Real-Time Piano Accompaniment Model Trained on and Evaluated According to Human Ensemble Characteristics

Kit Armstrong, Tzu-Ching Hung,
Ji-Xuan Huang & Yi-Wen Liu

Acoustics and Hearing Group, Dept. Electrical Engineering,
National Tsing Hua University, Hsinchu, Taiwan

NTHU Dept. of EE
Acoustics and Hearing Group

Context

("Western classical music")

Composer – Performer – Listener

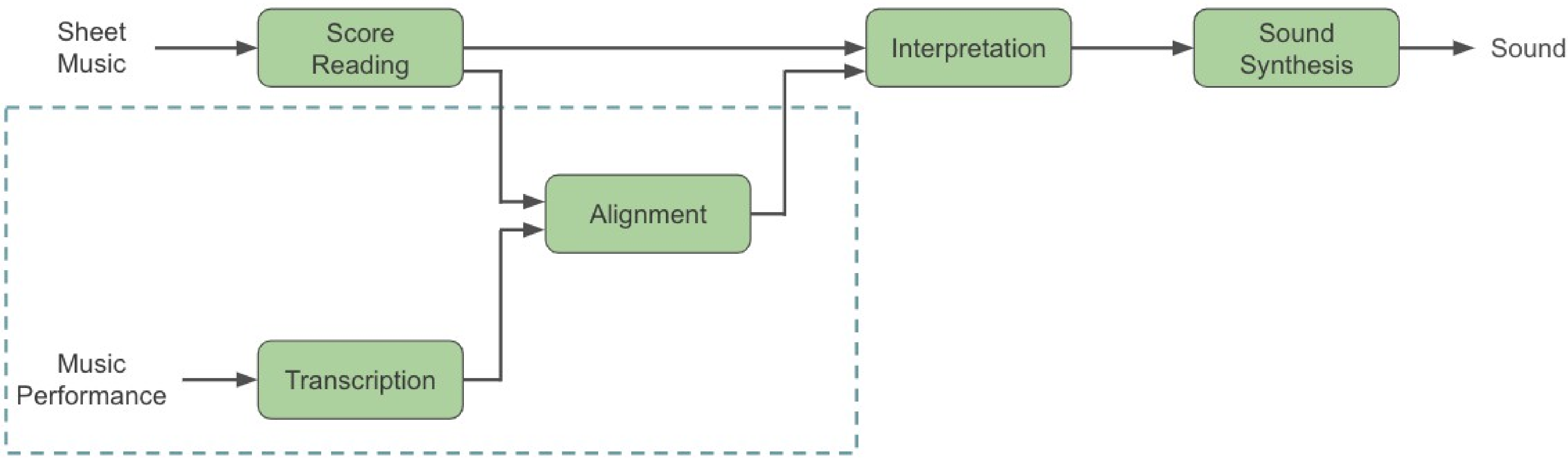
Context

("Western classical music")

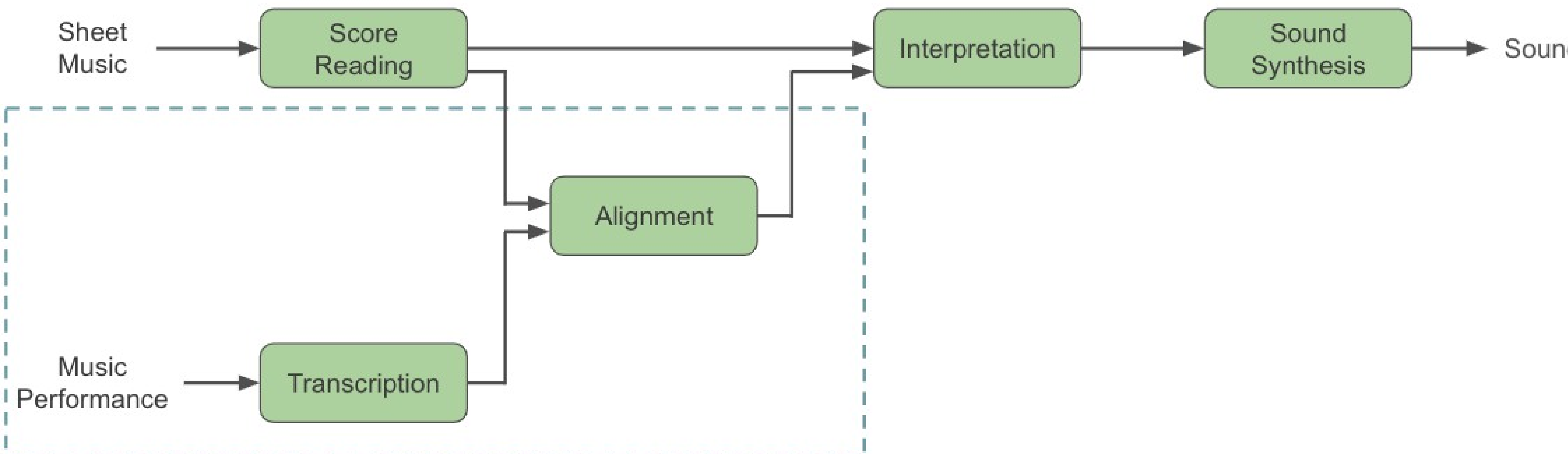
Composer – Performer – Listener

Create an AI performer

Context


Sheet
Music
Score
Reading
Interpretation
Sound
Synthesis
Sound
Alignment
Music
Performance
Transcription

Context



Score reading:
sheet music → a machine-friendly representation

Transcription:
musical performance → a machine-friendly representation

Score following:
musical performance + score → score-aligned performance

Interpretation:
score + score-aligned performance → MIDI or similar

Sound synthesis:
interpretation → sound

Focus: Interpretation

Prevalence:

- Music-notation programs
  like Finale, Sibelius, MuseScore, etc.
- Useful tool for composers
- Not suitable for performance

Focus: Interpretation

Prevalence:

- Music-notation programs
  like Finale, Sibelius, MuseScore, etc.
- Useful tool for composers
- Not suitable for performance

Levels of advancement:

- Basic, in essence MusicXML to MIDI
- Algorithmic expressiveness
- Ongoing attempts with machine learning

Focus: Interpretation

Challenges:

- Deformation for natural performance

Focus: Interpretation

Challenges:

- Deformation for natural performance
- Understanding score indications

Focus: Interpretation

Challenges:

- Deformation for natural performance
- Understanding score indications
- Synchronization in an ensemble

Model

Scope:

- Digital piano
- 1 person ("input") + AI "accompaniment"
- Precisely defined score

Goal:

- Human-like time synchronization

Model

Scope:

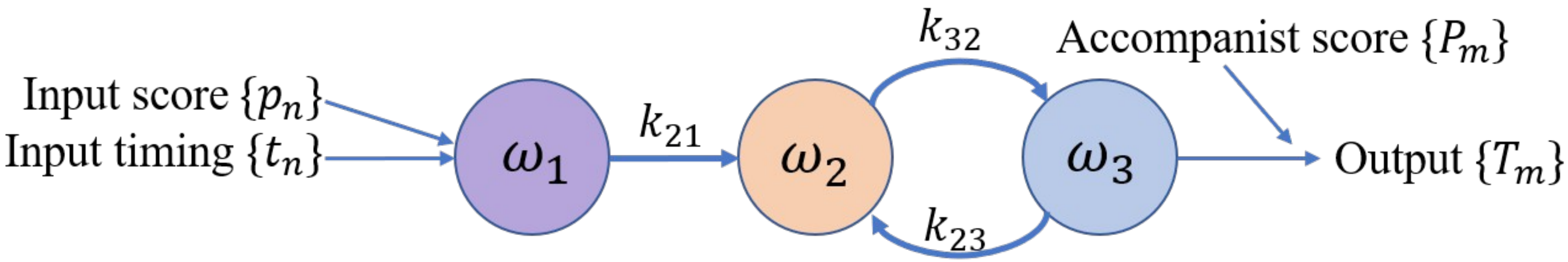
- Digital piano
- 1 person (“input”) + AI “accompaniment”
- Precisely defined score

Goal:

- Human-like time synchronization

<u>Collaboration</u> more than <u>synchronization</u>

Model

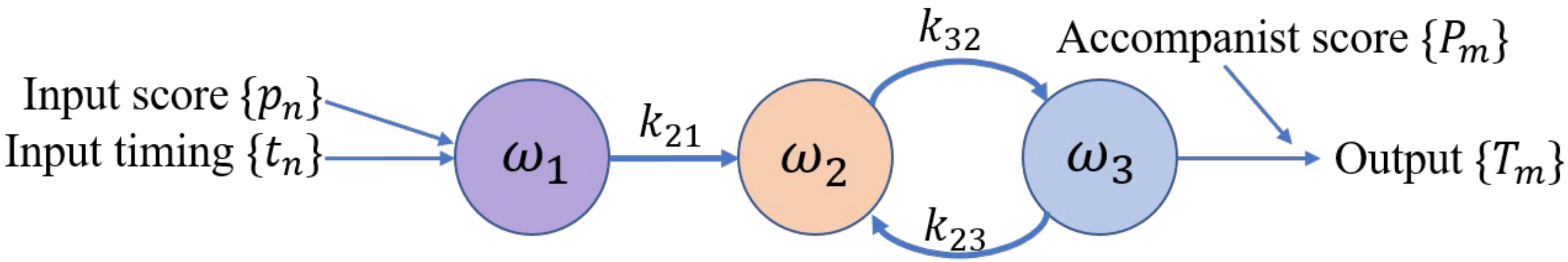


Intuition:

"I'm early" → Slow down
"I'm late" → Speed up

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

Model

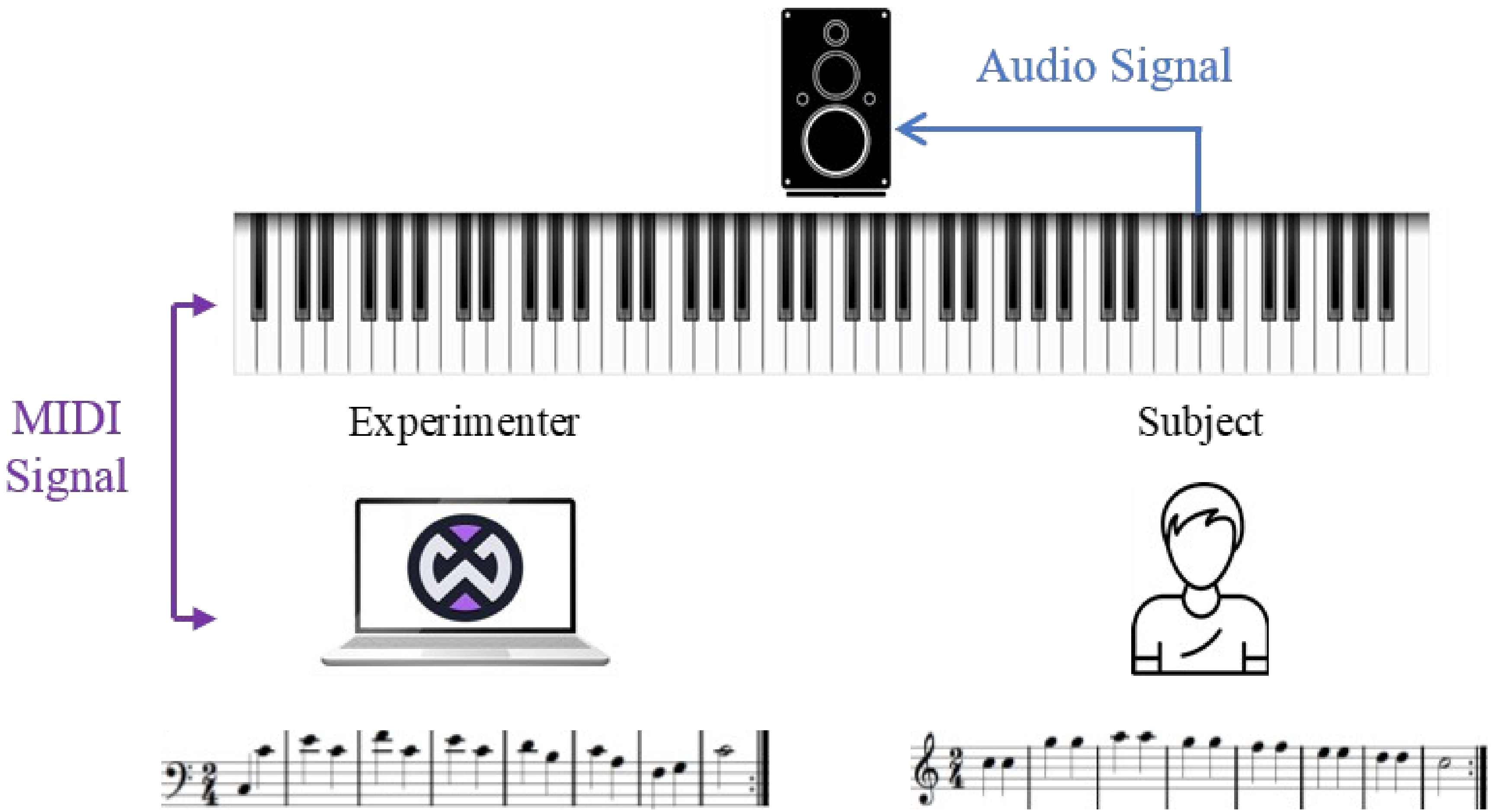


Predictive nature:

- extrapolate ω_3 position to predict next outputs
- each new input rewrites all future predictions
- additional learned parameter: “reaction time”

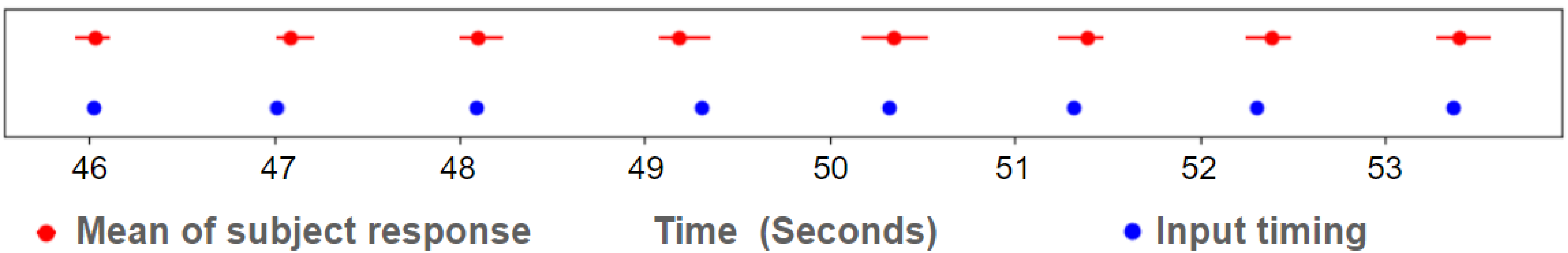
Model training

Capturing human behavior



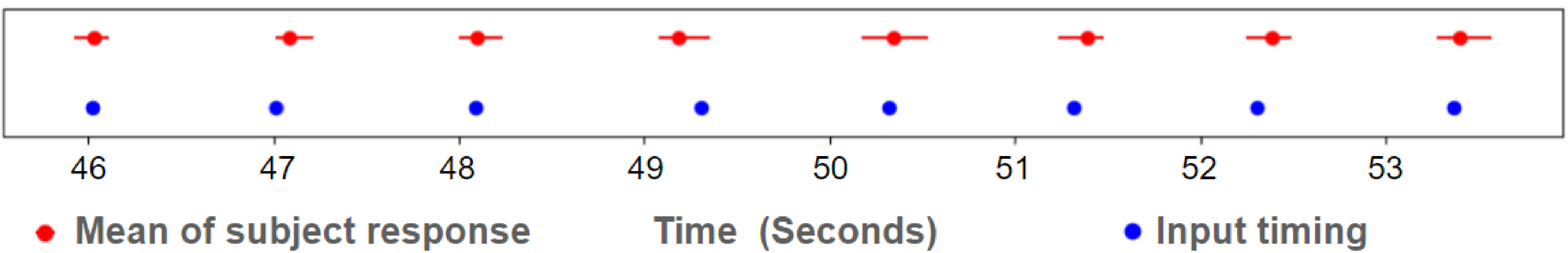
Model training

Capturing human behavior



Model training

Capturing human behavior



Choose parameters such that
the model performs most similarly

Model training

Additional elements

- MIDI Velocity matching (running average)
- Basic error handling

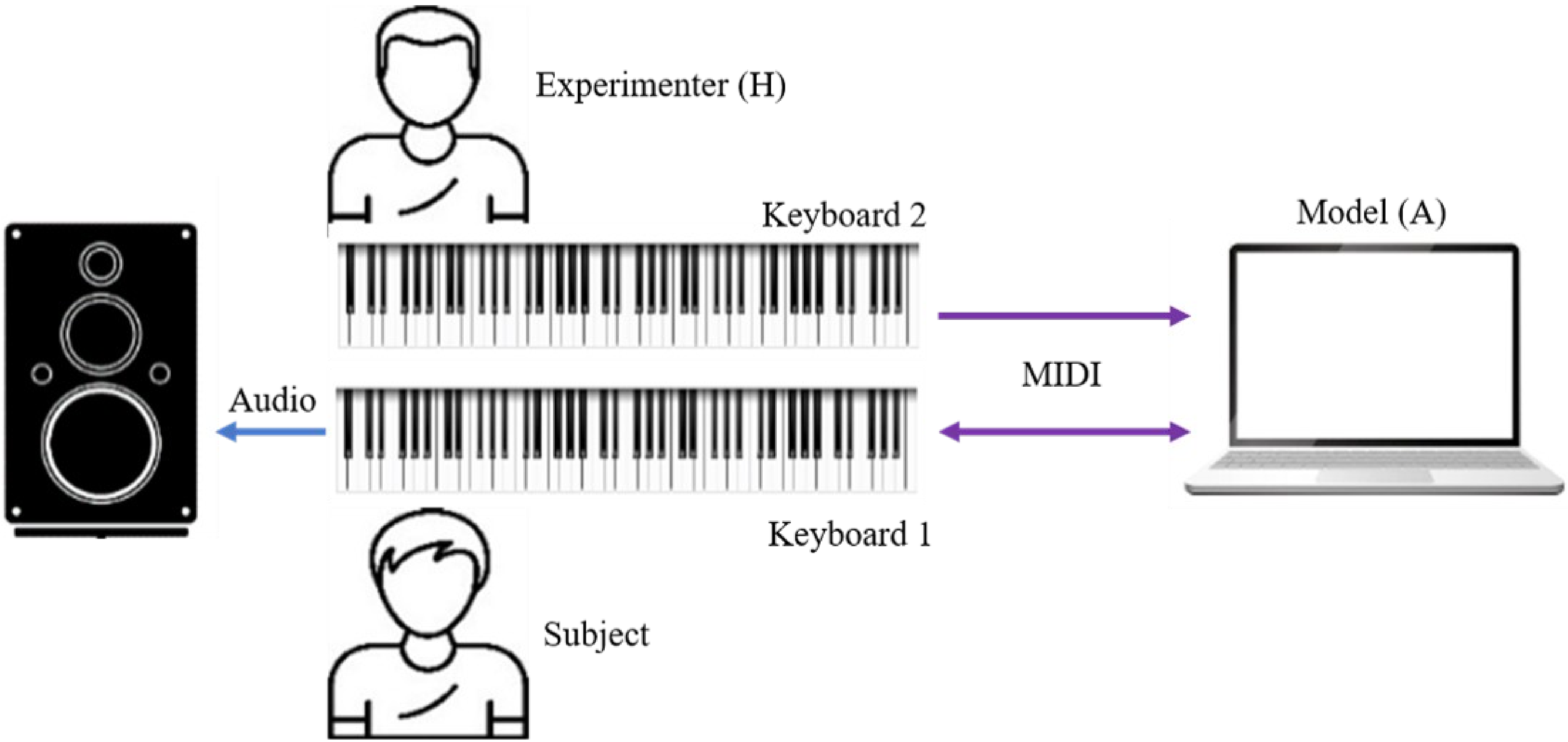
Testing

"Turing test"

Can people tell the difference between Human (H) and AI (A)?

Environment:
Create identical setup for H and A

Testing


Experimenter (H)
Keyboard 2
Model (A)
MIDI
Audio
Keyboard 1
Subject

Testing

- **1.**
  W. A. Mozart: *"Twinkle, Twinkle, Little Star"* K. 265, Theme
- **2.**
  W. A. Mozart: *"Twinkle, Twinkle, Little Star"* K. 265, Variation II
- **3.**
  J. S. Bach/C. Gounod: *"Ave Maria, Méditation sur le Prélude de Bach"*

Testing

"Turing test"

Can people tell the difference between Human (H) and AI (A)?

<u>Results (subjective):</u>

12 participants, 80 trials

Total: 59% correct guesses

Testing

“Turing test”

Can people tell the difference between Human (H) and AI (A)?

How different are really H and A?

Objective Discriminants

- Desynchronization
- Jerk
- Velocity curves

Objective Discriminants

Desynchronization

Human

Piece	Number of trials	μ_Δ	σ_Δ
1	13	0.0872	0.0385
2	13	0.0506	0.0921
3	11	0.0478	0.0684

Automatic

Piece	Number of trials	μ_Δ	σ_Δ
1	14	0.0656	0.0533
2	14	0.0420	0.0344
3	11	0.0254	0.0219

Objective Discriminants

$$\text{Jerk} = \sum(\text{position}''')^2$$

Human

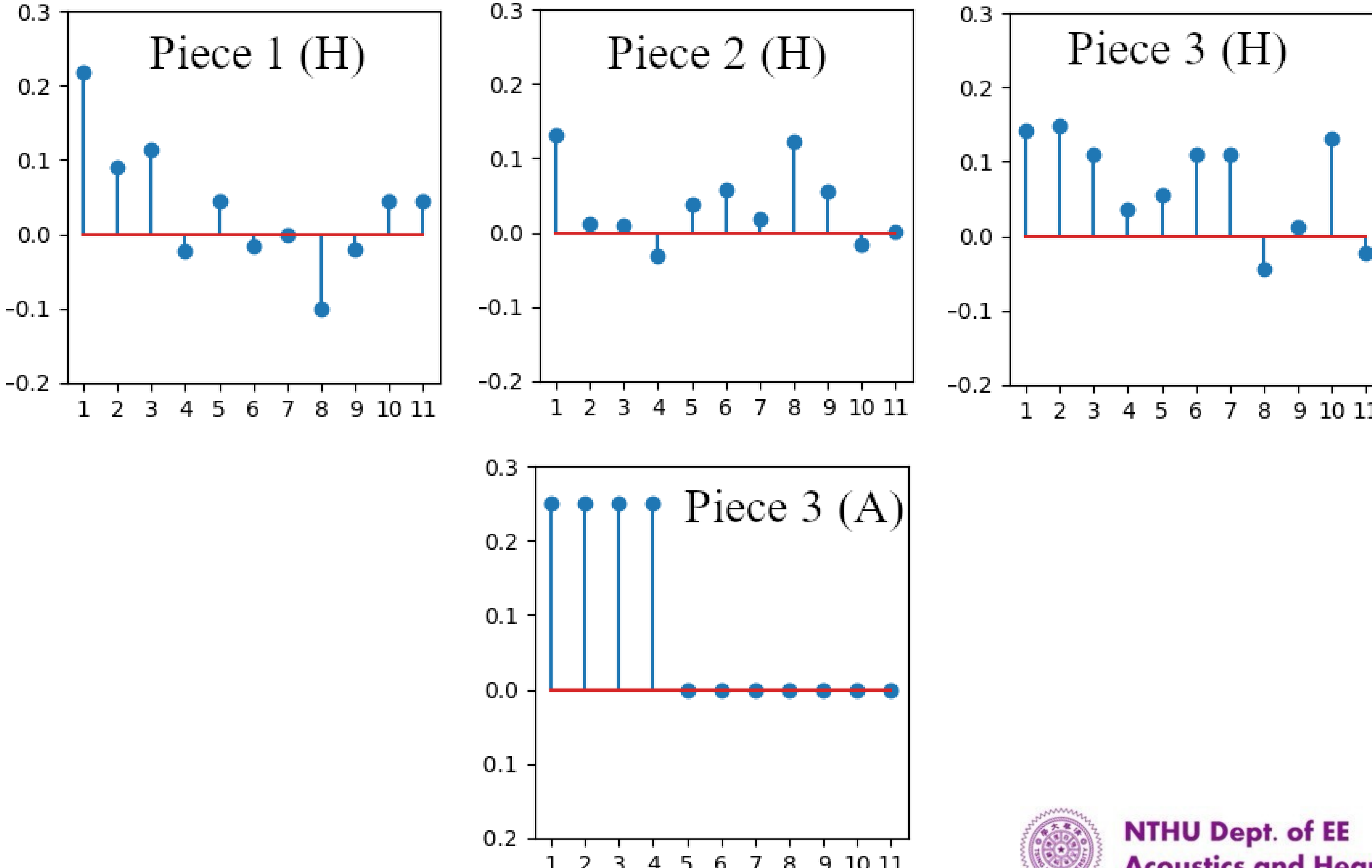
Piece	Trials	μ_J	σ_J
1 H	17	3.074×10^2	4.069×10^2
2 H	11	1.205×10^5	0.947×10^5
3 H	17	0.699×10^2	1.546×10^2

Automatic

Piece	Trials	μ_J	σ_J
1 A	19	0.719×10^3	1.405×10^3
2 A	22	2.817×10^5	2.556×10^5
3 A	24	0.716×10^2	1.355×10^2

Objective Discriminants

Velocity correlation



NTHU Dept. of EE
Acoustics and Hearing Group

Problems & Limitations

Problem: Output delay

Getting the OS to reliably send a MIDI message at a pre-determined future time

- Misrepresents the model
- Bumpy effect
- Cascading slow-down
- Worse when the output has many notes

NTHU Dept. of EE
Acoustics and Hearing Group

Problems & Limitations

Limitations:

- Only rhythm is taken into account

Problems & Limitations

Limitations:

- Only rhythm is taken into account

- Each note carries the same weight or salience

Problems & Limitations

Limitations:

- Only rhythm is taken into account

- Each note carries the same weight or salience

General limitations:

- No intrinsic musicality

- Primitive score reading

Conclusion

- This model tries to imitate certain intuitive aspects of ensemble performance.
- Although it is so far very simple and has limited scope, playing with it is surprisingly satisfying, especially compared to “perfect accompaniment”.
- Through objective analysis, we endeavored to understand external participants' responses.
- Recordings of human collaborative music-making can help further train machine “musicality”.