

Kurs: **DG8100 Själständigt arbete 15 hp**

2022

Konstnärlig kandidatexamen i musik, elektroakustisk komposition 180 hp

Institutionen för komposition och musikteori

---

Handledare: Ludvig Elblaus

Niklas Dahlqvist

# **Ambient**

Dataset, snedträning och simulerade kollektiv

Skriftlig reflektion inom självständigt arbete

Till dokumentationen hör även följande inspelningar och partitur:

*N\_orgl, N\_midi, N\_gamba, N\_lånky, JB22, Ettusentvåhundraepoker*



# Innehållsförteckning

|   |    |
|---|----|
| <b>1. Introduktion</b>                                      | 2  |
| 1.1 Inledning   | 2  |
| 1.2 Metod och teoretisk förankring                          | 2  |
| 1.3 Syfte och frågeställningar                              | 3  |
| <b>2. Bakgrund</b>  | 5  |
| 2.1 Algoritmisk komposition                                 | 5  |
| 2.2 Agens   | 6  |
| <b>3. Tekniker</b>  | 8  |
| 3.1 Algoritmer  | 8  |
| 3.1.1 Neuralsyntes  | 8  |
| 3.1.2 Symbolisk generering                                  | 9  |
| 3.1.3 Neuralt generativa strukturer                         | 9  |
| 3.2 Dataset   | 9  |
| 3.3 Snedträning   | 10 |
| 3.3.1 <i>Svaltträning</i>                                   | 10 |
| 3.3.2 <i>Kontaminerat dataset</i>                           | 11 |
| 3.3.3 <i>Feedback-kopplat dataset</i>                       | 11 |
| 3.4 Ramverk   | 11 |
| <b>4. Verk</b>  | 13 |
| 4.1 <i>N_midi</i>   | 13 |
| 4.2 <i>N_orgl</i>   | 13 |
| 4.3 <i>JB22</i>   | 14 |
| 4.4 <i>Etttusentvåhundraepoker</i>                          | 15 |
| 4.5 <i>N_gamba</i>  | 15 |
| 4.6 <i>N_lånky</i>  | 16 |
| <b>5. Resultat och Diskussion</b>                           | 17 |
| 5.1 Framtida arbete   | 19 |
| 5.2 Sammanfattning  | 19 |
| <b>6. Källor</b>  | 21 |
| 6.1 Tryckta källor och litteratur                           | 21 |
| 6.2 Elektroniska källor                                     | 22 |
| 6.3 Dataset   | 22 |
| 6.4 Resurser  | 23 |
| <b>7. Bilagor</b>   | 24 |
| 7.1 Ljudinspelning av <i>Etttusentvåhundraepoker</i> (2021) | 24 |
| 7.2 Ljudinspelning av <i>JB22</i> (2022)                    | 24 |
| 7.3 Partitur till <i>N_gamba</i> (2022)                     | 24 |
| 7.4 Ljudinspelning av <i>N_lånky</i> (2022)                 | 24 |
| 7.5 Ljudinspelning av <i>N_orgl</i> (2022)                  | 24 |
| 7.6 Ljudinspelning av <i>N_midi</i> (2022)                  | 24 |

# 1. Introduktion

## 1.1 Inledning

Algoritmisk komposition är någonting som upptagit en stor del av min studietid, då jag uppskattar det ospecifika i denna kompositionsmetod. Istället för att exakt specificera hur musiken ska låta kan man dra upp regler om vad som får och inte får hända. Detta resulterar i att musiken är ständigt föränderlig och varje iteration av stycket blir annorlunda från den innan. Då jag upplevde att jag hade för mycket kontroll över mina algoritmiska system började jag intressera mig för maskininlärningstekniker. Detta grundade sig i att mina system kändes förutsägbara och jag ville experimentera med tekniker som kunde utvecklas bortom min kontroll. Någonting som tilltalade mig med tekniken var att i och med att du använder den så tvingas du också ge upp en del av äran som kompositör, då metoden intuitivt påminner om ett slags medkomponerande. Detta fick mig att vilja utforska hur långt det var möjligt att driva detta, och undersöka i vilken mån det är möjligt att suddas ut sig själv från sin egen praktik. I projektet bör tesen förstås som ett estetiskt förhållningssätt till materialet snarare än att faktiskt utplåna sig själv från sin egen konst, då jag inser det omöjliga i att helt och hållet göra detta.

## 1.2 Metod och teoretisk förankring

Enligt Howard Becker kan kulturell produktion förstås som en gemensam aktivitet där ett större antal människor är mer eller mindre delaktiga.<sup>1</sup> Dels i form av administrativ personal som arbetar runt omkring konstnären, men också i form av den sociala och historiska kontext som konstnären verkar i.<sup>2</sup> Becker menar att

Works of art, from this point of view, are not the products of individual makers, "artists" who possess a rare and special gift. They are rather, joint products of all the people who cooperate via an art world's characteristic conventions to bring works like that into existence.<sup>3</sup>

Min egen estetik kan därför ses som en produkt av det sammanhang jag verkar i, samt av en bredare historisk och social kontext och min musik står i direkt relation till denna kontext. Man kan förstå den som ett möjligt utfall från en kollektiv estetisk grund och man kan förstå andras musik i denna

---

<sup>1</sup> Cluley, Robert, Art words and art worlds: The methodological importance of language use in Howard S. Becker's sociology of art and cultural production. *Cultural sociology* 6.2 (2012), s. 203

<sup>2</sup> *ibid*, s. 202-203

<sup>3</sup> Becker, Howard, *Art Worlds*, University of California Press, Berkeley, (2008), s. 35

kontext på samma sätt.<sup>4</sup> Det neuralt genererade materialet fungerar på ett liknande sätt där den kollektiva grunden istället utgörs av ett dataset. Ett dataset kan inom maskininlärning bestå av olika saker men i mitt fall så består dom uteslutande av musik, antingen i form av MIDI-data eller av faktiska ljudfiler. Det kan också bestå av en eller flera kompositörers musik, men även om datasetet består av en enskilds persons verk vill jag mena att det fortfarande kan ses som ett uttryck för den personens sammanhang och influenser och därför i någon mening också blir kollektivt.<sup>5</sup> Jag anser därför att material genererat med neurala nätverk bör ses som en generalisering av en kollektiv erfarenhet, vilket väcker frågor om hur man kan förhålla sig till materialet i termer av upphovsperson.

En konstnärlig praktik kan alltså förstås utifrån kontexten personen verkar i. I det här projektet ämnar jag ta denna teoretiska ingång och använda det som metod för min praktik för att diskutera idén om det individuella skapandet kontra den kollektiva. Med hjälp av maskininlärningstekniker hoppas jag kunna skapa simulerade kontexter genom att använda mig av olika dataset och på detta sätt *industrialisera* denna modell. Min metod är alltså en slags översättning från en social teori till en estetisk praktik, och är tänkt att fungera som en slags spegling av hur konstnärlig produktion uppstår. Min förhoppning är att dataseten kommer fungera som ett slags simulerad social och historisk kontext där nätverkets output blir ett uttryck för datasetets och de individuella medlemmarnas sammanhang.

Neurala nätverks förmåga att generalisera det kollektiva gör det möjligt att i viss mån modellera det komplexa systemet av sammankopplade praktiker och influenser som utgör ett konstnärskap på ett sätt som gör det till ett unikt lämpligt verktyg för projektet.

### 1.3 Syfte och frågeställningar

Syftet med projektet är att undersöka hur jag genom att simulera kontexter kan förändra min agens som kompositör i förhållande till materialet, samt vilka konsekvenser detta får för min konstnärliga praktik. I projektet vill jag också undersöka relationen mellan musik genererad med neurala nätverk och upphovsskap, för att diskutera idén om det individuella skapandet.

Genom att använda olika maskininlärnings-tekniker hoppas jag kunna närma mig dessa frågor och förskjuta min agens i förhållande till mitt konstnärskap. Med detta förhållningssätt vill jag undersöka vad som händer när gränserna mellan material och upphovsperson suddas ut. Hur kan man förhålla sig till ett material skapat med denna metod? Jag ämnar också undersöka hur jag

---

<sup>4</sup> Cluley, s. 203

<sup>5</sup> *ibid*, s. 203

genom att manipulera dataseten på olika sätt kan förändra och träda in i olika upphovsroller. Eftersom att jag alltid kan ses som kompositör för de stycken som jag presenterar menas med termen upphovsroll det skiftande förhållande jag som kompositör har till materialet beroende på *hur* jag har förhållit mig till det. Min målsättning med projektet är också att skapa en diskussion om hur vi som konstnärer kan tillgodogöra oss maskininlärningstekniker och formulera ett språk som gör det möjligt att prata om maskininlärning ur ett estetiskt perspektiv. Detta då jag har upplevt svårigheter med applicera den existerande datavetenskapliga begreppsapparaten på min egen praktik.

Jag kommer att utgå från följande frågeställningar:

- Vilka upphovsroller kan jag skapa genom att förändra signal/brus förhållanden i träningsdatan till mina maskininlärnings-algoritmer?
- Hur kan jag påverka min agens som kompositör genom att simulera kontexter med hjälp av maskininlärning.

## 2. Bakgrund

Artificiell intelligens har en lång historia och de första experimenten med neurala nätverk utfördes under 1950-talet. De senaste åren har det dock skett något av en AI boom och det har gjorts framsteg i flera olika fält, inte minst i form av musikalisk generering. I detta kapitel kommer jag att ge en kort historisk översikt över fältet, samt diskutera frågor om eventuell agens och kreativitet i algoritmiska system.

### 2.1 Algoritmisk komposition

Neural komposition kan betraktas som en underkategori till algoritmisk komposition och det blir därför relevant att först introducera några algoritmiska kompositörer som relaterar till projektet.<sup>6</sup> Ett av de tidigaste exemplen är kompositörerna Lejaren Hiller och Leonard Isaacsons stycke *Illiac Suite* (1957) som genererades med hjälp av regelsystem och Markov-kedjor.<sup>7</sup> Tonsättaren Iannis Xenakis använde sig i hög grad av stokastiska algoritmer för att generera material som han sedan använde i sitt komponerande.<sup>8</sup> Matematikern och kompositören Harriet Padberg skrev vad som kan anses vara den första avhandlingen om algoritmisk komposition. Hennes program använde sig av text input för att styra olika delar av kompositionen.<sup>9</sup>

Jag är långtifrån den första att använda mig av neurala nätverk inom musik och kommer därför presentera ett urval kompositörer som fokuserar på AI och musik som på olika sätt relaterar till mitt eget projekt.

Kompositören Holly Herndon har arbetat med AI i många olika former men är kanske mest känd för sitt AI-barn *Spawn* som medkomponerade hennes skiva *Proto* från 2019.<sup>10</sup> Hon har även utvecklat en applikation tillsammans med *Never before heard sounds* som hon kallar för *Holly+*. Applikationen är i princip en röstsyntesalgoritm som är tränad på hennes egen röst, vilket gör det möjligt för vem som helst att använda sig av den.<sup>11</sup> Herndon beskriver sin inställning till maskininlärning på följande sätt:

---

<sup>6</sup> Då algoritmisk komposition har en lång historia med otaliga exempel finns det inte plats att göra en ingående studie av ämnet och jag hänvisar istället till Fernández, Jose D och Francisco Vico, AI methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research* 48 (2013) för den som är intresserad av ämnet.

<sup>7</sup> Roads, Curtis. *Composing electronic music: a new aesthetic*. Oxford University Press, USA, (2015), s. 342

<sup>8</sup> Fernández, Jose D och Francisco Vico, AI methods in algorithmic composition: A comprehensive survey. *Journal of Artificial Intelligence Research* 48 (2013), s. 517

<sup>9</sup> *ibid* s.518

<sup>10</sup> <https://www.artnews.com/art-in-america/interviews/holly-herndon-emily-mcdermott-spawn-ai-1202674301/>, hämtad 2022-02-20

<sup>11</sup> <https://holly.mirror.xyz/54ds2liOnvthjGFkokFCoaI4EabytH9xjAYy1irHy94>, hämtad 2022-02-20

The major difference is that we see Spawn as an ensemble member, rather than a composer. Even if she's improvising, as performers do, she's not writing the piece. I want to write the music!<sup>12</sup>

Det finns alltså både likheter och olikheter mellan hennes förhållningssätt och det som jag använder mig av i projektet. Herndon har ambitionen att själv vara den som skriver musiken, medan jag har ambitionen att förskjuta min relation till verken.

Dadabots är en amerikansk duo som använder sig i princip uteslutande av neurala nätverk för att på olika sätt generera musik. De har genererat ett flertal album med en modifierad version av *sampleRNN*<sup>13</sup> och har även ett antal livestreams igång där de live-genererar allt ifrån metal till fri-jazz.<sup>14</sup> Dadabots fokuserar huvudsakligen på att generera musik från diverse subkulturer och de har ofta ett reproducerande ideal i deras projekt, vilket skiljer sig från mitt projekt på det sättet att jag försöker presentera nätverkets estetik.<sup>15</sup>

Den engelska tonsättaren och ljudkonstnären Vicky Clarke undersöker frågan "*How can concrete materials and neural networks project future sonic realities?*" som en del av Unsupervised, en grupp bestående av kompositörer och forskare vid PRISM och NOVARS.<sup>16</sup> I projektet använder hon sig av dataset med konkreta ljud som hon själv spelat in i anslutning till Manchesters gamla industriområden. Clarke använder sig av en implementation av *SampleRNN* som är gjord vid PRISM.<sup>17</sup> Clarke använder samma algoritm som jag använder men hennes projekt skiljer sig från mitt då hon explicit jobbar med dataset bestående av material som hon har spelat in själv. Detta är en aspekt som inte finns med i mitt projekt då jag arbetar med dataset bestående av andra kompositörers stycken, och explicit arbetar med det kollektiva på ett sätt som saknas i hennes praktik.

## 2.2 Agens

Fernández och Vico menar att man kan dela upp algoritmisk komposition i två kategorier.

*Computer-aided algorithmic composition* (CAAC) syftar på algoritmisk komposition där små

---

<sup>12</sup> <https://www.theguardian.com/music/2019/may/02/holly-herndon-on-her-musical-baby-spawn-i-wanted-to-find-a-new-sound>, hämtad 2022-02-20

<sup>13</sup> [https://github.com/Cortexelus/dadabots\\_sampleRNN](https://github.com/Cortexelus/dadabots_sampleRNN)

<sup>14</sup> <https://dadabots.com/faq.php>, hämtad 2022-03-10

<sup>15</sup> Zukowski, Zack & C. J. Carr, *Generating black metal and math rock: Beyond bach, beethoven, and beatles*. (2018), s.1

<sup>16</sup> <https://auramachine.blog/2022/01/04/residency-framework-why-machine-learning-musique-concrete>, hämtad 2022-05-19

<sup>17</sup> <https://vickyclarke.org/about/>, hämtad 2022-03-10



grader av kompositionen automatiseras. Med algoritmisk komposition menas istället system där automatiserings graden är väldigt hög och kompositören inte gör mer än att sätta parametrarna för stycket.<sup>18</sup> Då jag anser att denna terminologi är något förvirrande kommer jag istället använda begreppen *Low level algorithmic automation* (LLAA) och *High level algorithmic automation* (HLAA) för att referera till detta. De olika verken i projektet hamnar på olika platser mellan dessa två begrepp vilket är något som jag kommer diskutera längre fram i uppsatsen.

Algoritmiska systems eventuella kreativitet är en omdiskuterad fråga, då det är svårt att bedöma kreativiteten hos en algoritm:

It may seem simple to measure computational creativity against human standards: we can simply ask people to listen to human and machine compositions, and declare an algorithmic composition system as creative if these people cannot tell apart its compositions from human ones. As Ariza (2009) noted, this kind of “musical Turing Test” has been performed by many different researchers trying to validate their systems, but they are valid if the algorithmic composition system just aspires to imitate, not to be truly creative and create a truly innovative work of art.<sup>19</sup>

I det här projektet är det dock inte nödvändigtvis relevant att utvärdera eventuell kreativitet i algoritmerna, utan snarare systemets potentiella *agens*. Agens kan definieras i vid bemärkelse som en entitets förmåga att agera och få någonting att hända.<sup>20</sup> Problemet med denna definition är att den går att applicera på en för stor mängd entiteter för att riktigt betyda någonting. En snävare definition av begreppet är en medveten entitets förmåga att avsiktligt agera, och därmed också ha ett ansvar för sina handlingar.<sup>21</sup> I den här texten använder jag *agens* som ett mellanting mellan dessa två formuleringar: En entitets förmåga att agera och fatta självständiga informerade beslut. Fördelen med denna definition är att ett informerat beslut inte nödvändigtvis implicerar någon form av mänskligt kognitivt resonemang. Nackdelen är att definitionen fortfarande är väldigt vid.<sup>22</sup>

Det är alltså inte relevant för projektet ifall systemet är intelligent i någon mänsklig bemärkelse, utan enbart om det är kapabelt att ta självständiga beslut som jag inte har kontroll över och som får direkta konsekvenser för musikens estetik. Ifall dessa beslut kan anses kreativa eller inte lämnar jag åt lyssnaren att bedöma.

---

<sup>18</sup> Fernández och Vico, s. 515-516

<sup>19</sup> *ibid*, s. 560

<sup>20</sup> Dahlstedt, Palle, ”Musicking with Algorithms: Thoughts on Artificial Intelligence, Creativity, and Agency”. *Handbook of Artificial Intelligence for Music*. Springer, Cham, (2021), s. 881

<sup>21</sup> *ibid* s. 881

<sup>22</sup> För en mer ingående diskussion om *agens* och AI se Palle Dahlstedts artikel *Musicking with Algorithms: Thoughts on Artificial Intelligence, Creativity, and Agency*

### 3. Tekniker

Det finns en stor mängd tekniker tillgängliga som använder sig av maskininlärning, alltifrån färdiga mjukvaror och program till experimentella *notebooks*.<sup>23</sup> Det finns också en stor mängd olika användningsområden för tekniken, men i mitt projekt har jag fokuserat på algoritmer som genererar musik i en eller annan mening, antingen i form av symbolisk generering såsom MIDI-data eller i vågforms domänen. I kapitlet kommer jag att gå igenom de algoritmer som jag har använt mig av i projektet. Jag kommer också diskutera dataset och *snedträning*, samt hur jag har förhållit mig till materialet.

#### 3.1 Algoritmer

##### 3.1.1 Neuralsyntes

Den teknik jag använt mig mest av i mitt projekt är neuralsyntes, och jag har framförallt arbetat med algoritmen *PRISM sampleRNN* som är utvecklad vid The Royal Northern College of Music i Manchester.<sup>24</sup> Algoritmen är en implementation av *SampleRNN* som presenterades 2017 vid ICLR.<sup>25</sup> *SampleRNN* är ett *recurrent neural network*, vilket innebär att det är ett nätverk som används för att förutsäga sekvenser. Genom att upprepade gånger be nätverket att förutsäga vad som kommer närmast blir det möjligt att sätta ihop längre sekvenser av nätverkets förutsägelser för att skapa musik med vilken längd som helst.<sup>26</sup> Algoritmen är ursprungligen framtagen för att användas inom talsyntes områden och genererar ljudfiler en sampling i taget, vilket innebär att det för en ljudfil med en samplingtakt på 44100 ska förutsäga lika många samplingsar per genererad sekund.<sup>27</sup> Träningsperioden för nätverket varierar beroende på dataset och kan pågå mellan några minuter till oändligt länge. Det resulterar i att det är en långsam process att arbeta med algoritmen men det innebär också att potentialen i varje genererad fil undersöks noggrannare då processen för att generera den är så besvärlig.

---

<sup>23</sup> <https://colab.research.google.com/>

<sup>24</sup> <https://github.com/rncm-prism/prism-samplernn>

<sup>25</sup> Mehri, Soroush, et al. *SampleRNN: An unconditional end-to-end neural audio generation model*. (2016)

<sup>26</sup> Zukowski, Zack & C. J. Carr. *Generating albums with sampleRNN to imitate metal, rock, and punk bands*. (2018). s. 1

<sup>27</sup> *ibid*, s.1-2

### 3.1.2 Symbolisk generering

Jag har även arbetat med Magentas algoritm *Polyphony RNN* som tränar på och genererar MIDI-data.<sup>28</sup> Algoritmen är uppbyggd kring ett LSTM-nätverk och är kapabel att generera polyfonisk musik. Tränings- och genereringsprocessen är betydligt snabbare jämfört med *PRISM SampleRNN* vilket uppmuntrar till ett mer flexibelt och experimentellt arbetssätt.

### 3.1.3 Neurtalt generativa strukturer

Det finns en begränsning med tidigare nämnda algoritmer, nämligen att dem inte arbetar i realtid. Som ett komplement till dessa har jag därför även arbetat med Nick Collins SuperCollider klass *NeuralNet*.<sup>29</sup> Klassen är ett *feedforward neural network*, vilket innebär att informationen alltid rör sig i en riktning. Det medför att nätverket är dåligt på att förutsäga sekvenser men kan däremot användas till exempelvis satslogik.

## 3.2 Dataset

De dataset som jag har jobbat med består framförallt av samtida elektroakustisk musik, och har ofta en direkt koppling till utbildningen i elektroakustisk komposition vid Kungliga Musikhögskolan. Detta är ett medvetet val som jag har gjort utifrån framförallt två skäl.

Väldigt mycket av den musik som finns tillgänglig som genereras med hjälp av diverse maskininlärnings-tekniker tenderar att vara tränad på klassisk musik av framförallt dem stora mästarna såsom Bach, Mozart, Beethoven.<sup>30</sup> Att reproducera klassisk musik blir ett sätt att legitimera en ny teknologi, och det går att dra paralleller till tidigare liknande exempel på detta såsom Wendy Carlos skiva *Switched on Bach* där hon tolkar Bach på en Moog synthesizer.<sup>31</sup> Budskapet blir att om ett neuralt nätverk klarar av att reproducera Bach så klarar den av att reproducera vad som helst. Detta är någonting som jag anser är problematiskt då det antyder att den klassiska musiken är den måttstock som man kan mäta all annan musik utifrån. Dessutom tycker jag också att det är fruktansvärt oinspirerande att använda ny teknik för att reproducera gammal musik. Jag har därför valt att använda mig av samtida musik istället, med undantaget för Giovanni Pierluigi da Palestrina, och tänker mig att det kan fungera som en slags motpol till detta tänk. Det ska dock sägas att en annan möjlig anledning till att klassisk musik förekommer så mycket inom

---

<sup>28</sup> [https://github.com/magenta/magenta/tree/main/magenta/models/polyphony\\_rnn](https://github.com/magenta/magenta/tree/main/magenta/models/polyphony_rnn)

<sup>29</sup> <https://composerprogrammer.com/teaching/supercollider/advanced/index.html>, hämtad 2022-05-18

<sup>30</sup> Exempel på detta är: *Deepbach*, *Coconet*, *EMI*, *Beethovens 10:e symfoni*, *Googles Bach AI Doodle* etc.

<sup>31</sup> <https://www.wendycarlos.com/+sob.html>, hämtad 2022-05-18

maskininlärning är att det är förhållandevis lätt att få tag i stora dataset som inte är upphovsrättsskyddade.

Jag anser att kompositörer i en viss generation har ett ansvar att lyfta varandra i den mån som är möjligt och kanonisera sin generations musik. Genom att använda personer från min egen generation i dataseten kommer jag bidra till detta, även om jag förstår att mitt bidrag i det stora hela är väldigt litet.

### 3.3 Snedträning

En av de stora utmaningarna med att jobba med maskininlärningstekniker är att få algoritmen att producera material som inte är en direkt reproducering av datasetet. I forskningssammanhang är detta ofta någonting eftersträvansvärt, och fungerar som ett sätt att bedöma kvaliteten på algoritmen.<sup>32</sup> I en konstnärlig kontext däremot är det inte uppenbart att reproducering är det mest önskvärda. För att tackla denna problematik har jag utvecklat ett antal *snedtränings*-tekniker för att manipulera dataseten och därmed i förlängningen resultatet. Genom att manipulera dataseten på olika sätt uppstår också olika upphovsroller i relation till mig och algoritmen. I följande kapitel kommer jag att förklara hur de olika teknikerna fungerar samt vad det får för estetiska konsekvenser för det klingande resultatet.

#### 3.3.1 Svälträning

*Svälträning* är den mest grundläggande metoden för snedträning och innebär att träningsprocessen pausas för tidigt. Detta innebär att nätverket inte lyckas justera sina vikter ordentligt och resulterar i att diverse artefakter bevaras i den genererade filen. Genom att experimentera med att pausa träningen vid olika epoker går det att justera förhållandet mellan brus och signal i genereringen och därmed även med avvikelse från datasetet. En epok kan inom maskininlärning förklaras som ett varv igenom hela datasetet.<sup>33</sup>

Inom datavetenskapen refereras detta till som *underfitting* och är ett exempel på problematiken hos den existerande begreppsapparaten, då begreppet för med sig värderingar om vad som är en lyckad modell.

---

<sup>32</sup> se diskussion om dataset, s.9

<sup>33</sup> <https://deeptai.org/machine-learning-glossary-and-terms/epoch>, hämtad 2022-05-19

### 3.3.2 *Kontaminerat dataset*

En annan metod som har visat sig vara givande är att kontaminera datasetet med någonting som inte hör hemma där. I praktiken handlar det om att introducera en viss procent avvikande material i datasetet. Beroende på hur avvikande materialet är samt hur stor procent av datasetet som består av detta kan man manipulera det klingande resultatet. Om kontamineringen har stora likheter med det övriga materialet så kan det i bästa fall uppstå en slags synergi där den generade filen har drag av datasetets båda delar. Om kontamineringen däremot är väldigt avvikande resulterar det antingen i kraftiga brusartefakter, eller att kontamineringen våldsamt slår sig igenom och tvingar på sig delar av sin karaktär på ljudfilen. Genom att hitta platser mellan dessa ytterligheter går det att producera en mängd olika variationer på samma dataset med vitt skild karaktäristik.

### 3.3.3 *Feedback-kopplat dataset*

Slutligen har jag laborerat med att feedback-koppla datasetet. Detta innebär att det ursprungliga materialet i datasetet byts ut löpande mot genererade filer från nätverket, vilket resulterar i ett ständigt föränderligt dataset. Detta är någonting som jag bara har hunnit experimentera med flyktigt vilket innebär att jag inte riktigt har kunnat utvärdera teknikens potential. De inledande experiment som jag har gjort har dock genererat lovande resultat.

## 3.4 Ramverk

Med hjälp utav snedträningen tillåts maskinen utveckla en egen estetik. Henrik Frisk diskuterar i texten *Aesthetics, Interaction and Machine Improvisation* hur man kan förhålla sig till en musikalisk artificiell intelligens som har utvecklats bortom skaparens estetik:

One may imagine that in the near future there will be tools, like KA, but more advanced, that exhibits some notion of what we can call musical intelligence. Are we in that situation prepared to allow the machine to develop its own aesthetics, as we would with a human co-player?<sup>34</sup>

Även om algoritmerna jag har arbetat med inte uppvisar musikalisk intelligens i den bemärkelse som Frisk talar om, anser jag att det ändå är relevant att förstå algoritmens output som ett uttryck för en estetik. I projektet har jag strävat efter att respektera denna estetik i så stor mån som möjligt.

Då jag i det här projektet har strävat efter att experimentera med upphovsrollen har jag känt ett behov av att formulera ett antal regler för hur jag ska förhålla mig till det genererade material som

---

<sup>34</sup> Frisk, Henrik. *Aesthetics, Interaction and Machine Improvisation. Organised Sound 25.1* (2020), s.38

ligger till grund för kompositionerna. Som en konsekvens av syftet med projektet har jag behövt begränsa mitt eget inflytande på materialet, och försökt återge det såsom det är. Ramverket som jag har satt upp kan ses som en uppställning av min egen agens i förhållande till algoritmernas.

1. Jag har tillåtit mig att ändra tonhöjden och ibland även uppspelningshastigheten av de genererade ljudfilerna, men jag har dock begränsat mig till oktav och kvint transponeringar.
2. Jag har i vissa fall tillåtit mig att trimma bort klick och andra artefakter i de filer när jag har ansett att det har varit motiverat. Jag har dock strävat att i största möjliga mån återge filerna såsom de är och ofta låtit artefakter och klick vara kvar i filerna då jag dels tycker att det tillför någonting estetiskt och dels har resonerat att även dessa element av filerna är ett uttryck för nätverkets estetik.
3. Jag har tillåtit mig att processa materialet med digitala effekter såsom distortion och reverb, samt att EQ:a det.
4. Jag har i viss mån tillåtit mig att klippa ner vissa ljudfiler, men detta är dock något som jag har varit ganska restriktiv med, då jag har kommit på mig själv med att annars klippa i dem så att man förlorar essensen i filen.
5. Slutligen har jag förbehållit mig rätten att själv välja ut de genererade filer som jag själv tycker är bäst och använt mig av dem.

I arrangeringen av de genererade ljudfilerna har jag strävat efter att bevara så mycket som möjligt av identiteten i ljudfilerna och försökt organisera dem på ett sätt där de förstärker varandras respektive inneboende kvaliteter.

## 4. Verk

I uppsatsen kommer jag presentera sex verk som illustrerar mina frågeställningar som jag presenterade i inledningen. Detta är ett urval som jag har gjort då jag anser att dessa verk ger en bra överblick över de olika utfall som projektet har resulterat i. Verken finns också representerade som vidlagda ljudfiler i appendix.

### 4.1 *N\_midi*

**Titel:** *N\_midi*

**Durata:** 03:58

**Uruppförande:** LjudOljud

**Dataset:** Giovanni Pierluigi da Palestrina, *Missa Papae Marcelli* (1562), *Missa Brevis* (1570).<sup>35</sup>

**Datasetmanipulationer:** -

**Algoritm:** Polyphony RNN

Verket är baserat kring MIDI-data genererat av ett dataset bestående av två av Giovanni Pierluigi da Palestrinas mässor. Den genererade MIDI-datan processeras i SuperCollider, genom en algoritm som lyfter ut kortare fraser ur kompositionen, och styr sedan olika delar av kompositionen. All ljuddesign är gjord i SuperCollider.

Stycket undersöker hur mycket av stilen i en genre som är beroende av tempo, och MIDI-filen spelas därför upp på 0.25 gånger hastigheten mot det genererade original tempot.

### 4.2 *N\_orgl*

**Titel :** *N\_orgl*

**Durata:** 11:32

**Uruppförande:** Sound Of Stockholm

**Dataset:** Kali Malone, *The Sacrificial Code* (2019), Theodor Kentros, *Fractura* (2019), *approximately fifteen.five minutes surrounding f & c* (2021).<sup>36</sup>

**Datasetmanipulationer:** Svältträning.

**Algoritm:** *PRISM sampleRNN*

---

<sup>35</sup> Giovanni Pierluigi da Palestrina, *Missa Brevis* [Musik] 1570, *Missa Papae Marcelli* [Musik], 1562.

<sup>36</sup> Kali Malone, *The Sacrificial Code* [Musik], 2019, Theodor Kentros, *Fractura* [Musik], 2019, *approximately fifteen.five minutes surrounding f & c* [Musik] 2021.

Stycket består av ett material genererat med neuralsyntes och är ett av mina första framgångsrika experiment med denna teknik. Nätverket är tränat på ett dataset bestående av kontemporär orgelmusik och har därefter genererat en större mängd samlingar på mellan 30 och 60 sekunder. I styckets inledande del är ett mindre urval av dessa samlingar arrangerade fritt över varandra utan något externt kompletterande material. Del A börjar med en ljudfil från epok 10, där nätverket enbart lyckats producera brus, och gradvis introduceras filer från senare epoker, först 30, sedan 35 och till sist 40. Detta resulterar i att stycket blir en slags reflektion över den träningsprocess som lett fram till det, där man kan följa nätverkets utveckling över tid.

I styckets B-del har jag använt en något friare arrangerings metod, då jag har kompletterat det neuralt genererade materialet med ljud syntetiserat i SuperCollider. Jag har dock strävat efter att förstärka inneboende gestik och klang i det neuralt genererade materialet snarare än att lägga till externt komponerat material.

### 4.3 JB22

**Titel :** JB22

**Durata:** 05:01

**Uruppförande:** Ej uppfört

**Dataset:** Kablam, *Confusia* (2019), *skröp* (2020), Theodor Kentros, *Fractura* (2019), *approximately fifteen.five minutes surrounding f & c* (2021).<sup>37</sup>

**Datasetmanipulationer:** Kontaminerat dataset, svältträning.

**Algoritm:** PRISM sampleRNN

JB22 är komponerat med överblivet material som genererats till andra stycken. Jag har arrangerat det relativt fritt men har begränsat mig till ett mindre antal samlingar. Stycket skiljer sig alltså från de andra verken då jag inte har skapat ett specifikt dataset som stycket utgår ifrån utan istället har återanvänt gamla dataset. Stycket består också enbart av neuralt genererat material.

Stycket har två delar där den första delen har en dynamisk rörelse uppåt och den andra neråt, vilket resulterar i att den övergripande formen på stycket blir alltså en slags peak form.

---

<sup>37</sup> Kablam, *Confusia* [Musik], 2019, *skröp* [Musik], 2020



#### 4.4 Ettusentvåhundraepoker

**Titel :** Ettusentvåhundraepoker

**Durata:** 05:00

**Uruppförande:** LjudOljud.

**Dataset:** -

**Datasetmanipulationer:**

**Algoritm:** Verkspecifikt neuralt nätverk

Stycket är centrerat kring ett nätverk som under 1200 epoker lär sig att spela trummor. Till skillnad från de andra styckena generas det här stycket i realtid, därför hörs resultat och framsteg för varje epok direkt i kompositionen. Algoritmen fungerar på det sättet att nätverket från början inte har sett någonting av datasetet och därför gissar helt vilt på hur den ska agera. Efter varje epok får den se mer och mer av datasetet med resultatet att den utvecklas gradvis genom stycket. När 1200 epoker har passerat är stycket slut. Stycket är gjort i samarbete med Theodor Kentros och all ljudsyntes är gjord av oss medan struktur och form är skapat av nätverket.

#### 4.5 N\_gamba

**Titel :** N\_gamba

**Durata:** 05:24

**Uruppförande:** LjudOljud

**Dataset:** Giovanni Pierluigi da Palestrina, *Missa Papae Marcelli* (1562), *Missa Brevis* (1570)

**Datasetmanipulationer:** -

**Algoritm:** Polyphony RNN

Verket består av MIDI-data genererad från samma dataset som *N\_midi*. Jag valde jag ut två delar om 16 takter vardera som upprepades tre gånger var. Varje gång en del upprepades oktaverades den också. Stycket är skrivit för Theorbo och Viola da Gamba och har en kanon liknande form. Det inleds med att Viola da Gamba spelar ensam och efter 16 takter, när stämman oktaveras, inträder Theorben. Verket har en minimalistisk form och är uppbyggd kring ett ostinato.

## 4.6 *N\_lånky*

**Titel :** *N\_lånky*

**Durata:** 06:44

**Uruppförande:** Ej uppfört

**Dataset:** Linus Hillborg, *Magelungsverket* (2021), Kablam, *Confusia* (2019), *skröp* (2020), Theodor Kentros, *Fractura* (2019), *approximately fifteen.five minutes surrounding f & c* (2021), Erik Klinga, *Fältinspelning* (2021).<sup>38</sup>

**Datasetmanipulationer:** Svältträning, Kontaminerat dataset

**Algoritm:** PRISM sampleRNN

*N\_lånky* är uppbyggt kring ett material som är genererat från ett dataset bestående av Linus Hillborgs skiva *Magelungsverket*. I stycket är detta material presenterat såsom det är genererat och arrangerat på ett liknande sätt som i *N\_orgl*. Istället för att komplettera detta material med externt genererat material har jag istället förstärkt och kompletterat inneboende gestik i ljudfilerna med neuralt genererat material.

Formen i stycket påminner om formen i *JB22* då verket också har två delar där dynamiken i den första delen stegras medan den i andra delen faller av.

---

<sup>38</sup> Linus Hillborg, *Magelungsverket* [Musik], 2021, Erik Klinga, *Fältinspelning* [Musik], 2021.

## 5. Resultat och Diskussion

I projektet har jag undersökt på vilka sätt jag kan förskjuta min agens i förhållande till mina kompositioner. Jag kommer här kritiskt reflektera över hur detta har fungerat i relation till de verk jag presenterar i texten, och på vilka sätt min metod har påverkat styckena. Min agens som kompositör fungerar olika i verken, dels beroende på hur jag har manipulerat dataseten men också beroende på hur bra jag lyckats förhålla mig till mitt ramverk. I *N\_midi* gick jag ifrån ramverket genom att jag skrev en algoritm som slumpmässigt hoppade till olika platser i den genererade MIDI-filen. Detta resulterade i att slutresultatet kan likställas med brus då identiteten i det genererade materialet helt försvann. Konsekvensen blev att det neuralt genererade materialet lika gärna hade kunnat genereras med kvantiserade slumpgeneratorer. Jag betraktar därför stycket som misslyckat i förhållande till projektets syfte, då stycket blev ett uttryck för min egen estetik snarare än nätverkets. Min agens är inte heller förskjuten i någon annan mening än att stycket är ett resultat av en slumpmässig process.

Som en reaktion på detta hade jag ett mer restriktivt förhållningssätt i *N\_gamba*. Jag använde mig av samma dataset i stycket, och var den här gången noga med att bevara intervall och ackordföljder intakta såsom dem genererats. Jag tillät mig dock att i enlighet med det ramverk jag satt upp oktavera sekvenserna. I och med att jag höll mig från att involvera mig för mycket med materialet skulle jag argumentera för att min agens är förskjuten i stycket och min roll kan liknas vid en slags redaktörs-roll. Stycket kan ses som ett uttryck för datasetets estetik i den mening att jag enbart presenterar material genererat av nätverket, men det kan också ses som ett uttryck för min egen i någon mening då jag har valt ut ett begränsat urval av den genererade filen.

*Ettusentvåhundraepoker* är det stycke som skiljer sig mest från de övriga i projektet, dels då det är helt gjort i SuperCollider och dels för att det genereras i realtid. Verket relaterar heller inte till kontextuella relationer i lika hög grad som de andra styckena, då datasetet inte består utav andra kompositörers verk. Stycket kan ses som ett slags redovisande över nätverkets inlärningskurva och det som i vanliga fall skulle vara förarbetet lyfts istället fram och utgör själva stycket. Detta innebär att jag får en slags tekniker-roll som skiftar fokus mot det tekniska materialet snarare än att jag som kompositör visar ett slutresultat.

Manipuleringen av dataseten och träningsprocessen är ett försök att tvinga nätverket att generera material med en estetik som tar sin utgångspunkt i datasetet men likväl har en särart. Genom att justera signal-brus förhållandet justeras också avvikelserna från datasetet, något som också får

konsekvenser för min agens i förhållande till materialet. Desto högre brusratio i dataseten desto mindre förutsägbart blir också resultatet vilket innebär att jag får mindre kontroll över utfallet.

I *N\_orgl* svälttränades algoritmen vilket resulterade i brusiga ljudfiler med oförutsägbart rytmik. Detta medförde att de genererade filerna hade ett distinkt uttryck som skilde sig från materialet i datasetet. Materialet som stycket består av är presenterat såsom det är genererat, men med vissa externa ljudkällor som förstärker inneboende gestik i ljudfilerna. Detta resulterar i att min agens är förskjuten i förhållande till stycket då jag har presenterat materialet i en kontext snarare än att komponera med det. Min roll i stycket kan liknas vid en slags kurators-roll och ljudfilerna kan ses som de individuella verken medan formen representerar utställningen verken presenteras i. I stycket blir också nätverkets output en konsekvens av den simulerade kontext som jag skapat genom datasetet, då denna output presenteras såsom den är. Eftersom att ljudfilerna arrangeras av mig som på många sätt verkar i samma sammanhang som kompositörerna i datasetet uppstår en mängd kontextuella relationer. Dels mellan kompositörerna i datasetet, sedan i den simulerade kontext som datasetet utgör och därefter mellan mig, datasetet och kompositörerna. Det hela kan liknas vid ett komplext feedback-nätverk där varje punkt både refererar bakåt och framåt, vilket relaterar till Beckers tankar om kulturell produktion.

*JB22* är sammansatt av överblivet material från andra stycken och fungerar som ett slags collage-stycke. I verket har jag en slags producent roll i den meningen att jag sammanfogat dessa delar till en helhet. Kompositörerna i datasetet och jag har en liknande kontextuell relation som i *N\_orgl* men här fungera feedback-nätverket också i ytterligare en dimension, då Kajsa Blom vars musik är med i datasetet, också har använt sig av filerna. Konsekvensen av detta blir en typ av meta-relation, dels mellan Bloms musik som inkorporerar dessa ljudfiler samt den simulerade kontext bestående av Bloms musik som skapat filerna. En tredje relation uppstår också mellan filerna i mitt stycke och filerna i Bloms och nätverket blir därför i någon mening också *vertikalt*, i den mening att det inte längre bara refererar framåt och bakåt utan också mellan stycken skapade med samma material. Detta bidrar till att min agens i förhållande till verket förskjuts och blir grumlig då gränserna mellan material och upphovsperson tunnas ut.

I *N\_länky* har jag en liknande roll som i *N\_orgl* i den mening att jag även i detta stycke presenterat ljudfilerna såsom de är och snarare arrangerat än komponerat med dem. De kontextuella relationerna mellan mig datasetet och kompositörerna i datasetet är också komplexare i detta stycke. Den vertikala koppling som uppstod i relation till *JB22* återfinns även i detta stycke, då Bloms musik är med i datasetet, men det går också att spåra tydliga kontextuella kopplingar mellan två av de andra kompositörerna i datasetet, nämligen Linus Hillborg och Theodor Kentros som

tillsammans utgör duon *Sänkt*. Detta innebär en tydlig vertikal koppling även utanför datasetet vilket skapar ett komplext nätverk av refererande, både utanför och innanför datasetet.

Genom projektet sker en intensifiering av de kontextuella relationerna mellan mig, datasetet och kompositörerna och det uppstår kopplingar som påminner om Beckers tankar om kulturell produktion. Jag vill mena att det som händer i den virtuella kontexten också är det som händer i verklighetens kontexter. Därför menar jag att det inte går att tala om ett individuellt skapande på något meningsfullt sätt, då skapande inte uppstår i ett vakuum. Precis som i maskininlärningsalgoritmerna vill jag mena att skapande uppstår i relation till den kontext konstnären verkar i, och alltid influeras och formas av sitt sammanhang.

Detta aktualiserar också frågan om relationen mellan musik skapad med neurala nätverk och upphovsskap. Eftersom det genererade materialet kan ses som ett utfall från ett komplext nätverk av relationer är det inte uppenbart att jag kan tillskrivas någon äganderätt för filerna. Jag menar därför att filerna kan ses som en kollektiv produktion, en slags generalisering av en social kontext. Därför blir det inte relevant att tala i termer om upphovsskap angående neuralt genererat material. Inte på grund utav någon inneboende intelligens eller kreativitet hos maskininlärnings-systemen, utan just eftersom att filerna kan ses som en generalisering av en mänsklig erfarenhet.

## 5.1 Framtida arbete

Jag planerar att fördjupa mitt arbete med maskininlärnings-tekniker genom att utveckla sätt för att arbeta med neurala nätverk i realtid. Ett exempel på detta är att vidareutveckla min egna SuperCollider-klass *Neuron* som jag har implementerat i samarbete med Viktor Sandström. Neuron är en modulär byggsten bestående av en enskild neuron och därför går att konfigurera till en mängd olika nätverk. Min förhoppning är att med detta verktyg implementera *Reinforcement learning* system där nätverkets framsteg hörs direkt i kompositionen. Jag planerar också att utveckla mitt arbete med feedback-kopplade dataset för att på det sättet fördjupa mina snedträningstekniker.

## 5.2 Sammanfattning

I projektet har jag förhållit mig till det musikaliska materialet och träningsprocesserna på olika sätt. Genom relationerna mellan mig och materialet har jag kunnat laborera med min upphovsroll. Beroende på hur mycket jag involverat mig i de genererade filerna, samt hur jag har manipulerat träningsprocessen har jag kunnat förflytta min egen position i förhållande till stycket. Ett antal av verken som presenteras i uppsatsen skulle jag anse som lyckade utfall givet projektets syfte, och då

graden av inflytande jag har haft på kompositionen ser olika ut i styckena förändrar det också min relation till dem.

Slutligen har jag diskuterat hur dataseten kan förstås som en simulerad kontext och vad detta får för konsekvenser i relation till min agens i kompositionerna. Jag har också visat på de sammanflätade nätverk av kontextuella relationer som uppstår i musiken på grund av detta och jag har argumenterat för att detta påminner om hur kulturell produktion fungerar i praktiken.

## 6. Källor

### 6.1 Tryckta källor och litteratur

Becker, Howard, *Art Worlds*, University of California Press, Berkeley, (2008)

Cluley, Robert. "Art words and art worlds: The methodological importance of language use in Howard S. Becker's sociology of art and cultural production." *Cultural sociology* 6.2 (2012), s. 201-216.

Dahlstedt, Palle. "Musicking with Algorithms: Thoughts on Artificial Intelligence, Creativity, and Agency." *Handbook of Artificial Intelligence for Music*. Springer, Cham, (2021), s. 873-914.

Fernández, Jose D & Francisco Vico. "AI methods in algorithmic composition: A comprehensive survey." *Journal of Artificial Intelligence Research* 48 (2013), s. 513-582.

Frisk, Henrik. "Aesthetics, Interaction and Machine Improvisation." *Organised Sound* 25.1 (2020), s. 33-40.

Mehri, Soroush, et al. "SampleRNN: An unconditional end-to-end neural audio generation model." (2016), hämtad från <https://arxiv.org/pdf/1612.07837.pdf>, 2022-01-20

Roads, Curtis. *Composing electronic music: a new aesthetic*. Oxford University Press, USA, (2015)

Zukowski, Zack & C. J. Carr. "Generating black metal and math rock: Beyond bach, beethoven, and beatles." (2018), hämtad från <https://arxiv.org/pdf/1811.06639.pdf>, 2022-01-15

Zukowski, Zack & C. J. Carr. "Generating albums with samplernn to imitate metal, rock, and punk bands." (2018), hämtad från <https://arxiv.org/pdf/1811.06633.pdf>, 2022-01-20

## 6.2 Elektroniska källor

<https://dadabots.com/faq.php>, 2022-03-10

<https://vickyclarke.org/about/>, 2022-03-10

<https://auramachine.blog/2022/01/04/residency-framework-why-machine-learning-musique-concrete>, 2022-05-19

<https://www.theguardian.com/music/2019/may/02/holly-herndon-on-her-musical-baby-spawn-i-wanted-to-find-a-new-sound>, 2022-02-20

<https://holly.mirror.xyz/54ds2IiOnvthjGFkokFCoaI4EabytH9xjAYy1irHy94>, 2022-02-20

<https://www.artnews.com/art-in-america/interviews/holly-herndon-emily-mcdermott-spawn-ai-1202674301/>, 2022-02-20

<https://composerprogrammer.com/teaching/supercollider/advanced/index.html>, 2022-05-18

<https://deepai.org/machine-learning-glossary-and-terms/epoch>, 2022-05-19

<https://www.wendycarlos.com/+sob.html>, 2022-05-18

## 6.3 Musik

Dahlqvist, Niklas, *Etttusentvåhundraepoker* [Musik], 2021

Dahlqvist, Niklas, *JB22*, [Musik], 2022

Dahlqvist, Niklas, *N\_gamba*, [Musik], 2022

Dahlqvist, Niklas, *N\_lånky*, [Musik], 2022

Dahlqvist, Niklas, *N\_orgl*, [Musik], 2022

Dahlqvist, Niklas, *N\_midi*, [Musik], 2022



Hillborg, Linus, *Magelungsverket*, [Musik], 2021

Kablam, *Confusia* [Musik], 2020

Kentros, Theodor, *Fractura* [Musik], 2019, *approximately fifteen.five minutes surrounding f & c*  
[Musik], 2021

Klinga, Erik, *Fältinspelning* [Musik], 2021

Malone, Kali, *The Sacrificial Code* [Musik], 2019

Giovanni Pierluigi da Palestrina, *Missa Papae Marcelli*, [Musik], 1562, *Missa Brevis*, [Musik], 1570

## **6.4 Resurser**

[https://github.com/magenta/magenta/tree/main/magenta/models/polyphony\\_rnn](https://github.com/magenta/magenta/tree/main/magenta/models/polyphony_rnn)

<https://github.com/rncm-prism/prism-samplernn>

<https://colab.research.google.com/>

## **7. Bilagor**

**7.1 Ljudinspelning av *Ettusentvåhundraepoker* (2021)**

**7.2 Ljudinspelning av *JB22* (2022)**

**7.3 Partitur till *N\_gamba* (2022)**

**7.4 Ljudinspelning av *N\_lånky* (2022)**

**7.5 Ljudinspelning av *N\_orgl* (2022)**

**7.6 Ljudinspelning av *N\_midi* (2022)**

