2023 November 29

Deep Learning for Music Analysis and Generation

DDSP & Automatic Mixing

(audio → audio)

Yi-Hsuan Yang Ph.D.
yhyangtw@ntu.edu.tw

Outline

- Differentiable digital signal processing (DDSP)
  - Uses a neural network to convert a user's input into complex DSP controls that can produce realistic signals
- MIDI-DDSP
- Automatic mixing

Demo: Tone Transfer

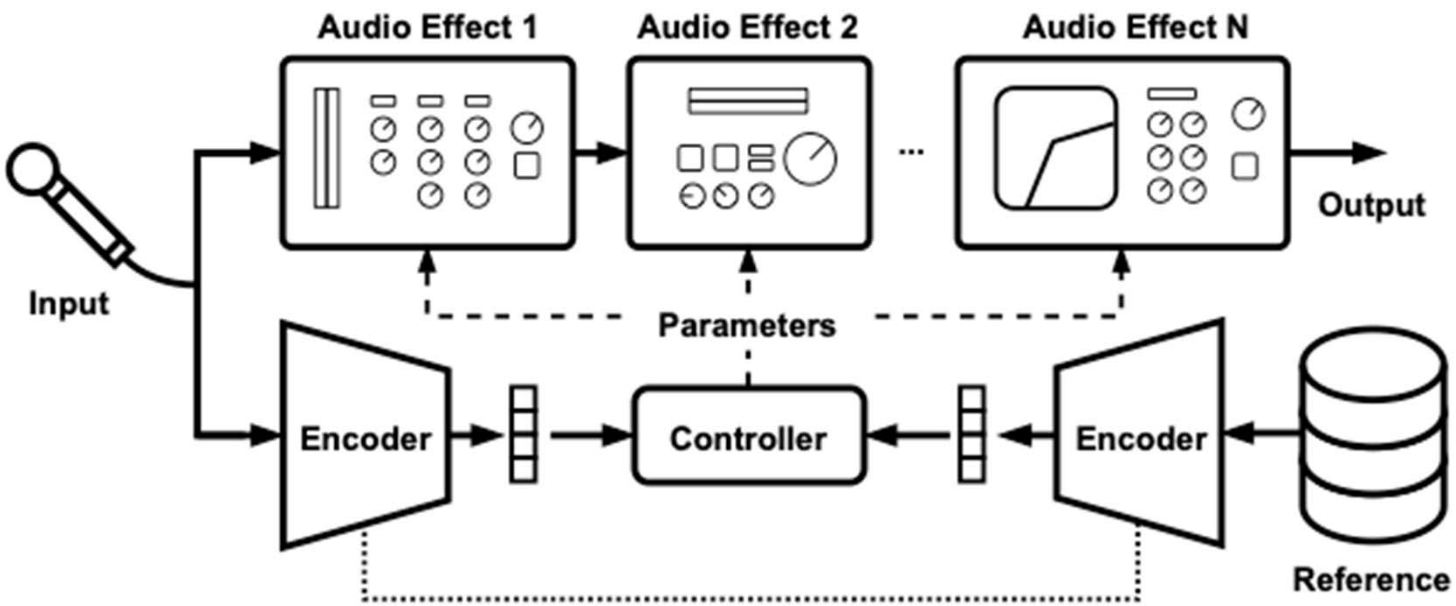
(Make music easier to play with)

https://sites.research.google/tonetransfer

Demo: Mixing Style Transfer

(Make musicians' life easier)

https://csteinmetz1.github.io/DeepAFx-ST/



Reference 1: ISMIR 2023 Tutorial

https://intro2ddsp.github.io/intro.html

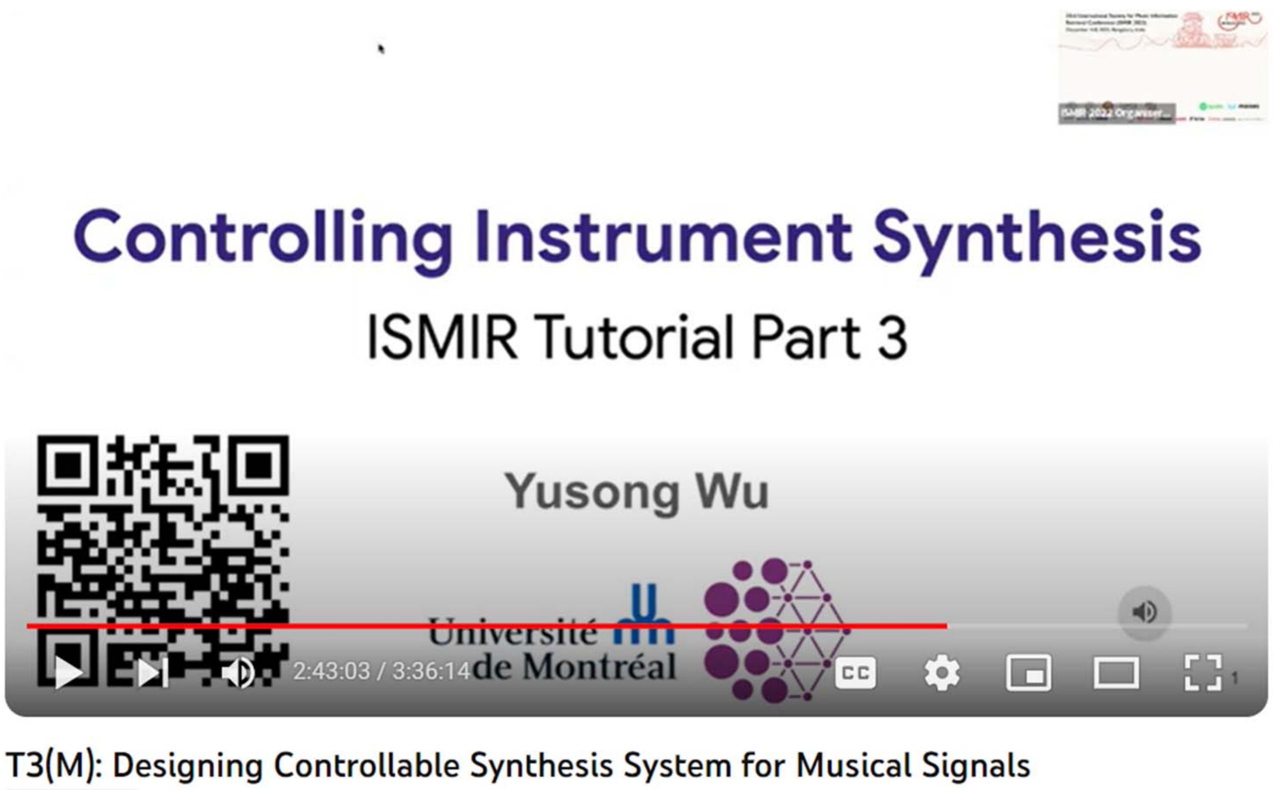
https://github.com/intro2ddsp/intro2ddsp.github.io

https://docs.google.com/presentation/d/1o9RWWmKX0yVVQii4-dtH3OlGZwqrfhEDgLo3582JnfM/edit#slide=id.p

Reference 2: ISMIR 2022 Tutorial

https://github.com/lukewys/ISMIR2022-tutorial

https://youtu.be/7U-zDL5con8?si=HcD7YDN66YPlyGCN&t=9783



Reference 3: ISMIR 2022 Tutorial

https://dl4am.github.io/tutorial/landing-page.html

\

https://docs.google.com/presentation/d/1UhBZucbGnX-ItOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.p

Deep Learning for Automatic Mixing

This is a web book written for a tutorial session of the 23rd International Society for Music Information Retrieval Conference, Dec 4-8, 2022 held at Bengaluru, India in hybrid format. The ISMIR conference is the world's leading research forum on processing, searching, organising and accessing music-related data.

Outline

- **Differentiable digital signal processing (DDSP)**
  - https://intro2ddsp.github.io/intro.html

- MIDI-DDSP

- Automatic mixing

DSP & Audio Synthesis

https://intro2ddsp.github.io/background/neural-audio-synthesis.html

What Is DDSP?

https://intro2ddsp.github.io/background/what-is-ddsp.html

- “For example, a neural network might **output a value which is used as the cutoff frequency of a filter**, which is implemented differentiably”
- “During training, a loss function is computed on the output of the filter and, using the backpropogation algorithm, its gradient with respect to the neural network’s parameters is computed.”
- “In order to perform this computation, the derivative of the filter’s output with respect to its cutoff frequency must be evaluated. That is to say, the filter forms a part of the computation graph, and its gradient is a factor of the chain rule decomposition of the loss gradient.”

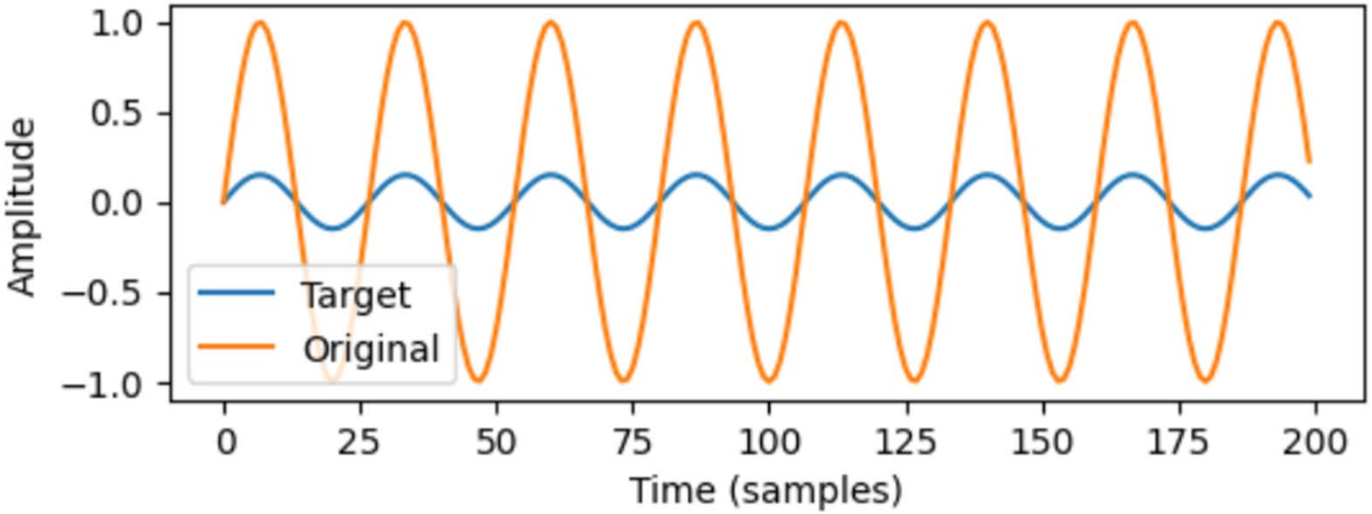
Why DDSP?

https://intro2ddsp.github.io/background/what-is-ddsp.html

1. We have prior knowledge about the class of signal we are interested in
2. We wish to infer the parameters of a particular signal processor or signal model
3. We are concerned about inference-time latency
4. We wish to allow human control over model outputs

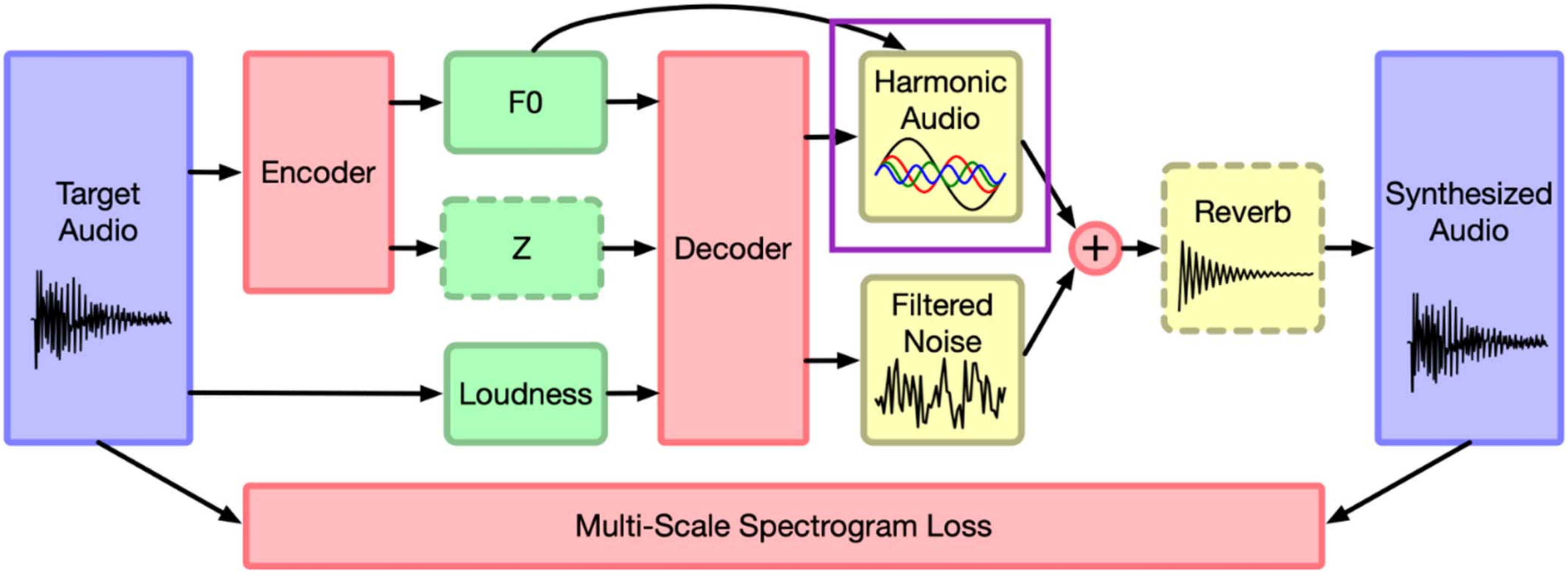
A Differentiable Gain Control

https://intro2ddsp.github.io/first-steps/diff_gain.html



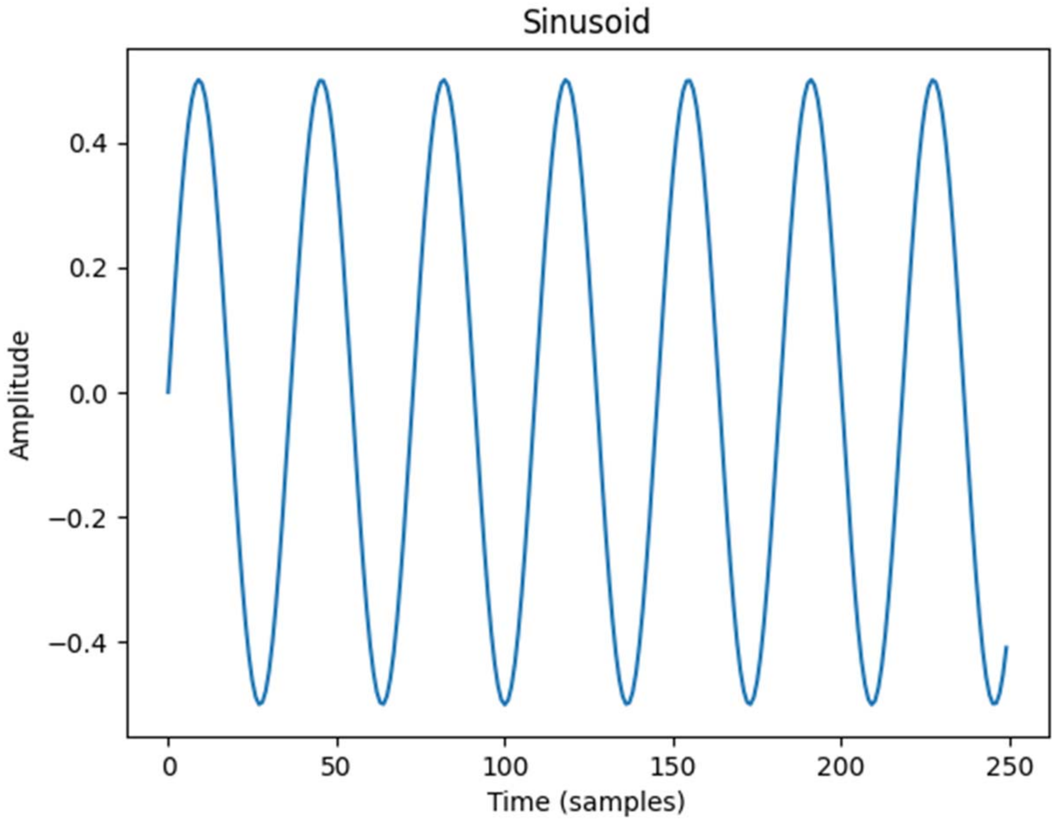
Sinusoidal Modelling Synthesis

https://intro2ddsp.github.io/synths/introduction.html



Writing a Differentiable Oscillator in PyTorch

https://intro2ddsp.github.io/synths/oscillator.html



Optimizing Parameters for the Differentiable Oscillator

https://intro2ddsp.github.io/synths/oscillator.html

- Optimizing amplitude → easy
- Optimizing frequency → difficult due to many local minima

Additive Synthesis

https://intro2ddsp.github.io/synths/additive.html

$$y[n] = \sum_{k}^{K} \alpha_k[n] \sin\left(\phi_k + \sum_{m=0}^{n} \omega_k[m]\right)$$

$$y[n] = \sum_{k=1}^{K} \alpha_k[n] \sin\left(\phi_k + k\sum_{m=0}^{n} \omega_0[m]\right)$$

$$\sum_{k=1}^{K} \hat{\alpha}_k[n] = 1 \text{ and } \hat{\alpha}_k[n] > 0$$

$$y[n] = A[n]\sum_{k=1}^{K} \hat{\alpha}_k[n] \sin\left(k\sum_{m=0}^{n} \omega_0[m]\right)$$

Harmonic Synthesizer

https://intro2ddsp.github.io/synths/harmonic_optimize.html

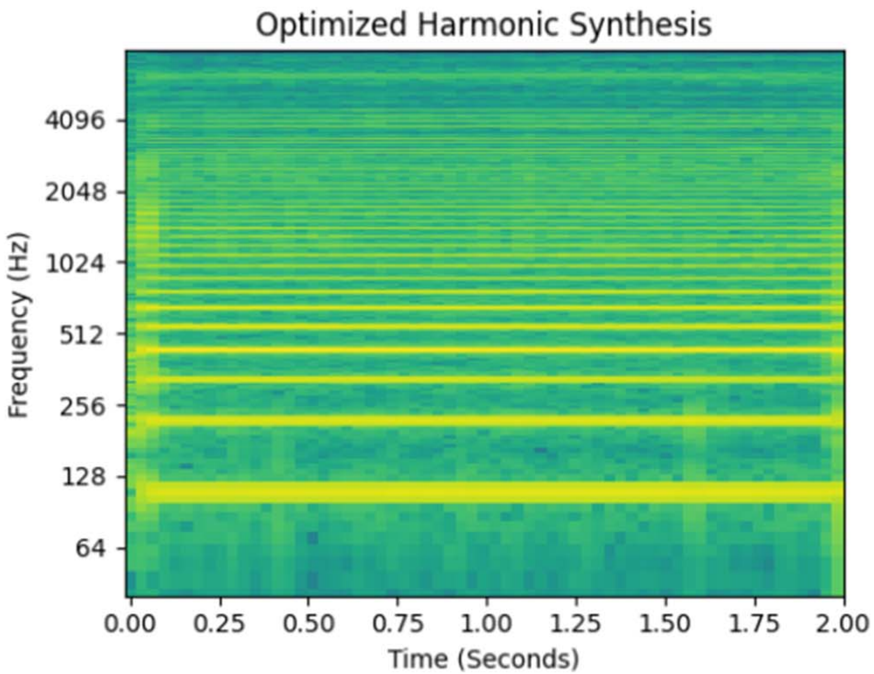
1. Constraining harmonic amplitudes to sum to one
2. Adding a global amplitude parameter
3. Parameter scaling to constrain the possible range of amplitudes
4. Removing frequencies above the Nyquist frequency which will result in aliasing

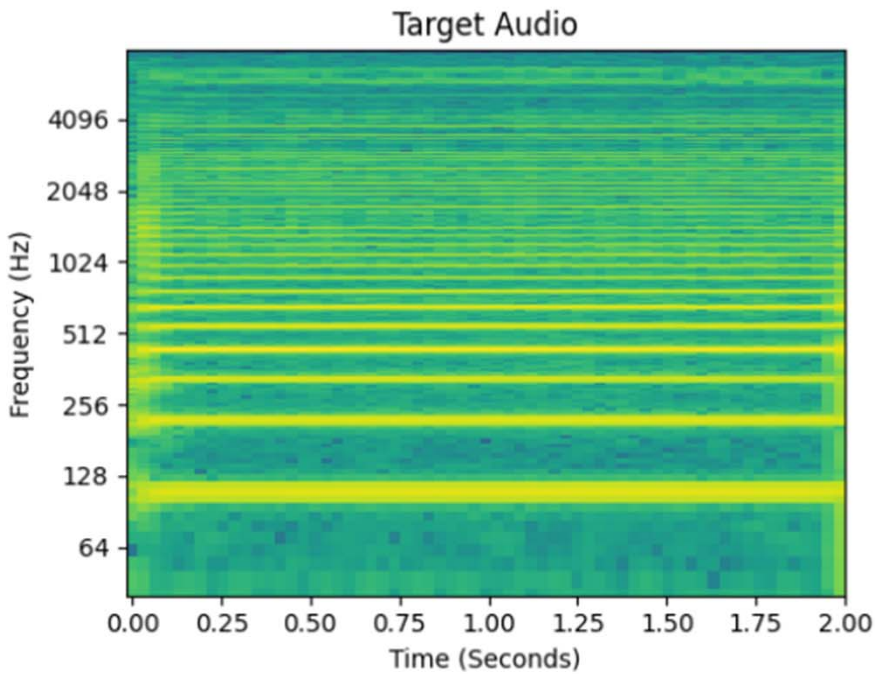
$$y[n] = A[n] \sum_{k=1}^{K} \hat{\alpha}_k[n] \sin\left(k \sum_{m=0}^{n} \omega_0[m]\right)$$

Optimizing a Harmonic Synthesizer

https://intro2ddsp.github.io/synths/harmonic_optimize.html

https://intro2ddsp.github.io/synths/harmonic_results.html





Differentiable Synthesis Libraries

https://intro2ddsp.github.io/synths/libraries.html

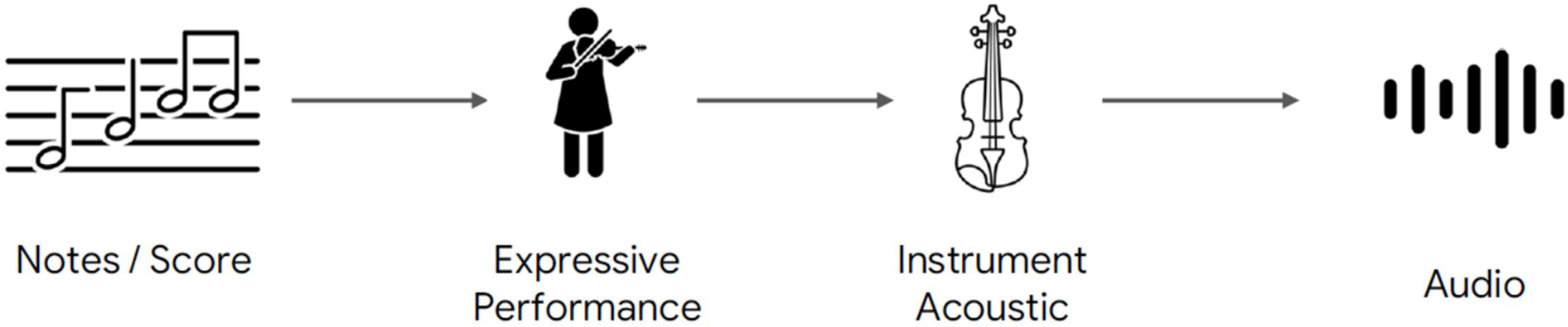
- https://github.com/magenta/ddsp
- https://github.com/acids-ircam/ddsp_pytorch
- https://github.com/torchsynth/torchsynth
- https://github.com/PapayaResearch/synthax
- https://github.com/csteinmetz1/dasp-pytorch

Outline

- Differentiable digital signal processing (DDSP)
  - Uses a neural network to convert a user's input into complex DSP controls that can produce realistic signals
- **MIDI-DDSP** **(ICLR'22)**
  https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925
- Automatic mixing

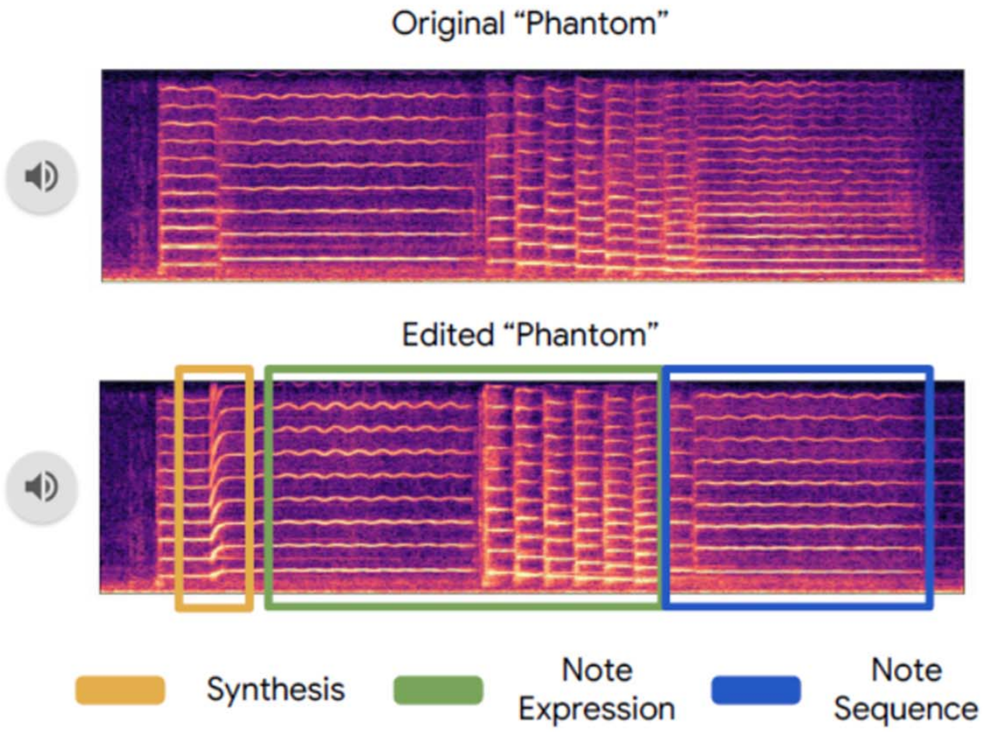
Human Instrument Performing Process

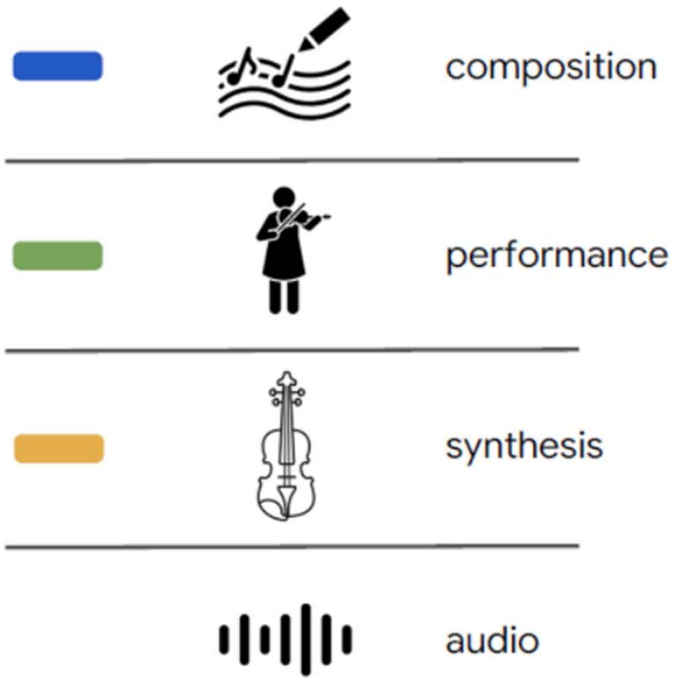
https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925



Controlling Instrument Synthesis

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925

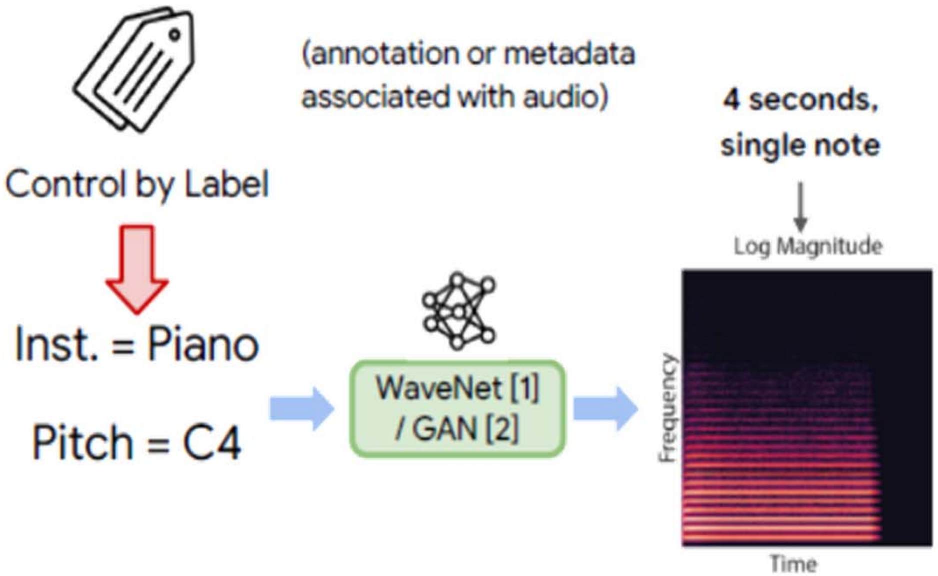




Note to Audio

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925

- Generate single note condition on **pitch and instrument**.
- Modelling mainly **timbre**.

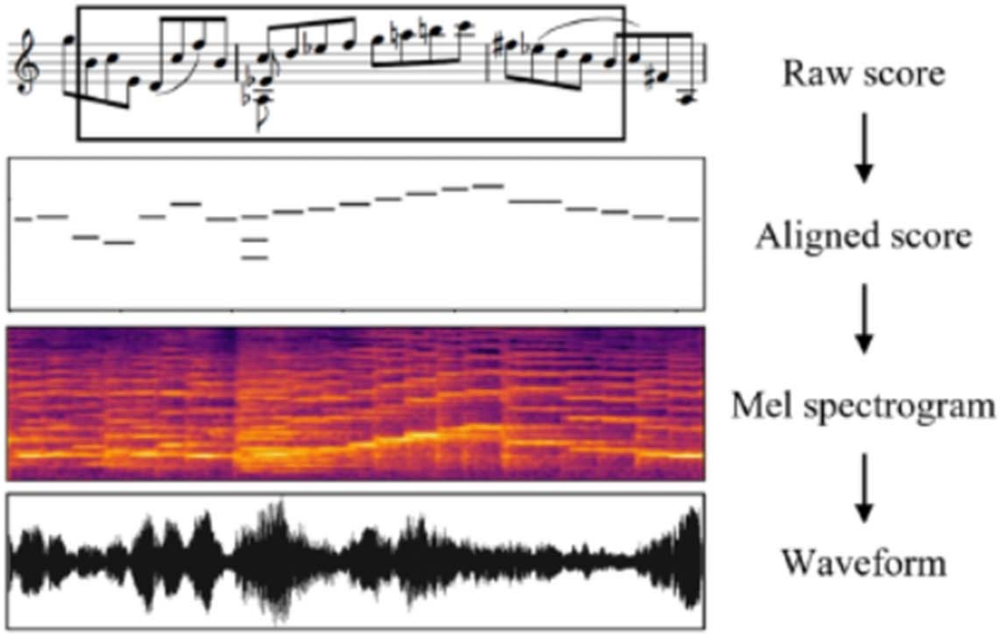


NSynth [1] & GANSynth [2]

Score to Audio

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925

- Music score → Audio
- Generates **timbre** and **expressive performance** together.



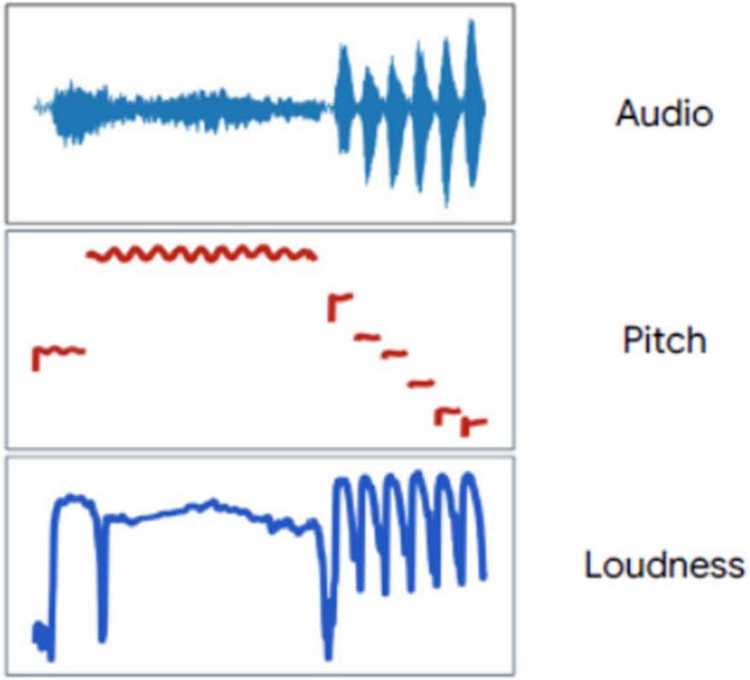
Deep Performer [3]

Other Aspects of Instrument Synthesis

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925

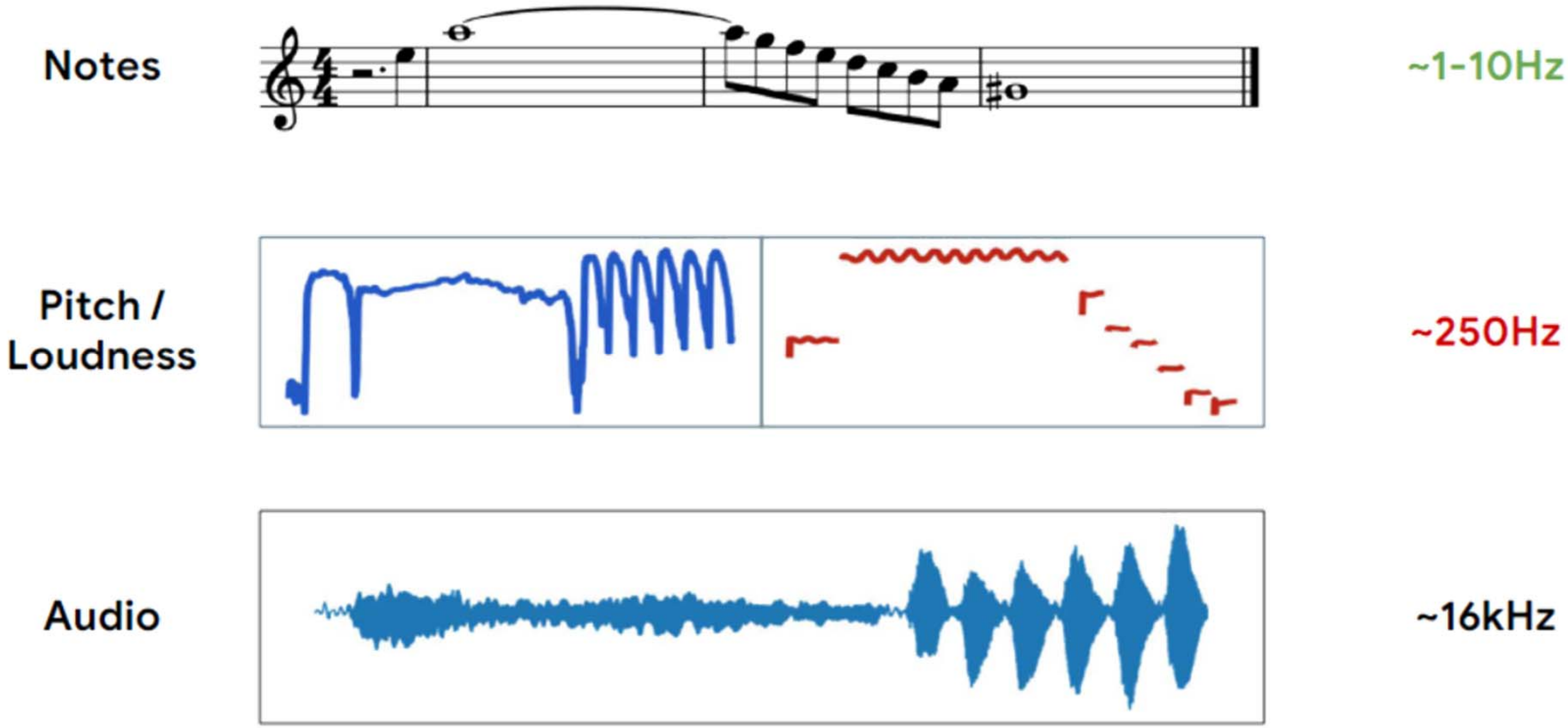
- **Low-level** quantities that changes **frequently**.
- E.g: pitch, loudness, expressive performance, etc.
- **No "labels"** available.

*labels: annotation or metadata associated with audio



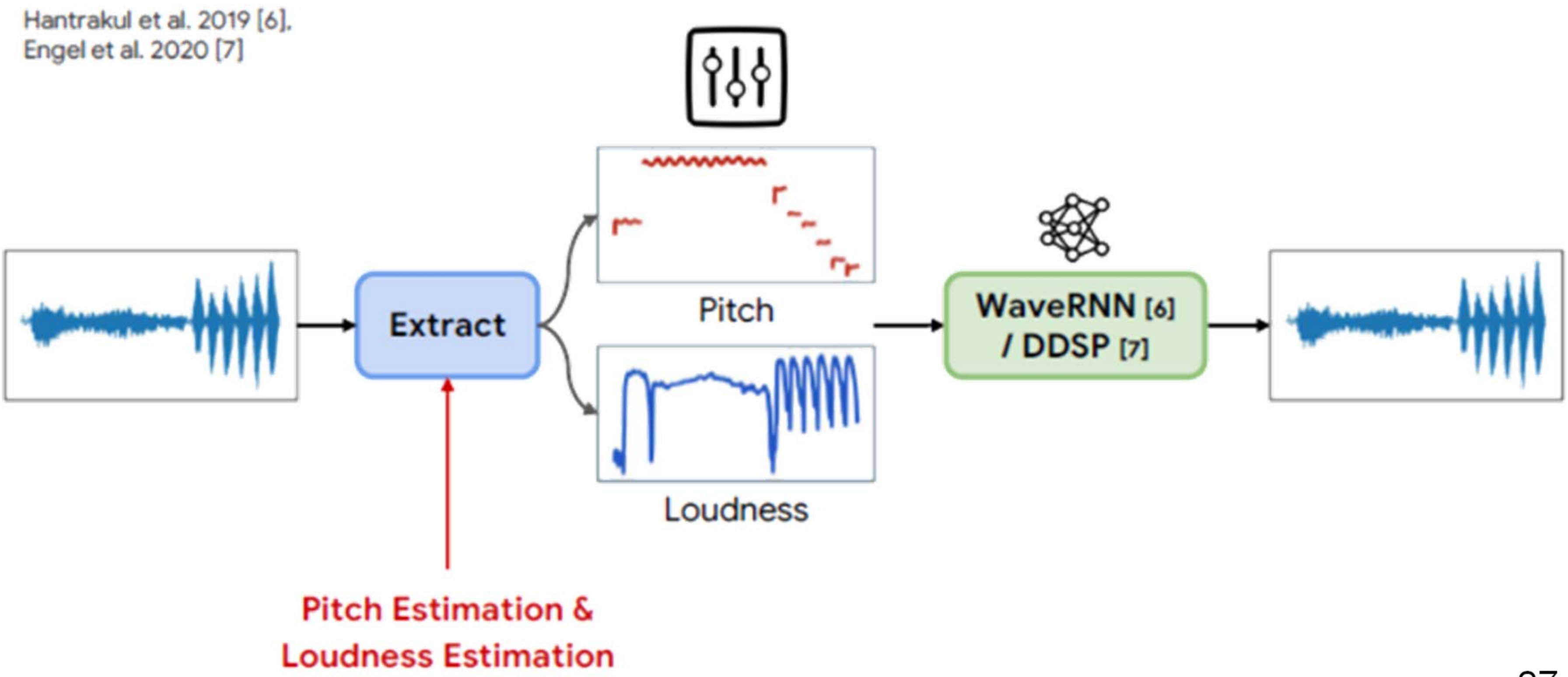
Low-level Quantities

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925



Extract the Label: Pitch and Loudness

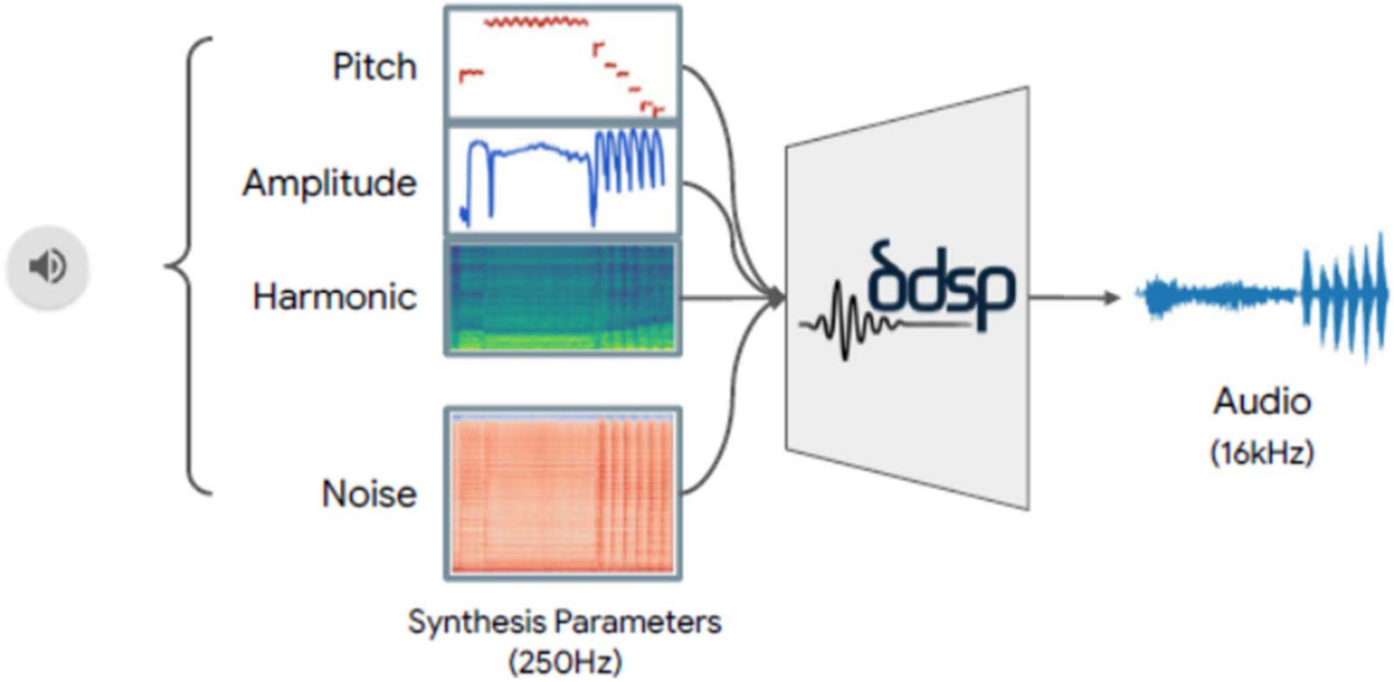
https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925



Learn to Extract Synthesis Parameters: DDSP

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925

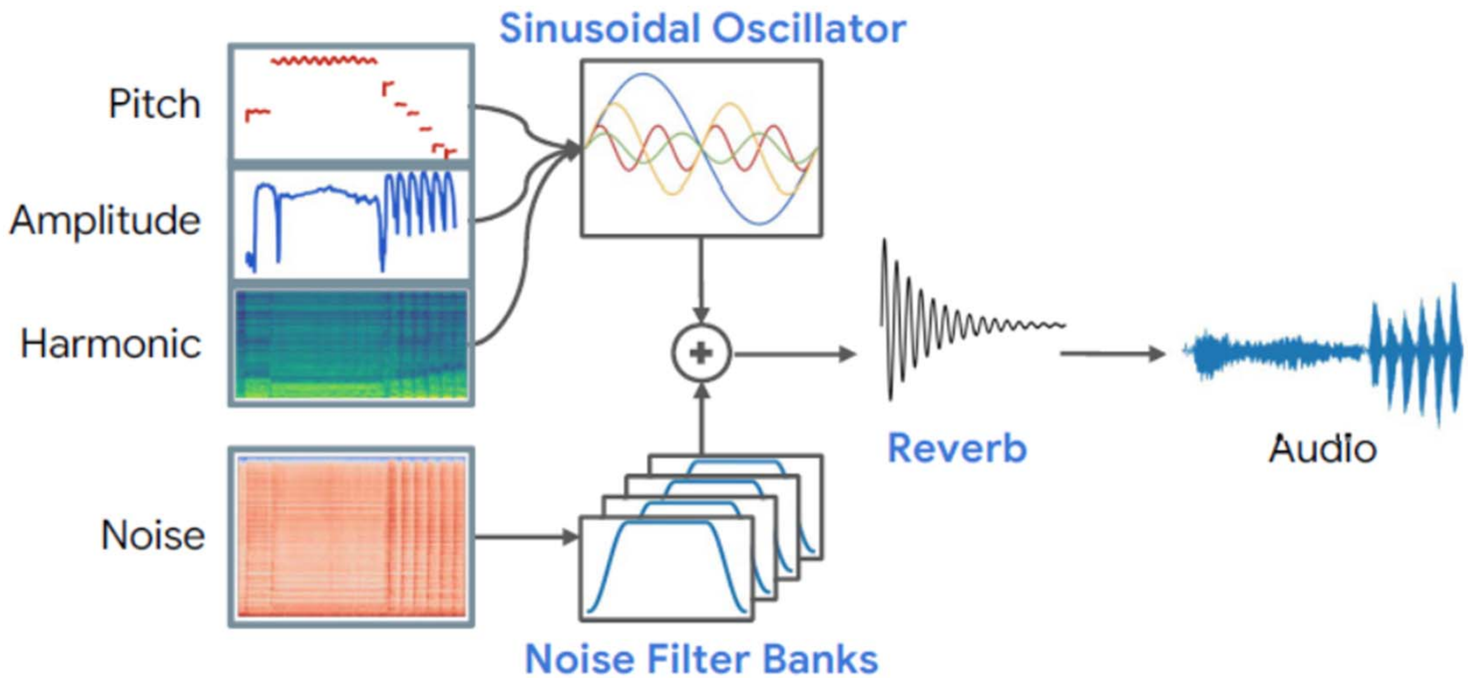
Engel et al. 2020 [7]



DDSP: Differentiable Digital Signal Processing

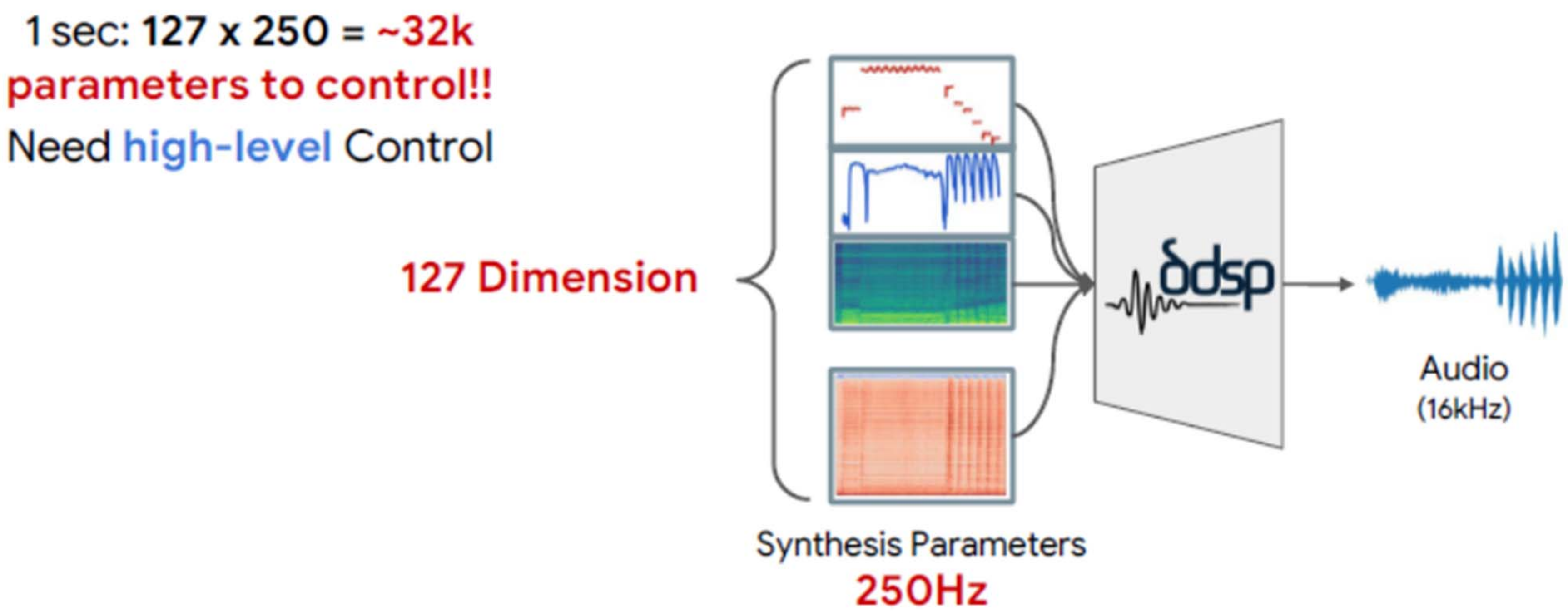
https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925

Engel et al. 2020 [7]



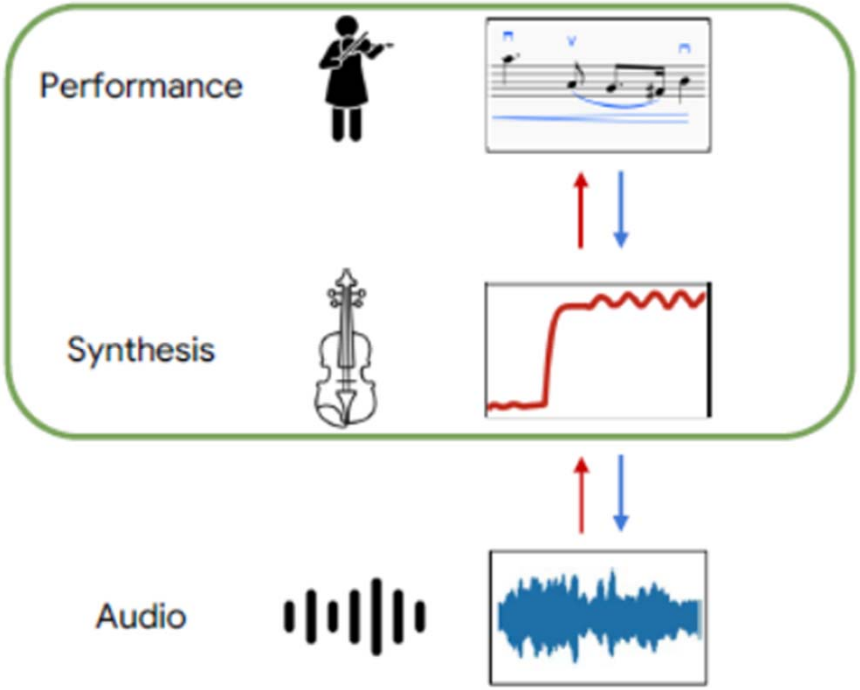
Problem of Low-level Control

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925



Extract Performance Parameter

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925

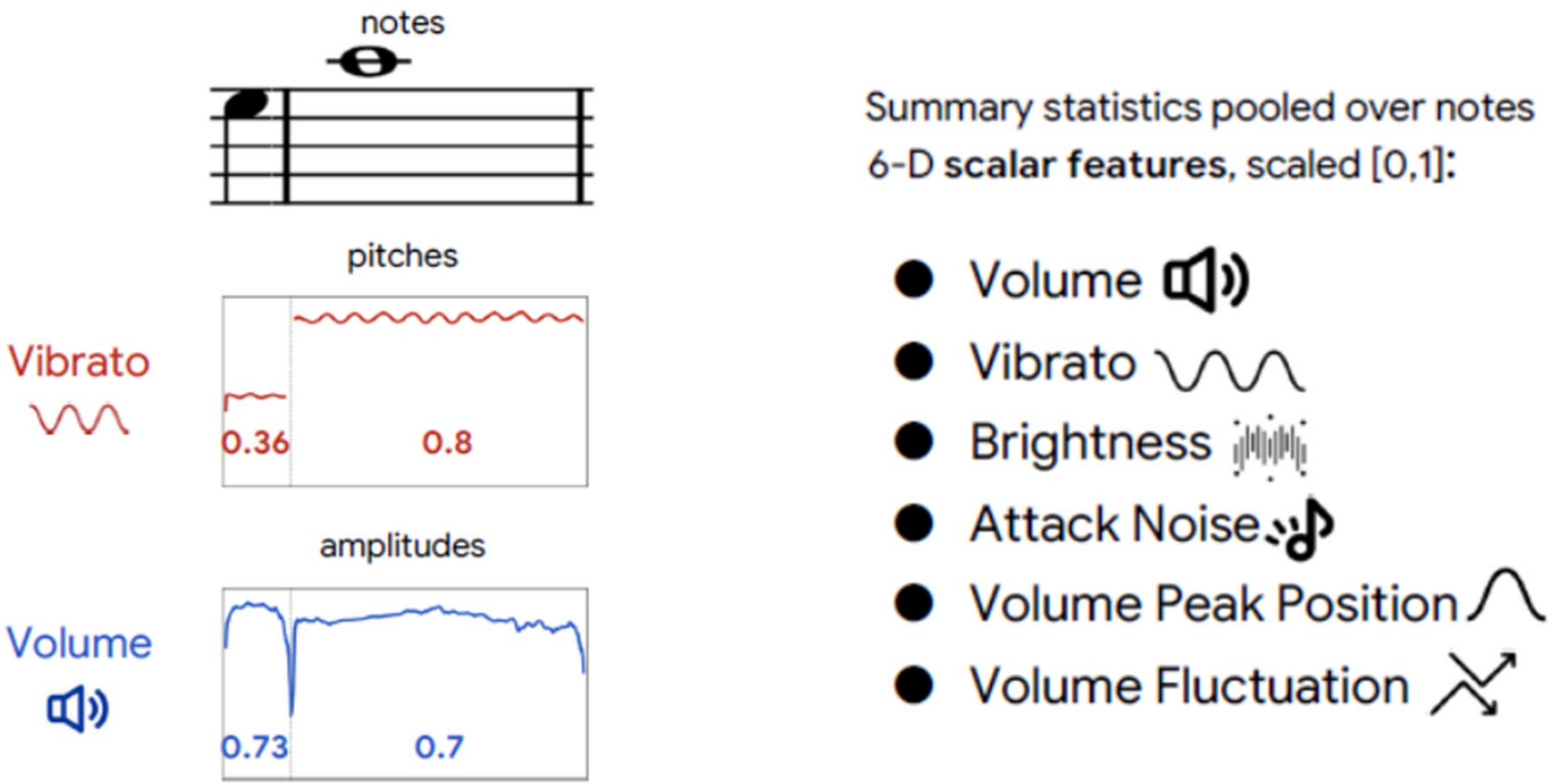


Summary statistics pooled over notes
6-D **scalar features**, scaled [0,1]:

- Volume
- Vibrato
- Brightness
- Attack Noise
- Volume Peak Position
- Volume Fluctuation

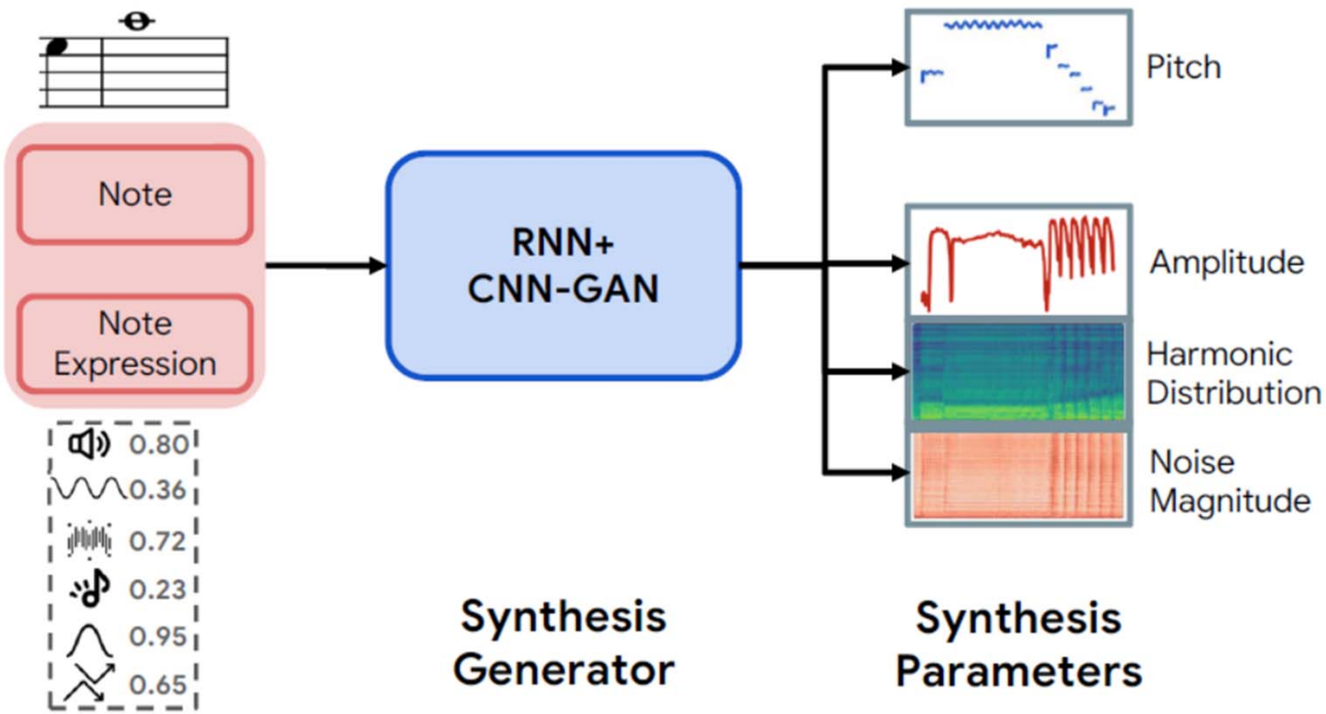
Extract Performance Parameter

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925



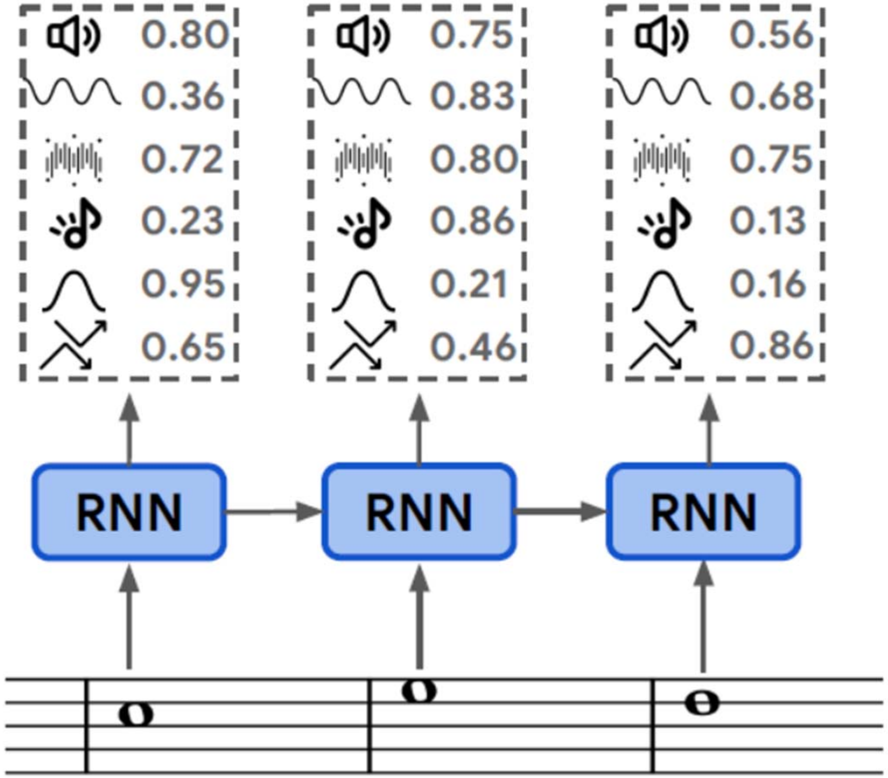
Synthesis Generator

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925



Autoregressive Prior on Expression Controls

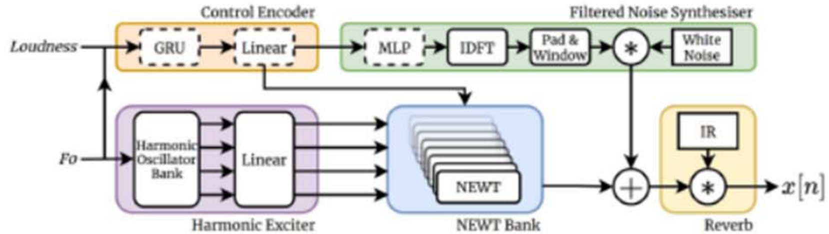
https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925



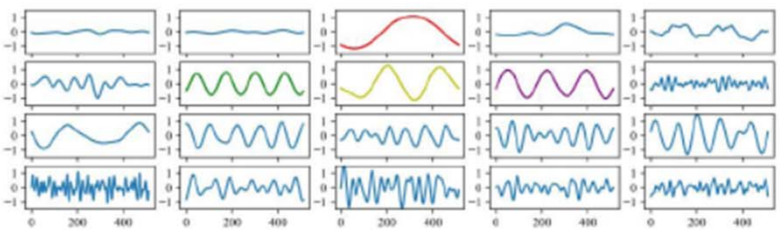
Other Differentiable Synthesis Works

https://docs.google.com/presentation/d/1xrzeAIMnVOumSql_L2oIfVMXcJxOKd3F2u4_DEIkmbY/edit#slide=id.g1a484a50b88_1_1925

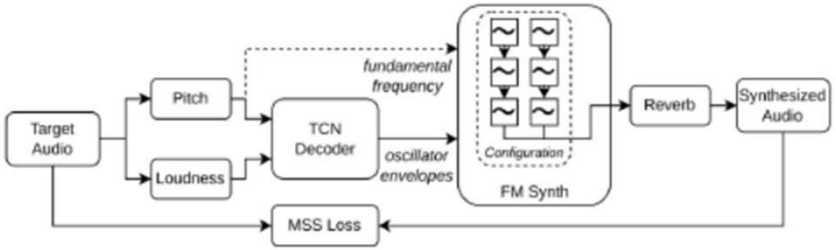
Waveshaping Synthesis [8]



Wavetable Synthesis [9]



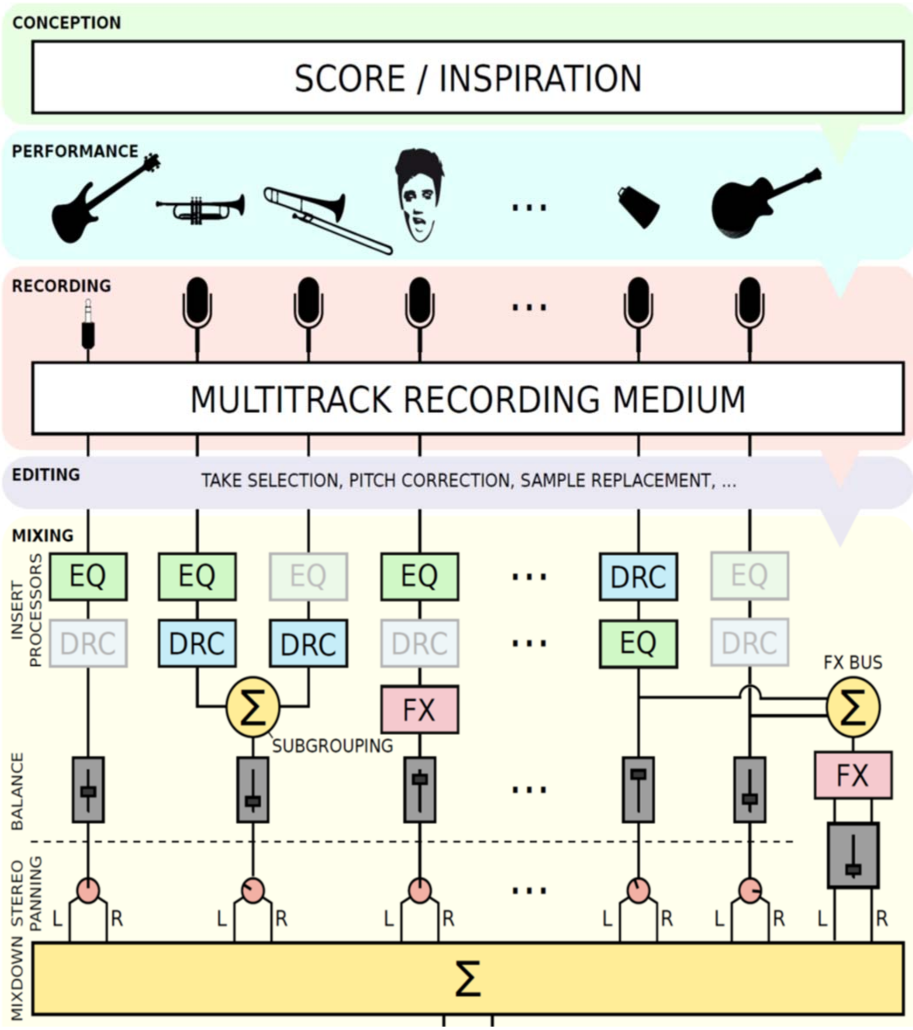
FM Synthesis [10]



Outline

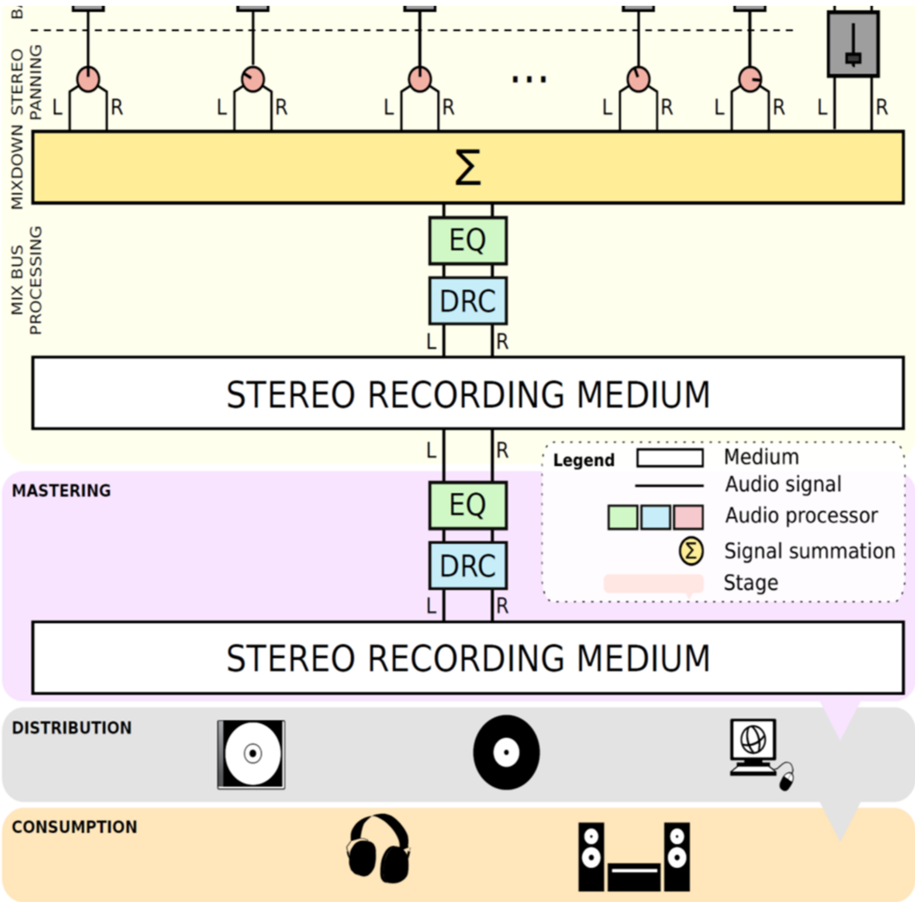
- Differentiable digital signal processing (DDSP)
  - Uses a neural network to convert a user's input into complex DSP controls that can produce realistic signals
- MIDI-DDSP
- **Automatic mixing**
  https://dl4am.github.io/tutorial/landing-page.html

Automatic Mixing



(Figure from the Internet)

Extension: Automatic Mixing



(Figure from the Internet)

Commercial Software (e.g., iZotope Neutron/Ozone)

AUDIO REPAIR

RX 10

Industry-standard audio repair tool used on movies and TV shows to restore damaged, noisy audio to pristine condition.

MIXING

Neutron 4

Mix smarter and faster while staying in your flow. Eight professional plug-ins combine to create your modern and intelligent mixing experience.

MASTERING

Ozone 11

Harness the power of Ozone 11, the industry-standard mastering suite. Featuring new processing like Clarity, Stem Focus, and Transient/Sustain for professional sound with ease and precision.

VOCAL PRODUCTION

Nectar 4

Get your vocals to sit in the mix with the most sophisticated set of tools for vocal production.

Mixing

https://docs.google.com/presentation/d/1UhBZucbGnX-ItOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252

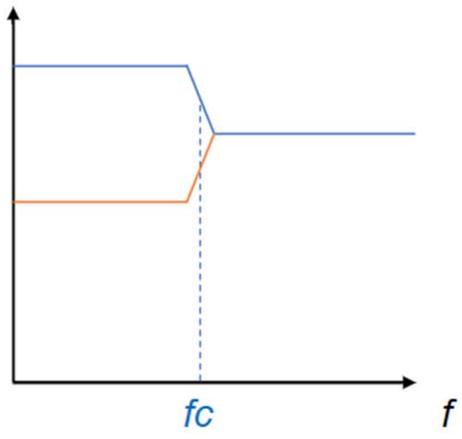


Fig. 9 Shelving filter resopnse.

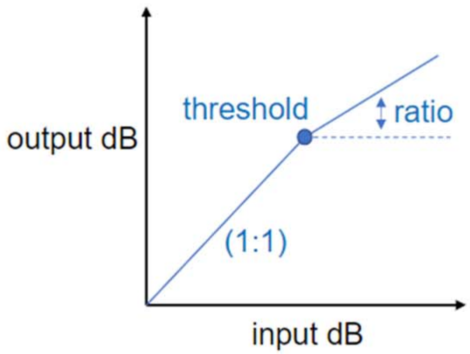


Fig. 12 Dynamic range compressor static characterist

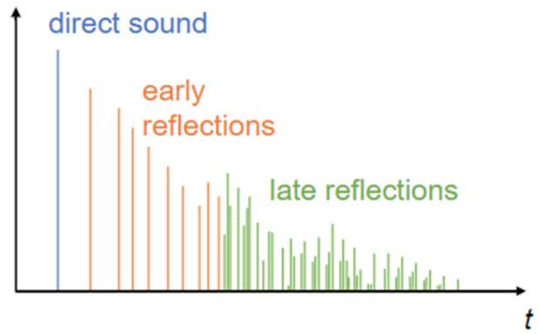
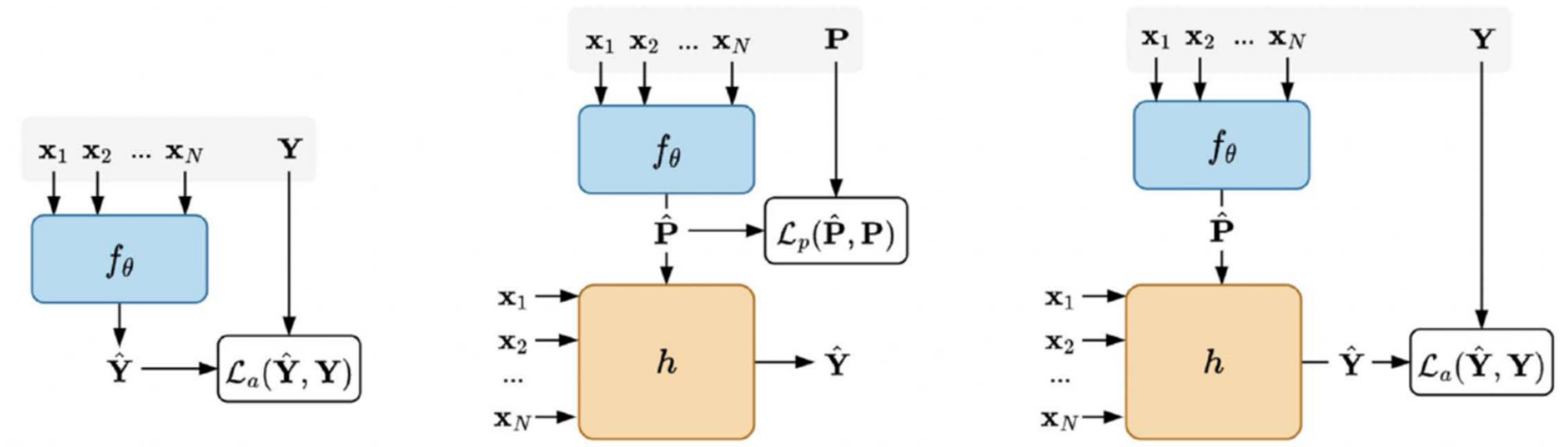


Fig. 17 Impulse response of an acoustic space.

- Gain
- Panning
- Equalization (EQ)
- Dynamic range compression (DRC)
- Artificial reverberation

Automatic Mixing: Problem Formulation

https://docs.google.com/presentation/d/1UhBZucbGnX-ItOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252



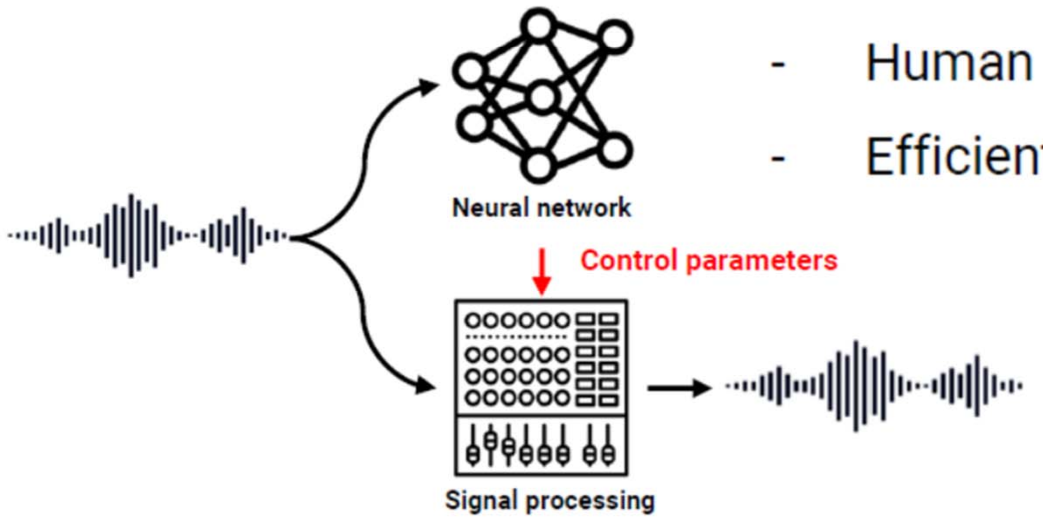
Direct Transformation

Parameter Estimation
(Parameter Loss)

Parameter Estimation
(Audio Loss)

Differentiable Signal Processing

https://docs.google.com/presentation/d/1UhBZucbGnX-ItOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252



- Leveraging existing DSP tools and knowledge
- High quality audio processing with few artifacts
- Human understandable outputs that can be adjusted
- Efficient and can easily run in real-time on CPU

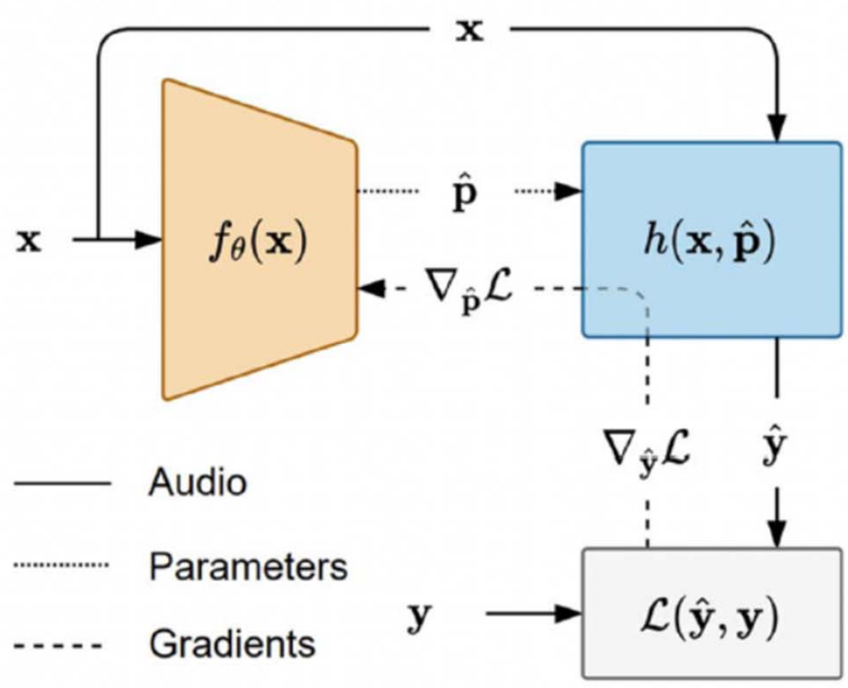
Differentiable Signal Processing

https://docs.google.com/presentation/d/1UhBZucbGnX-ItOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252

Non-differentiable

Discontinuous
(Discrete options)

Recursive operations



Backpropgation through the DSP is non-trivial

Techniques

https://docs.google.com/presentation/d/1UhBZucbGnX-ItOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252

- Automatic differentiation (AD)
- Neural proxies and hybrids (NP)
- Numerical gradient approximation (NGA)

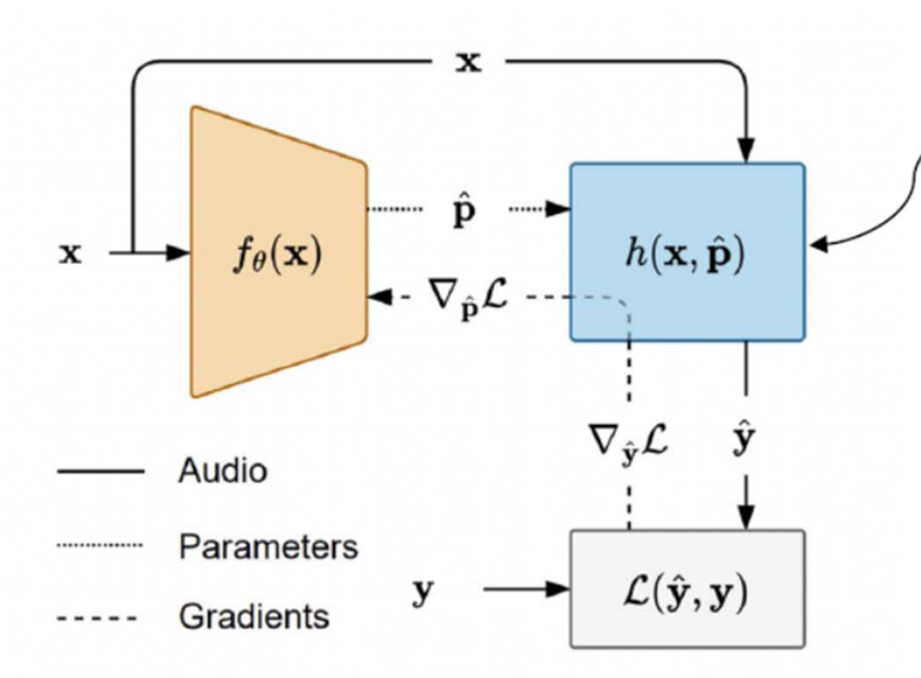
Automatic Differentiation

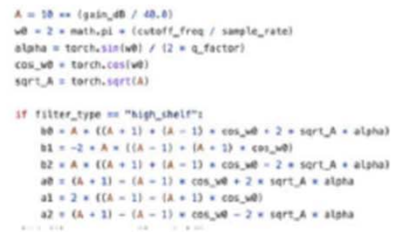
https://docs.google.com/presentation/d/1UhBZucbGnX-ItOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252

White-box

Requires hacks or tricks for each DSP

Doesn't work for all kinds of DSP





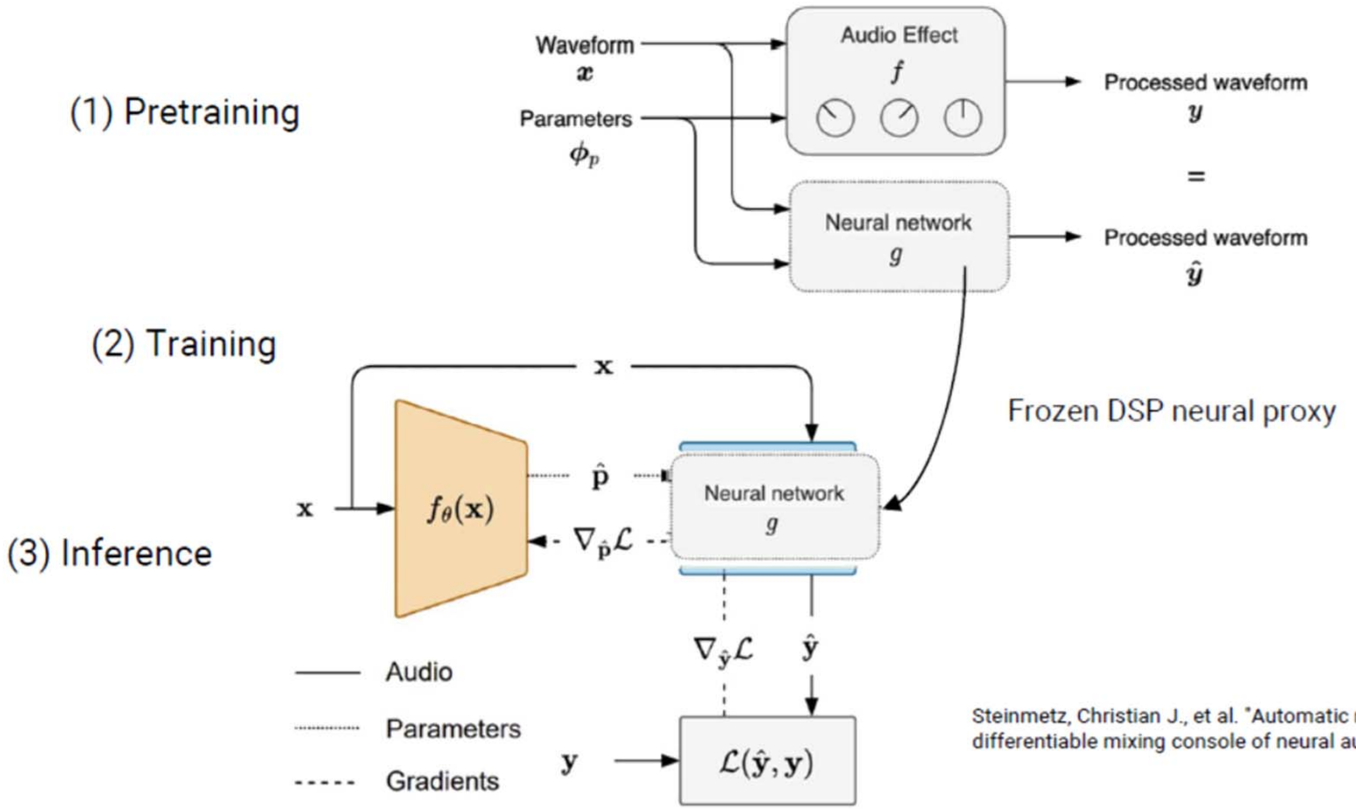
Explicitly define signal processing operations in autodiff framework

PyTorch TensorFlow JAX

Engel, Jesse, et al. "DDSP: Differentiable digital signal processing." *ICLR* (2021).

Neural Proxy

https://docs.google.com/presentation/d/1UhBZucbGnX-ltOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252



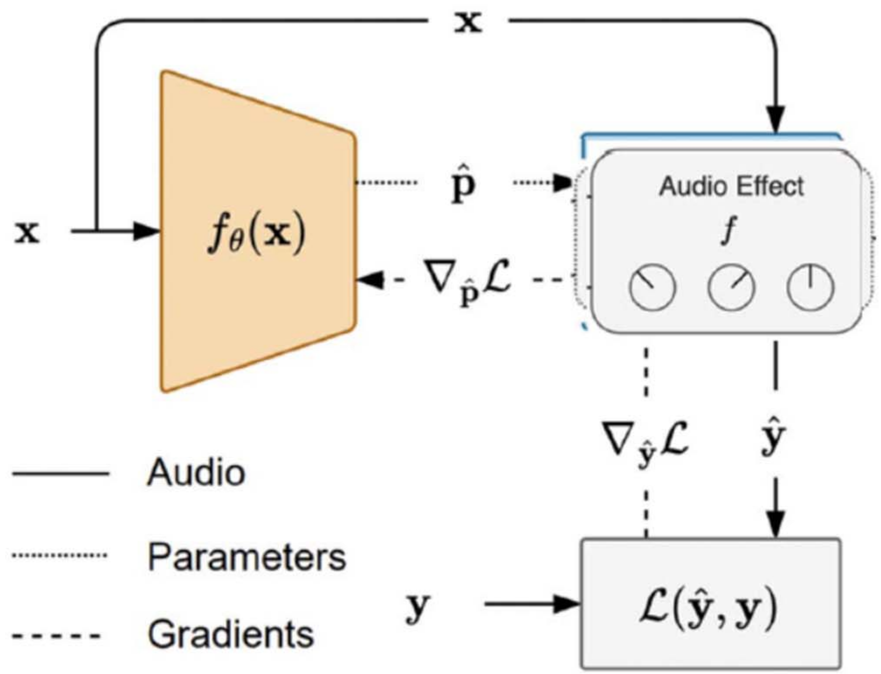
Steinmetz, Christian J., et al. "Automatic multitrack mixing with a differentiable mixing console of neural audio effects." ICASSP, 2021.

Neural Proxy Hybrid

https://docs.google.com/presentation/d/1UhBZucbGnX-ltOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252

(2) Training

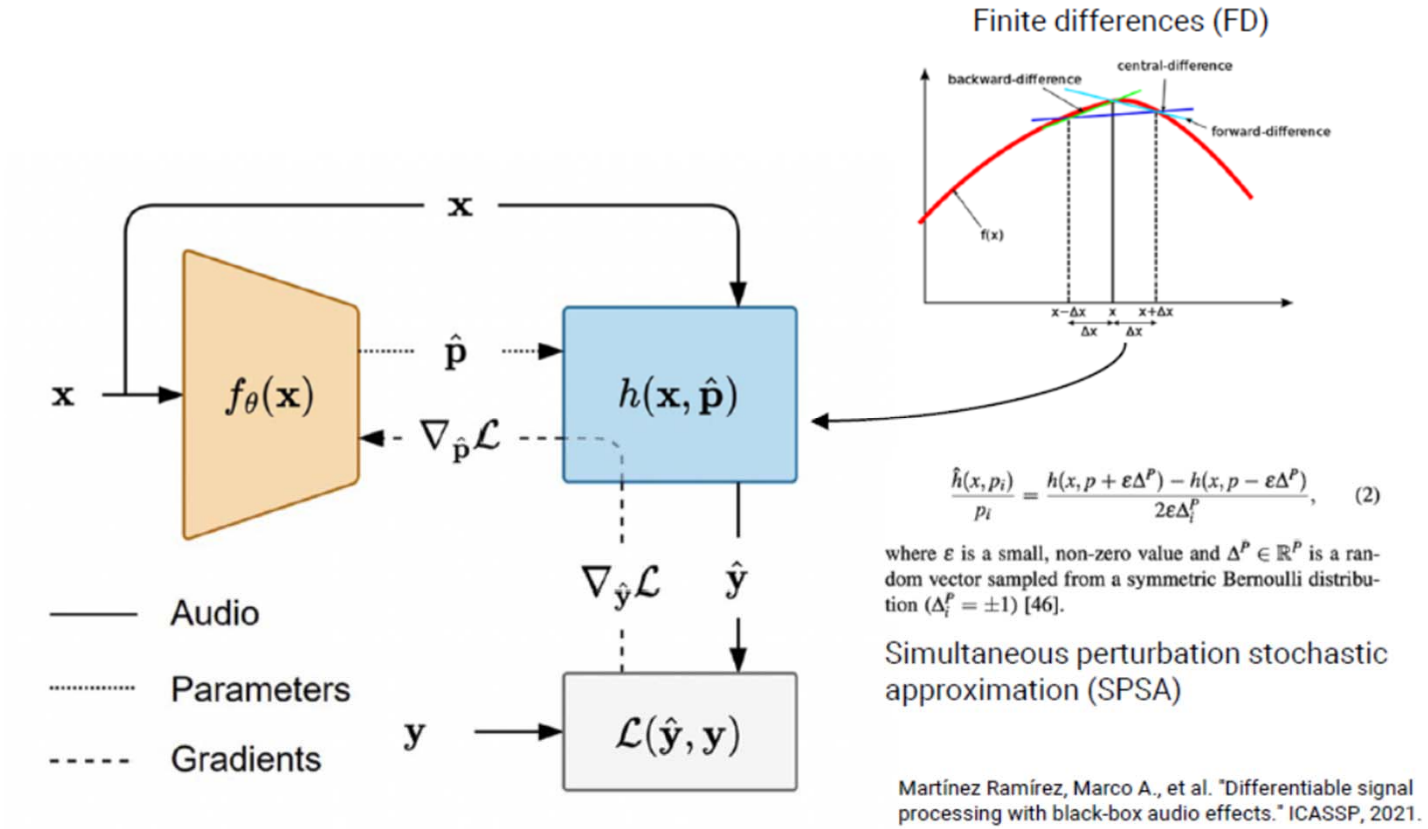
(3) Inference



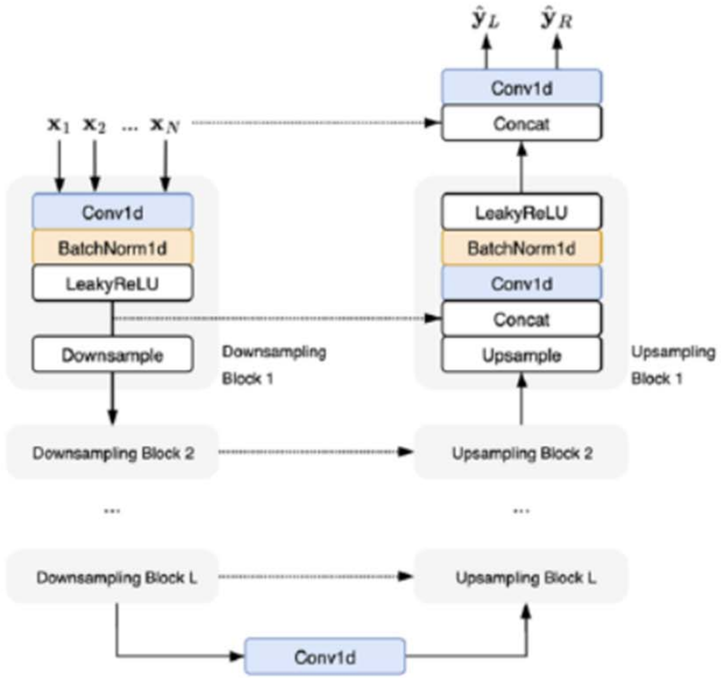
Use original DSP during inference

Gradient Approximation

https://docs.google.com/presentation/d/1UhBZucbGnX-ItOnVYkbdMoKrGfvHhM977OLyQ9h_Sxo/edit#slide=id.g174e8461521_0_252

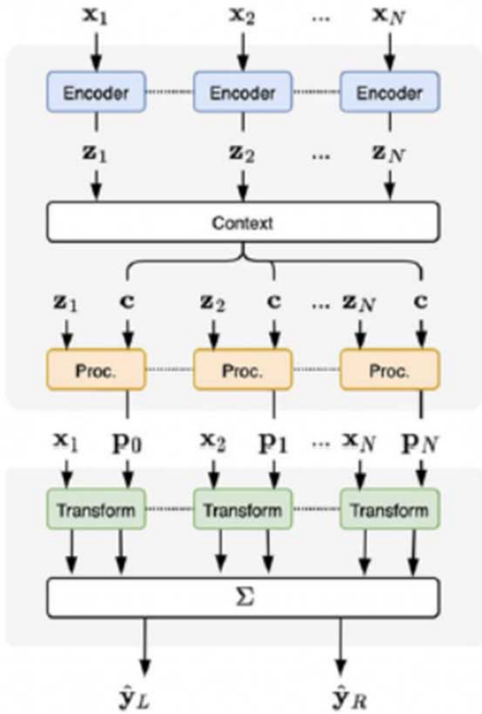


Deep Learning Models



Mix-Wave-U-Net

Direct Transformation



Differentiable Mixing Console

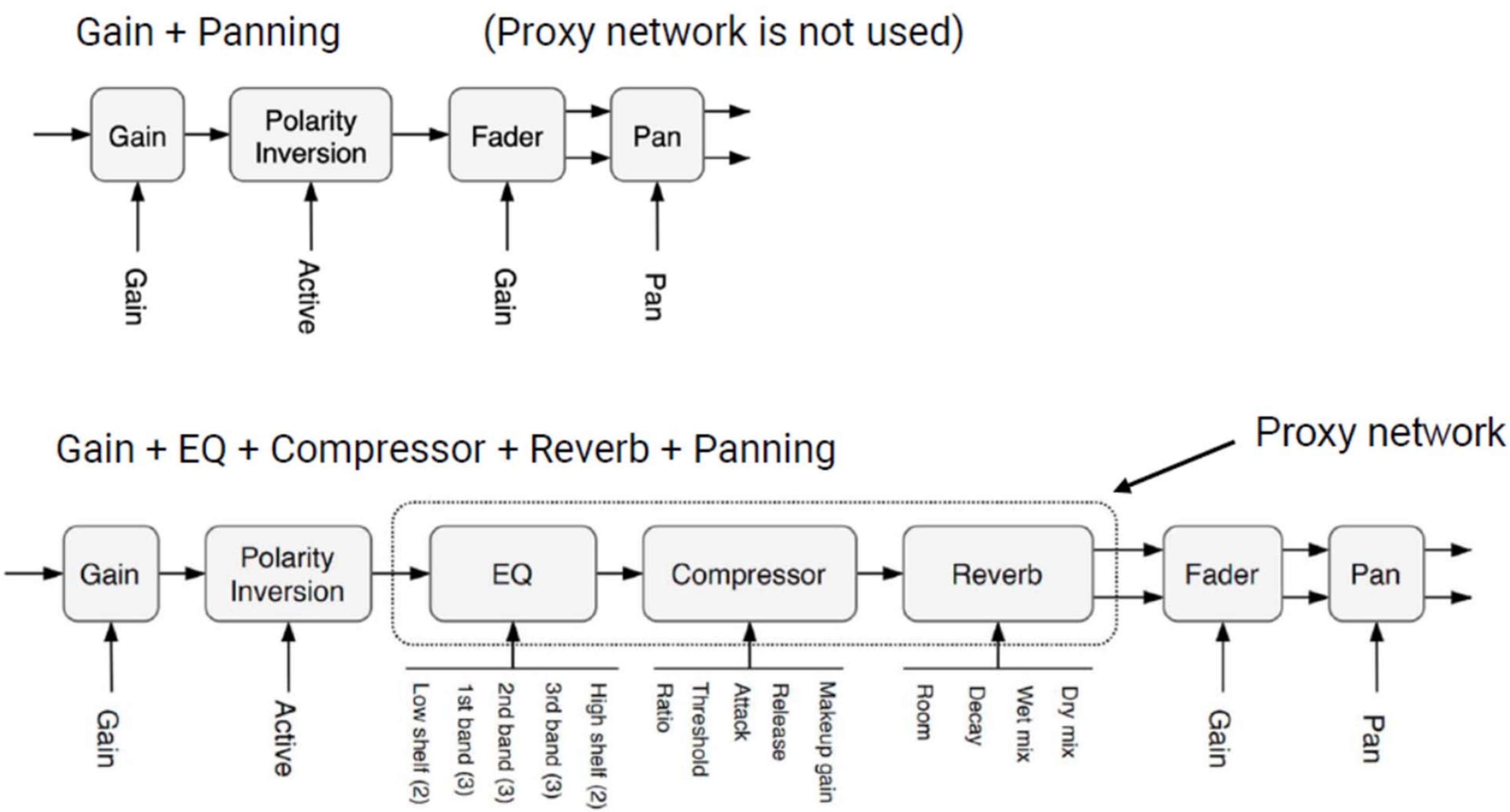
Parameter Estimation

"A Deep Learning Approach to Intelligent Drum Mixing with Wave-U-Net", Martínez Ramírez et al. 2021

"Automatic Multitrack Mixing with a Differentiable Mixing Console of Neural Audio Effects", Steinmetz et al. 2021

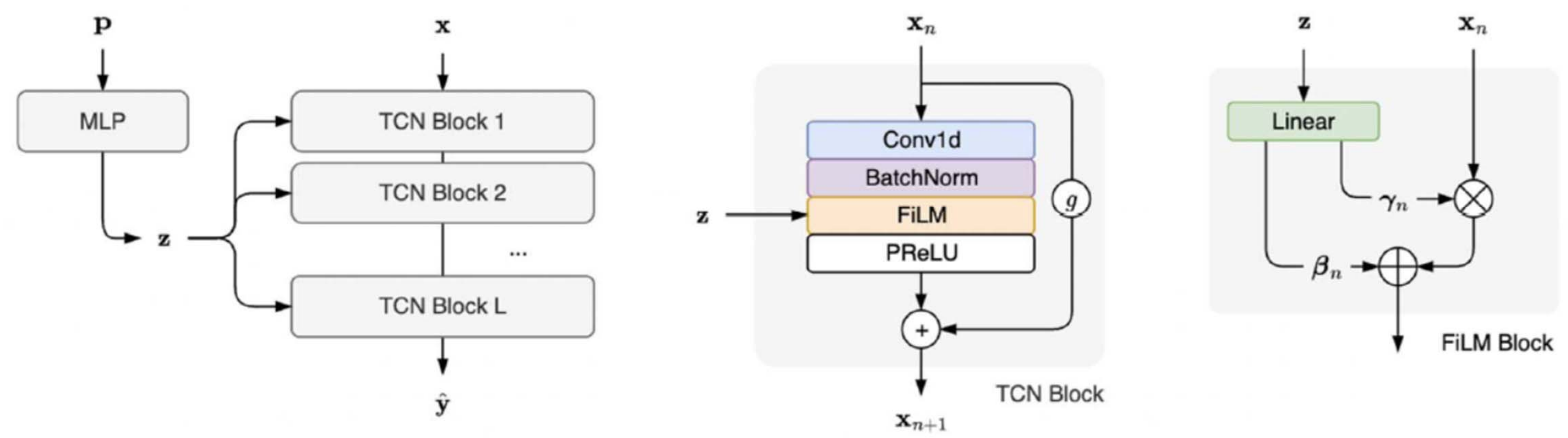
Differentiable Mixing Console

https://dl4am.github.io/tutorial/part_2/methods/02_dmc.html



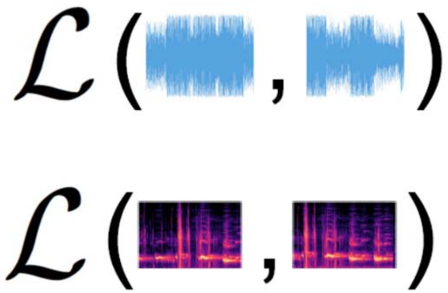
Differentiable Mixing Console

https://dl4am.github.io/tutorial/part_2/methods/02_dmc.html



Loss Functions

https://dl4am.github.io/tutorial/part_2/05_loss-functions.html



Time domain

Frequency domain

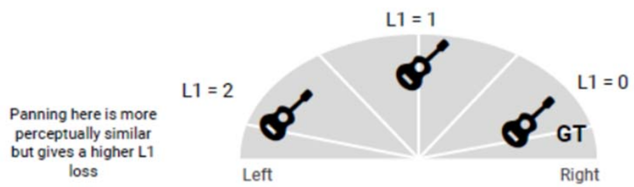
auraloss

A collection of audio-focused loss functions in PyTorch

https://github.com/csteinmetz1/auraloss

Stereo loss function

Loss function to encourage realistic mixes



L1 and L2 loss on stereo signals encourage panning all elements to the center.

$$y_{\text{sum}} = y_{\text{left}} + y_{\text{right}}$$

$$y_{\text{diff}} = y_{\text{left}} - y_{\text{right}}$$

$$\ell_{\text{Stereo}}(\hat{y}, y) = \ell_{\text{MR-STFT}}(\hat{y}_{\text{sum}}, y_{\text{sum}}) + \ell_{\text{MR-STFT}}(\hat{y}_{\text{diff}}, y_{\text{diff}})$$

Achieves invariance to stereo (left-right) orientation

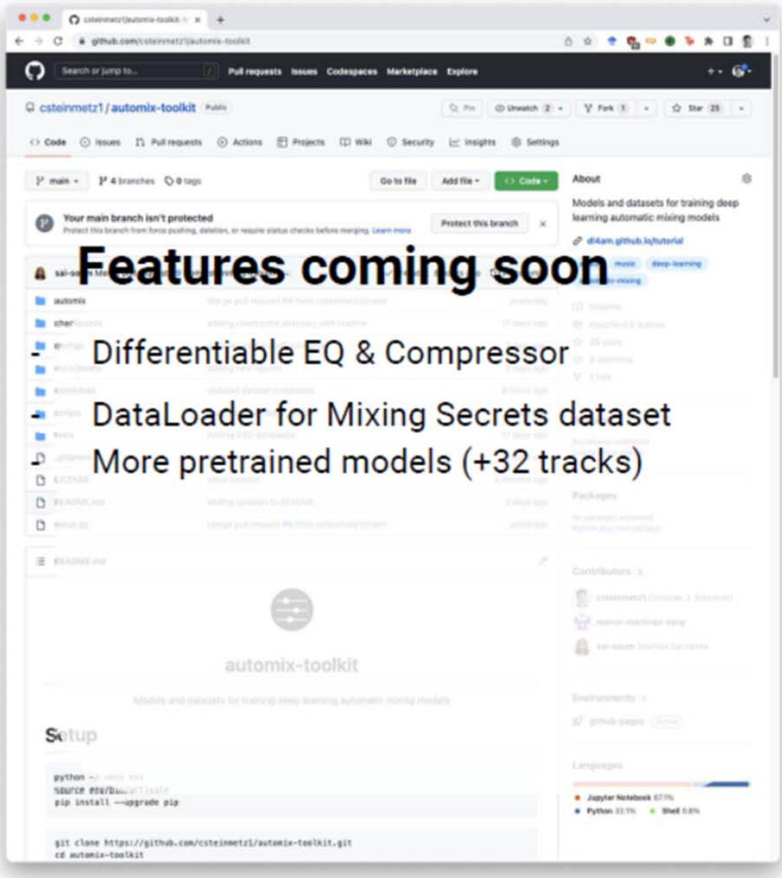
Library: automix-toolkit

https://github.com/csteinmetz1/automix-toolkit

automix-toolkit

https://github.com/csteinmetz1/automix-toolkit

★ Star   Star it on GitHub



Automatic Mixing Papers

https://csteinmetz1.github.io/AutomaticMixingPapers/

LEVEL | EQUALIZATION | COMPRESSION | PANNING | REVERB | MULTIPLE | MACHINE LEARNING | KNOWLEDGE-BASED | OVERVIEW | CLEAR

Show 10 entries

Search:

Year	Title	Author(s)	Category	Approach	Code
2021	A Deep Learning Approach to Intelligent Drum Mixing with the Wave-U-Net	M. Martinez Ramirez, D. Stoller and D. Moffat,	Multiple	ML	CODE
2021	Context Aware Intelligent Mixing Systems	M. N. Lefford, G. Bromham, G. Fazekas, and D. Moffat	Review	Multiple	
2020	One-shot parametric audio production style transfer with application to frequency equalization	S. I. Mimilakis, N. J. Bryan, and P. Smaragdis	Equalization	ML	
2020	Mixing with intelligent mixing systems: evolving practices and lessons from computer assisted design	M. N. Lefford, G. Bromham, and D. Moffat	Review	Multiple	
2020	Automatic multitrack mixing with a differentiable mixing console of neural audio effects	C. J. Steinmetz, J. Pons, S. Pascual, and J. Serrà	Multiple	ML	CODE
2019	An automatic mixing system for multitrack spatialization for stereo based on unmasking and best panning practices	A. Tom, J.D. Reiss, and P. Depalle	Panning	KBS	CODE
2019	Automatic mixing level balancing enhanced through source interference identification	D. Moffat and M. B. Sandler	Level	KBS	