

A close-up, low-angle shot of a man with a beard and mustache playing a double bass. He is looking down at the instrument, which has six strings. The lighting is warm and focused on his face and the instrument, creating a dramatic effect against a dark background.

MINGUS - Melodic Improvisation Neural Generator Using Seq2seq

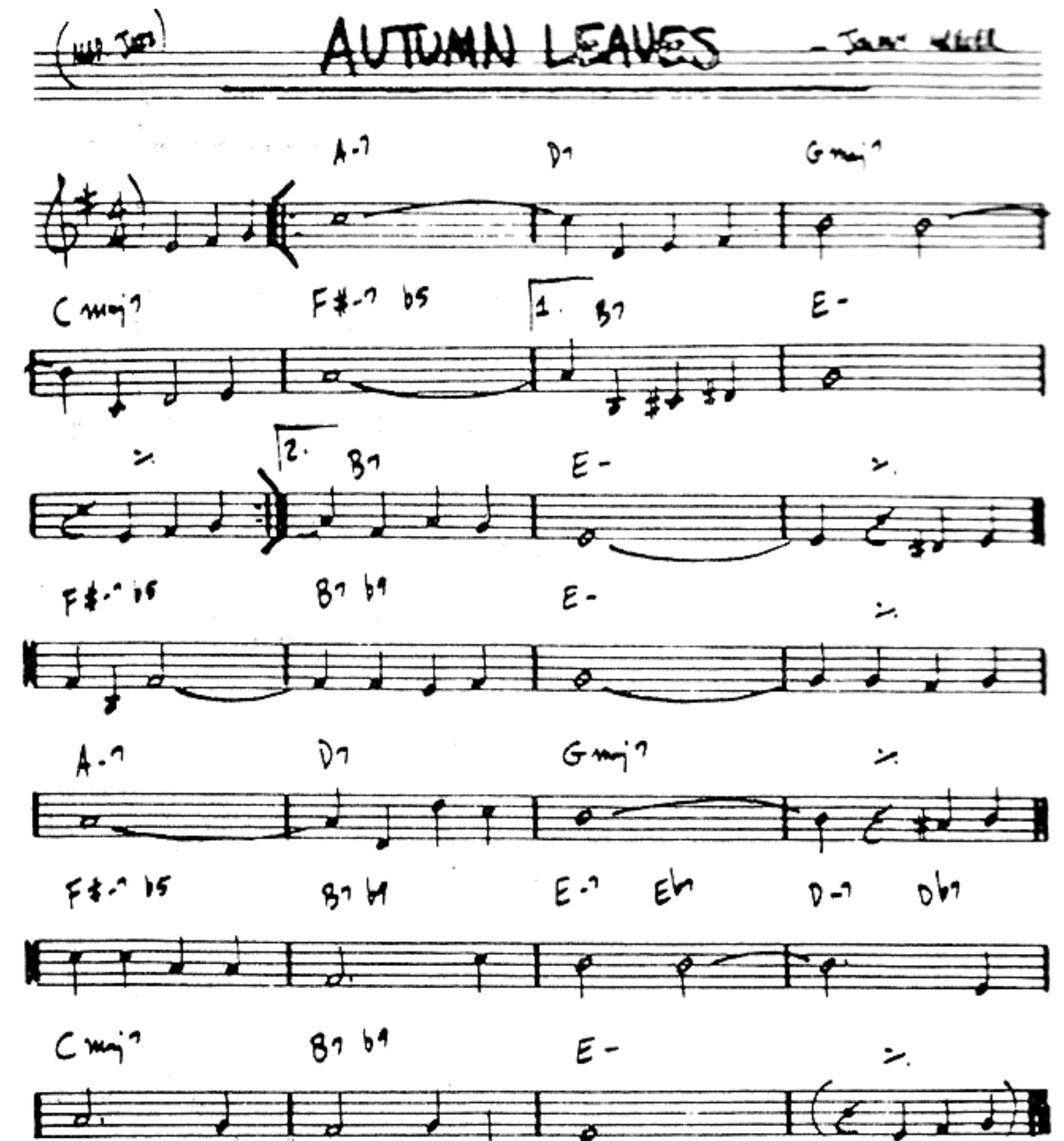
Fall 2020 project, Vincenzo Madaghiele

Outline

- Introduction
- State of the art
 1. Discovering Music Relations with Sequential Attention (Jiang, Xia, Berg-Kirkpatrick, 2020)
 2. Explicitly Conditioned Melody Generation: A Case Study with Interdependent RNNs (Genchel, Pati, Lerch, 2019)
- Approach
- Results
- Conclusions

Jazz standard structure

A jazz improvisation is a live sequence generation with constraint



Challenges of jazz music

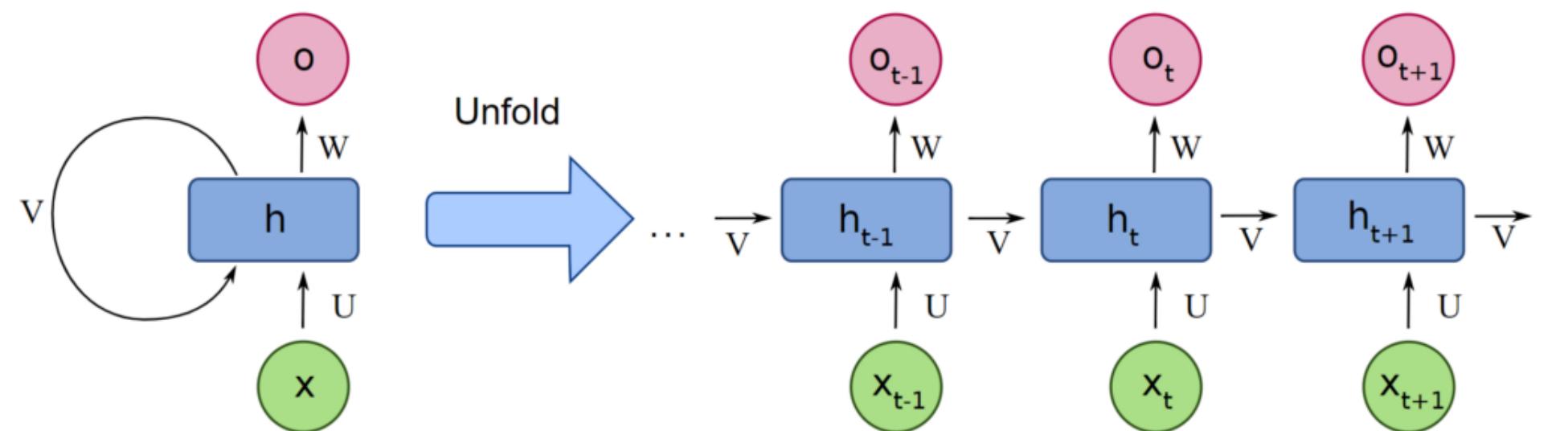
- Uneven time divisions (triplets, quintuplets...)
- Different time signatures
- Complex chord structures (many extensions)
- Complex harmony

State of the art

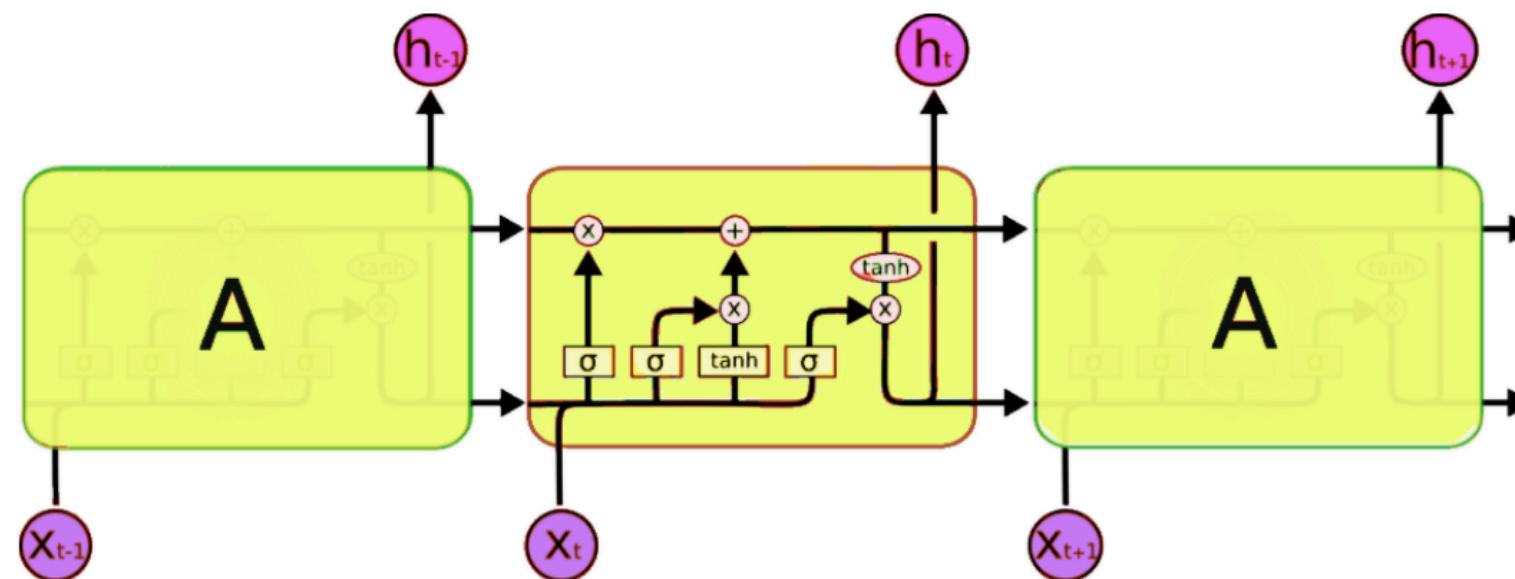
RNNs vs Transformer

RNN vs Transformer

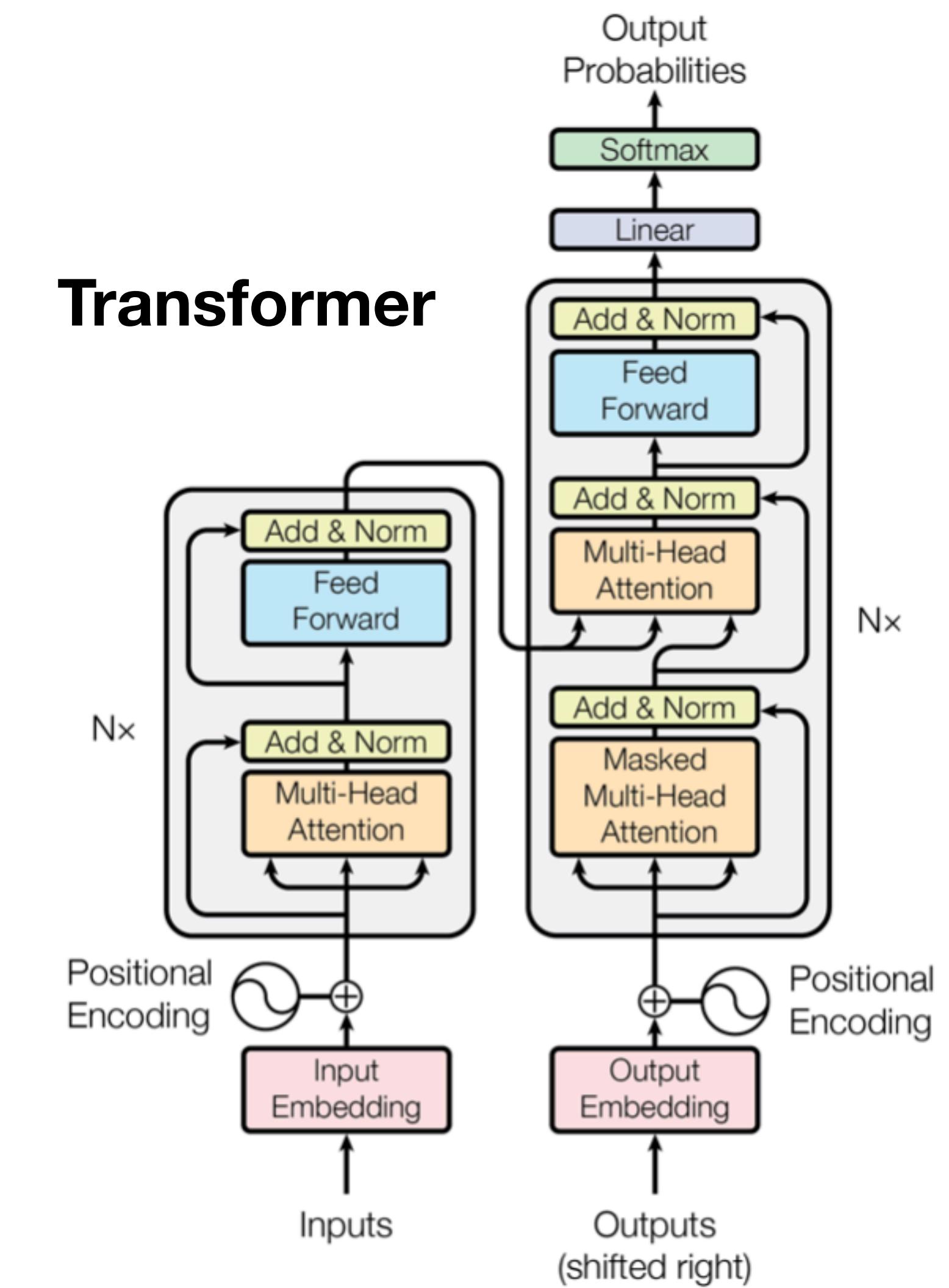
Recurrent Neural Network (RNN)



Long Short Term Memory (LSTM)



Transformer



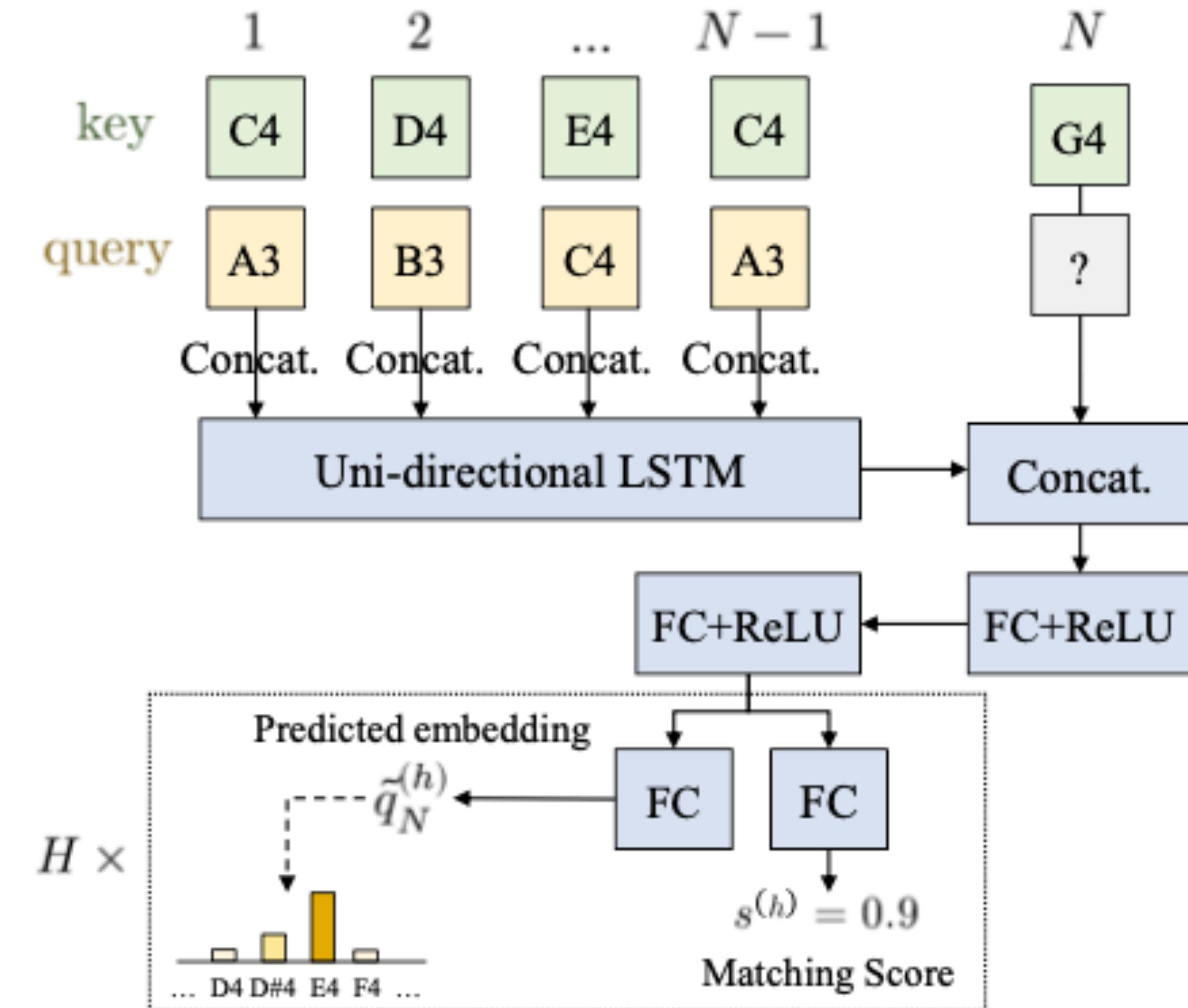
State of the art

1. Discovering Music Relations with Sequential Attention (Jiang, Xia, Berg-Kirkpatrick, 2020)

Data representation

Input Sequence	
(1)	A4 (s) B4 (s) C5 (s) G4 (s) F4 (s) (s) (s) E4 (s) (s) (s) A4 (s) B4 (s) C5 (s) G4 (s) F4 (s) (s) (s) ?
(2)	G4 (s) A4 (s) G4 (s) F4 (s) E4 (s) D4 (s) C4 (s) (s) (s) F4 (s) G4 (s) F4 (s) E4 (s) D4 (s) C4 (s) ?
(3)	C4 (s) D4 (s) E4 (s) F4 (s) G4 (s) E4 (s) C4 (s) G3 (s) D4 (s) E4 (s) F#4 (s) G4 (s) A4 (s) F#4 (s) D4 (s) ?

Sequential Attention unit



State of the art

2. Explicitly Conditioned Melody Generation: A Case Study with Interdependent RNNs (Genczel, Pati, Lerch, 2019)

Data representation

To be estimated

- Pitch sequence
- Duration sequence



Conditioning data

- Current chord (C)
- Next Chord (N)
- Bar-position sequence (B)
- Inter-conditioning between Pitch and Duration (I)

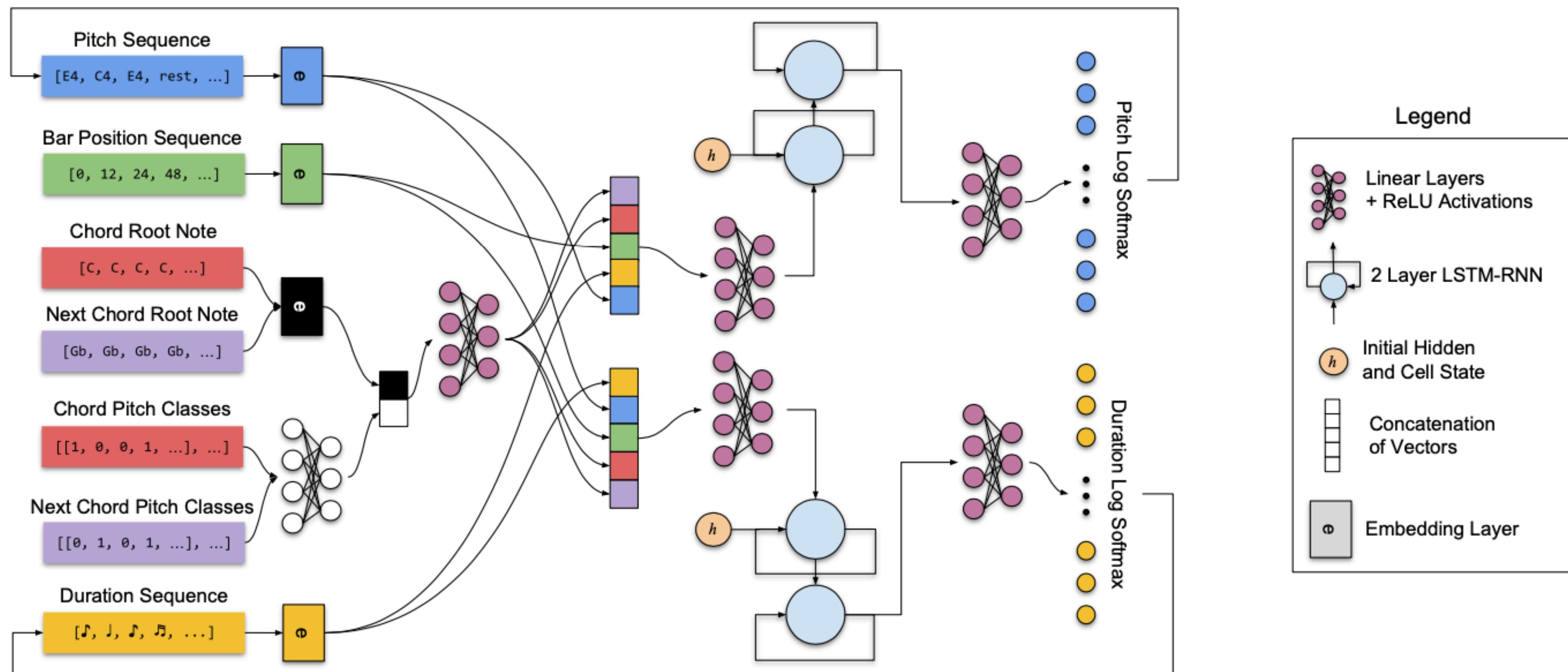
$$P = \{\text{rest}, \text{G4}, \text{Bb4}, \text{D5}, \text{rest}, \text{C5}, \text{Bb4}, \text{C5}, \text{A4}, \text{A4}, \text{F4}, \text{C4}, \text{Eb4}, \text{G4}\}$$

$$D = \{\text{J}, \text{J}, \text{J}\}$$

$$B = \{0, 12, 24, 36, 48, 60, 84, 0, 12, 24, 36, 42, 48, 60\}$$

$$C = \{\{\text{G}, [1,0,0,1,0,0,0,1,0,0,0,0]\}, \dots, 7 \text{ times}, \{\text{F}, [1,0,0,0,1,0,0,1,0,0,0,0]\}, \dots, 7 \text{ times}\}$$

Model structure



Approach

The experiment

Datasets

Weimar Jazz DB

Score Title Composer

Melody

Copyright

Nottingham DB

Score Title Composer

Track 0

Copyright

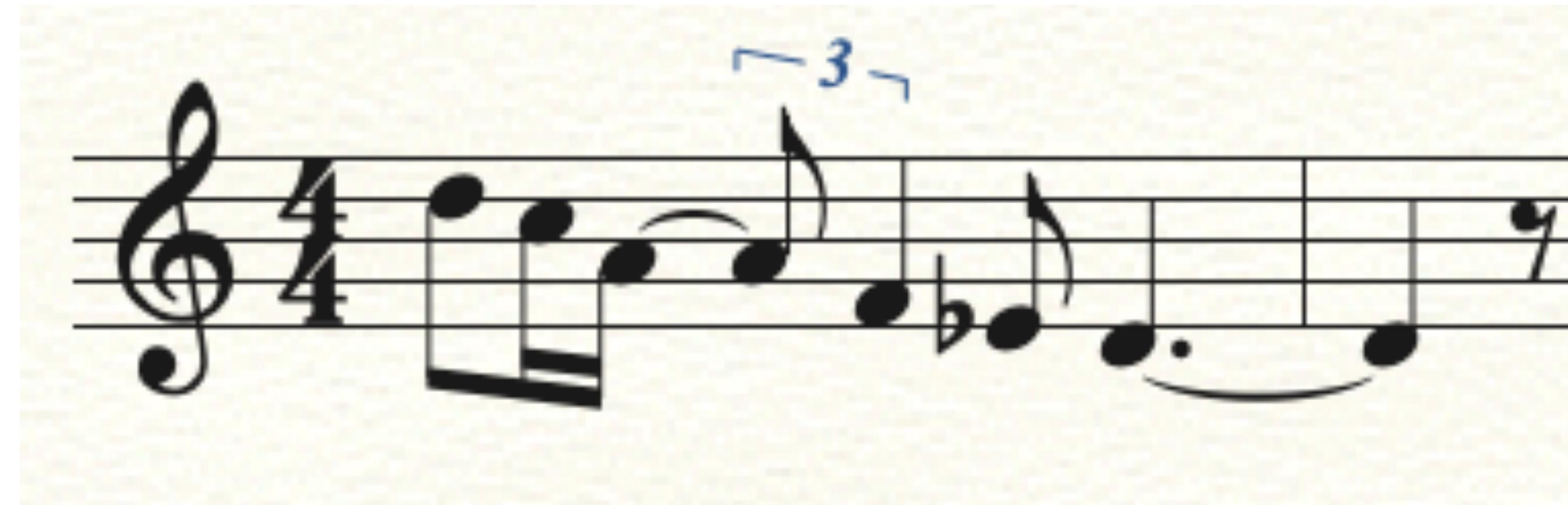
Folk DB

Score Title Composer

Track 0

Copyright

Data representation



Example melody, from Miles Davis improvisation on Oleo, Weimar Jazz DB

$$P = [74, 72, 69, 65, 63, 62, R]$$

$$D = [16th, 16th, 8th \text{ note triplet}, \text{dot } 16th, 8th \text{ note triplet}, \text{dot quarter}, 8th]$$

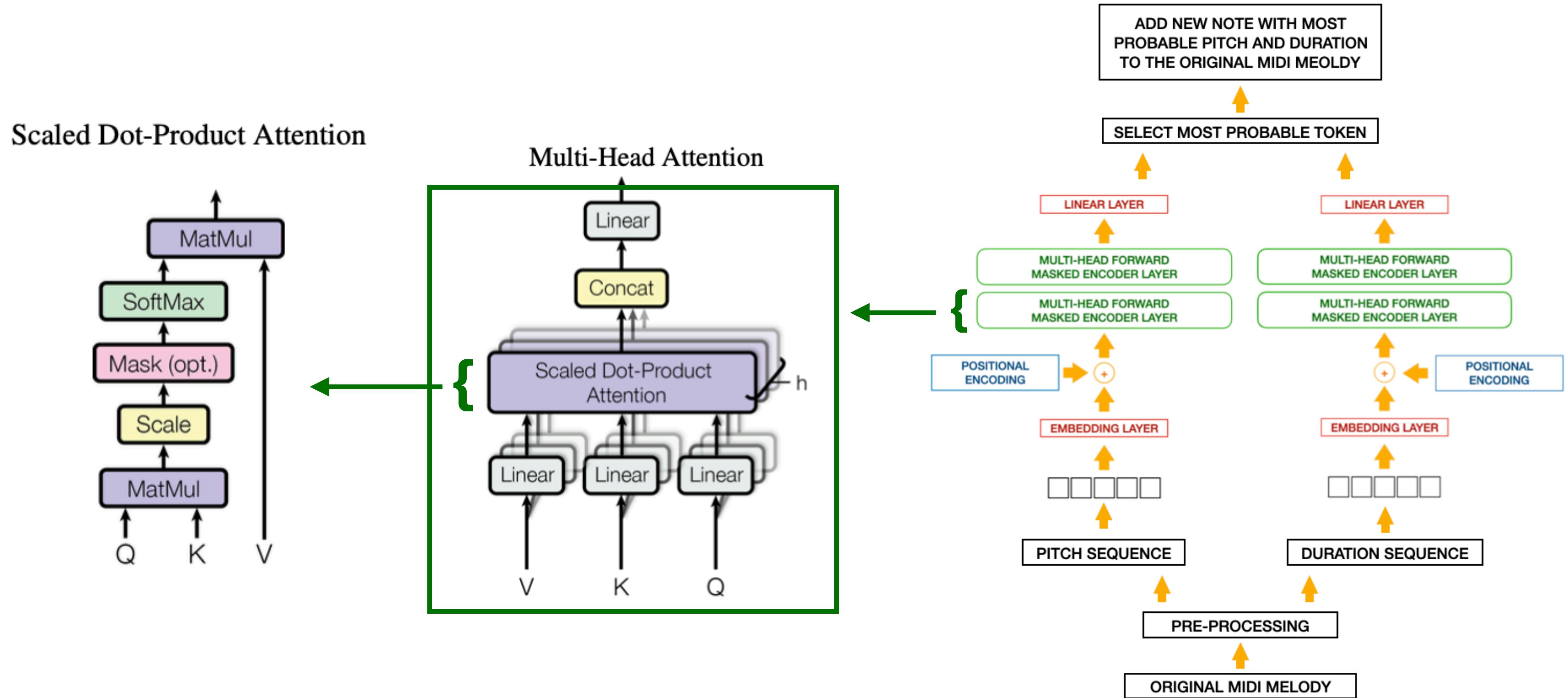
Melody segmentation

The image shows three staves of musical notation, likely from a jazz improvisation. The top staff is labeled "Melody". The middle staff begins at measure 5. The bottom staff begins at measure 9. Each staff is divided into segments by colored boxes. The first segment (measures 4-7) is highlighted with a red box. The second segment (measures 5-7) is highlighted with a green box. The third segment (measures 7-8) is highlighted with an orange box. The fourth segment (measures 9-10) is highlighted with a blue box. The fifth segment (measures 10-11) is highlighted with a yellow box. Each segment contains a blue bracket labeled "-3-", indicating a specific melodic pattern or segment length.

Example of melody segmentation, from Charlie Parker improvisation on Donna Lee, Weimar Jazz DB

MINGUS architecture

MINGUS



Results

Metrics comparison

Accuracy and Perplexity

Perplexity	MINGUS	SeqAttn
Dataset	pitch	duration
Weimar jazz	13.2	2.3
Nottingham	7.39	2.09
folkDB	7.02	1.86

Accuracy [%]	MINGUS	SeqAttn
Dataset	pitch	duration
Weimar jazz	13.92	58.62
Nottingham	28.91	75.46
folkDB	30.44	81.04

MGEval

MGEval	MINGUS		ECMG	
Measure	KL div	overlap area	KL div	overlap area
total used pitch	0.011	0.770	0.025	0.888
total pitch class histogram	0.069	0.707	0.242	0.686
pitch class transition matrix	0.137	0.813	0.026	0.714
pitch range	0.033	0.757	0.019	0.849
avg IOI	0.074	0.765	0.533	0.480

MGEval comparison
between MINGUS
and ECMG on folkDB

MGEval	Weimar jazz DB		Nottingham DB	
Measure	KL div	overlap area	KL div	overlap area
total used pitch	0.068	0.408	0.045	0.824
total pitch class histogram	0.059	0.280	0.151	0.718
pitch class transition matrix	0.186	0.413	0.577	0.711
pitch range	0.045	0.48	0.057	0.786
avg IOI	0.004	0.778	0.008	0.916

MGEval metrics by
MINGUS on
Weimar Jazz DB
and Nottingham DB

Generation Demo



Example of melody generation, from John Coltrane
improvisation on Mr P.C., Weimar Jazz DB

Conclusions

Discussion and future work

Discussion

Conclusions

- State of the art generation performance
- Good performance on complex music
- Motif-level coherence
- Highly adaptable architecture

Limitations

- No significant improvement with respect to RNN-based models
- Long-term coherence
- Pitch-class transition

Future work

- Conditional models
- Live interface
- Long-term coherence

Thank you!

Vincenzo Madaghiele
(vincenzo.madaghiele@eurecom.fr)

Code, datasets and generated music examples are available at:
<https://github.com/vincenzomadaghiele/MINGUS>