

Perspektiven in der Sprachtechnologie: Hörende, sprechende und verstehende Computer

Wolfgang Wahlster

Fachrichtung Informatik
Universität des Saarlandes &
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
Stuhlsatzenhausweg 3, Geb. 43.1
66123 Saarbrücken
Tel.: (0681) 302-5252/4162
Fax: (0681) 302-5341
E-mail: wahlster@dfki.de
WWW: <http://www.dfki.de/~wahlster>

Hauptziel: Wir wollen durch Künstliche Intelligenz die Technik menschenfreundlicher machen!

Speziell: Der Mensch soll sich nicht dem Computer anpassen mü
sondern umgekehrt soll sich der Computer dem Me
anpassen!

Bisher:

Für den Mensch schwer zu erzeugen: Für den Computer leicht zu versteh
Künstliche Sprache, z.B. Java



```
for ( int i = 2; i < finalBit; i++)
  if ( sieve.get( i ) )
    for ( int j = 2 * i; j < size; j +=
        sieve.clear( j );
```



Neu:

Für den Mensch leicht zu erzeugen: Für den Computer schwer zu versteh
Natürliche Sprache, z.B. Deutsch



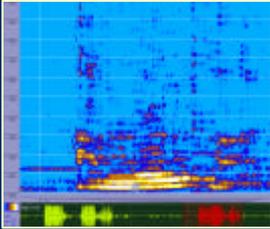
Ist 677 eine Primzahl?

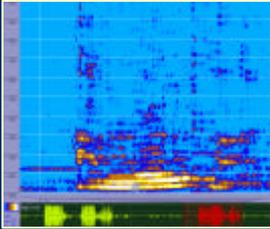


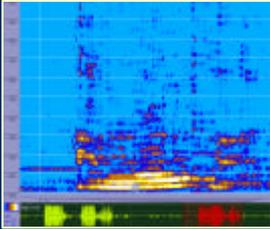
Warum ist Sprachverstehen für den Computer so schwer?

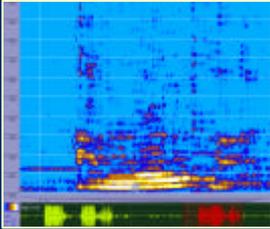
☞ Wortgrenzen gehen im Sprachfluß unter:

Beispiel: „amontag“ → „am Montag“

☞ Der Mensch  ma“

Beispiel:  klappt es“

Bedeutung  cht klappt es.

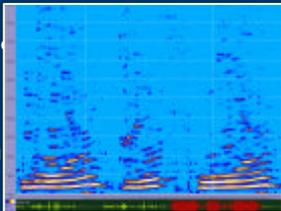
oder  ht. Klappt es?

© W. Wahlster, DFKI

Warum ist Sprachverstehen für den Computer so schwer?

☞ Gleiche Schallwellen werden je nach Kontext zu
verschiedenen Wörtern

Beispiel: „Urlauber wollen wieder me:r ans me:r“
→ Urlauber wollen wieder mehr ans Meer.

☞ Viele Mens 

Beispiel:

Bedeutung

oder

© W. Wahlster, DFKI

Warum ist das Sprachverstehen für den Computer so schwer?

☞ Bei spontaner Rede entstehen viele Versprecher

Beispiel: „Wir treffen uns dann am Mon, äh, am Dienstag.“

☞ Dialogpartner fallen dem Sprecher oft „ins Wort“

Beispiel: Sytem: „Können wir dann am Mittwoch zusammen zum Essen?“

Sprecher: „Da kann ich nicht.“

© W. Wahlster, DFKI

Warum ist das Sprachverstehen für den Computer so schwer?

☞ Der Redefluß leitet häufig in die Irre

Beispiel: „Die Staatssekretärin der
~~Objekt: Staatssekretärin~~ Prädikat: ~~Ministerpräsident.~~“

Subjekt: Ministerpräsident Prädikat: lobt Objekt: Staatssekre

☞ Viele Formulierungen sind mehrdeutig

Beispiel: „Wir telefonierten mit Freunden in Japan.“

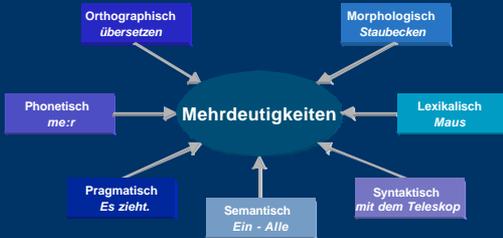
Bedeutung (1) Wir telefonierten (mit Freunden in Japan).

oder (2) (Wir telefonierten mit Freunden) in Japan.

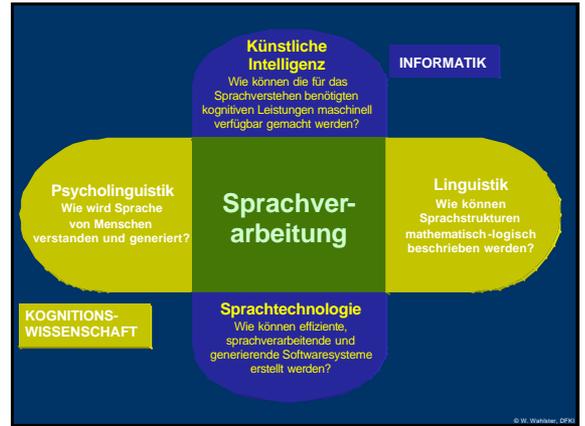
© W. Wahlster, DFKI

Disambiguierung: Auflösung mehrdeutiger sprachlicher Äußerungen

- Problem der kombinatorischen Explosion der Lesarten durch Propagierung von Alternativen über alle Verarbeitungsebenen
- Durch die Unsicherheit bei der Spracherkennung entstehen Wörtergitter mit alternativen Hypothesen, welche die Flut von Lesarten noch weiter erhöhen



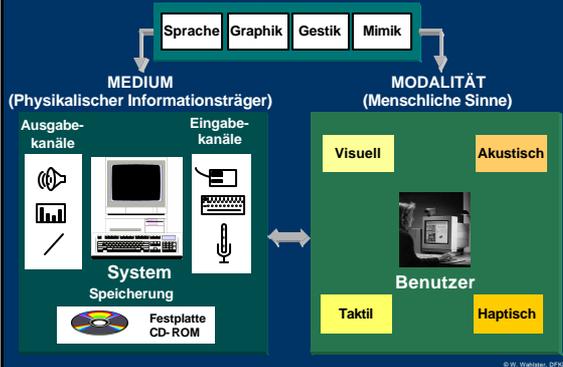
© W. Waikitor, DFKI



© W. Waikitor, DFKI

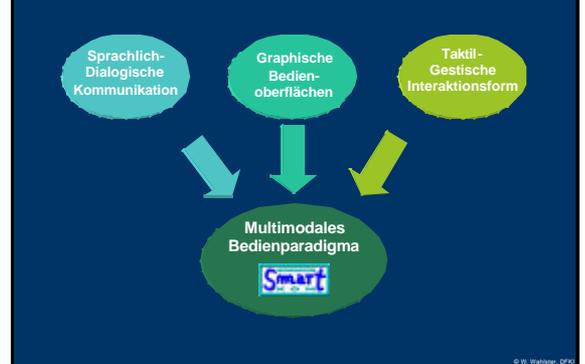
Code, Medium und Modalität

CODE (Symbolsysteme)



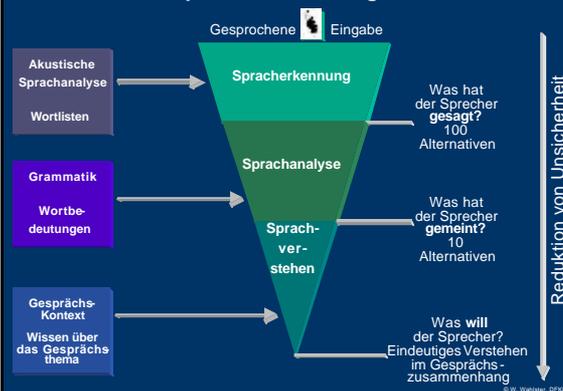
© W. Waikitor, DFKI

Die Leitvorstellung multimodaler Interaktion



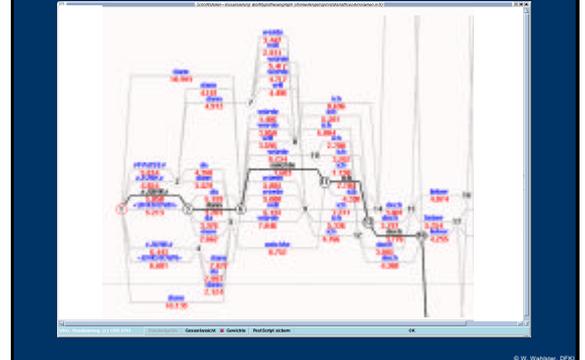
© W. Waikitor, DFKI

Drei Stufen der Sprachverarbeitung



© W. Waikitor, DFKI

Ergebnis der Spracherkennung: Worthypothesengraph



© W. Waikitor, DFKI

Verzögerte Disambiguierung Skopusmehrfachdeutigkeiten auf der Basis von Unterspezifikation

- (A) **Einen** Computer benutzen **alle** Informatikstudenten.
 (1) $Sx(\text{computer}(x) \cup y(\text{informatik-student}(y) \text{ benutzt}(y,x)))$
 (2) $y(\text{informatik-student}(y) \text{ benutzt}(y,x) \cup Sx(\text{computer}(x) \text{ benutzt}(y,x)))$

Unterspezifizierte Repräsentation (ohne Skopusdisambiguierung)

- (3) $\{Sx: \text{computer}, y: \text{informatik-student}\}(\text{benutzt}(y,x))$
 (B) Das ist der Zentralrechner PDP-10. <vor 20 Jahren>
 (1)
 (C) Oft bringen sie ihr Notebook mit in die Vorlesung. <heute>
 (2)

© W. Wahlster, DFG

Mobile Dialogübersetzung für Spontansprache



Lösung: Dreierkonferenz: Der VerbMobil-Server vermittelt zwischen zwei Mobilfunkteilnehmern

© W. Wahlster, DFG

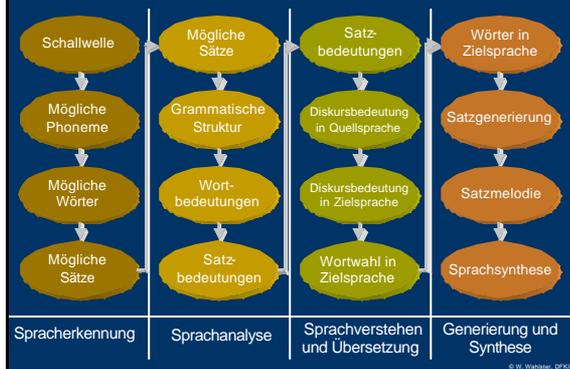
Mobile Dialogübersetzung über Handy mit dem System VERBMOBIL



Ergebnis von 8 Jahren Forschung mit einem Team von 100 Wissenschaftlern, 20 Produkte und 8 Spin-Off Firmen

© W. Wahlster, DFG

Von der Eingabeschallwelle zur Ausgabeschallwelle



© W. Wahlster, DFG

Deutscher Zukunftspreis 2001 – Preis des Bundespräsidenten für Technik und Innovation für Sprachverstehende Computer



© W. Wahlster, DFG

Vom Sprachdialog zum Multimodalen Dialog

Klassische Mobiltelefonie



Verbmobil
Reine Sprache

UMTS-Mobilgerät der dritten Generation



SmartKom
Sprache, Graphik, Gestik

© W. Wahlster, DFG

Deep Map: Multimodaler mobiler Touristenführer für Heidelberg
Kooperation u.a.: EML - DFKI - ISL



Mobile Dialogführung

Lokationsadaptive Interpretation von Benutzeranfragen



© W. Wahlster, DFKI

Deep Map: Multimodaler mobiler Touristenführer für Heidelberg



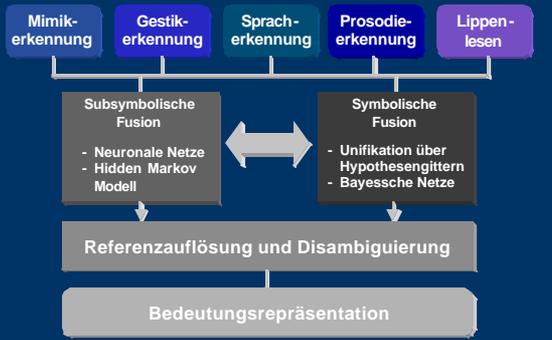
Sprachgesteuerte 3D-Visualisierung

Multimodale Präsentationsplanung (Text, Graphik, Bilder)



© W. Wahlster, DFKI

Symbolische und Subsymbolische Fusion von Eingabemodalitäten



© W. Wahlster, DFKI

Fokussierende Geste zur Disambiguierung der Spracheingabe (Wahlster 1991)

X	B	A	2
Y	A	C	1
Z	A	E	3

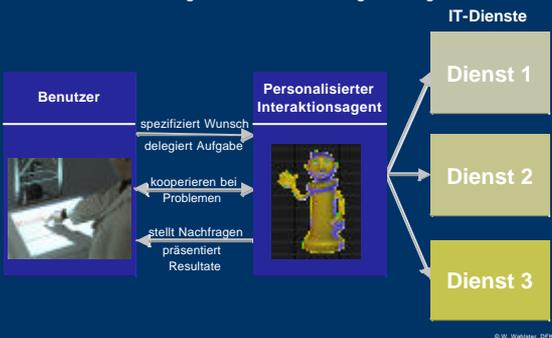


„Warum soll ich das ‚A‘ löschen?“

© W. Wahlster, DFKI

Die SDDP-Interaktionsmetapher für SmartKom

SDDP = Situated Delegation-oriented Dialog Paradigm



© W. Wahlster, DFKI

Kombination von Sprache und Gestik in SmartKom

Auf der Karte sind die Kinos markiert, in denen der Film „Eine kleine Weihnachtsgeschichte“ läuft.



© W. Wahlster, DFKI

Multimodale Ein- und Ausgabe in SmartKom



© W. Wabnitz, DFG

Wechselseitige Disambiguierung durch Multiple Eingabemodalitäten

Die kombinierte Sprach- und Bildverarbeitung erhöht die Robustheit und die Verstehensleistung multimodaler Benutzerschnittstellen



© W. Wabnitz, DFG

Disambiguierung durch Selektionsrestriktionen und Weltwissen

Vater zu einem Service-Roboter im Cyber-Restaurant:

(1) Die Apfelschorle trinkt meine Tochter, die Weinschorle meine Frau.

(A) trinkt (Agens: Apfelschorle, Objekt: Tochter) Ü
trinkt (Agens: Weinschorle, Objekt: Frau)

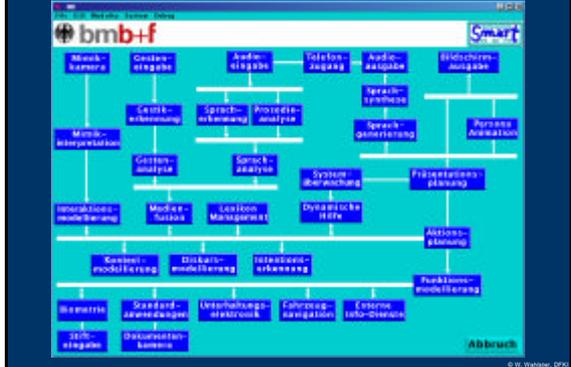
Weltwissen: Apfelschorle, Weinschorle \in Getränk
Tochter, Frau \in Mensch

Selektionsrestriktion: trinkt (Agens: Mensch, Objekt: Getränk)

(B) trinkt (Agens: Tochter, Objekt: Apfelschorle) Ü
trinkt (Agens: Frau, Objekt: Weinschorle)

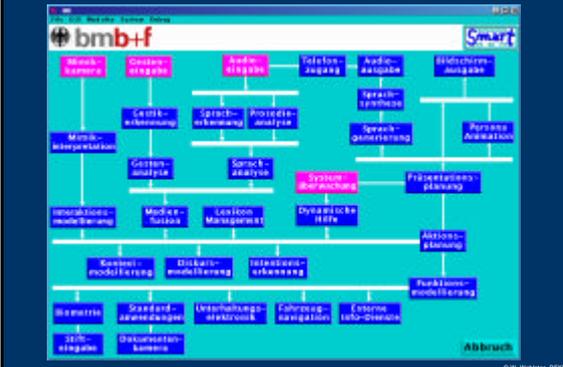
© W. Wabnitz, DFG

Fusion und Allokation multipler Modalitäten in SmartKom



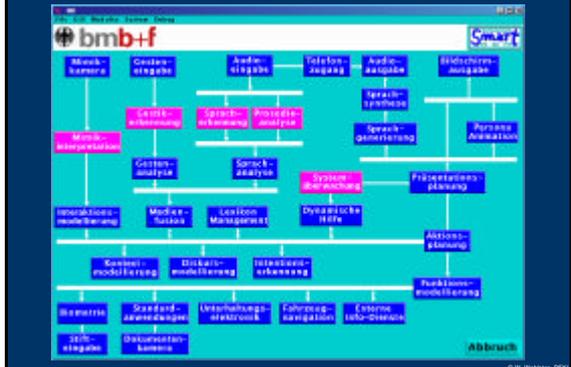
© W. Wabnitz, DFG

Fusion und Allokation multipler Modalitäten in SmartKom



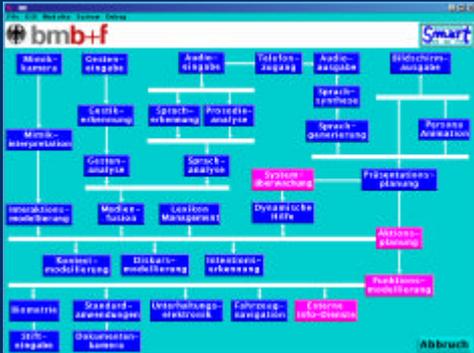
© W. Wabnitz, DFG

Fusion und Allokation multipler Modalitäten in SmartKom



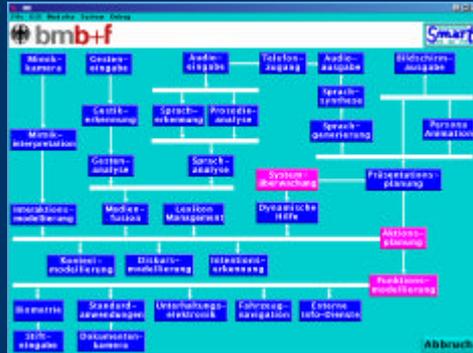
© W. Wabnitz, DFG

Fusion und Allokation multipler Modalitäten in SmartKom



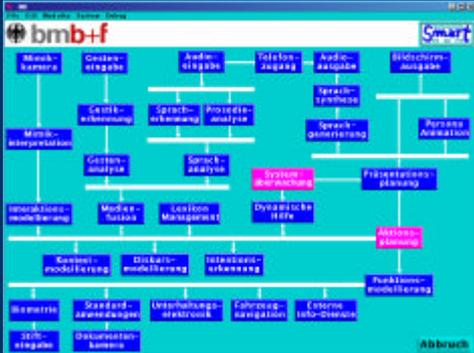
© W. Wabnitz, DFB

Fusion und Allokation multipler Modalitäten in SmartKom



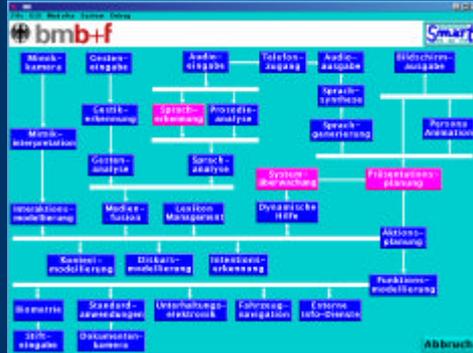
© W. Wabnitz, DFB

Fusion und Allokation multipler Modalitäten in SmartKom



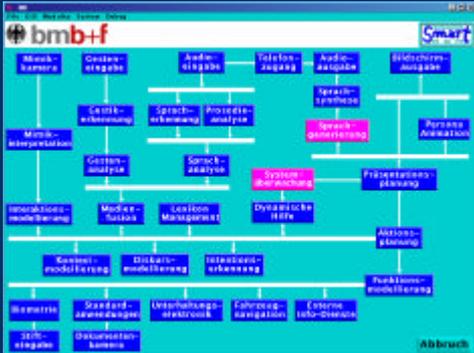
© W. Wabnitz, DFB

Fusion und Allokation multipler Modalitäten in SmartKom



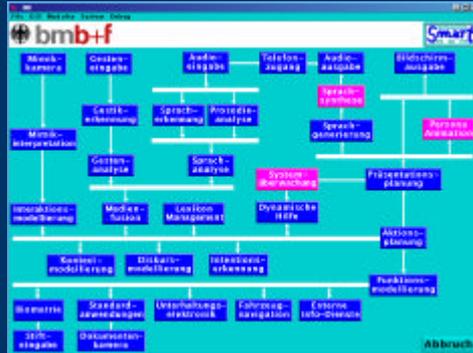
© W. Wabnitz, DFB

Fusion und Allokation multipler Modalitäten in SmartKom



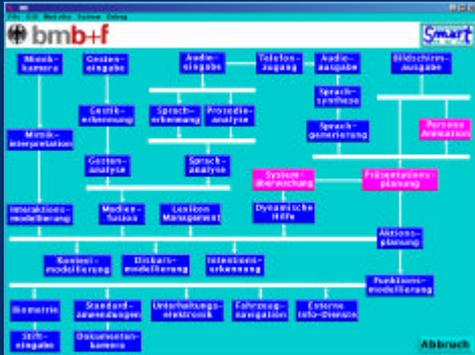
© W. Wabnitz, DFB

Fusion und Allokation multipler Modalitäten in SmartKom

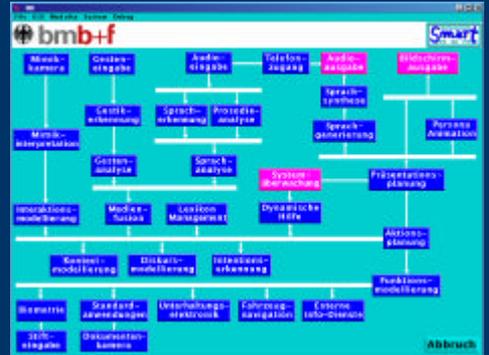


© W. Wabnitz, DFB

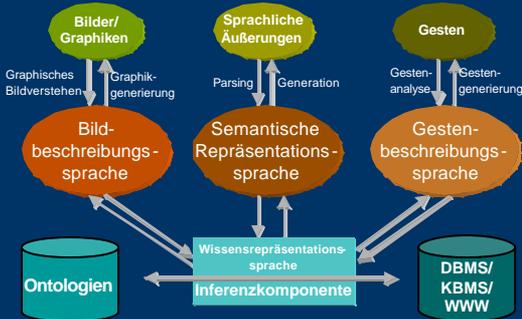
Fusion und Allokation multipler Modalitäten in SmartKom



Fusion und Allokation multipler Modalitäten in SmartKom



Modalitätsspezifische Repräsentations-sprachen als Zwischenstufe zur Medienfusion



Fusion von Sprach- und Mimikerkennung in SmartKom

Modifikation bis hin zur Negation der Standardsemantik (Ironie, Sarkasmus)

(1) Smartakus: Hier sehen Sie die Übersicht zum heutigen ZDF-Programm.

(2) Benutzer: Echt toll.



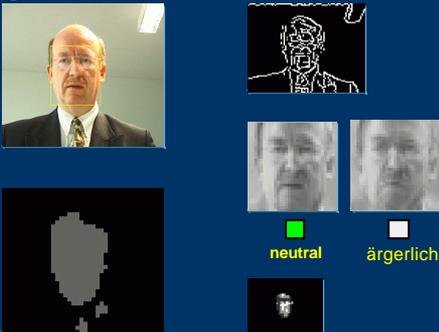
(3) Smartakus: Ich zeige Ihnen alternativ das Programm eines anderen Senders.

(2') Benutzer: Echt toll.

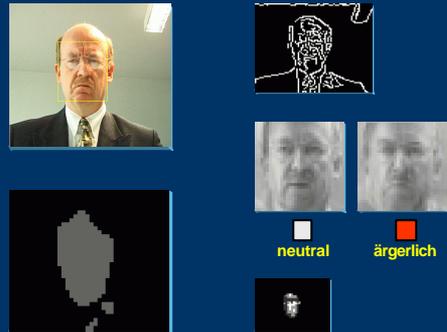


(3') Smartakus: Welche Sendungen wollen Sie aus dem ZDF-Programm sehen oder aufzeichnen?

Videobasierte Mimikerkennung auf der Basis von Eigenfaces



Sprecherunabhängige Emotionserkennung



Multimodale Dialoge mit Navigationssystemen für Autofahrer und Fußgänger

SmartKom bietet einen uniformen Navigationsdialog trotz unterschiedlicher Positionierungstechnologien



Benutzer: Ich möchte nach Heidelberg fahren.

Smartakus: Wollen Sie die schnellste oder kürzeste Strecke fahren?

Benutzer: Die schnellste.

Smartakus: Hier sehen Sie eine Karte mit der schnellsten Verbindung von Saarbrücken nach Heidelberg.

© W. Walkner, DFKI

Multimodale Dialoge mit Navigationssystemen für Autofahrer und Fußgänger



Smartakus: Wir sind jetzt in Heidelberg angekommen. Hier ist ein Stadtplan mit den wichtigsten Sehenswürdigkeiten.

User: Ich möchte mehr Information über diese Kirche.

Smartakus: Hier sieht die Webseite über die Peterskirche.

User: Wie komme ich zu Fuß am besten von diesem Parkplatz zu der Kirche.

Smartakus: Auf dieser Karte habe ich den Weg markiert.

© W. Walkner, DFKI

Multimodale Kommunikation mit Unterhaltungselektronik

Beispiel: Zugriff auf digitale Programmführer



Benutzer: Welche Sender zeigen gerade aktuelle Nachrichten?

Smartakus: CNN und NTV zeigen gerade Nachrichten.

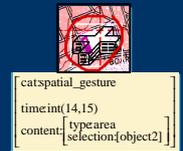
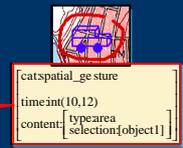
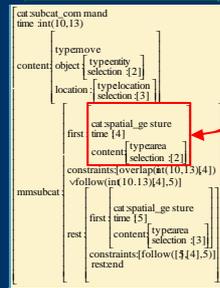
User: Zeige diesen Sender am Fernsehen und zeichne diese Nachrichtensendung auf.

Smartakus: Okay, CNN wird jetzt vom Videorekorder aufgezeichnet und hier sind die NTV-Nachrichten.

© W. Walkner, DFKI

Unifikationsbasierte Integration von Sprach- und Gestikeingabe (MVPQ, © Johnston 2000, AT&T)

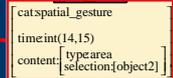
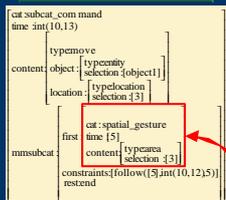
“Platziere das hier!”



© W. Walkner, DFKI

Unifikationsbasierte Integration von Sprach- und Gestikeingabe (MVPQ, © Johnston 2000, AT&T)

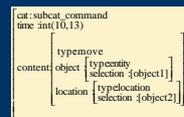
“Platziere das hier!”



© W. Walkner, DFKI

Unifikationsbasierte Integration von Sprach- und Gestikeingabe (MVPQ, © Johnston 2000, AT&T)

“Platziere das hier!”



© W. Walkner, DFKI

Sprachtechnologie im Alltag

Einen Capuccino in 10 Minuten bitte!



Sprachgesteuerte Kaffeemaschine

Bitte in die Winterbergstraße in Saarbrücken!



Sprachdialog mit Fahrzeugelektronik

Ich würde gerne Mozarts Klavierkonzert Nummer 3 hören!



Sprachbasierte Musikauswahl

Sende folgende Email an Meyer; Sehr geehrter Herr Meyer, Bitte senden Sie dringend die Agenda für Montag.



Diktat von Emails



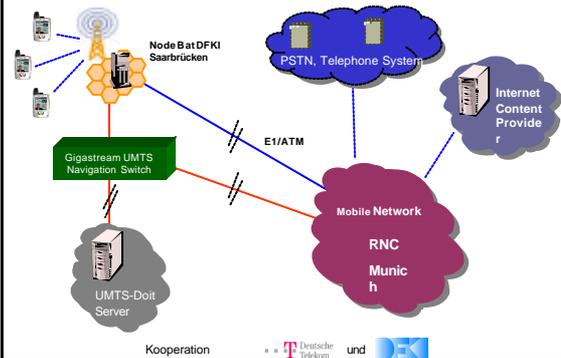
© W. Walkner, DFKI

Klassische Sprachsteuerung im Auto: Kein spontansprachlicher Dialog - Sprachaktivierungstaste



© W. Walkner, DFKI

Erstes UMTS-Anwendungstest- und Evaluationszentrum



Kooperation

Deutsche Telekom und



Embassy: Sprachbasierte Musikwahl



© W. Walkner, DFKI

ORBA Auskunftsdienst: „ShopFinder“

„Wo gibt es nächsten Geldautomaten?“

- ORBA greift auf online Produktkataloge, Shop- und Branchenverzeichnisse zu und lokalisiert die im jeweiligen Erreichbarkeitsradius liegenden Shops, von denen bekannt ist, dass sie das gesuchte Produkt vorhalten.
- ShopFinder ist kombinierbar mit:
 - multimedialer Produktinformation
 - Routenplanung und Zielführung



© W. Walkner, DFKI

ORBA Erinnerungsdienst: „ActiveList“

„Erinnere mich an die Batterie für die Uhr, wenn ich an einem entsprechenden Shop vorbeikomme!“

- ORBA erinnert an vorher notierte Kaufwünsche, sobald der Benutzer an dem explizit angegebenen Shop vorbeikommt bzw. sich irgendeinem Shop nähert, von dem bekannt ist, dass er das gesuchte Produkt vorhält.
- ActiveList ist kombinierbar mit:
 - multimedialer Produktinformation;
 - „mCoupon“ Aktionen;



© W. Walkner, DFKI

ORBA Ortungsdienst: „PartnerRadar“

„Wo (...) stecken Lisa und Tom“

- ORBA hilft beim Aufspüren und der Zusammenführung von Familien- Gruppenmitgliedern, sofern diese ebenfalls mit einem Mobilgerät ausgestattet sind, und ihre Ortung gestatten.
- PartnerFinder ist kombinierbar mit:
 - Routenplanung und Zielführung
 - Geschäftsempfehlungen, z.B. zentral gelegenes Restaurant als Treffpunkt.



© W. Wählner, DFKI

Sprachtechnologie im Alltag

Zeige mir alle Beiträge der Tagesschau, in denen Regierungsmitglieder zur Green Cardsprachen!



Sprachgesteuerte Suche in digitalen Fernseharchiven

Was hat der Ministerpräsident auf der COLING zur Sprachtechnologie gesagt?



Inhaltliche Suche in privaten Audioarchiven

Ich möchte gerne einen Termin mit Doktor Kurematsu nächste Woche in Kyoto ausmachen!



Dialogübersetzung

© W. Wählner, DFKI

Sprachtechnologie für die Post-PC Ara Personalisierte Benutzeroberflächen

• In der Hand *tragbar*

Mobile Computing



• Am Körper *anziehbar*

Wearable Computing



• Im Körper *implantierbar*

Biohybrid Computing



© W. Wählner, DFKI

Offene Forschungsprobleme

● Probleme der maschinellen Lernverfahren

- ⊗ Teure Datensammlung
- ⊗ Kognitiv unrealistische Trainingsdaten
- ⊗ Datenknappheit

● Probleme mit manuell erstellten Wissensquellen

- ⊗ Mangelnde Robustheit
- ⊗ Domänenabhängigkeit
- ⊗ Geringe Skalierbarkeit

© W. Wählner, DFKI

Vielen Dank für
Ihre Aufmerksamkeit