



Bild: Rudolf A. Blaha

Digitale Kinderstube

Sprachassistenten und ihre komplexe Beziehung zum Nutzer

Aus technischer Sicht ist ein Sprachassistent nur eine Software. In der Realität betrachten viele Nutzer Alexa & Co. hingegen als Mitbewohner, Familienmitglied oder gar Partnerersatz. Damit das Verhältnis nicht zerbricht, müssen sich die Entwickler mit Psychologie und menschlichem Miteinander beschäftigen.

Von Nico Jurrán

Dass Alexa, Cortana, Siri und Google Assistant als fast menschlich betrachtet werden, ist umso erstaunlicher, als sie nicht einmal eine körperliche Repräsentation eines Menschen bieten: Sie stecken

in vernetzten Lautsprechern und Mobiltelefonen, in manchen Fällen ertönt ihre Stimme sogar nur aus dem Fernseher oder der Stereoanlage.

Nun ließe sich zu Recht anmerken, dass Menschen offenbar schon immer den Drang haben, Lebloses mit Leben zu füllen. Viele geben etwa ihrem Auto einen Namen oder kümmern sich hingebungsvoll um Tamagotchis. Die Frage ist jedoch, ob die Entwickler der Sprachassistenten dieses Verhalten aktiv unterstützen.

Kein Zweifel besteht daran beim US-Hersteller CogniToys und dessen sprechendem Spielzeugdinosaurier „Stemosaur“, dem der Nutzer einen individuellen Namen geben kann. In starkem Kontrast steht dazu erst einmal Googles Herangehensweise, seine Assistentin schlicht „Google Assistant“ zu nennen. Dennoch gibt es auch hier, ebenso wie bei

Amazons Alexa, eine „quasi-zwischenmenschliche“ Ebene zum Nutzer – beispielsweise wenn sie dem Nutzer schöne Träume wünscht, wenn er sich abends mit „Gute Nacht“ von ihr verabschiedet.

Bezaubernde Jeannie

Amazon hätte die „Vermenschlichung“ seiner Assistentin beim Echo Show vorantreiben können – beispielsweise indem das Display ein passend zur Sprachausgabe animiertes Gesicht anzeigt. Die Entwickler sind diesen Weg bislang jedoch nicht gegangen.

Diese Entscheidung dürfte damit zusammenhängen, dass manche Anwender wahrscheinlich eine Visualisierung ablehnen würden, die nicht der Vorstellung entspricht, die sie sich von der Sprachassistentin gebildet haben. Mit diesem Problem sieht sich auch Hollywood immer

wieder bei der Verfilmung berühmter Romane konfrontiert.

Beim Echo Show bekommt der Nutzer lediglich zusätzlich den Text zur gesprochenen Antwort eingeblendet. Eine weitere Annäherung an den Nutzer ist aber dennoch zu verzeichnen: Auf Liebeserklärungen reagiert Alexa im Echo Show mit Antworten, die rosa Herzchen schmücken. Mehr als recht sachliche Reaktionen à la „Ich fühle mich geschmeichelt“ bekommt der Nutzer aber auch hier nicht.

Dass die deutschen Anwender Alexa trotzdem eine Seele zusprechen, zeigte sich im Sommer 2017, als eine Reihe von Nutzern besorgt meldeten, der Assistentin würde es „nicht gut gehen“. Auslöser war eine Änderung des Tonfalls, wenn die Sprachassistentin etwa auf eine Begrüßung mit „Hallo“ antwortete. Laut der Entwickler war diese Einfärbung nicht gewollt; vielmehr hätten mehrere kleinere Modifikationen an unterschiedlichen Stellen der Sprachsynthese unbeabsichtigt zu diesem Ergebnis geführt. Aufgrund der Rückmeldungen sah sich Amazon gezwungen gegenzusteuern und Alexa wieder zu einem lebensbejahenden Timbre zu verhelfen.

Stimmgewaltig

Deborah Harrison, die Schöpferin von Cortana, verriet der Süddeutschen Zeitung [1], mit ihrem siebenköpfigen Team die Persönlichkeit von Microsofts Sprachassistentin sogar wie eine Romanfigur entworfen zu haben – mit einer Geschichte, Eigenarten, Vorlieben, Meinungen.

Auffällig ist, dass sich praktisch alle Entwickler dazu entschieden haben, ihrem Sprachassistentensystem erst einmal eine weibliche Stimme zu geben – auch wenn man in den USA für den Google Assistant mittlerweile auf eine männliche wechseln kann. Für die weibliche Stimme lassen sich eine Reihe plausibler Gründe anführen – allen voran, dass viele Anwender bei dem Begriff des „digitalen Assistenten“ das klassische Bild der fleißigen Sekretärin im Hinterkopf haben dürften.

Die Entwickler der deutschen Alexa verrieten c't, dass sie sich noch von einer anderen Überlegung leiten ließen. Demnach holt man sich, wenn man den Gedanken der Sprachassistentin als menschliches Wesen konsequent weiterverfolgt, mit dem Assistenten eine zunächst völlig fremde Person ins Haus. Und hier kommt simple Psychologie ins Spiel: Ein Mensch, der sich für eine ihm unbekannt Person

von der Straße entscheiden muss, um diese mit in die eigene Wohnung zu nehmen, wählt eher eine Frau – weil die meisten Menschen (unabhängig vom eigenen Geschlecht) diese als weniger bedrohlich empfinden als einen Mann.

Der erwähnte Dino weicht von diesem Schema ab und spricht mit einer recht knöcheligen männlichen Stimme, die bestenfalls an Yoda aus Star Wars erinnert, für deutsche Nutzer aber teilweise schon etwas gruselig wirkt. Allerdings richtet sich der Dino als Lernspielzeug an amerikanische Kinder, die die Stimme wohl eher mit dem auch recht schräg klingenden Cookie Monster (Krumelmonster) aus der Sesamstraße in Verbindung bringen.

Der von IBM Watson mit künstlicher Intelligenz ausgestattete Dino betont zudem Satzteile wesentlich stärker als andere Sprachassistenten und wirkt damit augenblicklich wie ein Geschichtenerzähler – passend zum jungen Zielpublikum. Der Hersteller kündigte zur Markteinführung an, alternativ eine Jungen- und eine Mädchenstimme anzubieten. Ob dieser Plan umgesetzt wurde, entzieht sich allerdings unserer Kenntnis.

Persönlichkeit

Doch nicht nur bei der Stimmauswahl ähneln sich die Sprachassistentensysteme. Alle geben freundlich Auskunft und arbeiten Aufgaben ohne Murren ab, sind aber nicht devot. Sie passen also wirklich eher ins Bild einer freundlichen Sekretärin als einer bezaubernden Jeannie, die auf alle Befehle mit „Ja, Meister“ antwortet. Passend dazu verwehren sich Alexa & Co. mit einem bestimmten Auftreten gegen Beleidigungen.

Die Reaktionen der Assistentinnen beispielweise auf sexuell-aggressive und frauenfeindliche Sprüche gehen manchen Kritikern wie Leah Fessler von der Medien-Website Quartz News allerdings nicht weit genug, wie der Tagesspiegel berichtete [2]. In dem Beitrag kamen auch amerikanische Eltern zu Wort, die beklagten, dass Alexa weder ein „Bitte“ einfordere noch ein „Danke“ verlange, was Kindern beibringe, dass sie auch ohne höfliche Umgangsformeln das bekommen, was sie wollen.

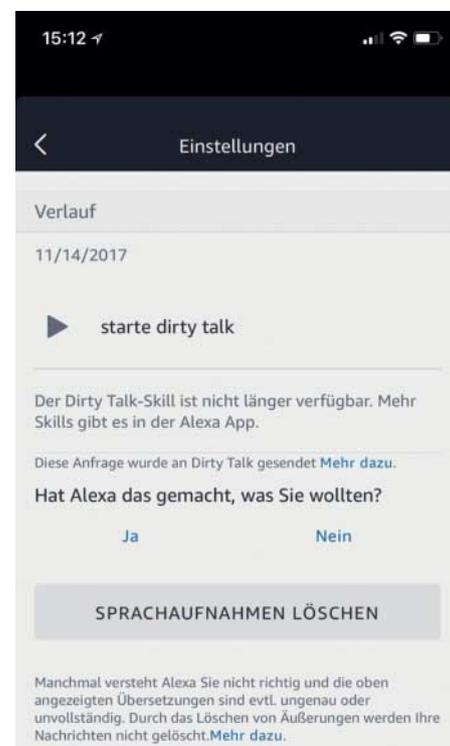
Es ist aber durchaus nicht so, dass die Entwickler Umgangsformen völlig außer Acht lassen: Bei jeder Lokalisierung nehmen Amazon, Google und Microsoft Anpassungen in dem Sinne vor, dass die Sprachassistenten passend zur jeweiligen

Kultur des Landes beispielsweise förmlicher, niedlicher oder lockerer auftreten. Nach Aussagen eines Entwicklers des Google Assistant gegenüber c't nimmt die Anpassung an lokale Gepflogenheiten tatsächlich die meiste Zeit in Anspruch. Denn was in einem Land als sympathische Formulierung gewertet werde, gilt in einem anderen als aufdringlich – und andernorts wiederum als unterkühlt. Hier gelte es also, die richtige Balance zu finden. Lediglich Fluchen gilt bislang übergreifend als Tabu.

Einen echten Hype gab es vor einiger Zeit um die Alexa-Skill „Dirty Talk“, nach deren Installation die Sprachassistentin haufenweise zweideutige Sprüche vom Stapel ließ. Das war aber eben deshalb lustig, weil man so etwas nicht von Alexa kannte. Mittlerweile ist der Skill aus Amazons Ökosystem verschwunden; wer ihn installiert hat, bekommt beim Aufruf nur noch eine Fehlermeldung.

Alexa, erzähle mir einen Witz

Apropos lustig: Zu den Standardaufgaben jedes Sprachassistenten in allen Ländern gehört das Erzählen einen guten Witzes.



Kein „Dirty Talk“ mehr von Alexa: Die Skill, die Amazons Sprachassistentin zweideutige Sprüche entlockte, ist aus der Alexa-App verschwunden. Wer sie bereits installiert hat, kann sie nicht mehr nutzen.



In den USA bietet CogniToys einen sprechenden Spielzeugdinosaurier an, der kindgerechte Antworten gibt, die sich nach dem jeweiligen Entwicklungsstand des Nachwuchses richten.

Nach Meinung vieler Nutzer scheitert Alexa gerade in dieser Disziplin kläglich.

Tatsächlich ist das Erzählen guter Witze gar nicht so einfach; der Erfolg hängt von einer Reihe von Vorüberlegungen ab, die Menschen meist unbewusst durchführen. So wählt der Vortragende aus seinem Repertoire im Idealfall den zur jeweiligen Hörerschaft und zur Situation passenden Witz aus. Talentierte Erzähler passen den Witz sogar noch auf die Gegebenheiten an, bauen mit Pausen Spannung auf und beziehen das Publikum aktiv mit ein. Zudem unterstützen sie ihren Vortrag mit Gesten.

Dem Sprachassistentensystem fehlt der ganze Kontext, weshalb es für ihn schwierig ist, einen wirklich zur Gelegenheit passenden Scherz rauszuhauen. Und auch hier gilt wiederum, dass der Witz die Grenze zur Versautheit nicht überschreiten darf – schon weil Kinder mit im Raum sein könnten.

Glaubt man den Aussagen der Entwickler, hat sich Amazon daher ganz bewusst dafür entschieden, Alexa schlechte Witze erzählen zu lassen – also aus der Not eine Tugend gemacht und den schlechten Witz zur Kunstform erhoben.

Angepasstes Verhalten

Eine weitere Intensivierung der Beziehung zwischen Mensch und Maschine ist schon dadurch zu erwarten, dass Sprach-

assistentensysteme künftig einzelne Nutzer an der Stimme identifizieren sollen. Der Google Assistant macht es vor, auch wenn die Erkennung aktuell noch nicht fehlerfrei funktioniert. Zu beobachten ist aber, dass Google die Identifizierung bereits nutzt, um den Namen des jeweiligen Anwenders in Antworten einzuflechten.

Betrachtet man, dass sich viele der aktuellen Entwicklungen bei Amazon und Google um die Auslieferung von Nachrichten an den Nutzer dreht, ist vorstellbar, dass Sprachassistenten künftig selbst einen Dialog initiieren, wenn für einen erkannten Benutzer eine Mitteilung vorliegt. Derzeit muss man als Nutzer stets ein Aktivierungswort nennen oder – im Falle des Dinos – eine Taste drücken.

Im nächsten Schritt könnten sich die Sprachassistenten stärker an den einzelnen Nutzer anpassen. Ein wohl nicht ganz ernstgemeinter Ansatz des Google Assistant ist die Reaktion auf das Jugendwort des Jahres 2017: Auf „I bims“ liefert er eine passende Antwort.

Eine sinnvolle Nutzung präsentiert aber bereits CogniToys Spielzeugdino, der seine Antworten dem Alter des Empfängers anpasst. Einem 9-Jährigen nennt der Assistent auf die Frage, wie weit der Mond entfernt ist, beispielsweise die genaue Entfernung, während er einem 5-Jährigen lediglich erklärt, dass die Distanz zu groß ist, um dorthin zu gehen.

Sprich mit mir!

Trotz ihrer künstlichen Intelligenz können die Assistenten aktuell kein richtiges Gespräch mit dem Menschen führen. Daran ändert auch im Falle des Google Assistant nichts, dass er in Antworten den Namen des Nutzers einstreut und mit Folgefragen zurechtkommt. „Es ist derzeit eher so, als würde man einem Hund Kommandos geben“, resümiert John Grötting, der als Group Design Director bei der Innovations- und Designberatung Fjord Unternehmen zum Thema digitale Sprachassistenten berät.

Wie weit die aktuelle Kommunikation mit Sprachassistenten noch von einem Gespräch unter Menschen entfernt ist, erkennt man laut Grötting unter anderem daran, dass sich nahestehende Personen bei einer Unterhaltung nicht ständig komplette Sätze formulieren müssen. Ein gemeinsamer Wissens- und Erfahrungsschatz erlaube es stattdessen, selbst komplexere Inhalte mit wenigen Schlüsselwörtern zu übermitteln.

Die zwischenmenschliche Kommunikation beschränkt sich zudem nicht auf einen Datenaustausch; ihre Leistung umfasst auch die Reaktion auf Emotionen des Gegenübers. Dass viele Entwickler gerne die Stimmungen des Nutzers – beispielsweise über die Stimme oder den per Kamera erfasste Gesichtsausdruck – erkennen würden, ist spätestens seit dem letzten Mobile World Congress bekannt. Dort hatten mehrere Unternehmen Prototypen präsentiert, die künftig genau dazu in der Lage sein sollen.

Doch wie werden Nutzer darauf reagieren, wenn Sprachassistenten wie alte Bekannte reden oder nachfragen, ob es ihnen auch wirklich gut geht? John Grötting geht davon aus, dass die Entwicklung der Sprachassistenten wie die Robotik früher oder später auf die als „Uncanny Valley“ (englisch „unheimliches Tal“) bekannte Akzeptanzlücke stößt. Hier wie dort würde sich die Empathie des Menschen gegenüber einem menschenähnlichen Wesen schlagartig in Abscheu wandeln, sobald dieses einen bestimmten Grad an Menschenähnlichkeit überschreitet, also ein zu perfektes Imitat darstellt. Bis dahin dürfte uns aber noch einige Zeit bleiben, in der wir Alexa & Co. alles anvertrauen können. (nij@ct.de) **ct**

Literatur

- [1] Kathrin Werner, Diese Frau bringt Computern Witze bei, Süddeutsche Zeitung, 10.10.2017, abrufbar unter ct.de/yfw8
- [2] Thomas Seibert, Alexa tyrannisiert Amerika, Der Tagesspiegel, 25.03.2017, abrufbar unter ct.de/yfw8



Sony präsentierte auf dem MWC 2017 den Prototyp seines „Xperia Agent“, der die Stimmung eines Menschen an dessen Gesichtsausdruck und anhand des Tonfalls der Stimme erkennen soll.