

Thomas Hörner

Marketing mit Sprachassistenten

So setzen Sie Alexa & Co strategisch
erfolgreich ein

2. Auflage



Springer Gabler

Marketing mit Sprachassistenten

Thomas Hörner

Marketing mit Sprachassistenten

So setzen Sie Alexa & Co strategisch erfolgreich ein

2., überarbeitete und erweiterte Auflage

Mit Beiträgen von Maria Müller



Springer Gabler

Thomas Hörner
Bamberg, Bayern, Deutschland

ISBN 978-3-658-40195-5 ISBN 978-3-658-40196-2 (eBook)
<https://doi.org/10.1007/978-3-658-40196-2>

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

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019, 2023

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 geographische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Imke Sander

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.

Der Teil I beleuchtet dazu erst einmal die Hintergründe eines Marketings auf Sprachassistenten. Kap. 1 fragt, was Sprachassistenten eigentlich 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 (2016) sowie in den Marketing-Mix vorgenommen. Ergänzt werden diese Darstellungen mit Überlegungen zu Sprache und Kommunikation auf Sprachassistenten.

Der Teil II im Buch wendet sich dann konkreten Marketing-Ansätzen auf Sprachassistenten zu. Als Erstes wird in Kap. 4 eine zentrale, grundlegende Marketingstrategie auf Sprachassistenten angesprochen: 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.

Das Kap. 6 ist dann 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.

Es folgt Kap. 7, das sich Sprachassistenten aus vertrieblicher Sicht ansieht und sich mit der Umsatzgenerierung bzw. dem direkten Verkauf mittels Sprachassistenten beschäftigt.

Waren bis zu diesem Kapitel eher eigenständige Voice-Apps betrachtet worden, geht Kap. 8 auf produktbegleitende Voice-Apps ein, die als einfache Produktzugaben dienen oder produktaufwertende Funktionen wahrnehmen können.

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.

Der Teil III liefert schließlich ganz praktische Hilfestellungen. Maria Müller wird auf den Prozess zur Umsetzung einer Voice-App (Kap. 10) eingehen und ausführlich aufgezeigen, wie konkrete Entwicklungsschritte aussehen können, um qualitativ hochwertiges Marketing auf Sprachassistenten in der Praxis umzusetzen (Kap. 11).

Abgerundet wird das Buch schließlich mit Kap. 12, 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, P., Keller, K. L., Brady, M., Goodman, M., & Hansen, T. (2016). *Marketing management* (3. Aufl.). Pearson.

Inhaltsverzeichnis

Teil I Hintergründe und Einordnung ins Marketing

1 Sprachassistenten – ein paar Hintergründe	3
1.1 Was sind Sprachassistenten eigentlich? Etwas Technik – auch für Nichttechniker	4
1.1.1 Eine kurze Historie der Mensch-Maschine-Schnittstellen	4
1.1.2 Macht Sprache Tastatur und Touchscreen überflüssig?	8
1.1.3 Sprach-Dialog-System oder Sprachassistent?	9
1.1.4 So funktionieren Alexa & Co	12
1.2 Ihr eigenes Unternehmen in Alexa & Co bringen	17
1.2.1 Voice-Apps (insb. Alexa Skills)	18
1.2.2 Voice-App-Aufrufe ohne Namensnennung	23
1.2.3 Nutzer identifizieren durch Account Linking	25
1.2.4 Das Konzept ist entscheidend	27
Literatur	31
2 Einordnung in die Unternehmensaktivitäten	33
2.1 Ist Ihr Unternehmen überhaupt betroffen?	35
2.2 Betroffene Unternehmensbereiche	41
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	45
2.2.6 Infrastruktur und sonstige Abteilungen	45

2.3	Wann ist der richtige Zeitpunkt?	46
	Literatur	49
3	Sprachassistenten im Marketing	51
3.1	Warum: Markt und Marketing auf dem Weg zu Sprachassistenten	52
3.2	Wofür: Ziele und Controlling im Voice-Marketing	56
3.3	Wo: Einordnung in Customer Journey, Marketing-Mix und Cross-Channel	62
3.3.1	Die Customer Journey	63
3.3.2	Sprachassistenten im Marketing-Mix	85
3.3.3	Cross-Channel-Marketing mit Sprachassistenten	91
3.4	Hintergründe: der Kommunikationskanal Sprachassistent	96
3.4.1	Spezifika des Medienkanals Sprachassistent	96
3.5	Kommunikation und Sprache in Voice-Apps	103
3.5.1	Satzlänge, Wortwahl und die richtige Sprache	103
3.5.2	Die Vielfalt der menschlichen Sprache	108
3.5.3	Emotionen und Sinne ansprechen	110
3.5.4	Eigennamen, Fremdwörter & Co	112
3.5.5	Wenn Unerwartetes passiert	113
3.5.6	Audio-Clips integrieren	116
	Literatur	117

Teil II Marketing-Ansätze konkret

4	Voice-Content-Marketing	123
4.1	Wie Content-Marketing zu seiner Rolle und Bedeutung kam	124
4.2	Was (Voice-)Content-Marketing eigentlich ist	126
4.2.1	Abgrenzung und Definition	127
4.2.2	Voice-Content-Marketing auf Sprachassistenten	128
4.3	Wie (Voice-)Content-Marketing wirkt – etwas Werbewirkungsforschung	130
4.3.1	Marketing-Ziel: Bekanntheit	135
4.3.2	Marketing-Ziel: positive Einstellungen und Image	137
4.3.3	Vertrauen erhöht Kaufwahrscheinlichkeit	139
4.3.4	Wirkung werblicher Marketing-Instrumente erhöhen	139

4.3.5	Konsequenzen für die Gestaltung von Voice-Apps	142
4.4	Gutes Voice-Content-Marketing konzipieren	144
4.4.1	Die passenden Themen und Lebensbereiche	144
4.4.2	Eignung von Inhalten für Voice	148
4.4.3	Qualitätskriterien für Voice-Content	150
4.4.4	Tipps und Tools für die Praxis	154
4.5	Konzeptionelle Ansätze für Voice-Content-Marketing	156
4.5.1	Interaktive Alltagshilfe	157
4.5.2	How-to und längere Anleitungen	160
4.5.3	Beratung	161
4.5.4	Nachschlagewerke, Lexika und Verzeichnisse	163
4.5.5	Unterhaltung und Spiele	164
4.5.6	Bildung und Lernen	166
	Literatur	167
5	Kundenservices leisten	171
5.1	After-Sales-Bestellabwicklung	174
5.1.1	Bestellungen abändern und verwalten	175
5.1.2	Versandstatusanfragen	176
5.1.3	Bezahlung, Adresse & Co	177
5.1.4	Retourenmanagement	178
5.2	Kundenkonto, Punktekonto und frühere Bestellungen	179
5.3	Inbetriebnahme, Aufbau oder Konfiguration von Produkten	181
5.4	Schulung und Trainings	182
5.5	Probleme bei der Produktnutzung	184
5.6	Produktberater und -information	185
5.7	Grundlegende Anforderungen an Kundenservices	186
	Literatur	190
6	Marken auf Sprachassistenten	191
6.1	Von der Markierung zum Voice-Branding	194
6.1.1	Zur Begrifflichkeit Voice Identity und Voice Branding	196
6.1.2	Einordnung von Voice Branding in die Markenführung	197
6.1.3	Grundlegendes für die Voice Identity	199
6.1.4	Die Markenstimme auf Sprachassistenten	202
6.2	Markenkommunikation per Voice-App	204

6.2.1	Wer (kommuniziert)?	205
6.2.2	Was und wie (wird kommuniziert)?	213
6.2.3	Mit wem (wird kommuniziert)?	218
6.3	Das „Kunden-fragen-Alexa“-Feature	220
	Literatur	221
7	Mit Sprachassistenten verkaufen?!	223
7.1	Grundlegende Fragen bei Voice-Verkäufen	225
7.1.1	Kauf nur noch direkt bei Amazon?	225
7.1.2	Das Problem der Kundendaten und Bezahlung	227
7.2	Strategien zur Umsatzerzielung	230
7.2.1	Direkte Bestellannahme	231
7.2.2	Wiederkauf	236
7.2.3	Deals und Angebote	237
7.2.4	Verfügbarkeits- und Preisabfragen	238
7.2.5	Verkaufs- und Produktberatung	240
7.2.6	Von der Produktinformation zum Kauf	243
7.2.7	Merkzettelfunktionen	245
7.2.8	Say and Collect	247
7.2.9	Direkte Conversion infolge von Content-Marketing	249
7.2.10	Bestellen auf Sprachassistenten aus rechtlicher Sicht	251
7.2.11	Geht Verkaufen über Sprachassistenten wirklich?	251
7.3	Indirekter Vertrieb	253
7.3.1	Angebote der Sprachassistentenbetreiber	253
7.3.2	Voice-Apps von Drittanbietern	255
	Literatur	255
8	Produktbegleitende Voice-Apps	257
8.1	Voice-App – die kostengünstige Produktzugabe	259
8.2	Leistungsbündel mit Voice-Apps	262
8.3	Produktbegleitende Leistungen	263
8.3.1	Vereinfachte Handhabung	263
8.3.2	Aufwertung des Produkts durch erweiterte Leistungen	265
8.3.3	Hilfestellung durch interaktive Handbücher	267
8.4	Voice-Apps als integraler Produktbestandteil	268
	Literatur	269

9 Sprachassistenten in eigene Produkte integrieren	271
9.1 Sprachbedienung Ihrer eigenen Produkte	272
9.2 Vollintegration von Sprachassistenten	274

Teil III Umsetzung in der Praxis

10 Prozess zur Erstellung einer Sprachanwendung	279
10.1 Phase 1: Immerse – Eintauchen in das Projekt	281
10.2 Phase 2: Deep Dive – Detailtiefe und Fokus	282
10.3 Phase 3: Convert – Von der Blaupause zur Anwendung	283
10.4 Phase 4: Vivify – Livegang und Hypercare	286
10.5 Phase 5: Learn – iteratives Nachschärfen	286
Literatur	287
11 Voice User Experience Design und Conversational Copywriting	289
11.1 User Experience – Ein Überblick	290
11.2 Die Tonalität einer Marke – Grundlagen für konsistente und wiedererkennbare Kommunikation	293
11.2.1 Was macht die Marke aus? – Werte und Attribute festhalten	294
11.2.2 Wie klingt die Marke? – Leitplanken definieren	296
11.3 Bot-Personas – Wer sie sind und wie sie sprechen	298
11.3.1 Uncanny Valley – Dort, wo es unheimlich wird	301
11.3.2 Bot Persona Steckbrief – Eine virtuelle Persönlichkeit erschaffen	302
11.3.3 Voices – Von menschlichen Sprechern und künstlichen Stimmen	307
11.3.4 Speech Synthesis Markup Language – Codebasiertes „Gestalten“ von Sprachausgaben	308
11.4 Voice UX Design – Konzepte erstellen und testen	311
11.4.1 Glückliche Pfade und Beispieldialoge – Ein Herantasten an die Mensch-Maschine-Interaktionen der Anwendung	312
11.4.2 Flussdiagramme – Darstellen, wie sich der Nutzer durch die Sprachanwendung bewegt	313
11.4.3 Ausarbeiten der Architektur – Detailarbeit jenseits der „glücklichen Pfade“	314

11.4.4	Screendesign – Die Sprachanwendung visuell anreichern	318
11.4.5	Testen, testen, testen – Denn der Nutzer steht im Mittelpunkt des Prozesses	318
11.5	UX Copywriting für Sprachbots – Maßgeschneiderte Kommunikationskomponenten	320
11.5.1	Kooperationsprinzip und Konversationsmaximen – Leitsätze für gelungene Gespräche	321
11.5.2	Best Practices –Dialoge in Sprachanwendungen intuitiv und angenehm gestalten	322
11.6	Nutzereingaben – Eine möglichst natürliche Interaktion gewährleisten	328
	Literatur	331
12	Weiteres für die Sprachassistentenpraxis	335
12.1	Ohne Voice-App auf Sprachassistenten präsent sein	335
12.2	Suchmaschinenoptimierung und Sprachassistenten	339
12.3	Geschäftsmodelle rund um Sprachassistenten	342
12.3.1	Voice-App als Umsatzquelle	343
12.3.2	Zubehör, Entwicklung und Dienstleistungen	345
	Literatur	346

Über den Autor

„Das Internet ist kein Netz von Computern, sondern ein Netz von Menschen“

Thomas Hörner – nach diesem Motto begleitet Thomas Hörner seit fast 30 Jahren Unternehmen aus Mittelstand und Industrie in den Online-Medien, schult Mitarbeiterinnen und Mitarbeiter in Agenturen und Unternehmen und beschäftigt sich im Rahmen seiner Hochschultätigkeit mit Themen des E-Commerce. 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 und des menschlichen Verhaltens 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, 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). In einer anderen Veröffentlichung hat er ein Werbewirkungsmodell sowie ein Controlling-Framework für das Content-Marketing entwickelt und veröffentlicht. Darüber hinaus beschäftigt er sich mit gesellschaftlichen Fragen wie der Ethik beim Einsatz digitaler Medien oder dem Eigentumsrecht an persönlichen Verhaltensdaten. Sie erreichen ihn unter kontakt@thomas-hoerner.de.

Über die Autorin von Kapitel 10 und 11

Maria Müller ist User-Experience-Designerin (oder auch kurz: UX-Designerin). Sie leitet das Conversational UX Design Team der Conversational Technology Agentur Frank Reply in Frankfurt am Main und beschäftigt sich seit mehr als sechs Jahren mit dem Erstellen von Sprachanwendungen, der Technologie und den kommunikationstheoretischen Hintergründen. Oder in ihren eigenen Worten: „Mein Herz schlägt für diese faszinierende Technologie, die den Nutzern – vom Kindergartenkind bis zur Seniorin – auf intuitivste Weise ermöglicht, Wissen zu erlangen, Inhalte zu nutzen, sich unterhalten zu lassen und sicherer und bequemer durch den Tag zu kommen. Ich bin begeistert von den Chancen und Möglichkeiten, die uns diese Schnittstellen bieten, und sehe den Mehrwert, den wir mit sinnstiftenden Sprachanwendungen in das Leben der Anwender bringen können, während wir gleichzeitig einen hohen Business-Value und eine nachhaltige Kundenbindung etablieren.“

Nach ihrem Studium an der Hochschule Darmstadt und ihrem Diplom in Kommunikationsdesign im Jahr 2012 war Maria Müller zunächst fünf Jahre lang als User-Experience-Designerin, hauptsächlich für Embedded Interfaces, in einer Berliner Digitalagentur tätig. Mittlerweile zurück im Rhein-Main-Gebiet und vom Conversation Design Institute als Conversation Designerin zertifiziert, taucht sie mit Neugier und Freude in immer neue Themengebiete rund um Sprachtechnologien ein und hält europaweit Vorträge zu Bereichen, die sie begeistern, wie beispielsweise zum Nutzen von künstlicher emotionaler Intelligenz (Emotion AI) im Zusammenhang mit Sprachanwendungen oder den Vorteilen von Sprachauthentifizierung im Consumer-Bereich. Ebenfalls europaweit betreut sie Kunden und Voice-Projekte im B2C-, B2B- und B2E-Kontext.

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
DCE	Direct Content Effect
ELM	Elaboration Likelihood Model
ERP	Enterprise Resource Planning
F&E	Forschung und Entwicklung
KPI	Key Performance Indicator
MS-DOS	Microsoft Disc Operating System
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	Speech 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

Teil I

Hintergründe und Einordnung ins Marketing

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, und immer mehr Unternehmen auch konkrete Anwendungen im Marketing umgesetzt haben.

Konkrete Unternehmensaktivitäten in solchen neu aufkommenden Medien – wie in letzter Zeit eben die Sprachassistenten – führen aber dazu, dass diese erst einmal grundlegend verstanden werden müssen. Und außerdem müssen solche neuen Marketing-Kanäle auch aus wissenschaftlicher Sicht grundlegend in das Marketing und seine existierende Systematik eingeordnet werden.

Dieser Aufgabe widmet sich der erste Teil dieses Buches. Es soll dabei aber nicht bei einer theoretisch-abstrakten Betrachtung bleiben. Vielmehr sollen diese Hintergründe und Einordnungen helfen, das Wissen zu bekommen, das Medium Sprachassistent besser zu verstehen und damit in der Praxis besseres Marketing auf und mit diesen zu gestalten. Schon deshalb sind die eher theoretischen Betrachtungen immer wieder mit Verweisen auf ganz konkrete Anwendungen und Beispiele verknüpft (die dann zum großen Teil in den Kapiteln der weiteren Buchabschnitte auch noch ausführlicher betrachtet werden).



Sprachassistenten – ein paar Hintergründe

1

Inhaltsverzeichnis

1.1	Was sind Sprachassistenten eigentlich? Etwas Technik – auch für Nichttechniker	4
1.1.1	Eine kurze Historie der Mensch-Maschine-Schnittstellen	4
1.1.2	Macht Sprache Tastatur und Touchscreen überflüssig?	8
1.1.3	Sprach-Dialog-System oder Sprachassistent?	9
1.1.4	So funktionieren Alexa & Co	12
1.2	Ihr eigenes Unternehmen in Alexa & Co bringen	17
1.2.1	Voice-Apps (insb. Alexa Skills)	18
1.2.2	Voice-App-Aufrufe ohne Namensnennung	23
1.2.3	Nutzer identifizieren durch Account Linking	25
1.2.4	Das Konzept ist entscheidend	27
	Literatur	31

Nutzt man Marketing-Instrumente für unternehmerische Aktivitäten, setzt das voraus, dass man zumindest grob verstanden hat, wie diese funktionieren. Insbesondere ist es wichtig, sich der technischen und insbesondere auch der kommunikativ-medialen Besonderheiten dieses neuen Mediums bewusst zu werden. Nur so kann ein Marketing-Instrument (hier: die Sprachassistenten) erfolgversprechend und fachlich korrekt in das Marketing allgemein sowie in eine unternehmerische Marketing-Strategie im Speziellen eingeordnet werden.

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.

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 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 Spitze 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 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.

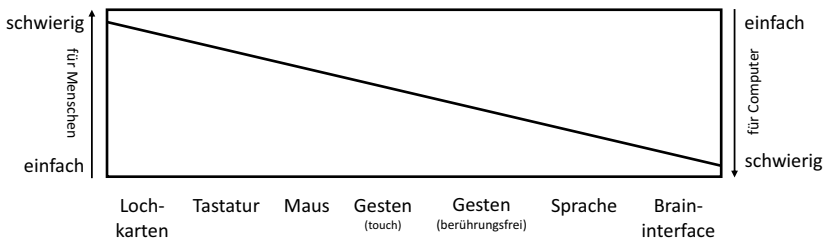


Abb. 1.1 Entwicklung der Mensch-Maschine-Interfaces

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 musste und dessen 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 und die menschliche Kommunikation 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 festerem 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 noch für lange Zeit 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 damals 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 fiel schließlich sogar das Tippen von Steuerworten zur Auslösung von bestimmten Funktionen weg. 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 1983 der Apple Lisa mit grafischer Benutzeroberfläche und 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 alle ermöglichten einer breiten Masse von Nutzern die Computerbedienung über Handbewegungen (Maus) und boten 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 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 brauchte es dieses Werkzeug (und den Mauszeiger am Bildschirm) jetzt nicht mehr. Es konnten die Elemente dort jetzt direkt berührt und manipuliert (angetippt, verschoben etc.) 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 deutlicheinfacher als nur die Steuerung mit der Maus. Es ist kein Zufall, dass der Touchscreen 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 angeschlossenen Schnittstellentechniken (wie Tastaturen, Mäuse und Bildschirm sowie auch Stromkabel etc.) befreit. Computer waren jetzt ortsunabhängig und 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, während sie zuvor vorwiegend in beruflichen Tätigkeiten und auf Schreibtischen zu finden waren. 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 der Fortentwicklung der Mensch-Maschine-Schnittstellen. Jetzt wird die Bedienung von Computern nicht nur von Tastatur und Maus befreit, sondern von jedem körperlichen Kontakt. Es sind jetzt noch nicht einmal mehr irgendwelche physischen Berührungen notwendig. Stattdessen werden einfach nur ausgesprochene Wörter zur Mensch-Maschine-Schnittstelle.

Diese Tatsache erhöht einerseits nochmals erheblich die von Smartphones bereits eingeführte Ortsunabhängigkeit der Computernutzung: Es ist nicht mehr 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-Maschine-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 Bedarfs eines physischen Kontakts zum Computer ist Sprache aber auch die natürlichste Art der Kommunikation, die von jedem Menschen intuitiv genutzt wird. Sie befreit von der Notwendigkeit, für die Nutzung von Computern erst eine spezielle Fähigkeit (Lochkarten lesen, Tastatur bedienen, Maus oder Gesten) zu erlernen, um mit diesen zu kommunizieren. Zwar ist die Entwicklung von Sprachschnittstellen noch nicht am Endpunkt echter menschlicher Kommunikation angekommen und hat heutzutage durchaus vielfältige Einschränkungen, aber Sprachassistenten kommen schrittweise (und schnell) einer natürlich-sprachlichen Kommunikation immer näher – gerade in Zeiten generativer KI.

Die Folge aus diesen beiden eben dargestellten Entwicklungen: Die Nutzung von Computern dringt nach der letzten Welle der Smartphone-Verbreitung 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 Mikrofone 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 – kann 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 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 (Sprach-) Dialog vermittelbar ist.

In diesem Sinn sollten Unternehmen auch den Einsatz von Sprachassistenten planen: Kundenservices, Marketing-Anwendungen und auch eine Gerätebedienung, die mithilfe von Sprache vorteilhaft 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 werden. 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 Computercodes umsetzen, Befehle auslösen und teils einfache 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 Lichtanschalten 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 Marketing-Anwendungen 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 ihren 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-App-Stores). 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 speaker“, also dem Amazon Echo, dem Google Home oder ähnlichen Geräten anderer Anbieter. Sie werden als meist runde und verschieden große Hardware verkauft, die zu Hause oder auch im Büro aufgestellt wird. 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 ermöglichen. 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 der Smart Speaker bzw. der 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 „Hey Google“ heißen (und vom Nutzer auch auf andere Worte umgestellt werden können bzw. ähnliche Aktivierungsworte auch erkannt werden). 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 über das Internet dorthin 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 in einem zentralen Rechenzentrum voraussetzen.

Die Software in der Cloud (dem Rechenzentrum) des Sprachassistentenanbieters trennt schließlich mithilfe komplexer Algorithmen und Techniken des maschinellen Lernens und künstlicher Intelligenz 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: schon 2015 heißt es „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). Heutzutage liegt dieser Wert deutlich höher.

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

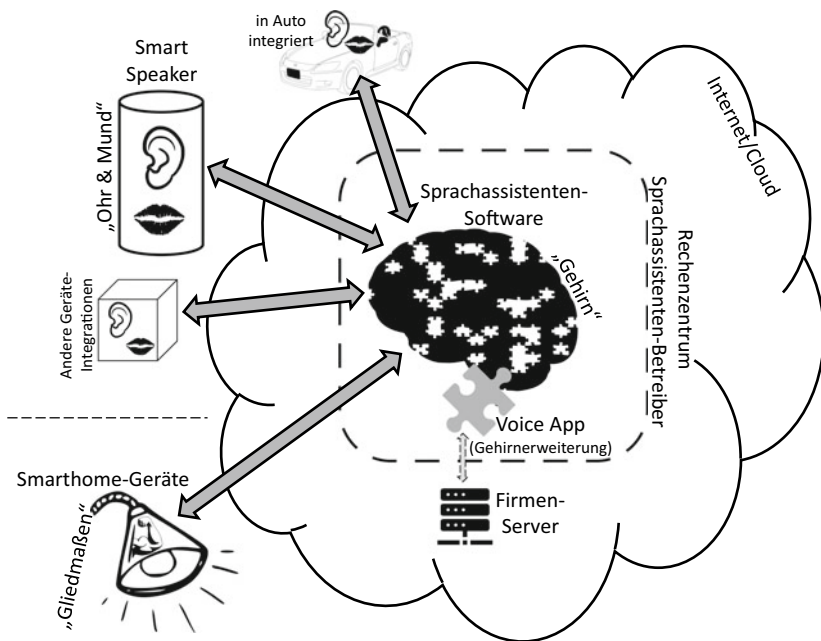


Abb. 1.2 So funktionieren Sprachassistenten – Prinzipskizze für Techniklaien

viel mehr Wissen zur Sprache und gegebenenfalls sogar zu kulturellen Hintergründen braucht. Analog ist – stark vereinfacht gesagt – die Herausforderung für Sprachassistenten, einerseits die Worte aus der Audioaufnahme zu extrahieren, andererseits den Inhalt dieser Worte auch 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.: „O. k., 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: „O. k., 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: Es wird die zur Absicht passende Gehirnregion von Alexa & Co gesucht und aktiviert. Ist der Sprachassistent direkt nach einer Antwort gefragt, greift er dafür 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.

Interessant und wichtig ist für Unternehmen und das Marketing an dieser Stelle aber: 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 eigenen Voice-App-Software also quasi ein zusätzlicher Teil des Gehirns der Sprachassistenten, der im passenden Moment (insbesondere wenn er direkt angesprochen wird) aktiviert wird.

Nach ihrer 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