Dialog Design für sprachliche und multimodale Mensch-Maschine-Interaktionen im Automobil

Dipl.-Psych. T. Noszko[1], Prof. Dr. A. Zimmer
Institut für Angewandte Psychologie
Universität Regensburg
Universitätsstr. 31, 93053 Regensburg
t.noszko@empirience.de

Zusammenfassung

Im Rahmen einer Simulationsstudie wurde eine sprachgesteuerte Anwendung zur Nutzung eines Autotelefons und eines elektronischen Adressbuchs untersucht, wobei Effekte der Dialog Gestaltung (hier: zeitliches Systemverhalten) auf das Nutzerverhalten und Nutzererleben im Mittelpunkt standen. Anschließend wurde im Rahmen eines Feldversuchs ein multimodales Bedien-/Anzeigekonzept zur Nutzung der Funktionen eines Autotelefons und Navigationssystems untersucht. Beobachtet wurden das Nutzungsverhalten der Probanden sowie ihr Blickverhalten in unterschiedlichen Verkehrssituationen. In einem Interview wurden zusätzlich subjektive Angaben erhoben.

Die Befunde zeigen, dass Sprachsteuerung in Kombination mit akustischer und graphischer Ausgabe für den Einsatz im Automobil „reif“ ist. Das Blickverhalten der Fahrer war unauffällig und an die jeweilige Verkehrssituation angepasst. Die Probanden empfanden die Sprachsteuerung als weniger beanspruchend und deutlich komfortabler als die Eingabe über ein manuelles Bedienelement. Wesentlich hierfür ist jedoch ein ausgereiftes Dialog Design, das insbesondere die folgenden Anforderungen erfüllt: kooperatives Systemverhalten, kein „Hetzen“ des Nutzers und ein durch den Nutzer kontrollierter Dialogverlauf.

Einleitung

Automobile werden zunehmend mit Navigationssystemen und Telematikgeräten ausgestattet. Die Bedienung dieser Geräte - insbesondere im fahrenden Automobil - stellt sehr spezifische Anforderungen an die verwendeten MMI-Technologien. Die Kombination von sprachlicher Eingabe und akustischer Ausgabe könnte eine sichere und komfortable Lösung darstellen, da die Augen auf der Straße und die Hände am Lenkrad verbleiben können.

Um den in die Sprachsteuerung gesetzten Erwartungen gerecht zu werden, ist es notwendig, neben der visuellen und motorischen auch die mentale Beanspruchung im Zusammenhang mit dieser Technologie zu berücksichtigen. Ein ausgereiftes Dialog Design ist notwendig, um die mentale Beanspruchung des Nutzers (und Fahrers) gering zu halten. Ein Mensch-Maschine Dialog, der den Nutzer überfordert und/oder verärgert, wird weder den vermuteten Komfort noch die erhofften Sicherheitsvorteile bieten.
Dialog Design umfasst neben den inhaltlichen Aspekten, wie z.B. die Formulierung der Sprachausgaben, auch dynamische Aspekte. Diese haben wesentlichen Einfluss

[1] Zum Zeitpunkt der hier berichteten Untersuchung noch am Institut von Prof. Dr. A. Zimmer tätig, mittlerweile bei *empirience – Ergonomische Forschung & Beratung* (siehe www.empirience.de)

auf die Flüssigkeit des Mensch-Maschine-Dialogs. Zu den dynamischen Aspekten zählt insbesondere das zeitliche Systemverhalten, welches im wesentlichen durch zwei Parameter bestimmt wird: zum einen die Zeitspanne zwischen Nutzereingabe und Systemrückmeldung (*feedback delay*), zum anderen die Zeitspanne, die das System auf eine Nutzereingabe „wartet“ (*time out*), bevor es eine Eingabeaufforderung (*prompt*) ausgibt. Das Dialog-Beispiel in Tabelle 1 soll verdeutlichen, dass die *prompts* eine wesentliche Rolle bei der Realisierung eines kooperativen Systemverhaltens spielen. Dabei wird davon ausgegangen, dass der Nutzer schweigt, wenn er sich über den nächsten Schritt im Dialogverlauf nicht sicher ist.

Mensch (Spracheingabe)	**Maschine (Sprachausgabe)**	**Anmerkung**
[startet per Taste]	[akustisches Startsignal]	*feedback*
Telefon	Telefon	*feedback*
Nummer wählen	Nummer wählen	*feedback*
[schweigt]		*time out*
	Bitte diktieren Sie die Nummer	***prompt***
0 9 4 1	0 9 4 1	*feedback*
[schweigt]		*time out*
	Diktieren Sie weitere Ziffern oder sagen Sie „Hilfe“	***prompt***
9 4 3	9 4 3	*feedback*
[schweigt]		*time out*
	Diese Nummer wählen?	***prompt***
Nein	Diktieren Sie weitere Ziffern oder sagen Sie „Hilfe“	*feedback*
3 8 4 3	3 8 4 3	*feedback*
Wählen	Die Nummer wird gewählt	*feedback*

Tabelle 1: Beispiel für den Einsatz von prompts in einem sprachlichen Mensch-Maschine Dialog.

Zielstellung der durchgeführten Untersuchungen war unter anderem die Bestimmung der idealen Länge des Parameters *time out*. Bei einem langen *time out* erfolgt der *prompt* möglicherweise zu spät, das heißt der Nutzer empfindet das System als „unkooperativ“. Bei einem kurzen *time out* könnte der *prompt* jedoch zu früh erfolgen, so dass sich der Nutzer möglicherweise bedrängt und gehetzt fühlt.

Untersuchung 1 (Simulatorstudie)

In dieser Untersuchung sollten die folgenden Fragen geklärt werden:

- Kann eine sprachgesteuerte Anwendung bei gleichzeitiger Fahrzeugführung effizient genutzt werden?
- Welchen Einfluss hat die Länge des *time out* auf das Nutzerverhalten und -erleben?

Für diese Untersuchung wurde eine rein sprachgesteuerte Anwendung für die Benutzung eines Autotelefons (Eingabe einer Telefonnummer, Auswahl aus einer Namenliste, Wahlwiederholung) verwendet. Die Systemausgaben dieser Anwendung wurden überwiegend akustisch realisiert. Das grafische Display bestand lediglich aus einer Anzeige für den Zustand der Spracherkennung (aktiv/inaktiv) sowie einer Zeile für die zuletzt erfassten Eingaben des Nutzers (siehe Abbildung 1).

Abbildung 1: Der sichtbare Anteil der sprachgesteuerten Anwendung (Untersuchung 1).

Zusätzlich kam ein auf Video-Projektion basierender Fahrsimulator der Universität Regensburg zum Einsatz. Die Versuchsstrecken umfassten sowohl Stadtverkehr als auch Überlandfahrten.

Unabhängige Variablen waren zum einen die Länge des *time out* (800 ms und 3000 ms), zum anderen die Zusatzaufgabe (mit und ohne Fahrzeugführung).

Die Stichprobe umfasste 16 Personen (12 Frauen und 4 Männer). Diese wurden auf 4 Gruppen verteilt (siehe Tabelle 2).

	Länge des *time out*	
Zusatzaufgabe	**800 ms**	**3000 ms**
mit Fahrzeugführung (Fahrsimulator)	n = 4	n = 4
ohne Fahrzeugführung	n = 4	n = 4

Tabelle 2: Die Aufteilung der Stichprobe auf 4 Gruppen.

Nach einer kurzen Einführung in die Bedienung der sprachgesteuerten Anwendung und den Fahrsimulator wurden die Probanden gebeten, mehrere Aufgaben zu bearbeiten. Die erste Aufgabe lautete jeweils: „Bitte rufen Sie bei [Ihren Eltern, im Büro, Ihrer eigenen Nummer, ...] an“. Um die Aufgabe zu bearbeiten, mussten die Probanden in das entsprechende Menü wechseln, dann die Nummer diktieren und schließlich den Wählvorgang auslösen (siehe hierzu auch das Beispiel in Tabelle 1).

Erhoben wurden subjektive Angaben zur Nutzerzufriedenheit mittels Fragebogen und Interview sowie die Aufgabenbearbeitungszeiten (*total task times*). Der Versuchsleiter protokollierte darüber hinaus das Bedienverhalten der Probanden.

Ergebnisse der Simulatorstudie (Untersuchung 1)

Die Befunde zu den *total task times* bei Bearbeitung der ersten Telefonaufgabe belegen, dass eine sprachgesteuerte Anwendung auch bei gleichzeitiger Fahrzeugführung effizient genutzt werden kann. Die *total task times* hängen nicht davon ab, ob die Probanden im Fahrsimulator oder am Schreibtisch mit der sprachgesteuerten Anwendung interagierten (siehe Abbildung 2).

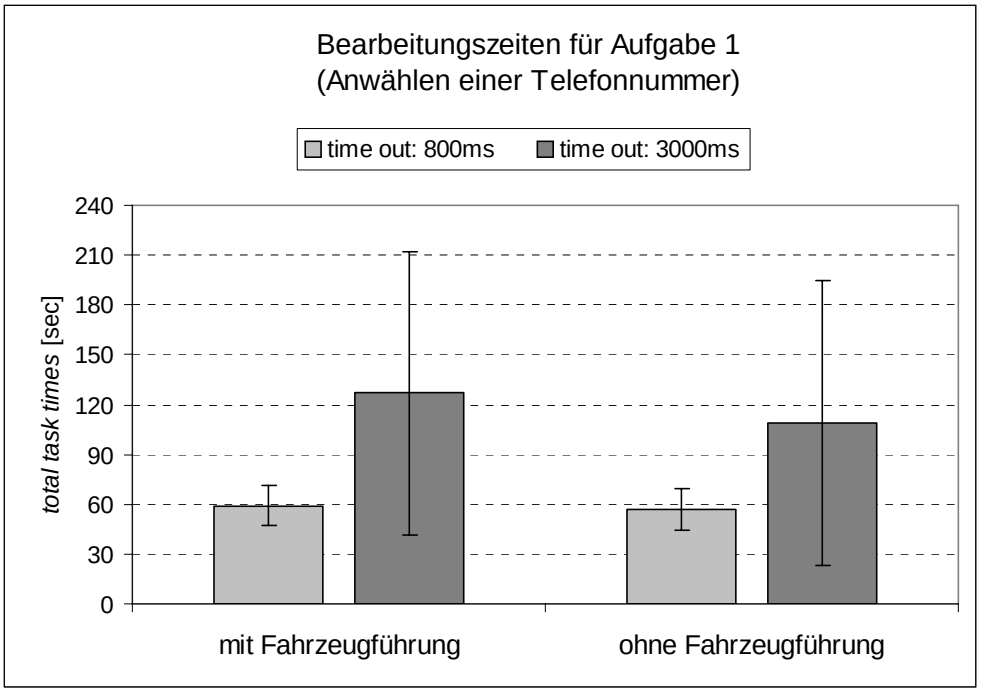

Abbildung 2: Aufgabenbearbeitungszeiten für Aufgabe 1 (Anwählen einer Telefonnummer).

Auffällig ist der enorme Einfluss, den die Länge des *time out* auf die Bearbeitungszeiten ausübt. Die Probanden, die mit einem *time out* von 800 ms konfrontiert wurden, erreichten deutlich kürzere *total task times* als jene, die mit einem *time out* von 3000 ms arbeiteten (siehe Abbildung 2).

Die kürzeren Bearbeitungszeiten sind darauf zurückzuführen, dass bei einem *time out* von 800 ms eine wesentlich straffere Nutzerführung resultiert als bei einem *time out* von 3000 ms. Sobald der Nutzer nicht mehr weiter weiß und schweigt, liefert ihm das System nach kurzer Zeit einen *prompt*. Die Probanden der anderen Gruppe, die relativ lange auf den *prompt* warten mussten (3000 ms), hatten oft nicht die notwendige Geduld und verfielen dann auf eine wenig effiziente Versuch-und-Irrtum Strategie.

Die Angaben im Fragebogen zeigten jedoch, dass die hohe Effizienz des kurzen *time out* (800 ms) mit der Beurteilung „Das System hetzt mich“ erkauft wurde. Die Probanden, die mit einem time out von 800 ms konfrontiert wurden, hatten stärker den Eindruck, sich bei der Befehlseingabe beeilen zu müssen, als jene, die mit einem time out von 3000 ms arbeiteten (siehe Abbildung 3). Dies gilt insbesondere für die Probanden, die zugleich ein Fahrzeug im Fahrsimulator führten.

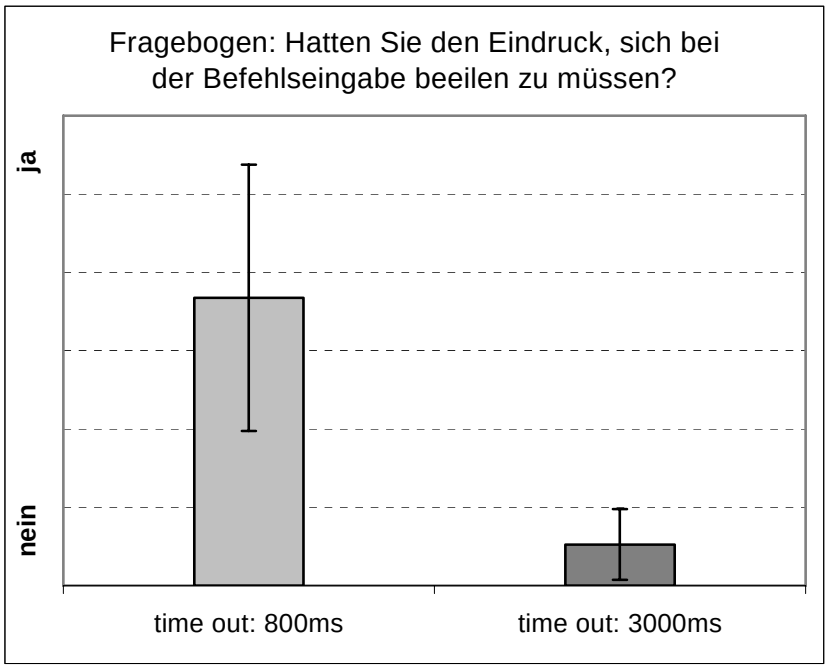

Abbildung 3: Angaben im Fragebogen zum Thema „Das System hetzt mich“.

In den Interviews kritisierten die Versuchspersonen u.a., dass die bereits eingegebenen Ziffern gelöscht wurden, wenn es zu einem (meist unfreiwilligem) Dialog-Abbruch durch das System kommt. Das heißt nach einem Dialog-Abbruch muss der Nutzer die Aufgabe ganz von vorne beginnen, da das System bei jedem Abbruch in seinen Ausgangszustand zurückkehrt. Dies stellt einen Verstoß gegen das Kriterium des nutzerkontrollierten Dialogverlaufs dar.
Auffällig war ferner, dass es einzelne Probanden gab, bei denen die Spracherkennung sehr häufig versagte. Diese Probanden fielen dem Versuchsleiter jedoch nicht durch eine undeutliche Sprache auf.

Schlussfolgerungen (Untersuchung 1)

Die Befunde zu den *total task times* (mit/ohne Fahraufgabe) zeigen, dass Sprachsteuerung auch während einer Fahrzeugführung effizient genutzt werden kann. Daher wurde als Folgeuntersuchung ein Feldversuch geplant. Für diesen sollte jedoch ein - anhand der nun vorliegenden Ergebnisse - verbessertes System erstellt werden.

Die Ergebnisse zu den Effekten der Länge des *time out* werfen jedoch Probleme auf: Entweder man „hetzt“ den Nutzer und verhilft ihm damit zu einer effizienten Aufgabenbearbeitung oder man lässt ihm viel Zeit, jedoch mit der Konsequenz, dass inakzeptable Bearbeitungszeiten und unvorhergesehenes Benutzerverhalten entstehen. Schließlich wurde das System um eine grafische Anzeige erweitert, so

dass zusätzlich zur akustischen (und damit zeitlich transienten) Nutzerführung auch eine grafische (und damit zeitlich permanente) Nutzerführung zur Verfügung steht. Es wurde davon ausgegangen, dass bei Vorhandensein einer permanenten grafischen Anzeige der *time out* für die akustischen *prompts* auf 7 sec erhöht werden kann.

Die im Interview erhobenen Kritiken führten dazu, dass das Dialog-Design so verändert wurde, dass die vom Nutzer bereits eingegebenen Daten auch im Falle eines Dialog-Abbruchs erhalten bleiben.

Die Tatsache, dass manche Personen enorme Schwierigkeiten mit der Spracherkennung hatten, führte dazu, dass das System um ein manuelles Eingabegerät erweitert wurde.

Zusammengenommen führen die geschilderten Verbesserungen von einem (unimodal) sprachgesteuerten zu einem multimodalen Bedien-/Anzeigekonzept.

Untersuchung 2 (Feldversuch)

In dieser Untersuchung sollten die folgenden Fragen geklärt werden:

- In welcher Weise wird die grafische Anzeige während der Fahrten genutzt?
- Bevorzugen die Nutzer die Sprach- oder die manuelle Eingabe?
- Ist der *time out* von 7 sec für den akustischen *prompt* eine sinnvolle Einstellung?

Für diesen Feldversuch wurde also ein multimodales Bedien-/Anzeigekonzept verwendet, das neben Spracheingabe und akustischer Ausgabe auch ein manuelles Eingabegerät sowie einen Monitor zur Verfügung stellte (siehe Abbildung 4).

Das manuelle Eingabegerät umfasst 4 Bedienelemente:

- die linke Taste führt zurück in die vorangegangene Menüebene; damit dient sie auch der Fehlerkorrektur
- die rechte Taste startet den Sprachdialog
- der rechte Dreh-Druck-Knopf reguliert die Lautstärke (drehen) und schaltet das Radio bzw. den CD-Player an oder aus (drücken)
- der linke Dreh-Druck-Knopf dient der Auswahl (drehen) und Bestätigung der Menüoptionen (drücken)
-

Abbildung 4: Multimodales System mit manuellem Eingabegerät und Monitor.

Die Menüoptionen sind auf dem Monitor bogen- oder kreisförmig angeordnet, um das *mapping* mit dem linken Dreh-Druck-Knopf zu gewährleisten (siehe Abbildung 5).

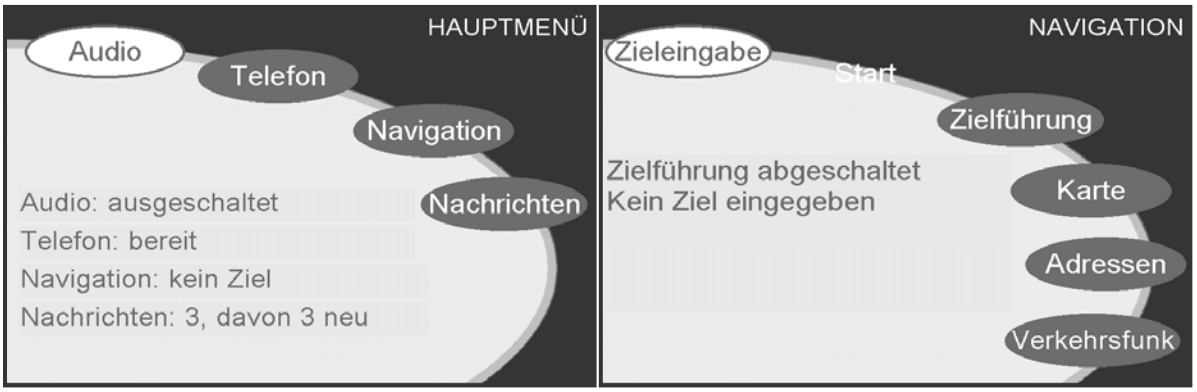

Abbildung 5: Zwei Beispiele für die Bildschirminhalte des multimodalen Systems.

Die 10 Versuchspersonen (∅ 32,3 Jahre, 2 Frauen und 8 Männer) des Feldversuchs waren ortskundige Fahrer. Alle wurden mit dem selben System konfrontiert. Als unabhängige Variable wurde lediglich die Streckenanforderung variiert, das heißt die Testroute führte sowohl über eine Ausbaustrecke als auch durch ein Wohngebiet (siehe Abbildung 6).

Abbildung 6: Zwei beispielhafte Abschnitte der Testroute (Ausbaustrecke und Wohngebiet).

Die Probanden erhielten eine kurze Einweisung in die Bedienung des multimodalen Bedien-/Anzeigekonzepts. Sie wurden darauf hingewiesen, dass sie zwischen den beiden Eingabemodalitäten jederzeit frei wählen können. Alle Probanden fuhren die Testroute (Rundkurs) einmal ab, um sich die Strecke einprägen zu können und sich an das Versuchsfahrzeug zu gewöhnen. Danach folgten zwei Testfahrten, während der die Probanden gebeten wurden, verschiedene Aufgaben zu bearbeiten (z.B. das Anwählen einer Telefonnummer, Eingabe von Stadt- und Straßennamen in das Navigationssystem).

Erhoben wurde das Blickverhalten (Frequenz der Blickzuwendungen zum Monitor und Blickdauern), das Bedienverhalten (bevorzugte Eingabemodalität) sowie die *user*

response time, das heißt die Zeitspanne zwischen Systemausgabe und Nutzereingabe. Diese *user response time* wurde zur Bestimmung der notwendigen Länge des *time out* herangezogen. Zusätzlich protokollierte der Versuchsleiter Auffälligkeiten im Bedienverhalten. Am Ende der Untersuchung wurde jeweils ein kurzes Interview durchgeführt.

Ergebnisse des Feldversuchs (Untersuchung 2)

Die Befunde zur Blickfrequenz belegen, dass das Blickverhalten an die jeweilige Verkehrssituation angepasst wurde, das heißt in dem unübersichtlichen Wohngebiet blickten die Versuchspersonen seltener auf den Monitor als auf der Ausbaustrecke (siehe Abbildung 7). Dieser Befund gilt für beide Testfahrten, geht also nicht auf einen Übungseffekt zurück.

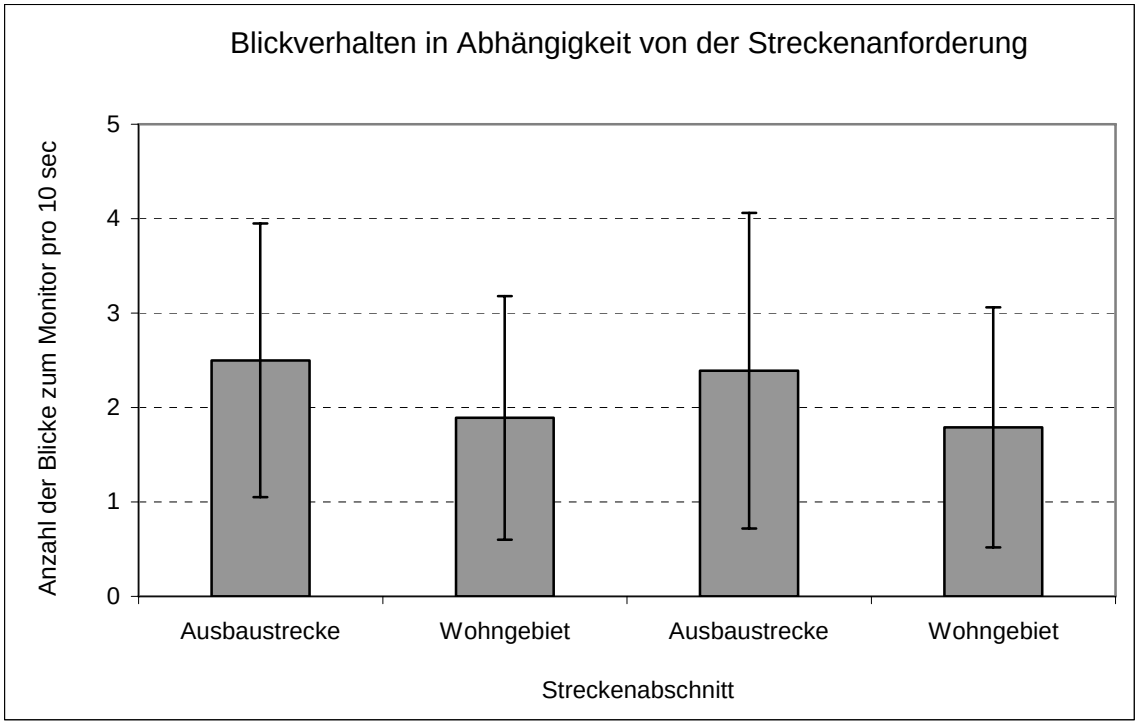

Abbildung 7: Blickfrequenz bei der Bearbeitung einer Telefon-Aufgabe in Abhängigkeit von der Streckenanforderung.

Die Dauern der Blickabwendungen vom Verkehrsgeschehen wurden für die Aufgaben „Telefon 1“ (Ausbaustrecke, erste Testfahrt) und „Telefon 4“ (Wohngebiet, zweite Testfahrt) ausgemessen. Die Befunde sind unauffällig, das heißt sie ähneln den Blickdauern, wie man sie auch bei der Bedienung eines konventionellen Autoradios vorfindet (siehe Abbildung 8).

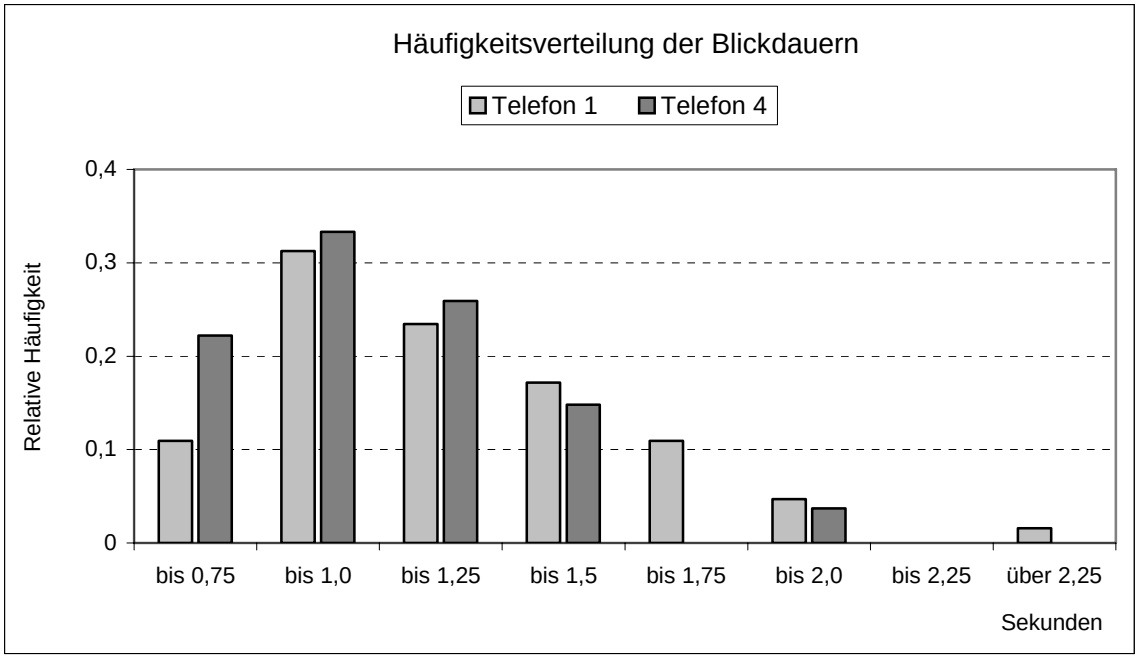

Abbildung 8: Dauern der Blickabwendungen vom Verkehrsgeschehen bei der Bearbeitung der Telefon-Aufgaben 1 und 4 (Häufigkeitsverteilung).

Die bevorzugte Eingabemodalität wurde vom Versuchsleiter protokolliert. Die Befunde weisen auf eine starke Bevorzugung der Spracheingabe hin (siehe Abbildung 9). Eine Versuchsperson fiel in die Kategorie „gemischt“ (Nutzung von Sprach- und manueller Eingabe in einem relativ ausgewogenen Verhältnis). Im Interview gab diese Versuchsperson jedoch an, dass sie nur aufgrund von Schwierigkeiten mit der Spracherkennung die manuelle Eingabe genutzt habe.

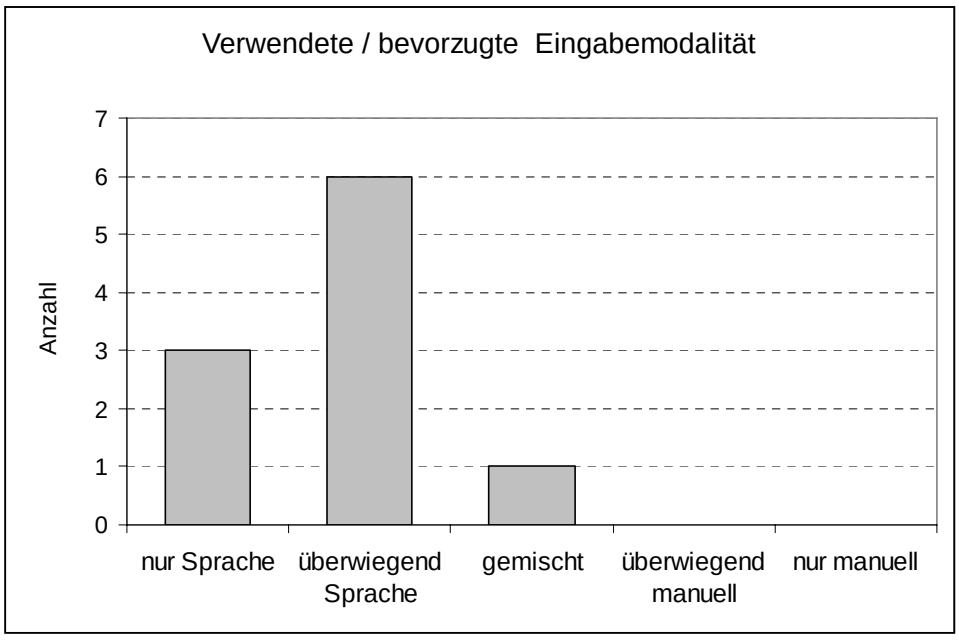

Abbildung 9: Die von den Probanden genutzte bzw. bevorzugte Eingabe-modalität.

Im Interview wurden die Probanden befragt, weshalb sie die Spracheingabe vor der manuellen Eingabe bevorzugten. Die Antworten der Probanden zeigen, dass sie die

Spracheingabe als die sicherere und auch komfortablere Eingabemodalität ansehen (siehe Abbildung 10).

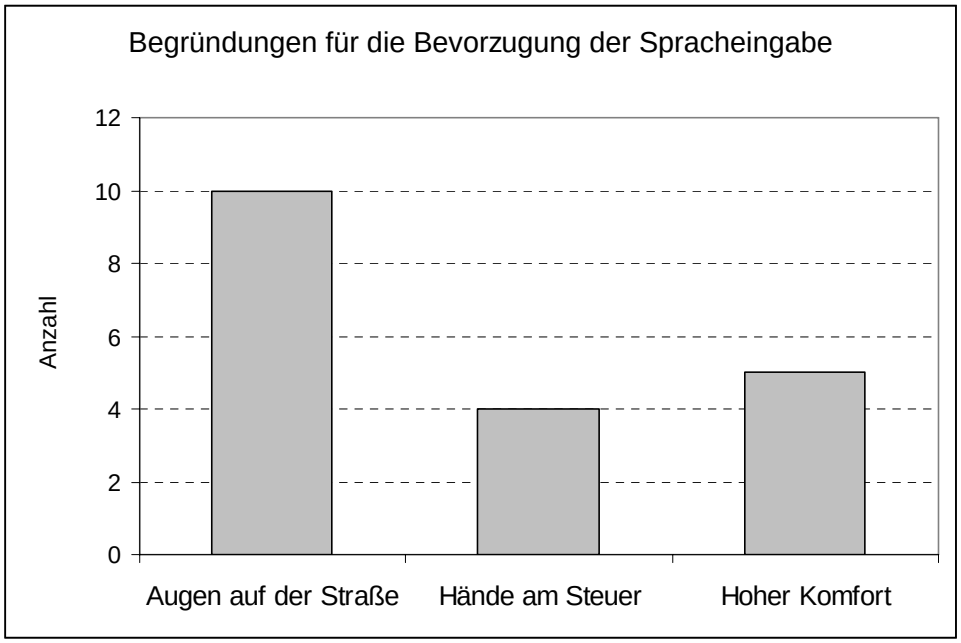

Abbildung 10: Die von den Probanden genannten Gründe für ihre Bevorzugung der Spracheingabe gegenüber der manuellen Eingabe.

Die erhobene *user response time* (in Anlehnung an die *system response time*) wird dazu genutzt, die geeignete Länge des *time out* zu bestimmen. Diesem Vorgehen liegt die Idee zugrunde, dass ein Nutzer, der die momentan relevanten Befehle kennt, innerhalb einer bestimmten Zeitspanne sein nächstes Kommando eingibt. Wird diese Zeitspanne überschritten, muss davon ausgegangen werden, dass der Nutzer den relevanten Befehl entweder nicht weiß oder aus sonstigen Gründen „den Faden verloren“ hat (z.B. aufgrund von Verkehrsanforderungen).

Der Großteil der Nutzereingaben erfolgt innerhalb der ersten 1,5 sec nach der letzten Systemausgabe (siehe Abbildung 11). Aufgrund von Denkpausen des Nutzers oder aufgrund von Verkehrsanforderungen kann sich die Nutzereingabe jedoch verzögern (bis zu 3,0 sec nach der letzten Systemausgabe). Auffällig ist, dass im Zeitraum von 3,0 bis 4.0 sec nach der letzten Systemausgabe keine Nutzereingabe stattfand. Die „späten“ Eingaben (nach 4,0 bis 4,5 sec) waren meist durch eine Verunsicherung der Nutzer gekennzeichnet (fragender Tonfall, geratene Kommandos, zahlreiche Häsitationen). Diese „späten“ Eingaben müssen daher einer Versuch-und-Irrtum-Strategie zugerechnet werden. Mit anderen Worten: Der in diesem Versuch gewählte *time out* von 7 sec ist zu lang. Etwa 3,5 sec nach der letzten Systemausgabe sollte im Falle eines schweigenden Nutzers ein *prompt* ausgegeben werden, um Verunsicherungen des Nutzers entgegenzuwirken und ein kooperatives Systemverhalten zu realisieren.

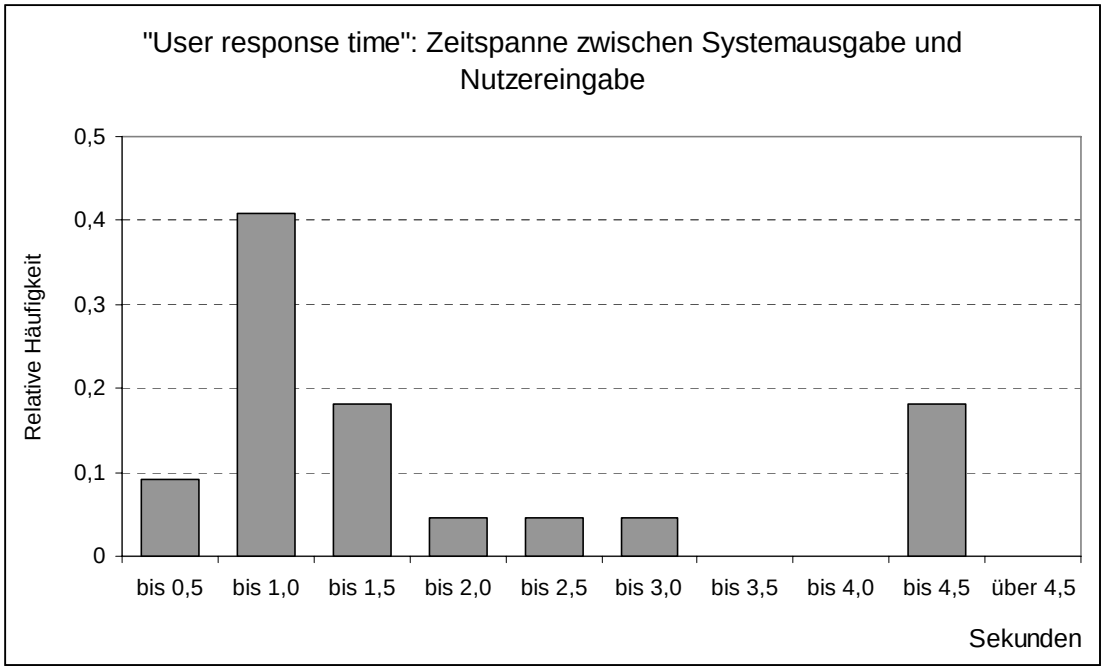

Abbildung 11: Häufigkeitsverteilung der „user response time“.

Ausgehend von den erhobenen *user reaction times* sollte der *time out* für den akustischen *prompt* daher etwa 3,5 sec betragen. Einerseits wird dem Nutzer genügend Bedenkzeit zugestanden (das heißt er wird durch das System nicht gehetzt), andererseits wird ihm relativ bald geholfen, wenn er nicht mehr weiter weiß.

Der Versuchsleiter konnte mehrfach beobachten, dass die Probanden in der Lage waren, die Fahrzeugführung und den Dialog mit dem System gleichzeitig durchzuführen. Einige Probanden zeigten beispielsweise mustergültiges Blickverhalten an rechts-vor-links-Kreuzungen obwohl sie gerade mit dem System interagierten. In schwierigeren Verkehrssituationen dagegen wurde der Dialog mit dem System einfach ignoriert, das heißt die Probanden wandten ihre ganze Aufmerksamkeit der primären Fahraufgabe zu. Anschließend wurde der Dialog problemlos wieder aufgenommen.

Schlussfolgerungen (Untersuchung 2)

Die Befunde zum Blickverhalten belegen, dass der Monitor des multimodalen Bedien-/Anzeigekonzepts keine inakzeptablen Blickabwendungen vom Verkehrsgeschehen provoziert. Die Blickfrequenz zeigt, dass die Nutzer ihr Blickverhalten an die jeweiligen Streckenanforderungen anpassten. Die Dauern der Blickabwendungen sind unauffällig.

Die Probanden bevorzugten die Eingabe per Sprache vor der manuellen Eingabe. Hierfür wurden vor allem sicherheitsrelevante Gründe genannt (geringere visuelle und geringere motorische Beanspruchung). Zusätzlich wird die Spracheingabe als sehr komfortabel empfunden.

Die Befunde zu den *user response times* lassen vermuten, dass sich die Kooperativität des multimodalen Systems erhöhen ließe, wenn der *time out* für den akustischen *prompt* auf etwa 3,5 sec gesetzt würde. Spracheingaben des Nutzers, die vor dieser Zeitmarke stattfanden, waren zielorientiert und korrekt. Spracheingaben, die erst danach stattfanden, wiesen Merkmale auf, die auf eine Verunsicherung des Nutzers schließen lassen.

Der Versuchsleiter konnte während des Feldversuchs immer wieder beobachten, dass die Probanden den Dialog mit dem System ignorierten, wenn die Verkehrssituation gesteigerte Aufmerksamkeit erforderte. Dies ist vermutlich auch darauf zurückzuführen, dass die Probanden keine negativen Konsequenzen hinsichtlich des Systemverhaltens erwarten mussten:

- es gehen keine bereits getätigten Eingaben verloren (nutzerkontrollierter Dialogverlauf)
- das System wartet „geduldig“, bis der Nutzer den Dialog wieder aufnimmt
- nach einer bestimmten Zeitspanne hilft das System dem Nutzer, den Dialog wieder aufzunehmen (kooperatives Systemverhalten)

Der Dialog erfüllt somit das Kriterium der Aufgabenunterbrechbarkeit. Der Nutzer kann sich jederzeit vom Dialog abwenden und voll auf das Verkehrsgeschehen konzentrieren.