

SPRACHBEDIENUNG IM AUTOMOBIL

Gerd Gruchalski

Audi AG

gerd.gruchalski@audi.de

Abstract: In diesem Beitrag wird die Sprachbedienung im Automobil beleuchtet, ein kurzer historischer Abriss gegeben, als Vergleich kurz auf die Sprachbedienungen in modernen Smartphones eingegangen und die speziellen Anforderungen und Randbedingungen im Automobil erläutert. Die Integration in das Fahrzeug-HMI wird am Beispiel des Audi MIB dargestellt.

1 Geschichte der Sprachbedienungen im Automobil

Sprachbedienungen im Automobil gibt es seit beinahe 20 Jahren seit Mobiltelefone eine immer größere Verbreitung insbesondere bei Geschäftsfahrzeugnutzern erfuhren. Die Autohersteller haben sehr früh das Risiko der Ablenkung bei manueller Bedienung eines Telefons während der Fahrt gesehen und erkannt, dass die Sprachbedienung eine Alternative darstellt, und dies dann über ihre Systeme ausgerollt. Der typische Funktionsumfang wird dabei durch technische Limitierungen begrenzt.

Freisprechanlage (bei Audi „universelle Handyvorbereitung UHV“)

- eine Nummer sprechen, korrigieren und wählen
- einen Namen mit Nummer speichern
- die Nummer eines gespeicherten Namens wählen
- PIN eingeben

Infotainmentsysteme bei Audi:

1. Generation: Radio-Navigations-Systeme RNS

- Command & Control für Radio, CD und Navigation

2. Generation: Multimedia Interface MMI an 2001

- Synchronisation mit dem Telefonbuch eines Mobiltelefons
- One-Shot-Eingabe für Anrufe („anrufen bei Peter Müller“)
- Navigations-Zieleingabe durch Buchstabieren

3. Generation: MMI 3G ab 2004, update MMI 3G+ ab 2008

- Ganzwort-Zieleingabe („Ingolstadt“ – „Ettinger Straße“ – „20“)
- Medienbedienung („Interpret Queen“ oder „Titel Yellow Submarine“)
- erste Online-Dienste per Sprache („Onlineziel Pizza in Nürnberg“)

4. Generation: Modularer Infotainment-Baukasten MIB ab 2012

- Multislot-Zieleingabe („Ingolstadt Ettinger Straße 20“)
- erweiterte Online-Dienste (Diktieren von Nachrichten)

5. Generation: MIB Evo ab 2014

- One-Shot-Eingabe für Navigation („fahre mich nach Ingolstadt in die Ettinger Straße 20“)
- One-Shot-Eingabe für Medien („spiele Bohemian Rhapsody von Queen“)
- Sprachbedienbarkeit aller Online-Dienste (z.B. „Wetter“)
- Natürlichsprachlichkeit

2 Sprachbedienung im Mobiltelefon

Die Sprachbedienungen in Smartphones, namentlich „Siri“, haben das Thema Spracherkennung in die Werbung und damit in den Blick der Öffentlichkeit gebracht.

- "Siri" hat einen Namen (und seit „Big Bang Theory“ auch ein Gesicht)
- sie gibt humorvolle Antworten auf bestimmte Fragen
- wenn sie keine Antwort weiß, will sie "im Netz danach suchen"

Das wirkt intelligent und weckt Erwartungen. Bei weiterer Nutzung erkennt der Kunde, dass „Siri“ derzeit eher beschränkte Fähigkeiten und Einsatzbereiche hat:

- man hält das Mobiltelefon meist in der Hand und kann es dann auch leicht manuell bedienen
- die Nutzung einer Sprachbedienung in Gesellschaft oder in einem Büro ist störend
- die Ablenkung schon beim zu Fuß Gehen ist groß, im Fahrzeug ist sie größer

Apple hat das erkannt und bietet für Siri den „eyes-free-mode“ an.

Durch die Entwicklung von Smart-Watches und anderen „Wearables“ eröffnen sich für die Sprachbedienung weitere Einsatzmöglichkeiten.

3 Sprachbedienung im Automobil

3.1 Primärfunktion Fahren

Die Primärfunktion eines Automobils ist das Fahren. Jeder Fehler des Fahrers kann zu einem Unfall führen. Daher liegt beim Auto und dessen Bedienung ein starker Fokus auf Sicherheit und der Reduktion der Ablenkung auf ein Mindestmaß.

Anforderungen:

- die Hände sollen möglichst immer am Lenkrad sein
- der Aufmerksamkeit soll auf die Fahrbahn gerichtet sein
- die Bedienung soll möglichst einfach und wenig ablenkend sein

Gleichzeitig wächst die Komplexität der Fahrzeugausstattung aber auch von Mobiltelefonen immer weiter. Die Sprachbedienung hat das Potential, die Ablenkung zu senken.

Ziel der Sprachbedienung im Kraftfahrzeug ist daher

- eine komfortablere, schnellere Bedienung komplexer Fahrzeugfunktionen
- die Unterstützung des Nutzers bei fahrrelevanten Funktionen
- die Verringerung der Anzahl der Bedienschritte

- die primäre Bedienung erfolgt über das Multifunktionslenkrad (eine multimodale Bedienung durch den Nutzer ist parallel möglich)
- die Verringerung der Ablenkung

3.2 Darstellung des Funktionsumfangs des Sprachdialogs

Jedes aktuelle Kraftfahrzeug ist ein hochkomplexes Gerät. Für die primäre Fahraufgabe haben sich weltweite Standards herausgebildet: Lenkung, Gas, Bremse, Kupplung usw. Aber schon die Sekundärbedienung variiert von Fahrzeug zu Fahrzeug: Licht, Zündung, Scheibenwischer, Handbremse usw.

Bei Infotainmentsystemen ist die Variantenvielfalt größer und jedes HMI erklärungsbedürftig. Bedienungsanleitungen sind mittlerweile dicke Bücher, die nur wenige Nutzer wirklich lesen und sich später daran erinnern. Die Kernfunktionen muss der Kunde daher ohne explizite Erklärung finden und verstehen können.

Das haptische HMI kann den Nutzer durch Menüs führen. Bei der Sprachbedienung muss er die Kommandos auswendig wissen oder sie während der Nutzung erlernen.

Wie kann man ihm die Funktionsumfänge und deren Bedienung darstellen?

3.2.1 Hilfe-System

Ein entsprechendes Hilfe-System kann den Nutzer unterstützen. Dabei muss es den Spagat schaffen, Neukunden zu unterstützen, ohne erfahrene Nutzer zu stören.

- ausführliche Prompts, Rückfragen und Bestätigungen
 - sind gut und hilfreich für Neukunden
 - verlangsamen die Sprachbedienung aber für den erfahrenen Nutzer
 - lassen den Sprachdialog auf Dauer langatmig und unkomfortabel wirken
- Kommandoanzeigen
 - überdecken die eigentliche Anzeige
 - können nur einen kleinen Teilbereich der Sprachfunktionen anzeigen
 - müssen gelesen werden und lenken daher stärker ab
- Unterstützung bei Fehlerkennungen
 - es ist für eine Sprachbedienung bereits sehr schwierig, eine Fehlerkennung hinreichend sicher festzustellen
 - ein Fehler hierbei ist für den Nutzer doppelt irritierend
- Korrekturmöglichkeiten
 - sinnvoller ist es daher, unabhängig von der Fehlerursache möglichst einfache Korrekturmöglichkeiten anzubieten
- explizit angeforderte Hilfe oder Tutorials
 - sehr wenige Nutzer sagen überhaupt „Hilfe“ oder ähnliches, wenn sie nicht wissen, wie ein bestimmtes Kommando lautet
 - auch diese Kommandos müssen letztlich auswendig gelernt werden

3.2.2 Dialogführung

Es gibt viele verschiedene Ansätze der Dialogführung. Die beiden häufigsten im Kraftfahrzeug sind:

- „speak what you see“: Jedes sichtbare Menü ist gleichzeitig auch ein Kommando und ein Kommando entspricht einem Tastendruck. Dies ist der einfachste Ansatz, insbesondere bei Touch-HMIs, nutzt aber kaum die Möglichkeiten einer Sprachbedienung aus und bringt nur geringen Nutzen.
- Parallele Modellierung von Sprachbedienung und haptischem HMI: Der Sprachdialog folgt den grundlegenden Abläufen und den Wortwahlen des HMI und nutzt auch dessen Screens, bildet aber nur für den Sprachdialog notwendige Schritte nach. Dies erlaubt eine flüssigere Bedienung. Der Nutzer kann das gemeinsame mentale Modell erkennen und aus der HMI-Darstellung auf die möglichen Kommandos schließen.

Globale Kommandos und Shortcuts können dabei die Grenzen zwischen verschiedenen Domänen des HMI auflösen und die Restriktionen des Dialogmodells verwischen.

Entsprechendes gilt für die Dialogführung innerhalb jedes Einzeldialogs. Es müssen von jedem Sprachdialog

- unvollständige Eingabe durch den Nutzer behandelt werden
- Mehrdeutigkeiten aufgelöst werden
 - per Listenauswahl durch den Nutzer
 - per Priorisierung durch das System aufgrund von Kontext und Wahrscheinlichkeiten
- bei Fehlern eine Korrektur möglich sein
 - Fehlerkennung durch das System
 - Falscheingabe durch den Nutzer
 - nicht eindeutige Datenbasis
 - andere Fehler
- Nicht-Erkennung und Out-of-Domain-Eingaben behandelt werden
 - Siris „...oder soll ich im Netz danach suchen?“ ist für das Fahrzeug untauglich
- Bestätigungen der Eingabe und der Aktion erfolgen

Dabei beschleunigt und vereinfacht jede nicht gestellte Rückfrage den Dialog.

3.3 Ziel: Intuitive Sprachbedienung

„Intuitiv“ ist hierbei ein Buzzword, ein Ideal, das jeder Designer für sich in Anspruch nimmt, aber keiner für alle vorhandenen Nutzergruppen erfüllt oder auch nur erfüllen kann.

Wie erreicht man eine möglichst weit gehende Annäherung an dieses Ideal?

Man zerschlägt den gordischen Knoten der Systemzwänge und erlaubt dem Nutzer einen möglichst natürlichsprachlichen Dialog weitgehend unabhängig vom haptischen HMI. Nur sein Start- und Endpunkt muss dort verankert sein. Dies erlaubt die größten Freiheiten und die flexibelste Sprachbedienung, verstärkt aber das Problem, wie die möglichen Kommandos dem Nutzer erklärt werden können.

Wesentliche Elemente sind:

- Spracheingabe
 - statistische Sprachmodelle
 - semantische Interpretation des gesprochenen
 - mindestens die Intention erkennen und den zugehörigen Dialog starten
- Sprachausgabe
 - verschiedene Formulierungen, keine starren Sätze
 - der Sprachdialog spricht als "ich" vs. bisherige Passiv-Konstrukte
 - die Eingaben des Nutzers werden in die Antwort aufgenommen
- Sprachdialogablauf
 - "Wie bitte?" vermeiden, sofern zumindest die Intention erkannt wurde
 - der Folgedialog kann durch entsprechende Rückfragen stärker geführt werden

Der subjektive Eindruck der Natürlichsprachlichkeit beim Nutzer weckt dabei allerdings Erwartungen, die der Sprachdialog erfüllen können muss.

Auch ein Sprachdialog hat ein „uncanny Valley“.

3.4 Serverbasierte Erkennung

Die Sprachdialoge in Smartphones basieren weitestgehend auf einer serverbasierten Erkennung und benötigen daher eine permanente Datenverbindung zu ihrem Server. Daraus ergeben sich grundsätzlich Probleme mit

- Netzabdeckung
- Verbindungsqualität
- Datenrate
- Roaming
- Verfügbarkeit des Servers

Die Ansprüche des Kunden an ein Kraftfahrzeug sind weitaus höher als die an ein Mobiltelefon. Die Kernfunktionen müssen auch in einer Garage, in einem Parkhaus, in einem Funkloch, im Ausland usw. erfüllt werden.

Die Lebensdauer eines Fahrzeugs ist um eine Größenordnung höher als die eines Mobiltelefons.

Serverbasierte Systeme können daher derzeit nur ergänzend integriert werden.

4 Funktionsumfänge eines Sprachdialogs am Beispiel Audi MIB

4.1 Domänen

Die Sprachbedienung im modularen Infotainmentbaukasten MIB bei Audi entspricht im Wesentlichen dem oben dargestellten. Sie enthält ein statistisches Sprachmodell für ca. 130 Topics, durch das eine natürlichsprachliche Eingabe in weitgehend freier Form möglich ist, sowie unterstützend klassische Grammatiken.

Die abgedeckten Domänen entsprechen dabei den Haupt-Use-Cases im Fahrzeug.

Telefon:

- anrufen bei einer Telefonnummer
- anrufen bei einem Kontakt
 - der Kontakt wird vom Mobiltelefon synchronisiert
 - jeder Kontakt kann mehrere Telefonnummern beinhalten
- gespeicherte Favoriten anrufen

Beispiel: „Ich würde gerne mit Peter Müller im Büro telefonieren“

Navigation:

- Zieleingabe als One-Shot oder schrittweise
 - zu einer Adresse
 - zu einem Sonderziel
 - zu einem Kontakt aus dem Mobiltelefon
 - zu einem gespeicherten Favoriten
- Steuerung von Routenoptionen, Kartenfunktionen usw.

Beispiele: „Fahre mich nach Ingolstadt in die Ettinger Straße 20“, „...zu einem italienischen Restaurant“, „...zu einer Aral Tankstelle“, „...zu Peter Müller“, usw.

Radio:

- Sender per Namen einstellen (RDS, DAB, SiriusXM)
- Frequenzen einstellen (FM, AM)
- Favoriten auswählen

Beispiele: „Stelle bitte Bayern 3 ein“ oder auch einfach nur „Bayern 3“

Medien:

- Jukebox, USB, SD-Karte, iPod, CD, DVD ...
- Album, Interpret, Titel per Namen
- Favoriten

Beispiele: „den iPod bitte“ oder „spiele Bohemian Rhapsody“

Online:

- Zieleingabe Online (via Google)
- Diktieren von Nachrichten (SMS, Mails, ...)
- Sprachbedienung der Audi Connect-Apps wie Wetter, Tankstellensuche usw.

Beispiele: „Onlineziel Pizzeria in Nürnberg“ oder für die Wetter-App z.B. „Wetter“

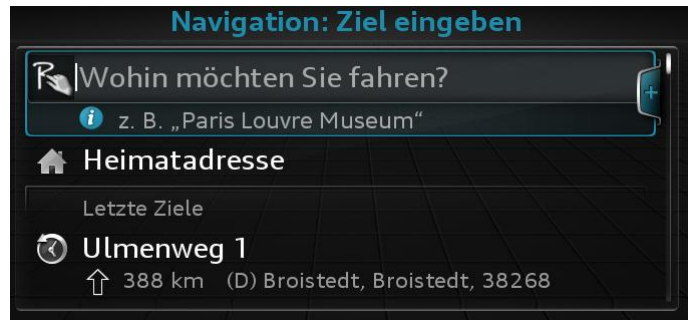


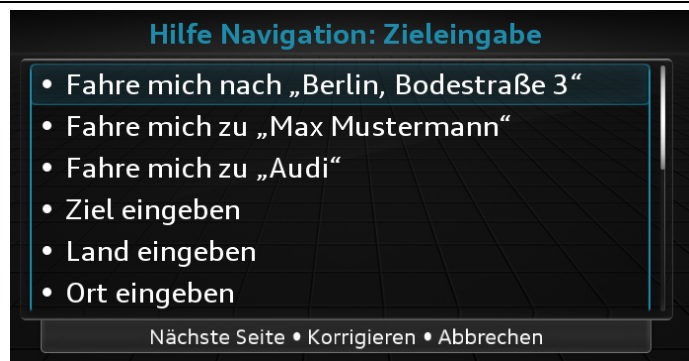

Hilfe:

- mehrstufiges Hilfe-System

Beispiel: „Was kann ich hier sagen?“

sowie ca. 130 weitere Topics in allen Domänen.

4.2 Beispiele-Screens des Audi MIB

	<p>Basisscreen der haptischen Zieleingabe</p> <ol style="list-style-type: none"> 1. Zeile: Freie Zieleingabe per Touchpad 2. Zeile: gespeicherte Heimatadresse 3. Zeile und alle weiteren Seiten: Letzte Ziele in zeitlicher Sortierung <p>Mit dem rechten Softkey [+] können kontextspezifische Menüs geöffnet werden.</p>
	<p>Basisscreen der Zieleingabe nach Kommando „Adresse eingeben“</p> <p>In der Fußzeile ist eine kleine Hilfeanzeige mit den für diesen Screen erwarteten Eingaben.</p>
	<p>Auswahlliste nach „fahre mich nach Ingolstadt Ettinger Straße 20“</p> <p>In Zeile 1 das Ergebnis mit der höchsten Konfidenz, in den folgenden Zeilen auch auf den nächsten Seiten mögliche Alternativen.</p> <p>Multimodale Bedienung ist möglich.</p>
	<p>Anzeige nach „Hilfe Navigation“</p> <p>Die Kommandos zur Zieleingabe werden vorgelesen und parallel angezeigt.</p> <p>In der Fußzeile werden die für diesen Hilfe-Screen möglichen Kommandos angezeigt.</p>
	<p>„Hilfe Sprachbedienung“</p> <p>Hinter den jeweiligen Überschriften folgen weitere Screens für die Teilbereiche.</p>

	<p>Startscreen von Audi Connect d.h. der Audi Onlinedienste („Apps“)</p>
	<p>„Hilfe Audi Connect“</p>

5 Quo Vadis?

Audi ist der festen Überzeugung, dass die Sprachbedienung die zentrale Instanz ist, um die Bedienung von Fahrzeugsystemen zu erleichtern und die Ablenkung zu reduzieren, und wird den Weg der Natürlichsprachlichkeit konsequent fortsetzen, um dem Ideal einer freien, intuitiven Bedienung näher zu kommen.