

Benutzererwartungen im Zusammenhang mit dem Sprachdialogsystem INSPIRE

Rosa Pegam¹, Jan Krebber², Ute Jekosch³

¹ *Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Deutschland, Email: rosa.pegam@rub.de*

² *Institut für Kommunikationsakustik, Ruhr-Universität Bochum, Deutschland, Email: jan.krebber@rub.de*

³ *School of Architecture, Rensselaer Polytechnic Institute, Troy, N.Y., USA, Email: ute.jekosch@rub.de*

Einleitung

Menschen haben bestimmte Vorannahmen über das, was die Gestaltung und den Verlauf eines Dialogs betrifft. Entsprechend der Einschätzungen, die die Interaktionspartner übereinander machen, passen sie auch ihre Redeweise an. Offensichtlich wird ein ähnliches Verhalten an den Tag gelegt, wenn Menschen mit Maschinen sprechen. Dies zeigt sich in der Sprechgeschwindigkeit, im gewählten Register, im Stil (knappe oder komplexe Satzbildung) und in einer Begrenzung der Konversationsthematik.

Zur Erreichung eines flüssigen Dialogverlaufes gilt in jedem Falle, den Benutzererwartungen zu einem gewissen Grade entgegenzukommen. Die Durchführung von Experimenten mit dem INSPIRE System hat es ermöglicht, klar zu bestimmen, inwiefern Erwartungen bezüglich des Dialogverlaufs nicht getroffen wurden, welche Gründe dafür zu suchen sind und welche Auswirkungen dies auf das Interaktionsverhalten und auf die anschließenden Qualitätsurteile seitens der Versuchspersonen hatte.

Das INSPIRE System

INSPIRE¹ ist ein Sprachdialogsystem, das in so genannten Smart-Homes installiert werden kann. Es ermöglicht seinen Benutzern, eine Reihe von technischen Geräten im Haushalt per Spracheingabe zu steuern. Das System soll gleichermaßen von naiven als auch von erfahrenen Benutzern bedient werden können.

Nutzererwartungen im Dialog

Während der Interaktion gibt es eine Vielzahl von Fällen, in denen beobachtet werden konnte, dass Erwartungen nicht getroffen wurden. Enttäuschte Erwartungen führen zunächst zu stockenden und langen Dialogen und können dadurch Verärgerung oder Frustration bei den Nutzern hervorrufen, weil diese sich etwa unverstanden fühlen oder nicht begreifen, was das System von ihnen erwartet. Erwartungen entstehen aus Erfahrungen, welche sich aus einer Vielzahl von Erlebnissen mit früheren Gesprächspartnern zusammensetzen (z.B. Erfahrungen im Austausch mit Ausländern, Hörgeschädigten oder Unterhaltungen mit störenden Hintergrundgeräuschen). Obwohl jeder Mensch individuelle Erfahrungen macht, gibt es typischerweise generelle Verhaltensmuster in wiederkehrenden Situationen: Benutzer sprechen z.B. langsamer, deutlicher und gedehnter, wenn sie meinen, der Spra-

cherkenner funktioniere unzureichend. Außerdem bleiben Benutzer im Allgemeinen bei einem bestimmten Register. Komplizierte Satzkonstruktionen werden vermieden, mit dem Ziel, den Dialogverlauf so flüssig wie möglich zu halten. Im Verlaufe der Interaktion entwickeln sich bei den Benutzern Lernprozesse, durch die sie im Umgang mit dem System gewandter werden und die es ihnen letztendlich erlauben, Vorgänge zusammen mit dem System zügiger und erfolgreich durchzuführen. Dabei passen Benutzer ihre Sprechweise an, indem sie z.B. solche Begriffe, die sie als Schlüsselwörter ausgemacht haben, gezielt einsetzen.

Trotz aller Vorannahmen und entsprechender Anpassung ihrer Sprechweise fanden Benutzer ihre Erwartungen dem Dialogverlauf gegenüber häufig nicht getroffen, was sich in den Bewertungen zur Systemqualität widerspiegelt, die sie nach den Interaktionen mit dem System abgegeben haben. Im Allgemeinen sind diejenigen Dialoge am anfälligsten, die zur Bedienung der Geräte nötig sind, bei denen das System die meisten Informationen vom Nutzer benötigt. Die Tests haben gezeigt, dass die Anzahl der zu füllenden Slots (d.h. die Komplexität der Aufgabe) in einem Zusammenhang mit der Beliebtheit der entsprechenden Geräte bei den Versuchspersonen steht. Die folgende Liste stellt eine Anordnung der Geräte nach Schwierigkeitsgrad dar, wie sie in Anlehnung an die Bewertungen der Versuchspersonen erstellt werden konnte:

1. TV/VCR/EPG
2. Lampen
3. Anrufbeantworter
4. Rollo
5. Ventilator

Je größer die Menge der obligatorischen Parameter, die das System vom Nutzer erfragen muss, umso höher empfanden auch die Versuchspersonen im Allgemeinen die kognitive Belastung, die mit der Bedienung der Apparate einherging. Es ist allerdings zu bemerken, dass die Struktur des Anrufbeantworters komplexer ist als die der Lampen.

<i>Anrufbeantworter:</i>		
<i>Alle Nachrichten abhören und dann die letzte löschen</i>		
1.	Apparat	Anrufbeantworter
2.	Aufgabe	Aktuelle Nachricht
<i>Neuer Dialog</i>		
3.	Aufgabe	Nächste Nachricht
<i>Neuer Dialog</i>		
4.	Aufgabe	Nachricht löschen
5.	Bestätigung	Ja

Tabelle 1: Obligatorische Parameter des Anrufbeantworters.

¹ **IN**fortainment management with **SP**eech **I**nteraction via **RE**remote-microphones and telephone interfaces, siehe zu diesem System auch Krebber (2005)

Zum Vergleich im folgenden eine Beispielstruktur zum Bedienen der Lampen.

<i>Lampe: Gelbe Stehlampe einschalten und dimmen</i>		
1.	Apparat	Lampe
2.	Standort	Gelbe Stehlampe
3.	Aufgabe	Einschalten
<i>Neuer Dialog</i>		
4.	Apparat	Lampe
5.	Standort	Gelbe Stehlampe
6.	Aufgabe	Dimmen

Tabelle 2: Obligatorische Parameter der Lampe.

Die Bedienung des Anrufbeantworters hat zwar eine kompliziertere Struktur, vor allem, weil es sich für den Benutzer als schwierig erwiesen hat, eine bestimmte Nachricht auszuwählen oder festzustellen, an welchem Punkt sich das System gerade befindet. Trotzdem entschieden mehr Benutzer, sie wollten die Lampen lieber auf herkömmliche Weise bedienen als den Anrufbeantworter.

Häufige Probleme

Im Folgenden werden die am häufigsten beobachteten Probleme aufgeführt. Eine Kategorisierung der Gründe für problematische/unrunde Interaktionen kann wie folgt erstellt werden. Eine Gruppierung in sprachliche und technische Systemgrenzen kann vorgenommen werden:

<i>Sprachliche Systemgrenzen</i>	<i>Technische Systemgrenzen</i>
Referenzen zu dynamischen Objekten	Modalität
Referenzen zu statischen Objekten	Fehlerkennungen
Ambiguitäten	Kein Multitasking

Tabelle 3: Kategorisierte Problembereiche.

Dynamische Objekte sind solche, die ständigen Änderungen unterliegen wie z.B. TV-Sendungen. Eine häufig vorkommende Dialogsituation, die während der Experimente mit dem System auftrat, war, dass Benutzer die *Titel* von Sendungen nannten (z.B. „Ich möchte gerne James Bond sehen.“), was nicht vom System interpretiert werden konnte, da Sendungstitel nicht im Lexikon aufgeführt sein können.

Statische Objekte sind solche, die fest installiert sind, wie etwa die Lampen. Da das INSPIRE System drei bedienbare Lampen kennt, ist eine Standortspezifizierung bei jedem Gebrauch einer Lampe notwendig („Gelbe Stehlampe einschalten“). Die Bedienung der Lampen gestaltete sich deshalb oft schwierig für die Testpersonen, unter anderem, weil sie in diesem Zusammenhang Referenzen gebrauchten, die vom System nicht aufgelöst werden konnten:

User: *Die gelbe Stehlampe einschalten.*

System: [gelbe Stehlampe wird eingeschaltet]

User: *Andere Stehlampe auch einschalten.*

System: Ich habe Lampe und einschalten verstanden. Welche Lampe meinen Sie?

Hier erwartet der Benutzer, dass das System ableitet, welche Stehlampe noch nicht bedient worden ist und das Wort *andere* entsprechend auflöst. Da das System aber stets explizite und vollständige Anweisungen benötigt, um eine Aktion durchzuführen, reagiert es mit einer Aufforderung an den Benutzer, den Standort der Lampe genau zu spezifizieren.

Ambiguitäten führen zu Konfusion, da das System solche Begriffe nicht sicher zuordnen kann (z.B. *Nachrichten* kann auf den Anrufbeantworter oder den Fernseher verweisen). Derartige Begriffe können dazu führen, dass das System einen unerwünschten Aufgabenfokus betritt und der Benutzer wiederum so lange neue Werte eingeben muss, bis der gewünschte Fokus erreicht ist.

Modalität bezeichnet die Eingabemodalität. INSPIRE ist unimodal, was den Benutzern nicht immer komplett bewusst war. So erwarteten sie vom System u.a., dass es Befehle wie „die Lampe rechts von mir“ interpretiert, wobei keine Möglichkeit besteht, den Benutzer zu orten.

Fehlerkennungen führen immer zu Störungen im Dialogfluss und ziehen Klärungs- und Korrektur-Turns des Benutzers nach sich.

Multitasking wird von INSPIRE nicht unterstützt. D.h., es kann immer nur ein Gerät pro Befehl bedient werden. Werden zwei verschiedene Geräte genannt, wird ein Klärungsdialog vom System angewendet („Meinen Sie Lampe oder Fernseher für Gerät?“).

Zusammenfassung

Erwartungen werden durch das Abrufen von früheren Erfahrungen hergestellt. Dies gilt sowohl für Mensch-Mensch als auch für Mensch-Maschine Interaktion. Die resultierenden Anpassungen betreffen sowohl Wortwahl als auch Aussprache. Die häufigsten Probleme in der Interaktion mit dem System, die durch enttäuschte Erwartungen hervorgerufen wurden, wurden gruppiert unter sprachliche und technische.

Literatur

[1] Krebber, J. (2005) *Hallo, ist jemand zu Hause? Ermittlung der notwendigen Worterkennungsrate eines Smart-Home-Systems*. Proceedings DAGA '05

[2] Homepage des INSPIRE Projektes URL: <http://www.inspire-project.org>

[3] Boland, H., Hoonhout, J., Möller, S., Krebber, J., Schuchardt, D., Rajman, M., und Smeele, P. (2004). *System Usability Evaluation Report, Deliverable 6.2, IST project INSPIRE (INfotainment management with SPeech Interaction via REmote-microphones and telephone interfaces, IST-2001-32746)*

[4] Möller, S. (2005). *Quality of Telephone-Based Spoken Dialogue Systems*, Springer, US-Boston MA.

[5] Bernsen, N. O., Dybkjær, H., und Dybkjær, L. (1998). *Designing Interactive Speech Systems*. Springer-Verlag, London.