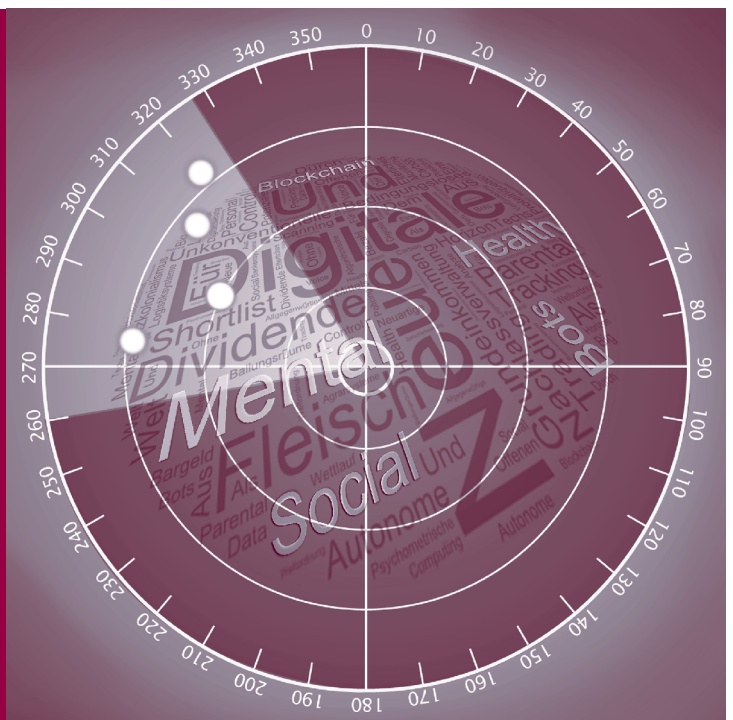




BÜRO FÜR TECHNIKFOLGEN-ABSCHÄTZUNG
BEIM DEUTSCHEN BUNDESTAG

Jan-Peter Ferdinand
Tobias Jetzke

Voice Computing – allgegenwärtige Spracherkennung





Voice Computing – allgegenwärtige Spracherkennung

Kurzdarstellung des Themas

Voice bzw. Vocal Computing beschreibt die Möglichkeit, über Sprache mit Computern, mobilen und stationären Endgeräten sowie deren softwarebasier-ten Anwendungen zu interagieren. Als intuitive und natürliche Schnittstelle bildet Voice Computing einen innovativen Entwicklungsschritt im Kontext von Mensch-Maschine-Interaktion (Nass/Brave 2005). Aktuell ist Voice Computing primär in virtuellen intelligenten Assistenten integriert, die in zwei wesentliche Anwendungszusammenhänge diffundieren: Entweder erweitert die Technologie als App-basierte Lösung die Funktionalität von Smartphones und anderen mo-bilen Endgeräten, oder sie trägt als Smart-Home-Anwendung dazu bei, das pri-vate Wohnumfeld zu vernetzen. Während Anwendungen wie Apples »Siri« oder Googles »Now« durch ihre Verbreitung über iOS und Android bereits weitge-hend in die Praxis der Smartphonennutzung integriert sind, bilden Artefakte wie Amazons »Alexa« oder Googles »Home Assistant« eine neue Kategorie von Produkten für den Heimgebrauch, die die digitale Durchdringung des Alltags weiter vorantreiben.

Die technologische Neuheit der benannten Anwendungen besteht darin, dass sie eine über natürliche Sprache gesteuerte Schnittstelle schaffen, die nicht auf einzelne Befehle reduziert ist, sondern ganze Sätze und den Kontext ihrer Äußerung erfassen kann. Die so entstehende intelligente Assistenzfunktion greift ihrerseits auf selbstlernende Algorithmen künstlicher Intelligenz (KI) zu-rück. Die Implikationen von Voice Computing sind dabei äußerst komplex und in ihren Folgen bisher nur schwer abzusehen. Ein vieldiskutierter Aspekt betrifft die Ambivalenz zwischen latenter Verfügbarkeit auf der einen sowie einer et-waigen Einbuße von Privatsphäre auf der anderen Seite: Indem die Assistenz-systeme über Sprache aktiviert werden, sind die Mikrofone zur Aufnahme der Signale latent empfangsbereit (The Economist 2017). Darüber hinaus wird durch die natürlichsprachliche Kommunikation mit den virtuellen Assistenten aufseiten der Nutzer mitunter eine irrationale soziale Nähe erzeugt, die den prinzipiell funktionalen Zweck der Interaktion emotional überhöht und eine künstliche Vertrautheit erzeugt (Turk 2016). Da die virtuellen Agenten zudem auf digitalen Plattformen großer Technologieanbieter aufsetzen, stellen ihre As-sistenzfunktionen, wie beispielsweise die Suche nach Informationen im Internet oder die Durchführung von Buchungen und anderen kommerziellen Transak-tionen, die Entscheidungssouveränität ihrer Nutzer zumindest infrage (Hackett 2017; Liptak 2017).



Hintergrund und Stand der Entwicklung

Ähnlich wie die Entwicklungen von visuellen Nutzungsoberflächen (insbesondere Microsofts Windows) oder haptisch bedienbaren Touchscreens von Smartphones, Tablets etc.) bilden auch Voice-Computing-Interfaces eine technologische Innovation im Kontext von Mensch-Maschine-Interaktion (Nass/Brave 2005). Ausgangspunkt dieser Entwicklung ist der Wunsch, Bedienschnittstellen möglichst intuitiv und niedrigschwellig zu gestalten. Da die Nutzung von Sprache als Eingabe- und Ausgabesignal eine nahezu bruchlose und somit natürliche Kommunikation mit Computerprogrammen ermöglicht, bildet Voice Computing in diesem Zusammenhang einen deutlichen Entwicklungssprung (Ebling 2016). Die aktuelle Relevanz der Technologie erschließt sich aus ihrer Anwendung in virtuellen Assistenten, die mobile oder stationäre Endgeräte mit KI und Deep-Learning-Algorithmen verknüpfen. Indem sich die Sprachsysteme unter Rekurs unzähliger Interaktionsbeispiele selbst trainieren, verbessert sich nicht nur deren Fähigkeit zur Spracherkennung, sondern auch die Qualität ihrer verbalen Ausgabe (The Economist 2017).

Auf Produktebene betten sich Voice Computing und virtuelle Assistenten in zwei Anwendungskontexte ein: Während Programme wie Siri oder Now bereits weitgehend in die Nutzungspraxis von Smartphones integriert wurden, sind Geräte wie Alexa oder Home Assistant noch relativ neu am Markt. Amazon startete den Verkauf des Echolautsprechers als Hardware für den Alexa-Voice-Service in den USA im Juni 2015, in Deutschland im Oktober 2016. Google verkauft seinen Home Assistant seit November 2016 bisher ausschließlich in den USA. Beide Geräte sind hinsichtlich ihrer Hardware sehr ähnlich: Als Kombination aus Lautsprechern, Mikrofonen und Schnittstellen, die sie digital vernetzen, bilden sie eine neue Produktkategorie im Schnittbereich aus Smart-Home-Komponenten und dem Internet der Dinge (IoT), welche die digitale Durchdringung von privaten Wohnräumen vorantreibt und das häusliche Umfeld mit globalen Datenströmen vernetzt. Konkrete Verkaufszahlen zu häuslichen Assistenzsystemen gibt es bisher nicht, jedoch gehen entsprechende Schätzungen von einem kommerziellen Erfolg aus. 2016 soll Amazon demnach weltweit ca. 5,2 Mio. Geräte verkauft haben (Dunn 2016).

In einer aktuellen Umfrage von Statista wurden 1.001 Personen, die bereits von Sprachassistenten (mobile und stationäre) gehört haben, nach den Zwecken befragt, für die sie entsprechende Lösungen nutzen würden (Brandt 2017). Die drei meistgenannten Funktionen bilden das Abrufen von Informationen über Suchmaschinen (48 % der Befragten), die Erinnerung an persönliche Termine (46 %) sowie das Erfragen der Wettervorhersage (45 %). Dass die primäre Nutzung der Assistenzfunktion bisher noch vergleichsweise eingeschränkt erfolgt, hat auch mit den technologischen Grenzen der Systeme zu tun. So gut sie bereits



darin sind, Sprache zu verstehen und zu verarbeiten, so schwer fällt es ihnen noch, das Gesagte kognitiv zu erfassen und in einen übergreifenden Sinnzusammenhang einzuordnen. Um die Potenziale von Voice Computing und digitalen Assistenten in praktische Anwendungen zu überführen, müssen Interaktionsprozesse daher noch responsiver gestaltet und die Systeme in die Lage versetzt werden, anschlussfähig und konsistent mit Menschen zu kommunizieren (The Economist 2017). Dass diese Herausforderung auch aufseiten der Technologieanbieter eine hohe Priorität besitzt, zeigt sich am Beispiel des von Amazon initiierten »Alexa Prize«. Hierbei schickt der Konzern insgesamt zwölf internationale Forscherteams mit Aussicht auf 2,5 Mio. US-Dollar ins Rennen, um einen Chat-Bot zu programmieren, der sich – in Anlehnung an den Turingtest – über eine Dauer von 20 Minuten kohärent mit Menschen zu aktuellen gesellschaftlichen Themen unterhalten kann (Amazon 2016).

Gesellschaftliche und politische Relevanz

Indem die über Voice Computing nutzbar gemachten Assistenzfunktionen über Sprache aktiviert werden – dazu genügen kurze Anreden wie »Hey Siri«, »Okay Google« oder »Alexa« –, sind die Systeme ständig in Bereitschaft. Für die Nutzer der Systeme resultiert daraus ein hohes Maß an Bequemlichkeit und eine Verfügbarkeit auf Zuruf. Damit steigt gleichzeitig das Ausmaß an Durchdringung der persönlichen Lebenswelt durch kommerzielle digitale Plattformen. Da die bisher verfügbaren Assistenten von großen Technologiekonzernen wie Amazon, Apple oder Google angeboten werden, bildet jedes der genutzten Endgeräte eine Schnittstelle zu deren globalen Datennetzen, die permanent mithören (Weddeling 2017). In diesem Zusammenhang geht die Nutzung entsprechender Dienste auch mit einer Einbuße an Privatsphäre einher, wobei vertieft zu klären wäre, ob und wie die Daten aus dem Umfeld der Nutzer von den Technologiekonzernen gespeichert werden bzw. zu welchen Zwecken die Auswertung der gewonnenen Daten erfolgt (Carroll 2015; Sauer 2017).

Auch wenn die digitalen Dienste bisher primär zu Auskunftszwecken genutzt werden, ist mittelfristig davon auszugehen, dass sie auch die Märkte für Konsumgüter verändern werden. Insbesondere mit Blick auf Amazons Heimasistent Alexa und die Einbindung in das Versandökosystem deutet sich an, wie Voice Computing die Abwicklung kommerzieller Transaktionen beeinflussen kann. So können Nutzer über Alexa direkt auf die Verkaufsplattform von Amazon zugreifen und Bestellungen tätigen. Für die Konsumenten stellt sich somit nicht mehr die Frage, wo oder über wen sie ihre Waren beziehen, was die Dominanz von E-Commerce mit allen absehbaren Folgen weiter zuspitzt und zu zusätzlichen Monopolisierungstendenzen führen kann (Kharpal 2017).



Sofern Forschung und Entwicklung im Bereich der Kontextualisierung von Mensch-Maschine-Kommunikation Voice Computing in die Lage versetzen, anschlussfähig und responsiv mit Menschen zu interagieren, gewinnt auch die soziale und emotionale Komponente digitaler Assistenzsysteme an Relevanz. So berichtet beispielsweise Daren Gill, der bei Amazon als Produktdirektor für Alexa arbeitet, dass bereits 500.000 Menschen seiner Alexa die Liebe gestanden haben. Darüber hinaus scheint es für viele Nutzer eine gängige Praxis zu sein, Alexa einen »Guten Morgen« zu wünschen oder sich nach der Interaktion für ihre Assistenz zu bedanken (Turk 2016). Auch wenn die Aussicht auf amouröse Beziehungen zwischen Menschen und virtuellen Assistenzsystemen (vergleiche hier z. B. den Film »Her« von Spike Jonze [Imdb 2014]) mittelfristig sicherlich keinen Nutzungsschwerpunkt der Systeme bilden wird, ist ein Wandel der Interaktionsmuster zwischen Menschen und Computern tendenziell absehbar. In diesem Zusammenhang gilt es, die sozialen Implikationen einer künstlichen Vertrautheit zu analysieren, die in der Vermenschlichung von Maschinen über Voice Computing ihre technologische Basis hat.

Über den derzeit dominanten Anwendungskontext persönlicher Assistenten hinaus kann Voice Computing auch für professionelle Unterstützungsfunktionen genutzt werden. Da die Technologie eine natürliche Schnittstelle für jegliche computerbasierte Anwendungen bietet, ist sie prinzipiell für alle Tätigkeitsfelder relevant, in denen die manuelle Bedienung computergesteuerter Geräte hinderlich ist. Aktuelle professionelle Anwendungsfelder liegen in der Medizin, wo Voice Computing beispielsweise die Nutzung von IBMs KI-Lösung »Watson for Oncology« unterstützt (Ebling 2016), oder in der Wissenschaft, in der Forscher die Assistenzfunktionen von Amazons Alexa für die Unterstützung ihrer Labortätigkeiten erweitern (Halford 2017). Erhöhen die prinzipiell vielfältigen Anknüpfungspunkte von Voice Computing in privaten wie professionellen Zusammenhängen einerseits deren wirtschaftliche und gesellschaftliche Relevanz, potenzieren sich andererseits auch die zuvor angedeuteten Herausforderungen, z. B. bezüglich des Umgangs und Schutzes der aufgezeichneten Daten.

Mögliche vertiefte Bearbeitung des Themas

Die Nutzung von natürlicher Sprache als Medium für die Interaktion mit Computern und softwarebasierten Anwendungen lässt weitreichende gesellschaftliche Implikationen erwarten, die bisher jedoch noch nicht systematisch erhoben und hinsichtlich ihres gesellschaftspolitischen Handlungsbedarfs bewertet wurden. Vor dem Hintergrund der Aktualität des Themas und der ansteigenden Diffusion von Voice Computing in verschiedene, breit gestreute Anwendungsfelder, wäre



für die vertiefte Bearbeitung des Themas eine Sondierungsstudie geeignet. Die Implikationen von Voice Computing sollten dabei in wirtschaftlicher, sozialer und politischer Dimension betrachtet werden, um mögliche regulatorische Handlungsbedarfe zu sondieren. Zu untersuchende Fragestellungen könnten sein, welchen Einfluss das Voice Computing auf Arbeits- und Lebenswelten hat, wie sich die Interaktionsmuster zwischen Menschen und Computern wandeln, inwieweit Aspekte der Entscheidungs- und Datensouveränität aber auch sicherheitsrelevante Aspekte (z. B. Überwachung, Stimmabgleiche) betroffen sind.

Literatur

- Amazon.com, Inc. (Hg.) (2016): The Alexa Prize. <https://developer.amazon.com/alexaprize> (7.4.2017)
- Carroll, R. (2015): Goodbye privacy, hello »Alexa«: Amazon Echo, the home robot who hears it all. 21.11., www.theguardian.com/technology/2015/nov/21/amazon-echo-alexa-home-robot-privacy-cloud (10.4.2017)
- Dunn, J. (2016): It's been a good year for the Amazon Echo. 28.12., www.businessinsider.de/amazon-echo-sales-figures-stats-chart-2016-12?r=US&IR=T (10.4.2017)
- Ebling, M. R. (2016): Can Cognitive Assistants Disappear? In: IEEE Pervasive Computing 15, 2016 (3), S. 4–6
- Hackett, R. (2017): Amazon Echo's Alexa Went Dollhouse Crazy. 9.1., <http://fortune.com/2017/01/09/amazon-echo-alexa-dollhouse/> (7.4.2017)
- Halford, B. (2017): Meet your new lab assistant. In: Chemical & Engineering News Volume 95(1), S. 26–27
- IMDb.com, Inc. (Hg.) (2014): Her. <http://www.imdb.com/title/tt1798709/> (10.4.2017)
- Kharpal, A. (2017): Amazon's voice assistant Alexa could be a \$10 billion »mega-hit« by 2020. 10.3., www.cnbc.com/2017/03/10/amazon-alexa-voice-assistant-could-be-a-10-billion-mega-hit-by-2020-research.html (10.4.2017)
- Liptak, A. (2017): Amazon's Alexa started ordering people dollhouses after hearing its name on TV. 7.1., www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse (7.4.2017)
- Nass, C. I.; Brave, S. (2005): Wired for speech. How voice activates and advances the human-computer relationship. Cambridge
- Sauer, G. (2017): A murder case tests Alexa's devotion to your privacy. 28.2., www.wired.com/2017/02/murder-case-tests-alexa-devotion-privacy/ (10.4.2017)
- Brandt, M. (2017): Dafür würden die Deutschen Siri und Co. nutzen. 31.3., <https://de.statista.com/infografik/8757/art-der-nutzung-von-digitalen-sprachassistenten/> (7.4.2017)
- The Economist (Hg.) (2017): Now we're talking. How voice technology is transforming computing. 7.1., www.economist.com/news/leaders/21713836-casting-magic-spell-it-lets-people-control-world-through-words-alone-how-voice (6.6.2017)
- Turk, V. (2016): Home invasion. In: New Scientist Volume 232(3104–3106), S. 16–17
- Weddeling, B. (2017): Alexa hört immer mit. 7.3., www.handelsblatt.com/technik/it-internet/amazon-sprachassistent-alexa-hoert-immer-mit/19485418.html (10.4.2017)



**BÜRO FÜR TECHNIKFOLGEN-ABSCHÄTZUNG
BEIM DEUTSCHEN BUNDESTAG**

KARLSRUHER INSTITUT FÜR TECHNOLOGIE (KIT)

Neue Schönhauser Straße 10
10178 Berlin

Tel. +49 30 28491-0
Fax +49 30 28491-119

buero@tab-beim-bundestag.de
www.tab-beim-bundestag.de