

# Dialogsysteme in der Praxis

Dr.-Ing. Bernd Plannerer

mail: [info@speech-recognition.de](mailto:info@speech-recognition.de)

[www.speech-recognition.de](http://www.speech-recognition.de)

## Inhalt

- Typischer Projektverlauf
- Zustandsbasierte Dialogsysteme
- Dialogdesign
- Spracherkennung
- Schwächen der Spracherkennung
- Lösungsansätze

## Typischer Projektverlauf

- Anforderungsanalyse / Spezifikation
- Dialogentwurf
- Definition der Lexika / Grammatiken
- Datenaufbereitung
- Interner Dialogtest mit Prototypen
- Aufnahme der Prompts
- Pilotbetrieb
- Systemoptimierung

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Projektdokumente

- Requirements Specification  
(zusammen mit dem Kunden)
- System Documentation  
(extern, für den Kunden)
- System Design Documentation  
(=> intern, Implementierung)

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Projektdokumente

- Testdokumente
  - Akzeptanzkriterien
  - Interne Testfälle
  - Testfälle für Abnahmetest
- Projektplan
  - mit Aufgaben und Terminen für:
    - Kunde
    - Auftragnehmer
    - Dritte (z.B.: weitere Lieferanten)

## Requirements Specification

- Geschäftsmodell
  - Am Geschäftsmodell orientieren sich die wichtigsten Designentscheidungen
- Dialogstruktur
- Telefonie-Integration
  - CTI Link Protokoll
  - Telefonieprotokolle (Analog, CAS, E1, T1 etc.)
- Interfaces zu Fremdkomponenten
  - Datenbanken, SMS, Fax etc.

## Requirements Specification

- Dimensionierung des Systems
  - Ausfallsicherheit / Ausfallverhalten
  - Lastverhalten / Zahl der Trunks
- Betriebskonzept
  - Update
  - Überwachbarkeit
  - Wartbarkeit (Fernwartung)
  - Betriebsstatistiken, z.B.
    - Transaktionsstatistiken
    - Erfolgsquote
    - Benutzerverhalten

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Dialogdesign

- Definition der Zielgruppe
  - Alter, Häufigkeit der Benutzung, Sprache etc.
- Definition der „Persona“
  - zielgruppenorientierte Auswahl eines Sprecherprofils (m/w, Alter, Ausdrucksweise etc.)
- Definition der Dialogstruktur
  - Dialogschritte
  - Navigation
  - Fehlerbehandlung

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Dialogdesign

- Definition der Ansagetexte
- Definition möglicher Benutzerantworten
  - Vokabular und Restriktionen (Grammatik)
- Definition der Daten
  - Eindeutigkeit
  - Vollständigkeit
  - Aktualität

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Verifikation des Dialogdesigns

- „Offline“ anhand der Spezifikation
  - Dialogstruktur
  - Formulierung der Ansagen
  - Navigation
  - Benutzerantworten
- „Wizard of Oz“ Simulation
  - Mitarbeiter testen den Dialog
  - System wird durch einen Mitarbeiter simuliert
  - Erste Beurteilung des Dialogs ohne Implementierungsaufwände

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Interner Dialogtest mit Prototypen

- Kleiner, informierter Benutzerkreis (Lieferant und Kunde)
- Aufnahme vorläufiger Prompts / Synthese
- Prototyp-Daten
- Test auf Dialogverlauf
- Test auf Bedienbarkeit / Navigation
- Test auf Fehlerbehandlung
- Qualifiziertes Benutzer-Feedback

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Aufnahme der Prompts

- Produktion der endgültigen Ansagen und Wartemusik
- Professionelle Studioumgebung
- Professionelle Sprecher gem. Persona-Definition
- Bei Systemen mit Mischung aus Prompts und Synthese: „corporate voice“ Ansatz: Ein Sprecher für Prompts sowie Synthesematerial

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Pilotbetrieb

- Pilot I
  - kleiner, öffentlicher Benutzerkreis
  - Sprachdatensammlung
  - Auswertung des Benutzerverhaltens (qualitativ)
  - Feedback von den Benutzern (Fragebogen)
- Pilot II
  - größerer Benutzerkreis
  - Werbung wird geschaltet
  - statistische (quantitative) Auswertungen
  - anschließende Optimierung

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Systemoptimierung

- Beobachtung der Systemperformance über einige Monate
- Adaption des Dialogs
  - typisches Benutzerverhalten
  - typische Benutzerfehler
- Adaption des Erkenners
  - Lexikon
  - Grammatik
  - ggf. Nachtrainieren der akustischen Modelle

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Dialogsysteme: Technik

- Zustandsbasierte Dialogsysteme
- Dialogdesign
- Sprachausgabe
- Spracherkennung
- Auswertung der Erkennungsergebnisse
- Datenaufbereitung

## Dialogsysteme: Technik

- Zustandsgesteuerte Dialogsysteme: „graph-based system“
  - Verhalten des Dialogsystems wird durch einen Zustandsautomaten (Graphen) beschrieben
  - Dialog besteht aus Abfolge von Dialogzuständen
  - gut geeignet für klare Menüstrukturen
  - Erkennung: Einzelwörter oder Phrasen
  - Einsatz komplexer Grammatiken für zB. Geldbeträge, Uhrzeit, Datum

## Dialogsysteme: Technik

- Alternative: „frame-based system“
  - „Ich möchte morgen um neunzehn Uhr von München nach Hamburg fahren“
    - Auffüllen der Felder in einem Formular („template-based system“)
    - User-input kann mehrere Fragen gleichzeitig beantworten
    - System fragt nach unbesetzten Feldern
    - kein vorgegebener Dialogablauf
    - natürlichsprachlicher Dialog („concept spotting“)

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Zustandsbasierte Dialogsysteme

- Zustände des Dialogsystems sind beschrieben durch
  - Ansage (Prompt)
  - Erkennungsvorgang mit bestimmten Restriktionen (Vokabular, Grammatik etc.)
  - ggf. interne Aktion des Systems (z.B. Datenbankabfrage, Prüfung des Konfidenzwertes)

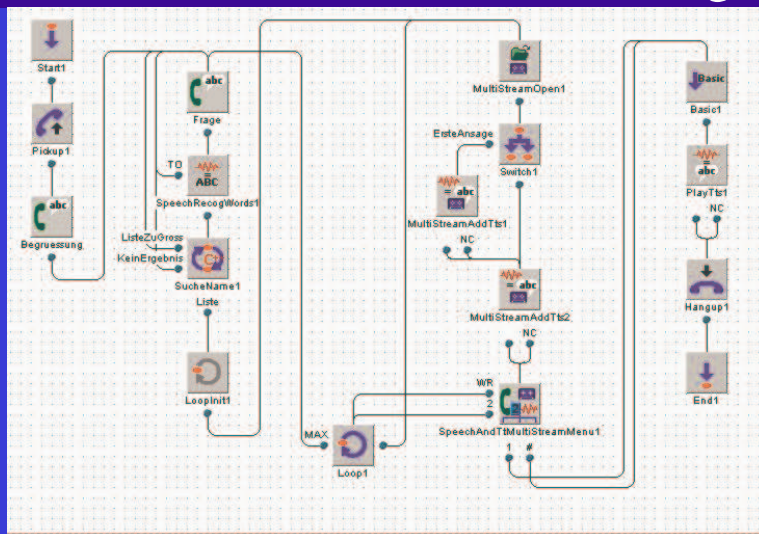
Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

# Zustandsbasierte Dialogsysteme

- Übergänge zwischen den Zuständen
  - abhängig vom Erkennungsergebnis
  - abhängig vom Resultat der internen Aktion
  - Fehlerbehandlung
- Zustand des Dialogs entspricht genau dem aktiven Zustand des Automaten

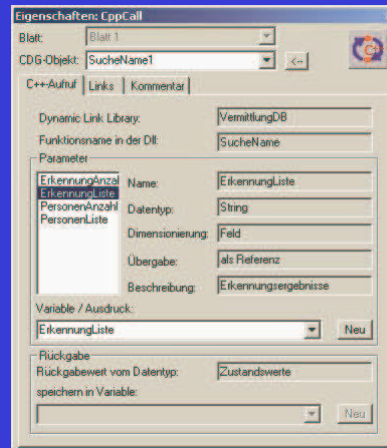
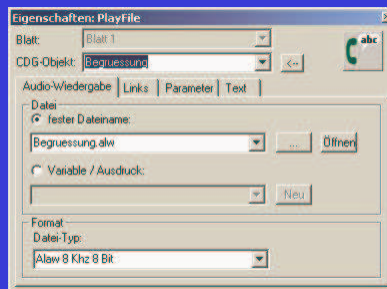
Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Entwurfstool „CDG“ von CreaLog



Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Eigenschaften der Knoten



Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Dialogdesign

- Zielgruppe
- Umgebung
- Dialogstruktur
- Usability
- Fehlerbehandlung
- Operator Fallback
- Unterstützte Sprachen
- Speech Samples vs. Synthese

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Dialogdesign

- Definition der Zielgruppe
  - Häufigkeit der Nutzung
  - regional / national / international
  - monolingual / multilingual / fremdsprachige Sprecher
  - Alter
- Umgebung
  - ruhige Umgebung: Home, Office
  - laute Umgebung: Auto, öffentl. Nahverkehr
  - Mobiltelefon / Festnetz
  - Telefon DTMF-fähig ?

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Dialogstruktur

- Topics
  - Definition des Ablaufs im „Gutfall“
  - Definition der Ansagen
  - Definition der möglichen Benutzerantworten (Vokabular und Grammatik)
  - Verzweigungstiefe
  - Usability
  - Fehlerbehandlung
  - Navigation / Hilfe
  - Dialogdauer

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Dialogstruktur

- Richtlinien zum Design
  - das Gedächtnis des Benutzers nicht überfordern!
  - nicht mehr als 5 Menüpunkte /geringe Verzweigungstiefe
  - den Dialogablauf durch Ansagen transparent machen
  - Feedback geben
    - nach erfolgreicher Eingabe
    - nach Timeout
  - kontextsensitive Hilfe anbieten

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Navigation

- Richtlinien zum Design
  - grundlegende Navigationsbefehle sind immer verfügbar („zurück“, „Hilfe“, „Abbruch“ etc.)
  - kontextabhängige Navigation („weiter“, „letzter bzw. nächster Eintrag“)
  - shortcuts
    - zusätzliche Navigationsbefehle
    - Barge-In
    - DTMF alternativ zur Spracherkennung
  - der Benutzer darf nie in einem Dialogschritt gefangen sein !
  - evtl. Rücksprung zu einem bekannten Startpunkt

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Sprachausgabe

- Definition der Persona
  - zielgruppenorientierte Personenbeschreibung, z.B.: „Anlageberaterin, 28 Jahre, engagiert, sportlich...“
  - Auswahl eines passenden Sprechers aus dem Sprecherpool

## Sprachausgabe

- Begrüßung
  - Dem Benutzer mitteilen, daß er mit einer Maschine spricht
  - Hinweis auf Hilfe
  - Hinweis auf Navigationsmöglichkeiten
  - Hinweis auf Expertenmodus

## Sprachausgabe

- Technische Aspekte
  - Ausgabe aufgenommener Ansagen
    - sehr gute Sprachqualität und Verständlichkeit
    - natürlicher Dialog, zielgruppenorientiert
  - Problem: Datenabhängige Ansagen
    - extrem viele aufgenommene Ansagen
      - hohe Aufwände
      - unflexibel, Problem der langfristigen Sprecherverfügbarkeit
    - Alternative: Synthese
      - unnatürlich, geringere Qualität
      - harte Übergänge im Dialog zwischen aufgenommenen Ansagen und Synthese

Dr. Bernd Plannerer

[www.speech-recognition.de](http://www.speech-recognition.de)

## Sprachausgabe

- Richtlinien zum Design
  - natürlich
  - gut verständlich
  - einfache Formulierungen
  - „Barge-In“ verwenden (erlaubt Expertenmodus)
  - Falls Synthese benutzt wird: Einen Satz zum „Einhören“ vorschalten !
  - Bei kurzen Wartezeiten geeignete Geräusche verwenden („audio hourglass“), keine Pausen !

Dr. Bernd Plannerer

[www.speech-recognition.de](http://www.speech-recognition.de)

## Ansagetexte steuern das Benutzerverhalten

- Richtlinien zum Design
  - immer erst die Aktion, dann den Befehl nennen („Goal --> Action“ Struktur)
  - keine „abwechslungsreichen“ alternativen Formulierungen für die benutzten Begriffe !
  - nicht: „Zum **Sichern** der Daten sagen sie **Speichern**“
  - sondern: „Zum **Speichern** der Daten sagen Sie **Speichern**“
  - die wahrscheinlichste Menüauswahl zuletzt nennen
  - offene Fragen vermeiden („was wünschen Sie?“)

## Ansagetexte steuern das Benutzerverhalten

- Richtlinien zum Design
  - durch die Ansage die Antwortvielfalt einschränken („welche Aktion wollen Sie durchführen ? **Speichern** oder **Löschen** ?“)
  - die Ansage mit einer kurzen Bestätigung beginnen („Gut ! Und jetzt die Kontonummer !“)
  - für lange Ziffernketten DTMF anbieten

## Verifikation von Benutzereingaben

- explizite Verifikation („Sagten Sie München?“)
  - zeitaufwendig, umständlich, geringe Akzeptanz
  - nur für kritische Dialogschritte verwenden
  - unmöglich bei großer Zahl von Alternativen
- implizite Verifikation („Bitte nennen Sie den Betrag, den Sie überweisen wollen !“)
  - kurzer Dialogverlauf
  - bei hoher Konfidenz anwenden
  - Klärungsdialog ist ggf. komplex

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Fehlerbehandlung

- Richtlinien zum Design
  - Fehler werden vom System verursacht, nicht vom Benutzer!  
(„Entschuldigen Sie bitte, ich habe Sie nicht verstanden“)
  - Auswertung von Konfidenzwerten verhindert unsinnige Verifikationsfragen
  - bei erneuter Rückfrage die Ansage wechseln
  - nicht mehr als zwei Rückfragen !
  - Navigationshilfe anbieten
  - Eventuell Operator-Fallback

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Hilfetexte

- Richtlinien zum Design
  - kontextabhängige Hilfe
  - unterbrechbar (barge-in)
  - Beispiele geben
  - positive Formulierungen wählen, nicht den Benutzer beschuldigen

## Literatur

**Michael F. Mc Tear**

"Spoken Dialog Technology: Enabling The Conversational User Interface"

ACM Computing Surveys, Vol. 34, No.1, March 2002, pp. 90-169

**Beasley, Farley, O'Reilly, Squire**

"Voice Application Development with VoiceXML"

SAMS Publishing, Aug. 2001, ISBN 0672321386

**Susan Weinschenk, Dean T. Barker**

"Designing Effective Speech Interfaces"

Wiley, April 2000, ISBN 0-471-37545-4

**Michael H. Cohen, James P. Giangola, Jennifer Balogh**

"Voice User Interface Design"

Addison-Wesley Pub Co, February 2004, ISBN 0321185765

## Probleme beim Einsatz von Spracherkennung

- „out of vocabulary“ (OOV)
- Hesitationen
- „Barge In“
- Erkennen-Unsicherheit aufgrund der Vokabulargröße

## Probleme beim Einsatz von Spracherkennung

- OOV
  - Erkennen-Unsicherheit ist bei kleinen Lexika mittels Konfidenzmaß erkennbar
  - Abhilfe: Rückfrage mit neuer Formulierung
- Hesitationen
  - Abhilfe durch Wortketten / Satzerkennung und Modelle für Hesitationen



## Probleme

- „Barge In“
  - Benutzer spricht vor Ende des Prompts
    - Erkennung noch nicht gestartet
    - Äußerung ist unvollständig und wird falsch erkannt
  - Abhilfe:
    - Erkennung läuft bereits während des Prompts
    - Echo cancellation ist notwendig
  - Sofortiger Abbruch des Prompts bei Störgeräuschen, Räuspern etc ist unerwünscht.
  - Syntaxgesteuertes Barge In
    - erst bei Erkennung sinnvoller Äußerungen wird der Prompt unterbrochen

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

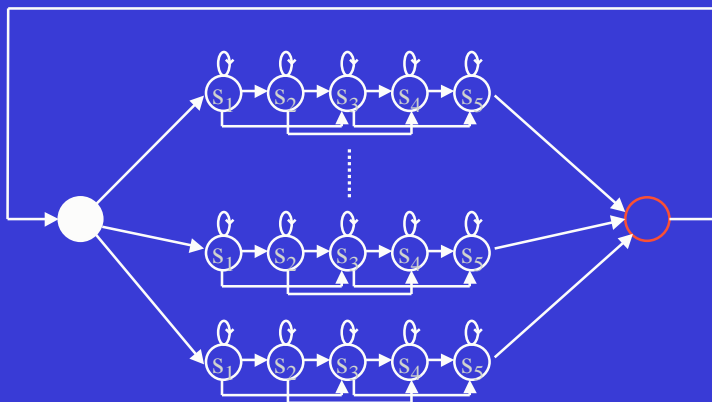
## Erkennung-Unsicherheit

- Sprechvariabilität
  - Alter
  - Geschlecht
  - Vokaltraktlänge
  - Sprechgeschwindigkeit
  - Dialekt
- Modelliert in der Trainingsphase der HMMs
- Diskriminanz zwischen Wortmodellen sinkt mit der Vokabulargröße

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

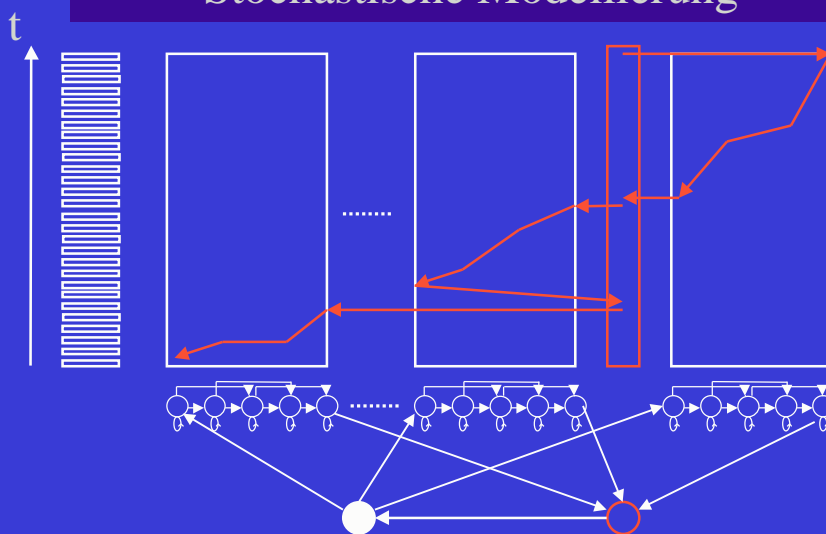
# Stochastische Modellierung

## Macro-HMM für Wortketten



Dr. Bernd Planerer [www.speech-recognition.de](http://www.speech-recognition.de)

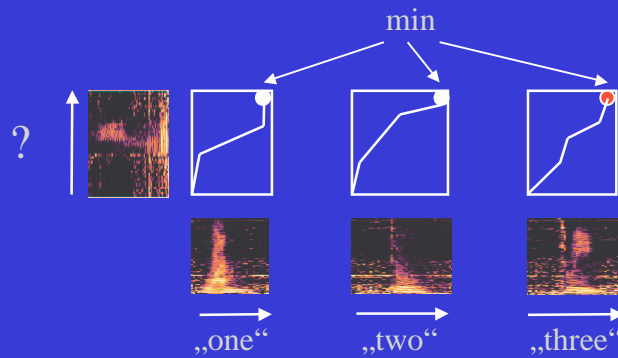
# Stochastische Modellierung



Dr. Bernd Planerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Pattern Matching und Dynamische Programmierung

$\tilde{X} \in \omega$ , wenn  $D(\tilde{X}, M_\omega) < D(\tilde{X}, M_j) \forall j=1..V, j \neq \omega$



Dr. Bernd Planerer [www.speech-recognition.de](http://www.speech-recognition.de)

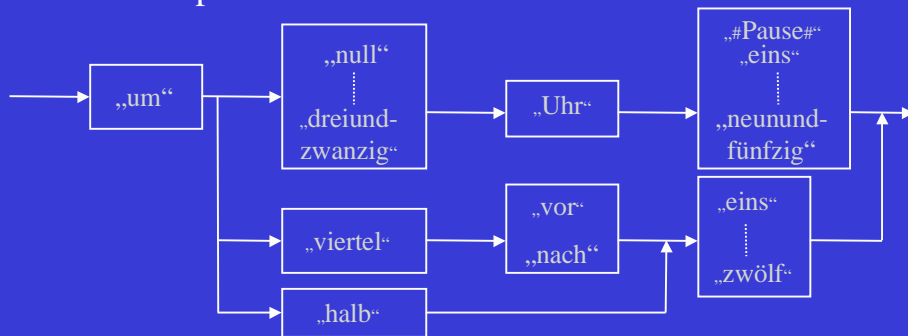
## Verbesserung der Erkennungsleistung

- Integration zusätzlicher Restriktionen
  - Finite State Grammatik
    - Beschreibung von Phrasen durch Grammatik
    - als „Fertigbausteine“ erhältlich
    - semantische Analyse ist einfach
  - Language Model
    - statistischer Ansatz
    - automatisch trainierbar

Dr. Bernd Planerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Restriktionen durch Grammatik

- Finite State Grammatik
  - Beispiel: Uhrzeit



Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Verbesserung der Erkennungsleistung

- Grundregel: Zu jeder Zeit ein möglichst kleines Vokabular !
- Lexikon und Grammatik werden abhängig vom Dialogzustand festgelegt
  - Aktives Vokabular wird minimiert
- Lexikon und Grammatik werden dynamisch geladen
  - Alle Firmennamen für einen bestimmten Ort
  - Alle Abteilungen einer Firma

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Auswertung der Erkennungsergebnisse

- Top-1 Ergebnisse sind selten korrekt
- Auswertung mehrerer Hypothesen
  - N-Best
  - Wortgraph
- Verifikationsschritt im Dialog
- Anwendung zusätzlicher Wissensquellen
  - z.B. Plausibilitätsprüfungen (Checksumme)
  - Disambiguierung über Datenbank

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Mehrdeutigkeiten

- Kontonummer: 700 500 01
- N-Best Erkennung
- „siebenhundert fünfhundert null eins“  
=> 700 500 0 1
- „sieben hundert fünf hundert null eins“  
=> 7 100 5 100 0 1
- „siebenhundertfünf hundert null eins“  
=> 705 100 01

Dr. Bernd Plannerer [www.speech-recognition.de](http://www.speech-recognition.de)

## Disambiguierung über Datenbank

- Beispiel Telefonauskunft
  - Datensätze mit Namen, Vornamen, Straße, Telefonnummer
  - Gesucht ist ein bestimmter Datensatz
  - Suchargumente
    - N-Best Liste der Ortsnamen
    - N-Best Liste der Namen
    - N-Best Liste der Vornamen
    - N-Best Liste der Straßennamen

## Verifikation im Dialog

- Explizite Verifikation
  - umständlich, zeitaufwendig
  - geringe Akzeptanz
  - unmöglich für große Zahl von Hypothesen
- Implizite Verifikation
  - benutzerfreundlich, bessere Akzeptanz
  - Klärungsdialog ggf. kompliziert
  - unmöglich für große Zahl von Hypothesen

## Disambiguierung über Datenbank

Alle Hypothesen werden zur Suche  
in der Datenbank eingesetzt

Manhattan NY or Manhasset NY or Madison NY ...  
and  
Milla or Miller or Murilla ...  
and  
Frank or Frankie or Fred ...

## Auswertung der Erkennungsergebnisse

- Datenbankanfrage  
Nur bestimmte Kombinationen existieren:
  1. Frank Miller, Manhasset NY
  2. Frank Milla, Manhattan NY
- Disambiguierungs-Dialog wird aus dem  
Anfrageergebnis dynamisch generiert

## Definition der Lexika / Grammatiken

- Für statische Lexika
  - automatische phonetische Transkription
  - ggf. manuelle Korrektur
  - Erkennungsperformance hängt stark von der Lexikongröße ab

## Datenaufbereitung

- Für dynamische Lexika
  - Lexikon / Grammatik des Erkenners wird dynamisch aus dem Ergebnis einer Datenbankabfrage erstellt und in den Erkennen geladen
  - Qualität des Datenbestandes ist essentiell für gute Performance, da das Lexikon aus den Daten erzeugt wird !

Weitere Informationen  
[www.speech-recognition.de](http://www.speech-recognition.de)

