Vincenzo Madaghiele

MINGUS - Melodic Improvisation Neural Generator Using Seq2seq

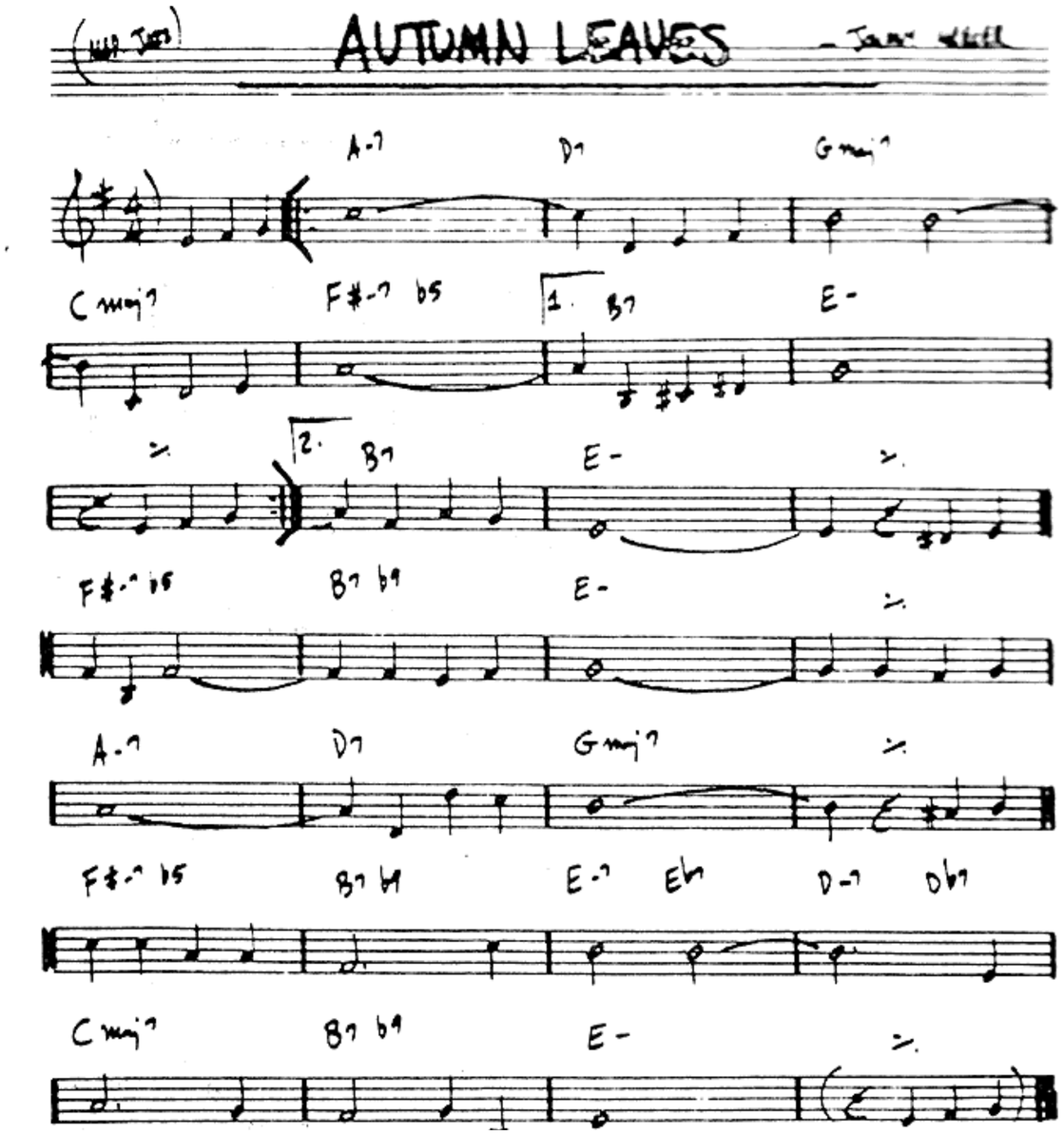
Fall 2020 project

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

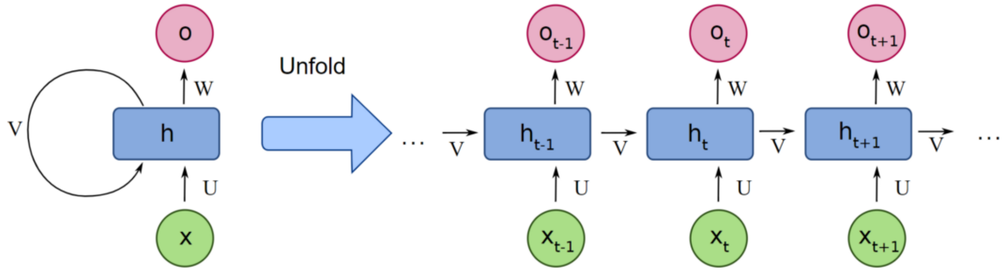
- Uneven time divisions (triplets, quintuplets…)
- Different time signatures
- Complex chord structures (many extensions)
- Complex harmony

State of the art

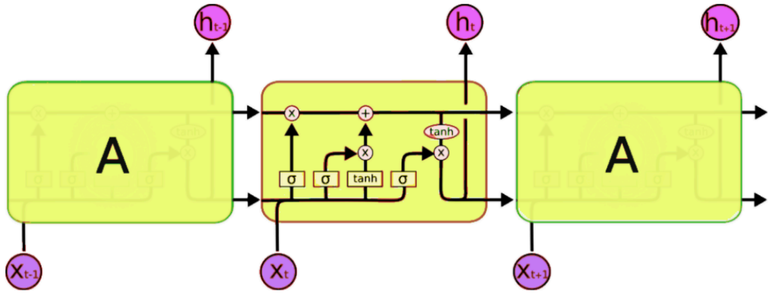
RNNs vs Transformer

RNN vs Transformer

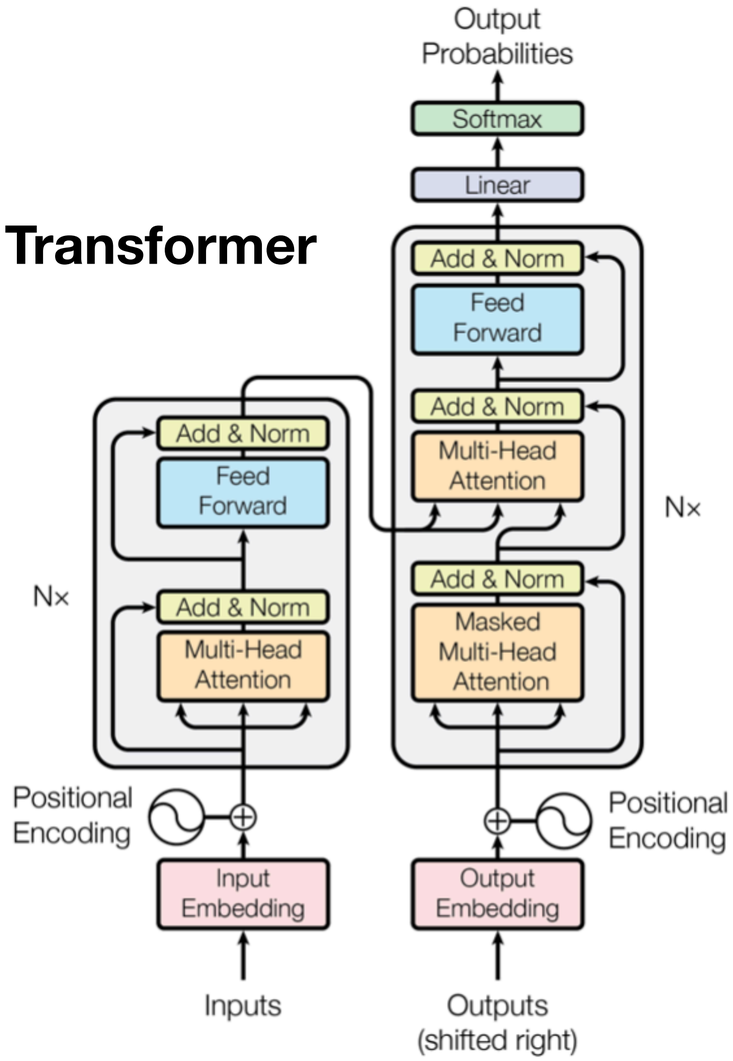
Recurrent Neural Network (RNN)



Long Short Term Memory (LSTM)



Transformer



State of the art

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

Sequential Attention unit

$$[s^{(1...H)}; \tilde{q}_N^{(1...H)}] = \text{SeqAttn}(\mathbf{q}_{1...N-1}, \mathbf{k}_{1...N}, e) \quad (1)$$

$$h_N = \text{LSTM}(f_1, f_2, ..., f_{N-1}) \quad (2)$$

$$[s^{(1...H)}; \tilde{q}_N^{(1...H)}] = \text{MLP}([h_N; k_N]) \quad (3)$$

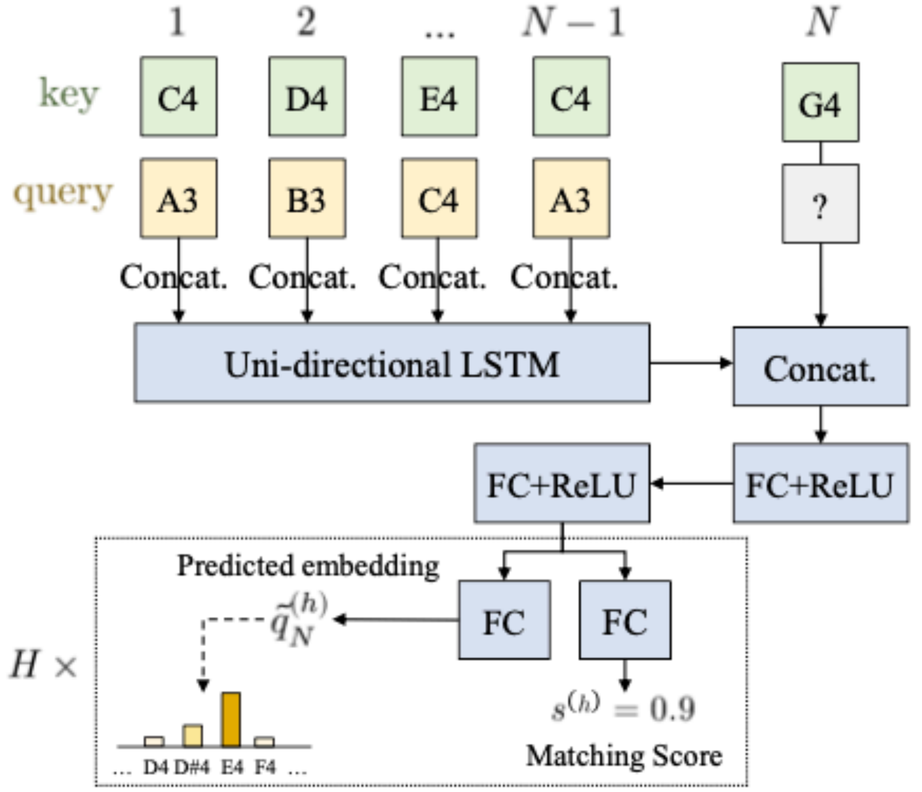
Self-Attention Layer

$$[s_i^{(1..H)}; \tilde{x}_{N,i}^{(1..H)}] = \text{SeqAttn}(\mathbf{x}_{1...N-1}, \mathbf{x}_{1-i...N-i}, e_i) \quad (4)$$

$$\hat{s}_i^{(h)} = \frac{\exp(s_i^{(h)})}{\sum_{i'} \exp(s_{i'}^{(h)})} \quad (5)$$

$$\tilde{x}_N^{(h)} = \sum_i \hat{s}_i^{(h)} \tilde{x}_{N,i}^{(h)} \quad (6)$$

$$\tilde{x}_N = \text{Linear}([\tilde{x}_N^{(1)}; ...; \tilde{x}_N^{(H)}]) \quad (7)$$



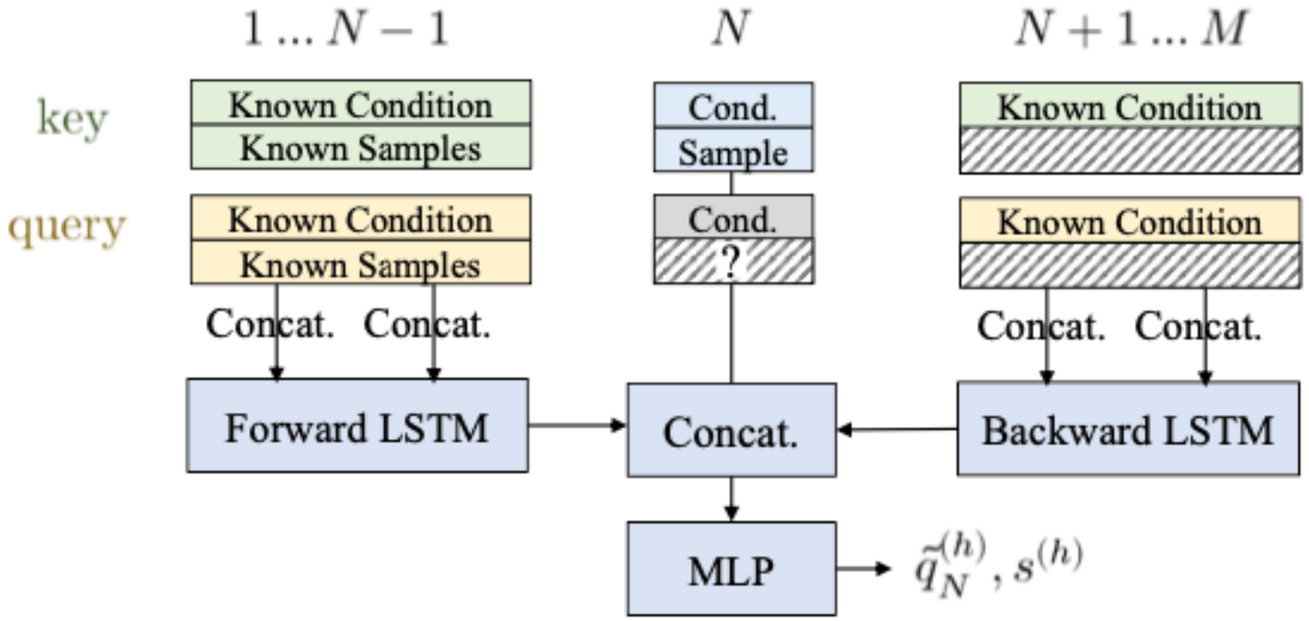
Conditional sequential attention

Used for chord conditioning

$$\overrightarrow{h}_N = \mathrm{LSTM}_{\mathrm{fw}}(f_1, f_2, ..., f_{N-1}) \quad (8)$$

$$\overleftarrow{h}_N = \mathrm{LSTM}_{\mathrm{bw}}(b_M, b_{M-1}, ..., b_{N+1}) \quad (9)$$

$$[s^{(1...H)}; \tilde{q}_N^{(1...H)}] = \mathrm{MLP}([\overrightarrow{h}_N; \overleftarrow{h}_N; k_N; q_N^c]) \quad (10)$$



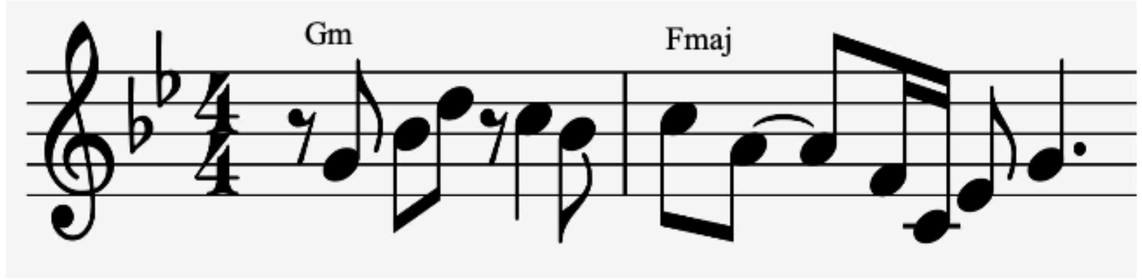
State of the art

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

Data representation

To be estimated

- Pitch sequence
- Duration sequence

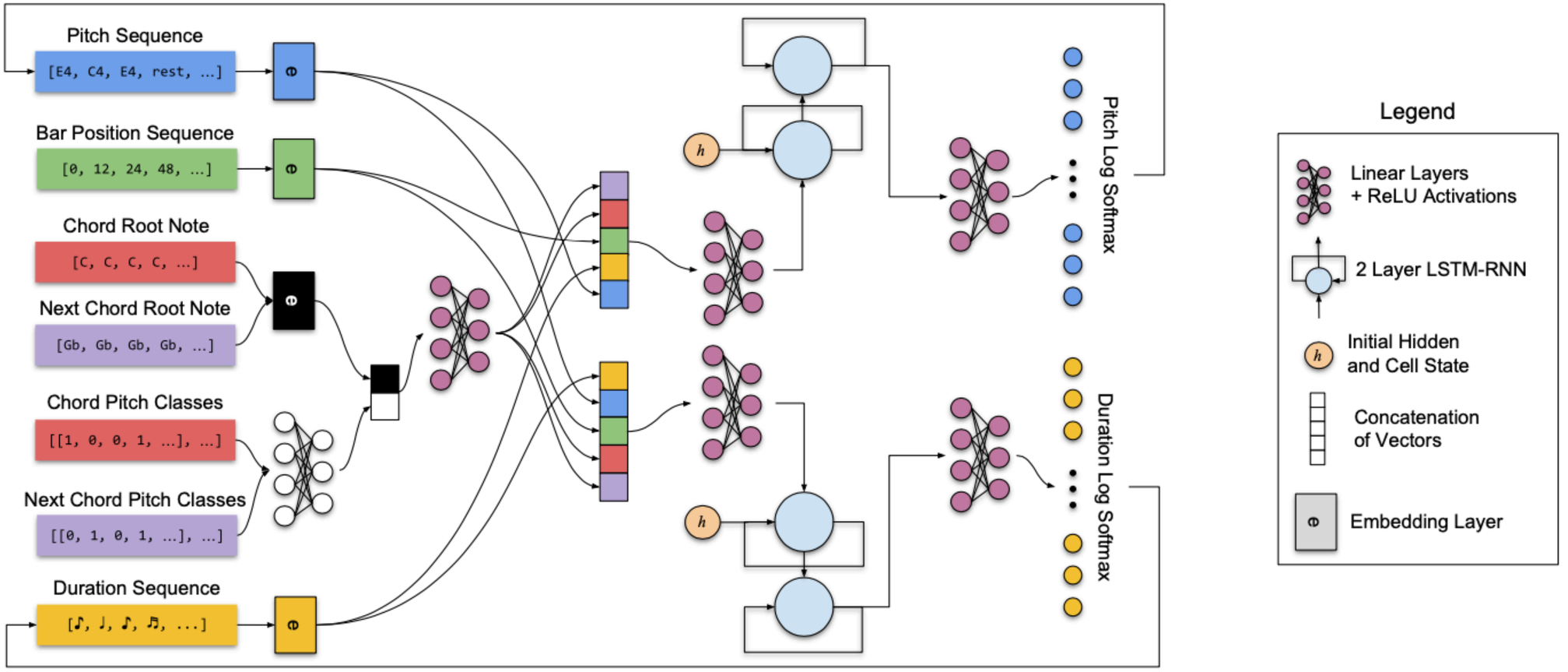


P = {rest, G4, Bb4, D5, rest, C5, Bb4, C5, A4, A4, F4, C4, Eb4, G4}
D = {♪, ♪, ♪, ♪, ♪, ♩, ♪, ♪, ♪, ♪, ♬, ♬, ♪, ♩.}
B = {0, 12, 24, 36, 48, 60, 84, 0, 12, 24, 36, 42, 48, 60}
C = {{G,[1,0,0,1,0,0,0,1,0,0,0,0]},...7 times, {F,[1,0,0,0,1,0,0,1,0,0,0,0]},...7 times}

Conditioning data

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

Model structure

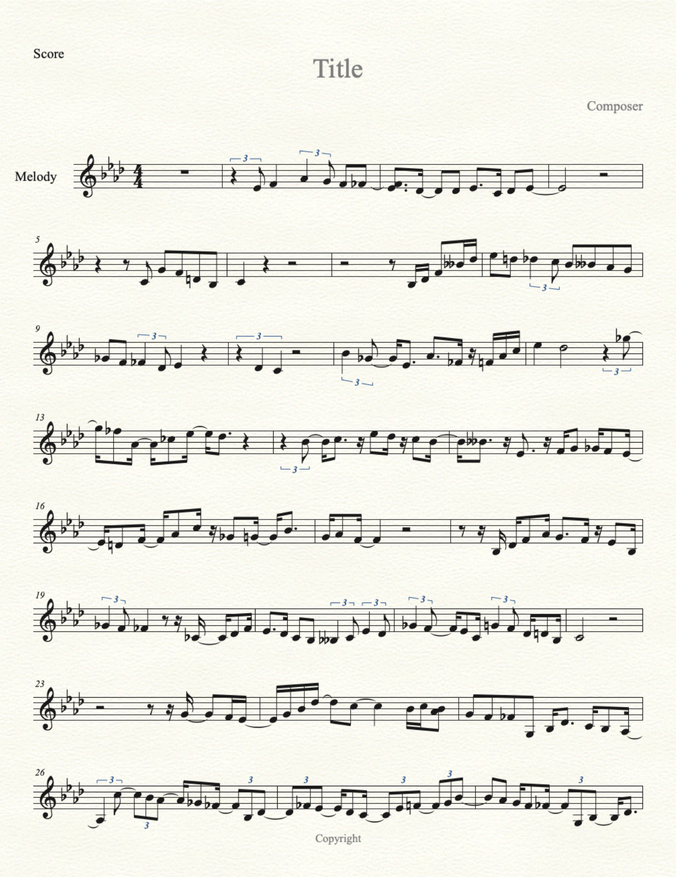


Approach

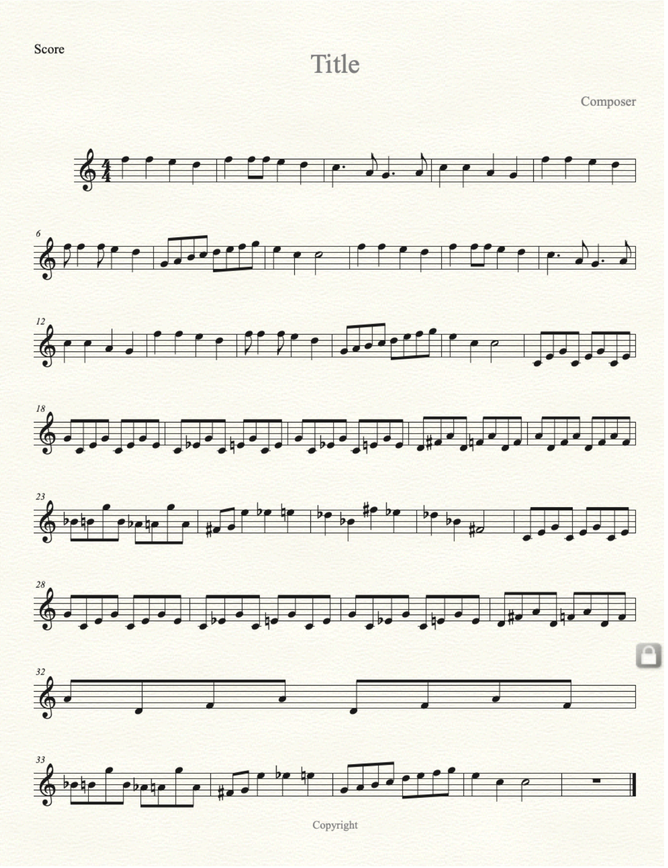
The experiment

Datasets

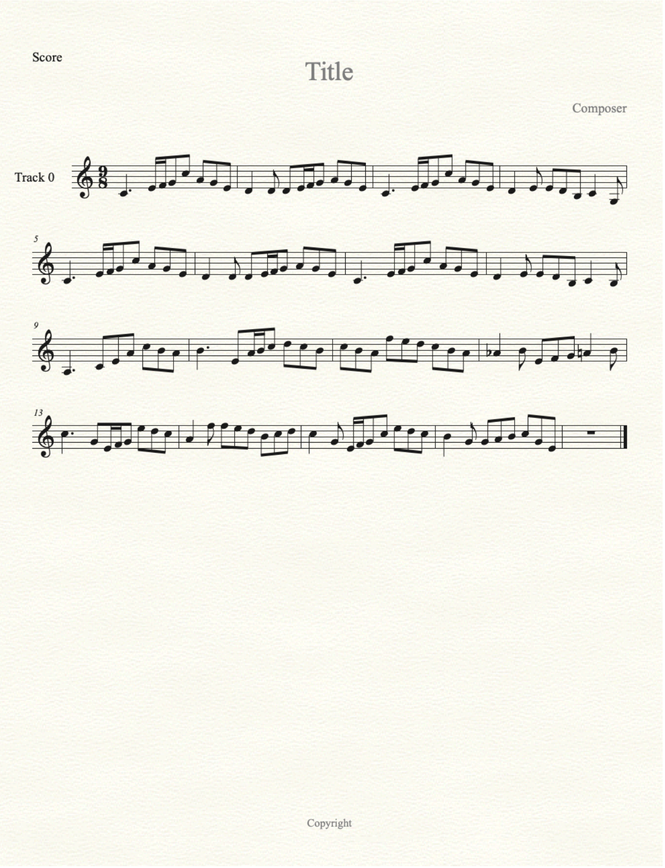
Weimar Jazz DB



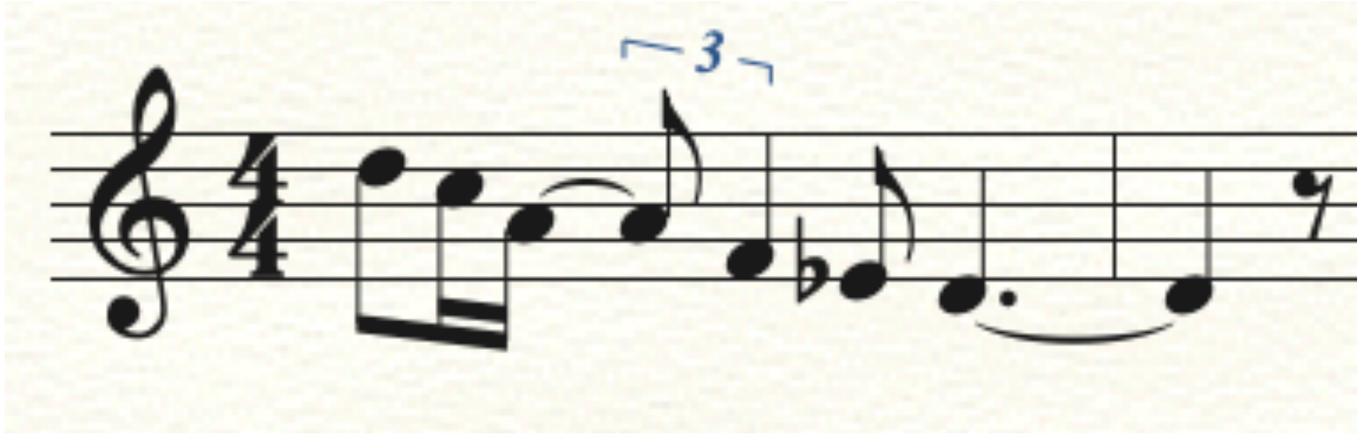
Nottingham DB



Folk DB



Data representation

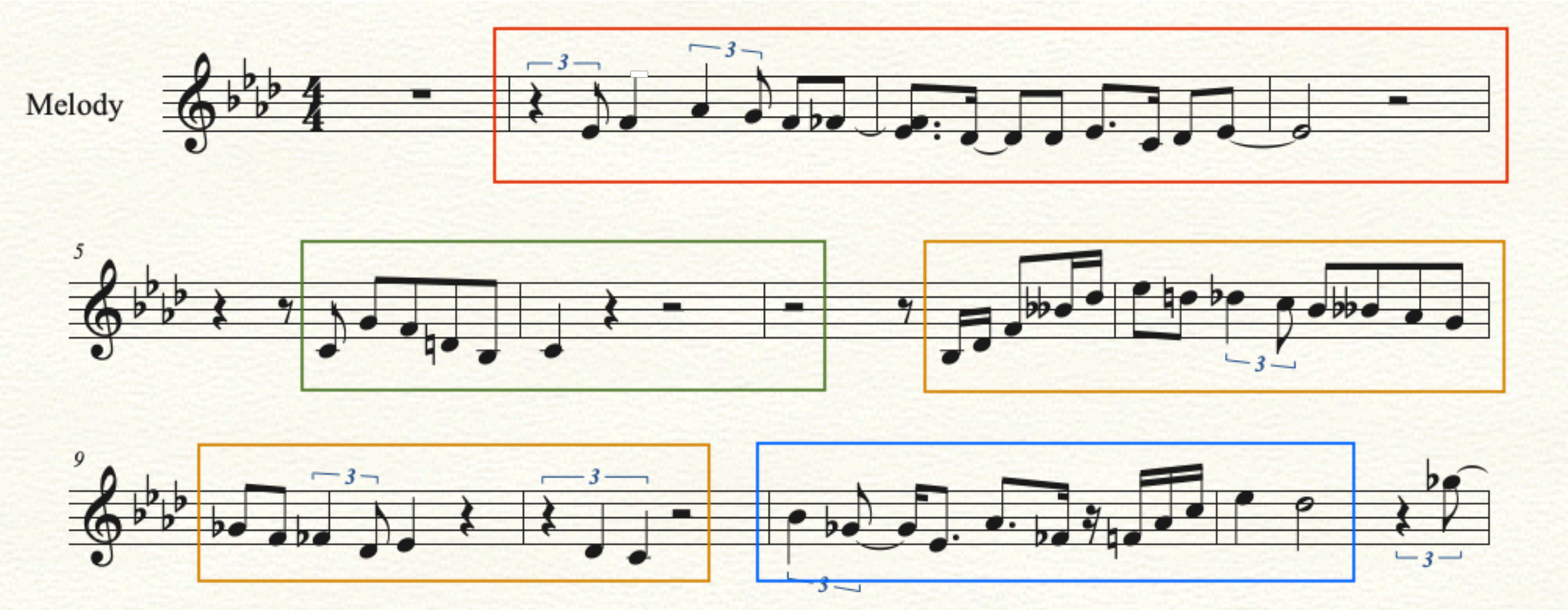


Example melody, from Miles Davis improvisation on Oleo, Weimar Jazz DB

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

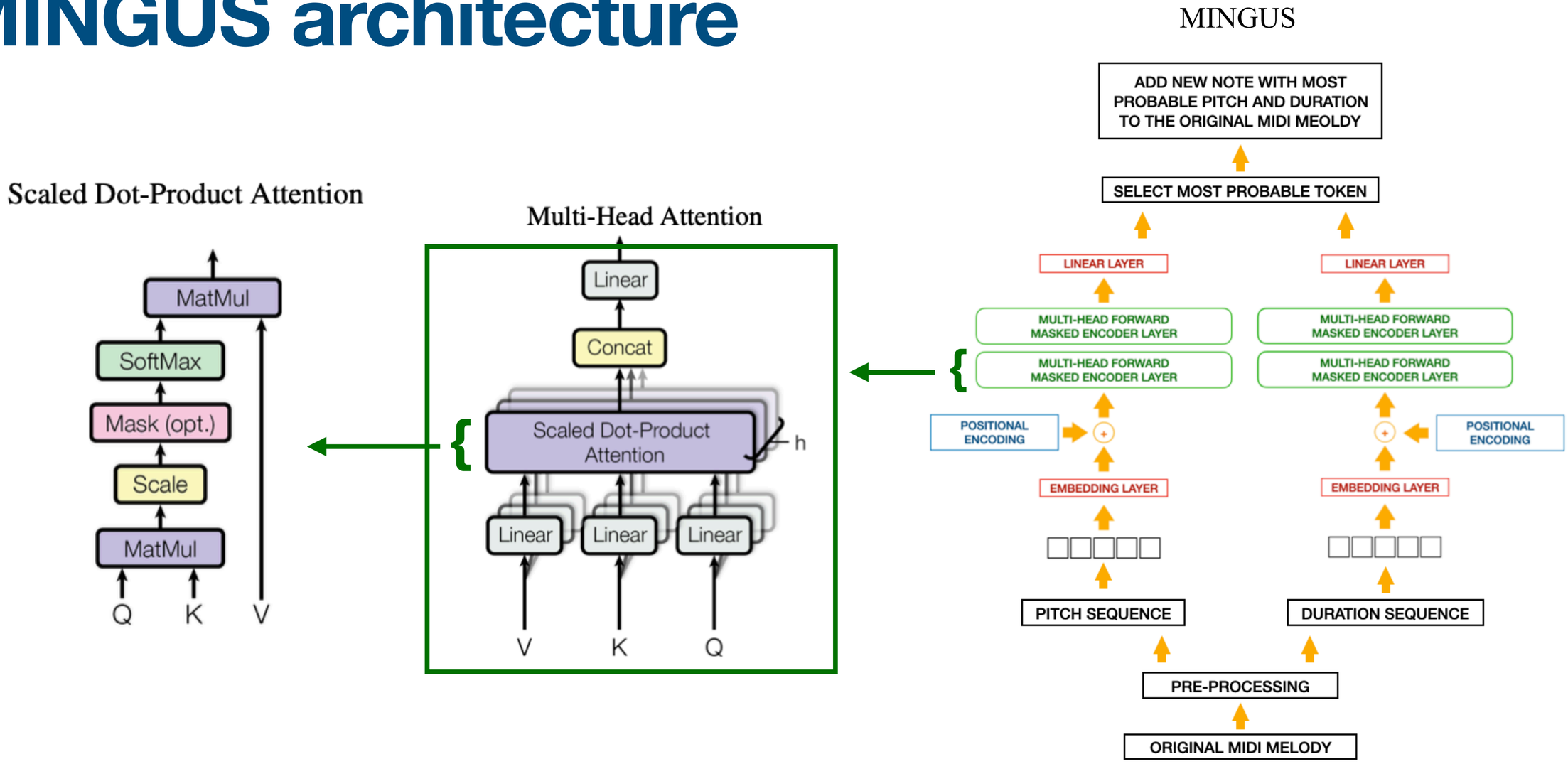
$$D = [16th,\ 16th,\ 8th\ note\ triplet,\ dot\ 16th,\ 8th\ note\ triplet,\ dot\ quarter,\ 8th]$$

Melody segmentation



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

MINGUS architecture



Results

Metrics comparison

Accuracy and Perplexity

Perplexity	MINGUS		SeqAttn
Dataset	**pitch**	**duration**	
Weimar jazz	13.2	2.3	6.29
Nottingham	7.39	2.09	1.54
folkDB	7.02	1.86	-

Accuracy [%]	MINGUS		SeqAttn
Dataset	**pitch**	**duration**	
Weimar jazz	13.92	58.62	47.26
Nottingham	28.91	75.46	88.23
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

Conclusions

Discussion and future work

Discussion and future work

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

Thank you!

Vincenzo Madaghiele
(vincenzo.madaghiele@eurecom.fr)

Code, datasets and generated music examples are available at:

https://github.com/vincenzomadaghiele/MINGUS