Part of the 2020 Joint Conference on AI Music Creativity (CSMC + MuMe), KTH Stockholm

AI Music Generation Challenge 2020

in the style of the Double Jigs in Sergeant Francis O'Neill's "1001"

What?

Build an artificial system that generates the most plausible *double jigs*, as judged against the 365 published in F. O'Neill "The Dance Music of Ireland: O'Neill's 1001" (1907).[1] Up to two prizes will be awarded, and a performance of the best ones will occur at the 2020 Joint Conference on AI Music Creativity (CSMC+MuMe), Oct. 19-23 in Stockholm, Sweden (http://kth.se/aimusic2020). The panel of judges consists of four (human) experts in Irish traditional music and performance.

Why?

This challenge has three aims. The first aim is to promote meaningful approaches to evaluating music AI.[2] The second aim is to see how music AI research can benefit from considering traditional music, and how traditional music might benefit from music AI research.[3] The third aim is to facilitate discussions about the ethics of music AI research applied to traditional music practices.[4]

How?

1. BEFORE JULY 17, register your intent to participate by notifying the organizer: bobs@kth.se.
2. Build a music AI that generates music. You can train your AI on anything, but remember that the results will be judged against the 365 double jigs in O'Neill's "1001". To facilitate judging, the music you submit must be rendered either as ABC notation,[5] staff notation, MIDI, or mp3-compressed audio files.
3. Have your AI generate 10,000 tunes.
4. Write a brief technical document describing how you built your system, presenting some of its features and outcomes, and linking to your code and models for reproducibility.
5. BEFORE SEPTEMBER 22, email bobs@kth.se a link to download your generated collection and your technical document.
6. **Only one submission** from each participant will be allowed, so choose your best one.

The evaluation of submissions will proceed in four stages:

1. For each submitted collection, five tunes will be selected at random. For example, given an ordered collection of 10,000 music files, five random integers will be drawn without replacement in [1,10000]. The files indexed by those numbers will be selected.
2. Each tune from each AI will be assigned to two of four collections. The ordering of the tunes in each collection will be randomized. Each collection of tunes will be assigned to a judge for assessment. Information relating to the participating AI will not be disclosed.
3. **Stage 1 (judges working individually):**
    a. For each tune in a collection, the assigned judge will gauge whether it is plagiarized. If it is, then the tune will not be considered further.

---

[1] See this cleaned and corrected version: http://www.norbeck.nu/abc/book

[2] See: Sturm & al., "Machine learning research that matters for music creation: A case study," *J. New Music Research* 48(1): 36-55, 2018; Sturm and Ben-Tal, "Taking the models back to music practice: Evaluating generative transcription models built using deep learning," J. Creative Music Systems, vol. 2, Sep. 2017.

[3] https://medium.com/the-sound-of-ai/folk-the-algorithms-traditional-music-in-ai-music-research-b19bf392d991

[4] See: Holzapfel & al., "Ethical dimensions of music information retrieval technology," Trans. Int. Soc. Music Information Retrieval 1(1): 44–55, 2018; Sturm & al., "Artificial intelligence and music: Open questions of copyright law and engineering praxis," *MDPI Arts* 8(3), 2019.

[5] http://abcnotation.com/wiki/abc:standard:v2.1

   b. Otherwise, the judge will gauge whether the rhythm is close to that of a double jig. If it is not, then the tune will not be considered further.
   c. Otherwise, the judge will determine whether the pitch range of the tune is characteristic. If it is not, then the tune will not be considered further.
   d. Otherwise, the judge will determine whether the mode and accidentals are characteristic. If it is not, then the tune will not be considered further.

4. **Stage 2 (judges working individually):**
   a. For each tune that has passed through Stage 1, the judge will rate it along multiple dimensions with reference to the double jigs in O'Neill's "1001":
      i. *melody*
      ii. *structure and coherence*
      iii. *playability on traditional Irish instruments*
      iv. *memorability*
      v. *interestingness.*
   b. Each judge will send to the organizer a report of their assessments.
5. **Stage 3 (judges working together):**
   a. Each judge will present to the other judges the best tunes from their collections.
   b. Together they will decide which is the best double jig (or to award no prize).
6. **Stage 4 (judges working together with organizer):**
   a. The participating music AI systems that have produced highly ranked double jigs will be subject to an analysis of their consistency of quality.
   b. The organizer and judges will query the submitted collections of these AI by selecting tunes at random and assessing their quality.
   c. The judges together will decide which AI is most consistent in producing double jigs of high quality (or to award no prize).

			Judge ID			
		Tune ID	1	2	3	4
System ID	1	1	1	1		
		2		1		1
		3			1	1
		4	1			1
		5	1		1	
	2	1		1		1
		2	1	1		
		3		1	1	
		4			1	1
		5	1			1
	3	1	1		1	
		2		1		1
		3	1	1		
		4		1	1	
		5			1	1
	4	1	1			1
		2	1		1	
		3		1		1
		4	1	1		
		5		1	1	
	5	1			1	1
		2	1			1
		3	1		1	
		4	1	1		
		5		1	1	

An example of the experimental design in the first two stages is shown at left for five submissions. In this case, judges 1 and 2 each get collections with 13 tunes, and judges 3 and 4 get collections with 12 tunes. Each tune is assessed by two judges, and so each music AI system gets 10 assessments. Judges 1 and 2 assess five of the same tunes, and all other unique judge pairs assess four of the same tunes.

The next pages show an example outcome of stages 1 and 2 completed by an individual judge for ten jigs generated by *folk-rnn* models.[6] These questions were arrived at through expert elicitation with one of the judges. The exact wording of the final questions may not be the same.

Financial support for this competition comes from ERC-2019-COG No. 864189 (*MUSAiC: Music at the Frontiers of Artificial Creativity and Criticism*).

[6] Sturm & al., "Music transcription modelling and composition using deep learning," in *Proc. Conf. Computer Simulation of Musical Creativity*, (Huddersfield, UK), 2016. https://github.com/IraKorshunova/folk-rnn