KULTUR NEU ENTDECKEN

SWR2

SWR2 Wissen

Konversation mit KI

Wie Maschinen mit uns reden

Von Christoph Drösser

Sendung: Montag, 30. November 2020, 8.30 Uhr
Redaktion: Sonja Striegl
Regie: Christoph Drösser
Produktion: SWR 2020

Wir werden in Zukunft zweimal hinhören müssen, um herauszufinden, ob die angenehm menschliche Stimme, die mit uns spricht, nicht in Wirklichkeit von einer Maschine stammt.

MANUSKRIPT

Computerstimme:
Ich bin eine Computerstimme. Das können Sie sicher hören. Ich klinge mechanisch, betone die Wörter nicht richtig. Keine Sorge, Sie müssen mir nicht eine halbe Stunde lang zuhören.

Musikakzent

Computerstimme:
Meine Stimme klingt schon erheblich angenehmer oder? Trotzdem haben Sie sicherlich schnell gemerkt, dass auch ich von einem Computer erzeugt worden bin. Aber wir werden menschlichen Stimmen immer ähnlicher, und Sie werden im Verlauf der Sendung noch bessere Beispiele zu hören bekommen. Stimmen, die Atempausen einlegen und “äh” oder “hmm” sagen wie ein Mensch. Sie werden staunen.

Ansage:
Konversation mit KI - Wie Maschinen mit uns reden. Von Christoph Drösser.

Sprecher:
Diese Ansage kam von einem Menschen aus Fleisch und Blut. Von der professionellen Sprecherin Katrin von Chamier. Das gerade war die neutrale Variante. Sie kann aber auch anders – wie wär’s mit einer aufgekratzten Version?

Katrin von Chamier:
Konversation mit KI – Wie Maschinen mit uns reden.

Sprecher:
Oder cool und erotisch?

Katrin von Chamier:
Konversation mit KI – Wie Maschinen mit uns reden.

Sprecher:
Sie kann aber auch betont sachlich:

Katrin von Chamier:
Konversation mit KI – Wie Maschinen mit uns reden.

Sprecher:
Die menschliche Stimme vermittelt viel mehr als den Text des Gesprochenen. Wir schließen aus Tonfall und Stimmmelodie auf die Gefühle der Sprecherin. Katrin von Chamier verfügt über eine sehr wandelbare Stimme, sie spricht Werbetexte, aber auch sachliche Informationen ein.

Collage:
Werbespots mit von Chamiers Stimme

Katrin von Chamier:
Es hat immer mir wahnsinnig Freude bereitet, mit all den Nuancen zu spielen. Stimme, gerade im Radio, am Mikro, ist ja immer so eine Form von Intimität. Und wenn es intim wird, dann zählt ja jede Nuance. Das kann ich glaube ich gut.

Sprecher:
Mit diesen Nuancen, für die jeder und jede von uns ein sehr feines Ohr hat, haben maschinelle Stimmen noch ihre Probleme. Aber wir treffen immer öfter auf sie – im Navigationssystem, als digitale Assistentinnen im Handy oder in einem der sogenannten Smart Speaker wie Amazons Alexa, zunehmend auch als "Kundendienstmitarbeiter" im Internet. Auch Katrin von Chamiers Stimme gibt es inzwischen in einer Maschinenversion. Und auch meine eigene Stimme habe ich für diese Sendung klonen lassen.

Aber zuerst reisen wir einmal 240 Jahre in die Vergangenheit, um die Ursprünge der maschinellen Sprache zu ergründen.

Musikakzent

Atmo:
Quakende Laute von Wolfgang von Kempelens Sprechmaschine (www.youtube.com/watch?v=k_YUB_S6Gpo)

Sprecher:
Im Jahr 1780 präsentierte der Ingenieur Wolfgang von Kempelen dem staunenden Publikum seine "Sprechmaschine". Von Kempelen ist vor allem für seinen "Schachtürken" bekannt – eine Schach spielende Maschine, die aber in Wirklichkeit von einem Menschen bedient wurde, der darin hockte. Die Sprechmaschine dagegen war kein Taschenspielertrick. Sie war die erste Maschine, mit der versucht wurde, den menschlichen Stimmapparat physikalisch nachzubauen. Ein Blasebalg war die Lunge, eine Zungenpfeife bildete die Stimmlippen nach, und ein großer Trichter entsprach dem Mundraum. Eine Zunge und Zähne besaß die Maschine nicht, deshalb konnte sie nur einige Vokale und Konsonanten wirklich überzeugend produzieren.

Das Tonbeispiel eben stammt von einer Replika der Maschine, die Fabian Brackhane vom Leibniz-Institut für deutsche Sprache gebaut hat. Haben Sie verstanden, was die Maschine "gesagt" hat? Wahrscheinlich nicht. Wenn von Kempelen sie vorstellte, verriet er seinen Zuhörern immer, welche Wörter sie gleich hören würden. Dann verstand man sie besser. Also: Hören Sie mal auf die Wörter "Mama", "Papa", "Oma" und "Opa".

Sprechmaschine:
Mama, Papa, Oma, Opa.

Sprecher:
So genial und unterhaltsam die Maschine auch war, der anatomisch korrekte Nachbau des menschlichen Sprechapparats erwies sich als Sackgasse. Erst als man Klänge elektronisch erzeugen konnte, weckte auch die menschliche Stimme wieder das Interesse der Ingenieure. Auf der Weltausstellung 1939 wurde “Voder” vorgestellt, eine Entwicklung des amerikanischen Ingenieurs Homer Dudley. Die Frauen, die Voder bedienten – und es waren nur Frauen – mussten ein Jahr üben, um den Synthesizer verständliche Sätze sagen lassen zu können.

Atmo:
Präsentation Voder (www.youtube.com/watch?v=5hyI_dM5cGo):
Mann: For example, Helen, will you let the Voder say “she saw me”?
Voder: She saw me.

Specher:
Und jetzt das Ganze mit unterschiedlichen Betonungen.

Atmo:
Präsentation Voder (www.youtube.com/watch?v=5hyI_dM5cGo):
Mann: That sounded awfully flat. What about a little expression? Say the sentence in answer to these questions. Who saw you?
Voder: *She* saw me …

Sprecher:
Das klingt schon fast modern, so wie die verfremdeten Stimmen in elektronischer Popmusik. Und tatsächlich konnte diese Synthesizerstimme auch singen.

Atmo:
Präsentation Voder (www.youtube.com/watch?v=5hyI_dM5cGo):
Singt

Atmo:
Raumpatrouille: 20, 19, 18, 17…

Sprecher:
Offenbar konnte man sich allerlei verrückte Zukunftstechnik vorstellen, aber keine Maschine, die nicht maschinell klingt. Das änderte sich mit Stanley Kubricks Film “2001 – Odyssee im Weltraum”. Der Computer HAL, der gegen die Besatzung des Raumschiffs rebelliert, ist nicht nur sehr intelligent, er spricht auch wie ein Mensch.

Atmo:
Film-Ausschnitt 2001: “2001 – Odyssee im Weltraum”
Astronaut Bowman: Hallo HAL, hörst du mich? Hörst du mich, HAL?
HAL: Jawohl, Dave, ich höre dich.
Bowman: Öffne das Gondel-Schleusentor.
HAL: Es tut mir leid, Dave – aber das kann ich nicht tun.

Sprecher:
Wenn die Computerstimmen, hier natürlich von menschlichen Schauspielern eingesprochen, so natürlich sein konnten, dann war es nur eine Frage der Zeit, bis es zu erotischen Verwicklungen zwischen Mensch und Maschine kam. Der Film “Her” von 2013 machte das zum Thema, und die Stimme von Scarlett Johansson verdrehte nicht nur dem männlichen Hauptdarsteller den Kopf.

Atmo:
Film-Ausschnitt Her:
I love you so much. And this is who I am now.
I need you to let me go.
As much as I want to. I can't live in your book anymore.

Sprecher:
Von so viel Sinnlichkeit sind die Computerstimmen von heute noch weit entfernt. Aber sie sind in unserem täglichen Leben zu Gesprächspartnerinnen und -partnern geworden. Es ist ganz normal, dass wir mit unserem Handy reden, ohne dass ein Mensch am anderen Ende der Leitung sitzt. Die Entwicklung begann, als Apple 2011 Siri vorstellte, die Sprachassistentin für das iPhone.

Atmo:
Präsentation Siri: Phil Schiller: For decades, technologists have teased us with this dream, that you're gonna be able to talk to technology, you know, do things for us. Haven’t we seen this before, over and over? But it never comes true. We have very limited capability. What we really want to do is just talk to our device. ask a simple question. What's the weather going to be like today? And get a response.

Sprecher paraphrasiert darüber:
Seit Jahrzehnten träumen die Technik-Apostel davon, dass wir mit unseren Geräten sprechen, sagt der Apple-Manager Phil Schiller. Aber bisher sei das Versprechen nie eingelöst worden. Wir wollen ganz einfach mit ihnen reden und zum Beispiel fragen: Wie wird das Wetter heute?

Dann betrat der Manager Scott Forstall die Bühne und präsentierte live die neue Sprachassistentin. Und fragte tatsächlich nach dem Wetter.

Atmo:
Präsentation Siri:
Scott Forstall: What is the weather like today?
Siri: Here's the forecast for today.
Forstall: It is that easy. (Begeisteter Applaus)

Sprecher:
Wir können uns vielleicht gar nicht mehr vorstellen, wie revolutionär das damals war: mit dem Handy in ganz normaler Sprache zu reden und nicht nur ein paar simple

Kommandos zu geben. Siri versteht in gewissem Umfang, was wir wollen, auch wenn wir nicht direkt nach dem Wetter fragen. Siri, muss ich heute lange Unterwäsche anziehen?

Siri:
Im Moment ist es wolkenlos und 24 Grad. Die Temperatur sinkt heute abend auf 14 Grad.

Sprecher:
Das ist natürlich keine Konversation, aber immerhin eine intelligente Antwort auf eine gestellte Frage. Wie erstellt man eine solche Stimme, die jeden beliebigen Text sprechen kann und nicht nur auf Textbausteine zurückgreift? Naiv stellt man sich vielleicht vor, dass ein menschlicher Sprecher oder eine Sprecherin einfach alle Laute der deutschen oder englischen Sprache einmal aufnimmt, und die setzt ein Computerprogramm zu beliebigen Wörtern zusammen.

Aber so einfach ist das nicht. Die Laute unserer Sprache beeinflussen einander, und die Puzzlestücke für eine gute Sprachsynthese sind nicht die einzelnen Phone, wie es linguistisch heißt, sondern Diphone – Paare von Lauten. Und die muss es dann noch in vielen Varianten geben, wenn man eine abwechslungsreiche Stimme bauen will, die Sätze in unterschiedlicher Weise betonen kann. Dass uns falsche Betonungen irritieren oder belustigen, dafür sind die Ansagen in Bahnhöfen oder Straßenbahnen ein gutes Beispiel. Hier werden Sätze aus ganzen Wörtern und Zahlen zusammengesetzt, die eine menschliche Stimme vorher aufgenommenen hat. Trotzdem stimmen ab und zu die Betonungen nicht und der Sprachfluss wirkt manchmal abgehackt.

Atmo:
Autofahrt mit Stimme vom VW-Navigationssystem

Sprecher:
Katrin von Chamier wurde vor ein paar Jahren zu einem Casting für ein großes Unternehmen eingeladen, das eine eigene künstliche Stimme produzieren wollte. Erst später erfuhr sie, dass daraus die Stimme gebaut werden sollte, die im Navigationssystem jedes Autos des VW-Konzerns steckt.

O-Ton Katrin von Chamier:
Als dann diese konkrete Anfrage kam, habe ich im allerersten Moment gesagt: Nee, das mach ich nicht. Das kann ich doch nicht bringen, weil ich das eben auch so als Selbstunterwanderung meines Berufs gesehen hab. Aber dann konnte ich das natürlich relativ schnell übersehen, was das eigentlich ist. Und dann hab ich erfahren, dass zehn Stimmen ungefähr am Schluss zur Auswahl standen, und es gab eine CD, auf der dann wohl verschiedene Textausschnitte von all den Stimmen waren. Und dann mussten die Vorstandsmitglieder der damaligen Zeit sich das eine Woche anhören. Und ich hab deshalb gewonnen, weil ich nach einer Woche immer noch nicht genervt hab.

Sprecher:

Noch heute muss sie immer wieder in ihrem Studio Texte für das Navi einlesen, die Entwicklung geht schließlich weiter.

O-Ton Katrin von Chamier:
Das Entscheidende ist, dass ich immer dasselbe Setup habe, dass sich hier nichts verändert, dass ich wirklich dann die vier, fünf Stunden am Tag mit dem richtigen Mikrofon Abstand, mit der richtigen Dynamik und mit der richtigen Spannung, aber trotzdem mit viel Spielraum innerlich das Ganze einlese und mich an die Regeln halte, die dann da auch noch gelten.

Sprecher:
Für ihre Kinder war es zunächst ungewohnt, dass die Stimme der Mutter aus dem Navi kam. Auch deshalb, weil das nicht so wirklich zu ihr passte.

O-Ton Katrin von Chamier:
Es ist halt absurd, weil ich bin wirklich in meiner Familie und in meinem Freundeskreis derjenige mit dem miesesten Orientierungssinn. Niemand würde auf die Idee kommen, mich nach dem Weg zu fragen. Und dann bin ich diejenige, die das tut. Das ist total absurd. Selbst für mich.

Atmo:
Sprachbefehlt Katrin von Chamier

Sprecher:
Die Konstruktion einer hochwertigen Stimme wie der ersten Generation von Siri oder auch der VW-Navi-Stimme war viel Handarbeit. Die Ingenieure müssen Texte generieren, die möglichst viele unterschiedliche Diphone enthalten, dann müssen sie all diese Aufnahmen in ihre Bestandteile zerlegen. Bei der Sprachsynthese schließlich werden sie mit einem komplexen Algorithmus wieder zusammengesetzt.

Atmo:
Sprachbefehlt Katrin von Chamier

Sprecher:
Im Jahr 2016 folgte die nächste technische Entwicklung: Ingenieure der zu Google gehörenden Firma DeepMind entwickelten die sogenannte Wavenet-Technologie. Die basiert, wie inzwischen alle modernen Sprachtechnologien, auf Künstlicher Intelligenz, dem sogenannten Deep Learning. Im Prinzip lernt der Computer dabei sprechen, indem er sich zunächst ganz viele Beispiele von Stimmen anhört und sie mit einem Transkript des Gesprochenen vergleicht. Die “Grundstimme”, die dabei entsteht, kann relativ schnell an alle möglichen Sprecher angepasst werden. Wie natürlich die damit erzeugten Stimmen klingen, konnte das Publikum 2018 erleben, als Google ein System namens Duplex vorstellte. Das ist gedacht als eine Art persönlicher Butler. Dem ruft man dann nur noch zu: “Google, mach einen Friseurtermin für mich” – und der Assistent ruft selbständig im Friseursalon an. Hören Sie selbst – die erste Stimme ist die menschliche Mitarbeiterin eines Friseursalons, die zweite ist Google Duplex.

Atmo:
Präsentation Google Duplex:
Friseurin: Hello, how can I help you?
Google Duplex: Hi, I'm calling to book a women's haircut for a client. I'm looking for something on May 3rd.
F: Sure, give me one second
GD: Mm-hmm

Sprecher:
An dieser Stelle lachen die Zuhörer. "Mm-hmm" hat die künstliche Stimme gesagt – wie ein Mensch, der im Gespräch Zustimmung oder Verständnis signalisiert. Solche Füllwörter kommen in diesem Dialog mehrmals vor. Und die Frau am anderen Ende merkt zu keinem Zeitpunkt, dass sie mit einer Maschine spricht.

Atmo:
Präsentation Google Duplex:
Friseurin: Ok, perfect. So I will see Lisa at 10 o'clock on May 3rd
GD: OK, great thanks.
F: Great. Have a great day. Bye.

Sprecher:
Diese neue Generation von Stimmen klingt nicht nur natürlicher, sondern es erfordert auch viel weniger Aufwand, eine neue Stimme zu bauen. Das bieten inzwischen nicht nur die großen Firmen wie Google und Amazon an, sondern auch kleine Startups.

Ich beschloss, den Test zu machen, und fand eine im kanadischen Toronto beheimatete Firma namens Resemble.ai. Die hat sich auf das "voice cloning" spezialisiert, die Erstellung von künstlichen Stimmen, die einem bestimmten Menschen möglichst ähnlich sein sollen. Der Firmengründer, Zohaib Ahmed, erklärt, wie man mit sogenanntem Transfer Learning auch mit einer Viertelstunde beliebiger Tonaufnahmen eine neue Stimme erstellen kann.

O-Ton Zohaib Ahmed:
So there are a lot of papers out there … and it's then putting that on top of the pretrained model.

Übersetzung:
Es gibt zum Beispiel Techniken, da nimmt man das Foto eines Zebras und macht daraus ein Pferd. Das nennt man Transfer-Lernen. Die Idee ist, dass man ein Basismodell hat, das man – in unserem Fall – mit Hunderten von Stunden von Aufnahmen mit vielen Sprechern trainiert hat. Es hat bereits Deutsch mit verschiedenen Akzenten und verschiedenen Stimmen verstanden. Für einen neuen Sprecher muss es nicht alles noch einmal lernen. Sie geben uns 10 Minuten Daten, und wir müssen nur noch Ihr Timbre, Ihren Sprachstil auf das vortrainierte Modell übertragen.

Sprecher:
Welche Eigenheiten sich das System herauspickt, ist nicht steuerbar – der Lernprozess läuft völlig ungesteuert ab, manchmal mit überraschenden Ergebnissen:

O-Ton Zohaib Ahmed:
At the beginning of every sentence … few words. So that's very natural.

Übersetzung:
Wenn Sie am Anfang jedes Satzes auf Holz klopfen, dann denkt die Maschine, das ist ihre persönliche Art zu sprechen, und sie wird versuchen, genau dasselbe zu tun. Sie weiß, dass ein Maschen nicht 20 Sekunden sprechen kann, ohne zu atmen. Und so kommt es vor, dass die KI manchmal kurz vor dem Ende eines Satzes eine Pause einlegt, einmal tief einatmet und dann die letzten Worte sagt. Das ist alles sehr natürlich.

Sprecher:
Und nun wollen Sie sicher das Ergebnis hören. Also gut – hier ist meine synthetische Computerstimme. Ich kann über eine Tastatur einen beliebigen Text eingeben und dann noch ein paar Parameter wie Tempo und Betonung einstellen.

Synthetische Sprecherstimme:
Hallo, ich bin die synthetische Stimme von Christoph Drösser und übernehme hier mal kurz. Wozu können diese Stimmen angewendet werden? Zohaib Ahmed sagt, sie sollen im Moment noch nicht unbedingt menschliche Sprecher ersetzen. Er gibt das Beispiel von computergenerierten Animationsfilmen, bei denen die Bildebene unendlich variiert werden kann, aber der meist hochbezahlte Schauspieler, der einer Figur die Stimme leiht, nur einmal kurz ins Tonstudio kommt. Mit einer synthetischen Stimme kann man das vorher schon alles simulieren.

O-Ton Zohaib Ahmed:
And when Kevin Hart shows up, … his character and his style.

Übersetzung:
Und wenn dann Kevin Hart ins Studio kommt, sagt man ihm: "Hier ist die synthetische Version deiner Stimme, sie spricht den Satz so, wie wir ihn uns vorgestellt haben. Kannst du das vielleicht noch besser?" Kevin Hart ist wahrscheinlich immer noch der synthetischen Version überlegen, aber man hat vorher schon viele Versionen ausprobieren können, um zu sehen, ob es zu der Figur passt und seinem Stil entspricht.

Sprecher:
So, hier ist wieder meine menschliche Stimme. Ahmed demonstriert noch weitere Anwendungen für die künstliche Stimme. Ein Franzose kann kein Spanisch, will aber etwas auf Spanisch sagen? Kein Problem – seine sprachlichen Eigenheiten werden einfach auf eine spanische Stimme projiziert.

O-Ton Zohaib Ahmed:
Now, this is still very early work, it requires … how many pickled peppers did Peter Piper pick?)

Übersetzung:
Das hier sind noch vorläufige Arbeiten, das ist erheblich Daten intensiver als unsere typischen Anwendungen. Aber die Idee ist, dass man eine Stimme nehmen kann, die Französisch spricht …
(Französischer Satz)
Diese Person hat vielleicht noch nie Spanisch gesprochen, aber wir können sie Spanisch sprechen lassen.
(Spanischer Satz)
Und auch Englisch.
(Peter Piper picked a peck of pickled peppers, Peter Piper picked a peck of pickled peppers, how many pickled peppers did Peter Piper pick?)

Sprecher:
Zohaib Ahmed führt bewusst Beispiele für Anwendungen an, bei denen die künstliche Stimme der menschlichen überlegen ist oder die für Menschen schlicht unmöglich sind. Aber die Systeme könnten auch missbraucht werden. Kann man zum Beispiel anderen Menschen die Stimme klauen und ihnen Worte in den Mund legen? Firmen wie Resemble legen Wert darauf, dass sie Stimmen nur klonen, wenn der Eigentümer ausdrücklich einwilligt. Aber was eine kleine Firma mit zehn Mitarbeitern kann, das können auch Bösewichte mit dunkleren Absichten.

O-Ton Zohaib Ahmed:
Now, the thing I'm more concerned … you can tell what's real, what's fake.

Übersetzung:
Was mir mehr Sorgen macht: Wenn wir das schaffen, dann könnten das auch andere Leute tun. Als die Fotobearbeitungssoftware Photoshop herauskam, entstanden ja auch Hunderte von Klonen. Wir haben aber auch eine Open-Source-Software namens Resemblizer entwickelt, mit der man gefälschte Sprache erkennen kann. Die erstellt einen Fingerabdruck oder Stimmabdruck von einer Audiodatei. Und kann damit echt von gefälscht unterscheiden.

Sprecher:
Wie in vielen Bereichen wird hier KI eingesetzt, um Delikten auf die Spur zu kommen, die nur durch KI möglich sind. Trotzdem ist zu erwarten, dass wir in Zukunft viele Beispiele von geklonten Prominenten in Bild und Ton sehen werden, denen man Worte in den Mund gelegt hat, die sie nie gesagt haben.

Soundcollage:
Fake Obama, Trump, Queen Elizabeth

Sprecher:

Denkbar ist es auch, dass wir die Stimme eines geliebten Menschen zu seinen Lebzeiten sampeln – und nach dem Tod können wir ihn oder sie akustisch wieder zum Leben erwecken.

Atmo:
Filmausschnit aus “Black Mirror”

Sprecher:
In der britischen Science-Fiction-Serie “Black Mirror” wurde dieser Gedanke schon einmal zuende gedacht, komplett mit einem humanoiden Roboter, der dem Verstorbenen aufs Haar gleicht.

Je ähnlicher die Computerstimmen der unseren werden, umso größer wird die Gefahr, dass wir sie mit echten Menschen verwechseln. Beim Navi im Auto wissen wir, mit wem wir es zu tun haben. Wenn wir beim Anruf in einem großen Unternehmen in einem Callcenter landen, könnte es bereits sein, dass uns eine Kunststimme begrüßt. Das ist vielen Menschen unheimlich. Google hatte das unterschätzt, als es 2018 seine Software “Google Duplex” vorstellte. Viele Menschen versetzten sich in die Rolle der Friseurin am anderen Ende der Leitung, die nicht wusste, dass sie mit einer Maschine redete, und auch nicht darüber aufgeklärt wurde. Die Firma versprach hoch und heilig, dass sich die künstliche Stimme in Zukunft bei allen Anwendungen als Chatbot identifizieren werde.

Trotzdem aber stellt sich die Frage: Warum muss die digitale Assistentin überhaupt täuschend echt wie ein Mensch klingen? Oder der Pflegeroboter, der einen alten Menschen daran erinnert, die Medikamente einzunehmen? Es gibt Forscher, die sich Gedanken machen darüber, wie Maschinenin Zukunft mit uns reden werden. Zu ihnen gehört Leigh Clark von der britischen Swansea University.

O-Ton Leigh Clark:
We've been talking about quite … to be using this language.

Übersetzung:
Wir diskutieren viel über Mimikry. Wollen wir Menschenähnlichkeit um ihrer selbst willen, als ultimatives Design-Ziel? Wenn man das bis zum Ende treibt, dann grenzt das schon an Täuschung. Ein Weg, das zu umgehen, ist, dass die KI sich immer vorstellt: “Ich bin der Google Assistant”. Aber wenn er dann sagt: "Ähm, fünf Uhr könnte klappen", dann fragen Sie sich: Warum redet der so? Es gibt keine psycholinguistische Notwendigkeit für diesen Assistenten, so zu sprechen.

Sprecher:
Insbesondere wenn die Fähigkeiten solcher Systeme noch begrenzt sind, erzeugt eine übermäßig menschenähnliche Stimme eine seltsame Dissonanz, die die menschlichen Gesprächspartner irritiert oder sogar verärgert. Das heißt nicht, dass die Stimmen wieder robotermäßig klingen müssen.

O-Ton Leigh Clark:
I guess we're reaching that point of… dubious route, I suppose.

Übersetzung:
Wir sind an dem Punkt angelangt, an dem die Stimmen klar und verständlich sind. Wie viel weiter müssen wir gehen? Mit diesen “hmms” und den Denkpausen begibt man sich auf einen ethisch fragwürdigen Weg.

Sprecher:
Das sieht auch Alex Waibel so, ein deutscher Pionier in der Computerlinguistik, der am KIT in Karlsruhe und an der Carnegie-Mellon-Universität in den USA forscht. Vielleicht wird sich in Zukunft eine Computer-Sprechweise herausbilden, die nicht jede Marotte der menschlichen Sprache nachahmt, aber trotzdem angenehm klingt.

O-Ton Alexander Waibel:
Ich könnte mir gut vorstellen, dass es irgendwann mal sozusagen von diesem pragmatischen Einsatz solcher Systeme es für Menschen sogar angenehmer ist, wenn sie so eine gewisse, eine gewisse Cuteness haben, die bewusst robotisch ist. Toll sind ja solche Filme wie z.B. die „Minions“, oder ich weiß nicht, ob Sie den Film „Wall-E“ gesehen haben, da haben die Roboter wirklich menschliche Züge und sprechen komplett fließend, aber sie haben dann trotzdem irgendwie gewisse Eigenarten, die menschliche Persönlichkeit reflektieren, aber eindeutig anders sind. Und wenn man das auf diese Weise eindeutig identifiziert als anders oder ein Bot, dann könnte das wahrscheinlich sogar zu besserer Akzeptanz führen, als wenn man versucht, das auf Teufel komm raus so ähnlich wie möglich zu machen.

Atmo:
KFC-Roboterstimme: What could possibly be better than fried chicken? Sign up to the Colonel's club and you will be able to get your hands on free food and drink plus exclusive offers from KFC …

O-Ton Leigh Clark:
This example we do talk about is this … performing role for you this evening.

Übersetzung:
Es gibt zum Beispiel diesen Roboter von Kentucky Fried Chicken, der für ein Drive-in entwickelt wurde. Man gibt seine Bestellung auf, und er reagiert auf eine sehr theatralische Art und Weise. Er hat eine Persönlichkeit, er will gar nicht menschlich erscheinen, aber trotzdem unterhalten. Als würde er sagen: Ich bin heute abend Ihre Bedienung, und ich spiele diese Rolle.

Sprecher:
Auch in den täglichen Sprach-Anwendungen sollten wir uns fragen, ob die Mimikry, also das Nachahmen des Menschen, immer der richtige Weg ist. Das findet auch die Profi-Sprecherin Katrin von Chamier.

O-Ton Katrin von Chamier:
Ich finde, das sind Maschinen und Roboter und sind keine Menschen und keine Mensch-Ersatzpersonen. Und die sollen bitte auch so klingen wie Maschinen klingen und sich auch damit deutlich machen, so dass man auch nicht geneigt ist, eine Kommunikation mit einem vermeintlichen Wesen zu führen.

Sprecher:
Apropos Nachahmen von Menschen: Warum zum Beispiel sind Siri und Alexa, die Sprachassistenten von Google und Amazon, zunächst einmal Frauen gewesen? Entspricht das nicht dem Klischee von der eifrigen Sekretärin, die ihrem – natürlich männlichen – Chef stets zu Diensten ist? Wenn die Maschinen menschenähnlicher werden, dann tragen sie auch sofort den Ballast von menschlichen Vorurteilen und Rollenmustern mit sich herum.

Deshalb gibt es Entwickler, die zum Beispiel versuchen, geschlechtsneutrale Stimmen zu schaffen. Die androgynen Stimmen sind irgendwo zwischen Mann und Frau angesiedelt, wir können sie nicht wirklich einordnen. Die folgende Stimme namens Q wurde von Menschen geschaffen, die sich selbst irgendwo in der Mitte des Gender-Spektrums ansiedeln.

Atmo:
Geschlechtslose Stimme Q: Hi, I'm Q, the world's first genderless voice assistant. … defined by audio researchers …

Sprecher:
Der Nutzer kann per Schieberegler die Frequenz von Q verändern und die Stimme so männlicher oder weiblicher machen. Q ist gedacht als eine Alternative zu den Standardstimmen, mit denen die Sprachassistentinnen in den smarten Lautsprechern von Amazon oder Google ausgeliefert werden.

Musikakzent

Sprecher:
Für die Technikkonzerne ist die Sprache die neue Schnittstelle zwischen Mensch und Computer. Die elektronischen Stimmen werden immer besser, wir beginnen bereits, sie zu vermenschlichen. Es soll Leute geben, die sprechen mit ihrem Staubsaug-Roboter wie mit einem Haustier: “Das hast Du ja wieder blitzblank geputzt.” Aber bis die Computerstimmen alle Feinheiten und Nuancen haben, zu denen die menschliche Stimme fähig ist, wird noch einige Zeit vergehen. Deshalb hat Katrin von Chamier auch noch keine Angst, durch einen Algorithmus ersetzt zu werden.

O-Ton Katrin von Chamier:
Ich habe aber auch wirklich viele, viele Fähigkeiten, die so eine künstliche Stimme niemals haben wird. Ich habe Kreativität, ich kann überraschen, ich hab Humor. Ich kann um die Ecke denken. Ich kann auch Sarkasmus und Ironie. Das können die alle nicht. Aber ich kann das. Ich kann Emotion. Also wenn jetzt irgendwie ein YouTube-

Tutorial vorgelesen wird von einer künstlichen Stimme, ist mir das relativ egal. Es kann im schlimmsten Falle sein, dass ich wie so ein Luxusgut werde, dass meine Arbeit Luxus ist. Wer sich noch eine echte Stimme leistet – wow.

Sprecher:
Schon heute sind menschliche Schreibkräfte und menschliche Übersetzer Luxus – Computer setzen unsere Stimme in Schrift um, und sie übersetzen inzwischen erstaunlich gut von einer Sprache in die andere. Aber wir werden in Zukunft immer öfter zweimal hinhören müssen, um herauszufinden, ob die angenehm menschliche Stimme, die mit uns spricht, nicht in Wirklichkeit von einer Maschine stammt.

*****