ual: creative computing institute

LSTM Modelling of FFT Magnitudes for Neural Audio Generation with Minimal Data

Louis McCallum

CCI, UAL, London

MusicRAI 2024

Training LSTM models on sequences of **spectral data** can allow for **high quality audio** to be generated from **minimal datasets** of raw audio

This, in conjunction with **short training times**, provides significant creative and accessibility advantages that make it a more than **viable alternative** to other state of the art audio generation methods.

Heavier Models

Whilst audio quality from trained models is high, it requires significant **equipment**, **data** and **time** to acquire these trained models

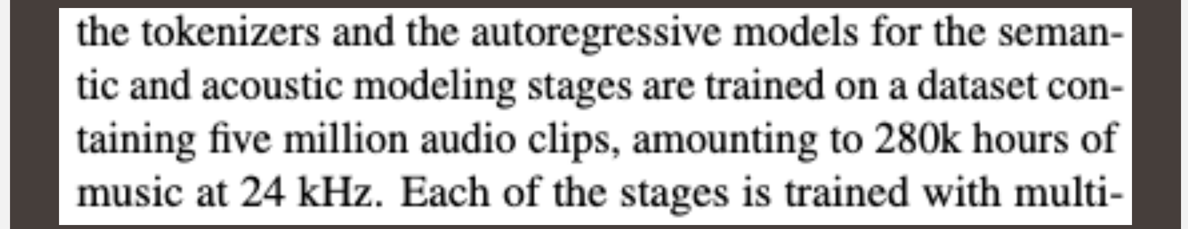
RAVE

Train for 1-2 days, min 3 hours of audio

ddsp

Train for 1-2 hours, 12 mins audio

AITK MusicLM

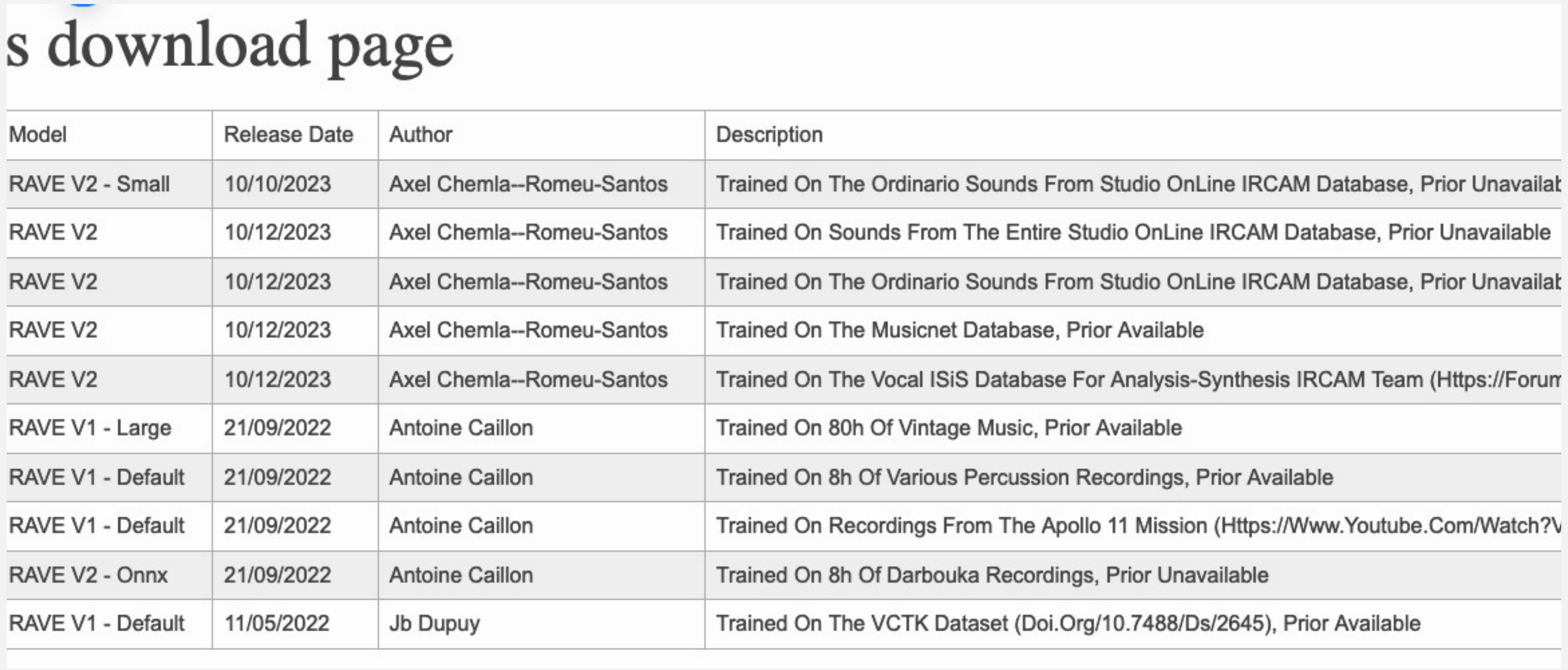


the tokenizers and the autoregressive models for the semantic and acoustic modeling stages are trained on a dataset containing five million audio clips, amounting to 280k hours of music at 24 kHz. Each of the stages is trained with multi-

IMPOSSIBLE?

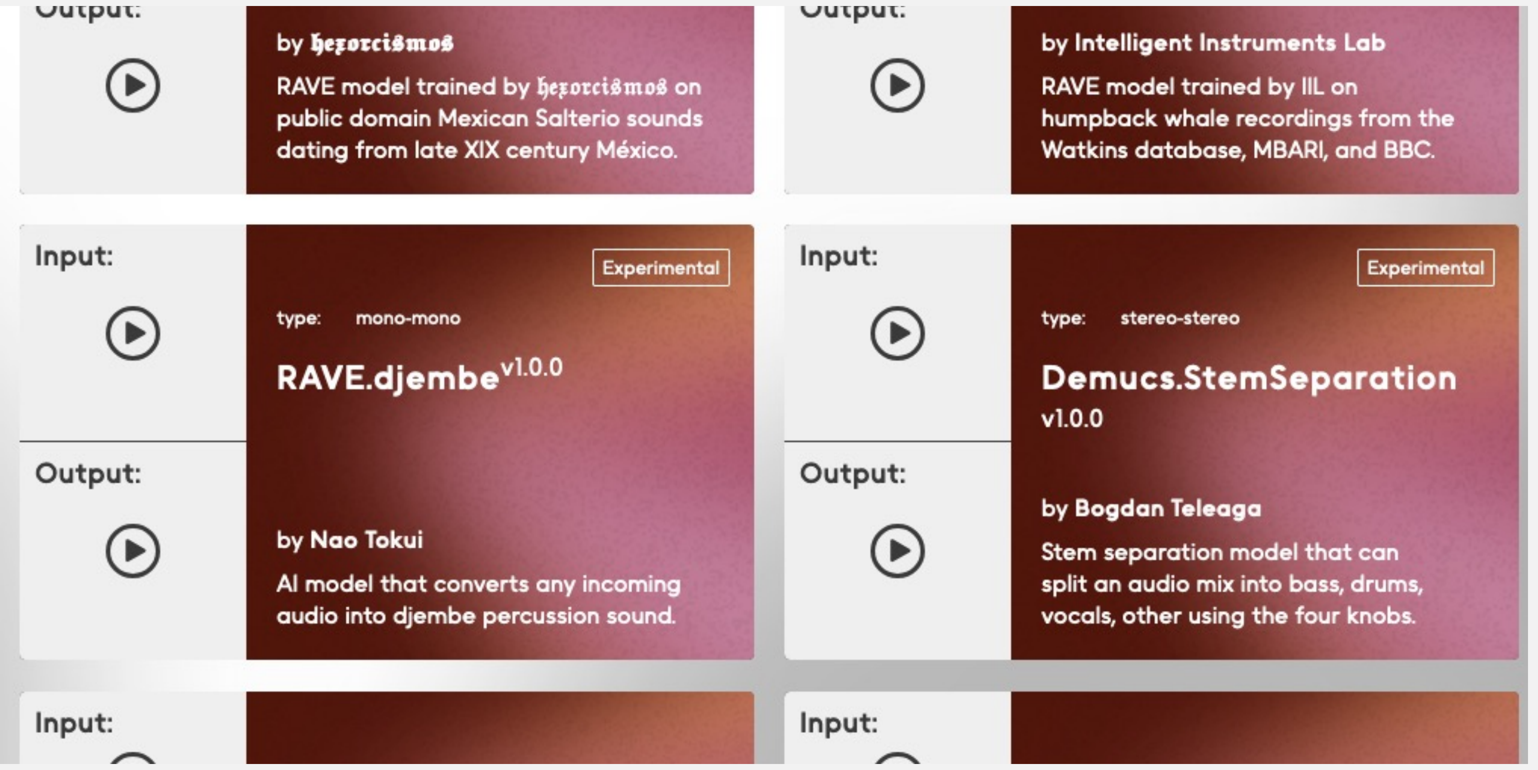
Heavier Models

Although using other’s **pretrained models** may be enough for some musicians (essentially using an novel off the shelf synthesiser), **being involved in the training** process can be a significant part of the creative process



s download page

Model	Release Date	Author	Description
RAVE V2 - Small	10/10/2023	Axel Chemla--Romeu-Santos	Trained On The Ordinario Sounds From Studio OnLine IRCAM Database, Prior Unavailat
RAVE V2	10/12/2023	Axel Chemla--Romeu-Santos	Trained On Sounds From The Entire Studio OnLine IRCAM Database, Prior Unavailable
RAVE V2	10/12/2023	Axel Chemla--Romeu-Santos	Trained On The Ordinario Sounds From Studio OnLine IRCAM Database, Prior Unavailat
RAVE V2	10/12/2023	Axel Chemla--Romeu-Santos	Trained On The Musicnet Database, Prior Available
RAVE V2	10/12/2023	Axel Chemla--Romeu-Santos	Trained On The Vocal ISiS Database For Analysis-Synthesis IRCAM Team (Https://Forum
RAVE V1 - Large	21/09/2022	Antoine Caillon	Trained On 80h Of Vintage Music, Prior Available
RAVE V1 - Default	21/09/2022	Antoine Caillon	Trained On 8h Of Various Percussion Recordings, Prior Available
RAVE V1 - Default	21/09/2022	Antoine Caillon	Trained On Recordings From The Apollo 11 Mission (Https://Www.Youtube.Com/Watch?V
RAVE V2 - Onnx	21/09/2022	Antoine Caillon	Trained On 8h Of Darbouka Recordings, Prior Unavailable
RAVE V1 - Default	11/05/2022	Jb Dupuy	Trained On The VCTK Dataset (Doi.Org/10.7488/Ds/2645), Prior Available



by hexorcismos

RAVE model trained by hexorcismos on public domain Mexican Salterio sounds dating from late XIX century México.

by Intelligent Instruments Lab

RAVE model trained by IIL on humpback whale recordings from the Watkins database, MBARI, and BBC.

type: mono-mono

RAVE.djembe$^{v1.0.0}$

by Nao Tokui

AI model that converts any incoming audio into djembe percussion sound.

type: stereo-stereo

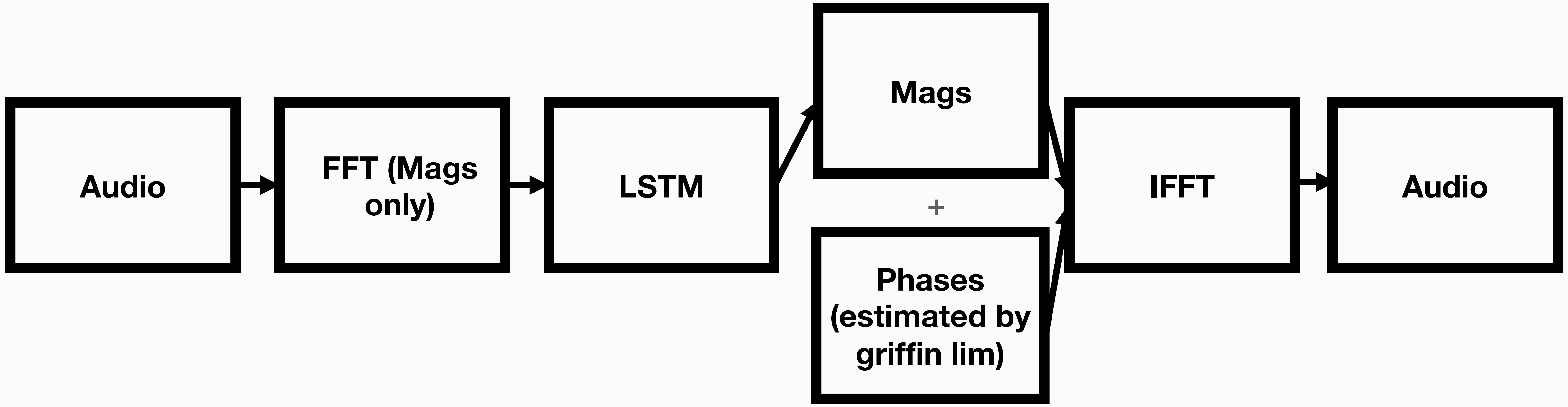
Demucs.StemSeparation v1.0.0

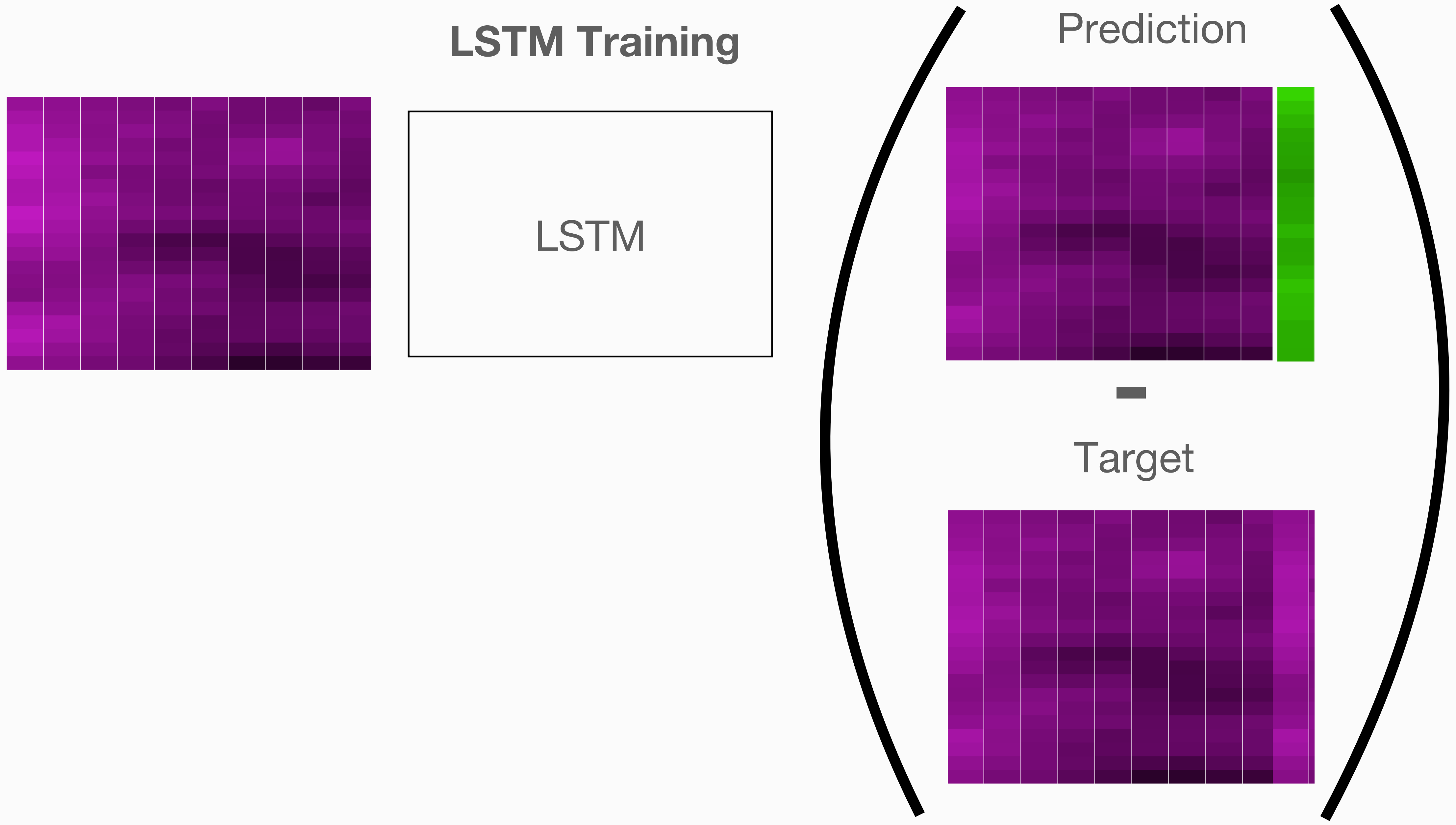
by Bogdan Teleaga

Stem separation model that can split an audio mix into bass, drums, vocals, other using the four knobs.

MAGNet

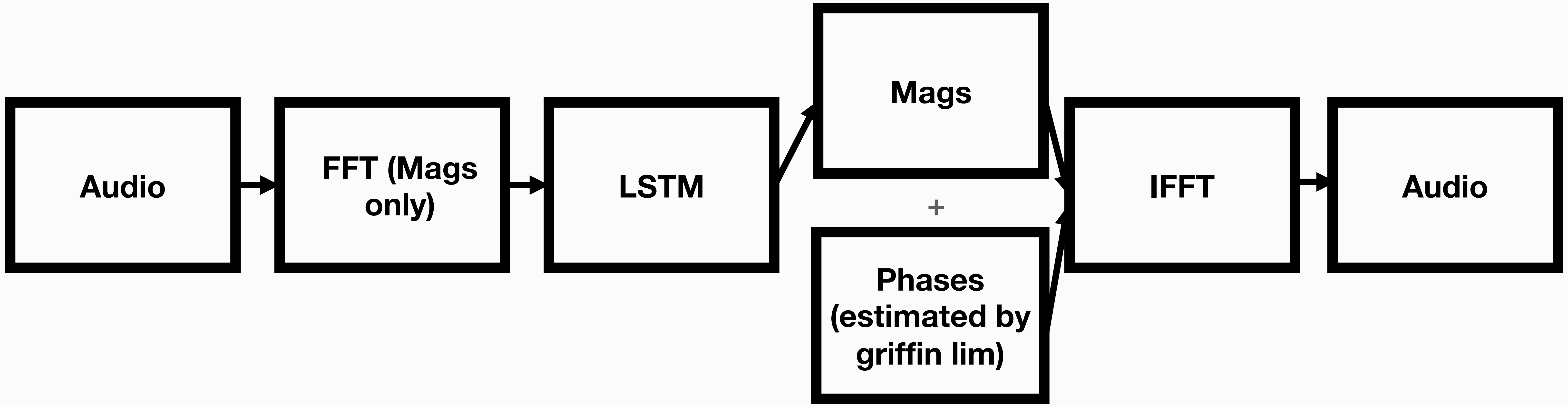
A small network with as little as **3 LSTM layers** is used to model the sequence of FFT **magnitudes**. A **phase vocoding** method is then used to estimate the phases and allow for conversion back into **CD quality** audio





MAGNet

Generation techniques involve **autoregressive generation** and run in **realtime,** allowing for creative intervention



Audio

FFT (Mags only)

LSTM

Mags

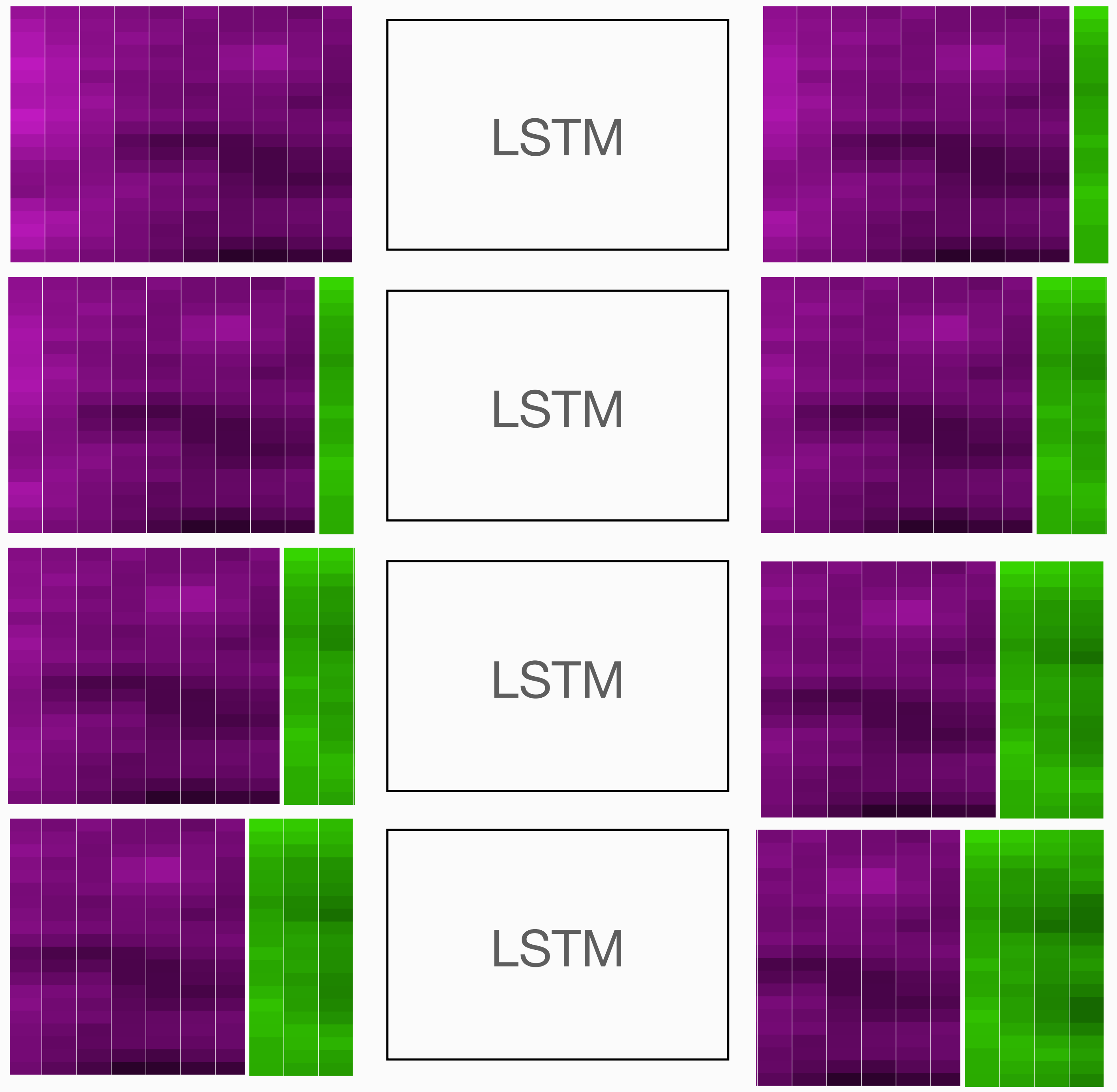
+

Phases (estimated by griffin lim)

IFFT

Audio

Auto Regressive Generation



Datasets can be as small as **30 seconds**. Training a model for **300 iterations** can allow for good results and can take **3 minutes** on free cloud platforms such as Google Colab, or personal computers such as the M1 Macbook Pro

Controllable Palette

This allows for musicians to not only be **very specific** in the audio palette they are looking to use, they also only need to find a **small amount** of it

Interactive Training

They can also train a model, **listen to the output**, see how it fits into the music they are working on and **update their dataset** or training parameters and embark again at little cost.

Train models
https://github.com/Louismac/MAGNet

Use in models in realtime with audio reactive drawing with Dorothy

https://github.com/Louismac/dorothy

Thank you

l.mccallum@arts.ac.uk