

Andre Alpar, Claudius Konopka

Beyond Touch – intelligente Assistenten: Die kommende Generation nach Desktop und Smartphone

Intelligente Assistenten verbreiten sich weltweit in rasantem Tempo. Immer schwindelerregendere Sphären werden in Studien durchbrochen. 100 Millionen, 500 Millionen, 1 Milliarde Devices, über die heute mit intelligenten Assistenten interagiert werden kann. Der Kampf um die neue Plattform ist vollends entfacht, das Potenzial gigantisch. Smart Speaker sind die ersten Schritte hinein in die vernetzte Welt von IoT. Schon bald werden Nutzer häufiger mit Assistenten als mit ihren Mitmenschen sprechen.

DER AUTOR



Andre Alpar ist seit über 20 Jahren unternehmerisch im Bereich Online-Marketing tätig. Heutzutage ist er als Investor, Fachautor, Keynote Speaker und Podcaster (#askOMR) aktiv.

DER AUTOR



Claudius Konopka ist Gründer und Geschäftsführer der Beyto GmbH, einer auf Voice und Conversational AI spezialisierten Beratungsagentur. Mit Beyto berät er Unternehmen zum Thema Voice und virtuelle Assistenten.



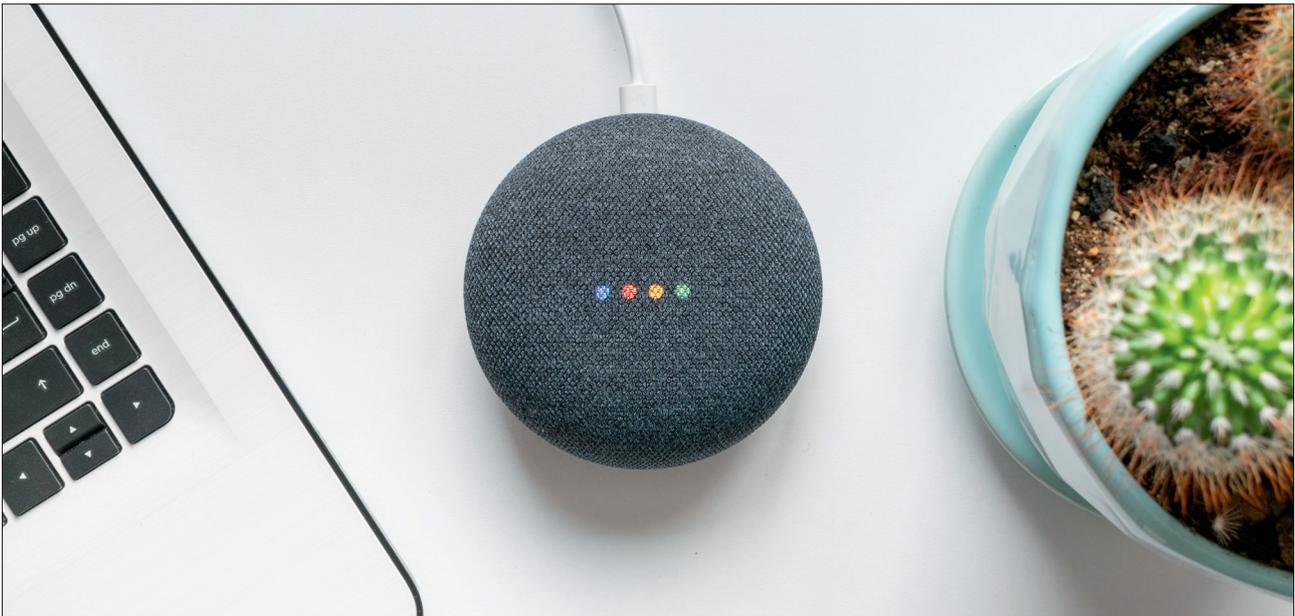


Abb. 1: Smart Speaker „Google Home Mini“ mit dem intelligenten Assistenten Google Assistant

Einordnung: Aktuelle Zahlen und Prognose

Amazons Alexa, der Google Assistant oder Siri von Apple – spätestens seit dem vergangenen Jahr halten Sprachassistenten in viel beachteter Geschwindigkeit Einzug in das Leben und die Haushalte von Millionen Menschen weltweit. Auf über drei Milliarden Devices sind die Plattformen der bekannten Tech-Giganten aufrufbar, allen voran von Google, mit dessen Google Assistant auf über einer Milliarde Geräten interagiert werden kann (<http://einfach.st/alpa1>). Vor wenigen Monaten erst gab Amazon bekannt, dass der eigene Assistant Alexa auf nunmehr über 100 Millionen Devices, von der eigenen Smart Speaker Echo-Serie bis hin zur Hardware von Drittanbietern wie Sonos oder Harman Kardon, verfügbar sei (<http://einfach.st/alpa2>). Zusammen mit den Smart Speakern von Google, dem Google Home sowie dem HomePod von Apple und einer Reihe weiterer Lautsprecher diverser Drittanbieter werden mittlerweile über 120 Millionen installierte Smart Speaker weltweit gezählt. Für dieses Jahr wird ein Wachstum von 84 % prognostiziert, was sodann mehr als 200 Millionen Smart Speaker

bedeutet. Über 50 % der amerikanischen Bevölkerung werden dann Zugriff auf einen Smart Speaker haben (<http://einfach.st/alpa3>). In Deutschland sind es derzeit über zwölf Millionen aktive Nutzer (<http://einfach.st/alpa4>). Das alles sind enorme Zahlen, die innerhalb weniger Jahre erreicht wurden und die auf ein rasantes Wachstum verweisen.

Der Grund, warum die Akzeptanz durch den Nutzer jetzt erfolgt, liegt sicher an der deutlichen Verbesserung der Sprachtechnologie. So wurde die Erkennungsrate von Automatic Speech Recognition (ASR) und Natural Language Understanding (NLU) gerade in den letzten Jahren enorm erhöht. Selbst Sprachmuster oder regionale Akzente werden mittlerweile problemlos erkannt. Neben der Sprachtechnologie ist ein weiterer Grund dem Smart Speaker selbst zuzuschreiben. Waren virtuelle Assistenten bis zum Markteintritt des Echos von Amazon in Smartphones und PCs integriert, wodurch sie stärker in der Rolle einer Bedienungshilfe wahrgenommen wurden, zogen sie mit dem Smart Speaker als eigenständiger Helfer – als ein richtiger persönlicher Assistent – in die Haushalte ein. Plötzlich konnten die Nutzer über den Assistenten andere Geräte steuern, Essen

bestellen oder sich Rezepte vorlesen lassen.

Dass das Potenzial dieser Plattformen gigantisch ist, sollte mittlerweile jedem ersichtlich sein. So wundert es dann auch nicht, dass in einer Umfrage von Pindrop unter 500 IT- und Wirtschaftsentscheidern in den USA, Großbritannien, Frankreich und Deutschland 84 % der befragten Unternehmen angaben, Voice-Technologien in Zukunft mit Kunden einsetzen zu wollen (<http://einfach.st/alpa5>). Wie die historische Entwicklung virtueller Assistenten verlief, welche Plattformen um Marktanteile kämpfen und welche Anwendungsfälle derzeit möglich sind, darum soll es in diesem Teil unserer dreiteiligen Serie „Intelligente Assistenten: Die kommende Generation nach Desktop und Smartphone“ gehen.

Die Anfänge von „Hey Siri“, „Alexa“, „OK Google“ und Co.

Der Ursprung heutiger Sprachassistenten ist wohl Anfang der 60er-Jahre zu finden, als William Dersch von IBM die Welt verblüffte, indem er im IBM-Pavillon der Weltausstellung 1962 die Technologie „Shoebbox“ vorstellte, den Vorläufer heutiger Spracherken-

nungssysteme. Die Software reagierte auf 16 gesprochene Wörter sowie die Zahlen 0 bis 9. Nach Shoebox gelang der nächste bedeutende Schritt mit dem Programm „Harpy“, das von der Carnegie Mellon University gemeinsam mit IBM und dem Stanford Research Institute entwickelt wurde. Basierte die Spracherkennung bis dahin auf einem reinen Matching-Prinzip, war Harpy in der Lage, ganze Sätze zu verstehen. Das Programm erkannte 1.011 Wörter, was dem Sprachschatz eines Dreijährigen entspricht. Damit war es das erste System, das über ein einfaches Matching hinaus unter Anwendung eines Sprachmodells Wortfolgen erkannte.

Infolgedessen brachte IBM Mitte der 80er-Jahre die sprachgesteuerte Schreibmaschine Tangora raus, die bereits 20.000 Wörter verarbeiten konnte. 1990 folgte Dragon mit „Dragon Dictate“ und 1997 „Dragon NaturallySpeaking“, das erste System kontinuierlicher Spracherkennung. Jetzt war es möglich, um die 100 Wörter je Minute zu verarbeiten. Ein Meilenstein in der damaligen Zeit. Fortan konnte Spracherkennung zur Dokumentenerstellung eingesetzt werden.

Im gleichen Jahr führte Microsoft den Assistenten „Clippy“ ein. Dabei handelte es sich um einen Office-Assistenten – der erste virtuelle Assistent, der den Benutzer bei Anwendungen von Microsoft Office unterstützen und die Benutzerfreundlichkeit verbessern sollte. Auch wenn die ersten Schritte Microsofts auf dem Weg zu einem virtuellen Assistenten scheiterten, sind sie ein weiterer Meilenstein. Denn zehn Jahre später sollte der von Apple veröffentlichte sprachbasierte Assistent „Siri“ das gleiche Ziel wie einst Clippy verfolgen.

Doch bevor Siri in den Vordergrund treten sollte, sorgte erneut IBM Schlagzeilen. Im Februar 2011 gewann das von IBM entwickelte Programm

„Watson“ in der Quizsendung Jeopardy! gegen menschliche Mitspieler. Ein Meilenstein auf dem Gebiet der künstlichen Intelligenz. Das Programm verarbeitete in natürlicher Sprache gestellte Informationen statistisch. Mehrere Algorithmen suchten gleichzeitig nach der korrekten Lösung und berechneten eine potenziell richtige Antwort. Die Leistung Watsons wurde dabei durch die „Fütterung“ des Programms mit riesigen Datenmengen ermöglicht.

In den folgenden Jahren sollten Programme über maschinelle Lern-techniken auf den Umgang mit riesigen Sprachdatensätzen trainiert werden. Der nächste große Schritt gelang dabei Google mit der Voice Search App für iOS. Dabei setzte Google auf Cloud Computing, was mit einem Male eine schier unglaubliche Rechenleistung für die Spracherkennung ermöglichte. Plötzlich war Google imstande, dazu große Datenanalysen der menschlichen Sprache durchzuführen. Dabei konnte Google auf Milliarden von Suchanfragen zurückgreifen. Im Jahr 2010 erweiterte Google die Sprachsuche um die personalisierte Erkennung auf der Android-Plattform. Mitte des Jahres 2011 folgte die Integration der sprachbasierten Suche in den Chrome-Browser.

Wenige Monate später, im Oktober 2011, launchte Apple mit dem iPhone 4s dann den eigenen virtuellen Assistenten Siri. Initial wurde der Sprachassistent als Applikation für iOS im Februar 2010 veröffentlicht, bevor Apple das Unternehmen Siri Inc. zwei Monate später im April 2010 akquirieren sollte.

Der nächste Schritt in der Evolution virtueller Assistenten erfolgte dann im Juli 2012, als wiederum Google seinen ersten persönlichen Assistenten „Google Now“ auf den Markt brachte. Ein virtueller Assistent, der basierend auf dem Suchverhalten sowie der Verwendung der Google Services dem Nutzer proaktiv Informationen bereitstellte.

Die Technologie galt als der erste virtuelle Assistent, der die Bedürfnisse des Nutzers wirklich zu antizipieren verstand.

Beinahe ein ganzes Jahr sollte daraufhin vergehen, bis Microsoft eine Antwort auf Siri und Google Now fand und im April 2013 mit Cortana einen eigenen virtuellen Assistenten auf der BUILD Developer Conference 2013 vorstellte sowie Anfang 2015 mit der Version von Windows 10 auf Desktop sowie den Mobile Devices einführte.

Kurz zuvor, im November 2014, sorgte wiederum Amazon für Furore an den Märkten. Hatte der Tech-Gigant gerade erst noch mit seinem Projekt „Fire Phone“ eine unerwartet heftige Bruchlandung hingelegt, wusste das Unternehmen die Märkte mit dem Launch von „Alexa“ und dem „Echo“-Device zu überraschen. Nach knapp vierjähriger Entwicklungszeit unter strengster Geheimhaltung wurde im für Innovationen zuständigen Lab 126 rechtzeitig zum Jahresendgeschäft der eigens entwickelte Sprachassistent für Prime-Mitglieder auf dem US-Markt gelauncht, bevor der offizielle Markteintritt im Juni 2015 erfolgte. Ab Oktober 2016 war Alexa mit dem Echo-Speaker in Deutschland verfügbar.

Ein knappes Jahr später, im Mai 2016, folgte die Antwort seitens Google mit dem virtuellen Assistenten „Google Assistant“. Ein intelligenter Assistent, der ursprünglich als Teil der Messenger-App Allo veröffentlicht wurde und jetzt auch Google Now ablöste.

Im April 2017 folgte der von Samsung entwickelte virtuelle Assistent „Bixby“, für den Samsung sich die Software und das Know-how im Oktober 2016 von Viv Labs gesichert hatte, einem von den ehemaligen Siri-Inc.-Gründern Dag Kittlaus und Adam Cheyer im Jahr 2012 neu gegründeten Unternehmen. Seit Anfang dieses Jah-

Anbieter	Amazon	Google	Apple	Microsoft	Samsung
Plattform	Alexa	Google Assistant	Siri	Cortana	Bixby
Verbreitung	~100 Millionen	~1 Milliarde	~500 Millionen	~500 Millionen	~500 Millionen
Anzahl Sprachen	~5	~28	~21 Sprachen	~18 Sprachen	~4 Sprachen
Anbieter Devices	Amazon Echo Serie, FireTV, Fire Tablet	Google Home Serie, Android Smartphones, Chromebook Pixel, Pixel Buds	HomePod, iPhone ab 4s, iPad, Apple Watch, MacBook Pro, Apple TV	Windows 10 PCs, Windows Phones, Xbox	Samsungs Produktpalette (Galaxy Smartphones, Galaxy Buds, Kühlschränke, Waschmaschine)
Schnittstelle	Alexa Voice Services (AVS)	Google Assistant SDK	HomeKit	Cortana SDK	Bixby SDK
Third-Party Devices	Windows 10 PCs, Harman Kardon, Libratone, Riva und Sonos	Panasonic, JBL, Sony, Onkyo, Sonos	Siri ist nur auf Apple-Geräten verfügbar	Amazon Echo Serie, Harman Kardon Invoice Speaker, HP Speaker, Windows 10 IoT Devices	N/A
Datenquellen	Bing Featured Snippets, Bing Answer Boxes, Bing Knowledge Graph, Wolfram Alpha	Google Featured Snippets, Google Answer Boxes, Google Knowledge Graph	Google Featured Snippets, Google Answer Boxes, Google Knowledge Graph, Wikipedia, Wolfram Alpha	Bing Featured Snippets, Bing Answer Boxes, Bing Knowledge Graph	Google Featured Snippets, Google Answer Boxes, Google Knowledge Graph, Google Translate, Pinterest
Third-Party Applikationen	Alexa Skills	Actions on Google	SiriKit	Cortana Skills	Bixby Capsule

Abb. 2: Übersicht der größten Plattformen in Europa/USA (Stand: Mai 2019)

res ist Bixby auf Deutsch verfügbar, nachdem der Assistent zuvor lediglich Englisch, Koreanisch und Mandarin-Chinesisch verstand. Noch für dieses Jahr hat Samsung einen eigenen Smart Speaker angekündigt.

Das „Ding im Moment“ (Smart Speaker) wird nicht „das Finale“ sein

Derzeit werden intelligente Sprachassistenten häufig mit Smart Speakern gleichgesetzt bzw. als diese wahrgenommen. Dabei stellt das Device, wie der Google Home oder die Lautsprecher der Echo-Serie von Amazon, lediglich die „Hülle“ des cloudbasierten Dienstes und auch nur einen weiteren Zwischenschritt in der Ausbreitung der neuen Plattformen dar. Wenn auch einen wichtigen Schritt, der die Entwicklung stark vorantrieb. Als Verantwortlicher für diesen Impuls ist sicher Amazon zu nennen, das nach dem Crash des eigenen Fire Phones zwingend ein neues Device für den entwickelten Sprachassistenten Alexa benötigte – die Geburtsstunde des Smart Speakers. Nun besitzt ein virtueller

Assistent selbst aber keine physische Form, sondern ist mehr ein komplexer Cloudservice, der in Smart Speakern, Autos oder Kühlschränken integriert ist. Im Zeitalter des Internet of Things werden virtuelle Assistenten über kurz oder lang in sämtliche Geräten Einzug finden und omnipräsent verfügbar sein. Auch wird es über eine reine Sprachingabe hinausgehen. Bewegungen, Mimik und Gesten werden ebenso zu einer natürlichen Interaktion mit dem Assistenten gehören, weshalb die derzeitige Bezeichnung Sprachassistent oder Voice auch etwas zu kurz gedacht ist. Besser ist da schon die Bezeichnung „Beyond Touch“. Weiter wird der Assistent zukünftig seinen Nutzer immer besser kennenlernen, seine Interessen und Bedürfnisse und damit sein Handeln verstehen. So wird der Assistent in einer fremden Stadt niemals das nahe gelegene (weltbeste) chinesische Restaurant empfehlen, wenn der Nutzer nach einer Restaurantempfehlung fragt, aber chinesisches Essen grundsätzlich nicht mag. Weiter wird der Assistent proaktiv werden, einer der relevantesten Schritte hin zu einem

richtigen Assistenten. Er wird seinem Nutzer proaktiv fragen, ob ein Taxi gerufen werden soll, zwei Stunden vor einem gebuchten Flug. Er wird ihn einchecken, eine favorisierte Platzwahl für ihn treffen. Darauf hinweisen, dass sich ein Freund in der gleichen Stadt aufhält, und ein Treffen vereinbaren, wenn beide keine Termine in ihren Kalendern haben. Wenn im Kühlschrank das Haltbarkeitsdatum von Lebensmitteln abläuft, wird er seinen Nutzer darauf hinweisen und gleich das passende Rezept dazu reichen. Doch das ist aktuell noch Zukunftsmusik. Derzeit befinden wir uns in einer Phase, in der virtuelle Assistenten durch Nutzer angelernt werden und die Anbieter für das Lernen so viele Kundendaten wie möglich benötigen – erheben und auswerten. Niedrige Preispunkte für Smart Speaker sowie aggressive Preispolitik am Markt sind daher ein effizientes Mittel, um an diese Daten zu gelangen. Denn klar ist, die Assistenten der Gegenwart müssen noch viel lernen, bis sie als intelligent zu bezeichnen sind und die zuvor skizzierten Eigenschaften Realität werden. Sodann sind

Smart Speaker auch nur noch Elektroschrott und virtuelle Assistenten omnipräsent eingebunden. Von der Toilette bis zur Glühbirne.

Übersicht der Anbieter am Markt

In der westlichen Welt sind es die üblichen Tech-Giganten, die mit ihren Assistenten derzeit um Marktanteile in Europa, USA und Teilen des asiatischen Raums ringen. Als relevant zu bezeichnen sind Alexa von Amazon, der Google Assistant, Siri von Apple sowie Cortana von Microsoft. Als einziges nichtamerikanisches Unternehmen ist Samsung hinzuzuzählen, das auch hierzulande seinen Assistenten Bixby stellt. In der östlichen Welt dominiert Yandex mit seinem Assistenten Alisa den russischen Markt und in China duellieren sich Baidu und Alibaba mit ihren Plattformen DuerOS sowie AliGenie. Ankündigungen für weitere Assistenten gibt es bereits von Facebook, das im April den Launch eines eigenen virtuellen Assistenten bekannt gab. Oder von der Deutschen Telekom, die mit „Hallo Magenta“ kurz vor dem Markteintritt steht. Neben den Tech- und Telekommunikationsunternehmen bewegen sich auch andere Branchen auf virtuelle Assistenten zu, wie beispielsweise die Automobilbranche, wo BMW mit „Hey BMW“ gerade die Freude am Fahren revolutioniert. Bei dieser Dynamik fällt es derzeit schwer, die Übersicht zu wahren. Wahrscheinlich ist sogar, dass die hier getätigte Auswahl an Anbietern und Plattformen bei Erscheinen des Artikels schon wieder veraltet ist. Denn jedes Unternehmen mit einer relevanten Reichweite arbeitet dieser Tage an einem eigenen Assistenten. Spannend zu beobachten wird sein, welche der Plattformen sich letztendlich durchsetzen. Hier wird es perspektivisch sicher Parallelen zur damaligen Entwicklung im Mobile-Sektor geben, wo sich letztendlich Android und iOS

behaupteten. Aktuell profitiert Amazon als Marktführer bei Smart Speakern weiterhin vom zeitlichen Vorsprung des frühen Markteintritts. Doch auch wenn Amazon derzeit Marktführer bei Smart Speakern ist, holt Google auf und hat zudem einen starken Wettbewerbsvorteil. Denn während Amazon sich die Nutzerbasis für Alexa gänzlich neu aufbauen muss, profitieren Google und die übrigen Anbieter von der vorhandenen Verbreitung ihrer Devices, allen voran ihrer Smartphones. Das erklärt auch den großen Unterschied von 100 Millionen Devices mit Alexa-Integration und einer Milliarde Devices, auf denen der Google Assistant läuft, oder die teilweise großen Unterschiede bei der Anzahl unterstützter Sprachen. Im Ringen um weitere Marktanteile haben die Anbieter ihre Plattformen geöffnet und ermöglichen die Integration ihrer Assistenten in die Geräte von Drittanbietern. Zudem ist eine weitere Parallele zum Mobile-Bereich von Android und iOS in der Möglichkeit zu finden, die Plattformen mit Third-Party-Applikationen – auch Skills und Actions genannt – erweitern zu können.

Use Cases und derzeitige Anwendungen

Zukünftige Use Cases für die Interaktion mit einem virtuellen Assistenten gibt es in unendlicher Anzahl. Mit Blick auf die Gegenwart und das Leistungsvermögen der Assistenten heutiger Generationen sind Use Cases in kleinerem Umfang überall dort zu finden, wo die Eingabe per Sprache gegenüber der Touch-Eingabe überlegen, weil bequemer und schneller ist. Das kann in Situationen wie dem Fahren eines PKWs sein (Adresseingabe ins Navigationssystem), bei Abfragen, die umfangreiche Informationen benötigen (Zugverbindungen), oder wenn der Nutzer seine Hände nicht verwenden kann (Smart-Home-Steuerung beim Halten eines Babys). Auch im Business-Umfeld

lassen sich zahlreiche Use Cases abbilden. Diese reichen beispielsweise von der Buchung eines Konferenzraumes bis zu schnellen BI-Abfragen in einer Besprechungsrunde. Gegenwärtig ist aber noch deutlich zu sehen, dass das Leistungsspektrum der Sprachassistenten in den Kinderschuhen steckt. Das zeigen aktuelle Umfragen zum Nutzungsverhalten. So gehören die Steuerung von Smart-Home-Geräten, Wetterabfragen, das Abspielen von Musik oder einfache Suchanfragen zu den häufigsten Anwendungsfällen. Ein ähnliches Bild zeichnet sich bei Third-Party-Applikationen ab. Zu den am besten bewerteten Anwendungen gehören seichte Entertainment-Angebote (Pups Generator) oder einfachste Soundausgaben (Tiergeräusche). Das erinnert alles ein wenig an die Anfänge des App Stores. Und hier liegt der springende Punkt. Denn im Gegensatz zu den Anfängen von Mobile-First, sollten Unternehmen dieses Mal den Einstieg in „Beyond Touch“ nicht verpassen. Das Potenzial ist gigantisch und es ist ratsam, frühzeitig einzusteigen. Als First Mover profitiert man von günstigen Akquisitionskosten, einem deutlichen Wettbewerbsvorteil in den Skill-/Actions-Stores und frühzeitigen Learnings.

Im zweiten Teil dieser dreiteiligen Serie „Beyond Touch – intelligente Assistenten“ erfahren Sie in der nächsten Ausgabe, wie die verschiedenen Plattformen über Applikationen (Skills und Actions) erweitert und damit intelligenter gemacht werden können, und erhalten einen Überblick zur aktuellen Lage im Skill-Store. Im dritten Teil erwartet Sie dann ein Ausblick zum Horizont und auf Szenarien, wie der perfekte intelligente Assistent sich in unser Leben einfügen könnte. ¶