

Appendix to “*Tradformer*: A Transformer Model of Traditional Music”

April 27, 2022

This appendix describes the outcomes several rounds of human evaluation of intermediate outputs of *Tradformer* as we developed it. In each round, 10 tunes were generated at random by *folk-rnn* (v2)¹ and 10 tunes generated by *Tradformer*. These were shuffled together and the collection was given to a human evaluator to rate in terms of good tunes and bad tunes. The number of good and bad tunes generated by each system served as a comparison of quality, and importantly whether engineering choices in developing *Tradformer* were having a positive impact on its performance.

¹Using the interface <https://folkrrnn.org/>

Round One

This was the evaluation made with the model in its first incarnation, with a reduced number of layers and no particular sampling technique.

#	Model	Grade	Comment
1	folkrnn	+	but boring (too much stepwise motion)
8	folkrnn	+	boring A part
9	folkrnn	+	no comment
10	folkrnn	+	but boring, and strange notated triplet
15	folkrnn	+	but boring
16	trad	+	my favourite!! I want to drop everything and play it. This will be next week's tune on AI Frontiers
17	folkrnn	+	odd jumps, but I think this one can be nice to work with
0	folkrnn	-	no comment
3	folkrnn	-	structurally ok, but very boring
6	trad	-	no comment
11	folkrnn	-	Interesting beginning with opportunities for ornamentation. Cutting out the last 8 measures would make it far better
12	trad	-	too repetitive (B part much better)
2	trad	- -	fragmented
4	folkrnn	- -	Another humpty dumpty what?
5	trad	- -	nonsense melody
7	trad	- -	4 bar 4 has seven quavers
13	trad	- -	What?
14	trad	- -	tritone!? Poor B part
18	trad	- -	structure is bad
19	trad	- -	seven bars in the A part (one bar with 7 quavers) and nine bars in the B part

Table 1: First run evaluation. + indicates good tunes, - bad ones and - - serious failures.

0



1



2



3

3



4



5



4

6



7



8

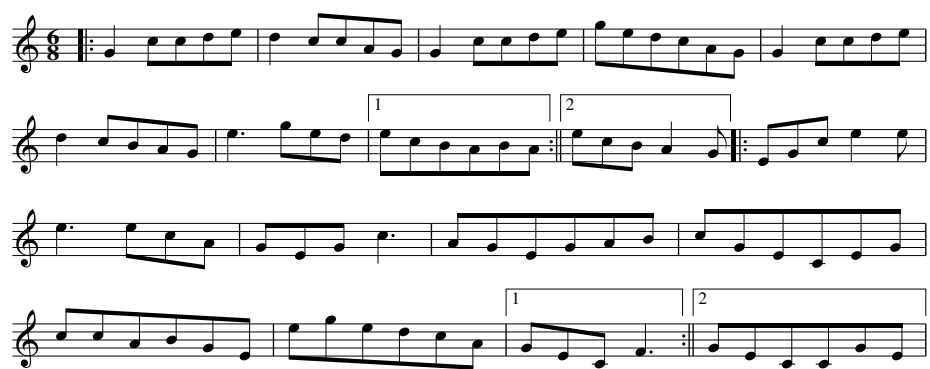


5

9



10



11



12



13



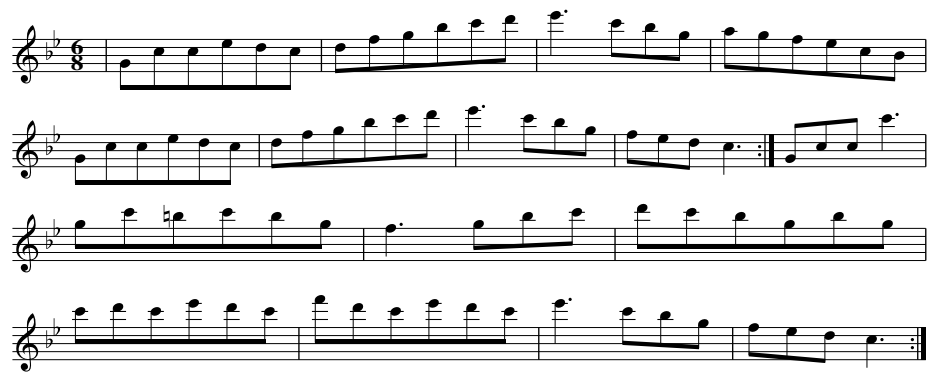
14



15



16



17



18



Round Two

The model in this evaluation featured the final number of layers and heads with a more conservative sampling strategy but no beam search yet.

#	Model	Grade	Comment
6	folkrnn	+	
7	trad	+	(booooooring tune one would play fast only to wow people, but ok)
8	folkrnn	+	(strange F#, but this tune is close)
10	trad	+	(okkk but can be made better and played as a slide)
11	folkrnn	+	barely ok
12	folkrnn	+	ready to be played! (Is this “the new kid’s new friend”?)
13	folkrnn	+	(but I would resolve on G not A)
15	folkrnn	+	I like it but want to improve it
1	trad	-	too repetitive
2	trad	-	booooring, parts don’t work together
4	folkrnn	-	this anacrusis is weird, melody aimless
9	trad	-	booooring
14	trad	-	too repetitive
16	folkrnn	-	too repetitive
17	folkrnn	-	melodically aimless
18	trad	-	too repetitive
3	trad	- -	(those semiquavers are in unexpected positions, but I want to work with this one because it sounds unique)
5	trad	- -	(structure is bad)
19	trad	- -	(but I want to work with it because it sounds unique)
20	folkrnn	- -	(but I want to work with it as an aire)

Table 2: Second run evaluation

1



2



3



4



5



6



7



8



9



10



11



14

12



13



14



15

15



16



17



16

18



19



20



17

Round Three

This model features the introduction of the beams search algorithm for sampling and mark a noticeable increase in quality.

#	Model	Grade	Comment
3	trad	++	two parts go well together, but borderline boring
4	trad	++	Great! only a few changes needed. I esp. like that B references A
13	trad	++	With some changes it would be nicer
17	folkrrnn	++	I like it!
18	folkrrnn	++	With some changes it would be nicer
5	trad	ok	but close to another tune (The Kesh)
8	trad	ok	a little repetitive
10	folkrrnn	ok	nice A part but not so much in B part
11	trad	ok	not much of a jig, maybe copied
12	trad	ok	but booring
14	trad	ok	but a little out of style with raised 7th
0	folkrrnn	--	B part is poor
1	folkrrnn	--	A part is poor, but I like the B part
2	trad	--	too much repetition
6	folkrrnn	--	A part is strange
7	folkrrnn	--	just strange
9	folkrrnn	--	weird, but I want to work with it
15	folkrrnn	--	big leap in A part, but I think that can be fixed to make a nice tune
16	folkrrnn	--	disjointed B part
19	trad	--	too repetitive in A and too aimless in B

Table 3: Third run evaluation

0



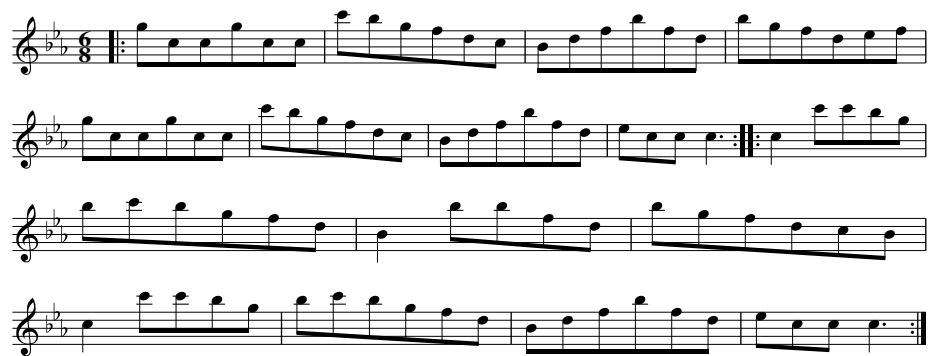
1



2



3



4



5



6



7



8



9



10



11



12



13

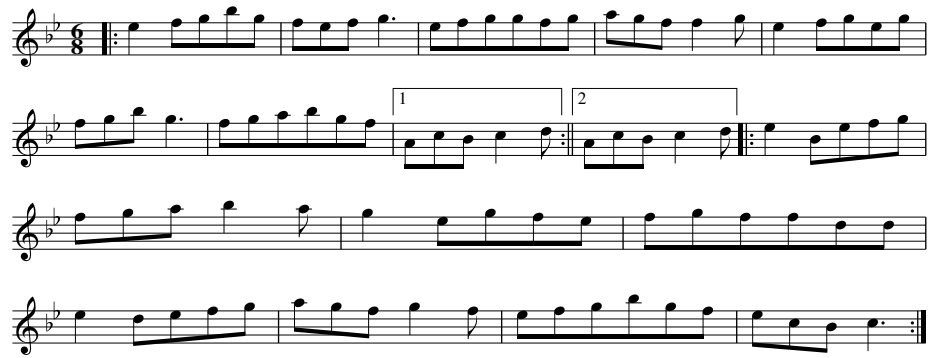


22

14



15



16

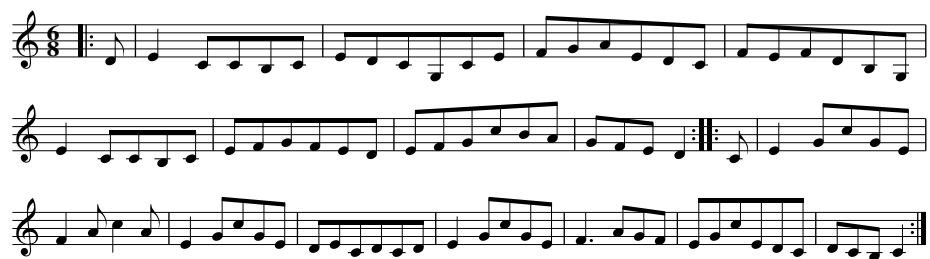


23

17



18



19



24

Round Three pt.2

We tried another batch with the same settings.

#	Model	Grade	Comment
0	trad	++	
6	folkrnn	++	
8	trad	++	
10	folkrnn	++	
12	folkrnn	++	
14	trad	++	
16	trad	++	
5	trad	ok	
1	folkrnn	--	
2	trad	--	
3	folkrnn	--	
4	folkrnn	--	
7	trad	--	
9	trad	--	
11	folkrnn	--	
13	trad	--	incorrect structure, but otherwise ok
15	folkrnn	--	
17	trad	--	bad structure, but I want to work with it
18	folkrnn	--	
19	folkrnn	--	but I want to work with it

Table 4: Round Three pt.2 evaluation

0



1



2



3



4



5



6



7



8



9



10



11



12



29

13



14



15



16



17



18



31



Round Four

We tested the final version of the tradformer with 4/4 tunes which are more challenging than 6/8

#	Model	Grade	Comment
2	trad	++	
8	trad	++	ha ha
9	trad	++	
1	trad	OK	
13	folk	OK	
19	folk	OK	with some changes can be made nice
0	trad	--	
3	trad	--	but could be great
4	trad	--	
5	trad	--	could make a nice single reel
6	trad	--	
7	trad	--	
10	folkrnn	--	
11	folkrnn	--	
12	folkrnn	--	
14	folkrnn	--	
15	folkrnn	--	
16	folkrnn	--	
17	folkrnn	--	
18	folkrnn	--	

Table 5: Round Four evaluation

0



1

Exercise 1 is a five-staff musical piece in 4/4 time with a key signature of two flats (B-flat and E-flat). The notation includes various rhythmic patterns such as eighth and sixteenth notes, as well as rests. The first staff begins with a treble clef and a key signature of two flats. The second staff ends with a double bar line and repeat dots. The third staff starts with a second ending bracket labeled '2'. The fourth staff continues the melodic line. The fifth staff concludes with a first ending bracket labeled '1' and a second ending bracket labeled '2'.

2

Exercise 2 is a four-staff musical piece in 4/4 time with a key signature of one flat (B-flat). The notation features a mix of eighth and sixteenth notes. The first staff starts with a treble clef and a key signature of one flat. The second staff continues the melodic development. The third staff includes a repeat sign followed by a key signature change to two flats (B-flat and E-flat). The fourth staff concludes the exercise with a final melodic phrase.

3



4



5



6



7



8



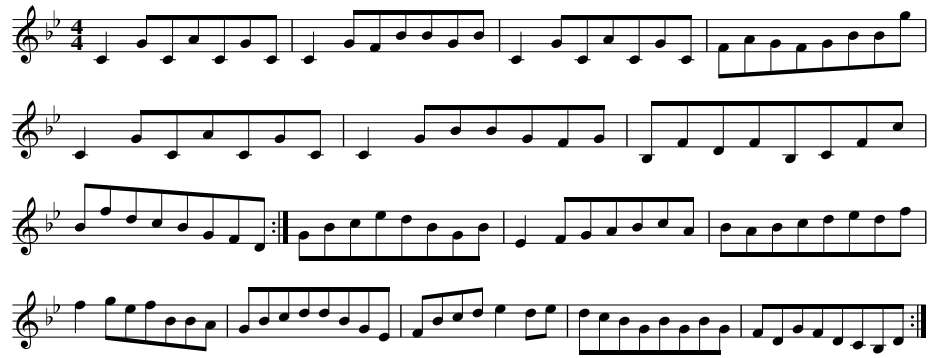
9



10



11



12



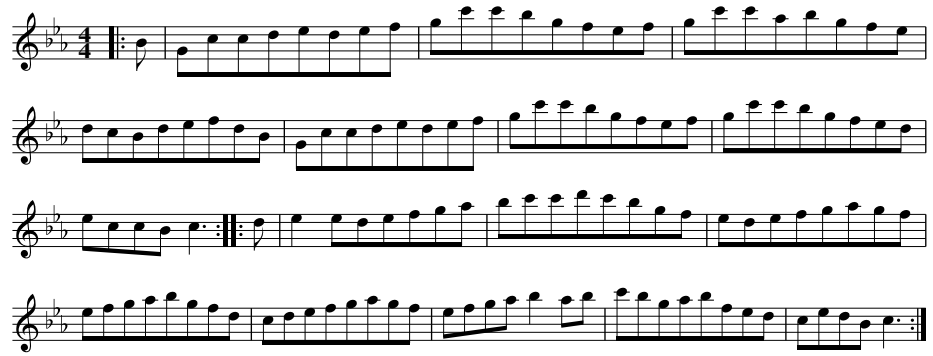
13



14



15



16



17



18



19



Round Five

Final test features tradformer against folk-rnn also with beam search. Strangely both models under-performed.

#	Model	Grade	Comment
17	folkrnn	++	slight changes needed, but great!
19	trad	OK	
0	trad	--	
1	trad	--	
2	trad	--	
3	folkrnn	--	
4	trad	--	
5	folkrnn	--	
6	folkrnn	--	but could be nice
7	folkrnn	--	
8	folkrnn	--	
9	trad	--	
10	folkrnn	--	
11	trad	--	but could be nice
12	folkrnn	--	
13	trad	--	
14	folkrnn	--	funny B part
15	folkrnn	--	
16	trad	--	poor imitation of Drowsy maggie
18	trad	--	

Table 6: Round Five evaluation

0



1



2



3



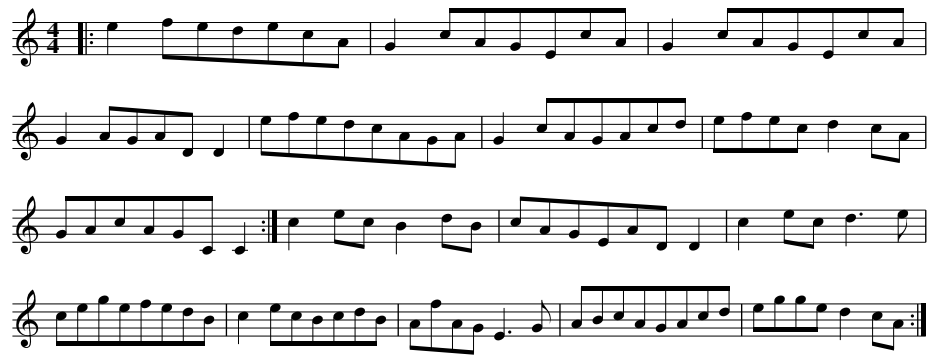
4



5



6



7



8



9



10



45

11



12



13



46

14



15



16



47

17



18



19



48

Results for The Ai Music Generation Challenge 2021

Below we present the results for The Ai Music Generation Challenge 2021. The Tradformer, with codename Vaxjo, won the competition. There were 5 judges that evaluated 10 tunes, 9 picked at random from the submission and 1 cherrypicked by each participant, according to four criteria: Danceability, Stylistic Coherence, Formal Coherence and Playability. Scores went from A (excellent) to F (fail). There were 5 other participants in the challenge. The table below list the sum of scores obtained by each participants.

Table 7: Sum of scores for each of the five participants. Tradformer (Vaxjo) in bold.

	Benchmark	Jönköping	Kalmar	Växjö	Småland	Oskarshamn
A	42	0	6	88	58	4
B	50	0	12	61	67	25
C	49	7	19	37	55	70
D	33	2	12	11	9	69
F	13	34	19	3	8	17

The following tables list the scores obtained by the Tradformer for each of the 10 tunes. Number 738 is the cherrypicked one. The lowest scores are mostly due to excessive repetition and in the case of 463 to the weird structure.

Judge 1				
Tune	<i>Danceability</i>	<i>S.Coherence</i>	<i>F.Coherence</i>	<i>Playability</i>
108	B	C	B	B
117	A	A	A	A
263	A	B	A	A
267	A	A	A	A
463	B	B	C	B
553	B	B	B	B
576	A	B	A	A
738	A	A	A	A
751	A	B	B	B
900	B	B	B	A

Table 8: Scores for the Tradformer model from Judge 1

	Judge 2			
Tune	<i>Danceability</i>	<i>S.Coherence</i>	<i>F.Coherence</i>	<i>Playability</i>
108	C	C	C	C
117	B	B	B	B
263	B	B	C	D
267	A	A	A	A
463	D	D	D	D
553	A	B	A	A
576	A	B	A	A
738	A	A	A	A
751	A	A	A	A
900	B	B	B	B

Table 9: Scores for the Tradformer model from Judge 2

	Judge 3			
Tune	<i>Danceability</i>	<i>S.Coherence</i>	<i>F.Coherence</i>	<i>Playability</i>
108	C	C	D	C
117	C	B	A	C
263	C	B	A	C
267	B	A	A	B
463	C	C	D	C
553	C	A	A	B
576	B	B	A	B
738	B	A	A	B
751	B	A	B	B
900	C	B	D	C

Table 10: Scores for the Tradformer model from Judge 3

Lastly, below is the sheet music for each of the 10 tunes.

108



	Judge 4			
Tune	<i>Danceability</i>	<i>S.Coherence</i>	<i>F.Coherence</i>	<i>Playability</i>
108	C	C	B	B
117	A	B	A	A
263	C	C	D	A
267	C	C	B	C
463	F	D	D	C
553	C	D	B	C
576	A	B	A	A
738	C	C	A	A
751	C	C	A	A
900	A	B	A	A

Table 11: Scores for the Tradformer model from Judge 4

	Judge 5			
Tune	<i>Danceability</i>	<i>S.Coherence</i>	<i>F.Coherence</i>	<i>Playability</i>
108	C	F	F	B
117	A	A	A	B
263	B	C	B	B
267	B	A	A	A
463	A	A	A	B
553	A	A	A	A
576	A	A	A	A
738	A	A	A	
751	A	A	A	A
900	B	A	B	A

Table 12: Scores for the Tradformer model from Judge 5

117



263



267



463



553



576



738



751



900

