

# Bachelorarbeit

## Disambiguierungsstrategien in Dialogsystemen

Lena Enzweiler

Universität des Saarlandes

5. Januar 2015

# Einleitung



- Während Autofahrt kognitive Belastung
- manuelle Bedienung des Systems zusätzlich belastend
- Sprachsteuerung sicherer und weniger belastend.  
Voraussetzung: durchdachte Dialogsysteme

# Dialogsystem in automobilen Anwendungen

Effiziente Dialogsysteme im Auto müssen folgende Punkte erfüllen



- Ablenkung während der Fahrt vermeiden
  - alle Informationen kurz und verständlich übermitteln
  - einfache und intuitive Bedienung garantieren
- Sprachäußerungen müssen raffiniert gestaltet werden

## Fokus der Studie

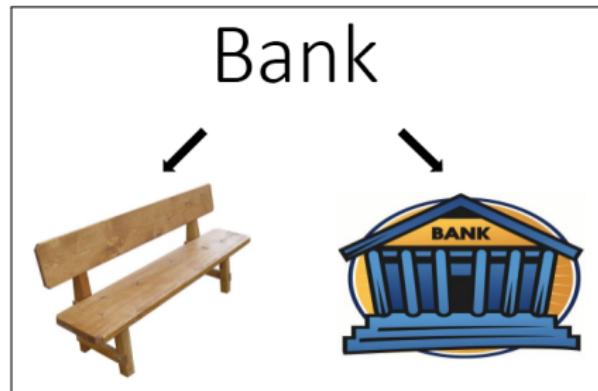
- ambige Eingaben des Benutzers möglich
  - System muss Mehrdeutigkeit der Eingaben auflösen
- Disambiguierung durch geschicktes Nachfragen beim Benutzer

### Fokus der Studie

Welche Disambiguierungsstrategie eignet sich für Dialogsysteme in einer automobilen Anwendung?

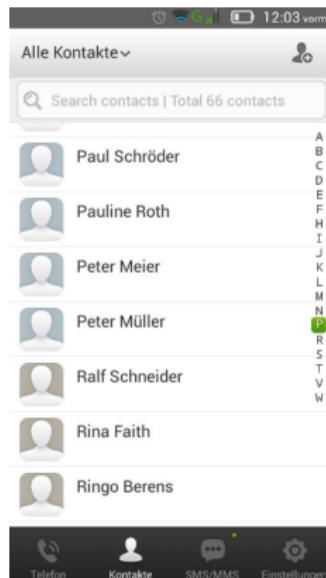
# Disambiguierung

Abgrenzung verschiedener Bedeutungen



# Disambiguierung in Dialogsystemen

- "Rufe Peter an!"
- System muss über Peter Meier und Peter Müller disambiguiieren
- unterschiedliche Disambiguierungsstrategien anwendbar



→ 3 Disambiguierungstrategien untersucht

# 1. Disambiguierungsstrategie: Aggregierte Auswahl ohne Pause

- alle möglichen Interpretationen in einer Sprachausgabe
- keine Pause zwischen Interpretationen
- auf Auswahl des Benutzers gewartet

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

## 2. Disambiguierungsstrategie: Aggregierte Auswahl mit Pause

- alle möglichen Interpretationen in einer Sprachausgabe
- Pause und Nummerierung zwischen Interpretationen
- auf Auswahl des Benutzers gewartet

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

### 3. Disambiguierungsstrategie: Sequentielle Auswahl

- alle möglichen Interpretationen in einer separaten Sprachausgabe
- auf Zustimmung/Ablehnung des Benutzer gewartet

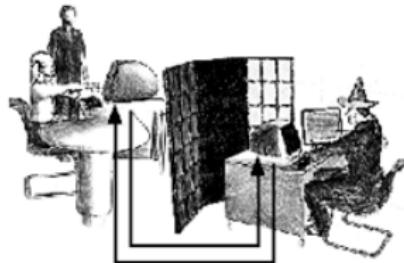
Akteur	Sprachausgabe
Benutzer	Rufe Peter an!
System	Meinst du Peter Meier?
Benutzer	Nein.
System	Meinst du Peter Müller?
Benutzer	Ja.
System	Ok, ich werde Peter Müller jetzt anrufen.

## Kurzbeschreibung

- Probanden fahren ein Rennspiel (hohe kognitive Belastung)
  - parallele Interaktion mit Dialogsystem
  - alle Disambiguierungsstrategien pro Versuchsperson untersucht
- 
- Probanden interagieren ohne Rennspiel (geringe kognitive Belastung)
  - eine Disambiguierungsstrategie zufällig getestet
- 
- Disambiguierungsstrategien auf Effizienz und Beliebtheit untersucht
- Ergebnisse mit und ohne Rennspiel werden miteinander verglichen

# Wizard-of-Oz

- Die Existenz eines funktionierenden Systems wird vorgetäuscht



- Versuchspersonen wird der Eindruck verliehen, sie würde mit einem echten Dialogsystem interagieren
- echtes Dialogsystem durch Versuchsleiter simuliert
- Control Panel entwickelt, mit welchem Sprachausgaben ausgeben werden können

## Testszenario

- Versuchspersonen sollen erfolgreich per Sprachsteuerung einen Anruf aufbauen
- insgesamt sollen vier Personen angerufen werden
- nach Anrufinitialisierung wird simuliert, dass die Spracheingabe zu unspezifisch ist
  - System stellt Rückfrage um zum Beispiel über mehrere mögliche Kontakte oder Telefonnummern zu disambiguiieren
- Nachfrage erfolgt in unterschiedlichen Strategien

### Beispiel

Benutzer: "Rufe Anke an"

System: "Meinst du Anke Meier oder Schuhmacher?"

# Testszenario

- relevante Personenangaben (Slots) werden über ein Personenprofil angezeigt.
- pro Anruf werden jeweils 2 Slots abgefragt.
- die zufüllenden Slots unterscheiden sich pro anzurufenden Kontakt
- Rückfragen sind so generiert, dass der Slot an zweiter Stelle der zu füllende ist

Anke Schumacher



Mobilnummer  
privat  
geschäftl.

Festnetznummer  
privat  
geschäftl.

 Mainzerstr. 23, 66121, Saarbrücken

 A.Schumacher86@gmx.de

# Versuchsaufbau

- Versuchspersonen fahren ein Rennspiel.  
→ Fahrsimulation
- Rennspiel: Need for Speed: Shift
- Rennspiel wird mit Lenkrad inklusive Gas- und Bremspedal gespielt  
→ realitätsgetreues Gefühl
- Es wird im Einzelrennen mit jeweils 5 Gegnern gespielt
- Versuchspersonen sollen möglichst hohe Platzierung erreichen  
→ Anstrengung und Konzentration soll hohe kognitive Belastung verursachen

# Versuchsaufbau - Rennspiel

Abbildung: Need for Speed - Shift



# Versuchsaufbau - Überblick

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

- Vorrunde zum Einspielen
- Runde 1-3: Rennspiel mit paralleler Systeminteraktion  
→ hohe kognitive Belastung
- Runde 4: nur Systeminteraktion  
→ geringe kognitive Belastung

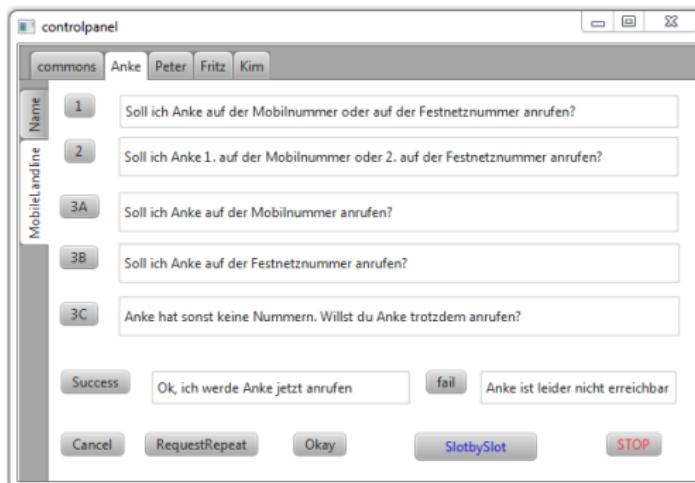
## Versuchsdesign

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

- 3 verschiedene Strecken, um Lerneffekt auszuschließen
- jede Strecke mit unterschiedlicher Disambiguierungsstrategie
- um Zeiten besser zu vergleichen:
  - Disambiguierungsstrategien werden auf Strecken verteilt
  - Versuchspersonen werden in Gruppen (1-3) aufgeteilt
- Die Strecken werden in gleicher Reihenfolge gefahren
- Gruppe 4 führt das Testszenario mit zufälliger Strategie aus.

## Control Panel

- entwickelt um ein laufendes Dialogsystem zu simulieren
- verschiedene Sprachausgaben können per Mausklick abgespielt werden



# Versuchspersonen

Versuchsablauf für eine Versuchsperson:

- ① Testrunde fahren
- ② Fragebogen über eigene Person ausfüllen
- ③ Strecke A fahren + Anke anrufen
- ④ Fragebogen über kognitive Belastung und Dialog ausfüllen
- ⑤ Strecke B fahren + Peter anrufen
- ⑥ Fragebogen über kognitive Belastung und Dialog ausfüllen
- ⑦ Strecke C fahren + Fritz anrufen
- ⑧ Fragebogen über kognitive Belastung und Dialog ausfüllen
- ⑨ Kim anrufen
- ⑩ Fragebogen über kognitive Belastung und Dialog ausfüllen

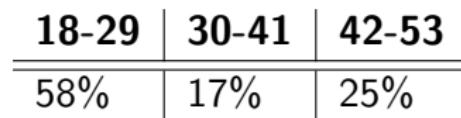
# Versuchsperson - Fragebogen

Fragebogen über eigene Person

Informationen über Person für spätere Auswertung benötigt

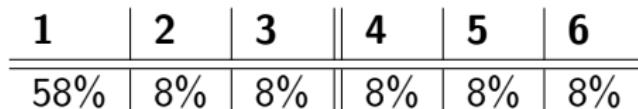
Wie alt sind Sie?

Altereingabe



Haben Sie Erfahrung mit Dialogsystemen?

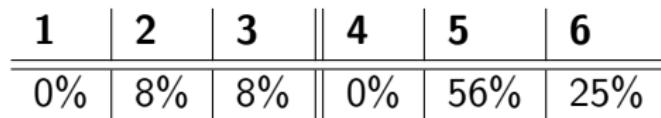
1: gar keine Erfahrung 6: viel Erfahrung



## Versuchsperson - Fragebogen

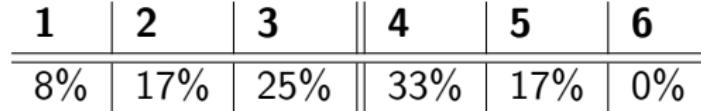
Spielen Sie oft Rennspiele?

1: sehr oft 6: nie



Wie technikaffin sind Sie?

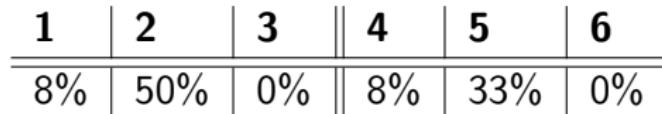
1: sehr technikaffin 6: gar nicht technikaffin



## Versuchsperson - Fragebogen

Wie schwer fiel Ihnen die Einführungsrunde?

1: sehr schwer 6: sehr einfach



# Auswertung

Folgende Punkte werden ausgewertet

- Zeiten werden gemessen
  - Rennzeiten
  - Dialogzeiten
- Fragebögen ausgewertet
  - Nasa-TLX
  - Strategien
- Task Completion
- Dialogverhalten

## Gemessene Zeiten - Rennzeiten

### Rennzeiten

Beeinflusst eine Disambiguierungsstrategie das Rennverhalten?

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	71,5 sek	93,0 sek	74,5 sek
Strecke B	68,8 sek	75,8 sek	91,5 sek
Strecke C	74,5 sek	58,4 sek	61,8 sek

Werte könnten durch unbalancierte Gruppen entstehen.

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

Zeiten statistisch nicht relevant und daher nicht aussagekräftig.

## Gemessene Zeiten - Rennzeiten

### Rennzeiten

Beeinflusst eine Disambiguierungsstrategie das Rennverhalten?

Rennzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	71,5 sek	93,0 sek	74,5 sek
Strecke B	68,8 sek	75,8 sek	91,5 sek
Strecke C	74,5 sek	58,4 sek	61,8 sek

Werte könnten durch unbalancierte Gruppen entstehen.

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

Zeiten statistisch nicht relevant und daher nicht aussagekräftig.

## Gemessene Zeiten - Dialogzeiten

### Dialogzeiten

Welche Strategie ermöglicht den kürzesten Dialog?

Nur die Zeiten von korrekt durchgeföhrten Dialogen bewertet.

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A	15,3 sek	20,4 sek	20,3 sek
Strecke B	14,3 sek	20,1 sek	22,1 sek
Strecke C	15,9 sek	21,0 sek	20,4 sek
ohne Strecke	14,9 sek	18,8 sek	17,6 sek

→ Strategie 1 ermöglicht den kürzesten Dialog.

## Gemessene Zeiten - Dialogzeiten

### Rennzeiten

Gibt es Unterschiede in den Dialogzeiten zwischen kognitiv hoch belasteten und kognitiv wenig belasteten Versuchspersonen?

Dialogzeiten	Strategie 1	Strategie 2	Strategie 3
Strecke A - C	15,2 sek	20,5 sek	20,8 sek
ohne Strecke	14,9 sek	18,8 sek	17,6 sek

- kürzere Dialogzeiten ohne Rennspiel erreicht
- Unterschiede nicht durch unterschiedliches Dialogverhalten erklärbar

→ bessere Reaktionszeit bei Dialoginteraktion ohne Rennspiel

# Fragebogen - Nasa-TLX

## Nasa-TLX

Bei welcher Strategie wurde eine höhere Belastung empfunden?

Gibt es Unterschiede in der Belastung zwischen den Runden mit und ohne Rennspiel?

# Fragebogen - Nasa-TLX

## Geistige Anforderung

1: Gering 6: Hoch

Strategien	Runde 1-4	Runde 1-3	Runde 4
Strategie 1	1,88	2,08	1,25
Strategie 2	2,06	2,42	1,00
Strategie 3	2,63	2,83	2,00

- Strategie 1 geringe geistige Anforderung
- Strategie 3 höchste geistige Anforderung
- Runde 4 weniger anfordernd als Runden mit Rennspiel

# Fragebogen - Nasa-TLX

## Anstrengung

1: Gering 6: Hoch

Strategien	Runde 1-4	Runde 1-3	Runde 4
Strategie 1	2,00	2,25	1,25
Strategie 2	2,13	2,55	1,00
Strategie 3	2,63	2,92	1,75

- Strategie 1 geringste Anstrengung
- Strategie 3 höchste Anstrengung
- Runde 4 weniger anstrengend als Runden mit Rennspiel

# Fragebogen - Nasa-TLX

## Weitere Fragen

### Körperliche Anforderung

1: Gering 6: Hoch

### Zeitliche Anforderung

1: Gering 6: Hoch

### Leistung

1: Gering 6: Hoch

### Frustration

1: Gering 6: Hoch

# Fragebogen - Strategien

## Strategien

Wie werden die Strategien von den Versuchspersonen bewertet?

Der Dialog lenkte mich stark vom Rennspiel ab

1: kaum 6: stark

Strategien	Runde 1-3
Strategie 1	2,25
Strategie 2	2,58
Strategie 3	2,58

→ Strategie 1 lenkte am wenigsten vom Rennspiel ab

## Fragebogen - Strategien

Wie gefiel Ihnen der Dialog insgesamt

1: Gering 6: Hoch

Strategien	Runde 1-4	Runde 1-3	Runde 4
Strategie 1	1,94	2,08	1,50
Strategie 2	2,50	2,67	2,00
Strategie 3	2,57	2,75	2,00

- Strategie 1 gefällt am besten
- Strategie 3 gefällt am wenigsten

## Fragebogen - Strategien

Fiel es Ihnen einfacher den Dialog ohne Rennspiel zu führen?

1: viel einfacher 6: nicht einfacher

Strategien	Runde 4
Strategie 1	2,00
Strategie 2	3,75
Strategie 3	2,25

→ der Dialog fiel im Durchschnitt ohne Rennspiel einfacher

## Fragebogen - Strategien

Welcher Anruf gefiel Ihnen insgesamt am besten

Anruf bzw. Strategie auswählbar

Strategien	nach Versuch
Strategie 1	75,0%
Strategie 2	16,6%
Strategie 3	8,3%

- Strategie 1 eindeutig am beliebtesten
- Strategie 3 am unbeliebtesten

# Fragebogen - Strategien

## Weitere Fragen

Die Nachfragen erleichterten es mir, den Anruf korrekt aufzubauen

1: erleichterte es 6: erschwerte es

Wussten Sie, wann das System Spracheingaben erwartete?

1: immer 6: nicht immer

- Strategie 1 am positivsten bewertet
- Strategie 3 am negativsten bewertet

# Task Completion

## Task Completion (TC)

Welche Strategie ist am erfolgversprechendsten?

- Die Task Completion wird für jeden Dialog wie folgt berechnet:
  - 0 Punkte, wenn kein Slot richtig gefüllt wird
  - 1 Punkt, wenn ein Slot richtig gefüllt wird
  - 2 Punkte, wenn alle Slots richtig gefüllt werden
- für jede Strategie wird die durchschnittliche Task Completion bewertet

## Task Completion

<b>Strategien</b>	<b>Runde 1-4</b>	<b>Runde 1-3</b>	<b>Runde 4</b>
1. Strategie	1,75	1,92	1,50
2. Strategie	<b>1,94</b>	1,92	2,00
3. Strategie	<b>1,63</b>	1,50	2,00
alle Strategien		<b>1,78</b>	<b>1,83</b>

- Strategie 2 am erfolgreichsten
- Strategie 3 am erfolgreichsten
- Runde ohne Rennspiel erfolgreicher als Runden mit Rennspiel

→ Unterschied gering: Mehr Werte benötigt für aussagekräftiges Ergebnis

## Auswertung - Zusammenfassung

- **kürzeste Rennzeit:** nicht aussagekräftig
- **kürzeste Dialogzeit:** Strategie 1
- Ergebnis **Nasa-TLX Fragebogen:**
  - Strategie 1 am unbelastetsten
  - Runde ohne Rennspiel weniger belastend als Runde mit
- Ergebnis **Fragebogen Strategie:**
  - Strategie 1 am positivsten bewertet
  - Dialog fiel im Durchschnitt ohne Rennspiel einfacher
- **Task Completion**
  - Strategie 2 > Strategie 1 > Strategie 3 (geringer Unterschied)
  - Dialog ohne Rennspiel erfolgreicher als Dialog ohne REnnspiel

⇒ **Strategie 1** am Effizientesten und Beliebtesten

# Vorahnung

*"The wise man avoids evil by anticipating it" (Publilius Syrus)*

Vorahnung ist lebenswichtig

- kein Halten von gefährlichen Tieren als Haustiere
- keine Spaziergänge bei Gewitter
- Reflexe ausüben

# Vorahnung

*"The wise man avoids evil by anticipating it" (Publilius Syrus)*

Vorahnung ist lebenswichtig

- kein Halten von gefährlichen Tieren als Haustiere
- keine Spaziergänge bei Gewitter
- Reflexe ausüben

# Vorahnung

*"The wise man avoids evil by anticipating it" (Publilius Syrus)*

Vorahnung ist lebenswichtig

- kein Halten von gefährlichen Tieren als Haustiere
- keine Spaziergänge bei Gewitter
- Reflexe ausüben

# Vorahnung

*"The wise man avoids evil by anticipating it" (Publilius Syrus)*

Vorahnung ist lebenswichtig

- kein Halten von gefährlichen Tieren als Haustiere
- keine Spaziergänge bei Gewitter
- Reflexe ausüben

# Vorahnung

*"The wise man avoids evil by anticipating it" (Publilius Syrus)*

Vorahnung ist lebenswichtig

- kein Halten von gefährlichen Tieren als Haustiere
- keine Spaziergänge bei Gewitter
- Reflexe ausüben

# Vorahnung

*"The wise man avoids evil by anticipating it" (Publilius Syrus)*

Vorahnung ist lebenswichtig

- kein Halten von gefährlichen Tieren als Haustiere
- keine Spaziergänge bei Gewitter
- Reflexe ausüben

# Vorahnung

*"The wise man avoids evil by anticipating it" (Publilius Syrus)*

Vorahnung ist lebenswichtig

- kein Halten von gefährlichen Tieren als Haustiere
- keine Spaziergänge bei Gewitter
- Reflexe ausüben