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

Auswertung der ergonomischen Evaluation des SmartKom Prototypen

–Szenario Public–

Befragungszeitraum 11. Juni 03 bis 13. Juni 03

Nicole Beringer, Irene Jacobi, Sibylle von Tiedemann

Ludwig Maximilians Universität München

Memo Nr. 17

Juli 2003

Juli 2003

Auswertung der Ergebnisse der Evaluations–Aufnahmen Szenario Public

Fragebogenversion vom 10.03.03

Nicole Beringer, Irene Jacobi, Sibylle von Tiedemann

Ludwig–Maximilians–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 Public/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
2.3 Geschlecht....5
2.6 Sonstige Variablen....6
3 Ergonomische Evaluation des Systems....6
3.1 Freie Fragen....6
3.2 Fragen mit kategorisierten Antwortmöglichkeiten....11
3.2.1 Erwartungen, Aspekte, Zustimmung....11
3.2.2 Semantisches Differential zur Funktionalität von SmartKom....13
3.2.3 Bewertung der Persona Smartakus....14
3.3 Allgemeines....15
4Fazit....17

1 Befragungszeitraum

Die Befragung wurde während des Evaluationszeitraumes für das Szenario Public vom 11.06.03 bis einschließlich 13.06.03 durchgeführt. Diese Memo befaßt sich ausschließlich mit den Befragten, die den SmartKom Prototypen Public getestet haben.

2 Anzahl und Zusammensetzung der Befragten

Insgesamt wurden 11 Personen befragt. Die folgenden Daten konnten alle bewertet werden und beziehen sich damit auf 11 gültige Fragebögen zur abschließenden Evaluation des Public–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 dem vorgestellten System. Alle Versuchspersonen sind allgemein an die Bedienung von Dialogsystemen gewöhnt und nahmen bereits an der abschießenden Evaluation Home als Probanden teil.

Die Befragung von ausschließlich nicht–naiven Versuchspersonen hat maßgebliche 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. 30.45 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	11
40–60	0	31.2	0
60 plus	0	26.9	0
Summe	100	100	n=11

Hier wird ersichtlich, daß die angestrebteRekrutierung 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

5 weibliche und 6 männliche Versuchspersonen teilen die Befragten ungefähr paritätisch.

2.4 Staatsangehörigkeit

2 der 11 Befragten gaben an, keinen deutschen Paß zu besitzen. Dies sind 27.27% aller Befragten dieser Welle. Damit ist die angestrebte Verteilung von ca. 8% „Inhaber eines nicht–deutschen Passes" um mehr als das Dreifache 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	11
Ohne Abschluß	0	–	0
Keine Angaben	0	–	0
Sonstige	0	–	0
Summe	100	100	11

[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 (von der) angestrebten Verteilung entsprechend den Daten des Statistischen Bundesamtes.

2.6 Sonstige Variablen

Mit 9 Personen sind die meisten Probanden deutsche Muttersprachler. Zwei Versuchspersonen gaben an, 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 sowie wichtige Zitate[1] 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	Anwendungen
Spontansprache: 7	Kinoprogramm: 1	Kinokarten reservieren: 2
Gestik: 5	Kultur allgemein: 1	Telefon/Fax: 2
	Wegbeschreibung: 4	
	Allgemeine Infos: 3	

Für die meisten Probanden war die Eingabemöglichkeit über Spontansprache wichtig. Die Bedienung durch Gestik wurde im Vergleich zu Home öfters (5 vs. 2 bei Home) genannt. Bei den Informationsmöglichkeiten bezogen auf die Anwendung lag der Schwerpunkt der Gewichtung auf der Wegbeschreibung. 3 der VPn wünschten sich allgemeine Informationen, je eine VP erwartete Informationen zu Kino und Kultur allgemein. Mögliche Anwendungen sahen die Probanden bei der Kartenreservierung (2 Nennungen) und bei der Bedienung von Telefon und Fax (ebenfalls 2 Nennungen). Zusätzlich zu Eingabe– und Informationsmöglichkeiten bzw. Gerätesteuerung lassen sich unter Systemeffizienz folgende

1 Eine automatische Rechtschreibkorrektur wurde auch auf die Zitate angewendet.

Erwartungen zusammenfassen:

- Geschwindigkeit: 5
- Benutzerfreundliche Bedienung: 7

Zusammenfassend läßt sich sagen, daß die Probanden alle ein System wünschen, das die jeweiligen Anforderungen 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: 2	Erwünschter Modus: Gestik oder Sprache 1
Sprache: 1	Wichtelhandlungen: 1
Display (wohin clicken) eindeutig: 1	verbale Aufforderung / durch Frage–Antwortrhythmus: 1
verbale Aufforderung / durch Frage–Antwortrhythmus: 1	Unklar, ob verstanden: 2
wg. eingeschränkter Mögl.: 1	Unklar, was möglich: 1
Allgemein: 5	Display (wohin clicken): 2
Feedback: 3	Ausgabe unklar: 3
keine Angabe: 1	Hat Sätze nicht verstanden:1

Eindeutigkeit wurde bei den Versuchspersonen erfüllt durch Vorwissen/Instruktion (2 Nennungen), durch Sprache (1), durch eindeutiges Display (1), durch die verbale Aufforderung bzw. durch Frage–/Antwortrhythmus (1) und sogar durch eingeschränkte Interatkionsmöglichkeiten (1). Drei Probanden sahen die Eindeutigkeit durch Feedback erfüllt. Fünf VPn fanden die Bedienung eindeutig, benannten aber das Wie nicht näher. Ein Proband machte keine Angaben.
Das System wurde von den Probanden als nicht eindeutig bedienbar empfunden, weil auf den erwünschten Modus nicht eingegangen wurde (1), die Wichtelhandlungen verwirrend sind (1), es Unklarheit über das Fehlen des Verständnisses durch das System (3) bzw. über die tatsächlichen Bedienmöglichkeiten (3) gab und die verbale Aufforderung durch Frage–Antwortrhythmus verwirrend war (1). Drei empfanden die Ausgabe als unklar. Z.B. ist nicht klar, was mit der Ausgabe, was das System verstanden hat, angefangen werden kann. Auch das Angebot der Alternativen wird nicht weiter erklärt (sprachliche Referenz darauf erforderlich, nur Zeigen funktioniert nicht). Kritisiert wurde von einer Person, daß es unklar sei, wann was gesagt wurde und das System „nicht versteht, was ich mit Karten reservieren meine, sondern nur auf Plätze reservieren reagiert".

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

ja	nein
Wegbeschreibung: 2	Nein: 8
Faxen: 1	

8 Probanden gaben an, keine Funktion genutzt haben zu können, die nicht explizit von der Versuchsleitung genannt wurde. Die restlichen Probanden hatten selbständigen Erfolg beim Erhalt einer Wegbeschreibung (2) und beim Versenden eines Fax (1), keine Angabe (3).

Inwiefern half SmartKom Ihnen, die Bedienung des Systems zu erlernen?

gar nicht	Folgendes half
Gar nicht: 7	Feedback: 3 (2 davon allerdings zu selten)
Keine Hilfefunktion: 1	
Schlechtes Feedback: 2	
Versuch und Irrtum: 2	

7 Versuchspersonen gaben an, keinerlei Hilfe bei der Bedienung durch das System erhalten zu haben. Ein Proband beklagte die immer noch fehlende Hilfefunktion. Bedienhilfen durch das System erfolgten durch Feedback (3), wenn es auch zu selten und schlecht war (2) bzw. durch „Versuch und Irrtum" (2). Ein Proband wies auf die fehlende Rückkehrfunktion hin. Diese Ironisierung bzw. Abweichung von der eigentlichen Fragestellung läßt den Schluß zu, daß die Versuchspersonen keine wirkliche Hilfe bei der Bedienung durch das System erfahren konnten.

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

erfüllt	nicht erfüllt
Gut bis sehr gut: 2	Syntheseausgabe: 'ich habe Sie nicht verstanden' 1
Mittel 3	Syntheseausgabe: falsche Antwort 3
Lässt sich unterbrechen: 1	Geschwindigkeit: zu langsam 1
	Gestikerkennung: 1
	Sitzplatzreservierung: 3
	Zurück(button) fehlt: 3
	Allg: 1

Insgesamt schienen die Versuchspersonen mit der Verarbeitung ihrer

Befehle zufrieden zu sein. So fanden 2 VPn die Antworten gut bis sehr gut erfüllt, 3 waren im großen und ganzen zufrieden und eine war froh, daß sich das System unterbrechen ließ. 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 (1) , der fehlende Zurückbutton (3), die nicht so wie erwartet funktionierende Sitzplatzreservierung (3) und die mangelnde Gestikerkennung (1) störend. Ein Proband sah seine Befehle und Wünsche allgemein nicht erfüllt. Und schließlich bemängelte eine VP, daß die Korrektur von ausgewählten Sitzplätzen, etc. nicht geklappt hat.

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

ja	nein
verbal (Wortwahl / Satzbau/ Artikulation / Geschwindigkeit) 1	Keine/ nicht bewußt: 2
Gesten „evtl. Antippen von Sitzen in der Reservierung nicht klar (zeigt man auf die Sitze, unter die Sitze?" 1	
Sprachausgabe nicht abgewartet 6	

Die Frage nach Bedienfehlern wurde von 2 VPn mit „nein" bzw. „nicht bewußt" beantwortet. In 7 Fällen gaben die VPn, sich verbaler „Fehler" (Wortwahl / Satzbau/ Artikulation / Geschwindigkeit, Räuspern, Sprachausgabe nicht abgewartet) bewußt zu sein, 1 verwies auf Gesten („evtl. Antippen von Sitzen in der Reservierung nicht klar (zeigt man auf die Sitze, unter die Sitze?")). Eine Versuchsperson gab folgende Aussage:
„Ich habe zu menschlich reagiert, das heißt, ich war trotz ewig lange Wartepausen noch zu schnell, aber eigentlich ist das nicht mein Fehler, sondern Fehler des Systems, die Ungeeignetheit im Feld menschliche Kommunikation eingestetzt werden zu können."

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

gar nicht	half
Gar nicht: 5	0 Nennungen
nur 'ich habe Sie nicht verstanden': 4	
ohne Angabe: 2	

Es zeigt sich bei dieser Frage wie auch schon in der vorangegangenen abschließenden Evaluation zum Anwendungsszenario Home, daß sich die VPn bei der Erkennung evtl. Fehler und deren Korrekturmöglichkeiten im Stich gelassen fühlen. 5

gaben daher an, keine Hilfe bekommen zu haben, 4 sahen im 'Ich habe Sie nicht verstanden" (Verb???) und 2 gaben keine Begründung. Zwei Probanden formulierten wie folgt:

„Die Anzeige, was er verstanden hat, ist schon mal ein guter Weg, aber was mit dieser Anzeige anfangen? Es gibt keine Hilfestellung, wie man weiterverfahren kann."

„Nur durch „Ich habe sie nicht verstanden", aber auch das war nicht hilfreich, wenn das System sich wie eine Schallplatte mit Sprung verhält und immer wieder dieselben Sätze wiederholt, ohne daß ich etwas gesagt oder getan habe."

Inwiefern reagierte SmartKom eventuell trotz Bedienfehler wie gewünscht?

reagierte nicht	reagierte
Gar nicht : 6	Trotz Unterbrechungen: 2
weiß nicht: 1	Sehr Verspätet: 3

Bei 6 Probanden reagierte SmartKom gar nicht wie gewünscht, bei 2 VPn trotz Unterbrechungen bzw. sehr zeitverzögert (3).
Eine VP wußte es nicht.

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

ja	nein
Display eindeutig: 2	Display unmotiviert/ Wichtel lächerlich: 2
Sprache eindeutig: 2	Sprache: Aussprache, Prosodie 5
	Sprache Bedeutung unklar "Ich berechne die Karte" 1 "Welche Alternative wollen Sie" 2
Allgemein: 1	Mangelnde Koordination zwischen Sprach–Graphik–Ausgabe 1
	Mangelndes Feedback: 2

Bei der Verständlichkeit der Ausgabe ist eine VP im allgemeinen zufrieden. Zumindest die Displaygestaltung scheint öfters eindeutig (2), wenn auch zwei VPn die Displayausgabe als unmotiviert und den Wichtel als lächerlich deklarieren. Bei der Sprachausgabe bemängelten 5 Probanden die prosodische Modellierung und Bedeutung (3: "Ich berechne die Karte"; "Welche Alternative wollen Sie"). Ein Proband beklagte sich über eine mangelnde Koordination zwischen Sprach–Graphik–Ausgabe und zwei über mangelndes Feedback. Eine VP machte keine Angabe.

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

ja	nein
Ja: 2	Zeit: zu langsam 8
	Aufwand zu hoch für zu wenig Infos: 2
	Plätze Reservieren: 3
	Aufwand: kein Zurückbutton 1
	Displayeingabe unklar: 1
	Falsche Ausgaben: 2

Diese Frage zeigt deutlich, daß die Probanden herkömmliche Methoden bevorzugen, weil sie schneller (8) und mit geringerem Aufwand (2) zum Ziel führen. Gerade bei der Platzreservierung sind herkömmliche Methoden für 3 Probanden effektiver. 4 Probanden beklagten sich über den Mehraufwand, den der fehlende Zurückbutton mit sich bringt (1), daß die Displayeingabe unklar ist (1) und daß falsche Ausgaben nicht zum Ziel führen (2).

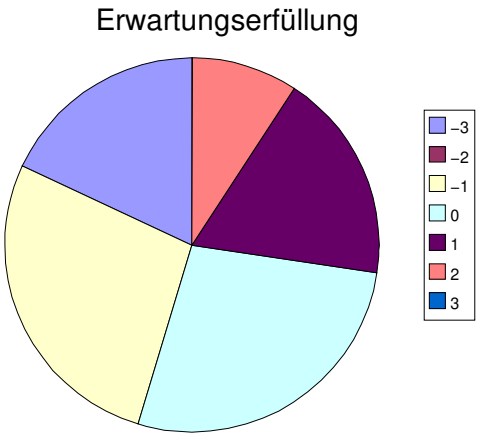
Im Gegensatz zur Vorgängerevaluation Home sahen sich jedoch 2 Probanden in der Lage, die ihnen gestellte Aufgabe im Vergleich zu herkömmlichen Methoden in einem dafür angemessenen Zeit– und Aufwandsrahmen zu erfüllen.

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 eine leicht negative Erwartungserfüllung durch das System. Der Median liegt mit 3 VPn bei einer Erwartungserfüllung von 0, 3 VPn vergaben –3, 3 vergaben eine Erwartungserfüllung von –1. Positive Erwartungserfüllungen waren bei 3 VPn zu verzeichnen (2 x 1 und 1 x 2), 5 VPn vergaben den Wert 0. Im Vergleich zur abschließenden Evaluation Home (Memo20) schienen die Versuchspersonen etwas positiver gegenüber dem System eingestellt zu sein. 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)	2	0	2	2	2	2	1	0.09	0
Displayausgabe (allgemein)	2	0	0	2	0	5	2	0.91	2
Spracheingabe (allgemein)	1	3	1	1	2	3	0	–0.18	0
Sprachausgabe (allgemein)	0	3	0	1	2	5	0	0.55	1
Die Möglichkeit mit dem System menschlich zu interagieren	3	1	0	2	2	2	1	–0.18	0
Das Design insgesamt	2	1	1	2	1	3	1	0.09	0
SmartKom reagiert auf meine Emotionen	6	0	1	3	1	0	0	–1.64	–3
Stifteingabe	nicht genutzt								

Besonders positiv zu werten sind hier die Ausgabemodalitäten, besonders negativ war die Bewertung der Reaktion auf Emotionen. Hier konnte nur von einer Person eine leicht positive Wertschätzung verzeichnet werden. Im Vergleich zur abschließenden Evaluation Home scheinen die Probanden insgesamt leicht positiver eingestellt zu sein.

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		
	2	1	2	1	3	2	0	–0.27	0

Auch die fehlende Möglichkeit zu unterbrechen wurde von den VPn bemängelt. Im Vergleich zur abschließenden ergonomischen Auswertung Home zeigt sich allerdings eine deutlich bessere Beurteilung (positive vs. negative Stimmvergabe hält sich die Waage).

Beim nachfolgenden Fragenkomplex wurde bei der Zustimmung zu folgenden Aussagen 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	2	1	2	1	3	2	0	–0.27	0
SmartKom hat zum größten Teil passende Antworten gegeben	1	2	0	1	5	0	0	0	1
Die Geschwindigkeit des Systems war angemessen	4	0	1	3	3	0	0	–0.91	0
Ich wußte während des Dialogs immer; was ich sagen konnte	2	1	2	2	3	1	0	–0.45	0
Die Eingabe über Zeigen hat funktioniert	2	1	3	2	2	1	0	–0.64	–1
Die Eingabe über Sprache hat funktioniert	2	1	0	1	6	1	0	0	1
SmartKom reagierte schnell auf meine Anfragen	4	0	0	5	2	0	0	–0.91	0
SmartKom ist inhaltlich einfach zu verstehen	1	0	0	3	5	1	1	0.64	1
SmartKom ist akustisch gut wahrnehmbar	0	1	2	2	2	2	2	0.73	1
Die Gestikeingabe hat funktioniert	0	3	1	4	1	1	0	–0.4	0
Die Aufgabe war einfach zu lösen	1	2	1	0	3	3	1	0.36	1

Leider immer noch ein Manko aus der Sicht der VPn ist das Verständnis durch das System, die Langsamheit und der Erfolg der Gestikeingabe. Positiv hervorzuheben im Vergleich zur abschließenden Evaluation Home ist auf jeden Fall eine überwiegend positive Bewertung der akustischen Wahrnehmbarkeit und der leichteren Aufgabenlösung. Auch bei der abschließenden Evaluation Public zeigt sich wie schon bei 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 (7 von 11 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	0	1	3	2	1	0.09	1
bot in Anzahl und Qualität angemessene Informationen an	1	0	2	2	4	1	0	–0.18	0
benötigt Eingaben nur einmal; um zum Ziel zu gelangen	3	1	3	2	2	0	0	–1.09	–1
bietet eine gute Hilfefunktion an	6	2	0	2	0	1	0	–1.82	–3
hat ein übersichtlich gestaltetes Display	0	1	1	1	2	1	4	1	1
reagierte schnell auf gesprochene Befehle	4	1	0	4	2	0	0	–1.09	0
reagierte schnell auf Zeigen	4	1	1	4	1	0	0	–1.27	–1

Tendentiell handelt es sich bei den vorliegenden Daten um eher neutrale bis positive Wertungen, die allerdings nicht ganz so gut ausfiel wie noch bei der abschließenden Evaluation Home. Dennoch ist auch 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.

Wie beurteilen Sie die Kompetenz von Smartakus? [2]	nicht Kompetent			–	Kompetent			Mean	Median
	–3	–2	–1	0	1	2	3		
	3	1	2	3	1	0	0	–1.2	–1
Wie natürlich wirkten die Aktionen der Persona auf Sie?	nicht Natürlich			–	Natürlich				
	–3	–2	–1	0	1	2	3		
	1	2	2	2	4	0	0	–0.45	0

2 1 Proband gab zur Kompetenz kein Urteil ab.

Die Persona wirkt mit –1.2 auf einer Skala von +3 bis –3 etwas inkompetenter als noch bei Home (Mean: –0.73) , was eine leicht negativere Tendenz im Vergleich zur Vorstudie (Memo20) darstellt.

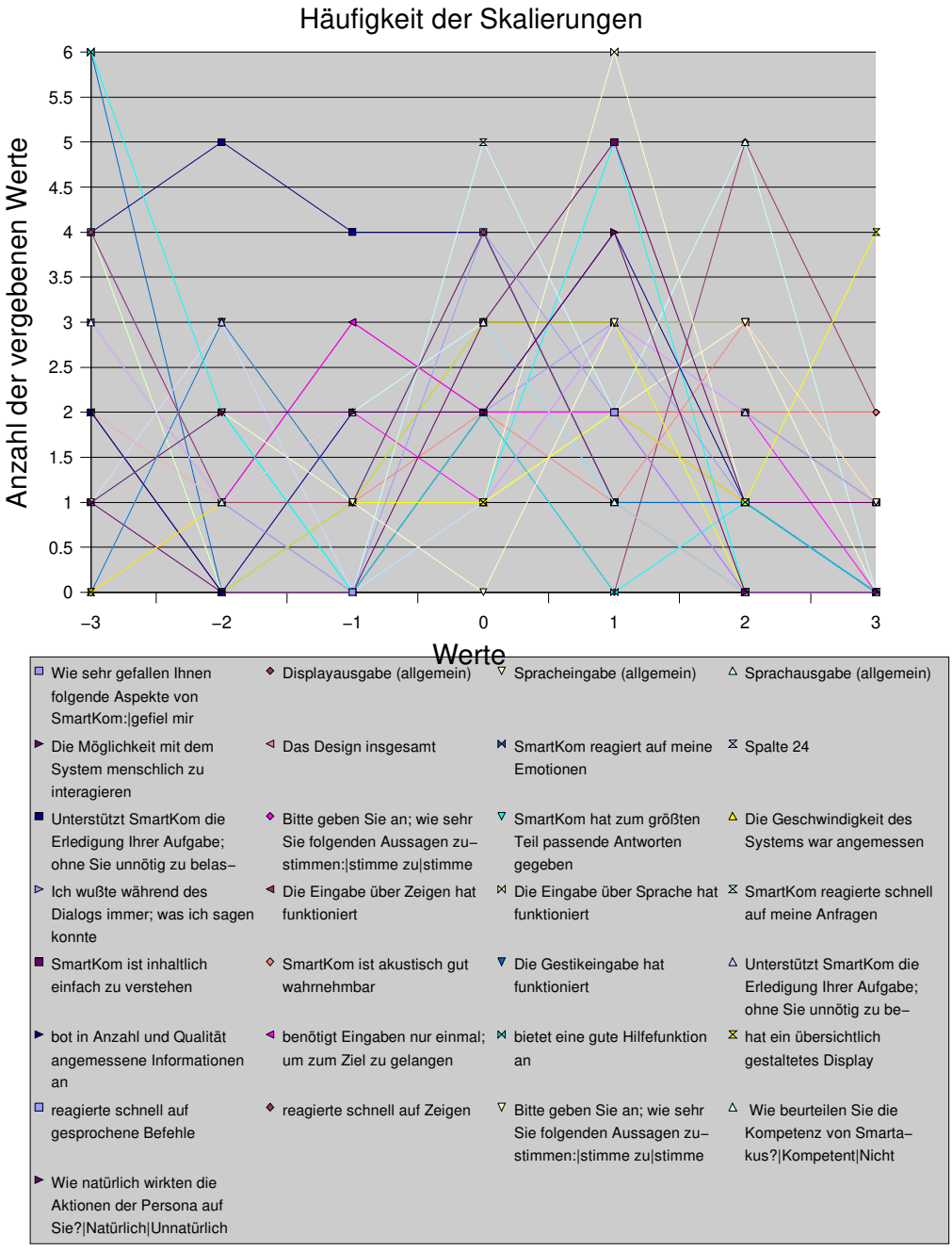
Auch die Natürlichkeit der Aktionen der Persona schnitt mit einem Mittelwert von –0.45 auf der selben Skala leicht schlechter ab als in der vorherigen Fragewelle.

Auffallend hierbei ist, daß keiner der Probanden eine bessere Wertung als 1 abgab, wohl aber bis –3 gewertet wurde.

Wie auch schon bei der abschließenden Home Evaluation läßt sich die eher negative Einstellung der Probanden zur Kompetenz und Natürlichkeit der Persona darauf zurückführen, daß es sich um eine reine Expertenevaluation handelt. Die Probanden haben insgesamt eine höhere Erwartungshaltung gegenüber Dialogsystemen.

3.3 Allgemeines

Die Häufigkeit der Skalierungen aller kategorisierbaren Fragen kann aus folgender 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 allen 11 Probanden bejaht. Für 8 VPn war diese Erklärung ausreichend, um das System im Rahmen der

Möglichkeiten zu bedienen, 2 beantworteten diese Frage mit „nein" und eine VPn mit „weiß nicht".

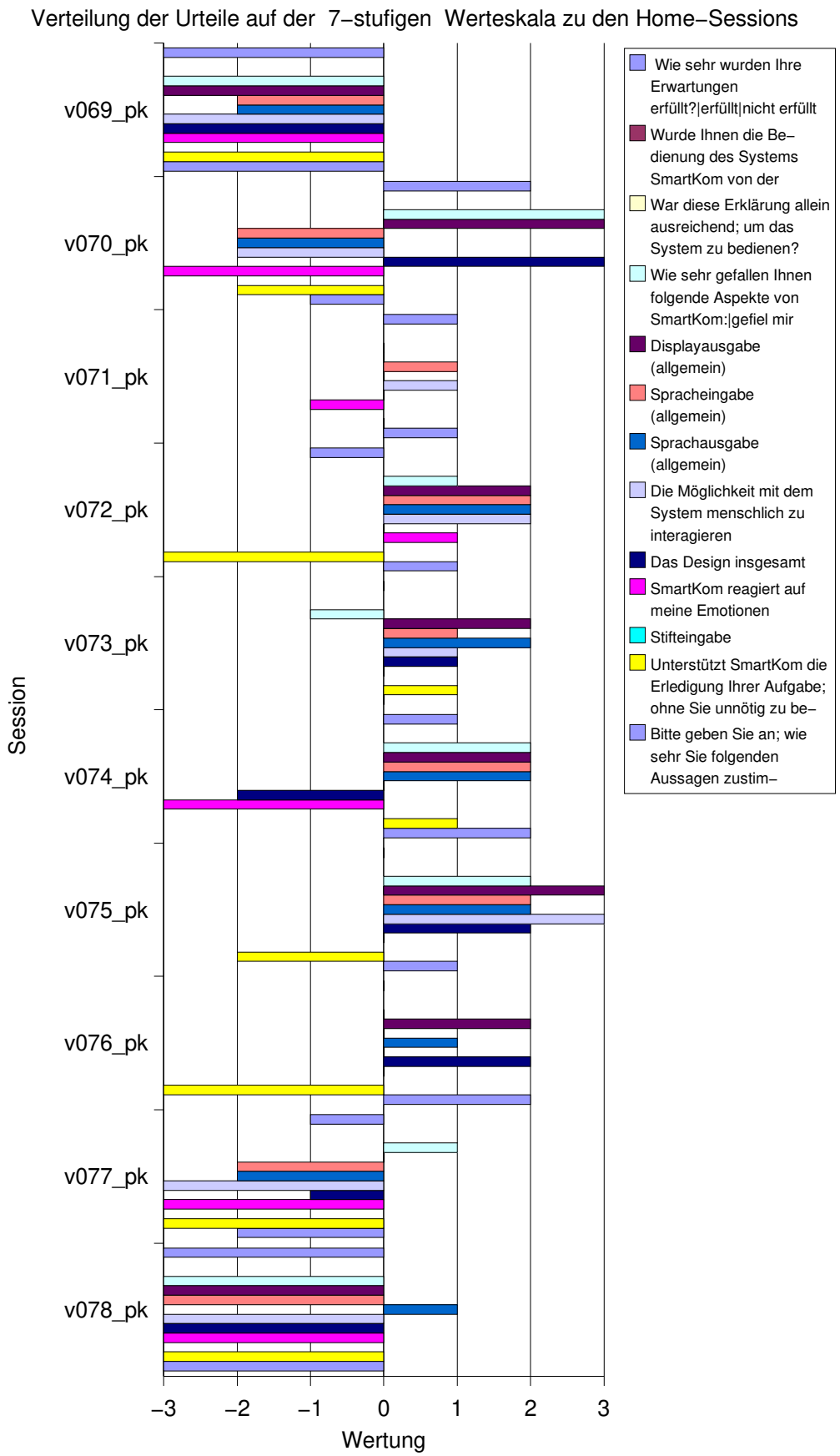
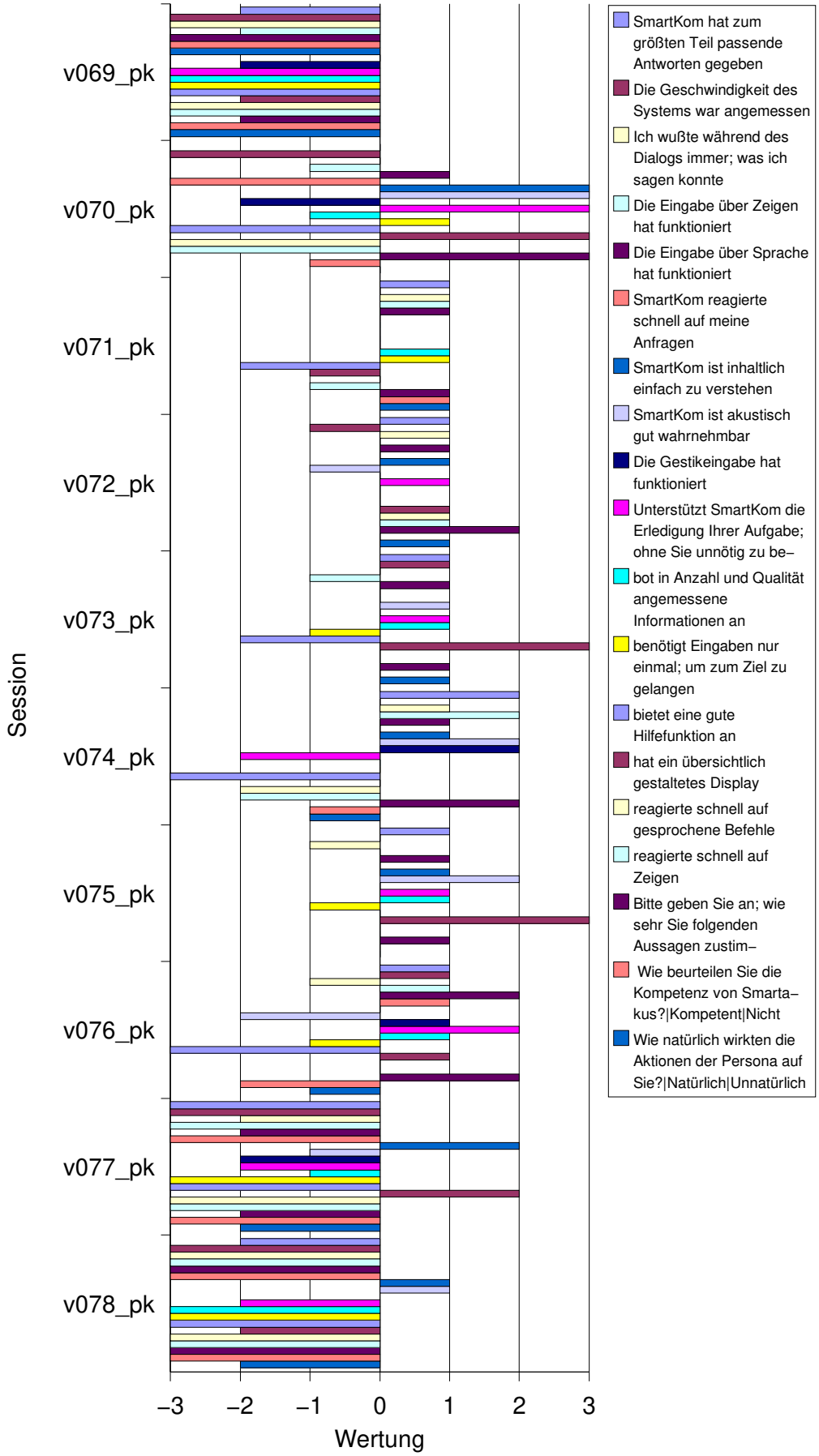
4 Fazit

Der SmartKom Prototyp, Szenario Public, wurde an 11 Versuchspersonen getestet. Die hier detailliert ausgewertete Benutzerbefragung ergab ein allgemein positiveres Bild des Prototypen als bei der begleitenden Evaluation Public.

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.

Abschließend sollte noch positiv vermerkt werden, daß sich zumindest 2 Probanden in der Lage sahen, die ihnen gestellte Aufgabe im Vergleich zu herkömmlichen Methoden in einem dafür angemessenen Zeit– und Aufwandsrahmen zu erfüllen.

Der SmartKom Prototyp scheint also trotz der immer noch nicht zufriedenstellendenInformationsdarbietung langsam in den Bereich der effizienten Bedienbarkeit zu gelangen.


Verteilung der Urteile auf der 7−stufigen Werteskala zu den Home−Sessions
v069_pk
v070_pk
v071_pk
v072_pk
v073_pk
v074_pk
v075_pk
v076_pk
v077_pk
v078_pk
Session
−3
−2
−1
0
1
2
3
Wertung
Wie sehr wurden Ihre Erwartungen erfüllt?|erfüllt|nicht erfüllt
Wurde Ihnen die Be−dienung des Systems SmartKom von der
War diese Erklärung allein ausreichend; um das System zu bedienen?
Wie sehr gefallen Ihnen folgende Aspekte von SmartKom:|gefiel mir
Displayausgabe (allgemein)
Spracheingabe (allgemein)
Sprachausgabe (allgemein)
Die Möglichkeit mit dem System menschlich zu interagieren
Das Design insgesamt
SmartKom reagiert auf meine Emotionen
Stifteingabe
Unterstützt SmartKom die Erledigung Ihrer Aufgabe; ohne Sie unnötig zu be−
Bitte geben Sie an; wie sehr Sie folgenden Aussagen zustim−

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