DDSP-BASED NEURAL WAVEFORM SYNTHESIS OF POLYPHONIC GUITAR PERFORMANCE FROM STRING-WISE MIDI INPUT
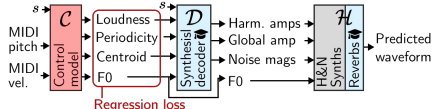
Nicolas Jonason[1], Xin Wang[2], Erica Cooper[2], Lauri Juvela[3], Bob L. T. Sturm[1], Junichi Yamagishi[2]

[1]KTH Royal Institute of Technology, Sweden [2]National Institute of Informatics, Japan
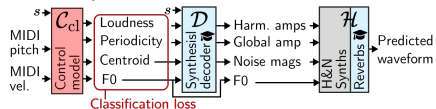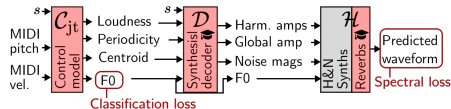[3]Aalto University, Finland
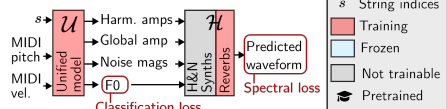
Systems
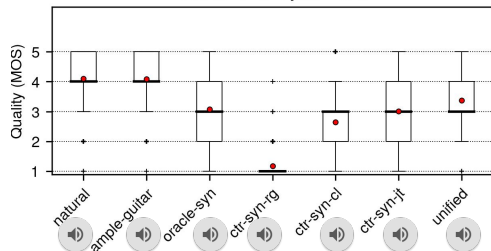
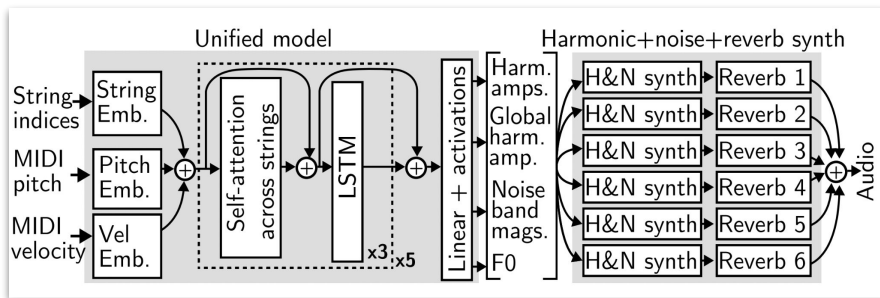1) `ctr-syn-rg (default configuration)`

2) `ctr-syn-cl`

3) `ctr-syn-jt`

4) `unified`

Legend
- s — String indices
- Training
- Frozen
- Not trainable
- Pretrained

Results for all systems, all

natural, ample-guitar, oracle-syn, ctr-syn-rg, ctr-syn-cl, ctr-syn-jt, unified

Architecture details

Unified model — Harmonic+noise+reverb synth

String indices → String Emb.
MIDI pitch → Pitch Emb.
MIDI velocity → Vel Emb.
Self-attention across strings — LSTM — Linear + activations
x3 x5

Harm. amps. — Global harm. amp. — Noise band mags. — F0

H&N synth → Reverb 1
H&N synth → Reverb 2
H&N synth → Reverb 3
H&N synth → Reverb 4
H&N synth → Reverb 5
H&N synth → Reverb 6
→ Audio

Why classify F0 rather than regressing F0?

Legend: target f0; f0 predicted w/ regression; f0 predicted w/ classification; MIDI input; MIDI input, neighbour strings

Paper & Audio Examples

Out of distribution examples

Related work   Dataset   Inharmonicity?

Listening Test   Conclusion