

Hafner, Nils; Henn, Harald

Article

Voice in der D2C-Journey – Sprachautomation als Erlebnis

Marketing Review St.Gallen

Provided in Cooperation with:

Universität St. Gallen, Institut für Marketing und Customer Insight

Suggested Citation: Hafner, Nils; Henn, Harald (2021) : Voice in der D2C-Journey – Sprachautomation als Erlebnis, Marketing Review St.Gallen, ISSN 1865-7516, Thexis Verlag, St.Gallen, Vol. 38, Iss. 6, pp. 46-53

This Version is available at:

<https://hdl.handle.net/10419/276167>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

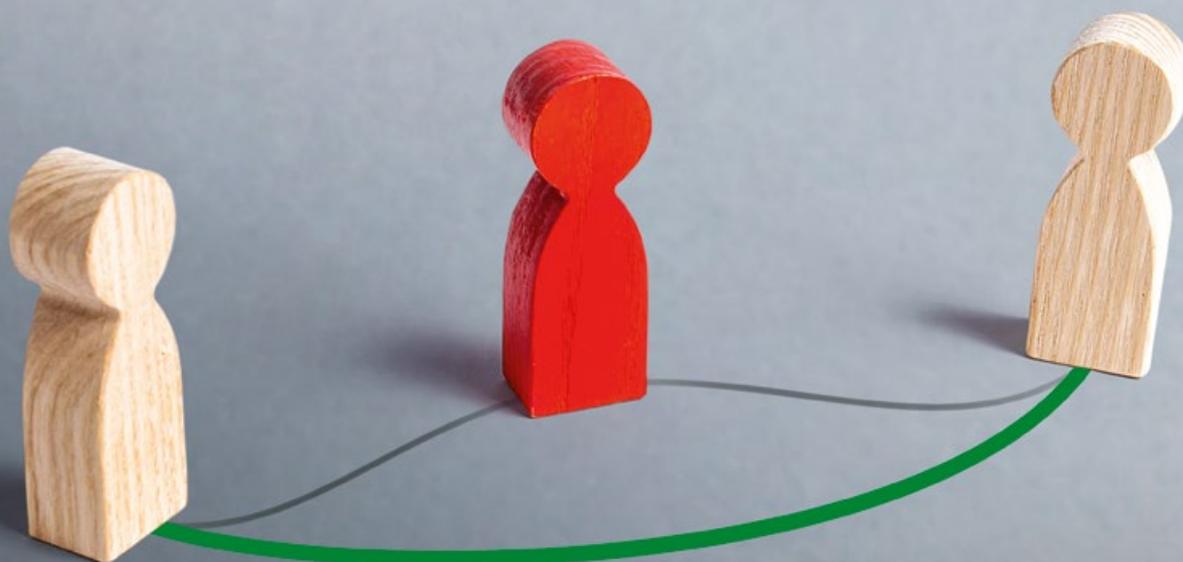
Marketing Review St. Gallen



6 | 2021 **SCHWERPUNKT** Mammut's Direct-to-Consumer (D2C)-Strategie • Drivers and Prerequisites for D2C Strategies • Erfolg durch D2C-Start-ups • Sustainable Corporate Incubation at a CPG Company • D2C-Lernstrategie einer FMCG-Marke • Voice-Systeme im D2C-Kontext **SPEKTRUM** Kundenansprache im persönlichen Verkauf • Objectives and Key Results-Methode und intrinsische Mitarbeitermotivation

www.marketing-review.ch

Direct-to-Consumer- Strategien



Voice in der D2C-Journey – Sprachautomation als Erlebnis

Dieser Artikel zeigt auf, welche Einsatzmöglichkeiten und Potenziale für Voice-Systeme in einem D2C-Kontext bestehen. Er skizziert, wie ein solcher Einsatz die Customer Journey verändern kann und welche Fähigkeiten im Hinblick auf die Dimensionen „Prozesse“ und „Technologien“ im Unternehmen notwendig sind. Schlussendlich werden Kosten und Nutzenpotenziale beschrieben.

Prof. Dr. Nils Hafner, Harald Henn

Voice-Systeme als Medien der Kundenkommunikation sind zurzeit in aller Munde. Das hat seine Berechtigung, da Sprache die natürlichste und gleichzeitig intensivste Form menschlicher Kommunikation ist (Lv, Zhang & Li, 2008). Sprache ist nicht nur natürlich, sondern auch schnell. Wir sprechen etwa 3-mal schneller als wir schreiben. In der unternehmerischen Kundenkommunikation – in Marketing, Vertrieb oder im Service – dominierten jedoch sehr lange grafische oder textlich orientierte Benutzerschnittstellen – zunächst per Maus oder Tastatureingabe, später per Touchscreen und Finger – da Sprache zu komplex war, um sie von Systemen verstehen und übersetzen zu lassen.

Die Konsequenz: Dialoge mit Kunden fanden entweder in schriftlicher Form (Fax, E-Mail, Webformular oder Chat) oder per Telefongespräch statt. Dieses Vorgehen ist in der Regel ressourcenintensiv, erfordert es doch einen zwischenmenschlichen Dialog. Gerade in Bereichen, in denen Unternehmen traditionell ihre Produkte über den Handel vertreiben, musste man sich also fragen, ob eine Kommunikation direkt vom Unternehmen an den Endkunden (Direct-to-Consumer, D2C) überhaupt wirtschaftlich zu rechtfertigen ist.

Dies wird nun auch zunehmend technisch möglich. Die ursprünglichen Defizite der Spracherkennung und -verarbeitung, wie

- die mangelhafte und langsame Spracherkennung,
- die Beschränkung auf vorgegebene und damit beim Kunden antrainierte Schlüsselwörter und dadurch bedingt rigide, vom System strikt vorgegebene Dialoge und Pfade sowie
- eine mangelhafte Erkennung des Kontextes und der Absicht des Kunden und
- die fehlende Anbindung solcher Systeme an CRM- und ERP-Systeme und

damit verbunden ein häufiger Medienbruch bei der fallabschliessenden Beantwortung von Bestellungen oder Anfragen sowie

- die fehlende Personalisierung sind mittlerweile beseitigt.

Einer der Gründe dafür liegt auch in der Standardisierung von Voice-Formaten. Initiativen wie das [openvoicenetwork](https://openvoicenetwork.org/) (https://openvoicenetwork.org/), zu deren Gründungsmitgliedern u.a. Microsoft oder die Deutsche Telekom gehören, sorgen dafür, dass die Akzeptanz von Voice-Systemen durch die Definition von Standards und verbindlicher Normen bei den Konsumenten weiter steigt. Dadurch entstehen auch Möglichkeiten der Automatisierung von barrierefreier Sprach-Kommunikation beispielsweise über die digitalen Assistenten Siri, Alexa oder Google Assistant. Unternehmen haben das längst erkannt und die Investitionen in derartige sprachbasierte Kommunikationssysteme gehen aufgrund des enormen Potenzials in die Milliardenhöhe (Kinsella, 2019). Um die Höhe dieser Investitionen nachvollziehen zu können, lohnt es sich, sich zunächst mit den Potenzialen und Einsatzgebieten spezifisch im B2C auseinander zu setzen.

Folgende Einsatzbereiche und Entwicklung bei der Nutzung von Voice-Systemen zeichnen sich ab

Der Einsatz von Voice-Systemen muss grundsätzlich auf der Grundlage einer strategischen Beurteilung des Aufwandes, Kundenakzeptanz, Nutzens erfolgen. Hier ist die Value Irritant Matrix (Price & Jaffe, 2006) als geeignetes Instrument zur Einordnung heranzuziehen. Diese wurde ursprünglich für den Internethändler Amazon in einer Only-Online-Channel-Umgebung entwickelt. Erst durch die zunehmende Digitalisie-

Prof. Dr. Nils Hafner

Institut für Finanzdienstleistungen Zug (IFZ), Hochschule Luzern Wirtschaft (HSLU)

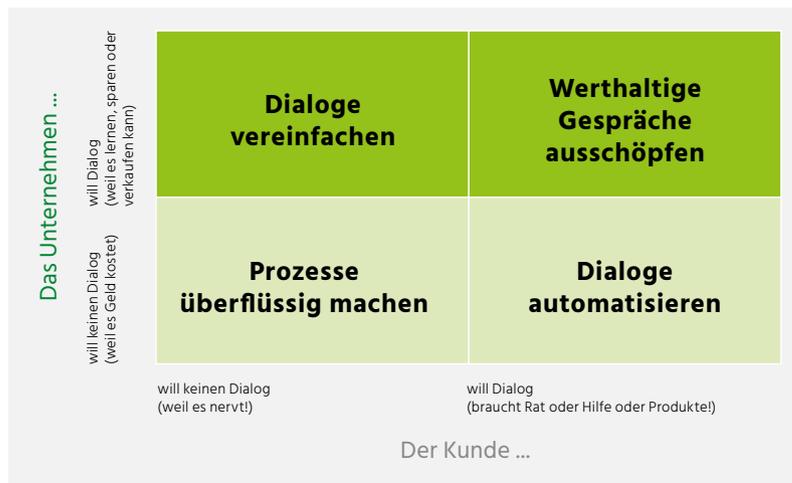
nils.hafner@hslu.ch

Harald Henn

CEO Marketing Resultant GmbH, Mainz, Deutschland

henn@marketing-resultant.de

Abb. 1: Value Irritant Matrix



Quelle: In Anlehnung an Price und Jaffe (2006).

Der letzten Jahre kann die Matrix auch auf komplexe Mehrkanalumgebungen sinnvoll angewendet werden (Hafner, 2019). Die Matrix wird in der folgenden Abbildung 1 dargestellt:

In der Praxis ist die Automatisierung von einfachen, homogenen und zugleich hochvolumigen Kundenanfragen per Sprache als erster Schritt im Kundenservice zu verzeichnen. Statusabfragen zu Lieferungen, Adress- und Kontoänderungen oder die Meldung von z.B. Verfügbarkeit von Produkten beim Hersteller. Beide Seiten – Unternehmen und Kunden – sind an einem persönlichen Gespräch für diese Geschäftsvorfälle nicht unbedingt interessiert. Die Mehrzahl der Voice-Anwendungen beginnt in diesem Quadranten der Value Irritant Matrix auch deshalb, weil Unternehmen sich so langsam an diese neue Technologie herantasten können, ohne entweder die Kundenbeziehung oder Umsatzpotenziale zu gefährden.

Gerade in den folgenden Anwendungsbereichen, bei denen die Sprache

spezifische Vorteile gegenüber textlichen oder visuellen Dialogen bietet, fokussiert sich momentan die Entwicklung von Voice-Systemen:

1. Multitasking-Situationen

Multitasking-Situationen sind Situationen, bei denen der Kunde die Hände nicht frei hat, weil er z.B. am Steuer des Autos oder auf dem Motorrad sitzt. Analog gilt das auch für Kunden, die in der Küche stehen, ein Gericht zubereiten wollen und sich per Sprachdialog über Zutaten und Arbeitsschritte sukzessive instruieren lassen. Hier bieten sich spezifische D2C-Interaktionsmöglichkeiten für Fahrzeughersteller, Nahrungsmittelproduzenten und Hersteller weisser Ware. Aber auch andere Use-Cases wie im Gesundheitssektor zeigen neue Potenziale für Voice-Systeme auf. Patienten im Krankenhaus können per Sprachbefehle aus dem Krankbett agieren; andere Dialogmöglichkeiten sind für manche Patienten zeitweise nicht möglich (Marxen, 2019).

Zusammenfassung

Zusammenfassend bleibt festzustellen, dass der automatisierte Einsatz von Voice-Systemen im Kundendialog, spezifisch aber in Multitasking-Situationen, unter der Anforderung grosser Schnelligkeit und unter Entertainment-Gesichtspunkten Potenziale gerade im D2C-Umfeld bietet, die Unternehmen für einen kostengünstigen automatisierten Dialog mit dem Endkunden nutzen sollten. Die Beispiele aus der Automobil- und der Sportbekleidungsbranche haben das aufgezeigt.

2. Zeitsparende Anwendungen

Die visuell oder textlich strukturierte Eingabe kann in einigen Anwendungsfällen sehr zeitaufwendig sein. Dies gilt z.B. für die gezielte Suche nach einem Produkt bei einem Anbieter von Sportschuhen und -bekleidung. Die Eingabe per Sprache ist um den Faktor 3 schneller als die strukturierte und geführte Dialogführung per Auswahlmenü oder Drop-down-Listen (Ruan et al. 2017). Wenn man die Suche per Sprache um weitere Anwendungsfälle erweitert, (Bestellung, Statusanfrage, Service) dann ist die Vision für Voice-Commerce – also des komplett per Sprache gesteuerten Dialoges – auch gerade im D2C-Kontext nicht mehr weit.

3. Entertainment-Dialog

Smart-Speaker-Systeme wie Alexa können durch innovative Voice-Systeme mit den Kunden Produkte konfigurieren. Eine im TV angeschauten Werbung für ein neues E-Auto kann im

Smart Speaker für eine individuelle Konfiguration des Wunschautos genutzt werden: „Alexa, was kostete der BMW genau in der Ausstattung wie in der Fernsehwerbung? Welche Extras muss ich dafür bestellen?“

Im Zusammenhang mit den Anwendungsmöglichkeiten sind quasis menschliche Dialoge der Kern einer für den Kunden überragenden und für das Unternehmen bezahlbaren Customer Experience (Lemon & Verhoef, 2016). McLean, Osei-Frimpong und Barhorst haben jüngst erst den positiven Einfluss von Voice-Assistants auf das Consumer-Brand-Engagement untersucht (McLean et al., 2021). Um die Kosten solcher Use-Cases abschätzen zu können, die eine oben beschriebene Veränderung zur Nutzung von Voice-Systemen gerade in einer beispielhaft skizzierten D2C-Umgebung nach sich zieht, ist daher ein grundsätzliches Verständnis der technologischen Komponenten und ein Verständnis des Aufbaus und der Gestaltung von Dialogen per Sprache notwendig.

Daraus resultieren nämlich notwendige Kompetenzen, über die im D2C-Umfeld aktive Unternehmen heute nicht verfügen, und die sie daher (unter Kostenfolge) zwingend aufbauen müssen, um erfolgreich zu sein.

Was ist das Besondere an Voice-Systemen?

Die Anforderungen an Voice-Systeme sind gerade im D2C-Kontext anders als an visuelle oder textliche Benutzeroberflächen. Die größte Herausforderung besteht darin, die Komplexität der menschlichen Sprache nachzubilden und das geeignete Modell für den Dialog mit dem Endkunden zu wählen. Um diese Herausforderung zu verstehen, hier eine kurze Erläuterung, wie die menschliche Sprache von Systemen „verstanden und verarbeitet“ wird und welche technologischen Komponenten dazu notwendig sind. Sprache kann nicht direkt von Systemen verarbeitet werden, sondern durchläuft folgende technische Einzelschritte (Aspect, 2017):

1. Spracherkennung (ASR: Speech Recognition)
2. Sprache in Text umwandeln, transkribieren (STT: Speech to Text)
3. Sprache verstehen (NLU: Natural Language Understanding)
4. Sprache generieren (NLG: Natural Language Generation)
- 5.. Text in Sprache umwandeln (TTS: Text to Speech).

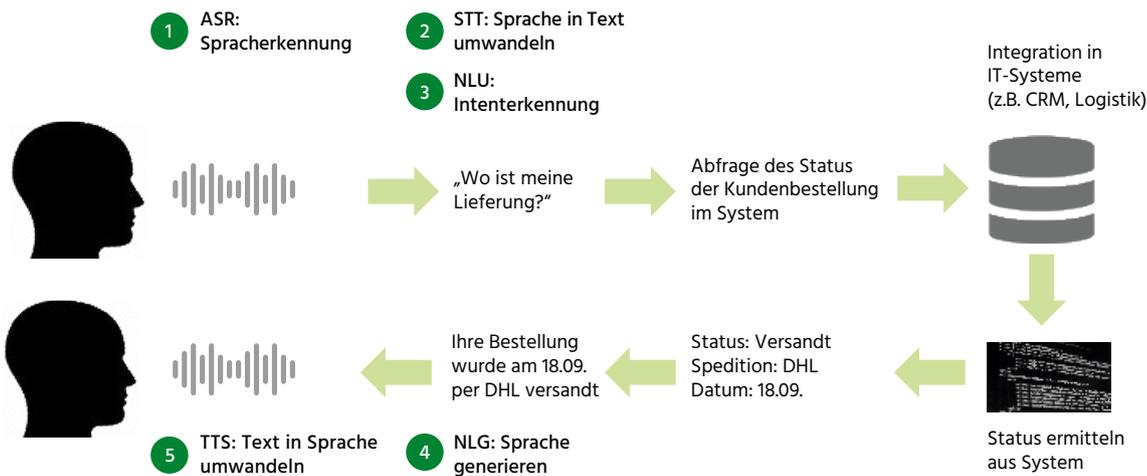
Diese fünf Schritte sind in der nachfolgenden Abbildung 2 des Unternehmens Aspect dargestellt:

In der Folge wollen wir die einzelnen Schritte kurz skizzieren, um das Verständnis für die grundlegende Technologie zu erhöhen:

ASR – Automatic Speech Recognition

ASR-Systeme zeichnen die Sprache auf und zerlegen sie anschließend in Phoneme, die später zu Text verarbeitet bzw. umgewandelt werden. Dabei

Abb. 2: Technische Komponenten einer Voice-Kommunikationslösung für Endkunden



Quelle: Aspect (2017).

ist ein Phonem – nicht Wörter oder Silben – die maßgebliche und grundlegende Masseinheit für die Spracherkennung. Hier sind die Anbieter mittlerweile auf einem technologischen Entwicklungsstand, der zwei wesentliche Anforderungen für die Akzeptanz beim Kunden im Dialog erfüllt:

1. Schnelligkeit, Verarbeitung in Echtzeit

Während die Systeme vor einigen Jahren noch mit einiger Zeitverzögerung Sprache verstehen konnten, geschieht dies heute ohne Zeitverzug; also in Echtzeit, sodass ein flüssiger Dialog zwischen Menschen und Systemen ermöglicht wird.

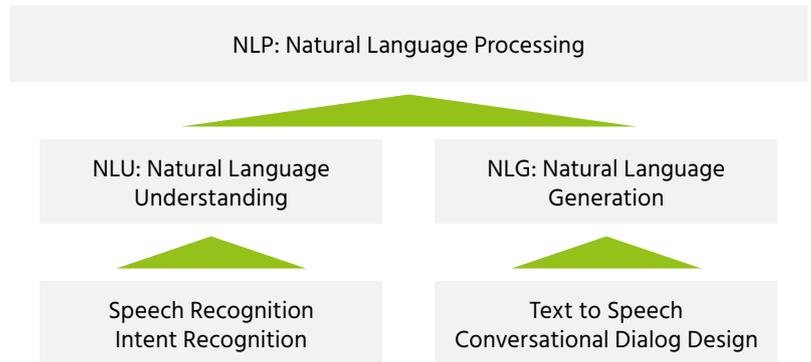
2. Es werden ganze Sätze und Freitext verstanden

Es gibt keine Beschränkungen oder rigide Vorgaben, die den Kunden in Dialogstrukturen oder eine Reihenfolge zwingen. Selbst leichte Dialekte stellen für Systeme kein Hindernis mehr dar. So wird momentan daran gearbeitet, dass sich sowohl Alexa von Amazon als auch Apples Siri sich in Schweizer Mundart bedienen lassen (Kedwes, 2021). Kunden müssen nicht mehr definierte Formulierungen nutzen oder sich an eine zuvor festgelegte Reihenfolge des Systems halten. Wir reden mit dem Voice-System wie mit dem guten Freund im Café.

So werden beispielsweise folgende Dialoge mit dem digitalen Assistenten Siri von Apple ermöglicht, die im Alltag einen hohen Nutzwert aufweisen, schnell durch Sprache einzugeben sind und deren Antwort nicht aufwendig recherchiert werden muss:

„Hey Siri, wie wird das Wetter morgen und muss ich einen Regenschirm mitnehmen?“

Abb. 3: NLP-Komponenten



Quelle: Eigene Darstellung in Anlehnung an Kabel (2020).

„Erinnere mich daran, die Tickets für das Kino zu buchen, wenn ich nach Hause komme.“

NLU, NLG und NLP: Verarbeitung natürlicher Sprache

Nachdem die Sprache vom System erkannt wurde, wird sie in Text umgewandelt, damit sie vom System verarbeitet werden kann. Nach der Umwandlung kommen die Funktionen NLU und NLG zum Tragen. NLU und NLG werden häufig zusammengefasst unter dem Oberbegriff NLP (Natural Language Processing); also dem Verstehen und Verarbeiten menschlicher Sprache:

NLU (Natural Language Understanding)

Hier geht es darum, die Bedeutung (den sogenannten Intent) der Kundenaussage oder einer Frage im Detail zu verstehen (Kannan & Bernoff, 2019). Mit Natural Language Understanding (NLU) können Systeme Bedeutungen von Begriffen lernen und die Intention des Kunden sowie den Kontext, aus dem der Kunde agiert, erkennen.

Das Wissen eines Unternehmens, welches für das Voice-System notwendig ist, kann dabei produkt-, kundenspezifisch oder organisationsspezifisch sein. Dabei ist festzuhalten, dass Voice-Systeme einen erheblichen Initialaufwand benötigen, um die spezifisch benutzten Wörter und Begrifflichkeiten sowie den situativen Kontext der Kunden zu verstehen und einzuordnen. Dies ist notwendig, da Menschen sich im Dialog mit einem System so verhalten, wie sie es auch mit einem menschlichen Gegenüber tun würden. Dies zeigt der Beispieldialog eines Kunden:

„Welche Modelle und Marken von Trekkingschuhen bieten Sie an?“
 „Gibt es die auch in Größe 46?“

Ohne NLU mit entsprechend aufgebauter Wissensdomäne und Kontexterkenner würde das Voice-System den zweiten Satz nicht einordnen können. Das Wort „die“ erschliesst sich nur aus dem Kontext. Und gibt ein Kunde das Wort Wanderschuhe anstelle Trekkingschuhe ein, ohne dass es eine Verknüpfung der beiden Begriffe Wanderschuhe und Trekkingschuhe gibt, bzw.

das Wort Wanderschuhe im Vokabular des Systems vorkommt, endet der Dialog abrupt. Der Aufbau spezifischer Wissensdomänen ist also die Basis für das Verstehen der Kundenabsichten und des Inhaltes.

Was für Menschen einfach und selbstverständlich ist, muss den Voice-Systemen also antrainiert werden. Dabei sinkt der Aufwand für das Training und den Aufbau branchenspezifischer Wissensdomänen wie z.B. für den Kundenservice eines Automobilunternehmens mit jeder neuen Implementierung des einmal trainierten Systems in der Branche. Die Wissensdomäne einer weiteren Automarke wird sich nur noch marginal im Vergleich zur ersten ändern. Damit profitieren Voice-Systeme vom Gesetz des zunehmenden Grenzertrags wie es auch für andere digitale Anwendungen gilt.

NLG und NLP: Warum Relevanz wichtig ist

Der zweite wesentliche Aspekt in der Sprachverarbeitung (NLP: Natural Language Processing) betrifft die Relevanz des Dialogs. Relevanz bedeutet in diesem Zusammenhang, dass alle notwendigen Informationen, die für einen individuellen Dialog benötigt werden, in Echtzeit vom Voice-System aus dem CRM-, Ticketing-, Logistik-, Wissensmanagement-System abgerufen und verarbeitet werden. Wenn man den Beispiel-Dialog „Trekkingsschuhe“ mit der Frage: „Sind die Schuhe sofort verfügbar?“ weiterführt, wird die Notwendigkeit der Echtzeitverknüpfung mit Daten aus diversen Systemen erkennbar. Das Voice-System wird bspw. durch eine Echtzeit-Abfrage an das Warenwirtschaftssystem die Antwort erhalten, dass die angefragten Trekkingsschuhe der Marke Meindl in Größe 46 aktuell nicht vorrätig sind.

Diese Daten fließen in den Dialog mit dem Kunden ein.

„Tut mir leid, dieses Modell ist in Größe 46 aktuell nicht lieferbar. Einen genauen Liefertermin kann Ihnen leider nicht mitteilen.“

Bis zu diesem Zeitpunkt ist dies ein ordentlicher Dialog des Voice-Systems. Für eine sehr gute Customer Experience kommt noch ein weiterer Aspekt hinzu, den wir als Conversational Design bezeichnen. Das System muss also durch Menschen trainiert werden, die ihm beibringen, wie man nutzstiftende und gute Dialoge auch im Sinne

des Unternehmens führt. Eine für den Kunden und das Unternehmen sehr gute Antwort wäre dementsprechend:

„Tut mir leid, dieses Modell ist in Größe 46 aktuell nicht lieferbar. Ich habe in Größe 46 ein Modell von Salewo, welches die gleiche Farbe hat und von unseren Kunden sehr gut bewertet wird. Zudem kann ich Ihnen auf dieses Modell einen Rabatt von 20% gewähren.“

So, oder so ähnlich würde ein Mensch – ein guter Verkäufer – in dieser Situation in einem Fachgeschäft agieren. Ein Voice-System kann einen solchen Dialog ebenfalls führen, wenn es die o.a. Voraussetzungen erfüllt. In unserem Beispiel ist neben der Abfrage an das ERP-System auch noch der Abgleich mit einem CRM-System erfolgt, bei dem der Kunde aufgrund seiner Bestellhistorie oder seines Kundenwertes einen Rabatt erhält. Voice in der D2C-Journey einzusetzen, bedeutet immer den Kontext des Kunden – seine aktuelle Position in der Customer Journey – also seinen Intent zu verstehen und die Antwort individuell und persönlich auszugestalten.

Es kann also festgestellt werden, dass Sprachverarbeitungssysteme durch zunehmend günstigere Datenverarbeitung und leistungsfähigere Machine-Learning-Anwendungen wie NLP (Dahm & Constantine, 2020) immer häufiger in der Lage sind:

- gesprochene Sätze in immer mehr Sprachen in Echtzeit zu verstehen,
- die Absicht (Intent), die der Kunde mit seiner Formulierung verbindet zu verarbeiten,
- aus dem Kontext heraus zu agieren,
- relevante Dialoge zu führen (Kannan & Bernoff, 2019) und
- darauf basierend, Prozesse zu verarbeiten.

Kernthesen

1. Voice ist besonders geeignet, um in einer D2C-Umgebung automatisierte, kostengünstige direkte Dialoge mit dem Endkunden zu führen.
2. Der Einsatz von Voice benötigt Wissen um Intent, Kontext und Dialogauspielung.
3. Der Einsatz von Voice setzt eine aufgabenorientierte Dialogführung voraus. – Welche Aufgabe will der Kunde erledigen?
4. Voice benötigt eine hohe Relevanz, muss also an den richtigen Stellen der Customer Journey stattfinden.
5. Voice-Systeme müssen nahtlos in die bestehende Systemlandschaft des Unternehmens integriert sein, um in Echtzeit funktionieren zu können.

Abb. 5: Dialoge nach den Prinzipien von Grice

Konversationsmaximen nach Grice		
	Ja	Ich habe Lust auf Eiscreme! Gibt es ein Geschäft in Nähe?
	Nikola Tesla Str. 4	Ok... Wo?
	Nein	Ah. Haben sie geöffnet?
	Sie öffnen um 10 Uhr.	Hm... Wann öffnen sie?
	Nein	Gut. Verkaufen sie dort Eiscreme?
	Ja	Oh. Gibt es in der Nähe einen Laden, der Eiscreme verkauft?
	Hauptstr. 10.	Argh, wo denn?
	Nein	Haben die geöffnet?
	Nein	Willst Du mich zum Narren halten?
		
Ich habe Lust auf Eiscreme! Gibt es ein Geschäft in Nähe?		
	Ja. Es gibt ein Geschäft, das Eiscreme verkauft in der Hauptstr 10	Haben die geöffnet?
	Nein. Sie öffnen um 10 Uhr.	
		

Quelle: Cimiano, Ph. (o.J.).

Es ist daher auch nicht länger notwendig, bestimmte, vorgegebene Wörter oder Formulierungen zu nutzen, so wie es die Sprachdialogsysteme älterer Generationen erforderten, sondern entspricht eher dem natürlichen Dialogverlauf zwischen Menschen. Kunden können mit Voice-Systemen und digitalen sprachgesteuerten Assistenten nahezu so kommunizieren wie mit dem menschlichen Gegenüber (McLean et al., 2021).

Unternehmen ist es so möglich – unabhängig vom Händler – direkte und bezahlbare Dialoge mit ihren Kunden und somit relevante Bestandteile der Customer Journey in Eigenregie beeinflussen zu können. Daraus resultiert in

nicht unerheblichem Masse Loyalität und Wiederkaufintention (Hernandez-Ortega & Ferreira, 2021). Um die geeignete Art und Weise des Conversational Designs verstehen zu können, zeigen wir im letzten Abschnitt dieses Fachartikels die Prinzipien auf, denen automatisierte Dialoge für eine herausragende Customer Experience folgen sollten.

Voice-Systeme benötigen Conversational Designer und Linguisten

Damit Dialoge zwischen Kunden und den Voice-Systemen der Marke wirkungsvoll und nützlich sind, müssen

sie jedoch bestimmten Prinzipien folgen, die für Unternehmen, die bislang primär über den Handel verkaufen und kommunizieren, bislang zu wenig im relevanten Kompetenzset waren und im Rahmen von D2C aufzubauen sind. Diese gehen auf den englischen Sprachwissenschaftler H. Paul Grice zurück. Dieser hat vier Grundprinzipien menschlicher Interaktion entwickelt, um einen wirkungsvollen Dialog zwischen Kunde und Unternehmen zu entwickeln (Grice, 1975):

- Maxime der Quantität,
- Maxime der Qualität,
- Maxime der Relation und
- Maxime des Stils/der Modalität.

Obwohl hier eine verhältnismässig alte Quelle vorliegt, können diese Punkte in die Gestaltung eines effektiven und nützlichen Dialogs mit einem Voice-System übersetzt werden:

- Das System muss genau die richtige Menge an Informationen geben.
- Die Informationen müssen korrekt und aktuell sein.
- Die Antworten müssen relevant sein.
- Die Antwort muss klar und verständlich sein.

Menschen nehmen textliche und visuelle Informationen anders wahr und verarbeiten diese anders als Sprache. Bei der visuellen Verarbeitung gehen Menschen selektiv vor. Unser Unterbewusstsein arbeitet mit Filterfunktionen (Poort et al., 2015). Unsere Aufmerksamkeit wird im ersten Schritt auf für uns wichtige Dinge gelenkt. Andere, eher unwichtigere Informationen, werden dann erst in einem zweiten Schritt verarbeitet und wahrgenommen. Sprache wird anders wahrgenommen und verarbeitet. Das, was Menschen hören, wird in der Reihenfolge des Gesagten

aufgenommen. Die hat Konsequenzen auf die Gestaltung des Dialogdesign von Voice-Systemen. Voice-Dialoge müssen in kleinere Bestandteile unterteilt werden, da Menschen auditiv weniger verarbeiten als visuell. Eine 1:1-Übertragung von Dialogen bspw. aus E-Mails auf Sprache ist daher nicht angezeigt. Amazon hat eine Voice-spezifische Ausgestaltung daher zur Grundlage für das Design von Alexa Skills erhoben. (Amazon, o.J.).

Nutzt man diese Prinzipien im Kundendienst-Kontext, kann die zwischenmenschliche Interaktion besser widergespiegelt werden. Ebenso lässt sich das Conversational Designern eines Voice-Assistenten merklich optimieren.

Beide dargestellten Dialoge geben dem Kunden dieselbe, richtige Antwort.

Die Kunst liegt darin, das Wissen um Kontext und Relevanz so anzuwenden, dass der Kunde weder zuviel noch zu wenig Informationen erhält. Diese Prinzipien, angewandt im Dialog mit den Kunden, spiegeln die menschliche Interaktion sehr viel besser wider und

sorgen für eine hohe Akzeptanz. Das Zusammenspiel von CX-Experten, Touchpointspezialisten und Conversational Designer sind demzufolge ein entscheidender Faktor für die erfolgreiche Umsetzung von Voice-Systemen für einen direkten Kundendialog. 

Handlungsempfehlungen

1. D2C-Unternehmen sollten Voice im Kundendialog nutzen.
2. D2C-Unternehmen sollten zusätzliche neue Fähigkeiten im Bereich der zugrundeliegenden Technologien, der Evaluierung von Intent der Kundschaft sowie der Gestaltung von Dialogen in Hinblick auf eine überragende Customer Experience aufbauen.
3. D2C-Unternehmen sollten die damit verbundenen Kosten sorgfältig und ganzheitlich abschätzen.

Literatur

Amazon (o.J.). "What Alexa says. Amazon Alexa. <https://developer.amazon.com/de/designing-for-voice/what-alexa-says/>

Aspect (2017). Voicebot. Blogs Aspect. <http://blogs.aspect.com/wp-content/uploads/2017/07/building-chatbots.png>

Cimiano, Ph. (o.J.). Mercury.ai Bielefeld. Mercury. <https://www.mercury.ai/de/>

Dahm, M. H. & Constantine, B. (2020). Machine Learning für den Mittelstand – Eine empirische Annäherung. In M. H. Dahm & St. Thode (Hrsg.), *Digitale Transformation der Unternehmenspraxis – Mindset, Leadership, Akteure, Technologien*. SpringerGabler.

Grice, H. P. (1975). Logic and conversation. In P. Cole & Morgan, J. L. (Hrsg.), *Speech acts (= Syntax and Semantics. Bd. 3)* (S. 41–58). Academic Press.

Hafner, N. (2019). Artificial Intelligence and Big Data im Kundenservice: Reality Check und Ausblick. In P. Gentsch (Hrsg.), *Künstliche Intelligenz für Sales, Marketing und Service – Mit AI und Bots zu einem Algorithmic Business* (S. 163–173). SpringerGabler.

Hernandez-Ortega, B. & Ferreira, I. (2021). How smart experiences build service loyalty: The importance of consumer love for smart voice assistants. *Psychology & Marketing*.

Kabel, P. (2020). *Dialog zwischen Mensch und Maschine*. SpringerGabler.

Kannan, P. V. & Bernoff, J. (2019). The age of intent: using artificial intelligence to deliver a superior customer experience. P.V. Kannan.

Kedwes, A. (2021). Alexa und Siri lernen Schweizerdeutsch: Jetzt ist Ihre Mundart gefragt. *Tagesanzeiger*. <https://www.tagesanzeiger.ch/alexa-und-siri-lernen-schweizerdeutsch-jetzt-ist-ihre-mundart-gefragt-681172479054>

Kinsella, B. (2019). Why tech giants are so desperate to provide your voice assistant. *Harvard Business Review*. <https://hbr.org/2019/05/why-tech-giants-are-so-desperate-to-provide-your-voice-assistant>

Lemon, K. N. & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing, AMA/MSI Special Issue*, 80, 69–96.

Marxen, R. (2019). Alexa im Krankenhaus? Voice Assistants im klinischen Alltag. *Healthrelations*. <https://www.healthrelations.de/alexa-im-krankenhaus-voice-assistants-im-klinischen-alltag/>

McLean, Graeme J., Osei-Frimpong, K. & Barhorst, J. B. (2021). Alexa, do voice assistants influence consumer brand engagement? Examining the role of AI powered voice assistants in influencing consumer brand engagement. *Journal of business research*, 124, 312–328.

Poort, J., Khan, A., Pachitariu, M., Nemri, A., Osolic, I., Krupic, J., Baura, M., Sahani, M., Keller, G., Mrsic-Flogel, T. & Hofer, B. (2015). Learning Enhances Sensory and Multiple Non-sensory Representations in Primary Visual Cortex. *Neuron*, 88(6), 1478–1490.

Price, B. & Jaffe, D. (2006). *The Best Service is No Service: How to Liberate Your Customers from Customer Service, Keep Them Happy, and Control Costs*. John Wiley & Sons.

Ruan, S., Wobbrock, J.O., Liou, L., Ng, A. & Landay, J. (2018). Speech Is 3x Faster than Typing for English and Mandarin Text Entry on Mobile Devices. *Hci Stanford*. <https://hci.stanford.edu/research/speech/paper.html>

Lv, X., Zhang, M. & Li, H. (2008). 2008 IEEE International Conference. *IEEE Explore*. <https://ieeexplore.ieee.org/Xplore/home.jsp>