



UNIVERSITÄT
DES
SAARLANDES

semv·x
semantic technologies and voice solutions

Disambiguierungsstrategien in Dialogsystemen

Bachelorarbeit

Fachrichtung Computerlinguistik

vorgelegt von

Lena Enzweiler

Saarbrücken, 9. November 2014

Meine Bachelorarbeit entstand im Zeitraum vom Juli 2014 bis Oktober 2014 bei SemVox GmbH unter der Leitung von Prof. Dietrich Klakow und der Betreuung von Pia Kuznik und Dr.-Ing. Markus Löckelt

Inhaltsverzeichnis

Abbildungsverzeichnis	VI
Tabellenverzeichnis	VII
Abkürzungsverzeichnis	IX
1 Einleitung	2
2 Related Work	5
3 Cognitive Load	6
4 Disambiguierungsstrategien	7
4.1 Disambiguierung	7
4.2 Disambiguierung in der Sprachverarbeitung	7
4.3 1. Strategie: Aggregierte Auswahl ohne Pause	8
4.4 2. Strategie: Aggregierte Auswahl mit Pause	9
4.5 3. Strategie: Sequentielle Auswahl	10
5 Versuch 1	11
5.1 Testszenario	11
5.2 Versuchsaufbau	13
5.3 Versuchsdesign	14
5.4 Auswertung	16
5.4.1 gemessene Zeiten	17
5.4.2 Fragebogen	19
5.4.3 Task Completion	26
5.5 Hypothesen	26

5.6	Versuchspersonen	26
5.7	Control Panel	28
5.8	Simulation und Durchführung	30
5.9	Resultat	32
5.9.1	Zeiten	32
5.9.2	Fragebogen	33
5.9.3	Task Completion	35
5.10	Qualitätssicherung	36
6	Versuch 2	36
6.1	Testszenario	36
6.2	Versuchsaufbau	37
6.3	Versuchsdesign	38
6.4	Auswertung	38
6.4.1	gemessene Zeiten	38
6.4.2	Fragebogen	40
6.4.3	Task Completion	40
6.5	Hypothesen	41
6.6	Versuchspersonen	41
6.7	Control Panel	41
6.8	Resultat	41
6.8.1	Zeiten	41
6.8.2	Fragebogen	44
6.8.3	Task Completion	46
6.9	Qualitätssicherung	47

7	Ergebnisse	47
7.1	Versuchsverlauf 1	47
7.1.1	Zeiten	47
7.1.2	Fragebogen	48
7.1.3	Task Completion	48
7.2	Versuchsverlauf 2	49
7.2.1	Zeiten	49
7.2.2	Fragebogen	50
7.2.3	Task Completion	50
8	Diskussion	51
8.1	Allgemeine Diskussion	51
8.2	Vergleichbare Studien	51
8.3	Future Work	51
9	Schlusswort	52

Abbildungsverzeichnis

1	Funktionsweise der ODP-S3 Plattform	3
2	Personenprofil: Anke	12
3	Fragebogen: Nasa-TLX	22
4	Fragebogen: Dialogverhalten	25
5	Fragebogen: Person	27
6	Controlpanel	29
7	Controlpanel	42

Tabellenverzeichnis

1	Interaktionsbeispiel Aggregierte Auswahl ohne Pause	8
2	Interaktionsbeispiel Aggregierte Auswahl mit Pause (Zahl) . .	9
3	Interaktionsbeispiel Aggregierte Auswahl mit Pause (Barge-In)	9
4	Interaktionsbeispiel Sequentielle Auswahl	10
5	Slotabfragen	13
6	Slotabfrage pro Person	13
7	Übersicht Versuchsablauf	14
8	Strecken- und Strategieverteilung	15
9	Anruf per Strecke	16
10	Durchschnittszeiten Strategie pro Strecke	17
11	Durschnittszeiten pro Strategie	18
12	Durchschnittsdialogzeiten	19
15	Durschnittszeiten Strategie pro Strecke	32
16	Durschnittszeiten Strategie pro Strecke	32
17	Durschnittszeiten mit Rennspiel versus ohne Rennspiel	33
20	Durschnittliche Task Completion (TC)	35
21	Slotabfragen	36
22	Slotabfrage pro Person	37
23	Übersicht Versuchsablauf	37
24	Strecken- und Strategieverteilung	38
25	Durschnittszeiten Strategie pro Strecke	39
26	Durschnittszeiten pro Strategie	39
27	Durschnittszeiten pro Strategie	40
28	Durschnittszeiten pro Strategie	40
29	Durschnittszeiten Strategie pro Strecke	43
30	Durschnittszeiten Strategie pro Strecke	43

31	Durschnittszeiten mit Rennspiel versus ohne Rennspiel	43
34	Durschnittliche Task Completion (TC)	46
35	Durschnittszeiten Strategie pro Strecke	47
36	Durschnittszeiten pro Strategie mit Rennspiel	47
37	Durschnittszeiten Strategie pro Strecke	48
38	Durschnittszeiten pro Strategie mit Rennspiel	48
39	Durschnittszeiten pro Strategie ohne Rennspiel	48
40	Durschnittliche Task Completion (TC)	48
41	Durschnittszeiten Strategie pro Strecke	49
42	Durschnittszeiten pro Strategie mit Rennspiel	49
43	Durschnittszeiten Strategie pro Strecke	50
44	Durschnittszeiten pro Strategie mit Rennspiel	50
45	Durschnittszeiten pro Strategie ohne Rennspiel	50
46	Durschnittliche Task Completion (TC)	50

Abkürzungsverzeichnis

Abstract

Die vorliegende Arbeit beschäftigt sich mit der Frage, welche Disambiguierungsstrategien in Sprachdialogsystemen für Benutzer bei hoher kognitiver Belastung am geeignetsten sind. Man fokussiert sich dabei auf Sprachdialogsysteme, welche speziell für die Bedienung während der Autofahrt konzipiert werden. Um der Frage der besten Disambiguierungsstrategie nachzugehen, werden in einem Wizard-of-Oz-Experiment Fahrszenarien simuliert, bei denen die Versuchspersonen mit einem Dialogsystem sprachlich interagieren. Dabei werden ambigge Eingaben des Benutzers simuliert worauf das System mit Disambiguierungsstrategien in Form von Nachfragen reagiert, welche eine entsprechende Benutzerreaktion verlangen. Anhand der Versuchsergebnisse wird analysiert, welche Strategien für den Benutzer am einfachsten und effektivsten waren. Insgesamt werden drei Disambiguierungsstrategien verfolgt. Aggregierte Auswahl ohne Pause, aggregierte Auswahl mit Pause, sowie die sequentielle Auswahl.

1 Einleitung

Dialogsysteme für das Auto müssen so gestaltet werden, dass sie den Fahrer so wenig wie möglich vom Fahren ablenken und ihm so gut wie möglich assistieren. Die Herausforderung für einen Dialog Designer besteht daher darin, Sprachäußerungen so raffiniert zu gestalten, dass dem Benutzer zum Einen alle relevanten Informationen in verständlicher Weise geliefert werden und zum Anderen, dass der Benutzer darauf möglichst einfach antworten und seine Anfragen und Wünsche effizient übermitteln kann. Die Funktionsweise eines Dialogsystems hängt von mehreren Komponenten ab, welche anhand der in Abbildung 1 dargestellten Funktionsweise der ODP S3 Plattform der SemVox GmbH kurz erläutert werden. Die ODP S3 Plattform ermöglicht die Umsetzung komplexer Sprachdialoge. Zunächst müssen die Spracheingaben des Benutzers zu semantischen Objekten verarbeitet werden. Dabei wird zunächst die Spracheingabe auf eine Grammatik gematched, welche alle möglichen Spracheingaben des Benutzers abfängt und semantischen Objekten zuweist. Diese werden dann von einem Backend-Server verarbeitet, woraufhin eine passende Sprachausgabe ausgelöst wird. In dieser Arbeit konzentriert man sich allein auf die Sprachgenerierung. Ein komplexer Dialog zwischen System und Benutzer führt häufig dazu, dass der Benutzer eine Eingabe macht, die das System nicht eindeutig zuordnen kann und mehrere Optionen für die Ausführung der vom Benutzer geäußerten Eingabe bestehen. Es muss an dieser Stelle vom System eine Rückfrage beim Benutzer erfolgen, sodass dieser seine vorherige Eingabe eindeutig übermitteln kann. Wenn der Benutzer zum Beispiel den Wunsch äußert einen Kontakt aus dem im System gespeicherten Adressbuch anzurufen, es allerdings zwei Kontakte mit diesem Namen gibt, muss das System eine Rückfrage stellen, um zu ermitteln, welcher dieser beiden Kontakte gemeint ist. Der Dialog Designer spricht in diesem Fall von Disambiguierung. Es wird in der vorliegenden Arbeit der Frage nachgegangen wie man konkret die Sprachausgabe einer solchen Disambiguierung

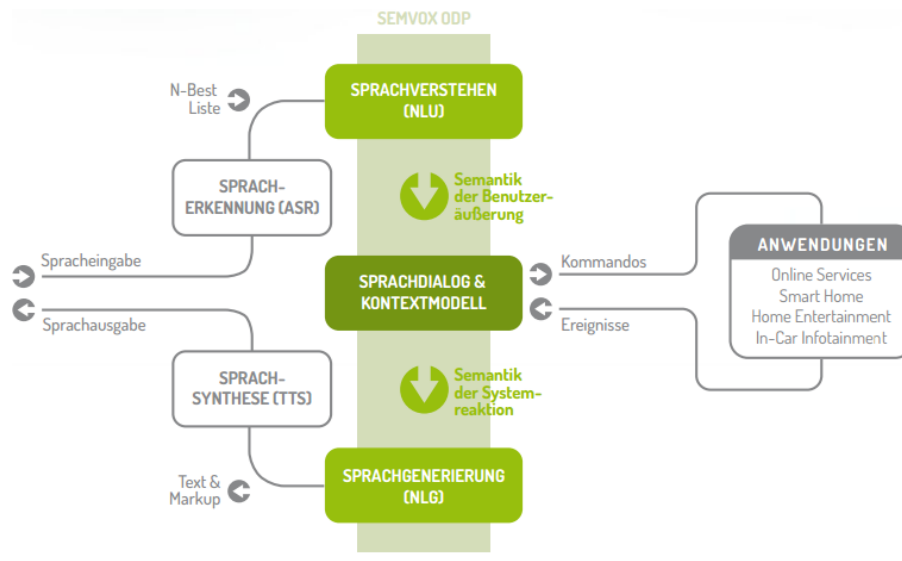


Abbildung 1: Funktionsweise der ODP-S3 Plattform

innerhalb eines Dialogsystems, das speziell für das Auto konzipiert wurde, gestaltet. Dabei werden drei verschiedene Strategien in einem Wizard-of-Oz-Experiment auf Effizienz und Beliebtheit unter den Versuchspersonen getestet. Diese Strategien werden im Kapitel 4 näher erläutert. Der Versuch in Kapitel 5 zeigt klar, dass bei einer Disambiguierung über wenige Optionen die Strategie **Aggregierte Auswahl ohne Pause** am beliebtesten unter den Versuchsperson ist. Man hat sich daher für einen zweiten Versuch entschieden, welcher sich lediglich in der Länge der Disambiguierung unterscheidet. Dieser in Kapitel 6 durchgeführte Versuch zeigt, dass bei einer Disambiguierung über mehrere Optionen die Strategie **Aggregierte Auswahl mit Pause** die beliebtere Strategie ist. Neben der Beliebtheit unter den Versuchspersonen wurden bei beiden Versuchen weitere Faktoren, wie Dialogzeit, erfolgreiches Abschließen des Dialoges (Task Completion) oder Unterschiede des Dialogverhaltens zwischen hoher und geringer kognitiver Belastung erforscht. Ein zusammenfassendes Ergebnis beider Versuche findet sich in Kapitel 7. Zum Schluss

erfolgt eine Diskussion über die ermittelten Ergebnisse und schließlich das Schlusswort.

2 Related Work

Dialogsysteme, Disambiguierungsstrategien in Interaktionen sowie kognitive Belastung in Videospielen und Dialogsystemen werden in weiteren Forschungsgebieten erforscht.

In einer früheren Studie ([Minker et al., 2002]) wurde eine weitere Disambiguierungsstrategie für Dialogsysteme untersucht. Bei dieser vorgestellten Strategie werden zusätzliche Informationen zur Disambiguierung vom User erfragt. Angewendet auf den Anrufen Task in dieser Studie würde das System zum Beispiel nachfragen, wo denn Fritz wohnt, anstatt zu fragen, ob der User Fritz aus München oder Ingolstadt meint. In weiteren Studien wurde die kognitive Belastung während Videospielen ([Ang et al., 2006], [Tsiakoulis et al, 2012]) und während einer Systeminteraktion untersucht ([Villing, 2009], [Tsiakoulis et al, 2012]). Dabei wurde festgestellt, dass eine kognitive Belastung das Dialogverhalten ändert. Die Ergebnisse zeigen, dass während einer hohen kognitiven Belastung längere Pausen zwischen zwei Sprachäußerungen eingelegt werden und die Anzahl der Sprachäußerungen geringer ist im Vergleich zur Anzahl während einer niedrigen kognitiven Belastung. ([Villing, 2009]) Außerdem kam man zu dem Ergebnis, dass die Anzahl der Bargeins während einer hohen kognitiven Belastung deutlich höher ist im Vergleich zu einer niedrigen Belastung. ([Tsiakoulis et al, 2012]) In ([Tsiakoulis et al, 2012]) hat man weiter herausgefunden, dass Versuchspersonen unter kognitiver Belastung Dialogabläufe mit einfachen Spracheingaben wie **ja** oder **nein** über solchen Dialogabläufen bevorzugen, in denen das System den zu füllenden Slot als Antwort verlangt. Daher wird vermutet, dass die dritte Disambiguierungsstrategie bei Versuchspersonen mit hoher kognitiver Belastung am effizientesten ist. In der erwähnten Studie führen die Versuchspersonen ebenfalls parallel zur Dialoginteraktion ein Rennspiel. Man hat dabei festgestellt, dass die Completion Rate des aufgestellten Task bei der alleinigen Interaktion mit dem Systems höher war als bei der Inter-

aktion parallel zum Rennspiel. Desweiteren konnte man durch die Ergebnisse des NASA-TLX Testes sehen, dass die Versuchspersonen einen Unterschied der kognitiven Belastung zwischen dem alleinigen Fahren, der alleinigen Interaktion und des Fahrens während der Interaktion gemerkt haben. Ähnliche Ergebnisse werden für die vorliegende Studie erwartet. In einer weiteren Studie wurde erforscht, dass eine kognitive Belastung, die durch eine parallele Interaktion mit anderen Spielern in einem Computerspiel ausgelöst wird, die Performance im Spiel verschlechtert ([Ang et al., 2006]). Die Kommunikation mit anderen Spielern im Computerspiel kann mit der Systeminteraktion aus dieser Studie verglichen werden, weshalb eine schlechtere Rennspielleistung während des Anrufen-Task im Vergleich zur Rennspielleistung ohne Systeminteraktion erwartet wird. In [Mishra et al., 2004] konnte festgestellt werden, dass Benutzer, die sich mehr auf einen anderen Task als auf die Systeminteraktion konzentrieren, eher unflüssige und abgehackte Sprachausgaben produzieren. In einer zukünftigen Arbeit könnte überprüft werden, ob solche Sprachäußerungen die angewendeten Disambiguierungsstrategien in einem echten System negativ beeinflussen. Desweiteren kann diese Erkenntnis dazu genutzt werden, um die Stärke der Ablenkung durch das Rennspiel der einzelnen Versuchspersonen zu bewerten.

3 Cognitive Load

- allgemein CL
- was in anderen Papern gesagt → was erwartet?
- Einfluss CL auf Dialogsystemen

4 Disambiguierungsstrategien

Insgesamt werden 3 Disambiguierungsstrategien auf Effizienz und Beliebtheit unter kognitiver Belastung getestet.

- Aggregierte Auswahl **ohne** Pause
- Aggregierte Auswahl **mit** Pause
- Sequentielle Auswahl

In den folgenden Unterkapiteln wird zunächst kurz auf das Prinzip der Disambiguierung eingegangen. Anschließend werden die Funktionsweisen der einzelnen Strategien erläutert und mögliche Vor- und Nachteile, sowie Präferenzen der Versuchspersonen diskutiert.

4.1 Disambiguierung

Bei einer Disambiguierung werden verschiedene Begriffsbedeutungen voneinander abgegrenzt bzw. differenziert. Dies gilt zum Beispiel für Nomen, welche den gleichen Begriff beschreiben aber ein anderes Konzept darstellen. Das Nomen **Bank** zum Beispiel kann sowohl ein Geldinstitut als auch eine Sitzmöglichkeit darstellen. Die Disambiguierung spielt bei der Sprachverarbeitung eine zentrale Rolle, da Spracheingaben nicht immer eindeutig formuliert werden und die dadurch entstehenden Mehrdeutigkeiten aufgelöst werden müssen.

4.2 Disambiguierung in der Sprachverarbeitung

Äußert ein Benutzer eines Dialogsystems eine ambige Spracheingabe, so muss das System diese disambiguieren. Diese Disambiguierung kann durch direkte Nachfrage der gewünschten Interpretation beim Benutzer erfolgen. Möchte der User zum

Beispiel einen Kontakt aus einem Adressbuch anrufen, dessen Vornamen mehrfach vorkommt, so wird eine Disambiguierung notwendig sein, wenn der Benutzer bei seiner Spracheingabe lediglich den Vornamen angibt. Um den gewollten Kontakt vom User zu erfragen kann das System einer der in dieser Arbeit behandelten Disambiguierungsstrategien verwenden.

4.3 1. Strategie: Aggregierte Auswahl ohne Pause

Bei dieser Strategie werden alle möglichen Interpretationen der ambigen Spracheingabe ausgegeben und auf eine Auswahl des Benutzers gewartet. In der folgenden Beispielinteraktion muss das System über den Nachnamen des von dem Benutzer adressierten Kontaktes disambiguieren. In der Sprachausgabe werden so alle möglichen Nachnamen (hier Meier und Müller) für den genannten Vornamen (hier Peter) zum Auswählen zur Verfügung gestellt. Der Benutzer kann während der Ausgabe mittels Barge-Ins antworten oder am Ende der Ausgabe mit dem gewünschten Nachnamen antworten.

Tabelle 1: Interaktionsbeispiel Aggregierte Auswahl ohne Pause

Akteur	Sprachausgabe
User	Rufe Peter an!
System	Meinst du Peter Müller oder Peter Meier?
User	Peter Müller.
System	Ok, ich werde Peter Müller jetzt anrufen.

Da diese Strategie einfach aufgebaut ist, sollte es für den Benutzer intuitiv klar sein, welche Antwort das System erwartet um die Interaktion weiter zu führen. Problematisch wird es wahrscheinlich bei einer hohen Anzahl an Disambiguierungsvorschlägen, da die Sprachausgabe entsprechend lang wird und der Benutzer

sich möglicherweise die komplette Sprachausgabe anhört, da die Möglichkeit zum Barge-In hier nicht auffällig ist.

4.4 2. Strategie: Aggregierte Auswahl mit Pause

Diese Strategie funktioniert im Prinzip wie die 1. Strategie. Der Unterschied liegt darin, dass diese Strategie die einzelnen Vorschläge durchnummeriert präsentiert und eine kurze Pause zwischen den Vorschlägen einlegt. Die Beispielinteraktionen zeigen die gleiche Situation wie in Strategie 1, allerdings antwortet der Benutzer im ersten Beispiel mit der Zahl, die der gewünschten Interpretation voran gestellt wurde und im zweiten Beispiel mit Hilfe eines Barge-Ins.

Tabelle 2: Interaktionsbeispiel Aggregierte Auswahl mit Pause (Zahl)

Akteur	Sprachausgabe
User	Rufe Peter an!
System	Meinst du [Pause] 1. Peter Müller [Pause] oder 2. Peter Meier?
User	den ersten.
System	Ok, ich werde Peter Müller jetzt anrufen.

Tabelle 3: Interaktionsbeispiel Aggregierte Auswahl mit Pause (Barge-In)

Akteur	Sprachausgabe
User	Rufe Peter an!
System	Meinst du [Pause] 1. Peter Müller [oder...]?
User	Ja.
System	Ok, ich werde Peter Müller jetzt anrufen.

Bei dieser Strategie ist die Möglichkeit zum Barge-In sichtbar und der User muss sich nicht die komplette Sprachausgabe zu Ende anhören. Allerdings könnte die Sprachausgabe bei einer kleinen Anzahl an Disambiguierungsvorschlägen durch die Pausen und die Nummerierung unnötig lang auf den Benutzer wirken. Daher bevorzugt der Benutzer vermutlich die 1. Strategie bei einer kleinen Anzahl an Interpretation und entsprechend die 2. Strategie bei einer hohen Anzahl an Disambiguierungsvorschlägen.

4.5 3. Strategie: Sequentielle Auswahl

Die Sequentielle Auswahl packt jeden Disambiguierungsvorschlag in eine separate Sprachausgabe und verlangt anschließend eine Bestätigung bzw. eine Ablehnung des angegebenen Vorschlags. Die ambige Spracheingabe wird dann mit der ersten Bestätigung des Benutzers aufgelöst.

Tabelle 4: Interaktionsbeispiel Sequentielle Auswahl

Akteur	Sprachausgabe
User	Rufe Peter an!
System	Meinst du Peter Meier?
User	Nein.
System	Meinst du Peter Müller?
user	Ja.
System	Ok, ich werde Peter Müller jetzt anrufen.

Diese Strategie ist wahrscheinlich besonders effizient, wenn der Benutzer einer hohen kognitiven Belastung ausgesetzt ist, da er das Tempo hier selbst bestimmen kann. Der Nachteil dieser Strategie liegt vermutlich darin, dass gerade bei vielen

Interpretationsvorschlägen die Interaktion sehr lange dauert und der User jedes Mal eine Spracheingabe zur Fortsetzung des Dialoges eingeben muss.

5 Versuch 1

Um zu testen, welche Disambiguierungsstrategie bei Versuchspersonen unter kognitiver Belastung am effizientesten ist, wird ein Wizard-of-Oz Experiment durchgeführt. Hierbei werden die Probanden ein Rennspiel fahren und parallel ein Testszenario durchführen, in welchem Sie per Spracheingabe erfolgreich einen Anruf aufbauen sollen. Desweiteren werden die Versuchspersonen dieses Testszenario ohne Rennspiel durchgehen, um mögliche Unterschiede der Ergebnisse zwischen kognitiv belastender und nicht kognitiv belastender Versuchsperson zu analysieren.

5.1 Testszenario

Während der Systeminteraktion sollen die Versuchspersonen erfolgreich einen Anruf ausführen. Insgesamt sollen vier Personen angerufen werden, welche dem User über Personenprofile angezeigt werden. Darin sieht die Versuchsperson welche Slots zu füllen sind. Abbildung 2 zeigt das Personenprofile von Anke aus welchem hervor geht, dass Anke auf der geschäftlichen Festnetznummer angerufen werden soll. Die Versuchspersonen werden am Anfang darauf hingewiesen, dass sie die Slots einzeln übergeben sollen. Nachdem der User spezifiziert hat, welchen Anrufer er anrufen möchte, fragt das System selbst die erforderlichen Slots ab. Diese Nachfrage wird in den unterschiedlichen Dialogstrategien erfragt. Pro Anruf gibt es insgesamt zwei zu füllende Slots, die mit der selben Disambiguierungsstrategie abgefragt werden. Beim nächsten Anruf muss die Versuchsperson andere Slots füllen und die Nachfrage erfolgt mit der nächsten Strategie. Die zu füllenden Slots sind in Tabelle 5 aufgelistet. Welche Slots pro Person abgefragt werden, zeigt Tabelle 6.

Anke Schumacher



Mobilnummer

privat

geschäftl.

Festnetznummer

privat

geschäftl.



Mainzerstr. 23, 66121, Saarbrücken



A.Schumacher86@gmx.de

Abbildung 2: Personenprofil: Anke

Tabelle 5: Beispiel Slotabfragen

Slot	erfragte Werte
Nummernotyp	privat oder geschäftlich?
Telephontyp	Mobilnummer oder Festnetznummer
Nachname	Meier oder Müller
Stadt	München oder Ingolstadt

Tabelle 6: Slotabfrage pro Person

Anke	Peter	Fritz	Kim
Nummerntyp		Nummerntyp	Nummerntyp
Telephontyp	Telephontyp		Telephontyp
	Nachname		
		Stadt	

5.2 Versuchsaufbau

Um eine möglichst realistische Fahrsimulation mit hoher kognitiver Belastung darzustellen, werden die Versuchspersonen ein Rennspiel mit einem Racing Wheel und den dazugehörigen Pedalen spielen. Bei dem Rennspiel handelt es sich um **Need for Speed: Shift**¹, welches im Einzelrennen - Modus mit jeweils drei Gegnern gefahren wird. Die Versuchspersonen bekommen neben der Systeminteraktion die Aufgabe, eine möglichst hohe Platzierung zu erreichen. Dies soll die Konzentration und damit die kognitive Belastung während dem Rennspiel steigern. Zu Beginn des

¹ http://www.needforspeed.com/de_DE/shift

Versuchs fahren die Probanden zunächst eine Testrunde. Mit dem Ergebnis dieser Runde kann man einschätzen wie gut die jeweiligen Personen im Rennspiel sind und weiter die Schwierigkeit des Spiels, und damit die Rennfähigkeiten der Gegner einstellen. In den nächsten drei Runden werden die Versuchspersonen parallel zum Rennspiel das Testszenario durchgehen und dabei drei Personen anrufen.

Der Anruf gilt nur dann als erfolgreich, wenn alle Slots korrekt gefüllt werden. In der letzten Runde findet nur eine Systeminteraktion statt, ohne paralleles Rennspiel und die dadurch verursachte kognitive Belastung. Tabelle 23 zeigt einen Überblick des Versuchsaufbaus.

Tabelle 7: Übersicht Versuchsablauf

1. Runde	2. Runde	3. Runde	4. Runde	5. Runde
Rennspiel	Rennspiel	Rennspiel	Rennspiel	
	Anruf Anke	Anruf Peter	Anruf Fritz	Anruf Kim

Während des Versuchs wird die Versuchsperson, das Rennspiel und die Dialoginteraktion aufgezeichnet. Dadurch wird sicher gestellt, dass man alle Reaktion einfangen und die Daten besser auswerten kann.

5.3 Versuchsdesign

Die Versuchspersonen fahren in den Runden 2-4 jeweils eine Strecke mit unterschiedlicher Disambiguierungsstrategie. Insgesamt werden diese auf drei unterschiedliche Strecken verteilt. Man hat sich für drei unterschiedliche Strecken entschieden, da man einen Lerneffekt bei einer gleichbleibenden Strecke ausschließen wollte. Parallel werden die Zeiten gemessen, die eine Versuchsperson für die Absolvierung einer Strecke bei der Interaktion mit einer bestimmten Disambiguie-

rungsstrategie benötigt (siehe Unterkapitel 6.4.1). Da man diese Zeiten miteinander vergleichen möchte, müssen die Disambiguierungsstrategien geschickt auf die Strecken verteilt werden, da die Strecke unterschiedlich lang sind und daher keine aussagekräftigen Vergleiche untereinander bieten. Um diesen Konflikt zu lösen, werden die Versuchspersonen in drei Gruppen aufgeteilt, sodass jede Gruppe jede Strecke mit einer unterschiedlichen Disambiguierungsstrategie fährt. Schließlich kann man so für jede Strecke die Zeiten für unterschiedliche Strategien sammeln und vergleichen, mit welcher Strategie eine bestimmte Strecke am schnellsten gefahren wurde (siehe Kapitel 6.4.1 Auswertung).

In der letzten Runde soll nur das Testszenario ohne Rennspiel durchgeführt werden. Hierfür gibt es Gruppe 4, welche aus allen Versuchsteilnehmern besteht. Diese wird jedoch nochmal in drei Zwischengruppen aufgeteilt, sodass ein Drittel der Versuchspersonen in der vierten Runde das Testszenario in Strategie 1, ein Drittel in Strategie 2 und das letzte Drittel in Strategie 3 durchführen. Ein Überblick der Strecken- und Strategieverteilung pro Gruppe ist in Tabelle 23 aufgelistet.

Tabelle 8: Strecken- und Strategieverteilung

Aufteilung	Strategie 1	Strategie 2	Strategie 3
1. Gruppe	Strecke A	Strecke B	Strecke C
2. Gruppe	Strecke B	Strecke C	Strecke A
3. Gruppe	Strecke C	Strecke A	Strecke B
4. Gruppe	keine Strecke	keine Strecke	keine Strecke

Jede Gruppe fährt die Strecken in der gleichen Reihenfolge (erst Strecke A dann Strecke B und schließlich Strecke C). Dadurch soll gewährleistet sein, dass die Streckenzeiten durch keinen Lerneffekt bei einer unterschiedlicher Reihenfolge beeinflusst werden. Wenn Strecke A mal zu Beginn und mal zum Schluß gefahren

wird, so könnten die Zeiten für die Runde am Schluß besser ausfallen, da die Versuchsperson durch die vorherigen Runden mehr an Spielererfahrung gewonnen hat und bessere Zeiten fährt. Die anzurufenden Personen sind auf bestimmte Strecken festgelegt und in Tabelle 9 gelistet.

Tabelle 9: Anruf per Strecke

Strecke	Anruf
Strecke A	Anke
Strecke B	Peter
Strecke C	Fritz
keine Strecke	Kim

5.4 Auswertung

Um herauszufinden, welche Disambiguierungsstrategie am effizientesten ist, werden verschiedene Auswertungen vorgenommen. Zunächst werden die Zeiten gemessen, die die Versuchsperson zum einen für das absolvieren der Strecke und zum anderen für das erfolgreiche abschließen des Testszenarios benötigt (Unterkapitel 6.4.1). Nach jeder Rennrunde wird die Versuchsperson außerdem einen Fragebogen ausfüllen, welche sich auf die subjektiv wahrgenommene kognitive Belastung und auf Merkmale der Disambiguierungsstrategien bezieht (Unterkapitel 6.4.2). Desweiteren wird die Task Completion ausgewertet um zu erforschen, wie erfolgreich ein Dialog geführt wurde. Schließlich wird durch aufkommende sprachliche Phänomene einer Disambiguierungsstrategie untersucht, ob sich das Dialogverhalten bei kognitiver Belastung von dem Dialogverhalten ohne kognitive Belastung unterscheidet.

5.4.1 gemessene Zeiten

Rennzeiten

Um zu analysieren, ob das Rennverhalten durch eine Disambiguierungsstrategie negativ beeinflusst wird, werden die Rennzeiten pro Runde gemessen. Für jede Strecke wird dann die durchschnittliche Zeit gebildet, die die Versuchspersonen mit paralleler Systeminteraktion in einer bestimmten Strategie benötigten. Das Ergebnis ist in Tabelle 10 aufgelistet.

Tabelle 10: Durchschnittszeiten Strategie pro Strecke

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	71,5 sek	93 sek	74,5 sek
Strecke B	68,75 sek	75,75 sek	91,5 sek
Strecke C	74,5 sek	58,38 sek	61,75 sek

Diesen Ergebnissen zufolge, gibt es keine Strategie, mit der eine Strecke besser oder schlechter gefahren wurde als mit anderen Strategien. Dies könnte jedoch daran liegen, dass einzelnen Werte durch schlechtere bzw. bessere Spieler in den Gruppen verfälscht wurden. Befindet sich zum Beispiel ein sehr schlechter Spieler in Gruppe 1 und ein sehr guter Spieler in Gruppe 2, so könnte die Durchschnittszeit für Strategie 1 auf Strecke A, durch die lange Zeit des schlechten Spielers, verschlechtert werden. Im Gegensatz dazu könnte die Durchschnittszeit für Strategie 3 auf Strecke A durch die guten Resultate des guten Spielers aus Gruppe 3 verbessert werden. Um dieses Problem zu lösen, wird pro Strategie der Durchschnitt aller mit dieser Strategie gefahrenen Rennzeiten berechnet. Zeiten von extrem guten bzw. schlechten Spielern sollten die Durchschnittszeiten ganzer Strategien dann nicht mehr beeinflussen. Die daraus resultierenden Werte geben dann eine Aussage dar-

über, mit welcher Strategie die Rennen am besten bzw. am schlechtesten gefahren wurden. Die endgültige Rennzeitberechnung für die Analyse der effizientesten Disambiguierungsstrategie ist in Tabelle 11 dargestellt.

Tabelle 11: Durchschnittszeiten pro Strategie

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Durchschnitt	71,58 sek	75,71 sek	75,92 sek

Die Unterschiede der Rennzeiten der einzelnen Strategien sind jedoch statistisch nicht relevant ($p = 0,79$). Daher kann hier nicht der Rückschluss gezogen werden, dass Strategie 1 die Versuchspersonen am wenigstens ablenkt. Desweiteren steht die Frage offen, ob die Rennzeiten überhaupt Ausschluss darüber geben können, welche Strategie sich am besten während der Autofahrt eignet. Dies könnte in zukünftigen Arbeiten durch einen umfangreicheren Versuch überprüft werden.

Dialogzeiten

Neben den Zeiten für das Rennspiel werden auch die Dialogzeiten berechnet. Anhand dieser Zeiten kann man sehen, mit welcher Strategie der kürzeste Dialog möglich ist. Desweiteren kann man die Dialogzeiten vergleichen, die einmal in der gleichen Strategie mit Rennspiel und einmal ohne Rennspiel erzielt wurden. Das könnte interessant sein, um die Unterschiede im Dialogverhalten zwischen einer kognitiv hoch belastenden Versuchsperson und einer weniger belastenden Person zu untersuchen. Eine längere Dialogzeit in einer gleichen Strategie ist möglicherweise auf eine längere Reaktionszeit zurückzuführen, weshalb bessere Zeiten in der vierten Runde, also ohne Rennspiel und damit ohne hohe kognitive Belastung, erwartet werden. Es werden alle Dialogzeiten aus den Runden mit Rennspiel gemessen und einmal für jede Strecke der Durchschnitt pro Strategie und einmal der

gesamte Durchschnitt pro Strategie gebildet. Diese Werte kann man dann gegen die Durchschnittszeiten aus der Runde ohne Rennstrecke vergleichen. Es werden allerdings nur die Dialogzeiten bewertet, bei denen der Dialog korrekt durchgeführt wurde, da sonst die Durchschnittszeiten verfälscht werden können. Tabelle 12 zeigt die Ergebnisse.

Tabelle 12: Durchschnittsdialogzeiten

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	15,34 sek	20,38 sek	20,28 sek
Strecke B	14,31 sek	20,05 sek	22,07 sek
Strecke C	15,97 sek	21,01 sek	20,35 sek
Strecke A - C	15,19 sek	20,52 sek	20,81 sek
ohne Strecke	14,9 sek	18,8 sek	17,59 sek

Die Unterschiede aus Strategie 1 sind statistisch signifikant gegenüber den Unterschieden aus Strategie 2 und 3. Die Unterschiede aus Strategie 2 und 3 sind zueinander jedoch nicht signifikant. Das Ergebnis zeigt, dass die Strategie 1 den kürzesten Dialog sowohl mit Rennspiel, als auch ohne Rennspiel ermöglicht. Die letzten beiden Zeilen der Tabelle zeigen, dass die Versuchspersonen einen deutlich kürzeren Dialog in jeder Strategie ohne Rennspiel ablegen. Die Ergebnisse aus dem Unterkapitel ?? lassen ausschließen, dass die Unterschiede aufgrund eines unterschiedlichen Dialogverhaltens zu erklären sind. Dies lässt vermuten, dass die Reaktionszeiten bei geringer Belastung kleiner sind als bei hoher Belastung und die zeitlichen Unterschiede dadurch zu Stande kommen.

5.4.2 Fragebogen

Neben den Zeiten wird nach jeder Runde ein Fragebogen ausgefüllt. Dieser besteht im ersten Teil aus einem Ausschnitt des NASA-TLX Testes zur subjektiven Einschätzung der empfundenen kognitiven Belastung. Im zweiten Teil werden Fragen über die zuletzt getestete Strategie gestellt und es wird die Möglichkeit gegeben positives oder negatives Feedback über den Dialog der letzten Runde zu geben. Zu Beginn des Versuchs wird ein allgemeiner Fragebogen ausgefüllt, der Informationen zur Versuchsperson liefert.

Nasa-TLX

Abbildung 3 zeigt den Nasa-TLX Teil des ersten Fragebogens. Die Ergebnisse dieses Testes werden zum Einen dafür genutzt um zu erforschen, bei welcher Strategie die Versuchspersonen eine höhere kognitive Belastung empfunden haben. Zum anderen kann man sehen, wie die Versuchspersonen ihre kognitive Belastung während einer Runde mit Rennspiel im Vergleich zur Runde ohne Rennspiel einschätzen. Die Nachfolgende Tabelle zeigt die Ergebnisse jeder Frage pro Strategie. Dabei werden jeweils die durchschnittlichen Antworten für alle Runden, der Runden mit Rennspiel und nur der Runde ohne Rennspiel aufgelistet.

Antwortenintervall	Strategien	Ergebnisse bestimmter Runden		
		1-4	1-3	4

Geistige Anforderung

1: gering 6: hoch	Strategie 1	1,88	2,08	1,25
	Strategie 2	2,06	2,42	1
	Strategie 3	2,63	2,83	2

Körperliche Anforderung

1: gering	Strategie 1	2	2,17	1,5
6: hoch				

Antwortenintervall	Strategien	Runden 1-4	Runden 1-3	Runde 4
	Strategie 2	1,44	2,25	1
	Strategie 3	2	2,25	1,25

Zeitliche Anforderung

1: gering	Strategie 1	1,87	2,09	1,25
6: hoch	Strategie 2	1,75	2	1
	Strategie 3	2,31	2,67	1,25

Leistung

1: gering	Strategie 1	4,25	4,5	3,5
6: hoch	Strategie 2	4,75	4,33	6
	Strategie 3	4,25	4	5

Anstrengung

1: gering	Strategie 1	2	2,25	1,25
6: hoch	Strategie 2	2,13	2,55	1
	Strategie 3	2,63	2,92	1,75

Frustration

1: gering	Strategie 1	1,69	1,83	1,25
6: hoch	Strategie 2	1,81	2,08	1
	Strategie 3	2	2,25	1,25

Die Unterschiede der Antworten einzelner Strategien sind nicht signifikant, so dass nicht eindeutig gesagt werden kann, welche Strategie die Versuchsperson am meisten bzw. am wenigsten belastet. Betrachtet man jedoch die Ergebnisse aller Fragen fällt auf, dass die Ergebnisse für die geistige Anforderung, die Anstrengung und die Frustration zeigen, dass die erste Strategie am unbelastendsten und die dritte Strategie am belastendsten gewertet wurde. Dies deckt sich auch mit den

Geistige Anforderung

Wie viel geistige Anforderung war bei der Informationsaufnahme und bei der Informationsverarbeitung erforderlich (z.B. Denken, Entscheiden, Rechnen, Erinnern, Hinsehen, Suchen ...)? War die Aufgabe leicht oder anspruchsvoll, einfach oder komplex, erfordert sie hohe Genauigkeit oder ist sie fehlertolerant?

1 2 3 4 5 6

Gering ☐ ☐ ☐ ☐ ☐ ☐ Hoch**Körperliche Anforderung**

Wie viel körperliche Aktivität war erforderlich (z.B. ziehen, drücken, drehen, steuern, aktivieren ...)? War die Aufgabe leicht oder schwer, einfach oder anstrengend, erholend oder mühselig?

1 2 3 4 5 6

Gering ☐ ☐ ☐ ☐ ☐ ☐ Hoch**Zeitliche Anforderung**

Wie viel Zeitdruck empfanden Sie hinsichtlich der Häufigkeit oder dem Takt mit dem die Aufgaben oder Aufgabenelemente auftraten? War die Aufgabe langsam und geruhsam oder schnell und hektisch?

1 2 3 4 5 6

Gering ☐ ☐ ☐ ☐ ☐ ☐ Hoch**Leistung**

Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?

1 2 3 4 5 6

Gering ☐ ☐ ☐ ☐ ☐ ☐ Hoch**Anstrengung**

Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?

1 2 3 4 5 6

Gering ☐ ☐ ☐ ☐ ☐ ☐ Hoch**Frustration**

Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?

1 2 3 4 5 6

Gering ☐ ☐ ☐ ☐ ☐ ☐ Hoch

Abbildung 3: Fragebogen: Nasa-TLX

Ergebnissen aus dem Strategien Fragebogen, in dem die erste Strategie als Favorit und die dritte Strategie als unbeliebteste Strategie gewertet wurde. Vergleicht man die Werte für Runde 1-3 mit den Werten von Runde 4 fällt auf, dass die Dialoge parallel zum Rennspiel mit Ausnahme der Frage nach der Leistung durchweg als belastender gewertet wurden als die Dialoge ohne Rennspiel. Dies zeigt, dass die Versuchspersonen einen Unterschied in der kognitiven Belastung gespürt haben.

Strategien

Der zweite Teil des Fragebogens ist in Abbildung ?? zu sehen.

In diesem Teil geht es darum, die einzelnen Strategie anhand verschiedener Kategorien zu bewerten. Die Nachfolgende Tabelle zeigt die Ergebnisse jeder Frage pro Strategie. Dabei werden jeweils die durchschnittlichen Antworten für alle Runden, der Runden mit Rennspiel und nur der Runde ohne Rennspiel aufgelistet.

Antwortenintervall	Strategien	Ergebnisse bestimmter Runden		
		1-4	1-3	4

Der Dialog lenkte mich vom Rennspiel ab

1: kaum 6: stark	Strategie 1	in Runde 4	2,25	nicht beantwortet
	Strategie 2	nicht	2,58	
	Strategie 3	beantwortet	2,58	

Die Nachfragen erleichterten es mir, den Anruf korrekt aufzubauen

1: erleichterte es 6: erschwerte es	Strategie 1	1,63	1,83	1,25
	Strategie 2	1,69	1,92	1
	Strategie 3	2,13	2,25	1,75

Wussten Sie, wann das System Spracheingaben erwartete?

1: immer 6: nicht immer	Strategie 1	1,19	1,17	1,25

Antwortenintervall	Strategien	Runden 1-4	Runden 1-3	Runde 4
	Strategie 2	1,38	1,33	1,5
	Strategie 3	1,5	1,67	1

Wie gefiel Ihnen der Dialog insgesamt?

1: sehr gut	Strategie 1	1,94	2,08	1,5
	Strategie 2	2,50	2,67	2
6: weniger gut	Strategie 3	2,57	2,75	2

Fiel es Ihnen einfacher, den Dialog ohne Rennspiel zu führen?

1: viel einfacher	Strategie 1	in Runden	nicht	2
	Strategie 2	1-3 nicht	beantwortet	3,75
6: nicht einfacher	Strategie 3	beantwortet		2,25

Welcher Anruf gefiel Ihnen insgesamt am besten?

Anruf bzw. Strategie auswählbar	Strategie 1	75%		
	Strategie 2	16,6%		
	Strategie 3	8,3%		

Die ersten vier Antworten dieses Fragebogen zeigen, dass die erste Strategie am positivsten und die dritte Strategie am negativsten gewertet wurden. Dies stimmt mit dem Ergebnis der letzten Frage überein, welche konkret nach der beliebtesten Strategie nachfragt. Die Vorletzte Frage zeigt, dass es dem Durchschnitt der Versuchspersonen einfacher fiel, den Dialog ohne Rennspiel zu führen. Dies bestätigt das Ergebnis aus dem Nasa-TLX Fragebogen, welches besagt, dass ein Unterschied in der kognitiven Belastung zwischen Dialog mit Rennspiel und ohne Rennspiel unter den Versuchspersonen bemerkbar ist.

Dialogverhalten

Wie zufrieden waren sie mit der Systeminteraktion

Der Dialog lenkte mich stark vom Rennspiel ab

Fiel es Ihnen schwer, das Rennspiel parallel zur Systeminteraktion zu spielen und so eine gute Leistung zu absolvieren?

1 2 3 4 5 6

lenkte mich kaum ab ☐ ☐ ☐ ☐ ☐ ☐ lenkte mich stark ab

Die Systemnachfragen erleichterte es mir, den Anruf korrekt aufzubauen

Hat das System dir dabei geholfen, die richtigen Personendaten einzugeben und somit eine Person korrekt mit den vorgegebenen Angaben anzurufen?

1 2 3 4 5 6

erleichterte die Eingaben ☐ ☐ ☐ ☐ ☐ ☐ erschwerte die Eingaben

Wussten Sie, zu welchem Zeitpunkt das System Spracheingaben erwartete?

Haben Sie gemerkt, wann das System auf eine Spracheingaben von Ihnen wartet um den Dialog fortzuführen?

1 2 3 4 5

habe die Stellen immer erkannt ☐ ☐ ☐ ☐ ☐ habe die Stellen nicht immer erkannt

Wie gefiel Ihnen der Dialog insgesamt?

1 2 3 4 5 6

Sehr gut ☐ ☐ ☐ ☐ ☐ ☐ Weniger gut

Gab es etwas was Ihnen an dem Dialog sehr gut gefiel?

Gab es etwas was Ihnen an dem Dialog nicht gefiel?

Abbildung 4: Fragebogen: Dialogverhalten

Person

In Abbildung 5 sind die Fragen dieses Fragebogens abgebildet. Die Fragen nach der Rennspiel- und Dialogerfahrung können für die spätere Auswertung der Zeiten interessant sein und eine mögliche Erklärung für stark abweichende Rennspiel- Und Dialogzeiten liefern.

5.4.3 Task Completion

Für jede Strategie wird die Task Completion ausgewertet, welche besagt, mit welchem Erfolg der Anruf ausgeführt wurde. Sie wird bemessen, in dem man für jeden richtig gefüllten Slot (siehe Tabelle 6) einen Punkt verteilt. Folgenden Punktzahlen sind also für jede Strategie möglich:

- 0 Punkte, wenn kein Slot richtig gefüllt wird
- 1 Punkt, wenn ein Slot richtig gefüllt wird
- 2 Punkte, wenn alle Slots richtig gefüllt wird

Zur Auswertung wird dann pro Strategie eine Durchschnittspunktzahl berechnet.

5.5 Hypothesen

beliebt: kurze Sprachausgabe, kurze Spracheingabe, wenig Aufmerksamkeit

unbeliebt: lange Sprachausgaben mit unnötigen Informationen, anstrengendes Zuhören

hoher CL: kurze knappe Eingabe (BargeIn, Füllwörter), Confirmations bevorzugt (Related Work) und möglichst viele Pausen (Strat 3)

kein/niedriger CL: weniger BargeIn und Füllwörter(Related Work)

Wie alt sind Sie?

Hast du Erfahrung mit Dialogsystemen?

1 2 3 4 5 6

gar keine Erfahrung ● ● ● ● ● viel Erfahrung

Spielst du oft Rennspiele?

1 2 3 4 5 6

sehr oft ● ● ● ● ● nie

Wie technikaffin sind Sie?

1 2 3 4 5 6

sehr technikaffin ● ● ● ● ● gar nicht technikaffin

Wie schwer fiel Ihnen die Einführungsrunde?

1 2 3 4 5 6

sehr schwer ● ● ● ● ● sehr einfach

Abbildung 5: Fragebogen: Person

5.6 Versuchspersonen

geplant: 10 bis max 30

Affinität zu Technik?

Studierende der CoLi?

Alter usw.

5.7 Control Panel

Um ein laufendes System zu simulieren wurde ein Control Panel entwickelt, welches verschiedene Sprachausgaben per Mausklick triggert. Damit kann der Versuchsleiter, der Wizard, die passenden Sprachausgaben auf entsprechende Benutzereingaben auslösen. Neben Ausgaben für die einzelnen Disambiguierungsstrategien sind weitere Sprachausgaben abgedeckt, welche oberflächlich zu jeder Eingabe des Benutzers eine Antwort bereit stellen und somit einen ungehindernten Ablauf des Dialogs gewährleisten. Zusätzlich dazu ist ein Stoppbutton enthalten, mit welchem per Klick alle aktiven Sprachausgaben abgebrochen werden können. Das Control Panel wurde mit JavaFx² entwickelt. Mit Hilfe des Programms JavaFX Scene Builder³ wurde zunächst das Design entwickelt und in einer .fxml Datei gespeichert. Diese wurde anschließend in Eclipse unter Installation des Plugins E(fx)clipse geladen und die Funktionen für die Sprachausgaben und des Stopp-Buttons implementiert. Die Sprachausgaben wurden online auf der Webseite <http://www.fromtexttospeech.com/> als .mp3 Datei generiert und anschließend zu .wav Dateien konvertiert. Abbildung 7 zeigt das Control Panel. Für jede anzurufende Person gibt es ein extra Tab mit speziellen Sprachausgaben. Die

² <http://docs.oracle.com/javase/8/javase-clienttechnologies.htm>

³ <http://www.oracle.com/technetwork/java/javase/downloads/javafxscenebuilder-info-2157684.html>

gemeinsamen Sprachausgaben wie **Cancel** und der Stoppbutton sind in jedem Personentab extra enthalten, damit eine schnelle Reaktion des Versuchsleiters möglich ist. Das Commonstab enthält die Begrüßungsausgabe. Zur Orientierung ist nach jedem speziellen Button die ausgelöste Sprachausgabe zu sehen.

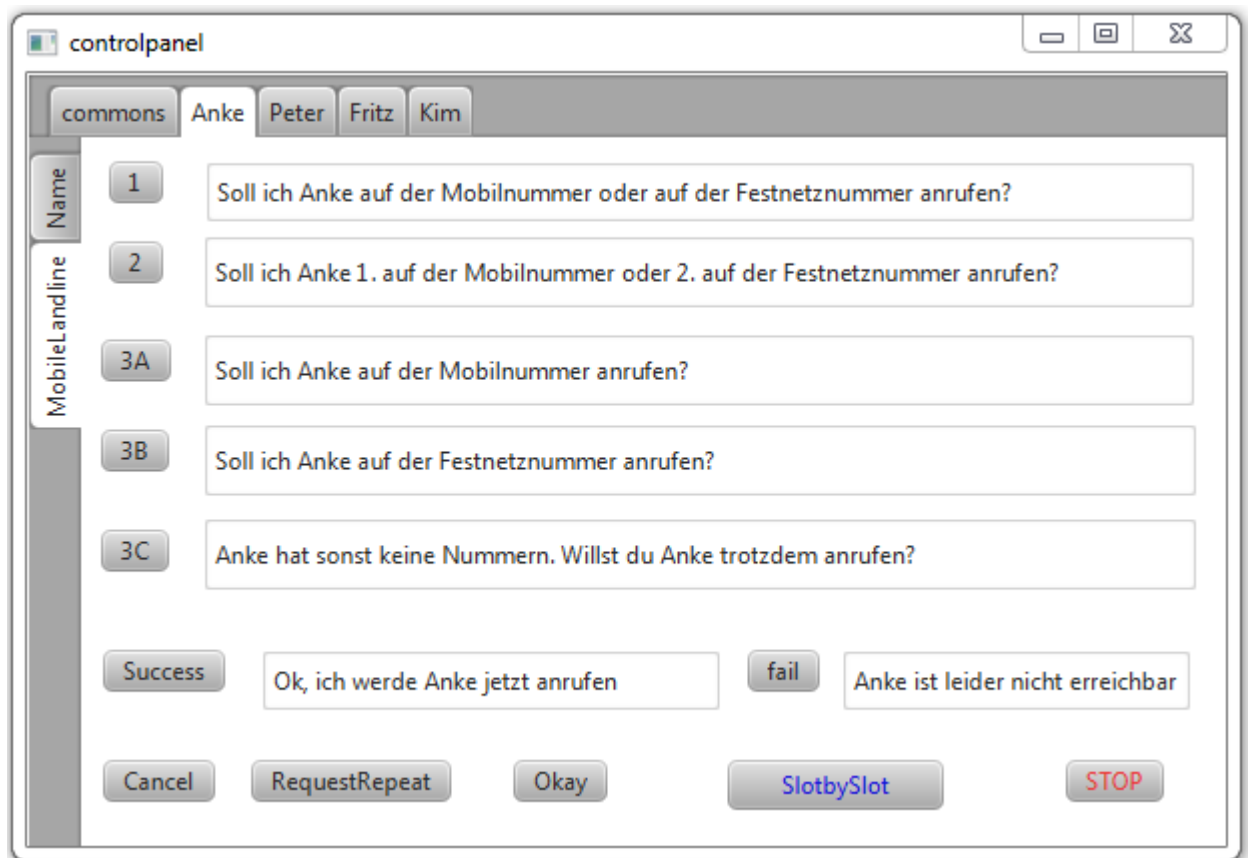


Abbildung 6: Controlpanel

5.8 Simulation und Durchführung

Versuchspersonen bekommen bestimmte Aufgabe

Ich habe eine System für das Auto gebaut, mit welchem ihr Telefonieren könnt. Bitte testet mal alle Funktionen des Systems

→ denken sie haben andere Aufgabe und wissen nicht was eigentlich getestet werden soll.

Versuchspersonen sollen also pro Runde einen Kontakt anrufen

Dauer einer Runde je nach Können ca. 2-3 Minuten

?? Versuchspersonen können/sollen sich vor jeder Runde kurz durchlesen, welche Interaktionen möglich sind

Bei jeder Interaktion gibt es Stellen, an denen ambige Spracheingabe getriggert werden.

Beispiele mit ambiger Satzeingabe und anschließender Disambiguierungsstrategie pro Aktion:

1. Strategie:

U: Rufe Paul an

S: Willst du Paul auf der Festnetznummer oder auf der Handynummer anrufen?

U: auf der Handynummer

2. Strategie mit Barge-in

U: Rufe Paul an

S: Willst du Paul auf 1. der Festnetznummer oder [..] anrufen.

U: Ja

2. Strategie mit deictic reference

U: Rufe Paul an

S: Willst du Paul auf 1. der Festnetznummer oder 2. auf der Handynummer anrufen.

U: ersteres **3.Strategie:**

U: Rufe Paul an

S: Willst du Paul auf der Festnetznummer anrufen?

U: Nein

S: Willst du Paul auf der Handynummer anrufen?

U: ja

Die zu testenden Disambiguierungsstrategien sind über die Runden verteilbar.

Jede Versuchsperson bekommt alle Disambiguierungsstrategien während des Testen präsentiert.

Am Schluß des gesamten Test soll die Versuchsperson einen Fragebogen ausfüllen (Nasa TLX (related work))

- Wie intuitiv war die Interaktion zu führen
- war die Interaktion während dem Fahren eher ablenkend oder störend?
- wie viel Aufmerksamkeit musste man dem System während der Interaktion schenken
- siehe nasa-tlx screenshot

??Anschließend über Disambiguierungsstrategien aufklären und über einzelne Strategien befragen. ?→ welche Strategie war am geeignetsten/einfachsten/intuitivsten für jeweilige Versuchsperson

5.9 Resultat

5.9.1 Zeiten

Rennzeiten

Tabelle 15: Durchschnittszeiten Strategie pro Strecke

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	71,5 sek	93 sek	74,5 sek
Strecke B	68,75 sek	75,75 sek	91,5 sek
Strecke C	74,5 sek	58,38 sek	61,75 sek
Insgesamt	71,58 sek	75,71 sek	75,92 sek

Dialogzeiten

Tabelle 16: Durchschnittszeiten Strategie pro Strecke

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	15,34 sek	20,38 sek	20,28 sek
Strecke B	14,31 sek	20,05 sek	22,07 sek
Strecke C	15,97 sek	21,01 sek	20,35 sek
Strecke A - C	15,19 sek	20,52 sek	20,81 sek
ohne Strecke	14,9 sek	18,8 sek	17,59 sek

Tabelle 17: Durchschnittszeiten mit Rennspiel versus ohne Rennspiel

Dialogzeiten	Runde 1-3	Runde 4
Durchschnitt	18,57 sek	17,59

5.9.2 Fragebogen

Nasa-TLX

Antwortenintervall	Strategien	Ergebnisse bestimmter Runden		
		1-4	1-3	4

Geistige Anforderung

1: gering 6: hoch	Strategie 1	1,88	2,08	1,25
	Strategie 2	2,06	2,42	1
	Strategie 3	2,63	2,83	2

Körperliche Anforderung

1: gering 6: hoch	Strategie 1	2	2,17	1,5
	Strategie 2	1,44	2,25	1
	Strategie 3	2	2,25	1,25

Zeitliche Anforderung

1: gering 6: hoch	Strategie 1	1,87	2,09	1,25
	Strategie 2	1,75	2	1
	Strategie 3	2,31	2,67	1,25

Leistung

1: gering 6: hoch	Strategie 1	4,25	4,5	3,5
	Strategie 2	4,75	4,33	6
	Strategie 3	4,25	4	5

Anstrengung

1: gering 6: hoch	Strategie 1	2	2,25	1,25
	Strategie 2	2,13	2,55	1
	Strategie 3	2,63	2,92	1,75

Frustration

Antwortenintervall	Strategien	Runden 1-4	Runden 1-3	Runde 4
1: gering 6: hoch	Strategie 1	1,69	1,83	1,25
	Strategie 2	1,81	2,08	1
	Strategie 3	2	2,25	1,25

Strategien

Antwortenintervall	Strategien	Ergebnisse bestimmter Runden		
		1-4	1-3	4

Der Dialog lenkte mich vom Rennspiel ab

1: kaum 6: stark	Strategie 1	in Runde 4	2,25	nicht beantwortet
	Strategie 2	nicht	2,58	
	Strategie 3	beantwortet	2,58	

Die Nachfragen erleichterten es mir, den Anruf korrekt aufzubauen

1: erleichterte es 6: erschwerte es	Strategie 1	1,63	1,83	1,25
	Strategie 2	1,69	1,92	1
	Strategie 3	2,13	2,25	1,75

Wussten Sie, wann das System Spracheingaben erwartete?

1: immer 6: nicht immer	Strategie 1	1,19	1,17	1,25
	Strategie 2	1,38	1,33	1,5
	Strategie 3	1,5	1,67	1

Wie gefiel Ihnen der Dialog insgesamt?

1: sehr gut 6: weniger gut	Strategie 1	1,94	2,08	1,5
	Strategie 2	2,50	2,67	2
	Strategie 3	2,57	2,75	2

Antwortenintervall	Strategien	Runden 1-4	Runden 1-3	Runde 4
--------------------	------------	---------------	---------------	---------

Fiel es Ihnen einfacher, den Dialog ohne Rennspiel zu führen?

1: viel einfacher	Strategie 1	in Runden	nicht	2
6: nicht einfacher	Strategie 2	1-3 nicht	beantwortet	3,75
	Strategie 3	beantwortet		2,25

Welcher Anruf gefiel Ihnen insgesamt am besten?

Anruf bzw. Strategie auswählbar	Strategie 1	75%		
	Strategie 2	16,6%		
	Strategie 3	8,3%		

5.9.3 Task Completion

Tabelle 20: Durchschnittliche Task Completion (TC)

Strategien	insgesamt	Runde 1-3	Runde 4
1. Strategie	1,75	1,92	1,5
2. Strategie	1,94	1,92	2
3. Strategie	1,63	1,5	2
Insgesamt		1,78	1,83

5.10 Qualitätssicherung

6 Versuch 2

Da die Ergebnisse des ersten Versuches sehr einheitlich gezeigt haben, dass bei einer Disambiguierung über zwei Möglichkeiten (zum Beispiel: Peter Müller oder

Peter Meier) die ersten Strategie am besten angekommen ist, hat man sich zusätzlich für einen weiteren Versuch entschieden. In diesem Versuch werden pro Disambiguierung mehr als zwei Möglichkeiten vorgeschlagen (zum Beispiel: Peter Müller, Peter Meier, Peter Lauer, Peter Fischer, Peter Schneider oder Peter Schmidt). Dabei will man herausfinden, ob die erste Strategie auch bei mehreren Vorschlägen bevorzugt wird.

6.1 Testszenario

Das Testszenario ist das gleiche wie im ersten Versuch. Die Versuchspersonen rufen jeweils Anke, Peter und Fritz bei parallelem Rennspiel an und anschließend Kim ohne Rennspiel. Der Versuch unterscheidet sich jedoch in den zu füllenden Slots, welche in Tabelle 21 aufgelistet sind. Die Anzahl der vorgeschlagenen Möglichkeiten für den jeweiligen Slot ist in Klammern angegeben. Welche Slots pro Person abgefragt werden, zeigt Tabelle 22.

Tabelle 21: Biespiel Slotabfragen

Slot	erfragte Werte
Typ(4)	geschäftliche Mobilnummer, geschäftliche Festnetznummer, private Mobilnummer oder private Festnetznummer?
Firma(6)	Kohlpharma, Möbel Martin, Globus, Sparkasse, Carglass oder Post
Nachname(6)	Meier, Bies, Schmidt, Bauer, Schuhmacher oder Schiller
Stadt(6)	Saarbrücken, Frankfurt, Köln, Berlin, Ingolstadt oder München

Tabelle 22: Slotabfrage pro Person

Anke	Peter	Fritz	Kim
Typ	Typ	Typ	Typ
Nachname			Nachname
	Firma		
		Stadt	

6.2 Versuchsaufbau

Der Versuchsaufbau ist identisch mit Versuch 1. Tabelle 23 zeigt einen Überblick.

Tabelle 23: Übersicht Versuchsablauf

1. Runde	2. Runde	3. Runde	4. Runde	5. Runde
Rennspiel	Rennspiel	Rennspiel	Rennspiel	
	Anruf Anke	Anruf Peter	Anruf Fritz	Anruf Kim

6.3 Versuchsdesign

Das Versuchsdesign wurde ebenfalls aus dem ersten Versuch übernommen. Ein Überblick der Strecken- und Strategieverteilung pro Gruppe ist in Tabelle 24 aufgelistet.

Tabelle 24: Strecken- und Strategieverteilung

Aufteilung	Strategie 1	Strategie 2	Strategie 3
1. Gruppe	Strecke A	Strecke B	Strecke C

Aufteilung	Strategie 1	Strategie 2	Strategie 3
2. Gruppe	Strecke B	Strecke C	Strecke A
3. Gruppe	Strecke C	Strecke A	Strecke B
4. Gruppe	keine Strecke	keine Strecke	keine Strecke

6.4 Auswertung

Wie in Versuch 1 werden die Zeiten gemessen, die die Versuchsperson zum einen für das absolvieren der Strecke und zum anderen für das erfolgreiche abschließen des Testszenarios benötigt (Unterkapitel 6.4.1). Nach jeder Rennrunde wird die Versuchsperson ebenfalls einen Fragebogen ausfüllen, welche sich zum einen auf die subjektiv wahrgenommene kognitive Belastung und zum anderen auf Merkmale der Disambiguierungsstrategien bezieht (Unterkapitel 6.4.2).

6.4.1 gemessene Zeiten

Rennzeiten

Tabelle 25: Durchschnittszeiten Strategie pro Strecke

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	sek Ø	sek Ø	sek Ø
Strecke B	sek Ø	sek Ø	sek Ø
Strecke C	sek Ø	sek Ø	sek Ø

Tabelle 43 ist nur eine Übergangstabelle, da einzelnen Werte durch schlechte Spieler in den Gruppen verfälscht werden können. Die endgültige Zeitberechnung für

die Analyse der effizientesten Disambiguierungsstrategie ist in Tabelle 26 dargestellt.

Tabelle 26: Durchschnittszeiten pro Strategie

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Durschnitt	sek Ø	sek Ø	sek Ø

Dialogzeiten

Neben den Zeiten für das Rennspiel werden auch die Dialogzeiten berechnet. In der Tabelle 27 sind die Durchschnittswerte der Dialogzeiten aus Runde 1-3 festgehalten. Diese Tabelle kann man gegen Tabelle 28 vergleichen, in der die Durchschnittswerte der Dialogzeiten von Gruppe 4, also der Dialogzeiten ohne Rennspiel, aufgelistet werden.

Tabelle 27: Durchschnittszeiten pro Strategie mit Rennspiel

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Durschnitt	sek Ø	sek Ø	sek Ø

Tabelle 28: Durchschnittszeiten pro Strategie ohne Rennspiel

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Durschnitt	sek Ø	sek Ø	sek Ø

6.4.2 Fragebogen

Die Fragebogen sind identisch mit diesen aus Versuch 1.

6.4.3 Task Completion

Für jede Strategie wird ebenfalls die Task Completion ausgewertet, welche besagt, mit welchem Erfolg der Anruf ausgeführt wurde. Folgenden Punktzahlen sind für jede Strategie möglich:

- 0 Punkte, wenn kein Slot richtig gefüllt wird
- 1 Punkt, wenn ein Slot richtig gefüllt wird
- 2 Punkte, wenn alle Slots richtig gefüllt wird

Zur Auswertung wird dann pro Strategie eine Durchschnittspunktzahl berechnet. Diese Vorgehensweise entspricht der aus Versuch 1.

6.5 Hypothesen

beliebt: kurze Sprachausgabe, kurze Spracheingabe, wenig Aufmerksamkeit

unbeliebt: lange Sprachausgaben mit unnötigen Informationen, anstrengendes Zuhören

hoher CL: kurze knappe Eingabe (BargeIn, Füllwörter), Confirmations bevorzugt (Related Work) und möglichst viele Pausen (Strat 3)

kein/niedriger CL: weniger BargeIn und Füllwörter(Related Work)

6.6 Versuchspersonen

geplant: 10 bis max 30

Affinität zu Technik?

Studierende der CoLi?

Alter usw.

6.7 Control Panel

Das Control Panel aus Versuch 1 wurde mit anderen Sprachausgaben ausgestattet und es wurden weitere Buttons für Strategie 3 hinzugefügt.

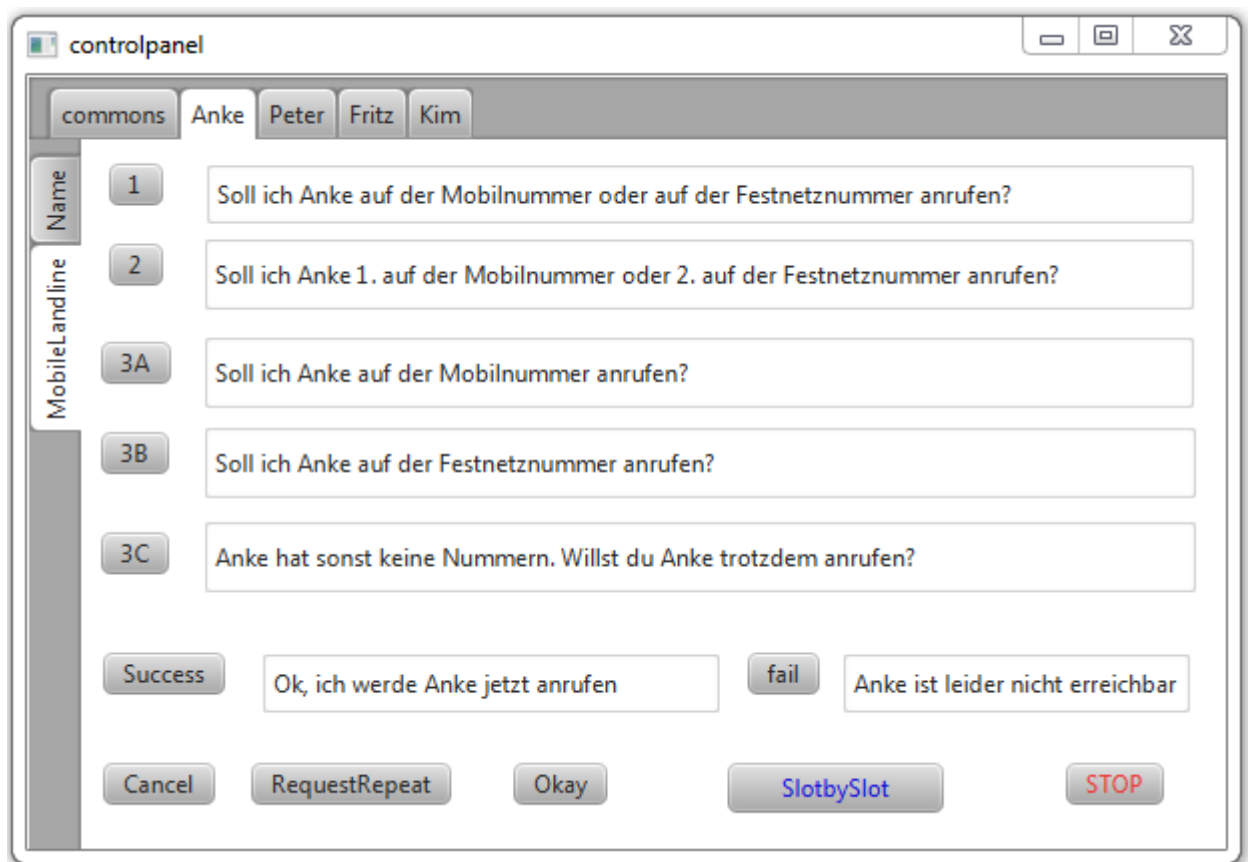


Abbildung 7: Controlpanel

6.8 Resultat

6.8.1 Zeiten

Rennzeiten

Tabelle 29: Durchschnittszeiten Strategie pro Strecke

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	81 sek	80,3 sek	74,25 sek
Strecke B	74,5 sek	84,25 sek	88 sek
Strecke C	75 sek	67,9 sek	65 sek
Strecke A - C	76,83	77,47	76,73

Dialogzeiten

Tabelle 30: Durchschnittszeiten Strategie pro Strecke

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	25,76 sek	38,32 sek	33,98 sek
Strecke B	31,41 sek	40,2 sek	36,61 sek
Strecke C	29,59 sek	37,9 sek	28,29 sek
Strecke A - C	29,55 sek	38,54 sek	34,34 sek
ohne Strecke	24,12	34,35	30,44

Tabelle 31: Durchschnittszeiten mit Rennspiel versus ohne Rennspiel

Dialogzeiten	Runde 1-3	Runde 4
Durchschnitt	34,64 sek	29,63 sek

6.8.2 Fragebogen

Nasa-TLX

Antwortenintervall	Strategien	Ergebnisse bestimmter Runden		
		1-4	1-3	4

Geistige Anforderung

1: gering 6: hoch	Strategie 1	2,31	2,67	1,25
	Strategie 2	2,31	2,58	1,5
	Strategie 3	2,56	2,83	1,75

Körperliche Anforderung

1: gering 6: hoch	Strategie 1	1,67	2,17	1
	Strategie 2	2,06	2,25	1,5
	Strategie 3	1,94	2,08	1,5

Zeitliche Anforderung

1: gering 6: hoch	Strategie 1	2	2,25	1,25
	Strategie 2	2,25	2,5	1,5
	Strategie 3	1,94	2,17	1,25

Leistung

1: gering 6: hoch	Strategie 1	5,13	4,83	6
	Strategie 2	4,88	4,67	5,5
	Strategie 3	4,56	4,08	6

Antwortenintervall	Strategien	Runden 1-4	Runden 1-3	Runde 4
--------------------	------------	---------------	---------------	---------

Anstrengung

1: gering 6: hoch	Strategie 1	2,38	2,83	1
	Strategie 2	2,44	2,67	1,75
	Strategie 3	2,31	2,67	1,25

Frustration

1: gering 6: hoch	Strategie 1	1,94	2,25	1
	Strategie 2	1,94	2	1,75
	Strategie 3	2,31	2,58	1,5

Strategien

Antwortenintervall	Strategien	Ergebnisse bestimmter Runden		
		1-4	1-3	4

Der Dialog lenkte mich vom Rennspiel ab

1: kaum 6: stark	Strategie 1	in Runde 4	3,5	nicht beantwortet
	Strategie 2	nicht	2,92	
	Strategie 3	beantwortet	3,17	

Die Nachfragen erleichterten es mir, den Anruf korrekt aufzubauen

1: erleichterte es 6: erschwerte es	Strategie 1	2,44	2,58	2
	Strategie 2	2,25	2,33	2
	Strategie 3	2,38	2,5	2

Wussten Sie, wann das System Spracheingaben erwartete?

1: immer 6: nicht immer	Strategie 1	1,63	1,67	1,5
	Strategie 2	1,5	1,5	1,5

Antwortenintervall	Strategien	Runden 1-4	Runden 1-3	Runde 4
	Strategie 3	1,57	1,5	1,75

Wie gefiel Ihnen der Dialog insgesamt?

1: sehr gut	Strategie 1	2,94	2,83	3,25
6: weniger gut	Strategie 2	2,44	2,42	2,5
	Strategie 3	2,38	2,3	2,5

Fiel es Ihnen einfacher, den Dialog ohne Rennspiel zu führen?

1: viel einfacher	Strategie 1	in Runden		2,75
6: nicht einfacher	Strategie 2	1-3 nicht	nicht	3
	Strategie 3	beantwortet	beantwortet	2,5

Welcher Anruf gefiel Ihnen insgesamt am besten?

Anruf bzw. Strategie auswählbar	Strategie 1	16,7%		
	Strategie 2	33,3%		
	Strategie 3	50%		

6.8.3 Task Completion

Tabelle 34: Durchschnittliche Task Completion (TC)

Strategien	insgesamt	Runde 1-3	Runde 4
1. Strategie	1,88	1,83	2
2. Strategie	1,81	1,75	2
3. Strategie	1,56	1,42	2
Insgesamt		1,67	2

6.9 Qualitätssicherung

7 Ergebnisse

7.1 Versuchsverlauf 1

7.1.1 Zeiten

Rennzeiten

Tabelle 35: Durchschnittszeiten Strategie pro Strecke

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	72,5 sek	93 sek	67 sek
Strecke B	68,75 sek	75,75 sek	94,5 sek
Strecke C	74,5 sek	58,37 sek	61,75 sek

Tabelle 36: Durchschnittszeiten pro Strategie mit Rennspiel

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Durchschnitt	71,92 sek	75,71 sek	74,42 sek

Dialogzeiten

Tabelle 37: Durchschnittszeiten Strategie pro Strecke

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	15,53 sek	20,38 sek	19,75 sek
Strecke B	14,31 sek	20,29 sek	18,08 sek
Strecke C	15,97 sek	21,01 sek	18,33 sek

Tabelle 38: Durchschnittszeiten pro Strategie mit Rennspiel

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Durchschnitt	15,27 sek	20,56 sek	18,72 sek

Tabelle 39: Durchschnittszeiten pro Strategie ohne Rennspiel

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Durchschnitt	14,9 sek	18,93 sek	17,59 sek

7.1.2 Fragebogen

7.1.3 Task Completion

Tabelle 40: Durchschnittliche Task Completion (TC)

Strategien	insgesamt	Runde 1-3	Runde 4
1. Strategie	1,64	1,76	1,25
2. Strategie	1,94	1,92	2

Strategien	ØTC samt	insge- ØTC Runde 1- 3	ØTC Runde 4
3. Strategie	1,55	1,38	2

7.2 Versuchsverlauf 2

7.2.1 Zeiten

Rennzeiten

Tabelle 41: Durchschnittszeiten Strategie pro Strecke

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	72,5 sek	93 sek	67 sek
Strecke B	68,75 sek	75,75 sek	94,5 sek
Strecke C	74,5 sek	58,37 sek	61,75 sek

Tabelle 42: Durchschnittszeiten pro Strategie mit Rennspiel

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Durchschnitt	71,92 sek	75,71 sek	74,42 sek

Dialogzeiten

Tabelle 43: Durchschnittszeiten Strategie pro Strecke

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	15,53 sek	20,38 sek	19,75 sek
Strecke B	14,31 sek	20,29 sek	18,08 sek
Strecke C	15,97 sek	21,01 sek	18,33 sek

Tabelle 44: Durchschnittszeiten pro Strategie mit Rennspiel

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Durchschnitt	15,27 sek	20,56 sek	18,72 sek

Tabelle 45: Durchschnittszeiten pro Strategie ohne Rennspiel

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Durchschnitt	14,9 sek	18,93 sek	17,59 sek

7.2.2 Fragebogen

7.2.3 Task Completion

Tabelle 46: Durchschnittliche Task Completion (TC)

Strategien	insgesamt	Runde 1-3	Runde 4
1. Strategie	1,64	1,76	1,25
2. Strategie	1,94	1,92	2

Strategien	ØTC samt	insge- ØTC Runde 1- 3	ØTC Runde 4
3. Strategie	1,55	1,38	2

8 Diskussion

was habe ich gemacht

wie waren die überlegungen warum wurden welche Entscheidungen getroffen warum wurden andere verworfen

8.1 Allgemeine Diskussion

Warum kann das Ergebnis verallgemeinert werden (cognitive load)

gilt nicht nur für Rennspielsimulation, sondern auch für andere Interaktionen(?)

8.2 Vergleichbare Studien

Vergleich mit anderen Studien möglich?

8.3 Future Work

VP in Gruppen unterteilen (je nach Wissenstand)

VP in Gruppen mit unterschiedlichen DisStrat aufteilen
andere DisSrat.

Unterschiede VP versch. Alters

9 Schlusswort

Literatur

- [Ang et al., 2006] Chee Siang Ang, Panayiotis Zaphiris, Shumalai Mahmood: *Cognitive Load Issues in MMORPGs* (2006).
- [Minker et al., 2002] W. Minker, U. Haiber, P. Heisterkamp, S. Scheible: *intelligent dialog strategy for accessing infotainment applications in mobile environments* ISCA Tutorial and Research Workshop (ITRW) on Multi-Modal Dialogue in Mobile Environments, Irsee (Germany) (June 2002).
- [Mishra et al., 2004] R Mishra, E Shriberg, S Upson, J Chen, F Weng, S Peters, L Cavedon, J Niekrasz, H Cheng, and H Bratt. *A wizard of Oz framework for collecting spoken human-computer dialogs.* (2004)
- [Tsiakoulis et al, 2012] P. Tsiakoulis, M. Henderson, B. Thomson, K. Yu, E. Tzir-
kel, S. Young: *The Effect of Cognitive Load on a Statistical Dialogue System* Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), pages 74–78, Seoul, South Korea, (July 2012).
- [Villing, 2009] Jessica Villing: *Dialogue behaviour under high cognitive load* Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue, pages 322–325,(2009)
- [Yin et al., 2007] Bo Yin, Natalie Ruiz, Fang Chen, M. Asif Khawaja: *Automatic cognitive load detection from speech feature* in OZ-CHI '07: Proceedings of the 19th Australasian conference on Computer-Human Interaction 249-255.