MINGUS: MELODIC IMPROVISATION NEURAL GENERATOR USING SEQ2SEQ –SUPPLEMENTARY MATERIAL–

Vincenzo Madaghiele **Pasquale Lisena** **Raphaël Troncy**
EURECOM, Sophia Antipolis, France
[vincenzo.madaghiele, pasquale.lisena, raphael.troncy]@eurecom.fr

1. SUMMARY

This supplementary material contains:

- This document reporting supplementary metrics result
- Working code used for the paper, WjazzDB converted to xml and a trained MINGUS model
- The sound samples used for blind listener evaluation

2. TRAINING PARAMETERS

Table 1 contains the parameters used for training MINGUS.

Parameter	Value
Sequence Length	35
Training batch size	20
Validation batch size	10
Testing batch size	10
Epochs	100
Hidden size	200
Encoder layers	4
Attention heads	4
Dropout	0.2
Pitch learning rate	0.5
Duration learning rate	0.05

Table 1: MINGUS parameters for pitch and duration network

3. EVALUATION RESULTS ON NOTTINGHAMDB

In this section we report the evaluation metrics obtained on NottinghamDB. The generation process has been the same used for generation with WjazzDB. Unfortunately, due to code incompatibilities, it was not possible to generate with BebopNet on NottinghamDB.

Table 2 shows the harmonic coherence obtained on NottinghamDB generations.

Harmonic coherence [%]	**Chord**	**Scale**
Original	75.63	95.34
MINGUS	70.32	**97.53**
BebopNet	-	-
SeqAttn	**74.06**	95.95

Table 2: Harmonic coherence on NottinghamDB

Table 6 shows the MGEval results obtained on NottinghamDB generations.

4. ABLATION STUDY

In this section we report the results of our ablation study. Out of all possible combinations of features that have been experimented, we reported here only the 9 highest scoring ones and the score obtained with no conditioning.

Table 3 summarises the features and their abbreviations used in the following tables.

Conditioning	Abbreviation
Pitch	P
Duration	D
Current chord	C
Next chord	NC
Current bass note	B
Current beat	BE
Offset in the bar	O

Table 3: Conditionings abbreviations

Table 4 reports the perplexity and accuracy obtained on different pitch models trained on WjazzDB, sorted by increasing accuracy.

Table 5 reports the perplexity and accuracy obtained on different duration models trained on WjazzDB, sorted by increasing accuracy.

Conditioning [pitch]	Perplexity	Accuracy [%]
No cond (P)	12.30	13.57
D-C-NC-B-BE-O	12.25	14.53
D-C-NC	12.19	14.54
D-B	11.90	14.60
D-C-NC-B-BE	12.02	14.61
D-BE	11.82	14.63
D-C-B	**11.71**	14.67
D-C-NC-BE-O	11.93	14.69
D-C-B-O	12.14	14.74
D-C-B-BE-O	11.96	**14.99**

Table 4: Metrics of MINGUS pitch models trained with different feature combinations on WjazzDB over 15 epochs. Only the baseline and the 9 best conditioning combinations are reported.

Conditioning [duration]	Perplexity	Accuracy [%]
No cond (D)	4.61	32.24
P-NC-B-BE-O	4.37	32.36
P-BE-O	4.37	32.36
C-BE-O	4.53	32.36
C-NC-B-BE-O	4.62	32.36
C-B-BE-O	4.39	32.39
BE-O	4.38	32.45
NC-BE-O	**4.36**	32.47
NC-B-BE-O	4.39	32.48
B-BE-O	4.38	**32.61**

Table 5: Metrics of MINGUS duration models trained with different feature combinations on WjazzDB over 15 epochs. Only the baseline and the 9 best conditioning combinations are reported.

MGEval	MINGUS		BebopNet		SeqAttn	
Measure	**KL div**	**overlap area**	**KL div**	**overlap area**	**KL div**	**overlap area**
total used pitch	**0.014**	0.805	-	-	0.084	0.860
total used note	**0.054**	0.666	-	-	0.082	0.822
avg IOI	0.036	0.860	-	-	**0.019**	0.636
avg pitch shift	0.0366	0.572	-	-	**0.031**	0.775
note length histogram	**0.006**	0.835	-	-	0.008	0.871
total pitch class histogram	0.080	0.863	-	-	**0.031**	0.869
note length transition matrix	0.077	0.745	-	-	**0.003**	0.900
pitch class transition matrix	0.101	0.853	-	-	**0.009**	0.828
pitch range	0.017	0.769	-	-	**0.010**	0.793

Table 6: MGEval comparison between MINGUS and SeqAttn on NottinghamDB