New Ethnic Becoming

Eine Annäherung an schlaue künstliche Stimmen

Friedemann Dupelius

Was für ein Eigenleben können Stimmen erhalten, wenn sie computergeneriert sind: Abgelöst von einer Sprecher*in werden sie auf besondere Weise Teil des täglichen Lebens und zum Tool künstlerischer Arbeiten. Dieser Text ist der neueste Teil einer von Friedemann Dupelius langfristig angelegten Recherche über künstliche Stimmen in Alltag, Technologie, Musik und Kunst.

Unsichtbar und körperlos haben sich die künstlichen Stimmen in unseren Alltag geschlichen. Sie sprechen zu uns aus Smartphones und schlauen Lautsprechern im Wohnzimmer, helfen beim Navigieren auf Bundesstraße und Datenautobahn. Sie wollen, dass wir weniger Screen Time verbringen, versprechen Freiheit der Bewegungen und Blicke, weniger viereckige Augen und mehr Intuition bei der Mensch-Maschine-Kommunikation. Dabei geben sie zunehmend vor, selbst ein Mensch zu sein – oder zumindest wie einer zu klingen. Das macht die Sache so heikel wie reizvoll. »Voice first!« heißt es seit rund zwei Jahren auf den Blogs und Newsportalen der Tech-Welt, nach »online first« und »mobile first« befinden wir uns demnach mitten in einem erneuten Paradigmenwechsel.

Nur folgerichtig setzen sich in jüngerer Zeit auch Musiker*innen und Künstler*innen mit synthetischen Stimmen auseinander. Deren künstlerische Verwendungszwecke sind so wandelbar wie die Entwicklung der Sprachsynthese an sich. Dabei docken diese Projekte immer auch an die technologischen, humanistischen und ethischen Fragen an, die uns jene Stimmen stellen.[1]

Industrie & Alltag

Im September 2016 begann mit der Einführung von Amazon Echo auch in Deutschland das Zeitalter der Smart Speakers. Einer Studie der Digital-Agentur Beyto aus dem Mai 2020 zufolge hat jede*r vierte Deutsche einen Smart Speaker zu Hause.[2] Der weltweite Absatz der intelligenten Sprachsysteme belief sich im 2. Quartal 2020 auf 30 Millionen. Marktführer ist Amazon mit 21,6 Prozent. Bis 2023 wird ein jährliches Wachstum von 35 Prozent pro Jahr prognostiziert. Smart Speakers wie Amazon Echo, Google Home, Google Nest oder der Apple Homepod werden der Studie nach vorrangig benutzt, um Alltag und Zuhause zu organisieren, Musik zu streamen oder Fragen zu beantworten. Hinzu kommen die Milliarden Smartphones, mit denen User*innen in verbale Kommunikation mit Siri (Apple), Google Assistant oder Cortana (Microsoft) in Verbindung treten können. Bereits 2020 erfolgt rund ein Drittel der Internetnutzung ohne Bildschirm.

Im Windschatten dieser Entwicklung sind neue Berufszweige und Unternehmen entstanden, die sich auf VUX (Voice User Interaction) und VUI (Voice User Interface) Design spezialisiert haben.

Die promovierte Linguistin Laura Dreessen hat hier ihre Nische gefunden und betreut als Teil der VUI.Agency die Entwicklung sprachbasierter Kommunikationssysteme für verschiedenste Kund*innen. Sie bezeichnet sich selbst als »Voice Architektin«. Ausgehend von ihrem linguistischen Fachwissen kann sie jeden Schritt von der Spracherkennung bis zur -ausgabe nachvollziehen. »Ich kann Maschinen das Sprechen beibringen, weil sich mein Studium ständig um das Abstrahieren von Sprache gedreht hat. Durch die linguistische Methodik hatte ich die optimale Vorbereitung, um Tausende von Sprachdaten zu sortieren und für Kommunikationsprozesse aufzubereiten.« In der VUI.Agency und den bislang wenigen anderen vergleichbaren Startups im deutschsprachigen Raum arbeiten interdisziplinäre Teams aus Informatik, Linguistik, (Sound) Design, Geisteswissenschaften und Kunst zusammen.

Computerstimme & Musik

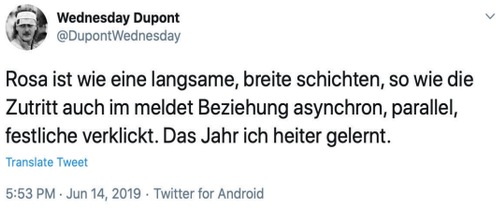
Seitdem Max Mathews 1961 in den Bell Labs einem IBM-Computer beibrachte, das Lied »Daisy Bell« zu singen, wurden eine Menge Phoneme über Lautsprecher entsendet. 1974 komponierte Charles Dodge den Zyklus *The Story of our Lives*, in dem eine synthetische Stimme das unspektakuläre Liebesleben eines Paares in Rezitativ-Weise singend erzählt. Die von Dodge benutzte Software wurde an der Columbia University extra für den musikalischen Gebrauch gestaltet. Gérard Grisey komponierte in *Les chants de l'amour* (1981-84) einen Dialog für synthetische Stimme mit einem Vokalensemble. Dafür verwendete er die am IRCAM entwickelte Software Chant. Parallel dazu hielt der Vocoder Einzug in die Popmusik, und ließ von Disco über Synth-Pop bis Electro und Techno kaum ein Genre unberührt. Allerdings basiert die in den 1930ern entwickelte Vocoder-Technologie auf dem Live-Input realer menschlicher Stimmen, im Gegensatz zu den vollsynthetischen Ansätzen von Mathews, Dodge und Grisey.

Auch das prägende massenkulturelle Phänomen artifizieller Stimmlichkeit der letzten Jahre

Die Twitter-Gedichte von Wednesday Dupont entstehen mittels Wischen auf der Handytastatur und gehen auf die Wort- und Auto-Correct-Vorschläge des Smartphones ein, die wiederum auf dem Wortschatz von Dupont basieren © Wednesday Dupont

ist nicht rein synthetischer Natur. Autotune ist ursprünglich ein Werkzeug, um intonatorische Ungenauigkeiten einer Singstimme zu glätten. Heute siedelt es sich in seiner Verwendung zwischen Effekt und Instrument an und darf in kaum einem Streaming-Hit oder Shisha-Bar-Soundtrack fehlen. Darüber, ob Autotune als Sinnbild für streamlinienförmige Berieselung oder doch als genialer Kunstgriff gelten darf, wird nicht nur in YouTube-Kommentarspalten unter HipHop-Videos diskutiert. Autotune changiert zwischen vermeintlicher Authentizität und ausgestellter Künstlichkeit. Bei einer Künstler*in merzt es musikalische Schwächen aus, bei der anderen fügt es ihrer Stimme eine neue Klangfarbe hinzu, mit Bedacht ausgewählt und in der Intensität ihrer Verwendung behutsam gesteuert.

Das allerneueste heiße Ding in Sachen künstliche Stimme und (elektronische) Musik ist KI. Die US-amerikanische Computermusikerin Holly Herndon hat mit dem Programmierer Mat Dryhurst die Stimme Spawn entwickelt, die Herndon als Erweiterung ihrer selbst betrachtet und eine Hauptrolle auf ihrem aktuellen Album *Proto* und den daran anknüpfenden Live-Performances einnimmt. Das Duo Amnesia Scanner ist mit der Stimme Oracle zum Trio angewachsen, hüllt den Sound seiner sogenannten »Deconstructed« Dance



Music damit in ein unverkennbares Gewand und sich selbst ins Schweigen darüber, wie Oracle genau entstand. Kurz vor Veröffentlichung steht das AI-Voice-Projekt von Mouse on Mars. Auch die zwei Berliner haben in den vergangenen Monaten an einer künstlichen Stimme gearbeitet, die eine prägende Rolle auf ihrem kommenden Album spielen wird und zugleich techno-ethische und postkoloniale Diskurse inkorporiert. Allein diese drei Stimmen unterscheiden sich in ihrer Klanglichkeit enorm voneinander. Im Gegensatz zu Autotune, das bei allem künstlerischen Potential auch für eine gewisse Vereinheitlichung und Komplexitätsreduktion des klanglichen Spektrums steht, zeichnen sich die musikalischen KI-Stimmen neuester Generation durch eine hohe Diversität aus, die in ihren klanglichen Kippmomenten mitunter unsicher und verletzlich erscheint.

All diesen Stimmen ist gemein, dass sie mittels Machine Learning entstanden sind, also: Eine Software lernt durch zahllose Proben humanen stimmlichen Klangmaterials, immer ähnlicher wie jenes zu klingen. Die Maschine entwickelt ein neuronales Netz, in dem sich alle ihrer zig Erfahrungen miteinander verknüpfen. Bringt man ihr auch bei, welchen Text das menschliche Stimmvorbild spricht, kann sie selbst irgendwann sprachliche Äußerungen vornehmen. Kurz: Je mehr Input, desto größer das neuronale Netz, desto hochwertiger das klangliche Ergebnis.

Klischee & Diversity

Welches Geschlecht hat eine Maschine? Kann einer künstlichen, körperlosen Stimme überhaupt ein Geschlecht zugeschrieben werden? Oder sind es wir, die automatisch versuchen, eine gehörte Stimme in eine Box einzusortieren? Im Consumer-Bereich sind die Zuschreibungen in ein binäres Geschlechterspektrum in der Regel gewollt und die damit einhergehenden Problematiken offensichtlich. Assistenzpersönlichkeiten von Alexa bis Siri sollen bewusst weiblich klingen und mit vermeintlich femininen Eigenschaften wie Hilfsbereitschaft und Einfühlsamkeit konnotiert werden. Stimmen, die Instruktionen geben, etwa bei Navigationsgeräten oder Sicherheits-Bots, haben öfter einen männlichen Klang. Strukturelle Sexismen in der Tech-Branche wie der Gesamt-Gesellschaft schreiben sich auch in diese Stimmen ein, was sie so gar nicht mehr künstlich, sondern auf unangenehme Weise sehr »authentisch« werden lässt. Sicher sind die oft männlich dominierten Entwickler*innen-Teams und deren oftmals männlichen Kapitalgeber*innen ein Grund dafür. »Natürlich spielen Markenaspekte eine große Rolle«, weiß Voice-Architektin Laura Dreessen aus ihrer Berufspraxis. »Kunden aus Brandabteilungen, die zum ersten Mal mit dem Thema Voice und Assistenzpersona konfrontiert sind, denken zunächst in Kriterien wie: ›Unsere Marke ist männlich, stark und innovativ.‹ Sie wünschen sich von uns die Entwicklung einer dementsprechenden Stimme. Wir können diesen Prozess aber steuern und auf andere Ebenen lenken.« Dreessen sieht sich und

ihre junge Branche in einer verantwortungsvollen und einflussreichen Position, von der aus über die bewusste Gestaltung von (Voice-)Technologie Wege in eine diversere Welt erschlossen werden können.

Die Voice-Architektin macht klar, dass ihr Diversität in der stimmbasierten Kommunikation ein Anliegen ist, sieht die Angelegenheit aber in einem komplexeren Kontext: »Es geht immer um Abstraktion. Was macht ein System als guten Konversationspartner aus? Wie reagiert es? Welche Signale möchtest du mit dieser Konversation senden? Wir haben die Verantwortung, einen ganzen Charakter zu erschaffen. Wortwahl, Sound, Stimmauswahl sind wichtige Parameter dabei. Wir versuchen, unsere eigene Philosophie in der Ausarbeitung einer Assistenzpersona zu prägen. Die Frage nach dem Geschlecht ist dabei ein Teil von vielen. Wir müssen vor allem für ein breites Verständnis dafür sorgen, was überhaupt in der KI-basierten Voice-Kommunikation passiert. Wenn die Menschen verstehen, dass sie selbst bestimmen und manipulieren können, wie eine Stimm-KI funktioniert, können wir auch dahin kommen, dass sie eine andere Einstellung zu Geschlechterrollen entwickeln. Die Maschinen spiegeln nur, was wir tun und sprechen.«

Eine Software lernt durch zahllose Proben humanen stimmlichen Klangmaterials, immer ähnlicher wie jenes zu klingen.

Ein Aspekt davon sei auch der bewusste Umgang mit Respekt und Höflichkeit in der Kommunikation, der sich in der Entwicklung bereits mitgestalten lässt. Dies beginne damit, dass das Mikrofon des Assistant am Ende einer Konversation noch drei Sekunden aufbleibt, damit sich die Userin bedanken kann – ein früher nicht berücksichtigter Aspekt, der für Dreessen aber eine Menge ausmachen kann. Auch werde das Thema Beleidigung und (die meist durch Männer vorgenommene) sexuelle Belästigung in der Entwicklung stärker berücksichtigt, und die Voice Assistants von heute reagieren darauf besser, ob selbstbewusst oder mit strikter Ignoranz.

Im März 2019 erhob Q zum ersten Mal die Stimme und stellte sich als »the first genderless Voice« vor. Q entstand in einer Kollaboration verschiedener Initiativen, darunter die NPO Equal AI und Copenhagen Pride. Die auf KI-Technologie basierende Stimme von Q bewegt sich im Frequenzspektrum zwischen 145 und 175 Hertz – ein Audiobereich, der empirischen Studien zufolge als weder eindeutig männlich noch weiblich wahrgenommen wird. Trainiert wurde die KI mit sprachlichem Lautmaterial mehrerer Personen unterschiedlicher Geschlechter. Eine Herausforderung dabei war, den Eindruck zu vermeiden, als würde man für Q lediglich eine männliche Stimme hoch- bzw. eine weibliche herunterpitchen, wofür viel an der spezifischen Zusammensetzung der Formanten gearbeitet wurde. Laut der Entwickler*innen will Q dauerhaft mit Genderklischees und deren Reproduktion durch KI-Assistenzsysteme aufräumen und für eine diverse, inklusive Zukunft der Tech-Branche kämpfen. Als dauerhafte »3rd option« will sich die genderneutrale Stimme auf den großen Plattformen etablieren.

Macht & Selbstermächtigung

Eine dezidiert weibliche Identität schreibt Holly Herndon ihrer KI-Stimme Spawn zu, die sie mit den Programmierern Mat Dryhurst und Jules LaPlace entwickelt hat und die ihr aktuelles Album *Proto* klanglich trägt. Anders als in vielen großen und kleinen Tonstudios (und genauso Tech-Konzernen) sollte diese weibliche Stimme eben nicht nach den Idealvorstellungen von

Männern geformt werden. Der »Code« von Spawn ist öffentlich zugänglich, womit sich Herndon für mehr Begegnungen zwischen KI-Technik und Bürger*innen stark macht.

Auch jenseits der Kunst kann moderne Stimmtechnologie die Wirklichkeit vieler Menschen verbessern. Menschen, die aufgrund von Krankheit oder Unfall keine Stimme (mehr) haben, können (wieder) eine erhalten. Auch die Konservierung von Stimmen, die beispielsweise aufgrund einer bevorstehenden Kehlkopf-OP bald nie wieder würden klingen können, ist mit KI-Technologie möglich. Solche Services kosten teils »nur« wenige 1.000 Euro und die Ergebnisse klingen durchaus lebensnah.

»Ich würde mir wünschen, dass wir Assistenzsysteme für immer kleinere Gruppen bauen«, sagt die Voice-Architektin Laura Dreessen. »Wir sprechen zwar von persönlicher Assistenz, aber etwas, das für Tausende von Nutzer*innen gebaut ist, ist nicht persönlich.« Sie nennt die Erkennung von Dialekt und Akzent, aber auch das Eingehen auf bestimmte Krankheitsbilder und Sprachfehler, kurzum: Barrierefreiheit, als offene Baustellen der VUI-Entwicklung.

Das US-Startup Modulate setzt sich gegen die Bullying-Unkultur im Gaming ein. Insbesondere Frauen oder sehr junge Spieler*innen werden von anderen, meist männlichen Gamern für ihr Geschlecht oder Alter gemobbt, auf welches sich im Online Gaming lediglich durch die via Headset offengelegte Stimme schließen lässt. Abhilfe sollen die von Modulate entwickelten Voice Skins schaffen. Damit kann sich die User*in wie mit einem optischen Avatar verkleiden und ihre akustische Identität verändern. Die Voice Skins sind bereits in einigen industriell verbreiteten Games integriert. Auch sie entstanden mit neuronalen Netzen auf der Grundlage menschlicher Sprachlaute. Die Echtzeit-Funktion ermöglicht, dass Sprechweise, Ausdruck und Emotion der Gamer*innen auch in ihrer Stimmverkleidung erhalten bleiben.

Das Phänomen Voice Skin wirft interessante Fragen auf: Warum richten wir uns morgens wie selbstverständlich Haar, Gesicht und Zähne, aber nicht die Stimme? Liegt es daran, dass uns im Alltag ein akustisches Pendant zum Spiegel fehlt? Könnte KI-Voice-Technologie so etwas ermöglichen? Alexa & Co. können bereits, in eingeschränkter Wirkung, anhand unseres Sprachklangs Emotionen oder auch Gesundheitszustand erkennen. Ist es uns egal, ob den Voice Assistenten unsere Stimme gefällt, und darf uns das überhaupt egal sein?

Einen Selbstversuch mit Voice-Services im Internet unternahm ich 2019/20 mit dem Kunstprojekt *The Voice of Wednesday Dupont*. Auf der Webseite des Anbieters lyrebird.ai, der mittlerweile unter Descript firmiert, trainierte ich eine KI mit Sprachsamples meiner eigenen Stimme und erhielt eine kostenlose, digitale LoFi-Version davon. Text-to-Speech-Eingaben im Browser wandelt lyrebird.ai in gesprochene Sätze um (am besten gelingt das auf Englisch), in denen bei aller Plastikhaftigkeit das Timbre meiner wirklichen Stimme zumindest anklingt. Ich fühle mich beim Hören von mir selbst entfremdet und zugleich irgendwo in den digitalen Tiefen repräsentiert, verzerrt gespiegelt. Gegen eine stattliche Summe verspricht lyrebird.ai/Descript eine realistischere KI-Version der Stimme der Kund*in zu generieren.

Fake & Wirklichkeit

Im April rezitierten Slavoj Žižek und Ayn Rand den 90er-Gassenhauer »Barbie Girl« von Aqua auf YouTube. Das mutet nicht nur deshalb seltsam an, weil Rand, gestorben 1982, Aqua nie gehört hat, sondern weil insbesondere Žižeks Stimme mitsamt slowenischem Akzent dem Original sehr nahe klingt und der Inszenierung damit eine feierliche Ernsthaftigkeit verleiht. Hinter dem Klamauk steckt der YouTube-Channel Vocal Synthesis, der damit öffentlichkeitswirksam den State of the Art von Googles KI-Voice-Technologie Tacotron 2 präsentiert und zugleich auf die Problematik akustischer Deepfakes sensibilisiert. Es ist schon gar nicht mehr so komplex, mit aufgezeichnetem Sprachmaterial eine täuschend echte KI-Kopie einer Stimme zu generieren.

Bild-, sondern auch Tonmaterial bei der Auswertung und vor der Ausstrahlung zu überprüfen. Eine Hilfe könnten akustische Wasserzeichen sein, die mittels spezieller Decoding-Algorithmen von Empfängerseite über den Echtheitsgehalt von Audio-, insbesondere von Sprach-Files Auskunft geben.

Anders als in vielen großen und kleinen Tonstudios sollte diese weibliche Stimme eben nicht nach den Idealvorstellungen von Männern geformt werden.

Noch hat diese Technologie ihre Schwächen: Sie funktioniert nicht in Echtzeit und tut sich schwer, Emotionen klanglich zu vermitteln.

Doch zeigt schon ein weiteres Beispiel die Wirkmacht der aktuellen Voice-Technologie: Auch Jay-Z hatte im Frühjahr einen Deepfake-Auftritt auf YouTube. »Er« rappt Meme-Texte und Passagen aus dem Buch *Genesis*. Erstaunlich hierbei ist, wie gut die KI Rhythmus und Betonungen seiner Rap-Stimme nachbilden kann. Nur Jay-Z selbst war davon nicht angetan und reagierte mit einer Klage. Doch wogegen eigentlich? Man kann nicht wirklich davon sprechen, dass er hier unerlaubt gesampelt wurde. Keine Millisekunde der Videos ist einem seiner Songs entnommen worden. Ist es geistiger Diebstahl, die Klangfarbe und Sprech-/Sing-/Rap-Weise eines Menschen zu appropriieren?

Deepfake ist natürlich ein ernstes Thema. Die Bilder komplett künstlich erzeugter Gesichter gingen um die Welt. Dass nicht nur Photoshop, sondern auch KI-Technologie Bilder fälschen kann, sickert allmählich ins allgemeine Bewusstsein. Audio hinkt da aber bislang hinterher. Aufklärungsarbeit wäre hier bitter nötig, denn die Technologie wartet nicht, ob morgen eine Trump-Ansprache über das Internet verbreitet wird, die ihm nie über die Lippen ging. Hier sind auch Nachrichtendienste und Medien in der Pflicht, nicht nur

Ob Deepfake oder »genuine« KI-Stimme: Die Echtheit einer Stimme ist 2020 nur noch schwer ohne technische Hilfe zu überprüfen. In der alltäglichen Voice-Technologie scheint die Maxime zu herrschen, die virtuellen Kommunikationspartner*innen so menschlich wie möglich klingen und wirken zu lassen. Sei es, um sich an der Simulation von Wirklichkeit zu erfreuen, sei es, um damit mehr Vertrauen herzustellen, als die User*innen (vermeintlich) einer blechernen Robo-Voice entgegenbringen würden.

In der Kunst scheint man diese Ideale nicht zu verfolgen. Das mag an einem Bewusstsein dafür liegen, dass bereits einer trainierten Singstimme inhärente Künstlichkeit innewohnt. Belcanto, Kastratengesang, Growling, aber auch die Sprechweise von Schauspieler*innen in Theater, Film und Hörspiel sind performte Acts, einstudierte Spezialformen des Gebrauchs menschlicher Stimme.

Eine Gesangsstimme ist nicht zwingend »echter« als eine singende KI, innerhalb derer Algorithmen es wiederum menschelt.

Singende KI-Stimmen wie Oracle von Amnesia Scanner oder Holly Herndons Spawn kokettieren mit einem Hybrid-Dasein aus Maschine und Mensch. Auch die Stimme, die Mouse on Mars für ihr kommendes Album entwickelt haben, zielt nicht darauf ab, als besonders menschlich wahrgenommen zu werden. Jan St. Werner und sein Duo-

Partner Andi Toma interessieren sich dafür, was in abstrakten synthetischen Sounds Reminiszenzen an menschliche Stimmklänge hervorruft: »Es gibt gewisse spektrale Bewegungen innerhalb von Klängen, bestimmte Resonanzen und Filterungen, die dafür sorgen, dass diese Klänge eine Vertrautheit im Menschen wecken, auf die wir stark und instinktiv anspringen«, so Werner. Das Interesse am Humanen im künstlichen Klang führte Mouse on Mars dazu, mit einem Entwickler*innen-Team (Rany Keddo, Derek Tingle, Birds on Mars) eine eigene synthetische KI-Stimme zu gestalten. Diese wurde mit der Stimme des Autoren und Forschers Louis Chude-Sokei trainiert und ähnelt dementsprechend seinem Timbre. Eine weitere Version der Stimme entstand aus Sprachmaterial der Mit-Entwicklerin Yagmur Uckunkaya – auch der Mensch, der die Maschine technisch mitgestaltet hat, sollte Teil davon werden. Die Stimme kann mittels Texteingabe verständliche Sätze äußern, aber auch singende Bewegungen (frei von tonalen Systemen) und abstrakte, lautmalerische Klänge hervorbringen. In der Musik von Mouse on Mars gliedert sie sich als eine Klangfarbe von vielen ein. Jan St. Werner beschreibt die Stimme als eine Art Synthesizer, der mit Parametern wie Pitch, Geschwindigkeit oder Samplingrate aufwartet, und somit als Instrument eingesetzt werden kann – als ein recht widerspenstiges allerdings. »Wir wollten eine anarchistische KI produzieren. Sie soll ablenkbar sein, unbelehrbar, sich wehren, ihre eigenen Wege gehen und uns überraschen.«

Bei den Trainingseinheiten las Louis Chude-Sokei aus einem eigens verfassten Sonic Fiction Essay, der das Miteinander von Menschen und Maschinen mit postkolonialen Visionen verknüpft. »In diesem Resonanzraum schwingt die ganze Platte«, sagt Werner. »Wenn sich Chude-Sokei mit Kreolismus und den dadurch entstehenden neuen Kulturformen auseinandersetzt, übertragen wir das auf unsere AI. Auch hier gibt es eine kulturelle Assemblage, entsteht Neues. Was ist Maschinenintelligenz? Was ist menschliche Kultur? Wo überlagert sich das und wie lernt das eine vom anderen? Es ist ja nicht so, dass ausschließlich die AI von uns lernt, auch wir lernen von ihr. Wenn wir mit diesem ›new ethnic becoming‹ umgehen lernen, entstehen in uns auch andere Sensibilisierungen dem Leben gegenüber. Wenn wir künstliche Intelligenz nicht als etwas verstehen, von dem wir lernen und mit dem wir auch ethisch umgehen müssen, wie sollen wir das dann mit dem Menschen tun?«

Zweifel & Angst

Die Furcht vor kommunizierenden Maschinen ist nichts Neues: Um das Jahr 1000 soll Gerbert von Aurillac, auch bekannt als Papst Silvester II., einen bronzenen Kopf entwickelt haben, der mit »Ja« und »Nein« auf politische und religiöse Fragen antworten konnte. Zeit seines Lebens rankten sich Gerüchte um Aurillac; angeblich habe er sich mit dem Teufel verbündet. Egal, ob es den Kopf wirklich gegeben hat – interessant ist, wie man eine künstliche, in eine Maschine integrierte Stimme schon damals in die Nähe des Okkulten rückte. Künstliche Intelligenz wird oft mit Unbehagen wahrgenommen, insbesondere im Voice-Bereich. Laut der Beyto-Studie empfindet die Hälfte der Nutzer*innen von Smart Speakers die Privatsphäre hierbei als größeres Problem denn in anderen technischen Bereichen.

Der Computermusiker Sam Kidel hat mit *Voice Recognition DoS-Attack* 2018 eine Art akustischen Schutzschild gegen lauschende Software entwi-

ckelt. Die künstlich erzeugten Phoneme von Kidels Software-Patch verwirren gängige Smart-Lautsprecher so sehr, dass sie nicht mehr entschlüsseln können, was echte Menschen parallel dazu im Raum sprechen.

Doch muss man denn Angst vor Sprach-KI haben? Laura Dreessen findet: »Es sind nicht Amazon oder Google – ich bin immer noch selbst dafür verantwortlich, meine Privatsphäre zu definieren. Ich kann immer entscheiden, ob mein Gerät auf stumm geschaltet ist, in welchem Raum es steht, welcher Konversation es beiwohnt, wie ich mich damit ausdrücke und in welchem Kontext. Man muss nicht alle technischen Details verstehen, aber das Bewusstsein darüber, dass diese Technologie nur von uns lernt und auf uns reagieren kann, dass sie sich nicht einfach so verselbstständigt, das muss man in der Gesellschaft noch schärfen.«

Jan St. Werner geht sogar noch weiter und mit offenen Armen auf die sprechenden, hörenden und lernenden Maschinen zu, denen der Mensch selbst durch seinen Input zur Intelligenz verhilft: »Über das Geben verändert man sich. Identität ist nichts, was man allein mit sich herumträgt und was einen schwerer macht. Mit der KI erscheinen andere Identitätskonstruktionen heute viel plausibler.« Vielleicht unterscheidet ein Kind, das im Dialog mit seinen Eltern zu sprechen lernt, gar nicht so viel von einer Maschine, der wir dasselbe lehren. Letztlich ist auch die Sprache, die jede*r von uns erlernt hat, nichts anderes, als ein über Jahrtausende entwickelter und verfeinerter Algorithmus, in dem das Wissen voriger Generationen wie unserer direkten Bezugspersonen steckt und an uns weitergegeben wird, die wir es wiederum mutieren lassen. Wo der genaue Sweet Spot zwischen umarmenden Baby-Analogien und vernünftiger Vorsicht liegen sollte (ganz sicher ist er kontextabhängig), darüber wird noch viel gesprochen werden. Idealerweise in synthetischen und ganz normalsterblichen Stimmen.

1. Die folgenden Zitate von Laura Dreessen und Jan St. Werner basieren auf Interviews, die der Autor mit ihnen führte.
2. Die Studie lässt sich kostenlos hier anfordern: https://www.beyto.com/smart-speaker-studie-2020/

Friedemann Dupelius (manchmal auch: Wednesday Dupont oder Friday Dunard) arbeitet mit Sprache und Sound und lässt das jeweilige Topic of Interest über die Form entscheiden: etwa Radiofeature, Essay, Interview, Audio Paper, Hörspiel oder Track.