

Thomas Hörner

Marketing mit Sprachassistenten

So setzen Sie Alexa, Google Assistant & Co strategisch erfolgreich ein

EBOOK INSIDE



Springer Gabler

Marketing mit Sprachassistenten

Thomas Hörner

Marketing mit Sprachassistenten

So setzen Sie Alexa, Google Assistant
& Co strategisch erfolgreich ein

Thomas Hörner
Bamberg, Bayern, Deutschland

ISBN 978-3-658-25649-4 ISBN 978-3-658-25650-0 (eBook)
<https://doi.org/10.1007/978-3-658-25650-0>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Gabler

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer Gabler ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Vorwort und Überblick über das Buch

Sprachassistenten sind inzwischen – im wahrsten Sinn des Wortes – in aller Munde. Das gilt für Konsumenten ebenso wie für die Marketingbranche. Allerdings hinken Unternehmen und deren Marketingabteilungen der Marktentwicklung – wie so oft beim Aufkommen neuer Medien – hinterher. Manche zeigen sich sogar überrascht von der scheinbar plötzlichen Bedeutung dieses neuen Medienkanals.

Das ist aber auch nicht verwunderlich, haben sich Sprachassistenten bei ihren Nutzern doch noch schneller verbreitet als das mobile Internet und Apps auf Smartphones. Schon deren Verbreitung ab etwa 2009 wurde von vielen als rasant empfunden. Und gerade einmal zwei Jahre nach dem Markteintritt von Amazon Echo sagen schon heute 24 % der Konsumenten einer Umfrage in den USA, Großbritannien, Frankreich und Deutschland zufolge, dass sie Sprachassistenten einer mobilen App oder Website vorziehen. In drei Jahren werden das voraussichtlich schon 40 % der Konsumenten tun (Capgemini 2018). Es bedarf also einer dringenden Beschäftigung von Wirtschaft und Wissenschaft mit diesem Marketinginstrument.

Fachliteratur zu Sprachassistenten ist bislang allerdings vergleichsweise wenig zu finden. In Publikumszeitschriften, die sich an die Nutzer von Sprachassistenten wenden, haben Amazons Alexa und der Google Assistant zwar schon fast eine Dauerpräsenz und auch Werbemagazine, E-Commerce- und Online-Magazine befassen sich – teils schon fast Hype-artig – mit Sprachassistenten. Die betreffenden Beiträge beschränken sich aber oft auf aktuelle Nutzerbefragungen oder geben sehr allgemeine und oberflächliche Tipps zur Nutzung. Weiterführende und vertiefende Fachliteratur ist bisher rar. Diese Lücke einer fundierten, fachlichen Einordnung und Behandlung des Einsatzes von Sprachassistenten in Marketing und Unternehmen soll dieses Buch schließen.

Das Kap. 1 fragt dazu erst einmal, was Sprachassistenten sind, wie sie grundsätzlich funktionieren und welche prinzipiellen Möglichkeiten es für Unternehmen gibt, sich selbst über Voice-Apps in diese zu integrieren. Ob Ihr Unternehmen betroffen ist und welche Abteilungen sich mit Sprachassistenten beschäftigen sollten, klärt Kap. 2.

Eine Einordnung von Sprachassistenten in die Marketingtheorie nimmt Kap. 3 vor. Ausgehend von den Spezifika des Medienkanals Sprachassistent und einer Festlegung von Zielen für das Sprachassistenten-Marketing wird u. a. eine Einordnung in den Kaufprozess nach Kotler (Kotler 2016) sowie in den Marketing-Mix vorgenommen. Ergänzt werden diese Darstellungen mit Überlegungen zu Sprache und Kommunikation auf Sprachassistenten.

Das Kap. 4 widmet sich dann einer zentralen Marketingstrategie auf Sprachassistenten: Voice Content Marketing. Es startet mit Überlegungen aus Sicht der Werbewirkungsforschung und gibt anschließend vielfältige Empfehlungen für die Marketingpraxis. Analog praxisnah geschieht dies für den Bereich des Kundenservice in Kap. 5.

Waren bis zu diesem Kapitel eher eigenständige Voice Apps betrachtet worden, geht Kap. 6 auf produktbegleitende Voice-Apps sein, die als einfache Produktzugaben dienen oder produktaufwertende Funktionen wahrnehmen können.

Die danach folgenden Kapitel stellen jeweils ein ausgesuchtes Teilthema in den Fokus: So widmet sich Kap. 7 der Frage nach Vertrieb und Umsatzgenerierung mit Sprachassistenten. Das Kap. 8 ist dagegen dem Thema Marke gewidmet und ordnet Sprachassistenten in die Markenführung ein. Diese wird um neue Aspekte ebenso erweitert und es werden auch wieder ganz praktische Hilfestellungen für den Marketingalltag gegeben.

Über das reine Marketing hinaus geht dann schließlich Kap. 9. Es wendet sich der Integration von Sprachassistenten und eigenen Produkten zu: von einer Steuerbarkeit bei Smarthome-Geräten bis zur wirklichen vollständigen Integration von Sprachassistenten in das eigene Produkt.

Abgerundet wird das Buch schließlich mit Kap. 10, das sich noch mit Fragen der Präsenz auf Sprachassistenten über Dritte (z. B. Portalanbieter) und einem kurzen Einblick in Voice-Suchmaschinenoptimierung (SEO) widmet.

An wen sich das Buch richtet

Mit seiner Mischung aus fachlicher Einordnung und Fundierung einerseits sowie vielen sehr praktischen Anregungen andererseits wendet sich dieses Buch an Entscheider und Mitarbeiter im Marketing von Unternehmen als auch in Werbe- und Internetagenturen. Außerdem soll es erste Ansätze für eine wissenschaftliche Einordnung dieses relativ neuen Medienkanals liefern.

Der Autor wünscht viele fachliche Erkenntnisse, gute Anregungen für die eigene Berufspraxis und natürlich viel Spaß beim Lesen.

Thomas Hörner
Berater, Autor, Redner und
Dozent für E-Commerce und
strategisches Onlinemarketing

Literatur

1. Capgemini Research Institute. (2018). Conversational Commerce – Why Consumers Are Embracing Voice Assistants in Their Lives. <https://www.capgemini.com/resources/conversational-commerce-dti-report/>.
2. Kotler, Ph., Keller, K. L., Brady, M., Goodman, M., & Hansen, T. (2016). Marketing Management (3. Aufl.). Harlow: Pearson.

Inhaltsverzeichnis

1	Sprachassistenten – ein paar Hintergründe	1
1.1	Was sind Sprachassistenten eigentlich? Etwas Technik – auch für Nichttechniker	2
1.1.1	Eine kurze Historie der Mensch-Maschine-Schnittstellen	2
1.1.2	Macht Sprache Tastatur und Touchscreen überflüssig?	6
1.1.3	Sprach-Dialog-System oder Sprachassistent?	7
1.1.4	So funktionieren Alexa & Co.	9
1.2	Ihr eigenes Unternehmen in Alexa, Google Assistant & Co bringen	15
1.2.1	Voice-Apps (Alexa Skills, Google Actions & Co)	16
1.2.2	Voice-App-Aufrufe ohne Namensnennung	20
1.2.3	Nutzer identifizieren durch Account Linking	22
1.2.4	Das Konzept ist entscheidend	24
1.3	Welcher Sprachassistent ist der Richtige?	28
	Literatur	30
2	Einordnung in die Unternehmensaktivitäten	33
2.1	Ist Ihr Unternehmen überhaupt betroffen?	35
2.2	Betroffene Unternehmensbereiche	40
2.2.1	Strategie und Unternehmensführung	41
2.2.2	Marketing und Vertrieb	42
2.2.3	Kundenservice	43
2.2.4	Produktentwicklung	44
2.2.5	Personalabteilung	44
2.2.6	Infrastruktur und sonstige Abteilungen	45

2.3	Wann ist der richtige Zeitpunkt?	46
	Literatur.	48
3	Sprachassistenten im Marketing	49
3.1	Markt und Marketing auf dem Weg zum Sprachassistenten	50
3.2	Ziele und Controlling im Voice-Marketing	54
3.3	Spezifika des Medienkanals Sprachassistent	60
3.4	Customer Journey und Cross-Channel.	66
3.4.1	Kaufprozess nach Kotler	68
3.4.2	AIDA- und ASIDAS-Modell	82
3.4.3	„Moments of Truth“.	84
3.5	Sprachassistenten im Marketingmix	87
3.5.1	Produktpolitik.	88
3.5.2	Preispolitik	90
3.5.3	Distributionspolitik	91
3.5.4	Kommunikationspolitik	91
3.5.5	Der Marketingmix für eine Voice-App	92
3.6	Cross-Channel-Marketing mit Sprachassistenten	93
3.6.1	Frage-Antwort-Dialoge im Cross-Channel-Marketing	93
3.6.2	Interaktive Cross-Channel-Funktionalitäten	95
3.6.3	Sprachassistenten in stationären Läden	96
3.7	Kommunikation und Sprache in Voice-Apps.	97
3.7.1	Satzlänge, Wortwahl und die richtige Sprache.	98
3.7.2	Die Vielfalt der menschlichen Sprache	101
3.7.3	Emotionen und Sinne ansprechen	105
3.7.4	Eigennamen, Fremdwörter und Co	107
3.7.5	Wenn Unerwartetes passiert.	108
3.7.6	Audio-Clips integrieren	110
	Literatur.	111
4	Voice Content Marketing	115
4.1	Entwicklung und Bedeutung des Content Marketing	117
4.1.1	Wie Content Marketing zu seiner Rolle kam.	118
4.1.2	Sprachassistenten im Content Marketing.	120
4.2	Wie Content Marketing wirkt – etwas Werbewirkungsforschung	121
4.2.1	Von Stimulus-Response- und Hierarchiemodellen zu Theory of Reasoned Action und Elaboration- Likelihood-Modell	122

4.2.2	Der Preparatory-Content-Effekt des Content Marketing	126
4.2.3	Konsequenzen für die Gestaltung von Content Marketing	128
4.3	Guten Voice Content konzipieren	130
4.3.1	Die passenden Themen und Lebensbereiche	130
4.3.2	Eignung von Inhalten für Voice	135
4.3.3	Qualitätskriterien für Voice Content	136
4.3.4	Tipps und Tools für die Praxis	141
4.4	Konzeptionelle Ansätze für Voice Content Marketing.	143
4.4.1	Interaktive Alltagshilfe.	144
4.4.2	How-to und längere Anleitungen	146
4.4.3	Beratung.	148
4.4.4	Nachschlagewerke, Lexika und Verzeichnisse.	149
4.4.5	Unterhaltung und Spiele.	150
4.4.6	Bildung und Lernen	153
	Literatur.	154
5	Kundenservices leisten	157
5.1	After-Sales-Bestellabwicklung	160
5.1.1	Bestellungen abändern und verwalten	161
5.1.2	Versandstatusanfragen	162
5.1.3	Bezahlung, Adresse und Co	163
5.1.4	Retourenmanagement.	163
5.2	Kundenkonto, Punktekonto und frühere Bestellungen	165
5.3	Inbetriebnahme, Aufbau oder Konfiguration von Produkten.	167
5.4	Schulung und Trainings	168
5.5	Probleme bei der Produktnutzung	169
5.6	Produktberater und -information	170
5.7	Grundlegende Anforderungen an Kundenservices.	171
	Literatur.	175
6	Produktbegleitende Voice-Apps.	177
6.1	Voice-App – die kostengünstige Produktzugabe	178
6.2	Leistungsbündel mit Voice-Apps.	181
6.3	Produktbegleitende Leistungen	182
6.3.1	Vereinfachte Handhabung	182
6.3.2	Aufwertung des Produkts durch erweiterte Leistungen	183
6.3.3	Hilfestellung durch interaktive Handbücher	186

6.4	Voice-Apps als integraler Produktbestandteil	187
	Literatur.	188
7	Mit Sprachassistenten verkaufen?!	189
7.1	Grundlegende Fragen bei Voice-Verkäufen	190
7.1.1	Kauf nur noch direkt bei Amazon?	190
7.1.2	Das Problem der Kundendaten und Bezahlung	192
7.2	Strategien zur Umsatzerzielung	195
7.2.1	Direkte Bestellannahme	196
7.2.2	Wiederkauf.	201
7.2.3	Deals und Angebote	202
7.2.4	Verfügbarkeits- und Preisabfragen	203
7.2.5	Verkaufs- und Produktberatung	204
7.2.6	Von der Produktinformation zum Kauf	207
7.2.7	Merkzettelfunktionen	210
7.2.8	Say and Collect	211
7.2.9	Conversion im Voice Content Marketing	213
7.2.10	Bestellen auf Sprachassistenten aus rechtlicher Sicht	214
7.2.11	Geht Verkaufen über Sprachassistenten wirklich?	215
7.3	Indirekter Vertrieb	217
7.3.1	Angebote der Sprachassistentenbetreiber	217
7.3.2	Voice-Apps von Drittanbietern	218
	Literatur.	219
8	Marken auf Sprachassistenten.	221
8.1	Von der Markierung zum Voice Branding	224
8.1.1	Zur Begrifflichkeit Voice Identity und Voice Branding	225
8.1.2	Voice Branding als Teil der Markenführung	227
8.1.3	Die Markenstimme auf Sprachassistenten	231
8.2	Markenkommunikation per Voice-App	233
8.2.1	Wer (kommuniziert)?	234
8.2.2	Was und wie (wird kommuniziert)?	240
8.2.3	Mit wem (wird kommuniziert)?	246
	Literatur.	247
9	Sprachassistenten in eigene Produkte integrieren	251
9.1	Sprachbedienung Ihrer eigenen Produkte	252
9.2	Vollintegration von Sprachassistenten	254

10	Was sonst noch zu bedenken ist	257
10.1	Ohne Voice-App auf Sprachassistenten präsent sein	257
10.2	Suchmaschinenoptimierung, Voice- und Sprachassistenten	261
10.3	Geschäftsmodelle rund um Sprachassistenten	264
10.3.1	Voice-App als Umsatzquelle	264
10.3.2	Zubehör, Entwicklung und Dienstleistungen	267
	Literatur	268

Über den Autor

„Das Internet ist kein Netz von Computern, sondern ein Netz von Menschen“ – nach diesem Motto begleitet Thomas Hörner seit über 25 Jahren Unternehmen aus Mittelstand und Industrie in den Online-Medien und beschäftigt sich mit den E-Commerce-Themen im Rahmen seiner Hochschultätigkeit. Über 20 Jahre führte er eine von ihm gegründete Internetagentur und ist heute als Berater, Redner, Autor und Dozent für E-Commerce und strategisches Online-marketing tätig. Er stützt sich dabei getreu seines obigen Mottos auf eigene unternehmerische Erfahrung, fundiertes Wissen der Psychologie als auch auf gute Kenntnisse der technischen Softwareentwicklung. Gerade die Kombination aus diesen drei Wissensgebieten führte ihn oft zu interessanten Projekten und Erkenntnissen.

Thomas Hörner veröffentlicht seit 2006 in den Verlagen Springer Gabler, C.H. Beck und im Eigenverlag sowie in verschiedenen Fachzeitschriften. Als Redner ist er auf Kongressen und anderen Fachveranstaltungen in Deutschland und dem angrenzenden Ausland ebenso gefragt, wie als Seminarleiter, für Inhouse-Schulungen oder als Coach. Daneben war er an verschiedenen Hochschulen aktiv und lehrt seit 2009 regelmäßig an der FOM Hochschule für Berufstätige (anfänglich in Hamburg, aktuell in Nürnberg). Bereits seit den 1990er-Jahren ist er aktives Mitglied im Bundesverband Digitale Wirtschaft (BVDW, anfänglich noch dmmv).

Mit dem Thema Sprachassistenten beschäftigt er sich seit deren erstem Aufkommen intensiv. Seine Schwerpunktinteressen liegen neben dem Voice Commerce im Content Marketing, bei Strategie und Trends des E-Commerce sowie bei Google (SEO, Ads, Analytics und weitere Tools). Darüber hinaus beschäftigt er sich mit gesellschaftlichen Fragen wie dem Eigentumsrecht an persönliche Verhaltensdaten oder der Datenethik. Sie erreichen ihn unter kontakt@thomas-hoerner.de

Abkürzungsverzeichnis

Abb.	Abbildung
AGB	Allgemeine Geschäftsbedingungen
AIDA	Attention – Interest – Desire – Action
ASIDAS	Attention – Search – Interest – Desire – Action - Share
ASO	App Store Optimization
B2B	Business-to-Business
B2C	Business-to-Consumer
CD	Corporate Design
CI	Corporate Identity
CLV	Customer Lifetime Value
CM	Content Marketing
CRM	Customer Relationship Management
ELM	Elaboration Likelihood Model
ERP	Enterprise Ressource Planning
F&E	Forschung und Entwicklung
KPI	Key Performance Indicator
NAO	Nameless App Optimization
PC	Personal Computer
PCE	Preparatory Content Effect
PI	Performance Indicator
PIM	Product Information Management
PR	Public Relations
SDS	Sprach-Dialog-System
SEO	Search Engine Optimization
S-R-Modelle	Stimulus-Response-Modelle
SSE	Speach Signal Enhancement

SSML	Speech Synthesis Markup Language
TRA	Theory of Reasoned Action
VCM	Voice Content Marketing
WLAN	Wireless Local Area Network
ZMOT	Zero Moment of Truth

Sprachassistenten – ein paar Hintergründe

1

Inhaltsverzeichnis

1.1 Was sind Sprachassistenten eigentlich? Etwas Technik – auch für Nichttechniker	2
1.1.1 Eine kurze Historie der Mensch-Maschine-Schnittstellen.....	2
1.1.2 Macht Sprache Tastatur und Touchscreen überflüssig?.....	6
1.1.3 Sprach-Dialog-System oder Sprachassistent?	7
1.1.4 So funktionieren Alexa & Co.	9
1.2 Ihr eigenes Unternehmen in Alexa, Google Assistant & Co bringen	15
1.2.1 Voice-Apps (Alexa Skills, Google Actions & Co).....	16
1.2.2 Voice-App-Aufrufe ohne Namensnennung	20
1.2.3 Nutzer identifizieren durch Account Linking	22
1.2.4 Das Konzept ist entscheidend	24
1.3 Welcher Sprachassistent ist der Richtige?	28
Literatur	30

Sprachassistenten haben sich in den letzten Jahren rasant verbreitet. Das Wachstum der Nutzerzahlen geschah sogar deutlich rasanter als noch vor etwa zehn Jahren bei der Verbreitung des Smartphones. Kein Wunder also, dass sich immer mehr Unternehmen mit diesem neuen Medium beschäftigen und überlegen, wie sie davon profitieren können.

Konkrete Unternehmensaktivitäten in solchen neu aufkommenden Medien – wie jetzt eben die Sprachassistenten – setzen aber voraus, sich der technischen und kommunikativen Besonderheiten eben dieses Mediums bewusst zu werden, als auch grob zu verstehen, wie diese funktionieren. Deshalb soll im Folgenden zu Beginn dieses Buchs kurz darauf eingegangen werden, wie die Entwicklung hin zu Sprachassistenten historisch einzuordnen ist, was technisch überhaupt

hinter Sprachassistenten steckt und insbesondere auch, wie sich Unternehmen in diese Systeme einklinken können. Alle nachfolgenden Kapitel widmen sich dann Marketing- und Unternehmensstrategien für dieses Kommunikationsmedium.

1.1 Was sind Sprachassistenten eigentlich? Etwas Technik – auch für Nichttechniker

Bekannt sind Sprachassistenten vorwiegend in Form von runden Boxen in verschiedenen Farben, die unter den Bezeichnungen Amazon Echo oder Google Home verkauft werden. Die gleichen Sprachassistenten existieren aber auch als Apps auf Smartphones und immer öfter sind sie direkt in Geräten von Auto bis Waschmaschine integriert.

Was aber steckt eigentlich hinter dieser Technik? Welche Abläufe produzieren eigentlich die Antworten, die wir als Nutzer von den Sprachassistenten bekommen? Ein kurzer Blick hinter die Kulissen soll helfen, Sprachassistenten und deren Technik etwas besser zu verstehen. Aber keine Angst: Sie müssen keine technische Vorbildung haben, um die folgenden Ausführungen zu verstehen.

1.1.1 Eine kurze Historie der Mensch-Maschine-Schnittstellen

Sprachassistenten sind die aktuelle Speerspitze einer Entwicklung, die bereits 1941 mit der Rechenmaschine von Conrad Zuse begann (Abb. 1.1). Bereits von den ersten Tagen der Computertechnik an stellte sich die Frage, wie Menschen Informationen mit Computern austauschen könnten. Der Mensch musste Computern ja Aufgaben und Anweisungen geben, die dieser bearbeiten sollte. Und daraus

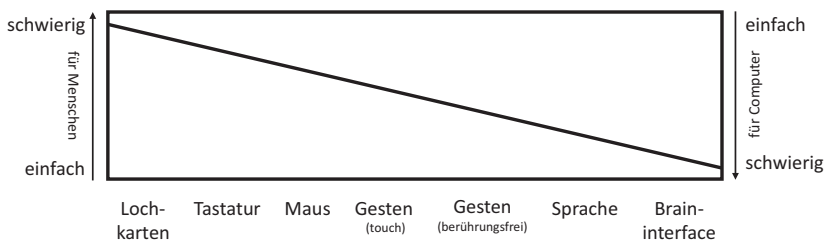


Abb. 1.1 Entwicklung der Mensch-Maschine-Interfaces

resultierende Ergebnisse mussten an den Menschen zurückgegeben werden. Diese Schnittstelle zwischen Menschen und Maschine brachte und bringt – neben Rechenleistung und neuen Algorithmen – immer wieder Innovationen hervor.

Insbesondere wurde diese Mensch-Maschine-Schnittstelle in ihrer bald 100-jährigen Entwicklung von Jahrzehnt zu Jahrzehnt immer menschenfreundlicher. Während sich zu Beginn des Computerzeitalters noch der Mensch an den Computer anpassen und seine Art, etwas zu lesen und zu schreiben übernehmen musste, ist es heute genau umgekehrt: die Computer passen ihre Kommunikation immer besser an die menschlichen Bedürfnisse an (zur Geschichte des Mensch-Maschine-Interface vgl. Hellige 2008).

Es fing mit Lochkarten an

In den ersten Computern wurden Anweisungen und Daten in Form von Lochkarten übergeben. Das war ein Papierstreifen oder eine einzelne Karte aus festem Karton, in die Löcher mit einem bestimmten Muster gestanzt wurden. Das jeweilige Muster repräsentierte die zu übertragenden Daten. Der große Nachteil: normale Menschen konnten mit diesen Lochkarten gar nichts anfangen. Neben Computern konnten nur Spezialisten diese Muster interpretieren und erstellen. Es brauchte großes Know-how, um sich Computern in dieser Kommunikationsform verständlich zu machen. Eine weite Verbreitung von Computern jenseits von Spezialistenkreisen war schon allein deshalb kaum möglich.

Es gab allerdings schon damals, in der Zeit der Lochkarten in den 1950er- und 1960er-Jahren, die Idee eines echten, natürlich-sprachlichen Mensch-Maschine-Dialogs (also eine Art Sprachinterface). Man dachte bereits damals darüber nach, dass sich eigentlich der Computer an die menschlichen Fähigkeiten anpassen sollte. Allerdings war die Technologie damals und für lange Zeit noch nicht in der Lage, diese Ideen auch wirklich zu realisieren.

Von Tastaturen und Textmonitoren zu Bildschirmen mit Mausbedienung

In der nächsten Stufe setzen sich erst einmal Interfaces mit Bildschirmen und Tastaturen durch. Jetzt hatten wesentlich mehr Menschen die Fähigkeiten, mit Computern zu kommunizieren. Informationen konnten jetzt durch das bereits von mechanischen Schreibmaschinen bekannte Tippen auf einer Tastatur sehr leicht in den Computer übertragen werden. Und durch getippte Steuerwörter konnten bestimmte Funktionen aufgerufen und Arbeitsanweisungen gegeben werden.

Diese jetzt wesentliche Vereinfachung des Informationsaustauschs zwischen Menschen und Computern war – neben der Einführung des Microsoft-Betriebssystems MS-DOS und der aufkommenden Personal-Computer (PC) als Hardware – auch eine wichtige Voraussetzung, die Computer aus den Forschungslaboren und

großen unternehmerischen Rechenzentren auf die Schreibtische in Unternehmen und immer häufiger auch in die Privathaushalte zu holen.

In einem weiteren Schritt entfiel schließlich sogar das Tippen von Steuerworten zur Auslösung von bestimmten Funktionen. Es wurde durch einfache Handbewegungen ersetzt: die Maus und grafische Benutzeroberflächen wurden als Mensch-Computer-Schnittstellen entwickelt. Der erste kommerzielle PC dieser Art war im Jahr 1981 der Xerox Star 1810. Es folgten 1987 u. a. die erste Version Windows 1.0 von Microsoft, das aber erst mit den Betriebssystemen Microsoft Windows 3.0 und Windows 95 in den Jahren 1990 bzw. 1995 eine weite Verbreitung fand. Sie ermöglichte einer breiten Masse von Nutzern die Computerbedienung über Handbewegungen (Maus) und bot ihnen eine flexible Bildschirmaufteilung (die sog. Fenster).

Gestensteuerung und Ortsunabhängigkeit

Mit dem Aufkommen von Touchscreens änderte sich die Mensch-Maschine-Schnittstelle erneut: Bisher hatte die Handbewegung bisher nur ein virtuelles Gerät bewegt (die Maus auf dem Schreibtisch steuert die virtuelle Maus auf dem Bildschirm), mit dem als eine Art Werkzeug Elemente auf dem Bildschirm manipuliert wurden. Mit dem Touchscreen konnten diese Elemente jetzt aber direkt am Bildschirm berührt und manipuliert werden.

Außerdem kamen vielfältigere Gesten zur Steuerung zum Einsatz: Wischen, Tippen, Finger auseinanderziehen (z. B. zum Zoomen) etc. Diese neue Art der Bedienung war nochmals viel einfacher als nur die Steuerung mit der Maus. Es ist kein Zufall, dass dies die erste Mensch-Maschine-Schnittstelle war, die Kinder intuitiv und oft sogar ohne jegliche Anleitung benutzen konnten. Diese Touch-Technologie gab es zwar rein technisch schon länger, den Startschuss für die massive Verbreitung gab aber Apple 2007 mit dem ersten iPhone.

Dieses und viele andere nachfolgende Smartphones unterschiedlichster Hersteller haben Computer darüber hinaus von ihrem festen Standort (z. B. auf dem Schreibtisch) sowie von den an sie angeschlossene Schnittstellentechniken (wie Tastaturen, Mäuse und Bildschirm sowie auch Stromkabel, etc.) befreit. Computer waren jetzt ortsunabhängig einfach durch Gesten nutzbar.

Für unsere Betrachtung aus dem Blickwinkel der Sprachassistenten erscheint an dieser Stelle der Hinweis wichtig, dass die eben beschriebene Entwicklungsphase auch dazu geführt hat, dass Computertechnik immer selbstverständlicher in den Alltag und das Privatleben integriert wurde. Eine Entwicklung, die Sprachassistenten nochmals fortsetzen und intensivieren.

Berührungsfreie Bedienung auf Entfernung: Sprache

Die Bedienung von Computern über natürliche Sprache ist an dieser Stelle historisch und technisch nur ein logischer weiterer Schritt in Fortentwicklung der Mensch-Maschine-Schnittstellen. Jetzt wird die Bedienung von Computern nicht nur von Tastatur und Maus befreit. Es ist für die Benutzung sogar keinerlei körperlicher Kontakt mehr notwendig. Diese Tatsache erhöht einerseits nochmals erheblich die von Smartphones bereits eingeführte Ortsunabhängigkeit der Computernutzung: es ist nicht einmal notwendig, zu einem Computer zu gehen oder diesen bei sich zu haben (wie es auch bei Smartphones noch notwendig war). Es genügt, wenn dieser in Hörweite ist.

So befreit diese neue Art der Mensch-Maschinen-Schnittstelle die Hände und den Blick des Nutzers. Er kann sich jetzt beliebigen anderen Aktivitäten zuwenden und gleichzeitig den Computer äußerst komfortabel nutzen – einfach auf Zuruf. Dieser wird damit in wesentlich mehr Alltagssituationen zu einer hilfreichen Unterstützung.

Über diese Aufhebung des physischen Kontaktbedarfs zum Computer ist Sprache aber auch die natürlichste Art der Kommunikation, die von jedem Menschen intuitiv genutzt wird. Es ist für die Nutzung von Computern jetzt nicht mehr notwendig, erst eine spezielle Fähigkeit (Lochkarten lesen, Tastatur bedienen, Maus oder Gesten) zu erlernen, um mit Computern zu kommunizieren. Noch ist diese Entwicklung zwar nicht am Endpunkt echter menschlicher Kommunikation angekommen, aber Sprachassistenten kommen dieser schrittweise (und schnell) immer näher.

Die Folge aus diesen beiden eben dargestellten Entwicklungen: Die Nutzung von Computern dringt nach der letzten Welle der Smartphoneverbreitung nochmals deutlich tiefer in Alltag und Privatsphäre der Menschen ein. Für Unternehmen lässt das natürlich einen Vorstoß in völlig neue Lebensbereiche ihrer Kunden zu und eröffnet vielfältige Chancen, die Kundenbeziehung zu intensivieren. Unternehmen und Marken können jetzt noch stärker ein Teil des Alltags der Kunden werden.

Gleichzeitig ist aber darauf hinzuweisen, dass Menschen in diesen intimeren Lebensbereichen viel empfindlicher sind, was ungewollte oder unangemessene Kommunikation angeht. Sie reagieren daher deutlich negativer auf aufdringliches Marketing oder scheinbare Störungen der Privatsphäre durch ungeeignete Kommunikation. Dieses Spannungsfeld sollte man sich immer bewusst machen, wenn man als Unternehmen Marketing- und Serviceanwendungen auf Sprachassistenten konzipiert.

1.1.2 Macht Sprache Tastatur und Touchscreen überflüssig?

Wenn nun aber Sprache eine für Menschen so extrem einfache und intuitive Art der Computernutzung ermöglicht, wird diese dann alle anderen Schnittstellen (von Tastaturen bis Touchscreens) überflüssig machen und vollständig ersetzen? Werden Computer in Zukunft ausschließlich über Sprache bedient werden?

Diese Alles-oder-Nichts-Sicht hört man leider oft, wenn sich neue Technologien verbreiten. Sollte doch das Fernsehen das früher dagewesene Kino eigentlich überflüssig machen; in Wirklichkeit haben beide sich nur für verschiedene Bedürfnisse und Verwendungen ausgeprägt. Und auch dem Internet wurde prophezeit, es würde in kürzester Zeit Ladengeschäfte überflüssig machen. Es zeigt sich aber, dass sich durch das Internet zwar Kundenverhalten und damit auch Art und Funktion von stationären Läden verändern, aber gerade die Kombination von Laden und Internet im Sinn eines Connected-Commerce- bzw. Cross-Channel-Marketings den erfolgreichsten Ansatz darstellt.

Ebenso ist ein kompletter Ersatz der bisherigen Mensch-Maschine-Interfaces durch Sprachassistenten nicht zu erwarten. Sicherlich wird sich ein Teil der Internetnutzung auf diese verlagern und es wird eine Anpassung des Nutzerverhaltens stattfinden: Alle Anwendungsfälle, bei denen Sprachassistenten klare Vorteile bieten, werden in absehbarer Zeit nur noch über diese stattfinden. Andere Anwendungsfälle, bei denen aber ein älteres Interface spezifische Vorteile hat, werden weiterhin mit diesem realisiert und weiterentwickelt. So können Texte beispielsweise schon lange auch über Mikrophone in Computer diktiert werden. Viele Menschen bevorzugen es aber weiterhin, gerade längere Texte über eine Tastatur einzugeben. Spätestens wenn Textteile markiert, kopiert oder verschoben werden sollen, sind und bleiben Tastatur und Maus (oder eventuell noch ein Touchscreen) deutlich einfacher, als dies allein mithilfe von Sprache zu tun. Auch die Erstellung von Präsentationen, Grafiken oder Zeichnungen wird sicherlich weiterhin mit der Maus oder mit Stiften auf Touchscreens erfolgen – so etwas ist über Sprache so gut wie nicht möglich. Und auch Monitore werden durch die Sprachinterfaces nicht überflüssig werden: so ist das Sehen einer der wichtigsten Sinne des Menschen und nicht zufällig werden Sprachassistenten immer häufiger auch mit Displays ausgestattet, um dem Nutzer dort etwas anzeigen zu können.

Alle Fälle jedoch, in denen Sprechen und Hören völlig ausreichend sind, werden sich in nicht allzu ferner Zukunft völlig auf Sprachinterfaces verlagern. Warum sollte ich für die Umrechnung von Pfund in Kilogramm lange auf Tastatur oder Touchscreen tippen, wenn ich einfach jederzeit und überall „wie viel sind 5 Pfund in Kilogramm“ fragen kann und sofort eine Antwort bekomme? Warum

sollte ich von der Couch aufstehen und auf ein Display tippen, nur um die Heizung um 2 °C höher zu drehen? Und jede Art von Wissen und Information – von Allgemeinwissen über Fachwissen bis zur Information, wann meine Webshop-Bestellung geliefert wird – können sehr einfach und zufriedenstellend nur mit Sprache beantwortet werden. Bildschirm, Tastatur oder Maus braucht es dazu nicht.

Etwas differenzierter wären beispielsweise die Schulung von Know-how oder Fähigkeiten sowie eine Unterstützung bei einer Problemlösung zu betrachten. Hier kann in manchen Fällen die Form eines sprachlichen Dialogs sicherlich völlig ausreichend sein, in anderen Fällen werden jedoch Papier oder Bildschirm nicht völlig überflüssig werden. So sind didaktisch aufbereitete Bücher mit vielfältigen Grafiken, Tabellen etc. für bestimmte Wissensbereiche effizienter – während ein anderer Bereich didaktisch wiederum besser in einem interaktiven Dialog vermittelbar ist.

In diesem Sinn sollten Unternehmen auch den Einsatz von Sprachassistenten planen: Kundenservices, Marketinganwendungen und auch eine Gerätebedienung, die mithilfe von Sprache von Vorteil sind, sollten schon bald auf Sprachassistenten verfügbar gemacht werden. Es würde sonst die Gefahr bestehen, am Markt gegenüber Wettbewerbern, eventuellen Start-ups oder schon allein hinsichtlich der Kundenerwartungen zurückzufallen. Die anderen Kanäle, von Print bis hin zur Website oder Mobile-App, sollten aber nicht vernachlässigt werden, sondern sich ihrerseits auf Anwendungsfälle konzentrieren, für die sie besonders geeignet sind. Und im Sinn eines Connected-Commerce- bzw. Cross-Channel-Marketings gilt es letztendlich, alle diese Kanäle geschickt zu verbinden, um Kunden ein bestmögliches Gesamterlebnis von Unternehmen, Marke und Produkten zu ermöglichen.

1.1.3 Sprach-Dialog-System oder Sprachassistent?

Im letzten Abschnitt haben wir die Entwicklung hin zu Sprachinterfaces aufgezeigt, mit denen eine Steuerung und Dateneingabe in Computer mithilfe natürlicher Sprache möglich wird. Ein solches Sprach-Dialog-System (SDS) ist für die Computerbenutzung natürlich schon sehr hilfreich, es ist aber eben nur auf dessen Bedienung beschränkt. Wirklich für den Alltag hilfreich werden die heute Sprachassistenten genannten Geräte erst, wenn sie weitergehende Anforderungen erfüllen und so von Sprach-Dialog-Systemen zu Sprachassistenten (engl. „smart assistants“) werden.

Die Antwort auf die Frage „Wie ist die Verkehrslage zum Büro?“ setzt nämlich viel mehr voraus, als nur die verwendeten Worte über ein Mikrofon in Text umzuwandeln. Es geht einerseits darum, den Sinn hinter den Worten zu verstehen

(Was ist gemeint mit Verkehrslage?) und weiteres Wissen damit zu verknüpfen (Wo ist eigentlich dieses Büro und wo befindet sich der Fragende aktuell?). Und genau hier unterscheiden sich einfache Sprach-Dialog-Systeme von Sprachassistenten: Während Erstere weitgehend nur gesprochene Worte in Computer-codes umsetzen, Befehle auslösen und teils einfach Dialogabläufe abarbeiten können (man denke nur beispielhaft an die bekannten Telefonsysteme „Wenn Sie X möchten drücken Sie 1, wenn Sie Y möchten die 2, ...“), verstehen Sprachassistenten auch die Intention der Sprechenden, sie stützen sich auf weitergehende Wissensdatenbanken, verknüpfen verschiedene Datenquellen und richten sich oft individuell und persönlich auf den jeweiligen Nutzer aus.

Ein Sprachassistent versteht in obigem Beispiel also, dass es um den Weg vom (in der obigen Frage ja gar nicht erwähnten) aktuellen Standort des Nutzers hin zu seinem Büro (und keinem anderen) geht und der Fragende mit dem Auto (und eben nicht mit Fahrrad oder Bus) fahren will. Der Sprachassistent kann so etwas z. B. aus früherem Nutzerverhalten oder Zugriff auf persönliche Daten schließen. Und er weiß auch, wo er sich Informationen über die Verkehrslage auf dieser Strecke holen kann. Assistenten gehen aber noch einen Schritt weiter (jetzt wird das Wort Assistent erst wirklich sinnvoll): Er wird eventuell sogar selbst aktiv und schlägt dem Nutzer eine alternative Fahrtroute oder eine gute Verbindung mit anderen Verkehrsmitteln vor, wenn die Verkehrslage auf dem üblichen Weg den nächsten Termin (im Kalender des Nutzers) gefährden würde. Außerdem weiß der Sprachassistent, dass eine wichtige Führungskraft im Büro heute Geburtstag hat und schlägt einen kleinen Umweg vor, der an einem Blumengeschäft vorbeiführt (nebenbei wird der passende Blumenstrauß schon einmal automatisch vorbestellt, damit keine Wartezeiten entstehen).

Sprachassistenten sind also im wahrsten Sinn des Worts Assistenten: Sie unterstützen Nutzer und Nutzerinnen in vielen Situationen des Alltags und gehen weit über eine reine Befehlsausführung (ein/ausschalten, merken, einfache Antwort geben usw.) hinaus.

Für die Anwendung von Sprache macht es im eben beschriebenen Sinn also einen großen Unterschied, ob diese als einfache Sprach-Dialog-Systeme oder echte Sprachassistenten ausgeführt werden. Das betrifft die Konzeption von Voice-Apps im Marketing genauso, wie die direkte Integration von Sprachassistenten in die eigenen Produkte: Es geht eben nicht nur um eine einfache Spracherkennung und darum, einen Touchscreen oder Bedientasten eines Geräts durch ein Sprachinterface zu ersetzen (z. B. bei einer Smarthome-Waschmaschine zu sagen: „Schalte Hemdenprogramm ein und starte Waschvorgang“). Vielmehr gilt es, darüber hinaus Assistenzleistungen zu konzipieren und den Nutzern der Sprachassistenten zu erbringen (z. B. „Ich habe Hemden mit Rotweinflecken – wie

wasche ich jetzt am besten?“). Diesen Gedanken gilt es immer bei der Konzeption von eigenen Applikationen auf Sprachassistenten zu berücksichtigen, um dem Namen Assistent auch gerecht zu werden.

Die Nutzung heute und morgen

Wenn wir ehrlich sind, werden viele Sprachassistenten in den Haushalten heutzutage sehr oft nur als reines Sprachinterface zum Musikhören, Fragen beantworten oder Lichtansalten genutzt. Aus Unternehmen ist daher immer wieder zu hören, dass Sprachassistenten keinen oder kaum Einfluss auf Verkauf und auf Umsätze hätten. Und oft reduzieren sich bis heute auch viele Marketinganwendungen auf einfache Frage-Antwort-Quizspiele. Das Potenzial von Sprachassistenten wird dabei völlig verkannt.

Die heutigen Nutzungsstatistiken zeigen aber sicherlich nur den ersten Schritt der aktuellen Entwicklung. Wie beim Aufkommen jeder neuen Technologie müssen auch die Nutzer von Sprachassistenten erst die neuen Möglichkeiten kennenlernen und sich an diese gewöhnen (sie in ihrem Alltag integrieren). Und viele Anwendungsmöglichkeiten sind bis heute noch nicht einmal realisiert oder technisch unkomfortabel. Diese Hürden fallen aber rasant und Nutzer finden immer mehr Anwendungen, die ihr Leben erleichtern. So zeigen Statistiken die Entstehung einer riesigen Anzahl Apps und Anwendungen für Sprachassistenten (analog den Millionen Apps in Smartphone-Appstores). Und selbst die direkt auf Sprachassistenten gemachten Umsätze sowie die durch sie beeinflussten Käufe in anderen Kanälen steigen deutlich.

Und nicht zu vergessen: Die Erkennungsgenauigkeit der gesprochenen Sprache ist deutlich angestiegen (sie liegt inzwischen bei über 95 % – was den Fähigkeiten eines menschlichen Zuhörers entspricht) und die Dialoge werden immer flüssiger, flexibler und menschlicher.

Das Morgen der Sprachassistentennutzung ist daher nicht erst in Jahrzehnten. Es ist in den nächsten Monaten und Jahren zu erwarten. Unternehmen tun daher gut daran, schon jetzt an sinnvollen Anwendungen für und auf Sprachassistenten zu arbeiten (und eben nicht nur einfach Sprachinterfaces zur Verfügung zu stellen). Der Markt und die Kunden von morgen (das schon bald heute sein wird) werden solchen Unternehmen zu Wettbewerbsvorteilen verhelfen.

1.1.4 So funktionieren Alexa & Co

Viele Menschen verbinden Sprachassistenten mit den „smart speakern“, also dem Amazon Echo, dem Google Home oder den analogen Geräten anderer Anbieter.

Sie werden als meist runde und verschieden große Hardware verkauft und zu Hause oder auch im Büro aufgestellt. Genau genommen sind diese Geräte aber gar nicht der eigentliche Sprachassistent. Sie sind nur eine Hardware (und etwas Software), die nichts anderes als den Zugang zum eigentlichen Sprachassistenten bieten. Der Sprachassistent selbst ist eine Software, die in den großen Rechenzentren der jeweiligen Anbieter läuft. Haben die vor Ort aufgestellten Geräte keine Verbindung mehr zu dieser zentralen Software, kann ein Sprachassistent nicht mehr genutzt werden. Funktioniert z. B. die Internetverbindung zu den Rechenzentren von Amazon, Google & Co nicht, sind die Geräte nur noch Lautsprecher und Mikrofon, die kaum etwas können.

Man könnte sagen: Die in Wohnzimmern, Küchen oder Büros stehenden Geräte sind letztendlich nur das Ohr und der Mund von Alexa, Google, Siri oder Cortana. Denn in den Gehäusen dieser Geräte sind u. a. spezielle Mikrofone verbaut, die Sprache in Räumen besonders gut erfassen. Die integrierte Elektronik macht dann eine grobe Vorverarbeitung all dessen, was die Mikrofone aus ihrer Umgebung aufnehmen. Insbesondere lauscht diese Elektronik laufend nach bestimmten Aktivierungsworten, die bei Amazon Echo „Alexa“ oder bei Google „Ok, Google“ heißen (und vom Nutzer auch auf andere Worte umgestellt werden können). Das Gerät muss daher ununterbrochen seiner Umgebung zuhören und das Gehörte verarbeiten. Die Hersteller versichern aber, dass bis zur Nennung des Aktivierungsworts die Kommunikation nicht aufgezeichnet und nicht ins Internet übertragen würde, sondern nach wenigen Sekunden vollständig vergessen (gelöscht) wird, sofern kein Aktivierungswort erkannt wurde.

Erkennt ein Sprachassistent aber festgelegte Aktivierungsworte, beginnt er, die nachfolgende Kommunikation (in ein spezielles Format komprimiert) über das Internet zur Serverfarm des jeweiligen Anbieters zu streamen (d. h. als digitale Aufnahme direkt zu übertragen). Man könnte sagen: Jetzt erst wird das Ohr mit dem Gehirn des Sprachassistenten, das sich im Rechenzentrum des Anbieters befindet, verbunden.

Es ist zwar zu erwarten, dass mit der technischen Weiterentwicklung bei zukünftigen Geräten auch einfache Standardfunktionalitäten (z. B. Uhrzeit ansagen, Umrechnungen vornehmen, Wetterinformationen geben, Musiksender starten oder eine erste Sprachvorverarbeitung tätigen, die dann auch offline zur Verfügung stünden) direkt von der integrierten Elektronik der Geräte vor Ort verarbeitet werden können. Jede weitergehende Kommunikation wird aber voraussichtlich auch zukünftig die Verarbeitung einer gestreamten Aufnahme im zentralen Rechenzentrum voraussetzen.

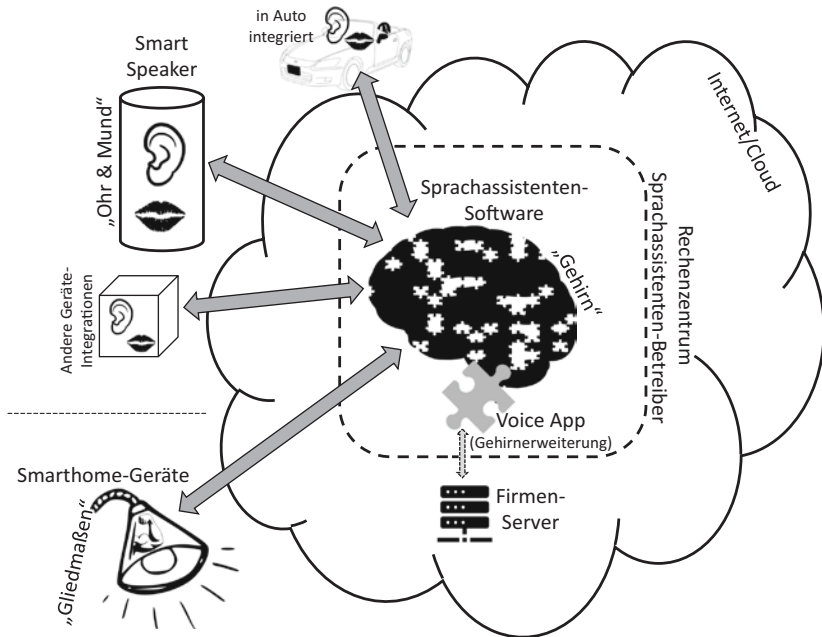


Abb. 1.2 So funktionieren Sprachassistenten – Prinzipskizze für Techniklaien

Die Software in der Cloud (dem Rechenzentrum) des Sprachassistenten-anbieters trennt schließlich mithilfe komplexer Algorithmen und Techniken des Maschinenlernens Hintergrundgeräusche von der Sprache und wandelt diese in computerlesbare Worte um. Technisch gesprochen geht es um Speech Signal Enhancement (SSE), Stimmbiometrie, Beamforming und Ähnliches. Und dazu braucht es eine riesige Rechenleistung: „Google betreibt neuronale Netze mit Abermillionen Neuronen und einer Tiefe von bis zu 30 Ebenen“ (Bager 2015), um diese Leistung vollbringen zu können (Abb. 1.2).

Es folgen als nächste Schritte die Textanalyse und Verarbeitungen zum inhaltlichen Verstehen. Das ist ähnlich, wie Menschen in einer fremden Sprache zwar einzelne Buchstaben in einem Text erkennen können, es aber für das Verstehen viel mehr Wissen zur Sprache und gegebenenfalls sogar zu kulturellen Hintergründen braucht. Analog ist – stark vereinfacht gesprochen – die Herausforderung für Sprachassistenten: Einerseits gilt es, die Worte aus der Audioaufnahme zu extrahieren, andererseits, den Inhalt zu verstehen.

Es geht bezüglich des Verständnisses insbesondere darum, Aufforderungen und Anweisungen von sonstigen Textteilen zu trennen: Es müssen die konkrete

Absicht des Nutzers (der sog. Intent) und zugehörige Parameter (die sog. Slots) extrahiert werden. Sagt man z. B.: „Ok Google, erhöhe die Raumtemperatur im Wohnzimmer um zwei Grad“, so wäre die Absicht (Intent) die Erhöhung der Raumtemperatur (also die Steuerung eines Thermostatventils) und die Parameter (Slots) wären Name des Zimmers und Gradzahl (also einmal Wohnzimmer und einmal zwei Grad als Parameterwerte). Hätte man gesagt „Ok Google, bitte Schlafzimmertemperatur um drei Grad hoch“ wäre das der gleiche Intent und die gleichen Slots (Zimmername und Gradzahl), letztere wären jedoch mit einem Wert belegt worden (Schlafzimmer statt Wohnzimmer und drei statt zwei Grad). Auch wenn die genaue Verarbeitung in der Praxis deutlich komplexer ist, zeigt dieses Beispiel sehr gut, wie die Sprachassistentensoftware Kommunikation interpretiert und maschinell verarbeitbar macht.

Nach der Ermittlung von Intent und Slots wird jetzt der zu diesem Intent passende Softwareteil aufgerufen – man könnte sagen: jetzt wird die passende Gehirnregion von Alexa & Co aktiviert. Ist der Sprachassistent direkt nach einer Antwort gefragt, greift er auf eine riesige Menge an Informationen in formalisierten Wissensdatenbanken zurück. Bei Google ist das u. a. der Knowledge Graph, der 40 Mrd. Fakten zu 570 Mio. Dingen in 15.000 Typen umfasst (Behshad Behzadi, der Director of Conversational Search bei Google laut Bager 2015) – und das sind eigentlich schon veraltete Zahlen aus 2014, die inzwischen nochmals deutlich höher liegen dürften.

Für Unternehmen und das Marketing ist an dieser Stelle wichtig: In dieser Verarbeitungsphase können sie sich auch als Drittanbieter einklinken. Sie können eine Voice-App integrieren, die sich ansprechen lassen kann und die Bearbeitung (und Beantwortung) bestimmter Anfragen (Intents) übernimmt (s. dazu im Detail Abschn. 1.2.1). Sie werden mit Hilfe Ihrer Voice App Software also quasi ein zusätzlicher Teil des Gehirns des Sprachassistenten.

Nach der Verarbeitung führt schließlich fast jede Anfrage zu einem Feedback. Dieses kann entweder wieder in Form einer sprachlichen Antwort, einer Rückfrage oder als Audio-Stream in Richtung des Nutzers erfolgen. Dazu wird technisch der auszugebende Text oder das zu streamende Audio-File vom zentralen Rechenzentrum über das Internet zurück zum Gerät vor Ort übertragen. Die dortige Elektronik gibt diese Daten dann über die integrierten Lautsprecher aus (um im obigen Bild zu bleiben: Das Gerät ist jetzt der Mund des Sprachassistenten, das vom Gehirn die Anweisungen bekommen hat, was gesagt werden soll).

Nicht jedes Feedback ist aber ein Sprachfeedback. Es gibt noch eine alternative Reaktionsform: Der Sprachassistent tut etwas, z. B. steuert er ein Gerät (z. B. ein Licht ein-/ausschalten, das Garagentor öffnen oder die Heizung um ein

paar Grad höher regeln). Dazu gibt die zentrale Sprachassistentensoftware entsprechende Befehle über vorher festgelegte Schnittstellen an die zu steuernden Geräte, die quasi die Hände und Gliedmaßen der Sprachassistenten sind und vom Gehirn entsprechende Befehle erhalten, wie sie sich verhalten sollen.

Zusammenfassend führt ein Sprachassistent die Vorgänge Sprachinformation entgegennehmen und an Rechenzentrum übermitteln, den Intent analysieren, Bearbeiten des ermittelten Bedarfs, Ausgabe als Audio oder Gerätesteuerung immer wieder aus. Manchmal reicht dabei ein einziger Durchgang, z. B. wenn nur eine einfach zu beantwortende Frage gestellt wird oder ein Gerät nur an- oder ausgeschaltet werden soll. Immer öfter entstehen aber auch ganze Dialoge, in denen mehrere solche Durchläufe erfolgen und eine inhaltliche Verknüpfung mehrerer Kommunikationsteile stattfindet. Ein Beispiel wäre die Anfrage: „Ok Google, ich brauche ein Taxi“. Der Google-Assistent antwortet dann: „Wann brauchst Du das Taxi? Jetzt?“ Sie antworten: „Nein, erst um 16 Uhr“. Google fragt dann weiter nach dem Ort fragt etc. So entsteht ein komplexer Dialog, bis schließlich Ihr Taxi durch den Sprachassistenten bestellt wird.

Sprachassistenten sind NICHT die kaufbaren Geräte

Es wurde bereits erwähnt und Sie sehen es anhand der obigen Darstellungen schnell: Man muss klar unterscheiden zwischen den kaufbaren Geräten Amazon Echo, Google Home, Apple HomePod, etc. auf der einen Seite und den Sprachassistenten Alexa, Google Assistant, Siri, etc. auf der anderen Seite.

Erstere sind – vereinfacht gesagt – nichts anderes als Mikrofone, Lautsprecher und etwas Elektronik. Sie sind die Ohren und Mänder der Sprachassistenten, die über das Internet mit dem eigentlichen Sprachassistenten im Rechenzentrum des jeweiligen Anbieters verbunden werden müssen und letztendlich nur den Zugang zu diesen ermöglichen.

Die Sprachassistenten selbst sind Software und Daten in den Rechenzentren der Anbieter. Dort ist Ihre Alexa, Ihr Google oder Ihre Siri beheimatet als Teil eines sehr komplexen Softwaresystems, sozusagen das Gehirn des Sprachassistenten.

Eine wichtige Folge aus dieser technischen Struktur: Ein Sprachassistent bleibt immer der gleiche, egal welches Gerät Sie als Zugang zu diesem Nutzen. Der Google Assistant, Alexa und andere können mehrere Ohren und Mänder haben. Einmal über den Google Home bzw. Amazon Echo, aber genauso auch über Smartphone-Apps, in verschiedensten Haushaltsgeräten oder im Auto. In allen Fällen sprechen sie mit dem gleichen – Ihrem – Sprachassistenten (dem einen Gehirn), nur jeweils über einen anderen Zugangsweg.