Learning to Control Signal Processing Algorithms with Deep Learning

WiMIR Project Guide, October 2021

Nicholas J. Bryan | Adobe Research
nibryan@adobe.com

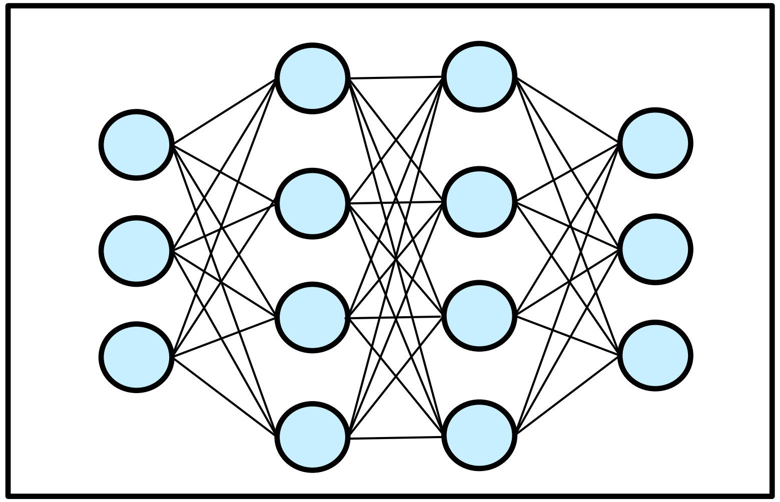
Sound Check

About Me

Undergraduate Experience

- University of Miami-FL
  - B.M. Music Engineering Technology, 2007
  - B.S. Electrical Engineering – Audio Emphasis, 2007

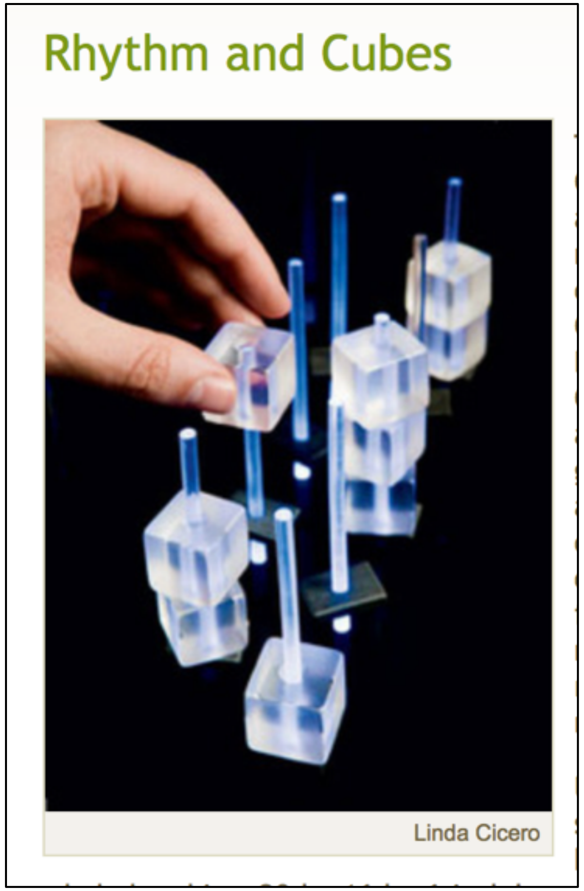
Undergraduate Experience



Graduate Experience

- Center for Computer Research in Music and Acoustics, Stanford University
  - M.A. Music, Science, & Technology, 2008
  - M.S. Electrical Engineering – Machine Learning Emphasis, 2011
  - Ph.D. Computer-Based Music Theory and Acoustics, 2014

Graduate Experience



"All the News That's Fit to Print"

The New York Times

National Edition

Northern California: Partly to mostly sunny. Highs in the middle and upper 50s, but 30s and 40s in the northern Sierra Nevada. Cooler tomorrow. Weather map, Page C10.

VOL. CLIX . . . . No. 54,880 SATURDAY, DECEMBER 5, 2009 $2.00

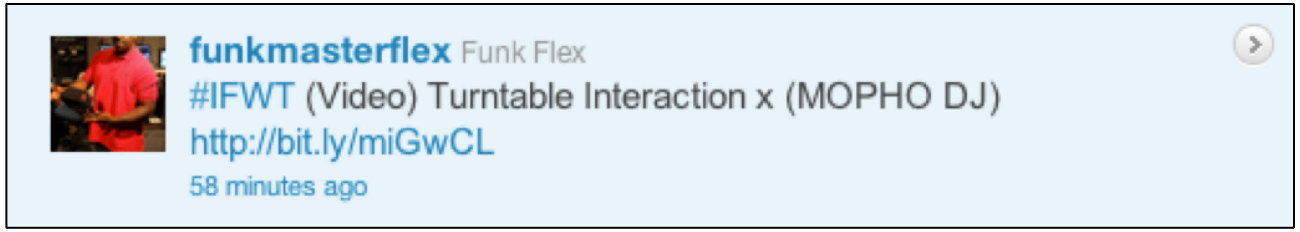
PETER DaSILVA FOR THE NEW YORK TIMES

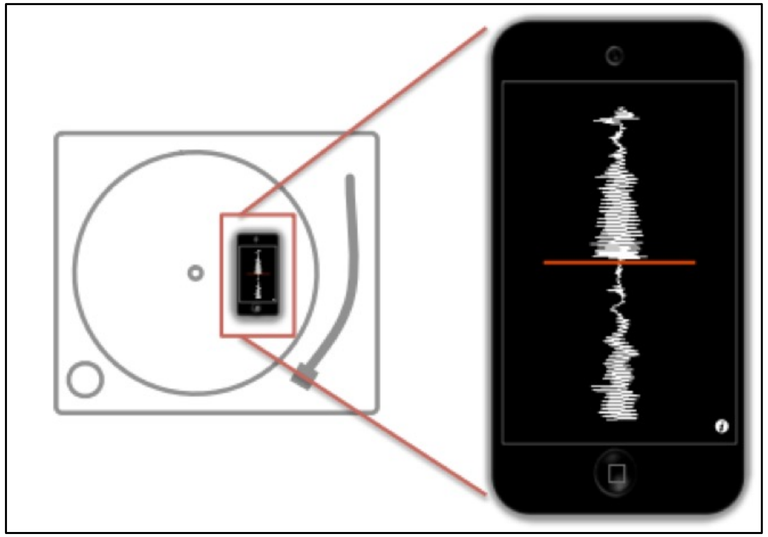
Nicholas Bryan plays multiple iPhones at a Stanford Mobile Phone Orchestra concert.

Ge Wang, a co-director of the Stanford Mobile Phone Orchestra, moved around the audience on Thursday, playing two iPhones.

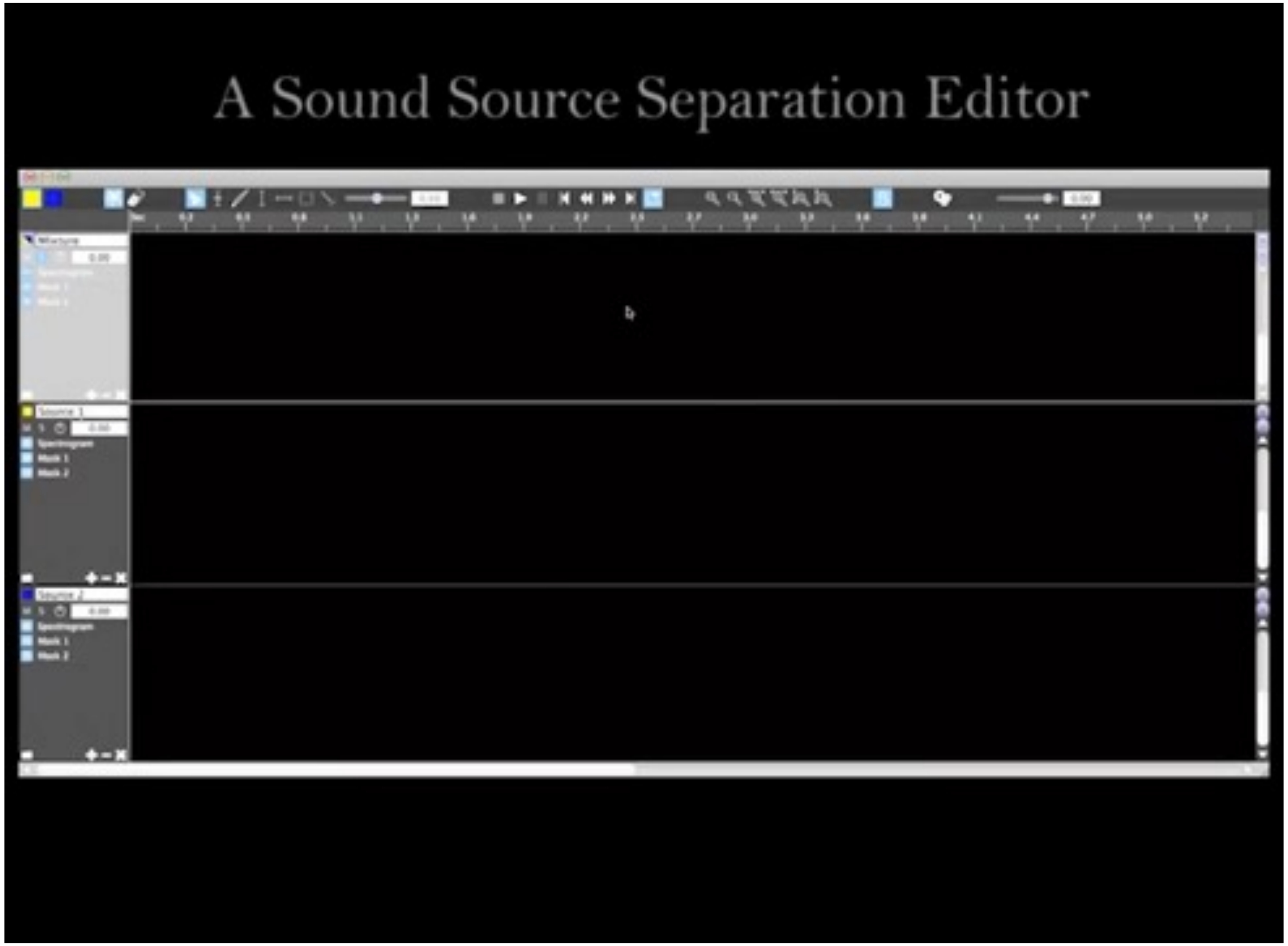
From Pocket to Stage, Music in the Key of iPhone

From Pocket to Stage, Music in the Key of iPhone





Inflection Point & PhD Thesis



ISSE: Interactive Sound Source Separation

isse.sourceforge.net

Post PhD Life – Interactive Media Group @ Apple

2014 – mid 2018


iPhone 6s


iPhone 7


iPhone 8


iPhone X

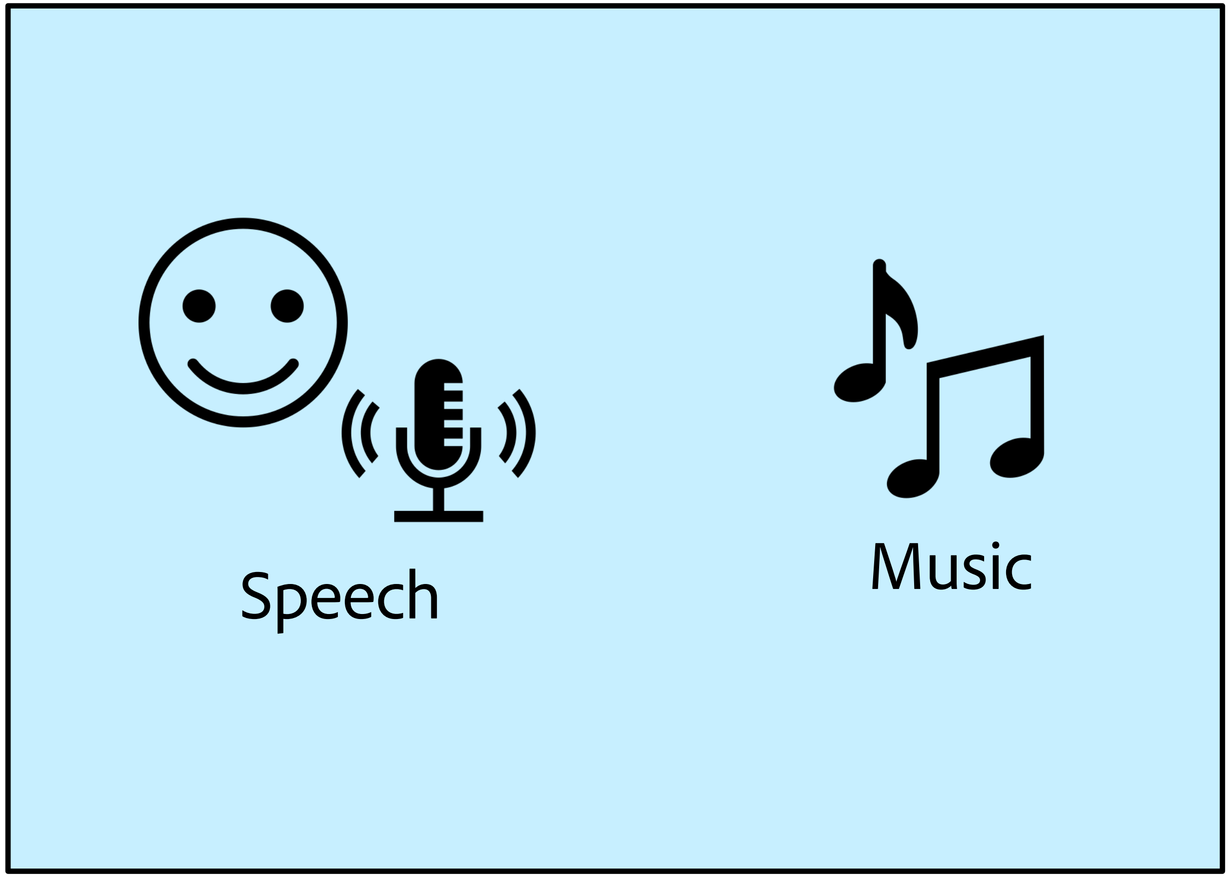
…


AirPods Pro

Post PhD Life – Adobe Research

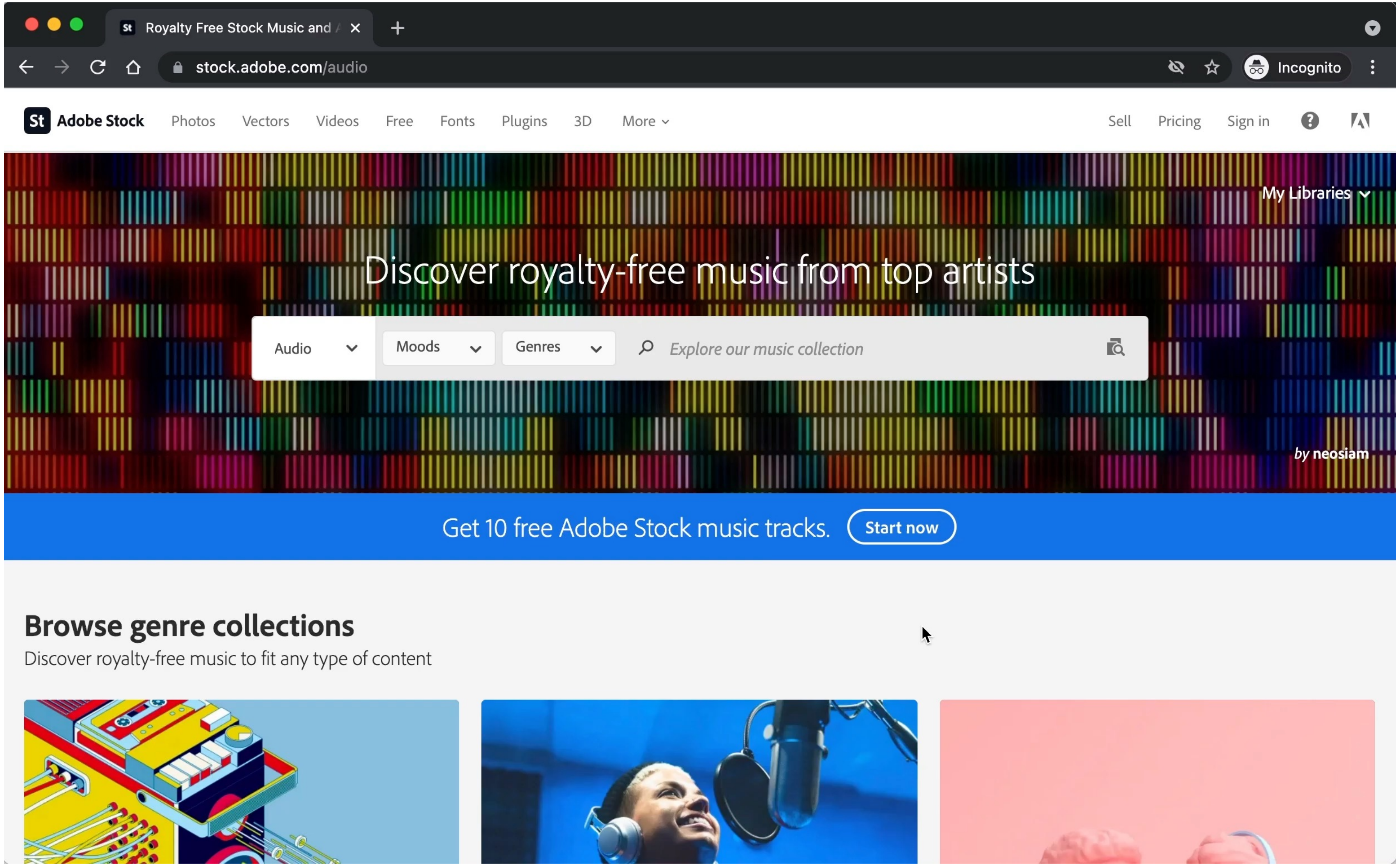
Search

Processing

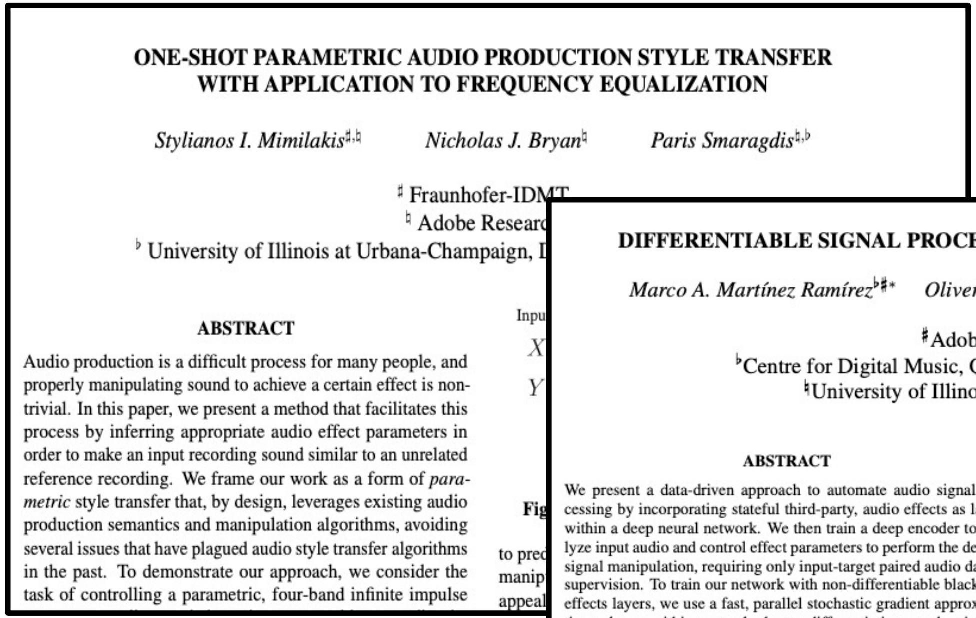


Music Similarity Search for Adobe Stock

New! Public Release!



Content-Based Audio & Music Processing



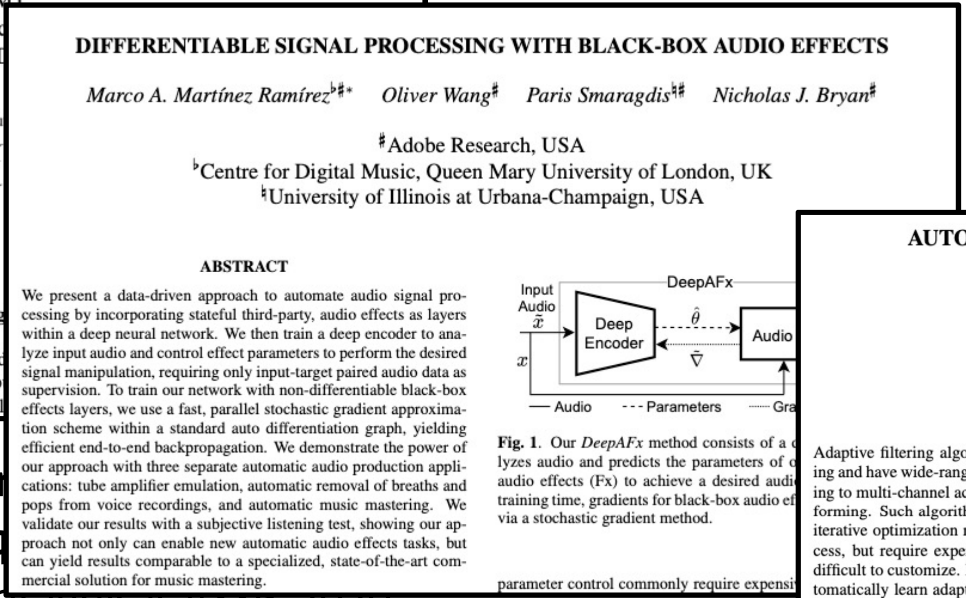
ONE-SHOT PARAMETRIC AUDIO PRODUCTION STYLE TRANSFER WITH APPLICATION TO FREQUENCY EQUALIZATION

Stylianos I. Mimilakis[♯,♮] *Nicholas J. Bryan*[♮] *Paris Smaragdis*[♮,♭]

[♯] Fraunhofer-IDMT
[♮] Adobe Researc
[♭] University of Illinois at Urbana-Champaign, I

ABSTRACT

Audio production is a difficult process for many people, and properly manipulating sound to achieve a certain effect is non-trivial. In this paper, we present a method that facilitates this process by inferring appropriate audio effect parameters in order to make an input recording sound similar to an unrelated reference recording. We frame our work as a form of *parametric* style transfer that, by design, leverages existing audio production semantics and manipulation algorithms, avoiding several issues that have plagued audio style transfer algorithms in the past. To demonstrate our approach, we consider the task of controlling a parametric, four-band infinite impulse



DIFFERENTIABLE SIGNAL PROCESSING WITH BLACK-BOX AUDIO EFFECTS

Marco A. Martínez Ramírez[♭♯*] *Oliver Wang*[♯] *Paris Smaragdis*[♮♯] *Nicholas J. Bryan*[♯]

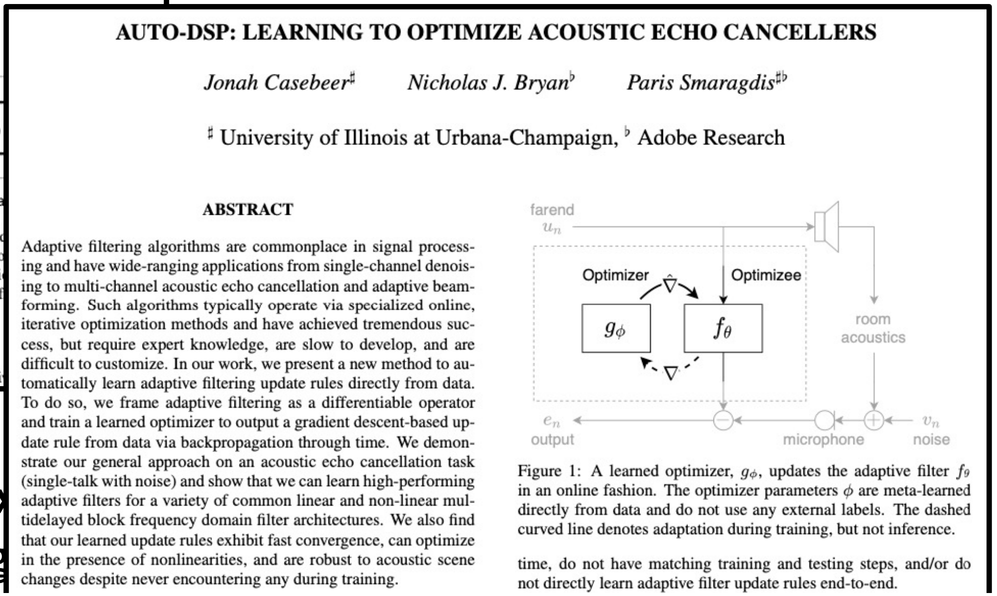
[♯] Adobe Research, USA
[♭] Centre for Digital Music, Queen Mary University of London, UK
[♮] University of Illinois at Urbana-Champaign, USA

ABSTRACT

We present a data-driven approach to automate audio signal processing by incorporating stateful third-party, audio effects as layers within a deep neural network. We then train a deep encoder to analyze input audio and control effect parameters to perform the desired signal manipulation, requiring only input-target paired audio data as supervision. To train our network with non-differentiable black-box effects layers, we use a fast, parallel stochastic gradient approximation scheme within a standard auto differentiation graph, yielding efficient end-to-end backpropagation. We demonstrate the power of our approach with three separate automatic audio production applications: tube amplifier emulation, automatic removal of breaths and pops from voice recordings, and automatic music mastering. We validate our results with a subjective listening test, showing our approach not only can enable new automatic audio effects tasks, but can yield results comparable to a specialized, state-of-the-art commercial solution for music mastering.

Fig. 1. Our *DeepAFx* method consists of a d lyzes audio and predicts the parameters of o audio effects (Fx) to achieve a desired audi training time, gradients for black-box audio ef via a stochastic gradient method.

parameter control commonly require expensi



AUTO-DSP: LEARNING TO OPTIMIZE ACOUSTIC ECHO CANCELLERS

Jonah Casebeer[♯] *Nicholas J. Bryan*[♭] *Paris Smaragdis*[♯♭]

[♯] University of Illinois at Urbana-Champaign, [♭] Adobe Research

ABSTRACT

Adaptive filtering algorithms are commonplace in signal processing and have wide-ranging applications from single-channel denoising to multi-channel acoustic echo cancellation and adaptive beamforming. Such algorithms typically operate via specialized online, iterative optimization methods and have achieved tremendous success, but require expert knowledge, are slow to develop, and are difficult to customize. In our work, we present a new method to automatically learn adaptive filtering update rules directly from data. To do so, we frame adaptive filtering as a differentiable operator and train a learned optimizer to output a gradient descent-based update rule from data via backpropagation through time. We demonstrate our general approach on an acoustic echo cancellation task (single-talk with noise) and show that we can learn high-performing adaptive filters for a variety of common linear and non-linear multidelayed block frequency domain filter architectures. We also find that our learned update rules exhibit fast convergence, can optimize in the presence of nonlinearities, and are robust to acoustic scene changes despite never encountering any during training.

Figure 1: A learned optimizer, g_ϕ, updates the adaptive filter f_θ in an online fashion. The optimizer parameters ϕ are meta-learned directly from data and do not use any external labels. The dashed curved line denotes adaptation during training, but not inference.

time, do not have matching training and testing steps, and/or do not directly learn adaptive filter update rules end-to-end.

”One-shot Parametric Audio Pr
w/Application to Frequency Eq
S. I. Mimilakis, N. J. Bryan, P. Smaragdis *ICASSP*, 2020.

"Differentiable Signal Proc. with Black-Box
M. A. Martínez Ramírez, O. Wang, P. Smarag
ICASSP, 2021.

"Auto-DSP: Learning to Optimize Acoustic Echo Cancellers."
J. Casebeer, N. J. Bryan P. Smaragdis,
WASPAA, 2021

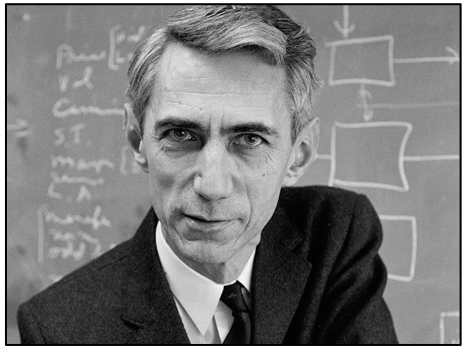
Learning to Control Signal Processing Algorithms with Deep Learning

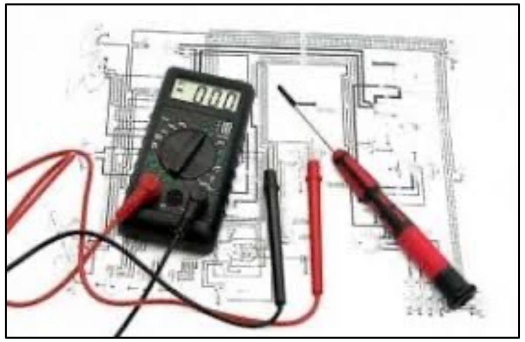
Is Signal Processing Still Useful?

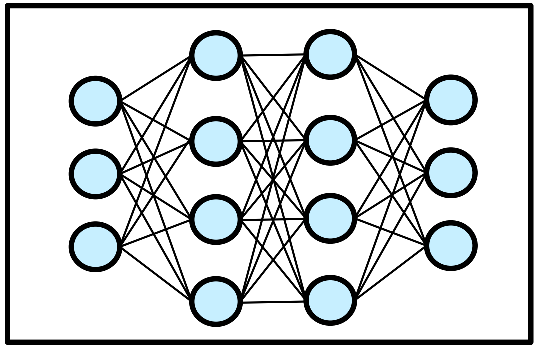
Imagenet classification with deep convolutional neural networks

Acoustic modeling using deep belief networks

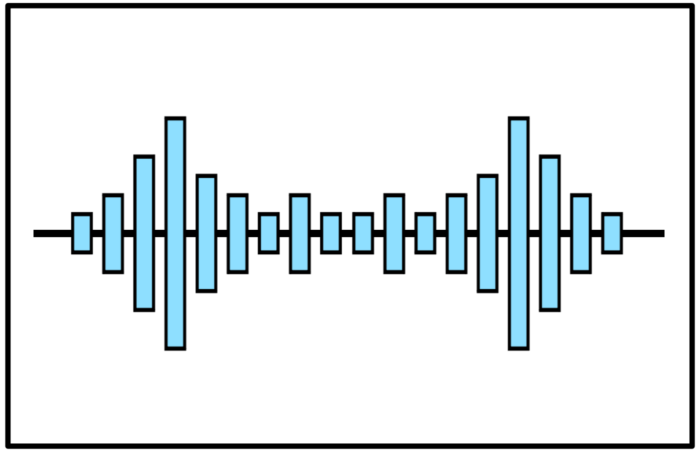
1940	2000	2012	Today
"Signal processing is the future" Probably Claude Shannon ☺	"DSP still the future" Probably most EE folks	"Deep nets are the future" Academia + industry	"Let's prosper together" Nervous DSP engineers

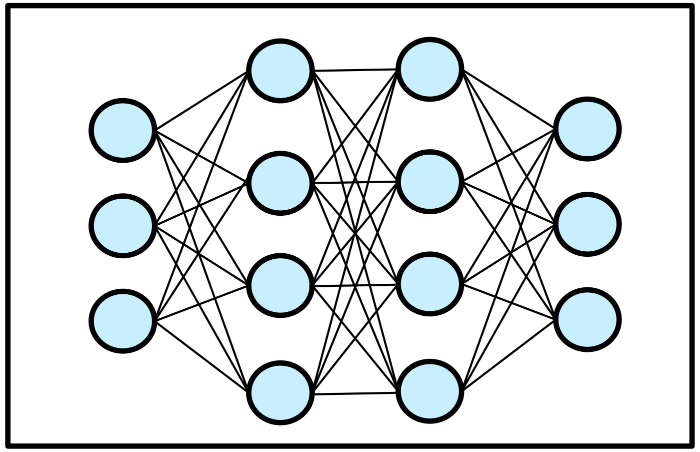




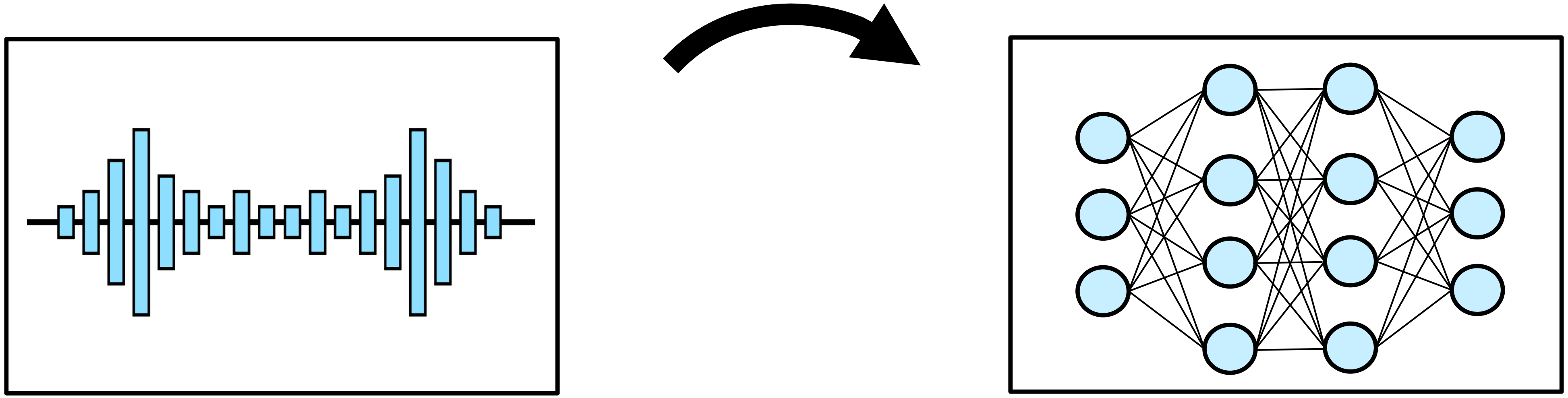


Signal Processing & Deep Learning



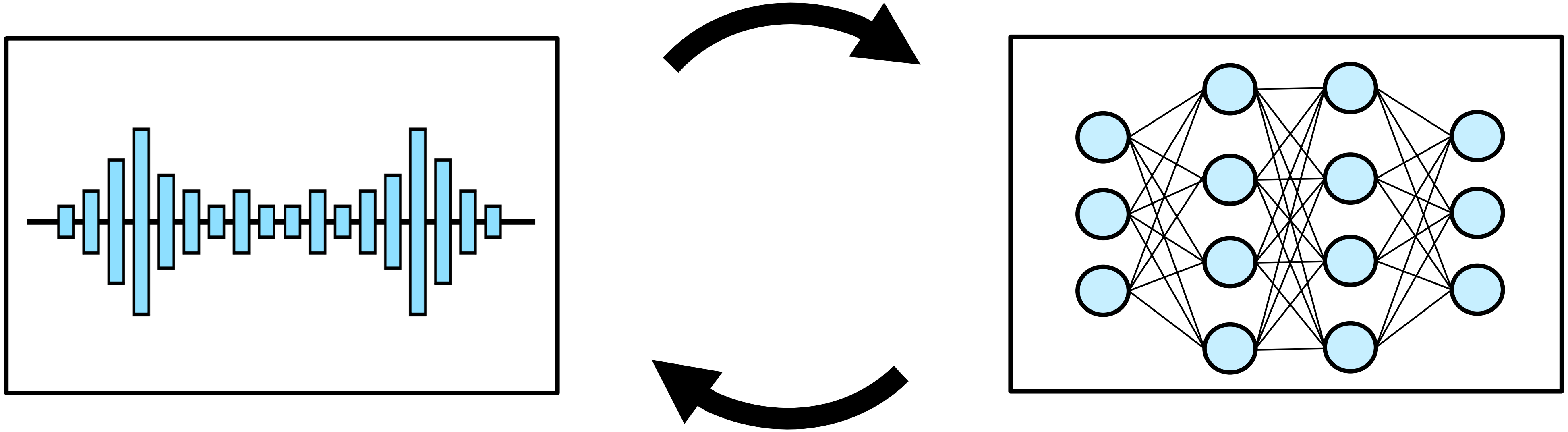


Option 1: Replace Signal Processing with Deep Learning



- Better versions of traditional algorithms (fit to data)
- Achieve previously unachievable tasks
- Emulate existing algorithms (e.g. analog hardware, software algorithms, etc.)

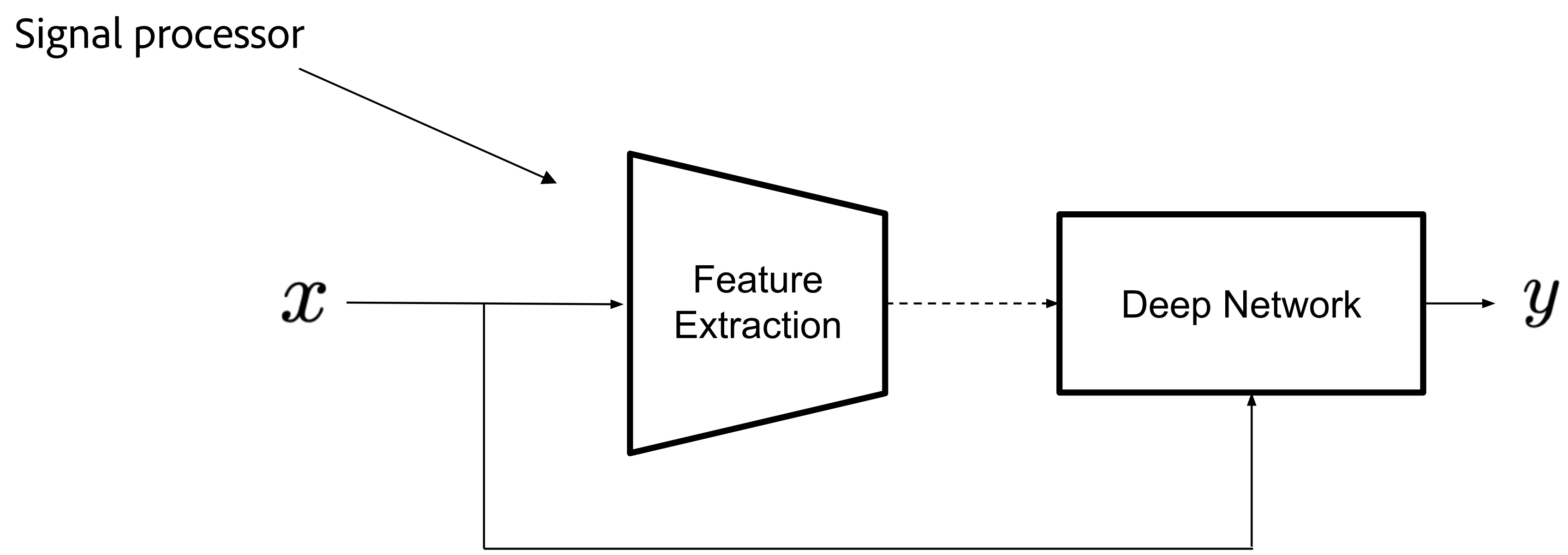
Option 2: Integrate Deep Learning into Signal Processing



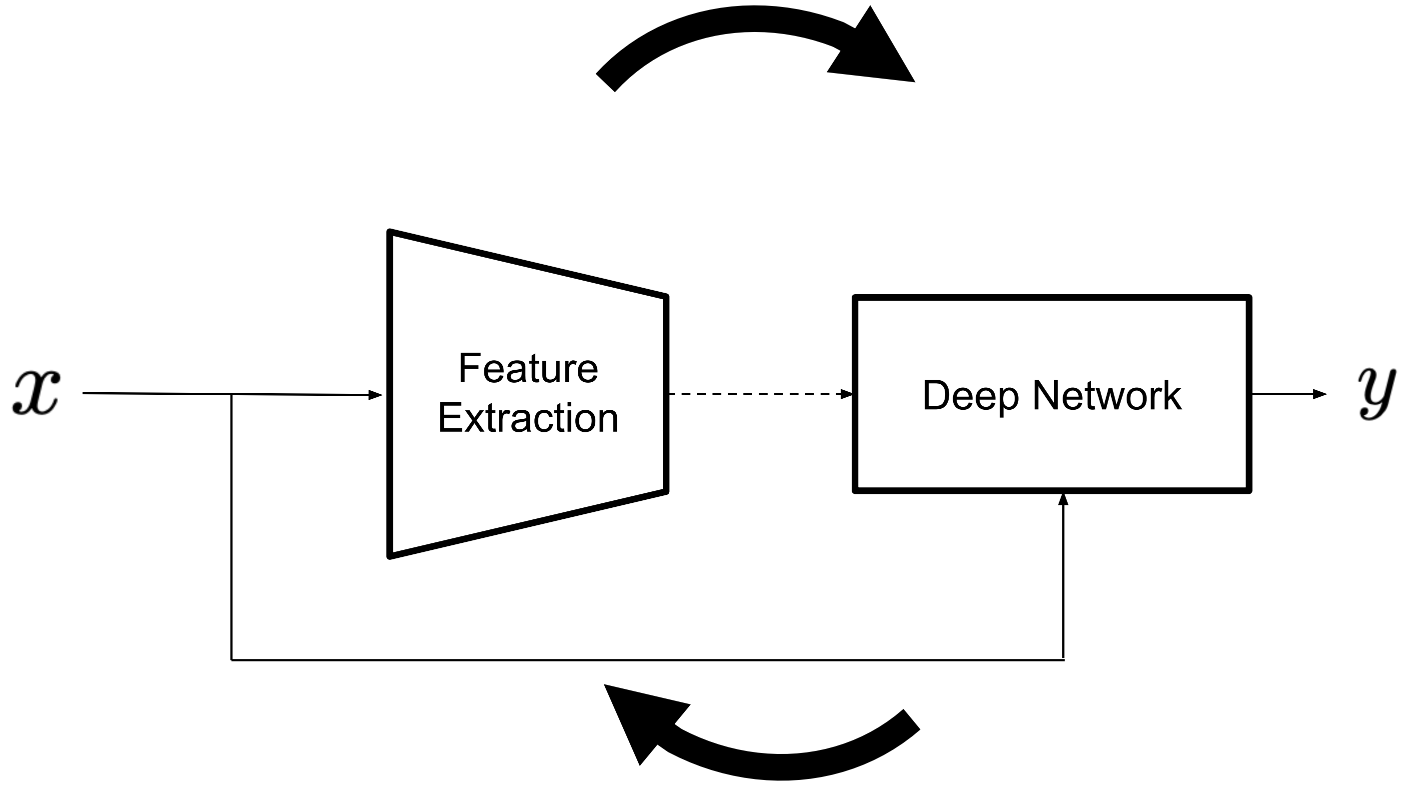
- **Inductive bias and expert knowledge**
- Faster, less complex, easier to implement
- More robust and generalizable
- More interpretable

Two Strategies

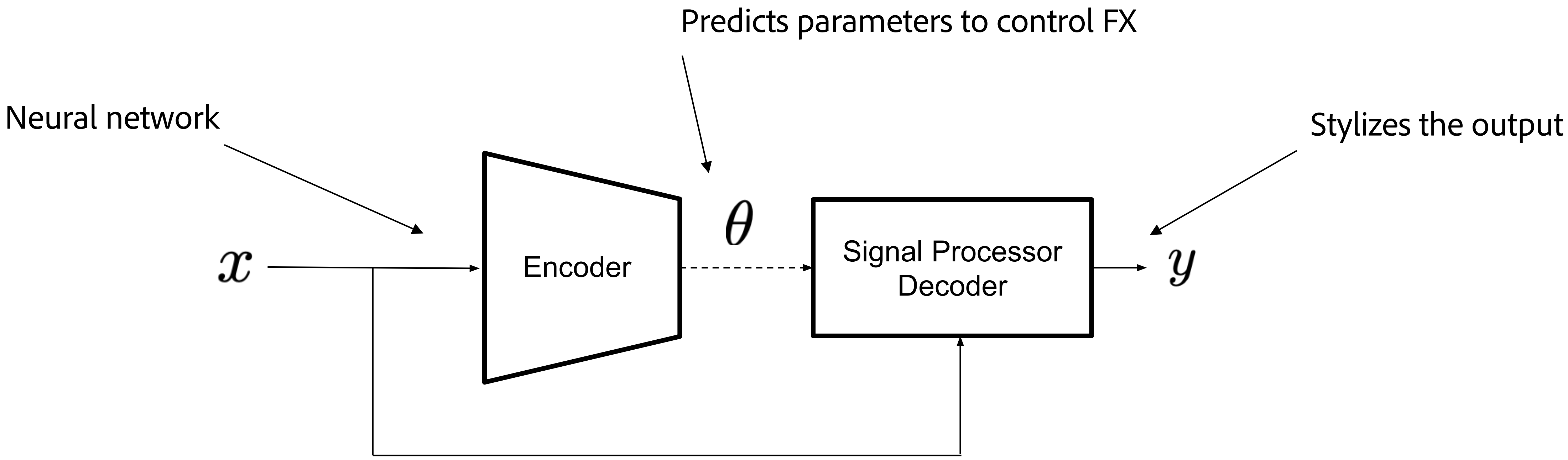
Signal Processing for Feature Extraction



Signal Processing for Feature Extraction



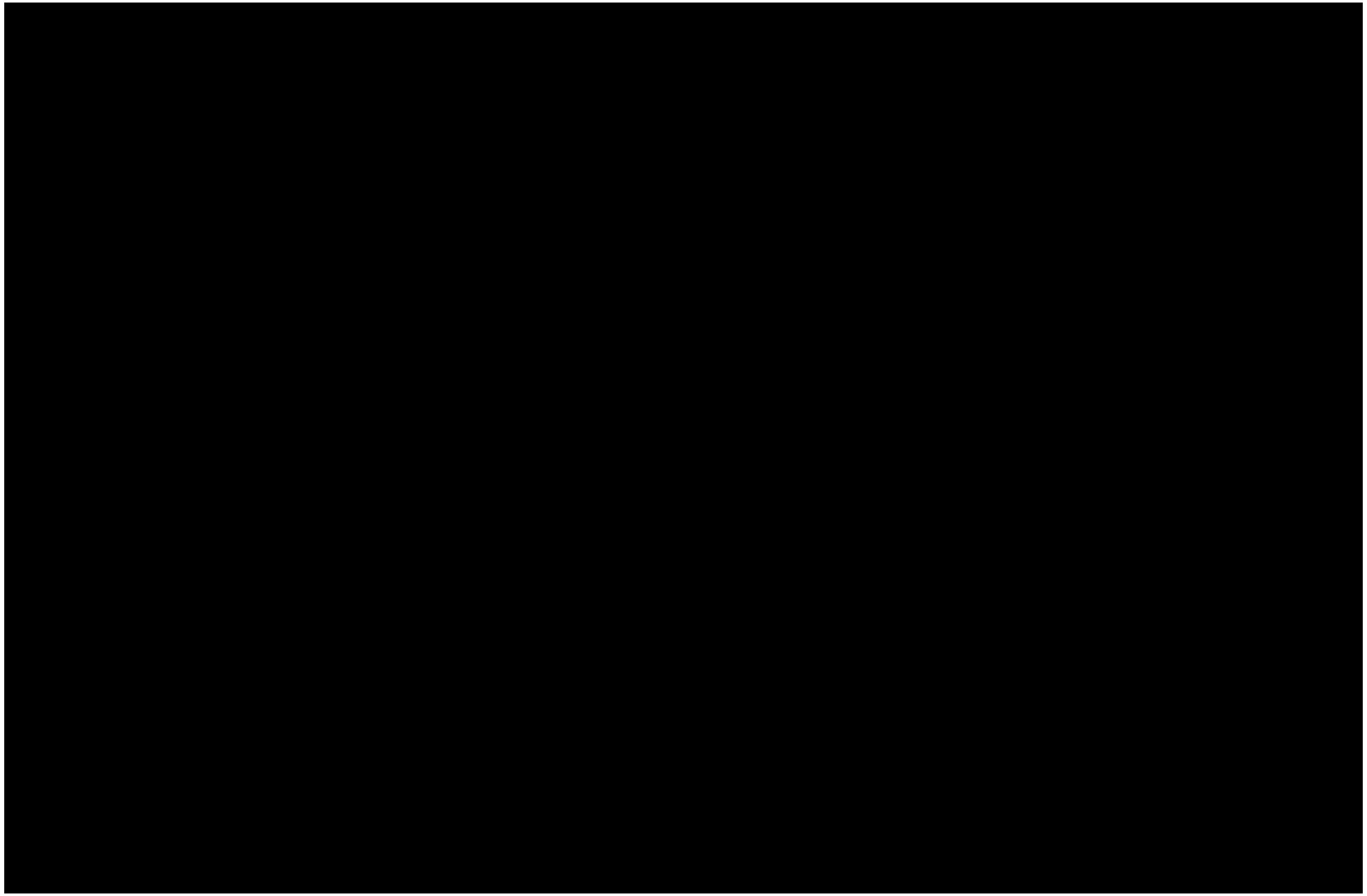
Signal Processors with Deep Control



Applications

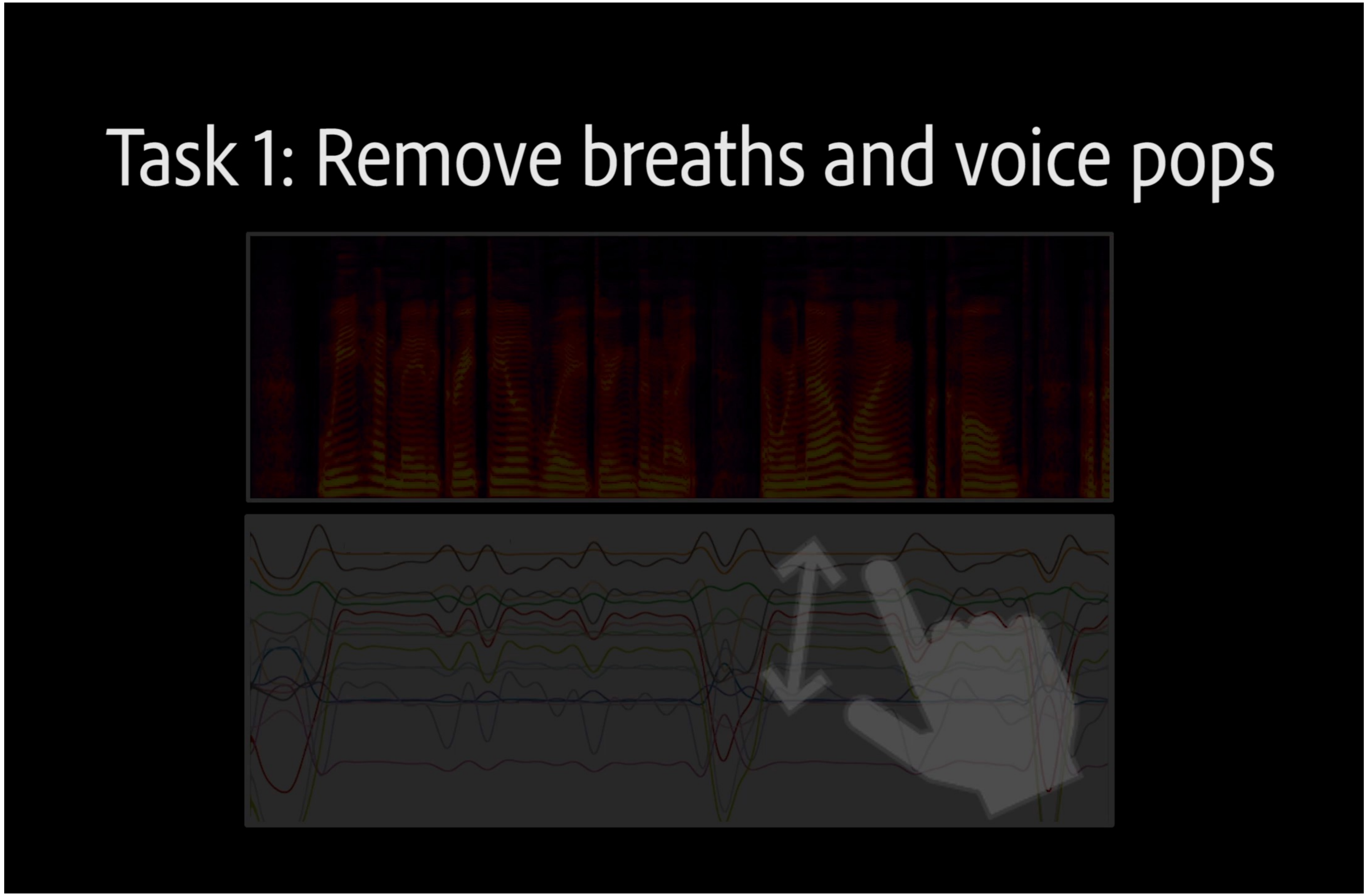
- Speech synthesis
- Automatic multi-track mixing
- Remove breaths from voice recording
- Emulate guitar distortion
- Automatic music mastering
- Optimal adaptive filters
- Automatic podcast production
- Learned optimization algorithms
- Music source separation

Remove Breaths



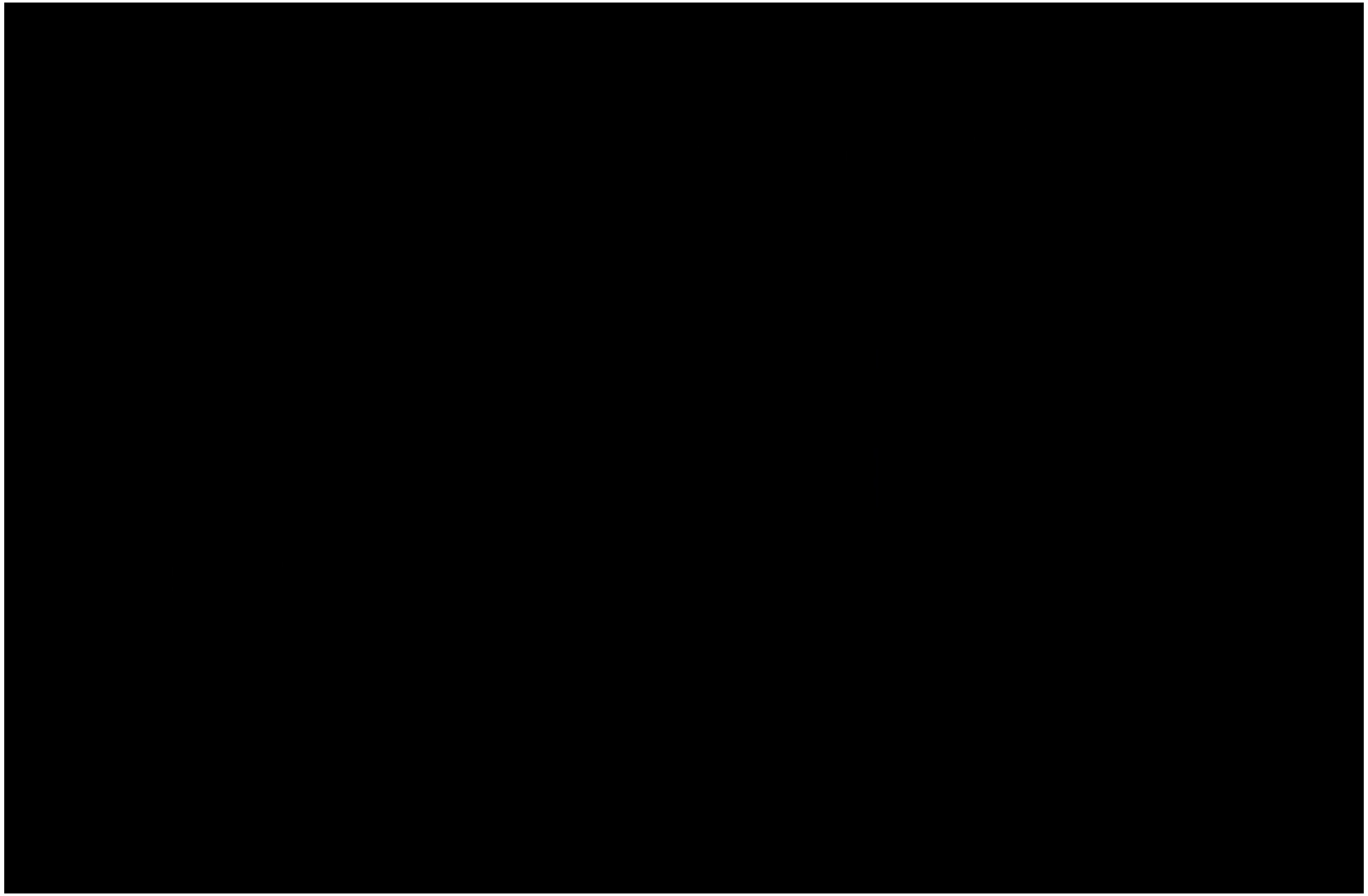
M. Martinez, O. Wang, P. Smaragdis, and N. J. Bryan, “Differentiable Signal Processing with Black-box Audio Effects”, ICASSP, 2021.

Guitar Distortion



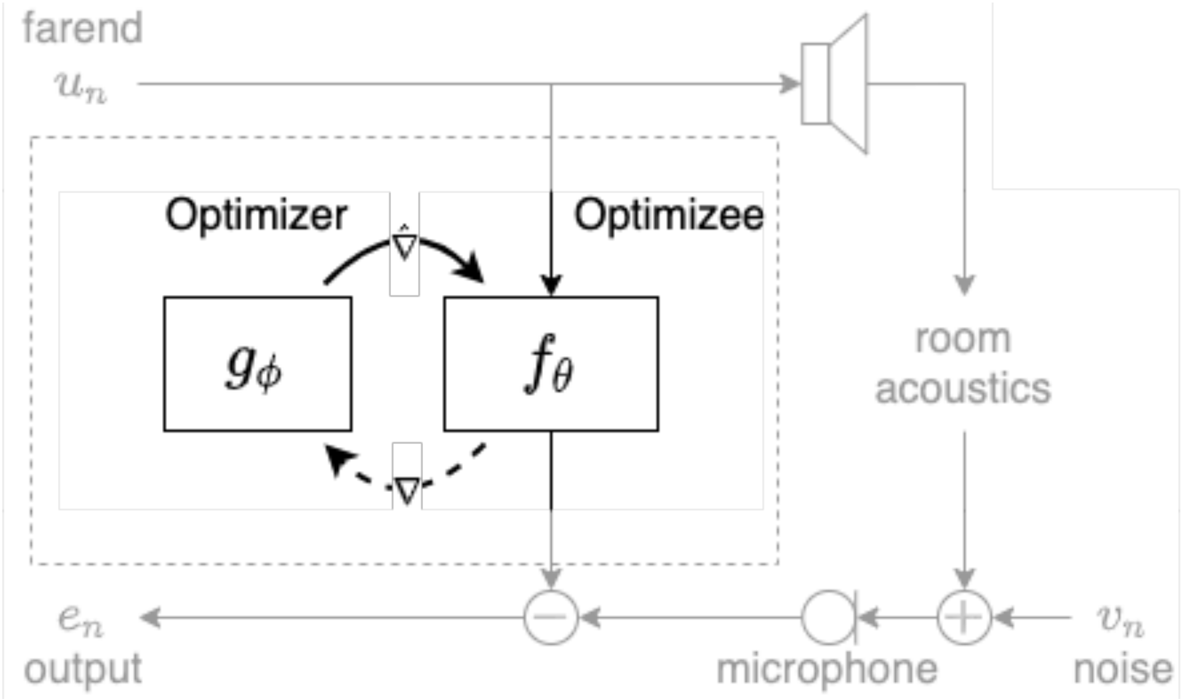
M. Martinez, O. Wang, P. Smaragdis, and N. J. Bryan, “Differentiable Signal Processing with Black-box Audio Effects”, ICASSP, 2021.

Music Mastering



M. Martinez, O. Wang, P. Smaragdis, and N. J. Bryan, “Differentiable Signal Processing with Black-box Audio Effects”, ICASSP, 2021.

Optimizing Acoustic Echo Cancellation



Mic Echo

Baseline

Ours

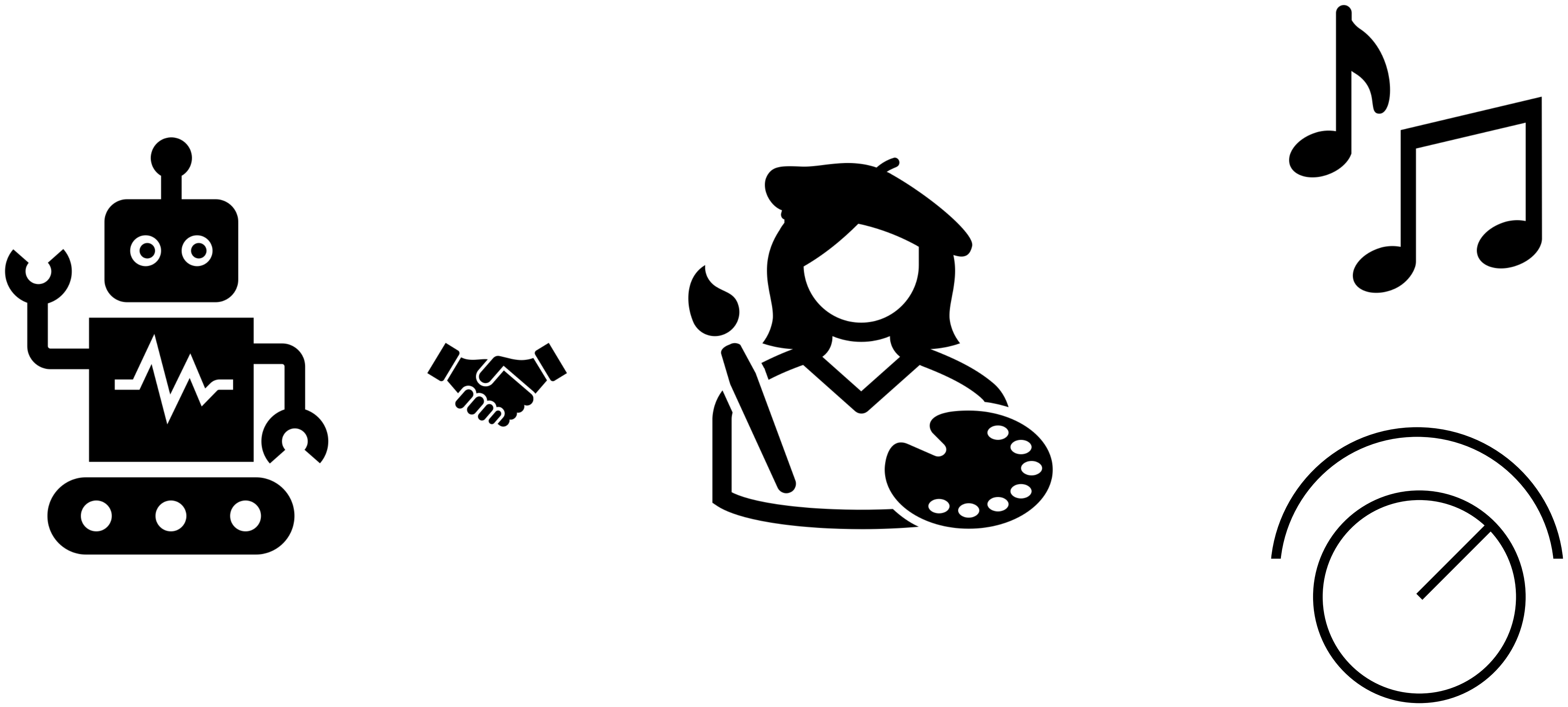
J. Casebeer, N. J. Bryan, and P. Smaragdis, "Auto-DSP: Learning to Optimize Acoustic Echo Cancellers", WASPAA, 2021.

Why is this Interesting?

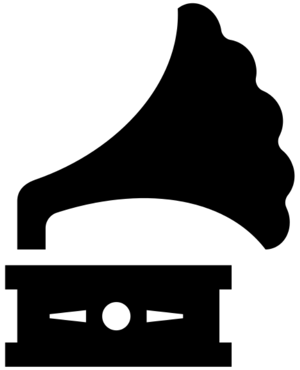
- Train a neural network to use audio FX/signal processors
- Benefits
  - Adaptive, signal-dependent signal processors
  - Tune signal processors to data
  - Harder to mess up audio quality
  - Estimate interpretable control parameters

Interpretable AI

- Blending deep learning with known algorithms let's us understand



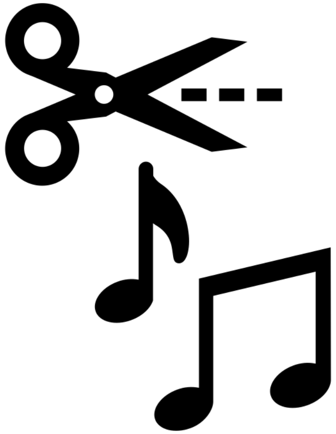
Applications for ISMIR



Music mastering

Guitar/FX modeling

Voice processing

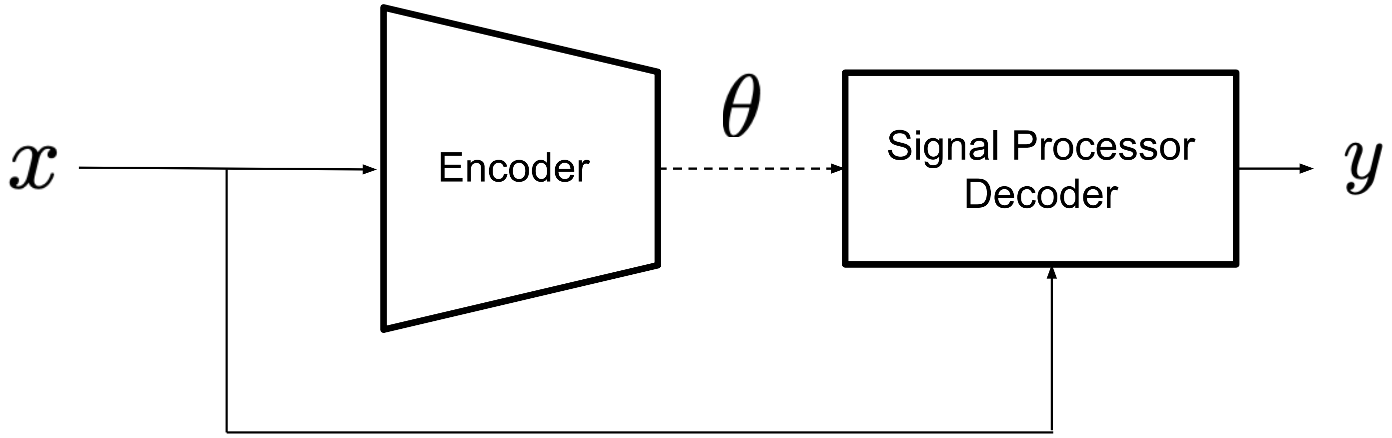


Music separation
& transcription

Music generation
& co-creation

Let's Dive Deeper

- Basic setup
- Advanced setups
- Future



Basic Setup

Must Be Differentiable

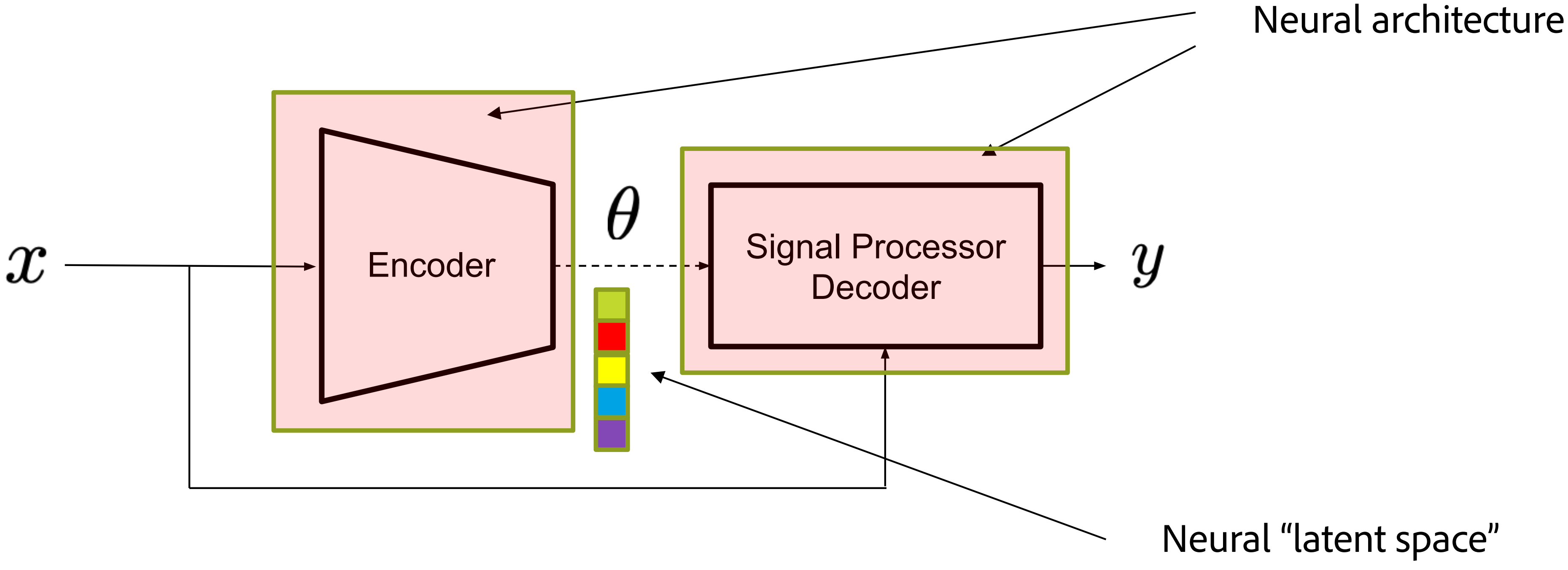
- Neural networks are trained using an algorithm called backpropagation
- All operations performed by the network must be differentiable and have a gradient

TensorFlow PyTorch JAX

Methods

- Neural architecture processor
- Auto-Diff signal processor

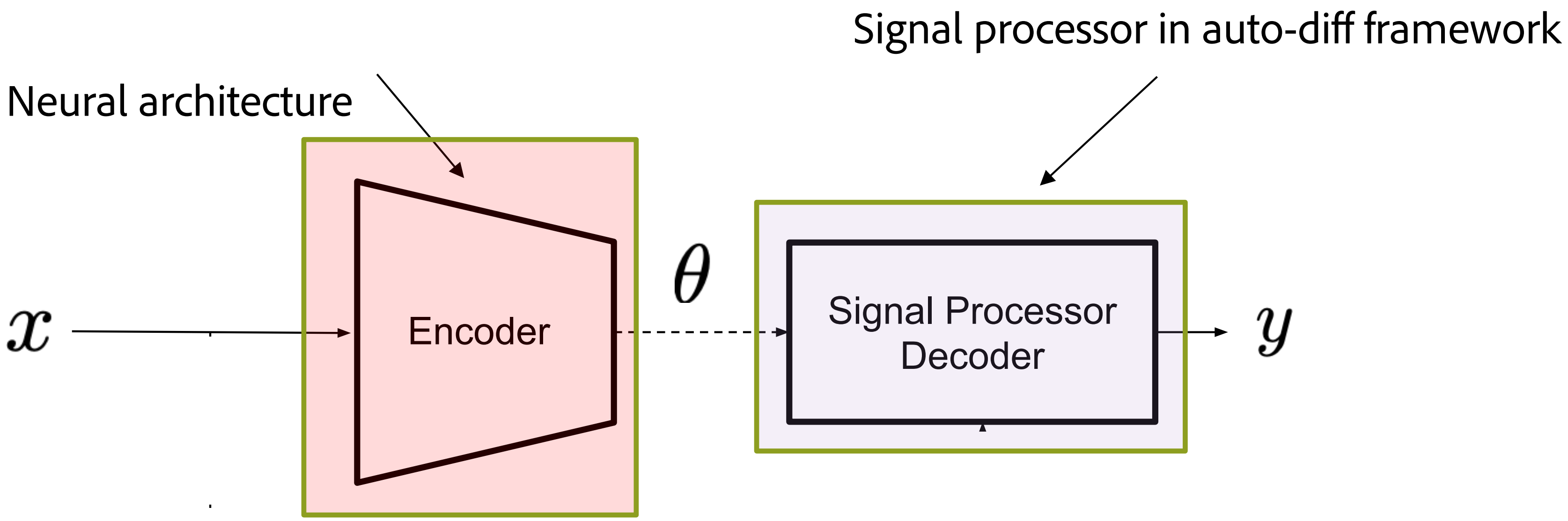
Neural Architecture Processor



- Mimic existing algorithms end-to-end with no interpretable control
- **Example**: Emulate analog music recording equipment

M. Martinez, et al.. Deep learning for black-box modeling of audio effects, Applied Sciences, 2020.

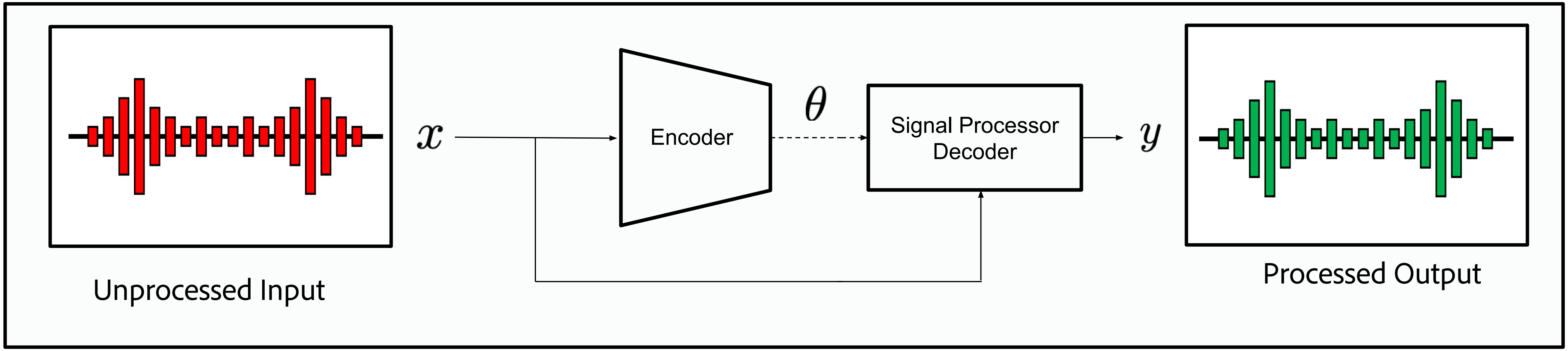
Auto-Diff Signal Processor



- Implement signal processing algorithms using differentiable operators
- **Example**: Speech synthesis & dereverberation

Engel et al., "DDSP: differentiable digital signal processing", ICLR, 2020.

Training



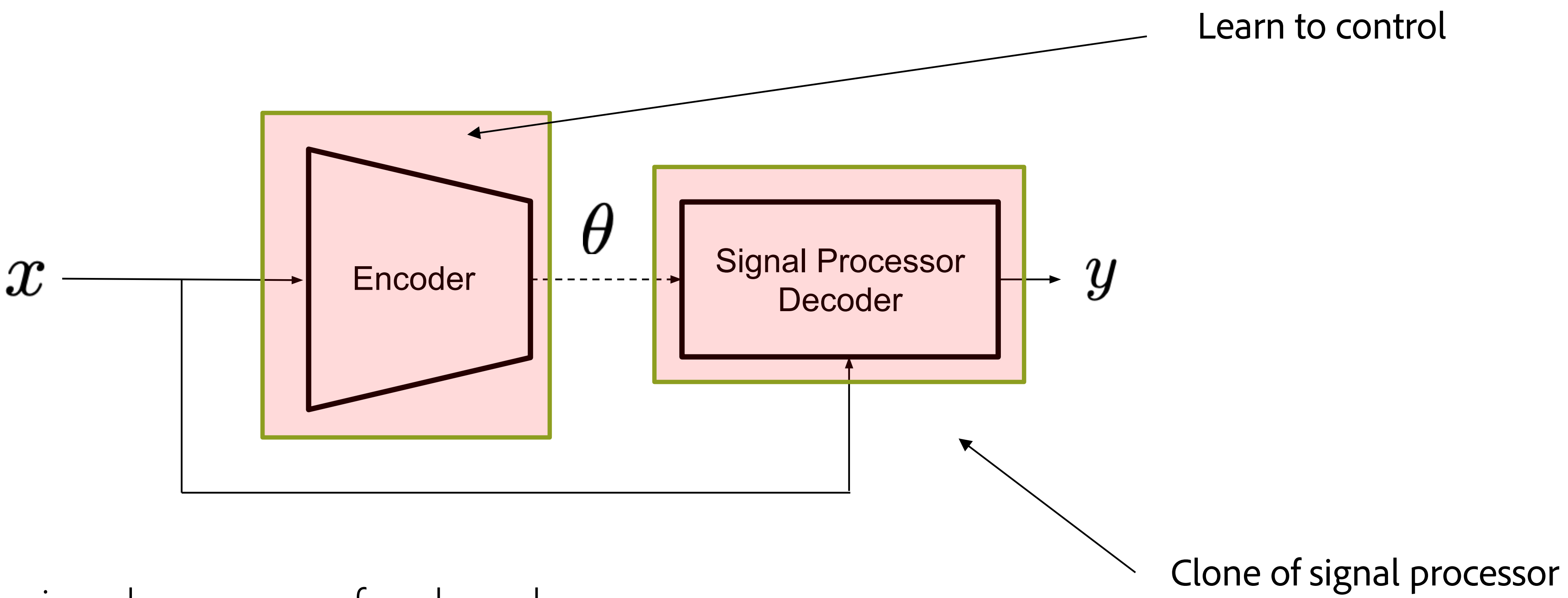
Only need input/output pairs. No parameter labels!

Methods

- Neural architecture processor
- Auto-diff signal processor
- Neural clone
- Neural proxy

No gradients special required!

Neural Clone

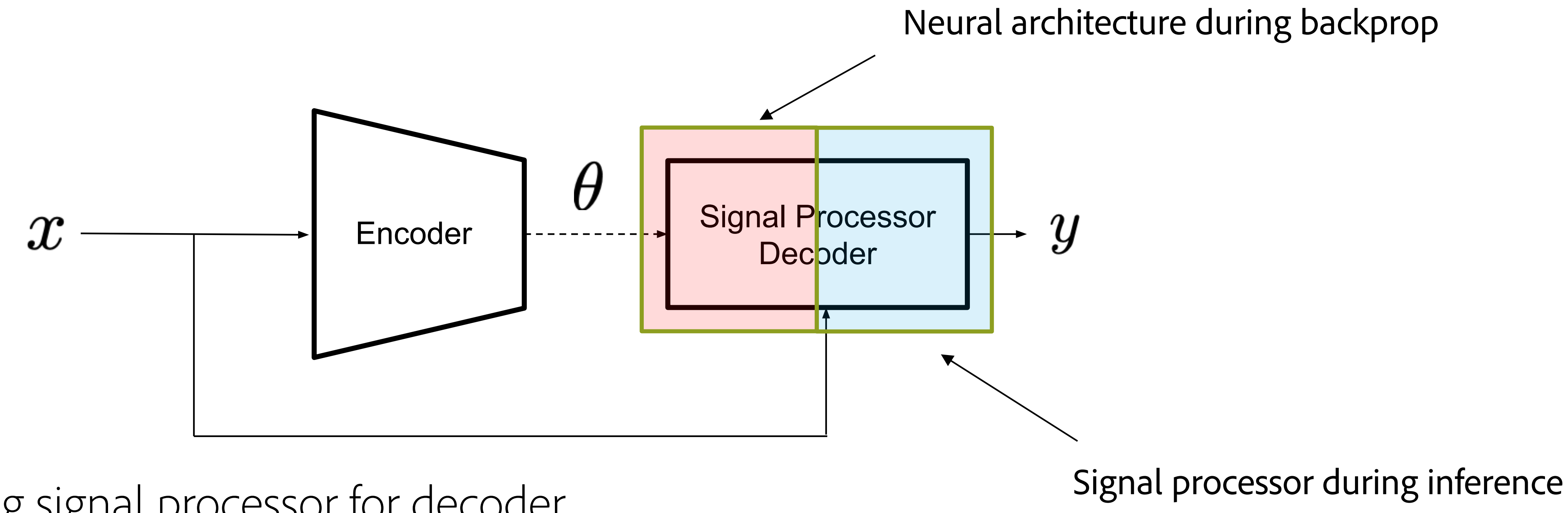


1. Clone, existing signal processor for decoder
2. Train encoder to control

- **Example**: automatic mixing of multi-track audio

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

Neural Proxy

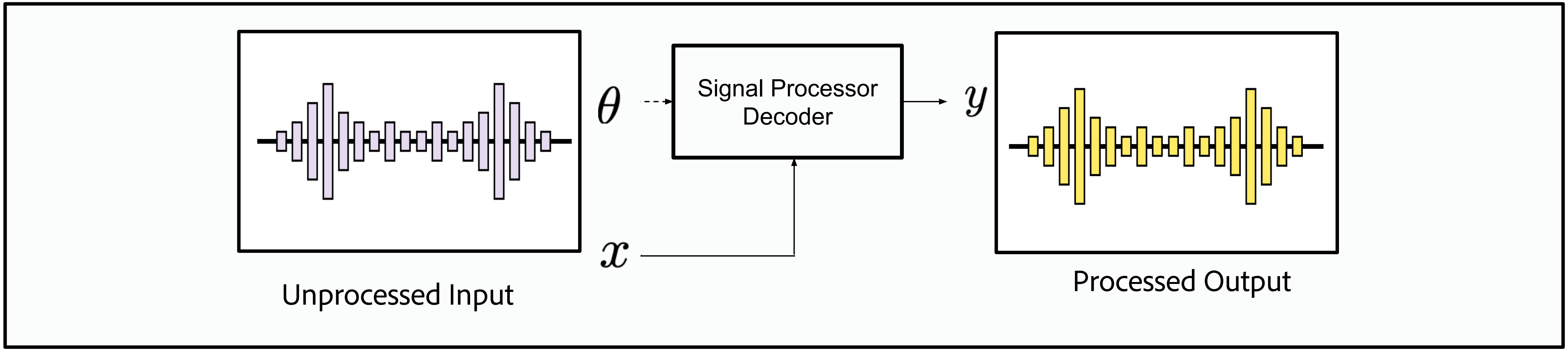


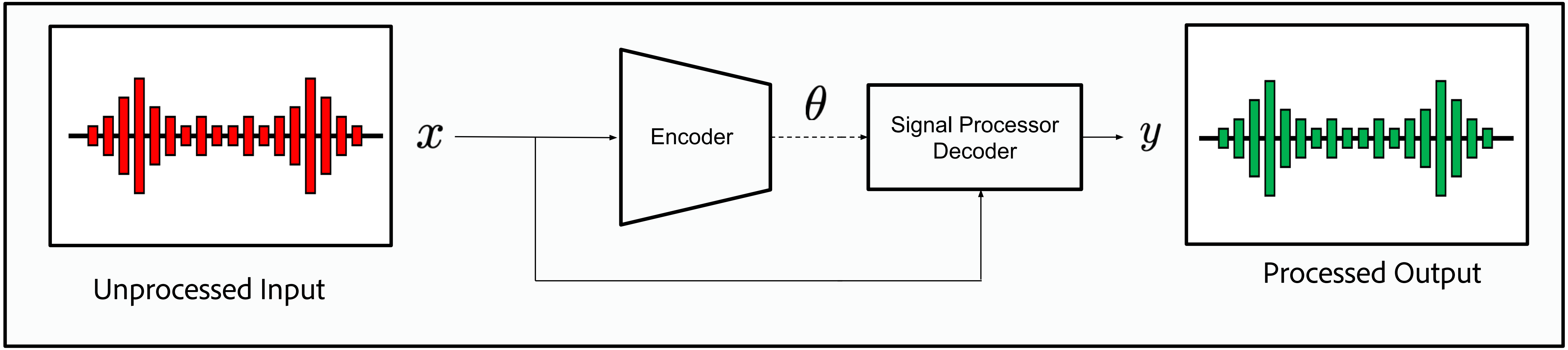
1. Clone existing signal processor for decoder
2. Train encoder to control the clone
3. Only use clone for training control, but use real processor for inference

Example: Optimally tune image processing pipeline

Tseng et al., "Hyperparameter opt. in black-box image processing using differentiable proxies", SIGGRAPH, 2019.

Two-Step Training





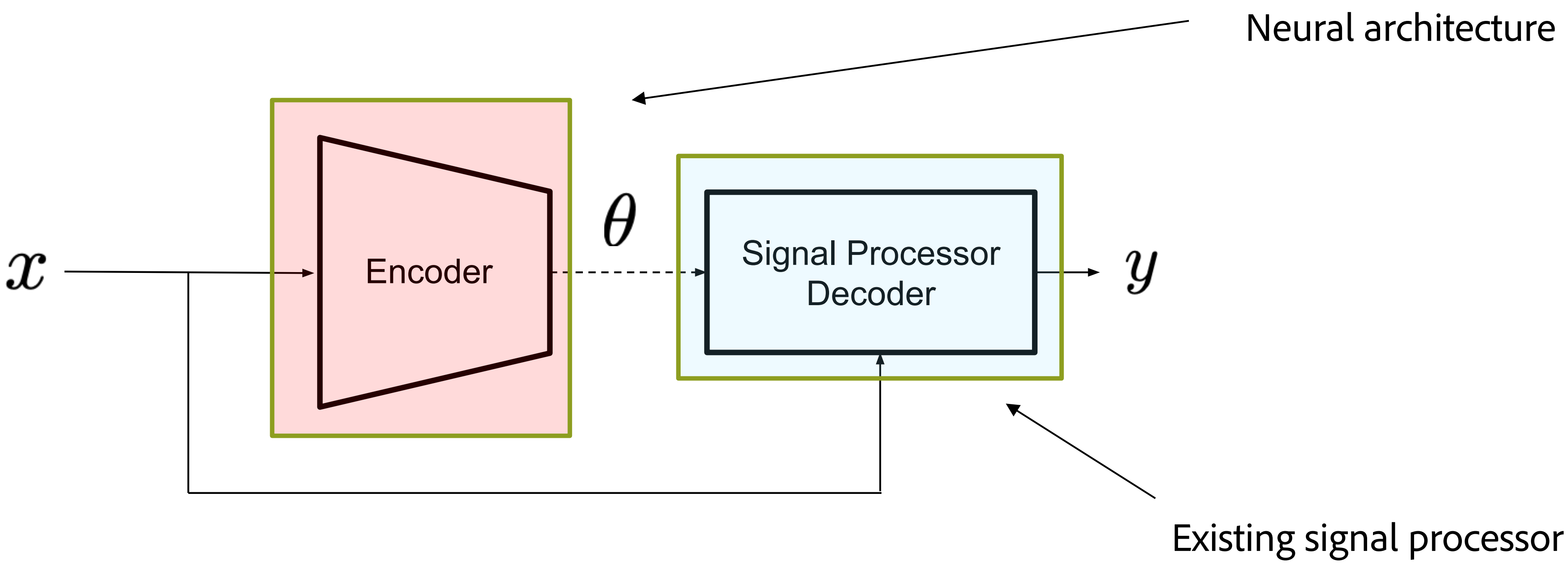
Methods

- Neural architecture processor
- Auto-diff signal processor
- Neural clone
- Neural proxy
- Numerical gradient approximation

No gradients special required!

Use third-party audio plugins!

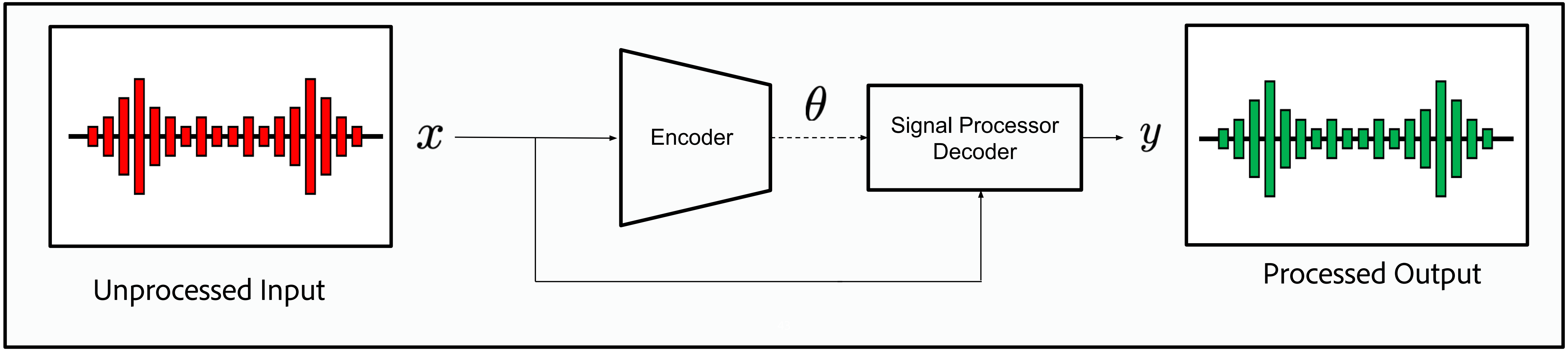
Numerical Gradient Approximation



- Signal processor as black-box function
- Use numerical gradient approx. within auto-diff
- **Example**: Use existing tools to remove breaths, master music, emulate guitar

M. Martinez, O. Wang, P. Smaragdis, and N. J. Bryan, "Differentiable Signal Processing with Black-box Audio Effects", ICASSP, 2021.

One-Step Training



Only need input/output pairs. No parameter labels!

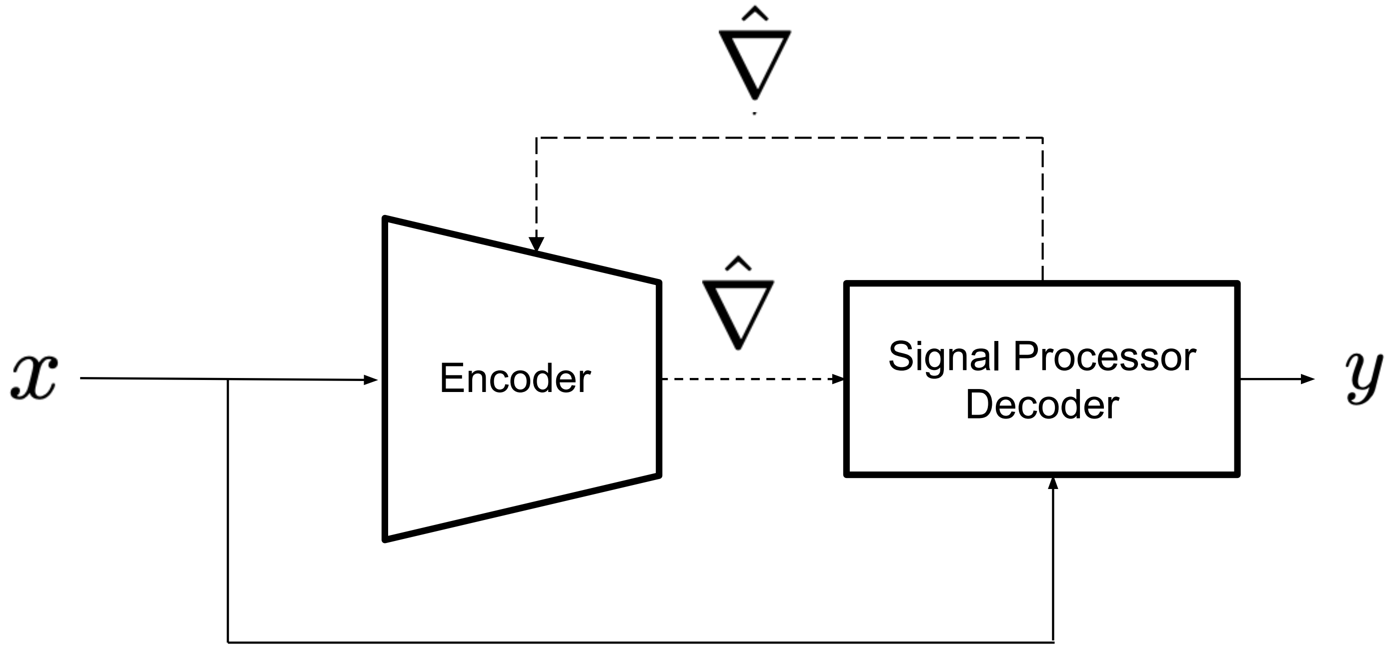
Works with third-party plugins and more!

Recap

- Use neural networks to control signal processing algorithms is powerful
- Many different strategies for differentiable signal processing, not just one
- Each strategy offers unique benefits and trade-offs
- Research area is wide open for exploration

Advanced Setups

Recent and Current Work



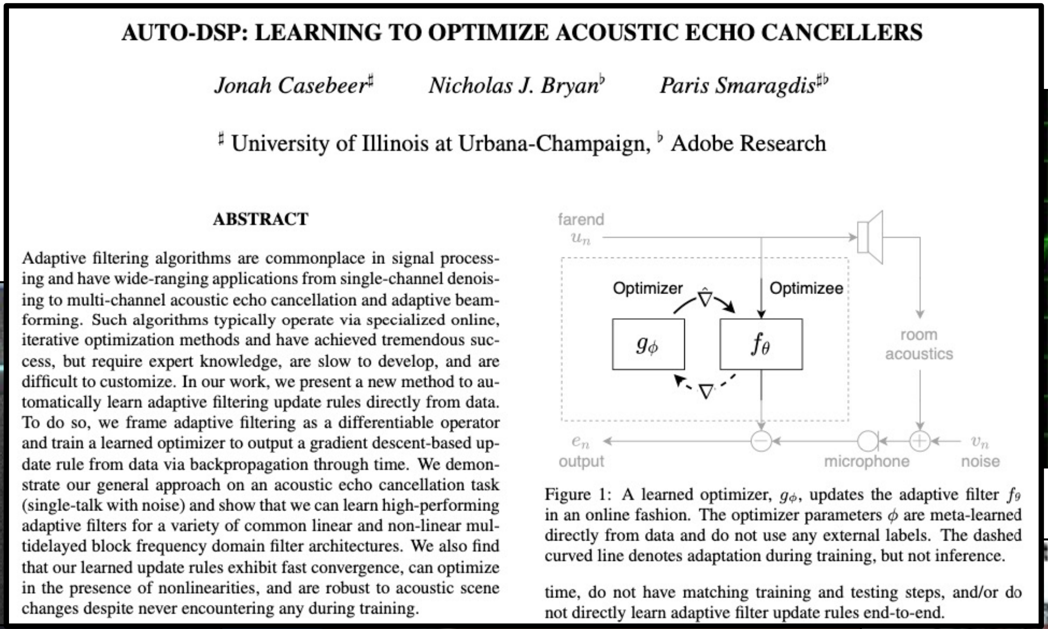
- Train stateful signal processors
- Train signal processors that are online optimization algorithms

J. Casebeer, N. J. Bryan, and P. Smaragdis, "Auto-DSP: Learning to Optimize Acoustic Echo Cancellers", WASPAA, 2021.

Adaptive Filters

"Auto-DSP: Learning to Optimize Acoustic Echo Cancellers."
J. Casebeer, N. J. Bryan P. Smaragdis, *WASPAA*, 2021

- Noise reduction
- De-reverberation
- Source separation
- Beamforming
- Echo Cancellation



AUTO-DSP: LEARNING TO OPTIMIZE ACOUSTIC ECHO CANCELLERS

Jonah Casebeer[♯] *Nicholas J. Bryan*[♭] *Paris Smaragdis*[♯♭]

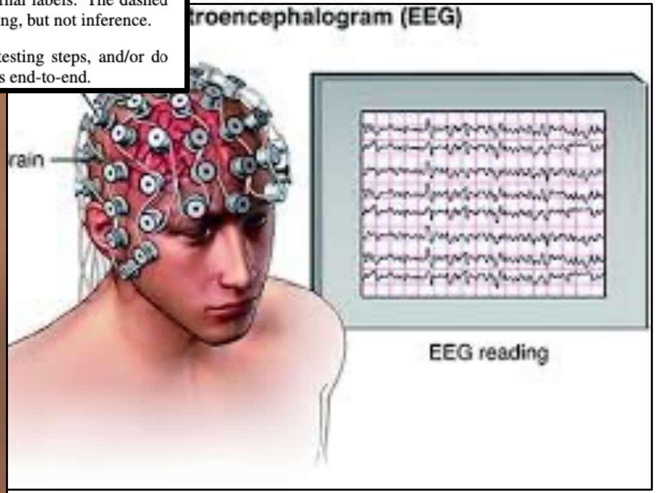
[♯] University of Illinois at Urbana-Champaign, [♭] Adobe Research

ABSTRACT

Adaptive filtering algorithms are commonplace in signal processing and have wide-ranging applications from single-channel denoising to multi-channel acoustic echo cancellation and adaptive beamforming. Such algorithms typically operate via specialized online, iterative optimization methods and have achieved tremendous success, but require expert knowledge, are slow to develop, and are difficult to customize. In our work, we present a new method to automatically learn adaptive filtering update rules directly from data. To do so, we frame adaptive filtering as a differentiable operator and train a learned optimizer to output a gradient descent-based update rule from data via backpropagation through time. We demonstrate our general approach on an acoustic echo cancellation task (single-talk with noise) and show that we can learn high-performing adaptive filters for a variety of common linear and non-linear multidelayed block frequency domain filter architectures. We also find that our learned update rules exhibit fast convergence, can optimize in the presence of nonlinearities, and are robust to acoustic scene changes despite never encountering any during training.

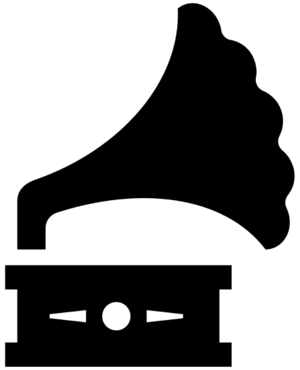
Figure 1: A learned optimizer, g_ϕ, updates the adaptive filter f_θ in an online fashion. The optimizer parameters ϕ are meta-learned directly from data and do not use any external labels. The dashed curved line denotes adaptation during training, but not inference.

time, do not have matching training and testing steps, and/or do not directly learn adaptive filter update rules end-to-end.



Future

Applications for ISMIR



Music mastering

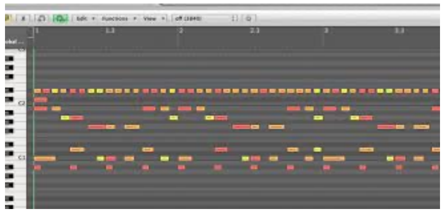
Guitar/FX modeling

Voice processing

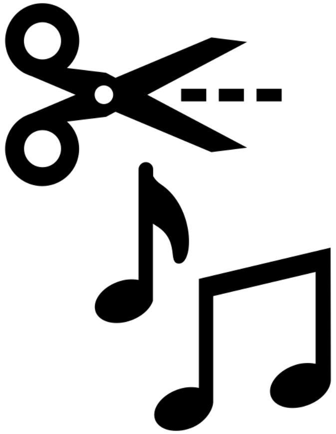
Music separation
& transcription

Music generation
& co-creation

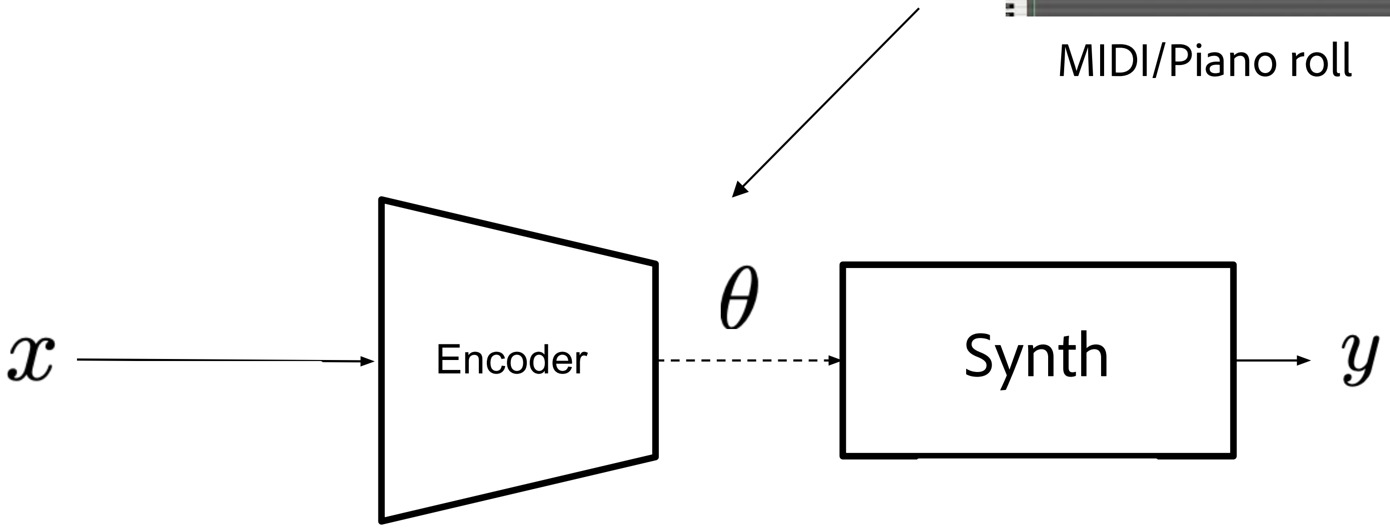
Music Separation & Transcription



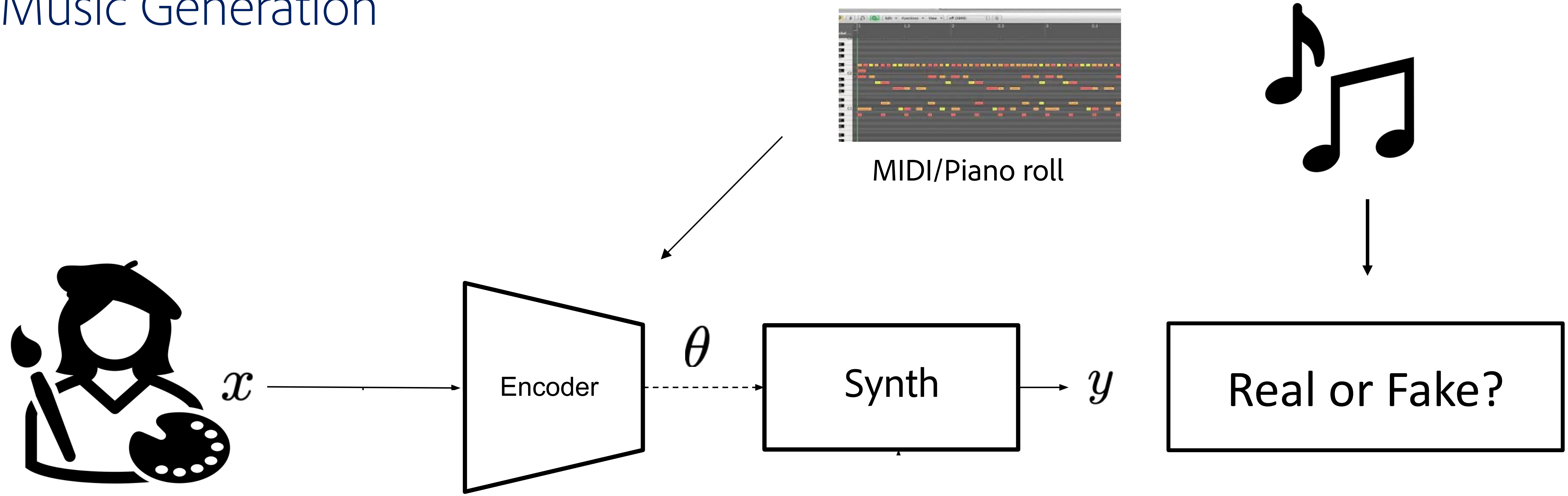
MIDI/Piano roll



Music separation
& transcription



Music Generation



Music generation
& co-creation

Conclusions

- Controlling signal processing algorithms with deep learning is powerful!
- Many methods for differentiable signal processing!
- Research area has many unsolved problems!
- Many music signal processing applications for WiMIR, ISMIR, and beyond!

Nicholas J. Bryan | Adobe Research
nibryan@adobe.com
https://njb.github.io

Thank you!

References

- S. I. Mimilakis, N. J. Bryan, and P. Smaragdis. "One-shot parametric audio production style transfer with application to frequency equalization." *ICASSP*, 2020.
- M. A. Martínez R., O. Wang, P. Smaragdis, N. J. Bryan. "Differentiable Signal Processing With Black-Box Audio Effects." *ICASSP*, 2021.
- J. Casebeer, N. J. Bryan, and P. Smaragdis. "Auto-DSP: Learning to Optimize Acoustic Echo Cancellers." WASPAA, 2021.
- M. A. Martínez R., E. Benetos, and J. D. Reiss. "Deep learning for black-box modeling of audio effects." *Applied Sciences* 10.2 (2020): 638.
- J. Engle, L. Hantrakul, C. Gu, A. Roberts, "DDSP: Differentiable Signal Processing", ICLR, 2020.
- C. J. Steinmetz, J. Pons, S. Pascual, J. Serra. "Automatic multitrack mixing with a differentiable mixing console of neural audio effects." *ICASSP*, 2021.
- E. Tseng, F. Yu, Y. Yang, F. Mannan, K. St. Arnaud, D. Nowrouzezahrai, J.F. Lalonde, F. Heide,. "Hyperparameter optimization in black-box image processing using differentiable proxies." *ACM Trans. Graph.* 2019.