PAT 498/598 (Winter 2025)

Music & AI

Lecture 12: Source Separation

Instructor: Hao-Wen Dong

SCHOOL OF MUSIC, THEATRE & DANCE
PERFORMING ARTS TECHNOLOGY
UNIVERSITY OF MICHIGAN

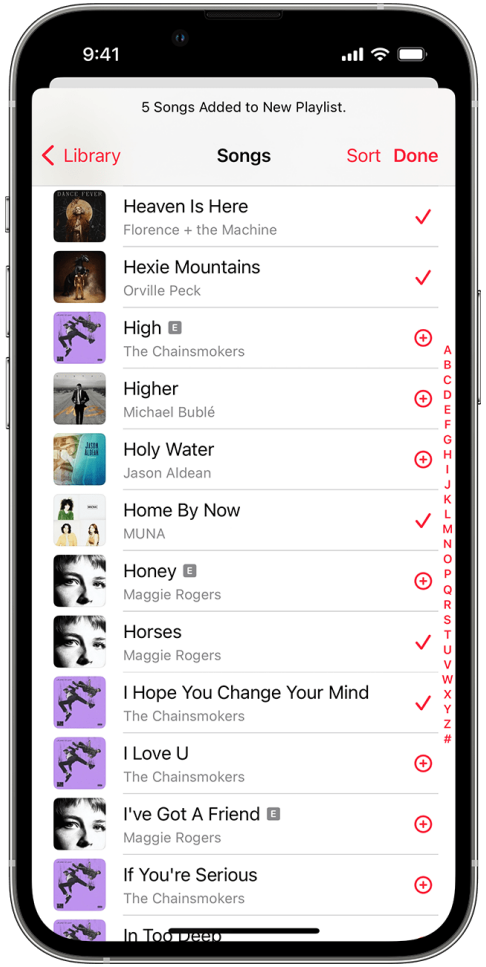
(Recap) Music Classification Tasks

- **Genre** classification (pop, rock, r&b, jazz, hip-hop, classical, etc.)
- **Mood** classification (happy, sad, calm, aggressive, cheerful, etc.)
- **Instrument** recognition
- **Composer** identification
- **Key** detection
- **Chord** estimation
- **Music tagging** → **Can cover everything above!**

(Recap) Music Classification for Recommendation

What to play next?

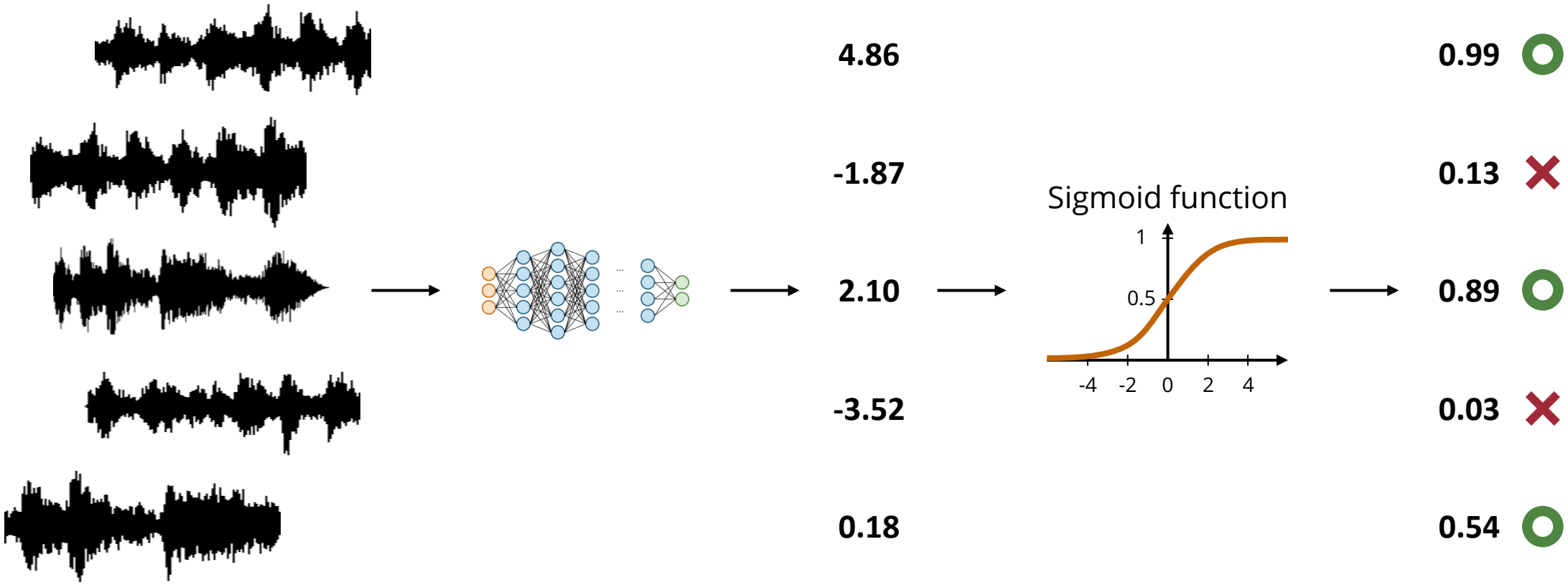
(Recap) Music Classification for Playlist Generation



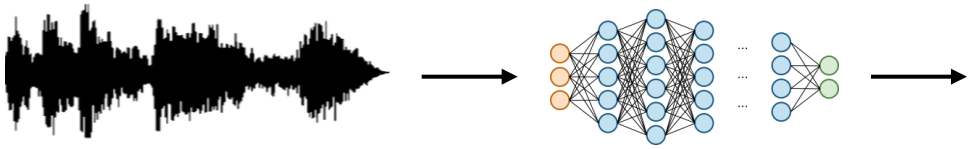
(Recap) Types of Classification Tasks

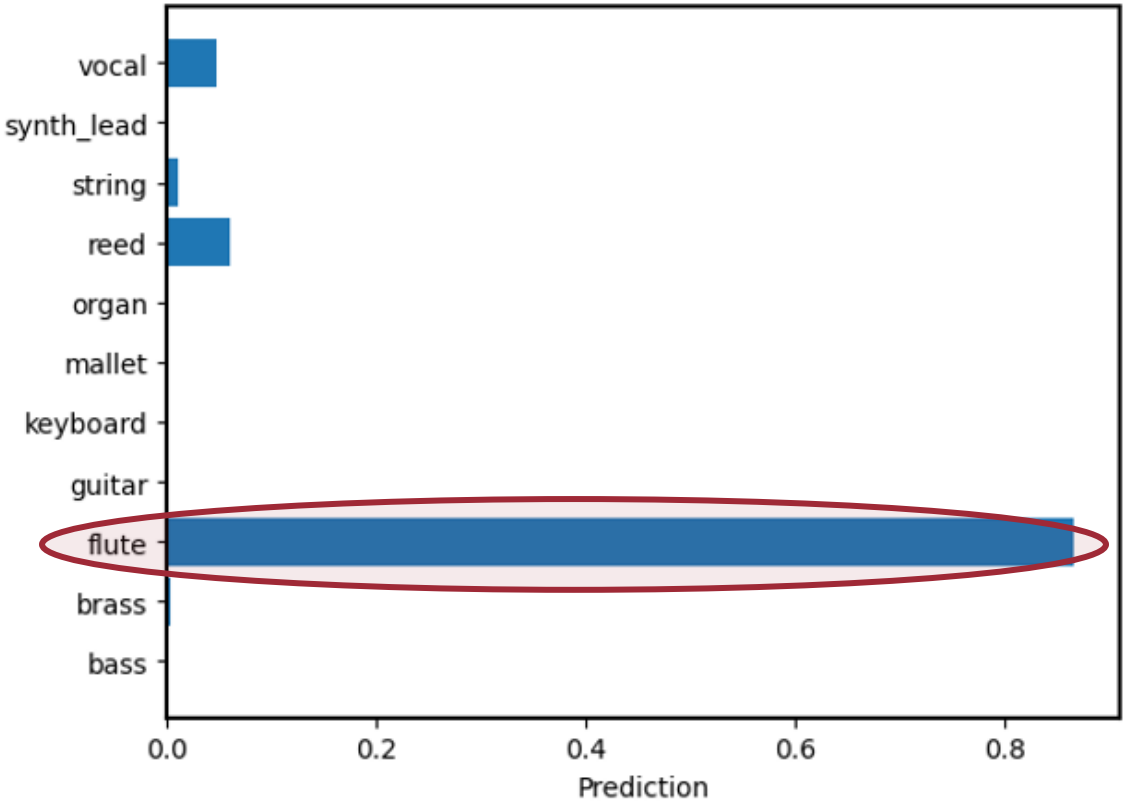
- **Binary** classification
- **Multiclass** classification
- **Multi-label** classification

(Recap) Binary Classification



(Recap) Multiclass Classification





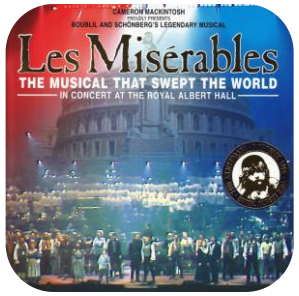
(Recap) Multi-label Classification

Soul, R&B, pop

Soundtrack, classical

Rock

Classical



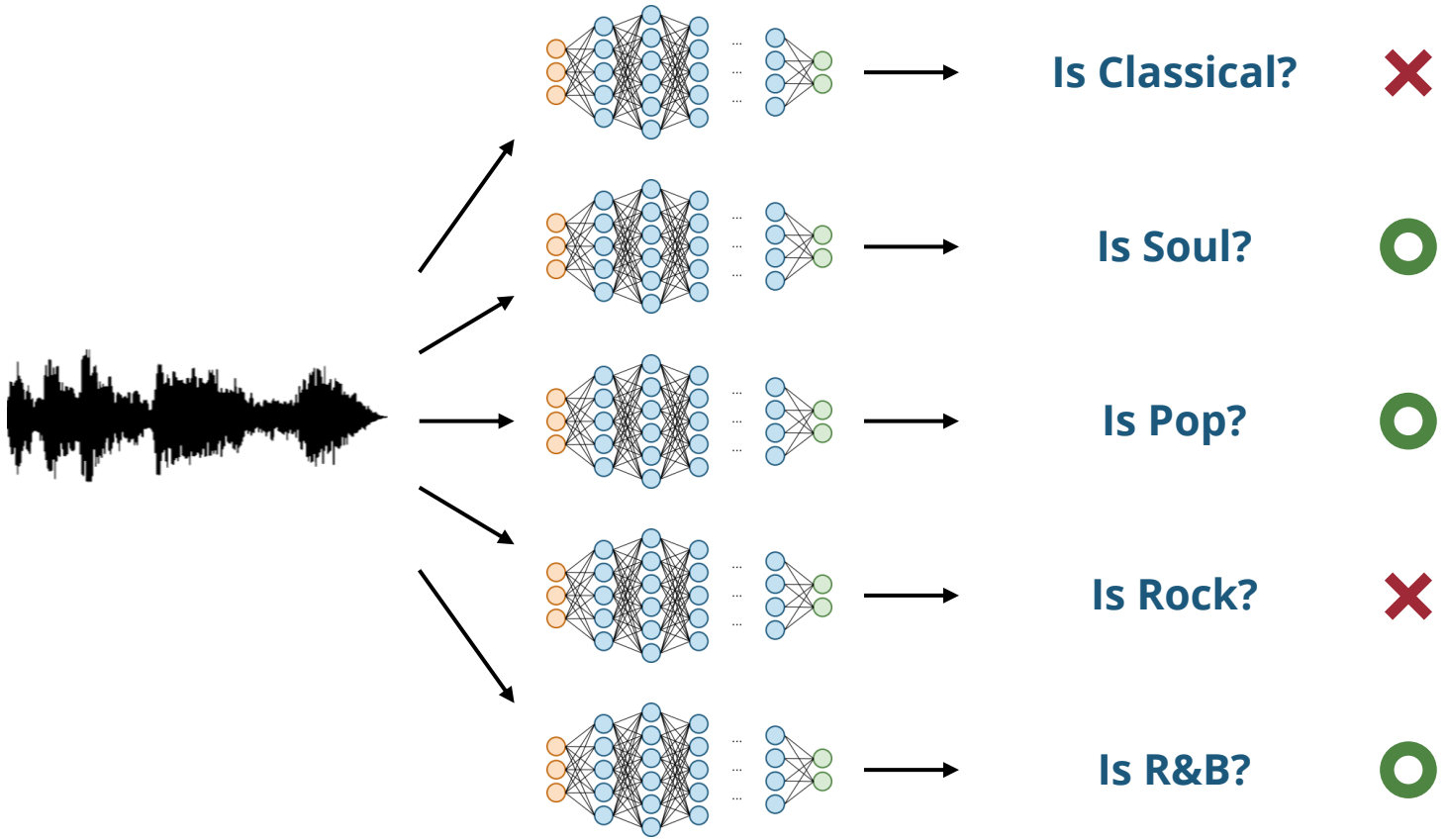
Musical

Alternative, rock

Pop

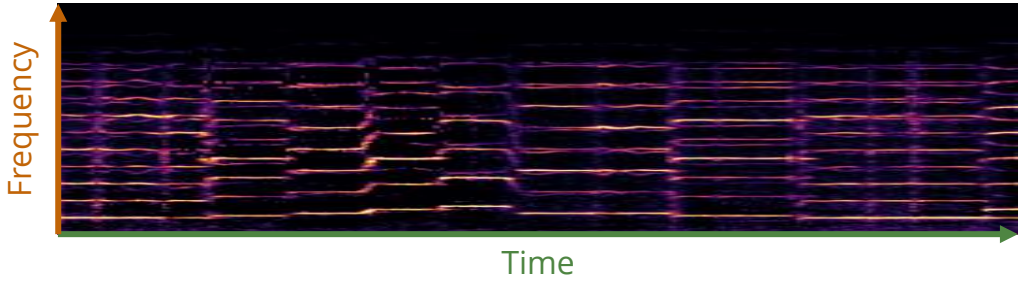
Pop, soul, R&B

(Recap) Multi-label Classification as Binary Classification



(Recap) Input Features

- **Waveform**
- Time-frequency representation (**spectrograms**)
- **Hand-crafted features** or **features provided in metadata**
  - Acoustic: loudness, pitch, timbre
  - Rhythmic: beat, tempo, time signature
  - Tonal: key, scale, chords
  - Instrumentation, expressions, structures, etc.



(Recap) Common Datasets

- **GTZAN**: 1,000 30-sec songs, 10 genres
- **MagnaTagATune**: 5,405 29-sec songs, 188 tags, 230 artists
- **Million Song Dataset (MSD)**: **1M** 30-sec songs, >500K tags, **tricky to access**
- **Free Music Archive (FMA)**: >10K **full** songs, 163 genres
- **MTG-Jamendo**: 55K **full** songs, 195 tags
- **AudioSet**: **1M** songs, YouTube URLs, **low-quality audio**
- **NSynth**: ~306K 4-sec instrument sounds

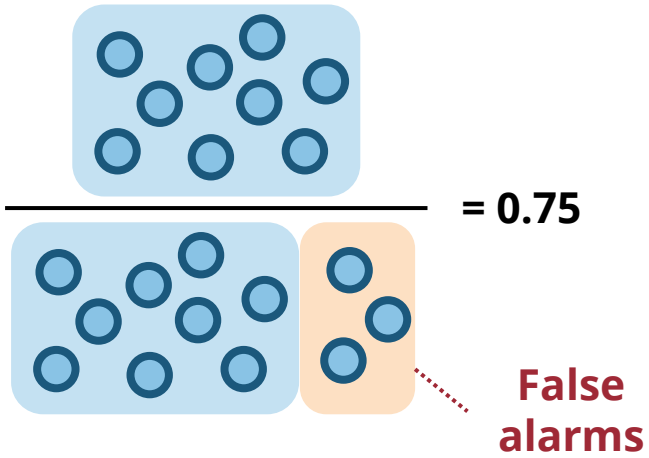
(Recap) Confusion Matrix for Binary Classification

Ground truth	Prediction: Positive	Prediction: Negative
Positive	True Positives (TP)  Cat	False Negatives (FN)  Not Cat
Negative	False Positives (FP)  Cat	True Negatives (TN)  Not Cat

(Recap) Precision vs Recall

Precision

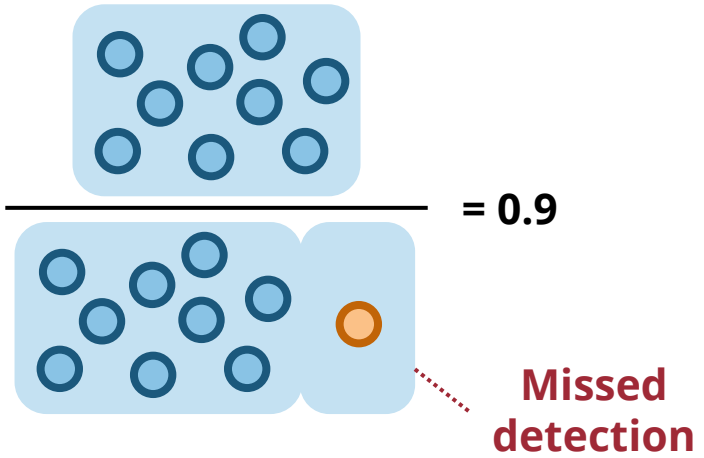
$$\frac{TP}{TP + FP} = \frac{\text{TP}}{\text{TP + FP}} = 0.75$$



How often predictions for the positive are correct

Recall

$$\frac{TP}{TP + FN} = \frac{\text{TP}}{\text{TP + FN}} = 0.9$$



How well the model finds all positive instances in the dataset

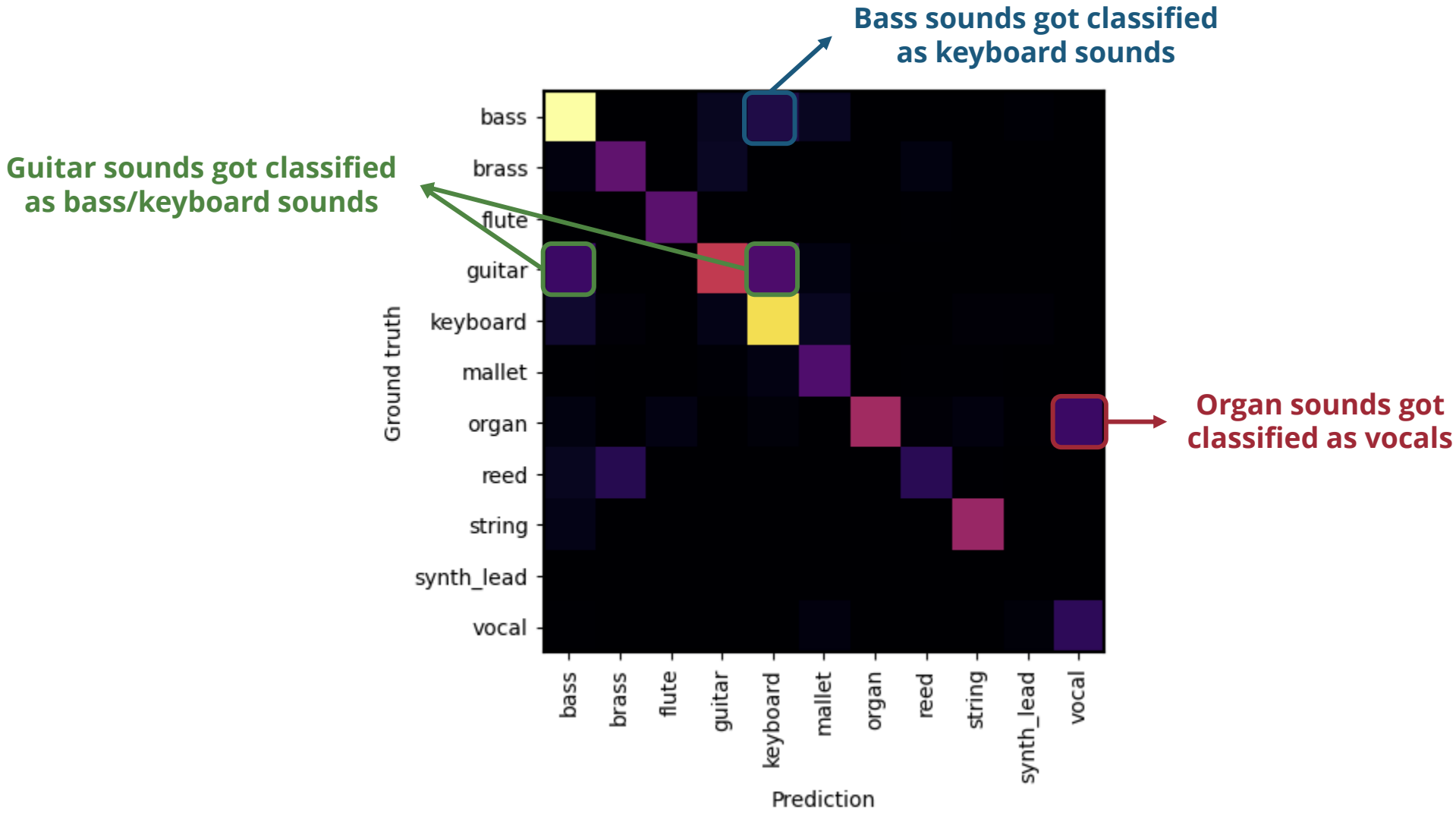
(Recap) F1 Score: Considering both Precision & Recall

- Particularly useful for **imbalanced datasets**
  - Work better than accuracy when the dataset is imbalanced
  - For example, music search, retrieval, and recommendation

$$F_1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

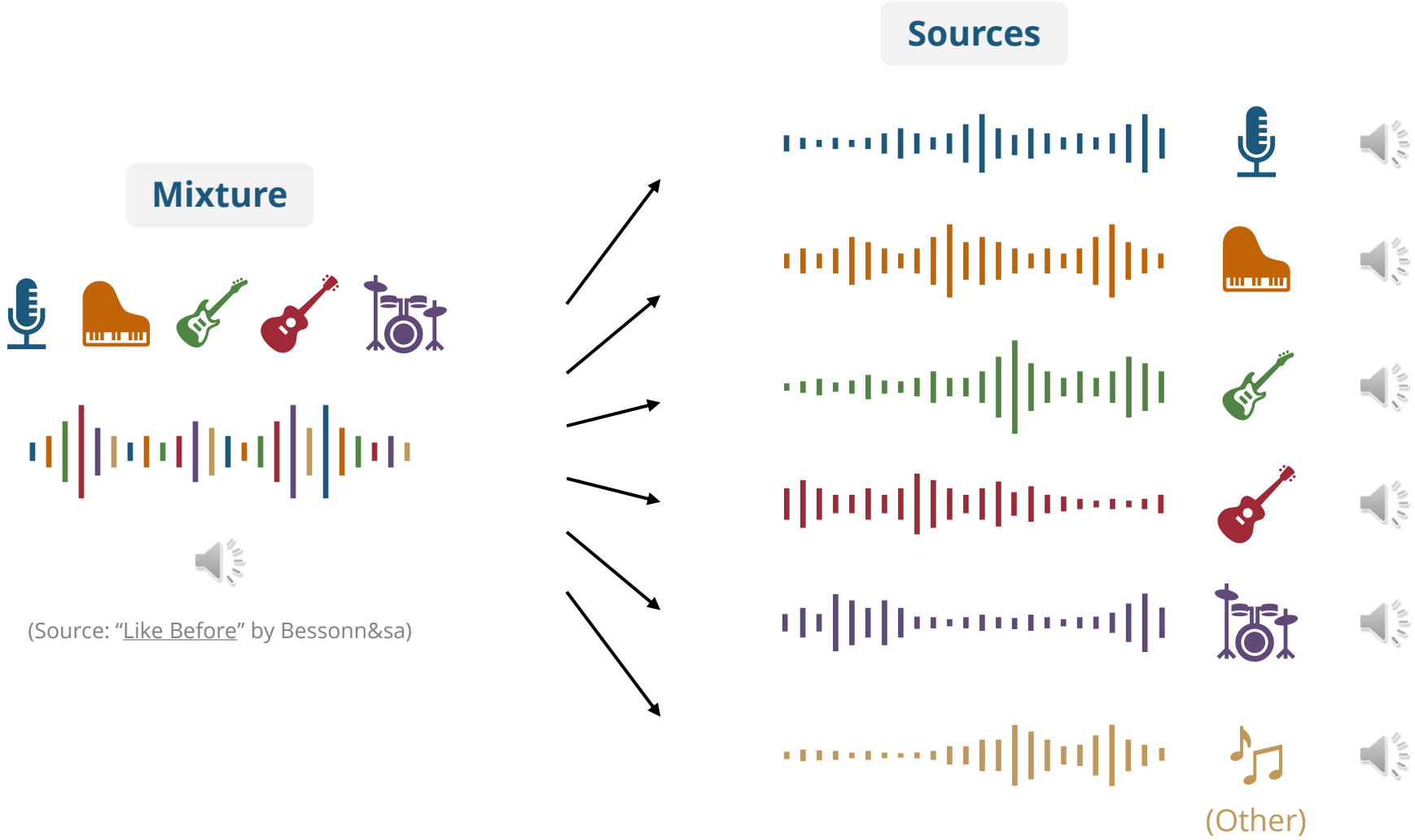
$$= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

(Recap) Confusion Matrix for Multi-label Classification



Source Separation

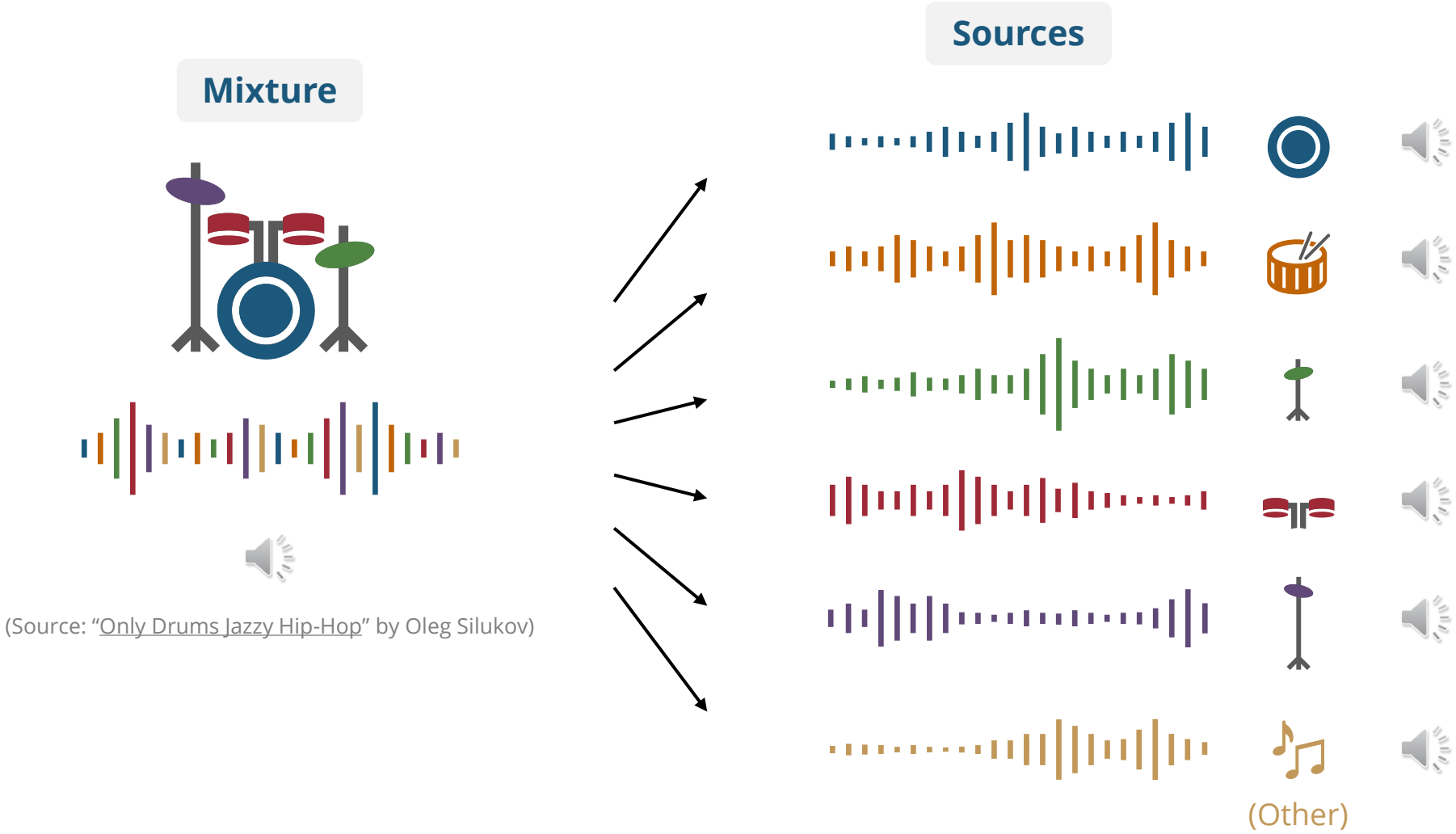
Source Separation



(Source: "Like Before" by Bessonn&sa)

Bessonn&sa, CC BY-NC-SA, via Jamendo
Illustration inspired by Ethan Manilow, Prem Seetharman, and Justin Salamon's Tutorial on "Open Source Tools & Data for Music Source Separation" at ISMIR 2020.

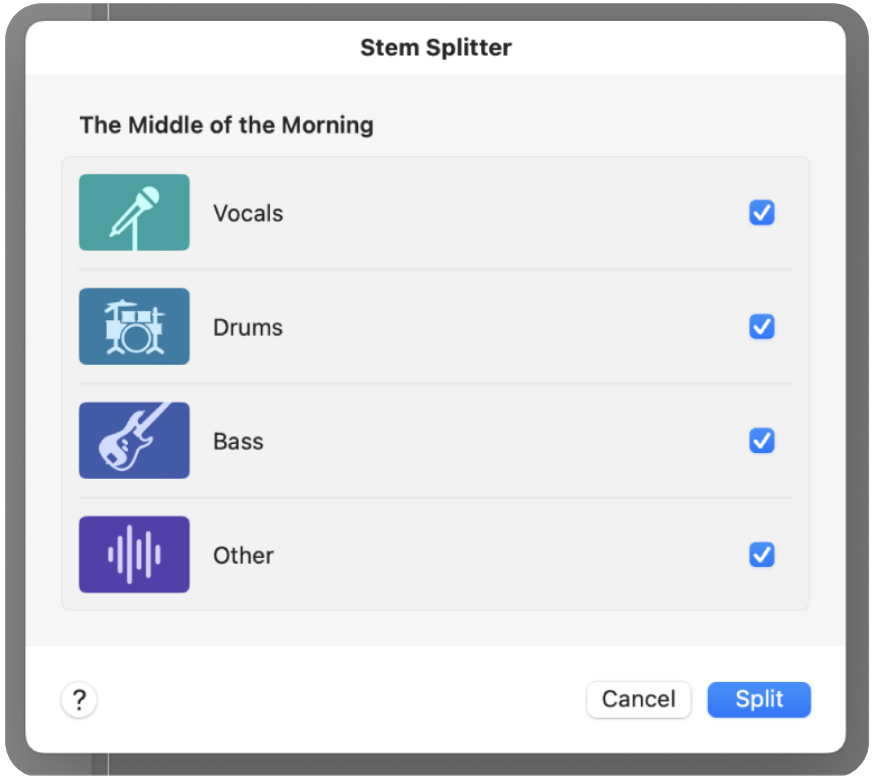
Source Separation



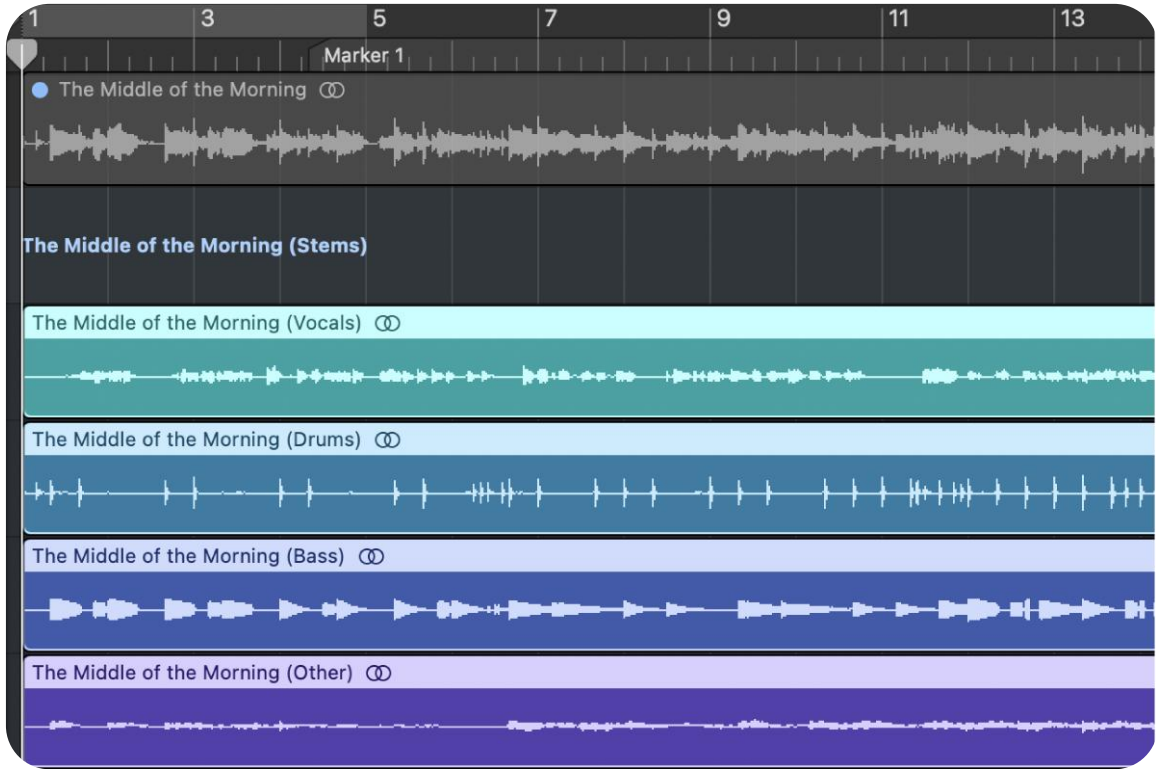
(Source: "Only Drums Jazzy Hip-Hop" by Oleg Silukov)

Oleg Silukov, CC BY-NC-SA, via Jamendo
Illustration inspired by Ethan Manilow, Prem Seetharman, and Justin Salamon's Tutorial on "Open Source Tools & Data for Music Source Separation" at ISMIR 2020.

Stem Splitter in Logic Pro



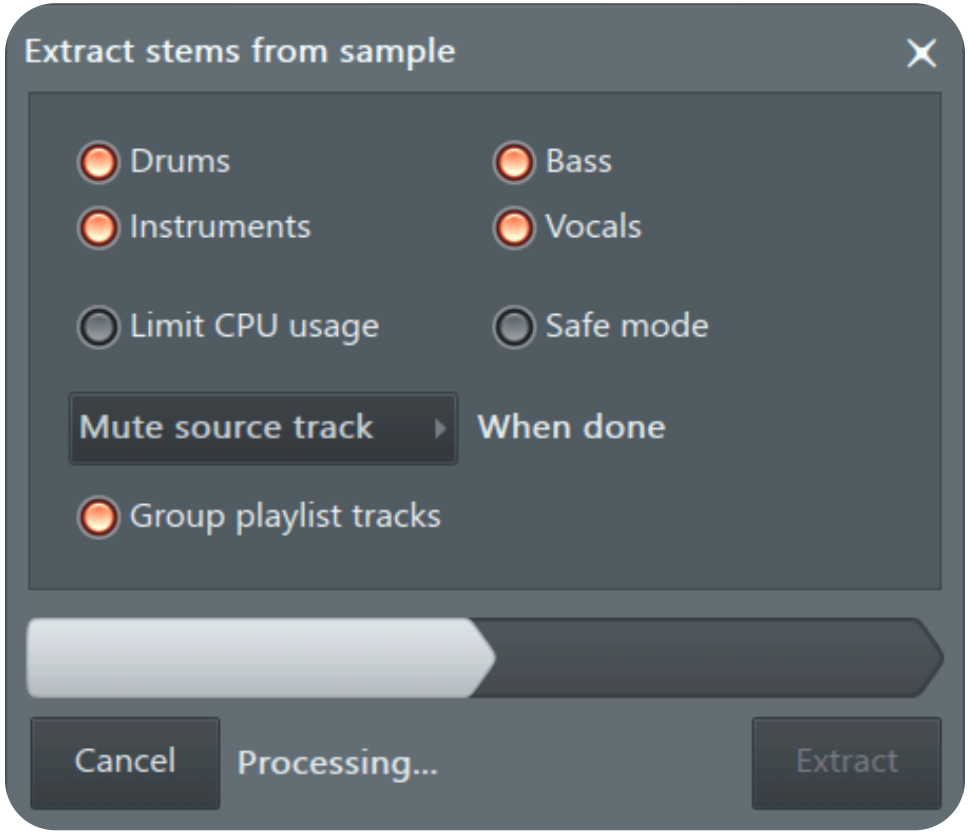
(Source: Logic Pro User Guide)



(Source: Logic Pro User Guide)

support.apple.com/guide/logicpro/extract-vocal-instrumental-stems-stem-lgcp61bae908/mac

Extracting Stems from Sample in FL Studio



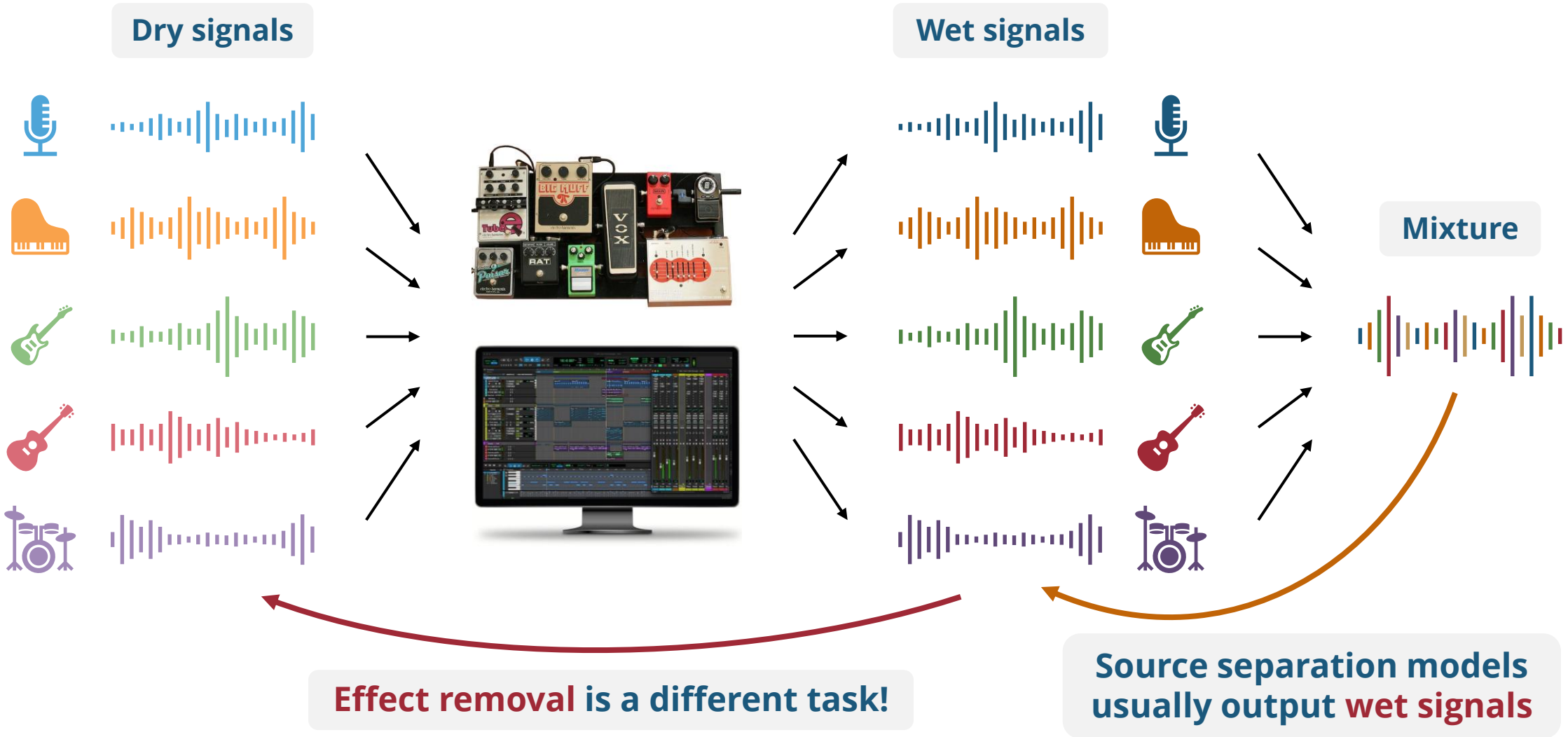
(Source: FL Studio Reference Manual)

image-line.com/fl-studio-learning/fl-studio-online-manual/html/playlist.htm#audio_clip_extractstems

Source Separation

- Also known as
  - **Stem separation**
  - **Stem splitter**
  - **Music demixing** → slightly different meaning
  - **Stem extraction** → slightly different meaning

Source Separation does NOT Remove Effects



Applications of Source Separation

- **Musical applications**
  - Remixing & sampling
  - Music practicing & education
  - Karaoke accompaniment generation
- **MIR tasks** (Oftentimes source separation is the first step)
  - Music transcription
  - Musical instrument & vocal detection
  - Singer identification
  - Lyric recognition
  - Lyric-to-music alignment

Ethan Manilow, Prem Seetharman, and Justin Salamon, "Open Source Tools & Data for Music Source Separation," *Tutorials of ISMIR*, 2020.

Moises

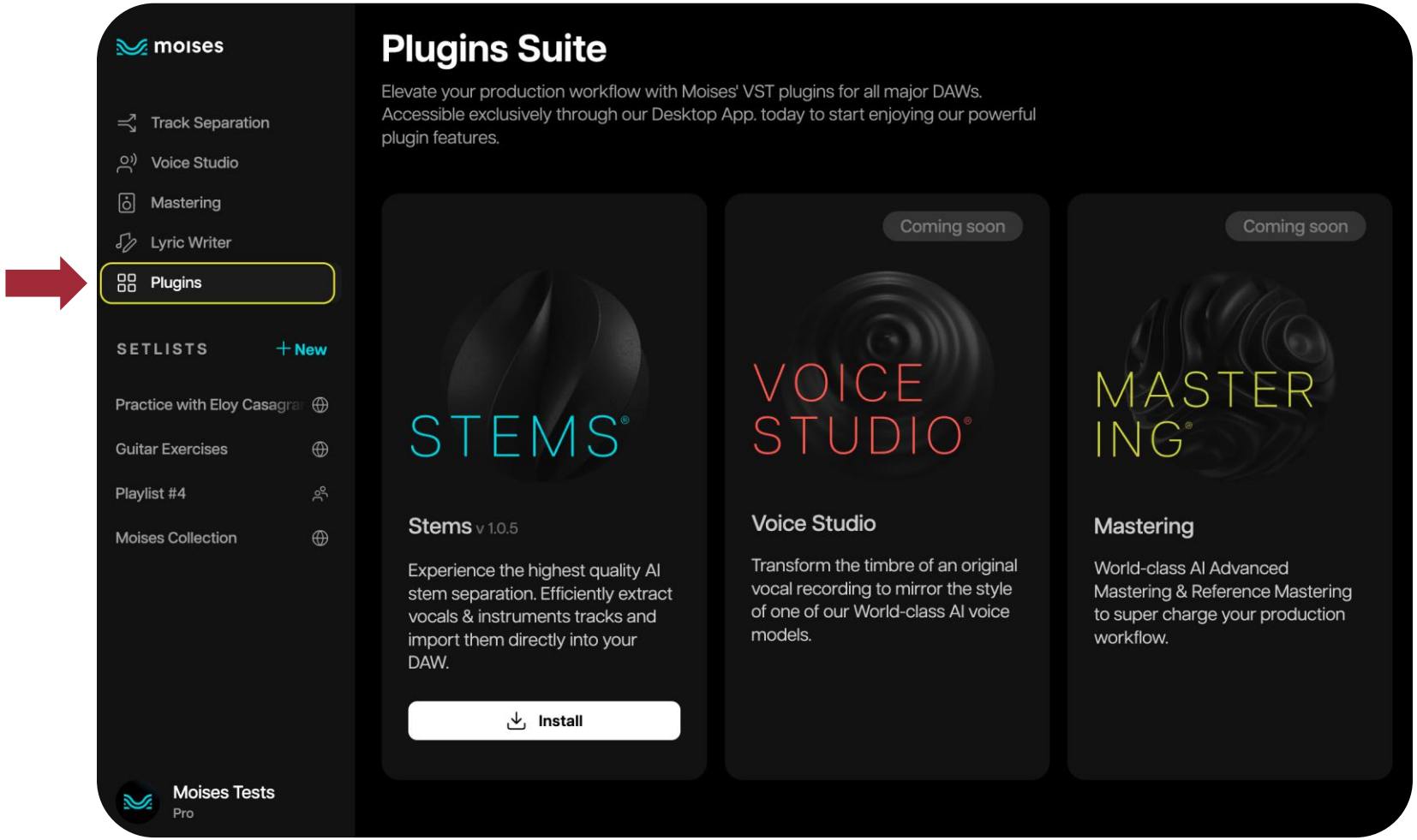
Moises

- 🤩 **Free Moises Pro license** until Summer 2025
- Register at studio.moises.ai/claim-trial/UMichFree/monthly/
  - Use your **U-M email** (@umich.edu)
  - Sign up in your **desktop browser**
  - Ignore the prompt to upgrade your account
  - **Deadline to sign up: March 14**

Moises Demo

youtu.be/cyXPgU5UiB8

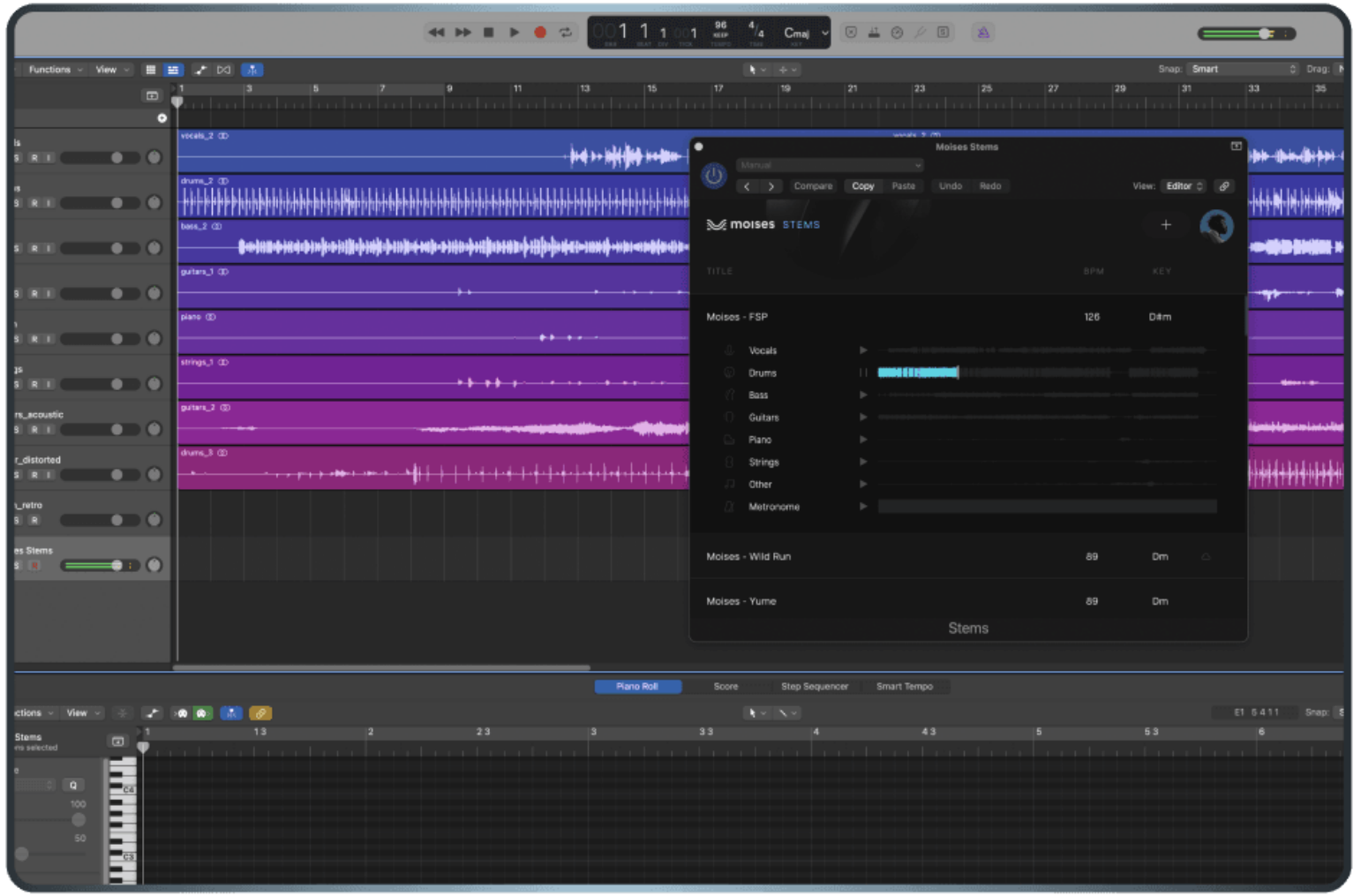
Moises VST Plugin



(Source: Moises)

help.moises.ai/hc/en-us/articles/11228246270876-How-do-I-set-up-Moises-Stems-on-a-DAW

Moises VST Plugin



(Source: Moises)

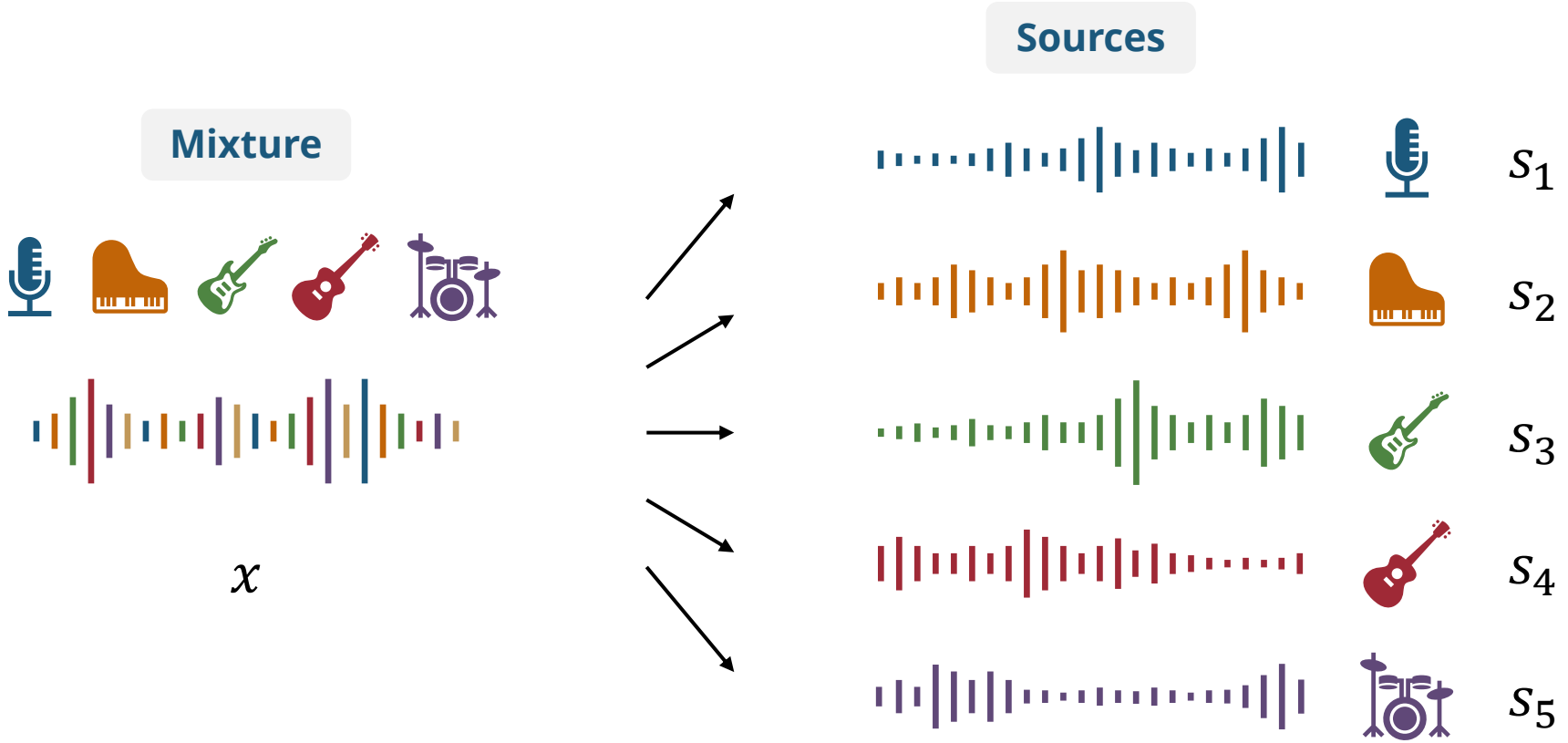
moises.ai/features/stems-vst-plugin/

How to Produce Music with Moises

- Part 1: “This is the way that a lot of tracks start”
- Part 2: Creating backing vocals from scratch
- Part 3: A new way to mix drums

How does it work?

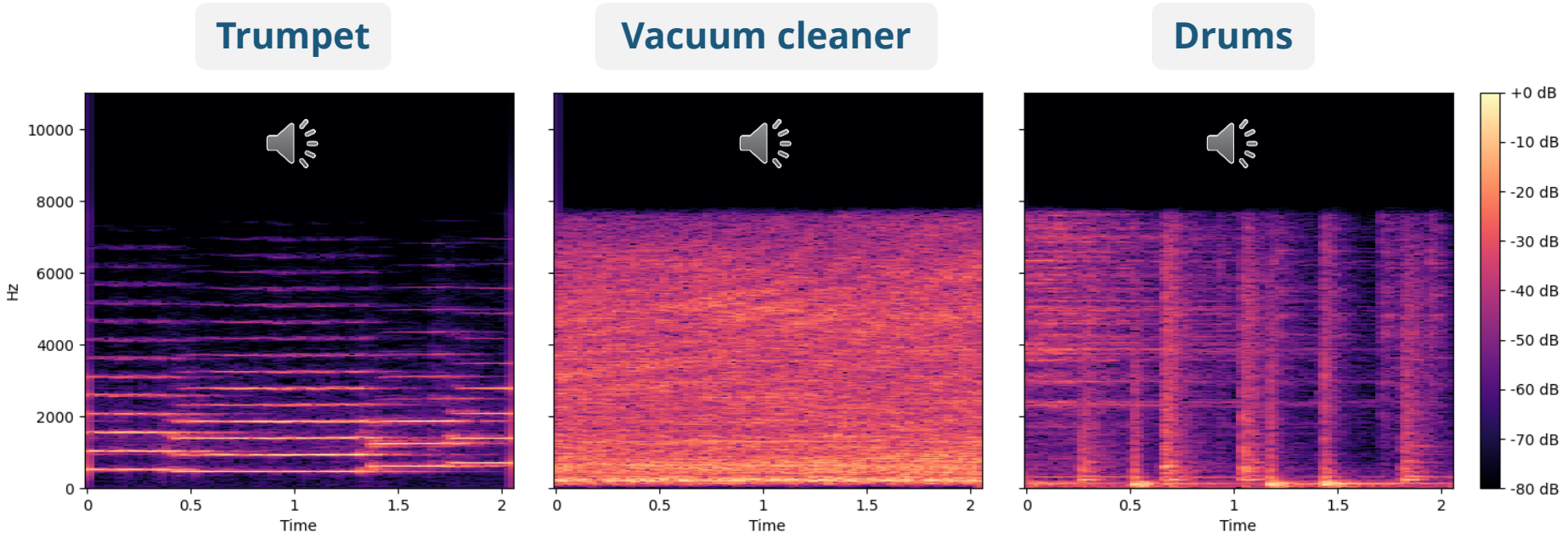
Mathematical Formulation



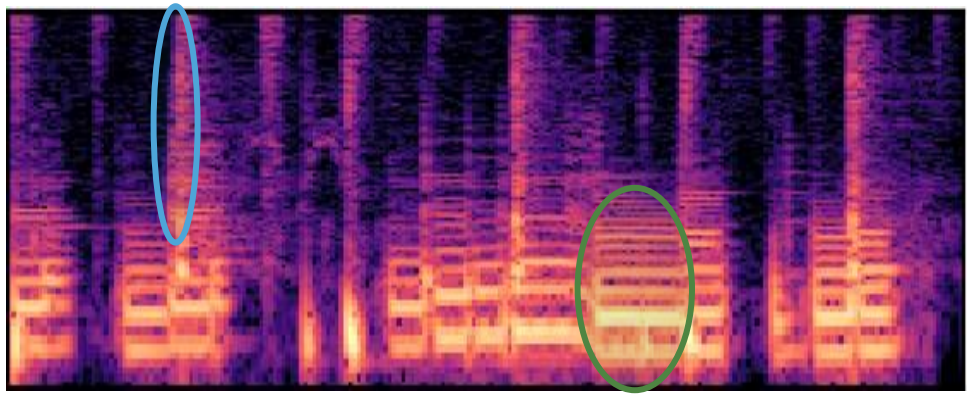
$$x = s_1 + s_2 + s_3 + s_4 + s_5 = \sum_i s_i$$

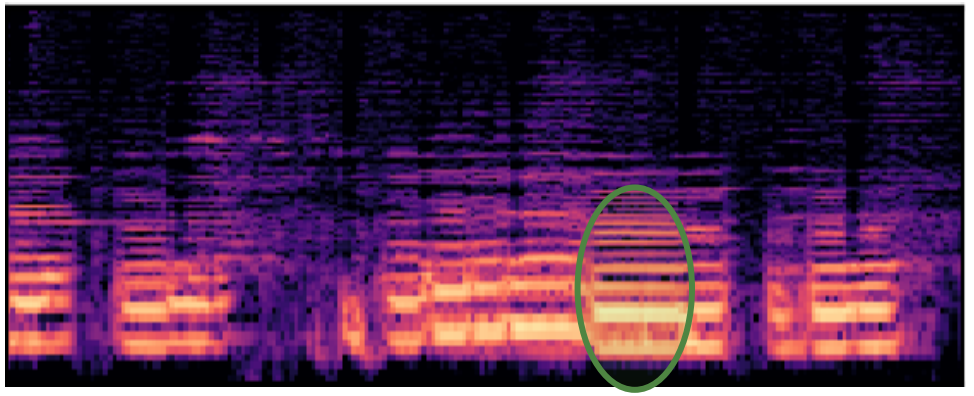
Source Separation is an Ill-posed Problem

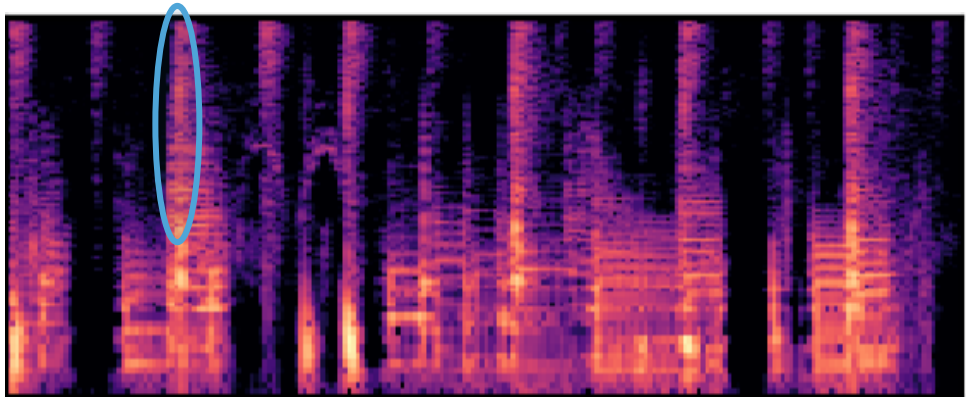
- There are **more than one solution** to $x = s_1 + s_2 + \cdots + s_N$
  - In fact, there are infinite possibilities
- However, we do know **what's more likely than another**!



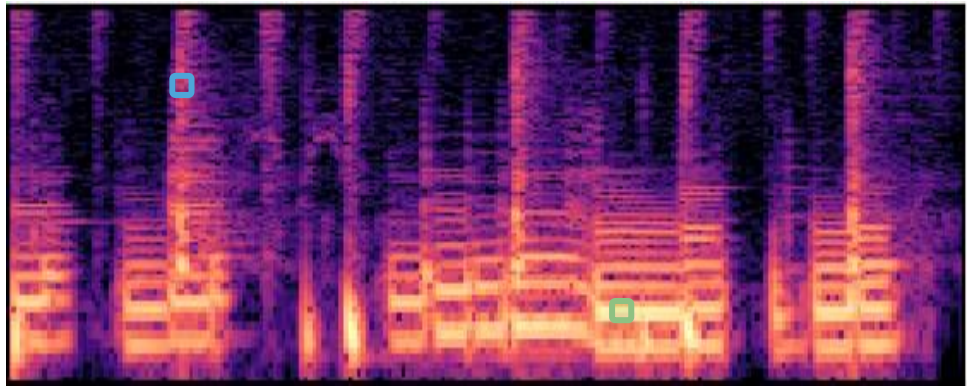
Time-Frequency Masking

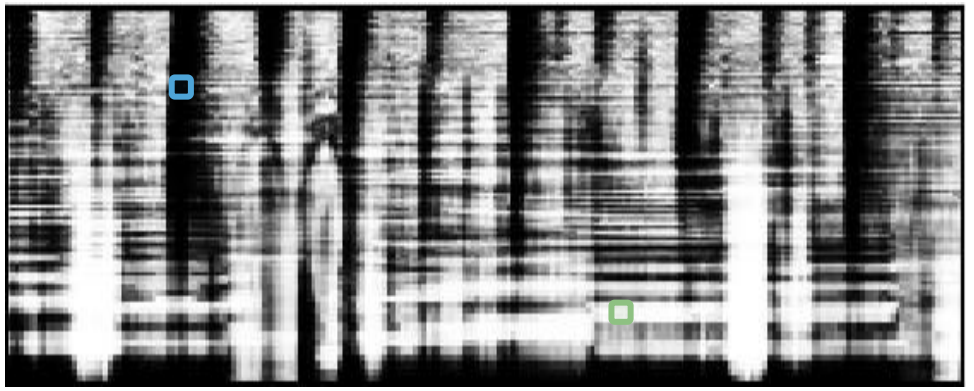


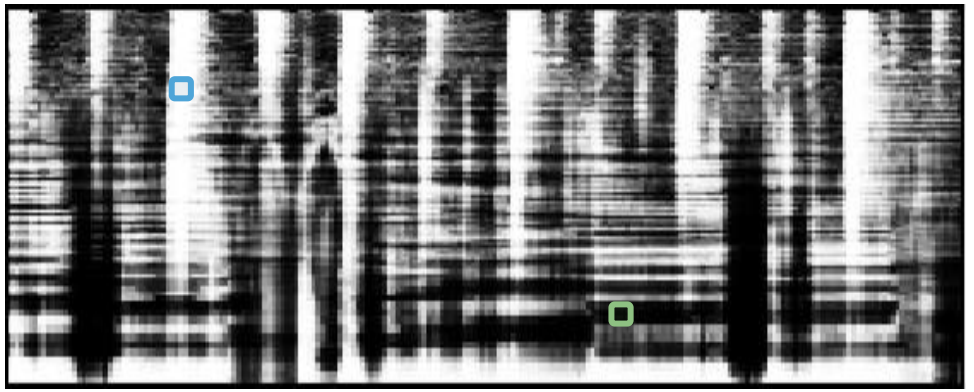




Time-Frequency Masking

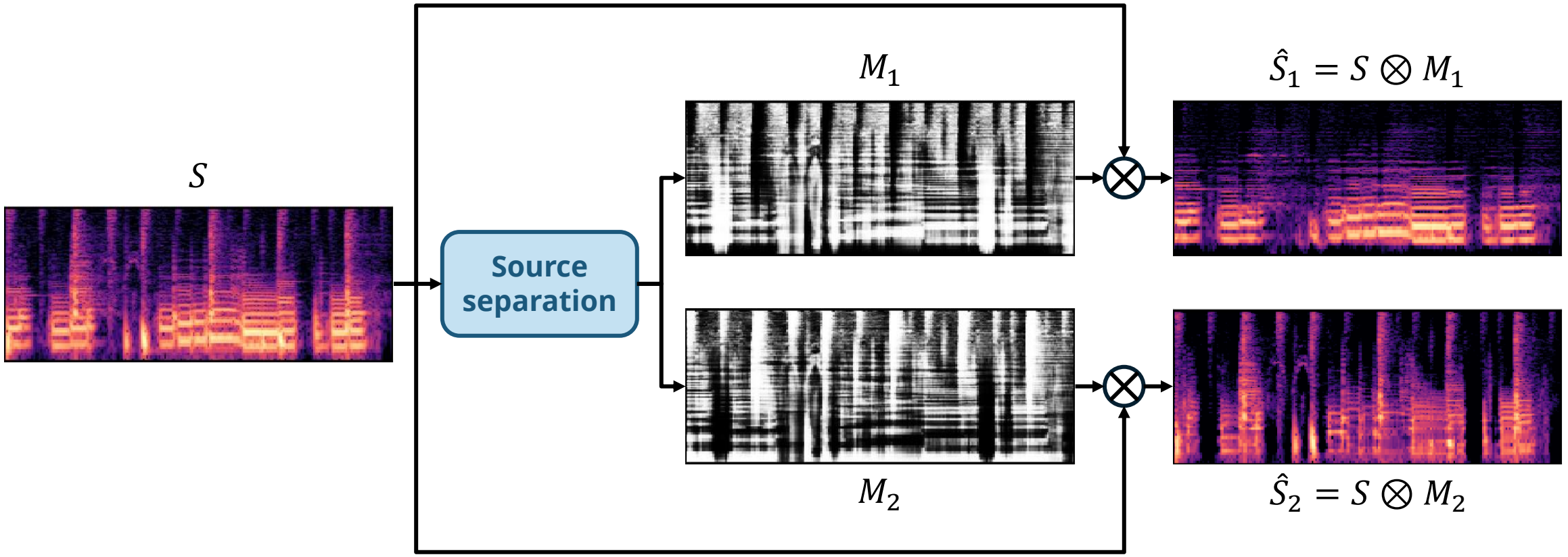




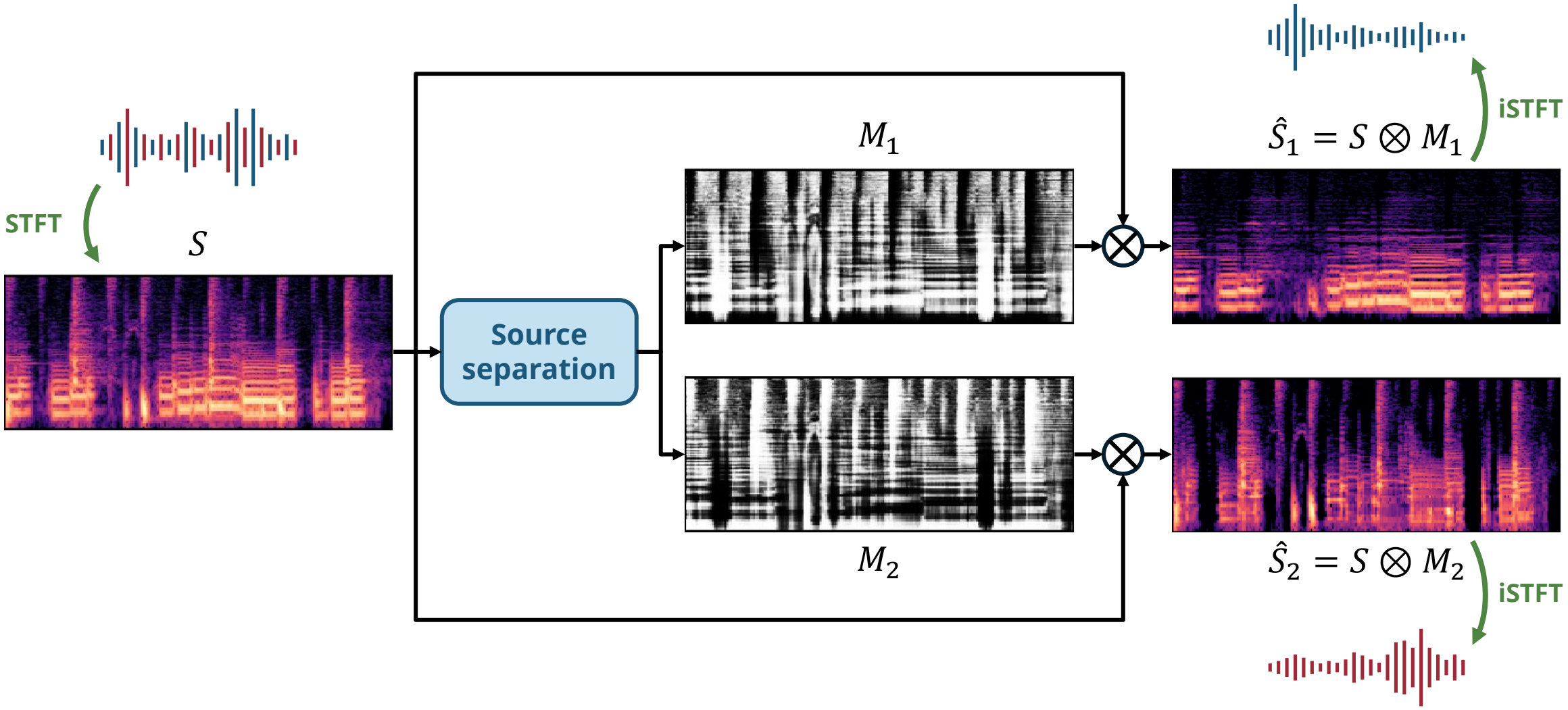


Classification on each time-frequency bin

Time-Frequency Masking



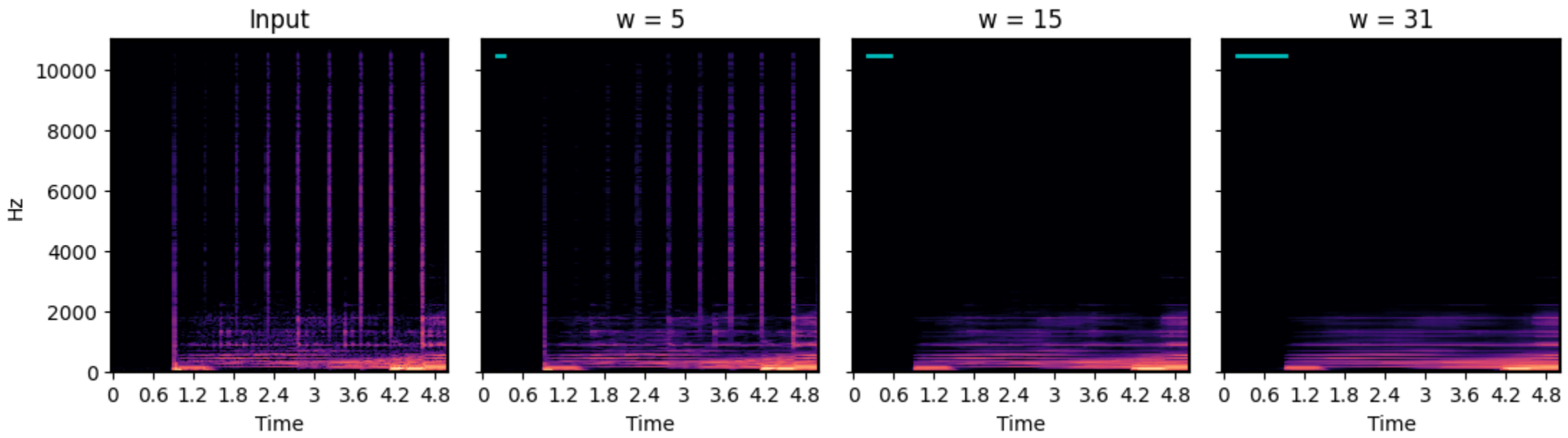
Time-Frequency Masking



Traditional Approaches

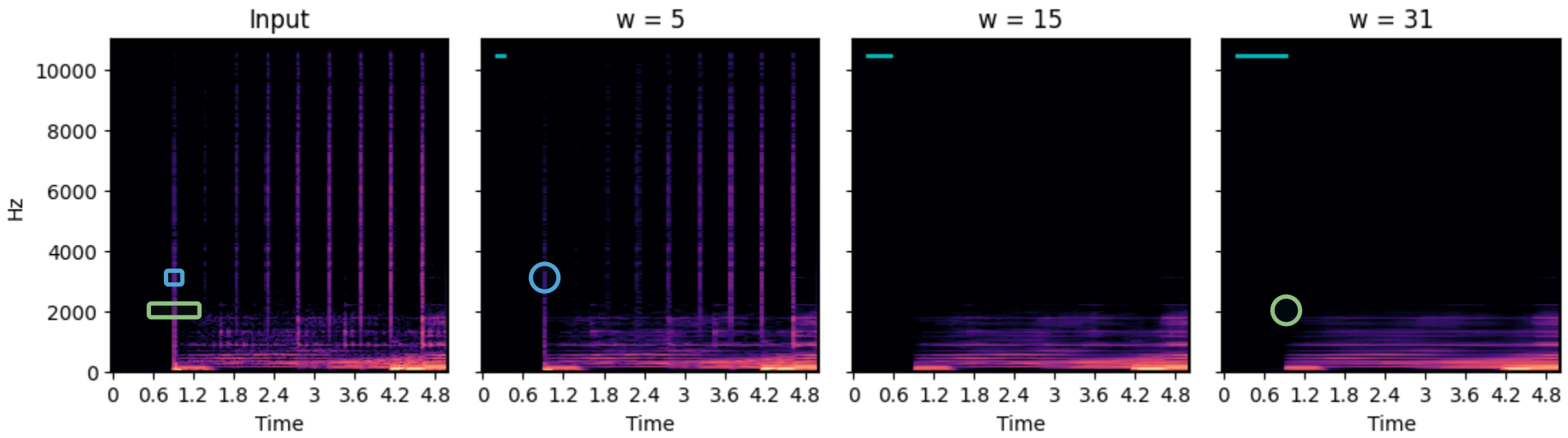
Percussive vs Harmonic Components

Harmonic-Percussive Separation (Fitzgerald et al., 2010)



Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

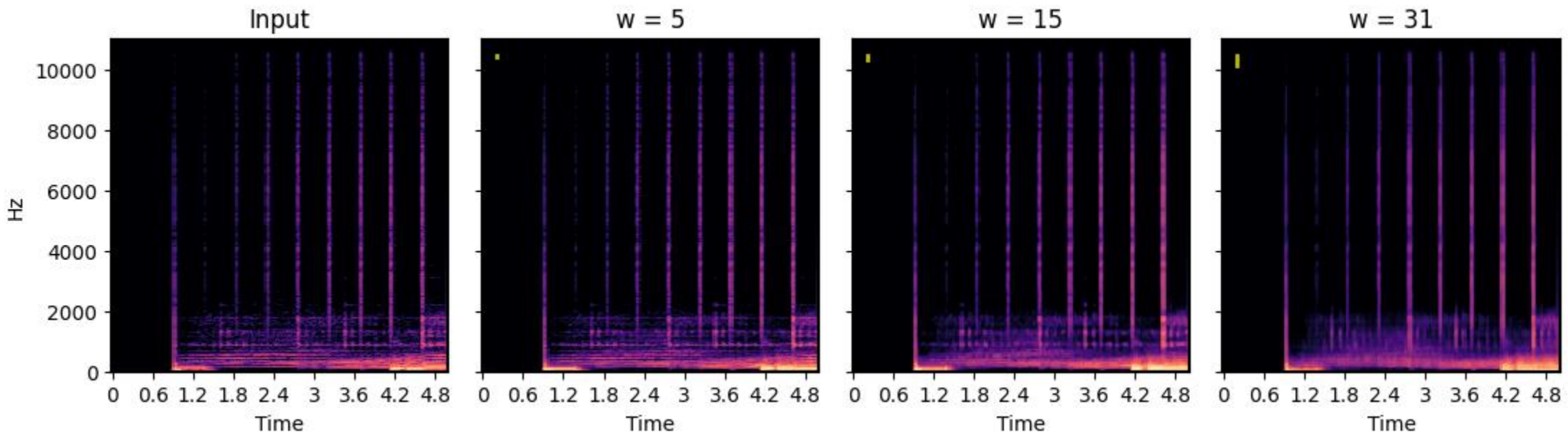
Harmonic-Percussive Separation (Fitzgerald et al., 2010)



Applying a median filter over the time axis makes percussive patterns less prominent!

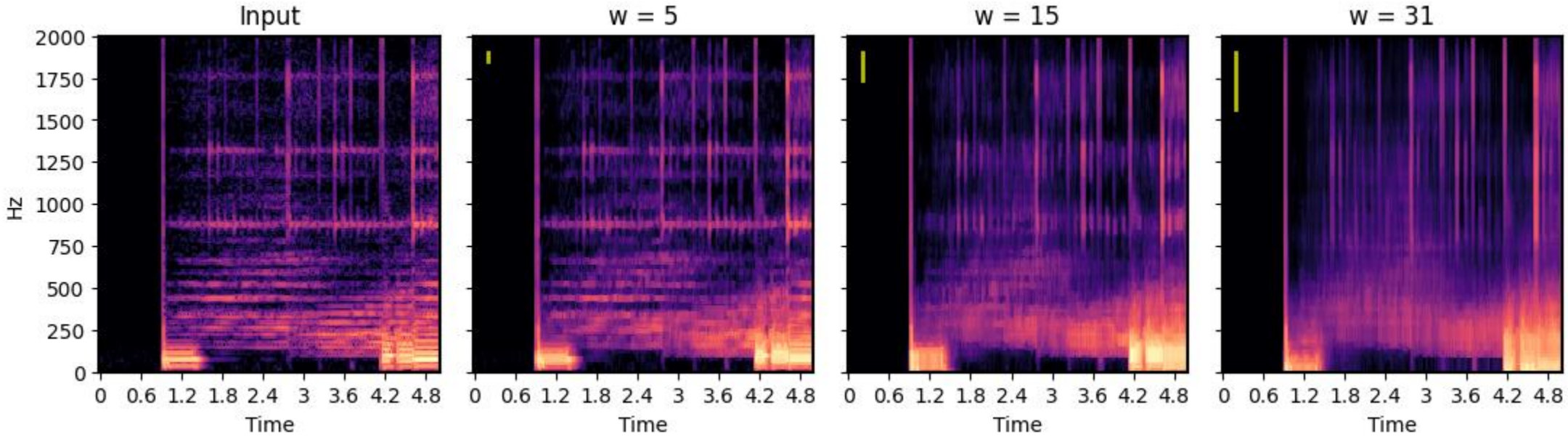
Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

Harmonic-Percussive Separation (Fitzgerald et al., 2010)



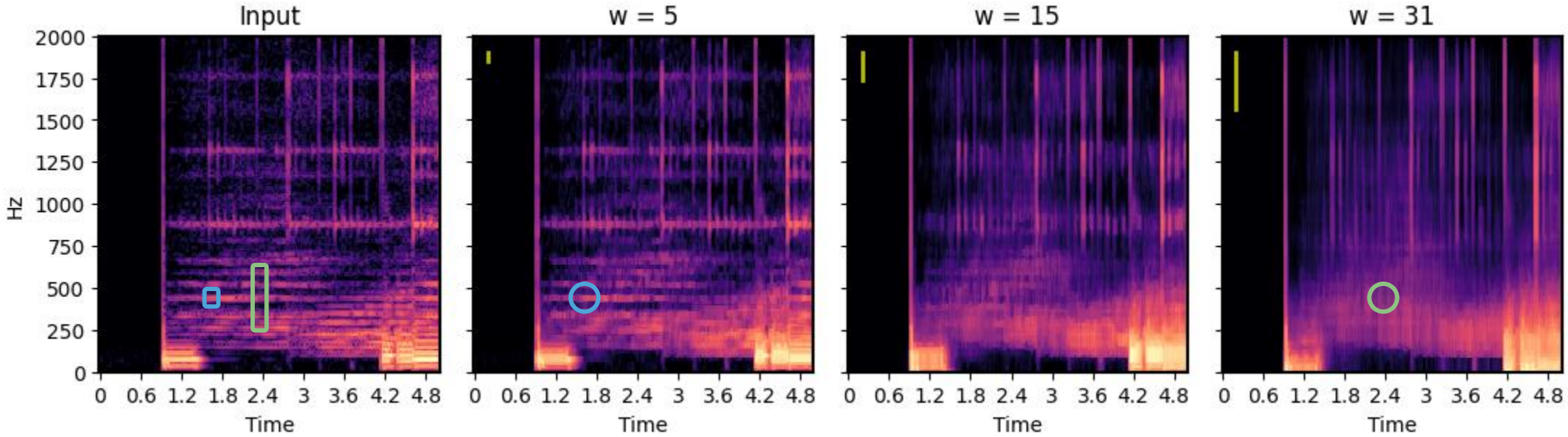
Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

Harmonic-Percussive Separation (Fitzgerald et al., 2010)



Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

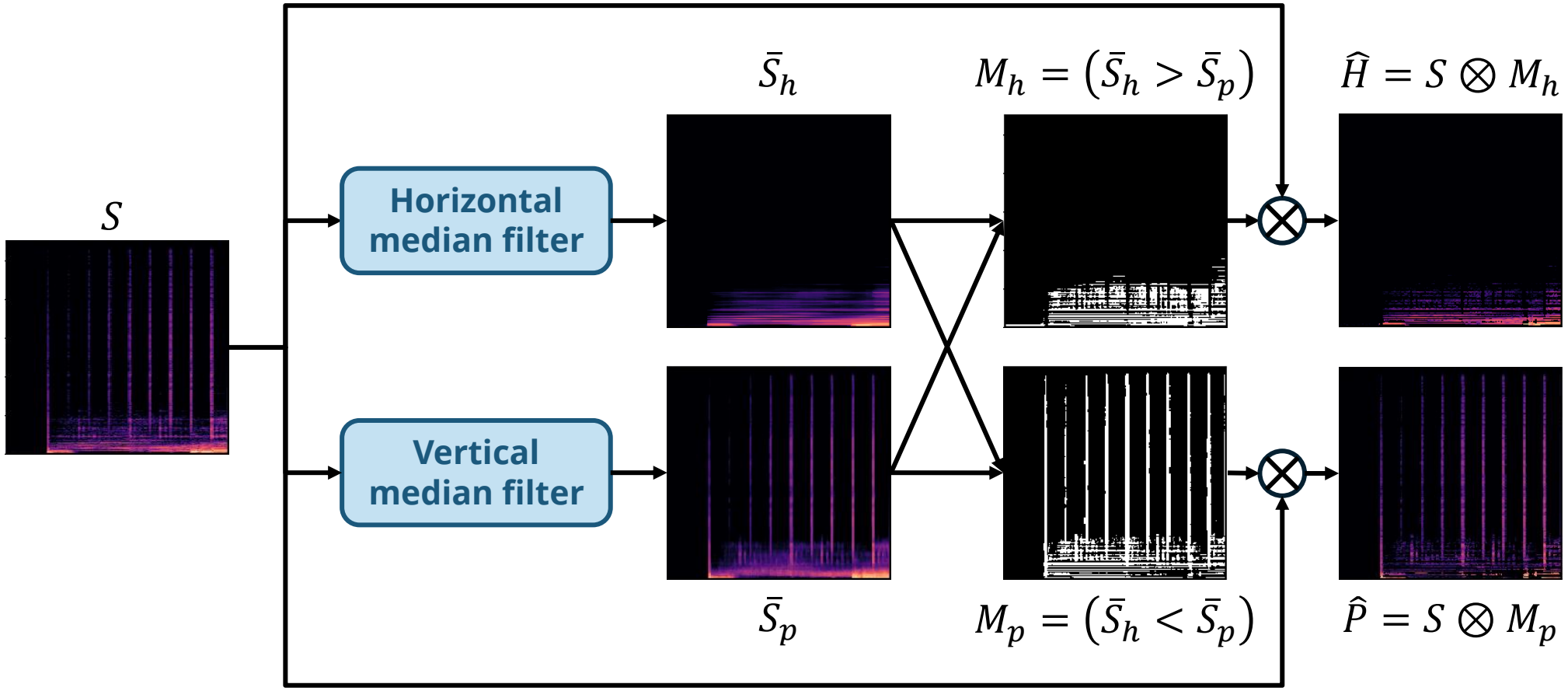
Harmonic-Percussive Separation (Fitzgerald et al., 2010)



Applying a median filter over the frequency axis makes harmonic patterns less prominent!

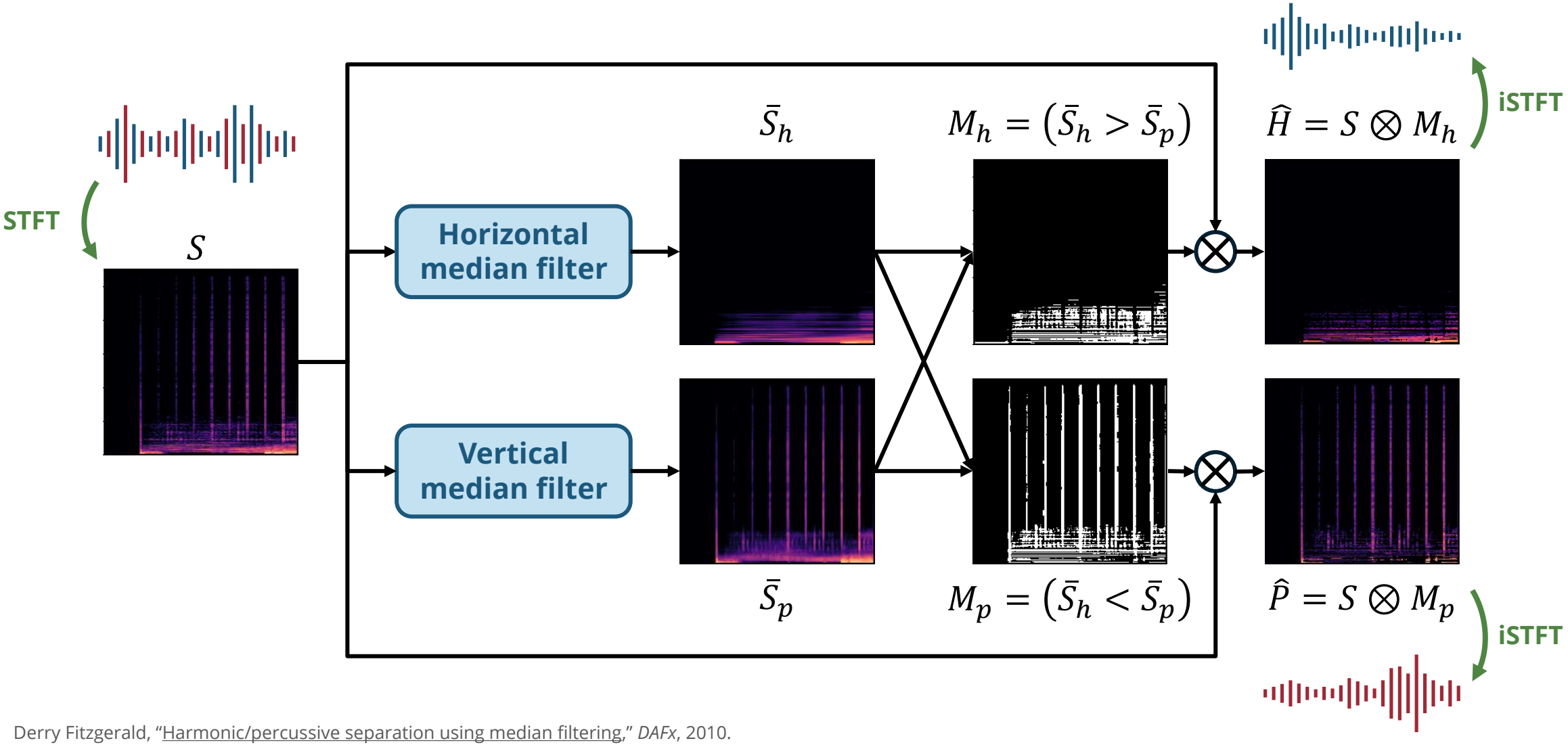
Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

Harmonic-Percussive Separation (Fitzgerald et al., 2010)



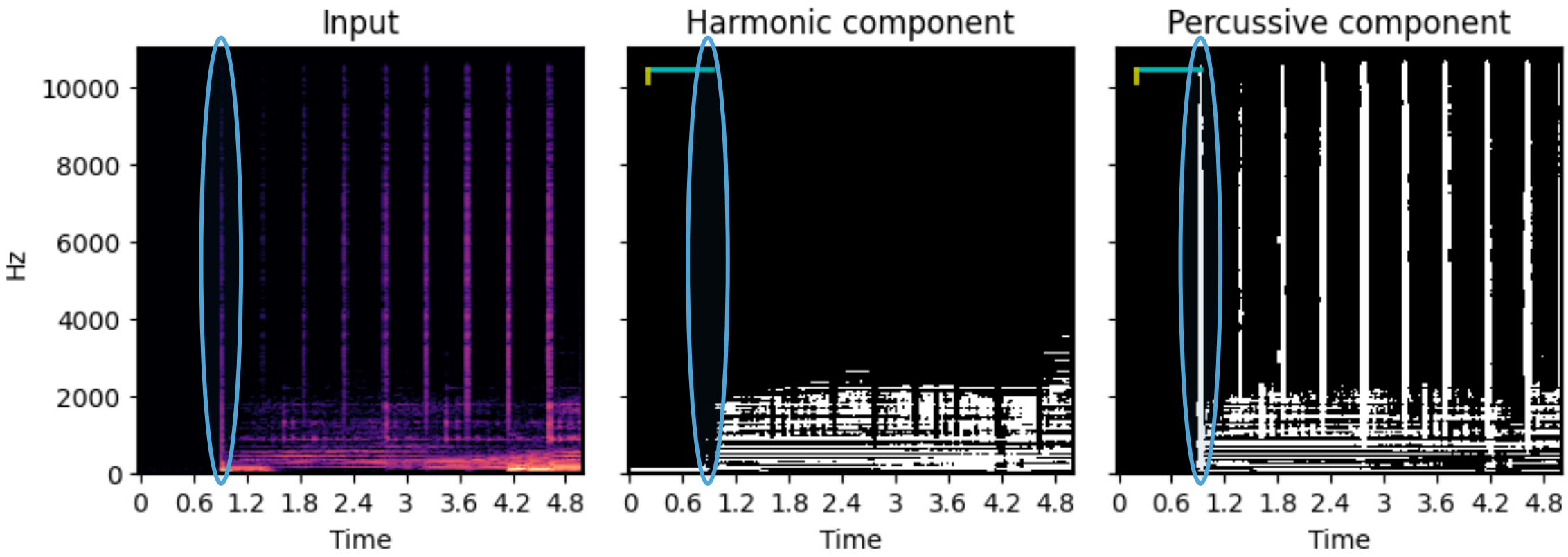
Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

Harmonic-Percussive Separation (Fitzgerald et al., 2010)



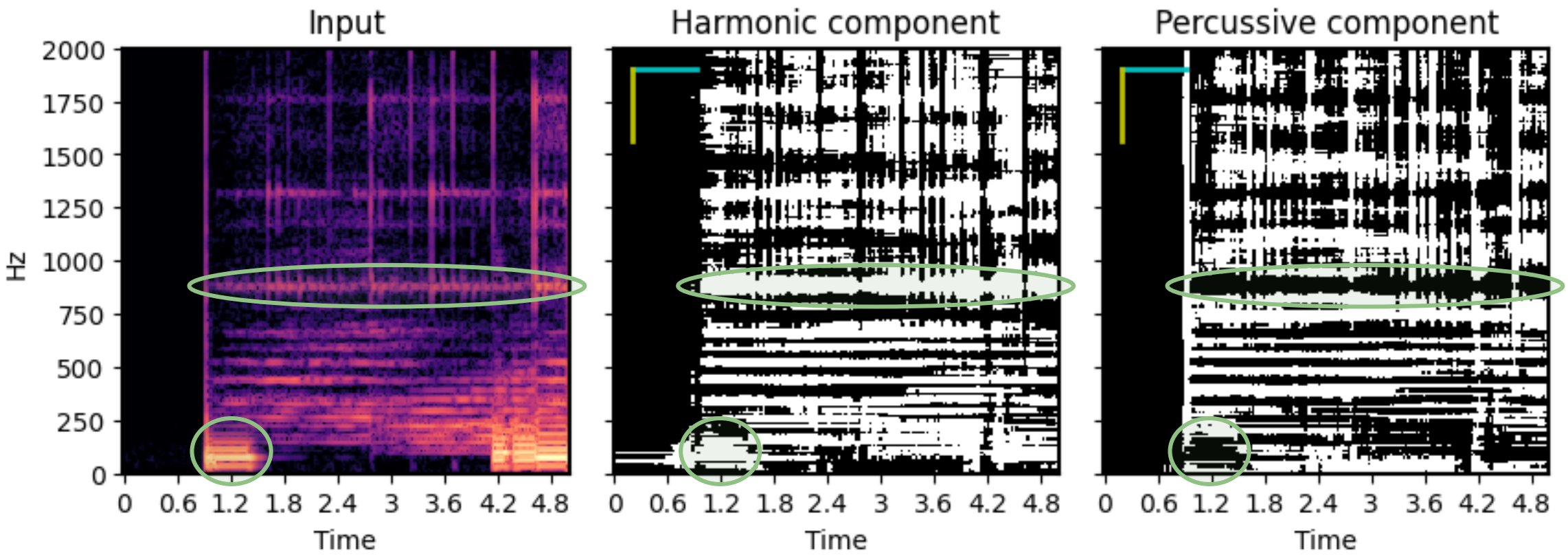
Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

Harmonic-Percussive Separation (Fitzgerald et al., 2010)



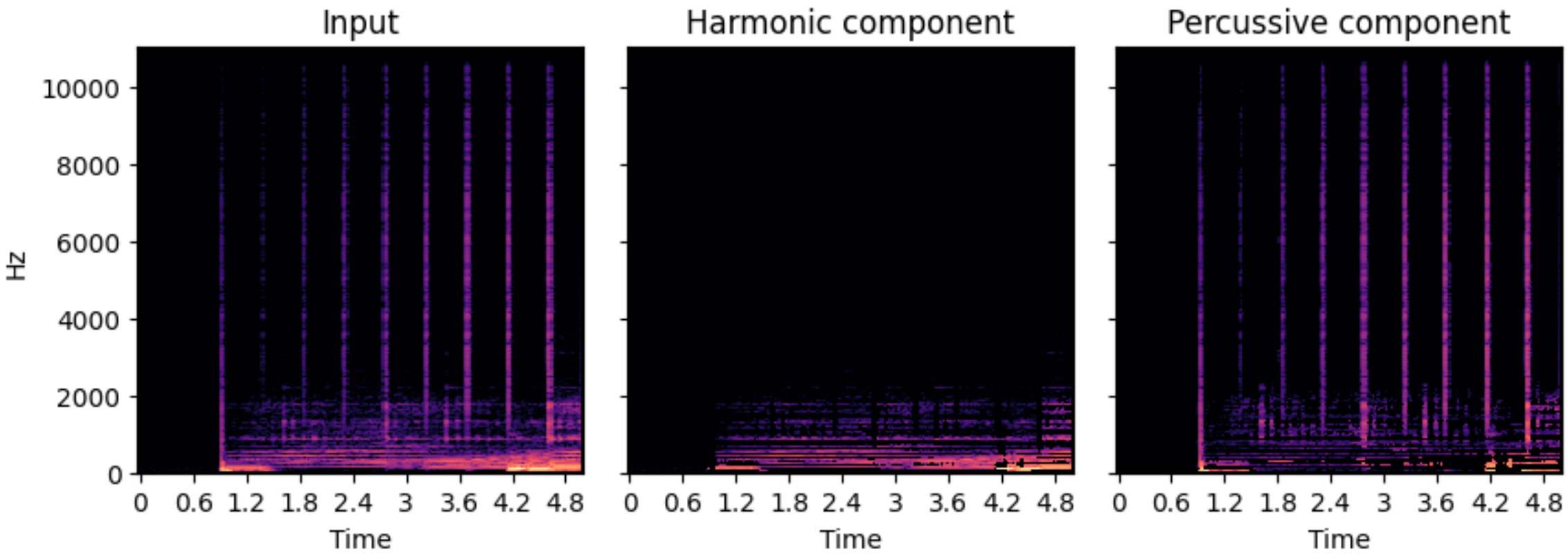
Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

Harmonic-Percussive Separation (Fitzgerald et al., 2010)



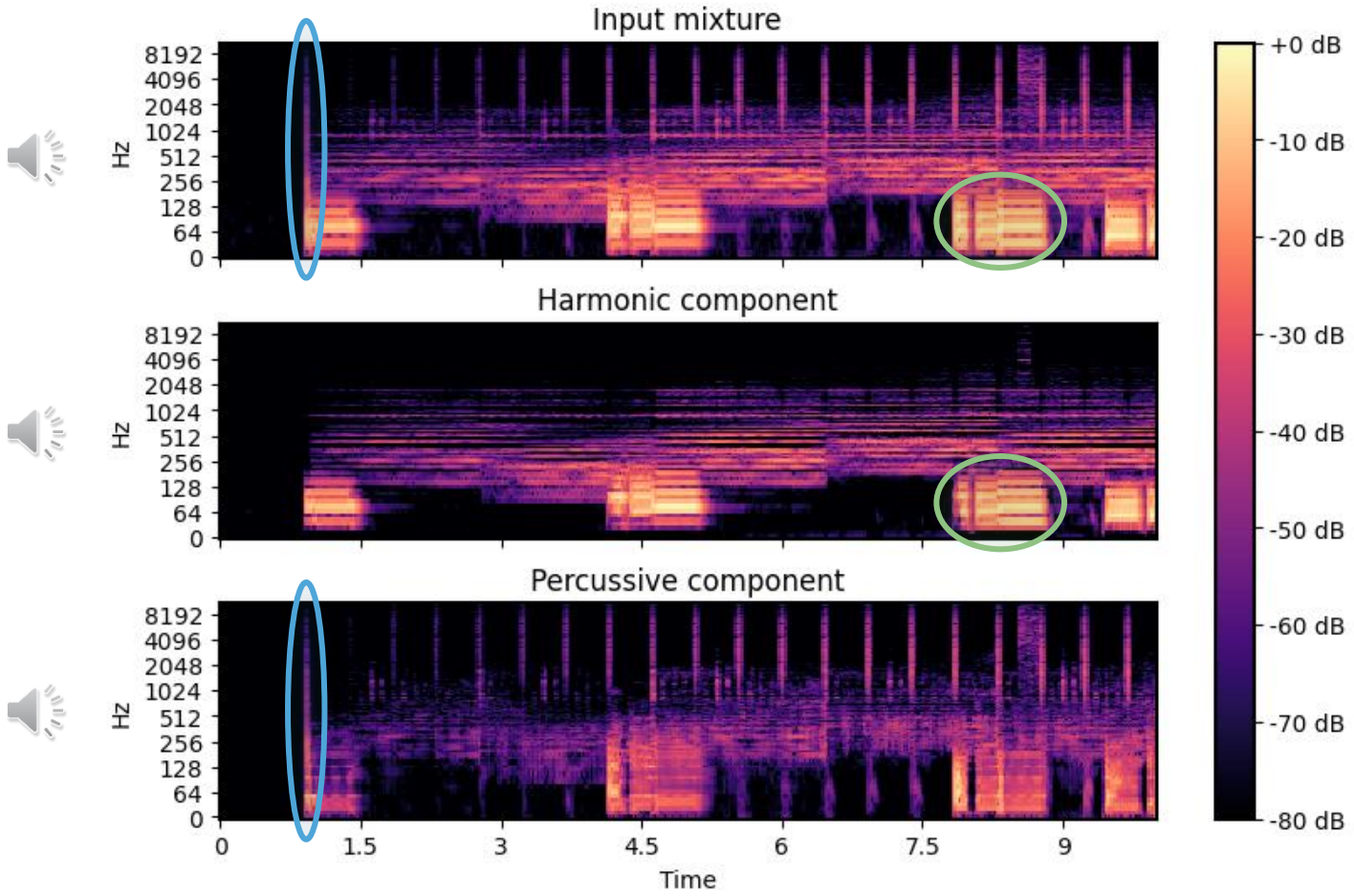
Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

Harmonic-Percussive Separation (Fitzgerald et al., 2010)



Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

HPSS: Example Result



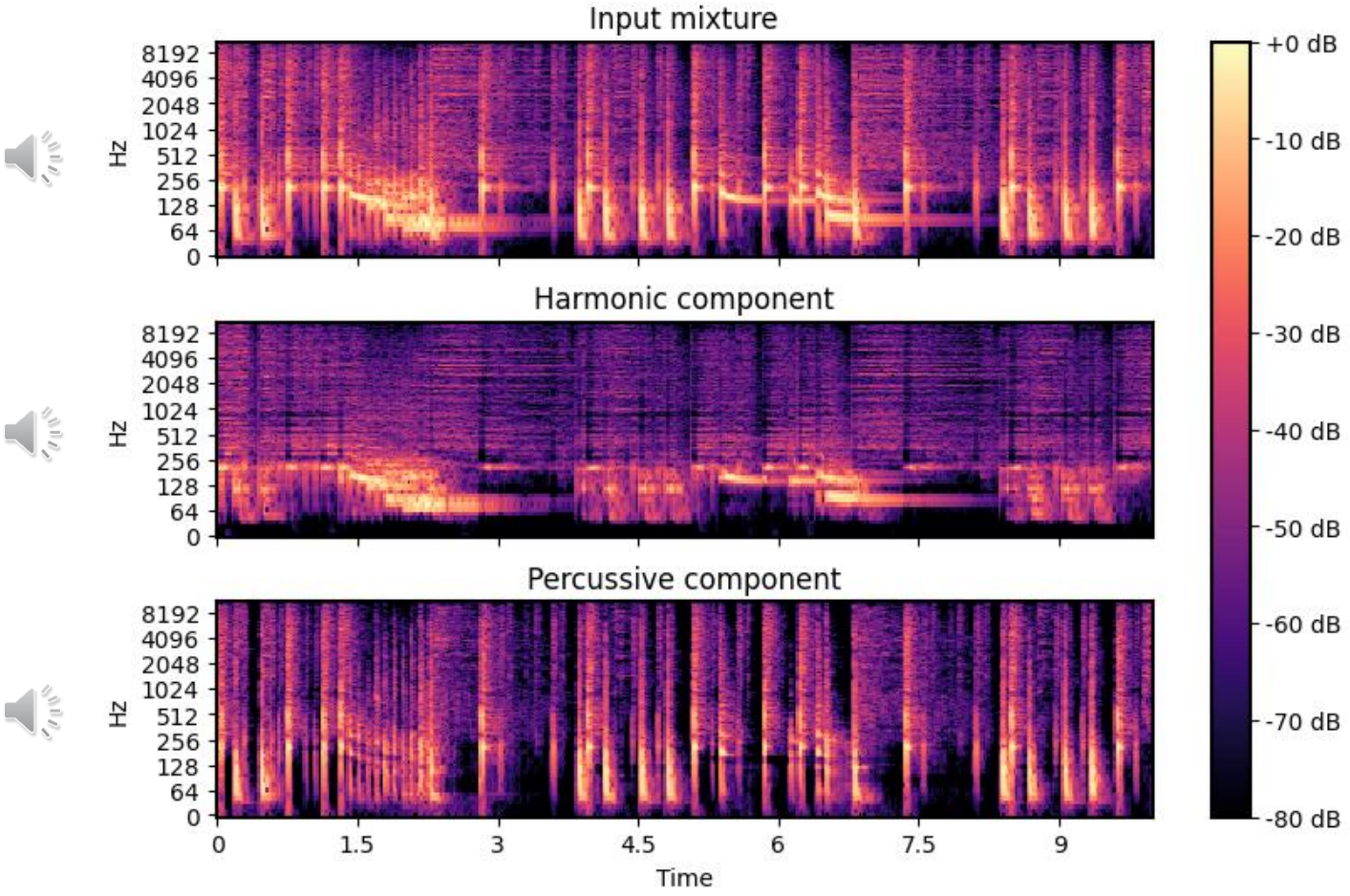
(Source: "Like Before" by Bessonn&sa)

Bessonn&sa, CC BY-NC-SA, via Jamendo

Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.

Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

HPSS: Example Result

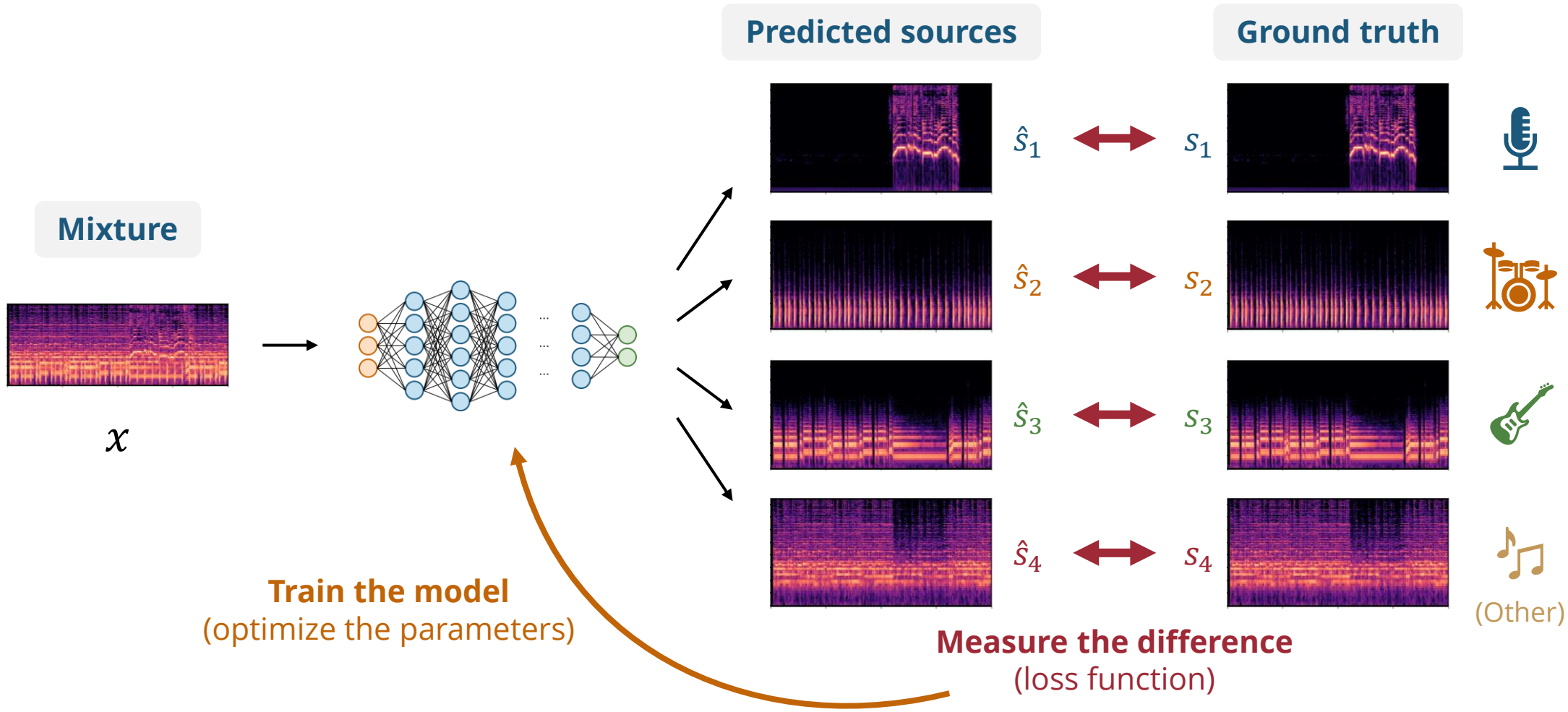


(Source: "Only Drums Jazzy Hip-Hop" by Oleg Silukov)

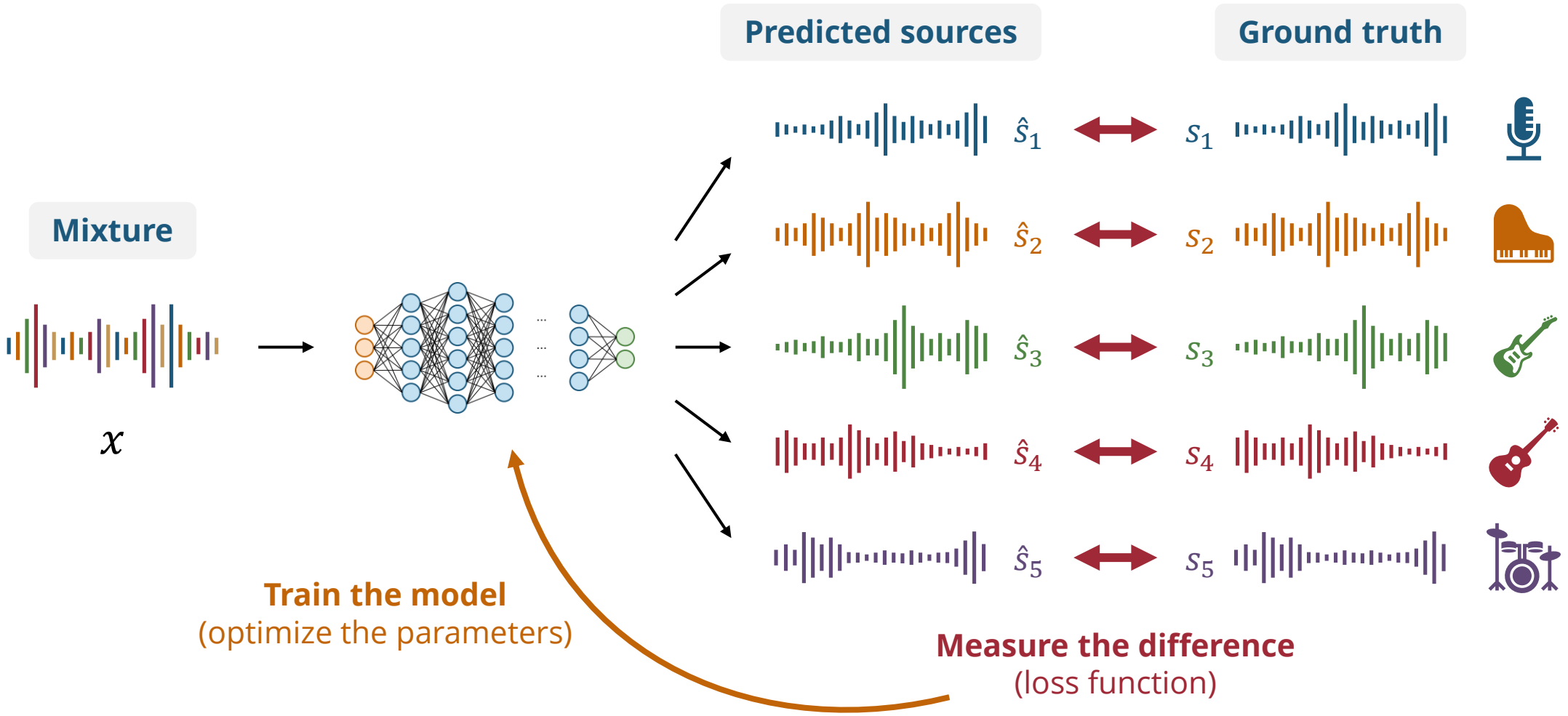
Derry Fitzgerald, "Harmonic/percussive separation using median filtering," *DAFx*, 2010.
Jonathan Driedger, Meinard Müller, Sascha Disch, "Extending Harmonic-Percussive Separation of Audio Signals," *ISMIR*, 2014.

Deep Learning Approaches

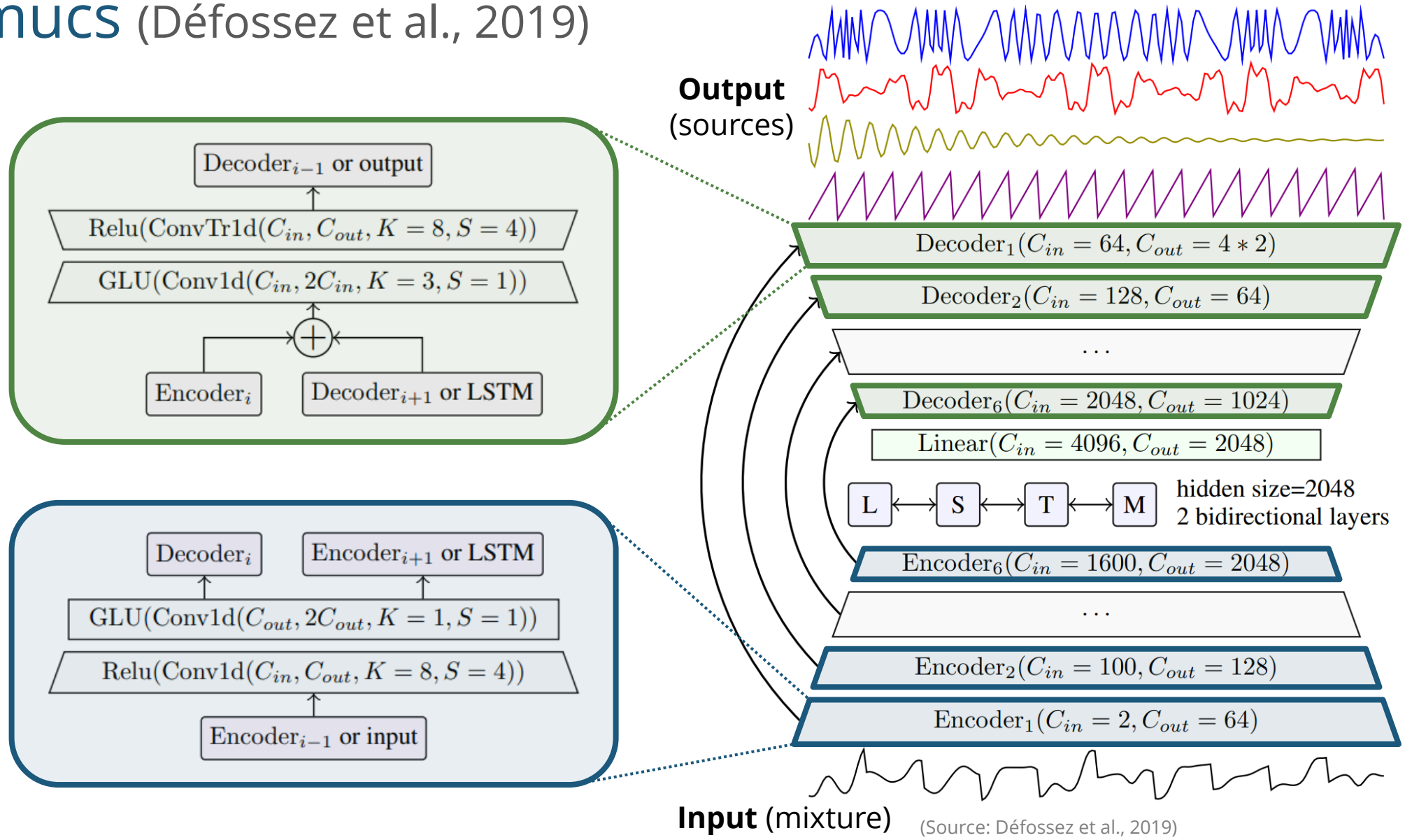
Deep Learning Based Source Separation



Deep Learning Based Source Separation



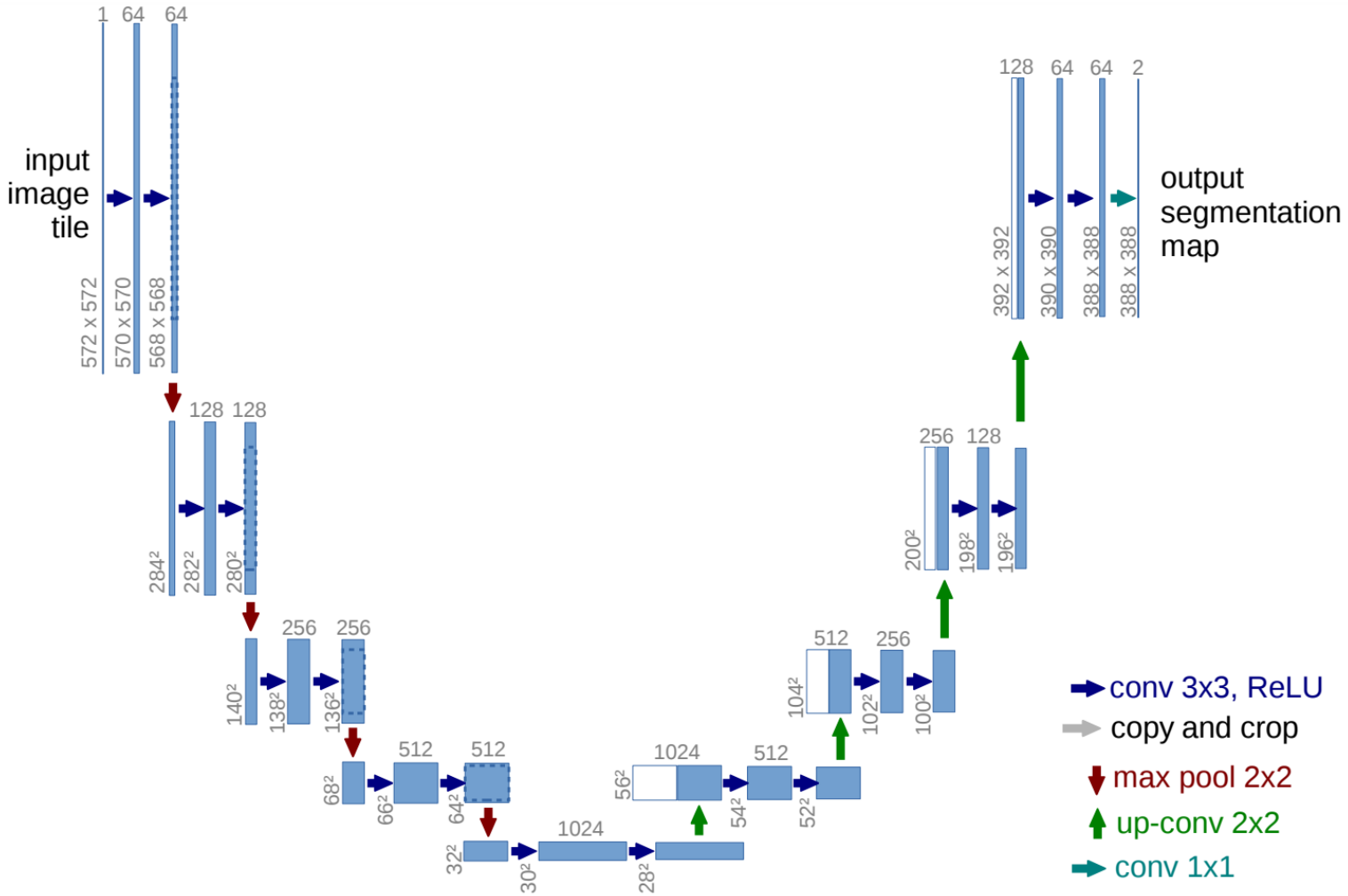
Demucs (Défossez et al., 2019)



(Source: Défossez et al., 2019)

Alexandre Défossez, Nicolas Usunier, Léon Bottou, and Francis Bach, "Music Source Separation in the Waveform Domain," *arXiv preprint arXiv:1911.13254*, 2019.

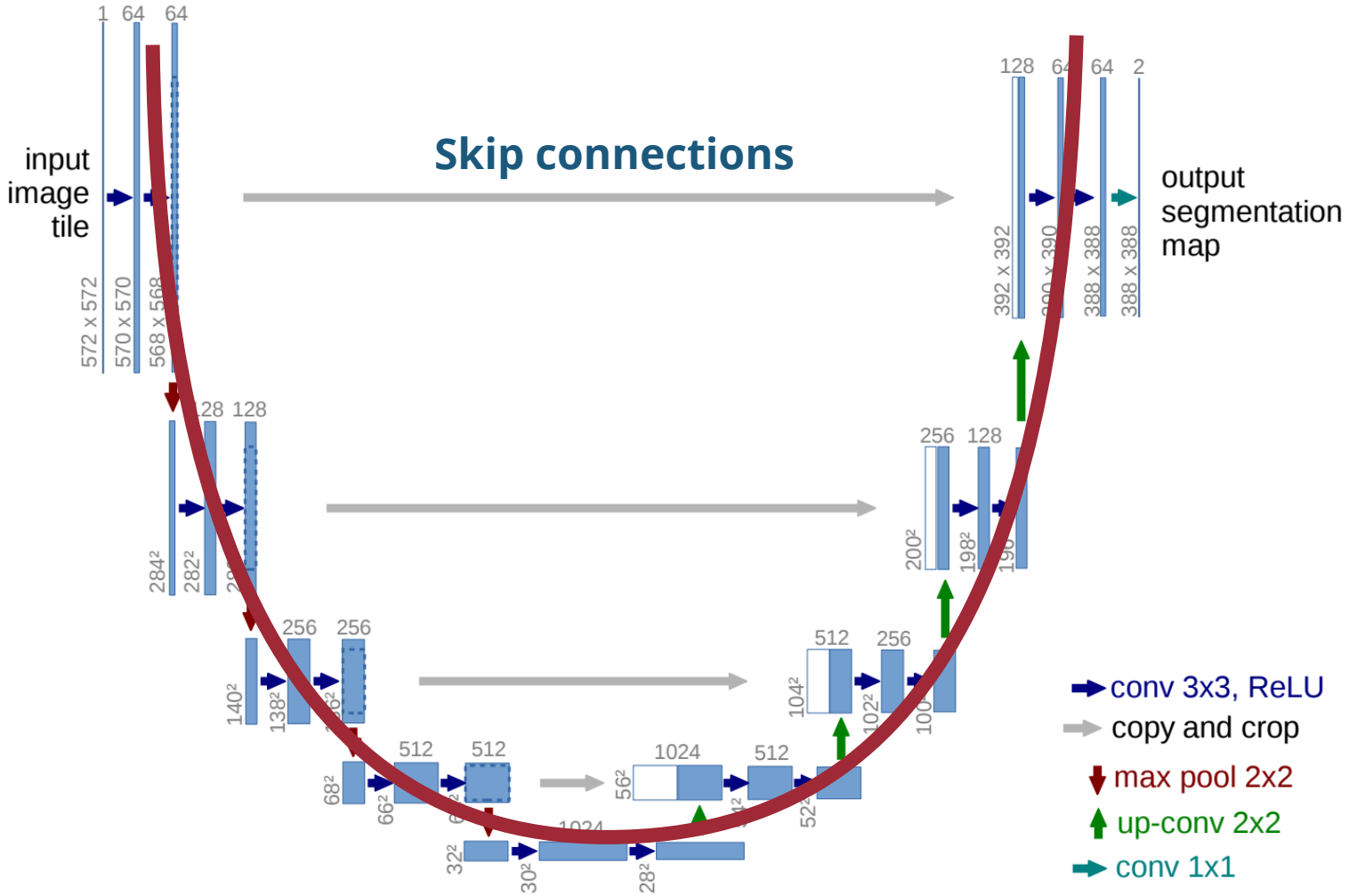
U-Net (Ronneberger et al., 2015)



(Source: Ronneberger et al., 2015)

Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *MICCAI*, 2015.

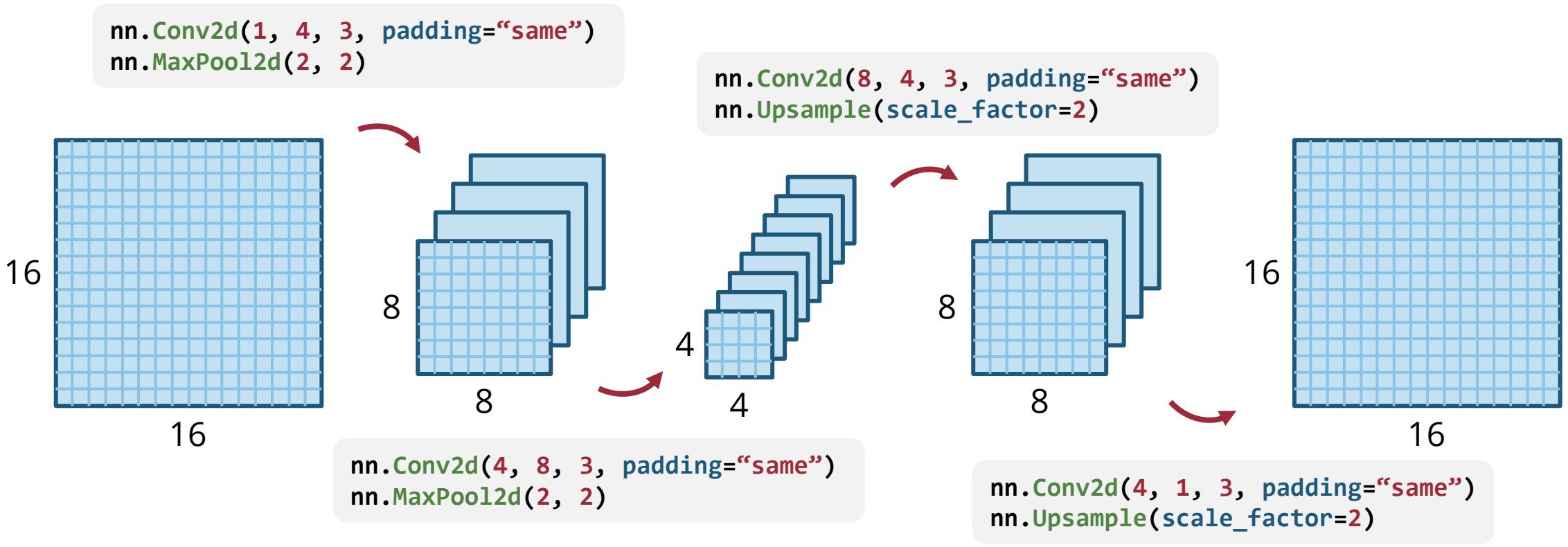
U-Net (Ronneberger et al., 2015)



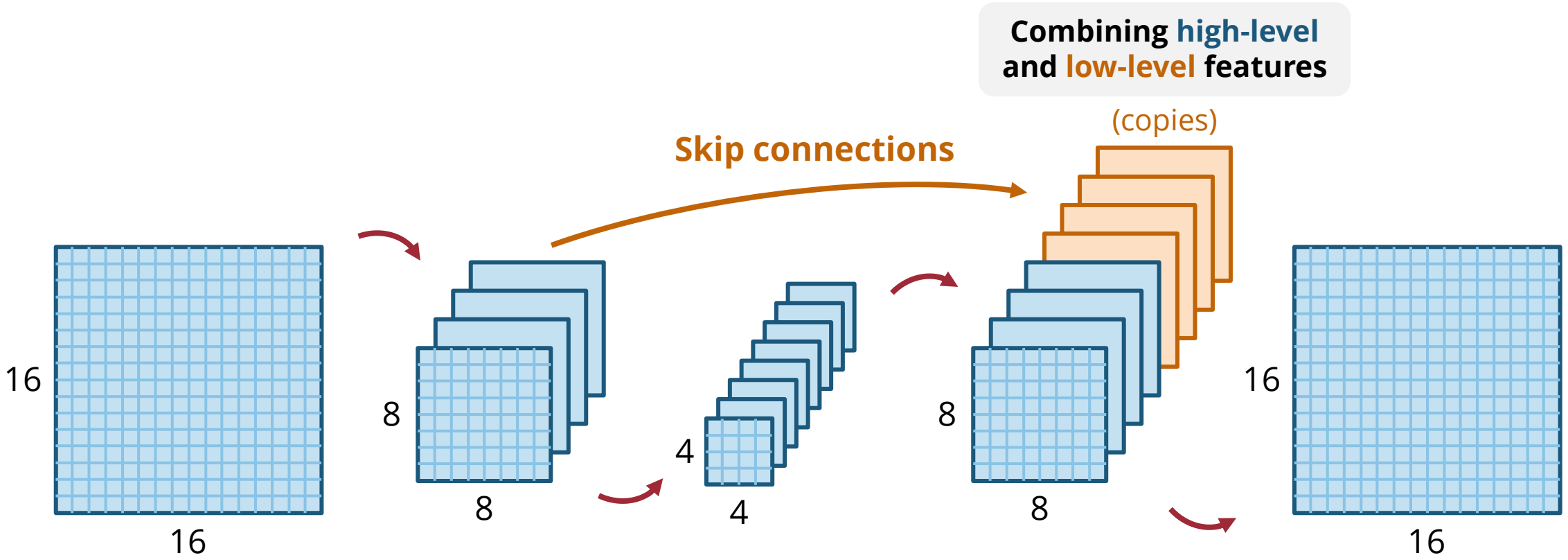
(Source: Ronneberger et al., 2015)

Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *MICCAI*, 2015.

A Toy Example of U-Net



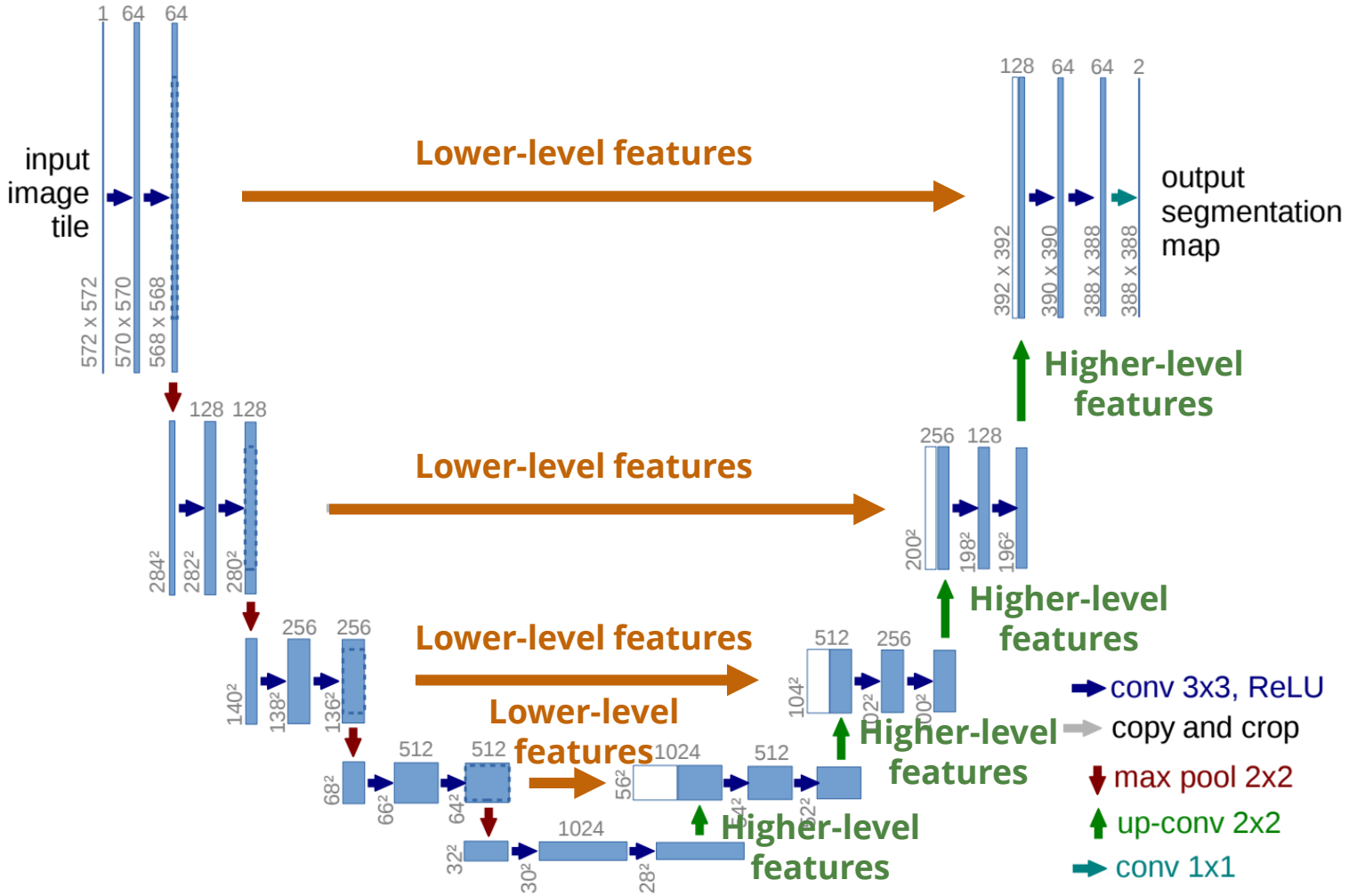
A Toy Example of U-Net




```
nn.Conv2d(8, 1, 3, padding="same")
nn.Upsample(scale_factor=2)
```

U-Nets are useful when the inputs and outputs have the same shape!

U-Net (Ronneberger et al., 2015)

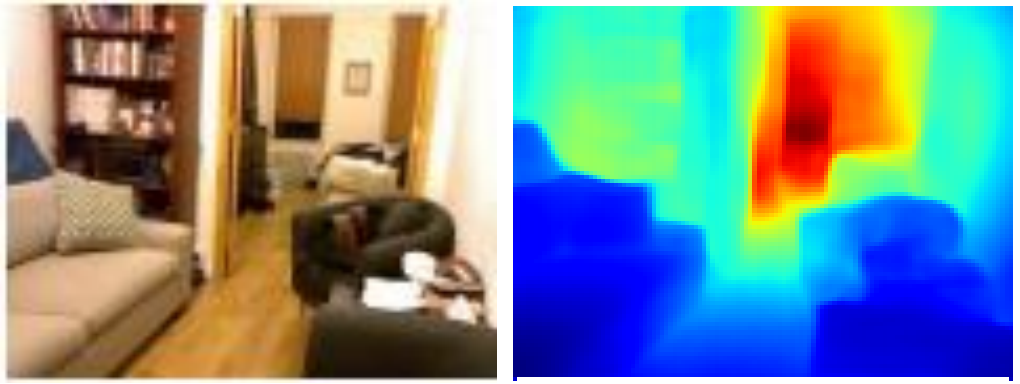


(Source: Ronneberger et al., 2015)

Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *MICCAI*, 2015.

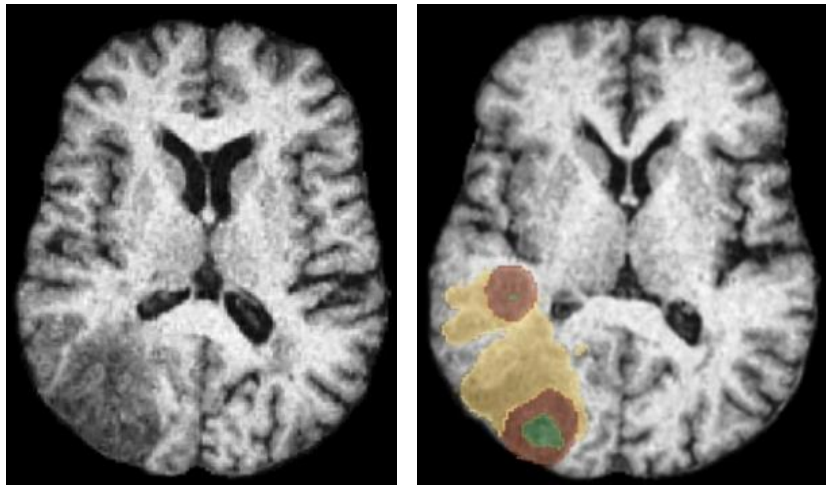
Applications of U-Nets

Depth Estimation



(Source: Barakat, 2018)

Tumor Segmentation



(Source: Kharaji et al., 2024)

Image Segmentation

(Source: Kirillov et al., 2023)

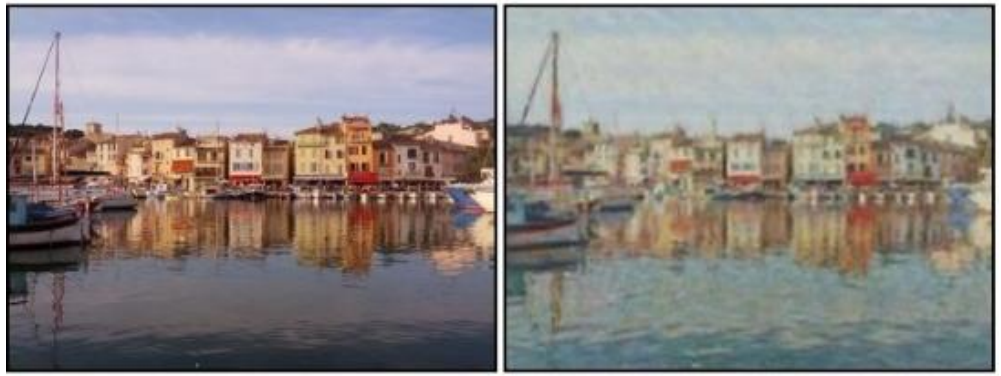
Omar Barakat, "Depth estimation with deep Neural networks part 1," *Medium*, January 11, 2018

Mona Kharaji, Hossein Abbasi, Yasin Orouskhani, Mostafa Shomalzadeh, Foad Kazemi, and Maysam Orouskhani, "nnU-Net for Brain Tumor Segmentation," *Neuroscience Informatics*, 2024.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick, "Segment Anything," *ICCV,* 2023.

Applications of U-Nets

Style Transfer

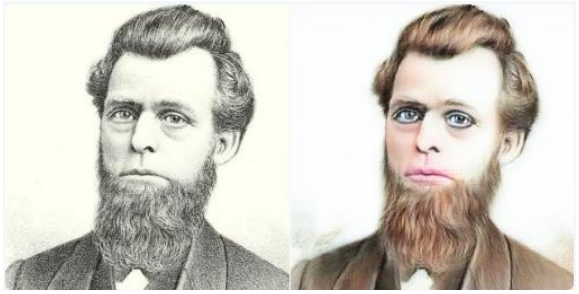


(Source: Zhu et al., 2018)

Sim2Real

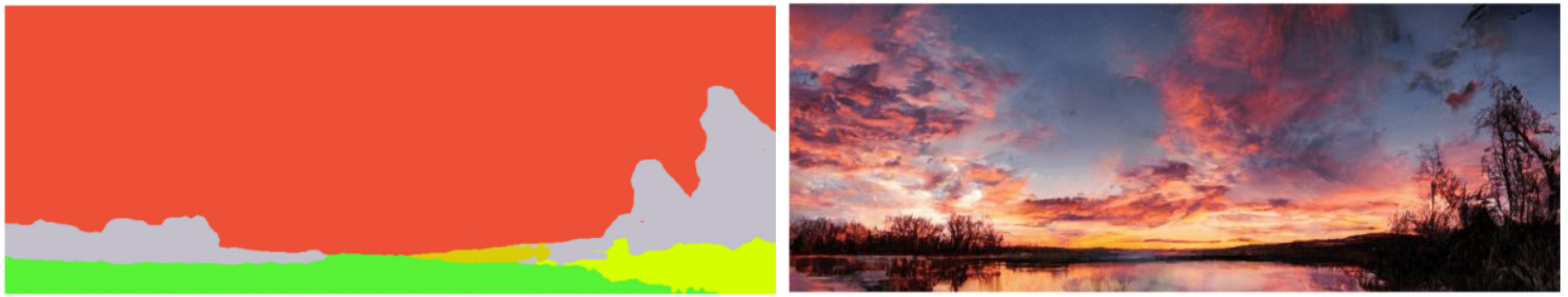
(Source: Zhu et al., 2018)

Colorization



(Source: Zhu et al., 2018)

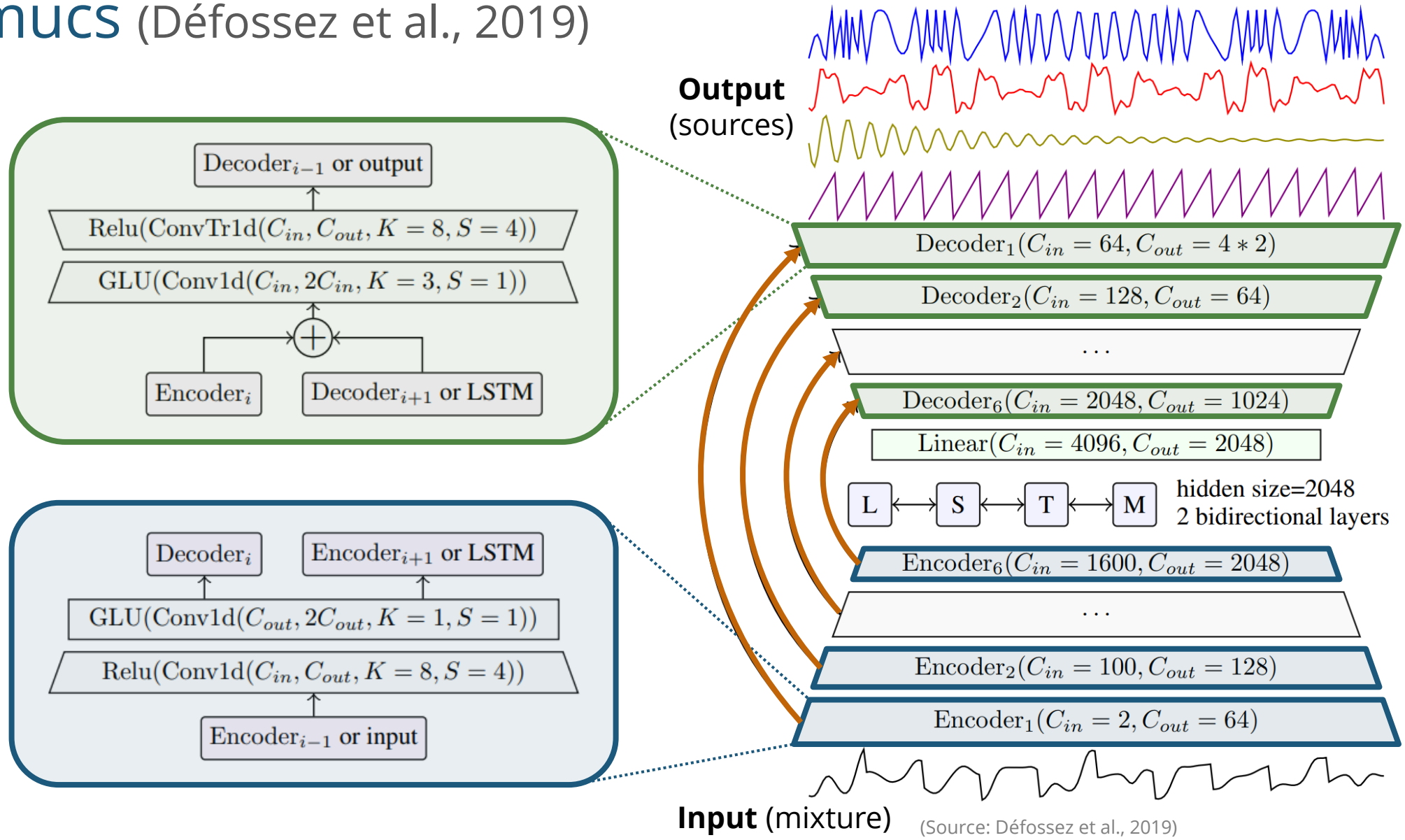
Semantic Synthesis



(Source: Rombach et al., 2022)

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," *CVPR*, 2022.
Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," *ICCV*, 2017.

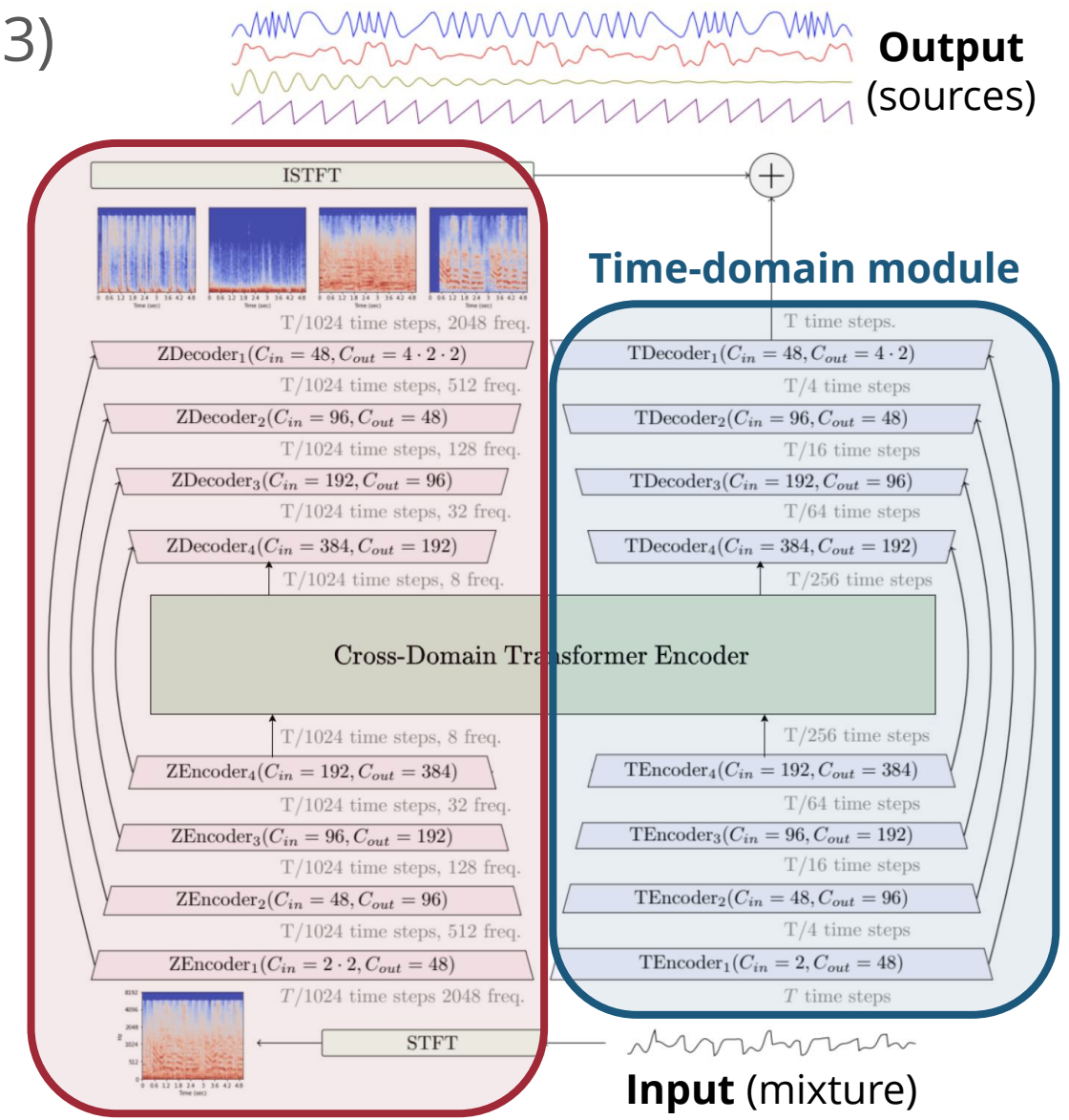
Demucs (Défossez et al., 2019)



(Source: Défossez et al., 2019)

Alexandre Défossez, Nicolas Usunier, Léon Bottou, and Francis Bach, "Music Source Separation in the Waveform Domain," *arXiv preprint arXiv:1911.13254*, 2019.

Demucs-Hybrid (Rouard et al., 2023)



(Source: Rouard et al., 2023)

Simon Rouard, Francisco Massa, and Alexandre Défossez, "Hybrid Transformers for Music Source Separation," *ICASSP*, 2023.

Datasets

- MIR-1K
- MedleyDB
- iKala
- DSD100
- MUSDB18 & MUSDB18-HQ
- MoisesDB
- Synthetic: Slakh2100, SynthSOD

Datasets

Dataset	Year	Tracks	Track duration (s)	Full/stereo?
MASS	2008	9	16 ± 7)	✗ / ✓
MIR-1K	2010	1,000	(8 ± 8)	✗ / ✗
QUASI	2011	5	(206 ± 21)	✓ / ✓
ccMixter	2014	50	(231 ± 77)	✓ / ✓
MedleyDB	2014	63	(206 ± 121)	✓ / ✓
iKala	2015	206	30	✗ / ✗
sigsep DSD100	2015	100	(251 ± 60)	✓ / ✓
sigsep MUSDB18	2017	150	(236 ± 95)	✓ / ✓
sigsep MUSDB18-HQ	2019	150	(236 ± 95)	✓ / ✓

(Source: SigSep)

sigsep.github.io/datasets/

Choral Separation (Chen et al., 2022)

Demo

Mixture		Soprano	Alto	Tenor	Bass
	→				
	→				

Data Augmentation

SoundFont	Standard	Expressive (vowels only)	Expressive (words)

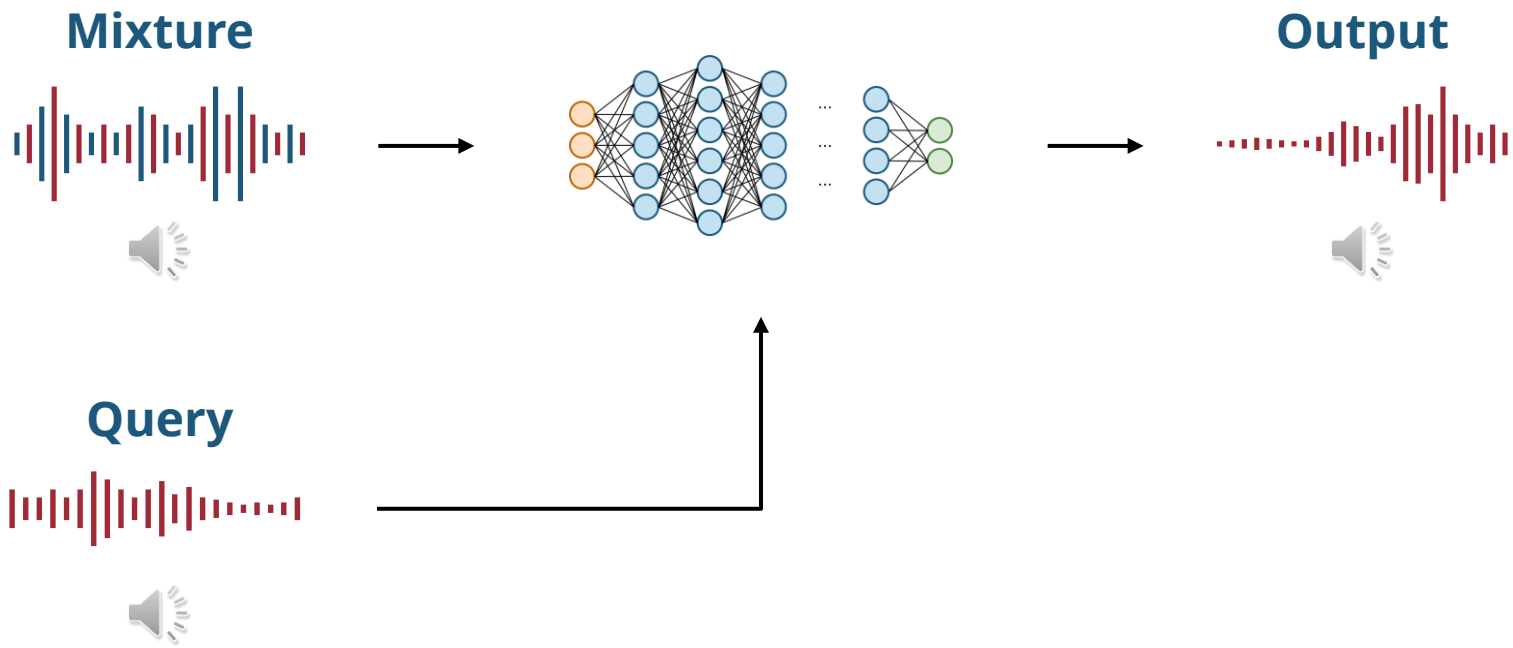
Ke Chen, Hao-Wen Dong, Yi Luo, Julian McAuley, Taylor Berg-Kirkpatrick, Miller Puckette, and Shlomo Dubnov, "Improving Choral Music Separation through Expressive Synthesized Data from Sampled Instruments," *ISMIR*, 2022.
retrocirce.github.io/cms_demo/

Blob Opera

- This is **NOT based on source separation**
- Sharing this simply **because it's cool**! 😎
- It's based on a **ML-based music harmonization model**!

artsandculture.google.com/experiment/blob-opera/AAHWrq360NcGbw

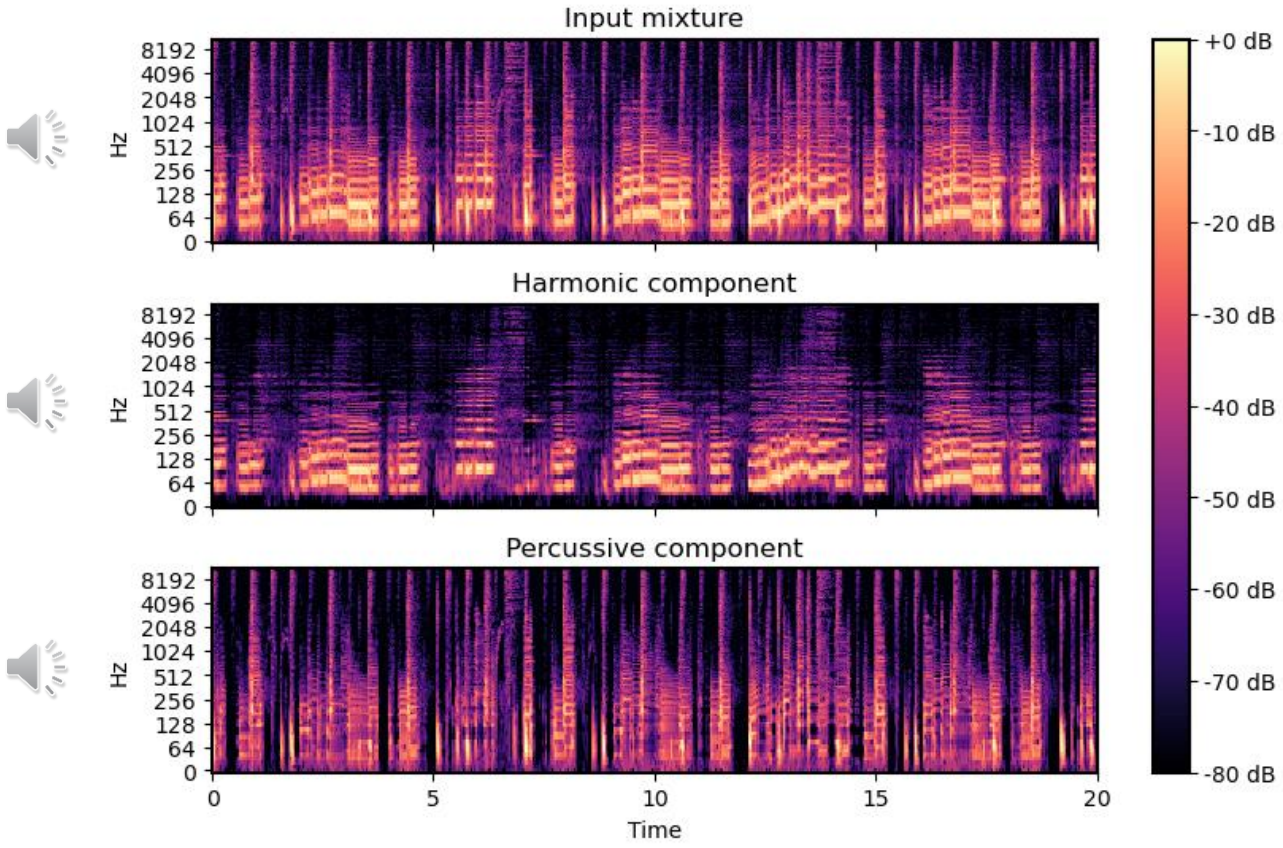
Beyond Known Sources: Query by Audio



Ke Chen, Xingjian Du, Bilei Zhu, Zejun Ma, Taylor Berg-Kirkpatrick, and Shlomo Dubnov, "Zero-shot Audio Source Separation through Query-based Learning from Weakly-labeled Data," *AAAI*, 2022.

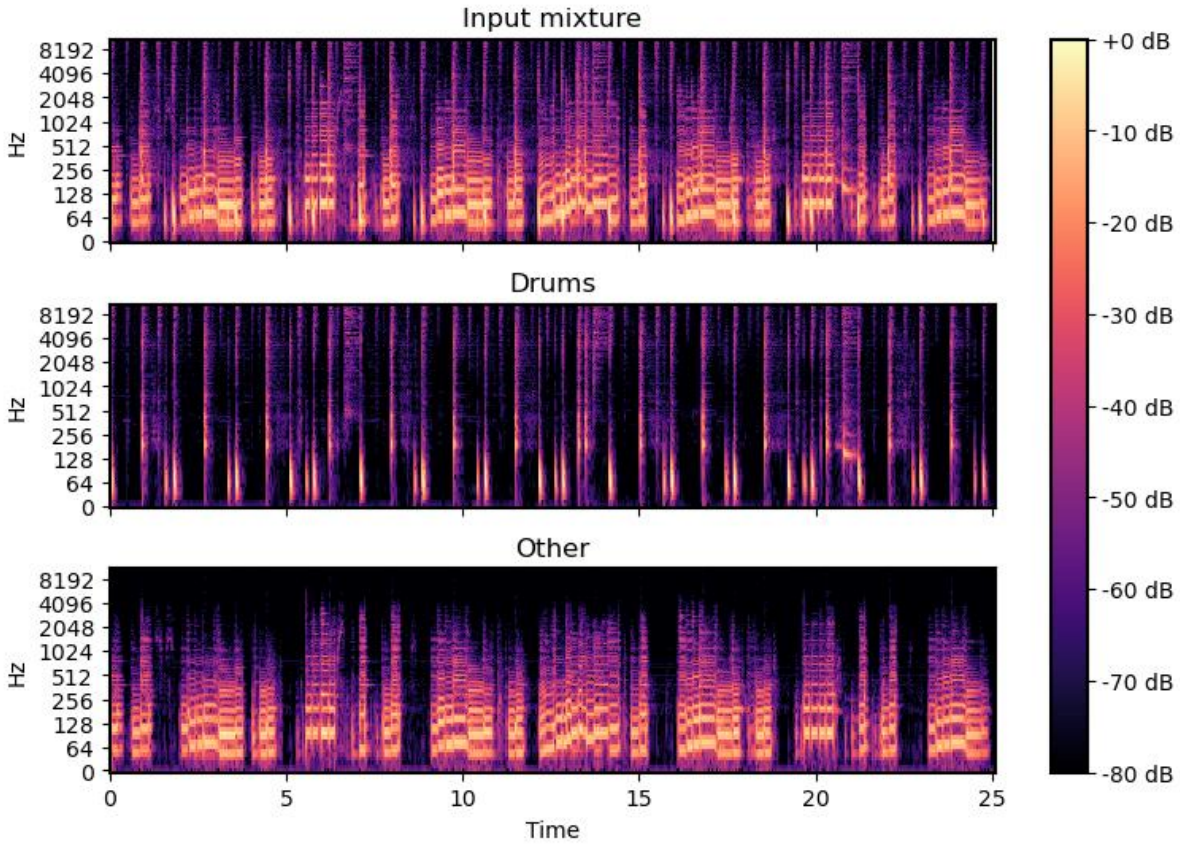
Homework 4: Source Separation

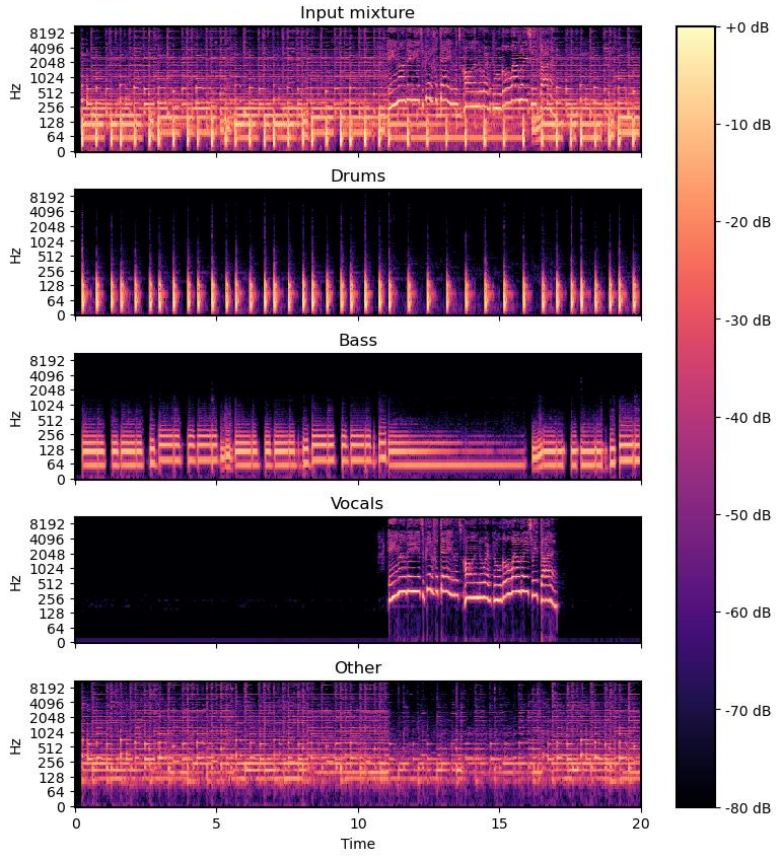
- **Part 1**: Harmonic-Percussive Source Separation (HPSS) using **librosa**



Homework 4: Source Separation

- **Part 2**: Music Source Separation using **Demucs**





Homework 4: Source Separation

- Instructions will be released on the course website
- Please submit your work to Gradescope
- Due at **11:59pm ET** on **February 28**
- Late submissions: **1 point deducted per day**
- No late submission is allowed a week after the due date

Optional Reading

- Ethan Manilow, Prem Seetharman, and Justin Salamon, "Open Source Tools & Data for Music Source Separation," *Tutorials of ISMIR*, 2020.