Meta-Human

Meta-Human ist ein Projekt von Peter Martin und Fynn Freist, das im Rahmen des Moduls “Musikinformatik” im Wintersemester 2021/22 am Institut für Musik und Medien der Robert Schumann Hochschule Düsseldorf entstanden ist.

Abstract: For MetaHuman two neural networks have been trained on YouTube songs and learned how to sing. A closed loop was created, where the two networks sing to each other and learn from what they hear. If this process is repeated often enough at some point there will be more machine-songs than human-songs in the training-set of the networks - at this point they've learned more from each other, then from humans. We hope that this leads to "meta-" or "post-human" music.

Warum machen Menschen Musik mit Computern? Selbstverständlich gibt es eine Vielzahl an Beweggründen. Alex McLean beschreibt das Spektrum dieser Beweggründe folgendermaßen:

> At one extreme lies the claim of the independently intelligent algorithm, in the form of computational agents which are deemed to be creative (...). At the other extreme, algorithms are treated more like musical notations, which humans work with and adapt as vehicles for their own creativity. (McLean 2018)

Einerseits können Computer also ähnlich wie herkömmliche Instrumente bzw. Notationssysteme als Werkzeuge genutzt werden, die Menschen erlauben ihrer Kreativität Ausdruck zu verleihen. Andererseits könnten Computer unter gewissen Umständen aber auch selbst als kreative Akteure auftreten und so zumindest zu “Co-Performern” eines Menschen werden.

Diese Ansicht scheint zunächst nicht zu einem Verständnis von Musik zu passen, das beispielsweise Jürgen Habermas vertritt. Laut Habermas kann Musik als eine Form der Kommunikation zwischen Menschen verstanden werden. Eine KünstlerIn drückt in ihrem Werk etwas aus und erschafft dadurch Bedeutung. Natürlich soll damit nicht unterstellt werden, KünstlerInnen und/oder Rezipienten könnten immer verbal ausdrücken, was die Bedeutung eines Werkes ist. Stattdessen soll nur die schwächere These vertreten werden, dass Kunstwerke eine Form der Kommunikation sind, die Bedeutung in irgendeiner Form transportieren. Ein kreativer Akt kann dann als eine Form der Kommunikation zwischen KünstlerInnen und RezipientInnen verstanden werden - oder auch noch

weiter zwischen KünstlerInnen - RezipientIn einerseits und zwischen RezipientInnen - RezipientInnen andererseits.

"Exploratory creativity is the process of exploring a given conceptual space; transformational creativity is the process of changing the rules which delimit the conceptual space." (Wiggings and Forth 2018)

Nach diesem Verständnis von Musik ist nun zunächst unklar, inwiefern ein nicht-menschlicher Akteur als Kommunikationsteilnehmer auftreten kann. Schließlich würde Habermas unterstellen, dass für einen Kommunikationsbeitrag bestimmte Geltungsansprüche erhoben werden müssen - beispielsweise auf Wahrheit oder Wahrhaftigkeit. Um einen sinnvollen Diskurs zu führen, müssen die Teilnehmer den Anspruch erheben, die Wahrheit zu sagen und das was sie sagen, auch "wirklich so zu meinen". Insbesondere letzteres kann kaum einem Computer unterstellt werden. Diese und andere Bedingungen, die Habermas an einen kommunikativen Diskurs erhebt, setzen eine gewisse Form der Intention bei den Kommunikationsteilnehmern voraus. Sie müssen mit ihrem Beitrag eine Absicht verfolgen. Auch hier darf die These nicht zu stark gelesen werden. Es soll nicht behauptet werden, eine KünstlerIn habe immer ein konkretes Vorhaben oder sogar ein Ziel, das ihr Werk erreichen soll. Es soll lediglich angenommen werden, dass ein kreatives Werk geschaffen wird, um damit Menschen etwas zu vermitteln.[1] Es scheint nicht plausibel einem Computer eine ähnliche Intention zu unterstellen. Computer haben schlicht keine Absichten.
Selbstverständlich können Computer syntaktisch korrekte kommunikative Beiträge formulieren. Das ist besonders ersichtlich im Feld der Synthese natürlicher Sprache. Neuronale Netze schreiben bereits ganze Artikel. Aber mit Habermas kann angenommen werden, dass zu einem semantisch sinnvollen Beitrag noch weitere Bedingungen erfüllt sein müssen, als nur syntaktische Korrektheit. Auch Wiggins und Forth scheinen die Annahme zu teilen, dass es gewissen Regeln oder Bedingungen gibt, die dafür erfüllt sein müssen:

"E is a set of rules which define the evaluation of the creative outputs resulting from the agent's activity, appropriately contextualized. The formalism does not specify what this context is; it might be the subjective judgement of the creating agent, or the subjective combined judgements of other agents, or comparison with some objective measure." (Wiggins&Forth 2018)

Eine plausible Regel oder Bedingung, scheint zu sein, dass für einen bedeutungsvollen bzw. sinnvollen kreativen kommunikativen Beitrag irgendeine Form von Intention oder Absicht notwendig ist. Es muss die Absicht bestehen, irgendeine Form von Bedeutung zu vermitteln - anderen oder sich selbst. Da Computer keine Absichten haben können, scheint es nicht möglich zu sein, dass Computer bedeutungsvolle kreative Beiträge produzieren.

McLean gibt einen Hinweis darauf, wieso Computer als kreative Akteure dennoch interessant sein könnten. Ihn interessiert:

---

[1] Das setzt keine Öffentlichkeit voraus. Es kann sich auch, um einen Austausch mit MitmusikerInnen oder sogar um ein "Selbstgespräch" handeln.

“the question of what new musics are now being found through algorithmic means which humans could not otherwise have made” (McLean 2018)

Es kommt gerade darauf an, Musik zu machen, die Menschen nicht hätten produzieren können. Hier wird der wichtige Gegenpart eines Kommunikationsprozesses erkennbar. Neben dem- oder derjenigen, die einen Kommunikationsbeitrag äußert - in unserem Fall dem Computer-, gibt es schließlich noch die RezipientInnen. Auch im Falle von Computer-Musik sind dies Menschen. Schließlich wird auch diese Musik von Menschen gehört. McLean scheint anzunehmen, dass das menschliche Publikum einen Bedeutungsgehalt aus von Computern erzeugter Musik generieren kann:

The listener may or may not transform their perceptions of such music into a cognitive framework that corresponds to the algorithm’s own methods, but in either case they may gradually assimilate the music into a meaningful whole. (McLean 2018)

Das wirft eine interessante Frage an das oben skizzierte Kommunikationsmodell von Habermas auf: Kann ein Rezipient einen Bedeutungsgehalt aus einem Kommunikationsbeitrag generieren, ohne dass dieser Beitrag mit der Intention gemacht wurde, Bedeutung zu vermitteln?

Wenn unsere oben getroffene Annahme richtig ist, dass Computer nicht die Absicht haben können, etwas zu vermitteln, dann stellt sich die Frage, ob dennoch ein Austausch zwischen ihnen und Menschen erzeugt werden kann, in dem an irgendeiner Stelle Bedeutung generiert werden kann. Da wir oben dafür argumentiert haben, dass diese Bedeutung nicht an der Stelle der Äußerung vom Computer erzeugt werden kann, bleiben nur die RezipientInnen als diejenigen, die Bedeutung erzeugen können. Kann das menschliche Publikum Bedeutung erzeugen aus einem kreativen Beitrag, der ohne die Intention Bedeutung zu generieren entstanden ist? Das entspräche einem interessanten Sonderfall in Habermas Kommunikationsmodell. Denn die Bedeutung eines kreativen Beitrags würde in diesem Fall nicht von demjenigen erschaffen, der ihn äußert, nämlich dem Computer. Wir haben oben argumentiert, dass das nicht möglich ist. Stattdessen wird die Bedeutung erst von den RezipientInnen - quasi im Nachhinein - geschaffen.
Es kann sogar angenommen werden, dass diese Situation ausschließlich mit Computern erzeugt werden kann. Zumindest, wenn man annimmt, dass Menschen gar nicht anders können, als mit einer Intention Äußerungen zu tätigen. Man stelle sich nur jemanden vor, der versucht die beschriebene Situation zu erzeugen. Diese Person müsste einen kommunikativen Beitrag erschaffen, ohne irgendeine Intention oder Absicht. Das muss scheitern - nämlich schon an der Absicht, eben das zu tun.

Der Grund, warum Menschen mit Computern Musik machen, kann in diesem Fall mit Wiggins und Forth folgendermaßen zusammengefasst werden. Im Idealfall erschafft ein Algorithmus selbstständig ein Werk, das an sich noch keine Bedeutung hat - daher war es auch für Menschen bisher unvorstellbar. Die RezipientInnen - zum

Beispiel die Programmierer des Algorithmus - sind dennoch in der Lage eine Bedeutung zu generieren. Ihnen gefällt die Musik, da sie ihnen etwas vermittelt.
Sie stellen drei Bedingungen, damit dieser Idealfall erreicht werden kann:

"In order to fulfill the potential that the above analysis suggests, we need three key ingredients. The first is the ability for our computer to relate the meaning of a program to its syntax. The second is for our computer to have a model of our coder's preference. The third is for our computer to manipulate the syntactic constructs available to our coder so as to take on some of the creative responsibility for the music." (Wiggings&Forth 2018)

Mit der obigen Argumentation sollte gezeigt werden, dass die erste Bedingung nicht erfüllt werden kann. Wir wollten zeigen, dass ein Algorithmus selbst keine Bedeutung in ein syntaktisch korrektes Programm bringen kann. In diesem Projekt soll der Frage nachgegangen werden, ob dennoch das Publikum eine Bedeutung generieren kann.

Um das zu untersuchen muss aber zunächst eine Musik geschaffen werden, die tatsächlich **vom** Computer und nicht **mit** einem Computer von einem Menschen erzeugt wurde. Schließlich wollen wir nicht untersuchen, ob Menschen **mit** Computern bedeutungsvolle Musik machen können. Das sollte außer Frage stehen. Viel eher wird ein Programm gesucht, das selbständig Musik produziert und in das kein Mensch eine Bedeutung "geschmuggelt" hat. Aber wie könnte ein Algorithmus aussehen, der eigenständig Musik produziert?

Der Idealfall wäre ein Programm, das sich selbst schreibt und dann Musik produziert. In diesem Fall wäre keinerlei menschlicher Input vorhanden und somit wäre sichergestellt, dass - falls eine Bedeutung entsteht - diese nicht von einem Menschen kommen kann.

Neuronale Netze scheinen einen Form von Algorithmen zu sein, die diesem Idealbild momentan am nächsten kommen. Im Gegensatz zu traditionellen Algorithmen, optimieren diese wichtige Komponenten ihres Codes selbst. Insofern wird häufig auch davon gesprochen, dass diese Netzwerke selbstständig etwas lernen. Es darf aber nicht vergessen werden, dass dieses Lernen immer darin besteht, einen Fehler zu minimieren, der vorher vom Menschen definiert wurde. Insofern ist es entscheidend, was als Fehler gilt. In der Regel wird die Abweichung von einem gegebenen Vorbild mathematisch quantifiziert und als Fehler definiert. Wenn beispielsweise ein neuronales Netz lernen soll Musik zu machen, dann muss vorher eine "Zielmusik" definiert werden. Das Netz optimiert die Parameter im Code dann so, dass Klänge erzeugt werden, die dieser "Zielmusik" möglichst nahe kommen.

Das stellt ein Problem für dieses Projekt dar. Denn es zeigt, dass auch neuronale Netze sehr abhängig von menschlichem Input sind. Einerseits müssen Menschen entscheiden, welche Musik dem Netzwerk als "Vorbild" dient. Andererseits wird die ausgewählte Musik wohl immer menschliche Musik sein müssen. Einfach weil - zumindest nach unserem Kenntnisstand - noch keine nicht-menschliche Musik existiert.

Wir haben versucht beide Probleme möglichst zu minimieren.

Die Auswahl der Songs, mit denen unsere Programme trainiert wurden, haben wir nicht selbst getätigt. Stattdessen haben wir uns dazu entschieden einen weiteren Algorithmus dafür zu verwenden: den YouTube Recommendations Algorithmus. Bekanntlich ist auf der Videoplattform YouTube ein neuronales Netzwerk im Einsatz, das zu jedem Video weitere ähnliche Videos vorschlägt. Wir haben diesen Algorithmus genutzt, um ausgehend von einem Start-Video über 3000 Songs zu sammeln, indem schlicht immer auf die nächste Empfehlung geklickt wurde und der entsprechende Song heruntergeladen wurde.[2]

Das zweite Problem ist schon etwas schwieriger zu lösen. Denn auch wenn die Musik von einem Algorithmus ausgewählt wurde, handelt es sich dabei doch um menschliche Musik. Ein Netzwerk, das mit menschlicher Musik trainiert wurde, wird versuchen, menschliche Musik nachzuahmen. Die einzige Möglichkeit das zu unterbinden, ist es, das Trainingsset nach und nach durch nicht-menschliche Musik zu ersetzen. Aus diesem Grund haben wir zwei neuronale Netze trainiert. Beide sind unabhängig voneinander entstanden und mit unterschiedlichen Datensätzen trainiert worden. Beide Netzwerke können Klänge erzeugen. Indem wir die Musik, die das eine Netzwerk erzeugt, dem anderen Netzwerk als weitere Trainingsdaten zur Verfügung stellen, wurde ein geschlossener Loop erzeugt. So wird nach und nach der Datensatz erweitert, indem die Netzwerke sich gegenseitig “vorsingen”. Die Netzwerke erinnern sich quasi an die Musik des jeweils anderen und lernen so voneinander. Wir erhoffen uns, dass ab einem bestimmten Zeitpunkt mehr maschinelle Musik im Datensatz der Netzwerke vorhanden ist, als menschliche Musik. In diesem Moment hätten die Netzwerke mehr von sich gegenseitig gelernt als von Menschen. Auf diese Weise soll Musik entstehen, die sich von menschlicher Musik emanzipiert und insofern als meta- oder post-human gelten kann.

Dieser Prozess dauert sehr lange. Das Projekt kann insofern als work-in-progress betrachtet werden.

[2] Natürlich wurde auch der YouTube Algorithmus für Menschen entwickelt, und anhand menschlicher Präferenzen und Verhaltensweisen trainiert. Zumindest konnten wir aber die menschliche Komponente um eine Ebene verschieben.

Bisherige Ergebnisse:

Hier kann auf zwei Audiodateien und mehrere Videodateien zugegriffen werden. Auf den Audiodateien ist zu hören, wie eines der Netzwerke Lieder nachsingt, mit denen es trainiert wurde - nämlich “Hey Jude” von den Beatles und “Ave Maria”.
Auf den Videodateien ist zu sehen, wie sich die Netzwerke austauschen:

1 Imitation of a human song
Hier rekonstruiert eines der Netzwerke - mit dem Namen Valerio - einen Song mit dem es trainiert wurde. Das zweite Netzwerk - Dennis - versucht hört diesen Song und versucht ihn möglichst genau zu imitieren. Valerio hört diese Imitation und versucht seinerseits wiederum sie zu imitieren. Und so weiter.

2 Variation on a Human Song 1
Auch hier beginnt Valerio damit, einen Song zu rekonstruieren, der Teil seines Trainingssatzes war. Dennis antwortet diesmal aber nicht mit einer Imitation dieses Songs, sondern mit einer Variation. Valerio hört diese Variation und antwortet seinerseits mit einer Variation davon. Und so weiter.

3 *Variation on a Human Song 2*
Dieses Video funktioniert nach demselben Prinzip wie das Video *2 Variation on a Human Song 1.* Der einzige Unterschied ist, dass Valerio mit einem anderen Song beginnt.

4 Composing a Machine Song
In diesem Fall komponiert Valerio selbst einen Song und beginnt diesen abzuspielen. Dennis entwickelt eine Variation davon, die wiederum Valerio hört. Er antwortet mit einer Variation seinerseits. Und so weiter.

Der Code für das Programm kann auf GitHub gefunden werden. Für die Dokumentation der technischen Umsetzung liegen mehrere Jupyter Notebooks auf GitHub vor, die den kommentierten Code enthalten und eine Erläuterung der technischen Idee. Eine Anleitung, wie auf die Notebooks zugegriffen werden kann, ist ebenfalls auf GitHub verfügbar.

McLean, Alex, 2018: Musical Algorithms as Tools, Languages, and Partners: A Perspective. In: The Oxford Handbook of Algorithmic Music

Wiggins and Forth, Geraint A. and Jamie, 2018: Computational Creativity and Live Algorithms. In: The Oxford Handbook of Algorithmic Music

Meta-Human GitHub Projekt:
https://github.com/doktorbanana/MetaHuman#

Meta-Human Beispiele:
https://drive.google.com/drive/folders/183Xb4NSxg_AnOehmYINbBRmoD4Kh4UoY?usp=sharing