Interconnected topics

Improving real-time collaboration

Expanding to other instruments

Applications

Improving real-time collaboration

Recall
S score, $I(S)$ interpretation

$$S = \{[i, o_i, \#_i, p_i, x_i]\}$$

$$I(S) = \{[i, o_i, \#_i, p_i, x_i, t_i, f_i]\}$$

Human plays S_1, machine plays S_2
Machine task: create $I(S_2)$
such that $I(S_1) + I(S_2)$ is a reasonable interpretation of $S_1 + S_2$

Improving real-time collaboration

Idea: use RNN to generate sequences $\{t_i\}, \{f_i\}$

Preliminary result: works for one score.
Fails to transfer to a new score.

Improving real-time collaboration

A visual analogy

Improving real-time collaboration

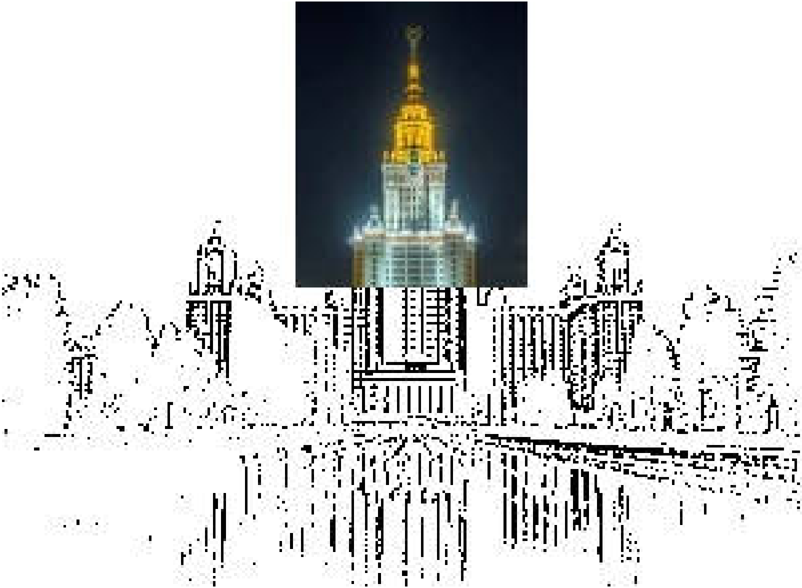
A visual analogy

Improving real-time collaboration

A visual analogy

Improving real-time collaboration

A visual analogy



Listening and understanding

How to represent the information received from real-time input?

Extreme 1:
Align to score
(what we did with Wild Rose Song)

Extreme 2:
Unmodified spectrogram
(have not tried this approach)

For MIDI Piano, our solution: score + 2 extra dimensions

Listening and understanding

Acoustic piano

We should conduct tests of existing transcription software, under varying acoustic conditions.

record simultaneously MIDI (ground truth T) and acoustic (S_1)
$S_2, \ldots$: acoustically record MIDI playback of T from various positions, with noise etc.
apply software in question to S_i, converting them to MIDI, then compare to T.

NB relevance to Jing-Heng's project

If results are not excellent, let's try to leverage score information.

Listening and understanding

Case study: Violin

Reduce to continuous graphs against time of
- fundamental frequency and
- loudness

From frequency, extract note onset times

Listening and understanding

Case study: Violin

Useful information for collaboration that we lose:

“tone colour”
- sul tasto = bow near the middle → more fundamental
- sul ponticello = bow near the edge → more higher partials
- bow pressure → different waveform
- (“articulation” $\subset$ bow pressure and speed vs. time)

vibrato

indications of a note about to change
- stop vibrato
- slide

Listening and understanding

Case study: Human voice

Match to a finite list of acceptable phonemes
Fundamental frequency (as before)
Loudness (as before)

It is difficult to extract good note onset times from the frequency and loudness graphs (cf our experience with Wild Rose Song).
Probably the listening must be based on a combination of melody and lyrics.

Listening and understanding

It is likely natural that every instrument should have its own protocol.

In chamber music, human musicians need to learn to play with each different instrument.

Applications

Concerts

Musical education

A Collaborative AI Performer in Western Classical Music

Kit Armstrong 2024-12-06

Scope: Western classical music

Fixed score, creative interpretation

“The Game”: create new and personal interpretations of classic scores, that listeners perceive as logical/convincing/etc.

Evaluation made in reference to:
score content + fashion + presentation quality + ...

General objectives

Enabling AI to make music

Helping humans to make music

Along the way: understand human music

General objectives

Enabling AI to make music

- So-called expressive performance
- Collaborative setting; chamber music
- “Turing test”

Helping humans to make music

- Instruments are difficult
- The difficulty of piano playing is partly mechanical
- Why turn a human into a machine?

Collaborative setting

Let two human musicians play with each other.
Can we replace one by a machine?

Experiment, Part 1: Data collection

Choose a score S and split it into two parts, T and U.
Invite 2 musicians. One plays T while the other plays U.

Let's call the interpretations $I(T), I(U)$.
Concept: can a machine reconstruct $I(U)$ from $I(T)$?

Collaborative setting

Experiment, Part 2: Modelling

We created a model to mimic how a human performs the task. Main points:

- Full knowledge of S
- Real-time action
  $\Rightarrow$ imperfect knowledge of $I(T)$, filling in as time goes on;
  $\Rightarrow$ cannot change the past

Collaborative setting

Experiment, Part 3: Testing

Choose more scores, $S_i = T_i + U_i$
(some new, some already in Part 1)

Invite musicians to play T_i.
Blind test: experimenter decides whether U_i is played by another human, or by the model.

Result: sometimes the human can't tell.
($33\% \leq$ accuracy $\leq 79\%$, depending on score chosen)

Model based on Kuramoto oscillators: Details

Focus: time-domain

Aim: to attempt to be human-like
⇒ imperfect synchronization

Parameters to be learned by using data collected in Part 1

Notation

$$S = \{[i, o_i, \#_i, p_i, x_i]\}$$

$$I(S) = \{[i, o_i, \#_i, p_i, x_i, t_i, f_i]\}$$

Model based on Kuramoto oscillators: Details

Original Kuramoto oscillator equation:

$$\frac{d\theta_i}{dt} = \sum_{j \neq i} k_{ij} \sin\left(\theta_j(t) - \theta_i(t)\right) + \Omega_i(t)$$

(concept: $\theta = p * 2\pi$)

Adaptations

Continuous → discrete

Interpolate between events

Event: score position p_1, time t_1
Event: score position p_2, time t_2

continuous "imaginary events" at (p, t) where

$$p = p_1 + t \cdot \frac{p_2 - p_1}{t_2 - t_1}, p_1 < p < p_2, t_1 < t < t_2$$

Model based on Kuramoto oscillators: Details

Adaptations (cont.)

Reaction time

Create imaginary events only once p_2, t_2 are known.

Real-time consequence: In order to calculate anything for $t > t_1$, we must wait until t_2.

Therefore, real-time accompaniment requires extrapolation.

Music $\neq$ Tapping

Music is directional, not cyclical ⇒ Remove sin from equation

No intrinsic speed

A score's speed is a result of interpretation, not intrinsic ⇒ Replace Ω with running average

New “reactive” model

Concept: eliminate interpolation
Condense the “effect” of each note into one action instead of spreading it over a time interval.
What is the effect?

$$\theta_2(t)' = k(\theta_2(t) - \theta_1(t)) + \theta_2(t-n)'$$

$$\theta_2(t)'' = K(\theta_2(t) - \theta_1(t))$$

New "reactive" model

With oscillators, we think of everything as functions of time.

But in the end we are actually trying to calculate time from other variables.

Monotonic increasing functions $\Rightarrow$ define $\tau(p)$ "=" $\theta_3^{-1}(p)$

$$\tau : p \mapsto t$$

The action induced by a note heard:

$$\tau''(p_0) = \alpha(t_0 - \tau(p_0))$$

New "reactive" model

Dynamically evolving interpretation

Model has a "current" interpretation I_t at time t.
It remains until an external event is perceived.

Event at time t_0:
It is written verbatim into I_{t_0}, and has ramifications in the rest of I_{t_0}.

Example: pure time-domain synchronization
Current interpretation I_{t_0}, with τ_0 being the relationship between t, p in I_{t_0}
Hear a note at t_1 with score position $p_1 \Rightarrow$

$$\texttt{concept:} \quad \tau(p_1)'' = \alpha(t_1 - \tau(p_1))$$

$$\tau_{t_1}(p) = \tau_{t_0}(p) + \alpha(t_1 - \tau(p_1)) \cdot (p - p_1)$$