Gefördert vom
bmb+f
Bundesministerium für
Bildung, Wissenschaft,
Forschung und Technologie

Mensch-Technik-Interaktion
SmartKOM

Auswertung der ergonomischen Evaluation des SmartKom Prototypen

–Szenario Home–

Befragungszeitraum 10. März 03 bis 20. März 03

Nicole Beringer, Irene Jacobi, Sibylle von Tiedemann

Ludwig Maximilians Universität München

Memo Nr. 16

Juli 2003

Juli 2003

Auswertung der Ergebnisse der Evaluations–Aufnahmen Szenario Home

Fragebogenversion vom 10.03.03

Nicole Beringer, Irene Jacobi, Sibylle von Tiedemann

Ludwig Maximilans Universität München
Schellingstr. 3
80799 München

Tel.: (089) 2180–5751
FAX: (089) 2800362

E–Mail: {beringer, jacobi, tiede}@phonetik.uni–muenchen.de

Dieses Memo gehört zu Teilprojekt 7: Intuitives Arbeiten mit dem Computer:
SmartKom Home/Office; AP 7.5 Ergonomische Evaluation

Das diesem Memo zugrundeliegende Forschungsvorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01 IL 905 gefördert. Die Verantwortung für den Inhalt liegt bei den Autorinnen.

Inhaltsverzeichnis

1 Befragungszeitraum....4
2 Anzahl und Zusammensetzung der Befragten....4
2.1 Naivität bzgl. des Systems....4
2.2 Alter....4
2.3 Geschlecht....5
2.4 Staatsangehörigkeit....5
2.5 Schulabschluß....5
2.6 Sonstige Variablen....5
3 Ergonomische Evaluation des Systems....6
3.1 Freie Fragen....6
3.2 Fragen mit kategorisierten Antwortmöglichkeiten....12
3.2.1 Erwartungen, Aspekte, Zustimmung....12
3.2.2 Semantisches Differential zur Funktionalität von SmartKom....14
3.2.3 Bewertung der Persona Smartakus....15
3.3 Allgemeines....16
4 Fazit....17

1 Befragungszeitraum

Die Befragung wurde während des Evaluationszeitraumes für das Szenario Home vom 10.03.03 bis einschließlich 20.03.03 durchgeführt. Diese Memo befaßt sich ausschließlich mit den Befragten, die den SmartKom Prototypen Home getestet haben.

2 Anzahl und Zusammensetzung der Befragten

Insgesamt wurden 20 Personen befragt. Die folgenden Daten konnten alle bewertet werden und beziehen sich damit auf 20 gültige Fragebögen zur abschließenden Evaluation des Home–Prototypen.

2.1 Naivität bzgl. des Systems

Aufgrund der im Vorfeld der begleitenden Evaluation festgestellten Probleme bei der Durchführung der Versuche mit naiven Versuchspersonen wurde für die abschließende Evaluation eine Expertenevaluation angestrebt, d.h. keine der Versuchspersonen ist naiv gegenüber des vorgestellten Systems. Alle Versuchspersonen sind allgemein an die Bedienung von Dialogsystemen gewöhnt.

Die Befragung von ausschließlich nicht–naiven Versuchspersonen hat massive Auswirkungen auf die Altersverteilung, den höchsten Schulabschluß bzw. die Nationalitätenverteilung.

2.2 Alter

Das Alter der teilnehmenden Versuchspersonen lag zwischen 22 und 40 Jahren (im Schnitt bei ca. 30 Jahren). In Anlehnung an die Alterskategorien des Statistischen Bundesamtes wurde folgende Verteilung erreicht:

Alter	***Prozent***	***Verteilung Statistisches Bundesamt*** *(in Prozent)*	***Anzahl Vpn***
14–18	0	5.2	0
18–21	0	4	0
21–40	100	32.7	20
40–60	0	31.2	0
60 plus	0	26.9	0
Summe	100	100	n=20

Hier wird ersichtlich, daß die erstrebte Rekrutierung der Versuchspersonen entsprechend der Altersstruktur der Bevölkerung der Bundesrepublik Deutschland nicht gelungen ist. Dies ist auf die erschwerten Bedingungen beim Aufbau des Systems SmartKom zurückzuführen, welche eine Nutzung durch komplett naive Versuchspersonen ausschloß.

2.3 Geschlecht

10 weibliche und 10 männliche Versuchspersonen teilen die Befragten paritätisch.

2.4 Staatsangehörigkeit

3 der 20 Befragten gaben an, keinen deutschen Paß zu besitzen. Dies sind 15.0% aller Befragten dieser Welle. Damit ist die angestrebte Verteilung von ca. 8% „Inhaber eines nicht–deutschen Passes" um ca. das Doppelte überboten.

2.5 Schulabschluß

Entsprechend der Kategorien des Statistischen Bundesamtes verteilten sich die Befragten wie folgt auf die verschiedenen Schulabschlüsse:

	Prozent	***Verteilung Statistisches Bundesamt (in Prozent)[1]***	***Anzahl Vpn***
Volks–/ Hauptschule	0	50.5	0
Realschule/ gleichwertiger Abschluß	0	28.7	0
(Fach–) Hochschulreife	100	20.8	20
Ohne Abschluß	0	–	0
Keine Angaben	0	–	0
Sonstige	0	–	0
Summe	100	100	20

[1]Verteilung beruht auf Personen mit allgemeinbildendem Schulabschluss. Personen ohne Abschluss wurden hier nicht erfaßt.

Erfragt wurde in diesem Zusammenhang der zuletzt erreichte allgemeinbildende Schulabschluss. Dies ist nicht zu verwechseln mit akademischen Graden oder Ausbildungsabschlüssen, die in dieser Verteilung nicht aufgelistet sind.

Wie oben bereits erwähnt, war eine Rekrutierung der Versuchspersonen ausschließlich nach statistischen Werten in dieser Phase der Evaluation nicht praktikabel. Deshalb differieren die Werte erheblich zur angestrebten Verteilung entsprechend den Daten des Statistischen Bundesamtes.

2.6 Sonstige Variablen

Mit 17 Personen sind die meisten Probanden deutsche Muttersprachler. Drei Versuchspersonen gaben an, Chinesisch, Russisch bzw. Französisch

als Muttersprache zu haben.

3 Ergonomische Evaluation des Systems

Zur ergonomischen Evaluation des Systems wurde den Probanden ein Fragenkatalog vorgelegt, der sich in zwei Bereiche aufteilt. Fragen mit kategorisierten Antwortmöglichkeiten (Werteskala von –3 ungenügend bis +3 sehr gut) lassen einen Vergleich mit objektiven Meßergebnissen der technischen Evaluation zu. Freie Fragen erlauben der VPn, ihre Wünsche und Einschätzungen stichpunktartig zu notieren und somit eine detailliertere qualitative Werteinschätzung abzugeben. Die folgenden Unterabschnitte beschreiben die einzelnen Antwortcluster zu den freien Fragen bzw. den kategorisierten Fragen.

3.1 Freie Fragen

Bei der qualitativen Beurteilung des Systems durch die Versuchsperson waren Mehrfachnennungen möglich. Für jede Frage (kursiv) werden in den jeweiligen Tabellen die Kluster mit der Anzahl der Nennungen wiedergegeben. Zusätzliche bzw. abweichende Nennungen werden nach der Tabelle aufgeführt.

Beschreiben Sie bitte in Stichpunkten Ihre Erwartungen an das System SmartKom bezüglich Funktionsumfang und Funktionalität (Was sollte es können und wie sollte es dies können)

Eingabemöglichkeiten	Informationsmöglichkeiten	Gerätesteuerung
Spontansprache: 9	Fernsehprogramm: 7	TV/Video/DVD/Spiele–konsole: 6
Gestik: 5	Organizer: 2	Haushaltsgeräte: 3

Für die meisten Probanden war die Eingabemöglichkeit über Spontansprache wichtig. Die Bedienung durch Gestik wurde insgesamt nur zweimal genannt. Bei den Informationsmöglichkeiten bezogen auf die Anwendung lag der Schwerpunkt der Gewichtung auf dem Fernsehprogramm. 2 der VPn wünschten sich die Möglichkeit eines Organizers. Bei der Gerätesteuerung waren die Nennungen anwendungsgebunden überwiegend im Bereich Unterhaltungsmedien. Eine Ansteuerung von Haushaltsgeräten durch das System wurde dreimal benannt.
Zusätzlich zu Eingabe– und Informationsmöglichkeiten bzw. Gerätesteuerung lassen sich unter Systemeffizienz folgende Erwartungen zusammenfassen:

- Bedienungshilfe: 3
- Lean–Back: 2
- Geschwindigkeit: 5
- Benutzerfreundliche Bedienung: 2

Zusammenfassend läßt sich sagen, daß die Probanden alle ein System

wünschen, das die jeweiligen Anforderungen – hier alles rund ums Fernsehen – schneller und effizienter als herkömmliche Methoden erfüllt und zwar mit unabhängigen Sprach– bzw. Gestikeingaben.

Inwiefern war die Bedienung des Systems SmartKom eindeutig? Bitte erläutern Sie Ihre Antwort.

eindeutig	nicht eindeutig
durch Vorwissen / Instruktion: 3 Hinweis auf (eingeschränkte) modale Mögl.) 2	Zurückbutton fehlt: 1
Multimodalität: 5	Wichtelhandlungen: 2
Display eindeutig: 2	Unklar, ob verstanden: 3
verbale Aufforderung / durch Frage–Antwortrhythmus: 3	unklar, was möglich: 6
wg. eingeschränkter Mögl.: 1	Ausprobieren 3

Eindeutigkeit wurde bei den Versuchspersonen erfüllt durch Vorwissen/Instruktion und Hinweis auf (eingeschränkte) modale Möglichkeiten (5 Nennungen), durch Multimodalität (5), durch eindeutiges Display (2), durch die verbale Aufforderung bzw. durch Frage–/Antwortrhythmus (3) und sogar durch eingeschränkte Interatkionsmöglichkeiten (1). Zwei VPn fanden die Bedienung eindeutig, benannten aber das Wie nicht näher.
Das System wurde von den Probanden als nicht eindeutig bedienbar empfunden, weil ein Zurückbutton fehlt (1), die Wichtelhandlungen verwirrend sind (2), es Unklarheit über das Verständnis durch das System (3) bzw. über die tatsächlichen Bedienmöglichkeiten (6) gab und einige sich nur durch Ausprobieren (3) helfen konnten.
Kritisiert wurde von einer Person: „Es ist nur das möglich auszuwählen, was auch zu diesem Zeitpunkt angezeigt ist."

Konnten Sie eine Funktion nutzen, ohne dass Sie die Versuchsleitung explizit darauf aufmerksam gemacht hat?

ja	nein
TV ein–/aus–/umschalten: 2	13
Allg.: 1	
keine Angabe: 3	

13 Probanden gaben an, keine Funktion genutzt haben zu können, die nicht explizit von der Versuchsleitung genannt wurde. Die restlichen Probanden hatten eigenständigen Erfolg beim TV ein–/ aus–/ umschalten (2), allgemeine Erfolge (1), keine Angabe (3) und einmal konnte das System auch mit von der Vorlage abweichenden Sätzen

bedient werden („Ich glaube nicht, dass ich ausschließlich die vorgegebenen Sätze benutzt habe").
Inwiefern half SmartKom Ihnen, die Bedienung des Systems zu erlernen?

gar nicht	Folgendes half
Gar nicht: 13	Feedback: 1
	Hinweis durch Wichtel, daß Präferenzen für einzelne Genre möglich sind. 1
keine Angabe: 1	

13 Versuchspersonen gaben an, keinerlei Hilfe bei der Bedienung durch das System erhalten zu haben. Bedienhilfen durch das System erfolgten durch Feedback (1) bzw. Hinweis auf Genrepräferenzen durch den Wichtel (1). Weitere „Bedienhilfen" wurden durch die Instruktion (1), durch „ungewolltes Hochschrauben der Nachrichtenpräferenz" (1) bzw. „Nur, durch die Aufforderung zur Auswahl." (1) gegeben. Diese Ironisierung bzw. Abweichung von der eigentlichen Fragestellung läßt den Schluß zu, daß diese Versuchspersonen keine wirkliche Hilfe bei der Bedienung durch das System erfahren konnten.

Könnte es sein, daß Sie Fehler bei der Bedienung des Systems gemacht haben? Wenn ja, welche?

ja	nein
verbal (Wortwahl / Satzbau/ Artikulation / Geschwindigkeit) 9	Keine/ nicht bewusst: 3
Räuspern 2	
Gesten 2	
Sprachausgabe nicht abgewartet 7	

Die Frage nach Bedienfehlern wurde von 3 VPn mit „nein" bzw. „nicht bewußt" beantwortet. In 18 Fällen gaben die VPn, sich verbaler „Fehler" (Wortwahl / Satzbau/ Artikulation / Geschwindigkeit, Räuspern, Sprachausgabe nicht abgewartet) bewußt zu sein, 2 verwiesen auf Gesten, die sie als falsch empfunden haben. In 4 Fällen vermuten die VPn Fehler in der Bedienung, die sie aber nicht explizit nennen können:

- nicht verstanden, daß Smartakus "wach auf " nicht verstanden hat
- „Nicht angefragte Informationen geliefert, angefragte Informationen nicht gezeigt"
- „Habe mich zweimal versprochen, aber das sollte nicht als Fehlbedienung gelten"

- „Ein– oder zweimal habe ich gesprochen, bevor eine verbale Ausgabe kam. Sollte aber bei einem System eigentlich möglich sein"

Inwiefern ging SmartKom auf Ihre Befehle und Wünsche ein? Bitte erklären Sie.

erfüllt	nicht erfüllt
Gut bis sehr gut:6	Syntheseausgabe: 'ich habe Sie nicht verstanden' 1
Mittel 4	Syntheseausgabe: falsche Antwort 3
schlecht / zu langsam 5	Geschwindigkeit: zu langsam 2
keine Angabe 4	Erkennung englischer Titel mangelhaft 1
	Kein Zurückbutton: 1
	Emotionserkennung:1

Insgesamt schienen die Versuchspersonen mit der Verarbeitung ihrer Befehle zufrieden zu sein. So fanden 6 VPn die Antworten gut bis sehr gut erfüllt, 4 waren im großen und ganzen zufrieden und nur 5 fanden die Ergebnisse schlecht bzw. zu langsam. 4 Probanden machten keine Angabe. Dennoch gab es einiges zu bemängeln. Die Syntheseausgabe 'ich habe Sie nicht verstanden' (1) und falsche Antworten (3) wurden als störend empfunden. Ebenso war die Langsamkeit der Verarbeitung (2) und die schlechte Verarbeitung englischer Titel (1), der fehlende Zurückbutton(1) und die Emotionserkennung (1) störend. Und schließlich bemängelte eine VPn, daß beim Lean Backward die Aktivierung/Deaktivierung nicht funktioniert. Detailliertere Hinweise geben folgende Aussagen:

- „Vorsprechen verschiedener Nachrichtensendungen war schön"(1)
- „Wenn man bespielsweise Gesten macht, wo sie nicht vorgegeben sind (also nicht in Verbindung mit bestimmten Sätzen), funktioniert das offenbar nicht und SmartKom kann den gleichzeitig geäußerten Satz (der ohne Geste auszuführen ist) nicht richtig verstehen."(1)
- „Nur ein mal verhing es sich in der Sparte "Serien (bzw. "Nachrichten")"(1)
- „Wunsch nach Liste von Nachrichten zwar ausgeführt, jedoch nachgefragt, ob die Präferenz für dieses Genre herabgesetzt werden soll."(1)

Hier zeigt sich, daß naheliegende Befehle (Sprache + Geste vs. nur Geste) für das System nicht gleichbedeutend zu sein scheinen bzw. daß das System Parallelen aufzubauen scheint (Nachrichtenliste vs. Herabsetzung der Präferenzen), die für den Menschen nicht nachvollziehbar zu sein scheinen.

Wie half SmartKom Ihnen; Ihre eventuellen Fehler zu erkennen und ggf. zu korrigieren?

gar nicht	half
Gar nicht: 11	0
nur 'ich habe Sie nicht verstanden': 8	

Es zeigt sich bei dieser Frage wie auch schon in beiden vorangegangenen begleitenden Evaluationen, daß sich die VPn bei der Erkennung evtl. Fehler und deren Korrekturmöglichkeiten im Stich gelassen fühlen. 11 gaben daher an, keine Hilfe bekommen zu haben, 8 sahen im 'ich habe Sie nicht verstanden', je eine VPn im Grinsen des Wichtels bzw. der „Wiederholung des Prozesses" einen Hinweis auf potentielle Fehler, die sie aber nicht beheben konnten.

Inwiefern reagierte SmartKom eventuell trotz Bedienfehler wie gewünscht?

reagierte nicht	reagierte
Gar nicht : 10	Manchmal trotz Sprachfehlern 2
	Nach Verbesserung: 1
keine Angabe: 3	Sehr Verspätet: 1

Bei 10 Probanden reagierte SmartKom gar nicht wie gewünscht, bei 2 VPn „manchmal trotz Sprachfehlern", bei je einer nach Verbesserung bzw. sehr zeitverzögert.
Weitere nicht kategorisierbare Antworten waren „reagierte normal", „trotz Räuspern" , „Hat mir erlaubt wieder ein Schritt zurück machen, um ein Programm nochmal auswählen können".

War der Systemoutput (Sprache und Display) jederzeit eindeutig zu verstehen? Bitte erklären Sie.

ja	nein
Display eindeutig 3	Display: zu wenig Informationsgehalt / Wichtel / roter Punkt: 5
	Sprache: Prosodie und Bedeutung 10
Allgemein 7	Mangelnde Koordination zwischen Sprach–Graphik–Ausgabe 4
	Mangelndes Feedback: 2

Bei der Verständlichkeit der Ausgabe sind sieben VPn im allgemeinen zufrieden. Zumindest die Displaygestaltung scheint öfters eindeutig (3), wenn auch fünf VPn den Informationsgehalt von Wichtel und rotem Punkt bemängeln. Bei der Sprachausgabe bemängelten 10 Probanden die prosodische Modellierung und Bedeutung (z.B. (SOf) statt /S@U/ „show"). Vier beklagten sich über eine mangelnde Koordination zwischen Sprach–Graphik–Ausgabe und zwei über mangelndes Feedback.

Konnten Sie die Ihnen gestellte Aufgabe in einem dafür angemessenen Zeit– und Aufwandsrahmen (im Vergleich zu herkömmlichen Methoden) erfüllen? Bitte erklären Sie.

ja	nein
	Zeit: zu langsam 17
	Aufwand: Mehrfacheingaben nötig 1
	Aufwand: wohlartikuliertes Sprechen nötig 1
	Aufwand: zu umständlich (kein Zurückbutton z.B.) 2
	Allgemein: 1

Diese Frage zeigt deutlich, daß die Probanden herkömmliche Methoden bevorzugen, weil sie schneller (17) und mit geringerem Aufwand (4) zum Ziel führen. Eine VP gibt keine Erklärung.

Folgende Zitate sind selbstredend für die Meinung der Probanden gegenüber dem SmartKom Prototypen Home:

„einziger Vorteil ist die Sortierfunktion, die die Fernsehzeitschrift hat"

... „ist sogar eine Zumutung und kostet unheimlich viel Nerven, zumal ich keine menschlichen emotionalen Reaktionen zulassen kann, da das System sonst mißverständlich darauf reagiert."

Nur eine VPn sucht den Grund für die Nichterfüllung der Aufgabe bei sich: „Nein, weil ein Fehler von meiner Seite gemacht wurde.".

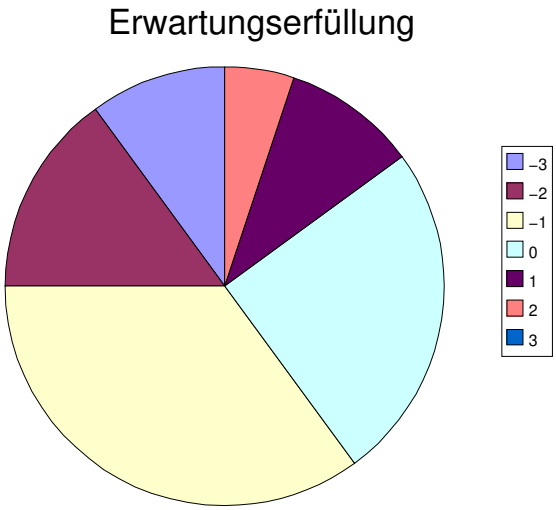
3.2 Fragen mit kategorisierten Antwortmöglichkeiten

3.2.1 Erwartungen, Aspekte, Zustimmung

Für den Vergleich der subjektiven Einschätzung des Benutzers mit den objektiv meßbaren Resultaten des Systems werden kategorisierte Antworten benötigt, die im folgenden näher beschrieben werden.

Zunächst wurden die VPn dazu aufgefordert, ihre Erwartungshaltung bzw. –erfüllung gegenüber dem System einzuschätzen. Die 7 Abstufungen (–3 gar nicht erfüllt bis +3 völlig erfüllt) ergaben folgende

Verteilung:



Die Graphik zeigt deutlich eine überwiegend negative Erwartungserfüllung durch das System. Der Median liegt mit 7 VPn bei einer Erwartungserfüllung von −1, 3 VPn vergaben −2, 2 sogar −3. Positive Erwartungserfüllungen waren bei 3 VPn zu verzeichnen (2 x 1 und 1 x 2), 5 VPn vergaben den Wert 0.

Des weiteren wurden Fragen zu einigen Aspekten von SmartKom gestellt. Auch szenariospezifische Detailfragen wurden implementiert.

Für multimodale Aspekte von Smartkom hatten die Probanden jeweils einen Aspekt von SmartKom von +3 (stimme zu/gefiel mir/funktionierte) bis −3 (stimme nicht zu/gefiel mir nicht/funktionierte nicht) zu bewerten. Diese wurden überwiegend neutral bis positiv gewertet, wie nachfolgende Tabelle zeigt.

Wie sehr gefallen Ihnen folgende Aspekte von SmartKom:	gefiel mir gar nicht			–	gefiel mir sehr			Mean	Median
	−3	−2	−1	0	1	2	3		
Displayeingabe (allgemein)	0	1	0	9	3	7	0	0.75	0.5
Displayausgabe (allgemein)	1	1	4	2	6	4	2	0.55	1
Spracheingabe (allgemein)	0	0	3	5	5	4	3	0.95	1
Sprachausgabe (allgemein)	3	2	6	2	3	1	3	−0.25	−1
Die Möglichkeit mit dem System menschlich zu interagieren	4	3	1	5	5	2	0	−0.5	0
Das Design insgesamt	0	5	2	5	3	3	1	0.1	0
SmartKom reagiert auf meine Emotionen	8	3	1	7	1	0	0	−1.5	−2
Stifteingabe	nicht genutzt								

Besonders positiv zu werten ist hier die Möglichkeit zur Displayeingabe, besonders negativ war die Bewertung der Reaktion auf Emotionen. Hier konnte nur von einer Person eine leicht positive Wertschätzung verzeichnet werden.

Unterstützt SmartKom die Erledigung Ihrer Aufgabe; ohne Sie unnötig zu belasten	läßt mich nicht unterbrechen			–	läßt mich unterbrechen			Mean	Median
	–3	–2	–1	0	1	2	3		
	4	5	4	4	1	1	0	–1.3	–1.5

Auch die fehlende Möglichkeit zu unterbrechen wurde von den VPn bemängelt.

Im Vergleich zur begleitenden ergonomischen Auswertung Home wurde bei der Zustimmung zu folgenden Aussagen besonders die passenden Antworten sehr positiv gewertet.

Bitte geben Sie an; wie sehr Sie folgenden Aussagen zustimmen:	stimme nicht zu			–	stimme zu			Mean	Median
	–3	–2	–1	0	1	2	3		
SmartKom hat mich verstanden	1	0	8	0	2	3	2	0.55	1
SmartKom hat zum größten Teil passende Antworten gegeben	3	5	5	5	0	8	5	0.4	0.5
Die Geschwindigkeit des Systems war angemessen	3	4	5	0	0	0	1	–1.65	–2
Ich wußte während des Dialogs immer; was ich sagen konnte	0	2	0	8	1	0	4	0	0
Die Eingabe über Zeigen hat funktioniert	2	4	4	1	2	7	0	–0.1	–0.5
Die Eingabe über Sprache hat funktioniert	0	2	0	5	9	3	1	0.7	1
SmartKom reagierte schnell auf meine Anfragen	8	3	3	4	2	0	0	–1.55	–2
SmartKom ist inhaltlich einfach zu verstehen	1	1	2	1	1	11	3	1.25	2
SmartKom ist akustisch gut wahrnehmbar	0	2	4	0	7	3	4	0.85	1
Die Stifteingabe hat gut funktioniert[1]	1	2	1	2	1	1	1	–0.22	0

1 Diese Funktion wurde von 11 Probanden nicht genutzt

Bitte geben Sie an; wie sehr Sie folgenden Aussagen zustimmen:	stimme nicht zu			–	stimme zu			Mean	Median
	–3	–2	–1	0	1	2	3		
Die Aufgabe war einfach zu lösen	1	2	1	2	1	1	1	0.9	1

Leider immer noch ein Manko aus der Sicht der VPn ist das Verständnis durch das System, die Langsamheit, der Erfolg der Stifteingabe und die Leichtigkeit der Aufgabenlösung. Positiv hervorzuheben im Vergleich zur begleitenden Evaluation Home ist auf jeden Fall eine überwiegend positive Bewertung der Spracheingabefunktion.

War bei der begleitenden Evaluation Home die durchweg negative Bewertung des Systems SmartKom im allgemeinen auffällig[2], so zeigt sich bei der abschließenden Evaluation Home insgesamt eine leicht positive Bewertung.

3.2.2 Semantisches Differential zur Funktionalität von SmartKom

In diesem Abschnitt wurden diverse Aspekte zur Funktionalität von SmartKom abgefragt. Auch hier sind positive Aussagen mit positiven Werten verknüpft. Die Skala reicht von +3 (Neigung zur positiven Antwort) bis zu –3 (Neigung zur negativen Antwort). Unentschiedenheit kann durch ankreuzen der mittleren Antwortmöglichkeit ausgedrückt werden. Bei der Aufgabenunterstützung gibt es zum Teil sehr weit gestreute Wertungen. So ist die Kompliziertheit des Systems von allen VPn exakt symmetrisch bewertet worden.

Einigkeit besteht allerdings in der nicht allzu berauschenden Informationsqualität, der wiederholten Anwendungen, der unzureichenden Hilfefunktion und der langsamen Reaktionen auf gesprochene und gestische Befehle für den Benutzer im eher negativen Sinn, im positiven Sinn wurde die Übersichtlichkeit des Displays gewertet (11 von 20 mit Werturteilen gut bis sehr gut!).

Unterstützt SmartKom die Erledigung Ihrer Aufgabe; ohne Sie unnötig zu belasten?	Sehr schlecht			–	Sehr gut			Mean	Median
	–3	–2	–1	0	1	2	3		
ist unkompliziert zu bedienen	1	3	3	4	3	3	1	–0.05	0
bot in Anzahl und Qualität angemessene Informationen an	3	2	5	5	5	0	0	–0.65	–0.5[3]
benötigt Eingaben nur einmal; um zum Ziel zu gelangen	3	4	3	5	3	2	0	–0.65	–0.5

2 „Die einzigen zwei positiven Durchschnittswerte beziehen sich auf die Aufgabe und die Verständlichkeit des Systems, wobei selbst die positive Bewertung nicht sonderlich extrem sind." (vgl. Memo12)

3 Mediane auf .5 endend liegen genau zwischen zwei Bewertungen

Unterstützt SmartKom die Erledigung Ihrer Aufgabe; ohne Sie unnötig zu belasten?	Sehr schlecht			–	Sehr gut			Mean	Median
	–3	–2	–1	0	1	2	3		
bietet eine gute Hilfefunktion an	12	2	3	3	0	0	0	–2.15	–3
hat ein übersichtlich gestaltetes Display	0	2	1	3	3	9	2	1.1	2
reagierte schnell auf gesprochene Befehle	6	3	6	1	2	1	0	–1.45	–1.5
reagierte schnell auf Zeigen	5	4	7	1	2	1	0	–1.3	–1

Im Gegensatz zur vorjährigen begleitenden Evaluation, wo überwiegend negative Bewertungen des Systems zu finden waren, ist die Tendenz bei den vorliegenden Daten eher neutral bis positiv. Auch ist hierbei die durchaus positive Bewertung der Übersichtlichkeit des Displays (hervorgehoben) zu beachten.

3.2.3 Bewertung der Persona Smartakus

In diesem Kapitel werden verschiedene Aspekte der Persona näher abgefragt. Dazu zählt die wahrgenommene Kompetenz und die Natürlichkeit der Aktionen.

Die Persona wirkt mit –0.73 auf einer Skala von +3 bis –3 etwas inkompetent, was eine leicht negativere Tendenz im Vergleich zur Vorstudie (Memo12) darstellt.

Auch die Natürlichkeit der Aktionen der Persona schnitt mit einem Mittelwert von –0.4 auf der selben Skala schlechter ab als in der vorherigen Fragewelle.

Wie beurteilen Sie die Kompetenz von Smartakus? [4]	nicht Kompetent			–	Kompetent			Mean	Median
	–3	–2	–1	0	1	2	3		
	2	4	2	2	5	0	0	–0.73	–1

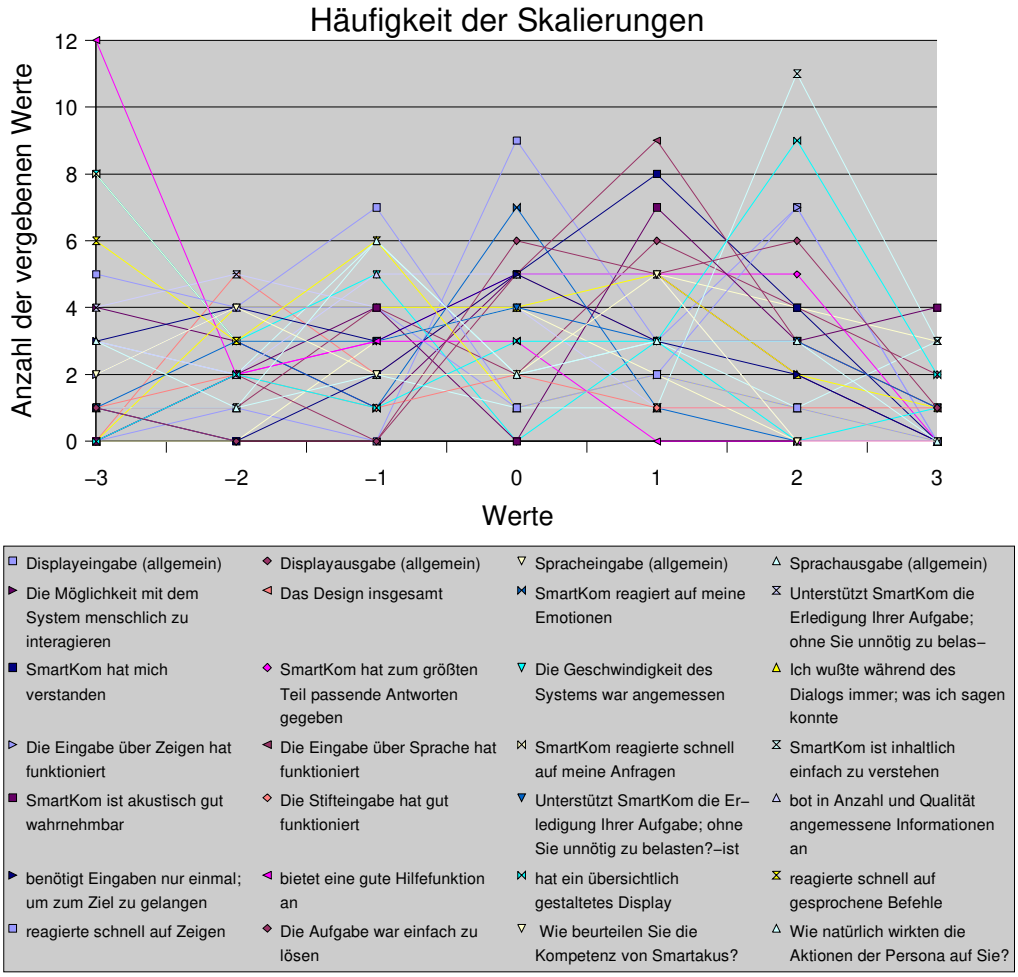
Wie natürlich wirkten die Aktionen der Persona auf Sie?	nicht Natürlich			–	Natürlich				
	–3	–2	–1	0	1	2	3		
	3	1	6	2	3	3	0	–0.4	–0.5

Daß die VPn zur Kompetenz und Natürlichkeit der Persona eher negativ eingestellt sind, liegt wahrscheinlich daran, daß es sich im Gegensatz zur vorangehenden Home–Evaluation diesmal um eine reine

4 5 Probanden gaben zur Kompetenz kein Urteil ab.

Expertenevaluation handelt. Die Probanden haben insgesamt eine höhere Erwartungshaltung gegenüber Dialogsystemen.

3.3 Allgemeines



Die Häufigkeit der Skalierungen aller kategorisierbaren Fragen kann aus vorangegangener Graphik zu jedem Fragekomplex entnommen werden. Das genaue Rating pro Session und Frage kann den beiden Graphiken am Ende des Dokuments entnommen werden.

Die beiden folgenden Fragen konnten mit kategorisierten Antworten „Ja/ Nein/ Weiß nicht" beantwortet werden, werden allerdings nicht mit den objektiv meßbaren Ergebnissen aus den Logfiles verglichen. Sie lauten:

Wurde Ihnen die Bedienung des Systems SmartKom von der Versuchsleitung erklärt?

und

War diese Erklärung allein ausreichend; um das System zu bedienen?

Die erste Frage wurde von 19 Probanden bejaht, eine Versuchsperson

verzichtete auf eine Erklärung durch die Versuchsleitung, da ihr das System bereits bekannt war. Sie verneinte daher die Frage. Für 17 VPn war diese Erklärung ausreichend, um das System im Rahmen der Möglichkeiten zu bedienen, 2 beantworteten diese Frage mit „weiß nicht" und eine VPn mit „nein".

4 Fazit

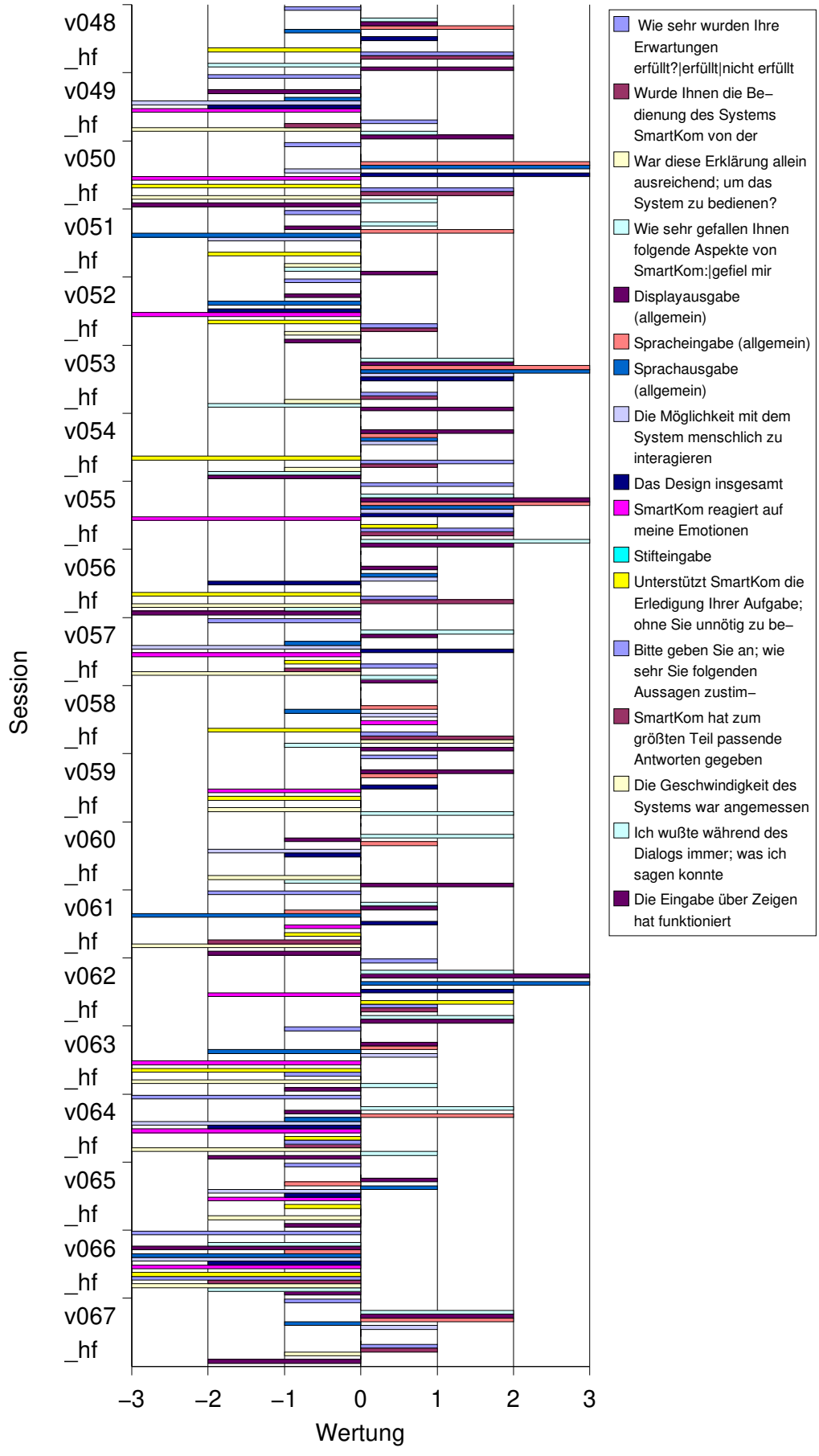
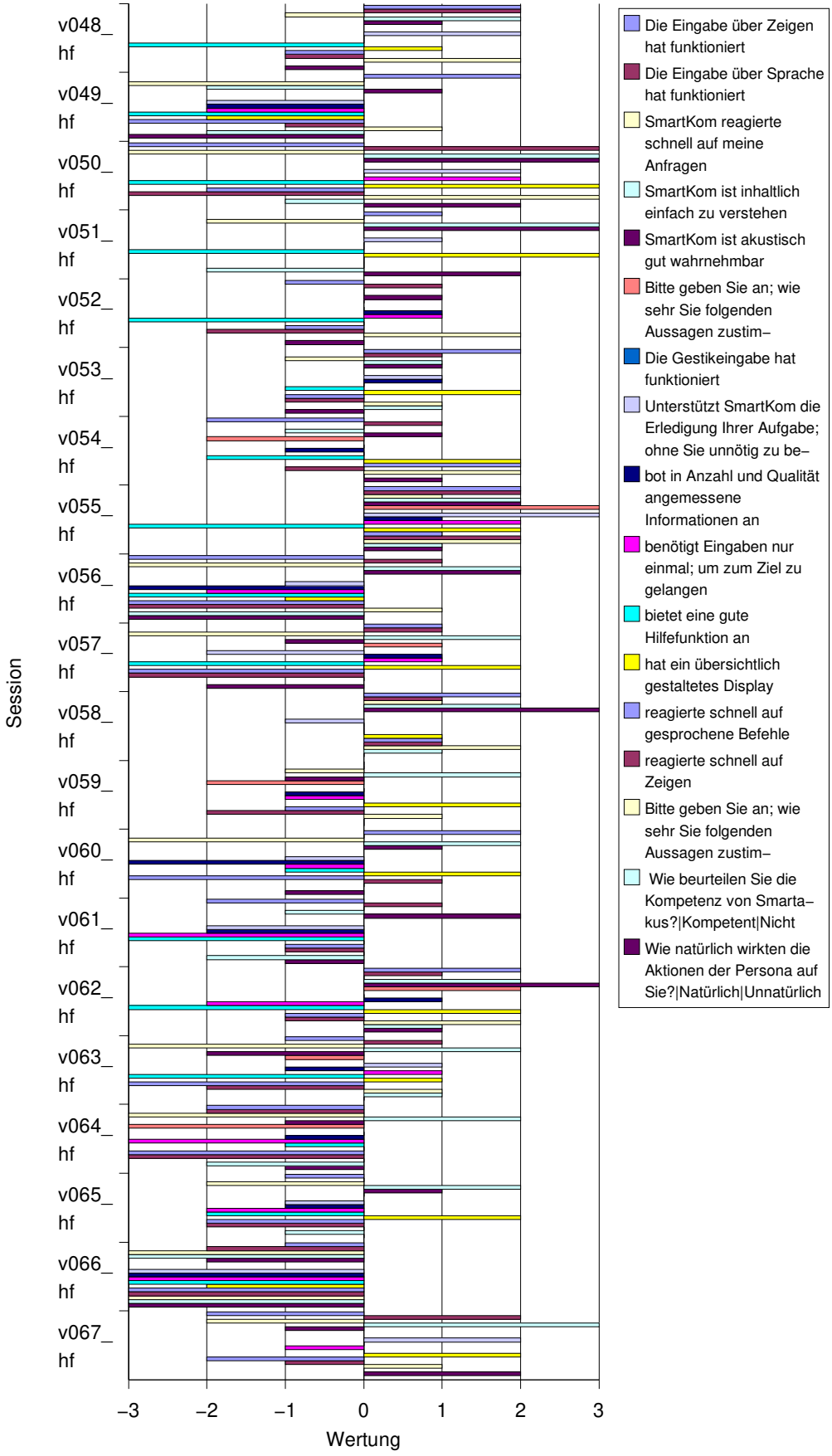
Der SmartKom Prototyp, Szenario Home, wurde an 20 Versuchspersonen getestet. Die hier detailliert ausgewertete Benutzerbefragung ergab ein allgemein positiveres Bild des Prototypen als bei der begleitenden Evaluation Home.

Allgemein schlecht wurde die lange Wartezeit zwischen Befehlseingabe und Reaktion des Systems beanstandet. Daran angeschlossen die Tatsache, daß es keine Benutzerführung gab. Die immer noch ungelöste Frage, was denn nun „nicht verstanden" wurde, ist bei den Probanden ebenfalls nicht gut gewertet worden.

Allerdings zeigte das System insbesondere in der Sprachverarbeitung (Sprachverstehen und weitere Dialogführung) sowie in der Reaktion auf gestische Eingaben einen enormen Fortschritt zur Vorgängerevaluation.

Im Hinblick darauf gilt es zu hoffen, daß die Weiterentwicklung des Basisystems im Bereich der Informationsdarbietung für weitere Prototypen ebenfalls noch bis Projektende wesentliche Erfolge verzeichnen kann.

Verteilung der Urteile auf der 7−stufigen Werteskala zu den Home−Sessions



Verteilung der Urteile auf der 7−stufigen Werteskala zu den Home−Sessions