Studi di estetica, anno XLVI, IV serie, 3/2018
ISSN 0585-4733, ISSN digitale 1825-8646, DOI 10.7413/18258646064

Luca Casini, Marco Roccetti[1]

The impact of AI on the musical world: will musicians be obsolete?

Abstract

Artificial intelligence (AI) is going through a period of renewed interest and success thanks to the rise of neural networks, the staple of the so-called deep learning. Creating a computer program capable of writing believable music has been tried since the 1960s, with lackluster results. Composing music seemed something beyond the potential of machines, but recent developments in the field are challenging this conception. This article will explore the latest developments falling at the intersection between artificial intelligence and music and then investigate what possible impact such new technologies may have on the musical world, from a technical as well as an aesthetic standpoint, trying to demystify some common misconceptions and worries.

Keywords
AI, *Deep learning*, *Music*

1. *Introduction*

This article proposes some considerations regarding the impact that AI could have on music, focusing on the aesthetics aspects of the matter with the intention of giving a common ground for discussion for the benefit of computer scientists, philosophers and musicians.

In the first paragraph, we briefly summarize what artificial intelligence, machine learning and neural networks are and how they may be employed in the context of musical applications. Such presentation is in no way exhaustive, it addresses the least technical reader, aiming at giving a basic knowledge of the terminology and the inner workings of the mentioned technologies.

The second paragraph provides a brief overview of the state-of-the-art research topics in music classification, generation and production

[1] luca.casini4@studio.unibo.it, marco.roccetti@unibo.it.

applications using a specific branch of artificial intelligence, namely deep learning.

Finally, a few considerations are made regarding the aesthetics of AI generated music, focusing on the implications this may have on the artistic experience. In our conclusion, we explain why we may doubt about the dystopian fears which are expressed by some, identifying a series of limitations, both aesthetic and technical, that we should prevent from the rise of AI composers in the near foreseeable future.

2. *Machines that learn*

AI is one of the most discussed topics of the recent years and it seems to be taking by storm every field it is applied to: from industry to entertainment. What most may be unaware of is the fact that the whole field of AI is very old and is built upon many different disciplines and technologies that have been used in the effort of creating a computer program able to reason. Since its beginning, interest in AI came and go as new technologies promised great and unparalleled results that in turn did not meet the expectations. We are now living an AI renaissance, powered by what is referred to as *deep learning.* This term comes from the fusion of the words *machine learning* and *deep neural networks*.

Machine learning (ML) is a field of Computer Science and Statistics that concerns the development of probabilistic models allowing computer to learn to perform some tasks (usually predictions) by looking at some data, without being explicitly programmed to do so. The procedure of feeding that data to the algorithms is called *training* as the performance of the program is expected to increase as it runs.

Artificial neural networks (ANN) are one of the tools of ML and, as the name suggests, are inspired by biological neurons. They were theorized in 1943 (McCulloch and Pitts 1943), but were soon dismissed because of their limited capabilities. The main issue with those initial models were the limited size and depth, and the lack of an efficient way to train those models. In the 1980s both of those problems have been addressed thanks to the availability of more computational power and the introduction of the *backpropagation* algorithm (Rumelhart *et al.* 1986). The results reestablished ANN as a viable model for ML. The scarcity of data and computational power, however, did not allow ANN to show their full potential. In fact, they seemed destined to oblivion until in the 2000s the Internet made enormous amounts of data available, a.k.a. *big data*, and advancements in parallel computation made computational time more tolerable. The neural networks used today are made of hundreds of neurons stacked on tens of layers, hence the term *deep neural network* and *deep learning.*

Deep learning models started beating state of the art solutions in many problems like machine translation, language and image recognition, prediction and content suggestion, anomaly detection to name a few of their fields of application. Those models often required no intervention that might require domain-specific knowledge of the data, apart from the necessity of building clean and coherent datasets.

Music creation, based on probabilistic techniques, has been attempted ever since the idea of machine learning has existed, and even before there has been research on generative models and algorithmic music. The first notable example is represented by the ILLIAC Suite written by Lejaren Hiller in 1957. Since then, most of the research has moved on implementing techniques and models borrowed from the world of natural language processing, as the two problems share many similarities: both are hierarchically structured flows of symbols that exhibit interdependence through time.

The initial works in the field were based on generative formal grammars, constraint programming and Markov models. Formal models based on rules have very limited applications, though: usually the rules implemented counterpoint in the style of Bach, and the resulting musical content is not always pleasant. Probabilistic models like Markov Chains on the other hand are more expressive, but also more complex and difficult to train and manipulate.

As machine learning progressed and artificial neural networks became its preferred technique, the performance of such models increased. The advent of *recurrent neural networks* (RNN) has been a game changer for many applications that deal with timeseries like music and text and swept away previous state of the art models based on Markov chains. RNN are an architecture in which each unit can retain a memory of its previous state and is thus able to learn patterns that repeat in time. It is usually represented as a normal neural network with the addition of a connection looping back to itself for each neuron. Although performance improved, the output of those models still lacks long-term direction and coherence and it works well only when limited to a genre and structure, as in the work of Eck and Sturm.

Right now, the real challenge is to create something that is coherent in the long term, while steering the creation of music according to some desired feature or style, in a similar fashion to the results obtained by Gatys with images. For the first goal, researchers are experimenting with hierarchical architectures in hope of abstracting higher-level features regarding structure and repetition. Style and interpretability are tackled with the use of generative models based on latent space representation like variational auoteconders and generative adversarial networks; this category of models can learn how data is distributed in the high dimensional space of the features which can also represent high level concepts like note density or style.

3. *The state of the art for AI in musical applications*

There is a lot of interest in the creation of deep learning models that can understand and work with music under various perspectives. The first to be investigated was the classification of music based on genre, instruments or any other feature that may be of interest. Prior to the advent of deep learning this kind of task, referred to as *music information retrieval* (MIR), was performed using complex systems that used manually engineered features based on acoustics and psychoacoustics considerations.

With neural networks the same results, if not better ones, may be achieved using the simple audio signal with little to no additional work. Spotify, for example, uses deep learning to power its recommendation system, combining data regarding user listening habits with data on the musical content of the song they just listened; Shazam is using it to recognize songs in no time with incredible accuracy, through the identification of the unique footprint each song has, as represented in their database.

On the other hand, we have models that can produce music, either as raw audio signals, or in a symbolic representation. Learning from and generating directly raw audio has been tried with very poor results; Google WaveNet (Van Den Oord *et al.* 2016) is the only experience that seemed to produce acceptable content, but the results were short burst of piano music, without any identifiable context, nothing more than a novelty from a system designed to do mainly vocal synthesis.

The vast majority of systems focus, instead, on the generation of symbolic music, that is music notated in some format that is then to be played by either a human or a machine. Symbolic music is represented in a variety of formats, the most common being *MIDI* and *abc notation*. Those two formats have been chosen mostly because of the vast availability of data online, but they are not the most practical formats out there.

Data availability has also greatly influenced the choice of musical genre to train those neural networks. Classical music is one of the most used as is highly available and does not pose any problem for what concerns copyright laws. In the context of classical music there is also an additional bias as the corpus of Bach's chorales has become one of the standard datasets for polyphonic music generation.

Another genre that is very present in the literature is Irish folk music, thanks to a vast database of abc-notated music that is free to download. Models using Irish folk music reached among the best results, mainly due to the constrained form of the genre and the reduced clutter of the abc notation compared to MIDI data that makes the task easier.

Jazz and rock music are less available and are also often more irregular in their structure, a characteristic that makes them more difficult to work with. For the same reason most models, up until now, have focused on the generation of monophonic music or at best polyphonic music for one instrument, like the piano. Hao-Wen Dong's work with MuseGAN is one of the first attempts at band arrangement and showed promising results: the instruments are actually playing "together" but the music starts to feel a little random after a few measures.

The best results have unsurprisingly been achieved by privately funded research groups with more resources than most universities. Google, with its project Magenta, is very active on this front and has put out many interesting models that can compose melodies, both monophonic and polyphonic, and drum grooves, play duet with a human, improvise over a chord progression, and, with their last model, interpolate between two melodies in a semantically correct manner as well as creating three-part music (drums, bass, melody) that plays significantly better than any other similar work. In a recent paper, they also explored a new model that learned the style of a certain genre of music, to transfer it to another piece with unprecedented efficacy.

The goal of the Magenta project is to provide the basis for the creation of a set of tools that can enhance musical creativity leveraging deep learning. Those tools are sometimes just an upgrade of more traditional musical equipment, for example the drum machine-like components of MusicVAE (Roberts *et al.* 2018), their latest model. Some other times they allow new for expressive possibilities that were not achievable before, as in the case of the NSynth, a model that is able to learn the features that characterize a certain sound at the wave level not only to reproduce it, but also to reproduce any hypothetical instruments that would exist somewhere between two real ones. For example, it can create a sound that shares the characteristic of a snare drum and flute, but that is a not a trivial mixing of the two original sounds.

Another project that achieved remarkable results is Flow machines, led by F. Pachet and backed by Sony. It garnered mainstream attention when they released a song called *Daddy's car*, the music and the lyrics of which were both written in the style of The Beatles by their AI composer. Looking more closely at their work it appears how the song was actually written using one of the tools they worked on, called *flow composer*, that is an intelligent assistant to a human composer and can create new songs as well as filling the blanks in a given piece or generate stylistically coherent accompaniment chords for a given melody; it is clear however that the song was arranged and performed by a human musician. This research group also created other interesting tools that focus on the interaction with musicians to create interesting performance. The *continuator* is a system that can listen and then play along with a pianist, learning its style and producing coherent musical responses to his playing. The *reflexive looper* can be seen as an evolution of this technology and it is an upgrade of the traditional looper, a device that repeats a short musical fragment after it is played once, where the musician's input is not only looped but also arranged according to the particular style of the song being played with the addition of other instruments if requested. All these projects are not based on neural networks, like all the others, but on a technology called *constrained Markov models* (Pachet 2011) that combines classical machine learning *Markov chains* with the field of *constraint programming;* the first offers the capability of learning style and the latter enforces long-term structure and enables creative control on the output.

The mentioned projects show a much more interesting approach than one that tries to substitute humans: the creation of new interactive instruments and tools that can allow musicians and creatives to enhance their expressive capabilities beyond their actual capabilities. Empowering the artist instead or replacing them and creating a feedback loop to collaboration that leads to better art and better tools. Tools like NSynth also shows how AI can be used to create a whole lot of new instruments that were simply not possible and, in this way, contribute to the creation of new aesthetics in the same way that synthesizers, drum machines and samplers were fundamental in the genesis of electronic dance music, hip hop and their related genres.

Two aspects that are rarely considered and not very researched, to the best of the authors' knowledge, are the application of machine learning to execution and production.

The first problem is about creating a model that starting from a symbolic representation of some piece of music like a MIDI file, the most common format for controlling virtual instruments and also used to represent sheet music, can output a realistic performance of it by playing the notes with appropriate timing and dynamics. An experiment in this direction was attempted by Google Magenta with their *performanceRNN* model (Simon and Oore 2017), that was trained on data from real performances of classical musicians and generated new music in the same performing style, but without following a musical sheet. If listened for more than a few seconds the wandering nature of the music becomes clear, but is still very believable as a performance and could be mistaken for a pianist doodling on the keyboard.

The second is learning all the parameters that concern the world of musical production, like equalization, compression and equalization. Those techniques are obviously important only in the context of recorded music, but it can be argued that for many genres, most notably electronic music, the experience of the music in its recorded form is as important, if not more, as the musical content itself. There is not as much interest in this topic as in the others we mentioned previously, but it is still possible to find researchers, like Mimilakis, who are working on their applications to sound separation, mixing and mastering.

4. *Technical limitations and aesthetic considerations*

When people hear about new progress in the world of AI it is inevitable that someone will think about how machines will substitute humans in every task in a matter of years, with a sense of impending doom on the human race; this kind of reaction, fueled by years of science fiction in the popular culture, are understandable but based on misconceptions on the true capabilities of AI. On the other hand, there are those, often artists, who proudly stand saying that a computer will never be able to replicate the product of a human artist, as what a computer does may not be regarded as an artistic product, but rather a mere calculation.

The truth is obviously somewhere in between. The current state of the art in musical generative models based on machine learning has already reached an impressive stage and for the first time there seems to be interest in the field coming from sources other than academia. However, it must be noted how we are far from a system that can autonomously create a composition that can be considered acceptable for the average listener. Still, we would not exclude the possibility that one day a machine could create music as proficiently as humans.

One of the biggest issues that has yet to be overcome is the fact that artificial neural networks struggle to grasp the long-term structure that is typical of music. This problem is shared with other fields like NLP, where generating a long text is difficult, even with a constrained structure, like poetry.

As hinted before, the way of representing music is also a limiting factor, as working with music that does not fall in the common 4/4 time-signature or features irregular rhythms like triplets means adding a significative computational load. Similarly, the number of subdivisions of each bar of music is a very important factor; many works do not exceed the limit of 16^{th}. Those limitations are a double-edged sword: while they reduce the complexity of the model, they avoid the creation of simple and regular rhythmic patterns that are more pleasing to the ear. Hence, the expressive power of those models is severely limited and will not be able to reproduce more nuance and uncommon music.

Finally, it must be noted how, while the output of the models is musical and coherent, high level features like style are not easily learned and reproduced. The only case in which style is actually learned is when the training set comprises only pieces of that specific style. However, creating, for example, a model that can transform a song by changing its arrangement so that it sounds like another genre of music is something that is beyond the current capabilities.

These limitations will probably be surpassed in the foreseeable future, but it is still unlikely that a human musician will be replaced, even by a very advanced AI composer. The reason for this is that at the center of the musical aesthetic experience the musician often prevails over the music s/he plays. This is true for both pop music and for niche genres that have the status of art, as jazz and classical music.

In pop music this can be seen by looking at how people do not care about the fact that the actual singers are backed up by a team of producers and are often not involved in the creation of the music that they perform, or even worse that their singing voice is altered in order to produce a more suitable performance. People appreciate the music and the figure of the pop star and those two things are bound together. If the producers of Madonna decided to write the same songs for another singer the result would not be the same because the character Madonna is as important, if not more, as the musician Madonna. This is akin to the famous example of the *Brillo boxes* made by Arthur C. Danto: the value of the box was not given by the box itself, but by the fact that they were by Andy Warhol. On the subject of pop music, it is probable that in the future AI tools will streamline the process of production of popular music by major labels. This would not really affect the fruition of pop music as the creation of it is already "industrialized" in the way it is carried out. For the listener knowing that there was a machine instead of a pool of 10 producers behind a singer would make no difference. The presence of musicians in each step of production would still be necessary as they are needed both to operate AI powered tools and to record and perform the music live.

The same can be made, and are even more true, for genres where the musicians play their own music and are perceived as legitimate artists by their public and the critic, like jazz, experimental rock and singer-songwriters. Enthusiasts are often more focused on the nuances of the artistic expression and tend to be less captivated by the personality and more by its skills. Saying this, we are not trying to downplay the importance of cultural factors, especially given that it was central in shaping genres like jazz and rock music, but just to highlight a different approach and a different sensibility from two different kinds of listeners that still share an underlying aesthetic mechanism.

A slightly different world is that of classical musicians that are appreciated, instead, for what they can contribute by interpreting in their own personal way a piece of music written by someone else; the aesthetic experience is split between the author and the performer. In this context the music generated by an AI can be played by real musicians, but it remains the loose end in the evaluation of experience: who should be considered the author, the machine or the human that designed it?

The importance of the human musician is reinforced by the fact that recorded music has not stopped people from enjoying live music. Even if the musical content is the same, the modality of its experience is so fundamentally different that the two things can be considered something totally different. The act of seeing a musician perform preserves the "aura" that is described by Walter Benjamin and that recorded music lacks. Similarly, we can say that there is another aura that comes not from witnessing the performer create on stage in front of the audience that AI in its immateriality lacks. If musicians survived the age of recorded music and easy reproduction, they will surely survive the advent of musical AI

Looking at the matter from the opposite perspective, however, we may not exclude that in the future AI generated music could become successful as music genre itself, as people may be interested and curious about the result of an AI that freely plays, also in settings where it interacts with human musicians. This kind of experience would obviously be fundamentally different from a concert with human performers and/or composers, and for this reason these two forms of art could coexist without problems. On the other hand, in 2017 a study by Chamberlain tested the aesthetic appreciation of art produced by a computer or a robotic agent and found out that there are negative biases towards the quality of computer-generated art and its ability to be a compelling work of art. They also observed that people responded differently according to the degree of anthropomorphic quality of the agent creating the work; seeing the robot create the artwork from scratch and having a robot with a more humanoid appearance both increased appreciation of the final product, the presence of mistakes in the realization of the artwork also conditioned the public. If there is ever going to be a culture around AI music, its success is bound to the way it is presented to the public.

Listening to music generated and played solely by a machine brings up the question of who the actual author of the music is. If experiencing art is experiencing what the artist wanted his audience to experience, can we really say that the machine had any will in its creation?

Or is the human that programmed it that is in a way responsible for the result? In addition, is the machine actually creating or is it just the emergent result of statistics applied to the learned data? Can it be considered completely new? Most of those questions are hard to answer and depend on the specific model; if we are talking about what is available today, then the authors' opinion is that we cannot talk about music that has the same value of human composed music, and we would not consider it art. This is not because it is not completely original, as it is arguable that human musicians are original in the first place and so is the value placed on originality, but for the lack of will in the act of creation: current artificial intelligence does not want to convey a message, but is instead outputting the most likely results under some metric, where variations are a consequence of the intrinsic randomness of those models.

This lack of will of the artificial intelligence makes it also unable to create anything like those styles of music that present a strong connotation to the environment they emerged from like songwriters that talk about political and social themes, hip hop and rap music that is deeply linked to the experience of living at the margins of society and dealing with racism or music connected to subcultures like punk music. In all those cases the work is packed with meaning that goes beyond musical content, but that still influences the musical form of it. AI can at its best emulate those things from the examples it has learnt, but they mean nothing without the intention behind them.

5. *Aesthetics of interactivity*

As stated before, works focused on interactivity are yielding the most interesting results and we can legitimately expect that these kinds of technologies will have a greater impact on the world of music in the short term. Pieces of music created and performed using those interactive tools and techniques entail a series of ulterior aesthetics consideration beside those of the previous paragraph.

A detailed account on this can be found in the work of Garnett. He analyzed the aesthetics of interactive computer music focusing on what interaction with a human performer implies in the context of computer music, without the pretense of giving a clear definition of aesthetic and music or formulating any theory. Most of his considerations were formulated before AI music was as relevant as it is today, but the point that he makes still stands. The way deep learning is used to create music is not different in its modalities from other algorithmic techniques and they only distinguish themselves in the results and the implications given by the learning process. We will now review the most important parts that are relevant to our analysis.

Garnett's work starts by first describing what a human performer brings to computer music, namely gestural nuances, physical and cognitive constraints and interpretation. Gestural nuances like dynamics, rubato (changes in tempo) and articulation, are very difficult to obtain in computer music, if not by painstakingly inputting every one of these subtle variations; AI systems are on their way to surpass the limitations of old computer music, but still, achieving this kind of flexibility is difficult and the presence of a human performer is an easy solution. By physical and cognitive constraints Garnett intends that computer music without a human performer runs the risk of being "unperformable" and thus "incomprehensible" because the figure of a performer puts a limit on the cognitive level: if a musician must play it then it must be able to understand it. While algorithmic music may incur this problem, deep learning models are trained on real "performable" music and if they learned correctly, they should generate music that retains this characteristic. The concept of interpretation presents two aspects though: first, the musician's performance has an inherent variability that makes each performance of the piece a unique experience for the listener; second, the whole of the possible interpretations adds to the aesthetic value of the work and surpasses its original intended meaning.

Garnett then looks at the opposite side of the spectrum and analyzes what the computer brings to human performance. He considers all the possibilities in which the computer becomes a "cyber-instrument", an extension of the player.

First, the ability of computers to process the sound of a live instrument in a passive way (like reverb or delay effects) or by giving the control of this processing to the same musician playing the instrument or to another musician on stage. This augments the skills of the artist that can obtain new results with different instruments while retaining his/her skills.

Second, modifying the instrument with sensors or other equipment can create new correlations between certain gestures and sounds and add to the expressive capabilities of the player, without taking away nothing. These new possibilities force the musician to re-evaluate the way he plays the instrument and to practice new things in order to obtain the desired results. This in turn conditions the way the instrument is played in this context and leads to new music that was not possible before.

Third, computers can be used to create completely new instruments that directly bridge the musician to the machine, presenting a very different interaction mode. They can also be used to create instruments that do not correspond to any existent paradigm, requiring musicians to put a great deal of effort to become proficient. While the authors are skeptic of those drifts which lead away from traditional paradigms, we cannot but also recognizes the potential that this may represent for the birth of new aesthetics.

We conclude stressing how the 20th century music has used electronic and computer sound as a symbol of alienation and this led to the “aesthetic of the machine” as something counterposed to humanity. It is clear now that technology is something that is part of our everyday lives and can be integrated harmoniously; new generations see technology as something natural, an extension of their lives. In this sense interactive computer music should surpass the aesthetic of the machine by not making technology the goal and instead putting the performance, and the performer, at the center and consider this performance as a “shared human endeavor”. Technology as an extension of the musician can avoid the audience becoming “detached and emotionally removed” as it often happens when there is no human presence on stage and can make the music relevant and alive even in the future and most importantly make the audience feel part of something they all experience together.

6. *Conclusions*

This article aimed to describe the current state of the art in the field of automatic music generation based on machine learning techniques and to give a common ground for computer scientist and philosophers to understand the implications of such a technology. Neural networks are becoming an important tool for computer scientist and software developers and they are slowly permeating every aspect of our daily lives. Still, they are a very new technology, and many have not a clear understanding of their functioning, current capabilities and potential.

The introductory part was intended to demystify some of the unrealistic expectancies put on neural networks by marketing campaigns and give an idea of the actual state of the art for deep learning in musical applications. After this, we looked at the implications of those new technologies to the musical world, focusing on whether or not an AI composer could ever replace human musicians. A series of considerations were made looking at the matter from both the technical and the aesthetic angle:

- there is a series of technical limits to the current models, namely the impossibility of creating long compositions that exhibit a coherent structure, the difficulty to learn high level features like musical style, the computational weight of symbolic representations that leads to datasets of less complex and varied music;
- speaking of aesthetics, we highlighted the importance of the human figure of the performer and of the composer over the musical content that is conveyed. What really matters is the experience of seeing and witnessing someone play or sing and knowing that there is a creative will behind their interpretation or composition. The necessity for the presence of the artist in the enjoyment of music remains even in the hypothesis that the technical limitations were to be surpassed;
- following the work of Garnett we delved into the implications of interactive computer music. While the article speaks about computer music in its most general sense, most of the consideration still hold true when talking about AI music. In particular, the authors conclude that performers and machines can augment each other on the aesthetic level and the aesthetic of interaction can overcome the "aesthetic of the machine" that characterized the 20th century by bringing back the human factor both in the performance and in the audience experience.

Finally, while AI is having and will continue to have a great impact on the world of music and art, the prospect of it replacing human musicians, or artists in general, is highly unlikely first and foremost for reasons of aesthetic nature that are not tightly bound to the technological advancement of our world. What is already happening and hopefully will push the boundaries of art is the emergence of new musical tools powered by AI that enhance the expressive capabilities of performers and composers, opening up new creative possibilities.

Bibliography

Benjamin, W., Underwood, J.A., *The work of art in the age of mechanical reproduction*, London, Penguin, 1998.

Briot, J.P., Hadjeres, G., Pachet, F., *Deep learning techniques for music generation. A survey*, arXiv preprint arXiv:1709.01620, 2017.

Casini, L., *Variational automated music generation using variational autoencoders*, M.Sc Thesis, Univeristy of Bologna, 2018.

Chamberlain, R., Mullin, C., Scheerlinck, B., Wagemans, J., *Putting the art in artificial: aesthetic responses to computer-generated art*, "Psychology of Aesthetics, Creativity, and the Arts", n. 2/12 (2017), pp. 177-92.

Danto, A.C., *The transfiguration of the commonplace: a philosophy of art*, Cambridge, Harvard University Press, 1981.

Dong, H.W., Yang, Y.H., *Convolutional generative adversarial networks with binary neurons for polyphonic music generation*, arXiv preprint arXiv:1804.-09399, 2018.

Eck, D., Schmidhuber, J., *Finding temporal structure in music: blues improvisation with LSTM recurrent networks*, in *Neural networks for signal processing, 2002. Proceedings of the 2002 12th IEEE workshop*, IEEE, pp. 747-56.

Garnett, G.E., *The aesthetics of interactive computer music*, in "Computer Music Journal", n. 25 (2001), pp. 21-33.

Gatys, L.A., Ecker, A.S., Bethge, M., *Image style transfer using convolutional neural networks*, in *Proceedings of the IEEE conference on computer vision and pattern pecognition*, Los Alamitos, IEEE, 2016, pp. 2414-23.

Goodfellow, I., Bengio, Y., Courville, A., *Deep learning*, vol. 1, Cambridge, MIT Press, 2016.

Hiller, L., Isaacson, L.M., *Illiac suite, for string quartet*, New Music Edition, vol. 30, no. 3, Bryn Mawr, Theodore Presser Company, 1957.

McCulloch, W.S., Pitts, W., *A logical calculus of the ideas immanent*, in *Nervous activity*, "The Bulletin of Mathematical Biophysics", n. 5/4 (1943), pp. 115-33.

Mimilakis, S.I., Drossos, K., Virtanen, T., Schuller, G., *Deep neural networks for dynamic range compression in mastering applications*, in *Audio Engineering Society convention 140*, Audio Engineering Society, New York, MIT Press, 2016.

Pachet, F., Roy, P., *Markov constraints: steerable generation of Markov sequences*, "Constraints", n. 16/2 (2011), pp. 148-72.

Papadopoulos, G., Wiggins, G., *AI methods for algorithmic composition: a survey, a critical view and future prospects*, in *AISB symposium on musical creativity*, vol. 124, Edinburgh, MIT Press, 1999, pp. 110-7.

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., Eck, D., *A hierarchical latent vector model for learning long-term structure in music*, arXiv preprint arXiv:-1803.05428, 2018.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., *Learning representations by back-propagating errors*, “Nature”, n. 323/6088 (1986), p. 533.

Simon, I., Oore, S., *Performance rnn: generating music with expressive timing and dynamics*, Magenta Blog, 2017.

Simon, I., Roberts, A., Raffel, C., Engel, J., Hawthorne, C., Eck, D., *Learning a latent space of multitrack measures*, arXiv preprint arXiv:1806.00195, 2018.

Smith, B.D., Garnett, G.E., *The education of the AI composer: automating musical creativity*, Urbana-Champaign, University of Illinois Press, 2012.

Sturm, B., Santos, J.F., Korshunova, I., *Folk music style modelling by recurrent neural networks with long short term memory units*, in *16th International Society for Music Information Retrieval Conference*, Malaga, ISMIR, 2015.

Van den Oord, A., Dieleman, S., Schrauwen, B., *Deep content-based music recommendation*, in *Advances in neural information processing systems*, Cambridge, MIT Press, 2013, pp. 2643-51.

Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kavukcuoglu, K., *WaveNet: a generative model for raw audio*, in *SSW*, Grenoble, International Speech Communication Association, p. 125.