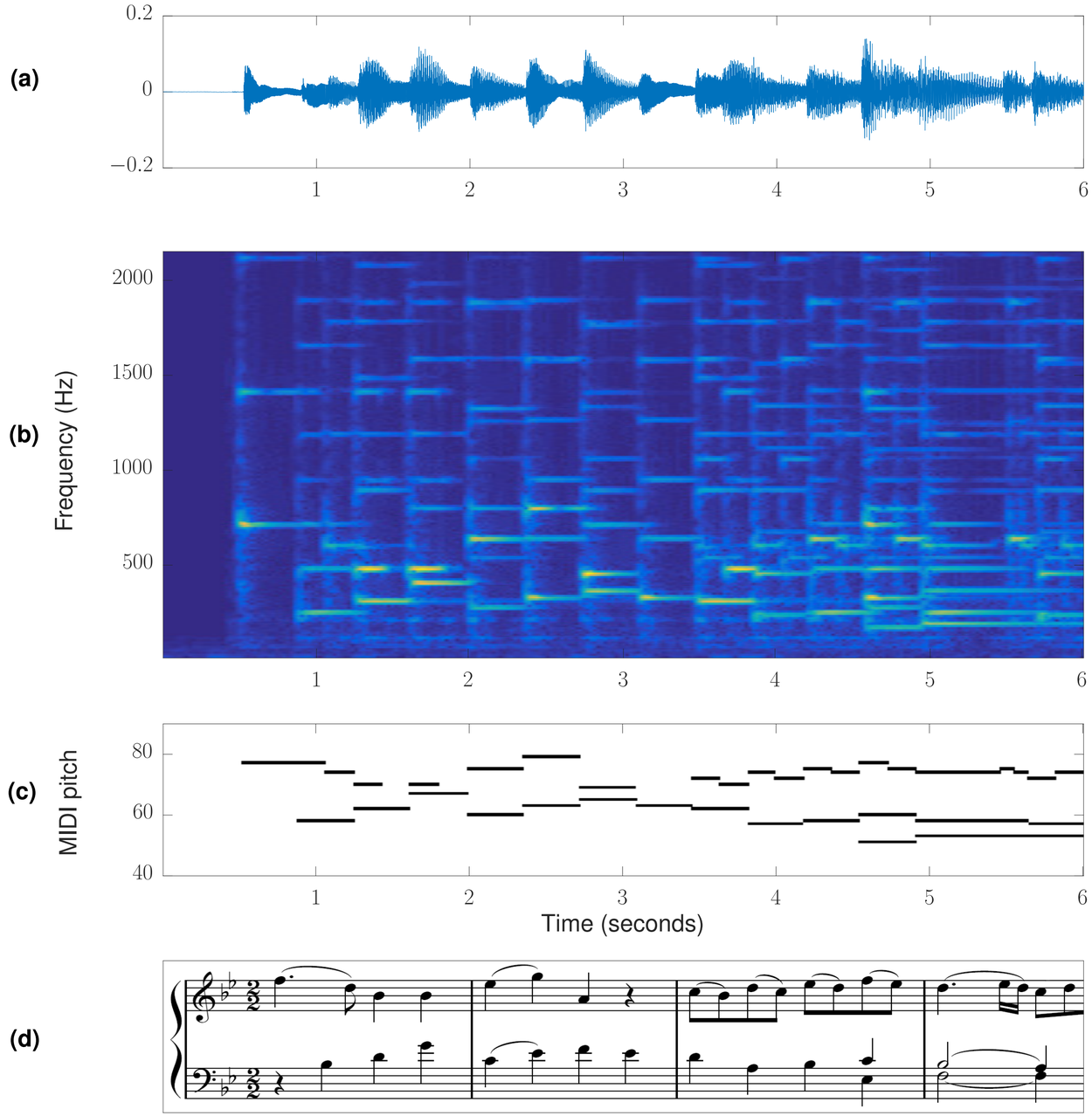
- From a time series
- To Fourier Magnitude
- To an event sequence



Music transcription

“speech to text” for music

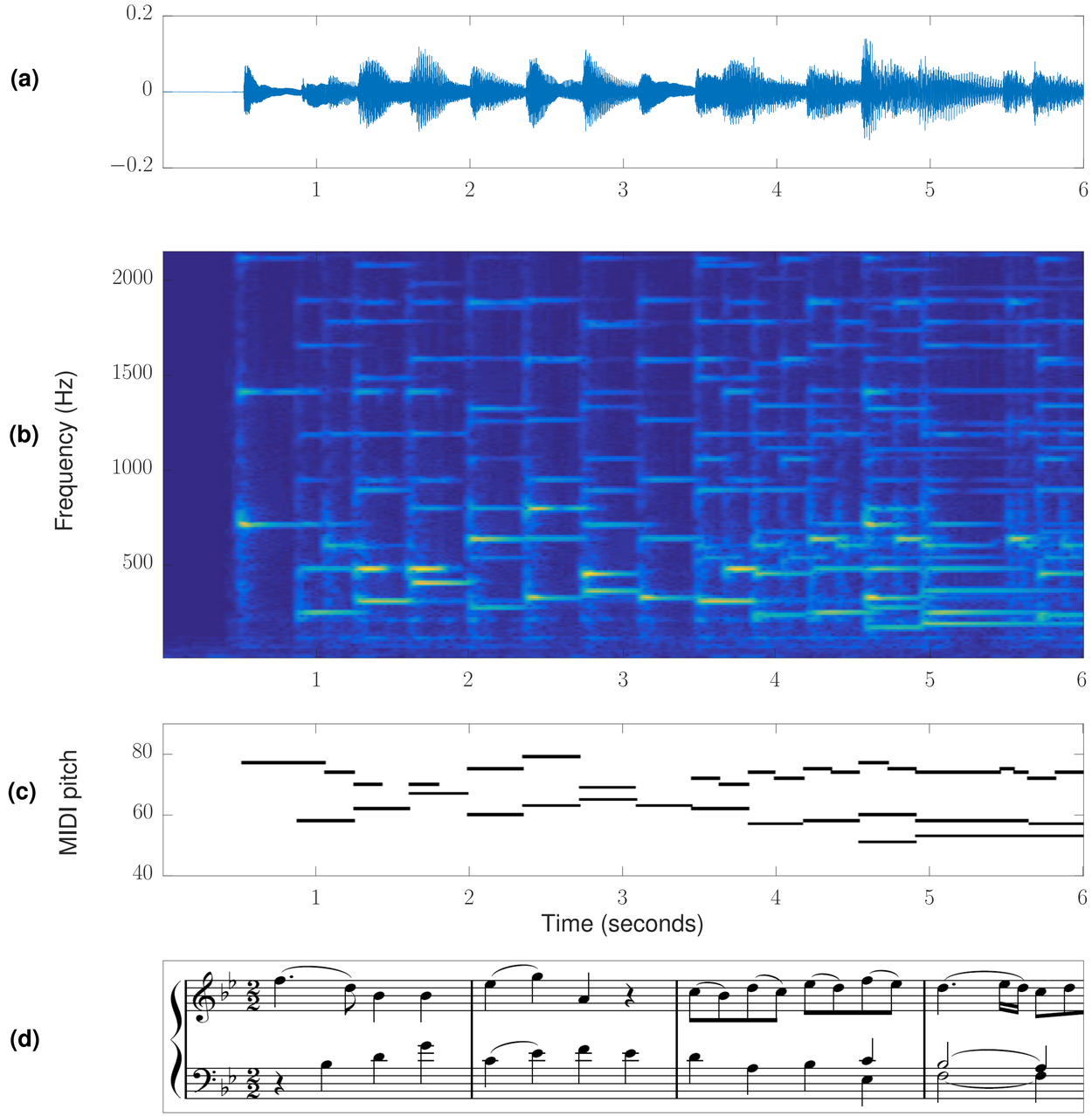
- From a time series
- To Fourier Magnitude
- To an event sequence
- To a symbolic representation



Music transcription

“speech to text” for music

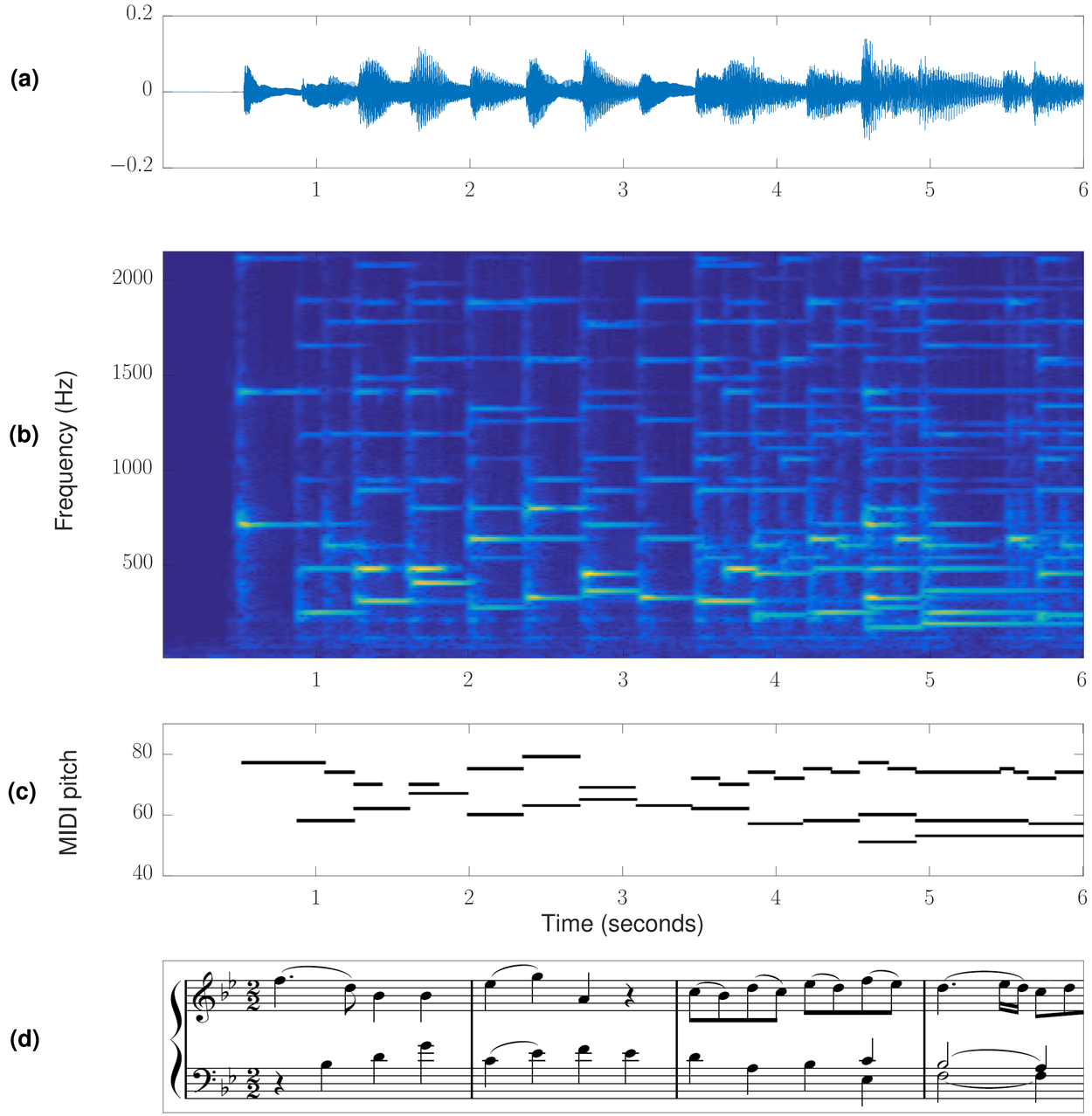
- From a time series
- To Fourier Magnitude
- To an event sequence
- To a symbolic representation
- Multiple pitches (overlapping in time and frequency)



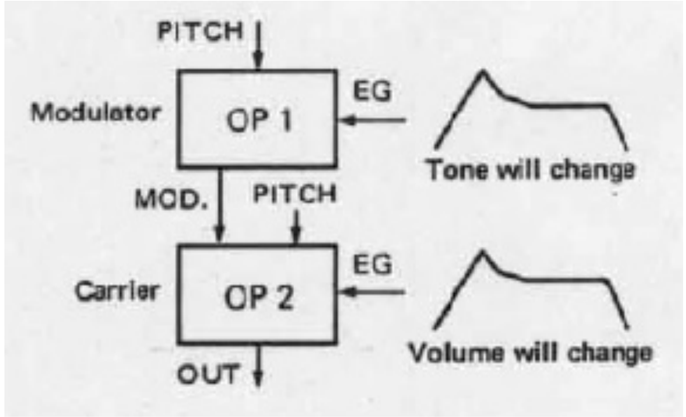
Music transcription

“speech to text” for music

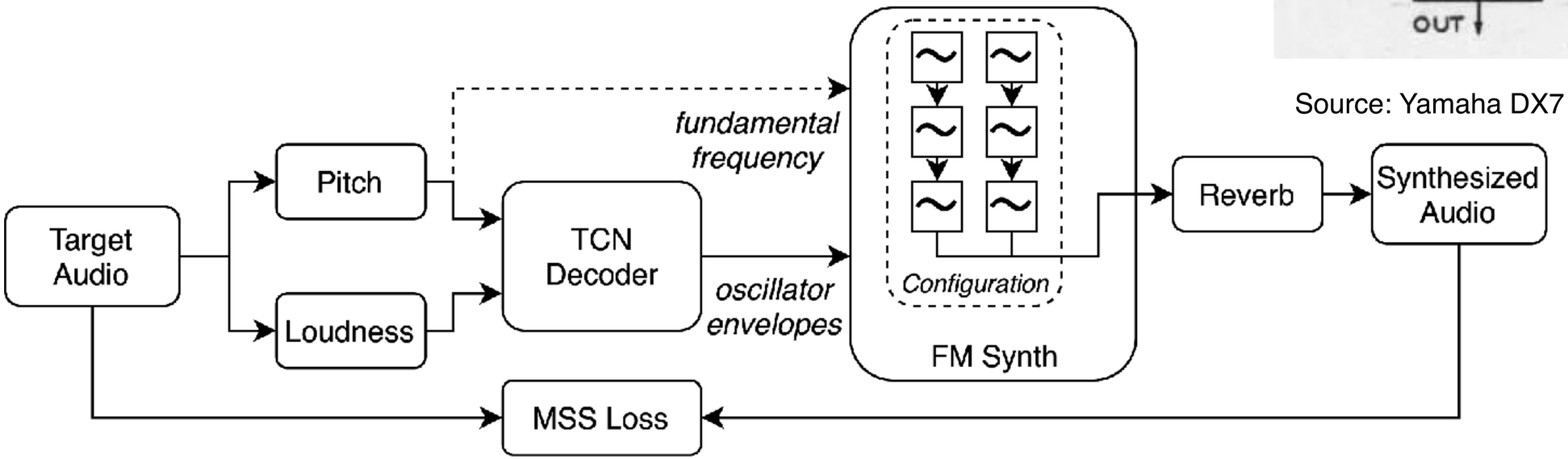
- From a time series
- To Fourier Magnitude
- To an event sequence
- To a symbolic representation
- Multiple pitches (overlapping in time and frequency)
- Instrument identification/ allocation



Music Synthesis
DDX7: A differentiable Yamaha DX7 model
Based on DDSP: Differentiable DSP



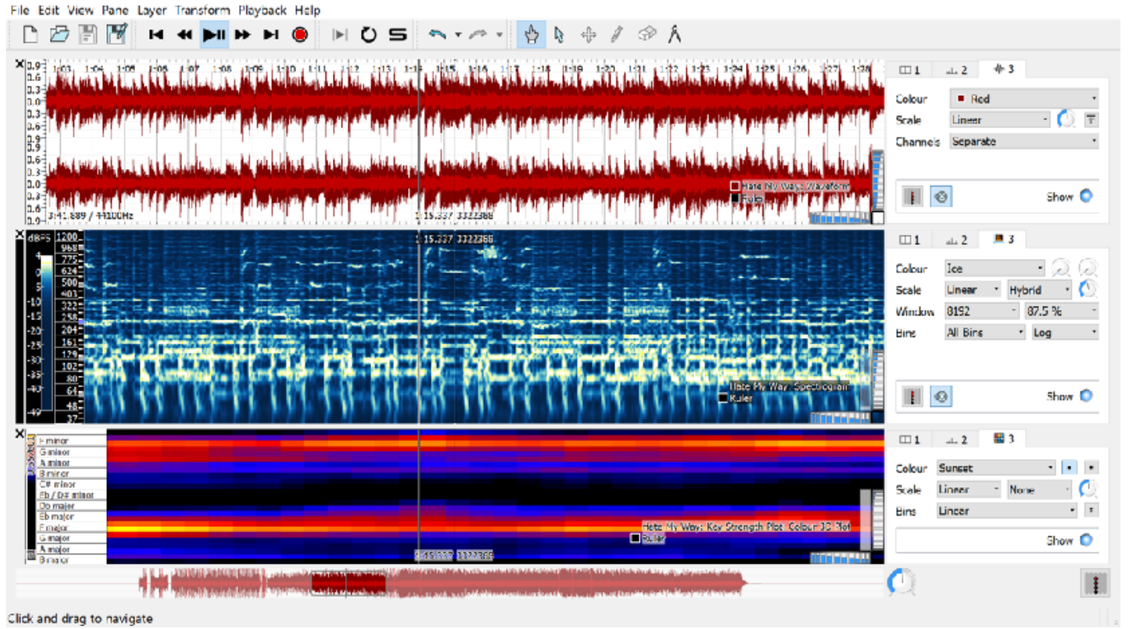
Source: Yamaha DX7 User's manual

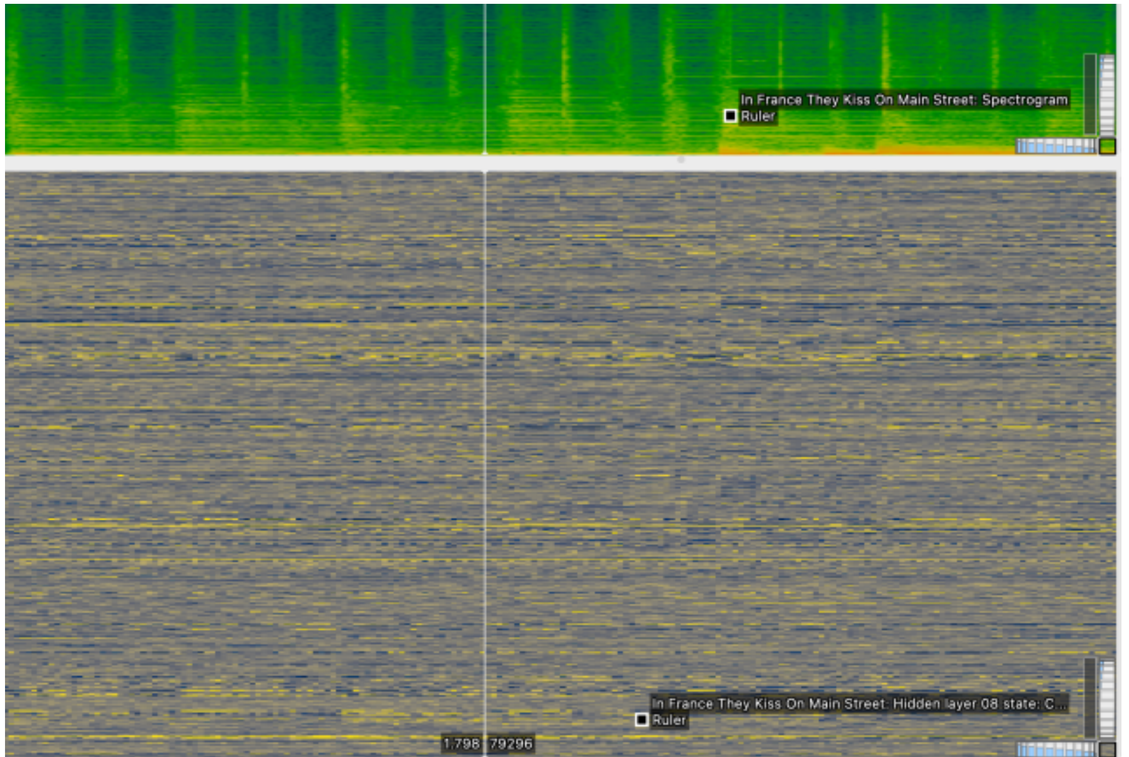


- Take inspiration from DX family of synths to constraint an optimization problem.
  - Fixed Oscillator Configuration.
  - Fixed Frequency ratios.
  - Few oscillators.
  - Envelope Generator controls *tone* and *volume*.
- Data-driven approach to an FM Synthesizer.
- **Result: A DX7 patch playable by an acoustic instrument**

Audio Features

AI Plugins for Sonic Visualiser

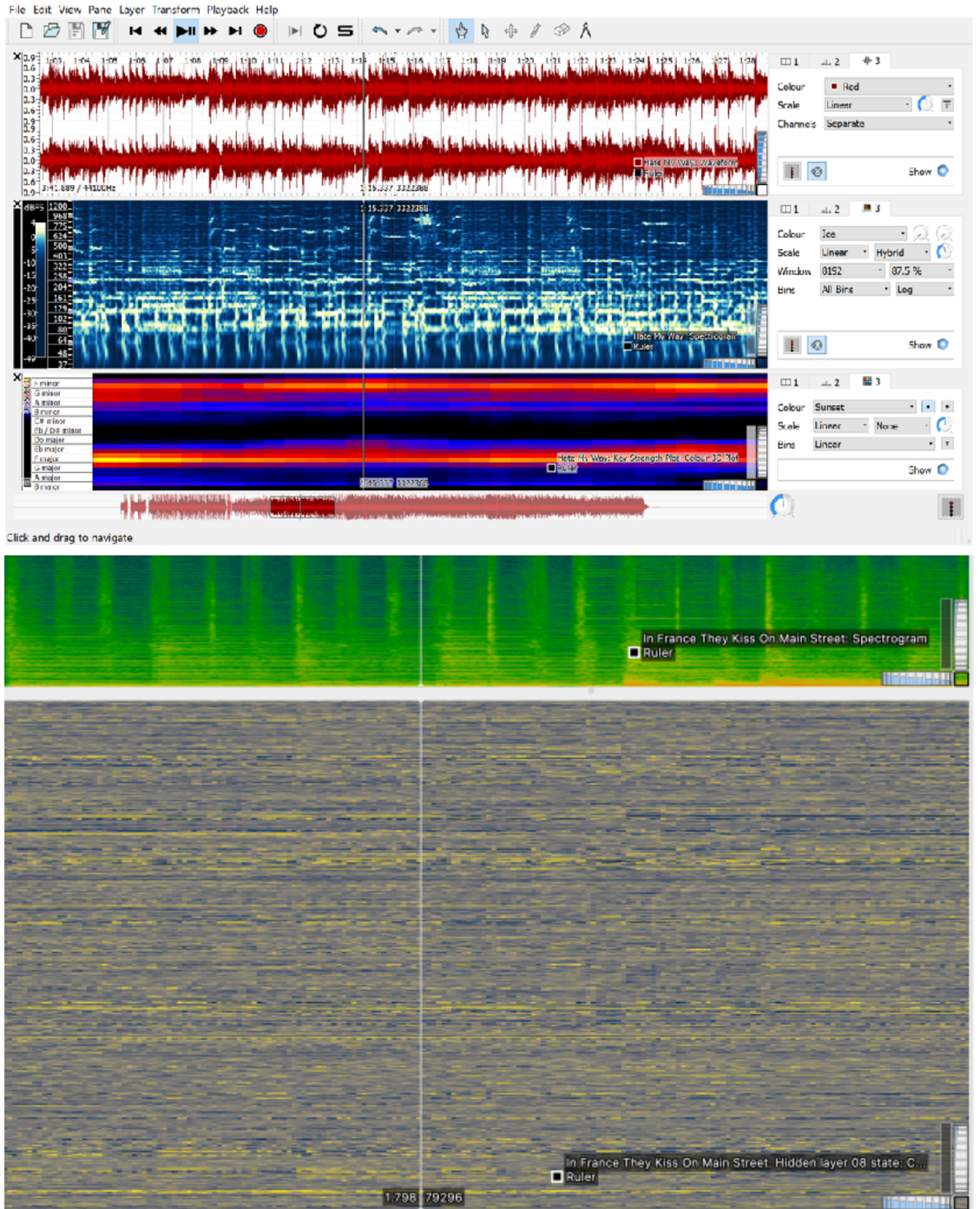




Audio Features

AI Plugins for Sonic Visualiser

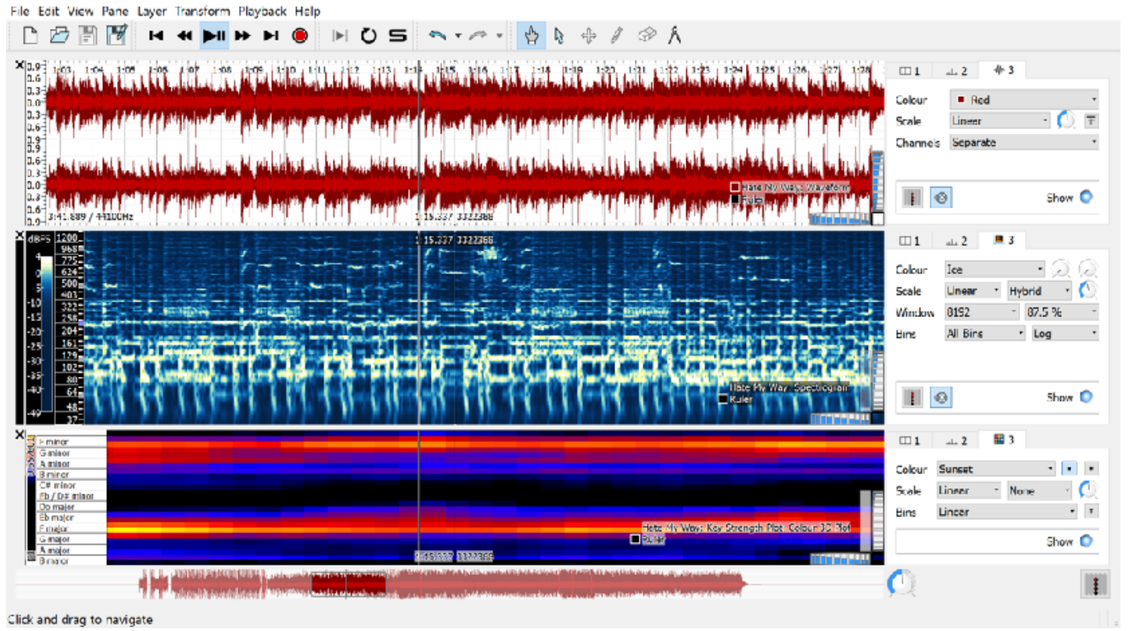
- Sonic Visualiser - open source tool from c4dm with an open plugin format

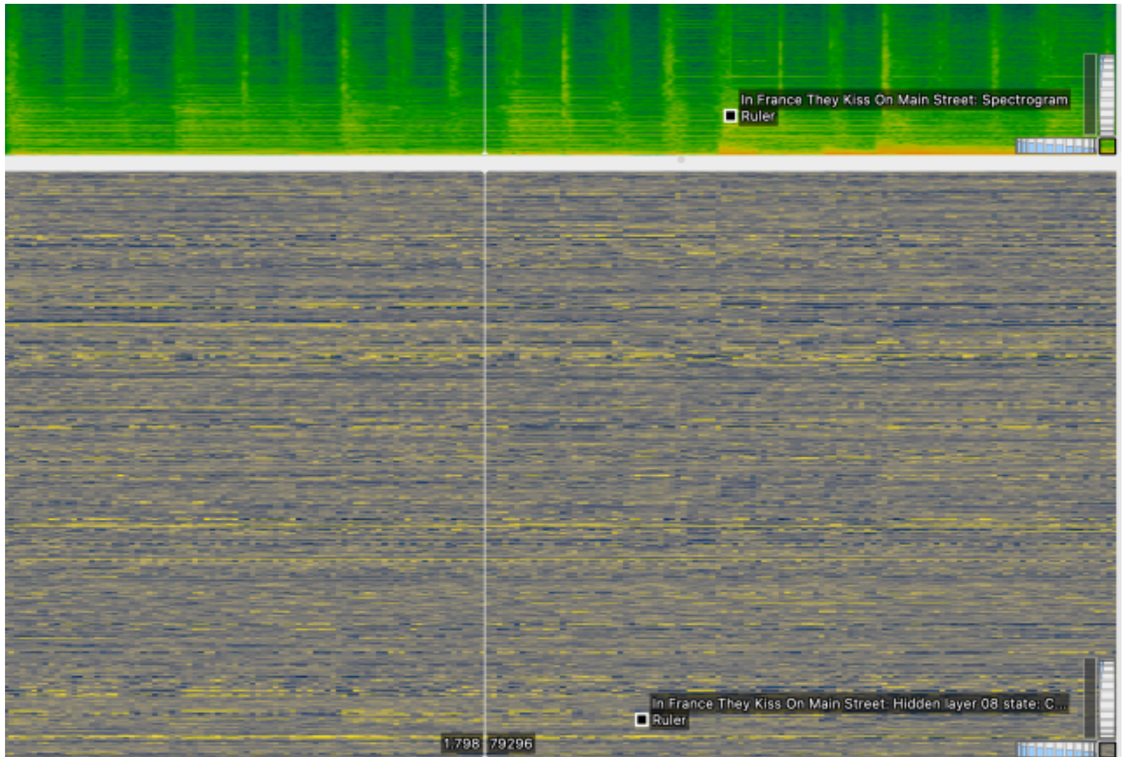


Audio Features

AI Plugins for Sonic Visualiser

- Sonic Visualiser - open source tool from c4dm with an open plugin format
- MERT Foundation Model plugin generates MERT audio features

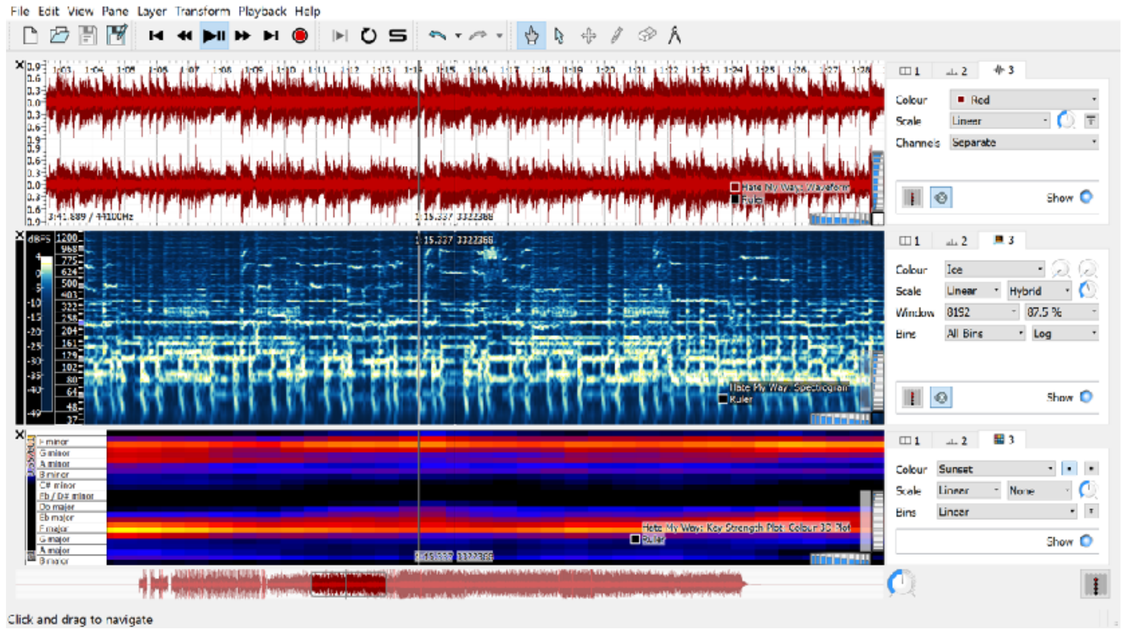


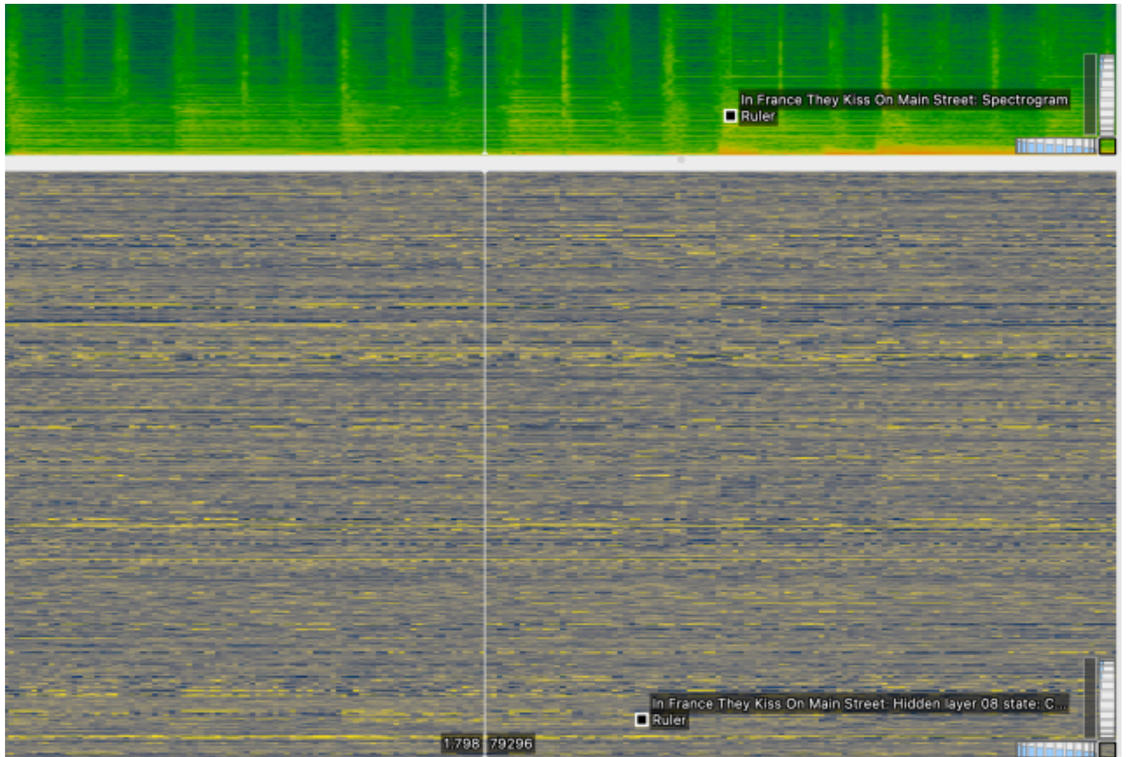


Audio Features

AI Plugins for Sonic Visualiser

- Sonic Visualiser - open source tool from c4dm with an open plugin format
- MERT Foundation Model plugin generates MERT audio features
  - >50k downloads May 2025 (huggingface), and > 750k total

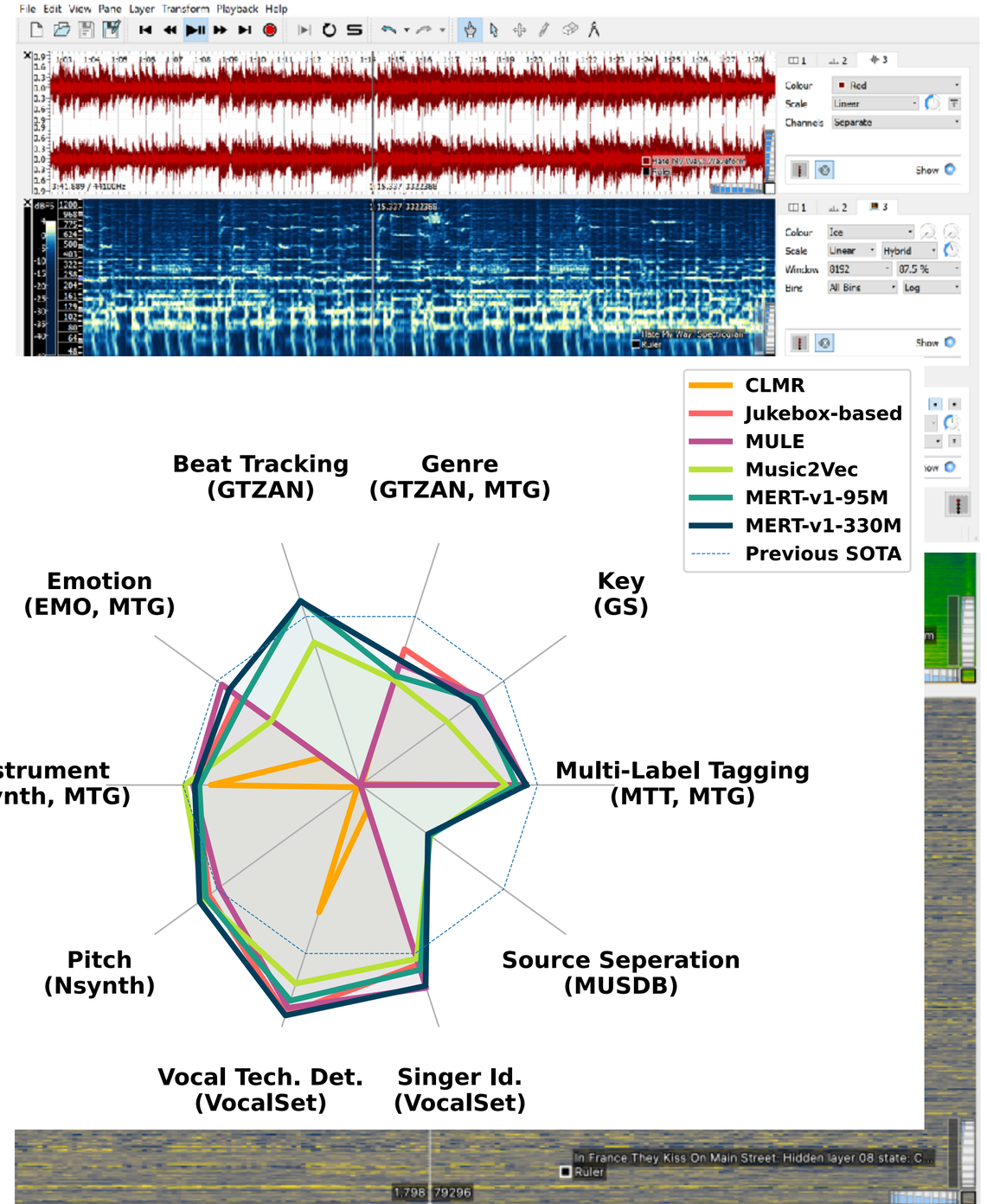




Audio Features

AI Plugins for Sonic Visualiser

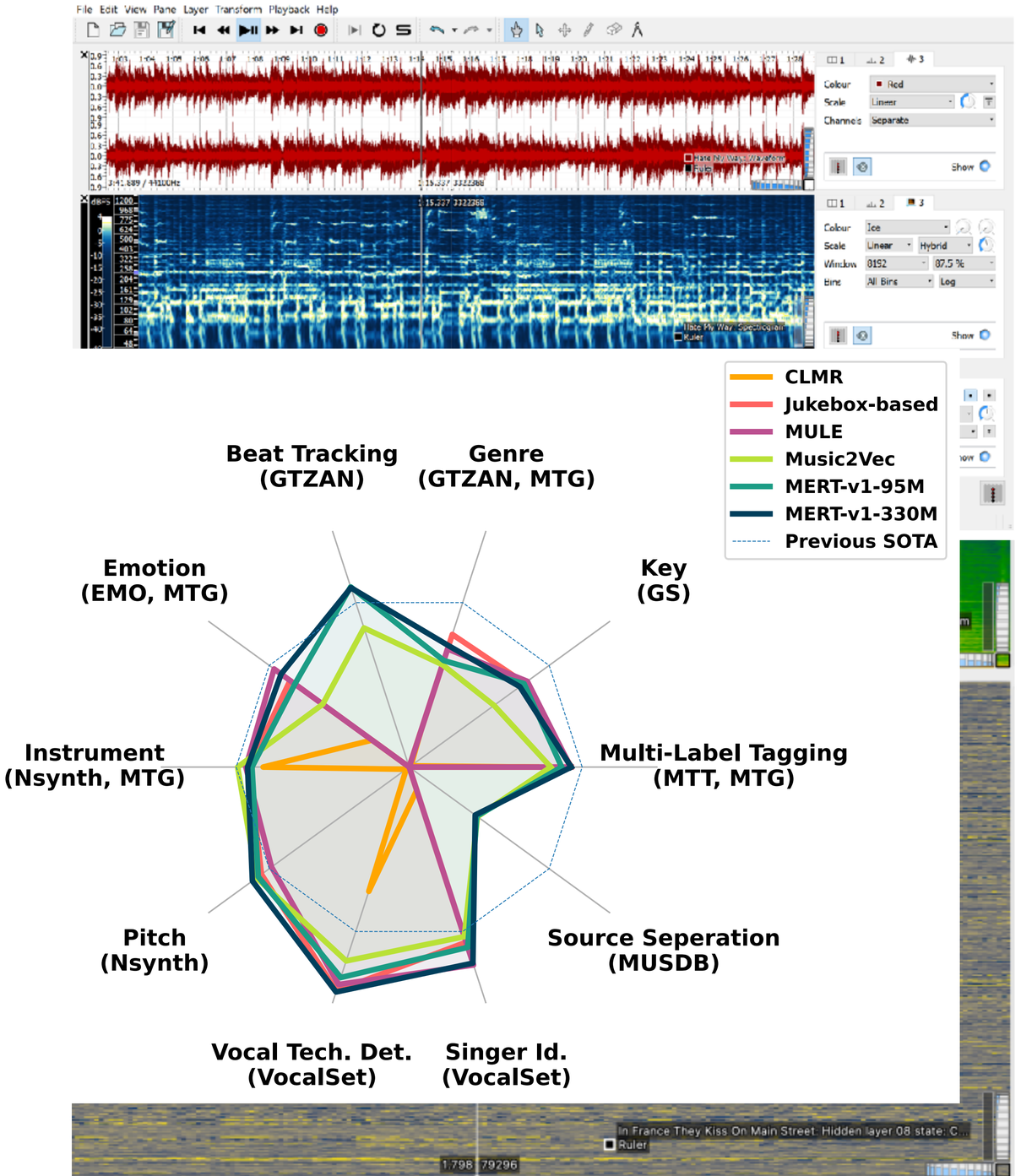
- Sonic Visualiser - open source tool from c4dm with an open plugin format
- MERT Foundation Model plugin generates MERT audio features
  - >50k downloads May 2025 (huggingface), and > 750k total



Audio Features

AI Plugins for Sonic Visualiser

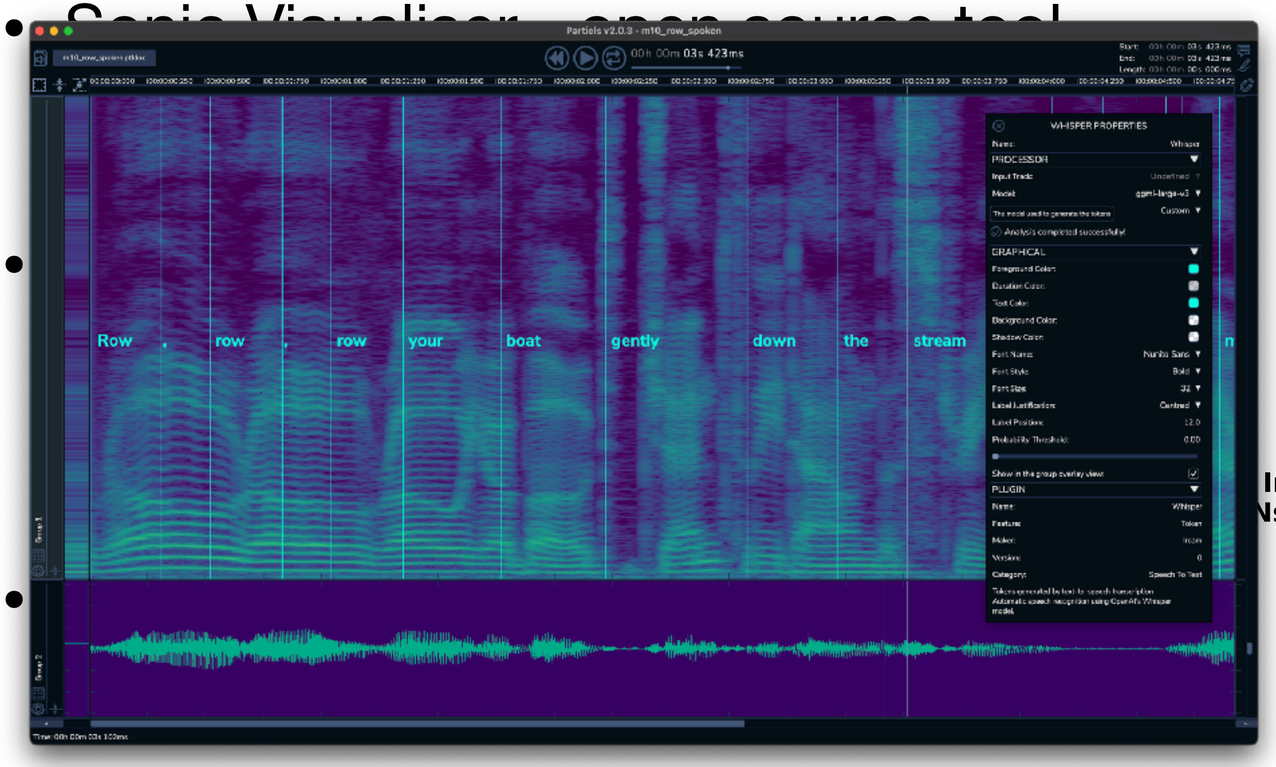
- Sonic Visualiser - open source tool from c4dm with an open plugin format
- MERT Foundation Model plugin generates MERT audio features
  - >50k downloads May 2025 (huggingface), and > 750k total
- WHISPER plugin useful for song lyrics transcription (from IRCAM in Paris)

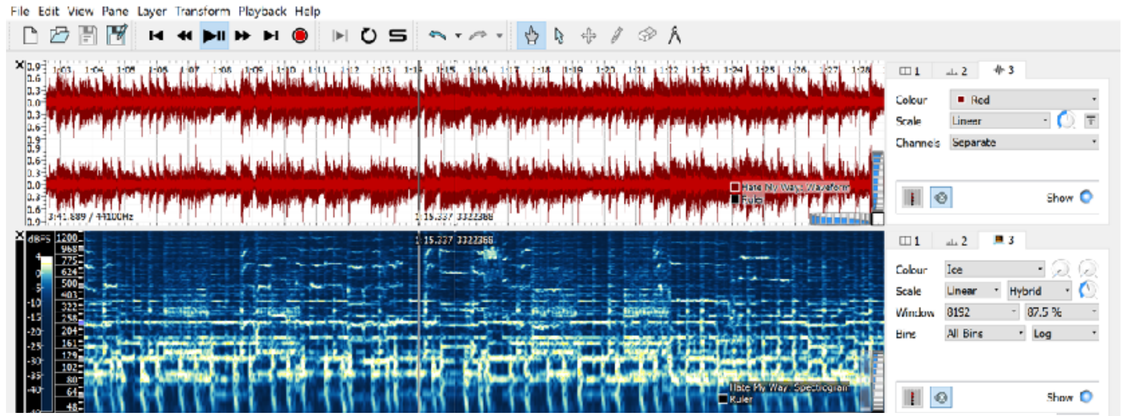


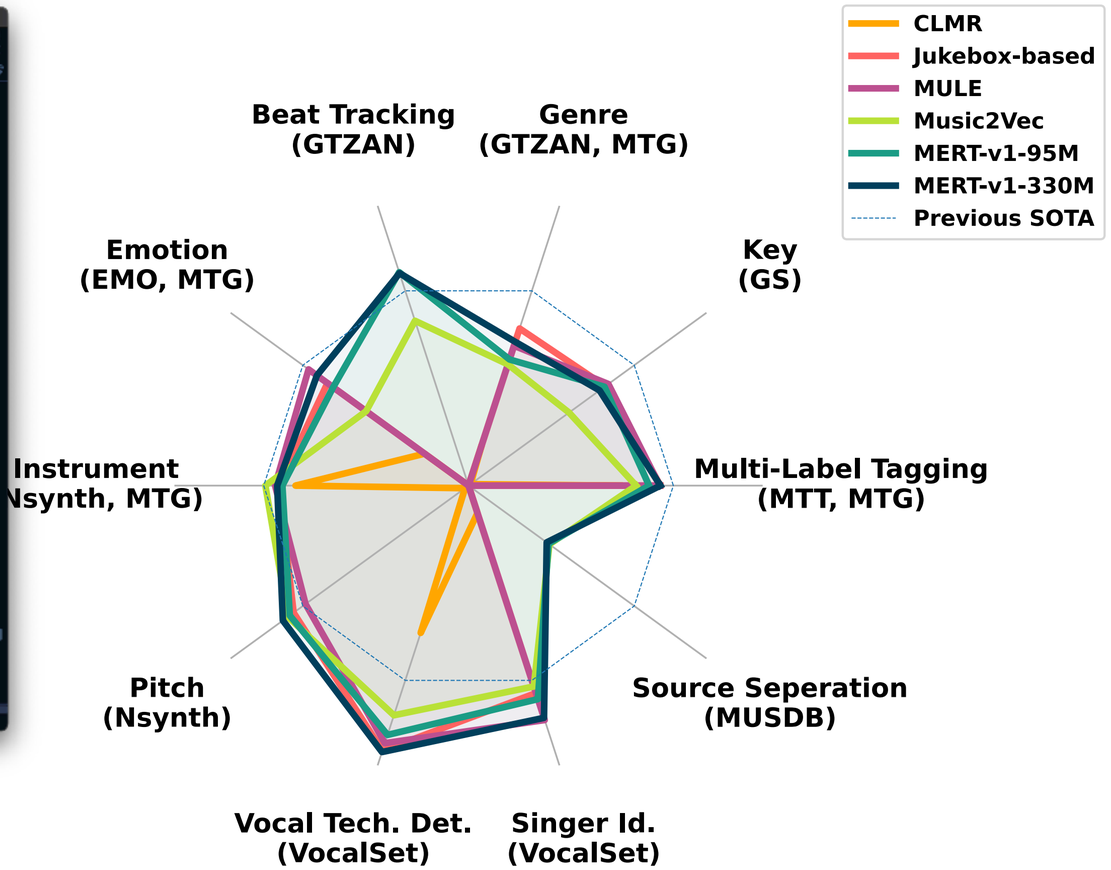
Audio Features

AI Plugins for Sonic Visualiser

- Sonic Visualiser - open source tool
-
-
-







A note on acceptable error in audio and music

A note on acceptable error in audio and music

- Decision, eg C-major vs C-minor vs E minor, violin vs viola vs tuba

A note on acceptable error in audio and music

- Decision, eg C-major vs C-minor vs E minor, violin vs viola vs tuba
  - ~ 90%, preferably better

A note on acceptable error in audio and music

- Decision, eg C-major vs C-minor vs E minor, violin vs viola vs tuba
  - ~ 90%, preferably better
- Reconstruction, ie rejection of interference, artefacts (source separation or generated audio)

A note on acceptable error in audio and music

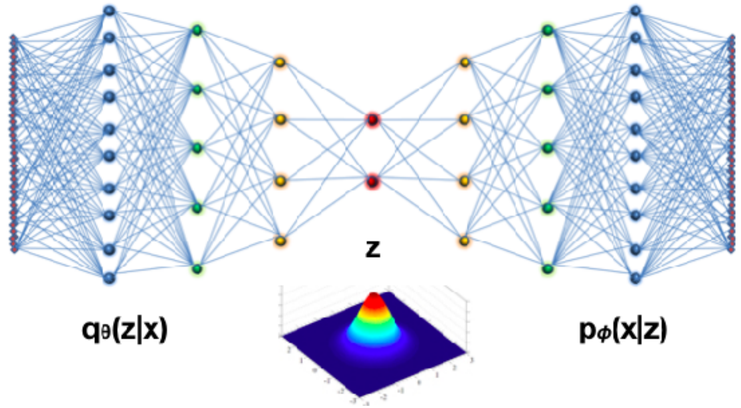
- Decision, eg C-major vs C-minor vs E minor, violin vs viola vs tuba
  - ~ 90%, preferably better
- Reconstruction, ie rejection of interference, artefacts (source separation or generated audio)
  - ~ 99.999% or higher!

A note on acceptable error in audio and music

- Decision, eg C-major vs C-minor vs E minor, violin vs viola vs tuba
  - ~ 90%, preferably better
- Reconstruction, ie rejection of interference, artefacts (source separation or generated audio)
  - ~ 99.999% or higher!
- Potentially simulaneously

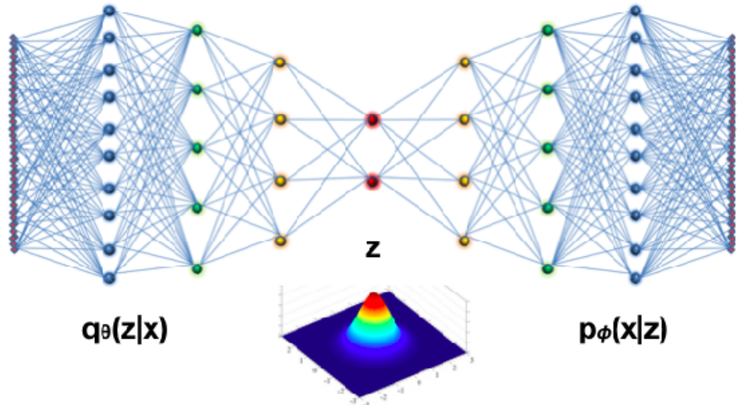
Applied Deep Learning - a critique

A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

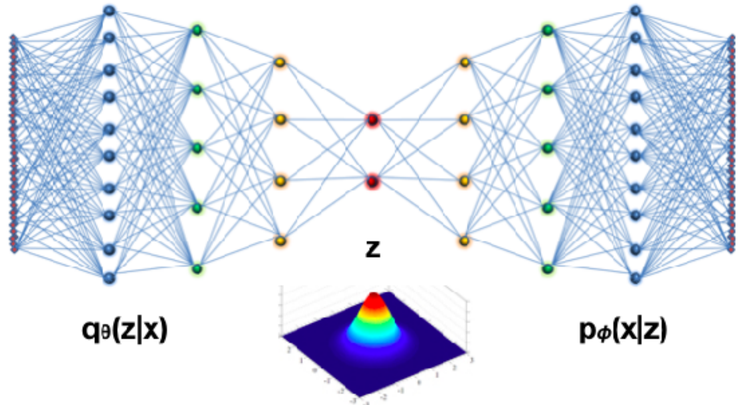
A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.

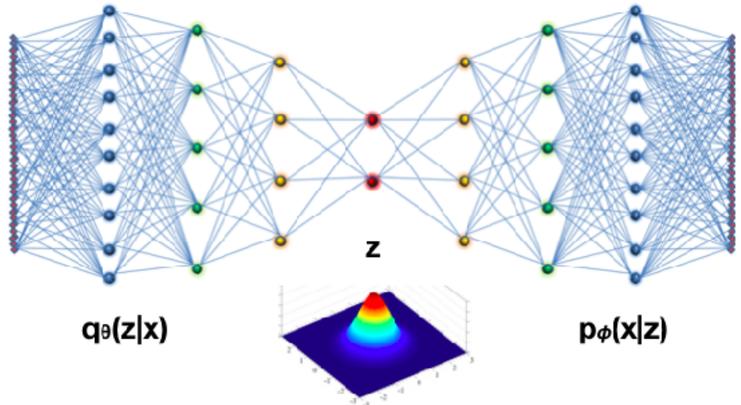
A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.

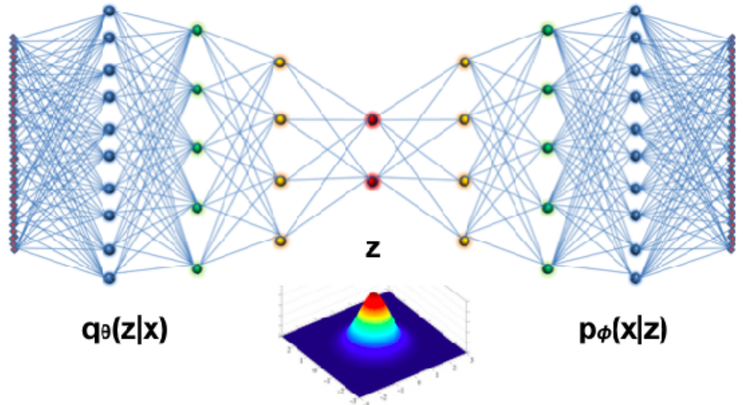
A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.
- Decide your input format: end-to-end (i.e. signal in) or transform in (e.g. Mel Spectrum, MFCC, Constant-Q, Wavelet, Wavelet of Wavelet,…)

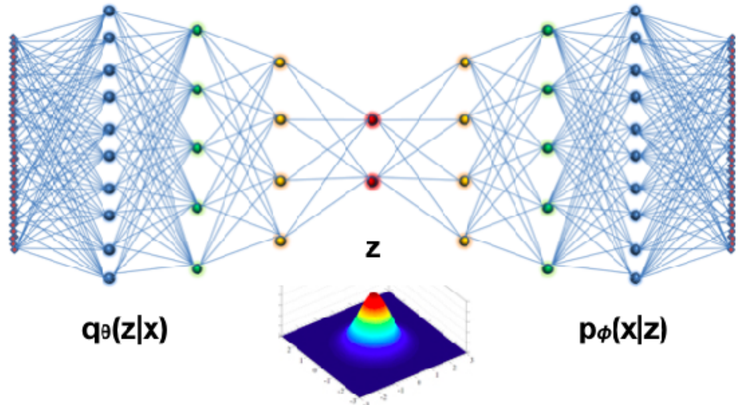
A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.
- Decide your input format: end-to-end (i.e. signal in) or transform in (e.g. Mel Spectrum, MFCC, Constant-Q, Wavelet, Wavelet of Wavelet,…)
- Choose a training regime and loss function: Triplet loss, L2 Norm etc., …

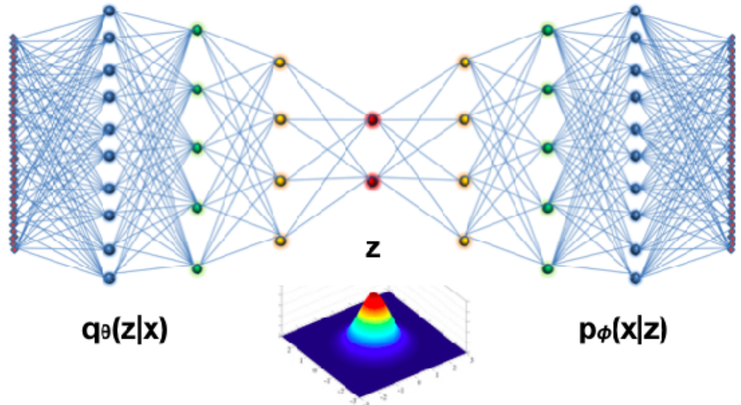
A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.
- Decide your input format: end-to-end (i.e. signal in) or transform in (e.g. Mel Spectrum, MFCC, Constant-Q, Wavelet, Wavelet of Wavelet,…)
- Choose a training regime and loss function: Triplet loss, L2 Norm etc., …
- Decide how many layers and how big each is

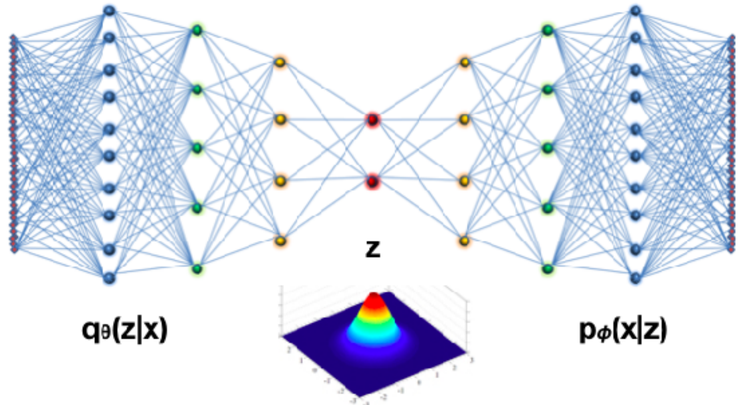
A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.
- Decide your input format: end-to-end (i.e. signal in) or transform in (e.g. Mel Spectrum, MFCC, Constant-Q, Wavelet, Wavelet of Wavelet,…)
- Choose a training regime and loss function: Triplet loss, L2 Norm etc., …
- Decide how many layers and how big each is
- Decide which framework and libraries to use

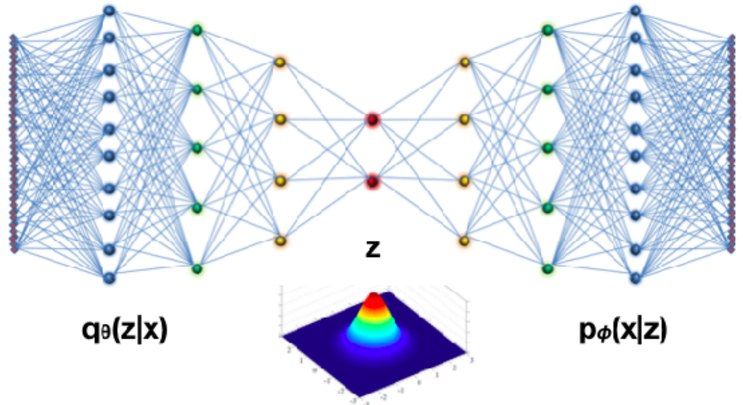
A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.
- Decide your input format: end-to-end (i.e. signal in) or transform in (e.g. Mel Spectrum, MFCC, Constant-Q, Wavelet, Wavelet of Wavelet,…)
- Choose a training regime and loss function: Triplet loss, L2 Norm etc., …
- Decide how many layers and how big each is
- Decide which framework and libraries to use
- Split the data set into training, (validation) and testing, possibly using F-fold validation

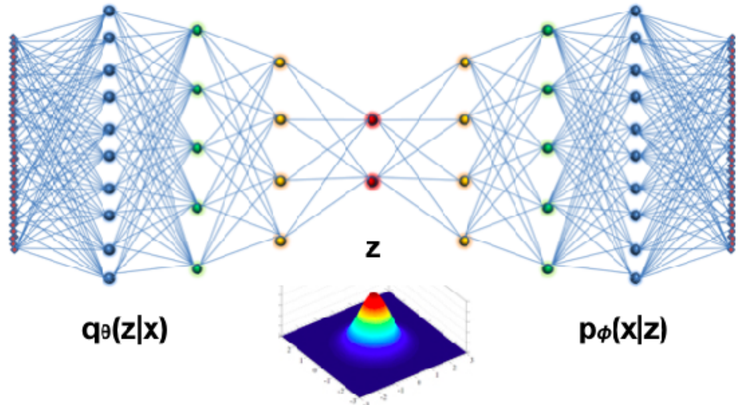
A typical AI/Music/Audio Deep Learning research pipeline



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.
- Decide your input format: end-to-end (i.e. signal in) or transform in (e.g. Mel Spectrum, MFCC, Constant-Q, Wavelet, Wavelet of Wavelet,…)
- Choose a training regime and loss function: Triplet loss, L2 Norm etc., …
- Decide how many layers and how big each is
- Decide which framework and libraries to use
- Split the data set into training, (validation) and testing, possibly using F-fold validation
- Run experiments for hours or days.

A typical AI/Music/Audio Deep Learning research pipeline

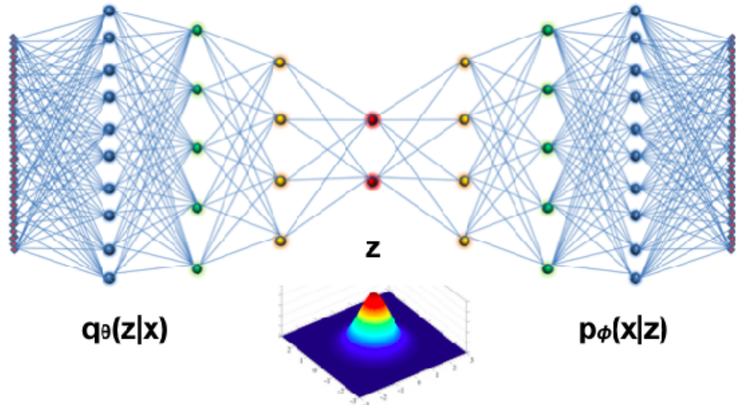


VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.
- Decide your input format: end-to-end (i.e. signal in) or transform in (e.g. Mel Spectrum, MFCC, Constant-Q, Wavelet, Wavelet of Wavelet,…)
- Choose a training regime and loss function: Triplet loss, L2 Norm etc., …
- Decide how many layers and how big each is
- Decide which framework and libraries to use
- Split the data set into training, (validation) and testing, possibly using F-fold validation
- Run experiments for hours or days.
- Evaluate against ground truth, generate statistics, prove your approach is this week's State of the Art.

A typical AI/Music/Audio Deep Learning research pipeline

What could possibly go wrong?



VAE Illustration by Stephen G. Odaibo, M.D.
From: https://medium.com/retina-ai-health-inc/variational-inference-derivation-of-the-variational-autoencoder-vae-loss-function-a-true-story-3543a3dc67ee

- Find or Create a dataset for your problem area. It needs some ground truth.
- Choose a specific DL architecture: Transformer, CNN, RNN, LSTM etc.
- Decide your input format: end-to-end (i.e. signal in) or transform in (e.g. Mel Spectrum, MFCC, Constant-Q, Wavelet, Wavelet of Wavelet,…)
- Choose a training regime and loss function: Triplet loss, L2 Norm etc., …
- Decide how many layers and how big each is
- Decide which framework and libraries to use
- Split the data set into training, (validation) and testing, possibly using F-fold validation
- Run experiments for hours or days.
- Evaluate against ground truth, generate statistics, prove your approach is this week's State of the Art.

All these could go wrong!

Or not be right enough

All these could go wrong!

Or not be right enough

- Quality and acoustic integrity of dataset. Quality and reliability of ground truth.

All these could go wrong!

Or not be right enough

- Quality and acoustic integrity of dataset. Quality and reliability of ground truth.
- What is the hypothesis being tested? Is the experiment well-formed?

All these could go wrong!

Or not be right enough

- Quality and acoustic integrity of dataset. Quality and reliability of ground truth.
- What is the hypothesis being tested? Is the experiment well-formed?
- Pick the most fashionable architecture and squeeze your problem onto it

All these could go wrong!

Or not be right enough

- Quality and acoustic integrity of dataset. Quality and reliability of ground truth.
- What is the hypothesis being tested? Is the experiment well-formed?
- Pick the most fashionable architecture and squeeze your problem onto it
- Reliability of framework and libraries. Who validates them? Who supports them?

All these could go wrong!

Or not be right enough

- Quality and acoustic integrity of dataset. Quality and reliability of ground truth.
- What is the hypothesis being tested? Is the experiment well-formed?
- Pick the most fashionable architecture and squeeze your problem onto it
- Reliability of framework and libraries. Who validates them? Who supports them?
- Is your test set truly representative of your “downstream” problem? Will you overfit because training and test data are too similar?

All these could go wrong!

Or not be right enough

- Quality and acoustic integrity of dataset. Quality and reliability of ground truth.
- What is the hypothesis being tested? Is the experiment well-formed?
- Pick the most fashionable architecture and squeeze your problem onto it
- Reliability of framework and libraries. Who validates them? Who supports them?
- Is your test set truly representative of your “downstream” problem? Will you overfit because training and test data are too similar?
- Who is paying your electricity bill and who is planting the trees to offset the carbon?

All these could go wrong!

Or not be right enough

- Quality and acoustic integrity of dataset. Quality and reliability of ground truth.
- What is the hypothesis being tested? Is the experiment well-formed?
- Pick the most fashionable architecture and squeeze your problem onto it
- Reliability of framework and libraries. Who validates them? Who supports them?
- Is your test set truly representative of your “downstream” problem? Will you overfit because training and test data are too similar?
- Who is paying your electricity bill and who is planting the trees to offset the carbon?
- Is your problem of real importance, or just a toy example?

Image from: https://www.scnsoft.com/blog/big-data-quality

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

Image from: https://www.scnsoft.com/blog/big-data-quality

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

- Data quality: how ecologically valid should it be?

Image from: https://www.scnsoft.com/blog/big-data-quality

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

Image from: https://www.scnsoft.com/blog/big-data-quality

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?
  - Quality of ground truth annotations: Keyboards = ?

Image from: https://www.scnsoft.com/blog/big-data-quality

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?
  - Quality of ground truth annotations: Keyboards = ?
    - inter-rater agreement is rarely examined

Image from: https://www.scnsoft.com/blog/big-data-quality

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?
  - Quality of ground truth annotations: Keyboards = ?
    - inter-rater agreement is rarely examined
  - Data augmentation: pitch-shift, white noise, MPEG artefacts, … Ecological validity?

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

Image from: https://www.scnsoft.com/blog/big-data-quality

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?
  - Quality of ground truth annotations: Keyboards = ?
    - inter-rater agreement is rarely examined
  - Data augmentation: pitch-shift, white noise, MPEG artefacts, … Ecological validity?
- Data set synthesis increasingly favoured

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

Image from: https://www.scnsoft.com/blog/big-data-quality

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?
  - Quality of ground truth annotations: Keyboards = ?
    - inter-rater agreement is rarely examined
  - Data augmentation: pitch-shift, white noise, MPEG artefacts, … Ecological validity?
- Data set synthesis increasingly favoured
  - Piecing together note samples to create bespoke training data (like pop!)

Image from: https://www.scnsoft.com/blog/big-data-quality

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?
  - Quality of ground truth annotations: Keyboards = ?
    - inter-rater agreement is rarely examined
  - Data augmentation: pitch-shift, white noise, MPEG artefacts, … Ecological validity?
- Data set synthesis increasingly favoured
  - Piecing together note samples to create bespoke training data (like pop!)
  - Physical modelling of instrument, mic and room for ecological validity

Image from: https://www.scnsoft.com/blog/big-data-quality

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?
  - Quality of ground truth annotations: Keyboards = ?
    - inter-rater agreement is rarely examined
  - Data augmentation: pitch-shift, white noise, MPEG artefacts, … Ecological validity?
- Data set synthesis increasingly favoured
  - Piecing together note samples to create bespoke training data (like pop!)
  - Physical modelling of instrument, mic and room for ecological validity
  - Almost fully provenanced data

Image from: https://www.scnsoft.com/blog/big-data-quality

Datasets in Music and Audio

~ #grains of sand at Bournemouth beach

- Data quality: how ecologically valid should it be?
  - Representative quality: enough genres? rhythmic variety? acoustic vs studio recordings? #instruments, #chord types, MPEG or uncompressed?
  - Quality of ground truth annotations: Keyboards = ?
    - inter-rater agreement is rarely examined
  - Data augmentation: pitch-shift, white noise, MPEG artefacts, … Ecological validity?
- Data set synthesis increasingly favoured
  - Piecing together note samples to create bespoke training data (like pop!)
  - Physical modelling of instrument, mic and room for ecological validity
  - Almost fully provenanced data
  - Leads to better trained, more generalisable models

Prospects for Artificial Neuroscience

Artificial what?

this talk, so far

Artificial what?

this talk, so far

- What’s the problem here?

Artificial what?

this talk, so far

- What’s the problem here?
  - lack of experimental rigour

Artificial what?

this talk, so far

- What’s the problem here?
  - lack of experimental rigour
  - lack of engineering

Artificial what?

this talk, so far

- What’s the problem here?
  - lack of experimental rigour
  - lack of engineering
  - Lack of mathematical models

Artificial what?

this talk, so far

- What’s the problem here?
  - lack of experimental rigour
  - lack of engineering
  - Lack of mathematical models
  - Paucity of ethical standards

Artificial what?

this talk, so far

- What's the problem here?
  - lack of experimental rigour
  - lack of engineering
  - Lack of mathematical models
  - Paucity of ethical standards
- "Artificial brains" created that

Artificial what?

this talk, so far

- What's the problem here?
  - lack of experimental rigour
  - lack of engineering
  - Lack of mathematical models
  - Paucity of ethical standards
- "Artificial brains" created that
  - Aren't understood - structures so huge and diverse. There is no "algebra" of DL

Artificial what?

this talk, so far

- What's the problem here?
  - lack of experimental rigour
  - lack of engineering
  - Lack of mathematical models
  - Paucity of ethical standards
- "Artificial brains" created that
  - Aren't understood - structures so huge and diverse. There is no "algebra" of DL
  - Aren't evaluated properly, rely on benchmarking (i.e. engineering not cognition-oriented). What about ecological validity?

Artificial what?

this talk, so far

- What's the problem here?
  - lack of experimental rigour
  - lack of engineering
  - Lack of mathematical models
  - Paucity of ethical standards
- "Artificial brains" created that
  - Aren't understood - structures so huge and diverse. There is no "algebra" of DL
  - Aren't evaluated properly, rely on benchmarking (i.e. engineering not cognition-oriented). What about ecological validity?
  - Are often trained with data that isn't optimal (get what you can)

Artificial what?

this talk, so far

- What's the problem here?
  - lack of experimental rigour
  - lack of engineering
  - Lack of mathematical models
  - Paucity of ethical standards

- "Artificial brains" created that
  - Aren't understood - structures so huge and diverse. There is no "algebra" of DL
  - Aren't evaluated properly, rely on benchmarking (i.e. engineering not cognition-oriented). What about ecological validity?
  - Are often trained with data that isn't optimal (get what you can)
  - Interact with each other and with humans, proliferating

Artificial what?

this talk, so far

- What's the problem here?
  - lack of experimental rigour
  - lack of engineering
  - Lack of mathematical models
  - Paucity of ethical standards

- "Artificial brains" created that
  - Aren't understood - structures so huge and diverse. There is no "algebra" of DL
  - Aren't evaluated properly, rely on benchmarking (i.e. engineering not cognition-oriented). What about ecological validity?
  - Are often trained with data that isn't optimal (get what you can)
  - Interact with each other and with humans, proliferating
  - Consume unsustainable amounts of energy

Artificial Neuroscience

What’s in a name?

Artificial Neuroscience

What's in a name?

- Neuroscience - several disciplines dealing with the structure, development, function, chemistry, pharmacology, and pathology of the nervous system (the brain, spinal cord, and peripheral nervous system).

Artificial Neuroscience

What's in a name?

- Neuroscience - several disciplines dealing with the structure, development, function, chemistry, pharmacology, and pathology of the nervous system (the brain, spinal cord, and peripheral nervous system).
- Combines physiology, anatomy, molecular biology, developmental biology, cytology, psychology, physics, computer science, chemistry, medicine, statistics, and mathematical modeling to understand the fundamental and emergent properties of neurons, glia and neural circuits

Artificial Neuroscience

What's in a name?

- Neuroscience - several disciplines dealing with the structure, development, function, chemistry, pharmacology, and pathology of the nervous system (the brain, spinal cord, and peripheral nervous system).
- Combines physiology, anatomy, molecular biology, developmental biology, cytology, psychology, physics, computer science, chemistry, medicine, statistics, and mathematical modeling to understand the fundamental and emergent properties of neurons, glia and neural circuits

- Artificial Neuroscience needs an equivalent definition and corresponding set of disciplines going beyond Computer Science.

Artificial Neuroscience

What's in a name?

- Neuroscience - several disciplines dealing with the structure, development, function, chemistry, pharmacology, and pathology of the nervous system (the brain, spinal cord, and peripheral nervous system).
- Combines physiology, anatomy, molecular biology, developmental biology, cytology, psychology, physics, computer science, chemistry, medicine, statistics, and mathematical modeling to understand the fundamental and emergent properties of neurons, glia and neural circuits

- Artificial Neuroscience needs an equivalent definition and corresponding set of disciplines going beyond Computer Science.
- Several branches of mathematics are vital. So are Engineering, VLSI & Circuits, Behavioural Sciences, Humanities, and of course, application domain knowledge

Holistic understanding …

- Artificial MRI
  - Mathematics for observing, measuring & understanding the learning and inference processes by observing and measuring
  - Mechanistic interpretability: exposing emergent structures and neural circuits
- Experimental Artificial Neuroscience
  - Beyond benchmarking: developing and testing behavioural hypotheses in ecologically valid experiments (incl. ablation and “surgery”)
  - Designing test data to fully probe behaviours
  - Exploring failure modes, not just accuracy
- Artificial Cognitive Development
  - Curriculum learning, transfer learning, domain adaptation, etc
- Machine Behavioural Science
  - Applying social sciences to collective behaviours of multiple AIs, AIs + humans

Artificial Psychology

- Critiques familiar practice in DL research
- Proposes methodologies and roles for psychologists
- Appropriate experimentation delivers insights into black-box systems -> XAI

Psychonomic Bulletin and Review (2021) 28:454–475
https://doi.org/10.3758/s13423-020-01825-5

THEORETICAL REVIEW

Artificial cognition: How experimental psychology can help generate explainable artificial intelligence

J. Eric T. Taylor[1,2] · Graham W. Taylor[1,2]

Accepted: 2 October 2020 / Published online: 6 November 2020

Abstract

Artificial intelligence powered by deep neural networks has reached a level of complexity where it can be difficult or impossible to express how a model makes its decisions. This black-box problem is especially concerning when the model makes decisions with consequences for human well-being. In response, an emerging field called explainable artificial intelligence (XAI) aims to increase the interpretability, fairness, and transparency of machine learning. In this paper, we describe how cognitive psychologists can make contributions to XAI. The human mind is also a black box, and cognitive psychologists have over 150 years of experience modeling it through experimentation. We ought to translate the methods and rigor of cognitive psychology to the study of artificial black boxes in the service of explainability. We provide a review of XAI for psychologists, arguing that current methods possess a blind spot that can be complemented by the experimental cognitive tradition. We also provide a framework for research in XAI, highlight exemplary cases of experimentation within XAI inspired by psychological science, and provide a tutorial on experimenting with machines. We end by noting the advantages of an experimental approach and invite other psychologists to conduct research in this exciting new field.

Artificial Psychology #2

- discover shape bias in a Comp Vis system by applying Cog Psych to a DNN.
- hence possibilities of ‘exposing hidden computational properties of DNN’
- Proceedings of the 34 th International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017

Cognitive Psychology for Deep Neural Networks: A Shape Bias Case Study

Samuel Ritter [*1] David G.T. Barrett [*1] Adam Santoro [1] Matt M. Botvinick [1]

Abstract

Deep neural networks (DNNs) have advanced performance on a wide range of complex tasks, rapidly outpacing our understanding of the nature of their solutions. While past work sought to advance our understanding of these models, none has made use of the rich history of problem descriptions, theories, and experimental methods developed by cognitive psychologists to study the human mind. To explore the potential value of these tools, we chose a well-established analysis from developmental psychology that explains how children learn word labels for objects, and applied that analysis to DNNs. Using datasets of stimuli inspired by the original cognitive psychology experiments, we find that state-of-the-art one shot learning models trained on ImageNet exhibit a similar bias to that observed in humans: they prefer to categorize objects according to shape rather than color. The magnitude of this shape bias varies greatly among architecturally identical, but differently seeded models, and even fluctuates within seeds throughout training, despite nearly equivalent classification performance. These results demonstrate the capability of tools from cognitive psychology for exposing hidden computational properties of DNNs, while concurrently providing us with a computational model for human word learning.

Machine Behaviour

- Many citations, none is mathematically oriented
- Argues for social science techniques to be applied to machine intelligence
- Out of MIT. But
  - Lovely web site, though no changes since 2019.

REVIEW

https://doi.org/10.1038/s41586-019-1138-y

Machine behaviour

Iyad Rahwan[1,2,3,34]*, Manuel Cebrian[1,34], Nick Obradovich[1,34], Josh Bongard[4], Jean-François Bonnefon[5], Cynthia Breazeal[1], Jacob W. Crandall[6], Nicholas A. Christakis[7,8,9,10], Iain D. Couzin[11,12,13], Matthew O. Jackson[14,15,16], Nicholas R. Jennings[17,18], Ece Kamar[19], Isabel M. Kloumann[20], Hugo Larochelle[21], David Lazer[22,23,24], Richard McElreath[25,26], Alan Mislove[27], David C. Parkes[28,29], Alex 'Sandy' Pentland[1], Margaret E. Roberts[30], Azim Shariff[31], Joshua B. Tenenbaum[32] & Michael Wellman[33]

Machines powered by artificial intelligence increasingly mediate our social, cultural, economic and political interactions. Understanding the behaviour of artificial intelligence systems is essential to our ability to control their actions, reap their benefits and minimize their harms. Here we argue that this necessitates a broad scientific research agenda to study machine behaviour that incorporates and expands upon the discipline of computer science and includes insights from across the sciences. We first outline a set of questions that are fundamental to this emerging field and then explore the technical, legal and institutional constraints on the study of machine behaviour.

Reasoning in LLMs

- Investigates how deep networks can learn abstract relational reasoning. Model behavior is evaluated against human performance on analogous cognitive tasks

Learning to Reason With Relational Abstractions

Andrew J. Nam*[1], Mengye Ren*[2], Chelsea Finn[1], James L. McClelland[1]
[1]Stanford University, [2]NYU

December 7, 2022

Abstract

Large language models have recently shown promising progress in mathematical reasoning when fine-tuned with human-generated sequences walking through a sequence of solution steps. However, the solution sequences are not formally structured and the resulting model-generated sequences may not reflect the kind of systematic reasoning we might expect an expert human to produce. In this paper, we study how to build stronger reasoning capability in language models using the idea of relational abstractions. We introduce new types of sequences that more explicitly provide an abstract characterization of the transitions through intermediate solution steps to the goal state. We find that models that are supplied with such sequences as prompts can solve tasks with a significantly higher accuracy, and models that are trained to produce such sequences solve problems better than those that are trained with previously used human-generated sequences and other baselines. Our work thus takes several steps toward elucidating and improving how language models perform on tasks requiring multi-step mathematical reasoning.

Interpretability

http://arxiv.org/abs/2208.06894

- Aware of visualisation and auralisation of layers and weights
- Improves on this using formal methods from Linear Algebra
- Links to interpretability but not to controlling network convergence

The SVD of Convolutional Weights: A CNN Interpretability Framework*

Brenda Praggastis† Davis Brown† Carlos Ortiz Marrero‡ Emilie Purvine†
Madelyn Shapiro† Bei Wang§

August 16, 2022

Abstract

Deep neural networks used for image classification often use convolutional filters to extract distinguishing features before passing them to a linear classifier. Most interpretability literature focuses on providing semantic meaning to convolutional filters to explain a model's reasoning process and confirm its use of relevant information from the input domain. Fully connected layers can be studied by decomposing their weight matrices using a singular value decomposition, in effect studying the correlations between the rows in each matrix to discover the dynamics of the map. In this work we define a singular value decomposition for the weight tensor of a convolutional layer, which provides an analogous understanding of the correlations between filters, exposing the dynamics of the convolutional map. We validate our definition using recent results in random matrix theory. By applying the decomposition across the linear layers of an image classification network we suggest a framework against which interpretability methods might be applied using hypergraphs to model class separation. Rather than looking to the activations to explain the network, we use the singular vectors with the greatest corresponding singular values for each linear layer to identify those features most important to the network. We illustrate our approach with examples and introduce the DeepDataProfiler library, the analysis tool used for this study.

Deep Learning Metrology

Journal of Machine Learning Research 22 (2021) 1-73 Submitted 4/20; Published 6/21

Implicit Self-Regularization in Deep Neural Networks: Evidence from Random Matrix Theory and Implications for Learning

Charles H. Martin CHARLES@CALCULATIONCONSULTING.COM
Calculation Consulting
8 Locksley Ave, 6B
San Francisco, CA 94122

Michael W. Mahoney MMAHONEY@STAT.BERKELEY.EDU
ICSI and Department of Statistics
University of California at Berkeley
Berkeley, CA 94720

Editor: Ohad Shamir

Abstract

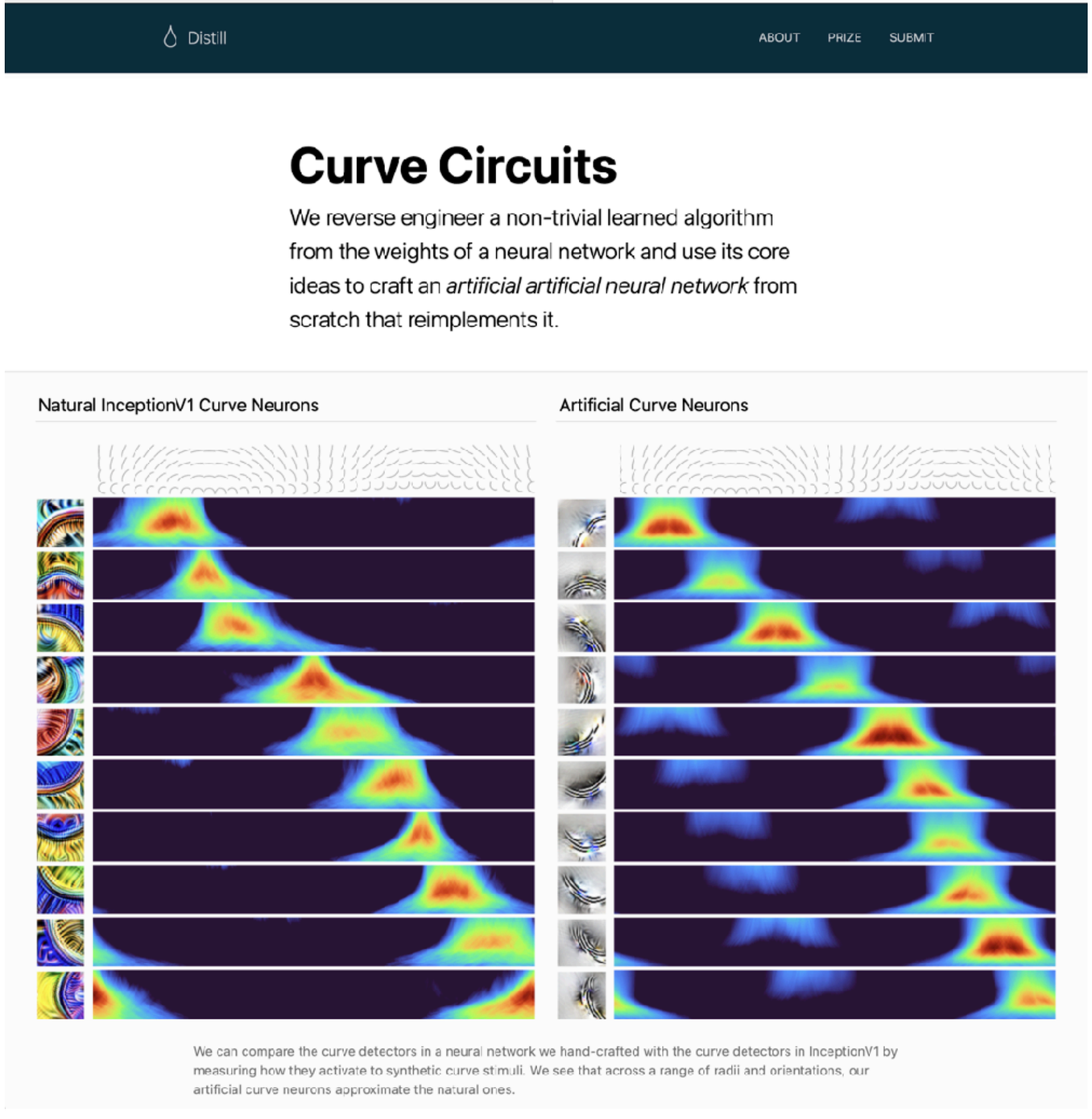
Random Matrix Theory (RMT) is applied to analyze the weight matrices of Deep Neural Networks (DNNs), including both production quality, pre-trained models such as AlexNet and Inception, and smaller models trained from scratch, such as LeNet5 and a miniature-AlexNet. Empirical and theoretical results clearly indicate that the DNN training process itself implicitly implements a form of *Self-Regularization*, implicitly sculpting a more regularized energy or penalty landscape. In particular, the empirical spectral density (ESD) of DNN layer matrices displays signatures of traditionally-regularized statistical models, even in the absence of exogenously specifying traditional forms of explicit regularization, such as Dropout or Weight Norm constraints. Building on relatively recent results in RMT, most notably its extension to Universality classes of Heavy-Tailed matrices, and applying them to these empirical results, we develop a theory to identify *5+1 Phases of Training*, corresponding to increasing amounts of *Implicit Self-Regularization*. These phases can be observed during the training process as well as in the final learned DNNs. For smaller and/or older DNNs, this Implicit Self-Regularization is like traditional Tikhonov regularization, in that there is a "size scale" separating signal from noise. For state-of-the-art DNNs, however, we identify a novel form of *Heavy-Tailed Self-Regularization*, similar to the self-organization seen in the statistical physics of disordered systems (such as classical models of actual neural activity). This results from correlations arising at all size scales, which for DNNs arises implicitly due to the training process itself. This implicit Self-Regularization can depend strongly on the many knobs of the training process. In particular, we demonstrate that we can cause a small model to exhibit all 5+1 phases of training simply by changing the batch size. Our results suggest that large, well-trained DNN architectures should exhibit Heavy-Tailed Self-Regularization, and we discuss the theoretical and practical implications of this.

- Distribution of eigenvalues is heavy tailed in large, well-trained networks
- Various stages of training identified by changing distribution
- Toolbox called ‘weightwatchers’

Discovering functional blocks

https://distill.pub/2020/circuits/curve-circuits/

- Image processing DLs learn curve detectors (and higher order function)
- Replace identified, learning ‘circuits’ with custom designed, low-power/ efficient circuits
- Performance is comparable
- Potential for commoditising DL models



Deep Neural surgery

- https://arxiv.org/abs/2110.11309
- “the largest existing models still make errors”
- “Producing such targeted edits [is] difficult”
- “Propose Model Editor Networks with Gradient Decomposition (MEND)”
- "MEND learns to transform the gradient obtained by standard fine-tuning, using a low-rank decomposition of the gradient to make the parameterization of this transformation tractable.”

FAST MODEL EDITING AT SCALE

Eric Mitchell, Charles Lin, Antoine Bosselut, Chelsea Finn, Christopher D. Manning
Stanford University
eric.mitchell@cs.stanford.edu

ABSTRACT

While large pre-trained models have enabled impressive results on a variety of downstream tasks, the largest existing models still make errors, and even accurate predictions may become outdated over time. Because detecting all such failures at training time is impossible, enabling both developers and end users of such models to correct inaccurate outputs while leaving the model otherwise intact is desirable. However, the distributed, black-box nature of the representations learned by large neural networks makes producing such targeted edits difficult. If presented with only a single problematic input and new desired output, fine-tuning approaches tend to overfit; other editing algorithms are either computationally infeasible or simply ineffective when applied to very large models. To enable easy post-hoc editing at scale, we propose Model Editor Networks with Gradient Decomposition (MEND), a collection of small auxiliary editing networks that use a single desired input-output pair to make fast, local edits to a pre-trained model’s behavior. MEND learns to transform the gradient obtained by standard fine-tuning, using a low-rank decomposition of the gradient to make the parameterization of this transformation tractable. MEND can be trained on a single GPU in less than a day even for 10 billion+ parameter models; once trained MEND enables rapid application of new edits to the pre-trained model. Our experiments with T5, GPT, BERT, and BART models show that MEND is the only approach to model editing that effectively edits the behavior of models with more than 10 billion parameters. Code and data available at https://sites.google.com/view/mend-editing.

Simplifying computation

Learning Low-rank Deep Neural Networks via Singular Vector Orthogonality Regularization and Singular Value Sparsification

Huanrui Yang[1], Minxue Tang[2], Wei Wen[1], Feng Yan[3], Daniel Hu[4], Ang Li[1], Hai Li[1], Yiran Chen[1]
[1]Department of Electrical and Computer Engineering, Duke University
[2]Department of Electronic Engineering, Tsinghua University
[3]Computer Science and Engineering Department, University of Nevada, Reno
[4]Newport High School, Bellevue, WA

http://arxiv.org/abs/2004.09031

Abstract

*Modern deep neural networks (DNNs) often require high memory consumption and large computational loads. In order to deploy DNN algorithms efficiently on edge or mobile devices, a series of DNN compression algorithms have been explored, including factorization methods. Factorization methods approximate the weight matrix of a DNN layer with the multiplication of two or multiple low-rank matrices. However, it is hard to measure the ranks of DNN layers during the training process. Previous works mainly induce low-rank through implicit approximations or via costly singular value decomposition (SVD) process on every training step. The former approach usually induces a high accuracy loss while the latter has a low efficiency. In this work, we propose **SVD training**, the first method to explicitly achieve low-rank DNNs during training without applying SVD on every step. SVD training first decomposes each layer into the form of its full-rank SVD, then performs training directly on the decomposed weights. We add orthogonality regularization to the singular vectors, which ensure the valid form of SVD and avoid gradient vanishing/exploding. Low-rank is encouraged by applying sparsity-inducing regularizers on the singular values of each layer. Singular value pruning is applied at the end to explicitly reach a low-rank model. We empirically show that SVD training can significantly reduce the rank of DNN layers and achieve higher reduction on computation load under the same accuracy, comparing to not only previous factorization methods but also state-of-the-art filter pruning methods.*

AI research harms the planet

- R. Couillet, D. Trystram and T. Ménissier, "The Submerged Part of the AI-Ceberg [Perspectives]," in IEEE Signal Processing Magazine, vol. 39, no. 5, pp. 10-17, Sept. 2022, doi: 10.1109/MSP.2022.3182938.
- Looking at energy consumption due to Deep Learning

Romain Couillet, Denis Trystram, and Thierry Ménissier

The Submerged Part of the AI-Ceberg

This article discusses the contradiction between the exploding energy demand of artificial intelligence (AI) and the information and communication (ICT) industry as a whole and the parallel strong request for energy sobriety imposed by the need to mitigate the impact of climate change and the anticipated collapse of civilization as we know it. Under the form of an open reflection on the goods and evils of AI, the article raises the suggestion of a drastic change in the AI paradigm, more in phase with the vital obligation to design a more resilient society.

Deep learning: The new Eldorado?

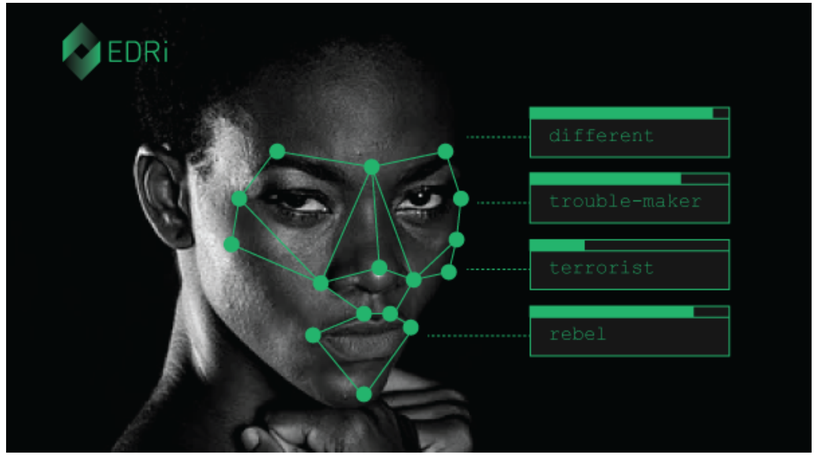
Over the past decade, the considerable growth of the digital world, propelled by AI, has had spectacular effects in a few scientific fields, such as computer vision and natural language processing, and given rise to many new technologies and consumer products. Today, this development even claims to revolutionize many other areas of our society. This revolution indeed concerns many aspects of our lives: we (and world with a few clicks, to name only a few [1], [2].

Deep neural network learning is at the forefront of this development and has spread rapidly, far beyond the confidential fields of its beginnings. In a matter of 10 years, this specific computer science tool—theorized as early as the 1980s [3]—has reached all levels of society: in companies, institutions, research laboratories, in virtually all engineering disciplines as well as life sciences. Easy to use as a black box thanks to an important software development effort—multiple "plug-and-play" solutions have been developed for engineers (and not only computer science experts), such as the popular TensorFlow library [4], [5]—deep learning has effectively replaced "conventional" tools (particularly in computer vision and natural language processing), imposing a form of radical monopoly on scientific domains. The radical monopoly of a tool is understood in the sense defined by Illich [34]: it alters the normative system of knowledge generation and sharing. Calls for projects, dedicated conferences, and job world really be on the way [6]? Of course, investing in deep learning and AI involves delegating to the machine the power of our decisions, which comes with many ethics and equity concerns [8]; as Stephen Hawking pessimistically stated in 2014, "The development of full artificial intelligence could spell the end of the human race.… It would take off on its own, and redesign itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded." [7] (As discussed next, this seemingly science-fictional statement is more profoundly explored by Illich [34] regarding the dangers of societal dependence on oil and machines, induced by an increasing loss of common knowledge and know-how that are moved from the population to computers and machines.) Yet, the many promises of AI clearly tip the scales toward increasingly more investment in the field [10]. Besides, researchers now deeply investigate the question of fairness in AI to smooth out these thorny angles [9].

Agency and AI
AI and Creativity

Human Rights vs the Machine

Image from: https://edri.org/our-work/facial-recognition-and-fundamental-rights-101/

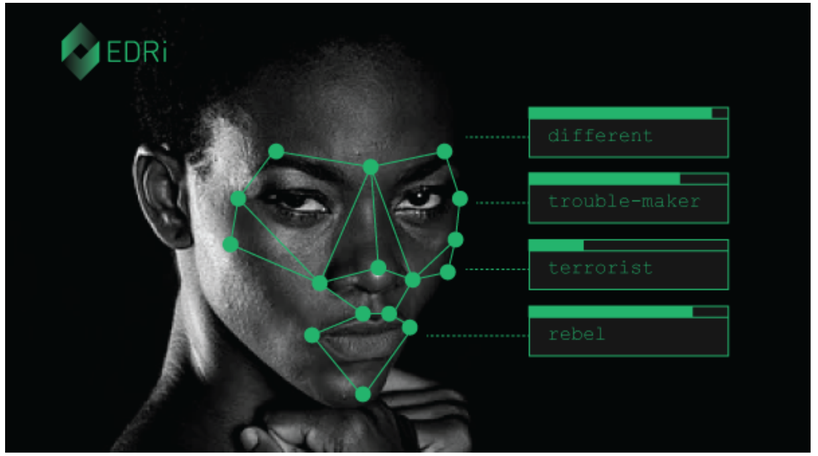


Agency and AI
AI and Creativity

Human Rights vs the Machine

- Links among algorithmic bias, copyright & ethics

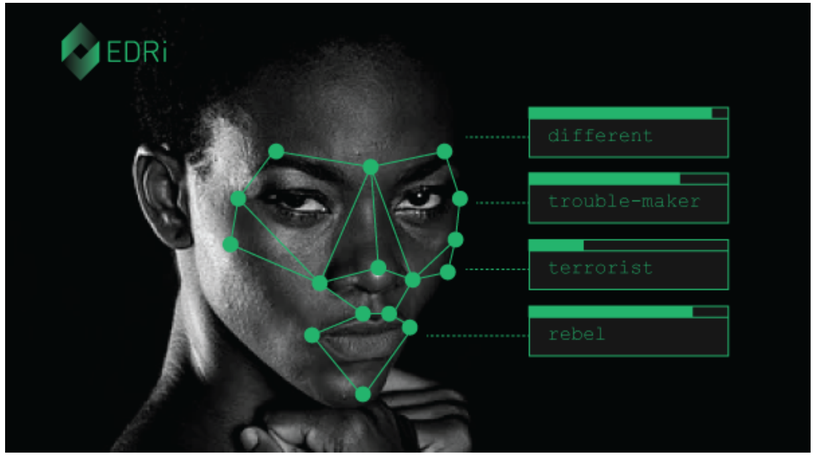
Image from: https://edri.org/our-work/facial-recognition-and-fundamental-rights-101/



Agency and AI
AI and Creativity

Human Rights vs the Machine

Image from: https://edri.org/our-work/facial-recognition-and-fundamental-rights-101/

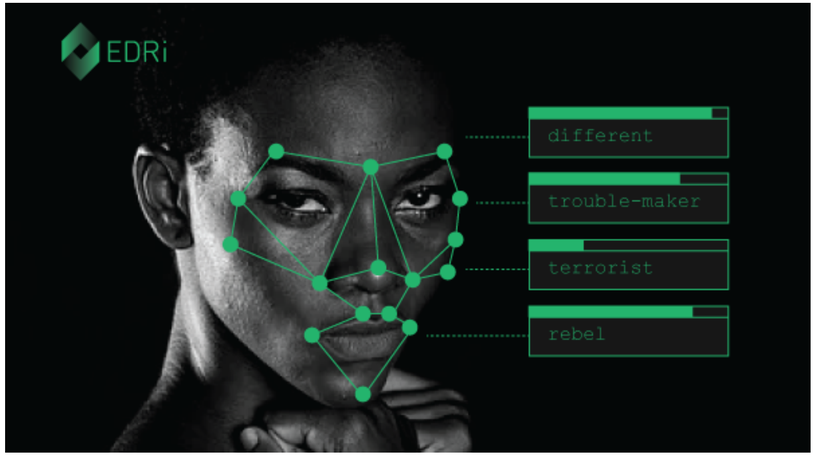


- Links among algorithmic bias, copyright & ethics
- Where to draw the line between human & machine in a cyber physical system?

Agency and AI
AI and Creativity

Human Rights vs the Machine

Image from: https://edri.org/our-work/facial-recognition-and-fundamental-rights-101/

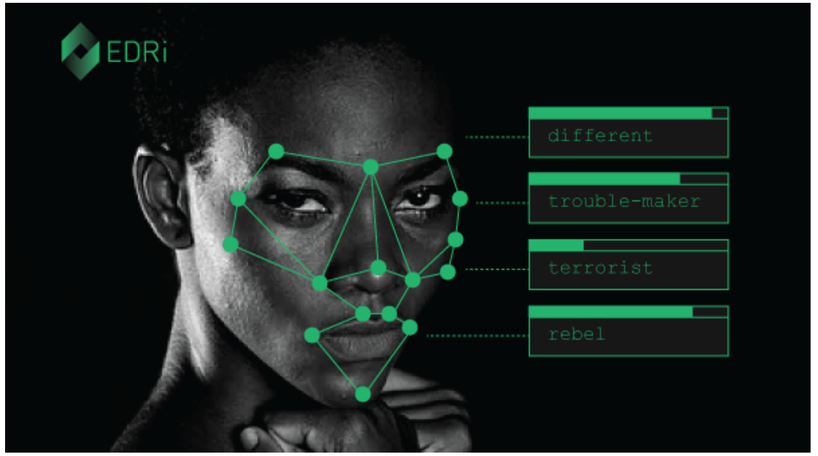


- Links among algorithmic bias, copyright & ethics
- Where to draw the line between human & machine in a cyber physical system?
- Do humans benefit by having to learn how to drive a system, thereby endowing agency?

Agency and AI
AI and Creativity

Human Rights vs the Machine

Image from: https://edri.org/our-work/facial-recognition-and-fundamental-rights-101/

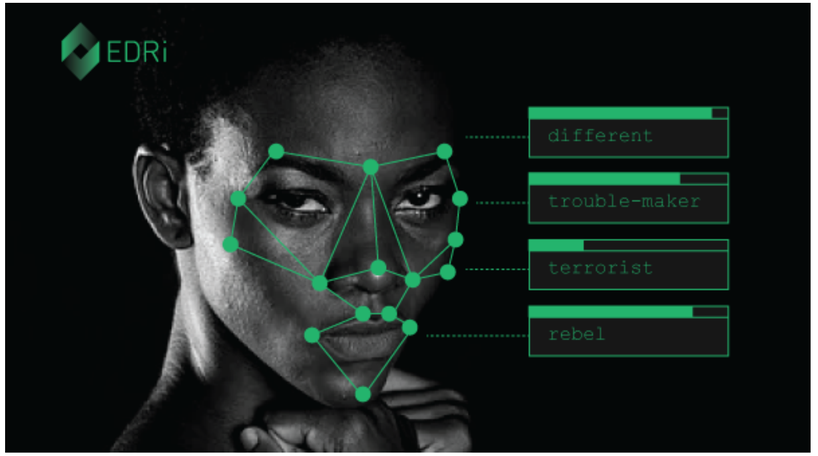


- Links among algorithmic bias, copyright & ethics
- Where to draw the line between human & machine in a cyber physical system?
- Do humans benefit by having to learn how to drive a system, thereby endowing agency?
- In a vocal percussion system to drive a drum synthesiser… does the system adapt to the person, or the person to the system?

Agency and AI
AI and Creativity

Human Rights vs the Machine

Image from: https://edri.org/our-work/facial-recognition-and-fundamental-rights-101/



- Links among algorithmic bias, copyright & ethics
- Where to draw the line between human & machine in a cyber physical system?
- Do humans benefit by having to learn how to drive a system, thereby endowing agency?
- In a vocal percussion system to drive a drum synthesiser… does the system adapt to the person, or the person to the system?
- In AI song-writing, does the machine suggest rhymes or write lines?

My own work

Supported by EPSRC Discipline Hopping Grant

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!
- I’ve hopped from EECS/c4dm to Maths

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!
- I’ve hopped from EECS/c4dm to Maths
- 18 month research agenda

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!
- I’ve hopped from EECS/c4dm to Maths
- 18 month research agenda
  - Explore & develop metrics for the internal state of a NN (during training)

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!
- I’ve hopped from EECS/c4dm to Maths
- 18 month research agenda
  - Explore & develop metrics for the internal state of a NN (during training)
  - Initialise NNs using low rank layers

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!
- I’ve hopped from EECS/c4dm to Maths
- 18 month research agenda
  - Explore & develop metrics for the internal state of a NN (during training)
  - Initialise NNs using low rank layers
  - Modify NN dynamics by tinkering with Singular Values

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!
- I’ve hopped from EECS/c4dm to Maths
- 18 month research agenda
  - Explore & develop metrics for the internal state of a NN (during training)
  - Initialise NNs using low rank layers
  - Modify NN dynamics by tinkering with Singular Values
  - Speed up training

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!
- I’ve hopped from EECS/c4dm to Maths
- 18 month research agenda
  - Explore & develop metrics for the internal state of a NN (during training)
  - Initialise NNs using low rank layers
  - Modify NN dynamics by tinkering with Singular Values
  - Speed up training
  - Explore on a Neural Audio model

My own work

Supported by EPSRC Discipline Hopping Grant

- I was recommended to watch LA + DL lectures by Gilbert Strang. A revelation!
- I’ve hopped from EECS/c4dm to Maths
- 18 month research agenda
  - Explore & develop metrics for the internal state of a NN (during training)
  - Initialise NNs using low rank layers
  - Modify NN dynamics by tinkering with Singular Values
  - Speed up training
  - Explore on a Neural Audio model
  - Develop a larger research agenda

Conclusions…

Holistic understanding …

- Artificial MRI
  - Linear Algebra, statistical mechanics non-linear dynamics, topology, geometry for
    - observing, measuring & understanding learning and inference processes
    - Manipulating the learning process to derive more efficient models
    - Mechanistic interpretability: exposing emergent structures and neural circuits
- Experimental Artificial Neuroscience & Artificial Cognition
  - Beyond benchmarking: developing and testing behavioural hypotheses in ecologically valid experiments (incl. ablation and “surgery”)
  - Designing test data to fully probe behaviours
  - Exploring failure modes, not just accuracy
- Artificial Cognitive Development
  - Curriculum learning, transfer learning, domain adaptation, etc
- Machine Behavioural Science
  - Applying social sciences to collective behaviours of multiple AIs, AIs + humans, role of humanities

…and a new approach to evaluation…

…and a new approach to evaluation…

- Explore failure modes

…and a new approach to evaluation…

- Explore failure modes
  - success/failure is not a binary decision

…and a new approach to evaluation…

- Explore failure modes
  - success/failure is not a binary decision
  - It’s where to learn lessons

…and a new approach to evaluation…

- Explore failure modes
  - success/failure is not a binary decision
  - It’s where to learn lessons
- Develop artificial cognition experimental procedures to complement benchmarking

…and a new approach to evaluation…

- Explore failure modes
  - success/failure is not a binary decision
  - It’s where to learn lessons
- Develop artificial cognition experimental procedures to complement benchmarking
- and relation to Mathematical Neuroscience

…and a new approach to evaluation…

- Explore failure modes
  - success/failure is not a binary decision
  - It’s where to learn lessons
- Develop artificial cognition experimental procedures to complement benchmarking
- and relation to Mathematical Neuroscience
- Leading to safer, explainable AI

… leading to better engineering…

… leading to better engineering…

- Neural Architecture Search and a new design algebra for semi-automated model generation and custom implementation
  - GPUs are general purpose AI chips! We can do better

… leading to better engineering…

- Neural Architecture Search and a new design algebra for semi-automated model generation and custom implementation
  - GPUs are general purpose AI chips! We can do better
- Low rank structures and algorithms for efficient learning and inference
  - Relationship with scalability and data set size
  - explore other fast matrix-vector techniques

… leading to better engineering…

- Neural Architecture Search and a new design algebra for semi-automated model generation and custom implementation
  - GPUs are general purpose AI chips! We can do better
- Low rank structures and algorithms for efficient learning and inference
  - Relationship with scalability and data set size
  - explore other fast matrix-vector techniques
- Custom hardware (incl. 1 bit – well-established in Signal Processing)
  - Relationship between 1 bit processing, ‘oversized’ layers & Universality

… leading to better engineering…

- Neural Architecture Search and a new design algebra for semi-automated model generation and custom implementation
  - GPUs are general purpose AI chips! We can do better
- Low rank structures and algorithms for efficient learning and inference
  - Relationship with scalability and data set size
  - explore other fast matrix-vector techniques
- Custom hardware (incl. 1 bit – well-established in Signal Processing)
  - Relationship between 1 bit processing, ‘oversized’ layers & Universality
- Re-engineer identified neural circuits (Mechanistic Interpretability) with purpose-defined sub-systems (e.g. curve detectors)
  - Towards a building block approach to neural networks
  - Relationships between circuits and low rank

… leading to better engineering…

- Neural Architecture Search and a new design algebra for semi-automated model generation and custom implementation
  - GPUs are general purpose AI chips! We can do better
- Low rank structures and algorithms for efficient learning and inference
  - Relationship with scalability and data set size
  - explore other fast matrix-vector techniques
- Custom hardware (incl. 1 bit – well-established in Signal Processing)
  - Relationship between 1 bit processing, ‘oversized’ layers & Universality
- Re-engineer identified neural circuits (Mechanistic Interpretability) with purpose-defined sub-systems (e.g. curve detectors)
  - Towards a building block approach to neural networks
  - Relationships between circuits and low rank
- Borrow from signal processing for new approaches, but don’t junk CMOS
  - virtual analog for digital equivalents to analog models of biological neural circuits: trainability ~ optimized circuit design
  - Non-linear Digital Wave Filters as compact, non-linear, convolutional building blocks in new, heterogeneous DL models