DDX7: DIFFERENTIABLE FM SYNTHESIS OF MUSICAL INSTRUMENT SOUNDS

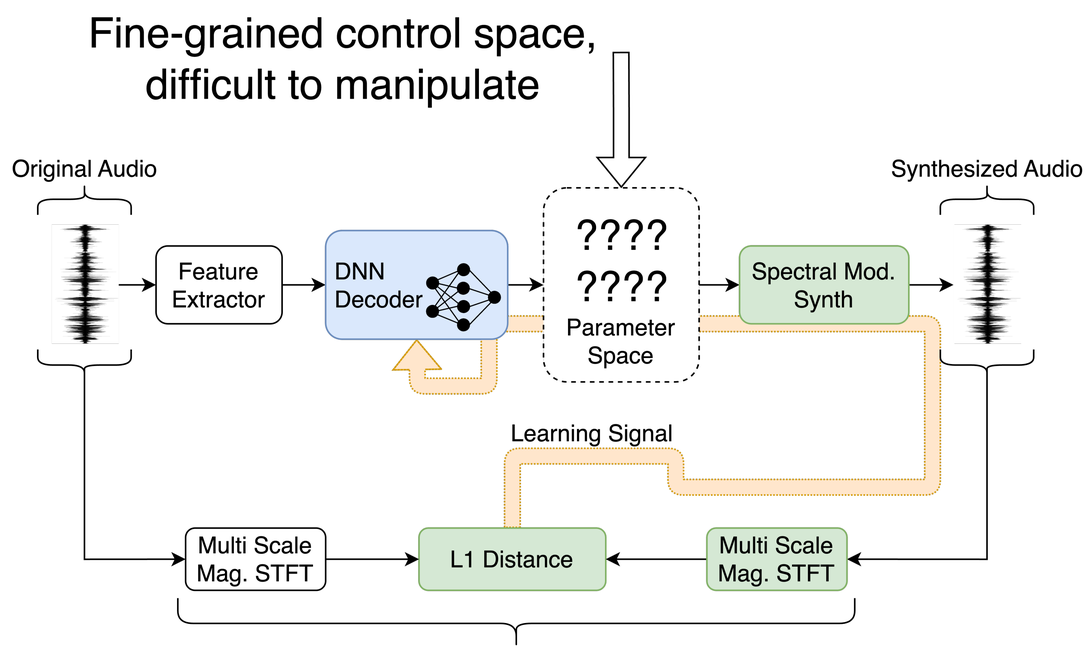
Franco Caspe, Andrew McPherson, Mark Sandler.
Centre for Digital Music - Queen Mary University of London

ABSTRACT

We present Differentiable DX7 (DDX7), a lightweight architecture for neural FM resynthesis of musical instrument sounds in terms of a compact set of parameters of a well-known FM synthesizer. We train the model on instrument samples extracted from the URMP dataset (Li et al., 2019), and quantitatively demonstrate its comparable audio quality against selected benchmarks. With this work, we enable data-driven, neural audio rendering in terms of classic sound design parameters that can be intervened in real-time and could be employed for live timbre manipulation.

MOTIVATION

Current Neural Audio Synthesis (NAS) algorithms, for instance (Engel et al., 2020), drive spectral modelling synthesizers that feature a complex control space, with parameters that are not usually employed by musicians for sound design.



These systems are trained from a corpus of audio data, and employ differentiable synthesizers models, that allow to propagate the learning signal through the generator blocks.

ENTER FM SYNTHESIS

With FM Synthesis we can generate complex spectra using a few common sound design primitives.

$$Out = A \cdot sin(2\pi f_c t + I \cdot sin(2\pi f_m t))$$

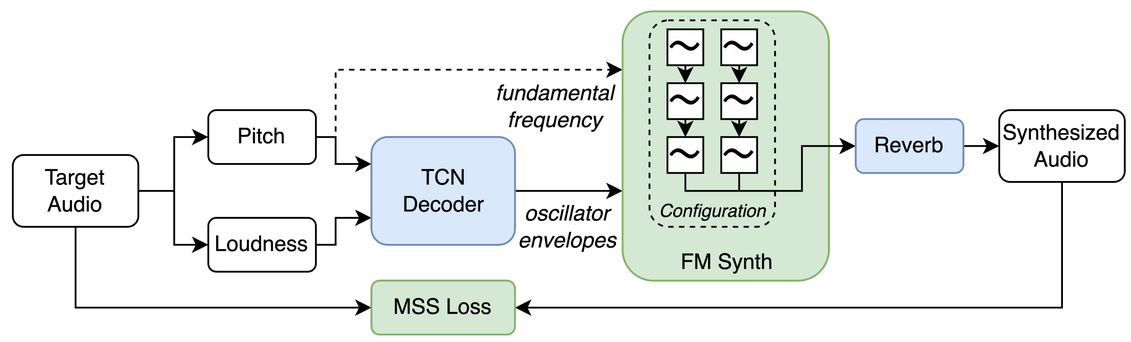
- Modulation Index *I* controls signal bandwidth.
- f_c controls position of spectrum.
- f_m controls distance between harmonics.

But optimizing the parameters of an FM synthesizer is not an easy task. Current losses employ by NAS algorithms cannot optimize well the position of harmonics in spectra (Turian and Henry, 2020).

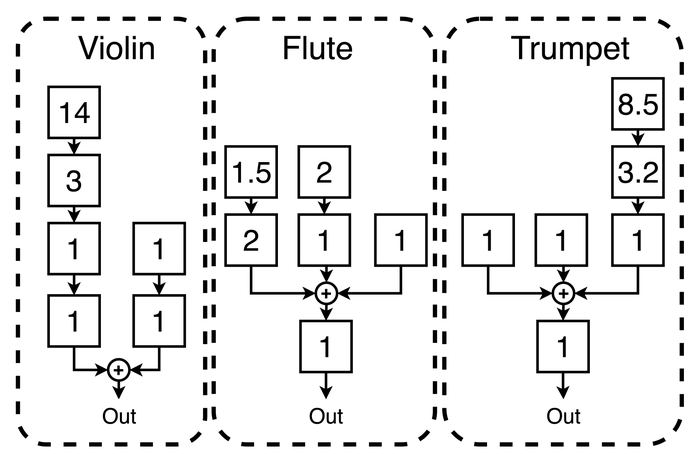
We take inspiration from the **Yamaha DX7** synthesizer, a 6-oscillator FM synth that generates dynamic changes on timbre by manipulating the **modulation indexes** while leaving *fixed* the frequency ratios and oscillator configuration.

DIFFERENTIABLE DX7

We pair a Temporal Convolutional Network (TCN) with a differentiable 6-oscillator FM synthesizer based on the DX7.



We train one DDX7 model for each instrument corpus extracted from the URMP dataset. We try with different fixed FM configuration and ratios copied from Yamaha DX7 patches.



We also test limiting the maximum modulation index, and conduct an oscillator ablation test. We train a Harmonic plus Noise model (Engel et al., 2020) as a baseline. We use the Fréchet Audio Distance (FAD) between test set resynthesis and complete audio corpus as evaluation metric.

RESULTS

We find that DDX7's performance is comparable to that of the baseline, while featuring a compact control space of well-known sound design primitives and a ten times smaller model size.

Model	***Flute***	***Violin***	***Trumpet***
Original Test Set	2.074	0.577	1.069
HpN Baseline	4.326	**0.795**	2.486
DDX7 (6 osc. $I_{max} = 2$)	**2.731**	1.618	4.941
DDX7 (2 osc. $I_{max} = 2\pi$)	3.364	8.270	**1.674**

FAD metric. Best results in bold, and best FM configurations underlined.

REFERENCES

Engel, Jesse et al. (2020). "DDSP: Differentiable Digital Signal Processing". In: *8th International Conference on Learning Representations*. Addis Ababa, Ethiopia.

Li, Bochen et al. (Feb. 2019). "Creating a Multitrack Classical Music Performance Dataset for Multimodal Music Analysis: Challenges, Insights, and Applications". In: *IEEE Transactions on Multimedia* 21.2, pp. 522–535. ISSN: 1520-9210, 1941-0077.

Turian, Joseph and Max Henry (Dec. 2020). "I'm Sorry for Your Loss: Spectrally-Based Audio Distances Are Bad at Pitch". In: *arXiv:2012.04572 [cs, eess]*. (Visited on 05/15/2022).

Scan me for more info, audio examples, and paper preprint.