Matthias Peissner

Ist Sprachinteraktion heute schon intuitiv?

Is Speech Interaction Intuitive Today?

Usability_Akzeptanz_Sprachinteraktion_Sprachapplikationen_Voice User Interface Design

Zusammenfassung. In einer empirischen Studie werden die Akzeptanz und Usability (Gebrauchstauglichkeit) von telefonbasierten Sprachapplikationen in Deutschland untersucht. Eine umfangreiche telefonische Befragung liefert Daten zur Nutzung und Akzeptanz von Sprachapplikationen. Detaillierte Erkenntnisse zur Gebrauchstauglichkeit aktueller Systeme und zu den relevanten Qualitätsaspekten der sprachlichen Benutzungsschnittstelle stammen aus Usability-Tests mit sechs ausgewählten deutschen Sprachapplikationen. Dieser Beitrag gibt einen Überblick über die wichtigsten Ergebnisse der Studie.

Summary. An empirical study investigates the acceptability and usability of telephone-based speech applications in Germany. Extensive telephone interviews provide data about the usage and acceptance of speech applications. Detailed insights into the usability of current systems and into the relevant quality aspects of the voice user interface stem from usability tests with six selected German speech applications. This paper presents an overview of the most important results of the study.

1. Einleitung

Sprache bietet viele Vorteile für die Kommunikation zwischen Mensch und Technik. Insbesondere für die Bereitstellungen von Informationen und Dienstleistungen über das Telefon ist Sprache eine attraktive Interaktionsform. Moderne Technologien und Standards ermöglichen heutzutage den Einsatz von effektiven Sprachapplikationen in vielen Bereichen. Seit einigen Jahren sagen Marktforscher dem kommerziellen Einsatz von Sprachtechnologien ein starkes Wachstum voraus (vgl. Kelsey Group 2002; Datamonitor 2003; Joscelyne & Lockwood 2003; Forrester 2000; Raebel 2002). Deutschland befindet sich dabei auf einer aussichtsreichen Position, großen Anteil an dieser Entwicklung zu haben (vgl. Joscelyne & Lockwood 2003). Dennoch werden telefonbasierte Sprachapplikationen im deutschsprachigen Raum weitaus weniger angeboten und genutzt als es die wirtschaftlichen und technologischen Potenziale erwarten lassen.

Eine im Oktober 2003 durchgeführte Studie des Fraunhofer Instituts für Arbeitswirtschaft und Organisation (IAO) zeigt, dass wesentliche Gründe hierfür im unzureichenden Bedienkomfort der sprachgesteuerten Systeme sowie in mangelnder Nutzungsbereitschaft und negativen Vorbehalten der potenziellen Benutzer zu suchen sind (Peissner, Biesterfeldt & Heidmann 2004). Die Studie liefert empirische Daten zur aktuellen Nutzung, Qualität und Akzeptanz von telefonbasierten Sprachapplikationen in Deutschland. Dabei umfasst der Begriff der Sprachapplikation Systeme mit akustischer Eingabe und akustischer Ausgabe. Tastengesteuerte Systeme (DTMF) sind ausgeschlossen. Die Studie nimmt stets die Perspektive des potenziellen Benutzers ein.

2. Die Benutzer – Ergebnisse einer telefonischen Befragung

Eine genaue Kenntnis der speziellen Nutzungsbedingungen und relevanten Eigenschaften der zukünftigen Benutzer ist unabdingbar für die ergonomische Gestaltung eines interaktiven Systems. Dies gilt in besonderem Maße, wenn neuartige Technologien eingesetzt werden und die Benutzer nicht auf umfangreiche Kenntnisse und Erfahrungen aus dem Umgang mit vergleichbaren Produkten zurückgreifen können. Als Grundlage für eine gezielte Analyse der Nutzeranforderungen eines speziellen Systems sind allgemeine Erkenntnisse über relevante Erwartungen und Erfahrungen der potenziellen Benutzer bezüglich Sprachapplikationen interessant.

Um den Bekanntheitsgrad von Sprachapplikationen in Deutschland zu erheben und um Nutzungsverhalten sowie typische Einstellungen potenzieller und tatsächlicher Benutzer zu erfassen, wurden im Oktober 2003 computerunterstützte telefonische Interviews (CATI) mit insgesamt 1005 Personen durchgeführt. Die Stichprobe ist repräsentativ für die deutsche Gesamtbevölkerung in Bezug auf Alter, Geschlecht und Bundesland. Die CATI wurde vom Fraunhofer Institut für Arbeitswirtschaft und Organisation (IAO) initiiert, konzipiert und

ausgewertet. Die Durchführung übernahm ein Marktforschungsinstitut.

2.1 Bekanntheitsgrad und Nutzungshäufigkeit

Knapp die Hälfte aller Befragten (46 %) gibt an, Sprachapplikationen nicht zu kennen. Nur 32 % können mindestens eine Applikation namentlich nennen. Diejenigen, die Sprachapplikationen nutzen, tun dies eher sporadisch. Die meisten von ihnen (40 %) geben an, im letzten halben Jahr weniger als fünfmal eine Sprachapplikation genutzt zu haben.

2.2 Wer benutzt Sprachapplikationen?

Die meisten Befragten, die angeben, bereits Sprachapplikationen genutzt zu haben, sind zwischen 25 und 45 Jahre alt (53,5 % der Nutzer). Die Gruppe der Nutzer unterscheidet sich von der bundesdeutschen Gesamtbevölkerung außerdem durch bessere Computerkenntnisse und durch einen höheren Anteil an Handybesitzern.

2.3 Zufriedenheit der Nutzer

Mehr als jeder Vierte (76 bzw. 26 %) der 289 Nutzer gibt an, „sehr unzufrieden" mit der von ihm am häufigsten genutzten Sprachapplikation zu sein. Nur gut jeder Dritte (102 bzw. 35 %) ist „sehr zufrieden" oder „eher zufrieden".

2.4 Bevorzugung alternativer Zugänge

Können Informationen und Dienstleistungen über andere Medien (wie z. B. über eine Internetseite, per E-Mail oder über eine grafische Benutzungsschnittstelle auf dem Mobiltelefon erreicht und genutzt werden), werden diese in der Regel der Sprachinteraktion vorgezogen.

2.5 Einstellungen und Erwartungen gegenüber Sprachapplikationen

Die detaillierte Betrachtung kritischer Einstellungen und Erwartungen tatsächlicher und potenzieller Nutzer kann als Ausgangspunkt genutzt werden, um mangelnde Nutzungszufriedenheit und Nutzungsbereitschaft besser zu verstehen. Einerseits können erste Hinweise daraus abgeleitet werden, in welchen Bereichen der Sprachapplikationen besonderer Verbesserungsbedarf besteht. Andererseits wird klar, welche Vorurteile in einer positiven Marktkommunikation und Öffentlichkeitsarbeit besonders berücksichtigt werden sollten. Insgesamt fällt die Bewertung verschiedener Eigenschaften von Sprachapplikationen wie etwa „übersichtlich", „einfach", „bequem" sehr heterogen und tendenziell eher positiv aus. Zwei der abgefragten Eigenschaften werden jedoch besonders schlecht bewertet: Die Befragten sind der Meinung, heutige Sprachapplikationen könnten eher nicht auf individuelle Bedürfnisse eingehen. Zweitens wird der Spaßfaktor als mangelhaft eingestuft, insbesondere von denjenigen, die bereits erste Nutzungserfahrungen sammeln konnten.

2.6 Einstellungen gegenüber Unternehmen, die Sprachapplikationen verwenden

Einige Unternehmen fürchten einen Imageverlust durch den Einsatz von Sprachapplikationen. Eine Technologie, die nicht perfekt funktioniert, könnte dem Kundenvertrauen schaden und damit nachhaltige Risiken für Kundenbindung und Marktstärke mit sich bringen. Tatsächlich stellt sich dieser Zusammenhang bei weitem nicht derart eindeutig dar. Zwar scheint der Einsatz von Sprachapplikationen auch mit negativen Folgen für das Firmenimage verbunden zu sein (weniger „einladend" und „persönlich"). Doch zeigt sich insbesondere, dass Unternehmen, die Sprachapplikationen einsetzen, als „innovativ" und „professionell" wahrgenommen werden.

3. Aktuelle Sprachapplikationen – Qualität und Erfolgsfaktoren

Um einen repräsentativen und zugleich detaillierten Überblick über die Qualität deutscher Sprachapplikationen zu erlangen, wurden im Oktober 2003 sechs ausgewählte Applikationen in Usability-Tests mit insgesamt 30 potenziellen Benutzern analysiert:

- **Berti:** Informationen zur Ersten Fußball-Bundesliga. 0 91 31/61 00 17
- **EVA:** Informationen wie News, Wetter, Verkehr etc. und Service-Angebote wie Mailbox, Kundenverwaltung und Adressbuch. 11 44, nur für e-plus-Kunden
- **Fahrplanauskunft der Bahn:** Abfrage von Zugverbindungen. 08 00/ 1 50 70 90
- **FAZ-Nachrichten:** Nachrichten zu verschiedenen Rubriken wie Sport, Politik, Wirtschaft, Kultur, Börse etc. 01 90/15 14 44
- **Postbank-Telefonbanking:** Einfache Transaktionen und allgemeine Information zu Wertpapieren oder Postbank-Produkten. 01 80/3 04 07 00
- **Quellefon:** Verschiedene Dienste des Quelle Versandhauses: Bestellung von Katalogartikeln, 01 80/5 33 66 22, Bestellung von Katalogen, 01 80/ 3 23 25 24, und Informationen zum persönlichen Kundenkonto, 01 80/ 3 68 78 88

Diese Auswahl bildet das Spektrum der zum Untersuchungszeitpunkt verfügbaren Sprachapplikationen repräsentativ ab. Vier der sechs getesteten Systeme können aufgrund der Ergebnisse der vorher durchgeführten Telefonumfrage zu den am häufigsten genutzten deutschen Applikationen gezählt werden. Außerdem sind die wichtigsten aktuellen Interaktions- und Dialogparadigmen vertreten (enge Menüführung bis hin zur Möglichkeit der nahezu freien Spracheingabe in ganzen Sätzen) und möglichst viele relevante Geschäftsfelder abgebildet.

Manche der getesteten Applikationen sind seit Durchführung der Usability-Tests im Oktober 2003 weiterentwickelt worden. Dadurch müssen die Testergebnisse teilweise relativiert werden.

3.1 Allgemeines Qualitätsniveau und negative Vorbehalte

Die sechs Sprachapplikationen, die in den Usability-Tests näher analysiert wurden, weisen sehr unterschiedliche Qualitätsniveaus auf. Besonders positiv werden die beiden Applikationen „Quellefon" und „Berti" bewertet.

Insgesamt scheinen die getesteten Sprachapplikationen besser zu sein als die allgemein verbreiteten Vorurteile und Vorbehalte es vermuten lassen. Die meisten Probanden haben nach den Tests ein positiveres Bild von Sprachapplikationen als vorher. Die Bewertung wesentlicher Eigenschaften wie Über-

sichtlichkeit, Anpassung an individuelle Bedürfnisse, bequeme Handhabung etc. fällt in den meisten Fällen nach der Benutzung positiver aus als vorher.

3.2 Qualitätskriterien

Die wichtigsten Problemfelder und Erfolgsfaktoren, die für die Schaffung von Akzeptanz relevant sind, können wie folgt beschrieben werden (vgl. Ballentine & Morgan 1999; Cohen, Giangola & Balogh 2004; Gardner-Bonneau 1999; Nass & Gong 2000; Weinschenk & Barker 2000; Yankelovich, Levow & Marx 1995):

Angebot an Inhalten und Funktionalitäten

Ein Kosten-Nutzen-Verhältnis, das dem Kunden vorteilhaft erscheint, ist grundlegende Motivation für die Nutzung einer Sprachapplikation. Um gegen alternative Informations- und Serviceangebote bestehen zu können, ist es entscheidend, die Potenziale und inhärenten Probleme der Sprachinteraktion zu erkennen (vgl. Peissner, Heidmann & Corves 2003) und auf dieser Grundlage gezielt geeignete Nutzergruppen und Nutzungssituationen zu identifizieren. Die heute verfügbaren Sprachapplikationen profitieren von einem klaren und thematisch eng begrenzten Inhaltsbereich. Komplexe Applikationen mit heterogenen Inhalten, wie z. B. EVA oder FAZ, erschweren die Orientierung und führen schnell zu kognitiver Überbelastung der Benutzer.

Automatische Spracherkennung

Angesichts aktueller Möglichkeiten sind die Spracherkennungsquoten der getesteten Applikationen zum Teil sehr enttäuschend. Nur zwei der sechs Applikationen erreichen insgesamt eine Erkennungsquote von über 80 %. Zwar ist eine zuverlässig funktionierende Spracherkennung ein wichtiges Merkmal einer guten Sprachapplikation; ihre Bedeutung für die Akzeptanz und Benutzungsfreundlichkeit wird jedoch meistens stark überschätzt (vgl. Peissner 2002). Der Einfluss der Spracherkennungsqualität auf die subjektiv empfundene Bedienqualität ist bei den verschiedenen Testapplikationen unterschiedlich ausgeprägt. Während bei Berti der Zusammenhang zwischen subjektiver Anstrengung und Spracherkennungsperformanz recht stark ist (Bestimmtheitsmaß $R^2 = 0.68$), besteht z. B. bei Quellefon und Postbank überhaupt kein Zusammenhang in dem Sinn, dass die Benutzer bei einer besseren Spracherkennung über ein geringeres Maß an Anstrengung berichteten.

Management von Erkennungsfehlern

Eine der wichtigsten Aufgaben des Voice User Interface Designs ist es, die Schwächen der automatischen Spracherkennung (ASR) zu kompensieren und die negativen Folgen von Spracherkennungsfehlern zu minimieren. Zunächst muss sichergestellt werden, dass der Benutzer das Auftreten eines Erkennungsfehlers bemerkt. Die Korrektur von Fehlern muss schnell und ohne „Umwege" möglich sein. Außerdem ist es wichtig, dass Erkennungsfehler keine schwerwiegenden negativen Konsequenzen wie etwa Datenverlust nach sich ziehen. Schließlich kann das Was und Wie der akustische Ausgaben beim Anrufer Akzeptanz für das Auftreten von Erkennungsfehlern schaffen und so ein kooperatives Nutzerverhalten fördern.

Unterstützung bei der Formulierung zielführender Benutzeräußerungen

Bei den meisten Sprachapplikationen ist der Kunde auf eine recht kleine Menge zulässiger Benutzeräußerungen festgelegt. Um den Benutzer dennoch sicher und effizient an sein Ziel zu führen, müssen die akustischen Ausgaben so gestaltet sein, dass dem Anrufer der Funktionsumfang klar wird, dass er also weiß, was er sagen kann und wie er sprechen muss. Dabei sind explizite Erläuterungen oft wesentlich langwieriger und weniger effektiv als eine implizite Beeinflussung der Sprachperformanz des Benutzers durch entsprechende Vorbildfunktion der akustischen Systemausgaben. Die hinterlegten Benutzeräußerungen müssen möglichst intuitiv sein, dem alltäglichen Sprachgebrauch der Zielgruppen entsprechen und leicht erlernbar, leicht erinnerbar und leicht sprechbar sind. Zugleich ist darauf zu achten, dass die Schwächen und Stärken der verwendeten ASR berücksichtigt werden, um eine optimale Erkennung zu begünstigen.

Navigationsmöglichkeiten und Dialogabläufe

„Mixed Initiative"-Dialoge zeichnen sich dadurch aus, dass geübte Benutzer bekannte Funktionen direkt aufrufen können, ohne lange Menüs durchlaufen zu müssen. Wenn der Benutzer jedoch nicht alle benötigten Informationen angibt, die Interaktion ins Stocken gerät (wiederholte Zurückweisung, kein Benutzerinput) oder wenn die Hilfefunktion aufgerufen wird, übernimmt das System die Initiative, präsentiert mögliche Benutzeräußerungen und führt so den Benutzer durch den Dialog. Nur ein derart flexibles Dialogdesign kann den Bedürfnissen heterogener Benutzergruppen gerecht werden und damit auch langfristig Akzeptanz und Kundenbindung schaffen. Außerdem sollte sich der Benutzer stets in der Lage sehen, den Dialog nach seinen Vorstellungen und Bedürfnissen zu kontrollieren und zu steuern, z. B. durch die Möglichkeit, mit sprachlicher Eingabe die Systemausgaben zu unterbrechen (Barge In). Schließlich sollten die wichtigsten und häufigsten Nutzungsszenarien besonders schnell und komfortabel bedienbar sein.

Akustische Informationsdarstellung

Der wichtigste Grundsatz für die akustische Ausgabe gesprochener Informationen ist: „Fasse dich kurz!". Gesprochene Texte müssen deutlich einfacher konstruiert sein als Schrifttexte: Die Sätze müssen kurz sein, die Satzkonstruktionen übersichtlich. Die Bezüge zwischen den Sätzen müssen klar sein. Außerdem müssen die verwendeten Begriffe, die Bezeichnungen spezifischer Funktionen und deren Oberbegriffe eindeutig und verständlich sein sowie direkt der Erfahrungswelt des Benutzers entsprechen. Auch nonverbale Gestaltungsaspekte sind von Bedeutung: Die akustische Tonqualität beeinflusst die Verständlichkeit, Earcons (nonverbale Audios – die akustische Entsprechung von Icons) und Hintergrundgeräusche können zur besseren Orientierung beitragen.

Anmutung und emotionale Wirkung

Das „Hear & Feel" einer Sprachapplikation ist nicht nur ein entscheidender emotionaler Faktor, sondern beeinflusst auch das Dialogverhalten des Benutzers und darüber schließlich Erfolg und Effizienz der Interaktion. Wichtigster Grundsatz

dabei ist, dass insgesamt ein in sich stimmiges System entsteht: Nutzergruppen, Anwendungsdomäne, Serviceprovider, Persönlichkeit, Stimme, Dialogansatz und Spracherkennungsqualität müssen zusammen passen.

4. Ausblick

Die dargestellte Studie veranschaulicht den aktuellen Stand von Sprachapplikationen in Deutschland. Bekanntheitsgrad, Nutzungshäufigkeit und Nutzungszufriedenheit sind noch relativ gering. Neben umfassender Öffentlichkeitsarbeit und Marketingkampagnen werden insbesondere gute Beispielapplikationen benötigt. Zwar scheinen die aktuellen Sprachapplikationen etwas besser als ihr Ruf zu sein, den hohen Anforderungen einer benutzergerechten Dialogführung entsprechen sie jedoch kaum. Die dargestellten Qualitätskriterien können als Ausgangspunkt einer ergonomischen Gestaltung dienen. Sie ersetzen jedoch nicht den gezielten Einsatz von Methoden des Usability Engineering, um in einem iterativen und benutzerzentrierten Entwicklungsprozess zu einer hochwertigen Sprachapplikation zu gelangen, die breite Zustimmung und Akzeptanz finden wird. Dies wird in der Zukunft eine stärkere und systematische Einbeziehung von Usability-Experten und User Interface-Spezialisten in die Entwicklung von Sprachapplikationen erfordern. Die Gestaltung und Evaluation sprachlicher Benutzungsschnittstellen stellen neue und spezifische Herausforderungen an die Fachkompetenzen des Usability Engineers. Jedoch wird die Beschäftigung mit speziellen Problemen der akustischen Mensch-Technik-Interaktion dazu beitragen, auch die Arbeit mit GUIs von einer neuen Perspektive zu betrachten. Lösungsansätze und Evaluationsaspekte können zum Teil direkt übertragen werden und bereichern so die tägliche Arbeit.

Literatur

Ballentine, B.; Morgan, D.: *How to Build a Speech Recognition Application. A Style Guide for Telephony Dialogues.* CA: Enterprise Integration Group, Inc., San Ramon, 1999.

Cohen, M.; Giangola, J.; Balogh, J.: *Voice User Interface Design.* Boston: Addison-Wesley, 2004.

Datamonitor: Voice Automation: Past Present and Future. http://www.speechtechmag.com/whitepapers/Intervoice.pdf, (Letzter Zugriff: 30.04.2004)

Forrester Research: The Web's Speech Impediments. http://www.forrester.com/ER/Research/Report/Summary/0,1338,9209,FF.html (Letzter Zugriff: 30.04.2004)

Gardner-Bonneau, D. (Ed.): *Human Factors and Voice Interactive Systems.* Norwell, MA: Kluwer Academic Publishers, 1999.

Joscelyne, A.; Lockwood, R.: *Benchmarking HLT progress in Europe.* Kopenhagen: EUROMAP Language Technologies 2003.

The Kelsey Group: Speech Technology revenues to reach $2 billion by 2006 according to the Kelsey Group. http://www.kelseygroup.com/press/-pr020430.htm (Letzter Zugriff: 30.04.2004)

Nass, C.; Gong, L. Social aspects of speech interfaces from an evolutionary perspective: Experimental research and design implications. *Communications of the ACM* **43**, No. 9 (2000) pp. 36-43.

Peissner, Matthias: What the Relationship between Correct Recognition Rates and Usability Measures Can Tell Us About the Quality of a Speech Application. In: *Proceedings of the 6th International Scientific Conference on Work With Display Units.* Berlin: ERGONOMIC Institut für Arbeits- und Sozialforschung 2002.

Peissner, M.; Biesterfeldt, J.; Heidmann, F.: *Akzeptanz und Usability von Sprachapplikationen in Deutschland.* Stuttgart: IRB Verlag 2004.

Peissner, M.; Heidmann, F. und Corves, L.: Usability Engineering von Sprach-Applikationen. In: *User Interface Tuning – Benutzungsschnittstellen menschlich gestalten* (Machate, J.; Burmester, M., Hrsg.). Frankfurt a.M.: Software & Support/PRO 2003.

Raebel, R.: Sprachtechnologie – Mehr als Spracherkennung. http://www.frankfurt-main.ihk.de/presse/ihk-wirtschaftsforum/2002/0201/sprachtechnologie/index.html (Letzter Zugriff: 04.10.2003)

Weinschenk, S.; Barker, D.T.: *Designing Effective Speech Interfaces.* New York: Wiley, 2000.

Yankelovich, N.; Levow, G.A.; Marx, M.: Designing SpeechActs: Issues in Speech User Interfaces. *CHI '95 Proceedings.* ACM Conference on Human Factors in Computing Systems. Denver, CO, May 7-11, 1995.

Matthias Peissner, Dipl.-Psych., hat 1999 sein Studium der Psychologie an der Universität Regensburg abgeschlossen. Seit Anfang 2000 arbeitet er im Competence Center Human-Computer Interaction des Fraunhofer-Instituts für Arbeitswirtschaft und Organisation (IAO) in Stuttgart. Momentan beschäftigt er sich als User Interface Designer und Usability Engineer mit verschiedenen Projekten zu Themen wie Sprachinteraktion, Imaging Software und Internet. Er ist Mitglied im Vorstand des „German Chapters der Usability Professionals Association".
E-Mail: Matthias.Peissner@iao.fraunhofer.de