Artificial Intelligence in Music

Fig 1 - Image generated using HyperGAN

Name: Jo Kallset
Student number: 15021283
Course: Creative Music Technology
Module Title: Individual Music Technology Project and Portfolio
Supervisor: Simon Waite

Acknowledgements

First and foremost I would like to thank Simon Waite for supervising this project. Throughout this project he has provided massive support and helped make this project come to fruition.
I would like to thank Aaron Dolton and Richard Harper for technical support throughout this project. I would like to thank Alex Hough for providing massive technical support for this project. Without Alex this project would not have been possible.
I would also like to thank all of the people that helped me evaluate the installation and my peers at university for providing feedback on this project.

Abstract

This dissertations objective is to investigate the use of Artificial Intelligence (AI) in the form of Neural Networks (NN) in interactive art. This project achieves this by making an interactive multimedia installation that uses NNs to generate audio. Along with research into interactive installations this is used to understand the interactive elements required in a performative autonomous computer system.

Research into contemporary AI use were done in order to get a broader understanding how how this technology can be implemented in media applications. Research into NN architecture and algorithms where also undertaken in order to understand how the technology works, allowing this project to use a suitable NN architecture.

Thorough testing into audio generation using NN was done. This helped the project better understand the unpredictable nature of NNs and how to best overcome the challenges posed when using this technology in an autonomous system. Testing into how an audience interacts were also done, allowing this project to get a better understanding of the challenges posed with interactive systems and how they are greatly affected by the audiences actions.

Portefolio

This project is published on my online portfolio at www.kmusiclab.com
On this page there will be published additional material including videos and further comments on the project. This website also contains previous and current projects that I'm working on that are related to sound technology. Some of these projects include:

- Building synthesizers from scratch using microcontrollers and electrical components.
- Programming plugins for ableton using the Max For Live environment (Ableton, 2018).
- Creating custom midi controllers.
- Creating interactive audio visual performances

Contents

1. Background **5**

2. Aim and Objectives **8**

2.1 Aim 8

2.2 Objectives 8

3. Research **9**

3.1 Neural Networks 9

3.2 Perceptrons 10

3.3 Convolutional Neural Networks 11

3.4 Recurrent Neural Networks 12

3.5 General Adversarial Network 13

3.6 Improvisational Musical AIs 15

3.7 Generating Musical AIs 16

3.8 Formatting and Interpreting Data 17

3.9 Interactive Installations 20

3.10 Evaluating A Performance 22

4. Method **23**

5. Results **25**

5.1 Neural Network Prototyping 25

5.2 Hardware 26

5.3 Visual Aesthetic 28

5.4 Programming Interactivity 29

5.5 Feedback 32

6. Conclusion **34**

7. Evaluation **35**

7.1 Neural Network Systems 35

7.2 Interactivity 35

7.3 Experimentation 36

7.4 Learning Outcomes 36

7.5 The Future of Interactive Arts 37

References **38**

Appendices **43**

1. Background

The arrival of advanced and powerful AI systems is reinventing the way society operates. From bank systems, national health services to governmental paperwork, AIs are proliferating in society (Sample, 2017). In 2014 Google spent £263 million on the AI startup *Deepmind* (Gibbs, 2014). Personal assistant systems such as Apple's *Siri* are helping people organise their lives (Apple, 2017), and in 2015 Tesla released a self-driving car that uses machine learning and AI to gather data in order to drive better and safer over time (Fehrenbacher, 2015). The use of this technology in the field of music is also becoming more prevalent with the release of tools such as *Izotope Neutron* and *Landr*. These are tools that use AI to help musicians mix and master music by analysing audio and applying appropriate effects and processes based on an analysis of pre existing released music (iZotope, 2017; LANDR n.d).

One approach on how to utilize AI in musical systems is through improvisational AIs. In 2017, *The Robotic Musicianship Group* at Georgia Institute of Technology presented a series of robots that play alongside humans as part of a performance. The AI system they made use machine learning to process a large amount of musical strophes and melodies. The system is then given a seed in the form of a four bar melody that it can use to create new melodies. It can then be used to play alongside other human performers and will listen to the raw data from the sound that their instruments make. After analysing this information, the system will automatically evolve and add that information into its memories in turn altering how it will play alongside human performers (Maderer, 2017). This allows artists to integrate an AI system into a performance without programming the systems individual actions beforehand, making it less time consuming for live musicians compared to programming an instrument. This also adds an interesting element of liveness through infinite variations in composition.

Another way of utilizing AI can be seen in the computer program *Wekinator*, made by Dr. Rebecca Fiebrink. The program uses machine learning to map human actions to control parameters for other musical applications. This can be used as a gesture recognizing tool and utilizes AI as a tool rather than the main component of a performance (Fiebrink, 2017). One example of this system in use is by using a webcam as a gesture controller to trigger the sounds of a drum machine. The system can be taught to play cymbals when a hand is in the web camera image and make it play a cowbell when your face is in the image. One would use Wekinator to record what pixels are what color at a given position and train it with more examples of the same action to make it more accurate.

Magenta is a research project started up by engineers and programmers from the *Google Brain* team. *Magenta* is not run by the google brain team alone and as such individual people from AI communities all over the world contribute to the development of their individual projects. *Magenta* started up with the goal of answering the question “can machine learning be used to create compelling art and music?”. With one of their latest published works, *Nsynth,* they try to answer this question by inventing a new form of audio synthesis. *Nsynth* utilizes deep neural networks to generate sounds one sample at a time. By learning directly from the raw data of an audio sample, *Nsynth* allows artists to generate sounds that use a combination of data from several audio samples. Artist will then be able to choose different sound characteristics from several sounds and create a new sound with a timbre and tone quality that would be unachievable with other forms of audio synthesis (Nsynth, 2017).

Another one of *Magenta’s* published works, *Performance RNN* is utilizing AI as a compositional tool. *Performance RNN* is a computer program that uses Recurrent Neural Networks (RNN) to compose polyphonic music with expressive timing and dynamics. By teaching the RNN musical compositions in the form of a database of MIDI files, Performance RNN takes on the expressive qualities from the source material and generates a new MIDI file (Oore et al, 2017). In addition to publishing the source code for *Performance RNN*, *Magenta* also published a web-browser version of Performance RNN. This allows artists to access a pre trained network that generates midi data in a web browser. The generated midi data is easily routed to a music software on a computer, allowing artists to utilize musical NNs without having the knowledge of how to create one (Thorat et al 2017).

Looking at ways that contemporary artist are using AI technology in creative applications inspired this project to explore how AI can be used as an expressive tool in musical performances. By looking at NN audio generation techniques and human interaction with computer systems this project will aim to create an interactive audio visual installation. The installation will analyse speech and sound patterns in real time and recreate audio based on this information. This will allow the project to look at how AI technology can be used in creative applications.

2. Aim and Objectives

2.1 Aim

This project aims to explore human interaction with AI technology. This will be done by creating a installation that uses NNs to recreate audio based on audio recordings of a surrounding audience

2.2 Objectives

- Research and evaluate contemporary interactive installations
- Implement NNs in a multimedia installation
- Implement interactivity in a multimedia installation
- Evaluate the project through audience feedback
- Create an online portfolio to showcase the installation and the process behind it

3. Research

The research will cover basic NN algorithms as well as how more complex NN systems work on a fundamental level. Topics related to data transformation and interactive systems in musical performances will also be covered within the context of art and computer programing in the field of music.

3.1 Neural Networks

Computer scientists have been using the human brain as a source of inspiration for a long time. In 1943, computer scientists Warren S. McCulloch and Walter Pitts came up with a conceptual idea for an artificial neural network. Their work describe a neuron as a single cell in a network of cells that receives and process inputs as well as generating an output. This artificial neural network was not meant as an accurate representation of the brain but rather a computational model, inspired by how the brain works (Shiffman, 2012).

This computational model proved important in solving one big problem in computing, that you have to give the computer specific instructions. If you asked a human to get a pencil on the table in front of it, the human would get the pencil. But if you asked a computer the same thing, you would have to give it specific instructions as to how many steps to take in a specific direction and so on. What deep learning and NNs allows you to do is to give the computer more abstract commands such as: Take a step forward in the direction that decreases the distance between the pencil, then, repeat until you reach the pencil. This would allow the computer to get to the pencil from anywhere in a room without a human giving it any other orders than to get the pencil (Serrano, 2016).

3.2 Perceptrons

NNs can be programmed in different ways, each with specific attributes that help them solve widely different tasks. However at the core of a neural network one will always find artificial neurons, mathematical models meant to represent biological neurons. There are several types of artificial neurons. One of the earliest models that was developed is the Perceptron.

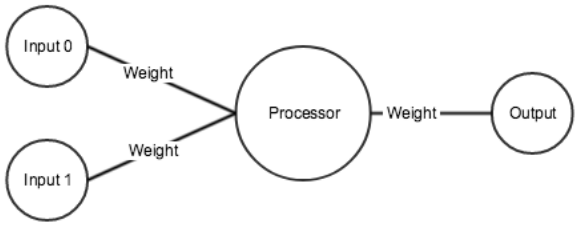


Fig 2 - A single Perceptron processing several binary inputs (Shiffman, 2012)

A Perceptron operates by analysing several binary inputs and then generating one binary output. The Perceptron was developed in 1950s and 1960s by Frank Rosenblatt. He proposed to use “Weights” in order to compute an output from the perceptron.
Weights represents the importance of individual inputs and determines how much of a specific input would affect the output of the NN (Shiffman, 2012).
The Perceptron uses a simple structure in its algorithm but when formed into a network, complex forms of decision making are made possible within the network of perceptrons (Nielsen, 2017; Picton, 2000).

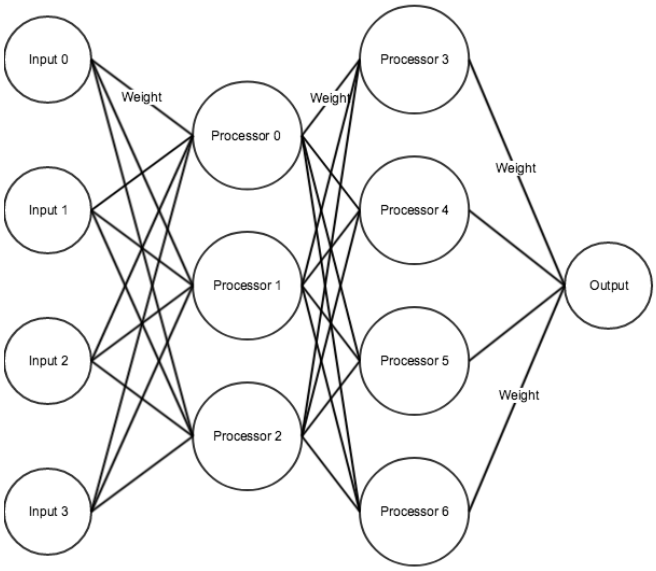


Fig 3 - A network of perceptrons allowing more complex decision making (Shifman, 2012).

3.3 Convolutional Neural Networks

As well as having different types of artificial neurons, NNs can use different algorithms to determine weights and have different network structures ranging from simple single layered networks to convoluted multi layer networks. One such type of NN is a Convolutional Neural Network (CNN). As shown in figure four a CNN uses several layers of artificial neurons to filter the incoming data, each layer with their own filtering techniques. By doing this a CNN is capable of analysing data and finding patterns in smaller sections of the data instead of looking at the input as just one section. When analysing an image of a cat the CNN would be able to recognize features of the cat like its paws or its fur. The CNN would combine the number of features found, generating a probability of the image containing a cat (Rohrer, 2016).

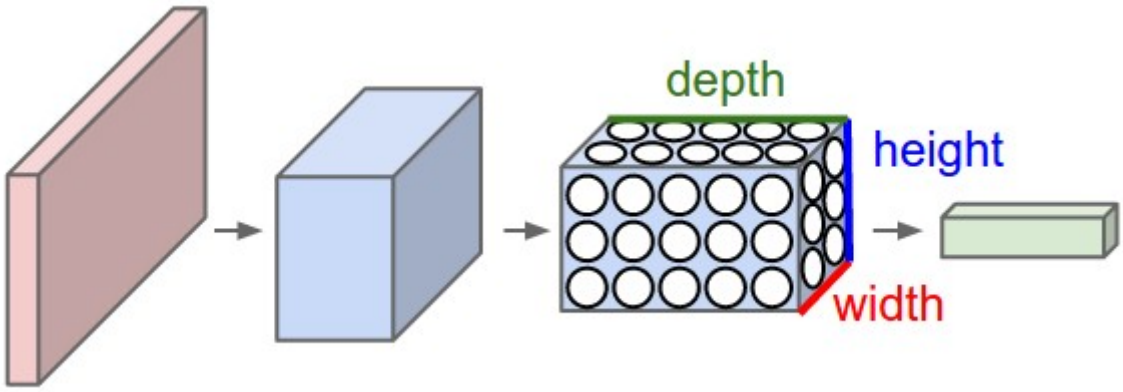


Fig 4 - CNN processing an in in 3 dimensions. Height and width can be used to analyse height and width of an image while the depth represents the different filters used to analyse the input (Karpathy, n.d.)

3.4 Recurrent Neural Networks

Another type of neural network is a Recurrent Neural Network (RNN). An RNN works in sequences storing the generated output of the network and using that to calculate future outputs. This becomes useful when working with a large database of data. One application would be to create character level language models. One could make a network that had the goal of generating the word ‘fish’. If the only data given to the network was two words, ‘fish’ and ‘fist’. The probability of the network placing an ‘i’ after ‘f’ will be 100% and the same applies to putting ‘s’ after ‘i’, but when it decides what comes after ‘s’ it will have a 50% probability of either generating another ‘h’ or ‘t’. However if you give the network more inputs, let’s say a sentence with three repetitions of ‘fish’ and seven repetitions of ‘fist’ the probability of the network generating ‘t’ after ‘s’ will change to 70% (Karpathy, 2015).

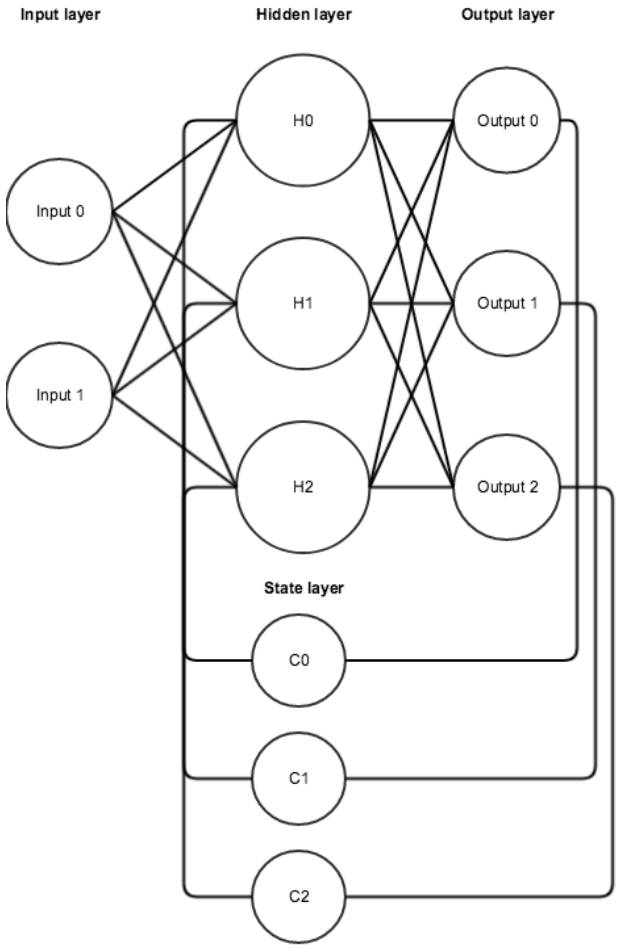


Fig 5 - A simple RNN model called a *Jordan* network, this uses the output stored in the state layer to affect the next input allowing you to apply the network to time-varying patterns of data (Jones, 2017).

3.5 General Adversarial Network

RNNs and CNNs are two neural network models that excel at generating a generalized output, a General Adversarial Network (GAN) will on the other hand create a more realistic or randomized output. To Illustrate this we will look at the generation of an image of a face. If trained with a CNN the neural net would look at the individual elements of a face, the eyes and the mouth and average these elements generating a face with perfect eyes that always use the same shape. With a GAN on the other hand the generated face would have variations in the shape of the eyes and mouth because the GAN doesn't look at these features with the intention of averaging them but rather to produce an output that would be within a certain range.

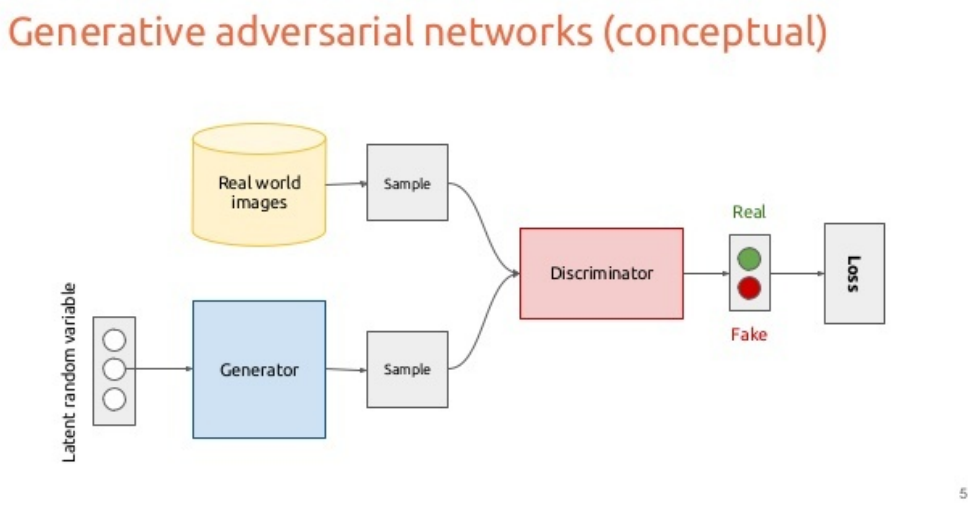


Fig 6 - GAN network structure (燕家猫, 2016)

As seen in figure 6 at its core a GAN works by combining two neural networks, One is a Discriminator and the other one is called Generator. The Discriminator is a classifier that is given an image and classifies the image by outputting a number between zero and one determined by how well the image matches the networks previous inputs. If you trained the network by inputting images of cats and the image you send to the input matches the patterns it learned from the previous inputs it will give you a number close to 1. The Generator is a network that uses random noise as input. The generator will use this noise to generate a pattern that it thinks matches the previous inputs given to the Discriminator. In order to know what to generate, the generator gives its output to the Discriminator which in turn will tell if the input matches the data you trained it with. Over time The Generator and Discriminator will then train each other, the generator feeding the Discriminator with what it believes is images of cats, and the Discriminator telling the generator if it think its a real cat or not, alternating between being fed real images and images generated by the generator.

The generator would at first generate random noise, then over time become better at generating images that matches the real data given to the Discriminator. The Discriminator would also learn whenever the generator generates an image the Discriminator think is real. If the generator manages to do that, the Discrimitor will try to find another feature or pattern in the data that will distinguish the real images from the generated ones.Theoretically, given enough time the result would be a NN that will be able to classify training data perfectly and a generator that can generate images indistinguishable from the real training data (Computerphile, 2017; Goodfellow et al, 2014).

3.6 Improvisational Musical AIs

In 2017 *Magenta* released *Performance RNN*, a Long Short Term Memory (LSTM) RNN. LSTM were first introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997. They designed the LSTM NN model in order to achieve better long term memory modeling than existing NN's. As shown in figure 9 at its core an LSTM use a convoluted algorithm for changing the cell state of each cell in the neural network. In order to change the cell state of the neuron the input has to go through three gates, each deciding how much of the input information should change the cell state of the cell. This helps regulate and control the cell state (Olah, 2015).

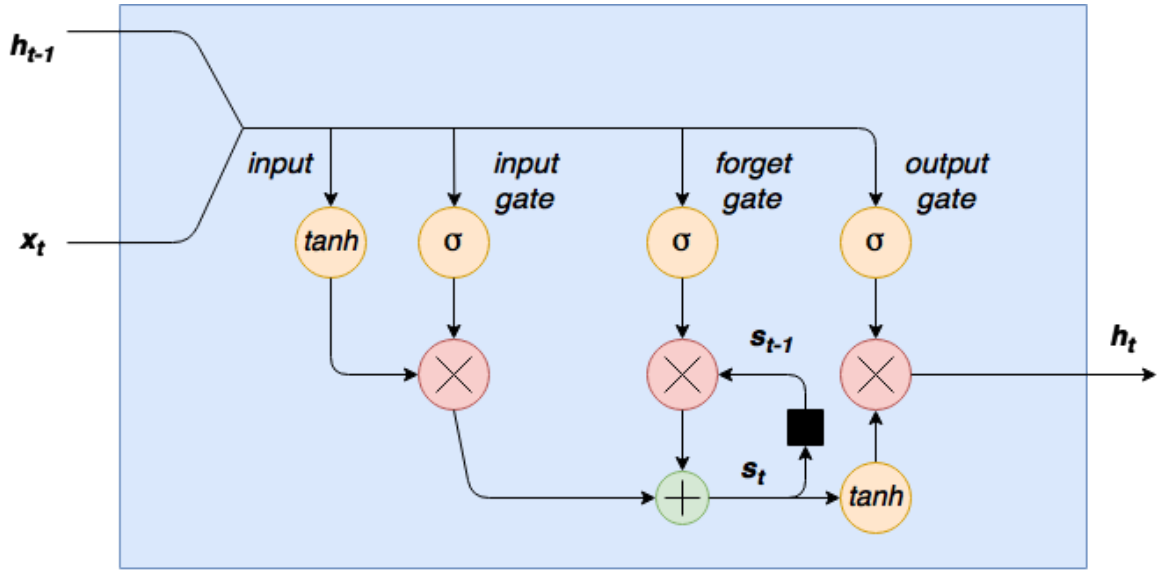


Fig 7 - Model of a common LSTM illustrating the three gates the network use to change the cell state (Thomas, 2017).

In addition to generating musical compositions, *Magenta* also wanted to be able to control the output of them, varying the randomness of the musical structure being generated. In order to achieve this they added a parameter they named "temperature" to the model, being represented by a value between zero and one, controlled by an external input such as a software fader. Having a high temperature, a number close to one would result in added randomness, while a number closer to zero would result in a more repetitive generated output. This was a sought after control since music naturally uses repetition. By varying the "temperature" control in small increments they they were able to achieve a more realistic generated output (Oore et al, 2017).

3.7 Generating Musical AIs

Magenta's Nsynth Is an audio synthesis method that uses NNs to synthesize audio by using a *WaveNet* based autoencoder. An autoencoder is a NN with the goal of reconstructing its input. This is important in applications where one want to generate a dataset that is similar but not the same as the one you give the neural network. Because of this an autoencoder is comprised of few layers and with less features compared to other more advanced neural network algorithms (Raval, 2018).

Autoencoders use a NN with unsupervised learning algorithms that applies backpropagation to the network. Backpropagation is simply put, a way of adjusting the neural network so that it produces a desirable output. By starting from the output of the neural networks you adjust how successful each preceding layer of the neural network is and adjust the weights of the network accordingly (Ng et al, n.d; Karpathy, 2016).
Magenta wanted *Nsynth* to be utilized by musicians as a creative tool to explore audio and bring excitement into the field of neural audio synthesis. To help achieve this *Magenta* created a dataset of musical notes consisting of approximately 300 000 different notes from 1000 different instruments. They created this dataset and made it public because there was a lack of comparable musical datasets of a size suitable for training current NNs (NSynth, 2017).

3.8 Formatting and Interpreting Data

In computer music, a widely used technique is to perform complex sound processing in the frequency domain is using a technique called Fast Fourier Transformation (FFT). FFT works by sampling a signal over time, dividing the signal into separate frequency components. As shown in figure 8 this enables one to break down a signal into several components letting one see the individual frequency amplitudes in an audio signal. One example of how to use this is to combine two signal sources with different frequencies into one signal. If one only look at the time domain of the signal one would see a single loudness curve. However if one apply a FFT on the signal one would be able to see these two frequency components as different events with different loudness levels (Smith, 1997).

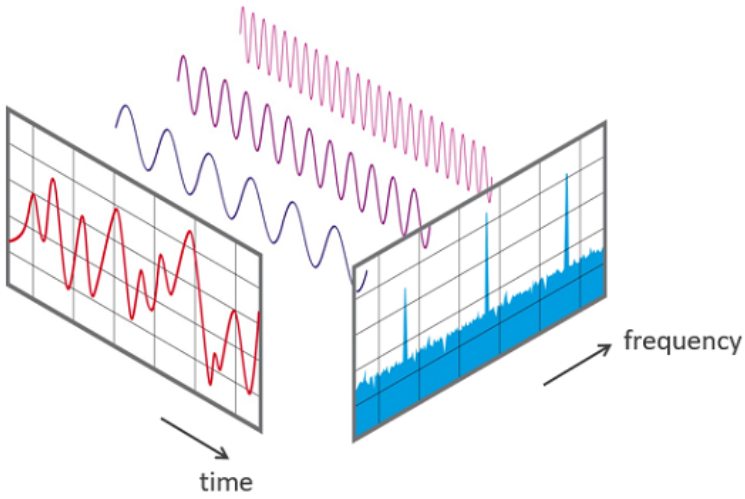


Fig 8 - A signal viewed in both the time and frequency domain (Phonical, 2017).

One tool that that utilize this technique is a phase vocoders. A phase vocoder allows you to analyse and resynthesize audio in the frequency domain by performing short overlapping fourier transformations to the audio allowing an artist to use techniques such as time stretching or control of the energy in a specific frequency spectrum. (Charles, 2008).

Max is a visual programming language meant to be a simplified language that would let artists use programming in multimedia projects (Cycling ’74, 2008). One task one can use *Max* for is shown in Jean-Francois Charles work on fft processing in *Max*.

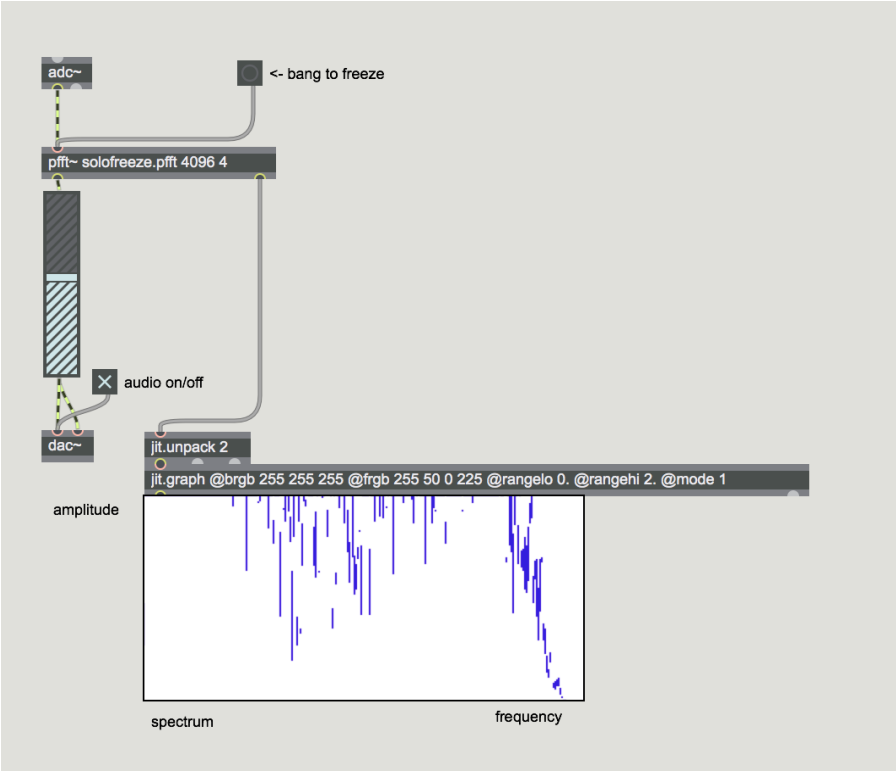


Fig 9 - *Solofreeze,* A Max patch allows users to hear a continuous tone from a sample of audio by applying FFT on the incoming audio (Charles, 2010).

In 2008, he created a Max patch that allows the user to stop a point in a stream of audio in real time, making the patch play a continuous tone based on the frequencies of the sound in that point in time. To achieve this Charles used FFT to analyse the frequency content of sounds coming in to the computer in real time. As seen in figure 9 he would be able to activate a button in the *Max* patch that would play back and resynthesize the last recorded frequency data as a continuous stream of audio. Charles also created a patch that would automatically display a sonogram from audio recorded into the program. As shown in figure 10, by using a combination of *Pfft~,* jit.matrix Charles was able to gather information of the frequency spectrum of the recorded sound, then taking the information and putting them in a matrix that would display the information as an image (Charles, 2008).

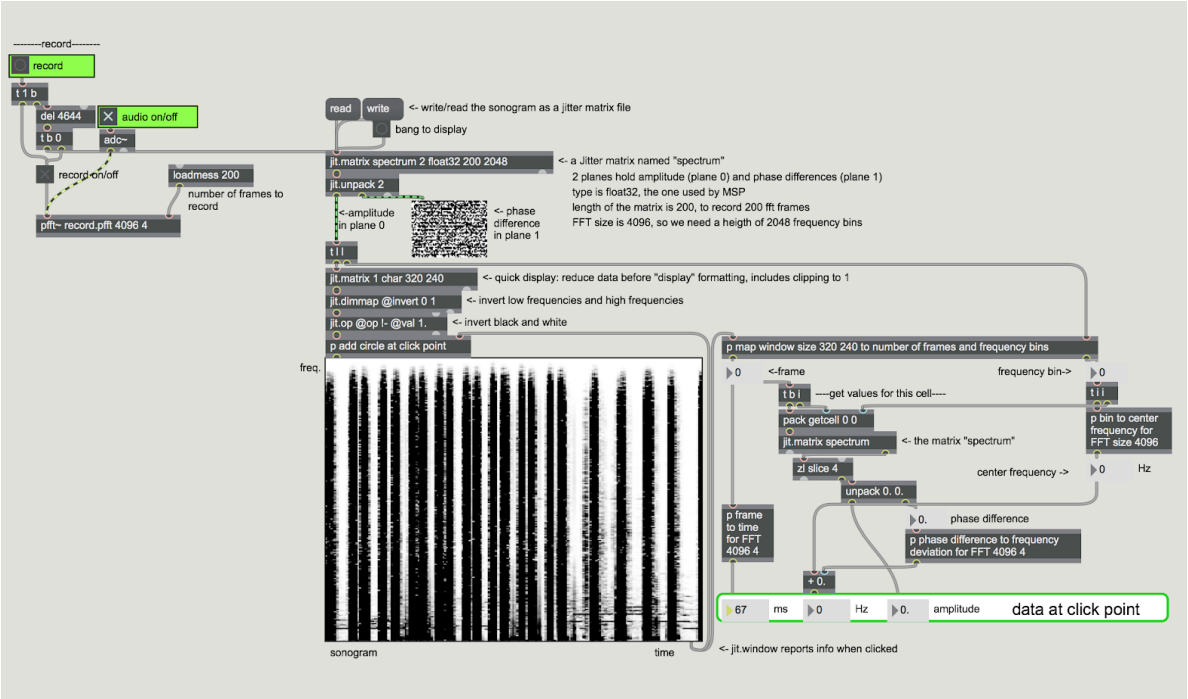


Fig 10 - *Explore the sonogram,* A *Max* patch allows users to see the frequency content of a sound (Charles, 2010).

3.9 Interactive Installations

Interactivity can be defined as the process of two people or things working together and influencing each other (Oxford University Press, 2018). Interactive experiences are qualitatively different from reactive or responsive works. Installations that use interactivity move control from both the audience and the installation and puts it in the hands of the conversation in between the two parts, neither having full control of the outcome (Bown, 2011). An example of an artist embracing this idea can be seen in *Eat* by Allan Kaprow. *Eat* was a performative installation that persuaded the audience to interact with the performance by eating food given by actors or placed in rooms. At the entrance to one room there were several apples strung from the roof on rough strings. The apples were placed just before the entrance to a room and if the audience wanted they could leave it be, eat one or just take a bite, leaving the apple with a bite mark (Kirby, 1965)

Interactive installations/systems can exhibit different degrees of interactivity, or have different amounts of interactive capacity. This is often linked to the complexity of the system. Simple systems that relates one input to one simple action can be seen as having low interactive capacity. One way to add complexity to such a system is with the use of autonomous behaviour. Making the system do a set of actions that the audience can answer to keeping the conversation between the two going. This would make both parts react to each other, having a transaction of information much like how a person uses an ATM (Bown et al, 2014).

One way to make use of autonomous systems is to add a performative nature to the system, not letting the autonomous actions be staged or predetermined but rather a product of the conversation with the audience. One example of this can be seen in the kinetic sculpture *The Senster* by Edward Ihnatowicz seen in figure 11.

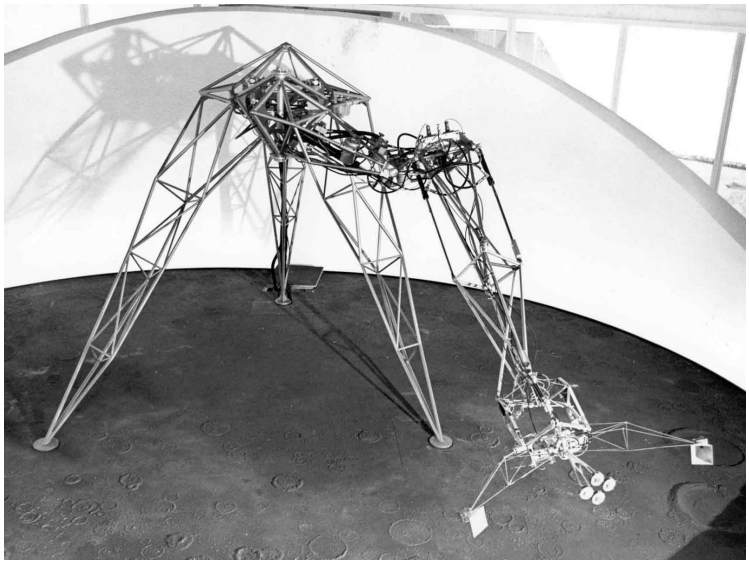


Fig 11 - *The Senster*, a kinetic sculpture created by Edward Ihnatowicz in 1970 (Ihnatowicz, n.d.).

The sculpture was made out of tubular steel with six hydraulic operated joints. The sculpture also used a Phillips P9201 computer, four microphones and two Doppler radar systems installed. This allowed the system to record an input from the audience that would control the hydraulic limbs. The system itself was reactive in the sense that it had several preprogrammed actions that would be triggered if the audience interacted with it. One of these was that whenever it detected a sound it would quickly turn its head toward the sound. Combining several such actions Ihnatowicz managed to create a system that would produce complex behavioral patterns. Because of the unpredictable behaviour of the audience and the acoustic characteristics of the hall it was displayed in, the sculpture seemed more intelligent than what Ihnatowicz thought. The audience would treat the sculpture as an animal that they could play with. After being given feedback on the installation from people who thought it was intelligent, Ihnatowicz was perplexed saying that the system had no intelligence, just pre programmed actions (Bown, 2011).

3.10 Evaluating A Performance

Interactive art allows artists to create systems that convey artistic expressions through the audiences interaction with it. Therefore in order to then evaluate Interactive art it is important to include audience research using both formal and informal approaches.
One way to gather informal data from a performance is through the use of qualitative data gathering. Unlike quantitative data which can be defined by numbers, qualitative data can not. Qualitative data cannot answer questions such as how many people got a specific grade in school. Qualitative data instead aims to, trough empirical data, understand relations in between non-quantifiable data. This would be useful in situations such as when the goal is to understand the social situation of certain communities as accurate as possible according to how the people in that community feel (Barbour, 2013).

Learning which methods are best suited to analyse and evaluate people's interactions with interactive artworks forms a great challenge. One way to achieve this would be to observe the audience over a period of time and immediately after the audience has interacted with the piece they would be engaged in recalling their interactions and describing how they interacted with the artwork. This was used in a study on the exhibition: *Light Logic* in Sheffield, UK. The exhibition consisted of drawings, paintings and interactive installations. After much thought the team behind the study gathered data on how to evaluate digital art. The team asked museums, galleries and artist communities in England, how they evaluated their digital art and how they used this evaluation to improve their works. The data they got from these sources included a range of methods for evaluating digital art. From Questionnaires to feedback gathered from Facebook.

Looking at each of the different evaluation methods, the research team choose to draw from different methods, creating a study practice that would be able to emphasise on the audiences inner thoughts on the interaction with the artwork while still allowing the audience to recall what they did at any point. This was thought to be one way to capture response on digital installations that would allow the artist to evaluate interactions with the art instead of evaluating just certain aspects of the exhibition which you would get by using techniques such as questionnaires.

The results of the study were that artist were left with a lot of both qualitative, quantitative and empirical data that they could reflect on. The study also concluded that the specific method used to evaluate the artipiece will work differently depending on the type of installation and for whom and what purpose the evaluation is done for. Where there is a big difference between methods that focus on institutional and policy driven data gathering versus individual artists and groups that put emphasis on collecting information on specific aspects of the artwork itself (Alarcon et al, 2014).

4. Method

The installation will consist of four computers each running individual NNs. Each of the computers will have one microphone, one speaker, two distance sensors and a computer monitor each. These computers will stand on top of pillars that would have all of the equipment built in to them. This would focus the attention of the audience towards the monitor screens.

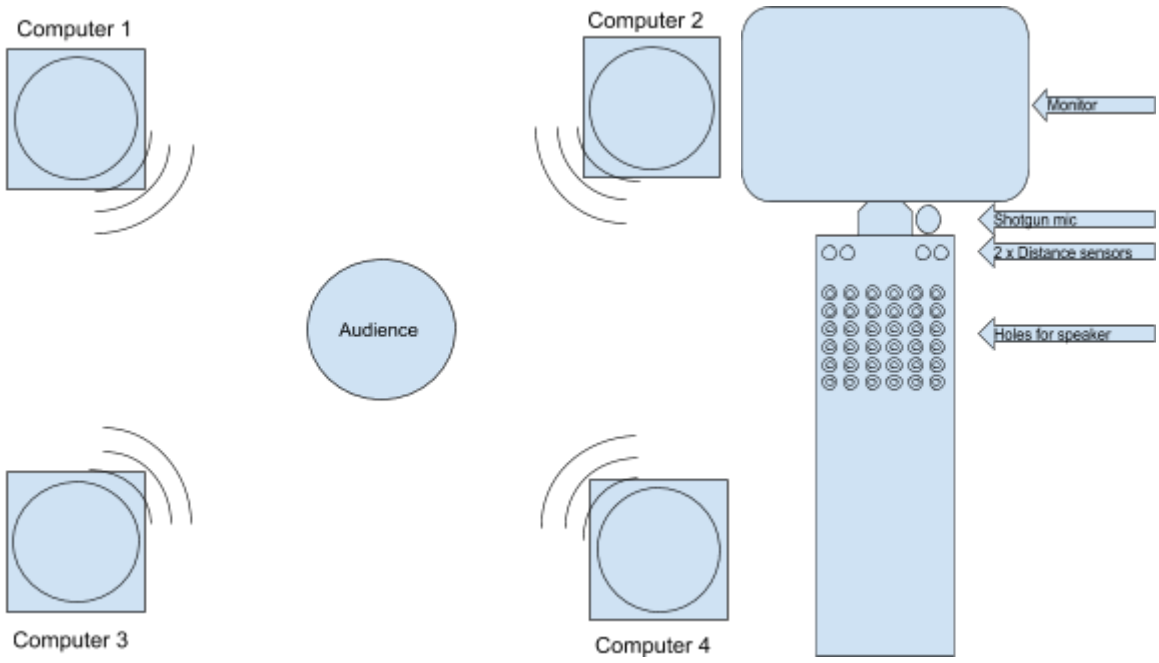


Fig 12 - Early draft of the installation setup, to the left four computers together in a room, to the right a model of how one of the computers would be like

After looking at NN architectures and how to utilize this technology in a musical performance, there was a number of issues that would need to be addressed in order to take the project from theory to practice. A major challenge was running a NN as efficiently as possible on available computer hardware. Running an RNN or CNN would be possible on datasets with few parameters, but since the goal was to use audio. In order to recreate audio at a suitable resolution another approach was deemed necessary. After looking at different types of NNs a GAN was deemed suitable as they excel at unsupervised training over time. HyperGAN is an Application Programming Interface (API) for Python that allows a GAN to run through a command line interface (255BITS, 2018).

HyperGAN works with images as its input so a way to convert the audio into image and back into audio again was needed. By using spectral processing it would be possible to process an audiofile as a 512 x 512 pixel image. By converting audio into images using FFT processing in *Max 7* would allow one to create images of a suitable resolution. This could be done with pfft~, storing frequency data from the audio input. The frequency data could then be stored in a matrix that would display its content as a 512x512 image, operating like a sonogram. Converting the Images generated by the neural network back into audio would be done by opening the image into a jit.matrix, harvesting data from each individual pixel of the image. This information can then be used to synthesize audio using the *oscbank~* object.

In order to add to the interactive capacity of the installation a set of sensors will be used to gather information from the audience. Using supersonic distance sensors combined with microphones, enough data would be gathered in order to track position, movement and loudness from the audience.The *Arduino Nano* microcontroller was used to interpret the distance information from the sensors, sending the results back to the computers running the NN programs using serial communication over USB.
To collect and process the data gathered from the audience trough these sensors, a program will be written using *Max 7*. This program will play the audio generated from the neural network back to the audience with the envelope and speed being controlled by interpreting the data from the sensors and microphone.

Looking at how one can evaluate an interactive artwork the method used with inb the evaluation of the exhibition *Light Logic* Will be used. Using the same technique for evaluation was deemed suitable for this project as well since both the *Light Logic* exhibition and this relies on interactive experiences. By setting up one camera inside of the room where the installation will be taking place the audiences interaction with the installation will be filmed. They will then be asked to come to a separate room where they will watch the just recorded film, narrating while seeing it, expressing the thoughts they had surrounding the experience. This will give a solid starting point in evaluating the installation. In addition the audience will be asked a small set of qualitative questions about the installation.

5. Results

5.1 Neural Network Prototyping

One of the main obstacles with using a NNs for this project, was the amount of computing power they required to run. The program would either use 24 hours of training to produce a sound that sounded close to the training material in terms of timbre and tonal quality, or spend 12 hours on a lower image resolution producing sound that was to far away from the training data. In order to overcome this, testing on different configurations for HyperGAN was done. HyperGAN allows one to change the type of GAN to use, the most common ones include: Wasserstein GAN, Improved GAN, DC GAN, LSGAN and BEGAN. The difference between them was how the NN algorithm runs, some with different models for how to calculate how close the GANs output was to the training data, while others used different approaches to how to normalise the data coming in to the network. Improved GAN was deemed most accurate in recreating a similar result to the training data over a short time period. The test done with these different types were over different amounts of time, some were trained for four hours, others 24, varying the timespan in order to see the effectiveness of the different types over different amounts of time. Below in figure 13 the training data and results from one of the tests with Improved GAN can be seen.

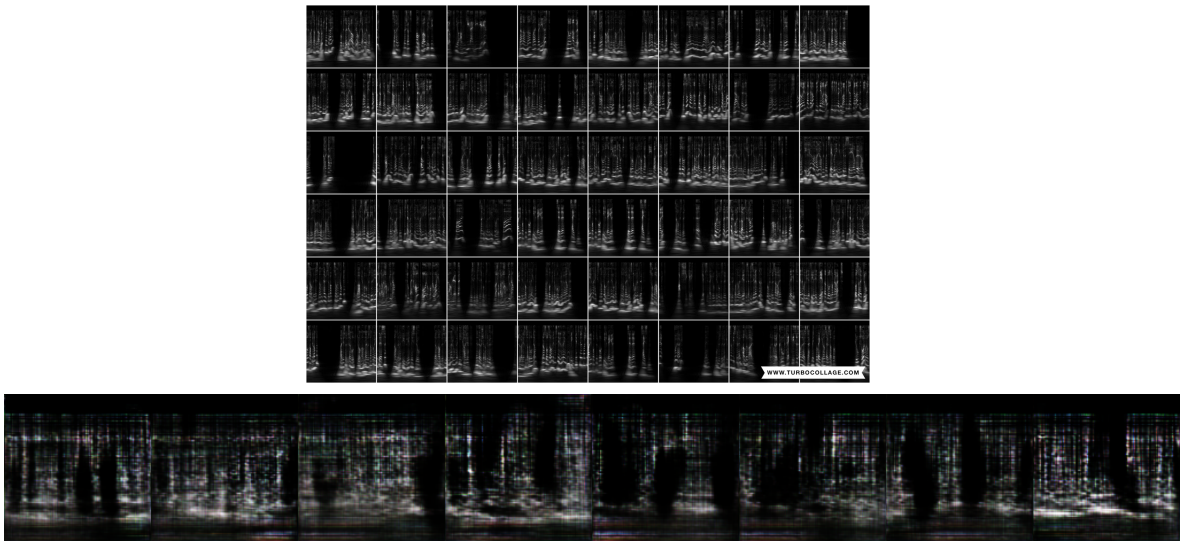


Fig 13 - On top, image set used to train NN. On the bottom, generated image after 6 hours of training.

Another solution that made HyperGAN more accurate over a shorter span of time was to run it at a lower pixel resolution. Running it at 128x128 made the output more accurate in addition to making HyperGAN run exponentially faster. This resulted in less training time required to get to the point where the NNs would produce sounds complex in tonality and timbre.

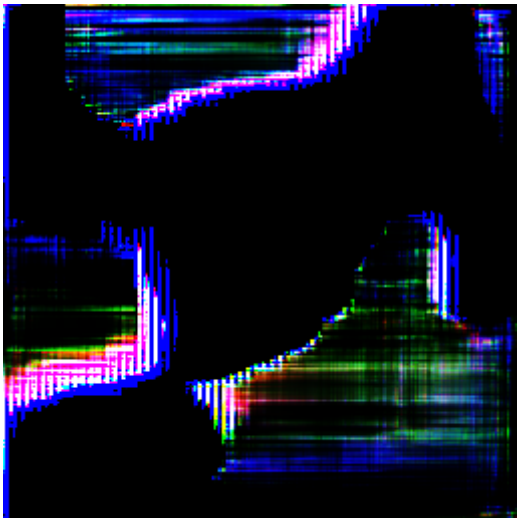


Fig 14 - Image generated from NN after the network had degraded.

Another obstacle faced was that HyperGAN would degrade over time, HyperGAN would generate images that got more and more accurate, but over time it would rapidly decline in its accuracy as seen in figure 14. To reach the point where it started to decline in accuracy would take longer if the training data given to it was larger, but working with training data of lower than 1000 images would make it start declining after approximately five hours. In order to solve this it was possible to restart HyperGAN by deleting its saves and restarting it. This was achieved by using terminal shells that would execute commands to delete only the save data. This would not delete the accumulated training data, and it would then be possible to run a program in *Max* that would record audio and save those audio files as training data. Then after a specified time the NN could restart itself, but with a larger database for training. The result of this was that the NN would degrade over time but when it had trained for to long it could restart itself, not letting itself degrade.

5.2 Hardware

One obstacle to the installation was the amount of hardware that was planned to be used in order to add to its interactive capacity. In order to make the installations interface organised and easier to interact with, custom made plinths were made to hold all of the sensors as well as speakers. Apple iMacs were also used as these could be placed on top of the plints, both acting as computers but also the main interface the installation had with the audience. This allowed the installation to focus the attention of the audience on to the iMac screens.

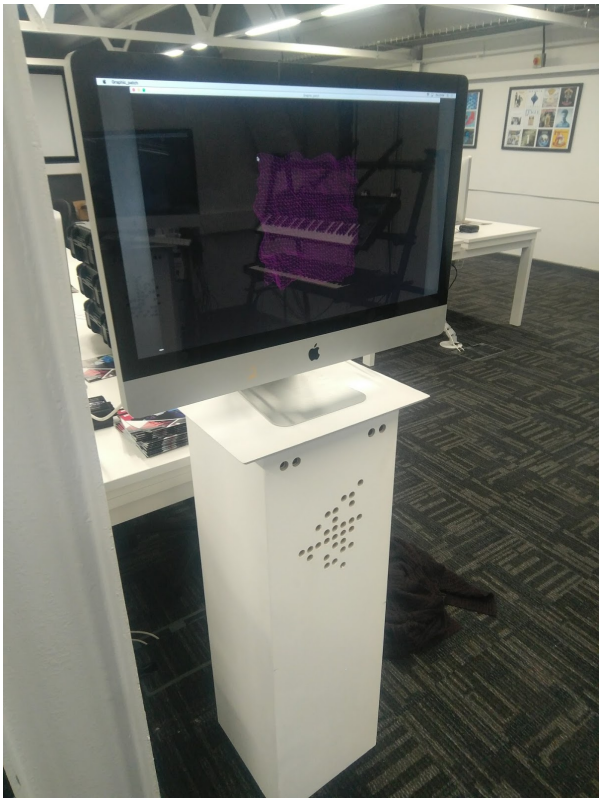


Fig 15 - iMac on top of a plinth. Inside there are two ultrasonic distance sensors and a speaker.

By using hardware sensors as a way for the audience to interact with the installation, the installation became more immersive and allowed the audience to express themselves in multiple ways. However a big obstacle was using ultrasonic distance sensors. The ultrasonic sensors used for the installation were tested thoroughly under a set condition, but once it came to test the sensors while fitted into the plinths they became inaccurate and would transmit error values. This is because these types of sensors use sound in order to measure distance, making them them slow and prone to bad readings. Because these sensors work by measuring an echo pulse, if an object is further away then even a short delay in the sound will affect the sensor readings. This type of sensor will also easily be disturbed by enclosed spaces causing echos and other sound sources that might interfere with the sensors **(**Morgan, 2014**).** To compensate for this the computer program receiving the information from the sensors would average the input over time as well as filter out certain ranges of numbers given by the sensors. This provided a more accurate reading but greatly affected the responsiveness of the sensors. The sensors would be able to detect people in front of it as well as approximate distance, but any rapid movements in front of the sensors would be averaged or filtered out.

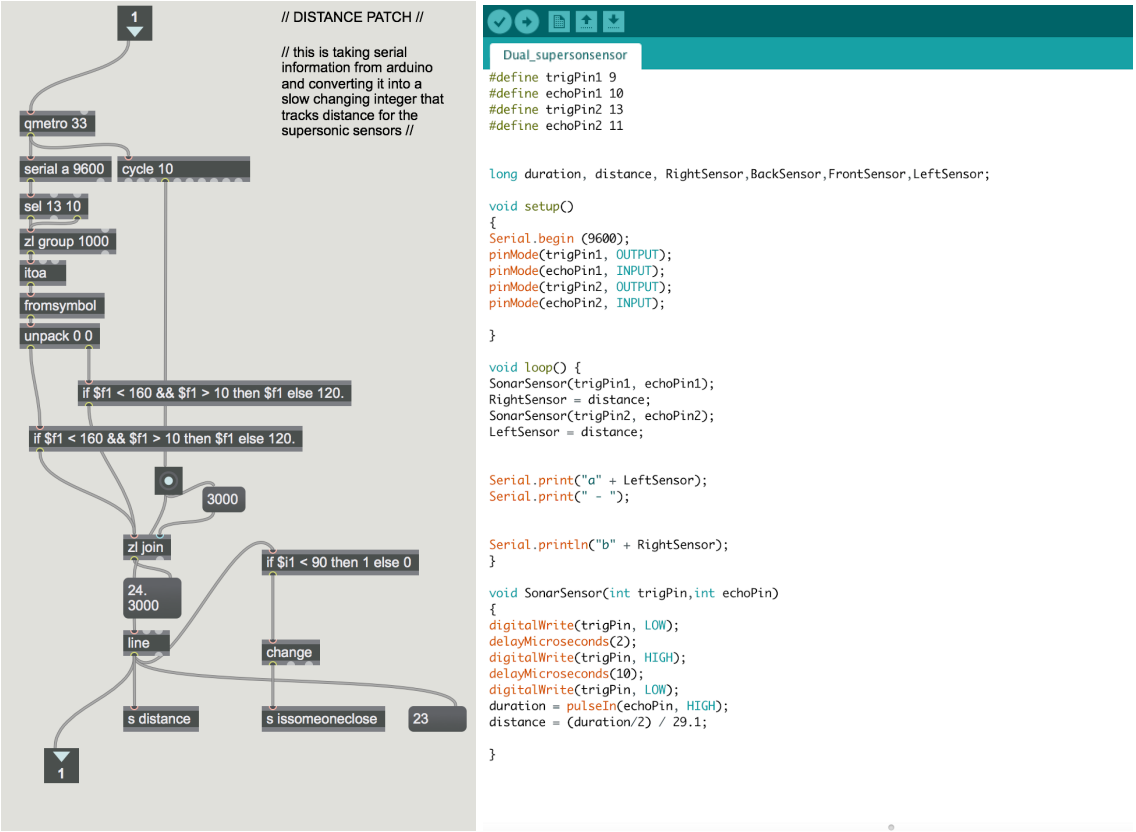


Fig 16 - To the left is the *Max* patch that smoothed and filtered the data coming from the ultrasonic distance sensors. To the right is the Arduino program that controlled the sensors.

5.3 Visual Aesthetic

One obstacle faced related to the interactive capacity of the installation was to provide a form of visual feedback to the audience. After getting feedback from peers on the installation a form of visual feedback from the installation was suggested. To create this Processing was used as it could create complex visual systems that wouldn't strain the computer hardware as much as similar visual programs created in *Max*. Looking at tutorials by *The Coding Train* (The Coding Train, 2016) a visual system that took audio amplitude as a modulator for a 3D triangle grid was used. This resulted in a program that would give visual feedback to the audience whenever the computers sent out audio. Since the main *Max* patch needed to be able to control the computer mouse on the screen, the window of the visual program in *Processing* was shrunk down and would take up only ½ of the screen. After getting feedback from peers this seemed ideal as the conversion of the audio files into images also added an aesthetic element of interest to the installation.

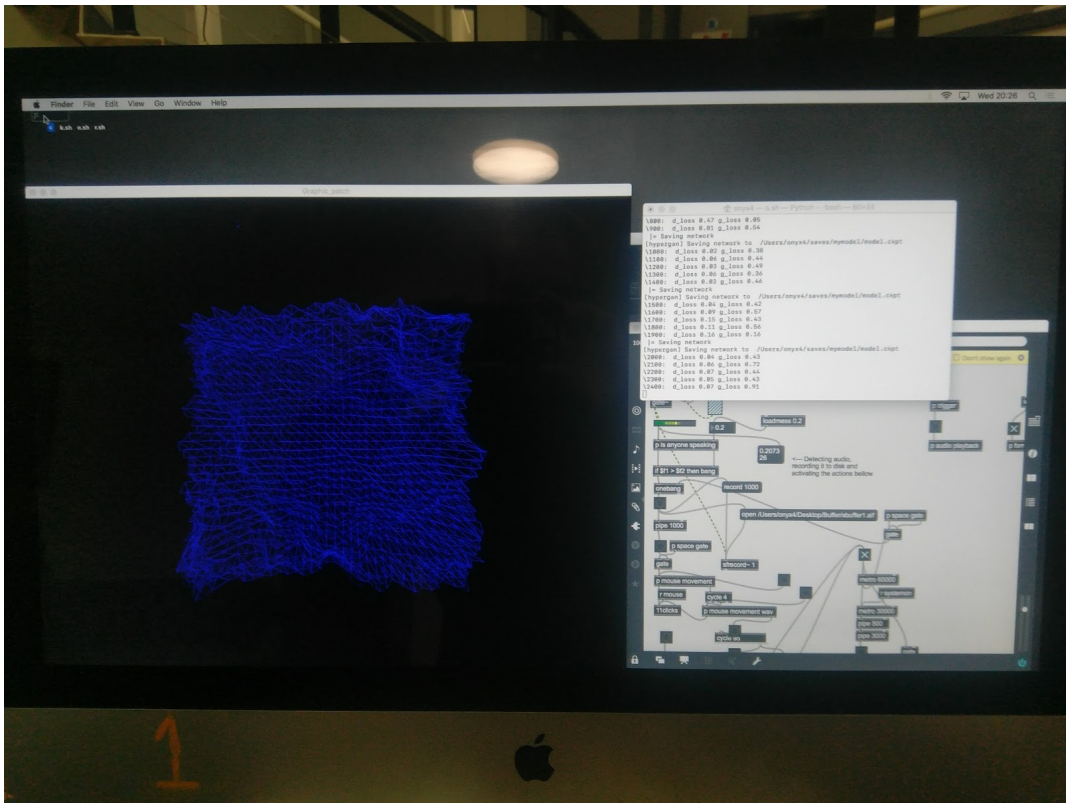


Fig 17 - Computer screen when the installation was running.

Another obstacle with the visual aspect of the installation was that the audience said that the different computers looked identical, not being able to differentiate them as individual elements of the installation. In order to achieve separation of the computers a single digit was written with a marker on each computer. This resulted in audience members giving identity to the different computers as the sounds they produced where divergent from each other.

5.4 Programming Interactivity

Using sensors to act as an interface between the computer and the audience would work by itself as a linear interaction point in between the audience and installation. However after testing the system, a stronger form of interactivity was deemed necessary. The use of several small autonomous behaviours was used in conjunction with randomised and convoluted reaction patterns from the system. This made the system more diverse and let the interaction be the attraction point of the installation rather than just using NNs for its novel idea.

Another obstacle was that the four separate computers acted individually, making the system as a whole behave like four individual systems instead of one system with four components in it. In order to connect each computers reaction to each other, the four computers were linked together over WIFI in order to share information from the sensors with each other as well as send a message telling the other computers if one of the computers played back audio. The computers also analysed the incoming audio using the *Max* external, Zsa descriptors (Malt et al, 2010). The analysed frequency information was also sent to every other computer. This allowed the four computers to trigger reactions from each other, making all four computers react as four individual pieces in a larger interactive system. This resulted in a system that would have a chance of triggering reactions from other computers if one computer was approached, making the individually NNs and computers act in unison as one. This also made the computers react more to each other, causing them to have a chance of "talking" with each other over longer periods of time such as 5-10 minutes.

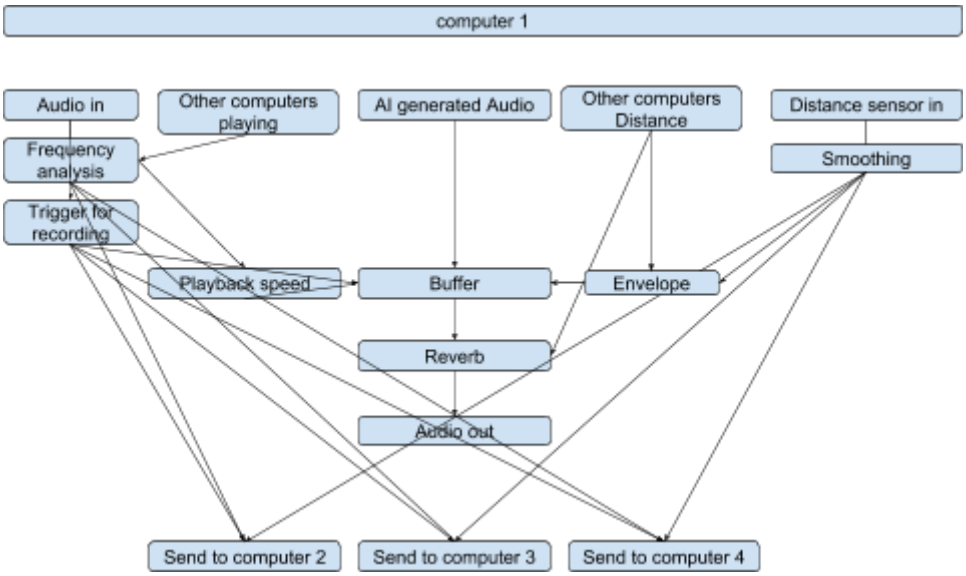


Fig 18 - Diagram of simplified signal flow in the *Max* patch.

Even though this made the computers have a connection between them, the system was still the same just duplicated by four. In order to differentiate the different computers probability function where programmed into each computer giving each computer have a higher probability of doing a certain action based on predetermined selections. This meant that computer two would have a higher chance of triggering sound if someone stood in front of computer four than if someone did the same to computer three. Probability functions like these where added to different aspects of the installation such as how often to play a sound, what pitch to play back the sound and what envelope it would play the sound in. The result of this was that the different computers would display more convoluted and differentiated behaviour. Some of these behaviours would be that the computers would if given the right input, interact with each other, having conversations in between them.

Making the program translate audio information into images for the NNs posed a challenge. Making use of FFT techniques to create a sonogram from the sound was possible, but using oscbank~ to convert the images back into audio resulted in unrecognisable audio not related to the original audio. In order to convert the audio and images the program *Photosounder* was used (Rouzic, 2008). This resulted in audio that would sound the same after being converted into images and back into audio again. After testing this with the NNs, a correlation between the training data and generated audio started to emerge as the NNs.

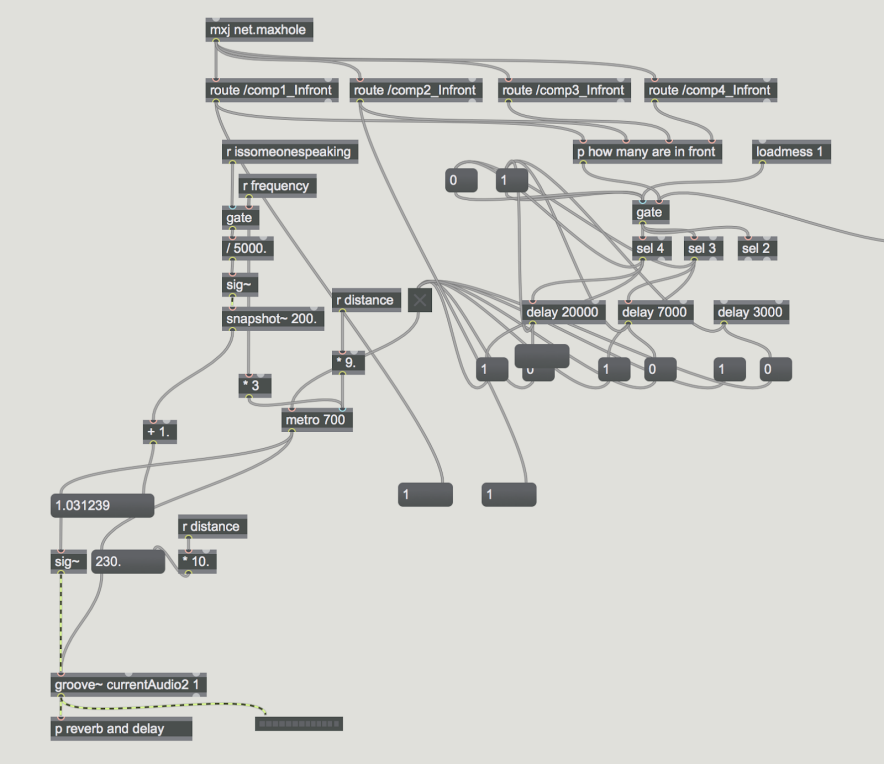


Fig 19 - Part of the max patch that would trigger playback of sound if there was a person in front of three or four computers.

However using *Photosounder* posed a new challenge as there was no simple way of automating the software. After testing out different methods including using *Script Editor* and automation using the *Terminal* in OSX, the best solution was to use the *11clicks* external object for *Max* (11olsen, 2014). *11clicks* allows one to control the computer mouse on the screen as well as using the mouse buttons. This made it possible to automate mouse movement in order to control the GUI of Photosounder.

5.5 Feedback

One of the biggest obstacles in this project was how to evaluate an interactive installation. Since Interactive installations are dependent on audience reactions thorough feedback was needed. The installation was evaluated by recording video footage of the audience interacting with it, then they were asked to comment on the interaction and were asked three questions.

Fig 20 - People interacting with the installation as part of evaluation.

- What did you think of this as an installation
- Did you think the installation was interactive
- Do you think this installation would be appropriate in a public space

In total eight participants participated in this evaluation. They went in to the room with the installations both one at a time and in groups.

- Person A and B went in together and were interviewed separately.
- C, D and E went in together and they were interviewed as a group.
- F, G and H went in by themselves and were interviewed by themselves.

This was done in order to get feedback on how the installation worked both in smaller and larger groups of people. This resulted in a variety of feedback on different aspects of the installation.

One of the obstacles with the evaluation was that the installation ended up developing its sound character over time. This caused the feedback to be focused on the installation in that point in time. To make the feedback cover the installation over a longer time perspective person A was asked to stay and evaluate the installation during the entire time the evaluation took. This was done in order to get a broader view of how the installation was perceived and resulted in feedback that were more varied and took the NNs development in time into account.

Person A expressed that the immediate reactions from the installation ended up going on for a longer period of time than expected, where person A would talk to the installation and the installation would react with a chain of reactions. This made the installation seem convoluted and made it hard to understand how to interact with it. Person A also gave the different computers personalities based on the sounds being played from them, where some of the computers reacted with sharp and loud sounds, others would play sounds that where quieter and had a continuous tone. Person A also expressed that the sounds being produced by the installation over a longer period of time would change drastically, adding to the interest of the piece.

Person C, D and E expressed that the installation lacked a direct reaction to their actions. They would try to make noise in front of the computers, but they didn't know if the computer was reacting to that specific noise or to the room itself. Person D also expressed that the moving mouse on the screen was confusing and directed attention away from the interactive experience. They all said that the main thing that linked the installations reaction to the audiences actions was the visual feedback from the screens. Person F expressed that The installation seemed interactive but that the interactivity was vague and person F wasn't sure if the computer was reacting to the person's actions or if it was just doing something automatic.

All of the people interviewed said that the installation was interesting and that they were curious about how the installation would react when they went into the room. However they also said that the installation was convoluted and would have gained from having a closer correlation between actions and reactions.

Looking at the feedback from the people interviewed there are some clear issues that would need to be resolved in order to make the installation seem more interactive. However there was also parts of the installation that piqued the interest of the audience. The feedback method used gave this project a lot of qualitative and empirical data that allowed the installation to improve itself and become more interactive.

6. Conclusion

In conclusion this project set out with the aim of exploring how AI can be used as a creative tool for musicians in interactive art. After discussing both contemporary and historical techniques for programming NNs as well as interactive installations this project resulted with an interactive multimedia installation.

Through exploring NN architectures and doing experimentation into how one can utilize this technology, this installation managed to recreate audio from preexisting audio recordings. The novelty of generating audio this way adds interest to the piece. However by preparing a set of training data beforehand and letting the audience see the NNs evolve instead of making them evolve by recording the audience, would have still added the same novelty to the piece. This would have allowed the project to focus on interactive aspects of the project, developing the interactive capacity of the project.

Because the installation was dependent on recording audio from the audience, the generated audio would be entirely dependent on the audience. This also meant that the space of the installation was crucial to the operation of it, requiring a quiet space. The final installation took place in a room where a lot of people would enter and leave, slamming doors behind them. This meant that most of the sounds recorded by the installation was slamming doors. This defeated the purpose of using the audience as source material for the audio generation.

One of the main objectives of the installation was creating a framework that focused on the relationship between the audience and computers. By also focusing on the relationship between the individual computers in the installation, the installation was able to both react to an audience but also react to itself making the installation react without minimal input from an audience. However this also made the interaction with the installation became vague, not letting the audience have as much to say in the interaction process.

7. Evaluation

7.1 Neural Network Systems

Researching common NN practice as well as historical ways of utilizing this technology gave an understanding on NNs that allowed this project to utilize NNs to recreate audio. Thorough testing involving different NN structures and settings of the NN was also conducted. This gave significant data and insight in how the NN networks behave over time and how they could be implemented into an interactive installation. Looking at these results allowed this project to run complicated NNs on consumer grade computers efficiently in order to recreate audio from a set of training data.

The final NN system managed to replicate sound but was reliant on a large database of training data. By doing further research and testing into how to one can use NNs to recreate audio more efficiently this project could have overcome this. This project was also not able to use audio as input data for the NNs. This would have made the project work without converting audio into images. Through further research into different NN architectures and how one can use different types of data with NNs this project could have gotten past that obstacle.

7.2 Interactivity

Researching interactivity and existing interactive systems created by contemporary artist gave insight in how one can implement interactivity in computer systems. This allowed the project to implement interactive programming into an installation, letting the NNs interact with an audience without direct human control of it. However feedback from peers and audience members gave insight into how the installations interactivity could have been improved. By doing further research into interactive systems and autonomous behaviour this project could have gotten insight into how to implement interactivity that to a higher degree reflected the actions of the audience.

This project researched evaluation methods for interactive installations and why it is important to get audience feedback. This allowed the project to implement evaluation methods that was able to gather a large quantities of data on how the interactive capacity of the installation could be improved and changed.

7.3 Experimentation

An important aspect of this project was the experimentation and testing done with NNs and the ultrasonic sensors. NN technology is a fast growing field of research and in order to uncover how the HyperGAN can replicate audio a lot of time went into testing out different configurations and different ways to convert audio into images. This allowed the project to discover how to most accurately recreate audio. The ultrasonic sensors used for this project required a comprehensive testing in order to optimise their functionality in different spatial locations. By looking at different blogs, forum posts and component datasheets the project gained knowledge in how to apply the sensors in different ways to maximise their efficiency. By combining this knowledge with comprehensive testing into how the sensors worked as part of the installation the project managed to utilize the sensors to track if people was standing in front of a computer or not. However The ultrasonic sensors used ended up producing a significant amount of error data and caused the installation to rely on the sound input more than the distance sensors. Further research into different types of sensors would have allowed the project to utilize a sensor that would accurately track the audiences actions and translate this into reactions from the system.

7.4 Learning Outcomes

By researching AI and NN architectures this project led to learning outcomes into how advanced computer systems can be used to replicate human behaviour and generate audio. The research into NNs also led to learning outcomes in how one can use this technology in different applications to make a computer system more intelligent and allow a system to be expressive without the intervention of a human. This project also led to learning outcomes in project management. As this project involved using a significant amount of hardware and software components a lot of effort went into researching how to utilize hardware and software best possible. This allowed the project to look for different solutions for problems such as converting audio into images.
The research into evaluation of interactive artworks gave insight into different approaches to evaluate artworks, and how most interactive artworks benefit from being evaluated on a project to project basis, as a common evaluation method can be hard to apply to interactive artworks. The project also looked into the importance of such evaluation and how, especially interactive artworks are dependent on audience feedback.

7.5 The Future of NNs in Interactive Art

This project explored one way to utilize AI in an interactive work of art. The approach of using NNs as an interactive can be further developed and implemented in multiple ways in order to add interest to the field of computer music.

One way to utilize AI in music would be as a new platform for audio synthesis. As mentioned, Nsynth is currently pionering NN audio synthesis, but as a concept it can be further expanded upon by creating synthesizers that use NNs to create complex audio waveforms that's impossible to recreate using existing synthesis techniques. By using NNs one can recreate instrument timbres based on the training data from several instruments, creating a new range of instruments combining timbral features several instruments into one.

Another way to utilize NNs in musical applications would be to create an interactive installation that use NNs to behave like the audience. This can be implemented as a humanoid robot that would use cameras to analyse human movement, recreating the movement of the audience.

References

11olsen (2014). *11clicks Max Object*. [Online]. 2014. Available from: http://www.11olsen.de/code/max-objects/11clicks. [Accessed: 10 May 2018].

255BITS (2018). *HyperGAN: A composable Generative Adversarial Network(GAN) with API and command line tool*. [Online]. 255BITS. Available from: https://github.com/255BITS/HyperGAN. [Accessed: 11 May 2018].

Alarcon Diaz, X., Akaroff, K., Candy, L., Edmonds, E., Farman, J. & Hobson, G. (2014). *Evaluation in Public Art: The Light Logic Exhibition. Interactive Experience in the Digital Age*.

Apple (2017). *iOS - Siri - Apple (UK)*. [Online]. 2017. Available from: https://www.apple.com/uk/ios/siri/. [Accessed: 11 May 2018].

Barbour, R.S. (2013). *Introducing Qualitative Research: A Student's Guide*. 2 edition. London: SAGE Publications Ltd.

Bown, O. (2011). Experiments in Modular Design for the Creative Composition of Live Algorithms. *Computer Music Journal*. 35. p.pp. 73–85.

Bown, O., Gemeinboeck, P. & Saunders, R. (2014). *Interactive experience in the digital age: evaluating new art practice*. Springer series on cultural computing. L. Candy & S. Ferguson (eds.). Cham: Springer.

Brandon Rohrer (2016). *How Convolutional Neural Networks work*. [Online Video]. 2016. Available from: https://www.youtube.com/watch?v=FmpDIaiMIeA. [Accessed: 11 May 2018].

Charles, J.-F. (2010). *Spectral Tutorials*. [Online]. 2010. Available from: https://cycling74.com/tools/charles-spectral-tutorials/. [Accessed: 9 April 2018].

Computerphile (2017). *Generative Adversarial Networks (GANs) - Computerphile*. [Online Video]. 2017. Available from: https://www.youtube.com/watch?v=Sw9r8CL98N0&t=539s. [Accessed: 22 January 2018].

Cycling'74 (2008). *MaxMSPHistory < FAQs < TWiki*. [Online]. 11 May 2008. History and background. Available from: https://web.archive.org/web/20080511162241/http://www.cycling74.com/twiki/bin/vi

ew/FAQs/MaxMSPHistory#Where_did_Max_MSP_come_from. [Accessed: 21 February 2018].

Fehrenbacher, K. (2015). *How Tesla is ushering in the age of the learning car*. [Online]. 2015. Fortune. Available from: http://fortune.com/2015/10/16/how-tesla-autopilot-learns/. [Accessed: 4 February 2018].

Fiebrink, R. (2017). *Wekinator | Software for real-time, interactive machine learning*. [Online]. Available from: http://www.wekinator.org/. [Accessed: 25 April 2018].

Gibbs, S. (2014). *Google buys UK artificial intelligence startup Deepmind for £400m | Technology | The Guardian*. [Online]. 2014. Available from: https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind. [Accessed: 5 February 2018].

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*. [Online]. Available from: http://arxiv.org/abs/1406.2661. [Accessed: 7 February 2018].

Ihnatowicz, E. (n.d.). *Cybernetic sculpture - The Senster*. [Online]. Available from: http://www.senster.com/ihnatowicz/senster/index.htm. [Accessed: 10 April 2018].

iZotope (2017). *Using the Track Assistant in Neutron | iZotope*. [Online]. 2017. Available from: https://www.izotope.com/en/community/blog/tips-tutorials/2016/10/using-the-track-assistant-in-neutron.html. [Accessed: 23 February 2018].

Jean-Francois Charles (2008). *Audio Freeze - Max/MSP/Jitter Spectral Sound Processing 2*. [Online Video]. 2008. Available from: https://www.youtube.com/watch?v=10-rtfQyMso. [Accessed: 21 February 2018].

Jones, J. (2017). *Recurrent neural networks deep dive*. [Online]. 17 August 2017. Available from: http://www.ibm.com/developerworks/library/cc-cognitive-recurrent-neural-networks/index.html. [Accessed: 9 April 2018].

Karpathy, A. (n.d.). *CS231n Convolutional Neural Networks for Visual Recognition*. [Online]. Available from: http://cs231n.github.io/convolutional-networks/. [Accessed: 9 April 2018].

Karpathy, A. (2015). *The Unreasonable Effectiveness of Recurrent Neural Networks*. [Online]. 2015. Available from: http://karpathy.github.io/2015/05/21/rnn-effectiveness/. [Accessed: 11 May 2018].

Karpathy, A. (2016). Yes you should understand backprop. *Andrej Karpathy*. [Online]. Available from: https://medium.com/@karpathy/yes-you-should-understand-backprop-e2f06eab496b. [Accessed: 15 February 2018].

Kirby, M. (1965). Allan Kaprow's 'Eat'. *The Tulane Drama Review*. 10 (2). p.pp. 44–49.

LANDR (n.d.). *LANDR: Instant Online Audio Mastering Software*. [Online]. LANDR. Available from: https://www.landr.com/en. [Accessed: 23 February 2018].

Luis Serrano (2016). *A friendly introduction to Deep Learning and Neural Networks*. [Online Video]. 2016. Available from: https://www.youtube.com/watch?v=BR9h47Jtqyw&t=583s+%5BAccessed+6+Jan.+2018%5D. [Accessed: 11 May 2018].

Maderer, J. (2017). *Robot Uses Deep Learning and Big Data to Write and Play Its Own Music*. [Online]. 2017. Available from: http://www.news.gatech.edu/2017/06/13/robot-uses-deep-learning-and-big-data-write-and-play-its-own-music. [Accessed: 11 May 2018].

Malt, M. & Jourdan, E. (2010). What is Zsa.Descriptors? *e--j dev*. [Online]. Available from: http://www.e--j.com/index.php/what-is-zsa-descriptors/. [Accessed: 11 May 2018].

Manzo, V.J. (2011). *Max/MSP/Jitter for Music: A Practical Guide to Developing Interactive Music Systems for Education and More*. 1 edition. New York: OUP USA.

Morgan, E.J. (2014). *HCSR04 Ultrasonic Sensor*. p.p. 6.

Murray-Browne, T. (2014). *The Cave Of Sounds: An Interactive Installation Exploring How We Create Music Together*. In: [Online]. 2014, NIME. Available from: http://www.nime.org/proceedings/2014/nime2014_288.pdf. [Accessed: 13 February 2018].

Ng, A., Ngiam, J., Yo Foo, C., Mai, Y., Suen, C., Coates, A., Maas, A., Hannum, A., Huval, B., Wang, T. & Tandon, S. (n.d.). *Unsupervised Feature Learning and Deep Learning Tutorial*. [Online]. Available from: http://ufldl.stanford.edu/tutorial/. [Accessed: 8 April 2018].

Nielsen, M.A. (2017). *Neural Networks and Deep Learning*. [Online]. Available from: http://neuralnetworksanddeeplearning.com. [Accessed: 11 May 2018].

NSynth (2017). *NSynth: Neural Audio Synthesis*. [Online]. 2017. Magenta. Available from: https://magenta.tensorflow.org/nsynth. [Accessed: 15 February 2018].

Olah, C. (2015). *Understanding LSTM Networks -- colah's blog*. [Online]. 2015. Available from: http://colah.github.io/posts/2015-08-Understanding-LSTMs/. [Accessed: 4 February 2018].

Oore, S. & Simon, I. (2017). *Performance RNN: Generating Music with Expressive Timing and Dynamics*. [Online]. 2017. Available from: https://magenta.tensorflow.org/performance-rnn. [Accessed: 4 February 2018].

Phonical (2017). *English: View of a signal in the time and frequency domain*. [Online]. Available from: https://commons.wikimedia.org/wiki/File:FFT-Time-Frequency-View.png. [Accessed: 10 April 2018].

Picton, P. (2000). *Neural Networks*. second. New York: Palgrave.

Raval, S. (2018). *Autoencoder Explained*. [Online Video]. 2018. Available from: https://www.youtube.com/watch?v=H1AllrJ-_30. [Accessed: 9 February 2018].

Rouzic, M. (2008). *Photosounder.com - Image-sound editor & synthesizer*. [Online]. 2008. Available from: http://photosounder.com/. [Accessed: 10 May 2018].

Sample, I. (2017). *Computer says no: why making AIs fair, accountable and transparent is crucial*. [Online]. 5 November 2017. the Guardian. Available from: http://www.theguardian.com/science/2017/nov/05/computer-says-no-why-making-ais-fair-accountable-and-transparent-is-crucial. [Accessed: 7 February 2018].

Shiffman, D. (2012). *The Nature of Code: Simulating Natural Systems with Processing*. 1 edition. s.l.: The Nature of Code.

Smith, S.W. (1997). *How the FFT works*. [Online]. 1997. The Scientist and Engineer's Guide to Digital Signal Processing. Available from: http://www.dspguide.com/ch12/2.htm. [Accessed: 4 February 2018].

The Coding Train (2016). *Coding Challenge #11: 3D Terrain Generation with Perlin Noise in Processing*. [Online Video]. 2016. Available from: https://www.youtube.com/watch?v=IKB1hWWedMk. [Accessed: 10 May 2018].

Thomas, A. (2017). *Recurrent neural networks and LSTM tutorial in Python and TensorFlow - Adventures in Machine Learning*. [Online]. 2017. Available from: http://adventuresinmachinelearning.com/recurrent-neural-networks-lstm-tutorial-tensorflow/. [Accessed: 10 April 2018].

Thorat, N. & Smilkov, D. (2017). Harness the Power of Machine Learning in Your Browser with Deeplearn.js. *Research Blog*. [Online]. Available from: https://research.googleblog.com/2017/08/harness-power-of-machine-learning-in.html. [Accessed: 13 February 2018].

燕家猫 (2016). *Generarive Adversarial Network*. [Online]. 2016. 知乎专栏. Available from: http://zhuanlan.zhihu.com/p/22549808. [Accessed: 10 May 2018].

Appendices

Additional Research

Max

Using Max one can create programs that can do anything within the constraints of the programming objects that exists within max. However Max as a programming language is specifically meant to be a programming language for musicians and media creators, with a simplified syntax. The Max syntax consists of square blocks that has the same a similar function to objects in other object-oriented programming languages. The blocks typically have input and output ports on them letting the user drag in between the in and outputs of objects to send data in between them. The blocks themself perform one or several computations depending on the object itself. One example of an object would be a ‘+’ object, which adds two numbers given to them in two inlets and outputs the result from its output (Manzo, 2011).

Processing

Processing is a object oriented programming language inspired by languages like BASIC and Logo with the intention of acting as a stepping stone into programming within the context of visual art. Processing where initially created with the purpose of acting like a software sketchbook and educational platform. Released in 2001, it has now grown into an open source alternative for visual artists to program professional graphical systems (processing.org, 2017). Processing is a derivative of Java, stripped of the parts of java that makes it hard to understand and use for beginners but retains some of the same functions and uses the same programming syntax. As a programming language processing is also optimised to let the user easily draw graphical content on the screen with functions and processes that are quickly formulated when writing code (Shiffman, 2015).

Interactive Experiences

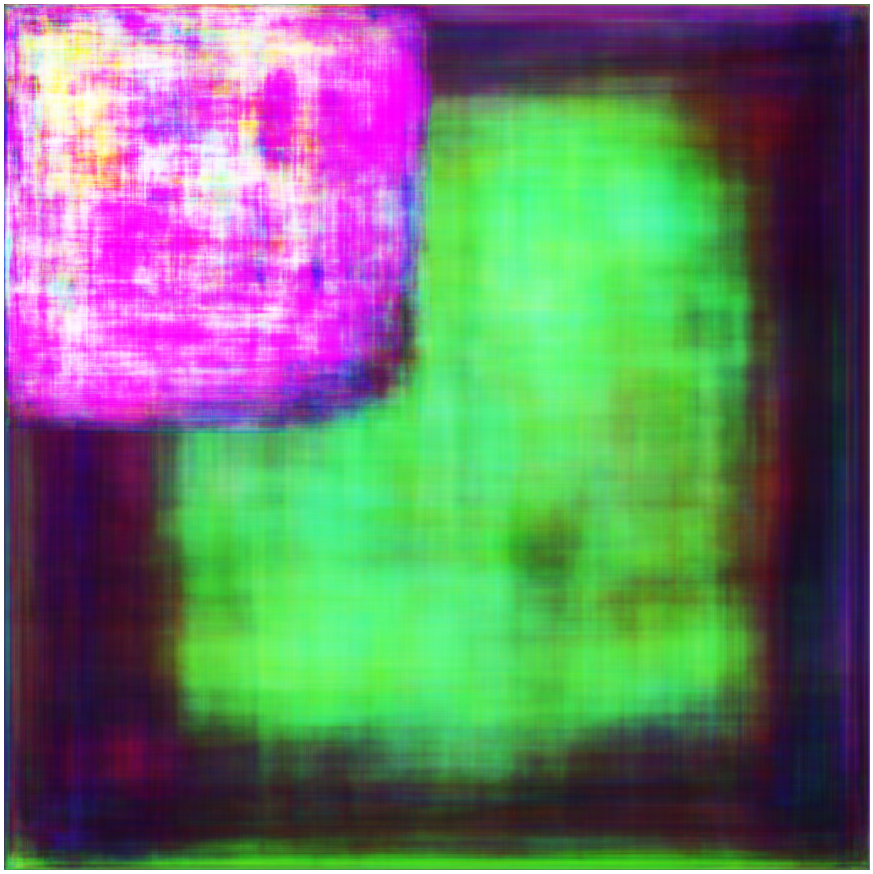
When classifying and comparing different interactive artworks you can look at how complex the system is and how that complexity relates to the artwork’s interactive capacity. Interactivity can be classified in different ways. One way to achieve a weak form of interactivity in a is by giving a system an input such as a button that the audience can press that will trigger an action or a sequence of actions. Using this form of interactivity combined

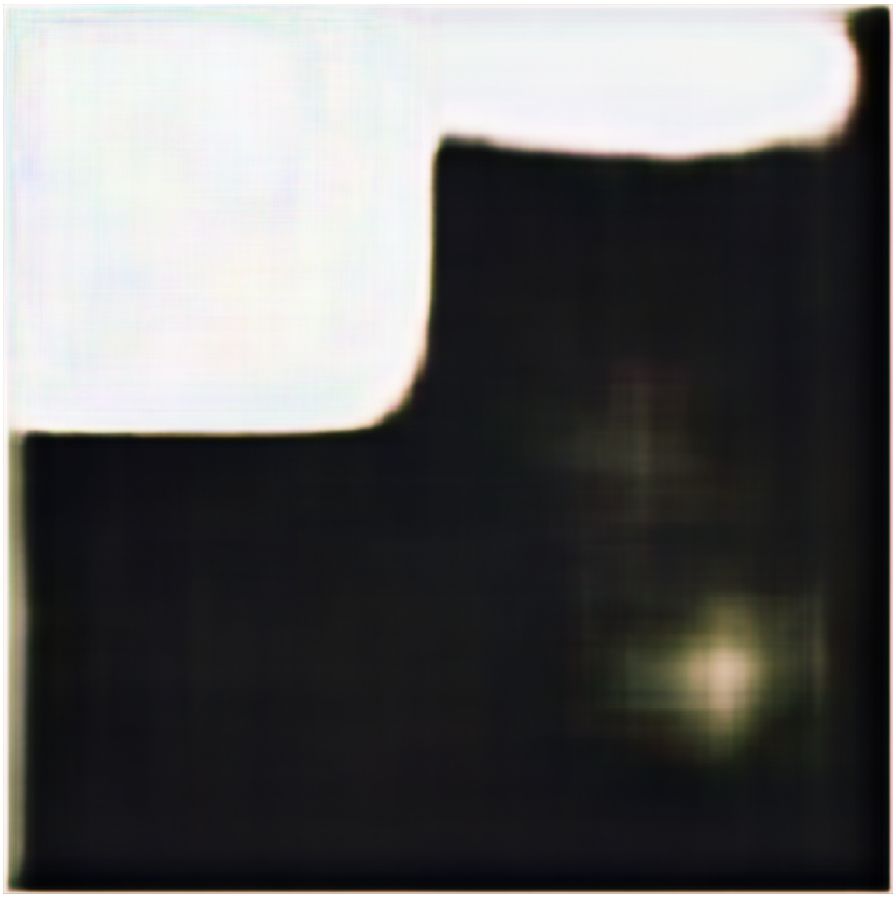
with autonomous actions can produce complex behaviours in a system. However if one relies on autonomous behaviour to invoke interactivity you strip away the system's ability to have a conversation with an audience and end up turning the installation from interactive to reactive (Bown, 2011).

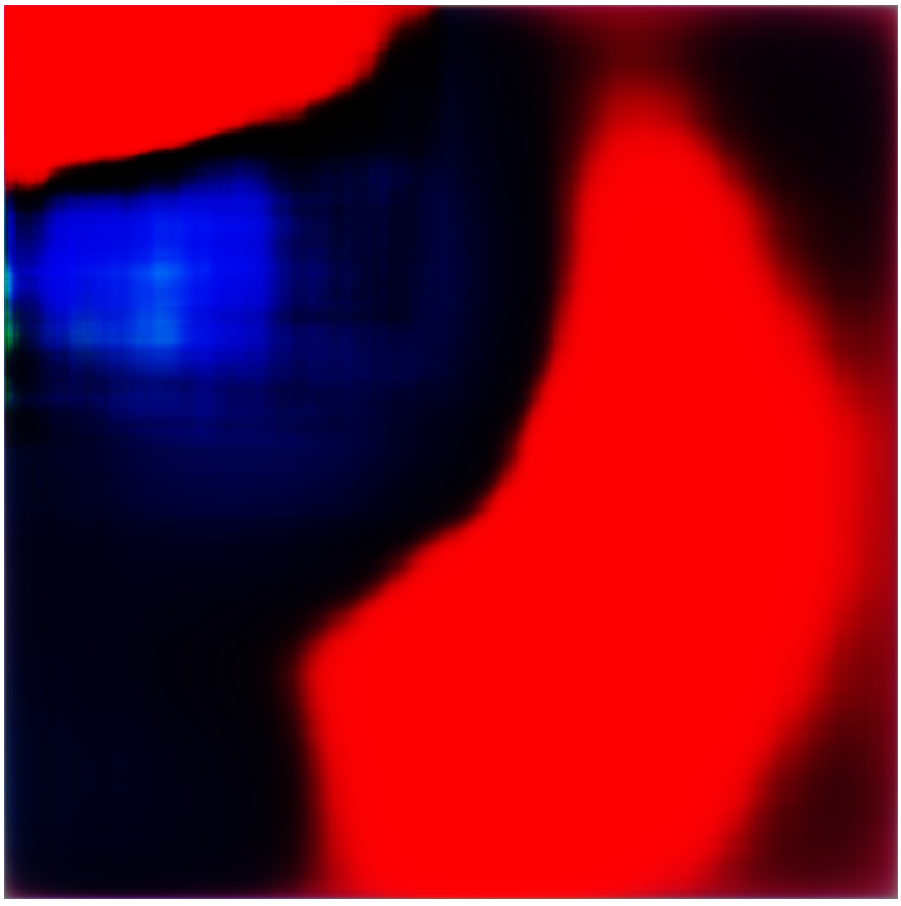
Cave Of Sounds is an interactive multimedia installation consisting of eight separate musical instruments. These instruments are placed around a circle forming a musical interface which the audience can interact with and play as a collective. The instruments were created by eight individual artists from the London based Music Hackerspace. Using a combination of hardware sensors, microcontrollers and software the installation encouraged the audience to interact with both the instruments they were playing but also all of the eight instruments in the installation (Murray-Browne, 2014).

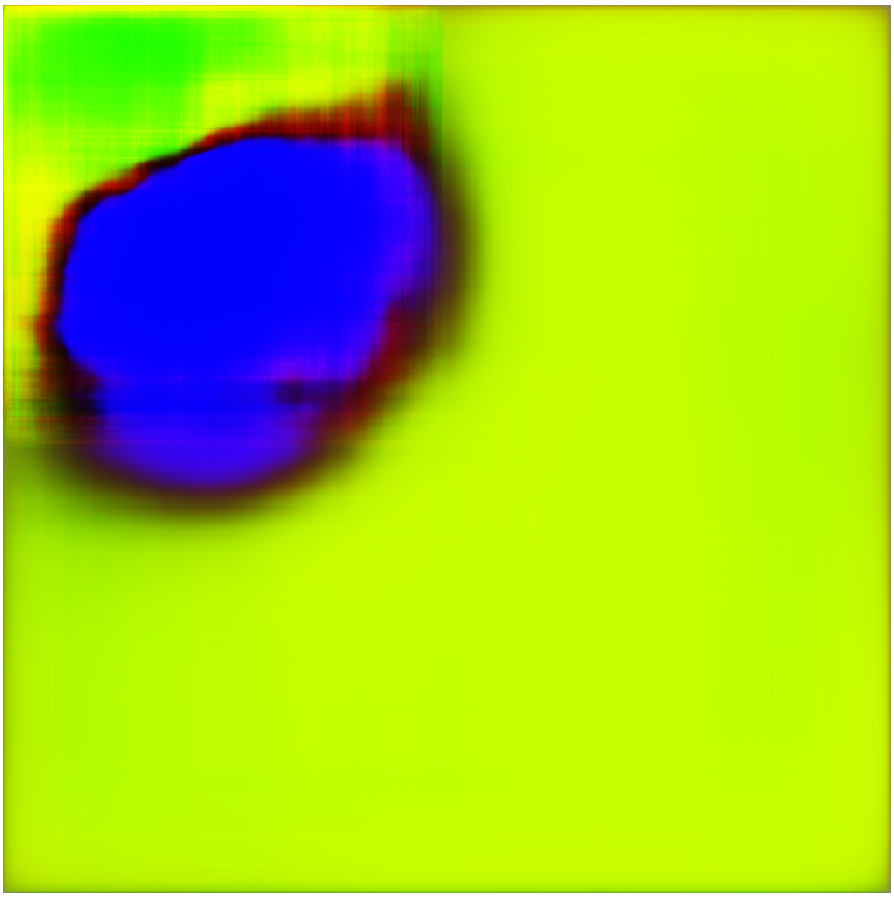
Neural Network Graphic Art

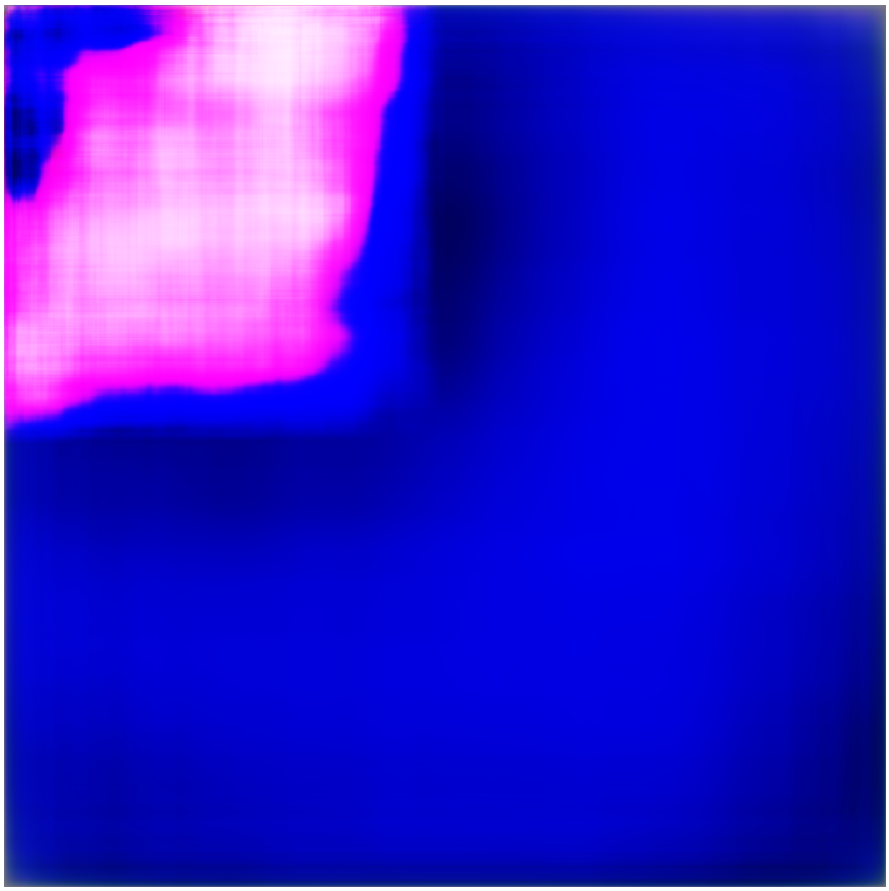
This project involved extensive testing into NNs and image generation trough HyperGAN. By adjusting the settings of hypergan and running it at a higher resolution the images generated were considered as artworks that will be used as visual art. Some of these will be included below:











Hardware Used

Shotgun mic, Røde NTG 2 x 4
http://www.rode.com/microphones/ntg-2

H6 portable recorder x 4
https://www.zoom-na.com/products/field-video-recording/field-recording/h6-handy-recorder

iMac x 4
https://www.apple.com/uk/imac/

Arduino Nano x 4 -
https://store.arduino.cc/usa/arduino-nano

Hypersonic distance sensor x 8
https://thepihut.com/products/ultrasonic-distance-sensor-hcsr04

Software Used

Max 7
https://cycling74.com/products/max-features

Zsa descriptors
http://www.e--j.com/index.php/what-is-zsa-descriptors/

Photosounder
http://photosounder.com/

HomeBrew
https://github.com/Homebrew/brew/blob/master/LICENSE.txt

HyperGAN
https://github.com/255BITS/HyperGAN/blob/master/LICENSE

Python 3.6.3
https://docs.python.org/3/license.html

Tensorflow
https://github.com/tensorflow/tensorflow/blob/master/tensorflow/core/lib/core/status.h

Pygame
https://www.pygame.org/docs/

NumPY
http://www.numpy.org/license.html#license

Hyperchamber
https://github.com/255BITS/hyperchamber/blob/master/LICENSE

Pillow
http://pillow.readthedocs.io/en/4.3.x/about.html#license

11clicks
http://www.11olsen.de/code/max-objects/11clicks

Building Plinth

For this project a plinth to hold all of the hardware for the installation was made.
This was made in the Staffordshire University wood workshop by me. In order to test the installation out as a prototype a single plinth was made by gluing together plates of MDF wood, then drilling the appropriate holes to let sound come out of the speakers as well as holes for the distance sensors. A hole for a shotgun microphone was also used but this ended up not working as the angle of the microphone made it record to low down, not recording the voice of people speaking to it. After the project was tested with a single machine four more plinths were made, taking into account changes needed in order to make it work with the hardware sensors.

Audio Synthesis

This project used NNs to synthesize audio that was processed in *Max,* then played through speakers as part of an installation. However throughout the testing and prototyping phase of this project HyperGAN would generate audio that didn't sound as intended. These sounds as well as the sounds generated through the installation will be uploaded to the portfolio website as an audio library

Consent Form

This project involved interviewing, filming and asking a group of volunteers questions about the project as part of the evaluation. The consent form given to the volunteers can be seen below.

Information Sheet

Project title: *Artificial Intelligence in Music*

My name is Jo Kallset and I am currently studying BSc (Hons) Creative Music Technology at Staffordshire University. As part of this award I am required to undertake a project.
My project is an investigation in using Artificial intelligence as a tool for creative music applications and more specifically how it can be used as a part of an audiovisual installation. I would like to ask for your permission to involve you in the project.

Information about the Project

This project uses Neural Networks to replicate sound and plays it back using information gathered from microphones and distance sensors.

The Testing Process

You will be asked to enter the room with the installation and then after coming out of the room you will transcribe your interactions with the installation while watching a video recording from when you interacted with the installation. You will then be asked qualitative questions about the installations interactive capacity.

Risks and Benefits of Taking Part

There are minimal risks, though if you suffer from epilepsy or have an aversion to sudden loud noises you will not be able to participate. By signing the consent form you are confirming that you are fit to participate.

The benefits are being an important part of a final year research project.

Participation and Confidentiality

Your participation in this study is completely voluntary and both) have the right to withdraw at any time. Please note that any data collected up to the point of withdrawal may be used within the study. Any information collected from this project will remain confidential and will be destroyed after the project has been assessed and the marks confirmed.

Further Questions and Contact Details

If you have any questions or would like further details regarding the project then please contact me:

Jo Kallset

jokallset@gmail.com

+47 40220967

If you have further questions or would prefer to contact a member of staff at the University then please contact my Project Supervisor:

Si Waite

S.J.Waite@staffs.ac.uk

+44 (0)1785 353717

Consent form

I have read the information sheet provided and understand all the information included in it.

I have had the opportunity to ask any questions about my participation in the project.

I confirm that I do not suffer from epilepsy nor have an aversion to sudden loud noises.

I confirm that video will be recorded during the duration of the test.

I confirm that audio will be recorded during the duration of the test.

My participation is voluntary and I understand that I can withdraw my data from the study at any time without giving any reason or having any rights affected.

I hereby give my consent to be a participant in this study.

Name (please print):

……

Date:

……

Signature:

……

Email Address (optional):

……

RESEARCH ETHICS

Proportionate Review Form

The Proportionate Review process may be used where the proposed research raises only minimal ethical risk. This research must: focus on minimally sensitive topics; entail minimal intrusion or disruption to others; and involve participants who would not be considered vulnerable in the context of the research.

PART A: TO BE COMPLETED BY RESEARCHER

Name of Researcher:	Jo Kallset

Student/Course Details (If Applicable)			
Student ID Number:			15021283
Name of Supervisor(s)/Module Tutor:			Simon Waite
PhD/MPhil project:	☐		
Taught Postgraduate Project/Assignment:	☐	Award Title:	Creative Music Technology
Undergraduate Project/Assignment:	X	Module Title:	Individual Music Technology project

Project Title:	Artificial Intelligence in music		
Project Outline:	The project will be carried out as an auditory installation using university property. The installation will be open to an audience and will include the audience interacting with the installation piece through microphones and distance sensors. The project will also be reviewed by select people and peers at Staffordshire university		
Give a brief description of participants and procedure (methods, tests etc.)	The project will be reviewed through group sessions where progress with the project will be presented throughout the year with multiple review sessions following the stages of the project.		
Expected Start Date:	30th August 2017	Expected End Date:	1st of May 2018

Relevant professional body ethical guidelines should be consulted when completing this form.

Please seek guidance from the Chair of your Faculty Research Ethics Committee if you are uncertain about any ethical issues arising from this application.

There is an obligation on the researcher and supervisor (where applicable) to bring to the attention of the Faculty Ethics Committee any issues with ethical implications not identified by this form.

Researcher Declaration

I consider that this project has no significant ethical implications requiring full ethical review by the Faculty Research Ethics Committee.	X

I confirm that:		
1.	The research will **NOT** involve members of vulnerable groups. Vulnerable groups include but are not limited to: children and young people (under 18 years of age), those with a learning disability or cognitive impairment, patients, people in custody, people engaged in illegal activities (e.g. drug taking), or individuals in a dependent or unequal relationship.	X
2.	The research will **NOT** involve sensitive topics. Sensitive topics include, but are not limited to: participants' sexual behaviour, their illegal or political behaviour, their experience of violence, their abuse or exploitation, their mental health, their gender or ethnic status. The research must not involve groups where permission of a gatekeeper is normally required for initial access to members, for example, ethnic or cultural groups, native peoples or indigenous communities.	X
3.	The research will **NOT** deliberately mislead participants in any way.	X
4.	The research will **NOT** involve access to records of personal or confidential information, including genetic or other biological information, concerning identifiable individuals.	X
5.	The research will **NOT** induce psychological stress, anxiety or humiliation, cause more than minimal pain, or involve intrusive interventions. This includes, but is not limited to: the administration of drugs or other substances, vigorous physical exercise, or techniques such as hypnotherapy which may cause participants to reveal information which could cause concern, in the course of their everyday life.	X
6.	The research **WILL** be conducted with participants' full and informed consent at the time the study is carried out: ● The main procedure will be explained to participants in advance, so that they are informed about what to expect. ● Participants will be told their involvement in the research is voluntary. ● Written consent will be obtained from participants. *(This is not required for self-completion questionnaires as submission of the completed questionnaire implies consent to participate)*. ● Participants will be informed about how they may withdraw from the research at any time and for any reason. ● For questionnaires and interviews: Participants will be given the option of omitting questions they do not want to answer. ● Participants will be told that their data will be treated with full confidentiality and that, if published, every effort will be made to ensure it will not be identifiable as theirs. ● Participants will be given the opportunity to be debriefed i.e. to find out more about the study and its results.	X N/A X

If you are unable to confirm any of the above statements, please complete a **Full Ethical Review**

Form. If the research will include participants that are **patients,** please complete the Independent Peer Review process.

Supporting Documentation

All key documents e.g. consent form, information sheet, questionnaire/interview schedule are appended to this application.	X

Signature of Researcher:	Jo Kallset	Date:	19th oct 2017

NB: If the research departs from the protocol which provides the basis for this proportionate review, then further review will be required and the applicant and supervisor(s) should consider whether or not the proportionate review remains appropriate. If it is no longer appropriate a full ethical review form **MUST** be submitted for consideration by the Faculty Research Ethics Committee.

Diagram and Risk Assessment for Installation

The installation that emerged from this project required a risk assessment and further planning to make sure equipment and staff/students were safe. This can be seen below.

Performance Location Document

Name: Jo Kallset
Course: Creative Music Technology
Supervisor: Si Waite

About

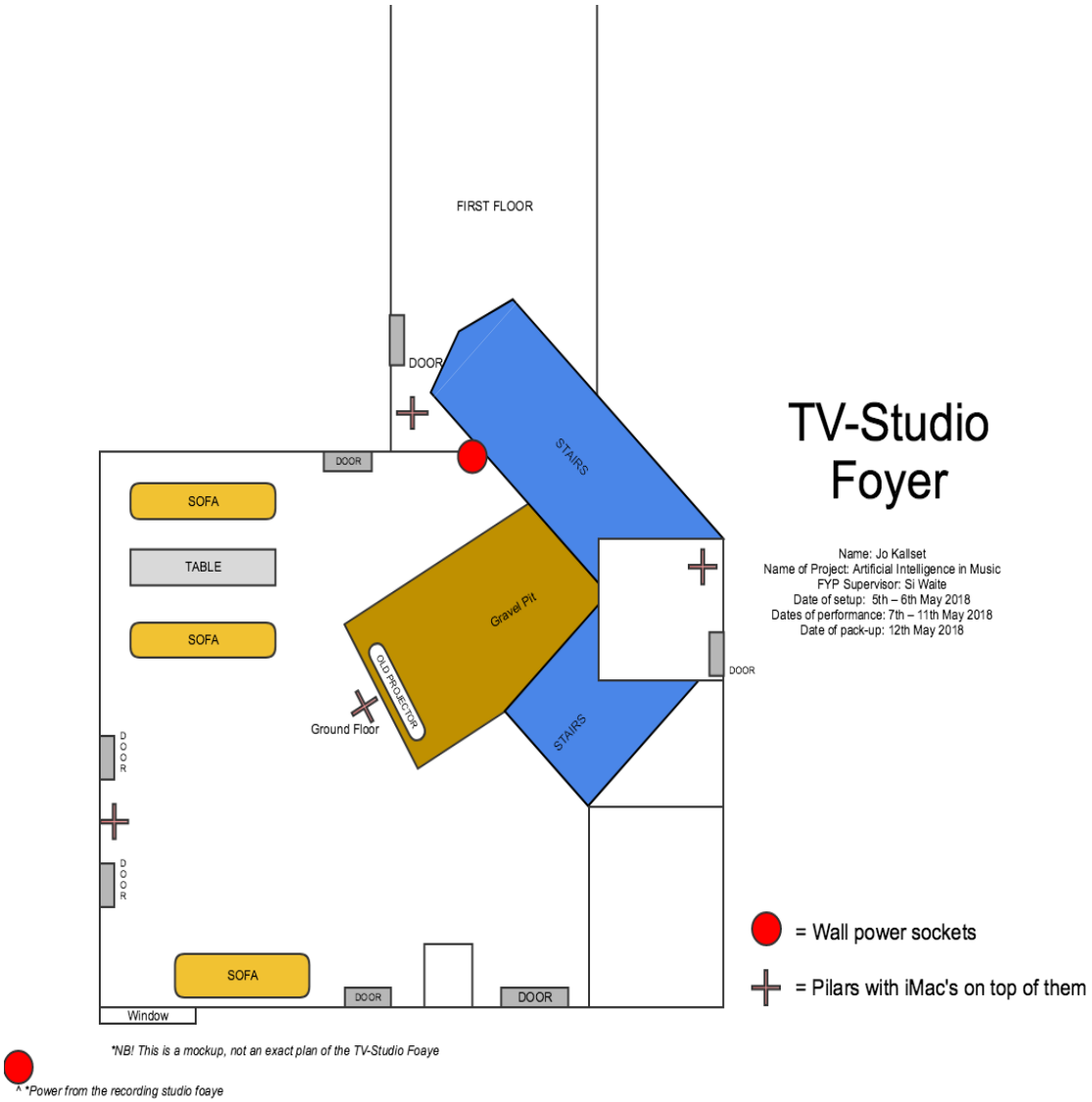
This document is a detailed plan on where and with what hardware that the installation is gonna be set up with. The installation is gonna take place during Noise Floor in the Cadman tv studio foyer.

Plan

The location of the installation will be in the foyer of the cadman studios tv studio the same week that Noise Floor is on (7th – 11th May 2018). The installation will be started up early in the morning each day and locked down at the end of the day, with the power cables going to the plinths moved inside of the building.
The hardware used in this installation will be the following but might be updated as the project is under development.

- Distance sensing module x 4
- iMac x 4
- Power extension cable x 5
- Audio interface x 4
- Xlr cable x 8
- Custom made Plinth enclosure x 4

The installation will be centered around 4 custom made wooden plinths that hold one speaker and 1 iMac each, because it is gonna be custom made a sketch will be included below to give a scale of reference.



*NB! This is a mockup, not an exact plan of the TV-Studio Foaye

^ *Power from the recording studio foaye

Risk of tripping, slipping and falling:

There will be a risk of the audience tripping on cables or other elements of the installation

To mitigate this each pilar will only have one cable going out of them each. This will be taped down or put inside of a cable bridge going to the closest power socket.

Equipment falling from plinths:

There will be a risk of people pushing the plinths used in the installation causing them to fall down, potentially breaking equipment and causing minor pain to audience if hit by parts of the installation.

To prevent this weights will be put inside of the plinths, making them bottom heavy.

People stealing the iMacs

There will be a risk of people coming in stealing the iMacs while no staff is in the room.

To prevent this the iMacs will be locked with a kensington lock each to the plinths.

Weather

Since there is a shortage of power sockets inside of the foyer a cable will be run from the foyer of the recording studios trough the windows this will be prone to weather damage.

To prevent ensure that safety requirements are met a power cable rated for outdoor use will be used for this and it will be secured properly to mitigate anyone from slipping or tripping in it.