

J. Ziegler & A. Schmidt (Hrsg.): Mensch & Computer 2010  
München: Oldenbourg Verlag, 2010, S. 301-310

# Mobile Dienste im Fahrzeug: Gestaltung von Sprachausgaben zur Reduzierung visueller Ablenkung

**Julia Niemann**

Deutsche Telekom Laboratories

Quality and Usability

**Jessika Reissland**

Technische Universität Berlin

Mensch-Maschine Systeme

**Anja Naumann**

Deutsche Telekom Laboratories

Quality and Usability

## **Zusammenfassung**

Fahrerinformationssysteme (FIS) mit graphischen Schnittstellen und ergänzender Sprachbedienung führen häufig zu visueller Ablenkung von der Straße, obwohl eine Interaktion vollständig über Sprachein-/ausgaben möglich ist. Ziel der vorliegenden Arbeit war es, anhand eines kognitiven Modells zur Aufmerksamkeitszuwendung zu erarbeiten, welche Faktoren die Blickzuwendungen zu einem Display begünstigen und welche Nachteile auf Seiten der Sprachausgaben zu finden sind. Anschließend wurden Gestaltungsmaßnahmen für einen mobilen Dienst (E-Mail) generiert. Diese Gestaltungsmaßnahmen haben eine Erhöhung der Auftretenshäufigkeit und Bandbreite von akustischen Informationen, sowie eine Minimierung des zeitlichen Aufwands von Sprachausgaben zum Ziel. Durch diese Gestaltungsmaßnahmen sollen die Nachteile von Sprachausgaben gegenüber Displays kompensiert und so die Blickabwendungen von der Straße minimiert werden.

## 1 Einleitung

Das Automobil hat seit seiner Erfindung Ende des 19. Jahrhunderts wie kein zweites Verkehrsmittel zur individuellen Fortbewegung der Menschheit beigetragen (Dragon, 2007). Zusätzlich zu dieser Entwicklung veränderte die mobile Telekommunikation mit all ihren Diensten (Telefon, Textnachrichten, E-Mails, etc.) die moderne Form der Kommunikation. Es erscheint offensichtlich, dass Fahrzeugnutzer auch während der Fahrt nicht auf ihre mobile Kommunikation verzichten möchten. Beim Einsatz solcher Dienste im Fahrzeug sollten jedoch die besonderen Anforderungen beachtet werden, die sich durch den sicherheitskritischen Kontext der Fahraufgabe ergeben. Die Interaktion mit Systemen, die dem Komfort, der Unterhaltung und der Informationsaufnahme, unabhängig von der Fahraufgabe dienen, sind nach Bubb (2003) lediglich Nebenaufgaben oder auch Tertiäraufgaben. Primäraufgabe ist immer die Fahrzeugführung, hinzukommen Sekundäraufgaben, die Handlungen beinhalten, die in Bezug zur Fahraufgabe stehen und bei der Primäraufgabe unterstützen. Ziel der vorliegenden Arbeit ist es, bei der Gestaltung von Schnittstellen mobiler Dienste im Fahrzeug die Besonderheit der Interaktion durch den Kontext der Fahraufgabe zu betrachten.

## 2 Sprachdialogsysteme im Fahrzeug

Eine Ablenkung des Fahrers durch die Nebenaufgabe sollte möglichst gering gehalten werden, um eine sichere Fahrzeugführung zu gewährleisten. Die Informationsaufnahme bei der Fahraufgabe erfolgt hauptsächlich über die visuelle Modalität während die Reaktionsausführungen vorrangig motorischer Natur sind. Visuelle Ablenkungen von der Fahraufgabe stellen eine der häufigsten Unfallursachen dar (Wierwille & Tijerina, 1995) und legen somit eine Erweiterung visuell-haptischer Schnittstellen von Informationssystemen um Sprachein-/ausgaben nahe, um genügend Ressourcen für die visuell-motorische Fahraufgabe bereitzustellen. Dies lässt sich auch durch die Annahmen des Multiplen Ressourcen Modells erklären, nach welchem zwei Aufgaben weniger stark miteinander interferieren, wenn sie unterschiedliche mentalen Ressourcen beanspruchen (Wickens, 1984). So konnte in Studien zu Ablenkungspotentialen von Fahrerinformationssystemen (Vollrath & Totzke, 2000) ein Vorteil akustischer Informationen gegenüber visueller Informationen in Bezug auf die Fahrperformance nachgewiesen werden. Da es jedoch nach Young et al. (2003) nicht nur zu visuellen und psychomotorischen Ablenkungen von der Fahraufgabe kommen kann, sondern auch akustische und kognitiv beanspruchende Nebenaufgaben die Leistung in der Primäraufgabe beeinträchtigen können, gilt es bei der Gestaltung von Sprachdialogsystemen diese Form der Ablenkung möglichst gering zu halten. So empfiehlt Jeschke (2008), dass Sprachausgaben nicht länger als unbedingt nötig sein sollten, sowie, dass einfache Formulierungen, kurze Sätze und übersichtliche Satzkonstruktionen vorteilhaft sind. Vilimek (2008) konnte in einer Untersuchung zur Ablenkungswirkung langer Sprachausgaben im Vergleich zu stichpunktartigen Sprachausgaben, zeigen, dass zusätzliche Informationen, die bedeutungsirrelevant waren und somit ignoriert werden konnten, zu Leistungseinbußen bei der Bearbeitung einer Primäraufgabe führten. Auf sprachästhetische Gesichtspunkte sollte somit eher verzichtet werden, wenn dies zu einer unnötigen zeitlichen Verlängerung der Sprachausgabe führt.

Wie erwähnt besteht die Zielsetzung beim Einsatz von Voice User Interfaces (VUIs) im Fahrzeug darin Blickabwendungen von der Straße zu vermeiden. Fraglich ist jedoch, ob

VUIs tatsächlich in der Lage sind, die Graphical User Interfaces (GUIs) vollständig zu ersetzen. Führt zudem der Einsatz von sprachbasierten Schnittstellen tatsächlich zwangsläufig dazu, dass die visuelle Aufmerksamkeit nicht mehr auf das Display des Informationssystems gelenkt wird? Kun et. al. (2009) konnten in einer Fahrsimulatorstudie zeigen, dass bei der Nutzung eines Portable Navigation Device (PND) mit akustischer Unterstützung Probanden häufiger den Blick von der Fahrbahn abwendeten, wenn ein Display vorhanden war, als wenn dieses abgedeckt wurde. Informationen wurden also visuell aufgenommen, obwohl die Nebenaufgabe auch ohne Blickzuwendungen zum Display erfolgreich bewältigt werden konnte. Außerdem resultierten die Blickabwendungen in stärkeren Variationen der Spurhaltung und dem Winkel des Lenkrads. Auch in den subjektiven Daten spiegelte sich wieder, dass trotz Sprachausgaben das visuelle Feedback für die Interaktion mit dem Informationssystem eine wichtige Komponente für den Fahrer darstellt. Das Ausblenden des Displays stellt insofern keine ausreichende Lösung dar, als dass hierbei keine Ursachen- sondern lediglich eine Symptombhebung erfolgt. Vielmehr sollten Sprachein/-ausgaben verbessert und überarbeitet werden, um Blickabwendungen von der Straße zu vermeiden.

### 3 Aufmerksamkeitszuwendung zu GUI und VUI

Welche Eigenschaften der graphischen Darstellung des Displays erhöhen die Wahrscheinlichkeit von Blickzuwendungen? Ein Modell, das Vorhersagen über die Aufmerksamkeitsausrichtung von Operateuren in dynamischen Umgebungen trifft, ist das SEEV-Modell von Wickens et al. (2003). Wickens et al. (2003) definieren vier Faktoren, die die Wahrscheinlichkeit einer Aufmerksamkeitsausrichtung auf einen Reiz beeinflussen: Salienz (Salience), Aufwand (Effort), Auftretenshäufigkeit/Informationsbandbreite (Expectancy) und Wert (Value). Die Faktoren Salienz und Aufwand repräsentieren hierbei die sogenannten Bottom-up Faktoren, die eine durch Umweltreize gesteuerte Aufmerksamkeitsverschiebung bewirken. Salienz bezeichnet die Eigenschaften des Reizes. Ein besonders salienter Reiz, beispielsweise ein hell blinkendes Licht, erhöht die Wahrscheinlichkeit der Blickzuwendung des Operateurs. Diese Wahrscheinlichkeit wird jedoch minimiert, wenn größere Distanzen durch Blick- und Kopfbewegungen zu überwinden sind (Aufwand). Die Auftretenshäufigkeit und der Wert hingegen sind Top-down Faktoren. Sie bezeichnen mentale Prozesse des Operateurs, die sich auf die Wahrscheinlichkeit der Blickzuwendung zu einer bestimmten Region auswirken. So beeinflusst die Auftretenswahrscheinlichkeit eines Reizes, die subjektiv erlebte Erwartung des Operateurs, in einer bestimmten Region ein Ereignis zu detektieren. Ist der Wert dieses Ereignisses hoch relevant, so erhöht auch dies die Wahrscheinlichkeit der Blickzuwendung. Hieraus ergibt sich die dargestellte Formel:

$$P(\text{Attention Allocation}) = \text{Salience} - \text{Effort} + (\text{Expectancy} + \text{Value})$$

Die Evaluation des SEEV-Modells und somit die Vorhersage der Blickpfade und Fixationen erfolgten sowohl für den Fahrzeug- als auch für den Luftfahrtkontext (Wickens & McCarley, 2007). Das Modell zeigt hierbei eine Prädiktionsgüte von  $r = 0.6-0.9$ . Nach Wickens's SEEV Modell kann also die Aufmerksamkeitsausrichtung auf bestimmte Bereiche im Fahrzeug durch Veränderung jedes einzelnen der vier Faktoren beeinflusst werden. Nachfolgend werden daher die Faktoren gesondert hinsichtlich ihres möglichen Einflusses auf die Wahrscheinlichkeit der Aufmerksamkeitsausrichtung auf GUI bzw. VUI betrachtet.

*Salienz.* „Saliency is a maximum for auditory icons“ (Wickens et al., 2008, S.6). Dies lässt vermuten, dass hierbei keine Vorteile auf Seiten der GUI gegenüber der VUI zu sehen sind.

*Wert.* Es kann zunächst davon ausgegangen werden, dass die Relevanz einer Aufgabe (z.B. die Eingabe eines Ziels in ein Navigationssystem) für den Fahrer prinzipiell unabhängig von der Art der Informationsdarbietung ist. Dem Modell von Wickens (1984) folgend, konkurriert jedoch eine visuelle Präsentation der Nebenaufgabe mit der Primäraufgabe, da diese bereits in starkem Umfang visuelle Ressourcen bindet (s.o.). Die Primäraufgabe, die für den Fahrer ebenfalls mit einem bestimmten Wert verbunden ist, wird also bei Zuwendung zur GUI vernachlässigt. Dementsprechend wird angenommen, dass die Bereitschaft des Fahrers, sich einer visuellen Nebenaufgabe zuzuwenden prinzipiell geringer ist, da dies zu Leistungseinbußen in der Fahraufgabe führen kann. Die Relevanz der Aufgabenbearbeitung steht demnach auch in Zusammenhang mit der erfolgreichen Bearbeitung und der Relevanz anderer Aufgaben (Fahraufgabe).

Die Faktoren Wert und Salienz sind demnach hypothetisch nicht für eine erhöhte Wahrscheinlichkeit der Aufmerksamkeitszuwendung zum Display verantwortlich.

*Auftretenswahrscheinlichkeit/Informationsbandbreite.* Sprachausgaben sind im Gegensatz zur Darstellung visueller Elemente nicht permanent sondern momentan und meist seriell repräsentiert. Die Wahrscheinlichkeit, dass bei Blickzuwendung zum Display des Fahrerinformationssystems die relevanten Informationen zu finden sind, liegt hiermit bei 100%. Eine Darbietung der akustischen Informationen dagegen erfolgt häufig nur einmalig, z.B. nachdem eine Eingabe durch den Fahrer erfolgte und ist anschließend nicht wiederholbar. Dies erfordert vom Operateur nach dem Arbeitsgedächtnismodell von Baddeley (1986) das mentale Aufrechterhalten der akustischen Informationen in der phonologischen Schleife. Kommt es zu für den Fahrzeugkontext typischen Unterbrechungen in der gegebenen Sekundäraufgabe, so kann dies eine Neuorientierung des Aufgabenbearbeitungszustands erforderlich machen, die durch die einmalige Repräsentation der akustischen Informationen jedoch nicht mehr möglich ist.

*Aufwand.* Der Vorteil von graphischen Informationen gegenüber Sprachausgaben und akustischen Signalen liegt darin, dass schneller mehr Informationen übermittelt werden können. Beim Lesen eines Textes können ca. 300 und mehr Wörter pro Minute erfasst werden, während sich dies beim Hören eines Textes auf nur 100 Wörter pro Minute beschränkt (Kozma, 1991). Der Zeitaufwand bezüglich Sprachsignalen ist alleine durch die Kosten der Repräsentation derselben (sequentiell) höher als bei visuellen Reizen. Durch gestalterische Maßnahmen der Informationen auf dem Display (bestimmte räumliche Anordnung, bestimmte Farbgebungen und Schriftgrößen der graphischen Elemente) können zudem auch indirekt schnell zusätzliche Informationen übermittelt werden, die die Bildung eines mentalen Modells der Menüstrukturen fördern (Hasebrook, 1995).

*Zielkonflikt zwischen Bandbreite an Informationen und Aufwand.* Die Darbietung aller Informationen der GUI durch die VUI würde zu einem hohen zeitlichen und somit kognitiven Aufwand für den Fahrer führen. Um dies zu vermeiden, erfolgt, wie bereits angesprochen, bei der Festlegung der Sprachausgaben häufig eine Einschränkung der Informationsbandbreite dieser im Vergleich zur VUI. Burmester et al. (2008) empfiehlt beispielsweise zur Entlastung des Arbeitsgedächtnisses, nicht alle Systemausgaben auch akustisch zu präsentieren und stattdessen Teile der Information auf das Display auszulagern. Häufig wird dies für die Prä-

sentation möglichen Sprachbefehle getan. Auf dem Display wird dem Nutzer angezeigt, welche gültigen Sprachbefehle in dem jeweiligen Interaktionsschritt genannt werden können (Say What You See- Prinzip, Weinschenk & Barker, 2000), während er in der akustischen Ausgabe lediglich inhaltliche Informationen oder eine eingeschränkte Auswahl der möglichen Sprachbefehle angesagt bekommt. Es handelt sich hierbei offensichtlich um einen Zielkonflikt zwischen den Parametern Informationsbandbreite und zeitlichem Aufwand.

## 4 Systemergonomische Gestaltungsmaßnahmen

Im Folgenden sollen Gestaltungsvorschläge für eine akustische Schnittstelle dargestellt werden, dessen Ziel es ist, die Auftretenshäufigkeit und Bandbreite von akustischen Informationen zu steigern und gleichzeitig den zeitlichen und kognitiven Aufwand für den Fahrer nicht zu erhöhen. Diese Gestaltungsmaßnahmen wurden bereits exemplarisch für das VUI einer E-Mail Applikation umgesetzt. Abbildung 1 zeigt Screenshots der Anwendung.



Abbildung 1.: Screenshots der E-Mail Applikation

Die prototypische E-Mail Applikation soll es dem Fahrer ermöglichen, während der Fahrt E-Mails abzurufen und sich diese mit Hilfe von Sprachsynthese vorlesen zu lassen. Weiterhin besteht die Möglichkeit auf E-Mails zu antworten, E-Mails weiterzuleiten und neue E-Mails zu generieren, indem eine Audiodatei aufgezeichnet wird, die anschließend an einen Kontakt aus dem Adressbuch versendet werden kann. Auch eine Suche nach Absendern wurde in der Applikation implementiert. Im Posteingang kann der Fahrer zwischen den empfangenen E-Mails mittels der Befehle „vorherige“ und „nächste“ navigieren. Die Applikation ist über Sprache vollständig bedienbar. Hierzu wurde ein Spracherkennungssystem sowie eine Aktionstaste an der Lenkradwippe integriert. Ausgehend von empirischen Ergebnissen, die eine Präferenz von Nutzern bezüglich kommandosprachlicher Schnittstellen im Fahrzeug belegen (Graham et al., 1999), wurde das System nicht natürlichsprachlich umgesetzt. Um den Spracherkennungssystem zu aktivieren, muss der Fahrer vor jeder Interaktion mit dem System die Aktionstaste kurz drücken, um anschließend einen der in der Spracherkennungssystem hinterlegten Sprachbefehle zu nennen. Als Feedback für eine getätigte Spracheingabe erfolgt eine systeminitiierte Sprachausgabe. Die Aktionstaste kann jederzeit gedrückt werden. So können auch Sprachausgaben des Systems unterbrochen werden, wenn dies in der aktuellen Fahrsituation nötig wird oder der Fahrer sich den Inhalt der Sprachausgabe nicht zu Ende anhören möchte.

### 4.1 Erhöhung der Auftretenshäufigkeit

Zur Erhöhung der Auftretenswahrscheinlichkeit und zur Entlastung der phonologischen Schleife sollte dem Fahrer die Möglichkeit gegeben werden, die Sprachausgaben jederzeit

wieder abrufen zu können, um sich u.a. nach Unterbrechung durch die Primäraufgabe wieder neu orientieren zu können ohne auf das Display zu schauen. Entscheidend ist hierbei die fahrerinitiierte Steuerung der Sprachausgaben, damit dieser die Aufmerksamkeit, wie auch bei visuellen Reizen, bewusst auf die Zweitaufgabe fokussieren kann. Somit wird eine aufgezwungene Interaktion mit dem System in beanspruchenden Fahrsituationen vermieden. Hierzu wurde die Aktionstaste zur Aktivierung des Spracherkenners um eine Funktionalität erweitert. Bei langem Drücken dieser Aktionstaste (Longterm Push Aktionstaste = LPA) sollen dem Fahrer alle relevanten Informationen der GUI dargeboten werden. Somit erfolgt eine Unterscheidung von direkten und systeminitiierten Sprachausgaben und fahrerinitiierten Sprachausgaben durch den LPA. Bei der aktuell implementierten E-Mail Applikation erhält der Fahrer über systeminitiierte Ausgaben ausschließlich Informationen zur Orientierung im Menü und inhaltliche Informationen (wie E-Mail Betreff, E-Mail Inhalt, Absender oder eingegebene Empfänger). Bei Aktivierung der LPA werden zusätzlich alle möglichen Sprachbefehle genannt („Say What You Hear“ (SWYH) Prinzip, analog zum „Say what you see“ Prinzip). Somit werden bei langem Drücken der Aktionstaste akustisch Orientierungsinformationen, Inhalte und gültige Sprachbefehle dargeboten. Dies hat den Vorteil, dass einen hohe Bandbreite an Informationen dargeboten wird, dass heißt alle Informationen, die auch auf dem Display zu finden sind, sind auch akustisch abrufbar.

Um den in Abschnitt 4 angesprochenen Zielkonflikt zwischen großer Bandbreite an Informationen und dem zeitlichen Aufwand durch lange Sprachausgaben auf dem LPA zu reduzieren, erfolgten weitere Gestaltungsmaßnahmen, die im Folgenden beschrieben werden.

## 4.2 Minimierung des zeitlichen Aufwands

Drei Lösungsansätze zur Minimierung des zeitlichen Aufwands wurden für die VUI der E-Mail Applikation umgesetzt:

- Nonverbaler Klänge
- Schneller abgespielte Sprachausgaben
- Informationsanreicherung

### 4.2.1 Nonverbale Klänge

Wie graphische Icons Text auf kleinerem Raum abbilden können, so ermöglichen nonverbale Klänge die schnellere Präsentation von akustischen Informationen (Brewster, 2002). Sie bieten somit eine Lösung zur Reduzierung der Zeit unter Beibehaltung der Informationseinheiten dar. Zu nonverbalen Klängen zählen Spearcons, Earcons und Auditory Icons (Brewster, 2002). Spearcons sind sehr schnell abgespielte und leicht erlernbare Sprachausgaben. Es handelt sich bei diesen Klängen somit nicht per se um nonverbales Material, doch Spearcons haben die Eigenschaft Sprache in kürzerer Zeit abzubilden. Zu dem werden die zu repräsentierenden Wörter häufig so schnell abgespielt, dass sie ohne Kontextinformationen oder Vorwissen nicht verstanden werden können. Jedes Wort behält sich aber seine eigene charakteristische Tonspur und stellt eine Art Fingerabdruck dar (Walker et al., 2009). Spearcons sind durch Audioprogramme sehr einfach zu produzieren und können in manchen Fällen auch Menühierarchien abbilden (vgl. Dingler et al. 2008). Earcons sind kurze musikalische Konzepte (Melodien), die ein Menüitem repräsentieren. Die künstlich generierten Töne, die sich durch Frequenz, Klangfarbe, Dauer, Tempo und Reihenfolge unterscheiden, eignen sich

besonders gut, um Menühierarchien ohne Sprache abzubilden. Allerdings erfordern sie immer eine Lernphase, da sie keine semantische Verknüpfung zum repräsentierten Wort vorweisen. Auditory Icons sind Klänge, die mit der Semantik des Wortes oder einem graphischen Icon assoziiert sind. Sie stellen direkte oder metaphorische Repräsentation eines Konzepts dar (Gaver, 1986). Ein Beispiel hierfür ist der Klang vom Zerknüllen eines Blatt Papiers, das bei Mac-Computern bei Verschieben einer Datei in den Papierkorb zu hören ist. Je stärker die Assoziation des Auditory Icons zu einem natürlichen Klang ist, desto intuitiver sind diese und müssen im besten Falle nicht erst erlernt werden, sondern erschließen sich dem Nutzer sofort. Abbildung 2 zeigt die Eignung von den nonverbalen Klängen auf den Dimensionen Erlernbarkeit (Identifizierbarkeit), Produzierbarkeit, Menühierarchien und Ästhetik (Wohlklang) in Anlehnung an Guski (1997).

	Auditory Icons	Earcons	Spearcons
Erlernbarkeit	+	-	o
Produzierbarkeit	-	o	+
Menühierarchien	-	+	o
Ästhetik	+	o	-

Abbildung 2.: Vergleich nonverbaler Klänge für Erlernbarkeit, Produzierbarkeit, Menühierarchien, Ästhetik

Für das VUI der hier beschriebenen E-Mail Applikation wurden vorrangig Earcons und Auditory Icons genutzt. Auf Spearcons wurde auf Grund des schlechten Wohlklangs verzichtet. Earcons sollen eine Orientierung im Menü unterstützen. Als Basis für die Earcons in der E-Mail Applikation wurde ein spezifischer E-Mail Klang generiert, der zur Darstellung der jeweiligen Menüunterpunkte um einen Ton mit unterschiedlicher Klangfarbe erweitert wurde. Die drei Bedienschritte zum Verfassen einer Sprachnachricht (Empfängereingabe, Aufnahme einer Nachricht und Zusammenfassung/Versenden) wurden durch ansteigende Tonhöhen repräsentiert. Bei jedem Bedienschritt, der den Nutzer näher an sein Ziel (das Versenden der Nachricht) bringt, wird der vorangegangene Klang um einen nächsthöheren Ton auf der Tonleiter ergänzt. Der Einsatz der Earcons ermöglicht es somit, beispielsweise auf die Sprachausgabe „Sie befinden sich im E-Mail Posteingang“ zu verzichten und so die akustische Ausgabe zeitlich zu verkürzen. Auditory Icons wurden eingesetzt, wenn eine intuitive Assoziation zu gegebenen Funktionalitäten der E-Mailapplikation gefunden werden konnten. So hören die Nutzer beim Versenden einer E-Mail nicht den Satz „Sie haben die E-Mail erfolgreich versandt.“ sondern den Klang eines durch die Luft fliegenden Objekts.

#### 4.2.2 Schneller abgespielte Sprachausgaben

Der Einsatz von schneller abgespielten Sprachausgaben erfolgt bei inhaltlichen Informationen, die dem Nutzer bereits als systeminitiierte Ausgabe präsentiert wurden. Betätigt der Nutzer beispielsweise bei der Empfängereingabe den LPA, so werden ihm die eingegebenen Empfänger noch einmal, aber dieses mal schneller, vorgelesen (30 %).

#### 4.2.3 Informationsanreicherung

Um indirekte Informationen, die durch graphische Mittel zur Gestaltung des Displays verwendet werden, auch akustisch abzubilden und einen hohen zeitlich Aufwand durch die Vertonung des Textes zu vermeiden, sind u.a. die Variation von Pausen, sowie unterschiedli-

che Stimmen bei der Verwendung von Sprachausgaben denkbar. Bei der E-Mail Applikation im Fahrzeug sind alle Bedienelemente graphisch durch magentafarbene Buttons von den restlichen Informationen unterscheidbar. Hierbei werden demnach Eingabemöglichkeiten, die der Fahrer in dem jeweiligen Interaktionsschritt hat, durch eine spezifische Farbe repräsentiert. Um diese Information auch akustisch abzubilden, wäre dies über Metainformation („Sie haben folgende Optionen...“) möglich. Um jedoch eine Zeitersparnis zu erreichen, wurde ein anderer Ansatz in der vorliegenden Arbeit gewählt. Bedienelemente (gültige Spracheingaben) der VUI werden mittels einer anderen Stimme (männliche Sprachsynthese) dargeboten als alle anderen Informationen (weibliche Sprachsynthese). Akustisch repräsentiert somit die männliche Sprachsynthese mögliche Spracheingaben. Unterschiedliche Pausenlängen wurden eingesetzt, um Informationen, die inhaltlich in Verbindung stehen, zu bündeln. So ist die Pause im Posteingang zwischen den Sprachausgabe „Antworten“ und „Weiterleiten“ zeitlich kürzer als zwischen „Weiterleiten“ und Elementen zur Navigation in der E-Mail Liste wie „Vorherige“ oder „Nächste“. Dies wird graphisch durch eine spezifische räumliche Anordnung der Elemente erreicht (vgl. Abbildung 1). Miller (1956) konnte zeigen, dass bis zu  $7 \pm 2$  Elemente im Arbeitsgedächtnis aufrechterhalten werden können, wobei diese Elemente in Länge und Komplexität variieren können. Durch die Bündelung bestimmter inhaltlich zusammenhängender Information mithilfe von unterschiedlichen Pausenzeiten könnte so eine bessere Memorierbarkeit der dargebotenen Informationen erreicht werden.

## 5 Zusammenfassung und Ausblick

Basierend auf einem kognitiven Modell zur Aufmerksamkeitsausrichtung wurden die geringere Auftretenshäufigkeit und der hohe zeitliche Aufwand von Sprachausgaben als mögliche Ursachen für eine Blickzuwendung zu Displays während der Fahrt identifiziert. Um diese Nachteile auszugleichen, wurden Gestaltungsmaßnahmen für das VUI einer E-Mail Applikation umgesetzt, die auf eine fahrerinitiierte Abrufbarkeit aller Informationen, die auf dem Display dargestellt werden, abzielen. Gleichzeitig wird durch den Einsatz von nonverbalen Klängen, schneller abgespielter Sprache und unterschiedlichen Stimmen und Pausenlängen versucht, mehr Informationen in kürzerer Zeit zu übermitteln, um eine Erhöhung des kognitiven und zeitlichen Aufwands durch die Repräsentation aller Information zu vermeiden. Ob diese Maßnahmen zu der erwarteten Minimierung der Blickabwendungen von der Straße führt soll in einer Fahrsimulatorstudie mit Erfassung von Blickdaten evaluiert werden. Weiterhin soll erhoben werden, welche der Gestaltungsmaßnahmen sich als besonders geeignet zur Reduzierung des kognitiven und zeitlichen Aufwands herausstellt und ob hierdurch auch die Bildung eines besseren mentalen Modells der Menüstruktur erreicht werden kann.

## 6 Literaturverzeichnis

Baddeley, A.D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Brewster, S.A. (2002). Non-speech auditory output. In Jacko, J.A. and Sears, A. (Eds) *Human-Computer Interaction Handbook*, Chap. 12. Mahwah, NJ.: Lawrence Erlbaum Associates, S. 220-239.

- Bubb, H. (2003). Fahrerassistenz – primär ein Beitrag zum Komfort oder für die Sicherheit? In: VDI-Berichte Nr. 1768: *Der Fahrer im 21. Jahrhundert. Anforderungen, Anwendungen, Aspekte für Mensch-Maschine-Systeme*. Düsseldorf, S. 25–44.
- Burmester, M., Graf, R., Hellbrück, J. & Ansgar, M.. (2008). Usability — Der Mensch im Fahrzeug. In *Infotainmentsysteme im Kraftfahrzeug*. Springer.
- Dingler, T., Lindsay, J., & Walker, B. N. (2008). Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. *Proceedings of the International Conference on Auditory Display (ICAD 2008)*, Paris, France (24-27 June).
- Donges, D. (1992). Das Prinzip der Vorhersagbarkeit als Auslegungskonzept für Maßnahmen zur Aktiven Sicherheit im Straßenverkehrssystem. In *VDI Bericht 948: Das Mensch-Maschine-System im Verkehr*. Düsseldorf: VDI Verlag.
- Dragon, L. (2007). Fahrzeugdynamik: Wohin fahren wir? In Schindler und Sievers (Hrsg.): *Forschung für das Auto von Morgen*; 1. Aufl. Berlin: Springer-Verlag. S. 239-260.
- Fricke, N. (2006). Semantic auditory icons as warning signals. In T. Stockman, L. V. Nickerson & C. Frauenberger (Hrsg.), *Proceedings of ICAD 2006 at Queen Mary, University of London. Published by the Department of Computer Science*, Queen Mary, University of London, UK.
- Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2(2), 167-177.
- Graham, R., Aldridge, L., Carter, C. & Lansdown, T.C. (1999). The design of in-car speech recognition interfaces for usability and user acceptance. In D. Harris (Hrsg.), *Engineering psychology and cognitive ergonomics: Job design, product design and human-computer interaction* (Vol. 4, S. 313-320) Aldershot: Ashgate.
- Guski, R. (1997). Psychological methods for evaluating sound quality and assessing acoustic information. *Acustica united with Acta Acustica*, 83(5), 765-774.
- Hasebrook, J. (1995). *Multimedia-Psychologie*. Heidelberg, Berlin, Oxford: Spektrum: Akademische Verlag.
- Jeschke, B. (2008). Sprachausgaben von Sprachdialogsystemen im Kfz. *ITG-Fachtagung Sprachkommunikation* (8.- 10. Oktober 2008 Aachen). VDE Verlag.
- Kozma, R.B. (1991). Learning with media. *Review of Educational Research*, 61, 179-211.
- Kun, A.L., Paek, T., Medenica, Z., Memarovic, N., & Palinko, O. (2009). Glancing at Personal Navigation Devices Can Affect Driving: Experimental Results and Design Implications. *Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2009)*, Sep 21-22 2009, Essen, Germany.
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on pur capacity for processing information. In: *Psychological Review* 63, S. 81–97.
- Vilimek, Roman (2008) *Gestaltungsaspekte multimodaler Interaktion im Fahrzeug. Ein Beitrag aus ingenieurpsychologischer Perspektive*. Düsseldorf: VDI Verlag.

- Vollrath, M. & Totzke, I. (2000). In-vehicle communication and driving: an attempt to overcome their interference. *Driver Distraction Internet Forum sponsored by the United States Department of Transportation*.
- Walker, B. N., Nance, A. & Lindsay, J. (2006). Spearcons: speech-based Earcons improve navigation performance in auditory menus. *Proceedings of the International Conference on Auditory Display (ICAD 2006)*, London, England, 20-24 June 2006, S. 63-68.
- Weinschenk, S. & Barker, D.T. (2000). *Designing effective speech interfaces*. New York: John Wiley & Sons.
- Wickens, C. D. (1984). Processing Resources in Attention. In Parasuraman, R. & Davies, D.R. (Hrsg.), *Varieties of Attention*, S. 63-97. New York: Academic Press.
- Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159-177.
- Wickens, C. D., Goh, J., Helleburg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multi-task pilot performance using advanced display technology. *Human Factors*, 45, 360–380.
- Wickens, C. D. & Holland, J. G. (2000) *Engineering Psychology and Human Performance*. New Jersey: Prentice-Hall.
- Wickens, C. D., & McCarley, J. M. (2007). *Applied attention theory*. Boca Raton, FL: CRC Press.
- Wickens, C. D., McCarley, J. S., Alexander, A. L., Thomas, L. C., Ambinder, M., Zheng, S. (2008), Attention-situation awareness (A-SA) model of pilot error. In D. C. Foyle & B. L. Hooey (Hrsg.), *Human performance modeling in aviation* (S. 213-239). Boca Raton, FL: Taylor & Francis.
- Wierwille, W. & Tijerina, L. (1995). Eine Analyse von Unfallberichten als ein Mittel zur Bestimmung von Problemen, die durch die Verteilung der visuellen Aufmerksamkeit und der visuellen Belastung innerhalb des Fahrzeugs verursacht werden. *Zeitschrift für Verkehrssicherheit*, 41, 164-168.
- Young, K., Regan, M. & Hammer, M. (2003). *Driver Distraction: A Review of the Literatur*. Clayton Campus: Monash University, Australia: Accident Research Centre.

### **Danksagung**

Dank geht an Stefan Seide, Thomas Scheerbarth, Frank Oberle, Felix Burkhardt & Jianshen Zhou.

### **Kontaktinformationen**

Julia Niemann (Deutsche Telekom Laboratories)

Telefon: +49 (30) 8353-58383

Telefax: +49 (30) 8353-58409

E-Mail: [julia.niemann@telekom.de](mailto:julia.niemann@telekom.de)