

Usability von Voice User Interfaces im Anwendungsbereich Car-Sharing

Carsten Wittenberg
Martin I. Tietze
Peter Passon
Steffen Kötter
Kai-Christoph Hamborg

Universität Osnabrück
Seminarstr. 20
49069 Osnabrück
{cwittenb, mtietze, ppasson, stkoette,
khamborg}@uos.

Abstract

Der vorliegende Beitrag berichtet über Vorgehensweise und Ergebnisse eines einjährigen Studienprojektes. Ziel des Projektes war es, Möglichkeiten und Grenzen des Einsatzes eines Voice User Interfaces für das Anwendungsfeld Car-Sharing unter besonderer Berücksichtigung der Usability an Hand einer Prototypenentwicklung zu untersuchen.

Die Entwicklung basiert auf umfassenden Anforderungsanalysen sowie einem Usability-Test und den daraus hervorgegangenen Designvorschlägen.

Über die Entwicklung des Prototypen hinaus deuten die Ergebnisse des Projektes darauf hin, dass Voice Applikationen in dem Anwendungsbereich gut realisierbar sind.

Aufgrund der flexiblen und stabilen Erreichbarkeit können entsprechenden Anwendungen durchaus Chancen gegenüber bestehenden Web- und Call-Center-basierten Anwendungen eingeräumt werden, sofern Usabilityaspekte bei der Entwicklung berücksichtigt werden.

Keywords

VUI, Car-Sharing, Usability, Evaluation, Anforderungsanalyse

1.0 Einleitung

Sprachbasierte Dialogsysteme werden in den letzten Jahren vermehrt in Bereichen standardisierter Serviceleistungen eingesetzt. Eine besondere Herausforderung aktueller Entwicklungen stellt dabei das Vordringen in halbstandardisierte Bereiche dar. Die Domäne des Car-Sharing bietet hier einen hervorragenden Anwendungsbereich.

Derzeit wird dieser Service noch jeweils etwa zur Hälfte über das Internet und über Call-Center abgewickelt. Der Einsatz sprachbasierter Dialogsysteme scheint aber aufgrund der starken Schematisierung weiter Teilbereiche der Interaktion auf den ersten Blick möglich und sinnvoll. So könnte der Service zusätzlich auch außerhalb der Call-Center Betriebszeiten zur Verfügung stehen und ortsunabhängig, flexibel und kurzfristig genutzt werden.

2.0 Übersicht

In dem Projekt wurde für den Bereich Car-Sharing ein sprachbasiertes System unter besonderer Berücksichtigung der Usability entwickelt. Das System sollte es Nutzern erlauben, typische Car-Sharing Aufgaben durchzuführen (Buchungen vornehmen, verändern und stornieren). Ausgangspunkt der Entwicklung war ein bereits vorliegender lauffähiger Prototyp, der in einem vorangegangenen Projekt erstellt worden war, in dem Aspekte der Usability aber nicht im Vordergrund standen.

Der Prototyp verfügte über ein 'frame'-basiertes Design, d.h. das Dialogmanagement basiert auf einer begrenzten Menge von Slots. Diese können im Idealfall in beliebiger Reihenfolge gefüllt werden, ohne explizit jede mögliche Permutation modellieren zu

müssen. So soll ein natürlicher Dialogfluss durch die Möglichkeit erreicht werden, dass der Nutzer die Eingabe in intuitiver Reihenfolge vornehmen kann. Die Inhalte der Slots bestehen aus den zur erfolgreichen Durchführung der Aufgabe notwendigen Daten und sind somit domänenspezifisch. Zusammen mit einer Regelmenge, die sowohl die Interaktion der Slots untereinander als auch mit der Umwelt beschreibt, bilden beide den Rahmen (frame), der zusätzlich auch als Modell der Aufgabe fungiert, wie im Folgenden für den Anwendungsfall Car-Sharing dargestellt wird.

Die Domäne Car-Sharing kann hinreichend mit vier Slots, dem Start und Ende der Buchung, dem Standort und dem Fahrzeug, modelliert werden. Die Regeln stellen einerseits die Konsistenz der Slots sicher, zum Beispiel, dass der Start vor dem Ende liegen muss. Ande-

rerseits repräsentieren sie Einschränkungen, denen der Nutzer in der Domäne unterliegt, zum Beispiel die maximale Buchungsdauer.

Für die Evaluation erfolgte eine Re-Implementierung auf der Plattform der VoiceObjects AG. Dieses Voice Application Management System ermöglicht die effiziente Verwaltung von automatisierten Sprachanwendungen über deren gesamten Lebenszyklus. Der objektorientierte Ansatz ermöglicht die einfache Erstellung von Sprachanwendungen, sowie deren Analyse. Als Basis für eine kontinuierliche Optimierung können bei Bedarf offene Standards mit bereits bestehenden Systemen und Sprachplattformen integriert werden.

Im Rahmen des Projekts wurden eine umfassende Anforderungs- und Bedarfsanalyse sowie ein Usability-Test durchgeführt, die in eine Respezifikation des Prototypen mündeten. Die Anforderungs- und Bedarfsanalyse bestand aus drei Komponenten: Erstens wurde ein Experteninterview mit dem Vorstand eines Car-Sharing Verbandes geführt und Beobachtungen in einem Call-Center vorgenommen. Zweitens wurde eine Nutzerbefragung unter Car-Sharing Nutzern durchgeführt. Dies geschah mit Hilfe eines Fragebogens, der von den Nutzern wahlweise als Papierversion oder Online im Browser bearbeitet werden konnte, sowie durch sechs Nutzerinterviews. Drittens wurde eine Analyse von Buchungsdaten vorgenommen.

3.0 Ergebnisse der Analyse

Resultat der Experteninterviews und der Beobachtungen im Feld war die Beschreibung typischer Buchungsprozesse und hierbei auftretender Probleme. Der Ablauf der drei wesentlichen Buchungsvorgänge, ein Fahrzeug zu Buchen, eine aufgegebene Buchung zu verändern, oder diese zu stornieren, erwies sich als stark schematisiert. Als wichtiges Ergebnis aus den Beobachtungen ging

hervor, dass die Call-Center Agenten in Abhängigkeit von ihrer Berufserfahrung bemüht waren, die durch die Buchungssoftware starr vorgegebene Abfolge der Eingaben zu mildern und dem Anrufer eine prioritäts gelenkte Eingabeabfolge zu ermöglichen. Weiterhin konnten auch auftretende Probleme identifiziert werden, wie zum Beispiel Lärmbelästigung bei Telefongesprächen mit den Call-Center Agenten. Auch Problemfälle, die nicht direkt im Zusammenhang mit dem Buchungsprozess stehen, wie Probleme am Fahrzeug oder mit dessen Sicherheitssystem, wurden ausgemacht.

Die Auswertung der Fragebögen (N=162, Rücklaufquote = 64,8%) und der Nutzerinterviews vermittelte Einblicke in die spezifischen Erwartungen der Nutzer an diese Dienstleistung sowie Aussagen über Befürchtungen, die die Nutzer mit der möglichen Einführung eines Sprachsystems verbinden. Beispielsweise wünschen sich viele Nutzer das Angebot eines ähnlichen Fahrzeugs, wenn das gewünschte nicht verfügbar ist oder Informationen über die Lage der Stellplätze. Weiterhin wurden Aspekte wie Freundlichkeit und Fachkompetenz des Call-Center-Agenten, telefonische Erreichbarkeit der Buchungszentrale und Zügigkeit des Gesprächs von vielen befragten Nutzern als 'wichtig' klassifiziert. Explizit bemängelt wurden z.B. lange Warteschleifen und inkompetente oder gereizte Call-Center Agenten. Ähnliche Aspekte beinhalteten auch die Befürchtungen der Nutzer. Diese richteten sich vor allem auf längere Gesprächszeiten, insbesondere einer zu langen Einleitung, auf die Angst, bei Fehleingaben in Schleifen zu geraten und auf Probleme bei nicht standardgemäßen Anforderungen, sowie eine zu künstliche Stimme.

Die letzte Komponente der Analyse, die Analyse der Buchungsdaten, diente dazu, Nutzertypen zu klassifizieren

und deren Anforderungen weiter zu spezifizieren. Die Analyse ergab, dass sich unter den Car-Sharing Nutzern überwiegend regelmäßige Häufig- und Seltennutzer befinden. Adressaten des Dienstes sind also erstens Nutzer, die in kurzen regelmäßigen Abständen, im Bezug auf den bevorzugten Stellplatz, auf den Buchungszeitraum und auf das gebuchte Fahrzeug, gleiche oder sehr ähnliche Buchungen durchführen. Zweitens sind es Nutzer, die in größeren, variablen Intervallen ebenso ähnliche Buchungen durchführen.

4.0 Gestaltungsmaßnahmen

Aus den einzelnen Teilen der Analyse konnten Anforderungen an das Systemdesign erarbeitet und gewichtet werden. Des Weiteren ermöglichte die Analyse die Gestaltung authentischer Nutzungsszenarien und die Entwicklung von Evaluationskriterien für den Usability-Test. Als Gestaltungsanforderung resultierte aus der Analyse die Anpassbarkeit des Systems an verschiedene Nutzungsgewohnheiten durch Nutzermodi oder -profile.

Speziell für die Seltennutzer, die nur in größeren Abständen Buchungen aufgeben, aber auch für Erstnutzer sind die schnelle Erlernbarkeit, intuitive Buchungsabläufe und eine klare Nutzerführung wichtig. Für regelmäßige Nutzer hingegen erscheinen Vorschläge des Systems unter der Berücksichtigung der vorherigen Buchungen angebracht, um repetitive Eingaben zu vermeiden. Dies könnte durch eine entsprechende Nutzermodellierung erreicht werden, aus der Informationen wie Hauptstandort oder das meist gebuchte Auto eines Benutzers bezogen werden können.

Außer den Standardbuchungsvorgängen sollten auch Informationen, beispielsweise über die Position von Stellplätzen, oder Beschreibungen von Fahrzeugen verfügbar sein. Weiterhin sollte die Nutzung des Systems nicht mehr Zeit in

Anspruch nehmen, als die Bedienung des entsprechenden WWW-Interfaces.

5.0 Evaluation

Der aus der Analysephase abgeleitete Gestaltungsbedarf wurde durch die Prototypen-Evaluation mittels eines Usability-Tests ergänzt. Hierbei wurde ein kombiniertes Vorgehen aus Verhaltensbeobachtung und einem problemfokussiertem, videogestützten Interview realisiert. Die Teilnehmer waren 34 Studierende der Universität Osnabrück im Alter von 19-36 Jahren ($M=22,85$). Alle Teilnehmer verfügten über Erfahrung mit Computern und dem Internet. Nur eine Person gab an, Mitglied in einem Car-Sharing Verein zu sein.

Nach der Bearbeitung des demographischen Fragebogens und der Erhebung von Computervorerfahrung und Technikaffinität der Teilnehmer, bearbeiteten diese fünf Aufgabenszenarien. Die Aufgabenszenarien bildeten die identifizierten typischen Aufgaben ab: Buchung, Stornierung und Umbuchung, sowie die Verlängerung einer laufenden Buchung. Zudem wurde eine Buchung durchgeführt, bei der Straßenlärm eingespielt wurde, um Reaktionen des Systems unter dieser in der Analyse als relevant identifizierten Bedingung untersuchen zu können. Im Anschluss an die Aufgabenszenarien wurde die subjektive Bewertung des Systems erhoben. Dies geschah durch den AttrakDiff 2 Fragebogen (Hassenzahl et al., 2003) sowie eine übersetzte Fassung des Subjective Assessment of Speech System Interfaces (SASSI; Hone & Graham, 2000). Abschließend wurde das problemfokussierte Interview zur Identifizierung von Nutzungsproblemen durchgeführt.

6.0 Ergebnisse: Evaluation

Für die vier Skalen des AttrakDiff 2 (Hedonische Qualität I und S, Pragmatische Qualität und Attraktivität) wie für den SASSI-Fragebogen, dessen Faktor-

struktur sich jedoch nicht gut replizieren ließ, liegen die Ergebnisse im mittleren Skalenbereich. Bei der Buchung unter der Lärmbedingung traten nicht mehr Interaktionsprobleme mit dem System auf, als bei der Buchung ohne Lärm.

Die in dem Interview erhobenen 296 Problembeschreibungen wurden von zwei unabhängigen Beurteilern den folgenden induktiv gewonnenen Kategorien zugeordnet: 1. "sprachliche Promptgestaltung", 2. "Anordnung der Menüpunkte", 3. "fehlende Funktionen", 4. "Rückmeldung und Anleitung", 5. "Probleme bei der Sprachein- und -ausgabe", 6. "unklares Eingabeformat", 7. "Planungsunterstützung", 8. "allgemeine Aussagen zum System", 9. "Sonstiges".

Den Kategorien ließen sich die folgenden Hauptprobleme entnehmen: Zunächst entstanden Schwierigkeiten vor allem durch die schwer zu erschließende Menüstruktur und unklare Benennungen der Menüitems. 62 Problemnennungen entfielen auf die Kategorie "sprachliche Promptgestaltung" und 14 auf "Anordnung der Menüpunkte". Beispielsweise konnten die Items zur Umbuchung von gerade laufenden und in der Zukunft liegenden Buchungen häufig nicht korrekt identifiziert werden.

Weiter entstanden schwerwiegende Probleme durch unklare Eingabeformate und eine sehr starre Eingabemaske für viele Prompts. So akzeptierte das System zur Auswahl des Autos lediglich die Autoklasse, nicht aber das Fabrikat als zulässige Eingabe. In dieser Kategorie fanden sich 27 Problemnennungen. Zudem erschwerte eine zu starre und gleichförmige Fehlerbehandlung den Umgang mit dem System, welches auf nicht erkannten Eingaben lediglich mit einer Aufforderung zur Eingabewiederholung reagierte. Auch bei erfolgreicher

Eingabe wurde die Rückmeldung vom System als unzureichend bemängelt. Insgesamt wurden 57 Problemnennungen der Kategorie "Rückmeldung und Anleitung" und 36 der Kategorie "Planungsunterstützung" zugeordnet.

Weiterhin wurde das Fehlen von Navigations- und Steuerungsmöglichkeiten, die dem Nutzer vom Umgang mit GUIs vertraut sind wie z.B. Abbrechen, Zurück, Wiederholen und Weiter bemängelt.

7.0 Respezifikation

Die Respezifikation des Prototypen konzentrierte sich der Anforderungsanalyse folgend auf die Optimierung des Systems im Hinblick auf die beiden vorherrschenden Nutzergruppen. Die resultierenden Gestaltungsmaßnahmen zielten dem entsprechend darauf ab, den Nutzern eine möglichst umfassende Kontrolle über das System durch die genannten transparenten und flexiblen Navigationsmöglichkeiten zu geben. Außerdem sollte das Wording an die Gewohnheiten der Nutzer angepasst werden. Zum Beispiel wurde der Begriff 'Buchung' von den Nutzern als nicht eindeutig aufgefasst, so dass die Prompts für unerfahrene Nutzer nicht aussagekräftig genug waren.

Das Problem der zu gleichförmigen Fehlerbehandlung scheint für die untersuchte Domäne durch eine 'plan'- an Stelle einer 'frame'-basierten Modellierung der Dialogkomponente behebbar. Denn der Modellierungsaufwand steigt bei einem frame-basierten System selbst für diese relativ einfache Domäne stark an, wenn komplexere Dialogstrategien implementiert werden sollen. Als Lösung bietet sich daher an, eine bestehende 'plan'-basierte Architektur einzusetzen, die sowohl ein Modell des Dialogs als auch ein explizites Aufgabenmodell enthält und auf diesen Repräsentationen selbstständig Schlussfolgerungen ziehen kann. Sie ermöglicht es rela-

tiv einfach Regeln zu definieren, dass z.B. fehlende Slots mit Hilfe von Domänenkenntnis über die Aufgaben und Nutzerverhalten ergänzt, bzw. Eingaben durch Initiative des Systems unter Berücksichtigung regelmäßiger Buchungsmuster des Nutzers vorgeschlagen werden. Außerdem könnte ein derartiges System bei inkonsistenten Inhalten der Slots selbstständig Alternativen anbieten, z.B. dann, wenn ähnliche verfügbare Fahrzeuge oder alternative Stellplätze gefunden werden müssen.

Was die Nutzeranforderungen nach der Behandlung von weitergehenden Problemen angeht, erwiesen sich diese Gebiete als sehr schwer zu strukturierende Bereiche, welche die Komplexität des Systems unverhältnismäßig gesteigert hätten und somit zunächst nicht zu realisieren waren.

8.0 Diskussion

Das im Projekt verfolgte Vorgehen hat sich im Nachhinein als sehr hilfreich erwiesen. Die unterschiedlichen Analysezugänge gestatteten die Betrachtung der Domäne aus verschiedenen Perspektiven. So wurde die Gestaltung der Prozesse und Aufgabenszenarien durch Beobachtungen und Dokumentenanalysen unterstützt, die Identifikation der Nutzertypen durch die Analyse von Buchungsdaten und Fragebögen. Befürchtungen und Probleme der Nutzer wurden durch Befragungen erhoben. Durch diese Strukturierung des Analyseprozesses war es möglich, Redundanz zu vermeiden und die Analyse effektiv zu gestalten. Die Feldbeobachtungen wurden als offene teilnehmende Beobachtungen durchgeführt, bei denen sofort relevante Ergebnisse festgehalten wurden. In den halbstrukturierten Interviews der Analysephase wurden relevante Aussagen sofort notiert und auf Tonbandaufzeichnung und Transkriptionen verzichtet. Die im Rahmen des Usability-Tests durchgeführten Interviews waren halbstrukturiert

und problemfokussiert. Die Auswertung erfolgte per Paraphrasierung der Intervieweraufzeichnungen und anschließender Kategorisierung der Problemnennungen. Das realisierte Vorgehen ermöglichte es, weite Teile der Arbeitspakete parallel zu bearbeiten und Abhängigkeiten zwischen diesen zu vermeiden. Das Projekt zeigt somit, dass es auch mit wenigen Ressourcen möglich ist, recht umfassende Usability-Aktivitäten zu realisieren.

Für die Prototypenentwicklung wurde deutlich, dass sich diese vor der Erstellung eines lauffähigen Prototypen zunächst besser auf eine sorgfältige Spezifikation der Dialogkomponente stützen sollte, die auf einer Anotierung bestehender Dialoge zwischen menschlichen Teilnehmern (beispielsweise aus einem Call-Center) beruht und per 'Wizard of Oz' Technik evaluiert wird.

Wie berichtet wurde das Dialogsystem durch die verwendeten Fragebogen in einem mittleren Bereich, d.h. weder positiv aber auch nicht negativ bewertet. Dies ist insofern bemerkenswert, da es sich um einen Prototypen und noch nicht um ein ausgereiftes System handelte und in den Tests eine nicht unerhebliche Menge von Usability-Problemen identifiziert wurde. Ob es sich hier um eine zu geringe Sensibilität der verwendeten Instrumente in der Domäne der Sprach-Dialogsystemen handelt oder der recht positiven Bewertung andere Faktoren zugrunde liegen, sollte in weiteren Untersuchungen geklärt werden.

Der große Vorteil sprachbasierter Dialogsysteme liegt im Vergleich zu WWW-Interfaces in der mobilen Erreichbarkeit, die zudem im Vergleich zum Call-Center rund um die Uhr eine gleich bleibende Nutzerumgebung mit verläSSLicher Performanz bietet. Zudem stellen lange Warteschleifen und inkompetente oder unfreundliche A-

genten - wie von befragten Nutzern als negative Erfahrungen mit Call-Centern geäußert - für Dialogsysteme kein Problem dar. Ob sprachbasierte Dialogsysteme neben Call-Centern und WWW-Interfaces bestehen können oder sich gar gegen diese durchsetzen werden, hängt sicherlich zuallererst von der Usability der Systeme ab. In unserem Projekt deutete sich an, dass diese mit vorhandenen Gestaltungskonzepten für die Standardfälle realisierbar ist. Der Entwicklungsaufwand bei der Aufnahme von Rand- und Sonderfällen ist allerdings ohne differenziertere Modelle, wie beispielsweise aus der KI oder der Computerlinguistik, als zu hoch anzusehen.

Danksagung

Für die aufgeschlossene Kooperation möchten wir besonders der Firma VoiceObjects danken, Herrn Zielstorff des Ökostadt/teilAuto Hannover, sowie dem WHD Heidelberg, insbesondere Herrn Macke.

9.0 Referenzen

Hone, K. S. & Graham, R. (2000): Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). *Natural Language Engineering*, 6, 287-303.

Hassenzahl, M., Burmester, M. & Koller, F. (2003): AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: J. Ziegler & G. Szwillus (Hrsg.): *Mensch & Computer 2003. Interaktion in Bewegung* (S. 187-196). Stuttgart: B. G. Teubner.

»Es ist erlaubt digitale und Kopien in Papierform des ganzen Papers oder Teilen davon für den persönlichen Gebrauch oder zur Verwendung in Lehrveranstaltungen zu erstellen. Der Verkauf oder gewerbliche Vertrieb ist untersagt. Rückfragen sind zu stellen an den Vorstand des GC UPA e.V. (Postfach 80 06 46, 70506 Stuttgart). Proceedings of the 4th annual GC UPA Track Gelsenkirchen, September 2006 © 2006 German Chapter of the UPA e.V.«