Voice Control um jeden Preis?

Theoretische und praktische Grundlagen für erfolgreiche Sprachsteuerungs-Angebote aus User Experience-Sicht

Sandra Schuster
Facit Digital GmbH
Neuhauser Straße 17
80331 München
s.schuster@facit-digital.com

Abstract
Als Technologie ist Sprachsteuerung seit SIRI fast schon zur Gewohnheit geworden. Doch nur weil etwas technisch machbar ist, heißt es noch nicht, dass es auch „sinnvoll" ist. Eine zentrale Grundvoraussetzung ist die Passung von Usecase und Technologie (Sprachsteuerung). Zudem soll der Nutzer die Anwendung intuitiv in der intendierten Weise bedienen können (und wollen). Dies verlangt bei der Umsetzung die besondere Berücksichtigung von Nutzungssituation und Nutzungskontext sowie adäquater Prinzipien der Interface-Gestaltung.
Ziel unseres Projektes war es demzufolge, sowohl theoretische als auch praktische Grundlagen für ein erfolgreiches Voice Control-Angebot zu erarbeiten. Dabei werden u.a. folgende Fragen beleuchtet: Welche inhaltlichen Konzepte eignen sich eigentlich für den Einsatz von Sprachsteuerung? Wann schafft Voice einen Mehrwert gegenüber Touch? Welche Anforderungen stellt das Medium Sprache an IA und UI-Design?

Keywords:
/// Voice Control
/// GUI
/// VUI
/// MUI

1. Ausgangslage und Erkenntnisziel

„Among speech interface designers, there's a credo: A good GUI and a good VUI are both a pleasure to use, a bad GUI is hard to use, but a bad VUI isn't used at all." **(Abbott, 2002)**

Wie wahr dieses Credo von Kenneth Abbott zu sein scheint, konnten wir Anfang des Jahres im Rahmen eines Forschungsprojektes für einen Automobilhersteller erfahren. Inspiriert durch steigende Nutzungszahlen von SIRI und Google Speech, sollte die Entwicklung eines mobilen, durch Sprache gesteuerten Fahrzeug-Konfigurators Innovationskraft und Technologiestärke der Marke belegen.

Ein entsprechender Prototyp für die geplante Smartphone-App befand sich zu diesem Zeitpunkt bereits in der Konzeption und sollte durch einen User Experience-Test auf Bedienbarkeit, Nutzerakzeptanz und Optimierungspotenziale hin überprüft werden. Allerdings war schnell klar, dass vor der Überprüfung der konkreten Ausgestaltung noch einmal die ganz grundsätzliche Frage beantwortet werden musste: Passt Voice Control hier eigentlich? Ist die Konfiguration eines Fahrzeugs (auf einem mobilen Endgerät) tatsächlich ein „sinnvoller" Usecase für Sprachsteuerung? Und erst dann: Wie sieht die konkrete Ausgestaltung aus? Welche Parameter sind zu beachten, um eine möglichst optimale User Experience und damit auch Marktakzeptanz zu erreichen?

Für unser Forschungsprojekt leiteten sich darauf folgende zentrale Fragestellungen ab:

- Welche technologischen und gestalterischen Aspekte sind zu berücksichtigen? Speziell: Wie sieht das optimale Zusammenspiel zwischen verschiedenen User Interfaces (z.B. GUI und VUI) aus?
- Trifft das Angebot überhaupt ein Nutzerbedürfnis und/oder eine potenzielle Nutzungssituation? Beziehungsweise: Für welche Usecases bietet sich Sprachsteuerung in besonderer Weise an, für welche gegebenenfalls eher nicht?
- Wann schafft Sprache einen („echten") Mehrwert, z.B. gegenüber Touch?

Um diese Fragestellungen fundiert zu beantworten, wählten wir ein mehrdimensionales Vorgehen: Anhand einer umfassenden Desk Research wurden zunächst wesentliche (theoretische) Grundlagen für die Konzeption und Gestaltung von Anwendungen, die auf Sprachsteuerung basieren, erarbeitet. Durch eine ergänzende Best Practice-Recherche konnten dabei erste allgemeine Regeln und Prinzipien im Sinne von Dos and Don'ts formuliert werden, welche sich sowohl auf technologische und gestalterische als auch auf Usecase-basierte Aspekte beziehen.

Darauf aufbauend wurde ein qualitativer Kriterienkatalog entwickelt, anhand dessen der aktuelle Konzeptstand (semi-funktionaler Prototyp der App) in Form einer heuristischen Evaluation überprüft werden sollte. Vertieft wurde diese Betrachtung durch einen (unabhängig voneinander durchgeführten) Cognitive Walkthrough durch zwei Usability Experten von facit digital. Dabei wurden typische Tasks und Konfigurationsprozesse des Voice Control-Konfigurators schrittweise durchgespielt und Optimierungsmöglichkeiten

identifiziert sowie konkrete Handlungsempfehlungen für unseren Auftraggeber formuliert.

Der folgende Beitrag skizziert pointiert und beispielhaft einige zentrale Erkenntnisse aus dem Forschungsprojekt. Nach einer groben thematischen Einordnung und Begriffsklärung, werden relevante Aspekte für die Gestaltung von sprachgesteuerten (Marketing-) Angeboten aus verschiedenen theoretischen und praktischen Blickwinkeln beleuchtet. Den Abschluss bildet unser (leider wenig optimistisches) Fazit für eine sprachgesteuerte Fahrzeug-Konfigurator-App.

2. Thematische Einordnung und Begriffsdefinition

Dass eine Fahrzeugkonfiguration nicht ohne visuelle Stützung auskommt, erklärt sich von selbst. Ein rein auf Sprache basierendes Angebot war demzufolge von vorneherein ausgeschlossen. Vielmehr galt es, das grafische User Interface (GUI) – an möglichst passender Stelle – mit Sprachsteuerung (VUI) zu verbinden. Werden wie in diesem Falle mehrere verschiedene Input-Methoden (z.B. Touch, Gesten, Sprache, etc.) miteinander kombiniert, ist in der Literatur die Rede von „Multimodalen Systemen" bzw. „Multimodalen User Interfaces" (MUI).

„Multimodal User Interfaces (MUI) process two or more combined user input modes (such as speech, pen, touch, manual gesture, gaze, and head and body movements) in a coordinated manner with multimedia system output. They are a new class of interfaces that aim to recognize naturally occurring forms of human language and behavior, and which incorporate one or more recognition-based technologies (e.g. speech, pen, vision)." **(Dumas et al., 2009)**

Die Stärke multimodaler Systeme liegt in der Kombination der Stärken der individuellen Modalitäten, in unserem Falle: der Sprach-Ein-und Ausgabe sowie der GUI-Ein- und Ausgabe. Dafür gibt es gängigerweise drei Strategien (vgl. Schnelle-Walke und Döweling, 2011):

- Substitutionsstrategie. Diese Strategie ist nützlich, wenn eine Modalität eine andere Modalität in einer Anwendung komplett ersetzt, die auf unterschiedlichen Endgeräten mit unterschiedlichen Eingabe- und Ausgabemöglichkeiten ausgestattet ist (Beispiel: Im Auto; reine Spracheingabe anstelle von manueller Eingabe).
- Redundanzstrategie. Wenn verschiedene Modalitäten in redundanter Art und Weise genutzt werden und damit letztendlich die gleiche Information zur gleichen Zeit übermitteln (Beispiel: Telefon: Klingelton und Vibrationsalarm)
- Komplementärstrategie. Gilt als bester Weg zur Umsetzung eines multimodalen Systems und ist dann gewährleistet, wenn jeweils die am besten geeignete individuelle Modalität eingesetzt wird, um die aktuelle anstehende Aufgabe zu lösen.

Wie oben bereits angedeutet, liegt die Komplementärstrategie als bester Lösungsweg zur Kombination von GUI und VUI im speziellen Anwendungsfall eines mobilen Fahrzeug-Konfigurators quasi auf der Hand. Nicht jedoch die Antwort auf die Frage: In welchen Fällen ist Spracheingabe tatsächlich besser geeignet als die Eingabe per Touch? Und in welchen Fällen ist Sprachausgabe tatsächlich besser geeignet als die Anzeige über das GUI?

3. Grundlagen für Konzeption und Gestaltung sprachgesteuerter Anwendungen

Unsere Annäherung an diese zentralen Fragen erfolgte unter Berücksichtigung folgender Leitfragen und Standpunkte: Was wissen wir über Sprache? Was sagen die (potenziellen) Nutzer? Was sagen die User Interface-Forscher? Was lehrt uns der Markt?

Die folgenden Abschnitte beleuchten ausgewählte Aspekte dieser Standpunkte.

3.1. Was wissen wir über Sprache?

Diese erste Leitfrage betrachtet die besonderen Spezifika der Sprache als Medium. Was sind eigentlich inhärente Eigenschaften der Sprache? Und was bedeuten diese für die Verwendung innerhalb eines multimodalen Systems und im Zusammenspiel mit einem grafischen User Interface? Schnelle-Walke und Döweling geben hier einige Anhaltspunkte (vgl. Schnelle-Walke und Döweling, 2011):

Zunächst einmal ist Sprache eindimensional. Das heißt das Ohr kann eine Reihe von Aufnahmen nicht so schnell erfassen wie das Auge Text und Bilder scannen kann. Auch ist Sprache flüchtig – und damit nicht das ideale Medium, um große Mengen an Daten und Informationen zu vermitteln. Sprache ist außerdem unsichtbar. Das macht es schwer, dem Nutzer (schnell) zu vermitteln, welche Aktionen durchgeführt und welche Formulierungen verwendet werden müssen, um diese Aktionen durchzuführen. Nicht zuletzt ist Sprache zudem asymmetrisch. Das bedeutet: Menschen sprechen schneller als sie tippen können, aber sie können wesentlich langsamer zuhören als sie lesen können.

Diese inhärenten Eigenschaften der Sprache lassen bereits erste Ableitungen zu, wenn es darum geht, welche Informationen eher über ein grafisches, und welche eher über ein sprachliches User Interface umgesetzt werden sollten.

So lässt sich festhalten, dass große Datenmengen wie Text, Bilder und Videos quasi zwangsweise über GUI dargestellt werden müssen. Um das (langsame) Tippen von längeren Texten zu vermeiden, sollte der Fokus der GUI-Eingabe auf kurzen textlichen Eingaben liegen. Andersherum sollte auf das Vorlesen längerer Textpassagen verzichtet und der Fokus der Sprachausgabe auf kurzen Bestätigungen oder Anweisungen liegen. Implizit ist diesen (wie auch den folgenden) Ableitungen die Grundvoraussetzung, dass dem Nutzer klar

und präzise vermittelt werden muss, welche Spracheingaben er machen kann bzw. darf und welche er nicht machen kann bzw. darf.

Zudem gibt es einige technische Faktoren der Sprachsteuerung, die aber – im Gegensatz zu den inhärenten Faktoren – beeinflusst werden können. Hierunter zählen vor allem die Qualität der Sprachsynthese, die Performance der Spracherkennung sowie der (individuell wählbare) Trade-Off zwischen Flexibilität und Genauigkeit.

In puncto Sprachausgabe ist die Qualität moderner Text-To-Speech-Systeme (TTS) nach wie vor als eher niedrig einzustufen. Im Allgemeinen ziehen es die Nutzer deswegen (noch) vor, zuvor eingesprochene und aufgenommene Sprachsignale zu hören, da sie natürlicher klingen. Bei der Spracheingabe bzw. -erkennung ist zu berücksichtigen, dass Sprache nicht mit einer 100%-igen Genauigkeit erkannt wird, selbst von Menschen nicht. Dennoch werden in der Nutzung höchste Ansprüche an die Performanz der Spracherkennung gestellt. Dies betrifft unter anderem auch die Gewährleistung von Flexibilität: In der gesprochenen Sprache kann das gleiche Anliegen sehr unterschiedlich ausgedrückt werden, allein durch unterschiedliches Vokabular, vor allem aber durch Referenzierung auf aktuelle Kontexte.

Ein (rein) sprachliches Interface muss demzufolge viele dieser Ausprägungen erkennen und unterstützen. Als gutes Beispiel dient die Datumsangabe: Ein flexibles System erkennt auch eine relationale Eingabe à la „Gestern" oder „Mittwoch, der zweite", ist in diesem Punkt jedoch gegebenenfalls anfälliger für Fehler. Dem gegenüber steht der Anspruch auf Genauigkeit, der (bis dato) eher durch die Eingabe einer starren Informationsabfolge erfüllt werden kann: „Bitte geben Sie das Jahr an… Bitte geben Sie den Monat an… Etc." Der optimale Trade-Off zwischen Flexibilität und Genauigkeit bleibt dabei in erster Linie Definitionssache (zumindest im Rahmen der technischen Möglichkeiten).

3.2. Was sagen die (potenziellen) Nutzer?

Verschiedene Nutzerstudien haben ergeben, dass der tatsächliche Gebrauch einer Modalität vor allem von den Faktoren Vertrautheit bzw. Expertise und Effizienz abhängt (vgl. hier und im Folgenden: Wechsung et al., 2009).

Speziell Touch stellte sich dabei wiederholt als die beliebtere Modalität im Vergleich zu Sprache heraus. Dies kann darin begründet sein, dass die Interaktionslogik bei grafischen Oberflächen weitgehend vertraut ist. Im Gegensatz müssen bei Sprachsteuerung noch viele interaktionsrelevante Kenntnisse erworben werden. Vor allem beim Lernen neuer Aufgaben kann man annehmen, dass zumeist gut bekannte Modalitäten eher zum Einsatz kommen (vgl. Seebode et al., 2009).

Auch ergaben experimentelle Untersuchungen, dass die Wahrscheinlichkeit der Nutzung einer Modalität davon abhängt, wie viele Interaktionsschritte zur Erreichung der gewünschten Reaktion durchzuführen sind, letztendlich also: wie effizient die Modalität ist (vgl. Schaffer, 2008). Eingabemodalitäten, welche die Lösung einer Aufgabe mit weniger Interaktionsschritten erlaubten, wurden dementsprechend häufiger genutzt. Dies lässt den Umkehrschluss zu, dass eine wenig vertraute und gegebenenfalls anspruchsvolle(re) Modalität dann erhöhte Nutzungschancen hat, wenn für den Nutzer klar erkennbar ist, dass dadurch ein deutlicher Anteil an Interaktionsschritten eingespart werden kann.

Die Kommunikation dieser Effizienz wird dabei in den meisten Fällen allerdings wieder dem GUI überlassen bleiben: Dessen initiale Aufgabe ist es dann, den Nutzer darauf hinzuweisen, dass durch die Verwendung der Spracheingabe eine schnellere Bedienung (insgesamt oder in bestimmten Teilbereichen der Anwendung) möglich ist.

3.3. Was sagen die User Interface-Forscher?

Eng verwoben mit den Anforderungen der (potenziellen) Nutzer, lassen sich auch aus Perspektive der User Interface-Forschung einige Richtlinien für multimodale Anwendungen formulieren (vgl. hier und im Folgenden: Larson und Oviatt, 2004; Schaffer, Schleicher und Möller, 2011).

Zunächst einmal gilt es, multimodale Systeme für eine möglichst große Bandbreite an Nutzern und Nutzungskontexten zu schaffen. UI-Designern ist also anzuraten, sich intensiv mit den psychologischen Charakteristika (kognitive Fähigkeiten, Motivation, etc.), dem Erfahrungsgrad der Nutzer sowie Fach- und Aufgaben-spezifischen Charakteristika befassen. Die empirische Identifikation von Personas und Usecases kann hier einen wichtigen Beitrag liefern.

Gerade letzteres impliziert auch sich verändernde Umgebungen (z.B. Nutzung zuhause, im Büro, während der Fahrt, an einem öffentlichen Ort wie Haltestelle oder Warteraum, etc.) und damit die Notwendigkeit zur Auswahl der dortig jeweils besten Kombination von Modalitäten. In diesem Zusammenhang werden auch Aspekte der Privatsphäre sowie Sicherheitsbelange relevant. In Situationen, in denen Nutzer ihre Privatsphäre schützen wollen, sollte ein sprachfreier Modus angeboten werden Dies gilt auch und vor allem für die Eingabe privater Daten (Passwörter, Adressen, etc.).

Die zentrale (und größte) Herausforderung bei der Gestaltung multimodaler User Interfaces stellt jedoch sicherlich die Optimierung der Interaktion auf die kognitiven und physischen Fähigkeiten der Nutzer hin. Für UI-Designer gilt es verlässlich herauszufinden, wie sie intuitive und effiziente Interaktionen schaffen können, welche auf primären menschlichen Wahrnehmungs- und Verarbeitungsfähigkeiten basieren (Aufmerksamkeit, Kurzzeitgedächtnis, Entscheidungsfindung) – und damit die kognitive Last des Nutzers im Umgang mit dem

System bzw. Angebot möglichst gering halten (zum „cognitive load"-Konzept vgl. Wickens, 2002).

In diesem Zusammenhang lassen sich (unter vielen anderen) zum Beispiel folgende Empfehlungen formulieren:

- Die visuelle Darstellung sollte mit manuellem Input gekoppelt sein, insbesondere für räumliche Informationen und parallele Verarbeitung; Sprachausgabe sollte an die Spracheingabe gekoppelt sein, insbesondere für Statusinformationen, serielle Verarbeitung, (Warn-) Hinweise oder Kommandoeingaben.
- Eine Dopplung von Sprach- bzw. Tonausgabe mit der visuellen Präsentation ist zu vermeiden, es sei denn, es müssen besonders wichtige Informationen übermittelt werden (Warnhinweise oder Systemmeldungen wie „Ich habe Sie nicht verstanden", „Bitte sprechen").
- Die Exploration von Inhalten sollte dem GUI und der Touch-Bedienung vorbehalten bleiben. Dies betrifft zum Beispiel folgende Interaktionsmöglichkeiten: Auswahl aus (längeren) Listen, Navigation (Wischen rechts/ links), Navigation in sichtbaren Elementen, Scrollen (hoch / runter), Vergrößern/ verkleinern, etc.
- Nutzung der Spracheingabe zur Abfrage des Systemstatus („Hilfe") bzw. auch zur Steuerung von Dingen in der „Peripherie" außerhalb des gerade sichtbaren GUIs/ „Area of Interest" („Zum Motor", „Wie hoch ist mein Preis?")
- Nutzung von Sprachdialogen für kurze Frage- / Antwort-Dialoge (System ergreift Initiative, z.B. „Meinten Sie schwarz?" „Ja."
- In der Sprachausgabe nur dann natürliche Sprache verwenden, wenn auch der Nutzer natürliche Sprache zur Steuerung des Systems verwenden kann. Falls die Spracheingabe nur einfache Befehle unterstützt, sollte auch die Sprachausgabe nur mit kurzen, klaren Hinweisen (Maschinensprache) erfolgen.
- Kombination von Sprache und GUI zur Behebung von Fehleingaben. Der Nutzer sollte die Modalität selbst auswählen können, um für die jeweiligen Inhalte/Aufgaben eine weniger fehleranfällige Modalität nutzen zu können. Falls ein Fehler passiert, sollte es Nutzern erlaubt sein, zu einer anderen Modalität zu wechseln.

3.4. Was lehrt uns der Markt?

Spätestens auf der Suche nach Best Practices lehrt uns der Markt, dass es kaum sprachgesteuerte bzw. multimodale Angebote gibt, die ähnlich komplexe Usecases abbilden wie es unser konkreter Anwendungsfall der mobilen Fahrzeug-Konfiguration tat bzw. tut.[1]

Die im Gros etablierten sprachgesteuerten Systeme lassen sich grob wie folgt typisieren und zugleich chronologisch ordnen:

- Automatische Auskunftssysteme (seit 1980): Reine Voice-User-Interfaces zum Beschwerde-Management (Self-service). Starke Verbreitung liegt in erster Linie an möglichen Einsparungen im Call Center.
- Diktieren, Text-To-Speech (seit 1990er Jahren): Multimodale Systeme, die meist im professionellen Umfeld genutzt werden (Journalisten, Mediziner, etc.). Mehrwert: Spracherkennung (inzwischen) teilweise fünfmal schneller als Tippen.
- Sprachsteuerung im Automobil (seit 2010): Multimodale Systeme, die verschiedenste Aufgaben übernehmen. Meistgenutzte Funktionen: Telefongespräch starten, annehmen, beenden, Zielführung, POI-Selektion, Vorlesen von Nachrichten. Mehrwert: Aufgaben für den Fahrer „mit den Händen am Lenkrad" ansonsten nicht bzw. nur schwer zu erfüllen.
- Smart TVS (seit 2011): Multimodale Interfaces, die von der Verbreitung der Smart TVs profitieren. Mehrwerte bislang für die breite Masse kaum erkennbar.
- „Mobile Helfer" (seit 2011): Multimodale Interfaces, die von starker Verbreitung der Smartphones sowie protegierter Technologien (Siri, Google Voice Search) profitieren. Meist genutzte Funktionen: Telefonanrufe tätigen, Textnachrichten eingeben, Voice-Based Search. Mehrwerte: Ermöglicht Sprachsteuerung im mobilen Umfeld, erlaubt effiziente Nutzung für regelmäßige, alltägliche und überschaubare Aufgaben, insbesondere dann, wenn Nutzer Hände und Augen nur teilweise frei hat.

Diese „mobilen Helfer" können als aktueller Benchmark gelten, an dem sich ein mobiler sprachgesteuerter Fahrzeug-Konfigurator messen lassen muss. Ihr hoher Nutzwert liegt vor allem im Charakteristikum der alltäglichen Handlungen, welche mit kurzen, überschaubaren Befehlen angesteuert werden können (Telefonanruf, Aufruf einer App, etc.) und schnell zu erlernen sind.

Hier zeigt sich bereits die erste Hürde zum „sinnvollen Usecase" der Sprachsteuerung.

4. Fahrzeug-Konfiguration als sinnvoller Usecase für Sprachsteuerung?

Bei der Fahrzeug-Konfiguration handelt es sich eben nicht um eine alltägliche Handlung, welche vom Nutzer regelmäßig durchgeführt. Allein dieser Umstand impliziert einige zentrale Nutzungsbarrieren:

Es muss davon ausgegangen werden, dass bei den (potenziellen) Nutzern kaum Vorwissen aus der realen Welt über den Prozess-Ablauf der Konfiguration existiert, schon gar nicht im nötigen Detailgrad. Ein Lerneffekt durch häufige Wiederholung kann dadurch nicht einsetzen.

Auch besteht – anders als im Bereich der „mobilen" Helfer – in den seltensten Fällen bereits zu Anfang der Konfiguration eine klare Zielvorstellung (wie zum Beispiel für das Tätigen eines Anrufs: „Ich möchte

Max Mustermann anrufen"), welche mit Hilfe von Sprachsteuerung effizient bedient erreicht werden kann. Bei der Fahrzeugkonfiguration ist zwar klar, dass ein Auto, vermutlich auch das konkrete Modell, konfiguriert werden soll, Detailausprägungen und einzelne Bestandteile der Konfiguration sind zu Beginn jedoch (im Normalfall) nicht klar.

Im Gegenteil, die Konfiguration lebt gerade von der (visuellen!) Exploration des Fahrzeugs. Nicht zuletzt deswegen ist auch ein Nutzungskontext „ohne Augen" (ebenfalls ein Erfolgskriterium der „mobilen Helfer") sehr unwahrscheinlich. In diesem Zusammenhang ist eher davon auszugehen, dass das GUI stets die bedeutendere Rolle spielen wird. Unter anderem aus oben genannten Gründen kann Sprachsteuerung hier allerdings nur bedingt unterstützen bzw. die Konfiguration tatsächlich effizienter (zum Beispiel als Touch) machen.

Die meistgenutzten Funktionen der Sprachassistenten profitieren vor allem vom schnelleren Verfassen von Texten durch Spracheingabe (E-Mail verfassen, Diktieren). Die Herausforderungen für die Umsetzung eines Fahrzeug-Konfigurators liegen nicht so sehr in Spracherkennung und Darstellung des eingegeben Textes auf dem Display, sondern im Sprachverständnis. Je länger der Sprachbefehl des Nutzers und je natürlicher die Formulierung, desto geringer ist die Wahrscheinlichkeit, dass das System das Kommando korrekt erkennt.

Unser Fazit? Die Fahrzeug-Konfiguration ist kein (sinnvoller) Anwendungsfall für Sprachsteuerung. Insbesondere, da sie aufgrund ihrer impliziten und implikativen Eigenschaften Konfigurationsprozess und -erlebnis eher behindert als fördert.

Dabei lässt sich das, was hier am Beispiel der Fahrzeug-Konfiguration dekliniert wurde, mühelos auf andere sprachgesteuerte Angebote übertragen und als beschränkende Bedingungen für die Usecase-Definition von Sprachsteuerung formulieren. Demnach schafft Sprache keinen Mehrwert (zum Beispiel gegenüber Touch):

- Für die Exploration von Inhalten und Elementen (z.B. in längeren Listen), die nicht a priori bekannt bzw. durch den Nutzer anzunehmen sind.
- Für die Exploration von (stark) visuellen Inhalten.
- Wenn sie für einmalige bzw. seltene Handlungen eingesetzt wird, für die kaum Vorwissen aus der realen Welt vorhanden ist (Nutzer haben keine Vorstellung über die vom System erwarteten Befehle und Eingabemöglichkeiten; Lerneffekt durch häufige Wiederholung kann nicht einsetzen).
- Wenn ein Nutzungsszenario einem festen Fahrplan folgt, der Schritt für Schritt durchgegangen werden muss – sondern erst dann, wenn durch Sprache Schritte übersprungen werden können.

Unserem Auftraggeber konnten wir die Weiterverfolgung des (damaligen) Konzeptansatzes nicht empfehlen. Allerdings leistete unsere Arbeit einen grundlegenden und wichtigen Beitrag dafür, weitere Ideen und Ansätze für sprachgesteuerte Angebote frühzeitig (das heißt vor allem: ohne „unnötige" Konzeptionsaufwände) zu bewerten sowie neue Anwendungsfälle, die tatsächlich einen sinnvollen Usecase für Sprachsteuerung darstellen, zu definieren.

Literatur

1. Abbott, K.R. (2002): Voice Enabling Web Applications: VoiceXML and Beyond. a!press, 2. Auflage.
2. Chandler, P. und Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. In: Cognition and Instruction, 8 (4), S. 293–332.
3. Dumas, B., Lalanne D. und Oviatt, Sh. (2009): Multimodal Interfaces: A Survey of Principles, Models and Frameworks. In: Human Machine Interaction, Heidelberg: Springer-Verlag Berlin, S. 3–26.
4. Larson, J.A. und Oviatt, S. (2004): Guidelines for Multimodal User Interface Design. In: Communications Of The ACM, Vol. 47, No.1, S.57–59.
5. Schaffer, S. (2008): Integration eines Spracherkenners in ein Rauminformationssystem. In: Quality.
6. Schaffer, S., Schleicher, R. und Möller, S. (2011): Simulation von Benutzerverhalten im Umgang mit multimodalen Diensten. 9. Berliner Werkstatt Mensch-Maschine-Systeme. VDI Verlag, S. 110–111.
7. Seebode, J., Schaffer, S., Wechsung, I., und Metze, F. (2009): Influence of User Characteristics on the Usage of Gesture and Speech in a Smart Office Environment. In: Proceedings of the 8th International Gesture Workshop 2009, Bielefeld.
8. Schnelle-Walka, D. und Döweling, S. (2011): Speech Augmented Multitouch Interaction Patterns. Darmstadt University of Technology.
9. Wechsung, I., Engelbrecht, K.-P., Schaffer, S., Seebode, J., Metze, F. und Möller, S. (2009): Usability-Evaluation multimodaler Schnittstellen: Ist das Ganze die Summe seiner Teile? In: Mensch & Computer 2009: Grenzenlos frei!?, München: Oldenbourg Verlag, S. 495–498.
10. Wickens C. D. (2002): Multiple resources and performance prediction. In: Theoretical Issues in Ergonomics Science, 3 (2), S. 159–177.

[1] Anmerkung: Seit Juli 2013 bietet AutoScout24 in der Android-Version der AS24-App eine sprachgesteuerte Fahrzeugsuche an und nähert sich damit dem Usecase „Fahrzeug-Konfiguration" zumindest thematisch an. Diese war zum Projektzeitraum noch nicht verfügbar. Auch liegen aktuell noch keine Nutzungsdaten vor. Nach erster Evaluation beschränkt sich die App jedoch auf die initiale Suche bzw. Selektion (Ersteingabe relevantes Modell, Farbe, Motorisierung, etc.). Systemrückmeldungen, Bestätigungen und Modifikationen der Suche werden weiterhin über das GUI abgebildet.