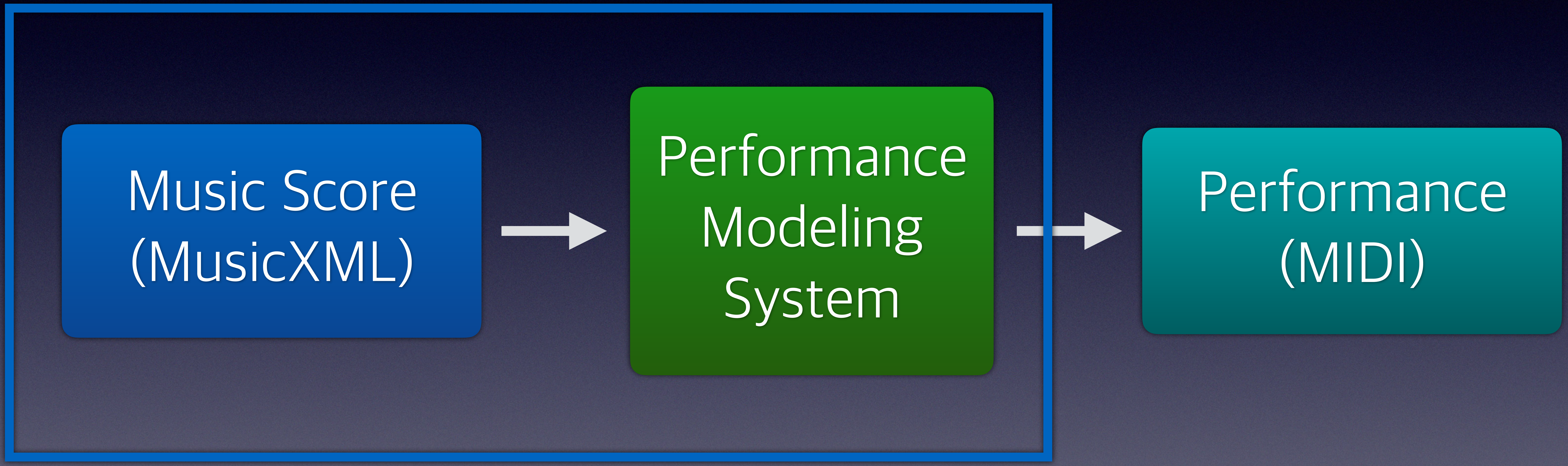
Graph Neural Network for Music Score Data and Modeling Expressive Piano Performance

Dasaem Jeong, Taegyun Kwon, Yoojin Kim, and Juhan Nam
Music and Audio Computing Lab
KAIST, Korea

Research Goal

- Modeling expressive piano performance (aka AI Pianist)

Research Goal

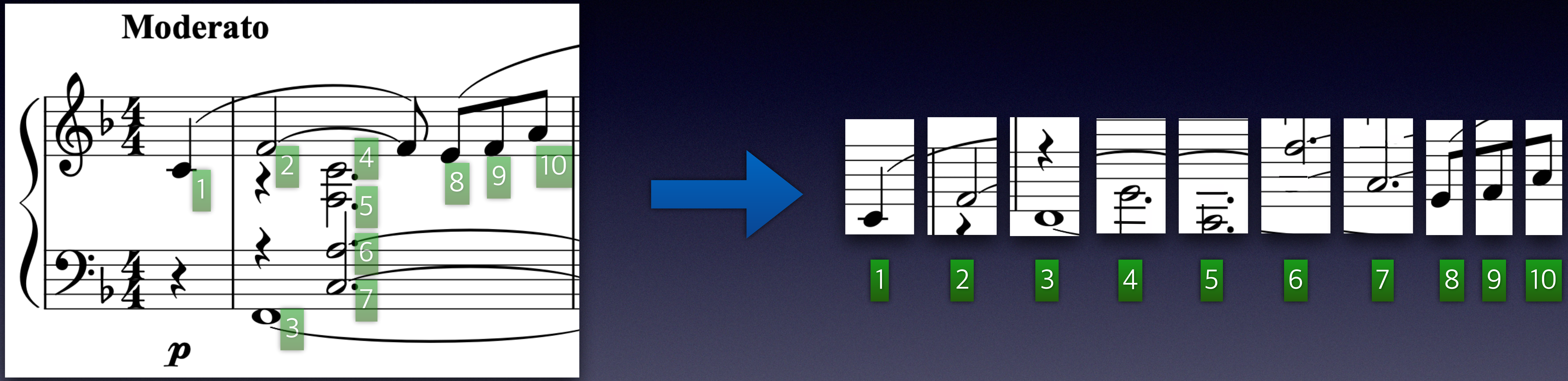


- The core part is embedding music score with neural network.

Previous Representations

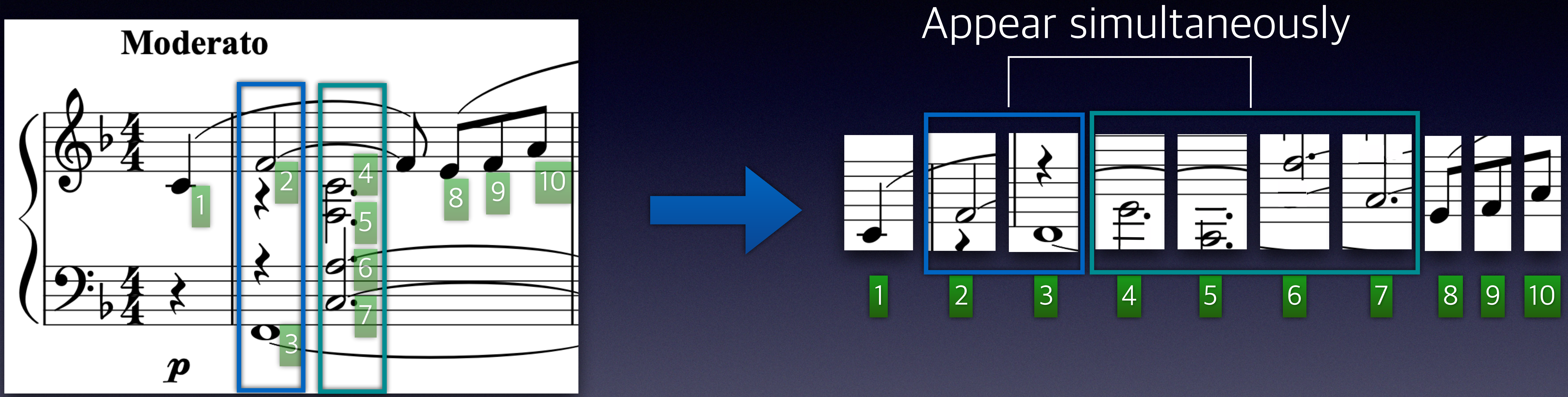
- Word-like sequence of notes
- 2D matrix of notes activation in time and pitch axis

Previous Representations



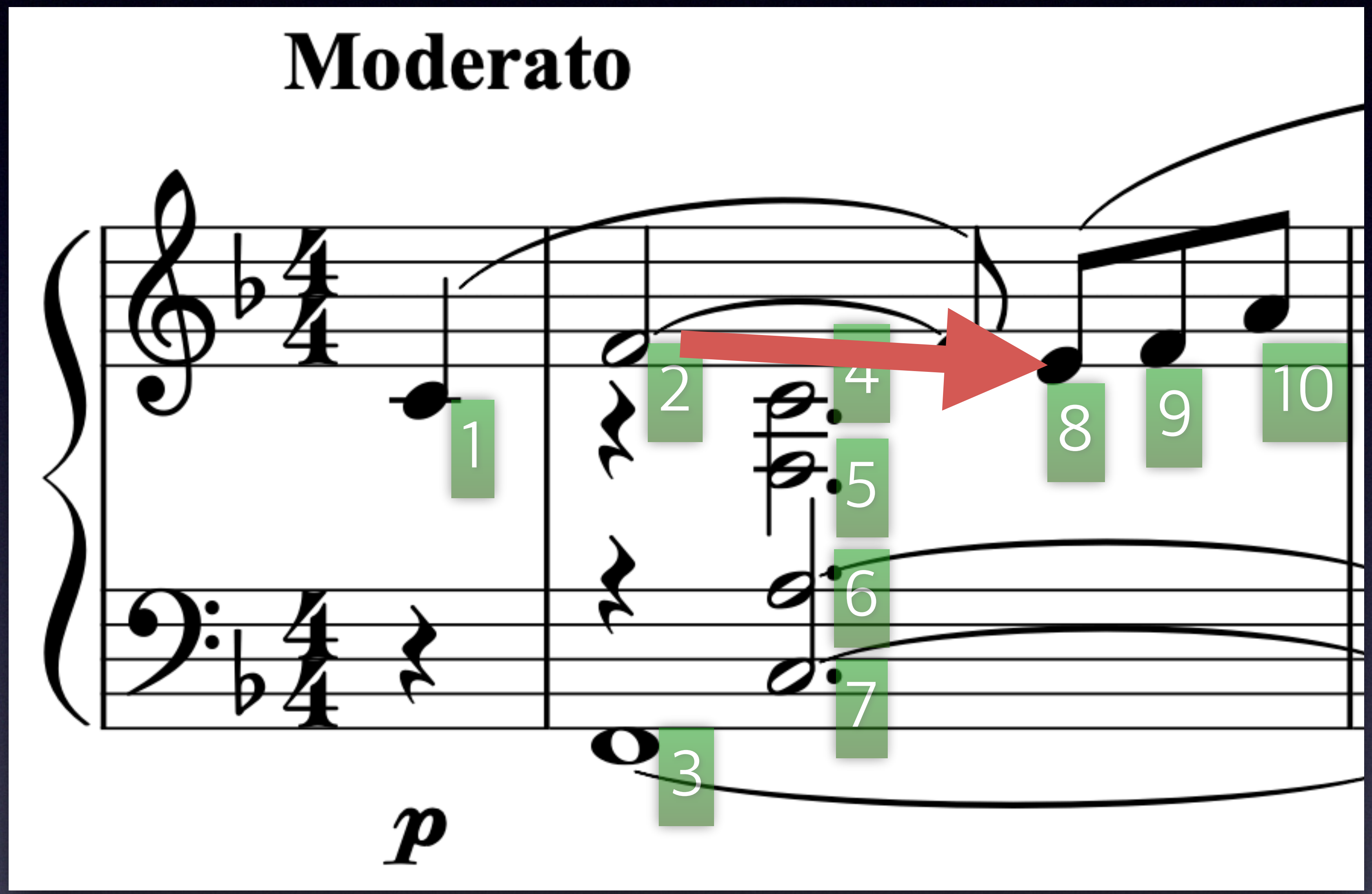
- Flatten music score as a word-like sequence of notes
- The relation of neighboring element in the sequence is not consistent

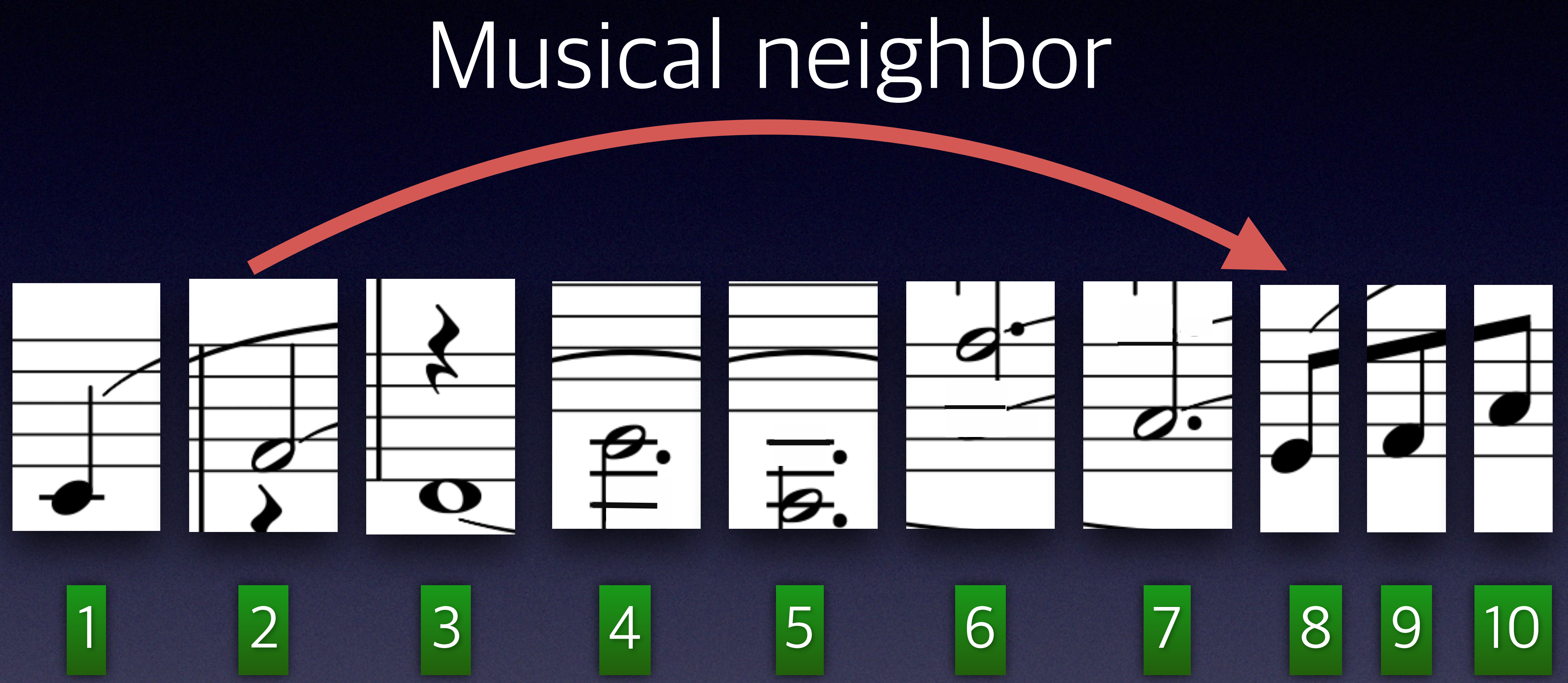
Previous Representations



- Flatten music score as a word-like sequence of notes
- The relation of neighboring element in the sequence is not consistent

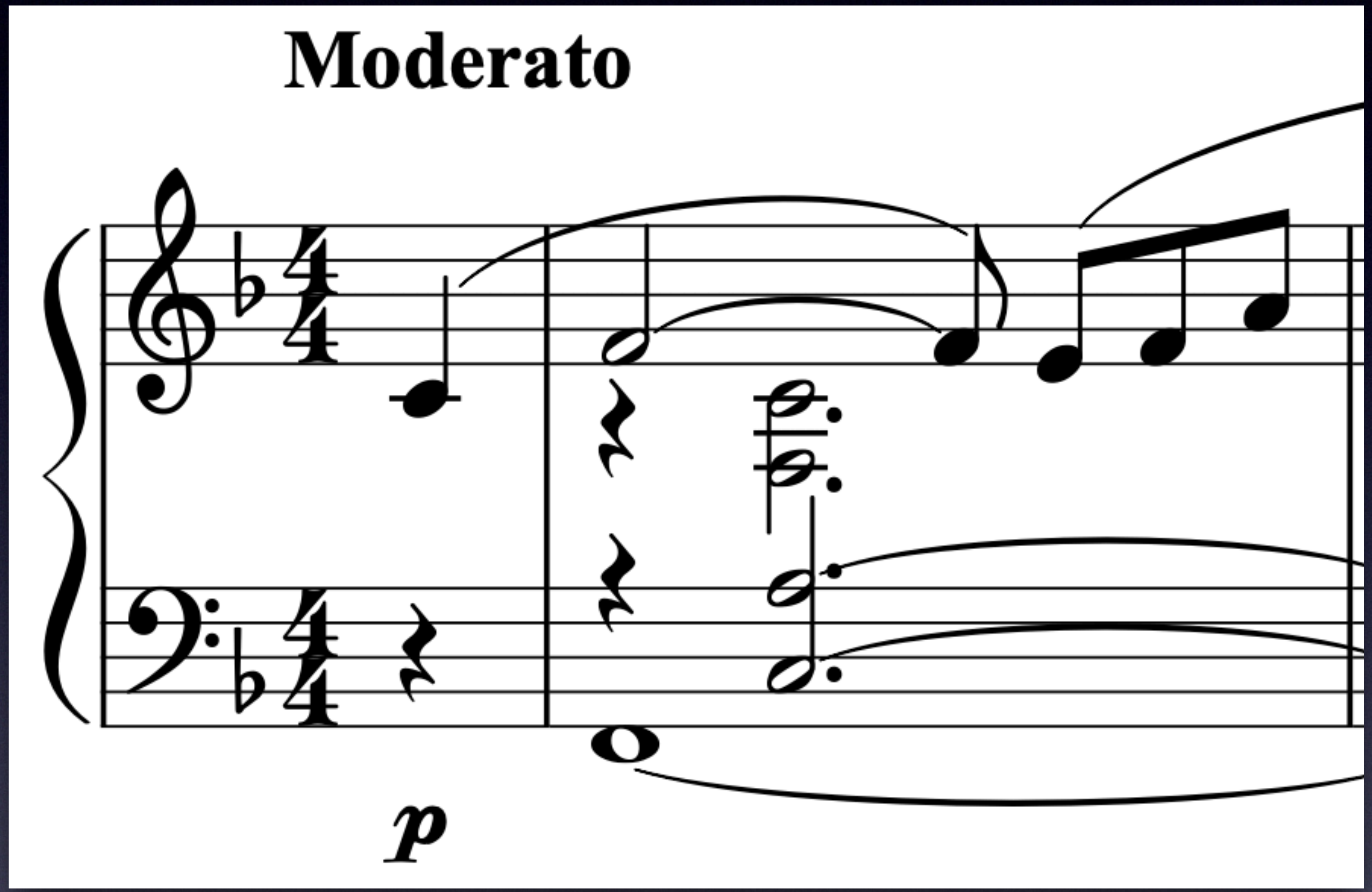
Previous Representations

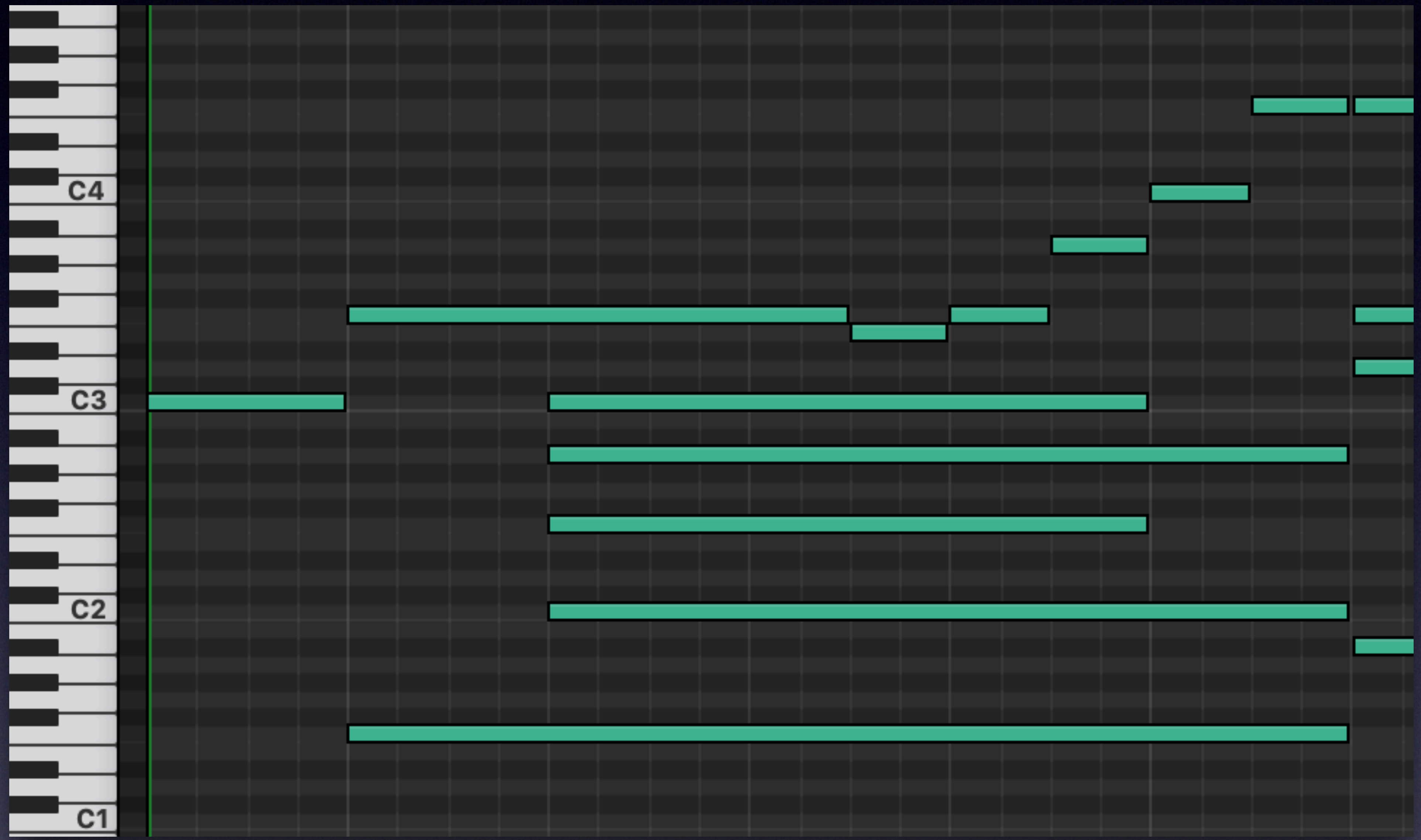




- Flatten music score as a word-like sequence of notes by time and pitch
- The relation of neighboring element in the sequence is not consistent

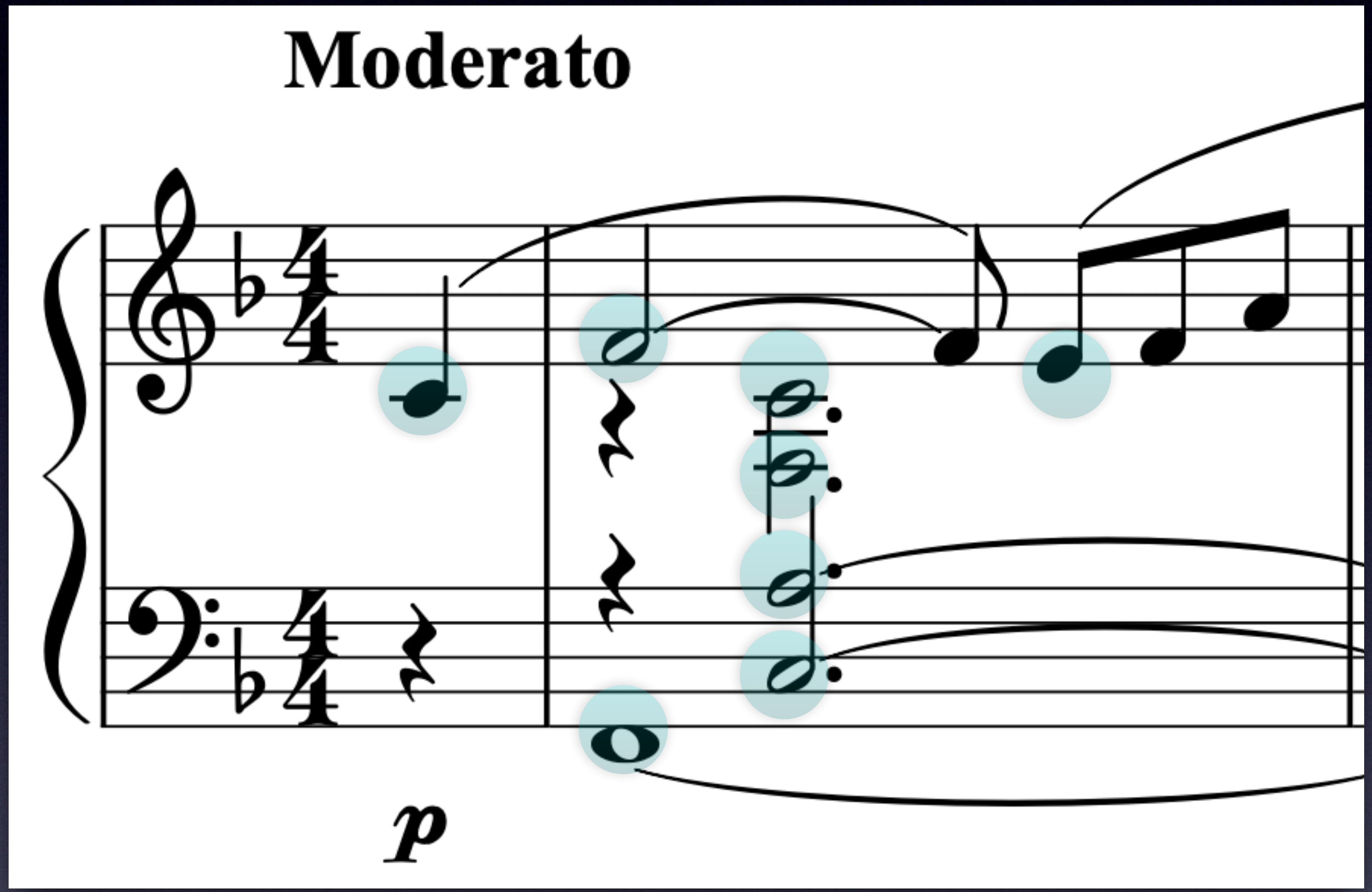
Previous Representations

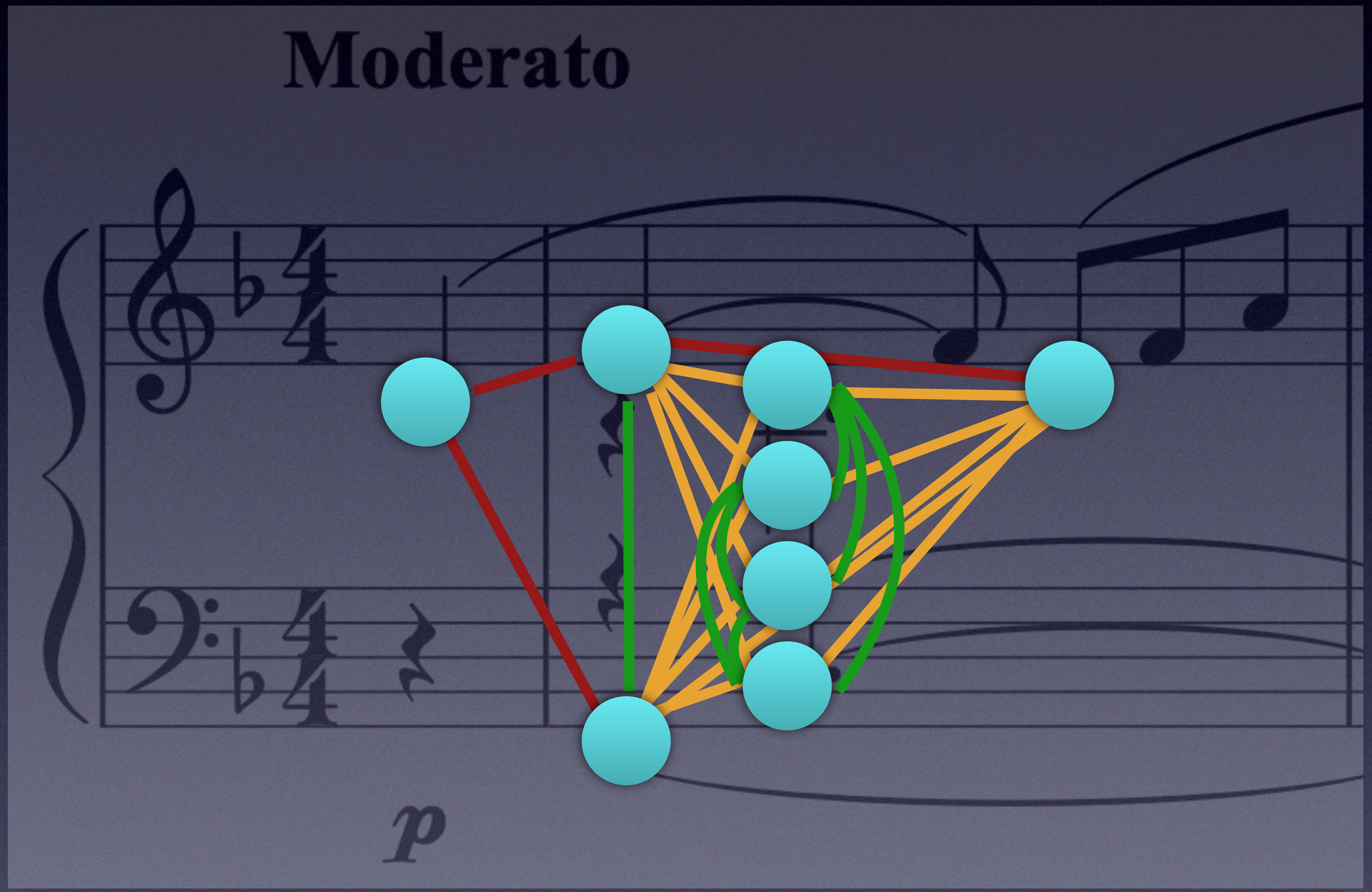




- Convert music score as a 2D matrix of note activation in time and pitch axis (piano-roll)
- Sampling-based representation rather than event-based

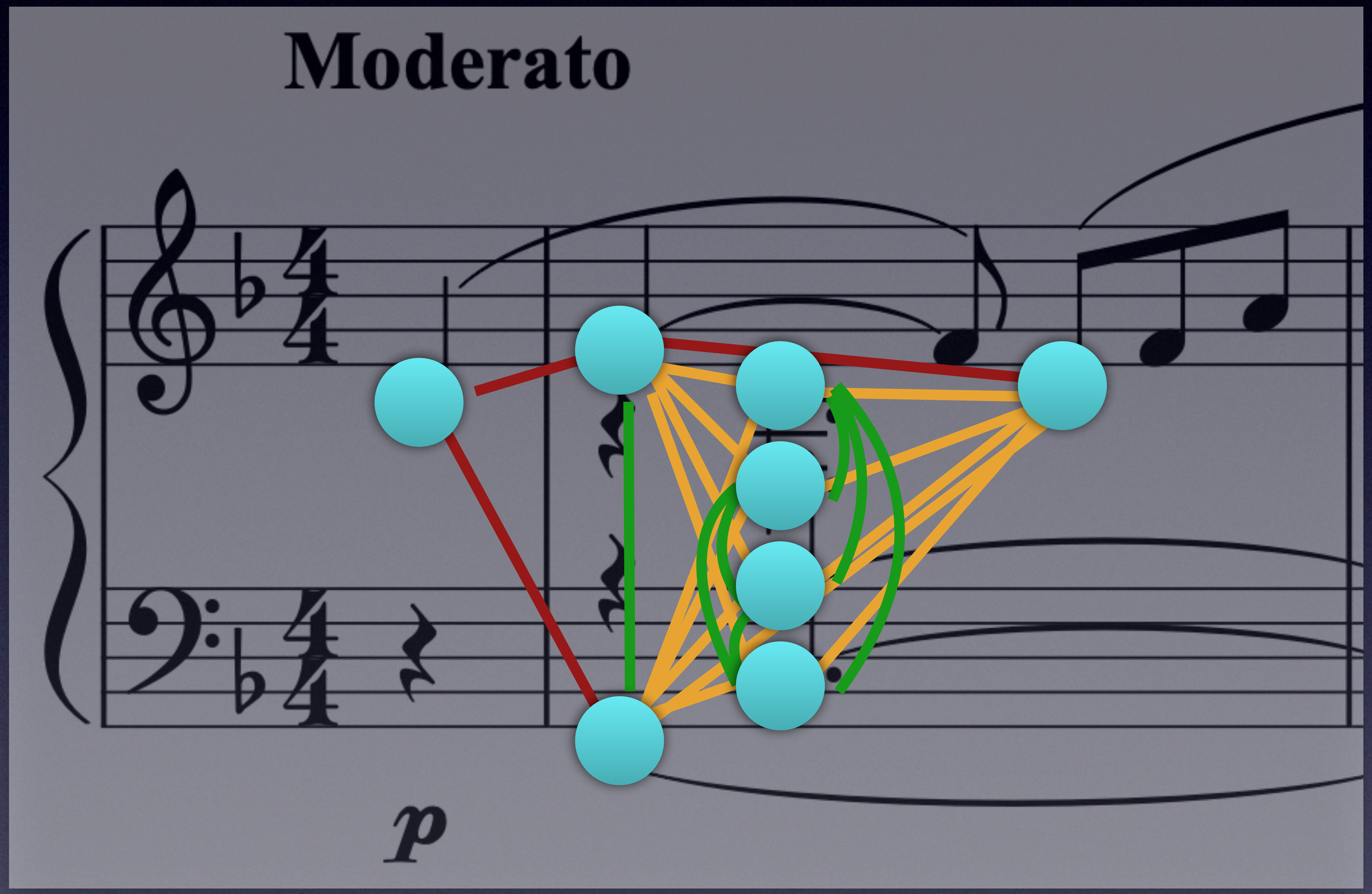
Our Idea: Music Score as Graph

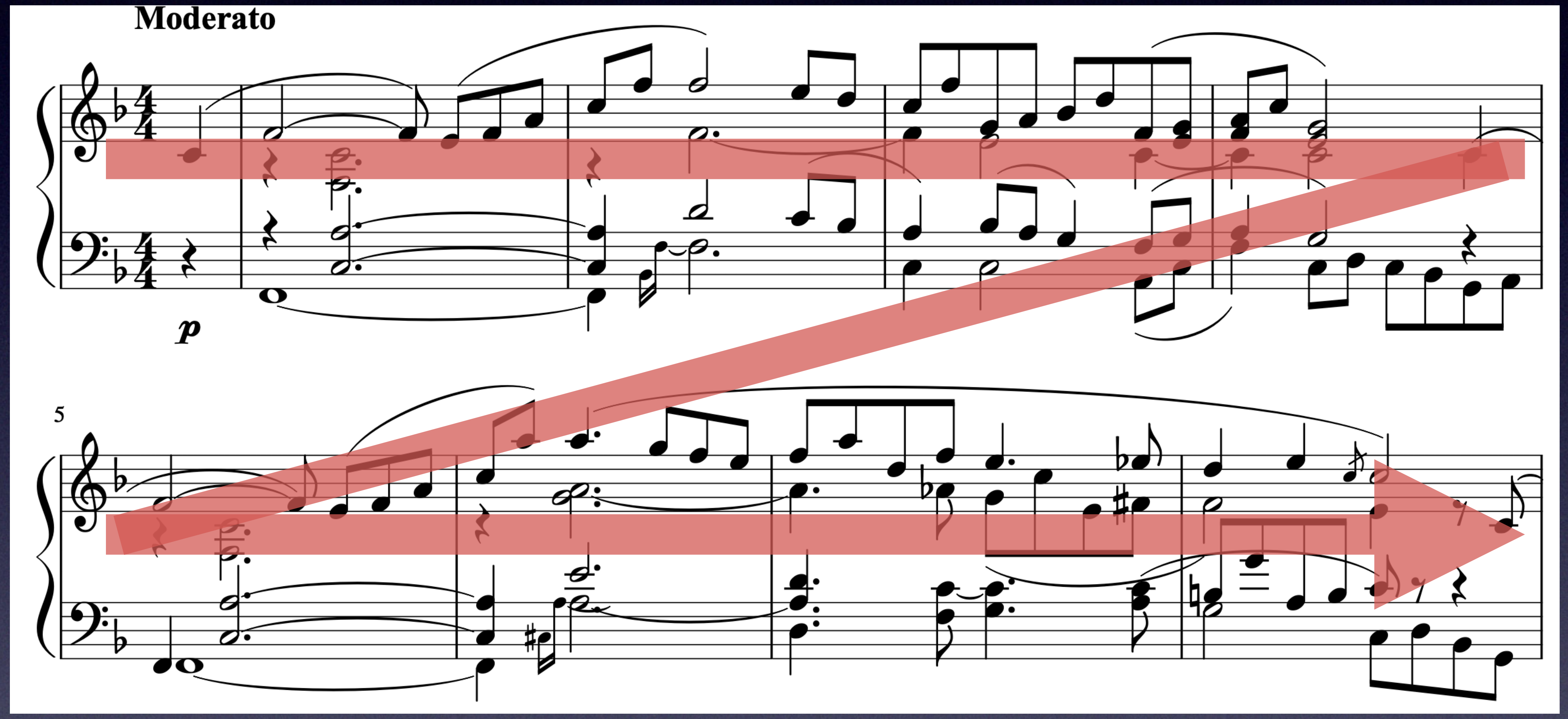




- Each note is considered as a graph node.
- Neighboring notes are connected by different types of edges
- Gated Graph Neural Network (GGNN)

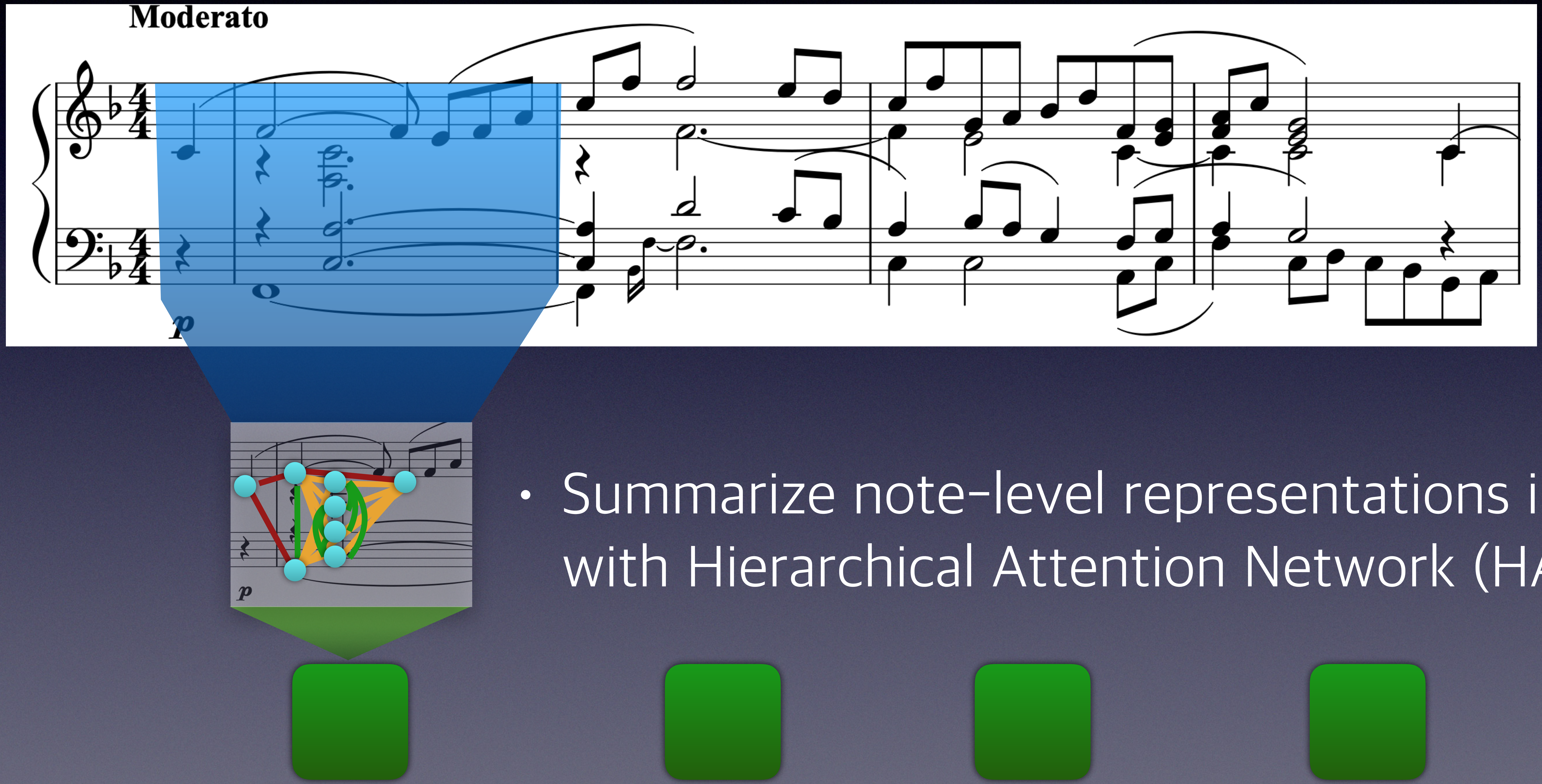
Music in Extended Context





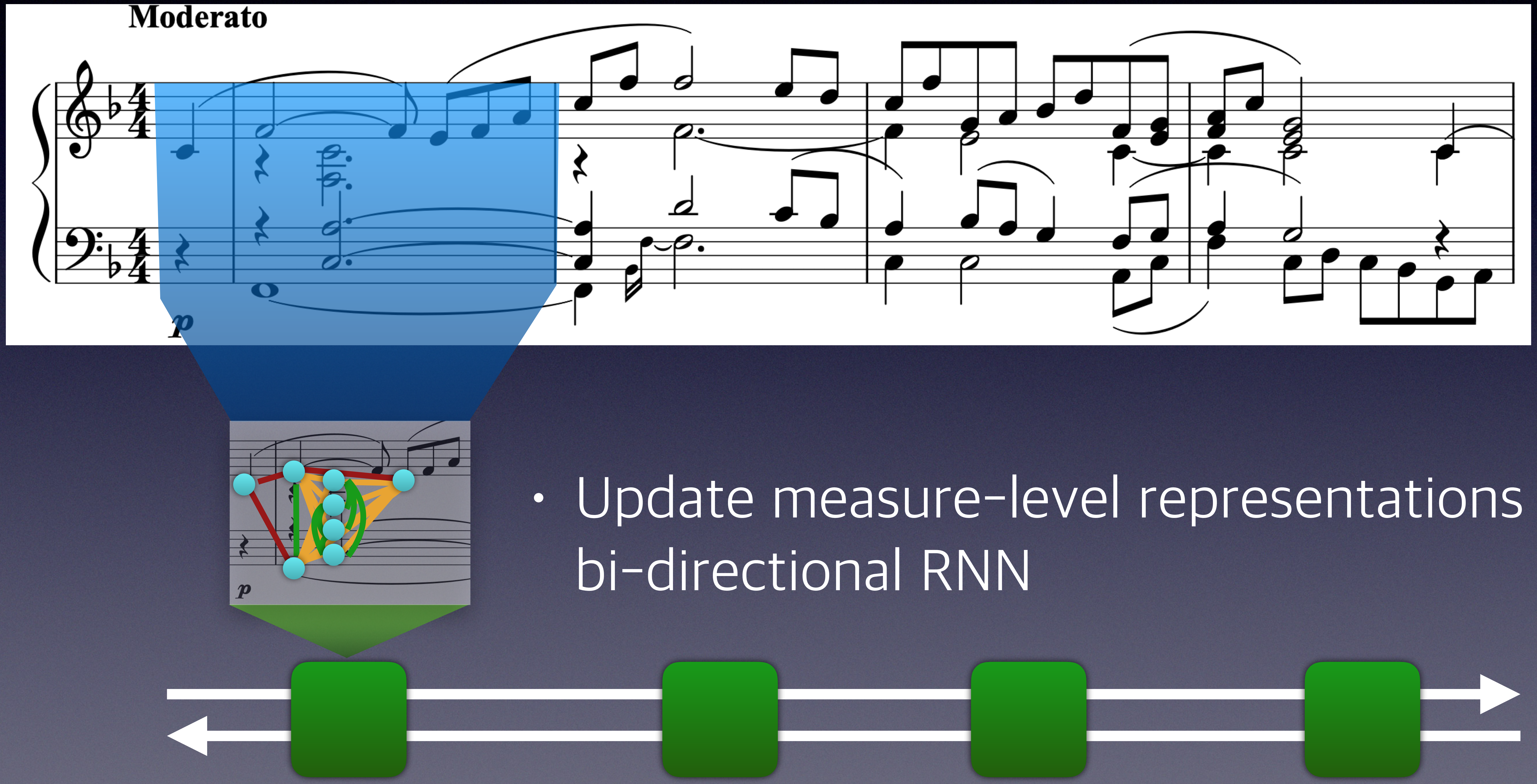
- GNN is suitable for handling the local context of each note.
- But music has sequence-like characteristics in extended context

Combining GNN and RNN



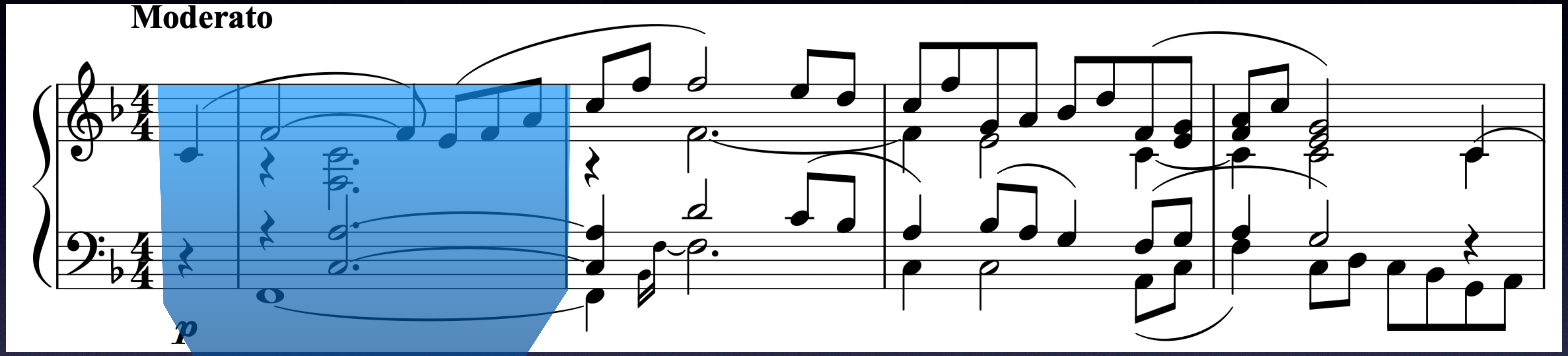
- Summarize note-level representations in a measure with Hierarchical Attention Network (HAN)

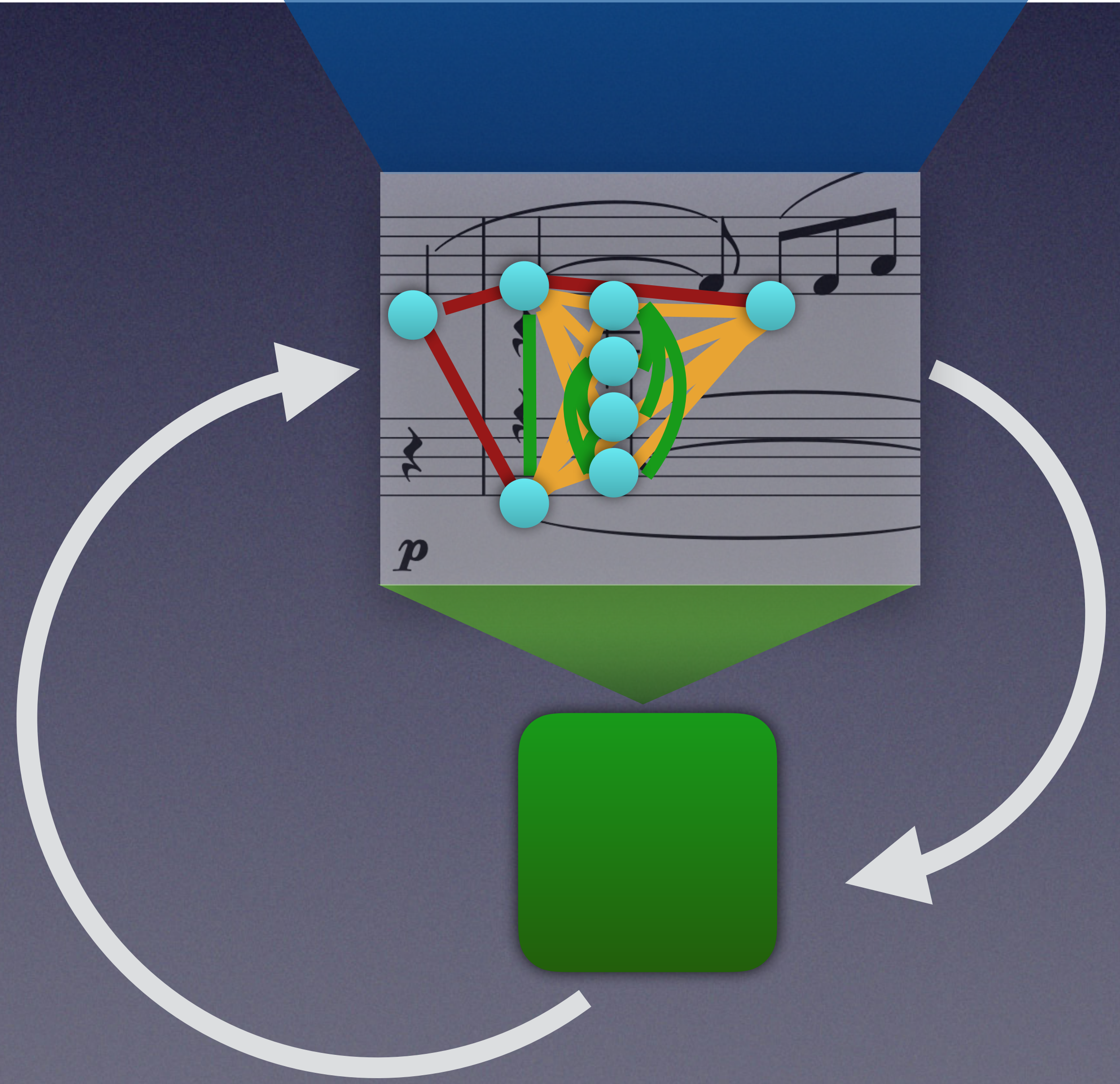
Iterative Update



- Update measure-level representations with bi-directional RNN

Iterative Update



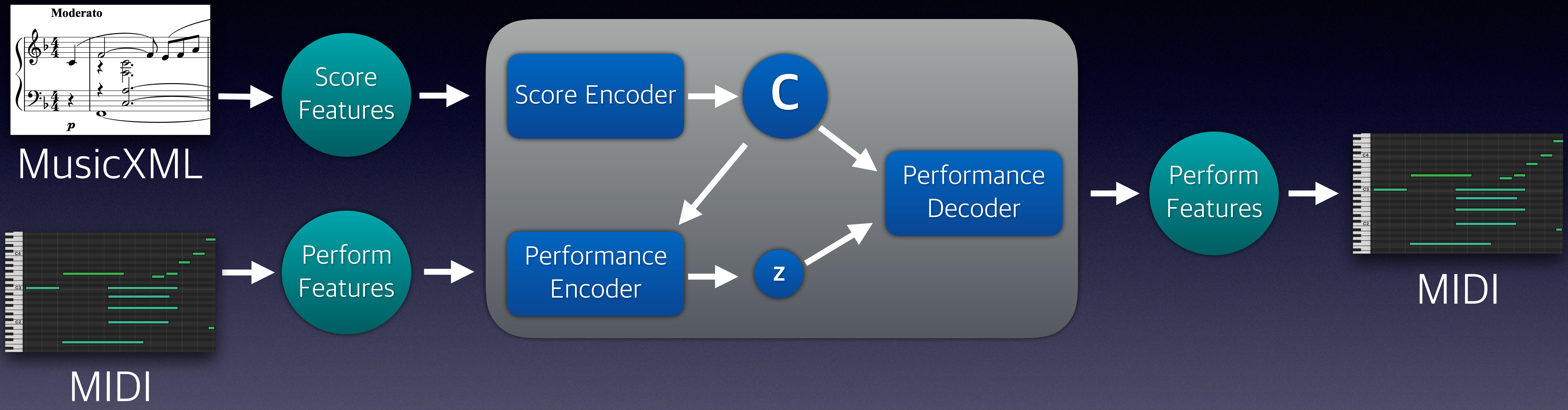


- Feed measure-level representations back into note-level representations
- Update note-level and measure-level representation again

Advantage of Iterative Update

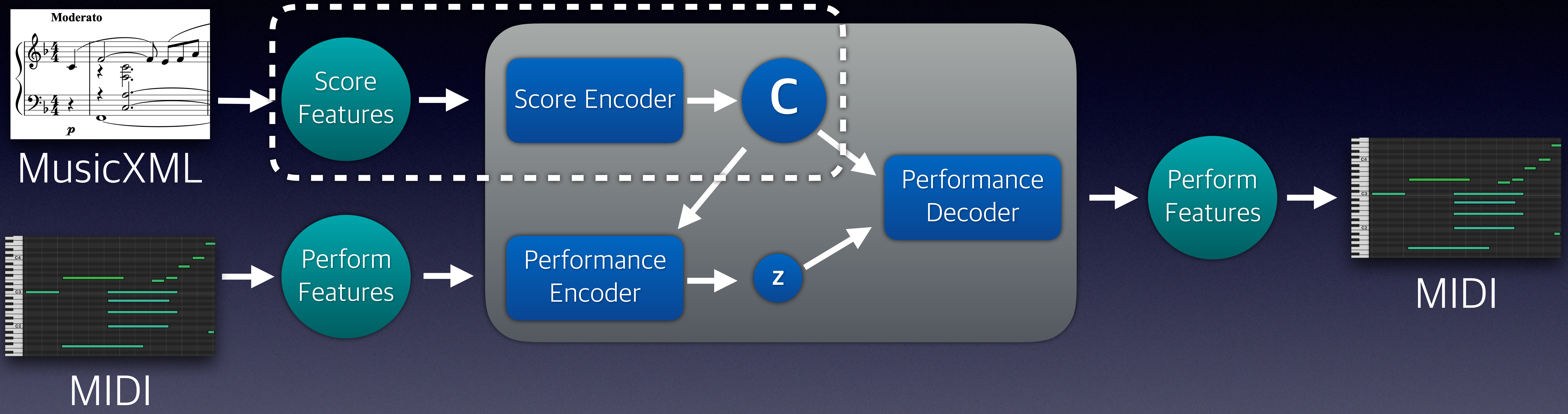
- Note-level representations can be updated considering the extended context
- It can compensate the lack of auto-regressive decoding in GGNN
  - Unlike RNN with sequence data, GNN cannot fix the output because of cyclic connection
- Named Iterative Sequential Graph Network (ISGN)

Performance Modeling System



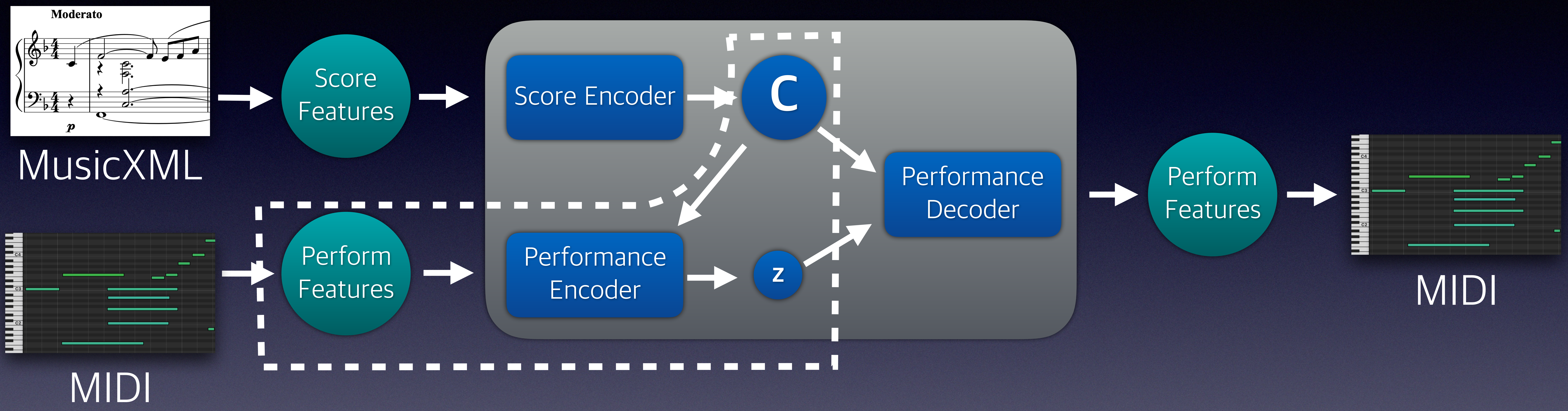
- Conditional Variational Autoencoder (CVAE)
- Takes music score and (optional) performance MIDI
- Input and output is a sequence of in note-level score and performance features

Performance Modeling System



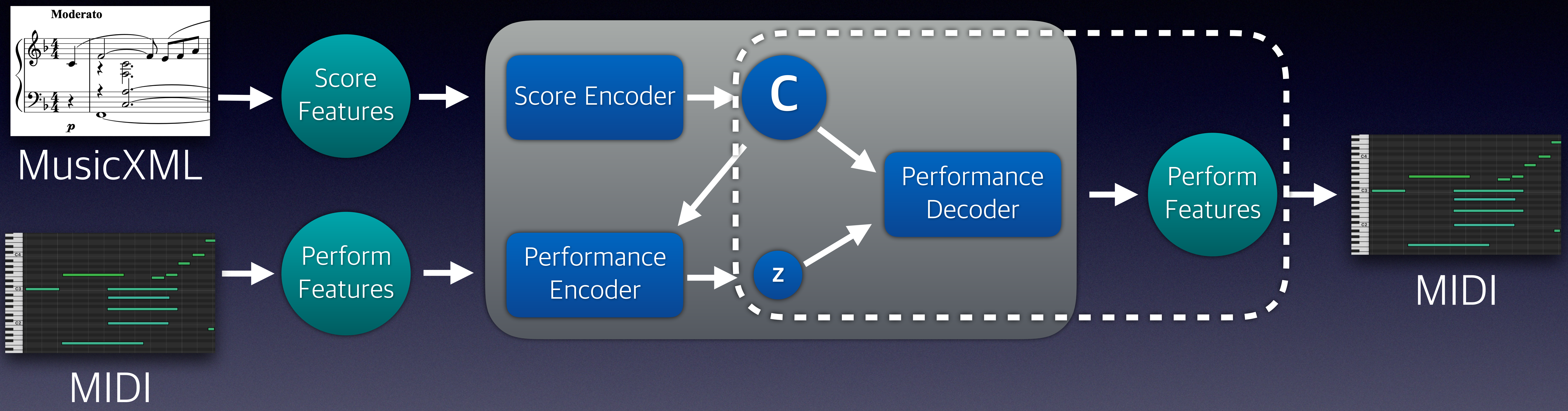
- Score Encoder takes score inputs and embeds it as a score condition **C**
- **C** is a sequence of note-level hidden representations.

Performance Modeling System



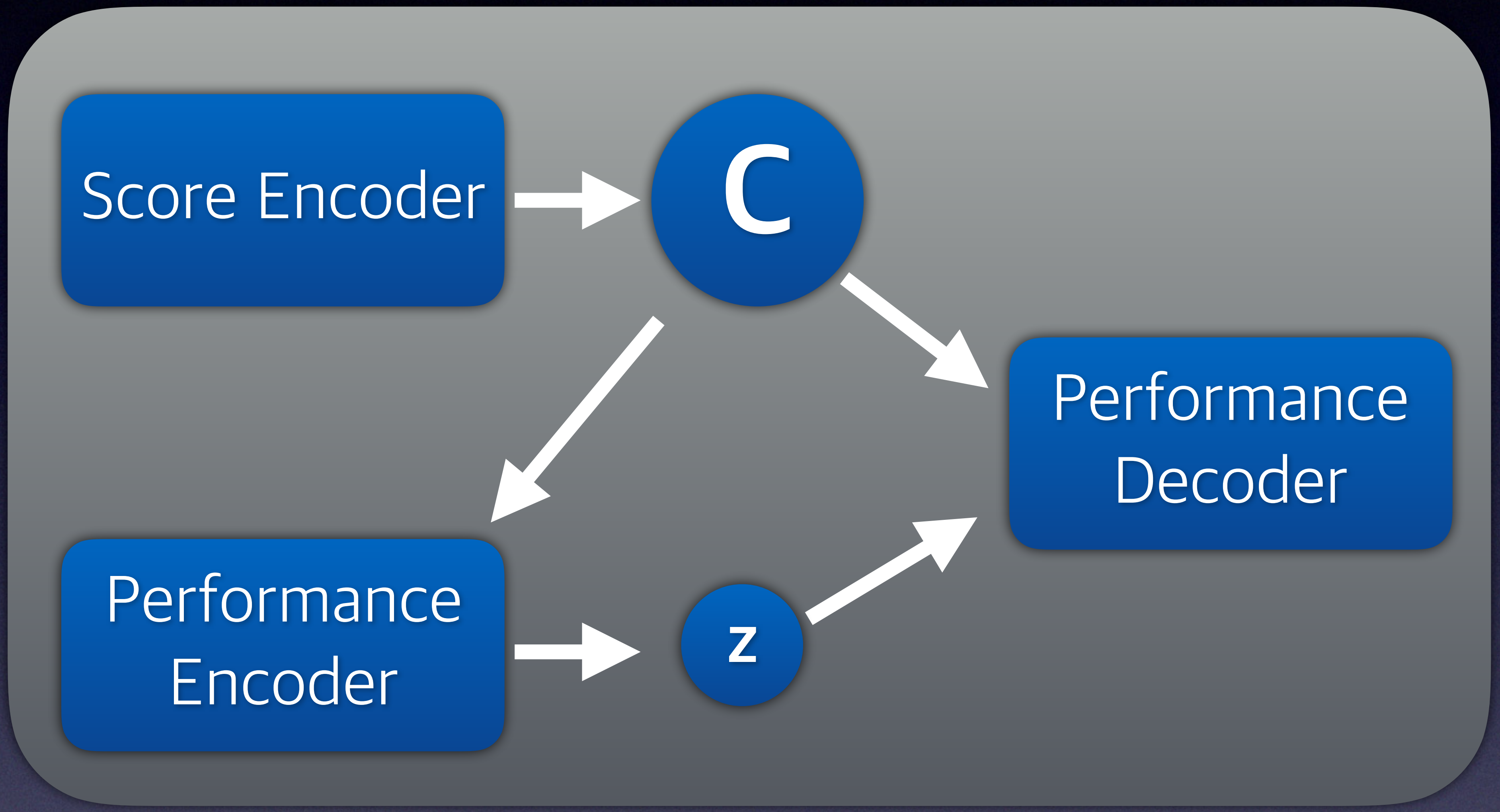
- Performance Encoder takes performance features and score condition as inputs and encode the probability of **z**
- **z** is a single vector that can be regarded as a 'performance style vector'

Performance Modeling System



- Performance decoder takes score condition **C** and performance style vector **z** and reconstructs the performance features.

Experiment

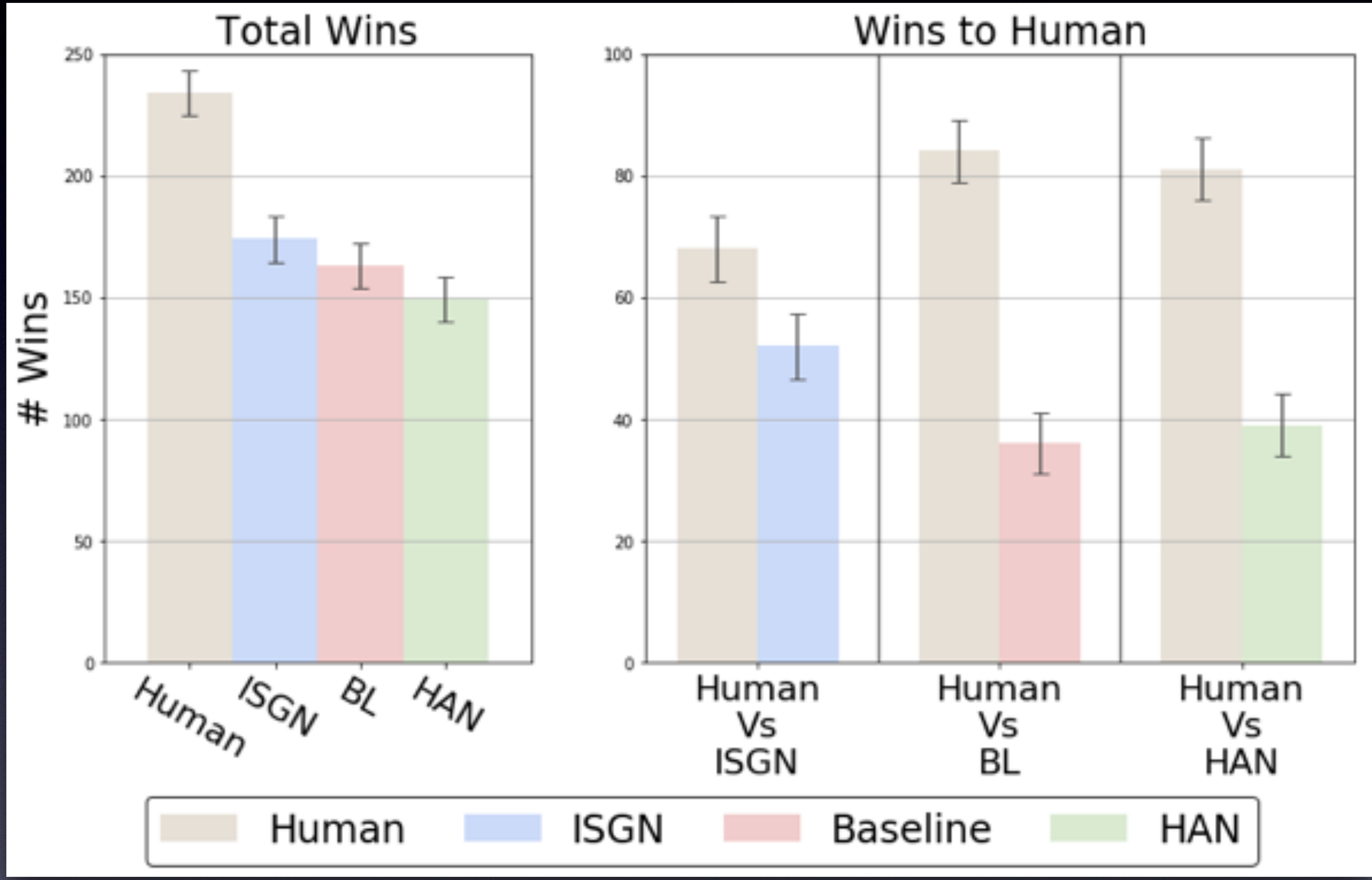


- Trained 4 models with same module structure but different NN architecture.
  - Baseline: Note-level LSTM only
  - HAN: Note-level LSTM, beat-level LSTM, measure-level LSTM
  - G-HAN: Note-level GGNN, beat-level LSTM, measure-level LSTM
  - Proposed: Note-level and measure-level ISGN

Experiment Result

Model	Tempo	Vel	Dev	Pedal	KLD
BL	0.2721	0.6011	**0.7678**	0.8056	**2.2581**
HAN	0.2380	0.6290	0.7938	0.7681	13.666
G-HAN	0.2785	0.6212	0.7705	0.8092	7.1113
Proposed	**0.2379**	**0.5877**	0.7978	**0.7544**	3.7247

Reconstruction loss on test set



Human listening test

- The proposed model showed better result than other models

https://github.com/jdasam/virtuosoNet

MACLab KAIST