Das Autoren-Trio: Matthias, Martina, Benjamin

Sonderdruck
aus IT Spektrum 06/2022

Ausgabe 06|2022
Deutschland € 12,90 Österreich € 13,90 Schweiz sfr 22,20

4 194164 112900 06

www.ITSpektrum.de

vormals OBJEKTspektrum

IT Spektrum

Digitaler Wandel & Software-Architektur für Profis

Multimodales Design – die Erfolgsformel für Sprachassistenten

Multimodales Design – die Erfolgsformel für Sprachassistenten

Effiziente Kommunikationsverfahren
Microservices unabhängig und mit hoher Verfügbarkeit umsetzen

Was Manager von Games lernen können
Ein Erfolgsfaktor für Führung und Agilität

Architektur-Porträt #3
Corona-Warn-App

MAIBORNWOLFF

G 6540F

Multimodales Design – die Erfolgsformel für Sprachassistenten

VUIs sind unsichtbar und überall

Sprachassistenten sind mittlerweile weit verbreitet. Integriert in Smartphones, Speakern oder Autos sind sie als „unsichtbare" Interfaces allgegenwärtig. Neben Smarthome und Entertainment spielen sie im Gesundheitswesen und vielen weiteren Wirtschaftszweigen eine immer größere Rolle. In diesem Artikel zeigen wir auf, entlang welcher Fragen Designer einen Sprachassistenten gestalten, wo Chancen und Grenzen von Voice-Assistenten liegen und wie Voice User Interfaces in eine multimodale Customer Journey eingebettet werden.

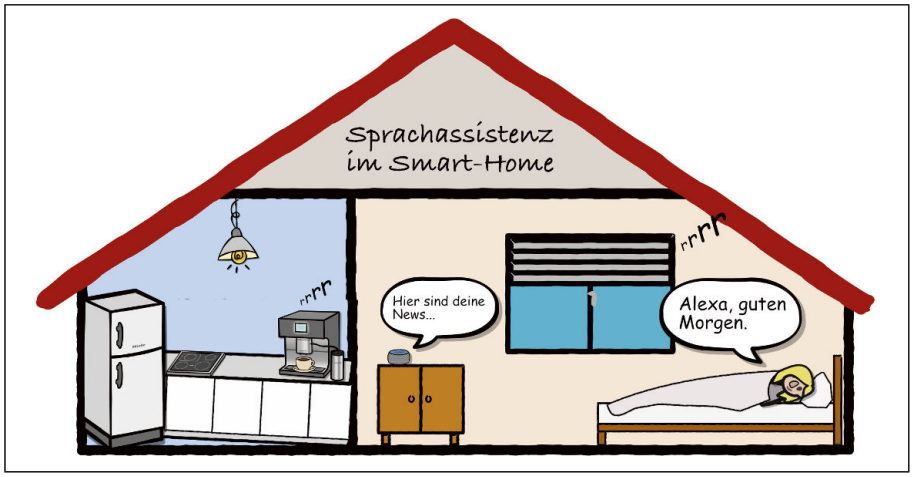


Abb. 1: Bedienung über Stimme im Smarthome

„Alexa, guten Morgen!" – die Rollläden fahren herauf, Verkehrsmeldungen werden ausgegeben, in der Küche wird das Licht eingeschaltet und die Kaffeemaschine heizt vor. All dies geschieht, bevor der Morgenmuffel seinen ersten Schritt aus dem Bett getan hat. Eine individuelle Sprachassistenz – mit smarten Endgeräten verknüpft – macht das möglich! (**siehe Abbildung 1**).

Komplexe Vorgänge, durch eine einzige Aussage ausgelöst, sowie das Führen ganzer Dialoge mit einem Sprachassistenten sind üblich. Doch wo setzt man Sprachassistenten am besten ein? Und wie gestaltet man solche Voice User Interfaces (VUI)?

Als Nutzer sind wir an den Umgang mit Grafischen User Interfaces (GUI) gewöhnt. Auch die Denke im Software-Ingenieurwesen ist noch stark von der Gestaltung visueller Schnittstellen geprägt. Aber VUI ist nicht GUI. Und damit ist vieles anders. Für Nutzer und Designer geht mit VUI ein Paradigmenwechsel einher.

Chancen von Voice-Schnittstellen

Warum sollten Unternehmen sich dem genannten Paradigmenwechsel und den damit einhergehenden Herausforderungen stellen? Weil mit anderen über die Stimme zu agieren, die natürlichste Form der Interaktion ist [Das19]. Und weil VUI Chancen eröffnen, die andere Technologien so nicht bieten.

VUI erschließt neue User-Gruppen

Gesprochene Sprache als Eingabemodalität öffnet neuen Gruppen die Tür in die digitale Welt. So können Menschen, die sich mit grafischen Oberflächen schwertun, dennoch digitale Services nutzen. Ältere Menschen interagieren mit einem Sprachassistenten oft besser als mit einem Computer oder Smartphone. Es fällt ihnen leichter, einen Assistenten nach dem Wetter zu fragen, als einzelne Buchstaben auf einer Tastatur zu suchen und das Wetter via App abzufragen.

VUI begünstigt Barrierefreiheit

Sprachsteuerung gestattet Menschen mit Einschränkungen, ihr Leben eigenständiger zu gestalten. So können Blinde Haushaltsgeräte wie die Waschmaschine, Kaffeemaschine, Spülmaschine usw. intuitiv steuern. Mobilitätseingeschränkte Personen können die Haustür für den Besuch per Sprachsteuerung öffnen lassen, ohne sich selbst zur Tür quälen zu müssen.

VUI bietet persönliche Assistenz

Durch die natürliche Sprache als Medium für die Interaktion mit dem Sprachassistenten wird das Nutzererlebnis individuell und persönlich. Die Nutzer können bei gesprochener Sprache ihre eigene Ausdrucksweise wählen. Die in den Sprachassistenten integrierte künstliche Intelligenz lernt mit und adaptiert nach und nach das Verhalten. Die Assistenz wird „persönlich", die Bedienung individuell.

VUI ermöglicht Multitasking

Generell bewerkstelligt ein VUI Sprechen und Agieren zugleich: Der Nutzer kann den Kuchenteig kneten und per Sprache die Temperatur des Ofens voreinstellen. Diese intuitive und freihändige Bedienung macht den Nutzer effizienter und steigert den Komfort.

VUI schafft neue Nutzungserlebnisse

Wird ein VUI in einer Customer Journey mit anderen digitalen Zugangsformen kombiniert, sprechen wir von multimodalem Design. Dieses bedenkt die Vorzüge aller Interfaces – grafisch, sprachlich sowie haptisch – und die technischen Eigenschaften der Geräte. Optimal kombiniert ermöglichen sie eine neue Form der User Experience.

Ein prominentes Beispiel für Multimodalität sind Navigationsgeräte. Eingaben können über Sprache getätigt werden, während der Weg angesagt und die Karte auf dem Display dargestellt wird. Multimodal verschmelzen haptische und visuelle Interfaces des Autos und der App mit dem Sprachinterface und werden zu einem Nutzererlebnis.

Der große Unterschied – sequenziell statt parallel!

Zunächst ist es den Nutzern vermutlich nicht bewusst: Die Interaktion zwischen Mensch und Maschine kann bei einer grafischen Bedienoberfläche hochgradig

parallel erfolgen, während sie bei einem VUI vorwiegend sequenziell stattfindet.
Wie **Abbildung 2** zeigt, können bei einer grafischen Darstellung auf einem Bildschirm verschiedene Informationsebenen dargestellt, abgefragt und übermittelt werden: Meta-, Struktur- und Bedieninformation. Bei entsprechender Aufbereitung entsteht zusätzlich eine Gewichtung der Informationen und die Aufmerksamkeit der Nutzer wird gelenkt.
Ähnlich verhält es sich mit haptischen Bedienelementen: Anhand der Beschriftungen weiß der Nutzer, an welche Position er den Drehknopf der Waschmaschine stellen muss. Ein Display bestätigt die Auswahl und informiert über die Programmdauer, während ein blinkender Startknopf signalisiert, was als Nächstes zu tun ist.
Bei einer reinen Sprachanwendung entfällt diese zusätzliche Metaebene, was zu einer sequenziellen Abfolge von Ausgaben, Rückfragen, Antworten und Handlungsanweisungen führt (**siehe Abbildung 2**). Hier ist der VUI-Designer mit einem vorausschauenden, robusten und effizienten Dialogdesign gefragt.

Herausforderungen von Voice-Schnittstellen

Für Nutzer und Gestalter ergeben sich weitere Herausforderungen, die es im Design zu adressieren gilt:

Gesprochene Sprache ist individuell
Natürliche Sprache lässt viel Freiraum, ist interpretationsfähig und kann Missverständnisse erzeugen. Wortwahl, Sprechgeschwindigkeit, Artikulation und Betonung unterscheiden sich bei den Nutzern erheblich, sodass das Verstehen der Nutzerabsicht für die Applikation zur anspruchsvollen Kernaufgabe wird. Wie individuell sprechen Nutzer? Sagen sie „Starte den Saugroboter“, „Saug die Wohnung!“ oder „Saugi, mach’ die Bude sauber!“?

Keine „sichtbare“ Unterstützung bei der VUI-Nutzung möglich
In Ermangelung eines „sichtbaren“ Layouts bei einem VUI erhalten Nutzer keine Hilfe, wann und wie sie mit dem System interagieren können. Bei Wartezeiten ist nicht erkennbar, ob das System noch rechnet. Darüber hinaus ist für Nutzer auch nicht – wie zum Beispiel bei GUI anhand eines neuen Buttons – erkennbar, dass die Applikation weiterentwickelt wurde und nun neue Funktionen enthält.

Das Kompetenzniveau der Nutzer ist sehr unterschiedlich
Der Umgang mit Voice-Technologie ist für VUI-Neulinge ungewohnt, für „alte Hasen“ selbstverständlich. Daher ist es eine Herausforderung, Erstere über eine gute Dialoggestaltung abzuholen und gleichzeitig Letztere nicht mit lästigen Dialogen und Hilfetexten abzuschrecken.

Während der Nutzer bei grafischen Darstellungen über die Hilfen hinwegscrollen kann, besteht bei VUI die Gefahr, lange „zugetextet“ zu werden.

Gutes VUI-Design – Materialkunde und multimodale Denkweise

Wie diese Herausforderungen im Design am besten adressiert werden, hängt vom Kontext und von der Erfahrung im Team ab. Initiatoren eines digitalen Vorhabens mit Voice-Bezug sollten von Anfang an Digital Designer mit entsprechendem Know-how und multimodaler Denke an Bord haben.
Denn während Gestalter von grafischen Anwendungen auf langjährig bewährte Designparadigmen und -praktiken zurückgreifen können, sind diese im VUI-Kontext erst im Entstehen. Die Entwicklung ist somit oft durch Trial & Error sowie inkrementelle Verbesserungen geprägt, jedoch machen eine solide Kenntnis des digitalen Materials rund um VUI-Technologie und das Wissen um ihre Möglichkeiten und Grenzen – wie in [Bit17] und [Bit18] gefordert – den Unterschied. Einen Überblick über die Grundlagen von VUI-Technologie und -Terminologie findet sich in [Bec22].

Womit startet das Design am besten?
Wie in jedem digitalen Vorhaben empfiehlt sich eine Orientierung am User Centered

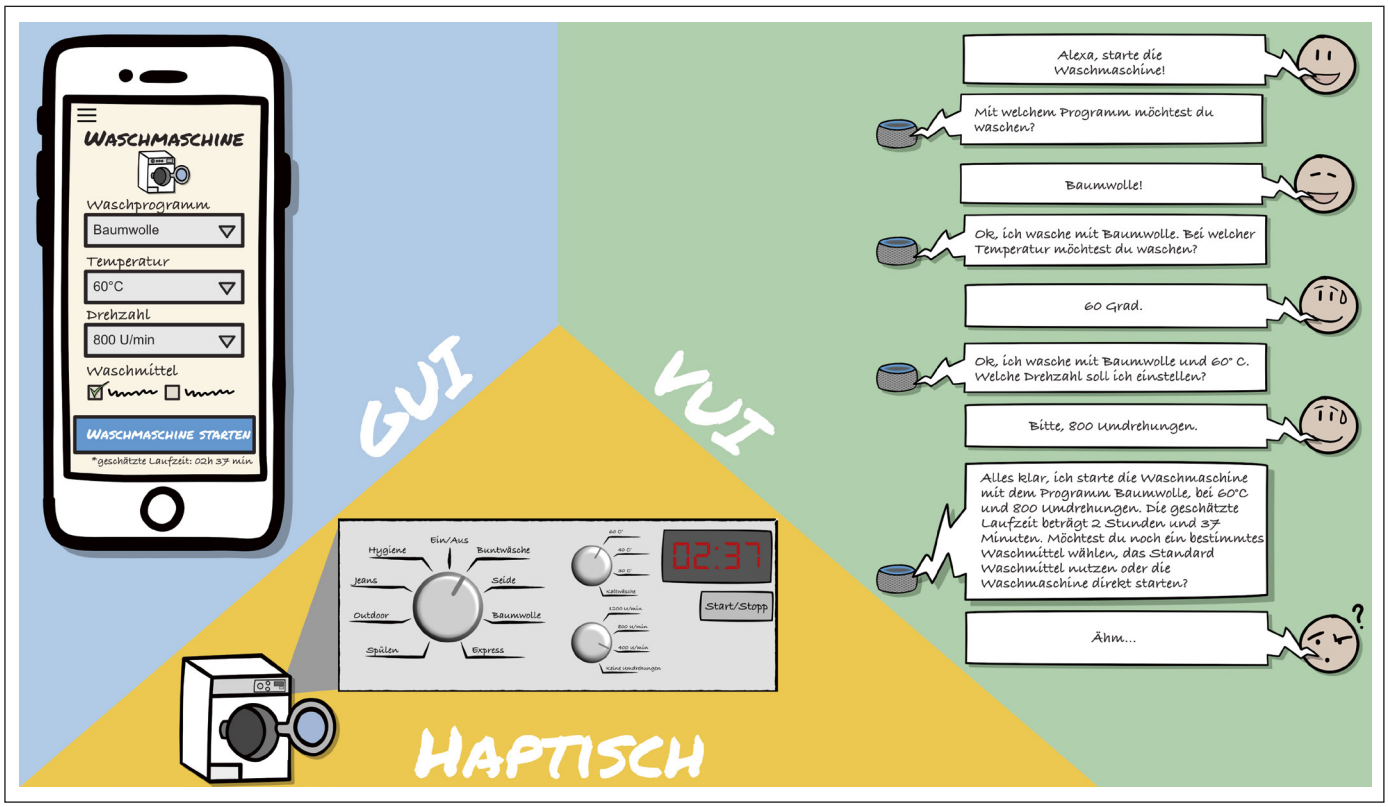


Abb. 2: VUI versus GUI und haptischer Bedienung

Angst vor Kontrollverlust?

Nicht alle Menschen stehen einer Sprachassistenz positiv gegenüber. Hinter Voice-Technologie steckt in der Regel eine künstliche Intelligenz, die ständig dazu lernt. Das führt zu einem Dilemma: Handelt eine Sprachassistenz mit der Zeit „zu schlau" und zu invasiv, empfinden das viele Nutzer als anstrengend oder unheimlich. Tut sie das nicht, wird die Sprachassistenz als „nicht smart genug" und unbrauchbar abgestempelt.

Die „gefühlt ewig lauschenden" Smart Speaker – also Mikrofone mit einem offenen Kanal ins Internet – machen vielen große Sorge: Wer hört da alles mit? Was ist mit Datenschutz? Gerade deutsche User treibt dies um. Dennoch besitzen ein Drittel aller deutschen Haushalte Smart Speaker. [Omd21]

Bei der Gestaltung von Sprachanwendungen sollten Designer dieses Empfinden der Nutzer berücksichtigen und eine Balance finden zwischen einem zu invasiven und einem zu defensiven Design.

Invasiv – ja, fast schon übergriffig – wirken Aussagen eines Sprachassistenten wie:

- „Hey, du hast schon lange nicht mehr gesaugt, willst du es nicht mal wieder tun?"
- „Alles klar, ich starte die Waschmaschine. Übrigens, Weichspüler ist im Sonderangebot. Möchtest du einen haben?"
- „Hey, ich habe gehört, ihr habt über Pizza diskutiert. Soll ich dir, wie immer, eine Pizza Funghi bestellen?"

Proaktives Handeln des Assistenten kann das Nutzererlebnis jedoch auch deutlich verbessern, wie beispielsweise die Nachricht: „Das Waschprogramm ist beendet!" oder „Blumengießen nicht vergessen!".

Die Bedürfnisse der Nutzer sind hier stark unterschiedlich. Es empfiehlt sich, dort, wo es möglich ist, Anwender das Verhalten des Assistenten nach eigenen Vorlieben konfigurieren zu lassen.

Design (UCD). Wir haben den Prozess in **Abbildung 3** auf multimodale Sprachanwendungen adaptiert.
Zuerst gilt es, das Vorhaben und sein Umfeld zu analysieren, um herauszufinden, ob und inwieweit eine Sprachschnittstelle die richtige Lösung ist.

Kontext:
- In welchem fachlichen Kontext bewegt sich das Vorhaben?
- Um welche Use Cases handelt es sich?
- Wie betten sich die Use Cases in die Customer Journey ein?
- Sollen physische Geräte in die Customer Journey eingebunden werden?
- Welche Ziele werden mit der Anwendung verfolgt (z. B. Barrierefreiheit)?

Zielgruppe:
- Wer ist die anvisierte Zielgruppe?
- Zielt die Applikation auf mobilitätseingeschränkte Menschen oder auf Menschen mit eingeschränktem Sehvermögen ab, die auf Sprachsteuerung angewiesen sind?
- Wie heterogen gestaltet sich die Zielgruppe?
- Welches Kompetenzniveau in Bezug auf den Umgang mit VUIs ist erwartbar?

Auch, wenn es sich vielleicht um eine „coole, neue Technologie" handelt, die manch Unternehmen gern im Einsatz hätte, ist Voice-Technologie für den Nutzer nicht in jedem Fall gebrauchsfreundlich und sinnvoll. Daher sollte verifiziert werden, ob Sprachsteuerung die beste Modalität für den Anwendungsfall ist.

Wann macht eine Voice-only-Anwendung Sinn?
Ein Paradebeispiel sind sogenannte *Wiki-Anwendungsfälle*. Der User stellt eine Frage, woraufhin die Applikation passende Antworttexte erstellt und diese akustisch ausgibt.

Beispiele für Voice-only-Anwendungen:
- *Informationsabfragen:* User: „Wie ist das Wetter heute in Frankfurt?"
  VUI: „In Frankfurt ist es heute sonnig bei Temperaturen zwischen 23 und 26 Grad."

- *Wikis:* User: „Wie lange kochen Kartoffeln?"
  VUI: „Kartoffeln müssen, je nach Größe, im Schnitt 30 Minuten köcheln."

- *Erinnerungen:* User: „Erinnere mich morgen um 18 Uhr daran, die Blumen zu gießen."
  VUI: „OK. Ich erinnere Dich morgen um 18 Uhr ans Blumengießen."
  VUI (am nächsten Tag): „Hier ist deine Erinnerung: Blumengießen."

- *Unterhaltungsanwendungen* wie interaktive Hörspiele.

Neben einfachen Frage-Antwort-Anwendungen eignen sich auch „One-Shot-Befehle" wie „Alexa, schalte das Licht an." Eigens konfigurierte Routinen, wie das anfangs beschriebene „Alexa, guten Morgen!", lösen mit nur einem Befehl eine Reihe von Aktionen aus und steuern gleich mehrere Geräte. Auch kurze und einfache Dialoge können rein über Voice modelliert werden. Folgende Kriterien finden sich dafür in [Das19]:

- Die Interaktion ist kurz, mit einem Minimum an Hin und Her in der Interaktion.
- Die Benutzer können die Aufgabe durch Konversation erledigen, auch wenn sie beschäftigt sind und nicht die volle Aufmerksamkeit aufbringen können.
- Die Nutzer empfinden es als sehr zeitaufwendig oder umständlich, dieselbe Aufgabe über ein grafisches Interface zu lösen.

VUI-Designer erkennen Use Cases, für die diese Kriterien nicht zutreffen und reine VUI-Applikationen an ihre Grenzen stoßen. Wenn Interaktionen über Sprache schwerfällig werden. Oder, wenn so große Mengen an Informationen ausgegeben werden, dass die Nutzer sie nicht mehr erfassen können. Die Folge: Die Akzeptanz sinkt! Beispiele hierfür sind das Vorlesen langer Zutatenlisten oder Anleitungen. Hier ist im Design multimodale Denkweise gefragt (siehe unten).

Welcher Sprachstil wird genutzt?
VUI-Designer sollten sich mit dem Sprachstil der Zielgruppe auseinandersetzen und Hypothesen dazu aufstellen:

- Welche Sprache sprechen die Nutzer?
- Welcher Sprachstil ist von der Zielgruppe als Eingabe zu erwarten?
- Welcher Sprachstil bietet sich für die Ausgaben der Sprachanwendung an?
- Gehoben, umgangssprachlich, Jugendsprache …?
- Ist die Zielgruppe hinsichtlich Dialekt oder Soziolekt eingrenzbar?
- Wird gesiezt oder geduzt?

Auf Basis dieser Hypothesen, die im Projektverlauf verifiziert oder falsifiziert werden, gestalten VUI-Designer die Applikation.

Mit viel Training lernen sich Sprachassistenz und Nutzer kennen
Für VUI-Gestalter hat es eine hohe Priorität, die Sprachassistenz – beziehungsweise die zugrunde liegende KI – permanent zu trainieren. Die Sprache der Nutzer und

ihre Anliegen schnell und präzise zu verstehen sowie diese in die richtigen Aktionen umzusetzen, wird zur Kernaufgabe. Nutzer und VUI-Applikation müssen sich „kennenlernen“, ihre Erwartungen artikulieren und eine „gemeinsame Sprache“ entwickeln.
Hier spielt das fortwährende Training der KI auf Basis echter Nutzungsdaten eine Schlüsselrolle. Es ist für die Gestaltung der Assistenz unerlässlich, so viel User-Feedback wie möglich zu erfassen, zu analysieren und die daraus gewonnenen Erkenntnisse in die Entwicklung einfließen zu lassen. Nur, wenn die Assistenz kontinuierlich verbessert wird, bietet sie so viel Mehrwert, dass sie aus dem Leben der Nutzer nicht mehr wegzudenken ist.

Wo sind die Grenzen des reinen VUI-Designs?
VUI-Designer wissen, wo *reine* VUI-Applikationen an ihre Grenzen stoßen. Folgende Fragen weisen den Weg Richtung multimodales Design:

- Sind die zu erwartenden Ausgabetexte zu lang (One-breath-test)?
- Wird der User in der Informationsaufnahme- und Merkfähigkeit überfordert, zum Beispiel durch zu viele Optionen?
- Werden Interaktionen unnötig kompliziert oder unnatürlich?
- Bieten sich weitere Interfaces, wie das Display des Smartphones oder eines IoT-Geräts, an?

Wenn hier die Antworten Richtung „ja“ tendieren, ist es sinnvoll, über *andere* oder *ergänzende* Zugangs- beziehungsweise Ausgabemöglichkeiten im Design nachzudenken.
Das folgende Beispiel zeigt dies deutlich: Fragt ein User über VUI bei einem Pizza-Service an, welche Pizzen es gibt, sollte der Sprachassistent keine 50 Pizzasorten vorlesen, sondern einfach antworten: „Ich habe dir die Menükarte auf dein Smartphone geschickt“.

Mehr als Voice: multimodale Customer Experience!

Wenn Voice-only-Lösungen an ihre Grenzen stoßen, ist multimodales Design gefragt. Letzteres führt zu Lösungen, in denen der User über *zusätzliche Ein- und Ausgabemöglichkeiten* (Display, Tastatur, Touch-Screen, ...) verfügt. Multimodale Interaktionen sind ganzheitlich und durchgängig zu modellieren. Dabei haben sich folgende Erfahrungen als nützlich erwiesen:

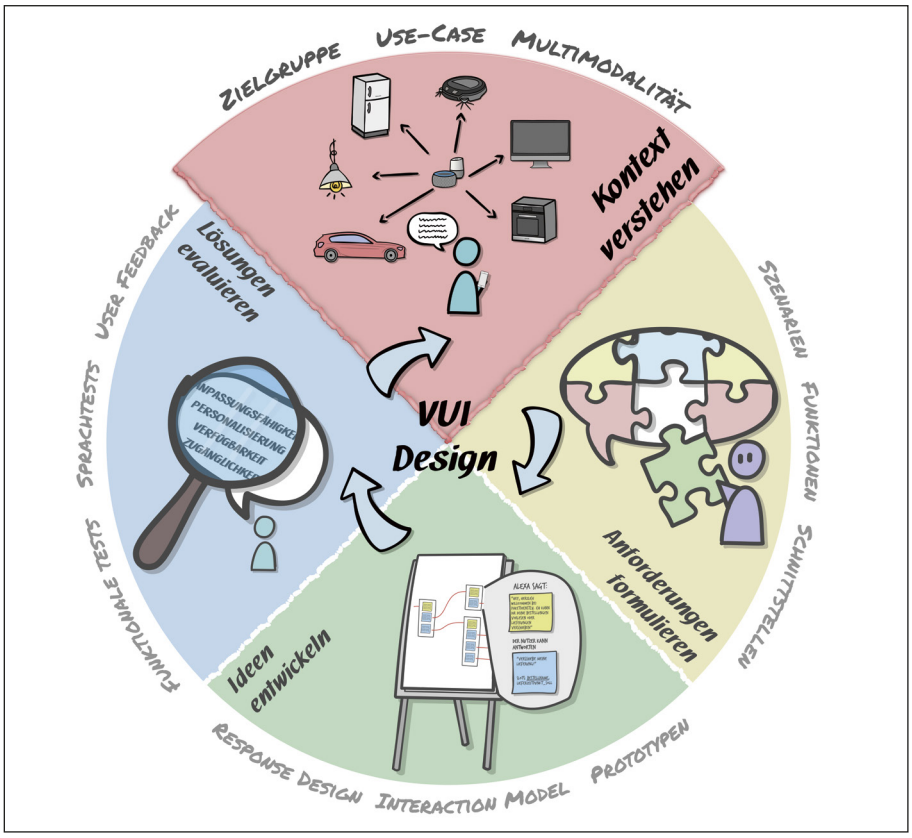


Abb. 3: User Centered Design Prozess (UCD) – adaptiert auf multimodale VUI-Applikationen

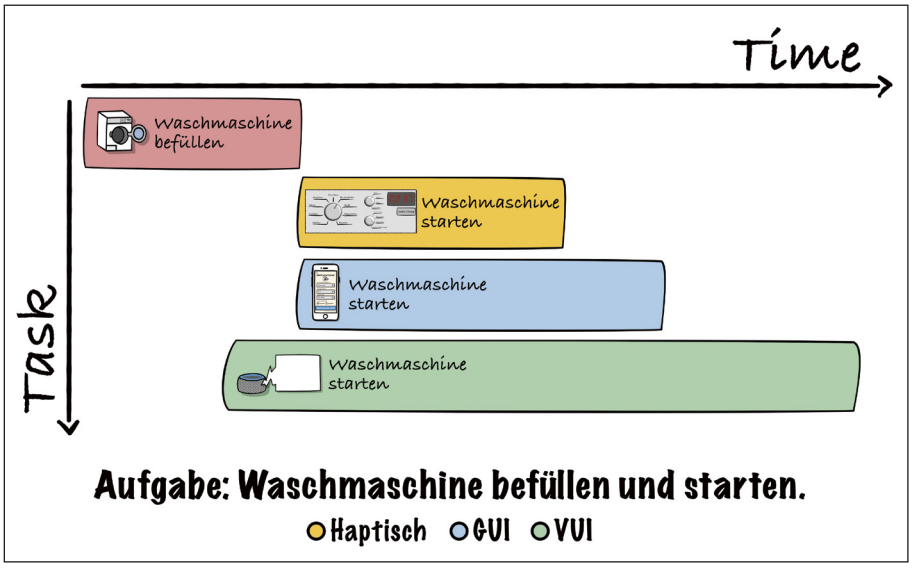


Abb. 4: Time-to-Task Gantt Chart am Beispiel Waschmaschine befüllen und starten

Empfehlung 1:
Beim Design an Time-to-Task orientieren
Um herauszufinden, welcher Teil einer Aufgabe sich als Voice-Use-Case eignet (und welcher nicht), ist die wichtigste KPI im multimodalen Design die *Time-to-Task:* Nutzer wählen das Interface, das sie am schnellsten zum Ziel bringt. In der Zeit, in der ein User in langen Dialogen jeden Parameter zur Steuerung seines IoT-Geräts via Voice an das Gerät sendet, hätte er längst drei Drehregler eingestellt und auf Start gedrückt.

Abbildung 4 veranschaulicht das in einem Gantt Chart. Zwar kann der User bereits während der Befüllung mit dem VUI interagieren, braucht für die Aufgabe jedoch deutlich länger.
Knetet der User einen Pizzateig und heizt per Sprachsteuerung den Ofen vor, hat sich seine Time-to-Task verkürzt. Tendenziell sind Menschen darauf aus, in Bezug auf die Zeit zu optimieren – insbesondere im Haushalt. Jeder Handgriff weniger wird gerne eingespart und ist Basis für das „smart“ in Smarthome.

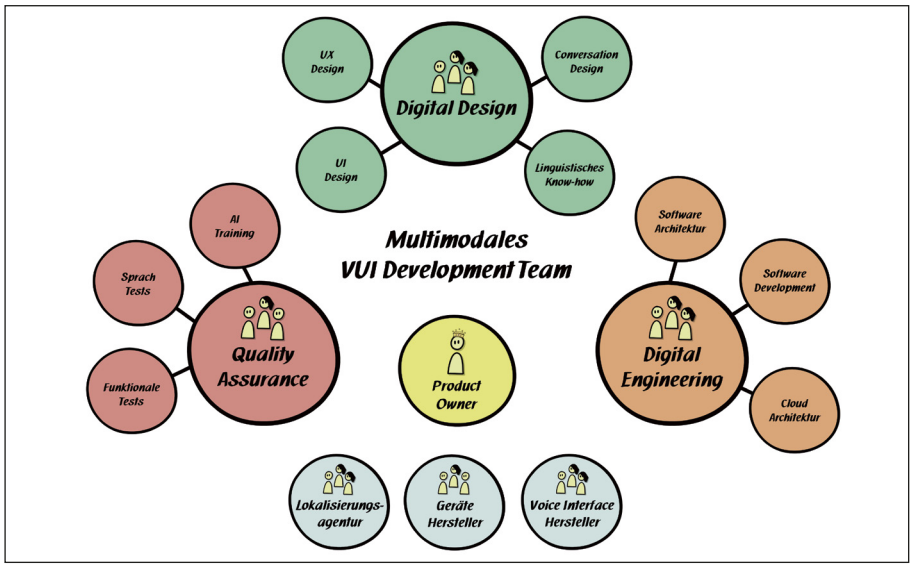


Abb. 5: Professionen und Spezialisierungen in multimodalen VUI Development Teams

Multimodales Design:

- Wie agieren unterschiedliche Modalitäten am sinnvollsten miteinander?
- Wie viel kann der User über welchen Weg/Sinn aufnehmen?
- Wie steht es um die Time-to-Task als Key-Metrik?
- Wie kommt der Nutzer am schnellsten an sein Ziel?
- Wie integrieren sich die Interfaces am besten in die Customer Journey?

Empfehlung 2:
VUI als ergänzendes Interface denken

VUI sollten im multimodalen Kontext ergänzend – nicht ersetzend – eingesetzt werden. Es gilt, die Stärken der verschiedenen Interfaces zu kombinieren und zu einer User Journey zu verschmelzen. Hier zwei Beispiele:

- *Einmalige Konfigurationsabfragen:* Wie konfiguriere ich mein Gerät? Über Sprache erhält der User den Hinweis auf ein einfaches Erklärvideo. Gerade einmalige Konfigurations-Use-Cases sind über Voice wenig sinnvoll.
- *Multimodale Kochassistenz:* Der Sprachassistent bietet Kochassistenz inklusive Rezeptbegleitung an. Wird das Design um visuelle Mengenangaben, Bilder des Gerichts und die Übermittlung der Programmparameter an das Küchengerät ergänzt, entsteht ein multimodales Erlebnis mit einem deutlich höheren Mehrwert für die Nutzer, als es sich aus der Summe der (isoliert betrachteten) einzelnen Anwendungsfälle ergäbe.

Empfehlung 3:
Innerhalb des Nutzungskontextes in einer konsistenten Sprache sprechen

Egal, ob direkt am IoT-Gerät, auf der Fernbedienung oder in der App, in Funktionen und Programmen: Alle Aufgaben, die dieselbe Aktion auslösen, sollten endgeräteübergreifend konsistent benannt werden:

- *Konsistente Benennung:* Wenn die höchste Gebläsestufe einer Dunstabzugshaube „Booster“ heißt, dann sollten Nutzer über Voice diese mit „Schalte den Booster an“ auch aktivieren können. Die Antwort sollte dann aber nicht sein: „Der Turbomodus wurde aktiviert“.
- *Einheitlicher Sprachstil:* Verwendet die App einen Sprachstil, in dem geduzt sowie flachsig und hipp gesprochen wird, sollte der User von der Sprachassistenz nicht gesiezt werden.

Generell sind Terminologie und Wortwahl zum Nutzer durchgängig zu gestalten und in Form von *Tonalitätsrichtlinien* festzuhalten. Ist ein VUI im Spiel, ist es umso wichtiger, dass visuelle und auditive Sprache übereinstimmen. Was sich über visuelle Mittel wie Icon-Sprache vermitteln lässt, ist über VUIs nicht immer einfach ausdrückbar. In Zeiten der Digitalisierung sollten daher Geräte-, App- und Voice-Entwicklung konzeptionell Hand in Hand laufen.

Empfehlung 4:
Wartezeiten kurzhalten und Signale senden

Langsame Technik bedeutet langsame Prozesszeiten. Dies führt bei Sprachsteuerung zu langem Warten auf Antwort und zu unangenehmen Pausen im Dialog. Beim User kann dies Verunsicherung und Ablehnung erzeugen. Daher spielt die technische Architektur der Geräte eine wesentliche Rolle dabei, ob ein Sprachassistent erfolgreich ist. Hersteller smarter Endgeräte sollten dies bereits bei der Entwicklung ihrer Modelle berücksichtigen, falls diese auch über cloudbasierte Voice-Interfaces angesteuert werden sollen.
Latenzzeiten können bei GUIs mit Ladebalken und Sanduhren überbrückt werden. Bei Voice muss dem Nutzer über Audio signalisiert werden, dass etwas im Hintergrund passiert. Für solche Wartezeiten müssen Designer schnelle und situativ angemessene Antworten gestalten:

- Je nach Dauer der Unterbrechung kann eine Art *Warteschleifen-Musik* integriert werden. Aber Achtung: Der Dialogfluss wird damit unterbrochen und der Spaß an der Bedienung könnte leiden.
- Veränderungen des Status – insbesondere Fortschritte – sollten dem User immer wieder signalisiert werden: nicht nur „aus“ oder „in Standby“, sondern auch „Gerät fährt hoch“ oder „Es findet eine Verarbeitung im Hintergrund statt“.

Beispiel: Eine Kaffeemaschine, die eine Minute lang hochfährt, bis sie betriebsbereit ist, sollte Statusmeldungen an den Nutzer absetzen. Sobald sie fertig ist, sollte sie signalisieren: „Ich bin bereit. Bitte stelle deine Tasse unter den Auslass.“
Gängige cloudbasierte Voice-Lösungen für Smarthome – wie Alexa und der Google Assistant – setzen in der Regel eine maximale Grenze für Antwortzeiten. Antwortet das smarte Endgerät nicht binnen weniger Sekunden, wird der Dialog beendet. Das lässt die Voice-Applikation und das Endgerät schlecht aussehen. Es ist wichtig, dass Designer und Product Owner die Gerätehersteller dahingehend sensibilisieren. Oder besser noch: Sie werden in den Entwicklungsprozess der Geräte involviert und können solche Anforderungen frühzeitig einfließen lassen.

Ein Team mit sich ergänzenden Professionen an Bord

Erst die Kombination unterschiedlicher Professionen, ganzheitlicher Denkweise und spezialisiertem Wissen ebnet den Weg zu einem erfolgreichen multimodalen Produkt (**siehe Abbildung 5**):

- Für das *Digital Design* im multimodalen Kontext braucht es grundlegendes Modellierungs-Know-how sowie ein Gespür für UX und UI-Design. Um

Literatur & Links

[Bec22] M. Beck, M. Linse, Voice-User-Interface-Design – wie ein Chatbot zu sprechen lernt, siehe: Voice-User-Interface-Design – ein Chatbot lernt sprechen, in: Informatik Aktuell (informatik-aktuell.de)

[Bit17] bitkom.org, Rollenideal Digital Design, siehe: **https://www.bitkom.org/Bitkom/Publikationen/Rollenideal-Digital-Design.html**

[Bit18] bitkom.org, Digital Design Manifest, siehe: **https://www.digital-design-manifest.de/**

[Das19] R. Dasgupta, Voice User Interface Design, Apress, Hyderabad, Telangana India, 2019

[Lau21] K. Lauenroth, Ganzheitliche Gestaltung der Digitalisierung erfordert eine neue Sprache, in: Jahrbuch Digital Design 2021, siehe: **https://www.bitkom.org/sites/default/files/2021-03/210318_digital-design_jahrbuch.pdf**

[Omd21] OMD Germany, Anteil der Haushalte in Deutschland, in denen es einen Smart Speaker gibt, von 2018 bis 2021, auf Statista veröffentlicht, siehe: **https://de.statista.com/statistik/daten/studie/1271603/umfrage/anteil-der-haushalte-in-deutschland-mit-smart-speaker/#professional**

möglichst natürliche Dialoge zu konzipieren, empfehlen sich Kenntnisse im Bereich Conversation Design und Linguistik. Hier zählt insbesondere das Wissen rund um die Sprechhandlungstheorie [Bec22, Bit17] sowie Dialoggestaltung.

- Bei IoT-Sprachanwendungen, die auf Cloud-Technologie basieren, sollte das *Engineering* mit ausreichend Cloud-Architektur-Expertise ausgestattet sein. Entwickler und Architekten sollten sich mit den Technologien rund um die Plattformen (z. B. Amazon Alexa oder Google Assistant) und Smarthome auskennen. Durch eine enge Zusammenarbeit mit den Designern entstehen Lösungskonzepte, die alle Geräte technisch einwandfrei einbinden.
- Multimodale Sprachanwendungen erfordern ein hohes Maß an *Qualitätssicherung*. Führt das Interface die Funktion am Endgerät korrekt aus? Versteht die künstliche Intelligenz, was der Nutzer sagt und meint? Die sprachlichen Ausgaben sind stetig auf Verständlichkeit und ein hochwertiges Hörerlebnis zu prüfen. Tester und KI-Trainer im Team verbessern und optimieren den Sprachassistenten gemeinsam.
- Ein starker *Product Owner*, der das Smarthome- und Voice-Thema lebt, ist zentral für den Erfolg. Er bildet die Schnittstelle zu Stakeholdern und stellt bei Bedarf die Verbindung zu externen Experten wie Lokalisierungsagenturen her. Der PO ist dafür zuständig, die Gerätehersteller in den Prozess einzubinden, sodass der multimodale Ansatz bereits bei der Konzeption neuer Geräte berücksichtigt wird.

Fazit – Von Beginn an multimodal denken

Sprachassistenten entfalten ihr volles Potenzial meist eingebettet in einen multimodalen Kontext. Bei der Gestaltung sollten Brüche in der Interaktion mit anderen Interfaces vermieden werden. Leitplanken für gutes multimodales Design lassen sich so zusammenfassen:

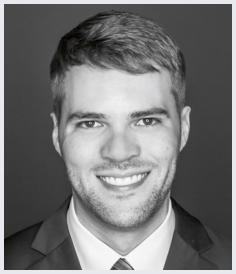
- Beim Design an Time-to-Task orientieren.
- Sprachassistenz als ergänzendes Interface denken.
- Innerhalb des Nutzungskontextes *eine* Sprache sprechen.
- Wartezeiten kurzhalten und Signale senden.

Leider werden in vielen Unternehmen physische Geräte und digitale Interfaces nicht als Einheit gedacht. Dies führt zu inkonsistentem Verhalten und steht dem Schaffen *eines* Nutzererlebnisses entgegen. Hier besteht großer Handlungsbedarf bei den Herstellern.
Für ein gutes VUI-Design im multimodalen Kontext ist ein interdisziplinäres Team zentral. Ein Team, das nicht auf seinem aktuellen Wissenstand stehen bleiben darf. Denn die Voice-Technologie entwickelt sich rasant. Da heißt es dranbleiben und: „Inspect and Adapt“! ||

Die Autoren

Dr. Martina Beck
(martina.beck@maibornwolff.de) ist seit über 25 Jahren Brückenbauerin zwischen Business und IT und Digital Designerin aus Leidenschaft. Begeisternde IT-Lösungen sind ihr Ziel. Sie liebt es, ihren Kunden die richtigen Fragen zu stellen und sich mit dem digitalen Material auseinanderzusetzen.

Matthias Linse
(matthias.linse@maibornwolff.de) entdeckte in einem Projekt zur Sprachsteuerung smarter IoT-Geräte seine Leidenschaft für Voice User Interfaces und Multimodalität. Als Digital Designer begeistert ihn das Zusammenspiel aus Fachlichkeit, Technik und Methodik.



Benjamin Vornholt
(benjamin.vornholt@miele.com) verantwortet als Product Owner Voice bei der Miele & Cie. KG die Entwicklung von Alexa Skills und Google Actions. Als Enthusiast für Sprachsteuerung und Smarthome ist er überzeugt von interdisziplinären Teams, nutzerzentriertem Design und agiler Softwareentwicklung.