

Multimodale Mensch-Roboter-Interaktion im privaten Haushalt

Sascha Herr, Tom Gross

Lehrstuhl für Mensch-Computer-Interaktion, Otto-Friedrich-Universität Bamberg

Zusammenfassung

Die Mensch-Roboter-Interaktion beschäftigt sich mit der Gestaltung, Implementation und Evaluation der Interaktion zwischen Benutzerinnen und Benutzern einerseits und Robotern andererseits. Dabei stellt ist Multimodalität von zentraler Bedeutung für den Erfolg dieser Interaktion. Wir berichten von unserer explorativen Interview-Studie zur Gewinnung von Anforderungen an die Multimodalität mit besonderem Blick auf Anforderungen an die Kombinierbarkeit und Nähe.

1 Einleitung

Die Mensch-Roboter-Interaktion (MRI) beschäftigt sich mit der Gestaltung, Implementation und Evaluation der Interaktion zwischen Benutzerinnen und Benutzern und Robotern (Dautenhahn 2014). Aus einer menschenzentrierten Perspektive ist es dabei wichtig, fundamentale Aspekte der sozialen Interaktion von Menschen untereinander zu verstehen, um diese bei der Gestaltung von MRI-Schnittstellen anzuwenden (Dautenhahn 2014) (Gross 2003). Einer dieser fundamentalen Aspekte ist die Multimodalität der Kommunikation, also der Kombination von Kanälen zum Austausch von Information (Kress 2009). Menschen bedienen sich ihrer durch Kombination von Sprache, Gesten, Körperhaltung, Blicken und Gesichtsausdrücken (Dautenhahn 2014). Der Haushalt stellt hierbei einen besonders interessanten Kontext dar, da dort die individuelle Umgebung und die beträchtliche Anzahl und Diversität möglicher Aufgaben vielschichtige Anforderungen mit sich bringen.

Der vorliegende Beitrag berichtet von einer explorativen Interview-Studie, in der wir Anforderungen, Bedürfnisse und Erwartungen von Benutzerinnen und Benutzern an die multimodale Interaktion mit Robotern für den Haushalt erhoben haben und davon Erkenntnisse zur Kombination von Modalitäten und deren Näheanforderungen ableiten.

2 Methode

Für unsere explorative Fragestellung setzten wir halbstrukturierte Interviews ein, welche uns erlaubten, flexibel und situativ im Gespräch reagieren zu können und tiefgehende Einsichten zu gewinnen.

Teilnehmerinnen und Teilnehmer: Wir rekrutierten fünf (zwei weibliche) freiwillige Interviewpartner im Alter von 30 bis 44 Jahren ($M=35,80$, $SD=7,05$), die einen eigenen Haushalt besaßen und diverse (nicht technische) berufliche Hintergründe hatten.

Material: Es wurde ein Interviewleitfaden entwickelt, der Instruktionen für den Interviewer enthielt und den Fragenkatalog in einer vorgeschlagenen Reihenfolge abbildete. Der Fragenkatalog wurde gemäß Interviewrichtlinien entwickelt und enthielt offene Fragen mit definierten Nachfragen zu Erwartungen und Anforderungen bezüglich der Interaktion mit einem Roboterarm im Haushalt mit Blick auf verschiedene relevante Modalitäten (Sprache, Gesten, Handführung, Blicke, PCs und mobile Geräte).

Interviewablauf: Die persönlichen Befragungen fanden im jeweiligen Haushalt der Teilnehmerinnen und Teilnehmer statt und dauerten zwischen 90 und 120 Minuten. Sie wurden nach Unterzeichnung der Einverständniserklärung für die nachfolgende Transkription auf Video aufgezeichnet. Die Teilnehmerinnen und Teilnehmer wurden zunächst in das Thema MRI im Haushalt eingeführt, indem ihnen drei Beispielbilder und ein Auszug eines Videos gezeigt wurde, worauf ein Leichtbauroboterarm (KUKA LBR) Haushaltsaktivitäten wie bügeln oder Löcher bohren, ausführt. Während der Interviews wurden fortwährend Notizen gemacht. Nach den Gesprächen erhielten die Teilnehmerinnen und Teilnehmer die Möglichkeit, weitere Kommentare hinzuzufügen und Fragen zu stellen.

Nach Abschluss aller Interviews wurden die Aufzeichnungen transkribiert. Im Zuge einer thematischen Analyse wurden Aussagen identifiziert und markiert, bedeutsame Kodierungen extrahiert und diese dann zu Hauptthematiken gruppiert.

3 Ergebnisse

Nachfolgend geben wir einen kurzen Eindruck über zentrale Aussagen zu verschiedenen Modalitäten.

Handführung: Dabei führen Benutzerinnen und Benutzer den Roboterarm in direktem Kontakt, sodass Bewegungsmuster aufgezeichnet werden können. Die Teilnehmerinnen und Teilnehmer empfanden diese Modalität als intuitiv und einfach, wobei Skepsis bezüglich der Drucksensibilität bei Greifaufgaben mit fragilen Objekten herrschte. Ein Teilnehmer sagte, „*Ich weiß manchmal selber nicht, wie viel Druck ich benötige, um eine Flasche Wasser aufzumachen oder ein Glas zu spülen, wie soll das dann der Roboter wissen?*“.

Demonstration: Dabei führen die Benutzerinnen und Benutzer Bewegungen vor, die der Roboter dann nachahmen soll. Die Teilnehmerinnen und Teilnehmer konnten sich diese

Modalität gut vorstellen, insbesondere um dem Roboter komplexe Bewegungsmuster beizubringen. Ein Teilnehmer formulierte, *„Es wäre das einfachste, eine Art Handschuh dafür zu haben und ich mache damit einfach etwas und der Roboter macht es nach. Zumindest für komplexe Dinge fände ich das angenehmer als den Roboter an die Hand zu nehmen.“*. Ein weiterer Teilnehmer sagte, *„Ich denke, dass die weitere Entwicklung dahin geht, dass der Roboter dich beobachtet und daraus von allein schlussfolgert, was zu tun ist.“*.

Gesten: Dabei führen Benutzerinnen und Benutzer abgeschlossene Armbewegungen durch, die der Roboter als Befehle erkennt und in Bewegungen umsetzt. Die Teilnehmerinnen und Teilnehmer empfanden Gesten als brauchbar, befürchteten jedoch eine gewisse Umständlichkeit. Eine Teilnehmerin meinte, *„Wenn ich auf den Stift am Boden zeige und dann eine Greifgeste mache...ich glaube, bevor ich das mache, hätte ich den Stift schon drei mal selber aufgehoben.“*. Ein Anderer war bezüglich der Privatsphäre besorgt, *„Wenn ich mit Gesten kommunizieren muss, dann heißt das auch, dass ich permanent überwacht werde. Ich denke, dass ich mich da unwohl fühlen würde.“*.

Blicke: Dabei werden Augenbewegungen der Benutzerinnen und Benutzer erfasst und interpretiert. Die Teilnehmerinnen und Teilnehmer bezweifelten, ob dies für die eindeutige Identifikation von Objekten akkurat genug sei. Ein Teilnehmer sagte, *„Nun, jetzt schau ich dich [den Interviewer] an, aber in Wirklichkeit fokussiere ich die Flasche hinter dir. Das wäre sicher auch schwierig für den Roboter.“*. Des Weiteren wurden Schwierigkeiten genannt, *„Ich denke, das, was du am schlechtesten im Griff hast, ist dein Blick...Blicksteuerung stelle ich mir ganz schwer vor.“*.

Sprache: Dabei werden Befehle an den Roboter durch verbale Kommandos gerichtet. Die Teilnehmerinnen und Teilnehmer bewerteten Sprache als die intuitivste Eingabemethode, allerdings unter der Voraussetzung, dass der Roboter über ein entsprechendes Domänenwissen verfügt. Ein Teilnehmer formulierte dies so, *„Der Roboter muss schon ein hohes Maß an Verständnis eingebaut haben. [...] Ich kann ihm nicht einfach sagen, ‚gehe an diese Position am Tisch‘, weil ich das selbst nicht genau weiß.“*.

PCs und mobile Geräte: Dabei wird der Roboter ferngesteuert (über Maus, Tastatur oder Touch). Diese Eingabemethode wurde als sehr leicht eingestuft. Die Wünsche reichten hierbei von auswahlbasierten Schnittstellen zu 3D-Simulationsumgebungen, wobei die Benutzerinnen und Benutzer die Gelenke des Roboterarmes durch Gesten, welche mit den Fingern oder einer Maus ausgeführt werden, steuern möchten. Die Teilnehmerinnen und Teilnehmer schätzten den Aufwand für diese Art von Eingabe als sehr gering ein, *„Ich kann das machen, ohne mich viel zu bewegen. Und ich kann die Bewegung erst simulieren bevor ich sie speichere und anwende.“*.

4 Diskussion und Ausblick

Die obigen Ausführungen und Zitate geben nur punktuelle Eindrücke wieder – unsere Auswertung der transkribierten 100 Seiten Interview-Aussagen zeigen aggregiert v.a. zwei Her-

ausforderungen für die tatsächliche Umsetzung von Multimodalität für MRI: die sinnvolle Kombination und die Näheanforderungen der Modalitäten.

Die Kombination verschiedener Modalitäten ist essentiell für eine flexible Interaktion mit Robotern im Haushalt. Durch Blickerkennung können relevante Bereiche eines Raumes und darin enthaltene Objekte identifiziert werden. Via PC oder mobiler Geräte können komplexe Aufgaben entwickelt und simuliert werden, bevor diese z.B. mit Handführung vor Ort fein abgestimmt werden. Sprache und Gesten können sich bei geometrischen (Distanzen zwischen Händen zeigen) oder programmatischen Informationen („Wiederhole das 15 mal“) ergänzen. Der simultane Einsatz von zwei oder mehr Modalitäten beim Training kann dieses vereinfachen und effizienter gestalten (z.B.: „*Stelle diese Tasse auf den Tisch*“ sagen, während man den Roboterarm dazu mit der Hand führt).

Die unterschiedlichen Anforderungen bzgl. räumlicher Nähe der verschiedenen Modalitäten offenbarten sich bei mehreren von den Teilnehmerinnen und Teilnehmer genannten Beispielen. Die höchste Näheanforderung zeigte sich bei der Handführung, wo ein direkter *physischer Kontakt* vorhanden sein muss. Eine niedrigere Näheanforderung haben Demonstration, Gesten und Blicke, die alle zumindest einen *visuellen Kontakt* benötigten. Sprache erfordert einen *auditiven Kontakt*, sodass insbesondere in kleinen Haushalten auch über Räume hinweg Befehle gegeben werden können. Die geringste Ortsabhängigkeit ist bei PCs und mobilen Geräten gegeben, die lediglich auf einem *virtuellen Kontakt* basieren, der die Benutzerinnen und Benutzer vom Roboter örtlich komplett entkoppelt.

Die sinnvolle Kombination von Modalitäten und deren Näheanforderungen sind potentielle Forschungsfelder, die bei der Gestaltung von multimodalen MRI-Schnittstellen beachtet werden sollten. In zukünftigen Arbeiten könnte durch quantitative Untersuchungen mit größeren Stichproben eine stärkere Generalisierbarkeit erreicht werden.

Danksagung

Wir danken den Teilnehmerinnen und Teilnehmern. Teile diese Arbeit wurde durch die TechnologieallianzOberfranken (TAO) gefördert.

Literaturverzeichnis

- Dautenhahn, K. (2014). *Human-Robot Interaction*. In Soegaard, M. und Dam, R.F., (Hrsg.). *The Encyclopedia of Human-Computer Interaction* The Interaction Design Foundation, Aarhus, Denmark.
- Gross, T. (2003). Universal Access to Groupware with Multimodal Interfaces. In *Proceedings of the Second International Conference on Universal Access in Human-Computer Interaction* (June 22-27, Crete, Greece). Lawrence Erlbaum, Hillsdale, NJ. pp. 1108-1112.
- Kress, G. (2009). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. Routledge, London.

Kontaktinformationen

Prof. Dr. Tom Gross, tom.gross(at)uni-bamberg.de, T. 0951-863-3940