Postprint

This is the accepted version of a paper published in *Journal of New Music Research*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Machine Learning Research that Matters for Music Creation: A Case Study

Bob L. Sturm [*][1], Oded Ben-Tal[2], Úna Monaghan[3], Nick Collins[4],
Dorien Herremans[5], Elaine Chew[6], Gaëtan Hadjeres[7], Emmanuel
Deruty[7], and François Pachet[8]

[1]Department of Speech, Music and Hearing, KTH Royal Institute of
Technology, Lindstedtsvägen 24 SE-100 44 Stockholm, Sweden,
bobs@kth.se
[2]Department of Performing Arts, Kingston University, UK,
Coombehurst House, Kingston University, Kingston-upon-Thames
KT27LA, UK, o.ben-tal@kingston.ac.uk
[3]Newnham College, Cambridge CB3 9DF, UK, umm21@cam.ac.uk
[4]Department of Music, Durham University, Palace Green, Durham,
DH1 3RL, UK, nick.collins@durham.ac.uk
[5]Information Systems Technology and Design Pillar, Singapore
University of Technology and Design, 8 Somapah Road, 487372
Singapore, dorien_herremans@sud.edu.sg
[6]Centre for Digital Music, Queen Mary University of London, Mile End
Road London E1 4NS, UK, elaine.chew@qmul.ac.uk
[7]Sony CSL, 6 rue Amyot, 75005 Paris, France,
firstname.surname@sony.com
[8]Spotify, 166 rue du Faubourg Saint Honoré, 75008 Paris France,
pachetcsl@gmail.com

Abstract

_____
[*]Corresponding author

1

Research applying machine learning to music modeling and generation typically proposes model architectures, training methods and datasets, and gauges system performance using quantitative measures like sequence likelihoods and/or qualitative listening tests. Rarely does such work explicitly question and analyse its usefulness for and impact on real-world practitioners, and then build on those outcomes to inform the development and application of machine learning. This article attempts to do these things for machine learning applied to music creation. Together with practitioners, we develop and use several applications of machine learning for music creation, and present a public concert of the results. We reflect on the entire experience to arrive at several ways of advancing these and similar applications of machine learning to music creation.

Keywords: applied machine learning, music generation, computational creativity, folk music

1   Introduction

The application of machine learning to music data aims to create machines that are beneficial to working with music. The uses of such systems span from the analytic, e.g., description and recommendation via music information retrieval (Schedl et al., 2014), to the synthetic, e.g., creative transformation and generation via algorithmic composition (Dannenberg et al., 1997; Pearce et al., 2002; Ariza, 2005; Nierhaus, 2008; Dean, 2018). The latter continues to be a very active research area (Fernández and Vico, 2013; Herremans et al., 2017), especially with deep learning methodologies (Briot et al., 2017), and has growing commercial interest.[1] However, how does any of this machine-learning research *matter*?

In her provocative paper presented at the International Conference on Machine Learning, Wagstaff (2012) argues that machine-learning research "matters" when it closes the loop between the design and application of its methods, and the use of the resulting technologies by real-world practitioners. She observes that this is a rare occurrence in machine-learning research, where many publications present experiments conducted using datasets that are only vague proxies for gauging real-world usefulness. For instance, a proposed learning method might result in a model that can predict with high accuracy the toxicity of mushrooms in some dataset, but this result does not imply that the learning method, model, or even problem it is thought to be solving, are meaningful or useful to a mycologist.

---

[1]Recent examples are Jukedeck (`www.jukedeck.com`), Aiva (`www.aiva.ai`) and Melodrive (`http://melodrive.com/`), three start-up companies creating systems that automatically compose royalty free music that can be used for sound tracks.

The same criticism might be said about much published machine-learning research applied to music modeling and generation. Typically, a researcher trains a machine-learning model on a collection of music recordings and measures how well it "fits" music held out from its training. This could involve measuring the likelihood of some "real" music in the model (Boulanger-Lewandowski et al., 2012; Greff et al., 2016), or looking at how the music generated by the model reflects high-level patterns (Jaques et al., 2016). One may look at or listen to generated music and qualitatively compare to the music in the training data using listening tests for lay listeners (Jaques et al., 2017) or experts (Collins and Laney, 2017). If these systems show a high degree of success in these "proxies", then that might be taken implicitly as a sign that they are useful; but who is the user, and what is the use? How do these technologies and the attendant research *matter* to music practitioners in the real world?

Wagstaff (2012) essentially proposes two principles to make research in machine learning matter:

1. Measure the concrete impact of the application of machine learning with practitioners in the originating problem domain;

2. With the results from the first principle, improve the particular application of machine learning, the definition of the problem, and the domain of machine learning in general.

Undoubtedly, crafting research with these principles is very hard. Involving practitioners brings into the fold numerous ethical and technical constraints, not to mention a real-world messiness obscured by cleanly labeled datasets, all-but-standardised train/test regimens, and straightforward statistical testing (Drummond and Japkowicz, 2010). Given the increase in the cost of such research, among other things, it is no wonder that applied machine-learning research that closes the loop between lab academics and real-world practitioners is rare.[2] However, this is precisely where Wagstaff (2012) argues that the most meaningful work begins: when the researcher measures how the technology they are developing actually impacts practitioners, and how that in turn can inform the research pursuit.

In this article, we employ the two principles of Wagstaff (2012) for machine learning applied to music creation. Together with practitioners, we apply a variety of machine-learning methods to different music datasets, and then employ the resulting systems

---

[2]Wagstaff's criticisms have not gone unaddressed. Some machine-learning researchers have mentioned that they are like manufacturers of engines, and so it is unfair to criticise them because they are not building vehicles. Even so, the engineering of useful engines should consider real-world use, constraints and specifications.

for music creation, from composition to performance. How does this technology impact the process of music creation? What is the significance of its use in the process and for the end result? What lessons can we learn from the experience to inform such applied machine-learning research? In this way, we link together machine learning applied to music creation, and the use of the technology by actual practitioners. We not only aim to gauge the real contributions to music creation, but also identify ways of advancing research applying machine learning to music creation. The entire experience contributes to an understanding of music creation *with* machine learning, and the ways each can inform the other.

In the next section, we describe several musical works composed by and co-composed with a variety of systems created using machine-learning methods. These works were premiered at a concert on May 23, 2017, the playbill of which is shown in Fig. 1.[3] In the third section, we discuss with composers and performers several questions related to the exercise: What does machine learning contribute to your work? What are the roles of human and machine creativity in your work? How do these roles matter for the audience of your work? Is it important to you to limit the human editing of generated results? Why or why not? In the fourth section, we discuss reactions from the audience in a questionnaire administered at the concert. In the fifth section, we discuss several ways forward in the application of machine learning to music creation.

2   Applications of machine learning to music creation

We now discuss several machine-learning systems and the music created with them and premiered at the public concert (Fig. 1). Table 1 summarises the details of the four different systems and their applications. Each of the following subsections are written by practitioners identified in the subsection title.

2.1   folk-rnn models

A *folk-rnn* model[4] is a long short-term memory (LSTM) network trained to model textual sequences of music transcriptions (Sturm et al., 2016). The specific representation
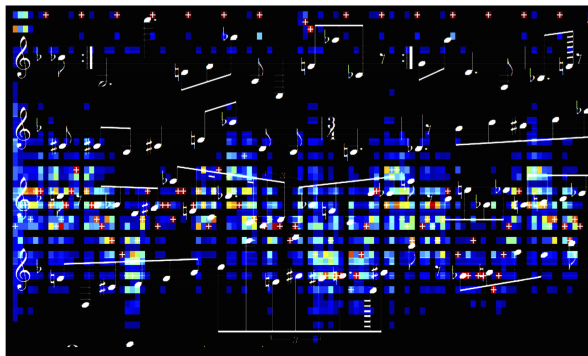
---

[3]Please see the Bottomless Tune Box YouTube channel for some of the works discussed below: `https://tinyurl.com/yd9uko74`

[4]`https://github.com/IraKorshunova/folk-rnn`

PARTNERSHIPS

Music composed by and co-composed with computers having "musical intelligence"

Pieces for organ: The Glas Herry Comment & X:7153 by *folk-rnn* + *DeepBach* (2017)	**Richard Salmon** *organ*
Traditional Irish Sets (with *folk-rnn* tunes in *italics*) • **Jigs** (The Cuil Aodha, The Dusty Windowsill, *The Glas Herry Comment*) • **Slow Reels** (Maghera Mountain, *X:2897*) • **Fast Reels** (The Rookery, *X:1068*, Toss The Feathers II)	**Daren Banarsë** **and Musicians**
March to the Mainframe, Interlude, *The Humours of Time Pigeon* by Bob L. Sturm + *folk-rnn* (2017)	**Ensemble x.y**
Ed SheerAI vs XenAkIs vs Aldele by Nick Collins (2017)	**Ensemble x.y**
3 morphed pieces from "A Little Notebook for Anna Magdalena" by J. S. Bach (1722) + MorpheuS (2017) **3 morphed pieces from "30 and 24 Pieces for Children"** by Kabalevsky (1937) + MorpheuS (2017)	**Elaine Chew** *piano*
Safe Houses by Úna Monaghan + *folk-rnn* (2017) ***The Choice*** by Úna Monaghan (2015) ***The Chinwag*** by Úna Monaghan (2015)	**Úna Monaghan** *Irish harp, concertina,* *electronics*
Pieces for organ: X:633 & The Drunken Pint by *folk-rnn* + *DeepBach* (2017)	**Richard Salmon** *organ*
Chicken Bits and Bits and Bobs by Bob L. Sturm + *folk-rnn* (2017)	**Ensemble x.y**
Bastard Tunes by Oded Ben-Tal + *folk-rnn* (2017)	**Ensemble x.y**

Figure 1: Playbill for the May 23, 2017 concert organised by Ben-Tal and Sturm.

it encodes is ABC notation,[5] which denotes meter, mode, bars and notes using text. The specific *folk-rnn* models we use for the compositions in the following subsections

---

[5] http://abcnotation.com/

Name	Machine learning approach	Training data	Applications
folk-rnn	Long short-term memory (LSTM) network	Symbolic transcriptions of folk music (much of which is from Ireland and the UK)	Material generation for curation and arrangement (Sturm, Banarsë); material co-creation (Ben-Tal, Monaghan)
DeepBach	LSTM	Bach chorale transcriptions	Chorale generation from melodies generated by *folk-rnn*
MorpheuS	Constraints (tonal and pattern)	none	Generation from template provided by existing composition
ArtIst	Variable-order Markov modeling	melody and chord information extracted from works of three different musicians/composers	chord sequence generation followed by melody generation

Table 1: Summary of machine learning methods and datasets used in this study.

are trained on over 23,000 ABC transcriptions of traditional tunes from the online crowd-sourced repository, `http://thesession.org`. Many of these tunes come from Ireland and the UK. The trained model generates new ABC sequences by iteratively sampling from and updating the posterior distribution over the transcription vocabulary at its output layer (Sturm, 2018). We use two different *folk-rnn* models. The first version (v1) was trained on sequences of characters from a text-document compilation of all the training tunes; the second version (v2) was trained on a tokenised version of the transcriptions. These models and their creation are described more completely in Sturm et al. (2016); Sturm and Ben-Tal (2017); Sturm (2018).[6] In the next subsections, we describe several ways in which we have used the models to create music.

---

[6]We have now developed an online application of these models for people to explore: `http://folkrnn.org`.

2.1.1 "Bastard Tunes" by Ben-Tal

A machine-learning model trained on music encodes some musical knowledge, at least in theory. When we examined outputs from the *folk-rnn* models (Sturm and Ben-Tal, 2017), we could see they are able to produce musically relevant repetition and variation of melodic patterns. So, I thought that the outputs generated by the *folk-rnn* v2 model when fed my own musical ideas would be interesting. I was wrong. But while I ended up using the model differently than I assumed when setting out to write my piece, both the capabilities and limitations of the model shaped the piece and the composition process that led to it. "Bastard Tunes", a four-movement ensemble piece (fl./pic., cl./bass cl., perc., piano, vl., db.), could not have been produced by either myself or *folk-rnn* without the other.[7]

As we have observed (Sturm and Ben-Tal, 2017), what *folk-rnn* v2 has actually learned to do is very limited, and tied to the specific context of the training data. The model lacks the ability to generalise, which means that when I tried applying it to my own musical ideas the results were unsuccessful. The model would simply ignore my input and revert to "noodling". To work effectively with this model I would need to adapt the parameters of my own musical idiom, and to meet the model half way. On a purely practical level the control I had over the model was rather limited. I can initialise it with a melodic fragment (usually a few notes to a few bars). If I am unhappy with the outputs, I can try to generate different outputs by changing the random seed, and repeat in the hope that the model yields something useful. I can change the temperature parameter of the generation process – which effectively flattens the sampling distribution and increases the chances of generating unlikely transcription symbols. (In other words, lower temperatures result in more "conservative" choices while higher temperatures lead to more "adventurous" outputs.) Finally, I can change the initial sequence that starts the generation process. None of these steps produce predictable changes in the outputs, however. I can pull these three levers, but even after many hours of working with the model I only have limited intuition about how these will steer it.

Fairly early on in the composition process I realised I was more interested in exploring the edges of the model, and not its more typical results. I achieved this primarily by constructing initialisation sequences from tokens that have low probability in the training data, e.g., meters such as 9/8 or 12/8, Mixolydian and Dorian modes, rests, dyads and very long notes. I also used the temperature parameter to shift the model toward producing atypical sequences.

---

[7]The performance of "Bastard Tunes" can be heard here: `https://goo.gl/S3PJvy`.

The first movement of "Bastard Tunes" is the only one in which I ended up where I planned to go originally. My idea was to generate melodies that start together and gradually diverge. There are four melodic lines that start almost identically, which I generated by setting the temperature to be low (0.2 instead of the default of 1.0). I then took the last measure of each melody and used it as a seed with a slightly higher temperature. I repeated this 15 or more times for each melodic strand. The first attempt was not successful. While the overall result matched my idea the details did not. The four melodies diverged too quickly initially, meaning there was no audible process. The melodies became separate after four of five bars and it all sounded the same until the final unraveling when high temperature parameters yielded strange results.

I restarted the process at an even lower temperature (0.05). I also set the meter to 9/8 to produce more interesting rhythms and tried several initial sequences until I got a promising beginning. The process unfolded in the same way: using the last bar as the initialisation for the next step. I raised the temperature more gradually at first, and I also edited some of the initialisation sequences. The v2 model can get stuck in rhythmic patterns, particularly dotted rhythms, which are typical of a specific dance form in some of its training data (e.g., hornpipe). When the final bar of a sequence had too much of those, I made small changes to drag the system away from sinking into such a trap. The result is substantially the material of the opening movement. I had to discard a significant portion of the generated material since what I had was too long. I also made various small edits such as replacing some notes with rests so the players can breath and adding articulations. I used octave transpositions to bring out interesting material at different points and spread the lines among the different instruments.

I wanted to have a slow and contemplative movement in the piece. While there are some ballads in the training data, the traditional tunes in the training data tend to be played fast and so have short notes. I experimented quite a lot with initialising the model using long notes and one of the unexpected results was a tune I use in the final movement, which is shown in Fig. 2. Eventually, I did get a melody that was meandering, but not totally aimless, and was less rhythmically rigid than most outputs of the model. Fig. 3 shows two sections from the generated melody, and the way I adapted those into the score of my piece. I increased the overall rhythmic variety with both longer and shorter notes and also produced local focus (e.g., highest f note in Fig. 3(b)). I also freely added and removed bars, transposed sections up or down, and added accompaniments primarily through the use of heterophony (e.g., Fig. 3(a)).

Both the second and fourth movements are based on tunes generated as part of the

Figure 2: Melody generated by the *folk-rnn* v2 model used for the fourth movement of "Bastard Tunes".

(a) Adding and removing measures; changing rhythms; extending to multiple lines

(b) Octave transpositions; microtonal inflections; changing rhythms; extending elements to create accompaniment

Figure 3: Editing in the composition process applied to a melody generated by the *folk-rnn* v2 model. "OBT" summarises Ben-Tal's treatment of the material generated by *folk-rnn*.

Figure 4: Melody generated by the *folk-rnn* v2 model used for the second movement of "Bastard Tunes"

long process of experimentation with the model v2. In total, I generated hundreds of tunes most of which were discarded immediately. I kept several dozen which held some potential interest. Two tunes grabbed my attention almost immediately suggesting a potential use. I recognised the potential of one tune with canonic treatment. This is somewhat surprising as I rarely use canons or canonic devices in my composition. But in this case the realisation came in a manner I can only describe as a moment of inspiration. When it came to composing this movement, I wanted a longer tune so I extended it by reseeding the model with the final measure. After a few trials and some further editing, I arrived at the melody in Fig. 4. The subsequent entries in the canon do not conform to the meter (e.g., 5 beat delay in 4/4). In the second half of the short movement, one of the voices starts on an off-beat to produce a nice hocketting effect because of the dotted rhythms. The final movement is the most straightforward arrangement of the four. The melody is almost unchanged from the output generated by the model. I added a somewhat unorthodox harmonisation and changed the meter in the middle part from 4/4 to 3/4.

"Bastard Tunes" is markedly different from previous pieces I composed. In that regard, my work with the model achieved its goal of creatively exploring new ideas by engaging with a "musical other". I treated the model outputs with considerable freedom, but other users can find a more dogmatic approach more useful. Wrestling with the model to arrive at material that I found acceptable forced me out of my existing habits, which is an exercise that every creative person should engage in periodically. There are, of course, other ways for composers to engage with "musical others", or to break habits. But using the model as a kind of "composition assistant" is different to using an arbitrary set of constraints, or to systematising certain aspects of the

Figure 5: X:488 generated by the *folk-rnn* v2 model is the melody in Sturm's "March to the Mainframe".

composition.

2.1.2 "Two short pieces with a short interlude" by Sturm

The principal mode in which I worked with the *folk-rnn* models involved curating from transcriptions generated by them. I did not limit myself to "cherry picking", but also challenged myself to work with generated material that was not immediately sensible. My composition consists of three movements: "March to the Mainframe", "Interlude" and "The Humours of Time Pigeon".[8]

"March to the Mainframe" (pic., cl., snare, piano, vln., db.) is an arrangement of X:488 from "The *folk-rnn* (v2) Session Book Volume 1 of 10".[9] Figure 5 shows the notation of the tune, which by and large with respect to the training data is a very successful output. It fits nicely across the rows of a G/D diatonic accordion, and with its limited selection of harmonies in the bass (push/pull in G: I/V, V/II, IV/IV, iii/vi). My 82-measure march is a straightforward arrangement of this 16-measure tune, with the full tune appearing twice, concluding with a repetition of its first eight measures.

---

[8]The score can be found here: `https://goo.gl/QGqBSK`. The performance at the concert can be heard here: `https://goo.gl/aCsynJ`.

[9]This and many more volumes can be found here: `https://highnoongmt.wordpress.com/2018/01/05/volumes-1-20-of-folk-rnn-v1-transcriptions/`. Syntheses of nearly 48,000 of these generated tunes can be auditioned at "The Endless *folk-rnn* Traditional Music Session" website: `https://goo.gl/EM7GmX`.

"Interlude" (fl., bass cl., snare with brush, piano, vl., db.) is built using material from the first four measures of "The Millennial Whoop Reel" — a piece I co-created with v2:[10]

Of the pieces I created with *folk-rnn* described here, this one is the furthest from the model-generated output.

I composed "The Humours of Time Pigeon" (pic., bass cl., snare w/o wires, piano, vl., db.) from a transcription I found in 72,000+ ABC sequences generated by v1.[11] I was specifically looking for failures of the model. Figure 6 shows how this melody suffers from counting errors and waywardness. The tune does not look or sound like much of the training data. Regardless, I found the title evocative enough to inspire working with the transcription — which is the second time I have worked with it (the first appears in my 2015 composition, "Eight Short Outputs ...",[12] where the transcription appears verbatim).

2.1.3 Tunes generated by folk-rnn models and harmonised with *DeepBach* by Hadjeres, Deruty and Pachet

DeepBach is an LSTM-based statistical model aimed at modelling chorale music (Hadjeres et al., 2017). After being trained on a dataset of 389 J.S. Bach chorales, *DeepBach* is able to generate four-part chorales in the style of Bach. This neural network approach learns everything from data, and uses no expert knowledge, such as rules of counterpoint. An online experiment conducted in 2016 assessed that *DeepBach* was highly competent in harmonizing the soprano lines of Lutheran chorales. A strength of this model is that it is also steerable: a user can impose positional constraints such as fixed notes, rhythm or cadences in the output generated by *DeepBach*. This enables us to think about musical composition as a dialogue between a system that proposes musical solutions and a user who interactively selects, tweaks or discards those proposals.[13] Other neural approaches have been proposed to solve this problem before, but

---

[10]https://goo.gl/zhBhka

[11]This appears in The *folk-rnn* (v1) Session Book Volume 15 of 20 (see footnote 9 for a link to this volume).

[12]https://www.youtube.com/watch?v=RaO4HpMO7hE.

[13]https://www.youtube.com/watch?v=OkkKjy3WRNo shows an example of such an interaction using a simple MuseScore plugin calling *DeepBach*.

Figure 6: Sturm's composition "The Humours Of Time Pigeon" is built around this melody of the same name generated by the *folk-rnn* v1 model. *folk-rnn* model v1 also learned to generate titles.

they use expert musical knowledge (Hild et al., 1992), or are less steerable.[14]

We applied *DeepBach* to harmonise several melodies generated by the *folk-rnn* v2 model.[15] Each tune provides the soprano part of the chorales, and *DeepBach* generates the three other parts. We took the freedom to adjust the timing of the given melodies, e.g., doubling the duration of the notes, and adapting the rhythm for pieces in 6/8, and to transpose them so that they fit within the (hypothetical) singer voice ranges. Figure 7 shows the first 8 measures of v2 transcription "X:633",[16] and its *DeepBach* harmonisation. Even when a melody features some unusual or rare melodic motions (compared to most of the Lutheran hymns used in J.S. Bach chorale harmonisations), *DeepBach* is still able to produce a fluid chorale texture and provides new insights and harmonic contours to the original melody. We can cite as examples of such melodic motions the natural E descending to C in m. 1, the descending sixth in m. 2, or the ending cadence with a Picardy third in F major. It is also worth noting how identical passages of the soprano part are harmonised differently (m. 1 and m. 9 for instance).

---

[14]https://github.com/feynmanliang/bachbot

[15]The performance of these pieces can be heard here: https://goo.gl/fKethA.

[16]See footnote 9 for the link to the volume from which this comes.

Figure 7: An example of how the first part of X:633 generated by the *folk-rnn* v2 model (top) is harmonised by *DeepBach* (bottom).

This highlights the variety of solutions proposed by *DeepBach*.

2.1.4 "Safe Houses" by Monaghan

This composition (for concertina and tape) is my response to "The *folk-rnn* (v2) Session Book Volume 1 of 10".[17] My approach was influenced by the following three factors. First, that the book is a computer-generated collection of notations, produced after a system learned from a database of Irish traditional dance music, also notated. This is an aural tradition, so the materials used and produced are by definition disconnected from the music they emulate. I wanted my piece to be mindful of Irish traditional music, the concept of machine learning, and the fact that an established cultural practice was being used by a machine. Second, the tunes produced vary in how successful they are as a piece of music in this style. Some of the pieces are impossible to identify as computer generated, whereas in others there are clear indications of their non-human origin. For example, there may be unusual note intervals and rhythmic figures, or the piece may be overly repetitive and formulaic. I made no attempt to hide this variation in my collaboration; I deliberately did not alter the computer generated tunes. I used a chance process to select which tunes to feature, as there were so many (3000 in volume 1 alone). I wanted to present a fair rather than idealized impression of the system's output. Third, I wanted to treat the computer as far as possible as an equal collaborator. The *folk-rnn* collection was already in existence when I began, so our respective contributions had to be made consecutively rather than in real-time. I decided my composition method and performance would mirror that of the computer system, as a way to enable a non-hierarchical collaboration.

Each tune in the collection is identified by a number, similar to a house address. I was struck by the sense of unease felt by almost all musicians to whom I described the work of the *folk-rnn* team (Sturm, Ben-Tal, et al.) and system. Most were wary of a computer system dipping into and using, (or perhaps infiltrating?) their "home" tradition. Change, electronic technology and fusion with other genres have previously been controversial topics in Irish traditional music (Vallely et al., 1999). As a reference to these reflections, I named the piece "Safe Houses" and used the house numbers of places I had lived in Belfast as a basis for selecting some tunes to work with.

I selected seven tunes using this idea: X:25, 28, 32, 34, 55, 82, and 279. I made audio recordings of my initial process of learning each tune from the notated music. I played through the melody slowly at first, working on particular sections in isolation that proved problematic, and then repetitions to get used to the piece and commit it to memory. This is a typical process I follow when learning Irish traditional music from

---

[17]See footnote 9. The performance of "Safe Houses" can be heard here: `https://soundcloud.com/unamonaghan`.

notation. I wanted to use it to directly influence the piece, in a way which mirrored the machine learning integral to the computer output. This produced seven audio files of different lengths that provided source recordings as a basis for the creation of a tape piece. I edited and manipulated the recordings in a digital audio workstation to produce an electroacoustic fixed media piece. When I could play each piece well and at performance speed, I re-recorded each with standard repeats (first part twice, second part twice, repeat the tune). This gave an estimate of length for the tape piece.

My compositions often explore new ways of presenting traditional dance music. A variable speed is one way to experiment, highlighting that in contemporary performance contexts for Irish traditional dance music, it is often presented without dancers, and therefore the speed does not need to be fixed. In "Safe Houses", the *folk-rnn* pieces are played one after the other, similar to a set of tunes, a familiar way to present Irish traditional music. However, there are not fixed cues in the tape part at which to start playing the melodies. The piece allows some leeway for the performer to choose the speed, to vary the speed, and to include short improvised links.

"Safe Houses" is now a 6-minute long piece for concertina and tape featuring tunes X: 82, 32, and 28 from the *folk-rnn* collection. (Figure 8 shows X:82.) The tape piece, played by computer, is composed by me from recordings of my playing as I learned the computer's compositions. We each play the other's work during the performance of the piece, and this work is made from the process we each went through, of learning the other's material.

Figure 8: X:82 generated by the *folk-rnn* v2 model, which appears in "Safe Houses" by Monaghan.

2.1.5 Three traditional sets

London-based Irish musician Daren Banarsë[18] assembled three sets of tunes by interleaving transcriptions generated by the *folk-rnn* v1 and v2 models with existing traditional tunes.[19] Banarsë made very few modifications to the generated transcriptions, an example of which is shown in Fig. 9. A "set" refers to a selection of pieces of dance music played one after another with no break. The organisation of tunes into sets is thought to be influenced by early recordings of Irish traditional music, such as those of Michael Coleman in the 1920s. One side of a 78rpm record had space for around 3 minutes of music. This was enough time for three tunes, which remains the standard size of a set of dance tunes to this day (Bradley, 1999). One set by Banarsë is of jigs, and includes "The Cuil Aodha," followed by "The Dusty Windowsill," and ends with "The Glas Herry Comment" from "The *folk-rnn* (v1) Session Book Volume 1 of 20" (see footnote 9). The second set is a slow reels set, and includes "Maghera Mountain" followed by X:2897 (Fig. 9) from "The *folk-rnn* (v2) Session Book Volume 1 of 10" (see footnote 9). The third set is of fast reels, and includes "The Rookery," followed by X:1068 from "The *folk-rnn* (v2) Session Book Volume 1 of 10", and ends with "Toss The Feathers".

2.2 MusAIcians by Collins

MusAIc is music created through artificial intelligence technology, and MusAIcians are programs manifested for a particular creative musical scenario, ideally aspiring to the status of musicians in their own right. We might credit them with greater intelligence where their final personality has grown more independent of the programmer, a situation much assisted by machine learning.

For the present composition, "Ed SheerAI vs XenAkIs vs AIdele" (fl., cl., pn., vl.), three artificially intelligent musAIcians were created. Each was derived from a corpus of example audio files from a given artist, respectively, Ed Sheeran's albums "+", "x" and "÷", a collection of Iannis Xenakis's choral, instrumental and electronic music, and Adele's "19", "21" and "25". Automatic analysis utilised the Melodia melody feature extraction algorithm (Salamon and Gómez, 2012) and Chordino chord detection (Mauch and Dixon, 2010), with post processing of vamp plugin outputs in SuperCollider. Feature extracted data was converted to integers for sequence modelling. In particular, chord symbols were parsed and pushed to seven basic chord classes, and

---

[18]http://www.darenbanarse.com

[19]The performance of these sets can be heard here: https://goo.gl/6R7H3e.

Figure 9: The *folk-rnn* v2 model generated X:2897 (top, see footnote 9), which Irish musician Daren Banarsë then modified (bottom).

the melodic material was interpreted as intervals within two octaves in product with two rhythm classes (long and short, following Fraisse, see for instance Clarke (1999)). Derived data was also prepared with pitch classes relative to the current chord at a
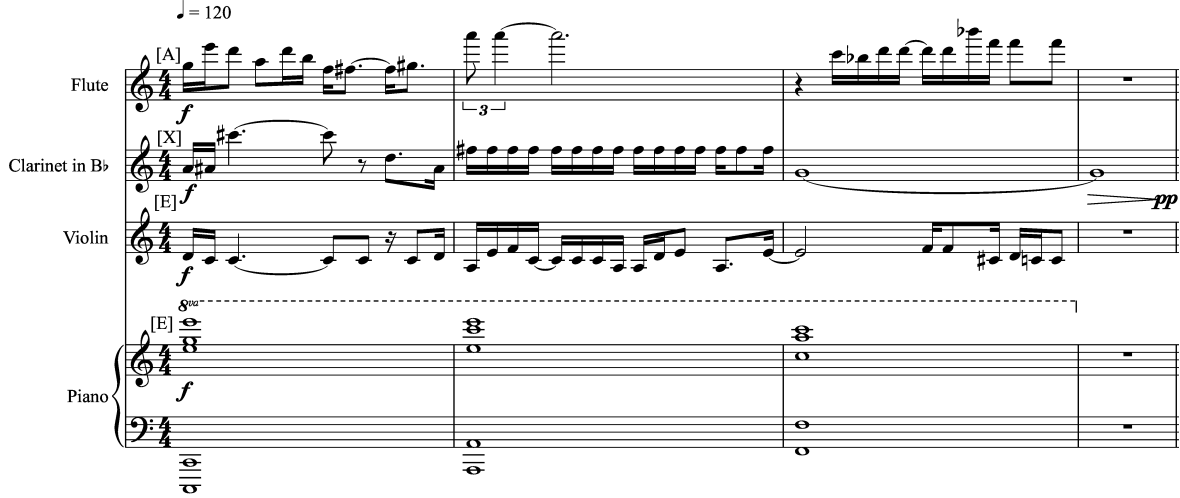
18

Figure 10: The first three bars of Ed SheerAI vs XenAkIs vs AIdele by Nick Collins. The [A] denotes material derived from the AIdele model, similarly [E] and [X]. The underlying harmonic structure (C, Am, F chords, one per bar) is spelt out in the piano and derives from analysis of chords within the Ed Sheeran corpus.

given time, to provide transposition free joint information.

After variable order Markov modelling over such material (Begleiter et al., 2004), the AIs could generate new chord sequences, and new melodies to a given chord sequence, in the style of their source material as mediated by a local Markovian process. The limiting factor in a system trained on information extracted directly from audio files is the quality of machine listening. The system can be revisited as more developed automatic transcription technology becomes available, but is already a creative response to the potential of machine learning over larger databases of material.

The concert work presented was a three minute non-pop song consisting of a series of sections, each based on chord material from one of the models (sometimes revealed in the piano), and changing allocation of the AIs to generate associated melodies for flute, clarinet and violin (see Fig. 10).[20] The difficulty of the rhythms varies, based on the level of quantisation (evening out) of the originally detected note events in the audio files on which novel generation depends. Except for a few range issues, breathing points and section edits, the material was not modified in rhythm or pitch from the computer output. The piece provokes on the nature of creative influence in music, and the sanctity of musical identity.

---

[20]A recording of the performance of the piece can be seen here: `https://youtu.be/9SQQwUMSGPQ`.

2.3 The MorpheuS system by Herremans and Chew

While many existing automatic composition systems perform well on a note-to-note level, one of the biggest challenges is automatically composing a new piece that has long-term structure (Collins, 2009; Herremans et al., 2015). MorpheuS (Herremans and Chew, 2016a, 2018) addresses this problem by generating a new piece from a template piece while preserving its long-term structure. MorpheuS takes two types of structure into account: tonal tension and recurrent patterns.

Tonal tension is calculated through the model for tonal tension (Herremans and Chew, 2016b) based on the spiral array, a mathematical model for tonality (Chew, 2014). This model captures three aspects of tonal tension for each time-slice of a piece: how dispersed the pitches are in tonal space (cloud diameter), a measure of dissonance; the tonal distance traversed from one time slice to the next (cloud momentum); and the distance from the current tonal context to the global key of the piece (tensile strain).

The second type of structure consists of recurrent patterns, which are detected from the template piece using the SIATECCompress and COSIATEC pattern-detection algorithms (Meredith et al., 2002; Meredith, 2015). These compression-based algorithms find clusters of notes with the same rhythm that reoccur (even those who are transposed) throughout the piece. By tuning the parameters of the compression algorithm, such as the maximum and minimum allowed pattern length, we can tweak the properties of the generated output. For instance, when constraining the generated music with a large number of short patterns, there is a strong resemblance to the original piece. When using a small number of long patterns, there are typically not enough recurring patterns in the new piece for it to sound coherent. MorpheuS therefore implements a balanced pattern length based on a series of experiments.

To optimize the fit to the template tension profile subject to the constraining patterns, MorpheuS starts by populating the rhythmic template with random pitches. It then optimizes the pitch of each note so that the target tonal tension profile is matched as closely as possible, while hard constraining long-term repetition patterns. This problem is efficiently solved (Herremans and Chew, 2016a) using a variable neighborhood search metaheuristic (Hansen et al., 2001).

For the concert, we used MorpheuS (Herremans and Chew, 2018) to morph six pieces, three based on selections from "A Little Notebook for Anna Magdalena" by Johann Sebastian Bach, and three based on selections from "30 Children's Pieces" and "24 Children's Pieces" by Dmitry Borisovich Kabalevsky.[21] We make no manual

---

[21] The live recording of the first performance and the accompanying scores can be heard and viewed here: `https://vimeo.com/234662284`.

adjustments to the generated notes, and present them as is.

Figure 11(a) shows a fragment of "Clowns" from Kabalevsky's "24 Children's Pieces"; Fig. 11(b) shows the same fragment with randomised pitches; and, Fig. 11(c) shows the final morphed version. Note that the patterns from the original piece in (a), such as the toggling eighth-note figure in the left hand, are preserved in the morphed piece (c), albeit with different pitches. Even though these same patterns are present in the random starting piece (b), without optimizing the tonal tension profile, the music contains strange dissonances and generally exhibit a sense of randomness. The improved coherence after matching the tonal tension profile confirms that this parameter is an important aspect in music generation. Solely by constraining the two aspects of tonal tension and recurring patterns, MorpheuS is able to produce interesting output.

There are currently still limitations to the MorpheuS system, most of those are due to the fact the tension model is based on pitch classes, not intervals, and the lack of a statistical model to guide note-to-note transitions. The latter leads to some unexpected transitions, such as large octave leaps and hand/finger crossings, that can be challenging for the performer. Because MorpheuS has no formal concept of cadences, this leads to some surprising endings. Some of these gaps will be addressed in a future version by integrating a statistical model, such as a recursive neural network, in the objective function.

3   Composer and musician responses

We now compile responses of composers and musicians to several questions about their work on this project. These questions were: What does machine learning contribute in your work? What are the roles of human and machine creativity in your work? How do these roles matter for the audience of your work? and, Is it important to you to limit the human editing of generated results? Why or why not?

## 3.1   Ben-Tal (Sec. 2.1.1)

The *folk-rnn* model contributed material that I treated in much the same way that past composers have when arranging folk tunes, e.g., Bartok, Britten, and Berio. The computer-generated material was "foreign" to me, but served as a source of inspiration as well as a a way of developing my own compositional language. The limitation of the model, in terms of the style it has learned as well as the mechanism of generating material, meant I had to develop different strategies of arriving at a result that I am happy with.

(a) Fragment from the original piece

(b) Randomised pitches
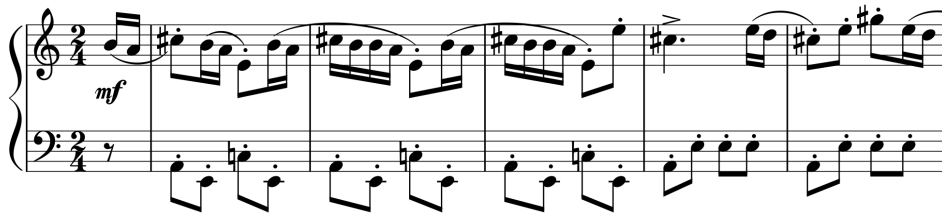
(c) Morphed fragment

Figure 11: A fragment from Kabalevsky's Clowns (a) is pitch randomised (b) and then morphed (c).

My own creativity was the dominant and final arbiter in "Bastard Tunes". But the material generated by *folk-rnn* model v2 both constrained my own creativity and suggested unusual paths, for example, the canonic movement. My own music rarely uses canonic devices, but I instantly recognised the potential in the tune generated by the model. This piece is very different to what I have composed so far, so the "creativity" of the system influenced the piece in significant ways.

On the one hand, the roles of the human and machine do not matter to the audience of my work. As I mention above, I am responsible for this piece as its composer. But there are always multiple ways of listening to any music. One of those modes of listening is to consider the play between human and machine that resulted in this piece. I think this process is evident and could be audible. Whether this is a rewarding way of listening depends on the individual.

Since the measure of success, for me, is the musical qualities of the piece, I don't view the amount of human intervention as an important factor. I also don't think the distinction between tweaking system design and system outputs is meaningful through

22

that artistic lens.[22] As composers, we are not obliged to be consistent or true to a system. Composers should aim to produce the best music they can and if it requires to deviate from whatever system they used to arrive at the result (whether the system is manifested in computer code or not is not relevant here) then so be it. If something isn't quite working (musically) in a piece saying, 'but this is what my system produced' is not a sufficient answer.

3.2 Sturm (Sec. 2.1.2)

Even when *folk-rnn* fails to generate a transcription that does not closely resemble those in its training data, it can still create compelling ideas that provide points of departure for composition. I also find it somewhat addictive exploring the massive amount of material that *folk-rnn* models have generated so far.[23] I enjoy the challenge of transforming some of the generated material into music. Meeting this challenge might involve composing contrapuntal subjects and contrasting melodies, producing harmonic motion, designing accompaniment and orchestration, and piecing together all my favourite bits into a form that produces the kind of music I like to hear. Often I find that my finished pieces arrive at places that I couldn't have imagined otherwise.

In my works here, then, I see the role of computer "creativity" as a source of "raw materials." To enjoy my music, though, I don't think one needs to understand the role played by machine learning. I am not opposed to making clear the role machine learning played in my composition. I think it can add allure to the end result, and invite a different kind of listening. A listener's curiosity and appreciation can be biased (positively or negatively) when they know something not human was involved in a creative work. In another light, I see my works as "advertisements" of my research. In this respect, I am compelled to identify exactly what machine learning contributed.

As regards to limiting the human editing of generated material, it depends on what I am trying to do. If I want to compose a piece of music that I think is successful, I find I can't shackle myself to "staying true" to the verbatim computer output. However, my enjoyment of and patience with this process are limited by the "quality" of the generated material with respect to my own musical voice. I am finding that some models trained on the same material as *folk-rnn* models, but using a different representation, do not generate very much that is interesting for such work.[24]

---

[22]This would be different if the focus is on validating a computational method.

[23]See footnote 9.

[24]Compare the *folk-rnn* material here `https://goo.gl/TyLnp9` with material generated by a model trained on a MIDI representation of the same training data `https://goo.gl/wGttMH`.

3.3 Hadjeres, Deruty and Pachet (Sec. 2.1.3)

Machine learning contributes the expressive power of the *DeepBach* system. It is able to accurately capture the style of Bach chorales and to generalise this style to unseen material in order to produce new contents in the learnt style. It can generate from scratch a totally new chorale or a novel harmonisation of a melody; but *DeepBach* is not restricted to these two specific cases. In fact, any section or any part can be regenerated while keeping all the other notes fixed. Furthermore, additional constraints, such as the key or the end of phrases, can be added by a user in order to have some control on the output of *DeepBach*. The result is that we are able to produce well-written music in no time and with little effort — even if we are not a Bach expert!

Each melody generated by a *folk-rnn* model is the soprano line around which *DeepBach* creates a chorale. We played with the *DeepBach* graphical user interface in order to see if the system managed to deal with such unusual chorale melodies. The human creativity here consists in choosing constraints and in selecting which parts to keep and which to make *DeepBach* regenerate. I don't consider this process as "cherry picking" since we do not choose amongst a bunch of generated solutions the best one. Instead, we can interact with the *DeepBach* system until we are satisfied. This approach is more satisfactory from the composer's point of view since we do play an active part in this computer-assisted compositional process. This creates a human-machine interaction.

As for the audience knowing the roles of the human and machine, it is as important as knowing how Bach produced his own chorale harmonisations. You can enjoy the purely musical content or try to understand more. Knowing the conditions of creation of a musical piece can be enlightening but I'm not sure it will change the way the music itself is perceived.

3.4 Monaghan (Sec. 2.1.4)

In this project, machine learning shaped my continuing consideration of the authenticity and ontology of folk and traditional music. Hence, machine learning contributed to the concept, character, form, length and content of the piece. Machine learning contributed musical material to use, and an emotional and cerebral response I wouldn?t have otherwise had. The system I used emulated a folk music, and the exploration of my response to that was just as important to me as using the musical material the machine produced. We did not directly affect one another's output in real time, but we both affected the product.

The collection generated by *folk-rnn* contains a remarkable number of tunes that are difficult to identify as computer generated. That is, the computer program is so

successful in emulating this music that many of the pieces could pass as belonging to the existing tradition. Some, however, contain nothing untoward in terms of melody, rhythm or structure, but are noticeably mundane and formulaic. Such tunes don't contain any, or enough points of interest, and are harder to learn because there is nothing memorable about them. The resulting pieces of music may be theoretically of the style, but as compositions, are mediocre.

To aim to "fix" what the computer produced did not feel collaborative or interesting; to incorporate its work with my own felt more collaborative. I constrained myself to present some of the computer-generated tunes in their original form. Therefore, I had to find a way to create an artwork I was happy with, despite finding these tunes lacking. In this work, machine learning forced me to engage with material I may otherwise have discounted as unoriginal or below par, prompting a different compositional method and output from myself.

I note that the composition was shaped by my encountering fixed results of machine learning, rather than being involved at an earlier stage. It would be interesting to compare with a project in which machine learning was happening in real time, or in which my actions influenced the machine output. I view "Safe Houses" as a collaboration between *folk-rnn* and myself. The melodies played on concertina have been composed by *folk-rnn*, having learned from collections of notated traditional music. The tape part is composed by me from recordings of my playing as I learned the computer's compositions, but which in performance is played by the computer. We each play the other's work, and the work is made from the process we each went through of learning the other's material.

For this particular work, I would prefer the audience to understand the role machine learning played. I did not compose the melodies played on concertina, and would not like to have them attributed to me. I would prefer the piece to be heard in the context of the collaboration and constraints that produced it. However, I am happy with it as a piece in its own right. Forging a relationship with the audience is a priority in my work, in addition to my own enjoyment of the performance of it, and the quality of the artwork itself. I don't think an audience necessarily needs to know about the process by which the music is produced, to enjoy it. They can connect with it as music without knowing how it was made. However, if they have been told machine learning is somehow involved, but don't understand how, that can often be problematic. I don't think an audience enjoys being confused. There are many other things I would prefer an audience to be doing, other than spending the performance trying to work out how a computer has been used. Where machine learning is used in conjunction with traditional music, I think it is important to be mindful of the sense of ownership,

and strength of feeling, that an audience may have.

Regarding human editing of generated results, my compositional practice is often a result of a multi-stranded process considering in tandem my priorities as composer, performer, music technologist and sound engineer. These roles are not easy to separate and exist on a continuum — I mention them individually here to highlight types of consideration that may be driving the process at any one time. I often use the input of technology — certain sensors, interfaces or software — as a way to shape my other ideas, or I may make compositional decisions based on a starting point from some computer-based sound or process. That is, technological contribution is dispersed throughout my work. In this case, the computer's direct input, the generated pieces, were finished before I began writing. I wanted to use its pieces in their entirety, because the nature of our relationship in this project meant that the computer was not in a position to edit or influence my contribution in a similar way. Also, given the personal connection I had with Irish traditional music, transcriptions of which were used to train the *folk-rnn* model, I was interested in making clear in this piece what part was contributed by the computer. Lastly I thought the audience might be interested to be able to identify aurally the role of the computer on this occasion. These considerations were more straightforward if the output was used unedited. In general, my decisions on human intervention in computer processes in music depend on the specific piece.

3.5   Banarsë (Sec. 2.1.5)

Most listeners of Irish traditional music would probably prefer that they where listening to music made by humans, with no machine involvement at all. It's seen as being created through an organic process — an aural tradition, going back centuries, using instruments made largely of "natural" materials. It currently serves an alternative to the computer-driven, pre-recorded or amplified music, popular today. In the case of the works on the concert, the machine is creating the music, and the human is shaping it to fit within the Irish traditional idiom. It could be shaped differently for another context, or left raw if that was your preference! But in this case, I think it's advisable to limit the human editing, to see exactly what the computer is capable of generating. I tend to edit as little as I can. The more editing I do, the more it would become a joint composition between me and the computer, which I'm not sure is my job here.

3.6 Collins (Sec. 2.2)

Machine learning contributes an ability to objectively analyse a larger audio corpus in a way that a human musicologist would find time-consuming and fatiguing. The objectivity arises because all modelling assumptions are necessarily built into a computer program. The indirect "understanding" of the computer program of the corpus repertoire, as arising via up front representational decisions and machine learning, moves away from the analyst's direct experience to alternative consequences. The corpus can even be vastly larger than any single human musician could listen to in their lifetime, and a transhuman listening state achieved.

As to limiting the human editing of any generated results, it depends on the goal of the algorithmic composition work: certain kinds of scientific modelling of compositional activity might require no editing, and many composition tasks can be very pragmatic. Aesthetically, I probably prefer less tweaking, since it seems purer to algorithmic work, but fully acknowledge that there are some tweaks in the current piece under discussion.

With this great capacity to reach new places available, I tried to use algorithm output relatively unfiltered by myself, excepting that I was of course as programmer able to adjust things until the general nature of output reached a workable form. I took first takes at all opportunities, according to a formal plan (allocating instruments to AIs, length of sections) that I plotted in advance. Leaving it more to the computer helped with stepping back from the music to appreciate the alternative world it had taken me to. I did tweak based on range, breathing issues and the odd annoying phrase, but have left in the majority, and it certainly doesn?t sound like any piece I would have composed directly myself. I can hear the algorithmic middleman quite clearly! There remains some mystery around the fine detail, since no human sat in on every formation detail of the Markov modelling nor feature extraction.

The audience doesn't have to appreciate the details of the learning, excepting that some sort of machine listening and learning process must have taken place to work off the audio corpora. The style emulation becomes a provocative core to the piece, especially in terms of database-savvy contemporary pop musicians and the clash of experimental music and mainstream chart work. New art often reflects new developments in technology, and machine learning and artificial intelligence provide inspiring currents in contemporary culture. The audience might listen for how the original sources have been transmuted via the computational intermediary.

3.7 Herremans and Chew (Sec. 2.3)

Automatic pattern extraction allows MorpheuS to generate music with some semblance of structure. This important step of machine intelligence is what enables the system to mimic complex structures such as themes, repetitions, and transposed variations of the themes from a template piece. A model of tonal tension further allows MorpheuS to generate music that closely replicates the tension profile of the exemplar piece.

Currently, MorpheuS only relies on the creativity of the human composer to create recurring patterns and the rhythmic template. The nature of the patterns which are detected from the original piece, however, depends on the compression algorithm. MorpheuS is responsible for generating entirely new pitches, using only the pattern template provided.

When purely enjoying the musical experience, it doesn't matter if the audience does not understand the role that machine learning played during the music generation. When listeners are evaluating these systems or their creativity, it is however, essential to understand which elements originated from human intervention.

In the pieces generated by MorpheuS, only a small handful of output pieces were created and presented without any manual edits. In order to fully assess the state of current automatic composition approaches, we feel it is essential to present the raw output of the system. While other systems may focus on being a computer-assisted composition system, MorpheuS aims to be fully autonomous. The current limitations of the system do require that an existing piece is taken as input to provide template repetition structures and tension profiles; in the future, some of these may be freely constructed or automatically generated.

4 Audience responses

At the concert of the works described in Sec. 2, we included with the printed program a questionnaire for audience members to complete and return. It asked participants open-ended questions about their favourite aspects, moments they found surprising, and how they were listening differently knowing that computers played a role in creating the music.[25] We received good participation, with 28 of about 50 attendees providing

---

[25]Specifically, the open-ended questions were: 1) What moments or aspects were your favourite? 2) What moments or aspects were most surprising to you? 3) What do you think of when someone says, "a piece of music is composed by a computer"? 4) How did you listen differently to a piece knowing that a computer played a role in its composition? 5) In the piece, "Ed SheerAI vs XenAkIs vs AIdele", what did you hear that you could identify of Ed Sheeran, Iannis Xenakis and Adele? We

written feedback.

In general, the comments given show an excitement around computer partnerships in music. One person wrote, "what matters is how good the music is — not the origin." There was some surprise too about what the results of such partnerships would sound like. About their expectations one person said, "before tonight I would have said mechanical." Another said, "It seems strange to me, and I am sure it is going to be 'modern music'" without specifying what they meant by "modern". Several people said that the concert made them think again about computer-generated music. This seems due in large part because we brought the material to expert musicians to work with and make music out of the system outputs. Some people thought this obscured or mitigated the role of the technology: "one of the strongest determining factors is how the musicians play. They can elevate dull material — and underplay good material." Others echoed the wider concerns people share about this new technology and its effects, "the computer generated pieces 'miss' something — would we call this 'spirit', emotion, or passion?" Another wrote, "I think the science is fascinating and it's important to explore and push boundaries, but I'm concerned for the cultural impact and the loss of the human beauty and understanding of music."

Several people commented on the diversity of the program, mentioning it as a strength, and particularly the fact that the audience heard multiple perspectives of the same system, i.e., *folk-rnn*. This was also evident in the fact that 17 people mentioned specific but different pieces as their favourite aspect of the concert. An interesting result from the questionnaire is the response to our question about how people listen differently when knowing a computer played a role in composing a piece. Some people mentioned that they do not listen any different; but others mentioned "trying to catch the computer," listening more for "messing up" or "something random", or paying more attention to the structure and stylistic consistency.

5 Machine-learning research that matters for music creation

We now discuss how our work applying machine learning to music creation relate to Wagstaff's two principles (Sec. 1).

---

included space for additional comments.

5.1 Concrete impacts for music creation

We have described several different ways in which the machine learning technologies in Table 1 were applied to music creation. These range from generation of an entire piece (*DeepBach*, MorpheuS and ArtIst), to curation from generated materials as well as co-creation (*folk-rnn*). Some of the concrete impacts of machine learning we can identify include: a sense of creating something that could not have existed otherwise; the offering of a "musical other" that could be poked and prodded to produce interesting results (and many that are not interesting); a way to break habits, challenge oneself with something new, and inspire ideas; a way to engage with particular idioms (Irish traditional music, Bach chorales, pop); a way to create arrangements (*DeepBach*) or variations (MorpheuS) in a short amount of time and with limited expertise. Some composers and musicians using these tools, however, want their audience to clearly understand the role machine learning played in their work. Some felt torn between staying true to what the system produced or what their own musical voice says to do.

Certainly, the impacts identified above could be accomplished by means other than machine learning, e.g., algorithmic composition with an expert system, listening to and learning to play traditional music with experienced players, hiring arrangers, and so on. We do not argue that machine learning is the best way to do these things. It must be highlighted though that a major contribution of machine learning to this domain is how its statistical foundation offers a way to mitigate the need to specify rules to bring about particular behaviours of computational systems. Machine learning allows a kind of "metaprogramming" of algorithms by giving examples of the kinds of behaviours one desires instead of requiring hard set rules. The is especially promising for the various under-specified rules embedded within training data that are difficult or even impossible to formally define.

5.2 Informing the research pursuit of machine learning

Our experience informs the research pursuit of machine learning in several ways. First and foremost, it shows that in the application of machine learning to music creation, we must consider a wider range of contributions than simply considering the dichotomy of success or failure of the system in producing that music. The "proxies" we discuss in the introduction — likelihoods of real sequences in a model (Boulanger-Lewandowski et al., 2012; Greff et al., 2016); compatibility of sequences with music-theoretic rules (Jaques et al., 2016); and qualitative listening tests by lay listeners (Jaques et al., 2017) or experts (Collins and Laney, 2017) — do not provide reliable indications of the usefulness of a machine learning technology for music creation. From our experience, we

see that despite a model having a good fit to validation data, and generating material that appears stylistically plausible and musically informed (e.g., stepwise motion, repetition and variation, cadences), its lack of generalisation can be surprising and present impediments to particular modes of music creation (Ben-Tal, "Bastard Tunes", Sec. 2.1.1). Seemingly paradoxically, the failure of a model to generate material that is similar to its training material may actually be a desired mode of operation. Ben-Tal's composition explores and takes advantage of the "naiveté" of the *folk-rnn* v2 model where it lacks training data. Collins' composition depends in its front-end upon the vagaries of chord and melody detection algorithms, which certainly do not perform to the standard of a human analyst on the complex sound stream mixture of full recordings. And despite a model producing an error-ridden output with poor stylistic similarity to its training data, it can still be useful for music creation (Sturm, "The Humours of Time Pigeon", Fig. 6).

In a real sense, we are getting ahead of ourselves with the suggestion that any of these models are learning style, or even something about music. This motivates the second contribution of our work: even with very good performance in these "proxy" evaluations, caution must be taken when discussing what these systems have actually learned to do. Even though a model may appear to be doing the right things, it may be working with concepts that are not very general (Sturm, 2014; Sturm and Ben-Tal, 2017). For instance, the *folk-rnn* models seem to be able to count time and repeat and vary material in ways that are stylistically plausible, but these abilities disappears when the models are pushed even a little outside of its training material. Involving domain experts in working with these models identifies these weaknesses and hence pinpoints areas in which models can be improved.

The above notions of success and failure are at odds with the training regimen of machine learning models, many of which are trained to reproduce validation sequences with high probability. However, that is not to say that this method of training is forever flawed, but that the measures of success in training a model do not translate to measures of success in the real-world use of the model. This motivates the third contribution of our work: models and systems for music creation should be designed such that they can be "calibrated" to human users. It is not easy to define one's preferred mode of operation for a model, but it might be a simpler matter for a model's behaviour to be reinforced in ways that the user finds useful. This moves toward building systems that are not necessarily meant to work "out of the box," but that do have the capacity to adapt to a user's peculiar needs. This leads to landscape of bespoke tools, catering to very specific needs that may be unique to one user; in fact, the history of algorithmic music is peppered with the active involvement of researchers that double as users of

the bespoke technologies they develop, e.g., Dannenberg et al. (1997); Biles (1999); Miranda and Biles (2007); Dean (2018).

A fourth contribution of our work is that it shows the training data of a model does not necessarily limit its application to music creation in ways one expects. Though the *folk-rnn* models are trained on Irish traditional music, our work shows how they are applicable to creating music that does not sound that way at all. Furthermore, even non-musical factors may contribute to a model's success, e.g., some of Sturm's works are inspired by the titles *folk-rnn* v1 generates. Given her knowledge of Irish traditional music, Monaghan's perception of the limitations of *folk-rnn* v2 provided parameters for the fixed media part of her composition (Sec. 2.1.4). She was inclined to make the tape part more layered, dynamic and prominent than it might have been, had it been written for combination with "better" computer-generated tunes. Monaghan's title and theme were inspired by the numbering of generated pieces, the cautious responses to the project from some traditional musicians, and her familiarity with existing research on authenticity, tradition and change in the Irish Traditional Music community (Vallely et al., 1999; O'Shea, 2008; Kaul, 2009).

This highlights a fifth contribution of our work: no music data is independent of its context and function, and the machine learning researcher must respect that. For instance, the MorpheuS works presented in Sec. 2.3 essentially borrow the rhythmic dimension of existing works. The training data of *folk-rnn* comes from a living tradition, which from the outset is misrepresented by a written, "short-hand" notation. Even though the training data is publicly available, it was crowd-sourced with different intentions of use, e.g., to share and preserve the tradition. The treatment of such data with statistical machinery has the high likelihood of being perceived as trivialising a tradition, which is closely related to notions of identity. If machine learning research applied to such a domain is to have a positive impact, it is imperative for the machine learning researcher to reflect on the ethics of what they are doing, and to build bridges to understand exactly how the technologies they are developing can harm or help a particular group (Holzapfel et al., 2018).

6 Conclusion

Our concert of music composed by or with machines (Fig. 1) comes soon after a similar one in London, billed as "the first concert ever in which all of the music played has

been written by a computer",[26] as well as the premier of the "world's first computer-generated musical" (Colton et al., 2016; Collins, 2016; Jordanous, 2017). Applying artificial intelligence to music creation has a long and rich history (Dannenberg et al., 1997; Pearce et al., 2002; Ariza, 2005; Nierhaus, 2008; Fernández and Vico, 2013; Dean, 2018; Herremans et al., 2017), from expert systems, generative grammars and Markov models in the late 1950s (Hiller and Isaacson, 1959), to evolutionary systems in the 2000s (Miranda and Biles, 2007), to the most recent data-driven deep learning systems (Briot et al., 2017). Music compositions fashioned from such methods, as well as books and articles describing them, are plentiful (Cope, 1991; Hiller and Isaacson, 1959; Miranda, 2000; Miranda and Biles, 2007; Roads, 1996; Todd and Loy, 1991). It is important to reinforce the fact that the application of machine learning, and artificial intelligence more generally, to music generation is already quite developed.

Our article contributes to all this work a case study of how machine learning can impact music creation, and how that experience informs machine learning research — two principles of machine learning research that matters (Wagstaff, 2012). We go beyond the "proxy" evaluations typical in machine-learning research, and test models directly in a composition-to-public-concert pipeline. We apply a variety of machine-learning methods and datasets (Table 1) to create models that can generate material independently, or together with a user. We create a variety of new musical works, and exhibit them in a public concert. This experience highlights several opportunities for improving machine learning research applied to music creation. When machine learning is to be developed and applied to music creation, then it should be evaluated in those terms. The holistic experience of domain practitioners working with machine learning technology provides valuable insight into the contexts in which it can be useful or not, and the ways in which its development and application can be improved.

7   Acknowledgments

---

[26]September 28, 2016, funded by Learning To Create (Lrn2Cre8), FET project number 610859 and ERC Advanced Grant 291156.

References

Ariza, C. (2005). Navigating the landscape of computer aided algorithmic composition systems: A definition, seven descriptors, and a lexicon of systems and research. In *Proc. Int. Computer Music Conf.*

Begleiter, R., El-Yaniv, R., and Yona, G. (2004). On prediction using variable order markov models. *J. Artificial Intell. Res.*, 22:385–421.

Biles, J. A. (1999). Life with GenJam: interacting with a musical iga. In *Int. Conf Systems, Man, Cybernetics*, volume 3, pages 652–656.

Boulanger-Lewandowski, N., Bengio, Y., and Vincent, P. (2012). Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. In *Proc. Int. Conf. Machine Learn.*, pages 1159–1166.

Bradley, H. (1999). *The Companion To Irish Traditional Music*, chapter Coleman, Michael, page 75. Cork University Press.

Briot, J.-P., Hadjeres, G., and Pachet, F. (2017). Deep learning techniques for music generation – a survey. *arXiv:1709.01620v1*.

Chew, E. (2014). *Mathematical and Computational Modeling of Tonality: Theory and Applications*. Springer, New York.

Clarke, E. (1999). *The Psychology of Music*, chapter Rhythm and timing in music, pages 473–500. Academic Press, San Diego, CA., 2 edition.

Collins, N. (2009). Form and algorithmic composition. *Contemporary Music Review*, 28(1):103–114.

Collins, N. (2016). A funny thing happened on the way to the formula: Algorithmic composition for musical theatre. *Computer Music J.*, 40(3):41–57.

Collins, T. and Laney, R. (2017). Computer-generated stylistic compositions with long-term repetitive and phrasal structure. *J. Creative Music Systems*, 1(2).

Colton, S., Llano, M. T., Hepworth, R., Charnley, J., Gale, C. V., Baron, A., Pachet, F., Roy, P., Gervás, P., Collins, N., Sturm, B. L., Weyde, T., Wolff, D., and Lloyd, J. R. (2016). The Beyond the Fence musical and Computer Says Show documentary. In *Proc. Int. Conf. Comp. Creativity*.

Cope, D. (1991). *Computers and Musical Style*. Oxford University Press.

Dannenberg, R. B., Thom, B., and Watson, D. (1997). A machine learning approach to musical style recognition. In *Proc. Int. Computer Music Conf.*, pages 344–347.

Dean, R., editor (2018). *Oxford Handbook of Algorithmic Composition*. Oxford University Press, Oxford, UK.

Drummond, C. and Japkowicz, N. (2010). Warning: Statistical benchmarking is addictive. kicking the habit in machine learning. *J. Experimental Theoretical Artificial Intell.*, 22:67–80.

Fernández, J. D. and Vico, F. (2013). AI methods in algorithmic composition: A comprehensive survey. *J. Artificial Intell. Res.*, 48(1):513–582.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Trans. Neural Networks and Learning Systems*, 28(10):2222–2232.

Hadjeres, G., Pachet, F., and Nielsen, F. (2017). DeepBach: a steerable model for Bach chorales generation. In *Proc. Int. Conf. Machine Learning*, pages 1362–1371.

Hansen, P., Mladenović, N., and Perez-Britos, D. (2001). Variable neighborhood decomposition search. *J. Heuristics*, 7(4):335–350.

Herremans, D. and Chew, E. (2016a). Morpheus: Automatic music generation with recurrent pattern constraints and tension proles. In *Proc. IEEE TENCON*.

Herremans, D. and Chew, E. (2016b). Tension ribbons: Quantifying and visualising tonal tension. In *Int. Conf. Technologies for Music Notation and Representation (TENOR)*, pages 8–18.

Herremans, D. and Chew, E. (2018). Morpheus: Generating structured music with constrained patterns and tension. *IEEE Trans. Affective Computing*.

Herremans, D., Chuan, C.-H., and Chew, E. (2017). A functional taxonomy of music generation systems. *ACM Computing Surveys*, 50(5):69:1–30.

Herremans, D., Weisser, S., Sörensen, K., and Conklin, D. (2015). Generating structured music for bagana using quality metrics based on markov models. *Expert Systems with Applications*, 42(21):7424–7435.

Hild, H., Feulner, J., and Menzel, W. (1992). HARMONET: A neural net for harmonizing chorales in the style of J. S. Bach. In *Neural Information Processing Systems*, pages 267–274.

Hiller, L. and Isaacson, L. (1959). *Experimental Music: Composition with an Electronic Computer*. Greenwood Press.

Holzapfel, A., Sturm, B. L., and Coeckelbergh, M. (2018). Ethical dimensions of music information retrieval technology. *Trans. Int. Soc. Music Information Retrieval (in press)*.

Jaques, N., Gu, S., Turner, R. E., and Eck, D. (2016). Generating music by fine-tuning recurrent neural networks with reinforcement learning. In *Deep Reinforcement Learning Workshop, NIPS*.

Jaques, N., Gu, S., Turner, R. E., and Eck, D. (2017). Sequence Tutor: Conservative fine-tuning of sequence generation models with kl-control. In *Proc. Int. Conf. Machine Learning*.

Jordanous, A. (2017). Has computational creativity successfully made it "Beyond the Fence" in musical theatre? *Connection Science*, 29(4):350–386.

Kaul, A. (2009). *Turning the Tune: Traditional Music, Tourism, and Social Change in an Irish Village*. Berghahn Books, New York.

Mauch, M. and Dixon, S. (2010). Approximate note transcription for the improved identification of difficult chords. In *Proc. Int. Soc. Music Info. Retrieval*.

Meredith, D. (2015). Music analysis and point-set compression. *Journal of New Music Research*, 44(3):245–270.

Meredith, D., Wiggins, G. A., and Lemström, K. (2002). Method of pattern discovery. *PCT patent application number PCT/GB02/02430, UK patent application*, (0211914.7).

Miranda, E., editor (2000). *Readings in Music and Artificial Intelligence.* Harwood Academic Publishers.

Miranda, E. and Biles, J. A. (2007). *Evolutionary Computer Music.* Springer-Verlag, London.

Nierhaus, G. (2008). *Algorithmic Composition: Paradigms of Automated Music Generation.* Springer.

O'Shea, H. (2008). *The Making of Irish Traditional Music.* Cork University Press, Cork, Ireland.

Pearce, M., Meredith, D., and Wiggins, G. (2002). Motivations and methodologies for automation of the compositional process. *Musicae Scientiae*, 6(2):119–147.

Roads, C. (1996). *Computer Music Tutorial.* The MIT Press.

Salamon, J. and Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Tran. Audio, Speech, Lang. Process.*, 20(6):1759–1770.

Schedl, M., Gomez, E., and Urbano, J. (2014). Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval*, 8(2-3):127–261.

Sturm, B. L. (2014). A simple method to determine if a music information retrieval system is a "horse". *IEEE Trans. Multimedia*, 16(6):1636–1644.

Sturm, B. L. (2018). What do these 5,599,881 parameters mean? an analysis of a specific lstm music transcription model, starting with the 70,281 parameters of its softmax layer. In *Proc. Music Metacreation workshop of ICCC*.

Sturm, B. L. and Ben-Tal, O. (2017). Taking the models back to music practice: Evaluating generative transcription models built using deep learning. *J. Creative Music Systems*, 2(1).

Sturm, B. L., Santos, J. F., Ben-Tal, O., and Korshunova, I. (2016). Music transcription modelling and composition using deep learning. In *Proc. Conf. Computer Simulation of Musical Creativity*, Huddersfield, UK.

Todd, P. M. and Loy, G. D. (1991). *Music and connectionism.* The MIT Press.

Vallely, F., Hamilton, C., Vallely, E., and Doherty, L., editors (1999). *Crosbhealach an Cheoil: The Crossroads Conference 1996: Tradition and Change in Irish Traditional Music*. Whinstone Music.

Wagstaff, K. L. (2012). Machine learning that matters. In *Proc. Int. Conf. Machine Learning*, pages 529–536.