A. M. Heinecke, H. Paul (Hrsg.): Mensch & Computer 2006: Mensch und Computer im Struktur*Wandel*.
München, Oldenbourg Verlag, 2006, S. 203-212

Hierarchische Stapelkarten und Sprachsteuerung: Neue Konzepte für multimodale PDA-Anwendungen

Hilko Donker, Oliver Stache

Dozentur Kooperative multimediale Anwendungen, Technische Universität Dresden

Zusammenfassung

Die multimodale Interaktion zwischen Mensch und Maschine stellt eine viel versprechende Basis für die Erstellung ergonomischer und gebrauchstauglicher PDA-Anwendungen dar. In diesem Beitrag wird die Kombination von Spracheingabe und stiftbasierter Steuerung der grafischen Benutzungsschnittstelle einer PDA-Anwendung untersucht. Zur Steuerung der Dialogabläufe wird die Metapher der hierarchischen Stapelkarten eingeführt. Zur Optimierung der Stiftinteraktion wird die grafische Benutzungsschnittstelle in einen Darstellungsbereich für Informationen und einen Navigationsbereich unterteilt. Alle Dialoge lassen sich sowohl durch eine Stifteingabe als auch durch Sprachkommandos steuern. Dieses Konzept wurde mit Hilfe einer Beispielimplementierung einer PDA-basierten Clientanwendung für den Webshop einer Online-Buchhandlung evaluiert.

1 Einleitung

In der Praxis werden unter multimodalen Benutzungsschnittstellen in der Regel hybride Benutzungsschnittstellen verstanden, die die visuellen und motorischen Kommunikationskanäle von Graphical User Interfaces (GUI) mit den auditiven von Speech User Interfaces (SUI) vereinen (Weinschenk & Barker 2000). Aufgrund der Kombination von GUI und SUI können die ergonomischen Vorteile beider Interaktionsformen in einer Benutzungsschnittstelle kombiniert und gleichzeitig einige ihrer Nachteile umgangen werden. Neben der Sprachausgabe kann eine auditive Ausgabe auch durch nichtsprachliche, gut differenzierbare akustische Signale (Earcons) (Brewster 1994) erfolgen. Sprachbasierte Benutzungsschnittstellen haben den Vorteil, dass sie die natürliche Sprache nutzen. Die natürliche Sprache ist variabel und abwechslungsreich, bei der Formulierung bietet sie dem Menschen einen Freiraum an Kreativität, ohne dabei an Effektivität zu verlieren. Das Hören von deutlicher, natürlicher Sprache stellt für den Menschen auch auf Dauer keine große kognitive Belastung dar. Die synthetisierte Sprachausgabe wird von den Nutzern jedoch auf Dauer als monoton und

unnatürlich empfunden. Bei der Spracheingabe stellt die Vermittlung der Sprachkommandos ein wesentliches Problem dar. Weder das Nachschlagen der Kommandos in der Programmhilfe noch das Vorlesen durch die Sprachsynthese stellen einen effizienten Lösungsansatz dar. Im praktischen Einsatz sprachbasierter Systeme hat sich gezeigt, dass Spracherkennungssysteme sehr empfindlich auf Umgebungsgeräusche reagieren, was insbesondere in mobilen Anwendungssituationen zu Problemen führen kann. Diese Probleme lassen sich bei multimodalen Benutzungsschnittstellen kompensieren, indem der Nutzer in einer lauten Umgebung eine alternative Eingabemodalität, zum Beispiel die Stiftsteuerung, wählt. Bei der Stiftsteuerung einer PDA-basierten Anwendung verdeckt der Nutzer oftmals wichtige Kontextinformationen mit seiner eigenen Hand, da auf den Displays der PDAs nur sehr wenig Platz für die Darstellung der Interaktionsobjekte und der Anwendungsinformationen für jeden Dialogschritt zur Verfügung stehen.

1.1 Gestaltungsgrundsätze für multimodale Benutzungsschnittstellen

Bei der Gestaltung einer multimodalen Benutzungsschnittstelle gilt es, bewährte Gestaltungsgrundsätze für grafische sowie sprachbasierte Benutzungsschnittstellen zu berücksichtigen. Darüber hinaus ergeben sich aus der Kombination der Modalitäten neue Grundsätze, die ebenso bei der ergonomischen Gestaltung dieser Schnittstellen beachtet werden müssen. Die folgenden Grundsätze basieren auf Ausarbeitungen von Weinschenk und Barker (Weinschenk & Barker 2000), sowie Pitt und Edwards (Pitt & Edwards 2002). **Grenzen der menschlichen Informationsverarbeitung**: Der Mensch hat Grenzen in Bezug auf die Quantität sowie die Qualität der von ihm zu verarbeitenden Informationen. Eine Benutzungsschnittstelle sollte diese Grenzen berücksichtigen. Zur Grenze der kognitiven Verarbeitung gibt George A. Miller an, dass ein Mensch sich zwischen fünf und neun Elemente für ungefähr 20 Sekunden merken kann (Miller 1956). Um diese Zahl zu erhöhen, muss die Information zum Beispiel aufgeteilt werden. Sanders und McCormick (Sanders & McCormick 1982) zufolge kann der Mensch sich eine Struktur von drei bis vier Elementen mit je drei bis vier Unterelementen am leichtesten merken. Die **Grenzen der visuellen Wahrnehmung** des Menschen äußern sich dadurch, dass ein Nutzer nicht alle dargestellten Informationen liest. Dieser Effekt verstärkt sich, wenn Informationen unübersichtlich dargestellt werden. Es sollten somit nur die wichtigsten Informationen – angemessen auf dem Bildschirm positioniert – dargestellt werden. Die **Verarbeitung auditiver Informationen** wird z.B. dadurch beeinträchtigt, dass eine wahrgenommene synthetisierte Sprache eine höhere Konzentration des Hörers verlangt als eine in der natürlichen Sprache gesprochene Äußerung. Durch die Verwendung einer synthetisierten Sprache wird der Transfer der Informationen ins Langzeitgedächtnis beeinträchtigt. Als Konsequenz sollten Sprachausgaben, die synthetisiert erfolgen, kurz und prägnant gestaltet sein. **Modale Integrität**: Modale Integrität bedeutet, dass für jede Aufgabe die effektivste Ein- oder Ausgabemodalität verwendet wird (Sanders & McCormick 1982). So sollte zum Beispiel das Selektieren eines Kartenausschnitts mit der Maus erfolgen. Das anschließende Vergrößern des Kartenauschnitts wird optimal durch das Sprechen des Kommandos „Vergrößern“ realisiert. Die Modale Integrität ist somit eine Voraussetzung für effektive multimodale Benutzungsschnittstellen. Damit Benutzungsschnittstellen auch resistent gegen äußere Einflüsse sind, sollte für jede Aufgabe ein alternativer Kommunikationskanal zur Verfügung gestellt werden. **Modale Konsistenz**: Wird ein und

dieselbe Information über verschiedene Modalitäten gleichzeitig präsentiert, müssen die jeweiligen Ausprägungen konsistent gestaltet sein. Dies bedeutet insbesondere, dass eine einheitliche Terminologie verwendet werden muss. **Nichtsprachliche auditive Ausgabe**: Earcons eignen sich, um die Aufmerksamkeit des Nutzers zu steuern. Wenn das System nach einer längeren Sprechpause eine Sprachinformation ausgeben möchte, sollte der Nutzer darauf vorbereitet werden ansonsten kann es passieren, dass er die ersten Worte der Ausgabe überhört. Am besten eignet sich dafür ein kurzer Ton, der sofort die Aufmerksamkeit des Nutzers erregt. **Privatsphäre und Sicherheit**: Bei der Eingabe sensitiver Daten sollte gewährleistet sein, dass die unmittelbare Umgebung diese nicht mithören kann. Darüber hinaus kann es einem Nutzer peinlich sein, wenn Meldungen über Bedienfehler mitgehört werden können. Für diese und ähnliche Fälle muss eine Alternative zur Spracheingabe und -ausgabe vorhanden sein.

2 Konzeption einer multimodalen PDA-Benutzungsschnittstelle

2.1 Hierarchische Stapelkartenmetapher

Die Navigationsstruktur der Anwendung wird dem Nutzer mit Hilfe der Bildschirmmetapher der Stapelkarten vermittelt. Bei der Stapelkartenmetapher (vgl. Microsoft Power Point Präsentationen) werden einzelne Bildschirmseiten durch Karten und die gesamte Anwendung durch einen Kartenstapel repräsentiert. Wird eine Karte vom Stapel genommen, sieht der Nutzer die darunter liegende Karte. In dem hier vorgestellten Konzept wurde diese Metapher so erweitert, dass der Kartenstapel hierarchisch strukturiert ist. Von einer Karte kann entweder zu einer darunter liegenden Hierarchieebene oder zurück zur darüber liegenden Karte gewechselt werden. Die Metapher der „hierarchischen Stapelkarten“ erlaubt es, die Dialoge der Anwendung sehr übersichtlich zu strukturieren. Beim Wechsel zwischen Karten wird deren Zusammenhang dadurch verdeutlicht, dass dieser Wechsel als Animation dargestellt wird. Beim Wechsel zu einer darunter liegenden Karte wird die aktuelle nach links aus dem Darstellungsbereich heraus geschoben und um zu einer Karte auf einer übergeordneten Hierarchiestufe zurückzukehren wird diese wiederum von links in den Darstellungsbereich hinein geschoben. Darüber hinaus entsteht mit Hilfe dieser Animation die Möglichkeit, zusätzliche Karten, die nicht Teil der Hierarchie sind, einzublenden, indem diese von einer anderen Seite ins Bild geschoben werden. Eine Karte, die Fehlermeldungen enthält, sollte zum Beispiel von der Hierarchie entkoppelt dargestellt werden. Die Aufmerksamkeit des Nutzers wird beim Eintreten von besonderen Dialogsituationen (Anzeige von Fehlermeldungen oder das Treffen einer kritischen Entscheidung) dadurch erhöht, dass mehrere Farbschemata für die Darstellung der Karten vorgesehen sind. Neben einem im Normalfall verwendeten grauen Farbschema werden zu diesem Zweck Farbschemata in Grün, Gelb und Rot als Signalfarben, sowie Blau als neutrale Farbalternative verwendet.

2.2 Layout der Benutzungsschnittstelle

Die Orientierung auf dem kleinen Display eines PDAs wird dadurch erleichtert, dass jede Stapelkarte in die beiden festen Darstellungsbereiche Information und Navigation untergliedert wird. Um zu vermeiden, dass der Nutzer bei der Bedienung mit einem Stift wichtige Informationen mit seiner Hand verdeckt, ist der Navigationsbereich unterhalb des Informationsbereichs angeordnet. Im Navigationsbereich sind Schaltflächen dargestellt, mit deren Hilfe der Nutzer sich zwischen den einzelnen Karten bewegen kann. Diese Schaltflächen sind untereinander angeordnet und erstrecken sich jeweils über die gesamte Displaybreite. Hierdurch wird gewährleistet, dass die Beschriftung der Buttons bei der Bedienung sowohl für Links- also auch für Rechtshänder nicht verdeckt wird. In Abbildung 1 wird die beschriebene Displayaufteilung schematisch verdeutlicht.

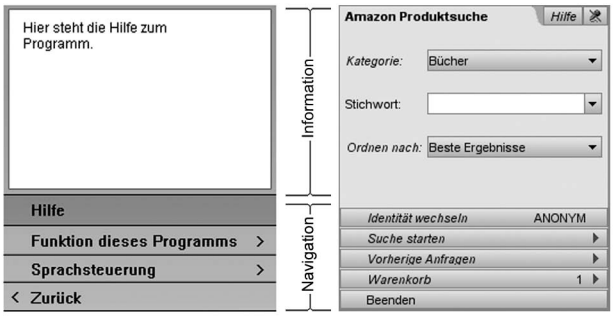

Abbildung 1: Darstellungsbereiche einer Bildschirmseite (Prototyp, Beispielanwendung)

2.3 Programmsteuerung mit Hilfe von Sprachkommandos

Die Vermittlung der in einer bestimmten Dialogsituation zur Verfügung stehenden Sprachbefehle wird dadurch realisiert, dass die Sprachbefehle besonders hervorgehoben in der GUI dargestellt werden. Auf diese Weise muss der Nutzer in der Einarbeitungsphase nicht sämtliche Befehle erlernen, sondern kann diese direkt vom Display ablesen. Als einheitliche Formatierung für Sprachbefehle werden diese kursiv gesetzt. Neben der Gestaltung der Sprachbefehle stellt die Ausgabe eines Feedbacks einen weiteren Problembereich der auditiven Programmsteuerung dar. Die Grundsätze der modalen Integrität fordern, dass die Reaktion auf eine auditive Eingabe ebenfalls auditiv übermittelt werden sollte. Nach einer Spracheingabe muss dem Nutzer zunächst eine Rückmeldung gegeben werden, ob ein gesprochenes Kommando vom System als solches identifiziert und zum Spracherkenner geleitet wurde oder ob der Nutzer sein Sprachkommando wiederholen muss. In einem zweiten Schritt muss dem Nutzer eine Rückmeldung über das Ergebnis der Erkennung gegeben werden. Der Nutzer benötigt eine Rückmeldung, ob ein Kommando erkannt wurde oder ob das Kommando zurückgewiesen wurde. Da beide Rückmeldungen nach jedem Sprachkommando gegeben werden, ist die effizienteste Möglichkeit für ein Feedback kurze aber prägnante Earcons zu verwenden. Nach jeder als Kommando interpretierten Spracheingabe wird ein kurzer Piepton

ausgegeben. Weist der Spracherkenner ein Kommando zurück, wird ein doppelter Piepton verwendet. Das Feedback auf eine erfolgreiche Kommandoerkennung wird entgegen der modalen Integrität nur implizit als visuelle Information dargestellt, indem die gewünschte Funktion ausgeführt und ihr Ergebnis auf dem Display dargestellt wird.

2.4 Multimodale Repräsentation klassischer Interaktionsobjekte

Interaktionsobjekte und Informationen werden in klassischen Anwendungen auf der grafischen Benutzungsschnittstelle präsentiert. Im Folgenden wird analysiert, welche Interaktionselemente sich für eine multimodale Interaktion eignen und auf welche Weise deren Sprachsteuerung realisiert werden können. Ein Button kann entweder auf die gewohnte Weise durch Antippen mit einem Stift bedient werden oder über ein Sprachkommando, welches seiner Beschriftung entspricht. Zur Auszeichnung dieses Sprachkommandos muss die Beschriftung entsprechend den Überlegungen in Abschnitt 2.3 kursiv hervorgehoben werden. Für Dropdown-Boxen ist eine als Sprachkommando gekennzeichnete Beschriftung der Box erforderlich. Über dieses Sprachkommando wird die Box ausgeklappt und die Elemente der Box werden dargestellt. Die Auswahl der Elemente erfolgt, indem entweder die Beschriftungen der einzelnen Elemente selbst als Sprachkommando dienen oder eine den Elementen vorangestellte Nummerierung als Sprachkommando verwendet wird. In Abbildung 2 sind diese beiden Varianten der multimodalen Dropdown-Box dargestellt. Grafische Listen dienen zur Darstellung von Elementmengen. Es können einzelne Elemente oder Teilmengen der Liste für eine Weiterverarbeitung selektiert werden. Um ein Element einer multimodalen Liste mittels Sprache selektieren zu können, werden sämtliche Einträge der Liste durchnummeriert. Die einem jeweiligen Element vorangestellte Nummer dient wiederum als Sprachkommando für die Selektion. In ähnlicher Weise können auch Teilmengen einer Liste über Sprachbefehle selektiert werden, allerdings wird eine solche Funktion vorerst nicht unterstützt. In Tabellen werden Elementmengen strukturiert dargestellt. Sie bieten dem Nutzer die Möglichkeit, einzelne Zellen oder ganze Zeilen zu selektieren. Im Rahmen dieser Arbeit wird nur die Selektion von Tabellenzeilen unterstützt. Um eine konsistente Bedienung zu gewährleisten, werden dafür, wie bereits bei den Listen und Dropdown-Boxen, die Tabellenzeilen durchnummeriert. Über das Sprechen der Nummer wird eine Zeile selektiert. Kartenreiter (oder Tabbed Panes) können alternativ zur Stiftsteuerung über Sprache ausgewählt werden. Ihre Beschriftung dient gleichzeitig als Sprachkommando.

Abbildung 2: Dropdown-Box mit Sprachsteuerung über einen Index bzw. Wortlaut des Eintrags

3 Beispielanwendung

Als Anwendungsbeispiel wurde ein mögliches zukünftiges Einsatzgebiet mobiler Anwendungen im Sinne des Ubiquitous Computing gewählt. Es wurde eine multimodale PDA-basierte Clientanwendung für den Webshop einer Online-Buchhandlung erstellt, die dem Nutzer eine weiterführende Onlinerecherche ermöglicht, während er sich in seiner Lieblingsbuchhandlung vor Ort befindet. Der Funktionsumfang entspricht nicht dem vollen Umfang dieses Online-Shops. Vielmehr liegt das Hauptaugenmerk auf der ergonomischen Gestaltung von wesentlichen Grundfunktionen. Die multimodale Benutzungsschnittstelle ist auf der Basis des frei verfügbaren Java-basierten GUI-Toolkits *Thinlet* (Bajzat 2005) und des Spracherkennungs- und Sprachsynthesesystems *jLab SpeechServer* entstanden, die miteinander kombiniert und um neue Funktionen erweitert wurden. Thinlet übernimmt die Darstellung der grafischen Benutzungsoberfläche, wickelt sämtliche GUI-basierten Interaktionen des Nutzers ab und steuert die daraus resultierenden Aufrufe der Programmlogik. Für die Beschreibung der grafischen Benutzungsschnittstelle dienen XML-Dateien. Der *jLab SpeechServer* wurde von der Fakultät Elektrotechnik und Informationstechnik der Technischen Universität Dresden entwickelt. Es handelt sich um einen auf Hidden-Markov-Modellen basierenden Sprachkommandoerkenner und dem auf Diphonkonkatenation basierendem Sprachsynthesesystem *DRESS*.

4 Evaluation der Beispielanwendung

4.1 Konzeption und Durchführung der Evaluation

Die Beispielanwendung wurde mit Hilfe eines Usability-Tests anhand der folgenden Kriterien und Fragestellungen analysiert: Welchen Einfluss hat die Metapher der hierarchischen Stapelkarten auf die Orientierung des Nutzers innerhalb der Anwendung? Wird eine bessere Übersichtlichkeit der dargestellten Informationen erzielt, indem das Display in einen Informations- und einen Navigationsbereich unterteilt wird? In welchem Umfang wird die Sprachsteuerung benutzt und welche Auswirkung hat dabei die Fehlerrate? Wird die Präsentation der Sprachkommandos in Form einer hervorgehobenen Darstellung in der GUI vom Nutzer ausreichend wahrgenommen? Sind die als Sprachbefehle gewählten Kommandos für den Nutzer verständlich und ist das Ergebnis dieser Anwendungsfunktionen vorhersehbar? Ist das Feedback, das die Nutzer im Anschluss an Sprachkommandos erhalten, ausreichend und zweckmäßig? Als Evaluationsmethoden wurden eine Videobeobachtung der Interaktion mit dem PDA und ein Interview im Anschluss an die Videobeobachtung gewählt. Im Rahmen der Videobeobachtung hatten die Versuchspersonen vier Aufgaben mit steigendem Schwierigkeitsgrad zu absolvieren. Die Evaluation wurde als Usability-Test mit anschließendem Interview durchgeführt. An der Studie haben sechs Versuchspersonen teilgenommen. Die Versuchspersonen gehörten zur Altersgruppe 20 bis 30 Jahre. Das Durchschnittsalter betrug 25,5 Jahre. Drei Versuchspersonen waren weiblich und drei männlich. Drei Personen hatten sehr gute Erfahrungen mit webbasierten Anwendungen, eine gute und zwei mittlere Erfahrungen. Keine der Versuchspersonen hatte Erfahrungen mit einem Sprachdia-

logsystem. Die folgenden vier Aufgaben wurden von den Versuchspersonen im Rahmen des Usability-Tests bearbeitet:

- Suchen nach einem konkreten Produkt innerhalb eines Beispielwebshops.
- Auslösen einer Bestellung innerhalb eines Beispielwebshops.
- Wiederholen einer bereits gestellten Suchanfrage.
- Freies Arbeiten unter Zuhilfenahme sämtlicher angebotener Funktionen.

4.2 Evaluationsergebnisse

Aus der Videoaufzeichnung der Versuche wurden die folgenden Größen betrachtet: die Bearbeitungszeit der Einzelaufgaben, die Anzahl der gegebenen Stift- und der Sprachbefehle, die Anzahl der nicht erkannten und der falsch erkannten Sprachbefehle, sowie die Anzahl offensichtlicher Navigationsfehler. Ergänzt wurden diese objektiven Größen durch subjektive Aussagen der Versuchspersonen aus den Interviews. Zu Beginn wird die **Orientierung innerhalb der Anwendung** betrachtet. Es wird analysiert, ob die benötigten Informationen und Funktionen gefunden wurden. Sämtliche Versuchspersonen erfassten den Aufbau der grafischen Benutzungsschnittstelle sofort und begannen jeweils unmittelbar mit der Erledigung der ersten Aufgabe. Die weitere Analyse des Videomaterials zeigt, dass die Unterteilung des Displaybereichs in Informations- und Navigationsteil ebenso schnell verinnerlicht wurde wie die hierarchische Strukturierung der Anwendungsseiten. Alle Versuchspersonen setzten zur Erfüllung der Aufgaben die Anwendungsfunktionen gezielt und effizient ein, was durch die geringe Rate an Navigationsfehlern belegt wird. Lediglich der Versuchsperson P3 sind zwei Fehler und der Versuchsperson P6 ist ein Fehler unterlaufen. Keine der Versuchspersonen griff zur Erfüllung der Aufgaben auf die Hilfefunktion der Anwendung zurück, was ebenfalls belegt, dass die Benutzungsschnittstelle intuitiv und übersichtlich gestaltet ist. Die Auswertung der ersten Phase der Arbeit mit dem System zeigt, dass sich die Metapher der hierarchischen Stapelkarten in Verbindung mit der Unterteilung der grafischen Benutzungsschnittstelle in einen Informations- und einen Navigationsbereich positiv auf die Übersichtlichkeit und die Bedienbarkeit dieser PDA-basierten Anwendung auswirkt. Die **Nutzung der Sprachsteuerung** bei der Erfüllung der Aufgaben war den Versuchspersonen freigestellt. In Abbildung 3 ist der Anteil der Spracheingaben an der Gesamtheit der Eingaben in Abhängigkeit von den Personen und Testaufgaben dargestellt. Darin wird deutlich, dass keine der Versuchspersonen ausschließlich die Sprachsteuerung oder aber ausschließlich die Stiftsteuerung verwendete, sondern die durch multimodale Benutzungsschnittstellen ermöglichte Kombination beider Modalitäten gewählt wurde. Abgesehen von Versuchsperson P3, die bis zu Beginn der vierten Aufgabe auf Stiftsteuerung verzichtete, ist bei allen anderen mit jeder Aufgabe ein Anstieg des Sprachanteils an der Interaktion festzustellen. Dies ist ein Indiz für eine steigende Akzeptanz der Sprachsteuerung. Für die Versuchspersonen P1, P3 und P5, welche die Sprachsteuerung besonders intensiv genutzt haben, ist jeweils in der vierten Aufgabe ein leichter Rückgang des Sprachanteils zu verzeichnen. Die Versuchspersonen wurden im Interview jeweils dazu befragt und begründeten diese Auffälligkeit mit der „zu langen Reaktionszeit des Spracherkenners“ und mit der „auf Dauer zu hohen Fehlerrate der Spracherkennung“. Der starke Einfluss der Fehlerrate der Spracherkennung auf die Verwendung der Sprachsteuerung wird bei Versuchsperson P4 besonders deutlich. Der Anteil der gesproche-

nen Befehle bei der vierten Aufgabe beträgt nur 38,9%, was auf die extrem hohe Fehlerrate bei der Spracherkennung von 64,3% zurückzuführen ist. Im Gegensatz dazu wurde bei Versuchsperson P2, die bei den Aufgaben 3 und 4 verstärkt Sprachbefehle einsetzte, jeweils eine Fehlerrate von 0% gemessen. Im Interview wurde von P2 ein besonderes Wohlgefallen der Sprachsteuerung bestätigt. Die Sprachsteuerung wurde insgesamt in hohem Maße genutzt, allerdings nur solange die Fehlerrate der Spracherkennung gering blieb.

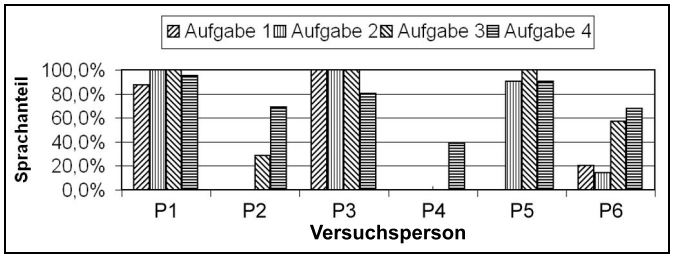

Abbildung 3: Anteil der Sprachinteraktion an der gesamten Interaktion

Die **Effizienz der Benutzungsschnittstelle** wurde anhand des Vergleichs der Bearbeitungszeiten der unterschiedlichen Eingabemodalitäten untersucht. Theoretisch ist die Sprachsteuerung der vorgestellten multimodalen PDA-basierten Clientanwendung weniger effizient als die Stiftsteuerung, da die Reaktion auf ein Stiftkommando unmittelbar folgt, ein Sprachkommando hingegen erst vom Spracherkenner mit einer durchschnittlichen Verarbeitungszeit von 1,1 Sekunden verarbeitet werden muss. Darüber hinaus treten bei der Sprachsteuerung Erkennungsfehler auf. Im Gegensatz dazu wird ein Stiftkommando stets korrekt erkannt, wenn der Nutzer auf die richtige Stelle tippt. Zur Bewertung der Effizienz der Modalitäten werden die Bearbeitungszeit der ersten drei Aufgaben von den Versuchspersonen P1 und P3 mit einem Sprachanteil von 98% im Vergleich zu denen der Versuchspersonen P2 und P4 mit einem Sprachanteil von 5% betrachtet (Abbildung 4). Bei der ersten Aufgabe ist die mittlere Bearbeitungszeit mit Sprachsteuerung niedriger als bei der stiftbasierten Steuerung. Für die Aufgaben 2 und 3 hingegen ist die Bearbeitungszeit mit der Sprachsteuerung geringfügig höher als die der Stiftsteuerung. Insgesamt ergibt sich aus diesen Beobachtungen, dass für die Zielgruppe der Einsteiger und Gelegenheitsnutzer keine signifikanten Unterschiede in der Effizienz der Eingabemodalitäten zu beobachten sind.

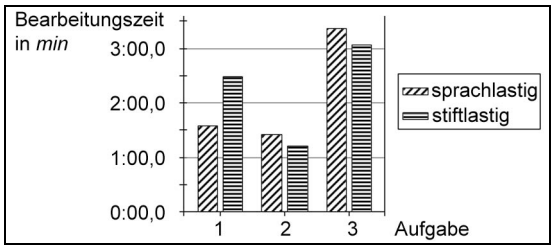

Abbildung 4: Bearbeitungszeiten bei stift- und sprachlastiger Interaktion

Die **Sprachbefehle** wurden von allen Versuchspersonen anhand ihrer Hervorhebung auf der grafischen Benutzungsschnittstelle sehr gut erkannt und gezielt zur Anwendungsbedienung genutzt. Die Analyse der Videoaufzeichnungen der Versuchspersonen P2 und P6, bei denen ein kontinuierlicher Anstieg des Sprachanteils zu verzeichnen ist, ergibt, dass zu Beginn ausschließlich Navigationsbuttons und erst später komplexe Interaktionselemente sprachgesteuert bedient wurden. Diese Versuchspersonen entwickelten schrittweise ein mentales Modell der sprachbasierten Programmsteuerung, indem sie zuerst die einfache mentale Verknüpfung eines Sprachkommandos mit dem Drücken eines grafischen Buttons verinnerlichten und sich darauf aufbauend die komplexere Sprachsteuerung von Dropdown-Boxen und Tabellen erschlossen. Die Auswertung der Interviews ergab, dass die Versuchspersonen die Vermittlung der Befehle als effektiv und die verwendeten Sprachkommandos überwiegend als angemessen und zweckmäßig empfanden. Sprachkommandos wurden jeweils durch ein **akustisches Feedback** quittiert, um dem Nutzer eine Rückmeldung über den Zustand des Spracherkennungssystems zu geben. Wurde ein Sprachkommado vom Spracherkennungssystems als gültiges Kommando identifiziert, so wurde dem Nutzer ein Feedback in Form eines kurzes Pieptons gegeben. Sofern ein Sprachkommando nicht als gültiges Kommando identifiziert wurde, wurde es mit einem doppelten Piepton zurück gewiesen. Die Videoaufzeichnungen dokumentieren, dass die Versuchspersonen bereits nach einer sehr kurzen Gewöhnungsphase korrekt auf das unmittelbar nach dem Sprechen eines Kommandos gegebenen Feedbacks reagierten. Blieb dieses aus, war den Versuchspersonen bewusst, dass das Kommando nicht als solches identifiziert wurde und das Kommando wurde erneut gesprochen. Alle Versuchspersonen verstanden ebenfalls das Feedback einer Rückweisung (doppelter Piepton). Ebenso problemlos haben die Nutzer die grafische somit eigentlich modal inkonsistente Reaktion auf ein erkanntes Sprachkommando erkannt. Problematisch war lediglich das Auftreten von Erkennungsfehlern, bei dem eine Funktion ausgeführt wurde, die vom Nutzer nicht gewollt war. Alle Nutzer benötigten beim ersten Auftreten eines Erkennungsfehlers eine Denkpause, um den Grund dieser unerwarteten Reaktion des Systems nachvollziehen zu können. Insgesamt wurde das Feedback, das im Anschluss an Sprachkommandos gegeben wurde, in den Interviews als verständlich eingestuft und der Zustand der Spracherkennung war den Benutzern dadurch stets transparent. Die Auswertung des Usability-Tests zeigt, dass die Versuchspersonen sehr gut mit der multimodalen Benutzungsschnittstelle zurechtkamen. Die alternative Programmsteuerung durch Sprachbefehle wurde von den Versuchspersonen sehr gut angenommen und effizient genutzt.

5 Ausblick

In diesem Beitrag wurden die Konzepte und Evaluationsergebnisse einer multimodalen Benutzungsschnittstelle für PDA-basierte Software vorgestellt. Die Metapher der hierarchischen Stapelkarten zur Strukturierung der Dialogabläufe wurde verwendet und auf jeder einzelnen Karte werden für eine bessere Übersichtlichkeit Informationen und Navigation räumlich getrennt dargestellt. Alle zur Verfügung stehenden Sprachkommandos waren ebenfalls auf der grafischen Benutzungsschnittstelle angegeben. Das vorgestellte Konzept wurde an einer PDA-basierten Clientanwendung für den Webshop einer Online-Buchhandlung erprobt. Diese Anwendung ist sehr stark durch den häufigen Wechsel zwischen Navigation

und Informationspräsentation geprägt. Die Übertragbarkeit und die Effizienz des vorgestellten Konzepts auf andere Anwendungsgebiete werden in weiteren Studien untersucht.

Literaturverzeichnis

Bajzat, R. (2005): GUI-Toolkit Thinlet. In: SourceForge.

Brewster, S. (1994): Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer-Interfaces. University of California, University of York.

Miller, G. A. (1956): The Magical Number Seven, Plus or Minus Two. In: The Psychological Review.

Pitts, I.; Edwards, A. (2002): Design of Speech-based Devices. London: Springer Verlag.

Sanders, M. S.; MacCormick, E. J. (1982): Human Factors In Engineering and Design. New York: McGraw-Hill.

Weinschenk, S.; Barker, D. T. (2000): Designing Effective Speech Interfaces. New York: John Wiley & Sons.

Kontaktinformationen

Oliver Stache: oliver.stache@t-systems.com
Dr. Hilko Donker: donker@inf.tu-dresden.de