

Ergebnisse einer initial-explorativen Studie zum Umgang von mittelalten und älteren Nutzern mit telefonbasierten Sprachdialogsystemen

Thomas Hempel
Siemens AG
Corporate Technology
User Interface Design
Otto-Hahn-Ring 6
81730 München
Thomas.Hempel@siemens.com

Abstract

Ob es ihnen bewusst ist oder nicht: eine zunehmende Anzahl von Verbrauchern hat im Alltag Kontakt mit modernen, telefongestützten Sprachdialogsystemen. Darunter finden sich zunehmend Personen mit sehr unterschiedlichen Lebens- und Erfahrungshintergründen.

Ziel dieser ersten explorativen Studie ist es, einen Teil dieser unterschiedlichen Erwartungen und Anforderungen hinsichtlich Bedienung und Persona-Design von Sprachdialogsystemen im Bereich der Informationssysteme mit Hilfe eines Usability-Tests zu dokumentieren.

Die Aufgaben, die die Studienteilnehmer im Usability-Test durchzuführen hatten, bestanden u. a. aus dem Aufspüren eines bestimmten Mobilfunktarifs im eigens gestalteten Sprachportal und die (sprach-) computergestützte Voranalyse eines simulierten Problems mit ihrer Internetverbindung, bevor schließlich eine Verbindung mit einem Service-Mitarbeiter hergestellt wurde. Das Sprachportal war speziell zur Vorklassifizierung solcher Nutzeranfragen ausgelegt worden. Es zeigt sich, dass Senioren erheblich mehr Schwierigkeiten bei der Bedienung von Sprachdia-

logsystemen zeigen als jüngere Erwachsene. Es besteht die Annahme, dass Senioren sich – im Gegensatz zu jüngeren Erwachsenen – stärker am bekannten, zwischenmenschlichen Dialog orientieren und ein vergleichsweise ungeeignetes Modell zur effektiven Bedienung von heutigen Sprachdialogsystemen besitzen. Es wird auf die Notwendigkeit nutzergruppenabhängiger Sprachdialoggestaltung hingewiesen.

Keywords

Sprachdialogsystem, Voice User Interface, Senioren, Akustik.

1.0 Einleitung

Der mittlerweile hohe Reifegrad heutiger Spracherkennungstechnologie hat zur Folge, dass Sprachtechnologie in Form automatisierter Standardanwendungen in eine zunehmende Zahl von Telefonielösungen integriert wird („Sprachportal“). Insbesondere eng umrissene Aufgaben bzw. sozial standardisierte Sprachdialoge bieten hier geeignete Einsatzmöglichkeiten (z.B. automatisierte Katalogbestellung, Ermitteln bzw. Wechseln von Telekommunikationstarifen, Kontoführung, Aktienhandel). Entsprechend breit und heterogen ist nun das Publikum, das mit dieser Art von Diensten konfrontiert wird. Aufgrund der Breite der genannten Anwendungen umfasst dies nicht nur junge Menschen, sondern auch ältere Mitbürger, aber auch Personen, für die der Umgang mit dem PC zum privaten oder beruflichen

Alltag gehört oder eben auch gerade nicht usw. Heute ist es immer noch Standard allen Anrufergruppen eines Sprachportals den gleichen Sprachdialog anzubieten. Anders ausgedrückt: Spezifische Gruppenmerkmale und deren entsprechende Reflexion im Sprachdialog werden bislang nicht berücksichtigt. Es besteht Grund zu der Annahme, dass ein solcher Einheitsdialog – trotz Breitenerfolgs – bei einigen Anrufergruppen zu einer deutlich geringeren Akzeptanz des Gesamtsystems führt als gewünscht (vgl. z.B. Hempel 2004, Stegmann et al. 2006).

Durch einschlägige Untersuchungen (z.B. Peissner et al. 2006) konnten bereits Akzeptanzkriterien von realen Sprachdialogsystemen identifiziert werden. Ziel der vorliegenden Studie war es nun, den Umgang unterschied-

licher Anrufergruppen mit einem einzigen telefonbasierten Sprachdialogsystem, das jedoch in unterschiedlichen Gestaltungsvarianten vorliegt, zu untersuchen, um hieraus Aussagen hinsichtlich der Gestaltung zukünftiger, zielgruppenangemessener Dialoge abzuleiten. Da zu diesem Thema äußerst wenig Literatur existiert und diese Studie daher keineswegs erschöpfend sein kann, sollen die gefundenen Ergebnisse dabei insbesondere potenzielle Herausforderungen für weitere Untersuchungen auf diesem Gebiet benennen und entsprechende Diskussionen stimulieren.

Die Untersuchung ist Teil einer Kooperation zwischen der *Siemens AG* und der *Deutsche Telekom AG*.

2.0 Durchführung der Untersuchung

2.1 Teilnehmer

An der Untersuchung nahmen 25 Personen teil. Jeweils 5 Teilnehmer entfielen auf die Gruppen Jugendliche (<15J), männliche Erwachsene, weibliche Erwachsene, sowie männliche bzw. weibliche Senioren (>62J).

Wegen der bewusst heterogenen Zielgruppe wurde der Test sehr „lebensnah“ durchgeführt, d.h. beispielsweise auf eine übliche Vorauswahl der Teilnehmer in punkto Hörverlust bewusst keine Rücksicht genommen. Alle Teilnehmer wurden jedoch gebeten, Ihnen bekannte Beeinträchtigungen des Gehörs mitzuteilen (z.B. Benutzung eines Hörgeräts, Tinnitus). Beeinträchtigungen wurden schließlich von einer weiblichen Erwachsenen und immerhin von drei von fünf männlichen Senioren berichtet.

2.2 Technikerfahrung

Im Test wurde von der Hypothese ausgegangen, dass der Grad der Erfahrungen, die die Teilnehmer im Umgang mit unterschiedlichen technischen und elektronischen Geräten in ihrem Leben gesammelt haben, einen Einfluss auf die Erfolgsrate im Umgang mit dem Sprachdialogsystem aufweist. Daher wurden die Teilnehmer gebeten Art und Nutzungshäufigkeit verschiedener technischer Geräte anzugeben.

2.3 Ort und Zeitraum der Studie

Die Untersuchung wurde im Frühjahr 2006 in München (Deutschland) durchgeführt. Alle Aussagen der Teilnehmer entstammen also diesem kulturellen und sozialen Umfeld. Eine Übertragung von Ergebnissen in andere Kulturkreise ist daher nicht ohne weiteres zulässig.

2.4 Aufgaben

Die Teilnehmer bearbeiteten plausible und alltagsnahe Aufgaben. Dazu gehörten z.B. das Ermitteln unterschiedlicher Telefonarife unter Benutzung eines automatischen Sprachportals sowie die eigenständige Vorermittlung eines fingierten Fehlers der heimischen Internetverbindung mit Hilfe des automatisierten Sprachdialogsystems.

Das dabei verwendete Sprachdialogsystem wurde von der Deutschen Telekom aufgesetzt. Es entstand auf der Grundlage verschiedener Voruntersuchungen, die im Rahmen einer Kooperation zwischen der *T-Systems* und dem *User Interface Design Center* der *Siemens Corporate Technology* durchgeführt wurden.

3.0 Ergebnisse

3.1 Ausfallrate bei Testteilnehmern

Routinehalber wurden Abbruchkriterien für den Usability-Test definiert. So z. B., wenn der Teilnehmer sichtbar und außergewöhnlich gestresst reagiert, so dass er durch einfache Mittel wie beispielsweise Pausen, Getränke oder andere Ablenkungsversuche während des Tests nicht mehr hinreichend zu beruhigen war.

Die durch obiges Abbruchkriterium im Test notwendigen Abbrüche traten ungewöhnlich häufig auf, aber variierten stark zwischen den Teilnehmergruppen. Während bei den Erwachsenen beiderlei Geschlechts keine Ausfälle zu verzeichnen waren, war diese Zahl bei den Senioren sehr groß¹:

- Alle 10 Erwachsenen haben den Test erfolgreich zu Ende geführt.

- Von den Senioren haben 6 von 10 den Test erfolgreich zu Ende geführt.

3.2 Gründe für den Abbruch

Von den Teilnehmern, die den Test nicht erfolgreich zu Ende geführt haben, haben alle Ihren Unmut über das System auch entsprechend ungefragt geäußert.

Die nachträgliche Analyse zeigt, dass gerade viele Senioren bei der Bedienung des Systems vergleichsweise lange Sprachäußerungen gemacht haben, sich manchmal während des Sprechens auch noch vergangene Angaben korrigiert haben – ein Umstand, der vom System nur selten mit Erfolg quittiert werden konnte. Auch jüngere Erwachsene haben anfangs oft längere Äußerungen getätigt, ihr Verhalten jedoch innerhalb der ersten Systemprompts an einen entsprechend erfolgreicherem Kommandostil angepasst, während viele der älteren Teilnehmer ihren Stil nicht geändert haben.

Auch verhielten sich einige Senioren sprachlich deutlich höflicher als andere Teilnehmergruppen und ihnen widerstrebt es geradezu einen sprachlichen Kommandostil zu verwenden. Einige haben sich dem System gegenüber auch mit Namen vorgestellt. Bei Jugendlichen und Erwachsenen konnte dies hingegen in keinem Fall beobachtet werden.

Trotz dieser gruppenspezifischen Unterschiede gaben alle Teilnehmer an, bereits vor diesem Test automatische Sprachdialogsysteme benutzt zu haben.

3.3 Persona: Zum Geschlecht und Alter des Systemsprechers

Vorausgehende hausinterne Voruntersuchungen gaben bereits Anlass zu der Vermutung, dass es seitens der Nutzer im Allgemeinen keine statistisch belastbare Aussage hinsichtlich der Be-

¹ Selbstverständlich wurden die entsprechenden Teilnehmer davon nicht in Kenntnis gesetzt.

vorzugung einer entweder männlichen oder weiblichen Stimme gibt. Dies konnte in der vorliegenden Untersuchung bestätigt werden.

Wie sich ebenfalls in vorausgehenden Untersuchungen andeutete, ist die Vorstellung der Teilnehmer hinsichtlich des Alters des Systemsprechers deutlich ausgeprägter:

- 43% der Teilnehmer bevorzugten ein Alter des Sprechers von 21-40 Jahren,
- 23% der Teilnehmer bevorzugten ein Alter des Sprechers von 41-59 Jahren.
- Ein Alter über 60 Jahre wurde im Test nie gewünscht, auch nicht von (gleichaltrigen) Senioren.

3.4 Technikerfahrung

Als ein einfaches Maß für die Technikerfahrung eines Teilnehmers wurde die Anzahl technischer Geräte, die ein Teilnehmer besitzt, aus einer vorbereiteten Liste ausgezählt und deren Nutzungshäufigkeit vom Teilnehmer mit „0“ für „nutze ich eigentlich nicht“, „1“ für „nutze ich gelegentlich“ und „2“ für „nutze ich täglich“ bewertet. Die Summe aus der Anzahl aller Geräte und deren Nutzungshäufigkeit wurde als grober Index zur Klassifizierung von Technikerfahrung angesehen. Der Index variierte im Test über alle Teilnehmer von 10-29.

Der Mittelwert der so berechneten Technikerfahrung und die entsprechende Anzahl Testabbrüche betrug für die einzelnen Gruppen:

- männliche Erwachsene: 23; 0 Testabbrüche
- weibliche Erwachsene: 24; 0 Testabbrüche;
- männlichen Senioren: 19; 1 Testabbruch
- weiblichen Senioren: 14; 3 Testabbrüche.

Aus den Daten zu Testabbrüchen und Technikerfahrung lässt sich also

möglicherweise die Hypothese ableiten, dass eine im Mittel geringere Technikerfahrung eine höhere Wahrscheinlichkeit für auftretende Schwierigkeiten im Umfang mit dem getesteten Sprachdialogsystem bedeutet. Oder andersherum ausgedrückt: Je technikerfahrener Teilnehmer sind, desto eher kommen sie mit dem vorliegenden Sprachdialogsystem zu recht.

Zu einer statistisch belastbaren Aussage zu dieser Hypothese sind natürlich weiterführende Arbeiten nötig.

4.0 Diskussion

Es war Ziel der vorliegenden initial-explorativen Studie verschiedene Nutzergruppen bei der Interaktion mit modernen Sprachdialogsystemen zu beobachten und in einem nachfolgenden Schritt die Möglichkeiten zielgruppenabhängiger Dialoge zu untersuchen und damit der Community weitere Forschungsfragen zu stellen. Da die Untersuchung auf Breite hin angelegt war, (z.B. verschiedenste Nutzergruppen, Interaktion mit dem Sprachdialogsystem so natürlich wie möglich), war es nicht möglich alle Umgebungsgrößen im Test geeignet zu kontrollieren. D.h., die gefundenen Ergebnisse gelten streng genommen nur für die vorliegende Untersuchung unter den beschriebenen Bedingungen und lassen sich ohne Validierung nicht in andere Zusammenhänge übertragen. Auch ist die Teilnehmeranzahl für statistische Zwecke natürlich nicht ausreichend.

Die Studie legt die Vermutung nahe, dass die Senioren im Test eine niedrigere Technikerfahrung aufweisen als die anderen Teilnehmergruppen. Es wird die Hypothese formuliert, dass diese niedrigere Technikerfahrung bei dieser Personengruppe zu einem feh-

lenden bzw. ungeeigneten mentalen Modell zur Bedienung von Sprachdialogsystemen führt. Dieser Umstand könnte die überdurchschnittliche Anzahl von Testabbrüchen für diese Personengruppe erklären.

Übertragen auf ein zukünftiges reales System wäre also zu überlegen inwiefern eine Vorklassifizierung von Anrufern z.B. nach Alter wünschenswert wäre, um hier mit entsprechend gestalteten automatischen Dialogen oder weiteren geeigneten Maßnahmen eine höhere Kundenzufriedenheit zu erzielen.

5.0 Schlussfolgerungen

Telefonbasierte Sprachdialogsysteme besitzen bislang große Potenziale für Anbieter wie Nutzer: z.B. wirtschaftliche Geschäftsmodelle für Betreiber und hohe Kundenzufriedenheit für definierte Anwendungsszenarien. Einer adaptiven, zielgruppenangemessenen Sprachdialoggestaltung könnte dabei in Zukunft gerade bei Systemen mit heterogener Benutzergruppe eine Schlüsselrolle zukommen.

Danksagung

Die Studie erfolgte im Auftrag des Siemens COM Chief Technology Office und ist Teil einer Kooperation zwischen der Siemens AG und der Deutschen Telekom AG.

6.0 Referenzen

Hempel, T. (2001): Usability von Sprachapplikationen – Woran hakt's? *Forum Ware*, 32(1-4), 24-26.

Stegmann, J.; Burkhard, F.; Oberle, F.; Ecker, M.; Englert, R.; & Müller, C. (2006): Einsatz der Sprecherklassifizierung in Sprachdialogsystemen. 7. *ITG-Fachtagung Sprachkommunikation*. Kiel: Christian-Albrechts-Universität.

Peissner, M.; Biesterfeld, J.; & Heidmann, F. (2004): *Akzeptanz und Usability von Sprachapplikationen in Deutschland*. Stuttgart: Fraunhofer IRB Verlag.

»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.«



