

Painting Music: Using Artificial Intelligence to create music from live painted drawings

Andrew Starkey¹, Kate Steenhauer², Jack Caven

¹School of Engineering, University of Aberdeen, UK, AB24 3UE. Email: a.starkey@abdn.ac.uk

²Kate Steenhauer Limited, web: <http://katesteenhauer.com> Email: katesteenhauer@hotmail.com

Abstract

This paper describes the development of Artificial Intelligence (AI) techniques that monitor a painting or drawing evolving in real time and produce musical notes that relate to the individual elements of art as the artwork develops on the canvas. The paper describes the practical approach required to capture the artwork unfolding in real time, and then describes the framework used to develop the correlations between visual art and music. The AI technique exploits these areas of similarity within the two distinct artforms in order to respond to the live-painted elements and produce musical notes that reflect the development of the evolving artwork. A prototype of this system was implemented in a live stage performance at Aberdeen May Festival 2019 whose narrative centred on the question *Is AI good or bad?* Other outputs of this project are a 20-minute film and a body of (tangible) visual artwork for digital platforms and gallery environments informed and inspired by AI. The integration of these disciplines through AI transforms a static artform into one that is dynamic, interactive, transformational, transient and temporal.

Keywords: Artificial Intelligence, live drawing, music improvisation, drawing performance

Introduction

The combination of visual art and music has an unequivocally compelling nature, with remarkable areas of similarity between the two artforms. Many artists [Kandinsky, 1910; Vi et al, 2017] have explored multisensory ways of perceiving and to a certain extent have examined the analogy of mark-making to sound. This project however uses cutting-edge Artificial Intelligence (AI) techniques to examine correlation and harmonious balance between calligraphic-style ink drawings and music, leading to the translation (conversion) of tangible marks into audible sounds (and in theory vice versa however that is not described in this paper).

Calligraphic-style mark-making (using the fluid medium of Japanese ink) lends itself greatly for creating a musical dialogue on paper consisting of a dynamic sequence of movements unfolding in time and pulsating in rhythm, while capturing the essence of a pose, gesture or any other subject matter.

AI has previously been applied to for example the problem of individual letter recognition [Graves, and Schmidhuber 2009, Graves et al 2009], and this project extends this in an intelligent manner to some of the fundamental parameters underlying visual art. However, the project also looks to explore and reflect on the impact of AI on society using the visual and aural outputs of the process. Previous studies [Fast and Horvitz 2016, Kelley et al 2020] have shown how the perception of AI by the public has changed over time, and that although

the public has been more optimistic than pessimistic when discussing AI, the fear of loss of control of AI has been increasing in recent years.

In practical terms, this means that the goal of the project is not to produce a process that always results in harmonious music and melodies, but that the AI can make mistakes. This question is interesting in a musical context since purely harmonious notes may be considered uninteresting by some people but beautiful by others. By the same token, “mistakes” in the musical process could result in an interesting and novel musical form for some listeners but be jarring for others to listen to. The exploration of this question is at the heart of this project and intends to tap into the public’s unconscious fear – or excitement – of AI and the change that this will bring to all aspects of society.

Similarly, in terms of the visual component, whilst the purity of the drawing discipline is not compromised, conflict transpires through the subject of each drawing, but also through the application of dramaturgical drawing gestures which ‘destroy’ or transform each drawing through either rolling, brushing paint, pouring water, or scrunching up the drawing itself. The choice depends on what type of energy, drawing effect and ultimately what meaning in the scene is called for. The drawing can be partially or completely obliterated by this process, or more often is transformed into abstract marks which forces the audience to consider the meaning of the drawing and also the emotional reaction produced by the development of the drawing.

Transforming a relatively conventional artwork that an audience has ‘invested’ in, has a pronounced effect, often evoking a strong (and sometimes immediate) emotional reaction, for example expressed as a gasp or shock running through the show. This also serves to reinforce the underlying philosophy of the drawing process combined with the musical output: i.e. that the drawing is not a static artform but is dynamic, interactive, transient, transformational and temporal.

This project has been developed with a focus on live stage performance with the outputs from the artist unfolding on a projector screen in real time and the musical outputs from the AI process amplified through loudspeakers. This multi-disciplinary collaboration celebrates the art of live drawing and highlights the dynamic capacity and interplay between drawing and music.

The application of AI in this novel manner reveals new insights into connections between seemingly unrelated compositional elements of these two art forms, exploring harmonic or dissonant comprehensions of the constituent forms and attributes of the pieces.

Drawing in a performative setting

Within the project the artwork consists of painted drawings using calligraphic style mark-making in Japanese ink, as can be seen in Figure 1 to 4, which show scenes from the performance at the 2019 MayFest Festival at the University of Aberdeen. The system is initially trained and tested through digitised static photos of previously developed artworks, and accordingly a system is developed that is able to interact with live-evolving (and thus wet) painted drawings (where the artist is the instigator with AI outputs to follow).

Whilst promoting drawing as a significant discipline in its own right, the oriental drawing materials and techniques are crucial to the goals and performative nature of this project, possessing the following characteristics:

- Minimalistic, 'clean', one-layer mark-making allows a good basis to break down and analyse a piece of art according to the fundamental elements and principles of art, and useful as a framework to correlate them to its musical counterparts as described in more detail below.
- These inks have a relatively high saturation/ purity or strength. Each drawing is monochromatic in nature with a range of tonal values obtained by diluting the ink with water.
- The ink's fluidity and calligraphic tool enable dynamic, varied and descriptive mark-making, which therefore lends itself to a varied musical output.



Figure 1: Top dog (Japanese ink, 700 x 560mm)

Real-time data capture

Within the live performative setting data is gathered in the following way:

- **Image Capture:** The system requires a laptop to be connected to a camera system that is situated immediately above the canvas. The camera used in the project was a GoPro Hero 4 Silver. This is connected to the laptop via BlueTooth or local WiFi and allows the laptop to take high-quality (static) photographs of the canvas (12MP). This process is controlled by the operator of the laptop, who monitors the artist and takes photos (following a predetermined signal by the artist) after each element in the drawing is painted. This is to ensure that the hands of the artist (or hair or head) are not captured in the image.
- **Image Analysis:** The laptop is running concurrently a computer program that has been developed in Matlab [Matlab 2019], a powerful software package commonly used within academia and the engineering field. This program is continually running and monitors for new images to be saved into a specified folder used by the GoPro camera, and will then subsequently begin the process of analysing the images. The first step of the entire process is to manually identify where the boundaries of the canvas lie, so that anything outside of this boundary is ignored in subsequent image analysis.

The program then analyses the image following the removal of anything outside boundary of the canvas and also compares the image against the previous image (if there is one) in order to clarify the details of the new element added to the canvas. The image is then reduced to the size of the new element rather than analysing the image in the context of the whole canvas, this allows individual elements to be more easily compared against each other. Image analysis of this new element is undertaken and includes the following image processing outputs:

- **Analysis of the amount of whitespace (i.e. no paint) in the element.** This is calculated by literally counting the number of pixels in the image that can be considered to be white, and can be converted into a ratio so that this is independent of the size of the painted element.
- **Analysis of the amount of painted canvas in the element.** This is calculated by subtracting the number of white pixels in the image from the total number of pixels in the image, and can be converted into a ratio so that this is independent of the size of the painted element.
- **Analysis of the variance of density of colour in the element.** This is accomplished by analysing the image once it is converted into a greyscale, and binning the output so that the first bin represents the colour that is closest to white, with other bins representing increasing strength in the painted colour. As before, this can be normalised so that the bins represent a ratio of the whole image so that this is independent of the size of the painted element.
- **Two-dimensional (2D) frequency analysis of the element.** This is accomplished by undertaking 2D Fourier transform of the element [Frigo and Johnson 1998]. The subsequent output is then binned into 10 separate bins which reflect the frequency density of the painted element in 10 equally spaced bins which range from low frequency to high frequency.

Musical notes are then subsequently produced based on this image analysis and also based on previous elements painted on the canvas (discussed further below).

This process continues until the operator indicates that the last element in the drawing has been done and the drawing is complete. This gives a final section of musical notes to be added and played by the computer, and then the music will stop.



Figure 2: *The Physical brain (Japanese ink, 600 x 550mm)*

Correlation between visual art and music

Basic (classical) definitions of the elements and principles of art are used to establish an initial framework that can subsequently be correlated to music theory. Examples of the relationships that are proposed by the authors between these worlds is shown in Table 1.

Table 1: *System attributes used in correlating art with music*

Measured Artwork Attribute	Music Counterpart
White (non-painted) Space	Silence
Painted Pixels	Velocity
Tonal Value	Pitch Range
Frequency	Number of Notes and Duration

There are clearly many more attributes that can and should be used, and this will form the focus of further work. However, the attributes described in Table 1 allow some of the fundamental properties of the painted elements to be found, and then matched against elements that are considered by the authors to be equivalent in the musical world.

The image analysis gives an output that is a ratio of the amount of white space (or non-painted area) for the new element. This is considered to be equivalent to the amount of silence in any new notes to be created. Similarly, for the ratio of painted pixels for the new element, this is considered to be equivalent to the velocity of the musical notes to be created. The tonal value in the image, measured by the strength of colour, can be considered to be inversely related to the pitch range of the musical notes, so that strong tonal strength represents low pitches, and light tonal strength represents high pitches. Finally, the frequency analysis of the element can be used to give a guidance to the system for the number of notes to produce when creating a musical motif, and the duration of the notes. This means that higher frequencies in the painted element (i.e. a more intricate image) will result in more notes in a musical motif that will be played at a faster pace, and vice versa.

Clearly, some of these correlations are subjective, with different interpretations possible. However, even if different approaches were to be applied, the resultant musical output would be similar as long as the individual elements identified above were matched against each other in some manner.

The program was then able to determine for any given painted element the attributes in music using the above mapping, and could then generate a sequence of notes to be played in real time.



Figure 3: Child's play (Japanese ink, 560 x 700mm)

Development of AI solution

The requirements for the real-time solution are for a system that can interpret the visual (physical) components of an element within a work of art and to match those elements against their musical equivalents. This process is made up of several steps, namely: image analysis of captured photograph of canvas; understanding and categorisation of image

analysis; re-interpretation of this categorisation into musical notes; production of musical notes in combination with other painted elements.

This leads to the development and implementation of a Self-Organising Map [Kohonen 2001], a type of neural network and AI technique that is loosely based on the way that our own brains work and that is able to automatically categorise input data which can then subsequently be used to make decisions based on the categorisation made.

Since the correlations between the visual art and music worlds has been defined as in Table 1, this means that the inputs into the SOM allow it to categorise the painted element that has just been created, and to then use that categorisation to define musical outputs using the correlations defined above. For example, an element such as that encircled in the top left panel of Figure 4 is a relatively simple element with a relatively large amount of whitespace and a block type structure. This implies that the musical element should also be simple, with few notes, with the possibility of silence, and due to the size and tone of the painted element should be of low frequency. However, an element as shown in Figure 4 has a more complex structure, indicating more musical notes to make up a harmony. As the tonal value of this second piece is higher, this means that the notes produced will begin at a higher pitch.



Figure 4: Community (Japanese ink, 1100 x 550mm). Circles in top panels illustrate art elements in image analysis with (left) dominance of low-frequency and (right) dominance of high-frequency

The analysis of the SOM using the Quantisation Error (Kohonen 2009) of the node that the element maps to (i.e. the decision made by the AI of which previously seen images the current image most closely resembles) allows a range to be defined for each of the musical outputs that are equivalent to the image analysis undertaken, since the image features that act as inputs into the AI process have direct musical equivalents as described and discussed above. Within these ranges the AI can then choose values for each of the musical equivalences defined above so that the process can begin the construction of the musical notes.

In this manner the SOM was initially trained by feeding it over more than 200 digitised static art images (captured by camera) previously created by Steenhauer. It should be noted that these artworks were different from those painted in the live performance, meaning that the SOM had not been trained on artworks or elements that were used in the real time environment. Each artwork itself is broken up into many individual components as well as the whole composition generating an image database of over 1000 pieces.

The categorisations (or attributes) described in Table 1 were applied within these pieces of art using image analysis techniques, obtaining numerical data for each painted element that was used to train a SOM. The SOM was made up of 25 neurons in total in a 5x5 arrangement, trained and tested, simulating the task that will be carried out in a live production.

Construction of musical notes

Once the process has defined the musical parameters with which to construct the music, the process needs to use a different technique to produce the actual notes. This uses a Markov Chain (MC) [Gagnius, 2017] model which is a statistical approach for giving a probability of the next state space. This can be used in conjunction with historical pieces of music in order to learn the probabilities of different notes following each other, or for which notes to use when constructing chords and harmonies, and has for example previously been applied in this manner for the analysis of Beethoven's Ode to Joy [Linskens, 2014].

This approach requires a number of pieces of music to exist (for example, several songs by the Beatles) and these are then analysed in terms of the notes that make up the musical piece. The MC model can then analyse all the notes within the songs presented to give a probability of choosing a second note given a first note. The first note is determined from the musical outputs of the SOM process as defined above, and then a random number generator is used to select the second note according to the MC model. This process can be continued for as many notes as are required for this image, and the required number of musical notes to be created.

The MC model approach can also be used to define which notes are used for harmonics or for chords. The MC model will learn the probability of notes being used as a chord or harmonic from the existing songs analysed, which can then be used in a similar way to give a choice of note to be used in real time by the process.

It should be noted that due to the random nature of this process that the same notes are not chosen each time. In addition, the probability of notes being chosen can also be altered by choosing different songs or compositional styles in the training of the MC model. For example, in the 2019 Aberdeen May Festival live performance, separate MC models were trained on collections of songs by the Beatles, by Phillip Glass, and by Beethoven for different scenes within the show. Each scene within the performance therefore had music created in real time by an MC model using songs from different artists.

This step was undertaken to short cut the musical knowledge that would ordinarily be required to create musical notes in real time that would have a harmonious sound. However, this is a limitation of the current approach and further work is required to develop an AI alternative that will have a rudimentary understanding of musical rules to decide what notes to produce. This would also enhance the musical outputs of the project and increase the merit of the work to the audience.

The process then will choose whether to create a new melody, or to create harmonies or chords, depending on whether painted elements have already been analysed, and if so then on a random choice so that the process does not choose the same each time. Again in the future this will be improved to include analysis of the evolving composition of the artwork in order to make this decision.

The system thus developed is therefore capable of exploiting the areas of similarity defined in Table 1 in order to produce novel musical notes that reflect the painted elements in real time.

A Musical Instrument Digital Interface (MIDI) synthesising toolbox is then used to produce the final aural output.

However, as discussed earlier in the paper the goal of the project is not always to produce perfectly harmonious music, and the random nature of the choice of notes in the creation of a musical motif, the length of notes within this motif and the notes chosen when playing chords or harmonies will sometimes result in music that is not pleasant to listen to. This is entirely by design and is to force the listener to consider the question of whether the AI is doing a “good” or “bad” job in creating the music. One listener may react differently to another since this is an entirely subjective question, and will also depend on the listener’s view on AI itself – i.e. it will bring to the surface the individual’s excitement or fear over the increasing usage of AI in all parts of our daily lives.

In addition, the development of this system gives the ability to the artist to create a multi-faceted and multi-disciplinary artistic expression through both visual and aural dimensions, over which the artist has full control, as well as the ability to explore the relationship between the drawings, the musical output of the system, and the audience themselves. This means that mark-making has a direct implication not only for the composition of the drawing but also on the development of the musical piece, and in turn the audience. For example, composing the same drawing but starting and ending with different elements will result in a very different musical output, depending on the characteristics of these elements. Therefore although drawing is often considered a static art form, the musical output that is based on the mark-making transforms this into a temporal and transient output which is dynamic and depends on the order and choice of mark-making by the artist.

The artist can control the musical output by the choice of compositional elements – for example the scene shown in **Error! Reference source not found.** is comprised of a number of figures which forces the AI to create a number of musical motifs and chords that are similar in their tonal range whereas the scene shown in **Error! Reference source not found.** has very distinct compositional elements that will result in a much different musical output.

The nature of the drawing process means that the mark-making will always be different even if the same drawing is undertaken. This is also the case for the musical output due to the built-in random process for the development of the notes and chords. This lends the entire system to a performance with no two performances being alike, and results in additional dimensions to the generated artistic piece.

Conclusion

Project **Painting Music** uses AI to create music from live-painted drawings. The system employs a SOM to analyse individual elements of an artwork as they are being drawn, and

from this produces music in real time and unique for each production. The system shows that it is possible to create a truly multi-disciplinary artform where tangible visual marks in one domain are used to drive the novel creation of audible sounds/music in another.

Painting Music is a live stage performance, but other outputs of the project include a 20-minute film and a body of visual artwork. This work is a prototype that will be built upon by incorporating a greater number of art elements, as well as the basic principles of art, and subsequently enhancing the correlation to the fundamental parameters of music theory. In addition, it will look to enhance the process of musical improvisation using theory and practice from this field and building AI approaches that will remove the need for heuristics in the music generation process.

The system developed thus far allows the artist to directly control the musical output of the system by the choice of compositional elements, and the sequence in which they are drawn and thus in turn their relationship with the audience. This gives an additional dimension to the artistic capabilities of the piece and moves visual art as a traditionally static experience into a dynamic interactive artform that is temporal and transient in nature. This is emphasised in the production by the transformation of the artwork once it is complete in a traditional sense, as illustrated in the Figures in this paper, giving an additional dimension to the artistic capabilities of the overall production.

Further work is required to enhance the integration of both art and music theory into the process, and will look to involve experts from both fields to develop the AI engine that drives the improvisation of the music that is created in real time. This is also important in terms of the question of novelty and innovation in the artistic piece. Currently the focus on novelty is with the artist's process and product, however the project would greatly benefit from integration of musical theory into the AI process. The integration would also create the possibility of undertaking the reverse of the system described in this paper; to create novel painted drawings in real time from the sounds/music as they are being played.

Painting Music offers a unique experience and taps into current demands for interactive and evolving 'stories', that contain elements of risk and reveal the artistic process to the audience. The production process brings attention to the art of live drawing and highlights drawing's dynamic capacity with other artforms, technology and their relationship with its audience. Intrigue within each production is created hooking the audience through our own questions centred around the narrative *Is AI good or bad?*, and exploring thoughts and fears over the application and impact of AI, and its prevalence both today and in the future.

Links to online outputs **Painting Music**

[1-minute highlights live stage performance](#)

[1-minute highlights of film](#)

[20-minute film](#)

[Project website and visual artwork](#)

References

Gagniuc, P. A. *Markov Chains: From Theory to Implementation and Experimentation*. USA, NJ: John Wiley & Sons. pp. 1–235. ISBN 978-1-119-38755-8. 2017

Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, J. Schmidhuber. A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, 2009.

Graves, A., and Schmidhuber, J.. Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks, in Bengio, Yoshua; Schuurmans, Dale; Lafferty, John; Williams, Chris K. I.; and Culotta, Aron (eds.), *Advances in Neural Information Processing Systems 22 (NIPS'22)*, December 7th–10th, Vancouver, BC, Neural Information Processing Systems (NIPS) Foundation, 2009, pp. 545–552, 2009

Fast, Ethan, and Eric Horvitz. “Long-Term Trends in the Public Perception of Artificial Intelligence.” (2016): n. pag. Print.

Frigo, M., and S. G. Johnson. “FFTW: An Adaptive Software Architecture for the FFT.” *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. Vol. 3, pp. 1381-1384, 1998

W. Kandinsky, *Concerning the spiritual in art*, Dover Publications, 1910

Kelley, Patrick Gage et al. ““Happy and Assured that life will be easy 10 years from now.”: Perceptions of Artificial Intelligence in 8 Countries.” *ArXiv abs/2001.00081* (2020): n. pag.

Kohonen, T., Ilari T. Nieminen, and Timo Honkela. On the Quantization Error in SOM vs. VQ: A Critical and Systematic Study. In *Proceedings of the 7th International Workshop on Advances in Self-Organizing Maps (WSOM '09)*. Springer-Verlag, Berlin, Heidelberg, 133–144. DOI: 10.1007/978-3-642-02397-2_16., 2009

Kohonen, T. and Schroeder, M. R. and Huang, T. S., *Self-Organizing Maps*, Springer-Verlag, Berlin, Heidelberg, 3rd edition, 2001

LINSKENS, E. *Music Improvisation using Markov Chains*, Maastricht University, 2014

MATLAB Release 2019, The MathWorks, Inc., Natick, Massachusetts, United States, 2019

Vi, C.T., Ablart, D., Gatti, E., Velasco, C., and Obrist, M. Not just seeing, but also feeling art: Mid-air haptic experiences integrated in a multisensory art exhibition. *International Journal of Human-Computer Studies* 108, 2017, 1–14. ISSN 1071 5819; <http://dx.doi.org/10.1016/j.ijhcs.2017.06.004>