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

SmartKom
Mensch-Technik-Interaktion

Auswertung der ergonomischen Evaluation des SmartKom Prototypen

–Szenario Mobil–

Befragungszeitraum 04. Juli 03 bis 11. Juli 03

Nicole Beringer, Irene Jacobi, Sibylle von Tiedemann

Ludwig Maximilians Universität München

Memo Nr. 18

Juli 2003

Juli 2003
Auswertung der Ergebnisse der Evaluations–Aufnahmen Szenario Mobil
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 Mobil/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

Die Befragung wurde während des Evaluationszeitraumes für das Szenario Mobil vom 04.07.03 bis einschließlich 12.07.03 durchgeführt. Diese Memo befaßt sich ausschließlich mit den Befragten, die den SmartKom Prototypen Mobil getestet haben.

2 Anzahl und Zusammensetzung der Befragten

Insgesamt wurden 10 Personen befragt. Die folgenden Daten konnten alle bewertet werden und beziehen sich damit auf 10 gültige Fragebögen zur abschließenden Evaluation des Mobil–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.

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 26 und 40 Jahren (im Schnitt bei ca. 30.8 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	10
40–60	0	31.2	0
60 plus	0	26.9	0
Summe	100	100	n=10

Hier wird ersichtlich, daß die angestrebte 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

5 weibliche und 5 männliche Versuchspersonen teilen die Befragten paritätisch.

2.4 Staatsangehörigkeit

Einer der 10 Befragten gab an, keinen deutschen Paß zu besitzen. Dies sind 10% aller Befragten dieser Welle. Damit ist die angestrebte Verteilung von ca. 8% „Inhaber eines nicht–deutschen Passes" gegeben.

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	10
Ohne Abschluß	0	–	0
Keine Angaben	0	–	0
Sonstige	0	–	0
Summe	100	100	10

[1]Verteilung beruht auf Personen mit allgemeinbildenden 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 9 Personen sind die meisten Probanden deutsche Muttersprachler. Eine Versuchsperson gab an, Russisch 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 Cluster mit der Anzahl der Nennungen wiedergegeben. Zusätzliche bzw. abweichende Nennungen (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)

Eingabemög-lichkeiten	Informations-möglichkeiten	Anwen-dungen	System-ausgabe
Spontansprache: 2	Sehenswürdigkeiten: 5	Terminkalender :1	Display: gute Lesbarkeit bzw. Zoommög–lichkeit: 2
Gestik: 2	Routenplanung/–berechnung: 2	Abspeichern von Karten + allg.Infos: 1	
Allgemeine Nennung:1	GPS: 2	Tagesprogramm erstellen: 1	
	Kultur allgemein: 2		
	Wegbeschreibung: 6		
	Allg. Infos: 4		

Für die meisten Probanden war die Eingabemöglichkeit über

1 Alle Zitate wurden einer automatischen orthographischen Korrektur unterzogen.

Spontansprache nicht mehr so wichtig wie noch in der Publicversion. Die Bedienung durch Sprache und Gestik wurde jeweils nur zweimal genannt. Bei den Informationsmöglichkeiten bezogen auf die Anwendung lag der Schwerpunkt der Gewichtung auf der Wegbeschreibung (6 Nennungen), gefolgt von Informationen über Sehenswürdigkeiten (5), Routenplanung/–berechnung, GPS und Kultur allgemein lagen je bei 2 Nennungen. Allgemeine Informationen wurden von 4 VPn gefordert. Insgesamt wurden von den Probanden 3 Anwendungen genannt: Terminkalender, Abspeichern von Karten und allgemeinen Informationen, Tagesprogramm erstellen. Zusätzlich sollte das Display gut lesbar sein und eine Zoommöglichkeit geboten werden.
Wie auch schon bei den anderen Szenarien können unter dem Begriff Systemeffizienz folgende Erwartungen zusammengefaßt werden:

- Geschwindigkeit: 5
- robuste/benutzerfreundliche Bedienung: 6

Zusammenfassend läßt sich sagen, daß die Probanden alle ein System wünschen, das „lieber nur wesentliche Infos" gibt (1) (dazu gehört auch eine Maßstabangabe (1)) als zuviel unnötige Detailangaben.

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

EINDEUTIG	NICHT EINDEUTIG
durch Vorwissen / Instruktion: Hinweis auf (eingeschränkte) modale Mögl.) : 2	Spracheingabe: 1
Displayausgabe: 1	Wichtelhandlungen: 3
Display (wohin clicken) eindeutig: 1	verbale Aufforderung / durch Frage–Antwortrhythmus: 4
	Unklar, ob verstanden: 2
wg. eingeschränkter Mögl.: 2	Unklar, was überhaupt möglich: 2
Allgemein: 1	Display (wohin clicken):1
Feedback: 1 (*): "Das System erklärt das, was man sowieso schon dargestellt sieht".	System reagierte nicht: 1

Eindeutigkeit sahen die bei den Versuchspersonen gegeben durch Vorwissen/Instruktion und Hinweis auf (eingeschränkte) modale Möglichkeiten (2 Nennungen), durch die Displayausgabe allgemein (1) bzw. wenn eindeutig klar war, wo man für weitere Informationen hinklicken mußte (1) und sogar durch die eingeschränkten Möglichkeiten (2). Eine VP fand die Bedienung eindeutig, benannte aber das Wie nicht näher. Ein anderer Proband sah das Feedback als eindeutig an: "Das System erklärt das, was man sowieso schon dargestellt sieht".
Das System wurde von den Probanden als nicht eindeutig bedienbar empfunden, weil die Wichtelhandlungen unklar sind (3), die verbale

Aufforderung durch Frage–/Antwortrhythmus unklar war (4), es Unklarheit über das Verständnis durch das System (2) bzw. über die tatsächlichen Bedienmöglichkeiten (2) gab (speziell beim Anklicken des Displays (1)), das System nicht reagierte (1) und die Spracheingabe nicht so funktionierte (1) wie erwartet. Hinweise über entstandene Unklarheiten gaben zwei Probanden: "Die Ausgaben waren oft etwas unklar – z.B. kamen z.B. mehrere Sprachausgaben und nur wenn man dies wußte, war klar, wann man weitermachen kann". Es war "oftmals unklar, warum er eine neue Karte berechnet hat".

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

JA	NEIN
Zoomen: 1	Nein: 9

9 Probanden gaben an, keine Funktion genutzt haben zu können, die nicht explizit von der Versuchsleitung genannt wurde. Eine VP konnte selbständig zoomen.

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

GAR NICHT	FOLGENDES HALF
Gar nicht: 8	rechtzeitiges Hinweisen, wenn nicht verstanden:1
schlechtes Display (wohin clicken?): 1	
Schlechtes Feedback: 1	

8 Versuchspersonen gaben an, keinerlei Hilfe bei der Bedienung durch das System erhalten zu haben. Die Bedienung wurde besonders beim schlechten Feedback (1) bzw. beim schlechten Display (Klickmöglichkeiten unklar) (1) beanstandet. Als Bedienhilfe wurde von einer VP der rechtzeitige Hinweis auf nicht verstandene Eingaben aufgefaßt.

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

ERFÜLLTHEITSGRAD	NICHT ERFÜLLTES SPEZIFIZIERT
Gut bis sehr gut: 1	Syntheseausgabe: 'ich habe Sie nicht verstanden': 1
Mittel: 4	Displayausgabe unübersichtlich: 1
schlecht/zu langsam: 1	Geschwindigkeit: zu langsam: 2
	Gestikerkennung: 1 (Verschieben Landkarte)
	schlechtes Feedback: 1
	Abbruch/Zurück(button) fehlt: 2
	Mauszeiger nicht dort, wo die Stiftspitze ist: 1

Insgesamt schienen die Versuchspersonen mit der Verarbeitung ihrer

Befehle zufrieden zu sein. So fand eine VP die Antworten gut bis sehr gut erfüllt, 4 waren im großen und ganzen zufrieden und nur eine fand die Ergebnisse schlecht. Die Syntheseausgabe 'ich habe Sie nicht verstanden' (1) wurde wie auch schon bei den anderen Szenarios kritisiert. Ferner wurde die Unübersichtlichkeit der Displayausgabe (1), die Langsamheit (2), die falsche Gestikerkennung (Verschieben der Landkarte; 1), das schlechte Feedback insgesamt (1), der fehlende Zurückbutton (2) und die schlechte Übereinstimmung von Mauszeiger und Stiftspitze (1) kritisiert.

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

JA	NEIN
Sprachausgabe nicht abgewartet :2	Keine: 6
„Ja, ich habe zwei oder drei Wörter gesagt, die nicht auf dem Plan standen." 1	schlechte Mauszeiger–Stift–Kalibrierung:1

6 VPn waren der Meinung, keine Bedienfehler gemacht zu haben, eine VP nannte als Grund für potentielle Bedienfehler die schlechte Mauszeiger–Stift–Kalibirierung. 2 Probanden vermuteten in ihrem Nicht–Abwarten der Sprachausgabe und ein Proband in der Tatsache „Ja, ich habe zwei oder drei Wörter gesagt, die nicht auf dem Plan standen."eigene gemachte Fehler.

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

GAR NICHT	HALF
Gar nicht: 9	durch Feedbackfragen: 1
nur 'ich habe Sie nicht verstanden': 2	

Es zeigt sich bei dieser Frage wie auch schon in beiden vorangegangenen abschießenden Evaluationen, daß sich die VPn bei der Erkennung evtl. Fehler und deren Korrekturmöglichkeiten im Stich gelassen fühlen. 9 gaben daher an, keine Hilfe bekommen zu haben, 2 sahen im 'ich habe Sie nicht verstanden' einen Hinweis auf potentielle Fehler, die sie aber nicht beheben konnten. Nur ein Proband fand Feedbackfragen hilfreich.

Inwiefern reagierte SmartKom eventuell trotz Bedienfehler wie gewünscht?

REAGIERTE NICHT	REAGIERTE
Gar nicht: 8	Trotz Bedienfehler:1
	Sehr Verspätet: 1
	Trotz schlechter Kalibrierung:1

Bei 8 Probanden reagierte SmartKom gar nicht wie gewünscht, bei einer VP reagierte das System trotz Bedienfehler. Ein anderer Proband erhielt eine sehr verspätete Reaktion und eine VP erhielt trotz

schlechter Kalibrierung eine Reaktion.

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

JA	NEIN
Display eindeutig: 1	Wichtel lächerlich/rote Lampe:5
Sprache eindeutig: 2	Sprache: Aussprache, Pausen, Prosodie: 3
Sprach–Display–Koordination: 1	Display: 1
Allgemein: 3	Mangelndes/schlechtes Feedback: 2
	Zuordnung Eingabe–Ausgabe unklar: 1

Bei der Verständlichkeit der Ausgabe sind drei VPn im allgemeinen zufrieden. Das Display wurde von einer VP explizit als eindeutig genannt, die Sprache sogar von 2 VPn. Auch die Sprache–Display–Koordination fand ein Proband eindeutig zu verstehen. Als uneindeutig empfanden die VPn den Wichtel („lächerlich") und die rote Lampe (5), auf der Sprachebene die Aussprache, unpassende Pausen und komische Prosodie (3), das Display (1), das mangelnde und schlechte Feedback (2) und die Zuordnung der Ein– und Ausgabe (1).

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: 5 (Routenplanung 2x positiv erwähnt)	Zeit: zu langsam:4 Zitat:"Das Teil würde ich nicht mal geschenkt haben wollen. Mehr als uneffektiv."
	Aufwand zu hoch: 2
	Menüübersicht fehlt: 1
	Aufwand: kein Abbruch/Zurückbutton: 1
	Allgemeine Nennungen: 2
	keine Hilfefunktion: 1

Im Gegensatz zu allen vorangegangenen Evaluationen konnte die Hälfte der Probanden die Aufgabe in einem dafür angemessenen Zeit– und Aufwandsrahmen erfüllen. Dabei wurde die Routenplanung explizit zweimal positiv erwähnt.

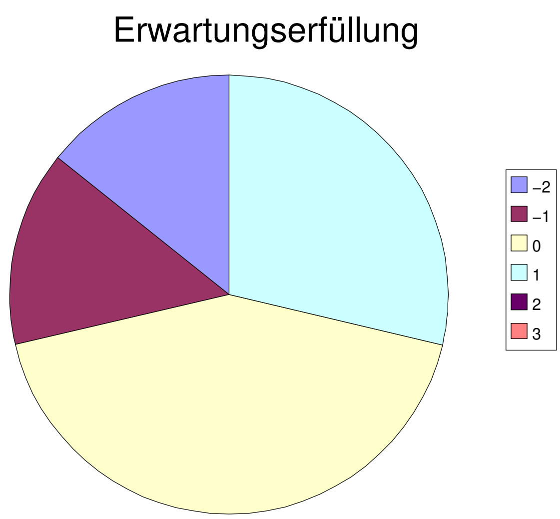
Als Gründe für die Nichterfüllung der Aufgabe in einem angemessenen Zeit– und Aufwandsrahmen führten die VPn die Langsamkeit des Systems (4) (Zitat: ***"Das Teil würde ich nicht mal geschenkt haben wollen. Mehr als uneffektiv."***), der zu hohe Aufwand (2), die fehlende

Menüübersicht (1), die fehlende Abbruchmöglichkeit bzw. der fehlende Zurückbutton (1) und die fehlende Hilfefunktion (1) an.

3.1.1 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 noch negative Erwartungserfüllung durch das System. Der Median liegt mit der Hälfte der VPn bei einer Erwartungserfüllung zwischen 0 und –1, 3 VPn vergaben –3, je eine –2 bzw. –1. Positive Erwartungserfüllungen waren bei 2 VPn zu verzeichnen (2 x 1), 3 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. [2]

2 Mediane auf .5 liegen genau zwischen den beiden Bewertungen

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)	1	0	2	2	2	2	1	0.4	0.5
Displayausgabe (allgemein)	2	1	0	1	4	1	1	0.1	1
Spracheingabe (allgemein)	0	2	1	0	5	1	1	0.5	1
Sprachausgabe (allgemein)	0	2	1	0	4	1	2	0.7	1
Die Möglichkeit mit dem System menschlich zu interagieren	3	0	1	3	3	0	0	–0.7	0
Das Design insgesamt	1	1	2	2	2	1	1	0	0
SmartKom reagiert auf meine Emotionen	7	1	1	1	0	0	0	–2.4	–3
Stifteingabe[3]	0	1	3	0	3	1	1	0.33	1

Sprachein– und –ausgabe wie auch Displayein– und –ausgabe im allgemeinen werden relativ positiv gesehen, besonders negativ war die Bewertung der Reaktion auf Emotionen (–3!). Hier konnte nur von einer Person eine neutrale Wertschätzung verzeichnet werden. 70 % der Befragten vergaben sogar den Wert –3!

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	3	0	2	1	0	0	–1.7	–2

Auch die fehlende Möglichkeit zu unterbrechen wurde von den VPn bemängelt. 70% der Befragten vergaben die Wertung –2 und schlechter.

Im Vergleich zur begleitenden ergonomischen Auswertung Mobil 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	2	0	1	3	3	1	0	–0.2	0
SmartKom hat zum größten Teil passende Antworten gegeben	1	1	0	4	3	1	0	0	0

3 eine VP gab an, das Feature nicht genutzt zu haben.

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 Geschwindigkeit des Systems war angemessen	5	0	3	0	2	0	0	–1.6	–2
Ich wußte während des Dialogs immer; was ich sagen konnte	2	3	1	2	1	1	0	–1	–1.5
Die Eingabe über Zeigen hat funktioniert	1	1	1	1	4	1	1	0.3	1
Die Eingabe über Sprache hat funktioniert	2	1	1	1	2	2	1	0	0.5
SmartKom reagierte schnell auf meine Anfragen	5	1	2	0	2	0	0	–1.7	–2.5
SmartKom ist inhaltlich einfach zu verstehen	0	1	0	4	3	1	1	0.6	0.5
SmartKom ist akustisch gut wahrnehmbar	0	0	1	1	2	2	4	1.7	2
Die Stifteingabe hat gut funktioniert	nicht genutzt								
Die Gestikeingabe hat funktioniert	nicht genutzt								
Die Aufgabe war einfach zu lösen	1	2	1	1	2	2	1	0.1	0.5

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. Im Vergleich zu den vorangegangenen Evaluationen muß hervorgehoben werden, daß die Probanden während des Dialogs oft nicht wußten, was sie sagen sollten. Positiv hervorzuheben im Vergleich zu vorangegangenen Evaluationen ist eine überwiegend positive Bewertung der akustischen Wahrnehmbarkeit der Ausgaben.

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

3.1.2 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	2	1	1	4	0	1	–0.1	0.5
bot in Anzahl und Qualität angemessene Informationen an	2	0	2	3	2	0	1	–0.3	0
benötigt Eingaben nur einmal; um zum Ziel zu gelangen	2	2	1	4	0	0	1	–0.8	–0.5
bietet eine gute Hilfefunktion an	9	0	0	1	0	0	0	–2.7	–3
hat ein übersichtlich gestaltetes Display	2	0	0	4	1	2	1	0.2	0
reagierte schnell auf gesprochene Befehle	4	2	1	1	2	0	0	–1.5	–2
reagierte schnell auf Zeigen	2	2	1	3	2	0	0	–0.9	–0.5

Im Gegensatz zu den beiden vorangegangenen abschließenden Evaluationen, wo die Tendenz der Daten eher neutral bis positiv war, zeigt sich hier ein deutlicher Rückgang bei der positiven Bewertung der Übersichtlichkeit des Displays (rot hervorgehoben). Die fehlende Hilfefunktion (blau hervorgehoben) erhielt mit 90% der Wertungen bei –3 die bis jetzt schlechteste Wertung!

3.1.3 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? [4]	nicht Kompetent			–	Kompetent			Mean	Median
	–3	–2	–1	0	1	2	3		
	2	2	1	4	0	0	0	–1.22	–1
Wie natürlich wirkten die Aktionen der Persona auf Sie?	nicht Natürlich			–	Natürlich				
	–3	–2	–1	0	1	2	3		
	1	3	1	0	4	1	0	–0.4	0

4 Ein Proband gab zur Kompetenz kein Urteil ab.

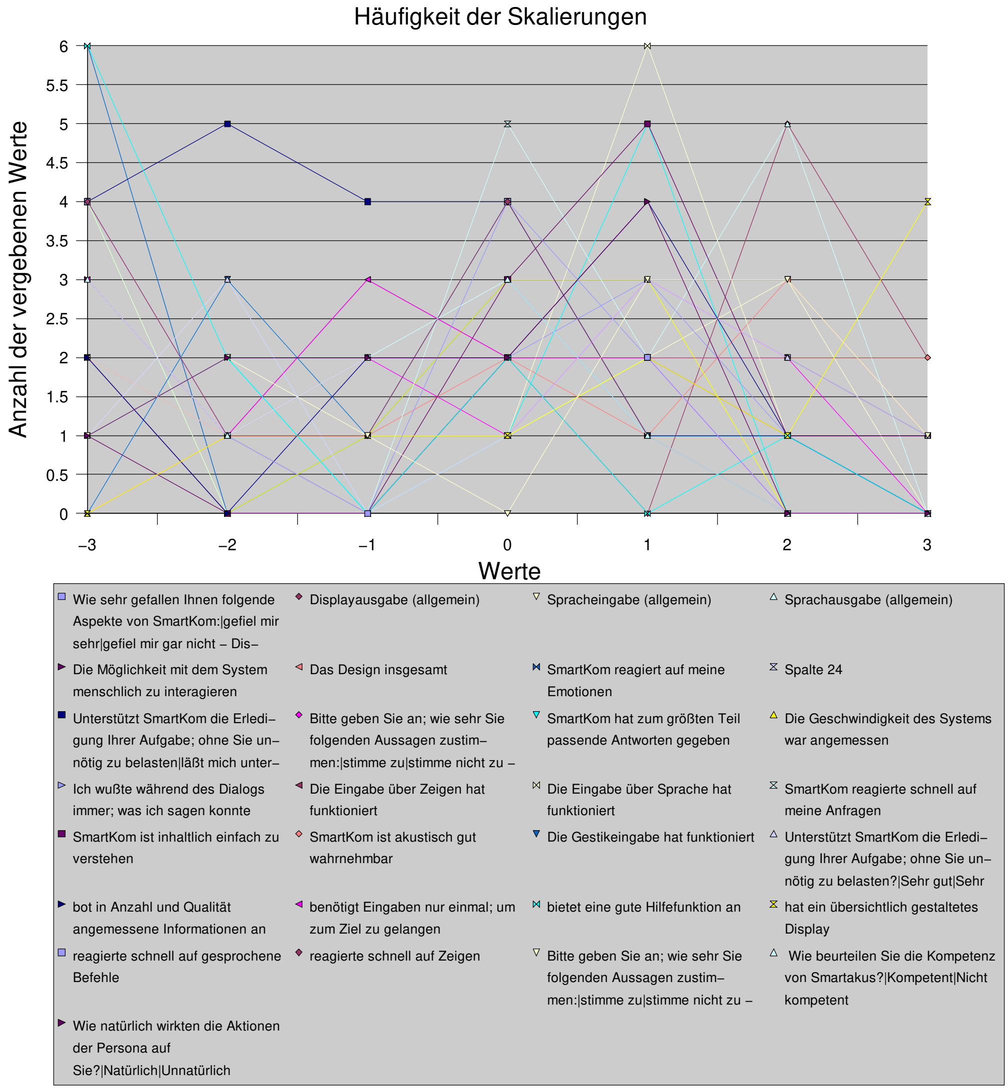
Die Persona wirkt mit –1.22 auf einer Skala von +3 bis –3 etwas inkompetenter als in den Evaluationen der anderen Szenarios,.

Auch die Natürlichkeit der Aktionen der Persona schnitt mit einem Mittelwert von –0.4 auf der selben Skala nicht gerade überragend ab.

Daß die VPn zur Kompetenz und Natürlichkeit der Persona eher negativ eingestellt sind, liegt wahrscheinlich daran, daß es sich im Gegensatz zur vorangehenden Mobil–Evaluation diesmal um eine reine Expertenevaluation handelt. Die Probanden haben insgesamt eine höhere Erwartungshaltung gegenüber Dialogsystemen, die zudem noch durch die beiden abschließenden Evaluationen erhöht wird. Bei Home und Public bereits negativ gewertete Features wurden bei der letzten abschließenden Evaluation Mobil natürlich noch negativer gesehen.

3.2 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 10 Probanden bejaht. Für 6 VPn war diese Erklärung ausreichend, um das System im Rahmen der Möglichkeiten zu bedienen, 4 VPn beantworteten diese Frage mit „nein" und keine mit „weiß nicht".

4 Fazit

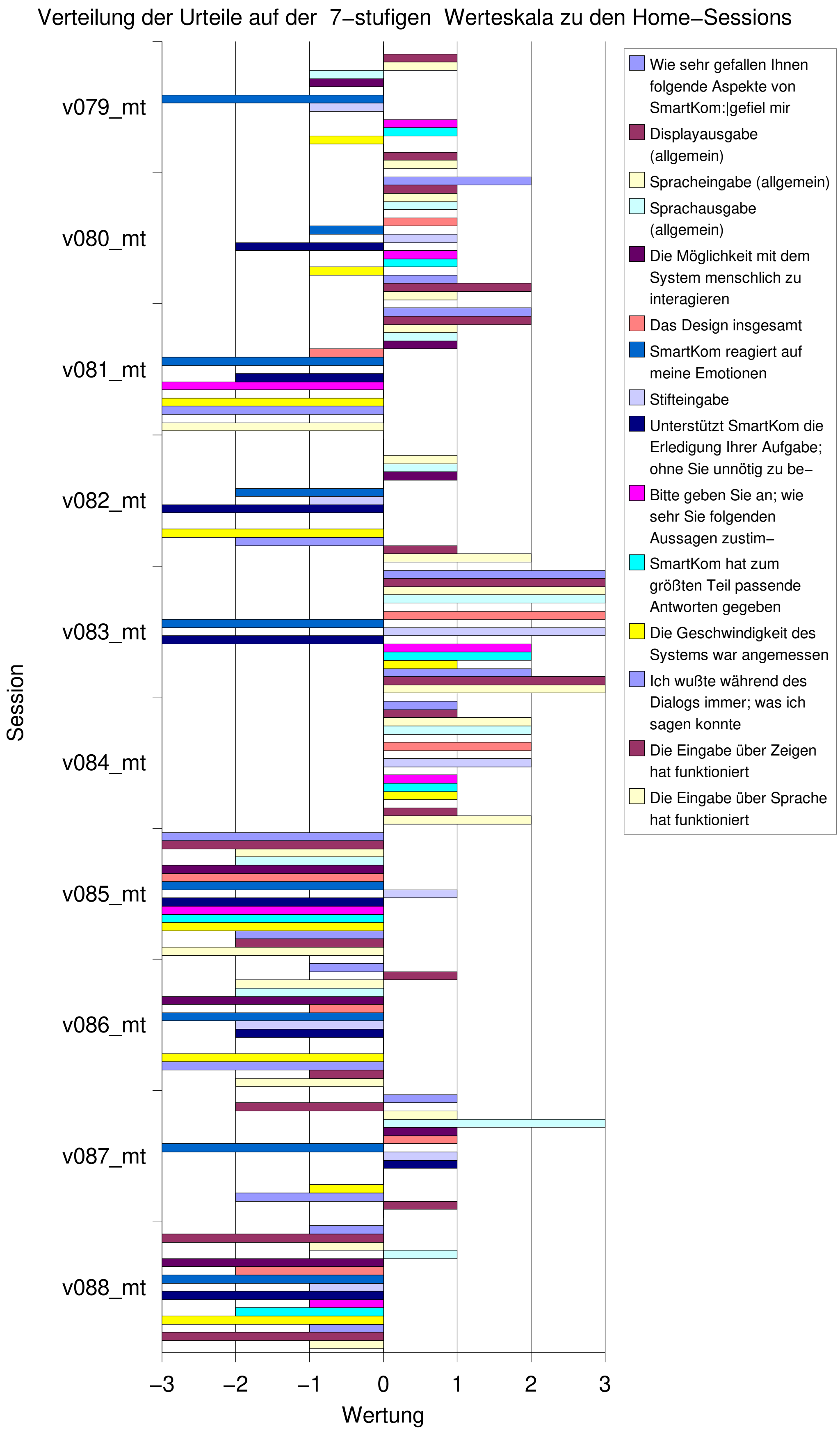
Der SmartKom Prototyp, Szenario Mobil, wurde an 10 Versuchspersonen getestet. Die hier detailliert ausgewertete Benutzerbefragung ergab ein allgemein positiveres Bild des Prototypen als bei der begleitenden Evaluation Mobil. Insgesamt schloß Mobil allerdings als schlechtestes Szenario bei der abschießenden Evaluation ab.

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.

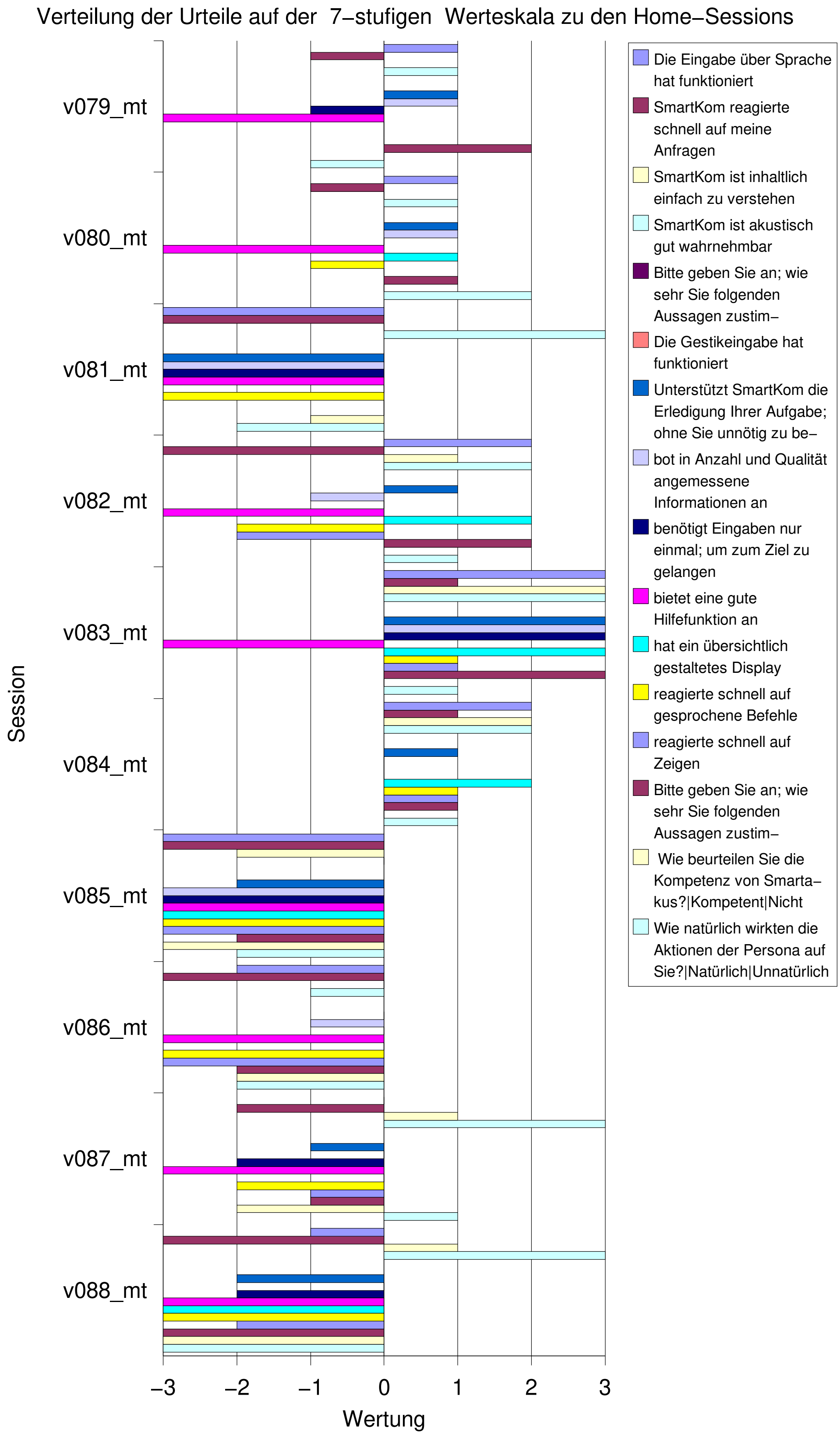
Im Gegensatz zu den beiden anderen Szenarien wurde die Displaygestaltung wesentlich negativer gewertet.

Allerdings zeigte das System insbesondere in der Sprachverarbeitung (Sprachverstehen) sowie in der Reaktion auf gestische Eingaben einen enormen Fortschritt zur begleitenden Evaluation Mobil. Leider war jedoch die Dialogführung wesentlich schlechter, sodaß die meisten Probanden oft nicht wußten, wie sie weiter verfahren sollten. Diese schlechtere Einschätzung seitens der VPn kann allerdings auch auf eine Hochsetzung der Bewertungskriterien aus einem Gewöhnungseffekt heraus zurückzuführen sein.

Abschließend läßt sich jedoch sagen, daß der Prototyp SmartKom zum Zeitpunkt der abschließenden Evaluation wesentliche Fortschritte zu Vorgängerdemonstratoren aufwies. Hätte die abschießende Evaluation zu einem späteren Zeitpunkt stattfinden können, hätte das System durchaus bessere Werturteile erhalten können.



Verteilung der Urteile auf der 7−stufigen Werteskala zu den Home−Sessions
Session
v079_mt
v080_mt
v081_mt
v082_mt
v083_mt
v084_mt
v085_mt
v086_mt
v087_mt
v088_mt
−3
−2
−1
0
1
2
3
Wertung
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−
SmartKom hat zum größten Teil passende Antworten gegeben
Die Geschwindigkeit des Systems war angemessen
Ich wußte während des Dialogs immer; was ich sagen konnte
Die Eingabe über Zeigen hat funktioniert
Die Eingabe über Sprache hat funktioniert



Verteilung der Urteile auf der 7–stufigen Werteskala zu den Home–Sessions
Session
v079_mt
v080_mt
v081_mt
v082_mt
v083_mt
v084_mt
v085_mt
v086_mt
v087_mt
v088_mt
–3
–2
–1
0
1
2
3
Wertung
Die Eingabe über Sprache hat funktioniert
SmartKom reagierte schnell auf meine Anfragen
SmartKom ist inhaltlich einfach zu verstehen
SmartKom ist akustisch gut wahrnehmbar
Bitte geben Sie an; wie sehr Sie folgenden Aussagen zustim–
Die Gestikeingabe hat funktioniert
Unterstützt SmartKom die Erledigung Ihrer Aufgabe; ohne Sie unnötig zu be–
bot in Anzahl und Qualität angemessene Informationen an
benötigt Eingaben nur einmal; um zum Ziel zu gelangen
bietet eine gute Hilfefunktion an
hat ein übersichtlich gestaltetes Display
reagierte schnell auf gesprochene Befehle
reagierte schnell auf Zeigen
Bitte geben Sie an; wie sehr Sie folgenden Aussagen zustim–
Wie beurteilen Sie die Kompetenz von Smarta–kus?|Kompetent|Nicht
Wie natürlich wirkten die Aktionen der Persona auf Sie?|Natürlich|Unnatürlich