Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

M-Commerce Mobile-Commerce

S-Commerce Social-Commerce

SPSS Statistical Package for the Social Sciences

AMOS Analysis of Moment Strucutres

TAM Technology Acceptance Model

SDK Software Development Kit

NLP Natural Language Processing

LSTM Long Short-Term Memory

C-Commerce Conversational Commerce

RNN Recurrent Neural Network

DNN Deep Neural Network

ASK Alexa Skills Kit

G2P Grapheme-to-Phoneme

CNN Convolutional Neural Network

SSML Speech Synthesis Markup Language

API Application Programming Interface

TTS Text to Speech

STT Speech to Text

SaR Speaker Recognition

SeR Speech Recognition

HMM Hidden Markov Model

VA Virtual Assistant

GMM Gaussian Mixture Model

CEC Constant Error Carousel

NN Neural Network

TFIDF Term Frequency-inverse Document Frequency

TBRU Transition Based Recurrent Unit

IFTTT If This, Then That

AWS Amazon Web Services

ASK Alexa Skills Kit

EM Expectation-Maximization

MS Multi-Style

Symbolverzeichnis

α	Cronbachs Alpha
δ	Effektstärke
$\bar{r}$	durchschnittlicher Korrelationskoeffizient
X^2	Modellpassung
r	Korrelationskoeffizient

1. Smart Speaker als neue Technologie im E-Commerce

Smart Speaker sind eine neue Technologie im E-Commerce, welche in den letzten Jahren eine rasante Verbreitung gefunden hat. Wie es zu dieser Entwicklung kam und welche Untersuchung angedacht ist, wird im Folgenden dargestellt.

1.1. Problemstellung und Zielsetzung

Ansätze und Methoden zur Mensch-Maschine-Interaktion, welche auf der Verarbeitung natürlicher Sprache (vgl. Natural Language Processing (NLP)) aufbauen, existieren bereits seit den 60er Jahren des vergangenen Jahrhunderts (Dale 2016, S. 812-815). Ein bekanntes Anwendungsbeispiel sind die sogenannten „Chatbots“. Diese waren bereits in den 90er Jahren ein Thema, ihren Entwicklungen wurde jedoch zwischenzeitlich kaum Beachtung geschenkt. Durch neue technologische Entwicklungen und Veränderungen im menschlichen Verhalten sind diese jedoch in verschiedenen Formen seit 2016 erneut auf dem Vormarsch (Dale 2016, S. 811, 812).

Ein Grund für diese Veränderung ist die weiter anhaltende Digitalisierung, welche zu einer vermehrten Nutzung des Internets durch die Allgemeinheit führte (van Eeuwen 2017, S. 2). Durch diese Entwicklung entstanden neue Methoden, Anwendungen und Geschäftsmodelle für den E-Commerce. Die in den letzten Jahren stark wachsenden Verbreitung von mobilen Endgeräten und steigende Nutzung von Instant-Messengern führte zu einem veränderten Verhalten der Menschen (van Bruggen et al. 2010, S. 333). So verwenden 81% der rund 62 Millionen Internetnutzer in Deutschland ihr Handy bzw. Smartphone als bevorzugtes Gerät. Desktop-Computer mit 65% sowie Laptops und Netbooks mit 69% wurden als meist genutztes Gerät abgelöst und werden dementsprechend auch immer weniger für den Online-Handel genutzt (Statistisches Bundesamt 2016). Durch diese Entwicklung ist auch ein verändertes Verbraucherverhalten zu erkennen (van Eeuwen 2017, S. 2; van Bruggen et al. 2010, S. 333). Kunden wollen wieder vermehrt in einem Dialog mit Unternehmen stehen, anstatt reine Informationen über Produkte zu erhalten. In Verbindung mit Weiterentwicklungen in der Verarbeitung natürlicher Sprache stehen Unternehmen neue Wege bereit, um mit ihren Kunden in Kontakt treten zu können (Dale 2016, S. 815).

Die Wahl, welche Art der Kommunikation Unternehmen nutzen sollten, ist dabei nicht eindeutig. Es existieren verschiedenen Formen, wie z. B. autonome Roboter (Panasonic Corporation 2018), Chatbots auf verschiedenen Plattformen (Facebook 2016) oder Smart Speaker. Eine ansteigende Verbreitung und erste Untersuchungen belegen ein anhaltendes Verlangen nach letzterer Technologie (Levin und Lowitz 2017; Buvat et al. 2018).

Welche Faktoren beim Einsatz im E-Commerce oder bei der Entwicklung von entsprechenden Anwendungen relevant sind, ist aktuell nicht belegt. Aus diesem Grund ist Ziel dieser Arbeit erste Aussagen über Faktoren und deren Relevanz für die Nutzung von Smart Speakern im Einsatz für den E-Commerce zu treffen. Auf diese Weise kann z. B. eine Grundlage für die spätere Ableitung von Erfolgsfaktoren gelegt werden. Die folgende Forschungsfrage gilt es in diesem Zusammenhang zu beantworten:

- Welche Faktoren sind für den Einsatz von Smart Speakern im E-Commerce wichtig?

1.2. Methodik

Nachfolgend wird dargestellt, welche Überlegungen bei der Auswahl eines geeigneten Modells zur Analyse der Akzeptanz gemacht wurden und wie die weitere Datensammlung aussehen wird.

1.2.1. Modellauswahl

Grundlage für die Identifikation relevanter Faktoren bildet eine qualitative Untersuchung von Smart Speakern und ihrer Anwendung im E-Commerce mit Hilfe einer Akzeptanzanalyse. Bei dieser Methode handelt es sich um eine empirische Fallstudie, welche mit Hilfe der Grundlagen des Technology Acceptance Model (TAM) nach Davis (1989) durchgeführt wird. Alternativ denkbar wäre auch die Nutzung der Grounded Theory-Methode (Glaser und Straus 1967) oder anderer Akzeptanzmodelle gewesen. Auf diese Weise wäre es z. B. möglich gewesen neue Theorien über einzelne Sachverhalte von Smart Speakern durch intensive Beobachtung der einzelnen Sachverhalte, zu gewinnen. Aufgrund der aktuell nur wenigen Anwendungen von Smart Speakern im E-Commerce und des sich schnell verändernden Umfelds, wurde eine derartige Untersuchung jedoch nicht angewendet. Auch die Verwendung anderer Modelle, wie z. B. das Prozessakzep-

tanzmodell nach Müllerleile et al. (2015), wurde in Erwägung gezogen. Eine Übertragung des Modells auf den Anwendungsfall der Smart Speaker im E-Commerce, erwies sich in einer ersten Betrachtung als schwierig, da nicht eindeutig war, wie die benötigten Daten erzielt werden könnten.

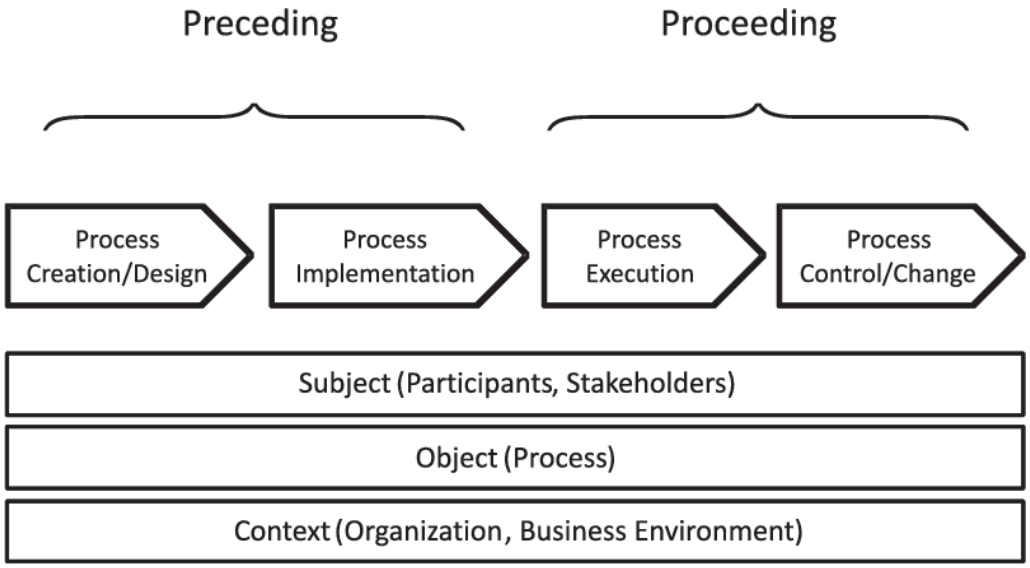

Abb. 1.1 Struktur der Prozessakzeptanz nach Müllerleile et al. (2015)

Das Ziel des TAM ist es, eine allgemeine Erklärung der Determinanten der Akzeptanz von IT-Technologien zu liefern, die in der Lage ist, das Nutzerverhalten in einem breiten Spektrum von Endbenutzer-Computing-Technologien und Benutzerpopulationen zu erklären und gleichzeitig theoretisch gerechtfertigt zu sein (Davis et al. 1989, S. 983). Es baut in seinen Annahmen auf der Theory of Reasoned Action von Ajzen (1980) auf. Nach dieser ist das Verhalten einer Person ein Resultat aus seinem eigenen Verlangen ein bestimmtes Verhalten vorzuzeigen, welches wiederum von subjektiven Normen und seiner Einstellung beeinflusst wird.

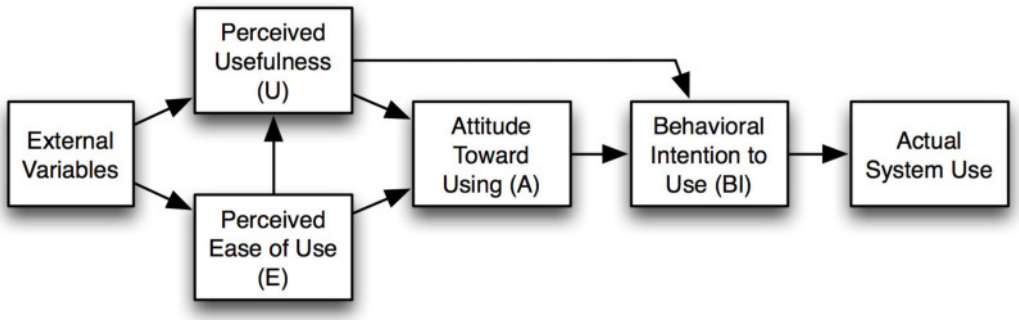

Abb. 1.2 Technology Acceptance Model (Davis et al. 1989)

Die Einschätzung, dass Usefulness und Ease of Use nach Davis (1989) die Hauptfaktoren der Akzeptanz sind, wurde in verschiedenen Studien unterstützt. Als problematisch wird jedoch angesehen, dass das TAM durch seinen Bezug zur Theory of Reasoned Action darauf aufbaut, dass Personen freiwillige eine Technologie nutzen (Brown et al. 2002). Oft kommt es jedoch dazu, dass Nutzer gezwungen sind neue Technologien zu nutzen und man in diesem Zusammenhang eher die Zufriedenheit der Nutzer mit dem System anstatt die Akzeptanz untersucht. Auch Ajzen ist bereits durch die Theory of Planned Behaviour auf dieses Problem eingegangen und hat entsprechende Anpassungen durchgeführt.

So ist ein weiterer Einflussfaktor, welcher durch die Theory of Planned Behaviour definiert wurde, die wahrgenommene Kontrolle. Sie beeinflusst dabei alle Faktoren, welche von der Theory of Reasoned Action definiert wurden (Ajzen 1991, S. 181-184). Da in dem Fall von Smart Speakern davon auszugehen ist, dass es nicht zur gezwungenen Verwendung kommt, hat die Kritik von Brown et al. (2002) keinen weiteren Einfluss auf die Untersuchung.

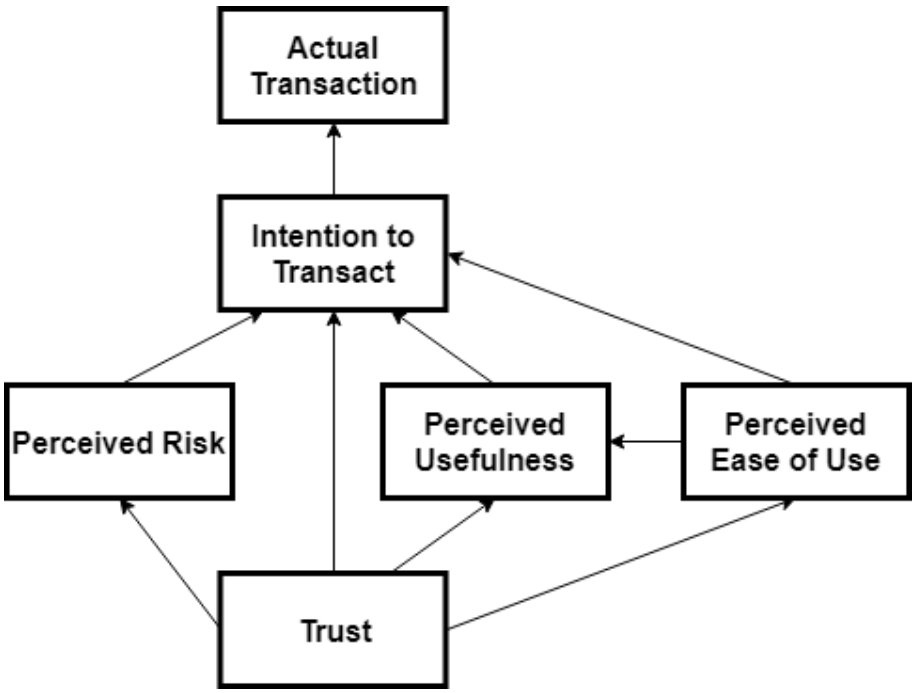

Abb. 1.3 Konzeptuelles Modell; Vgl. mit Pavlou S. 72 (2003)

Ausgehend auf dem Modell von Davis entstanden weitere Abwandlungen, welche jeweils verschiedene bestehende Probleme zu beantworten versuchten. Ein solches ist jenes konzeptuelle Modell für E-Commerce Akzeptanz von Pavlou (2003). Das TAM ist in seiner Anwendung nur für Office-Produkte ausgelegt gewesen. Pavlou passt das Modell entsprechend an, damit es auf einen allgemeinen E-Commerce-Transaktionsprozess anwendbar wird. Er erweitert jenes um die Faktoren der umweltbendingten Unsicherheit (Vertrauen

und wahrgenommenes Risiko), um das Kundenverhalten im E-Commerce genauer untersuchen zu können. Eine entsprechend auf den Transaktionsprozess mit Smart Speakern angepasste Version dieses Akzeptanzmodells wird im weiteren Verlauf der Untersuchung verwendet.

1.2.2. Datensammlung

Um für die zur weiteren Analyse benötigten externen Faktoren, welche auf Perceived Ease of Use, Perceived Usefulness, Perceived Risk und Trust wirken, umfassend abzudecken und um den zu untersuchenden Anwendungsfall genau zu definieren zu können, werden in einem State-of-the-Art die aktuellsten Lösungen und Anwendungsgebiete sowie -fälle dargestellt. Grundlage dieser Untersuchung sind Entwicklerblogs, Software Development Kits (SDKs), Whitepaper, sowie die Anwendungs-Plattformen aktueller Anbieter im Smart Speaker Markt.

Zur Identifikation der für die Akzeptananalyse benötigten externen Faktoren werden zum einen aktuelle Studien mit dem Untersuchungsziel Smart Speaker oder Virtual Assistant (VA) betrachtet und zum anderen eine Literaturanalyse nach Fettke (2006) durchgeführt. Grundlage der zweiten Untersuchung ist die Suche in akademischen Datenbanken mithilfe von Suchstrings. Als solche wurden „Smart Speaker“ und „Virtual Assistant“ zusammen mit „Akzeptanz“ und „Faktor“ verwendet. Von diesen wurden außerdem nur aktuelle Quellen[1] betrachtet, um Arbeiten auszuklammern, welche sich um Entwicklungen weit vor dem „Boom“ der Smart Speaker drehen.

Aufbauend auf den auf diese Weise identifizierten Faktoren werden entsprechende Hypothesen aufgestellt, welche Zusammenhänge zwischen Faktoren und Nutzungsverhalten von Benutzern aufdecken sollen. Ausgehend von den Hypothesen ist das Ziel Variablen und Fragen abzuleiten, welche Teilnehmer während eines Experiments, in welchem sie anhand einer Beipielanwendung die Problematik dargestellt bekommen haben, beantworten sollen. Mathematische und statistische Methoden in Kombination mit diesen Ergebnissen legen die Basis für die Diskussion der Hypothesen und der Beantwortung der Frage, welche Faktoren relevant für die Anwendung von Smart Speakern im E-Commerce sind.

[1] nicht älter als 2008, alles nach dem „Neural Winter“

1.3. Aufbau

Die Untersuchung des Themengebiets und die Ergebnisse dieser Arbeit sind dabei wie folgt aufgebaut:

In diesem ersten Kapitel wurde die Arbeitsgrundlage für die darauf folgenden Betrachtungen gelegt, indem Problemstellung, Zielsetzung und Vorgehensweise dargestellt wurden. Das zweite Kapitel wird darauf aufbauend die wichtigsten Zusammenhänge zwischen den bestehenden Technologien aufzeigen, relevante Definitionen geben und eine allgemeine Übersicht des Marktes der Smart Speaker bereitgestellt.

Im dritten Kapitel werden dann die Ergebnisse der Literaturanalysen aufgezeigt. In einem State of the Art wird dabei der aktuelle Stand der Technologie dargestellt und führende Lösungen aus verschiedenen Blickwinkeln betrachtet. Anwendungsgebiete und einzelne Anwendungsfälle werden in diesem Zusammenhang außerdem dargestellt.

Die Ergebnisse der zweiten Analyse, die für Smart Speaker relevanten Akzeptanzfaktoren, werden dann im vierten Kapitel dargestellt. An dieser Stelle wird außerdem Bezug dazu genommen, ob diese in die vom TAM definierten Faktoren übertragbar sind, oder eine Anpassung des Modells vorgenommen werden muss. Im vierten Kapitel wird dargelegt, wie die Analyse der Technologieakzeptanz von Smart Speakern durchgeführt wird. An diesem Punkt aufgezeigt wird, wie die Untersuchung entworfen wurde, auf welche Art und Weise die Erhebung der Daten stattfindet und welche Methoden genutzt wurden, um zu Erkenntnissen über die Akzeptanz von Smart Speakern zu gelangen.

Das fünfte und letzte Kapitel der Arbeit wird die Ergebnisse und Betrachtungen in einer Kurzfassung darstellen und zusammen mit der Vorgehensweise kritisch betrachten. Aufgrund der rasanten Entwicklung der Technologie wird an dieser Stelle in einem Schlusswort aufgezeigt, welche Entwicklungen in Zukunft von Smart Speakern zu erwarten sind.

2. Grundlagen, Definitionen und Einordnung des Themengebiets

Durch das relativ junge Alter der Technologie und die vielen Überschneidungen zu anderen Themengebieten, wird in diesem Kapitel dargestellt, wo Smart Speaker in den aktuellen Entwicklungen des E-Commerces einzuordnen sind.

2.1. Marktlage im E-Commerce

Der E-Commerce ermöglicht ein schnelllebiges Geschäft. Neue Technologien entstehen in immer kürzeren Abständen, begründen neue Geschäftsmodelle und ermöglichen es neuen Unternehmen bereits bestehende Strukturen leichter anzugreifen (Lee 2001, S. 349). E-Commerce hatte dabei als disruptive Innovation angefangen, ist aber mehr als nur eine neue Möglichkeit bestehende Praktiken oder Geschäftsmodelle zu unterstützen geblieben. Es handelt sich hierbei vielmehr um einen Paradigmenwechsel (Lee 2001, S.349-350), welcher mit dem Kauf und Verkauf von Waren und Leistungen über elektronische Verbindungen startete (Gabler Wirtschaftslexikon 2018a).

Begründet durch die Entwicklungen des E-Commerce entstanden Möglichkeiten auf weiteren Wegen elektronischen Handel durchzuführen. Der Social-Commerce (S-Commerce) entwickelte sich durch die ansteigenden Nutzerzahlen des Internets sowie Chancen, welche das Web 2.0 mit sich zog. Die Kunden traten wieder vermehrt in den Mittelpunkt der Unternehmen, wodurch neue Systeme zur Interaktion entwickelt wurden, wie z. B. Empfehlungs- und Bewertungsplattformen sowie Möglichkeiten zur weiteren Individualisierung von Produkten (Ickler et al. 2009, S. 51-53).

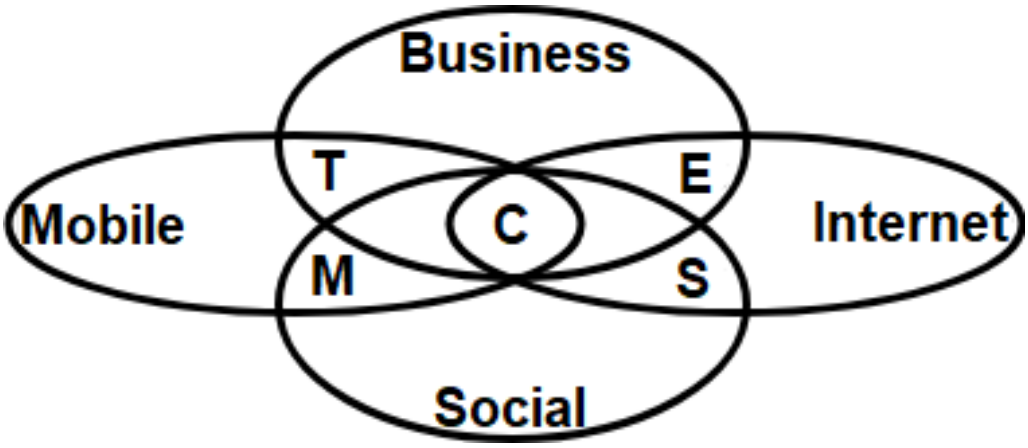

Abb. 2.1 Abgrenzung C-Commerce; in Anlehnung an Grami und Schell S. 12 (2004)

Durch die anhaltende Verbreitung von Smartphones und steigende Bedeutung dieser bei der Nutzung des Internets, entstanden neue Lösungen welche man dem Mobile-Commerce (M-Commerce) zusprechen kann. Kontext-, zeit- und standortsensitive Applikationen, sowie weitere Möglichkeiten hochpersonalisierte Angebote zu bieten, entstanden. Bei diesen handelt es sich um eine Spezialform des E-Commerce, bei dem mobile Endgeräte bei der Anbahnung, Abwicklung und Aufrechterhaltung von Leistungsaustauschprozessen mittels elektronischer Kommunikationsnetzwerke genutzt werden (Grami und Schell 2004, S. 1-3; Gabler Wirtschaftslexikon 2018b).

Eine der aktuellsten Entwicklungen in der Reihe der Technologien, welche durch E-Commerce bedingt wurden, stellt C-Commerce dar. Bei diesem handelt es sich um die Anwendung von künstlicher Intelligenz zur Kommunikation mit kommerziellen Hintergrund im Bereich des E-Commerces (van Eeuwen 2017, S. 3). In diese Entwicklung fallen die aktuellen Anwendungen von Chatbots und Smart Speakern [1].

2.2. Systemlandschaft im C-Commerce

Chatbots und Smart Speaker sind nur ein Teil der Systeme und Technologien des C-Commerce, welche man unter dem Begriff Conversational Agents zusammenfassen kann. Bei diesen handelt es sich um Softwareprogramme, welche auf die natürliche Sprache von Benutzern reagieren, diese interpretieren und mit ihr antworten (van Eeuwen 2017, S. 1-4).

Conversational Agents sind in zwei Kategorien einteilbar. Diese sind die embodied und disembodied Agents. Beide haben das Ziel, Gespräche mit Nutzern so natürlich wie möglich zu gestalten. Embodied Agents sind anthropomorphische Schnittstellen, welche animiert oder z. B. in Form von Robotern durch natürliche Sprache sowie Gestik und Mimik mit Nutzern in Konversation treten können. Diese dadurch ermöglichte Nutzung von wortlosen Kommunikationskanälen können gesellschaftliche Signale, wie Zustimmung oder Aufmerksamkeit, signalisieren und sind entsprechend hilfreich in einem sozialen Dialog (Hoffmann et al. 2009, S. 1-2). Sie finden durch diese Möglichkeiten z. B. Anwendung beim Lernen von neuen Sprachen (Wik und Hjalmarsson 2009).

Disembodied Agents sind kategorisierbar nach der Art der zu verarbeitenden Daten bzw. Information und können im Gegensatz zu den embodied Agents keine Gestik bzw. Mimik nutzen. Chatbots verarbeiten auf diese Weise die textuelle Eingabe von Nutzern, Smart

[1] T-Commerce ist die Anwendung der Technologien des E-Commerce auf die Möglichkeiten des TV (Jensen 2005)

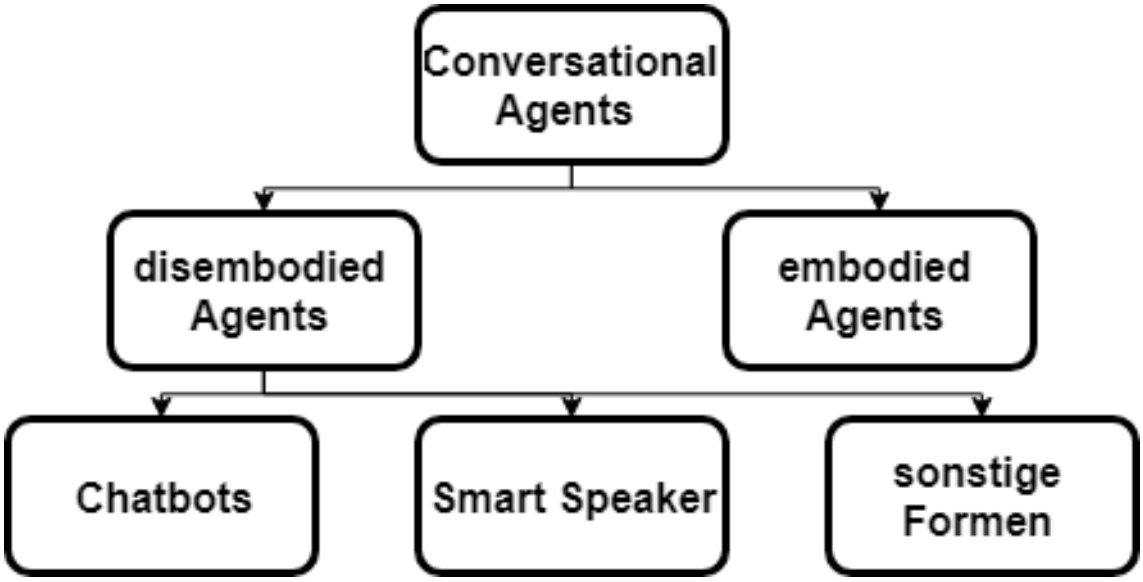

Abb. 2.2 Übersicht Conversational Agents; in Anlehnung an Radziwill und Benton (2017) S. 4

Speaker die gesprochen Sprache und zu den sonstigen Formen gehören einzelne Lösungen zur Verarbeitung anderer Daten, wie z. B. Bilder und Grafiken. Exakter sind Chatbots Softwareprogramme, welche entwickelt wurden, um eine intelligente Konversation mit einem Nutzer über dessen Eingaben zu simulieren (Mikic et al. 2009, S. 2). Eindeutige Definitionen für Smart Speaker existieren noch nicht, im Allgemeinen wird aber von einer Art Freisprecheinrichtung mit VA gesprochen, welcher bidirektionale Sprachverarbeitungstechnolgie auf Basis von Cloud-Technologien einsetzt (Martin 2017). Der VA ist dabei eine Software, welche mit Personen meist durch einen Browser interagiert und textuelle Fragen ohne menschliche Hilfe beantwortet (Duizith et al. 2004, S. 1), hier jedoch im Gegensatz zur Definition von Duizith (2004) aber auf Cloud-Technologien setzt. Im Zusammenhang dieser Arbeit wird die Definition von Smart Speakern um einen weiteren Faktor erweitert.

„Smart Speaker bezeichnen Freisprecheinrichtungen mit VA, welche bidirektionale Sprachverarbeitungstechnologien auf Basis von Cloud-Technologien anwenden und dauerhaft im eigenen Heim genutzt werden.“

Ohne diese letzte Eigenschaft wäre eine klare Abgrenzung zu der Nutzung von VAs in Smartphones nicht möglich und auch die besondere Eigenschaft von Smart Speakern, dass sie dauerthaft im eigenen Wohnort stehen, nicht ausreichend dargestellt.

Neben der Verarbeitung von Text und Sprache bestehen auch weitere Möglichkeiten. VA sind wie bereits erwähnt auch als Anwendungen in Smartphones anzutreffen und können dort neben Text und Sprache auch weitere Daten verarbeiten. Samsungs „Bixby“ bietet

entsprechend auch die natürliche Verarbeitung von Bildern an. Beispielweise Funktionen für das automatische Liefern von Information zu betrachteten Orten oder die Übersetzung von Sprache, welche im sogenannten „Viewfinder“ dargestellt wird, sind vorhanden (Samsung 2018).

2.3. Funktionsweise von Smart Speakern

Eine Definition, um was es sich bei Smart Speakern handelt wurde im vorherigen Abschnitt gegeben. Auf welche Art und Weise diese Geräte arbeiten wird nachfolgend dargestellt. Die natürliche Sprache eines Nutzers in der Nähe eines Smart Speakers wird von diesem aufgenommen und an den VA weitergeleitet. Durch diesen wird die Sprache anschließend analysiert und identifiziert, ob es sich um eine relevante Anfrage handelt. Falls nicht, werden diese Daten wieder verworfen.

Die Analyse der Sprache (auch Natural Language Processing genannt) beinhaltet bei Smart Speakern verschiedene Aufgaben. Die Speaker Recognition (SaR) hat z. B. die Erkennung der sprechenden Person als Ziel. Dies bedeutet, dass eine Anfrage, welche ein bestimmtes Schlüsselwort (z. B. „Hey Siri“) beinhaltet, auch bei mehreren gleichzeitig sprechenden Personen oder störenden Geräuschen eindeutig identifiziert und einer Person zugeordnet werden muss. Auch eine spätere Zuordnung weiterer Anfragen der selben Person fällt in diesen Aufgabenbereich (Richardson et al. 2015, S. 1671, 1672).

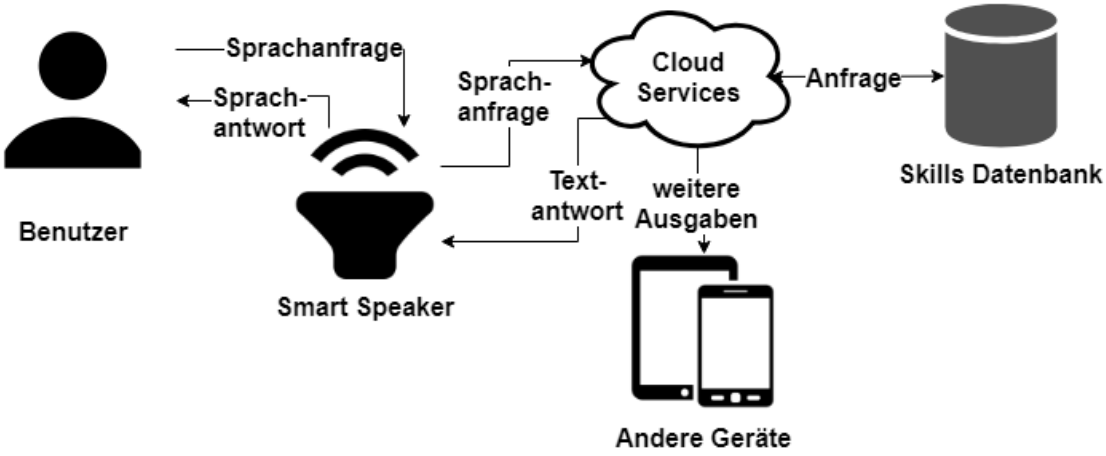

Abb. 2.3 Grobe Funktionsweise von Smart Speakern; in Anlehnung an Haack (2017) S. 2

Sind Sprecher und zugehörige Anfrage erkannt wurden, ist Aufgabe der Speech Recognition (SeR) zu untersuchen, was gesprochen wurde und diese mit Hilfe von Speech to Text (STT) in eine weiter zu verarbeitbare, textuelle Form, zu bringen (Siri Team 2018c).

Liegt diese vor, ist es Aufgabe des Natural Language Understandings aus den Daten weitere Informationen zu ziehen. Für diese Aufgabe können zwei verschiedene Vorgehensweisen genutzt werden, welche beide das Ziel haben das eigentliche Verlangen (im weiteren Verlauf als Intent beschrieben) des Nutzers zu identifizieren.

Information Extraction ist ein Teilgebiet der Wissensentdeckung und des Data Minings, welche das Ziel hat nützliche Informationen aus Texten natürlicher Sprache zu identifizieren (Gaizauskas und Wilks 1998, S. 70). Verschiedene Information Extraction Methoden existieren, im Zusammenhang mit Smart Speakern werden jedoch hauptsächlich jene genutzt, welche Schlagwörter bzw. Phrasen erkennen (Chowdhury 2003, S. 61). Diese Informationen können dann mit Hilfe von „Tagging“ als weitere Parameter in der textuellen Form der natürlichen Sprache hinzugefügt werden (Amazon.com Inc. 2018a).

Die zweite Möglichkeit, neben den Methoden der Information Extraction, ist einfacher Abgleich der Textform der Anfrage mit von Skills definierten „Invocation Names“ und „Intents“ (Amazon.com Inc. 2018a, Vgl. „Understand How Users Interact with Skills“). Eine Anfrage wie „Alexa, sag dem Lieferdienst ich möchte eine Pizza Hawaii“ enthält Beispielsweise das Schlüsselwort „Alexa“, wodurch der Smart Speaker und Prozess gestartet wird, den Invocation Name “Lieferdienst“, welcher den Skill eindeutig identifiziert und natürlich die weiteren, für den Skill relevanten Informationen „möchte eine Pizza Hawaii“, die daraufhin bearbeitet werden müssen.

Diese Verarbeitung der Daten erfolgt durch die sogenannten „Skills“. Sie erhalten eine Nachricht vom VA mit allen relevanten Information und senden eine entsprechende Antwort an diesen zurück. Neben einer einfachen Antwort, kann dies auch der Beginn eines Dialogs oder die Authentifizierung in einem anderen System sein sowie Daten für andere Geräte enthalten (Amazon.com Inc. 2018a, Vgl. „Create Intents, Utterances, and Slots“). Zu den in Grafik 2.3 dargestellten weiteren Geräten gehört bei jedem Smart Speaker mindestens ein Smart Phone. Dieses wird zur Einrichtung benötigt und dient als weiteres Ausgabemedium von Informationen für den Lautsprecher bei der Ausgabe von Antworten.

Handelt es sich bei dieser eine einfache Antwort, wird sie nur an den Smart Speaker weitergeleitet, wo es dann die Aufgabe der Speech Generation ist, die Daten in natürliche Sprache umzuwandeln. Dies erfolgt mit Hilfe von Text to Speech (TTS)-Methoden, wobei nach Ausgabe der Nachricht, im Normalfall, der Prozess beendet wird. Für Dialoge oder weitere Authentifizierungen geht der Smart Speaker in einen Wartezustand, um so die Antwort des Nutzers abzuwarten und im Kontext der selben Session bearbeiten zu können (Haack et al. 2017, S. 1, 2).

2.4. Technologien im Einsatz

Verschiedene Aufgaben wurden im vorigen Absatz aufgezeigt. Welche Technologien bei diesen zum Einsatz kommen, wird als nächstes dargelegt. Eine der ersten Aufgaben von Smart Speakern war die SaR. Die grundlegenden Technologien, auf welche diese Aufgabe aufbaut, sind identisch mit der der SeR. NNs in verschiedenen Formen bilden seit einigen Jahren die Basis. Für eine lange Zeit kamen jedoch Hidden Markov Models (HMMs) in Kombination mit Gaussian Mixture Models (GMMs) zum Einsatz. HMMs wurden genutzt, um die Variabiliät der natürlichen Sprache zu modellieren und GMMs überprüften die Güte der so entstandenen Lösungen. Diese Modelle konnten mit Hilfe des Expectation-Maximization (EM)-Algorithmus traniert werden und erhielten so die Fähigkeit natürliche Sprache zu verstehen bzw. Sprecher zu identifizieren (Hinton et al. 2012, S. 1-2).

Erste Alternativen zu diesem Aufbau gab es bereits 1993 mit der Anwendung von flachen NNs (Bourlard und Morgan 1993). Die Rechenleistung und Algorithmen zum Anlernen des Netzwerks fehlten jedoch zu dieser Zeit, um HMM ernsthafte Konkurrenz machen zu können. Neue Fortschritte im Bereich Machine Learning und leistungsfähigere Hardware ermöglichen die effektive Nutzung von Deep Neural Networks (DNNs) zur Verarbeitung natürlicher Sprache. Bei DNNs handelt es sich um künstliche NNs mit vielen nichtlinearen Layern, welche die Darstellung von horizontalen, sowie vertikalen Verbindungen zwischen Elementen des Netzwerks ermöglichten. An DNNs existieren zwei vorwiegende Ausprägungen, das Convolutional Neural Network (CNN) und Recurrent Neural Network (RNN). RNNs besitzen die Fähigkeit von einem Elemente des n-ten Layers Rückkopplungen zum gleichen, parallelen oder vorhergehenden Elementen zu definieren. CNNs zeichnen sich hingegen dadurch aus, dass sie, inspiriert von der Großhirnrinde, durch ein Pooling-Layer die Möglichkeit besitzen, nicht relevante Informationen zu filtern und sich diese Filterung zu merken (Bengio 2009, S. 82-84; Hinton et al. 2012, S. 1-4, 16-18).

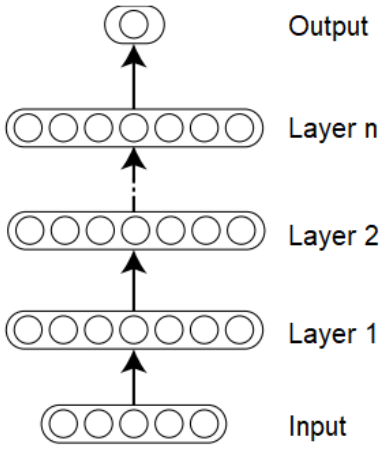

Abb. 2.4 Multi-Layer NN; in Anlehnung an Bengio (2009) S. 17

Long Short-Term Memory (LSTM) RNN) bilden die neuste Anwendung von NNs. Die grundlegende Idee hierbei ist, dass neben normalen auch „Constant Error Carousel (CEC)“ Elemente existieren. Diese CEC Elemente besitzen eine Aktivierungsfunktion, welche auf sie selbst gerichtet ist und ein Gewicht von 1,0 besitzt. Durch dieses Gewicht ist es möglich, dass erkannte Fehler unverändert im Netz sich rückwärts verbreiten und so deren Vermeidung gelernt wird. Angebunden sind die CECs an mehrere anpassungsfähige, nicht lineare Elemente. Durch diesen Aufbau ist es möglich, dass das LSTM Netzwerk die Bedeutung von Ereignissen identifizieren und sich merken kann (Schmidhuber 2015, S. 86-89, 95).

Diese Fähigkeiten werden genutzt, um innerhalb des NN zu lernen, welche Klangfolgen auf welche Wörter passen, damit eine für die Sprache sinnvolle Wortfolge entsteht. Dieser Prozess wird Language Modelling genannt und ist notwendig, damit neben der Sprache an sich auch Nuancen dieser richtig verstanden werden (Beaufays 2015) [2].

Neben der Erkennung des Sprechers und der Umformung der natürlichen Sprache existieren verschiedene Auszeichnungssprachen, welche helfen die in Textform vorliegenden Daten mit weiteren Informationen auszuschmücken. VoiceXML ist eine XML-basierte Sprache, welche Hilfestellungen bei der Sprachsynthese, -digitalisierung, -aufnahme, -verständnis und vielen weiteren Aspekten interaktiver Dialoge bietet (W3C 2004). Weitere Standards, wie z. B. Speech Synthesis Markup Language (SSML), für die Auszeichnung und Unterstützung der Sprachsynthese, existieren aufbauend auf dieser (W3C 2010).

Nach Verarbeitung durch entsprechende Skills werden die Antworten vom VA an den Smart Speaker gesendet. Dafür müssen die in textueller Form vorliegenden Nachrichten durch TTS-Synthese in für den Lautsprecher ausgebbare Sprache umgewandelt werden. Für diese Aufgabe der Speech Generation existieren aktuell drei verbreitete Methoden: Unit Selection und Parametric Synthesis sowie Mischformen dieser beiden Methoden wie z. B. Waveform Modeling. Unit Selection nutzt eine große Datenbank mit Sprachelementen, den sogenannten „Units“. Diese können mehrfach mit verschiedener Intonation oder spektraler Charakteristik vorliegen. Je nach Größe dieser Datenbank ist eine qualitativ hohe Synthese erzielbar, weswegen Unit Selection eine weite Verbreitung erfährt (Siri Team 2018a; Hunt und Black 1996, S. 71-96).

Bei Parametric Synthesis handelt es sich um eine Methode zur Sprachsynthese, welche auf HMM aufbaut. Es existiert auch hier eine Datenbank mit Sprachelementen, jedoch wird die Sprache durch ein trainiertes Model als Durchschnitt von ähnlich klingenden

[2] Auf diese Weise lernt das NN auch Spracheingaben wie „Wir essen jetzt, Opa!“, richtig zu verstehen und nicht falsche Schlüsse zu ziehen.

Sprachsegmenten gebildet. Parametric Synthesis erreicht dadurch nicht die hohe Qualität von Unit Selection, erstellt jedoch fließendere und leichter verständliche Sprache (Zen 2009).

Durch neuste Fortschritte in Deep Learning konnten mit Waveform Modeling qualitativ hochwertige Snythesen ähnlich zu Unit Selection, bei gleichzeitiger Flexibilität der Parametric Synthesis erzielt werden. Um dies zu erreichen wird direkt auf rohen Audio-Waveforms gearbeitet. Zur Tonerzeugung wird eine Architektur genutzt, welche auf Dilated Casual Convolutions baut. Diese sind eine besondere Art linearer Filter, welche unabhängig von zukünftigen Eingabedaten, sondern nur ausgehend von historischen Daten, eine Ausgabe erzielten (van den Oord et al. 2016, S. 1-2; Siri Team 2018a).

3. State of the Art im Bereich der Smart Speaker

Die durch die Entwicklung von Smart Speakern neu entstandenen Märkt für Geräte und dazugehörende Skills stehen in ihren ersten Jahren. Amazon hält mit seiner Lösung den größten Anteil im Markt der Geräte, jedoch versuchen viele Unternehmen diese Position mit ihrem Produkt anzugreifen.

3.1. The Big Four

Anhand der vier größten Unternehmen im Markt (aktueller Marktanteil oder allgemeine Unternehmensgröße) und ihrer Smart Speaker-Lösungen wird nachfolgend, ausgehend von den Technologien zugrundeliegenden VAs, ein Überblick gegeben, wie sich aktuelle Lage um Smart Speaker gestaltet.

3.1.1. Amazon Alexa

Amazons Smart Speaker-Lösung besteht nutzt den VA „Alexa“, welcher in verschiedenen hauseigenen Geräten wie z. B. dem Echo, Echo Dot und Echo Show integriert ist. Je nach Gerät bestehen diese aus Hochtonlautsprecher und Mikrofon, sowie optionale Woofer oder Displays. Ein ähnlicher Hardwareaufbau ist auch bei der Konkurrenz zu erkennen, wobei geringe Unterschiede in Anzahl der Lautsprecher und Mikrofone bestehen.

Mit einem Marktanteil von ca. 70% [1] steht Amazon mit seiner Smart Speaker-Lösung an der Spitze der Anbieter. Dieser Umstand ist zum Teil dadurch bedingt, dass Amazon als erstes Unternehmen das Potenzial erkannte und bereits 2015 Echo auf den Markt brachten.

Alexa nutzt in der allgemeinen Funktionsweise und an Technologien einen ähnlichen Aufbau wie in Kapitel 2.3 und 2.4 dargestellt. Der Smart Speaker sendet die durch das Mikrofon aufgenommenen Geräusche an Alexa weiter. Zur Identifikation des Sprechers und dem Inhalt der Sprachanfrage kommen zwei LSTM RNNs zum Einsatz. Als erstes wird ein LSTM Encoder genutzt, um das sogenannte „Wake Word“, im Normalfall „Alexa“, zu identifizieren. Darauf aufbauend existiert ein LSTM Decoder, welcher die restlichen

[1] verschiedene Zahlen existieren, wie z. B. 69,1% nach TrendForce (2018) oder 70% nach Routley (2018)

Sprachmerkmale der Anfrage identifiziert. Dabei wird zum einen der Intent identifiziert und weitere Merkmale der Anfrage mit diesem zusammen in einem Vektor in textueller Form gespeichert (Strom 2017, S. 16-22).

Als Endpunkte von Sprachanfragen an Alexa stehen „Skills“ zur Verfügung. Diese werden zum einen von Amazon bereitgestellt, zum anderen können sie aber auch mit Hilfe des Alexa Skills Kit (ASK), in den Sprachen Java, C#, Python, Go und NodeJS, selbst erstellt werden. Dem Entwickler ist es möglich aus verschiedenen Vorlagen als Hilfestellung zu wählen oder frei von Vorgaben individuelle Skills zu entwickeln. Eine Anbindung von eigenen Webseiten und -services ist durch das sogenannte „Account-Linking“ außerdem möglich. Aktuell existieren mehr als 25000 Skills in verschiedenen Sprachen mit einer steigenden Tendenz (Amazon.com Inc. 2018a).

Abb. 3.1 Amazon Echo (Amazon.com Inc. 2018b)

Ist eine Sprachanfrage durch einen Skill verarbeitet, werden mit Hilfe von Deep Learning die entstandenen Antworten in Sprache umgewandelt. In einem ersten Schritt wird der Text normalisiert. Dies bedeutet, dass z. B. €20 in „zwanzig euro“ für die weitere Verarbeitung umgewandelt werden. Darauf aufbauend wird Grapheme-to-Phoneme (G2P) Konversion angewendet. Diese wandelt die textuelle Form in eine entsprechende Darstellung der Lautform dieser um (Thu et al. 2016). Aufbauend auf diesen Daten kann eine Waveform generiert werden, welche dann als Sprache vom Smart Speaker ausgegeben wird (Amazon.com Inc. 2018a, S. 15)

3.1.2. Microsoft Cortana

Microsofts VA-Lösung trägt den Namen „Cortana“ und ist aktuell Bestandteil jedes PCs mit dem Betriebssystem Windows 10. Die Smart Speaker Variante „Harman Kardon Invoke“, welche mit diesem VA von Haus aus geliefert wird, ist seit Oktober 2017 in den Vereinigten Staaten von Amerika exklusiv erhältlich.

Der späte Einstieg und exklusive Vertrieb in den USA sind mögliche Gründe für die bisher nur geringe Verbreitung des Smart Speakers von Microsoft. Mit 1,3% Marktanteil in 2017 ist nur ein Bruchteil der Abdeckung, verglichen mit Konkurrenten, erreicht worden (Kinsella und Mutchler 2018). Aufgrund der Verbreitung von Cortana in weiteren Windows 10 Geräten und den Möglichkeiten diese untereinander zu integrieren, ist eine Untersuchung der Möglichkeiten dieser Lösung dennoch angebracht.

Abb. 3.2 Harman Kardon Invoke (Microsoft Corporation 2018a)

Bei der Verarbeitung der Sprachanfragen setzt Microsoft mit Cortana auf CNN komibiniert mit bidirektionalen LSTMs. Trainiert wurden diese Netzwerke mit Hilfe des Microsoft Cognitive Toolkit 2.1, welches bei der Modellerstellung und Optimierung der Parameter Unterstützung bot. Durch die Nutzung dieser Technologien konnte die Fehlerrate bei der Verarbeitung natürlicher Sprache auf menschenähnliche 5,1% reduziert werden (Huang 2017; Xiong et al. 2017a, S. 1-2). Weitere Informationen werden der Textform von Anfragen mit SSML hinzugefügt (Microsoft Corporation 2018b).

Nach einer derartigen Aufbereitung der Sprachanfrage und Identifikation des Intents, werden entsprechende Anfragen an „Bots“ in der Microsoft Azure Cloud gesendet. Aktuell existieren für Cortana ca. 240 Bots, speziell für die Smart Speaker-Lösung Kardon Invoke sind knapp 100 dieser ausgelegt [2]. Auch hier existieren zum einen von Microsoft bereitgestellte Bots oder die Möglichkeit als externe Person mit Hilfe des Cortana Skills Kit eigene Fähigkeiten für den VA zu erstellen. Als mögliche Programmiersprachen stehen aktuell C# und NodeJS zu Verfügung. Auch hier ist es durch das Connected Account Feature möglich, wie bereits bei Alexa, Cortana auf anderen Webseiten und -services zu authentifizieren, um so ein breiteres Spektrum an Funktionen bereitzustellen.

3.1.3. Apple Siri

Apple war einer der Vorreiter bei dem Thema VAs, als sie bereits im Oktober 2011 ihre Lösung mit dem Namen „Siri“, integriert im iPhone 4S, auf den Markt brachten.

Die Smart Speaker-Lösung „HomePod“, in welcher Siri integriert ist, wird seit Febraur 2018 vertrieben und hat einen dementsprechend bisher noch vernachlässigbar geringen Marktanteil unter den verfügbaren Geräten (Kinsella und Mutchler 2018).

Abb. 3.3 Apple HomePod (Apple Inc. 2018b)

In seiner Funktionsweise nutzt Siri bereits besprochene Technologien. Auf das Schlagwort „Hey Siri“ hört ein Scanner, welcher auf diese Phrase wartet und sie nutzerunabhängig abhört. Nutzerabhängig wird dann anhand der restlichen Sprachanfrage weitergearbeitet. D. h., es wird, wie bei Cortana, ein Vektor erstellt, welcher mit Informationen ausgezeichnet wird. Bei diesen kann es z.B. sich um Informationen über die Umgebung oder wie die Tonlage des Sprechers war, handeln (Siri Team 2018c).

Für Spracherkennung kommen erneut RNNs zum Einsatz, welche nach der Multi-Style (MS) Methode und mit Hilfe von Curriculum Learning trainiert wurden (Siri Team 2018c). Bei Curriculum Learning handelt es sich um eine Methode des maschinellen Lernens, bei dem das Training hoch organisiert ist und

[2] Stand 03.05.2018

auf einem Bildungssystem ähnlichen Curriculum aufbaut (Bengio et al. 2009, S. 1). Liegen die Daten in textueller Form vor, müssen diese weiter normalisiert werden, um z. B. den Intent eindeutig identifizieren zu können. Dabei müssen Passagen wie „dreiundzwanzigster Oktober Zweitausendachtzehn“ in „23.10.2018“ umgewandelt (Siri Team 2018b). Anfragen, welche diese Schritte durchlaufen haben, werden dann an eine passende Anwendung weitergeleitet und bearbeitet. Eine genaue Anzahl an verfügbaren Skills für den Smart Speaker ist noch nicht bekannt, da eine passende Übersicht (z. B. in Form eines Stores) nicht existiert. Durch das SiriKit ist es möglich geworden, Apps zu erstellen oder bestehende so anzupassen, dass sie durch Siri steuerbar werden. Zur Verfügung stehen einem dafür die IDE Xcode und die Programmiersprachen Swift sowie Object-C. Mit Hilfe dieser ist es z. B. möglich VoIP Anrufe, Zahlung oder die Verwaltung von Listen zu implementieren. Funktionen zur Authentifizierung, wie sie mit Alexa oder Cortana möglich sind, bestehen aktuell nicht. Auch die Steuerung von Smart Home Geräten ist eingeschränkt. Aktuell werden nur Geräte unterstützt, welche mit dem HomeKit kompatibel sind. Es handelt sich dabei um viele verschiedene Geräte, von Lichtern bis Garagenöffnern (Apple Inc. 2018a).

Der TTS-Prozess von Siri entspricht grob der allgemeinen Vorgehensweise. In einem Textverarbeitungsprozess wird dieser analysiert und durch ein Intonationsmodell diesem eine entsprechende Satzmelodie hinzugefügt. In einem zweiten Signalverarbeitungsprozess werden dann aus einer Einzelstückedatenbank Elemente gewählt und zu einer Waveform aneinandergereiht. Diese wird dann als Sprache ausgegeben (Siri Team 2018a).

3.1.4. Google Assistant

Googles VA-Lösung „Google Assistant“ debütierte 2016 in der Instant-Messenger-App Allo und später im gleichen Jahr in dem Smart Speaker Google Home. Neben dem hauseigenen Google Home existieren Smart Speaker von Harman, Sony, Panasonic und vieler weiterer Anbieter, welche den Assistant integrieren. Über das Google Assistant SDK for Devices ist es außerdem möglich, den VA in alle Geräte mit einer linux-armv7l oder linux-x86_64 Architektur zu integrieren.

Mit einem aktuellen Marktanteil zwischen 21% und 25% [3] Beginn 2018 ist Google der zweitgrößte Anbieter auf dem Markt und konnte seinen Anteil im Gegensatz zu der Konkurrenz am meisten vergrößern. Ein weiterer Anstieg des Marktanteils wird aktuell prognostiziert (TrendForce 2018; Routley 2018).

Googles Sprachverarbeitung begann mit „Voice“, welches 2009 aufbauend auf GMMs arbeitete. Seit 2012 wurde auf LSTM RNNs umgestellt, welche auch die Grundlage für den Google Assistant darstellt. Antrainiert wurden NNs, welche zum einen akustische und zum anderen sprachliche Umwandlungen durchführen können (Beaufays 2015).

Für das eigentliche Verständnis der in textuelle Form umgewandelten Sprachanfragen kommen neben Methoden, welche aus Information Retrieval bekannt sind, wie Term Frequency-inverse Document Frequency (TFIDF) und Googles „SyntaxNet“ zum Einsatz. Dies arbeitet mit Hilfe von sogenannten Transition Based Recurrent Units (TBRUs), eine Kombination bestehend aus RNNs und Transition Systems. Diese Technologie zeichnet sich dadurch aus, dass sie es erlaubt dynamisch neue Verbindungen im Netzwerk zu erstellen, abhängig von Aktivierungen einzelner Elemente. Besonders bei der bei Syntaxanalyse liefert diese Vorgehensweise, verglichen mit anderen Methoden, besonders gute Ergebnisse (Kong et al. 2017). Eine Auszeichnung der textuellen Form der Sprachanfrage wird auch an dieser Stelle mit Hilfe von SSML durchgeführt (Weiss und Petrov 2017; Petrov 2016).

Abb. 3.4 Google Home (Google LLC 2018a)

Verarbeitet werden die Nutzeranfragen bei Google von den sogenannten „Actions“. Diese definieren Intents, worauf reagiert wird und Fulfillments, die wiederum die Logik enthalten, wie reagiert wird. Zur Entwicklung eigener Actions steht Nutzern die Sprache NodeJS zur Verfügung. Aktuell rühmt sich der Google Assistant damit, über 1 Millionen Actions Nutzern zur Verfügung zu stellen[4]. Eine genaue Anzahl, welche dieser für den

[3] nach Trendforce 21,4% 2018 und VisualCapitalist 25% 2018
[4] nach Suchleiste auf https://assistant.google.com/explore?hl=en_us

Google Home Smart Speaker ausgelegt sind, existiert nicht, einzelne Seiten berufen sich jedoch darauf, dass es ca. 1830 seien sollen (Segan 2018). Eine Authentifizierung von Actions mit anderen Services ist, wie es auch bei der Konkurrenz unterstützt wird, möglich. Unterschiede bestehen jedoch darin, dass diese bereits bei der Installation durchzuführen ist, nicht erst bei Ausführung der Action bzw. des Skills (Google LLC 2018b).

3.2. Skill-Stores und Auffindbarkeit

Smart Speaker haben eine rasante Verbreitung erfahren (IDC 2018). Passende Möglichkeiten neue Fähigkeiten für diese Geräte zu entwickeln und diese zu finden, wurden von einigen Anbietern bereitgestellt. Cortana, Alexa und Assistant besitzen eigene „Läden“, in welchen man Anwendungen für verschiedene Aufgaben finden kann. Siri hängt an dieser Stelle noch hinterher und besitzt aktuell [5] nur eine Übersicht, welche Möglichkeiten bereitgestellt werden.

Was in den einzelnen Läden dargestellt wird unterscheidet sich auch drastisch von Anbieter zu Anbieter. Z. B. fehlen Cortanas Lösung eine Suchfunktion, sowie die Möglichkeit Bewertungen von Nutzern einzusehen oder eigene hinzuzufügen. Auch eine Übersicht, wieviele Personen eine bestimmte Anwendung bereits genutzt haben ist nicht vorhanden. Diese Information fehlt auch in den Läden von Alexa und Assistant, hier jedoch sind Suche und Bewertungsfunktionen vorhanden. Als Besonderheit gilt außerdem, dass Google auch hauseigene Anwendungen in seinem Store anbietet und somit die Möglichkeit bereitstellt auch diese zu löschen. Amazon hingegen verbirgt diese und weitere von Haus aus integrierte third party Apps, wie z. B. Spotify, wodurch diese nicht löschbar sind.

Ein einheitliches Vorgehen bzw. Standard hat sich entsprechend noch nicht etabliert. Dies könnte einer der Gründe sein, weswegen eine Studie von voicebot.ai mit 1057 befragten US-Amerikanern als Ergebnis zog, dass 48,2% der Smart Speaker Nutzer keine third party App nutzen bzw. nie eine weitere Funktion ihren Geräten hinzugefügt haben. Als weitere Wege, wie sie an neue Anwendungen kommen sind aktuell die Läden mit 17% noch hinter Empfehlungen von Freunden mit 22,5% nur die zweitwichtigste Möglichkeit (Kinsella und Mutchler 2018, S. 24).

Entgegen diesen Problemen, vor denen die Läden der Smart Speaker Anbieter stehen, hinterlassen eine überraschend hohe Anzahl an Personen, die eine Third-Party-App genutzt haben, eine Rezension zu dieser. 11% der Anwender nutzen diese Funktion, welche

[5] Stand. 07.06.2018, vgl. https://www.apple.com/ios/siri/

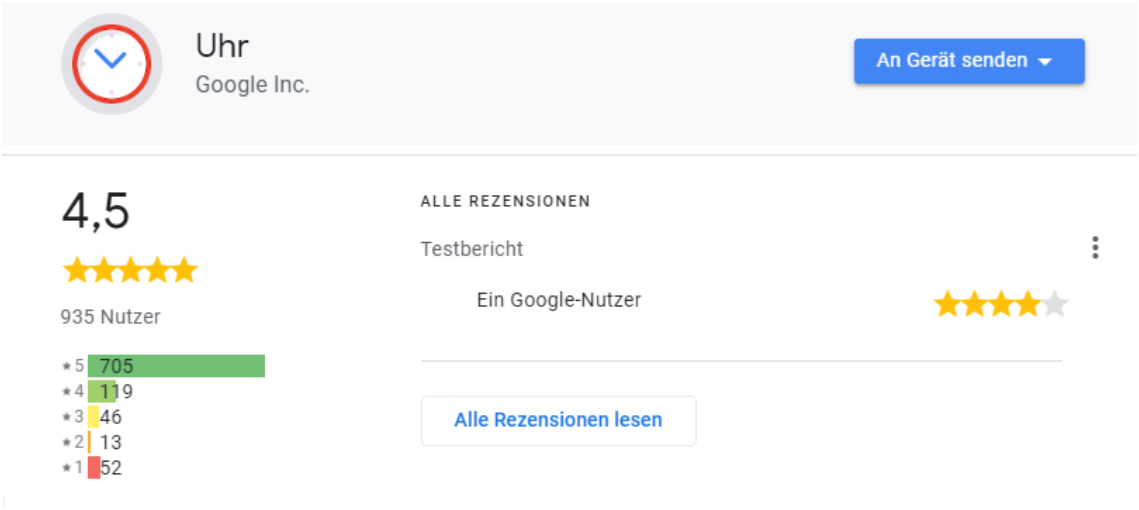

Abb. 3.5 Verkürzte Übersicht einer Anwendung bei Google Assistant Discovery, vgl. https://assistant.google.com/services/a/uid/0000005d8a63d90c

verglichen mit 0,5% der Käufer von Produkten in Online-Shops und weniger als 1% von Benutzern von Mobile Apps überraschend groß scheint (Kinsella und Mutchler 2018, S. 23).

Dieser Umstand wird in den weiteren Untersuchungen genutzt, um eine Übersicht der Nutzung von Anwendungen darzustellen. Aufgrund der geringen allgemeinen Nutzung der Läden und fehlenden Nutzerzahlen von Skills, wird deshalb je Kategorie nur der Extremwert der Rezensionszahlen, d. h. die Anwendung mit den meisten Rezensionen, aufgeführt. Da diese Information gänzlich für Anwendungen von Cortana fehlt, werden außerdem nur die für Assistant und Alexa angegeben. Eine Besonderheit hierbei ist, dass es sich bei Alexa um sprachspezifische und bei dem Assistant um einen ganzheitlichen Store handelt. Dies bedeutet, dass Alexa für jede Sprache einen eigenen Store mit entsprechenden Informationen und Daten bereitstellt, der Assistant hingegen nur einen Store besitzt und für diesen je Sprache eine unterschiedliche Sicht darstellt.

Zur Kategorisierung der Anwendungen dienen zwei Studien als Grundlagen. Zum einen Capgeminis Studie von 2018 zu Conversationale Commerce (Stichprobe umfasst 2500 Smart Speaker Nutzer in Europa und den USA), zum anderen von PwCs Befragung zu Digitalen Assistenten von 2017 (Stichprobe umfasst 1012 Personen aus Deutschland). In diesen wurde untersucht, welche Funktionen, wie stark genutzt werden und wo ein allgemeines Interesse der Nutzer liegt. Wenig überraschend sind Funktionen zur Information und Wiedergabe von Musik mit jeweils 82% und 67% bei Capgemini sowie 48% und 52%

bei PwC führend unter den meist genutzten bzw. gefragten Funktionen der Befragten. Bereits 35% und 22% haben jedoch ihren Smart Speaker bereits für Einkäufe jeglicher Art genutzt bzw. würden sich durch diese Unterstützung in ähnlichen Aufgaben wünschen.

3.3. Anwendungsgebiete

Smart Speaker besitzen verschiedene Anwendungsgebiete und -fälle. Obwohl es sich grundlegend weiterhin um einen Lautsprecher handelt, konnten durch VAs die Funktionen erweitert werden. Um welche es sich hierbei handelt wird nachfolgend dargestellt.

3.3.1. Information

Mit 82% der befragten Personen war nach Capgemini (2018) das Suchen nach Informationen die Funktion, welche die meisten Nutzern von Smart Speakern mindestens ein Mal genutzt hatten. 48% bzw. 44%, laut PwC (2017), hatten den Wunsch Informationen oder Wissensfragen durch ihren Smart Speaker zu erhalten oder beantwortet zu bekommen. Das große Verlangen nach dieser Art Anwendungsfälle wird auch durch einen zweiten Teil der PwC-Studie belegt. In diesem sollten die Befragten angeben, worin sie die Vorteile von Smart Speakern sehen. Mit 46% wurde der schnelle Zugang zu Informationen als Hauptvorteil angesehen.

Anwendungsfall	Assistant-Bewertungen	Alexa-Bewertungen
Wetter	6613	79
Informationssuche	5975	116
Übersetzungen	1203	46
Nachrichten	741	68

Tab. 3.1 Übersicht Informationsanwendungen

Die jeweiligen Stores der Skillanbieter definieren für diesen Anwendungsbereich oft verschiedene Kategorien. So werden Skills für Nachrichten, Bildung und Wetter getrennt voneinander bei Alexa, Google Assistant und Cortana angeboten. Apple hat aktuell noch keinen speziellen Store für die Möglichkeiten von Siri bereitgestellt, es ist eher von ei-

ner Übersicht der von Haus aus bereitgestellten Funktionen zu reden. Diese umfasst die Kategorien „Out and About“, „Tips and Tricks“ sowie „Getting Answers“, welche verschiedene Funktionen für das Informationsverlangen des Nutzers bereitstellen.

Neben diesen für Information und Wissen dedizierten Kategorien existieren weitere Überschneidungen in anderen Bereichen. Als Beispiel wäre der „Sixt Autvermietung“[6] Skill für Alexa anzubringen. Dieser ermöglicht es dem Nutzer über den Smart Speaker einen Leihwagen zu reservieren, hat aber außerdem eine Informationsfunktion für die nächstgelegene Sixt-Station integriert. Auf diese Weise kombinieren viele Skills ihre Kernfunktion mit weiteren Informationsmöglichkeiten.

3.3.2. Unterhaltung

Die Kategorie Unterhaltung stellt eine Kombination vieler Anwendungsfälle dar, wobei das Hauptaugenmerk dieses Bereichs eindeutig auf jeglicher Wiedergabe von Musik liegt. 67% der Nutzer von Smart Speaker haben ihr Gerät bereits für diese Funktion genutzt und 52% und damit die meisten der Befragten in der Studie von PwC (2017), sehen einen konkreten Wunsch darin, Musik mit ihrem Gerät abzuspielen (Buvat et al. 2018). Diese Ergebnisse sind, trotz der vielen Möglichkeiten die ein VA bietet, verständlich, da es sich in erster Linie weiterhin um einen Lautsprecher handelt und auch die Werbung einzelner Anbieter diesen Umstand gezielt hervorhebt[7].

Anwendungsfall	Assistant-Bewertungen	Alexa-Bewertungen
Spiele (Quiz, Trivia, usw.)	3685	234
Musikstreaming	1528	483
Radio	217	106

Tab. 3.2 Übersicht Unterhaltungsanwendungen

Entsprechend dieser Kundenwünsche bestehen bei allen vier bisher besprochenen Anbietern Skill-Kategorien für die Unterhaltung des Nutzers. Alexa, Google Assistant und Cortana definieren hierbei noch eine spezielle für Musik. Die zwei Hauptanwendungen in dieser Kategorie sind zum einen das Musikstreaming und zum anderen die Wiedergabe

[6] vgl. https://alexa.amazon.de/spa/index.html#skills/dp/B076CFZK2S/?qid=1528267457
[7] Apple vermarktet den Homepod an erster Stelle anhand seines Klangs https://www.apple.com/de/homepod/

von Radiosendern. Neben diesen können auch die Anwendungen der Kategorie „Spiele“ zur Unterhaltung gezählt werden. Auf diese Weise werden die Anwendungsfälle um z. B. interaktive Hörbucher bzw. Geschichten, Quiz, Trivia, Witze usw. erweitert.

3.3.3. Smart Home

31% der Befragten in der Studie von Capgemini (2018) hatten ihren Smart Speaker bereits in Kombination mit Smart Home Geräten genutzt. 23% definierten dies laut PwC (2017) als konkreten Anwendungswunsch. Die jeweiligen Anbieter ermöglichen dabei den Zugriff zu unterschiedlichen Geräten. Die Stores von Alexa, Google Assistant und Cortana besitzen eine Vielzahl von Skills bzw. Actions, um Geräte von Smart Home Anbietern in das eigene System einzubinden. Seit Anfang des Jahres funktioniert Cortana auch mit Geräten auf der If This, Then That (IFTTT)-Plattform und kann damit auf die Konkurrenz aufholen, liegt jedoch immernoch bei den verfügbaren Geräten klar hinter diesen zurück (IFTTT 2018). Siri und damit der HomePod können nur HomeKit-Geräte über natürliche Sprache bedienen und begrenzen dadurch die Auswahl benutzbarer Smart Home Geräte. Da in diesem Bereich keine große Unterscheidung zwischen den Anwendungsfällen existiert, wird an dieser Stelle eine Darstellung der Anzahl von unterstützten Geräten der Smart Speaker Anbieter als sinnvoller erachtet.

Virtual Assistant	Unterstützte Geräte
Alexa	302
Siri	63 (nur HomeKit)
Google Assistant	57
Cortana	13 (IFTTT unterstützt)

Tab. 3.3 Übersicht unterstützte Smart Home Geräte

3.3.4. Alltägliche Anwendungen

78% der Befragten von PwC denken, dass Smart Speaker viele Vorteile bieten (Zuberer 2017, S. 6). Einige dieser sind z. B. die Steuerung über die Sprache (43%), die bequeme Bedienung und Möglichkeit mehrere Sachen gleichzeitig zu machen (36% und 48%

nach Capgemini (2018) S. 14) , selbstständige Erinnerungen bzw. Hinweise zu erstellen (35%) oder das Delegieren von Aufgaben (27%). Smart Speaker ermöglichen diese Vorteile durch das Bereitstellen von Skills für alltägliche Aufgaben (Zuberer 2017, S. 6).

Anwendungsfall	Assistant-Bewertungen	Alexa-Bewertungen
Erinnerungen / Termine	1122	143
Zeiterfassung	935	17
Listenverwaltung	150	291
Kalenderverwaltung	137	565

Tab. 3.4 Übersicht Anwendungen für alltägliche Aufgaben

Diesen Umstand machen sich auch die Store Anbieter zu Nutze und versuchen entsprechende Skills in Kategorien wie „Produktivität“, „Werkzeuge“ oder „Dienstprogramme“ zu fokussieren. Unter diesen Skills haben sich verschiedene Anwendungen als grundlegend für jeden Smart Speaker herauskristallisiert. Erinnerungen bzw. Termine erstellen, Listen und Kalender verwalten und jegliche Art von Zeiten erfassen sind Aufgaben, die jeder Smart Speaker für seinen Nutzer erledigen kann.

3.3.5. Kommunikation

Eine in den Studien nicht aufgetretene Anwendung von Smart Speaker ist die Kommunikation mit anderen Personen. Ein Grund dafür ist, dass z. B. Alexa als in Deutschland am weitestens verbreiteter und bekannter Smart Speaker diese Funktion nicht bereitgestellt hatte und eine Kommunikation zwischen Alexa-Geräten erst nach Durchführung der jeweiligen Studien möglich ist. Eine Kommunikation zu Festnetzgeräten ist weiterhin nicht möglich. Cortana und Siris Funktionalität sind in dieser Hinsicht auch eingeschränkt. Während Cortana über den Smart Speaker Skype-Telefonate über Sprachbefehle starten kann ist keine Art der Kommunikation über den HomePod möglich. Nur der Google Assistant bietet umfassende Anwendungen durch SMS, Sprach- und kurznachrichten und Telefonie. Die genauen Möglichkeiten sind in der folgenden Übersicht dargestellt:

Anwendungsf.	Alexa	Assistant	Cortana	Siri
Kommunikation zwischen VA-Geräten	x	x	-	-
Kommunikation über Skills	-	x	x	-
Kommunikation über Telefonnetz	-	x	x (via Skype)	-

Tab. 3.5 Übersicht Möglichkeiten zur Kommunikation

3.3.6. E-Commerce

Die Kategorie, welche für weitere Untersuchungen am wichtigsten ist, stellt E-Commerce dar. Bereits 35% der von Capgemini befragten 2558 Nutzer haben über VAs Produkte gekauft. Weitere 34% bestellten Essen, 28% leisteten eine Art Zahlung oder bestellten sich ein Taxi. Dies sind geringfügig mehr, als die von PwC ermittelten 22% der Befragten, welche einen ausgeprägten Wunsch haben, durch Smart Speaker beim Shoppen oder Aufgeben von Bestellungen unterstützt zu werden.

Capgemeni hat in seiner Studie die Frage gestellt, bei welchen Produkten das größte Interesse besteht, diese über VAs zu kaufen. Durchweg waren alle Kategorien ähnlich interessant für die Befragten, mit 41-56%, gewesen. Elektronik, Spezialprodukte wie z. B. Tiernahrung und Kleidung waren die Kategorien von Einzelhandelprodukten, auf welchen das größte Augenmerk liegt. Lieferdienste und Spezialdienstleistungen wie z. B. Taxis hingegen sind unter den Dienstleistungen am gefragtesten. Trotz dieses ausgeprägten Interesses an Bestellfunktionen sind aktuell in den deutschen Skill Stores der Smart Speaker diese nicht existent [8] Die in den entsprechenden Kategorien geführten Anwendungen beschränken sich bisher auf Produkt- und Angebotsinformationen sowie Hilfefunktionen, wie z. B. die Suche nach dem nächsten Laden in der Nähe (Buvat et al. 2018, S. 6; Zuberer 2017, S. 5).

[8] Der Alexa Lieferando.de Skill besitzt eine Bestellfunktion, diese ist jedoch aktuell sehr problembehaftet. Vgl. https://alexa.amazon.de/spa/index.html#skills/dp/B06XTJGL63/reviews

Aufgrund des zuvor beschriebenen Sachverhalts wird an dieser Stelle eine Gegenüberstellung der bisher vorhandenen Skills und der Anzahl ihrer Bewertungen, sowie das Verlangen bzw. Nutzungsverhalten der Befragten der Capgemini- und PwC-Studien, vollzogen. Diese dient zur Verdeutlichung, welches Potenzial vorhanden ist und wie der bisherige Grad der Abschöpfung dieser aussieht.

Anwendungsf.	Assistant-Bewertungen	Alexa-Bewertungen	% Nutzer mit $\geq$ 1 Nutzung	% Nutzer mit Interesse
Produktbestlg.	-	72	35%	22%
Dienstleistung	-	-	28-34%	18-22%

Tab. 3.6 Übersicht E-Commerce Anwendungen und Studiendaten

Es ist jedoch zu beachten, dass es trotz eines Mangels an Angeboten Drittter, möglich ist über Alexa Amazon Prime Produkte zu kaufen und bereits durchgeführte Bestellungen zu wiederholen. Diese Art der Produktbestellung muss, nimmt man das bestehende Angebot aller Anwendungen zum Einkauf über Smart Speaker in Betracht, der absolute Marktführer in dieser Kategorie sein.

Drittanbieter fassen ihr E-Commerce Angebot entsprechend weiter, um gegen dieses bestehen zu können. Viele bieten, angelehnt an die Hauptwendungsgebiete von Smart Speakern, Möglichkeiten Informationen jeglicher Art bezüglich ihres Angebots bereitzustellen. Auf diese Weise sind Produkt-, Rabatt- und Angebotsinformation häufig anzutreffen. Auch die Verwaltung von Kundeninformationen und Wunschlisten sind Funktionen, welche bereits angeboten werden.

3.4. Positionierung der Lösungen

Nach einer kurzen Einführung in die Marktlage der Smart Speaker und der zu diesen gehörenden Shops und Skills, wird an dieser Stelle zusammenfassend eine Gegenüberstellung angebracht.

Kategorie	Alexa	Cortana	Siri	Assistant
Marktanteil	69-71%	1,3%	<1%	21-25%
Shop	Kategorien, Suche und Bewertung	nur Kategorien	Kategorien, Suche und Bewertung	nicht vorhanden
Skills	25000	100	?	1830
Sprachverarbeitung	LSTM RNN und Waveform Modelling	LSTM CNN	RNN und Waveform Modelling	LSTM RNN und TBRU
Entwicklungsmöglichkeiten	Java, C#, Python, Go und NodeJS	C# und NodeJS	Swift und Object-C	Java, C#, Go, NodeJS und Ruby
Positionierung	E-Commerce	Integration	Musik	Information

Tab. 3.7 Übersicht „Big Four“

Der Markt ist weiterhin in der Hand von Amazon, jedoch ist durch anhaltend größeren Absatz des Google Assistants im Vergleich zu Alexa in Zukunft eine Veränderung in dieser Hinsicht zu erwarten. Auch die Entwicklung von Cortana und mögliche weitere Konkurrenten die auf den Markt dringen[9], könnten dies beeinflussen. Es ist jedoch davon auszugehen, dass es bei dem Zweikampf zwischen Google und Amazon bleibt.

Die Shops der einzelnen Anbieter sind aktuell für Kunden kaum relevant und werden entsprechend wenig genutzt. Neue Wege der Information von Nutzern über Skills oder Möglichkeiten der Aktivierung dieser, welche nicht den Shop voraussetzen, könnten die Attraktivität stark erhöhen.

Durch die Möglichkeit, über Alexa entsprechende Artikel aus dem Amazon Prime Sortiment zu bestellen, hat sich Amazon mit seinem Smart Speaker eindeutig in Richtung E-Commerce mit seiner Lösung positioniert. Das Angebot der Konkurrenz ist in dieser Hinsicht kaum ausgebaut. Eine entsprechende andere Ausrichtung der weiteren Smart Speaker Anbieter ist bereits zu erknnen.

Die Verbreitung von Cortana in durch Windows 10 betriebenen Computern und aktuelle Pläne, wie die Integration mit Alexa (Mukherjee und Vengattil 2017) und Möglichkeiten Geräte mit IFTTT anzusprechen (IFTTT 2018) steuern eindeutig daraufhin, dass Cortana der zentrale Punkt für die Integration verschiedenster Technologien werden soll.

[9] z. B. die Telekom plant einen eigenen Smart Speaker (Deutsche Telekom AG 2018)

HomePod wird von Apple mit besonderen Klang und als „revolutionärer“ Lautsprecher angeworben. Waren die bisherigen Verwendungsmöglichkeiten des Gerätes, verglichen mit seiner Konkurrenz gering, so wird es mit iOs12 möglich Telefonate abzuwickeln. Auch die Integrationsmöglichkeiten werden immer vielfältigern, auch wenn sie durch die Beschränkung auf HomeKit-Geräte noch nicht an die der Wettstreiter heranreicht. Durch diese Entwicklung entfernt der HomePod sich langsam von dem Bild eines Musikgerätes mit wenigen „smarten“ Funktionen und hat die Chance in Zukunft seinen Platz auf dem Markt der Smart Speaker zu finden.

Google bietet seine Actions, ähnlich wie Alexas Skills, in einem Store an. Neben der Möglichkeit diese zu aktivieren, ist es jedoch auch möglich die meisten (Actions, welche keine weitere Verknüpfung oder Freigaben benötigen) durch einen einfachen Aufruf des Namens mit dem Google Home zu aktivieren. Dem Assistant ist es so möglich intuitiver auf Anfragen von Nutzern zu reagieren und es ist so leichter die nötigen Information zu erhalten, ohne vorher Stores durchsuchen zu müssen.

4. Akzeptanzmodell für Smart Speaker im E-Commerce

Im folgenden Kapitel werden Faktoren aufgezeigt, welche für die Akzeptanz von Smart Speakern identifiziert wurden und aufbauend auf diesen Hypothesen aufgestellt, die es anschließend zu untersuchen gilt.

4.1. Akzeptanzfaktoren

Ausgehend von der in Kapitel 1.2 definierten Vorgehensweise und nach dem ensprechenden Modell werden Ergebnisse der Literaturanalyse und Untersuchung von Studien dargestellt. Die Faktoren werden dabei in die nach Pavlou und Davis definierten Bereiche der Akzeptanzmodelle Perceived Ease of Use, Perceived Usefulness, Perceived Risk und Trust aufgeteilt.

4.1.1. Perceived Ease of Use

Als Definition von Perceived Ease of Use (wahrgenommene Bedienungsfreundlichkeit) wird in dieser Arbeit die von Davis (1989), nach Radner und Rothschild (1975) aufgestellte, verwendet. Diese definiert Perceived Ease of Use als den Grad, zu dem eine Person glaubt, dass die Nutzung eines bestimmten Systems ohne Aufwand möglich ist. Dies ergibt sich aus der Definition von „Ease“ (Leichtigkeit): „Freiheit von Schwierigkeiten oder großen Anstrengungen.“ Anstrengung ist eine endliche Ressource, die eine Person den verschiedenen Tätigkeiten, für die sie verantwortlich ist, zuweisen kann (Radner und Rothschild 1975). Aufbauend darauf beruht die Behauptung, dass eine Anwendung, welche in allen anderen Faktoren gleich einer zweiten ist, eher akzeptiert wird, wenn sie durch eine Person als leichter zu benutzen wahrgenommen wird (Davis 1989, S. 320).

4.1.1.1. Sprachverarbeitung

Smart Speaker bieten verschiedene Vorteile, welche eine einfache Bedienung in verschiedenen Lagen ermöglichen. Gegenüber Webseiten oder Smartphone-Apps finden Nutzer sie natürlicher, mehr auf ihre Bedürfnisse angepasst und sie finden es vorteilhaft, dass es

sich anfühlt als spräche man zu einer realen Person. Gegenüber dem Umgang mit Menschen hingegen sehen Nutzer den Vorteil, dass man den menschlichen Kontakt vermeiden kann und trotzdem seine natürliche Sprache nutzen kann (Buvat et al. 2018, S. 14, 15).

Entsprechend dieser Einschätzungen wurde menschenähnliches Verhalten als die Hauptkomponente, welche den Umgang mit Smart Speakern verglichen mit anderen Möglichkeiten des E-Commerces erleichert, identifiziert. Jenes anthropomorphes Verhalten war bereits bei der Einteilung der Lösungen, welche VAs integrieren, in Kapitel 2.2 kurz thematisiert worden. Die Eigenschaft, natürliche Sprache verarbeiten zu können, wurde an dieser Stelle als eine menschenähnliche Eigenschaft erkannt. Wie eine solche Verarbeitung ein maschinelles System menschlich machen kann wird nachfolgend dargestellt.

Einen Ansatz dafür bieten Edlund et al. (2008), welche für dieses Problem mentale Modelle bzw. Metaphern, nach Norman (1998), definieren. Mit Hilfe dieser ist es leichter unbekanntes bzw. komplexe Sachverhätlnisse verständlicher darzustellen. Die Metaphern sind zum einen aus der Entwicklerperspektive zu sehen, welche darstellt, was Nutzer wahrnehmen sollen und aus der Nutzerperspektive, welche wiedergibt, was dieser vom System versteht (Edlund et al. 2008, S. 4-5).

Aufbauend auf dieser Vorgehensweise wurden drei Metaphern entwickelt. Die Interface Metaphor, durch welche Nutzer ein Dialogsystem auch als eine solche maschinelle Schnittstelle wahrnehmen und mit dessen Hilfe erreichen, wozu die normalen Eingabemethoden sonst verwendet wurden. Bei der zweiten handelt es sich um Human Metaphor. Bei dieser wird das Dialogsystem als ein Gesprächspartner anerkannt. Es ist ein „Wesen“ mit Fertigkeiten der natürlichen Sprache. Eine dritte Metapher, die Android Metapher, wird als eine solche angesehen, bei der ein System zu bestimmten Zeitpunkten nach der Human Metapher und zum anderen nach der Interface Metapher agiert. Das Ziel ist es, dass eine Mensch-Maschine-Interaktion zu den Zeitpunkten, wo es als Human Metapher wahrgenommen wird, eine Mensch-Mensch-Interaktion imitiert. Ein „natürliches“ Verhalten wird dann erreicht, wenn das Verhalten der Maschine dem von der Metapher ausgehend erwarteten des Nutzers entspricht (Edlund et al. 2008, S. 4-5, 10-11).

Wie in Kapitel 3.3 dargestellt, haben Smart Speaker verschiedene Anwendungsgebiete und Aufgabenbereiche. Von der einfachen Verwaltung von Listen, über Informationssuche über Dialoge bilden diese Geräte verschiedenste Anwendungsfälle mit der natürlichen Sprache ab. Entsprechend entwickelt werden diese Anwendungen, sodass man im Umgang mit dieser Technologie davon reden kann, dass die Menschenähnlichkeit mit der Android Metapher dargestellt werden kann. Ausgehend von der gewählten Metapher sprechen Edlund et al. (2008) davon, dass ein System dem Nutzer dargestellt und auch

intern diesem entsprechend aufgebaut werden muss. Smart Speaker haben auf diese Weise zum einen als einfacher Ersatz für gewöhnliche Eingabemethoden zu dienen, sind aber auch in einigen Fällen ein Gesprächspartner.

Verschiedene Faktoren sind Grundlage für die natürliche Verarbeitung von Sprache. *Spracherkennung*, *Intent-Erkennung*, sowie *Dialoge*, *Aufgabenerfüllung* und *Sprachsynthese* sind wichtige Grundsteine. Welche Methoden und Technologien für diese Aufgabenbereiche genutzt werden, wurde bereits im Kapitel 2 dargestellt. Durch die Verwendung von LSTM RNN konnte für die Aufgabenbereiche Spracherkennung und Intent Erkennung bereits eine menschenähnliche 5,8% Fehlerrate von Xiong et al. (2017) nachgewiesen werden.

Für die Erfüllung der vom Nutzer gestellten Aufgaben und dem Aufbau von Dialogen stehen Entwicklern von allen Smart Speaker Anbietern verschiedene Bibliotheken und Hilfestellungen zur Verfügung, welche in Kapitel 3.1 dargestellt wurden. Auf diese Weise ist es möglich Mixed-Initiative Dialoge und Reaktionen auf Fehlerfälle abzubilden. Diese sind wichtige Bestandteile von menschenähnlichen Dialogen (Porzel 2006, S. 11-12; Allen et al. 2001, S. 6). Bei ersteren handelt es sich um solche, bei denen die Kontrolle über den Dialogablauf zwischen Nutzer und System wechselt, wodurch ein natürliches Gespräch mit dem System entsteht (Allen et al. 2001, S. 6). Für die Behandlung von Fehlern in der Sprachverarbeitung sind Entwicklern wiederum Möglichkeiten der Smart Speaker Hersteller gegeben, müssen aber entsprechend implementiert werden.

Mit Hilfe von in Kapitel 2.3 und 2.4 dargestellen Möglichkeiten der Synthese kann immer flexiblere und natürlichere Sprache erstellt werden. Der Grad, wie menschenähnlich die erstellte Sprache ist, hängt dabei von verschiedenen Faktoren ab, wie z. B. der genutzten Sprachdatenbank, den Sprechern, welche diese gefüllt haben und natürlich den verwendeten Methoden (Zen 2009, S. 1). Es besteht außerdem das Problem, in das sogenannte „Uncanny Valley“ (Mori 1970) abzudriften, also, dass die erstellte Sprache so menschenähnlich und natürlich ist, dass es nicht mehr erstrebenswert ist. Menschen können von zu „menschlichen“ Maschinenen abgeschreckt werden und diese dadurch nicht benutzen (Edlund et al. 2008, S. 9-10). Entsprechend müssen für Smart Speaker Anwendungen identifizieren, welchen Grad natürlicher Sprache es zu nutzen gilt, um von dem Nutzer akzeptiert zu werden.

4.1.1.2. Sprachsteuerung

Neben der von Davis (1989) definierten wahrgenommenen Bedienungsfreundlichkeit, zeigten auch andere Untersuchungen, dass ein Zusammenhang zwischen der Nutzungshäufigkeit von Systemen und der Zufriedenheit der Nutzer bei dieser zu erkennen ist (Bokhari 2006, S. 222-224). Koo et al. (2017) sprechen sogar davon, dass speziell Smart Speaker nicht nur Möglichkeiten bieten müssen, Aufgaben zu erleichtern und den Nutzer damit zufrieden zu stellen, sondern sie müssen durch diese Erleichterungen sogar neuen Nutzen schaffen, um erfolgreich zu sein.

Aufbauend auf den Methoden der Verarbeitung von natürlicher Sprache basiert der großteil der Steuerung eines Smart Speakers auf der gesprochenen Sprache. Mit dieser Art der Bedienung des Systems scheinen Benutzer zufrieden zu sein. Sie wird als eine Erleichterung gegenüber der gewohnten Verwendung von Maus und Tastatur angesehen (43% der Befragten der PwC Studie (2017) teilten diese Einschätzung). Die bequemere Bedienung von Geräten (36% der Befragten der PwC Studie (2017)), bzw. eine allgemein leichtere Bedienung als eine App oder Webseite (52% der Befragten der Capgemini Studie (2018) und der angenehmere Umgang als mit Menschen bzw. Call-Centern (47% der Befragten der Capgemini Studie (2018)), unterstrichen erneut, dass die Stärken von Smart Speakern in der Bedienung liegen.

Besonders die bequemere Bedienung von anderen Geräten ist ein Beispiel, welches nach Koo et al. (2017) durch seine Verbesserung von bestehenden Anwendungen Smart Speaker erfolgreich machen könnten. Diese Anwendungen ermöglichen die Steuerung von verbundenen Geräten durch die natürliche Sprache und erleichtern damit alltägliche Aufgaben durch neue Möglichkeiten der Bewältigung.

Neben diesem Faktor zeigen Koo et al. außerdem auf, dass aufgrund der besonderen Eigenschaften von Smart Speakern (vgl. Definition in Kapitel 2.2), diese nicht nur den Nutzer unterstützen und dessen Aufgaben erleichtern, sondern auch die Familie dessen in Betracht ziehen müssen (Koo et al. 2017, S. 7-8). Besonders die Funktionen der Kommunikation zwischen den Geräten kann an dieser Stelle als möglicher Anwendungsfall angebracht werden.

4.1.1.3. Sichtbarkeit

Ein in Kapitel 3.2 identifiziertes Problem ist die geringe Nutzung von Drittanbieter-Anwendungen, welche einhergeht mit der Unbekanntheit der Skill Stores und ihrer spärlichen Verwendung zum Finden von neuen Anwendungen. Obwohl die Stores von Alexa und Assistant eine passende Übersicht des Angebots bieten, müssen Anbieter von entsprechenden Skills bzw. Actions andere Wege nutzen, um Nutzer auf ihre Lösungen aufmerksam zu machen.

Erste Ideen, wie dieses Problem umgangen werden könnte, zeigt z. B. Google. Sagt man dem Assistant, dass er eine Action starten soll, welche noch nicht aktiviert ist, sucht er diese automatisch und falls vorhanden, wird diese aktiviert und gestartet (Schulze 2017). Wollen Nutzer eine bestimmte Anwendung starten, von der sie sicher sind, wie das Schlagwort ist, auf das gehört wird, dann kann auf diese Weise ein Öffnen des Stores umgangen werden und direkt der Skill aktiviert werden. Diese Funktion ist jedoch nur für Anwendungen möglich, welche keine weiteren Berechtigungen oder Accountverknüpfung benötigen, da jene bei der Aktivierung im Browser oder der App durchgeführt werden.

Verglichen mit dem Verhalten beim Finden neuer mobilen Apps, ist das Verhältnis zwischen Freunden und Stores als Anlaufmöglichkeit zwar ähnlich[1], jedoch ist die Kategorie der Personen welche keine mobilen Apps von Drittanbietern nutzen nicht existent. Auch die Suchmaschinen, welche eine wichtige Funktion bei der Suche von Apps darstellen, sind für Smart Speaker noch nicht relevant geworden. Im Speziellen für Anwendungen des E-Commerces ist dies besonders problematisch, da nur Alexa von Haus aus Möglichkeiten bietet Einkäufe zu tätigen und der Assistant diese Anwendungen nicht automatisch aktivieren kann, da sie im Normalfall weitere Berechtigungen benötigen. Auch Methoden des Reengagements, wie z. B. Push Notifications von mobilen Apps, sind bei Smart Speakern nicht möglich und somit nach der Aktivierung eines Skills bzw. einer Action nicht gegeben, sodass das Risiko größer ist, dass Nutzer eine Anwendung nach der ersten Verwendung vergessen (Kinsella und Mutchler 2018, S. 23-24; Tiongson 2015).

Durch diesen Umstand bleibt die Frage offen, ob die schlechte Auffindbarkeit von Anwendungen auch das Nutzerverhalten beeinflusst, indem Smart Speaker für die Nutzung im E-Commerce nicht nützlich erscheinen, weil das Wissen über mögliche Skills und Actions fehlt.

[1] 22,5% „Freunde“ und 17% „Stores“ bei Smart Speakern (Kinsella und Mutchler 2018) gegen 52% „Freunde“ und 40% „Stores“ bei mobilen Apps (Tiongson 2015)

4.1.2. Perceived Usefulness

Bei der Definition von Perceived Usefulness (wahrgenommene Nützlichkeit) wird grundlegend Davis' (1989), welche auf Peffer (1982), Schein (1980) und Vroom (1964) aufbaut, verwendet. Die Perceived Usefulness wird definiert als "der Grad, in dem eine Person glaubt, dass der Einsatz eines bestimmten Systems ihre Leistungsfähigkeit verbessern würde". Dies ergibt sich aus der Definition des Wortes useful (nützlich): "Fähigkeit, vorteilhaft eingesetzt zu werden."(Pfeffer 1982; Schein 1980; Vroom 1964). Ein System mit hoher Usefulness (Nützlichkeit) wiederum ist ein solches, bei dem ein Anwender an die Existenz eines positiven Nutzen-Leistungs-Verhältnisses glaubt (Davis 1989, S. 320).

4.1.2.1. Geschwindigkeit

Ein wichtiger Vorteil der sich durch die Nutzung von Smart Speakern gegenüber der Verwendung von Webseiten oder dem Umgang mit anderen Menschen ergibt, ist eine schnellere Bewältigung von Aufgaben (Buvat et al. 2018; Zuberer 2017). 49% der Befragten der Capgemini Studie (2018) und damit der am meisten genannte Vorteil gegenüber dem Umgang mit Menschen oder Call Centern, war die höhere Geschwindigkeit von VAs in der Benutzung. Ebenso war mit 46% der größte Vorteil von Smart Speakern laut der PwC Studie (2017) der schnellere Zugang zu Informationen.

Eben diese Schnelligkeit wird auch bei E-Commerce Anwendungen der Smart Speaker versucht beizubehalten. Sind alle Vorbereitungen[2] getroffen ist es z. B. über alexagesteuerte Smart Speaker möglich durch einfaches Zurufen eines Prime-berechtigten Artikelnamens diesen zu bestellen oder eine bereits getätigte Bestellung zu wiederholen. Auch eine Stornierung ist direkt nach Abgabe der Bestellung leicht möglich oder das Erhalten von Informationen über den Fortschritt der Bestellung[3]. Drittanbieter von Skills und Actions konzentrieren sich in Deutschland aktuell auf die schnellen Informationsmöglichkeiten von Smart Speakern (vgl. Kapitel 3.3.6).

Es bestehen aber auch grade durch diese hohe Geschwindigkeit bei möglichen Bestellungen über Smart Speaker rechtliche Probleme. Da es sich bei Verträgen über diese Geräte um Fernkommunikationsmittel handelt, kommen besondere Informationspflichten aus dem Fernabsatz zum Tragen (vgl. § 312c Abs. 2 BGB). Es müssten auf diese Weise In-

[2] im Normalfall Verknüpfung von Accounts und Verteilung von Berechtigungen
[3] vgl. https://www.amazon.de/gp/help/customer/display.html?nodeId=201807230

formation über Ware, Identität, Beschwerdewege, Rücktritts- und Widerrufsrechte usw. sprachlich ausgegeben werden. Ein Umschwenken auf weitere Angaben in App oder Display werden nicht geduldet (Koch und Schmidt-Hern 2018). Es wurde jedoch bereits die Informationspflicht durch § 312d Abs. 1. S. 1 BGB i.V.m. Art. 246a § 3 EGBGB abgeschwächt. Hiernach müssen nur noch Kerninformationen nach Art. 246a § 3 EGBGB wie Ware, Gesamtpreis, Identität des Unternehmens und Widerrufsrecht sprachlich mitgeteilt werden (Bockmeyer und Vogt 2018, S. 3-4; Specht und Herold 2018, S. 42).

4.1.2.2. Automation, Integration und Multitasking

Automation stellt einen wichtigen Faktor dar, wodurch Benutzer von Smart Speakern bei der Bewältigung ihrer Aufgaben, ohne weitere Eingaben, unterstützt werden. 39 der PwC bzw. 41% der Befragten der Capgemini Studie (2018) empfinden die Möglichkeiten von Smart Speakern zur Automation von Aufgaben als einen Grund, weswegen sie die Verwendung dieser dem Umgang mit Menschen, Call Centern oder Webseiten vorziehen. Auch die damit verbundenen selbständigen Erinnerungen und Hinweise sind nach 35% der deutschen Nutzer, welche in der PwC Studie (2017) befragt wurden, ein wichtiger Faktor. Von einem vollkommen autonomen Verhalten der Smart Speaker kann jedoch nicht ausgegangen werden, sondern es besteht weiterhin eine teilweise Mitwirkung des Menschen.

Zur Realisierung der Automation von Smart Speakern stellen die Schnittstellen zur *Integration* eine wichtige Grundlage dar. Alexa bietet z. B. Möglichkeiten der Verbindung mit Kalendern, Listen sowie die Verknüpfungen mit anderen Webservices. In Kombination mit der Verwaltung und Vergabe von Berechtigungen, welche es ermöglichen personenbezogene Daten des Nutzers durch Alexa zu nutzen, können verschiedene Teilbereiche von Aufgaben automatisiert werden. Durch die Nutzung entsprechender Schnittstellen können im E-Commerce, z. B. bei der Nutzung von Alexa für den Einkauf, bestimmte Schritte automatisiert werden, indem Daten von angeschlossenen Systemen übernommen werden. Nach Auswahl des Artikels werden auf diese Weise Lieferadresse und Zahlungsinformationen automatisch gewählt und dem Nutzer eine vorbereitete Bestellung präsentiert.

Neben diesen Faktoren stellen die Möglichkeiten von Smart Speakern Grundlagen dar, sodass Benutzer mehrere Aufgaben gleichzeitig bewältigen können. 48% der Befragten der Capgemini Studie (2018) sahen Multitasking als insgesamt zweitwichtigsten Vorteil gegenüber der Nutzung von traditionellen Webseiten oder Apps. Die Steuerung über die

eigene Sprache ermöglicht es z. B. gleichzeitig mit den Händen andere Aufgaben zu erledigen und man ist außerdem nicht mehr an einen Ort gebunden. Entsprechend gaben auch 43% der Teilnehmer der PwC Studie (2017) dies als zweitwichtigsten Vorteil bei der Nutzung von Smart Speakern an.

Zur Realisierung dieses Vorteils spielt die *Integration* erneut einen wichtigen Faktor. Durch die Verbindung mit anderen Systemen sind viele verschiedenen Anwendungsgebiete durch Smart Speaker abgedeckt (vgl. Kapitel 3.3). Eben diese Auswahl sehen auch 36% der Befragten Deutschen der PwC Studie (2017) als Vorteil der Nutzung von Smart Speakern, wodurch sie ihre Leistungsfähigkeit erhöhen.

4.1.3. Perceived Risk

„Perceived Risk“ (wahrgenommenes Risiko) stellt den subjektiven Eindruck eines Nutzers dar, dass bei dem Streben nach einem gewünschten Ziel ein Verlust erlitten werden kann (Bauer 1960). Ein hohes wahrgenommenes Risiko bei Transaktionen mit Smart Speakern wird als Verlust von aufgefasster Kontrolle verstanden und beeinflusst entsprechend negativ die Intention eine Transaktion durchzuführen. Ein geringes Risiko hingegen würde diese Kontrolle wiederherstellen und Nutzer geschäftsfreudiger machen (Pavlou 2003, in Anlehnung an S. 77).

Im E-Commerce, durch die unpersönliche Natur der Online-Welt und die implizite Unsicherheit, die bei der Verwendung einer globalen Infrastruktur für Transaktionen entsteht, ist ein solches wahrgenommenes Risiko ein unabdingbarer Faktor (Pavlou 2003, S. 77). Ring und Van de Ven (1994) klassifizieren in diesem Zusammenhang, dass Risiken entweder technologiebedingt sind (Environmental Risks) oder vom Handelspartner ausgehen (Behavioral Risk).

4.1.3.1. Environmental Risks

Umweltbedingte Risiken entstehen hauptsächlich aufgrund der unberechenbaren Natur des Internets, welches nie komplett durch den Händler oder Nutzer kontrolliert werden kann (Pavlou 2003, S. 77). Die Verteiber der Smart Speaker und Entwickler der VAs bzw. Anwendungen können dennoch Einfluss auf das System und seine Sicherheit nehmen, um Angriffe dritter Parteien zu verhindern.

Beispiele umweltbedingter Risiken im E-Commerce sind z. B. der Diebstahl von persönlichen oder Kreditkarteninformationen. Bekannte Vorgehensweisen sind das Hacken dieser Informationen oder Phishen mit Hilfe von dem Original ähnlich aussehenden Nachrichten oder Webseiten. Smart Speaker bieten neue Angriffsszenarion durch die Verwendung von natürlicher Sprache. Haack et al. (2017) haben in einer Analyse der Sicherheit von Amazon Echo Produkten erste Erkenntnisse über sprachbasierte Angriffe erhalten. Alexa nimmt jegliche Art von Sprache entgegen, d. h. auch indirekte Sprache, welche über Computer generiert wurde. Dies kann speziell bei einer bösartigen Bestellung ausgenutzt werden. Ausgelöst kann eine solche z. B. auch über eine auf einem anderen System erhaltene und vorgelesene Sprachmail werden. Alexa benötigt eine Bestätigung der Bestellung und einen vorher definierten 4-stelligen PIN. Die Bestätigung kann erneut auch von computegenerierter Sprache erfolgen. Bei der Eingabe des PIN-Codes konnten Haack et al. nachweisen, dass die Möglichkeit diesen durch ein entsprechendes Script zu „bruteforcen“ besteht (Haack et al. 2017, S.5-7).

Neben sprachbasierten Angriffen sind durch die Architektur von Smart Speakern und Nutzung von VAs auch Netzwerkangriffe denkbar. Entsprechende Szenarion wurden von Haak et al. getestet und Ergebnisse festgehalten. Die Anwendungsfälle umfassten dabei einen „Man in the Middle“-Fall, das Extrahieren von Informationen aus Paketen, das Wiederholen von Paketen und Application Programming Interface (API)-basierte Angriffe. In keinem dieser Fälle konnten sie Möglichkeiten für Ansatzpunkte bösartiger Verwendungen nachgeweisen (Haack et al. 2017, S. 8-9).

Ein weiteres Risiko, welches bei der Verwendung von Smart Speakern existiert, ist die Fragestellung der Privatsphäre. In der Arbeit von Caire et al. (2016) wurden verschiedene Herausforderungen von uns umgebenden intelligenten Systemen, anhand von drei verschiedenen Definitionen der Privatsphäre, identifiziert.

Ebendiese kann als ein Recht, welches durch Gesetze geschützt werden muss, gesehen werden. Eine weitere Sicht wäre diese als „Enabler“. Anhand dieser Sichtweise ist es einer Person möglich Kontrolle über seine Privatsphäre auszuüben, indem er selbst entscheidet, in welchem Maße seine persönlichen Informationen verarbeitet werden dürfen. Aufbauend auf diesen Eigenschaften der Privatsphäre besteht die dritte Definition, wodurch sie als „Commodity“ angesehen wird. Private Informationen können auf diese Weise gehandelt werden, z. B. im Tausch gegen bessere Dienstleistungen oder personalisierte Angebote (Caire et al. 2016, S. 628-631).

Besonders die letzte Sicht auf die Privatsphäre ist für die Nutzung von Smart Speakern relevant. Mit der Nutzung von Geräten wie z. B. dem Amazon Echo willigt man ein, dass jegliche Anfragen, welche man an das System stellt, aufgezeichnet und von Amazon weiter verarbeitet werden[4]. Wie in Kapitel 3.1.3 dargestellt, werden z. B. von Siri auch weitere Informationen über die Umgebung oder mögliche Gemütslage eines Nutzers gesammelt, um personalisierte Antworten geben zu können (Siri Team 2018c).

Entsprechend diesen Vorgaben ist es möglich, dass zum einen durch Angriffe persönliche Informationen, welche von Smart Speaker gesammelt wurden, abzugreifen, oder durch Fehlfunktion der Geräte diese preisgegeben werden. Erste Fälle, in denen Smart Speaker persönliche Informationen an dritte freigegeben haben, sind bereits aufgetreten[5] und werden verstärkt durch den Umstand, dass das Zuhause ein Ort der Entspannung und Ruhe ist, in dem Personen ihre Privatspähre ausleben können (Friedewald et al. 2005, S. 222-225).

Ein weiteres Risiko besteht außerdem für Personen neben dem eigentlichen Nutzer des Smart Speakers, welche außerdem gegenwärtig sind. Wird eine Smart Speaker Anwendung gestartet und die Sprache der Nutzer aufgenommen, ist es möglich, dass auch weitere Personen durch die Speaker Recognition (vgl. Kapitel 2.3) identifiziert und ein entsprechender Vektor erstellt wird. Eine weitere Einwilligung, welche vor der ersten Benutzung erteilt wird, kann an dieser Stelle nicht angefordert und somit mögliche persönliche Informationen oder Profile von dritten erstellt.

4.1.3.2. Behavioral Risks

Verhaltensrisiken können entstehen, da für Online-Händler aufgrund der entfernten und unpersönlichen Natur des E-Commerces und die fehlenden Möglichkeiten des Staates alle Transaktionen entsprechend zu überwachen, die Chance zum opportunistischen Handeln besteht (Pavlou 2003, S. 77).

Beispiele solchen Verhaltens wären im Allgemeinen E-Commerce Fällen Produktfehldarstellungen, Annehmen falscher Identitäten, Preisgeben von privaten Informationen, irreführende Werbung und die Kündigung von Garantien (Pavlou 2003, S. 77-78). Im speziellen Fall der Smart Speaker bestehen Unterschiede zu diesen allgemeinen Risiken. Werbung in jeglicher Form wurde bisher von Smart Speaker Anbietern noch nicht im-

[4] vgl. https://www.amazon.com/gp/help/customer/display.html?nodeId=201602040
[5] vgl. https://www.cnbc.com/2018/05/24/amazon-echo-recorded-conversation-sent-to-random-person-report.html

plementiert, erste Pläne für zukünftige Angebote jedoch bereits von Amazon gestartet[6]. Drittanbietern ist es möglich Werbung innerhalb ihrer Anwendung in Form von Sprachnachrichten oder auf angeschlossenen Geräten zu schalten. Diese Möglichkeit wird selten genutzt, es besteht aber besagtes Risiko der irreführenden Werbung.

Die Annahme falscher Identitäten könnte in Zukunft ein größeres Problem darstellen, falls die bestehenden Möglichkeiten von Alexa und Co. ausgebaut werden, bzw. wenn weitere Anbieter das Gerät für E-Commerce Anwendungen nutzen. Die Erstellung eines eigenen Skills ermöglicht es z. B. bei Amazon den „Amazon Pay“ Service, oder den entsprechenden „Google Pay“ Service von Google zu verwenden. Es wird dem Entwickler hierbei nicht erschwert die Identität einer anderen Person oder eines Unternehmens anzunehmen. Durch die Auswahl des Namens eines Skills, sowie durch die Wahl von entsprechenden Logos und Beschreibungen, ist es möglich jemand anderen zu mimen.

Für die Darstellung von Produkten, sowie Möglichkeiten Rechte oder Garantien mit Hilfe von Smart Speakern dem Nutzer wiederzugeben, bestehen aktuell verschiedene Möglichkeiten. Zum einen ist es denkbar Informationen über natürlich Sprache auszugeben, welche jedoch in sich fehlerbehaftet ist[7]. Eine Ausgabe von Informationen über angeschlossene Geräte, wie z. B. das Smartphone, ist mit Alexa und Assistant möglich. Cortana kann sogar über den Computer weitere Informationen bereitstellen. Über diese Schnittstellen kann Alexa Grafiken und Text, der Assistants außerdem Links und Listen, darstellen. Es ist entsprechend möglich Produkte, Rechte oder Garantien darzustellen bzw. zu kündigen.

4.1.4. Trust

Das Vertrauen zwischen Nutzer und Unternehmen stellt einen wichtigen Faktor im E-Commerce dar (Gefen und Straub 2003, S. 7). Es ist als solcher unabdingbar in unsicheren Situation, da es notwendig ist, um Risiken einzugehen und sich selbst anfällig gegenüber vertrauten Parteien zu machen (Hosmer 1995, S. 213-214). Vertrauen reduziert dadurch das Risiko Opfer von opportunistischem Handeln zu werden und ermöglicht den Handel in risikoreichen Situationen (Fukuyuma 1995).

[6] vgl. https://www.cnbc.com/2018/01/02/amazon-alexa-is-opening-up-to-more-sponsored-product-ads.html

[7] menschliche Worterkennung hat eine Fehlerrate von ca. 5,9% (Huang 2017)

Es existieren theoretische und empirische Untersuchungen, welche den Faktor „Trust“ und seine Integration in das TAM betrachten und unterstützen, sowie seinen Zusammenhang mit anderen Faktoren untersuchen (Pavlou 2003; Gefen und Straub 2003; Chircu et al. 2000).

Das Vertrauen ist ein wichtiger Einflussfaktor auf den wahrgenommenen Nutzen einer Schnittstelle, speziell in einer online Umgebung, da die Garantie, dass Nutzer den erwarteten Nutzen von einer Webseite erhalten, von den Personen hinter dieser abhängt (Gefen und Straub 2003, S. 9-10). Kann dem Anbieter aus Sicht des Nutzers nicht vertraut werden, dann ist für ihn nicht zu erwarten, dass man von jener Schnittstelle jeglichen Vorteil erwarten kann (Pavlou 2003, S. 78-79). Eine Übertragung dieser Ergebnisse auf Smart Speaker scheint trivial. Wird den Anbietern dieser Geräte bzw. der Anwendungen dieser nicht vertraut, dann ist es aus Sicht des Nutzers nicht vorstellbar aus der Nutzung von Smart Speakern einen Vorteil zu erzielen. Das Vertrauen wird außerdem notwendig, damit über diese neue Schnittstelle E-Commerce betrieben werden kann, da nur wenn es vorhanden ist, Nutzer gewillt sind das benötigte Risiko einzugehen.

Chircu et al. (2000) argumentieren in ihrer Arbeit, dass das Vertrauen im E-Commerce die wahrgenommene Nützlichkeit erhöt. Die Logik dahinter ist, dass das Vertrauen die Notwendigkeit des Nutzers verringert, die jeweilige Situation zu verstehen, zu überwachen und zu kontrollieren. Dadurch würde eine Transaktion erleichtert und müheloser gestaltet werden können (Pavlou 2003, S. 79). Empirische und theoretische Nachweise dieses Verhaltens sind rar, Jarvenpaa, Tractinsky und Vitale (1999) haben ihre Untersuchung des zwischenbetrieblichen Vertrauens auf Kundenverhalten erweitert, um darzustellen, dass das Vertrauen in einen Internet Store das Risiko dort zu kaufen verringert(Pavlou 2003, S. 79). Angewendet auf die Möglichkeiten und Eigenschaften der Smart Speaker könnte erhötes Vertrauen in Anwendungen bzw. die Geräte z. B. die Notwendigkeit Prozesse zur Freigabe von Berechtigungen und zum Verknüpfen von Accounts erleichtern, indem eine stete Kontrolle, Verständnis und Überachung dieser nicht weiter notwendig ist. Auf diese Weise würden die E-Commerce Anwendungen für den Nutzer müheloser wahrgenommen werden.

4.2. Intention to Transact und der Transaktionsprozess

Die Intention an Transaktionen teilzunehmen ist definiert als die Absicht eines Verbrauchers in eine online Handelsbeziehung mit einem Internethändler zu treten. Dabei kann es sich um das Teilen von Geschäftsinformationen, den Aufbau von Geschäftsbeziehungen und um Transaktionen an sich handeln (Zwass 1998).

Pavlou (2003) hat den Transaktionsprozess in seiner Arbeit in die drei Schritte „Information Retrieval“, „Information Transfer“ und „Product Purchase“ aufgeteilt (Pavlou 2003, S. 72).

Abb. 4.1 Online Transaktionen nach Pavlou (2003), vgl. S. 72

Der erste Schritt umfasst Teilbereiche wie das Sammeln von Informationen und Vergleichen mit Alternativen (Pavlou 2003, S. 72). Ein großes Problem des E-Commerces besteht darin, potenzielle Kunden in bezahlende umzuwandeln. Nur 3,6% der Besucher eines Online-Händlers kaufen im Durchschnitt etwas am Ende einer Sitzung (intelliAd 2018). Online existieren außerdem eine Großzahl an Vergleichsmöglichkeiten, welche einem die Informationsbeschaffung und das Vergleichen erleichten.

Besonders für Smart Speaker ist, dass die Aufgabenbereiche und Möglichkeiten dieses ersten Schrittes stark eingeschränkt sind. Auf E-Commerce ausgelegte Skills besitzen zum einen oft keine Möglichkeiten weitere Informationen über ihre Produkte zu erhalten (bevor man sie kaufen möchte) und bieten zum anderen selten einen Warenkorb an. Auch Funktionen oder Angebote um Produkte zu vergleichen existieren nicht in dem Maße der Online-Angebote und sind durch die Restriktion, hauptsächlich natürliche Sprache oder angeschlossene weitere Systeme zu nutzen, eingeschränkt. Auch die Auswahl eines passenden Händlers passiert bereits vor Auswahl des eigentlichen Produktes, da entweder der passende Skill bzw. die Action aktiviert werden muss, oder der von Haus aus aktivierte Shopping-Skill das hauseigene Angebot nutzt[8]. Die Hauptaufgaben dieses ersten Schrittes können entweder nicht durchgeführt werden oder müssen mit Hilfe von anderen Systemen bewältigt werden.

[8] nach aktuellem Stand (18.07.18) existieren keine Anwendungen, welche als Marktplatz mehrerer Anbieter agieren

Schritt zwei „Information Transfer“ umfasst die Registrierung bei dem ausgewählten Händler sowie die Freigabe von privaten Informationen. Im dritten Schritt „Product Purchase“ werden dann weitere nötige Informationen, die für die Zahlung und Lieferung notwendig sind, bereitgestellt, sodass der Kauf durchgeführt werden kann (Pavlou 2003, S. 72).

Diese in zwei Teile aufgeteilten Bereiche sind bei Smart Speakern oft in einem Schritt behandelt, oder im Fall von Google Assistant Actions bereits im 1. Schritt, beim Aktivieren einer entsprechenden Anwendung, durch das benötigte Account-Linking, bereitgestellt. Entsprechende Informationen werden auch bei dem hauseigenen Alexa Skill bereits vor Verwendung der Anwendung freigegeben. Als Entwickler eines eigenen Skills stehen einem nur durch das Alexa Skills Kit verschiedene Zeitpunkte zur Verfügung, wann eine entsprechende Abfrage stattfinden soll, wodurch ein ähnliches Verhalten wie in Abb. 4.1 dargestellt, möglich ist.

4.3. Hypothesen

Das positive Verhältnis zwischen einer Verhaltensintention und der endgültigen Aktion bzw. dem wirklichen Verhalten wurde bereits in der Theory of Reasoned Action und der Theory of Planned Behavior untersucht (Ajzen 1991; Ajzen und Fishbein 1980). Aufbauend auf diesen Theorien und empirischen Beweisen wird eine Beziehung zwischen der „Intention to Transact“ und der „Actual Transaction“ untersucht.

H1 Die Kaufabsicht des Verbrauchers hat einen positiven Einfluss auf das wirkliche Transaktionsverhalten mit Smart Speakern.

Dass Einflüsse von „Perceived Usefulness“ und „Perceived Ease of Use“ auf die Intention eines Nutzers bzw. diese zwischen den beiden Faktoren bestehen, konnte in weiteren Arbeiten, wie z. B. von Davis (1989) dargestellt werden. Eine Untersuchung dieser für Smart Speaker steht noch aus.

H2 Die wahrgenommene Bedienungsfreundlichkeit von Smart Speakern hat einen positiven Einfluss auf die Kaufabsicht über diese.

H3 Die wahrgenommene Bedienungsfreundlichkeit von Smart Speakern hat einen positiven Einfluss auf deren wahrgenommene Nützlichkeit.

H4 Die wahrgenommene Nützlichkeit von Smart Speaker hat einen positiven Einfluss auf die Kaufabsicht über diese.

Die Theory of Reasoned Actions hat bereits progonistiziert, dass Verbraucher eher bereit wären an Transaktionen teilzunehmen, wenn ihr wahrgenommenes Risiko gering ist (Ajzen und Fishbein 1980). Pavlou (2003) hat dies in seiner Untersuchung allgemeine E-Commerce Transaktionen bereits untersucht, eine weitere für Smart Speaker wird in dieser Arbeit durchgeführt.

H5 Das wahrgenommene Risiko bei der Nutzung von Smart Speakern hat einen negativen Einfluss auf die Kaufabsicht.

Mögliche Einflüsse von Vertrauen auf risikobehaftete Situation sowie Nützlichkeit und Bedienungsfreundlichkeit von Smart Speakern wurden bereits in Kapitel 4.1 dargestellt. Dabei liegt zugrunde, dass Verbraucher annehmen, dass ein vertrauenswürdiger Anbieter nicht opportunistisch verhalten wird (Gefen 2000).

H6 Das Vertrauen in Smart Speaker hat einen negativen Einfluss auf deren wahrgenommenes Risiko.

H7 Das Vertrauen in Smart Speaker hat einen positiven Einfluss auf deren wahrgenommene Nützlichkeit.

H8 Das Vertrauen in Smart Speaker hat einen positiven Einfluss auf die wahrgenommene Bedienungsfreundlichkeit.

H9 Das Vertrauen in Smart Speaker hat einen positiven Einfluss auf die Kaufabsicht mit diesen.

Diese Hypothesen und identifizierte Faktoren aus Kapitel 4.1 werden integriert in das von Pavlou (2003) aufgestellte TAM, in Abb. 4.2 dargestellt. Positive Relationen sind dabei grün und negative rot dargestellt.

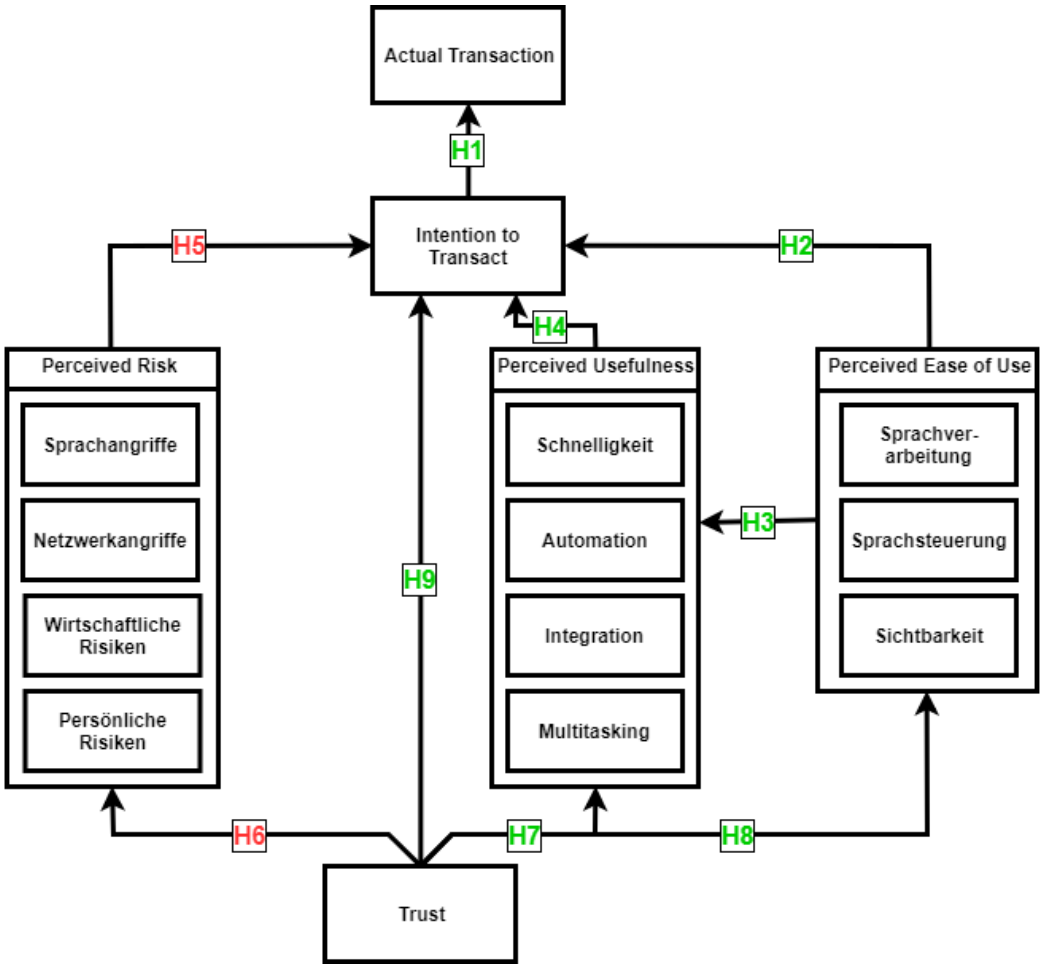

Abb. 4.2 Angepasstes TAM mit Hypothesen und Indikatoren

5. Akzeptanzanalyse von Smart Speakern im E-Commerce

In diesem Kapitel werden zum einen der Fallstudienaufbau dargestellt, die darauf aufbauende Durchführung der empirischen Fallstudie erläutert und gezeigt, welche Ergebnisse entstanden sind und wie diese zu den Hypothesen, welche in Kapitel 4.3 aufgestellt wurden, stehen.

5.1. Aufbau des Experiments

Begonnen wird mit der Definition von Variablen und der anschließenden Ableitung von Fragen für den Fragenkatalog, welcher bei der anstehenden Untersuchung verwendet wird. Außerdem wird der Smart Speaker Skill, welcher zur Demonstration für die empirischen Fallstudie entwickelt wurde, vorgestellt.

5.1.1. Variablen und Fragenkatalog

Aufbauend auf den durch das Technologieakzeptanzmodell definierten Faktoren ist das Ziel Variablen aufzustellen und mit Hilfe dieser in der empirischen Fallstudie Zusammenhänge darzustellen oder zu widerlegen. Entsprechend der Faktoren sehen die Variablen wie folgt aus:

AT	Actual Transaction	**PU**	Perceived Usefulness
IT	Intention to Transact	**PR**	Perceived Risk
PEO	Perceived Ease of Use	**TRU**	Trust

Bei der Auswahl der abhängigen und unabhängigen Variablen, sowie der Erstellung der Fragen und ihrere Skalen, wurde auf bereits bestehende Anwendungen des TAM zurückgegriffen. Grundlage dafür waren die Arbeiten von Pavlou (2003) und van Eeuwen (2017) sowie Davis und Fred (1989). Entsprechend wurden als abhängige Variablen AT und IT, als unabhängige PEO, PU, PR und TRU festgelegt. Die Fragestellungen und Aussagen des Katalogs beziehen sich auf die in Kapitel 4.1 aufgestellten Faktoren, welche Indikatoren der latenten Faktoren des von Pavlou (2003) erstellten TAMs darstellen.

Den Einstieg in den Fragebogen machen allgemeine, bzw. demographische Fragen, welche spätere Einordnungen der Ergebnisse und weitere Analysen ermöglichen und erleichtern sollen.

Fragebereich	Antwortmöglichkeiten
Geschlecht	männlich | weiblich
Alter	jünger als 22 | 22 - 32 | 33 - 45 | älter als 45

Tab. 5.1 Fragenkatalog Demographie

Im Anschluss an diese Fragestellungen folgen erste Untersuchungen, welche auf das Nutzungsverhalten zielen. Diese dienen zur allgemeinen Einordnung der Ergebnisse und sollen Aufschluss über die erste Variable AT bringen.

Fragebereich	Antwortmöglichkeiten
Smart Speaker im Besitz	ja | nein
Häufigkeit der Nutzung	täglich | wöchentlich | monatlich | einmalig | nie |
Skill von Drittanbieter aktiviert	ja | nein
Häufigkeit dessen Nutzung	täglich | wöchentlich | monatlich | einmalig | nie |
(AT 1) Smart Speaker für E-Commerce genutzt	ja | nein
(AT 2) Häufigkeit der Nutzung	tägliche | wöchentliche | monatlich | einmalig | nie |

Tab. 5.2 Fragenkatalog Nutzungsverhalten

Anschließend an diesen Teil wird der Befragte darauf hingewiesen, wie man den Smart Speaker Skill startet und entsprechend der Beispielprozess einmal durchgeführt. Dieser soll die Grundlagen bieten, um die darauf folgenden Aussagen besser einordnen zu können.

Der Hauptteil der Befragung behandelt die Zustimmung bzw. Ablehnung von Aussagen über Grundeinstellungen gegenüber Smart Speakern. Die Antwortmöglichkeiten bestehen dabei aus einer Skala mit fünf verschiedenen Haltungen, welche auf einer Likert-Skala aufbaut und wie folgt aussehen:

stimme zu			stimme nicht zu	
stark	mäßig	neutral	mäßig	stark
2	1	0	-1	-2

Tab. 5.3 Antwortmöglichkeiten auf Grundeinstellungen

Diese Skala dient als Grundlage, um die Haltung der Befragten gegenüber verschiedenen Faktoren der Verwendung von Smart Speakern im E-Commerce zu identifizieren. Zur weiteren Bearbeitung werden die gegebenen Antworten mit Werten (zweite Zeile Tab. 5.3) repräsentiert. Die Aussagen zur Analyse der Akzeptanz sehen dafür wie folgt aus:

Variable	Aussage
IT 1	Ich plane Smart Speaker in Zukunft für Anwendungen des E-Commerces zu nutzen.
IT 2	Ich denke ich werde Smart Speaker in Zukunft öfters für Anwendungen des E-Commerces nutzen.
IT 3	Ich würde anderen Leuten empfehlen Smart Speaker für Anwendungen des E-Commerces zu nutzen.
PEO 1	Smart Speaker zum Einkaufen zu nutzen ist einfach.
PEO 2	Die Sprachsteuerung erleichtert mir den Einkauf über Smart Speaker.
PEO 3	Der Smart Speaker versteht was ich ihm sage.
PEO 4	Der Smart Speaker hat passende Antworten auf die Fragen die ich ihm stelle.
PEO 5	Der Smart Speaker antwortet in einer natürlichen Sprache.
PEO 6	Ich finde passende Skills für die Aufgaben die ich für den Smart Speaker habe.
PU 1	Mit Smart Speakern kann ich schnell Einkäufe tätigen.
PU 2	Mit Smart Speakern kann ich meine Einkäufe leichter automatisieren.

Tab. 5.4 Grundhaltungen gegenüber Smart Speakern, Teil 1

Als Befragtengruppe werden hauptsächlich Personen in der Altersgruppe von 25 bis 34 Jahren erwartet. Dies begründet sich auf Art der Verteilung der Umfrage. Sie wird als Teil eines Experiments zusammen mit Mitarbeitern des Unternehmens dotSource GmbH durchgeführt. Ziel dabei ist es die größte Nutzergruppe widerzuspiegeln[1], da bei dieser erwartet wird, den höchsten Anteil an Nutzern zu erhalten, welche bereits Einkäufe über

[1] zwischen 22 bis 32 Jahren mit fast 50% Nutzeranteil (Buvat et al. 2018, S. 22)

ihre Geräte durchgeführt haben, um auf diese Weise entsprechende Aussagen über das Verhältnis zwischen AT und IT aufstellen zu können. Die geringere Repräsentativität einer solchen Stichprobe wird dabei in Kauf genommen, um entsprechende Aussagen tätigen zu können (Bortz und Döring 2006, S. 483).

Variable	Aussage
PU 3	Die Integrationsmöglichkeiten von Smart Speakern helfen mir meine Einkäufe leichter zu tätigen.
PU 4	Der Smart Speaker erleichtert es mir Einkäufe gleichzeitig, während ich andere Aufgaben erledige, zu tätigen.
PR 1	Ich bin beunruhigt, dass jemand anderes über meinen Smart Speaker Einkäufe tätigen kann.
PR 2	Ich bin beunruhigt, dass meine persönlichen Daten und Gespräche durch Smart Speaker preisgegeben werden.
PR 3	Ich bin beunruhigt, dass persönliche Daten und Gespräche von anderen Personen in meinem Haushalt durch Smart Speaker preisgegeben werden.
PR 4	Ich bin beunruhigt, dass mir durch die Nutzung von Smart Speakern für Einkäufe höhere Kosten entstehen.
TRU 1	Ich vertraue Smart Speakern meine Sprache richtig zu verarbeiten.
TRU 2	Ich vertraue Smart Speakern meine persönlichen Daten sorgsam zu verwenden.
TRU 3	Ich vertraue Smart Speakern passende Maßnahmen zum Schutz meiner Daten einzusetzen.

Tab. 5.5 Grundhaltungen gegenüber Smart Speakern, Teil 2

Das Ende der Befragung bilden zwei offen gestellte Fragen, welche den Befragten die Chance geben, weitere Risiken oder Vorteile von Smart Speakern anzugeben, die in dem Fragebogen nicht dargestellt wurden.

5.1.2. Smart Speaker Skill

Zur Veranschaulichung des Sachverhaltes (Smart Speaker im E-Commerce) wurde eine Anwendung entwickelt, welche einen einfachen Prozess darstellt und auf mehrere Eigenheiten von Smart Speakern eingeht. Zur Reduzierung der Anlernzeit, wie für Smart Speaker Anwendungen entwickelt wird, wurde ein Anbieter entsprechend meiner vorherrschenden Programmierkenntnisse gewählt. Da diese sich hauptsächlich auf Java be-

schränken fiel die Wahl auf einen Alexa Skill. Dieser wurde als Teil des Experiments jedem Befragten während der Bearbeitung des Bogens zur Vermittlung von Grundlagenwissen bereitgestellt.

Als Beispielprozess wurde die Bestellung einer Pizza über einen Alexa Smart Speaker implementiert. Dessen Ziel es ist, dem Nutzer bestimmte Faktoren und kritische Teilprozesse darzustellen und für die möglichen Probleme zu sensibilisieren.

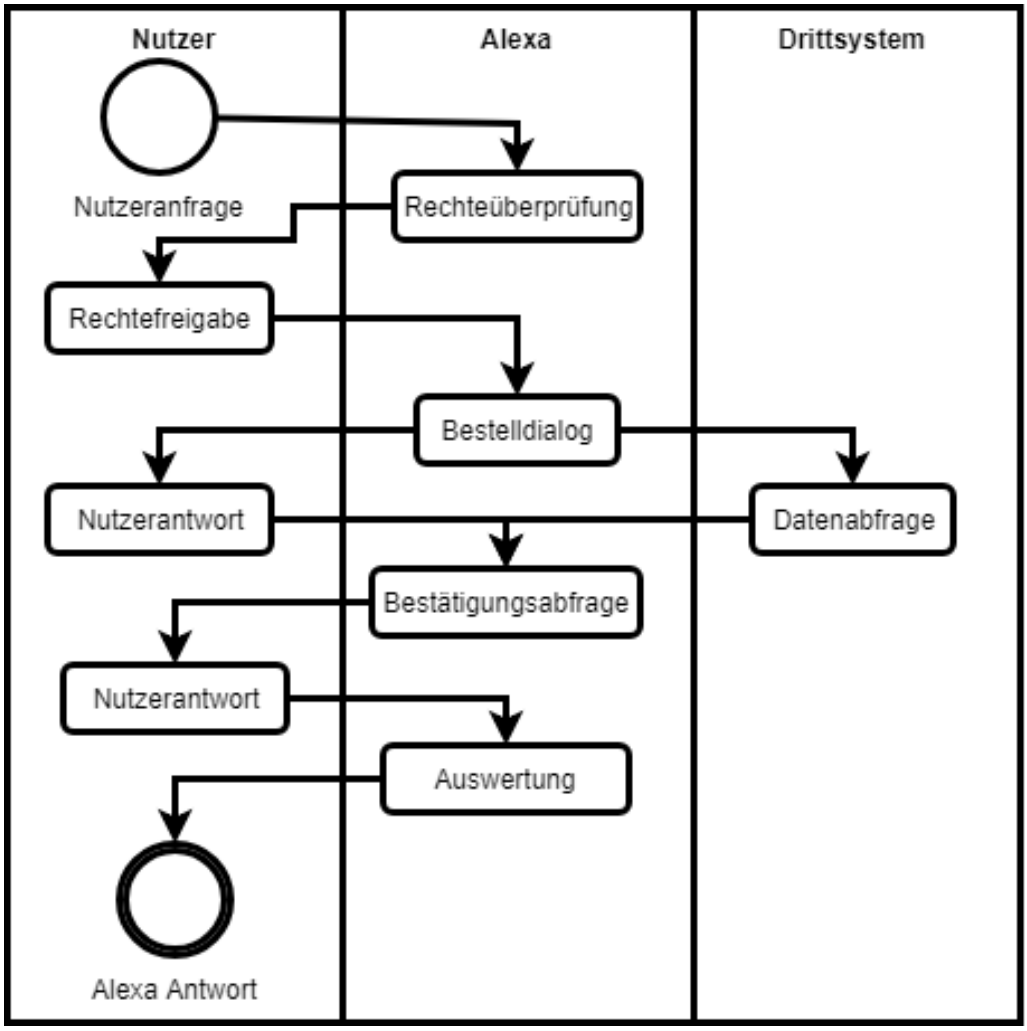

Abb. 5.1 Ablauf des Prozesses

Aus diesem Grund implementiert der Skill beide Möglichkeiten, welche Amazon bereitstellt, um persönliche Daten des Nutzers zu verarbeiten. Zum einen wird über das Account-Linking der Smart Speaker mit dem Amazon-Konto des Benutzers verknüpft, um auf diese Weise an den vollen Namen des Nutzers zu gelangen. Durch eine zweite Rechtefreigabe wird es dem Smart Speaker ermöglicht, nach einer Erlaubnis zu fragen, um an die hinterlegten Daten des Geräts zu gelangen. Sind z. B. Ort und Straße, wo der

Smart Speaker steht, hinterlegt, kann das System durch diese Berechtigung sie auslesen und weiter benutzen. Neben der Verarbeitung der persönlichen Daten wurde außerdem ein Dialog implementiert, welcher die Inhaltsstoffe der Pizza abfragt. Desweiteren war das Ziel, sich an die in Kapitel 4.1.2.1 dargestellten rechtlichen Vorgaben zu halten. Es werden entsprechend § 312d Abs. 1. S. 1 BGB i.V.m. Art. 246a § 3 EGBGB die Kerninformationen Ware, Gesamtpreis und Widerrufsrecht sprachlich mitgeteilt.

Der Prozess in einfacher Form in Abb. 5.1 dargestellt. Den Anfang dessen macht eine Nutzereingabe. In der Alexa Entwicklerkonsole ist dafür ein Schlagwort oder Text anzugeben, auf welchen gehört wird. In diesem Fall wurde „den Test“ festgelegt, sodass der Prozess anfängt, wenn man z. B. „Alexa, starte den Test“ sagt.

Empfängt der Smart Speaker diese Sprachanfrage wird sie zuerst weiter an ein bei Amazon Web Services (AWS) liegendes Amazon Lambda weitergeleitet. In diesem ist ein Handler hinterlegt, welcher auf Anfragen des ASKs (Sprachanfragen mit Hilfe von Alexa gestellt) reagiert. Dessen Aufgabe ist die Überprüfung, ob das Lambda auch von einem autorisierten Alexa Skill aufgerufen wird. Falls nicht wird die Anfrage abgelehnt und verworfen.

Handelt es sich um eine autorisierte Anfrage, wird diese von einem Speechlet weiter verarbeitet. Dieses überprüft in erster Linie, um welche Art von Anfrage es sich handelt. Es können dabei vier Arten unterschieden werden. *SessionStarted* wird automatisch als erstes bei einer Anfrage aufgerufen. Eine entsprechende Methode bearbeitet diese und initialisiert weitere notwendige Daten. In diesem Fall sind dies alle erstellten Handler-Klassen, welche zur weiteren Bearbeitung notwendig sind.

Die zweite mögliche Anfrage trägt die Bezeichnung *Launch*, welche nach dem Aufruf des Schlagwortes bzw. der Phrase aufgerufen wird. Diese enthält im Normalfall eine Begrüßung des Nutzers und legt weitere Möglichkeiten dar. In dem entwickelten Skill wird außerdem überprüft ob der Account verknüpft ist oder Berechtigungen freigegeben wurden und falls nicht, entsprechende Schritte zur Freigabe dieser gestartet.

Die Bearbeitung aller individuellen Anfragen von Nutzern wird durch eine *Intent-Anfrage* durchgeführt. Wie bereits in Kapitel 2.3 erläutert, geben Intents das Anliegen des Benutzer wieder. In der Entwicklerkonsole sind ihr Name und Wörter bzw. Phrasen, welche dieses Verlangen repräsentieren, hinterlegt. In diesem Fall ist der Intent mit dem Namen „Bestellung_Vorbereitet“ hinterlegt und hört auf „Pizza bestellen“. Der Name des Intents wird in der *Intent-Anfrage* mit übergeben und ermöglicht eine gezielte Bearbeitung.

Mit Hilfe einer Dialogstruktur werden weitere Informationen des Nutzers abgefragt. Insgesamt ist es möglich eine Pizza mit drei Inhaltsstoffen zu bestellen, der Dialog wird beendet, wenn Nutzer mindestens einen angegeben. Synonyme, wie z. B. Pizza Hawaii für eine solche mit Schinken und Ananas, werden nicht erkannt. Sind diese Informationen vorhanden wird erneut eine *Intent* Anfrage mit dem Namen „Bestellung_Vorbereitet“ aufgerufen, da jedoch alle Dialogmöglichkeiten erschöpft sind wird der Dialogstatus auf beendet gesetzt.

Dies signalisiert dem System, dass eine Bestellung mit den vorliegenden Daten vorzubereiten ist. Zur Anreicherung dieser wird mit Hilfe der zuvor gegebenen Berechtigungen die Adresse des Gerätes abgerufen und durch die Accountverknüpfung der Name des Nutzers identifiziert. Diese Daten werden dann, zusammen mit einem zufälligen Preis zwischen 4 und 8 € dem Nutzer sprachlich durch Alexa und in einer Karte auf der Alexa App ausgegeben. Außerdem wird der Nutzer darauf hingewiesen, dass eine entsprechende Karte ausgegeben wurde und in ihr weitere Schritte zum Abschließen oder Abbrechen der Bestellung bestehen.

Durch das Sprechen von „Bestellung abbrechen“ bzw. „Bestellung durchführen“ kann entsprechend weiter vorgegangen werden. Eine *Intent-Anfrage* mit dem Namen „Bestellung_Abgebrochen“ wird bei der ersten Phrase geliefert. Dieser ruft eine entsprechende Klasse auf, welche eine einfache Alexa Sprach- und Kartenantwort liefert, dass die Bestellung abgebrochen wurde.

Möchte der Benutzer die Bestellung durchführen enthält die *Intent-Anfrage* Anfrage den Namen „Bestellung_Durchgefuehrt“ als Namen. Die Daten der Bestellung wurden zuvor der Session als Attribut hinzugefügt, um eine weitere Nutzung zu ermöglichen. Dies wird von einer durch den Intent aufgerufenen Klasse genutzt und eine Nachricht mit identischen Informationen, wie die zuvor vorbereitete Bestellung, sowie der Nachricht, dass die Bestellung durchgeführt wurde, ausgegeben.

Die letzte Art der Anfrage trägt den Namen *SessionEnded* und wird entsprechend aufgerufen, wenn die Bearbeitung durch Alexa endet. In der dafür vorgesehenen Methode sind im Normalfall Aufräumarbeiten durchzuführen, welche in dem entwickelten Skill jedoch nicht notwendig sind.

Der Skill orientiert sich an dem in Kapitel 4.2 dargestellten Prozess für Transaktionen und den dargestellten Anpassungen für Smart Speaker. Eine Auswahl des Unternehmens wird durch die Aktivierung des Skills dargestellt. Der Transfer von Informationen und persönlichen Daten findet, wie auch bei dem durch Amazon implementierten Einkaufsprozess,

bereits vor der eigentlichen Bestellung statt. Auch die Informationsmöglichkeiten sind zu einem Minimum gehalten. Es steht nur eine Hilfefunktion zur Verfügung, welche Anregungen gibt, welche Phrasen zur welchem Ergebnis führen.

Es wurde außerdem versucht natürliche Sprache nach der Android Metapher von Edlund et al. (2008) zu implementieren, sowie Mixed-Initivative Dialoge darzustellen. Ersteres wurde dadurch realisiert, dass nach Aufruf der Phrase „Alexa, starte den Test“, die Antwort von Alexa sowie mögliche Hilfestellungen menschenähnlich gestaltet wurden. Entsprechend wurden je nach Aufgabe voll ausformulierte Sätze oder kurze Antworten gebene. Startet man daraufhin den Dialog zur Bestellung und Auswahl von Inhaltsstoffen, verhält sich Alexa wie ein Interface, gibt dem Nutzer an welche und wieviele Eingaben möglich sind und wartet auf eine entsprechende Antwort. Nach dieser Frage durch den Smart Speaker wird dem Nutzer die Initiative des Dialogs übergeben, indem er die Bestellung abbrechen kann, oder ihr zustimmt und sie damit durchführt. Auf diese Weise soll ein natürlicher Dialogverlauf nach Kapitel 4.1.1.1 dargestellt werden.

Durch diese Maßnahmen, die Abfrage von persönlichen Daten durch Berechtigungen und Accountverknüpfungen, sowie die Wahrung der Informationspflicht nach § 312d Abs. 1. S. 1 BGB i.V.m. Art. 246a § 3 EGBGB, werden durch den Skill wichtige Faktoren der Akzeptanz von Smart Speakern im E-Commerce abgebildet. Auf diese Weise soll gewährleistet werden, dass die befragte Person ausreichende Informationen über den Sachverhalt erhält.

5.2. Empirische Ergebnisse

Mit Hilfe einer Fokusgruppe (N=5) wurde der allgemeine Rahmen der Untersuchung validiert. In einem geschlossenem Umfeld wurde das Experiment mit Alexa-Skill und Umfrage mit diesen Probanden durchgeführt und erste Ergebnisse erzielt. Es handelte sich dabei um ein geführtes Experiment, bei dem Hilfestellungen bereitstanden, falls im Umgang mit der Technologie Probleme auftraten.

Die Ergebnisse wurden verwendet, um mit Hilfe von Cronbachs α (Alpha), die interne Konsistenz der verwendeten Skala zu untersuchen. Dieses wurde mit Hilfe des Programms IBM Statistical Package for the Social Sciences (SPSS) Statistics ermittelt und beträgt α = 0,793. Dieser Wert spricht für eine gute interene Konsistenz der zugrundegeliegenden Daten und Skalen. Weitere α der einzelnen Variablen sehen wie folgt aus:

Variable	Cronbachs α
IT	0,929
PEO	0,827
PU	0,774
PR	0,669
TRU	0,620

Tab. 5.6 Cronbachs α der einzelnen Variablen

Die Werte für IT, PEO und PU stellen akzeptable bis sehr gute Werte für die interne Konsistenz dieser Variablen dar, sodass keine weitere Betrachtungen notwendig sind. Werte zwischen 0,6 und 0,7 sprechen von einer ansatzweise problematischen Datenreihe, welche eine weitere Betrachtung bedingt. Durch Elimination des Indikators PR4 konnte der Wert des Alphas auf einen Wert von $\alpha = 0{,}821$ gehoben werden. Ein ähnlicher Effekt konnte identifiziert werden, wenn man den Indikator TRU1 eliminiert, wodurch der Wert auf $\alpha = 0{,}768$ gehoben wurde. Entsprechend wurden beide Indikatoren für weitere Betrachtungen nicht mehr berücksichtigt.

Die verwendete Umfrage wurde erstellt und durchgeführt mit den Funktionen von Google Forms. Der Skill wurde auf ein Amazon Echo, welches von dem Unternehmen dotSource GmbH bereitgestellt wurde, installiert und auch mit diesem in dem Experiment bereitgestellt.

Als weitere Ergebnisse der Fokusgruppe wurde eine Frage, welche sich auf das Uncanny Valley bezog, aus dem Katalog entfernt. Probanden der Fokusgruppe zeigten Probleme bei dem Verständnis der Frage und dadurch bei der Beantwortung dieser auf und erst nach weiterer Erläuterung während des geführten Experiments, konnten Antworten gegeben werden.

Außerdem wurde als Reaktion auf wiederholt auftretende Fragen während des Experiments die Grundhaltung PU1 in zwei Teilbereiche aufgeteilt. Zum einem in die Aussage, dass Smart Speaker schneller als Webseiten zu benutzen sind und zum anderen, dass sie schneller als ein Einkauf im Einzelhandel sind. Erste Probleme beim Aufbau des Dialog-Modells konnten auf diese Weise auch identifiziert werden. Nachdem unerfahrene Probanden die ersten Worte mit dem Skill gewechselt hatten, wirkten sie oft hilflos und benötigten Unterstützung, um den Dialog voranschreiten zu lassen. Eine Anpassung, dass Alexa dem Nutzer in der Begrüßung mitteilt, welche Aussage weiterführend ist, konnte dieses Problem im weiteren Experiment entfernen.

Als letztes Ergebnis der Fokusgruppe, wurde die Stichprobengröße der Hauptuntersuchung festgelegt. Diese wurde nach Bortz und Döring (2006) mit N = 22 festgelegt. Hierbei war grundlegend eine gewählte Effektstärke δ mit 0,5 (ein mindestens großer Effekt wird zugrunde gelegt) und $\bar{r}$ aus der Fokusgruppe mit einem Wert von 0,49, sowie die Verwendung eines α mit einer Signifikanz von 95%. Diese Wahl in Kombination mit der Tabelle der optimalen Stichprobenumfänge von Bortz und Döring (2006), führte der entsprechenden Stichprobenwahl (vgl. S. 628 Tab. 9.7, 2)). Als Folge dieser Wahl können jedoch nur Zusammenhänge mit einer hohen Korrelation, also $r \leq -0,5$ oder $r \geq 0,5$ als signifikant gewertet werden (Cohen 1988, S. 40-42).

Von den 22 Teilnehmern des Experiments waren exakt 11 männlich sowie 11 weiblich. Die Altersgruppe von 22 bis 32 Jahren war die, wie bereits im Design geplant, am meisten vertretene mit 15 Teilnehmern. Nur zwei Probanden waren jünger als 22 Jahre und fünf weitere lagen in der Altersgruppe von 33 bis 45 Jahren lagen. Personen im Alter über 45 Jahren wurden nicht befragt.

	Anzahl	**%**	**% Summiert**
Geschlecht			
männlich	11	50%	50%
weiblich	11	50%	100%
Alter			
jünger als 22	2	9,09%	9,09%
22 bis 32	15	68,18%	77,27%
33 bis 45	5	22,73%	100%
älter als 45	0	0	100%

Tab. 5.7 Demografische Übersicht der Ergebnisse

Nach den allgemein demografischen Fragen folgten erste, welche das bisherige Nutzungsverhalten mit Smart Speakern untersuchten. Von den 22 befragten Personen besitzen sechs mindestens einen Smart Speaker. Dies spiegelt das Nutzungsverhalten aktueller Studien, wie jene von Voicebot (2018) wieder, welche ca. 20% der der Befragten mit Zugang zu Smart Speakern identifiziert haben. In der selben Untersuchung konnte auch dargestellt werden, dass Smart Speaker Nutzer oft mehr als ein Gerät besitzen[2]. Ein entsprechendes Verhalten konnte auch hier identifiziert werden, wobei von den sechs Smart Speaker

[2] 19,3% zwei Geräte, 8% drei Geräte und 7% vier Geräte (Kinsella und Mutchler 2018)

Nutzern vier mehrere Geräte verwenden. Drei besaßen einen Echo und Echo Dot, sowie eine Personen einen Google Home und Google Home Mini. Kein Befragter hatte mehrere Geräte verschiedener Anbieter.

Die Nutzungshäufigkeit der Smart Speaker lag etwas über den bisher dargestellten Ergebnissen anderer Studien. Von den sechs Smart Speaker Nutzern verwenden fünf ihr Gerät täglich und eine Person wöchentlich. Keiner der Probanden nutzt sein Gerät nur selten, wie z. B. 12,7% der Befragten der Voicebot Studie (2018) angaben.

	Anzahl	%	% Summiert
Besitz eines Smart Speakers			
ja	6	22,73%	22,73%
nein	16	77,27%	100%
Smart Speaker Lösung			
Amazon Echo	3	30%	30%
Amazon Echo Dot	4	40%	70%
Google Home	1	10%	80%
Google Home Mini	2	20%	100%
Nutzung ihres Smart Speakers			
täglich	5	83,33%	83,33%
wöchentlich	1	16,67%	100%
Nutzung Drittanbieter-Skills			
täglich	1	20%	20%
wöchentlich	2	40%	60%
einmalig	1	20%	80%
nie	1	20%	100%
Wege der Skills-Entdeckung			
Stores	2	$33.\overline{3}\%$	$33.\overline{3}\%$
Freunde	2	$33.\overline{3}\%$	$66.\overline{6}\%$
Werbung	1	$16.\overline{6}\%$	$83.\overline{3}\%$
Newsletter	1	$16.\overline{6}\%$	100%

Tab. 5.8 Smart Speaker Nutzung, Besitz und Skillaktivierung

Auch bei der Aktivierung der Skills von Drittanbietern lagen die Befragten dieser Studie in ihren Antworten über den Ergebnissen der bisherigen Studien. Von den sechs Nutzern von Smart Speakern hatten alle außer einer Person bereits einen weiteren Skill selber aktiviert, was gegen die sonst beschriebenen 48%, welche von Voicebot (2018) identifiziert wurden, einen starken Unterschied aufweist.

Die Art und Weise, wie neue Skills entdeckt werden, ist aufgrund der geringen Antwortzahlen nur schwer mit bisherigen Ergebnissen vergleichbar. Bei der Beantwortung der Frage waren Mehrfachantworten möglich, sodass als Ergebnisse zwei Mal Freunde, zwei Mal die Stores, zwei mal Werbung und ein Mal Newsletter angegeben wurden. Eine Ähnlichkeit zu anderen Ergebnissen (Kinsella und Mutchler 2018, vgl. S. 24) ist bereits durch den Umstand, dass alle Nutzer auch weitere Skills aktiviert haben nicht zu erkennen. An dieser Stelle unterscheiden sich die empirischen Ergebnisse zu denen der durch Studien bisher aufgezeigten eindeutig.

Die Nutzung der Skills von Drittanbietern unterscheidet sich zu den von Haus aus in die Geräte integrierten stark. Nutzen fünf der sechs Smart Speaker Besitzer ihr Gerät täglich, verschiebt sich dieses Verhalten bei den aktivierten Skills hin zu einer geringeren Nutzung. Nur noch eine Person gab bei dieser Frage eine tägliche Nutzung an, zwei weitere gaben eine wöchentliche Verwendung an, eine weitere hat den aktivierten Skill nur einmalig genutzt und es kam sogar dazu, dass ein solcher Skill in einem Fall nie genutzt wurde. Entsprechende Betrachtungen wurden in keiner der drei bisherigen Studien durchgeführt, sodass ein Vergleich hier nicht möglich ist.

Ein weiteres Ergebnis des Experiments ist, dass unter den 22 Teilnehmern keine Person war, welche bisher eine Transaktion mit ihrem Gerät durchgeführt hat. Drei der Befragten, welche täglich ihren Smart Speaker nutzen und auch mindestens wöchentlich einen Skill von Drittanbietern nutzen, hatten dafür verschiedene Begründungen. Es war entweder zu umständlich, nicht für die Produkte die sie kaufen wollen passend oder nicht möglich (im Falle eines Google Nutzers).

Als direktes Ergebnis dieses Umstands konnte kein Zusammenhang zwischen den Variablen AT und IT untersucht werden. Die Aussage IT2 zur Grundeinstellung von Smart Speakern und des zukünftig steigenden Einkaufsverhalten mit diesen Geräten, entfiel dadurch vollständig. Die restlichen Variablen, ihre Indikatoren sowie Relationen untereinander sind in Abb. 5.2 zu sehen. Da es sich um eine symmetrische Matrix handelt wurde der Übersicht halber nur eine Seite angegeben. Grün markierte Zellen enthalten signifikante positive ($r \geq 0,5$) und rot markierte entsprechend signifikante negative Zusammenhänge ($r \leq -0,5$). Die Ergebnisse wurden mit Hilfe der von Google Docs bereitgestell-

ten CORR-Funktion erzielt und mit der entsprechenden Analysemethode von IBM SPSS Statistics verglichen. Das Ergebnis beider Methoden ist der Pearson product-moment correlation coefficient, welcher mit dem Symbol *r* dargestellt wird. Er gibt an, wie groß der Grad des linearen Zusammenhangs von zwei Merkmalen ist.

	IT1	IT3	PEO1	PEO2	PEO3	PEO4	PEO5	PEO6	PEO7	PU1	PU2	PU3	PU4	PR1	PR2	PR3	TRU2	TRU3
IT1	1.00	0.87	0.50	0.50	0.25	0.33	0.14	0.65	0.53	-0.21	0.50	0.42	0.44	-0.44	-0.48	-0.36	0.62	0.52
IT3		1.00	0.67	0.64	0.42	0.47	0.33	0.73	0.53	-0.30	0.53	0.34	0.50	-0.31	-0.50	-0.38	0.64	0.52
PEO1			1.00	0.68	0.47	0.33	0.63	0.66	0.61	0.00	0.50	0.31	0.67	-0.01	-0.33	-0.22	0.50	0.54
PEO2				1.00	0.57	0.51	0.67	0.66	0.54	-0.09	0.54	0.31	0.45	-0.15	-0.17	-0.01	0.35	0.28
PEO3					1.00	0.72	0.42	0.32	0.15	0.22	0.34	0.10	0.04	0.21	-0.02	0.08	0.23	-0.13
PEO4						1.00	0.38	0.54	0.31	0.00	0.28	0.30	0.19	0.12	-0.24	-0.08	0.53	0.23
PEO5							1.00	0.45	0.45	0.31	0.46	0.20	0.53	-0.05	-0.21	0.03	0.24	0.39
PEO6								1.00	0.67	0.12	0.41	0.48	0.51	-0.20	-0.56	-0.35	0.70	0.65
PEO7									1.00	0.18	0.47	0.49	0.50	-0.30	-0.41	-0.35	0.43	0.63
PU1										1.00	0.09	-0.09	0.10	0.47	0.28	0.89	0.12	-0.10
PU2											1.00	0.28	0.42	-0.29	-0.22	-0.07	0.14	0.23
PU3												1.00	0.30	-0.06	-0.13	0.06	0.49	0.48
PU4													1.00	-0.10	-0.13	0.05	0.38	0.52
PR1														1.00	0.46	0.47	0.11	-0.43
PR2															1.00	0.90	-0.37	-0.64
PR3																1.00	-0.28	-0.56
TRU2																	1.00	0.66
TRU3																		1.00

Abb. 5.2 Korrelationsmatrix

5.3. Diskussion der Hypothesen

Die zuvor dargestellten empirischen Ergebnisse und das in Kapitel 4.3 dargestellte Modell wurden mit Hilfe des Programms IBM SPSS Analysis of Moment Strucutres (AMOS) 25 modelliert, um weitere Erkenntnisse zu erlangen. Zu diesem Ziel wurde die Maximum-Likelihood-Methode verwendet, um Beziehungen zwischen den latenten Variablen zu identifizieren und Rückschlüsse auf Indikatoren zu ziehen.

Die neun in Kapitel 4.3 aufgestellten Hypothesen können mit Hilfe der auf diese Weise ermittelten Ergebnisse, welche in Tab. 5.9 dargestellt werden, diskutiert werden. Die Regressionsgewichte zwischen den entsprechenden Variablen geben dabei an, in welcher Art der Relation diese Variablen stehen und die Stärke dieser. Eine weitere Betrachtung, welcher der Indikatoren relevant für diese Ergebnisse ist, wird mit Hilfe der Korrelations-

koeffizienten aus Abb. 5.2, zusammen mit den Regressionsgewichten der Indikatoren aus Tab. 5.10 durchgeführt. Letztere zeigen auf, welcher Faktor relevant zur Erklärung einer latenten Variable ist und wie stark eine Änderung dessen auf die Variable einwirkt.

Hypothese	Variablen	Regressionsgewichte
H1	$IT \rightarrow AT$	-
H2	$PEO \rightarrow IT$	0,192
H3	$PEO \rightarrow PU$	0,636
H4	$PU \rightarrow IT$	0,688
H5	$PR \rightarrow IT$	-0,327
H6	$TRU \rightarrow PR$	-0,633
H7	$TRU \rightarrow PU$	0,506
H8	$TRU \rightarrow PEO$	0,385
H9	$TRU \rightarrow IT$	-0,296

Tab. 5.9 Hypothesen und Relationen der entsprechenden Variablen

Variablen	Gewichte	Variablen	Gewichte
$IT1 \rightarrow IT$	0,873	$IT2 \rightarrow IT$	0,997
$PEO1 \rightarrow PEO$	0,800	$PEO2 \rightarrow PEO$	0,897
$PEO3 \rightarrow PEO$	0,600	$PEO4 \rightarrow PEO$	0,521
$PEO5 \rightarrow PEO$	0,672	$PU1 \rightarrow PU$	0,863
$PU2 \rightarrow PU$	0,745	$PU3 \rightarrow PU$	0,491
$PU4 \rightarrow PU$	0,521	$PU5 \rightarrow PU$	0,692
$PR1 \rightarrow PR$	0,507	$PR2 \rightarrow PR$	0,959
$PR3 \rightarrow PR$	0,931	$TRU2 \rightarrow TRU$	0,609
$TRU3 \rightarrow TRU$	1,023		

Tab. 5.10 Relationen zwischen Indikatoren und Variablen

Die erste Hypothese befasst sich mit dem Verhältnis zwischen den Variablen AT und IT, welche mit Hilfe der Indikatoren AT1 und AT2 sowie IT1 und IT2 überprüft werden sollten.

H1 Die Kaufabsicht des Verbrauchers hat einen positiven Einfluss auf das wirkliche Transaktionsverhalten mit Smart Speakern.

Aufgrund der fehlenden Probanten, welche Transaktionen mit ihrem Smart Speaker durchgeführt haben, kann keine Aussage zu dieser Hypothese getroffen werden.

H2 Die wahrgenommene Bedienungsfreundlichkeit von Smart Speakern hat einen positiven Einfluss auf die Kaufabsicht über diese.

Zur Darstellung der Bedienungsfreundlichkeit wurden fünf Indikatoren aufgestellt, welche die allgemeine Einfachheit der Bedienung (PEO1), die Verwendung der Sprachsteuerung (PEO2), die SeR (PEO3), die Verarbeitung der Anfragen durch Skills (PEO4) und die natürliche Sprachsynthese (PEO5) untersuchten. Die Korrelation dieser Faktoren mit den beiden Indikatoren der Intention to Transact sind in Abb. 5.3 dargestellt.

	PEO1	PEO2	PEO3	PEO4	PEO5
IT1	0.50	0.50	0.25	0.33	0.14
IT3	0.67	0.64	0.42	0.47	0.33

Abb. 5.3 Korrelationsmatrix Hypothese 2

Mit Regressionswerten von 0,800 und 0,897 sind PEO1 und PEO2 jene Indikatoren, welche die stärksten Änderungen in der wahrgenommenen Bedienungsfreundlichkeit hervorrufen. Die weiteren Faktoren sind jedoch auch alle noch in einem Bereich, in welchem sie Einfluss auf diese Wahrnehmung ausüben können. Den größten Zusammenhang zur Absicht E-Commerces über Smart Speaker zu betreiben, bilden erneut Faktoren PEO1 und PEO2 dar mit Korrelationswerten von 0,5 und 0,67 bzw. 0,5 und 0,64.

Mit einem Regressionsgewicht von 0,192 zwischen den latenten Variablen PEO und IT ist ein leichter positiver Zusammenhang identifiziert, welcher die Grundlage bildet, dass H2 als belegt gewertet wird. Als wichtigste Faktoren für Smart Speaker im E-Commerce werden von der wahrgenommenen Bedienungsfreundlichkeit eine allgemeine einfache Bedienung und die Verwendung der Sprachsteuerung erklärt. Die weiteren Indikatoren sollten jedoch nicht vernachlässigt werden, da sie zum einen grundlegenden Einfluss auf die wahrgenommene Bedienungsfreundlichkeit besitzen (vgl. Regressionswerte von PEO3-5) und auch die Korrelationswerte von PEO3 und PEO4 nur gering unter der als signifikant definierten Schwelle lagen.

H3 Die wahrgenommene Bedienungsfreundlichkeit von Smart Speakern hat einen positiven Einfluss auf deren wahrgenommene Nützlichkeit.

Die wahrgenommene Nützlichkeit wurde mit Hilfe verschiedener Indikatoren untersucht. Diese sind Schnelligkeit gegenüber Webseiten (PU1), Schnelligkeit gegenüber Einzelhandel (PU2), Automation (PU3), Integration (PU4) und Multitasking (PU5).

	PU1	PU2	PU3	PU4	PU5
PEO1	0.66	0.61	0.50	0.31	0.67
PEO2	0.66	0.54	0.54	0.31	0.45
PEO3	0.32	0.15	0.34	0.10	0.04
PEO4	0.54	0.31	0.28	0.30	0.19
PEO5	0.45	0.45	0.46	0.20	0.53

Abb. 5.4 Korrelationsmatrix Hypothese 3

Mit einem Regressionsgewicht von 0,636 zwischen den latenten Variablen PEO und PU ist eindeutig ein positiver Zusammenhang zwischen diesen dargestellt. Die stärksten Ladungen auf PEO wurden bereits zuvor dargestellt, für PU stellen diese PU1 mit 0,863 und PU2 mit 0,745 dar. Auch bei den Korrelationswerten stellen diese, erneut mit den zuvor identifizierten Indikatoren für PEO, die wichtigsten dar. PU5, mit einer Ladung von 0,692 und starken Korrelationen mit PEO1 und PEO5 sollte jedoch auch nicht vernachlässigt werden. Mit diesen Ergebnissen ist festzuhalten, dass die schnelle Verwendung von Smart Speakern zusammen mit ihren Möglichkeiten Multitasking zu betreiben, den größten Zusammenhang mit der wahrgenommenen Bedienungsfreundlichkeit aufweisen, wobei diese sich dabei hauptsächlich in einer allgemein einfachen Bedienung und durch die Sprachsteuerung widerspiegelt.

H4 Die wahrgenommene Nützlichkeit von Smart Speaker hat einen positiven Einfluss auf die Kaufabsicht über diese.

Die Indikatoren zur Darstellung der Kaufabsicht und der wahrgenommenen Nützlichkeit wurden bereits vorgestellt. Die Korrelationswerte zwischen diesen sind aus Abbildung 5.5 entnehmbar.

	PU1	PU2	PU3	PU4	PU5
IT1	0.65	0.53	0.50	0.42	0.44
IT3	0.73	0.53	0.53	0.34	0.50

Abb. 5.5 Korrelationsmatrix Hypothese 4

Mit einem Regressionsgewicht von 0,688 ist der stärkste positive Zusammenhang zwischen zwei latenten Variablen dargestellt. Zwischen fast allen Indikatoren konnten hierbei signifikante positive Korrelationen berechnet werden. Am meisten abhängig ist die Kaufentscheidung von den bereits zuvor als entscheident identifizierten Faktoren PU1 und PU2. Im Zusammenhang mit der geplanten Kaufentscheidung sind die Indikatoren PU3 und PU5 auch relevant, wodurch die Kaufentscheidung hauptsächlich durch die Geschwindigkeit beeinflusst wird, jedoch Integration und Multitasking auch eine Rolle spielen.

H5 Das wahrgenommene Risiko bei der Nutzung von Smart Speakern hat einen negativen Einfluss auf die Kaufabsicht.

Mit einem Regressionsgewicht von -0,296 konnte eine leichte negative Relation zwischen den latenten Variablen PR und IT nachgewiesen werden, was Hypothese 5 unterstützt. Als Indikatoren für das wahrgenommene Risiko wurden dafür Einkäufe durch Dritte (PR1), Freigabe eigener persönlicher Daten (PR2) und die Freigabe der persönlichen Daten anderer Personen im Haushalt (PR3), verwendet.

	PR1	PR2	PR3
IT1	-0.44	-0.48	-0.36
IT3	-0.31	-0.50	-0.38

Abb. 5.6 Korrelationsmatrix Hypothese 5

Die stärksten Ladungen auf PR haben die Indikatoren PR2 mit 0,959 und PR3 mit 0,931. Dieses Ergebnis wird zum Teil durch die Korrelationsergebnisse aus Abb. 5.6 unterstützt, indem PR2 als einziger Faktor signifikante negative Korrelationen erzielen konnte. Die entsprechenden Werte von PR1 und PR3 zeigen einen negativen Zusammenhang auf, welcher jedoch nicht siginifikant ist. Damit ist der Hauptgrund, welcher gegen eine Kaufentscheidung spricht, das Risiko, dass die eigenen Daten freigegeben werden könnten.

H6 Das Vertrauen in Smart Speaker hat einen negativen Einfluss auf deren wahrgenommenes Risiko.

Mit einem Regressionsgewicht von -0,633 und dem damit stärksten negativen Ergebnis zwischen zwei latenten Variablen, ist ein negatives Verhältnis zwischen PR und TRU nachgewiesen und damit Hypothese 6 unterstützt. Als Indikatoren für das Vertrauen wurden die Verwendung der persönlichen Daten (TRU2) und der Schutz der Daten (TRU3) verwendet.

Mit Ladungen von 0,609 für TRU2 und 1,023 für TRU3, ist letzterer eindeutig als wichtigster Faktor des Vertrauens im Zusammenhang mit dem wahrgenommenen Risiko identifiziert. Dies wird auch durch die Korrelationswerte verdeutlicht, wobei TRU3 zwei signifikant negative Korrelationen mit PR2 und PR3 besitzt. Mit einem Ladungswert von ca. 1 ist eine Veränderung in dem Schutz der Daten gleich mit einer allgemeinen Änderung des Vertrauens gegenüber Smart Speakern im E-Commerce gleichzusetzen.

	TRU2	TRU3
PR1	0.11	-0.43
PR2	-0.37	-0.64
PR3	-0.28	-0.56

Abb. 5.7 Korrelationsmatrix Hypothese 6

H7 Das Vertrauen in Smart Speaker hat einen positiven Einfluss auf deren wahrgenommene Nützlichkeit.

Das positive Verhältnis zwischen dem Vertrauen und der wahrgenommenen Nützlichkeit wird mit Hypothese 7 untersucht. Dieser Zusammenhang konnte mit einem Regressionsgewicht von 0,506 zwischen TRU und PU eindeutig nachgewiesen werden. Als wichtige Indikatoren konnten bisher TRU3, PU1 und PU2 sowie in Teilen PU5 nachgewiesen werden.

	TRU2	TRU3
PU1	0.70	0.65
PU2	0.43	0.63
PU3	0.14	0.23
PU4	0.49	0.48
PU5	0.38	0.52

Abb. 5.8 Korrelationsmatrix Hypothese 7

Dieses Ergebnis wird auch durch die Korrelationsmatrix der für Hypothese 7 relevanten Indikatoren unterstützt. TRU3 hat erneut mit fast jedem Faktor der wahrgenommenen Nützlichkeit signifikant positive Relationen. PU1 und PU2 zeigen signifikante Zusammenhänge zum Vertrauen auf und auch PU5 kann dies nachweisen. Als interessanter Faktor der Nützlichkeit kann an dieser Stelle auch PU4 aufgenommen werden. Mit Werten

von 0,49 und 0,48 liegt dieser Indikator nur gering unter dem Schwellwert, sodass die Integrationsmöglichkeiten von Smart Speakern leichte Zusammenhänge mit dem Vertrauen in diese aufzeigen können.

H8 Das Vertrauen in Smart Speaker hat einen positiven Einfluss auf die wahrgenommene Bedienungsfreundlichkeit.

Die Untersuchung des positiven Zusammenhangs zwischen der wahrgenommenen Bedienungsfreundlichkeit und dem Vertrauen in Smart Speakern ist Gegenstand von Hypothese 8. Mit einem Regressionsgewicht von 0,385 konnte dieser dargestellt und somit H8 belegt werden.

	TRU2	TRU3
PEO1	0.50	0.54
PEO2	0.35	0.28
PEO3	0.23	-0.13
PEO4	0.53	0.23
PEO5	0.24	0.39

Abb. 5.9 Korrelationsmatrix Hypothese 8

Für diesen Zusammenhang am relevantesten sind nach Abb. 5.9 PEO1 und PEO4. Dies legt die Vermutung nah, dass eine Steigerung des Vertrauens des Nutzers dazu führt, dass er die Nutzung von Smart Speakern zum Einkaufen im Allgemeinenen als leichter ansieht und auch die Antworten in den Dialogen mit den Geräten als passender empfindet.

H9 Das Vertrauen in Smart Speaker hat einen positiven Einfluss auf die Kaufabsicht mit diesen.

Die letzte Hypothese untersucht den Zusammenhang zwischen dem Vertrauen der Nutzer in Smart Speaker und der geplanten Kaufabsicht. Die Korrelationsmatrix zwischen den vier untersuchten Indikatoren zeigt ein klares Bild, indem alle *r* ein positiv signifikantes Ergebnis aufzeigen. Der Schluss liegt nahe, dass die Hypothese bestätigt werden könnte.

	TRU2	TRU3
IT1	0.62	0.52
IT3	0.64	0.52

Abb. 5.10 Korrelationsmatrix Hypothese 9

Die Regressionsgewichte zwischen den latenten Variablen TRU und IT sprechen jedoch entgegen der Hypothese von einem negativen Verhältnis mit einem Wert von -0,296. Dies scheint im ersten Moment widersprüchlich, dass positive Korrelationen zwischen den Indikatoren einen negativen Zusammenhang zwischen den Variablen zur Folge haben. Erklärbar ist dieses Verhalten mit der Negativ Net Suppression nach Cohen (1975). Die Indikatoren verringern gegenseitig ihre Varianz und ermöglichen so eine positive Korrelation bei einer negativen Regression (Cohen und Cohen 1975, S. 84-91). Somit muss Hypothese 9 widerlegt und ein sogar umgekehrtes Verhalten als vorher angenommen bezeugt werden.

Abschließend werden alle relevanten Faktoren und in einer Übersicht dargestellt und ihre Relevanz für die Verwendung von Smart Speakern im E-Commerce diskutiert. Dafür werden die einzelnen Indikatoren der Wichtigkeit absteigend von oben nach unten in Tab. 5.11 dargestellt.

PEO	PU	PR	TRU
PEO1	PU1	PR2	TRU2
PEO2	PU2	PR3	TRU1
PEO5	PU5	PR1	
PEO4	PU4		
PEO3	PU3		

Tab. 5.11 Übersicht Variablen und Indikatoren

Mit den wichtigsten der Freigabe und dem Schutz der Daten als wichtigste Faktoren des wahrgenommenen Risikos und Vertrauens, liegt die Hauptverantwortung dieser in der Hand der Smart Speaker Hersteller. Es gilt eine Grundlage zu schaffen, auf welchen die Vorteile und Stärken der Geräte wahrgenommen werden können, um mögliche weitere Anwendungen im E-Commerce zu ermöglichen.

Den Nutzern ist es wichtig schnell und einfach über diese Geräte einkäufe tätigen zu können. Diese Faktoren müssen Entwickler in Betracht beziehen, wenn sie entsprechende Anwendungen für diese Geräte implementieren. Es gilt außerdem, die besonderen Möglichkeiten der Sprachsteuerung und dadurch entstehenden Multitasking Chancen auszuschöpfen. Diese bilden die wichtigstens Faktoren, welche im Zusammenhang mit Smart Speakern im E-Commerce identifiziert werden konnten. Ob dieser Zusammenhang nur eine Korrelation ist, oder es sich wirklich um eine Kausalität handelt, muss Gegenstand weiterer Untersuchungen werden.

6. Zusammenfassung

In einem letzten Kapitel werden die Arbeit und ihre Ergebnisse noch ein mal zusammengefasst, die verwendete Vorgehensewise kritisch betrachtet und untersucht, ob das Ziel der Arbeit erreicht wurde.

6.1. Überblick

In dieser Masterarbeit wurde das Thema der Smart Speaker als Anwendung im E-Commerce betrachtet und dafür im Speziellen untersucht, welche Faktoren relevant sind.

Für dieses Ziel wurde im ersten Kapitel eine Grundlegende Einführung in die Problematik gegeben indem eine kurze Zusammenfassung der Entwicklung von Smart Speakern dargestellt und davon ausgehend die Zielsetung abgeleitet wurde. Die Auswahl der Methode verlief nach zwei Gesichtspunkten. Zum einen wurde die Wahl eines passenden Modells dargestellt und wie relevante Daten gefunden werden sollten. Als Ergebnis davon wurde das TAM nach Pavlou (2003) gewählt und mit Hilfe einer Literaturanalyse nach Fettke (2006) Dokumente identifiziert.

Das zweite Kapitel befasste sich mit der grundlegenden Definition, um was es sich bei Smart Speakern handelt und eine Einordnung dieser in aktuelle Entwicklungen des E-Commerces sowie die Darstellung entsprechender Technologien. Dafür wurden Begriffe wie Conversational Commerce, Virtual Assistant und Neural Networks definiert und im Kontext von Smart Speakern betrachtet.

Mit Hilfe einer State of the Art Betrachtung wurde im dritten Kapitel der aktuelle Stand des Marktes der Smart Speaker und ihre Anwendungen dargestellt sowie Anwendungsgebiete und erste Implikationen für ihre Anwendung im E-Commerce aufgestellt. Zu diesem Zweck wurden die vier größten Unternehmen, welche eine Smart Speaker Lösung anbieten, untersucht und in Kurzform dargestellt. Des Weiteren wurden die Plattformen, über welche weitere Anwendungen für die Geräte beziehbar sind, betrachtet und in einer Positionierung der vier Konkurrenten diese zusammenfassend gegenübergestellt.

Das vierte Kapitel legte die Grundlage für die Akzeptanzuntersuchung der Smart Speaker, indem aufbauend auf dem definierten Akzeptanzmodell und mit Hilfe der Ergebnisse der Literaturrecherche nach Fettke (2006) mögliche Faktoren der Geräte aufgestellt wurden. Der Transaktionsprozess über Smart Speaker wurde daraufhin verdeutlicht, um passende Hypothesen über diesen und die zuvor besprochenen Faktoren aufstellen zu können.

Der Aufbau der weiteren Akzeptanzanalyse und die entstandenen Ergebnisse waren Hauptaugenmerk des fünften Kapitels. Zu diesem Zweck wurden zuerst latente Variablen aus den Hypothesen abgeleitet und Indikatoren dieser aufgestellt. Deren Verwendung in einem Fragenkatalog und wie dieser im Aufbau des Experiments genutzt wurde, ist im Zusammenhang mit dem für Alexa-Smart Speaker entwickelten Skill dargestellt wurden. Es folgte eine Betrachtung empirischer Ergebnisse einer Fokusgruppe. Diese waren die Überprüfung der internen Kosinstenz der Skala sowie die Elimination von nicht zielführenden Indikatoren mit Hilfe des Cronbachs α. Die Daten des Hauptexperiments wurden daraufhin in Demografie, Smart Speaker Nutzung und Korrelation aufgeteilt. Die Diskussion der Hypothesen wurde mit der Hilfe der Korrelationsmatrix und den Ergebnissen der Strukturanalyse, welche mit IBM SPSS AMOS 25 durchgeführt wurde, abgehandelt. Als Ergebnis konnten sieben von neun Hypothesen belegt werden. Eine Hypothese konnte nicht untersucht werden und eine weitere stellte sich als umgekehrt zur aufgestellten Hypothese heraus. Anhand der wichtigsten Faktoren, welche den Zusammenhang zwischen den latenten Variablen und zur Erklärung dieser herangezogen werden konnten, wurde die Diskussion auf die relevanten Faktoren der Smart Speaker im E-Commerce ausgeweitet, um so die Zielsetzung der Arbeit zu beantworten. Dabei konnten mehrere Faktoren für dieses Themengebiet als relevant identifiziert werden.

6.2. Kritische Würdigung

In diesem Abschnitt soll die in dieser Arbeit verwendete Vorgehensweise kritisch betrachtet werden. Beginnen möchte ich dabei mit der in der Methodik dargestellten Modellauswahl. Es wurde dargestellt, welche Probleme mit TAMs bestehen und wie verschiedene Autoren diese versuchten zu bewältigen. Das Modell von Pavlou (2003) erwies sich dabei als am besten für die Zielstellung geeignet. Zu beachten ist jedoch, dass durch das Modell und die dargestellte angepasste Version nur die Verbraucher- bzw. Benutzersicht betrachtet wurde. Mit Hilfe des Prozessakzeptanzmodells nach Müllerleile et al. (2015) wäre eine Betrachtung verschiedener Sichtweisen möglich gewesen, aber wie bereits in Kapitel 1.2.1 dargestellt wurde sich dagegen entschieden, da die Identifikationen von Ex-

perten problembehaftet war. Dennoch muss die Kritik stehen bleiben, dass die Unternehmensseite in dieser Arbeit und mit dieser Vorgehensweise nicht weiter betrachtet werden konnte.

Zur weiteren Informationsgewinnung und Literaturrecherche wurden zum einen ein State of the Art verwendet und zum anderen eine Analyse nach Fettke (2006). In einem State of the Art wird versucht ein Überblick über ein Thema darzustellen, wobei es dabei nicht um die Berücksichtigung aller Arbeiten eines Themengebiets geht, sondern eine kritische Wiedergabe von Quellen. Die Berufung auf die vier größten Anbieter und ihrer Entwicklungsinformationen sowie Blogs sollten dieser Vorgehensweise gerecht werden. In einer zusammenfassenden Übersicht wurde Stellung bezogen, wie eine Einschätzung zum jeweiligen Anbieter ausfällt.

Der Aufbau der Literaturanalyse wurde dargestellt und eine gezielte Aktualität der Ergebnisse erreicht, indem keine Quellen vor 2008 verwendet wurde, jedoch konnte keine einheitliche Qualitätsschwelle verwendet werden, da viele verschiedene Arten von Literatur zum Einsatz kamen. Die Auswahl der Literatur konnte entsprechend nicht weiter eingeschränkt werden, wie z. B. mit Hilfe des VHB-Jourqual3 und einer gewählten Qualitätsschwelle. Inwiefern dies die Qualität der identifizierten Faktoren von Smart Speakern beeinträchtigt hat, ist an dieser Stelle nicht ermittelbar.

Weiterer Bewertung unterzogen wird die Vorgehensweise bei der Experimentplanung und -durchführung. Eine entsprechende Festlegung der Stichprobengröße, wie in dieser Arbeit durchgeführt, birgt Probleme bei der Identifikation von Zusammenhängen. Wie bereits dargestellt, durften entsprechend nur $r \geq 0,5$ oder $r \leq -0,5$ als signifikant gewertet werden. Trotz dieser Einschränkung konnten eine Vielzahl an signifikanter Relationen ermittelt werden.

Das Fehlen von Probanden, welche Transaktionen mit ihren Smart Speakern durchgeführt haben, war ein unerwartetes Ergebnis. Das Ziel der Fokussierung auf Personengruppen zwischen 20 und 40 Jahren, welche in der IT tätig sind (Mitarbeiter der dotSource GmbH), sollte dies verhindern. Die Nutzung von Smart Speakern für Transaktionen im scheint in Deutschland noch nicht verbreitet zu sein. Dies hatte zur Folge, dass eine Hypothese verworfen werden musste, jedoch traten dadurch auch an anderer Stelle Probleme auf. Die Modellierung des TAM mit Hilfe von IBM SPSS AMOS bedarf deshalb einer weiteren Betrachtung. Das Modell konnte entsprechend nicht vollständig dargestellt werden, da die Relation der Hypothese 1 komplett fehlt. Dies hatte zur Folge, dass Gütekriterien der Modellierung, wie z. B. X^2 entsprechend schlechte Werte aufweisten. Nur durch eine

weitere Untersuchung, mit Probanden welche bereits Transaktionen mit Smart Speakern durchgeführt haben, könnte dieses Problem behoben werden. Jedoch ist nicht absehbar wann dies in Deutschland möglich wird.

Eine weiterer Punkt der Überlegung, wenn mit Ergebnissen von Korrelationsbetrachtungen gearbeitet wird, ist natürlich der Zusammenhang zwischen Korrelation und Kausalität. Dies bedeutet, dass eine hohe Korrelation zwischen zwei Faktoren, wie z. B. bei der wahrgenommenen Nützlichkeit und der Transaktionsabsicht dargestellt, nicht automatisch auch ein Zusammenhang zwischen diesen bestehen muss. Viel mehr gibt die Korrelation erste Anhaltspunkte bzw. einen Hinweis, wo Relationen bestehen könnten. Eine Übersicht solcher ersten Ansätze ist entsprechend das Ziel dieser Arbeit.

6.3. Ausblick

Ausgehend von den zuvor genannten Punkten wurde die Zielstellung der Arbeit erreicht und die am anfang gestellte Frage beantwortet. Es gilt in der Zukunft die identifizierten Faktoren weitergehende Betrachtungen zu unterziehen. Eine Durchführung weiterer Experimente bzw. einer reinen Befragung nur mit Personen, welche bereits Transaktionen durchgeführt haben, wäre ein möglicher Ansatz. Auf diese Weise könnten die noch offenen Fragen dieser Arbeit beantwortet werden. Entsprechend gilt es in Zukunft noch weitere Betrachtungen über Smart Speaker durchzuführen. Gerade ein solcher Markt, welcher seit mehreren Jahren stetig im wachsen ist, aber noch in vielerlei Hinsicht in den Anfängen steht (z. B. Entwicklung und Verbreitung der Stores), wird sich noch vielen Änderungen durchziehen.

Eine besondere Betrachtung könnte auch die Entwicklung in Deutschland darstellen, da die Intention mit diesen Geräten E-Commerce bzw. Einkäufe zu tätigen aktuell noch gering ist. Verglichen zu anderen europäischen Ländern und den USA ist dies ein besonderes Verhalten, da bereits viele Menschen diese Möglichkeiten getestet haben. Ob es sich dabei um einfache kulturelle Unterschiede oder eine langsamere Innovationsakzeptanz handelt, könnte Thema einer weiteren Untersuchung sein.

Auch eine Betrachtung der Marktentwicklung könnte in Zukunft interessant werden. Amazons bisheriger Marktanteil scheint durch Google zu schrumpfen und neue Teilnehmer drängen auf den Markt. Neue Ideen, wie z. B. die Einführung von Werbung bei Amazon

werden bereits diskutiert und möglicherweise könnten die westlichen Anbieter gewillt sein auf dem chinesischen Markt zu agieren, welcher aktuell in der Hand von drei lokalen Smart Speaker Herstellern ist.

Abschließend ist festzuhalten, dass Smart Speaker als Werbung für die VAs der großen Hersteller ein wirksames Werkzeug waren. Siri und Cortana sind durch die Integration im iPhone bzw. in Windows 10 Rechnern vielen ein Begriff. Hingegen haben Alexa und der Google Assistant durch die Integration in ihren entsprechenden Geräte einen Schub im Bekanntheitsgrad erhalten.

Literaturverzeichnis

Ajzen I, Fishbein M (1980) Understanding Attitudes and Predicting Social Behavior. Prentice-Hall, Englewood Cliffs, NJ.

Ajzen I (1991) The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes 50(2): 179–211.

Allen JF, Byron DK, Dzikovska M, Ferguson G, Galescu L, Stent A (2001) Toward conversational human-computer interaction. AI Magazine 22(4): 27–37.

Amazon.com Inc. (2018a) Dokumentation für Entwickler. https://developer.amazon.com/de/documentation. Abruf am 2018-4-22.

Amazon.com Inc. (2018b) Product Images - Amazon Echo. http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-imageproduct41. Abruf am 2018-5-2.

Apple Inc. (2018a) Home Zubehör. Die Liste wird immer smarter. https://www.apple.com/de/ios/home/accessories/. Abruf am 2018-9-13.

Apple Inc. (2018b) HomePod. https://www.apple.com/homepod/specs/. Abruf am 2018-5-3.

Bauer RA (1960) Risk Taking and Information Handling in Consumer Behavior. Division of Research, Graduate School of Business Administration, Harvard University.

Beaufays F (2015) The neural networks behind Google Voice transcription. https://research.googleblog.com/2015/08/the-neural-networks-behind-google-voice.html. Abruf am 2018-4-26.

Bengio Y (2009), *Learning Deep Architectures for AI*, Université de Montréal.

Bengio Y, Louradou J, Collobert R, Weston J (2009), *Curriculum Learning*, Université de Montréal.

Bockmeyer H, Vogt V (2018), *Alexa, Siri & Google Assistant - was ist erlaubt? Sprachassistenten und das Recht*, Institut für Informations-, Telekommunikations- und Medienrecht (ITM) Westfälische Wilhelms-Universität.

Bokhari RH (2006) The relationship between system usage and user satisfaction: a meta-analysis. The Journal of Enterprise Information Management 18(2): 211–234.

Bortz J, Döring N (2006) Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Springer Meidzin Verlag, Heidelberg.

Bourlard HA, Morgan N (1993) Connectionist Speech Recognition: A Hybrid Approach. Kluwer Academic Publishers, Norwell, MA.

Brown SA, Massey A, Montoya-Weiss MM, Burkman JR (2002) Do I Really Have to? User Acceptance of Mandated Technology. European Journal of Information Systems 11(4): 283–295.

Buvat J, Taylor M, Jacobs K, Khadikar A, Sengupta A (2018), *Conversational Commerce*, Capgemini Digital Transformation Institute.

Caire P, Moawad A, Efthymiou V, Bikakis A, Le Traon Y (2016) Privacy Challenges in Ambient Intelligence Systems. Journal of Ambient Intelligence and Smart Environments 8(6): 619–644.

Chircu AM, Davis GB, Kauffman RJ (2000), „Trust, Expertise, and E-Commerce Intermediary Adoption“, In: *AMCIS 2000 Proceedings*.

Chowdhury G (2003) Natural Language Processing. Annual Review of Information Science and Technology 37(1): 51–89.

Cohen J, Cohen P (1975) Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Wiley, New York.

Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, New York.

Dale R (2016) The return of the chatbots. Natural Language Engineering 22(5): 811–817.

Davis FD, Bagozzi RP, Warshaw PR (1989) User acceptance of computer technology: A comparison of two theoretical models. Management Science 35(8): 982–1003.

Davis FD (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technolog. MIS Quarterly 13(3): 319–340.

Deutsche Telekom AG (2018) Hallo Magenta! Mit dem Smart Speaker hört das Zuhause aufs Wort. https://www.telekom.com/de/medien/medieninformationen/detail/mit-dem-smart-speaker-hoert-das-zuhause-aufs-wort-507994. Abruf am 2018-5-29.

Duizith JL, Gouveia F, Tagliassuchi G, Brahm DR, da Silva LK (2004), *A Virtual Assistant for Websites*, Lutheran University of Brazil.

Edlund J, Gustafson J, Heldner M, Hjalmarsson A (2008) Towards human-like spoken dialogue systems. Speech Communication 50(8-9): 630 ff.

Facebook (2016) Introducing Bots on Messenger. https://developers.facebook.com/videos/f8-2016/introducing-bots-on-messenger/. Abruf am 2018-5-14.

Fettke P (2006) State-of-the-Art des State-of-the-Art. WIRTSCHAFTSINFORMATIK 48(4): 257–266.

Friedewald M, Da Costa O, Punie Y, Alahuhta P, Sirkka H (2005) Perspectives of ambient intelligence in the home environment. Telematics and Informatics 22(1): 221–238.

Fukuyuma F (1995) Trust:The Social Virtues and the Creation of Prosperity. Free Press, New York.

Gabler Wirtschaftslexikon (2018a) E-Commerce. https://wirtschaftslexikon.gabler.de/definition/e-commerce-34215/version-257721. Abruf am 2018-4-25.

Gabler Wirtschaftslexikon (2018b) M-Commerce. https://wirtschaftslexikon.gabler.de/definition/mobile-commerce-37243/version-260684. Abruf am 2018-4-25.

Gaizauskas R, Wilks Y (1998) Information extraction: beyond document retrieval. Journal of Documentation 54(1): 70–105.

Gefen D (2000) E-Commerce: The role of familiarity and trust. Omega: The International Journal of Management Science 28(6): 725–737.

Gefen D, Straub D (2003) Managing User Trust in B2C e-Services. eService Journal 2(2): 7–24.

Glaser BG, Straus A (1967) The Discovery fo Grounded Theory: Strategies for Qualitative Research. Aldine, Chicago.

Google LLC (2018a) Google Home. https://store.google.com/product/google_home. Abruf am 2018-5-3.

Google LLC (2018b) Was kann Google Assistant für dich tun? https://assistant.google.com/explore?hl=de_de. Abruf am 2018-5-22.

Grami A, Schell B (2004), *Future Trends in Mobile Commerce: Service Offerings, Technological Advances and Security Challenges*, University of Ontario.

Haack W, Severance M, Wallace M, Wohlwend J (2017), *Security Analysis of the Amazon Echo*, Allen Institute for Artificial Intelligence.

Hinton G, Deng L, Yu D, Dahl GE, Mohamed Ar, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B (2012) Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Processing Magazine 29(6): 82–97.

Hoffmann L, Krämer N, Lam-chi A, Kopp S (2009), „Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent?“, In: *Intelligent Virtual Agents, 9th International Conference*.

Hosmer LT (1995) Trust: The connection link between organizsational theory and philosophical ethics. Academy of Management Review 20(3): 213–237.

Huang X (2017) Microsoft researchers achive new conversational speech recognition milestone. https://www.microsoft.com/en-us/research/blog/microsoft-researchers-achieve-new-conversational-speech-recognition-milestone/. Abruf am 2018-6-2.

Hunt AJ, Black AW (1996), *Unit Selection in a concatenative Speech Synthesis System using a large Speech Database*, ATR Interpreting Telecommunications Research Labs.

Ickler H, Schülke S, Wilfling S, Baumöl U (2009), *New Challenges in E-Commerce: How Social Commerce Influences the Customer Process*, FernUniversität in Hagen.

IDC (2018) New IDC Smart Home Device Tracker Forecasts Solid Growth for Connected Devices in Key Smart Home Categories. https://www.idc.com/getdoc.jsp?containerId=prUS43701518. Abruf am 2018-6-27.

IFTTT (2018) "Hey Cortana," welcome to IFTTT. https://ifttt.com/blog/2018/02/Hey-Cortana-welcome-to-IFTTT. Abruf am 2018-5-22.

intelliAd (2018), *IntelliAd E-Commerce Branchenindex Q3/2017*.

Jarvenpaa SL, Tractinsky N, Vitale M (1999) Consumer trust in an Internet store. Information Technology and Management 1(12): 45–71.

Jensen J (2005), „Interactive Television: New Genres, New Format, New Content“, In: *Proceedings of the Second Australasian Conference on Interactive Entertainment*.

Kinsella B, Mutchler A (2018), *Smart Speaker Consumer Adoption Report*, voicebot.ai.

Koch B, Schmidt-Hern K (2018), *Alexa, wo bitte geht es hier zum BGH?*

Kong L, Alberti C, Andor D, Gobatyy I, Weiss D (2017), *DRAGNN: A Transition-based Framework for Dynamically Connected Neural Networks*, Carnegie Mellon University.

Koo H, Kim S, Nam C (2017), „Speaker Wars begins: Which applications will be the killer content for smart speaker?“, In: *14th International Telecommunications Society (ITS) Asia-Pacific Regional Conference*.

Lee CS (2001) An analytical framework for evaluating e-commerce business models and strategies. Internet Research: Electronic Networking Applications and Policy 11(4): 349–359.

Levin MR, Lowitz JN (2017), *Home Automation Device Market Grows Briskly, to 27 Million*, Consumer Intelligence Research Partners.

Martin E (2017) How Echo, Google Home, and Other Voice Assistants Can Change the Game for Content Creators. EContent 40(2): 4–8.

Microsoft Corporation (2018a) Harman Kardon Invoke with Cortana by Microsoft (Graphite). https://www.microsoft.com/en-us/store/d/harman-kardon-invoke-with-cortana-by-microsoft/8rl7xlnwn95v/9H91. Abruf am 2018-5-2.

Microsoft Corporation (2018b) Microsoft Speech Platform. https://msdn.microsoft.com/en-us/library/jj127902.aspx. Abruf am 2018-5-3.

Mikic F, Berguillo J, Llamas M, Rodriguez D, Rodriguez E (2009), *CHARLIE: An AIML-based Chatterbot which Works as an Interface among INES and Humans*, University of Vigo.

Mori M (1970) The Uncanny Valley. Energy 7(4): 33–35.

Muellerleile T, Ritter S, Englisch L, Nissen V, Joenssen DW (2015), „The Influence of Process Acceptance on BPM: An Empirical Investigation“, In: *IEEE 17th Conference on Business Informatics*.

Mukherjee S, Vengattil M (2017) Alexa allies with Cortana to take on Google Assistant, Siri. https://www.reuters.com/article/us-amazon-com-microsoft-partnership/alexa-allies-with-cortana-to-take-on-google-assistant-siri-idUSKCN1BA1RX. Abruf am 2018-7-4.

Norman DA (1998) The design of everyday things. MIT, London.

Panasonic Corporation (2018) Panasonic Conducts Demonstration Experiment of Autonomous Signage Robot, „HOSPI(R)“ at Narita Airport. http://news.panasonic.com/global/topics/2018/54270.html. Abruf am 2018-5-14.

Pavlou PA (2003) Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. International Journal of Electronic Commerce 7(3): 69–103.

Petrov S (2016) Announcing SyntaxNet: The World's Most Accurate Parser Goes Open Source. https://ai.googleblog.com/2016/05/announcing-syntaxnet-worlds-most.html. Abruf am 2018-5-17.

Pfeffer J (1982) Organizations und Organization Theory. Pitman, Boston, MA.

Porzel R (2006), „How Computers (Should) Talk to Humans“, In: *How People Talk to Computers, Robots, and Other Artifical Communication Partners*.

Radner R, Rothschild M (1975) On the Allocation of Effort. Journal of Economic Theory 10(3): 358–376.

Radziwill NM, Benton MC (2017) Evaluating Quality of Chatbots and Intelligent Conversational Agents. CoRR abs/1704.04579(1).

Richardson F, Reynolds D, Dehak N (2015) Deep Neural Network Approaches to Speaker and Language Recognition. IEEE Signal Processing Letters 22(10): 1671–1675.

Ring PS, Van de Ven AH (1994) Developing processes of cooperative inter-organizational relationships. Academy of Management Review 19(1): 90–118.

Routley N (2018) Amazon vs. Google: The Battle for Smart Speaker Market Share. http://www.visualcapitalist.com/smart-speaker-market-share/. Abruf am 2018-5-2.

Samsung (2018) Bixby - A smarter way to use your phone. https://www.samsung.com/us/explore/bixby/. Abruf am 2018-5-29.

Schein EH (1980) Organizational Psychology. Prentice Hall, Englewood Cliffs, NJ.

Schmidhuber J (2015) Deep learning in neural networks: An overview. Neural Networks 61(1): 85–117.

Schulze P (2017) Google Home Actions: Neue Funktionen mit Skills nachrüsten. https://www.turn-on.de/tech/ratgeber/google-home-actions-neue-funktionen-mit-skills-nachruesten-324703. Abruf am 2018-9-20.

Segan S (2018) Google Assistant now Has 1,830 Actions: Here They Are. https://www.pcmag.com/article/353240/200-things-to-ask-your-google-home. Abruf am 2018-6-29.

Siri Team (2018a) Deep Learning for Siri's Voice: On-device Deep Mixture Density Networks for Hybrid Unit Selection Synthesis. Apple Machine Learning 1(4).

Siri Team (2018b) Inverse Text Normalization as a Labeling Problem. Apple Machine Learning 1(3).

Siri Team (2018c) Personalized Hey Siri. Apple Machine Learning 1(9).

Specht L, Herold S (2018), *Roboter als Vertragspartner?*, MMR.

Statistisches Bundesamt (2016) 81 % der Internetnutzer gehen per Handy oder Smartphone ins Internet. https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2016/12/PD16_430_63931.html. Abruf am 2018-5-14.

Strom N (2017), „Amazon Alexa Technologies", In: *AWS Stockholm Summit*.

Thu YK, Pa WP, Sagisaka Y, Iwahashi N (2016), „Comparison of Grapheme-to-Phoneme Conversion Methods on a Myanmar Pronunciation Dictionary", In: *Proceedings of the 6th Workshop an South and Southeast Asian Natural Language Processing*.

Tiongson J (2015) Mobile app marketing insights: How consumers really find and use your apps. https://www.thinkwithgoogle.com/consumer-insights/mobile-app-marketing-insights/. Abruf am 2018-7-5.

TrendForce (2018) Apple Joins the Competition of Smart Speaker, but Amazon Remains the Market Leader, Says TrendForce. https://press.trendforce.com/node/view/3063.html. Abruf am 2018-5-28.

van Bruggen GH, Kersi AD, Jap SD, Reinartz WJ, Pallas F (2010) Managing Marketing Channel Multiplicity. Journal of Service Research 13(3): 331–340.

van den Oord A, Dieleman S, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K (2016), *WaveNet: A generative Model for raw Audio*.

van Eeuwen M (2017), *Mobile conversational commerce: messenger chatbots as the next interface between businesses and consumers*, University of Twente.

Vroom VH (1964) Work and Motivation. Wiley, Oxford, England.

W3C (2004) Voice Extensible Markup Language (VoiceXML) Version 2.0. https://www.w3.org/TR/voicexml20/. Abruf am 2018-5-16.

W3C (2010) Speech Synthesis Markup Language (SSML) Version 1.1. https://www.w3.org/TR/speech-synthesis11/. Abruf am 2018-5-16.

Weiss D, Petrov S (2017) An Upgrade to SyntaxNet, New Models and a Parsing Competition. https://ai.googleblog.com/2017/03/an-upgrade-to-syntaxnet-new-models-and.html. Abruf am 2018-5-17.

Wik P, Hjalmarsson A (2009) Embodied conversational agents in Computer Assisted Language Learning. Speech Communication 51(10): 1024–47.

Xiong W, Wu L, Alleva F, Droppo J, Huang X, Stolcke A (2017a), *The Microsoft 2017 Conversational Speech Recognition System*, Microsoft AI und Research.

Xiong W, Droppo J, Huang X, Seide F, Seltzer M, Stolcke A, Yu D, Zweig G (2017b) Toward Human Parity in Conversational Speech Recognition. IEEE/ACM Transactions on Audio, Speech, and LAnguage Processing 25(12): 2410–2423.

Zen H (2009) Statistical Parametric Speech Synthesis. Speech Communication 51(11): 1039–1064.

Zuberer S (2017), *Digitale Assistenten Bevölkerungsbefragung*, PwC Communications.

Zwass V (1998) Structure and Macro-level Impacts of Electronic Commerce: From Technological Infrastructure to Electronic Marketplaces. Irwin/McGraw-Hill, Boston, Mass.