

Sprachdialogsystem mit robuster automatischer Spracherkennung

Hans-Günter Hirsch

**Hochschule Niederrhein
Fachbereich Elektrotechnik und Informatik
Digitale Nachrichtentechnik**

<http://dnt.kr.hs-niederrhein.de>

- Sprachdialogsystem
- Verfahren zur robusten Spracherkennung
- Anwendungsbeispiele



Ziel

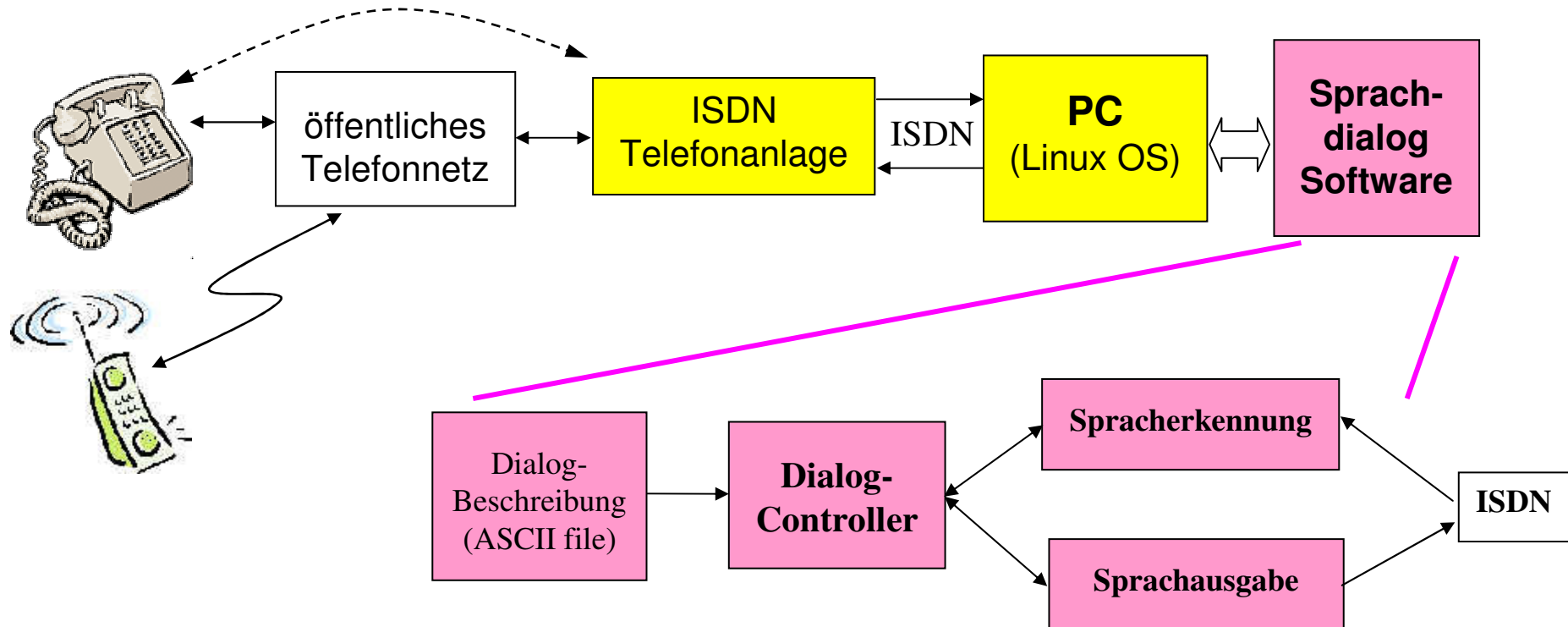
Ausgangspunkt

- Langjährige Untersuchungen zur robusten Spracherkennung
 ➔ Verfahren mit einer Adaption der spektralen Parameter

Ziel

- Realisierung eines Sprachdialogsystems zur
 - Demonstration der Robustheit des Erkennungsverfahrens
 - Ermittlung der in praktischen Anwendungen auftretenden Störeinflüsse

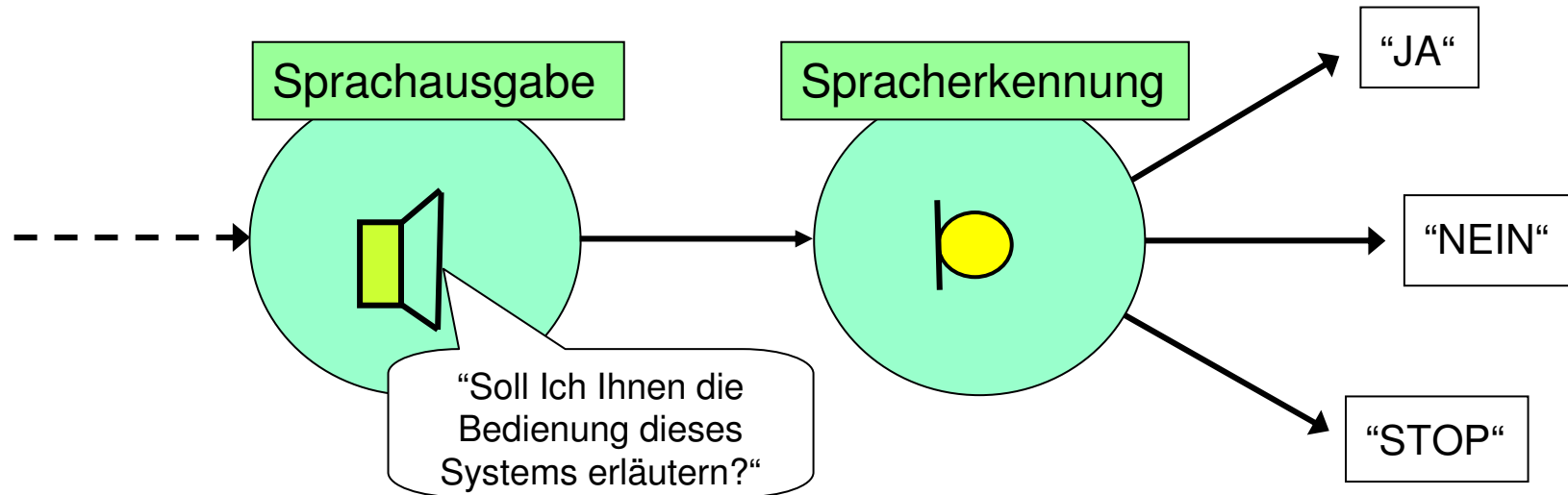
Sprachdialogsystem



Dialog-Software

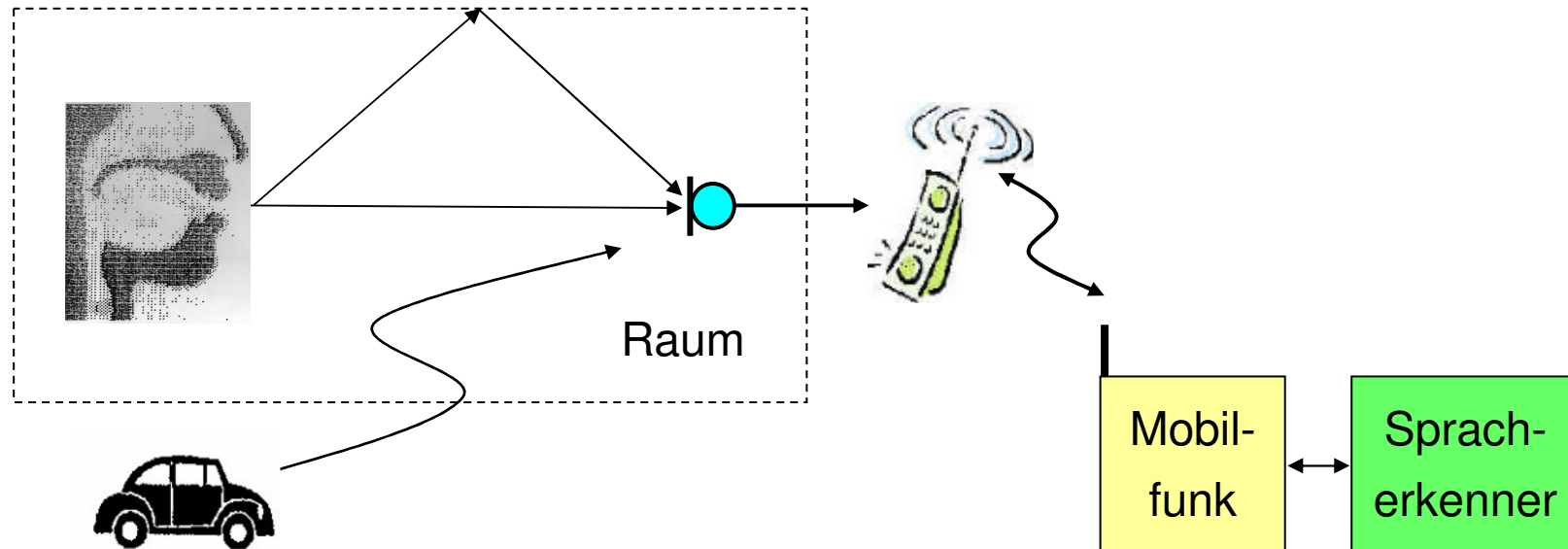
- Beschreibung des Dialogs in Textform
- Schnelle Realisierung und Modifikation einfacher Dialoge möglich

Dialog- Darstellung



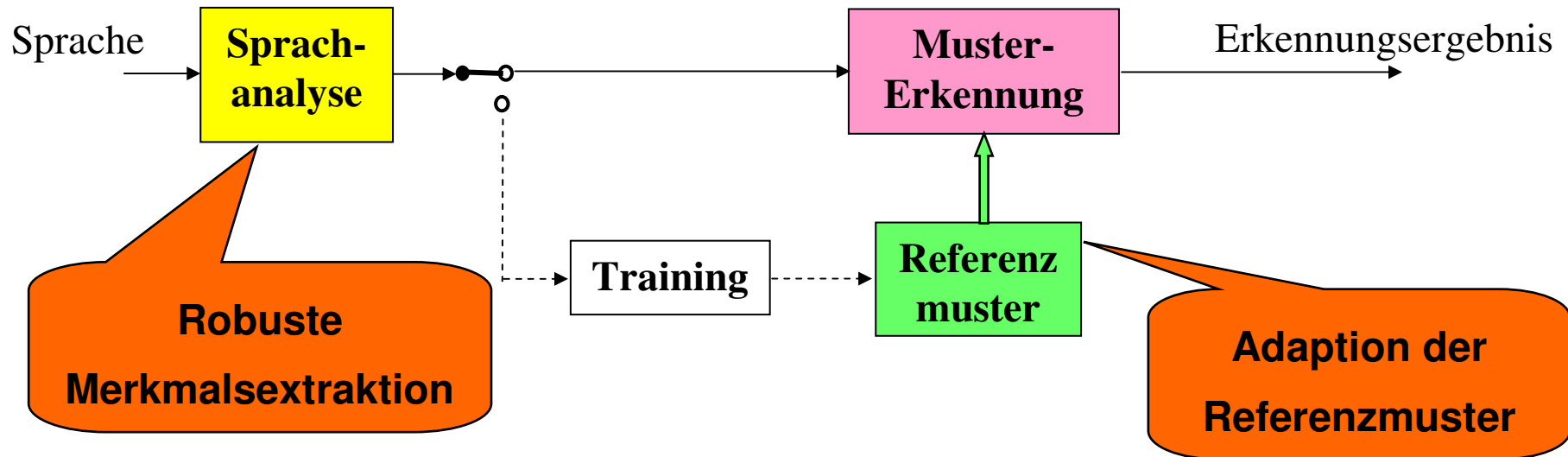
- Darstellung des Dialogs als Verkettung von Netzwerkknoten
- Definition eines einzelnen Knotens:
 - Bestimmte Aktion (Aufruf einer Funktion)
 - Aktionsparameter, z.B. <Filename> für Sprachausgabe
 - Festlegung der Übergänge in weitere Knoten in Abhängigkeit der Rückgabeparameter der aufgerufenen Funktion

Störeinflüsse



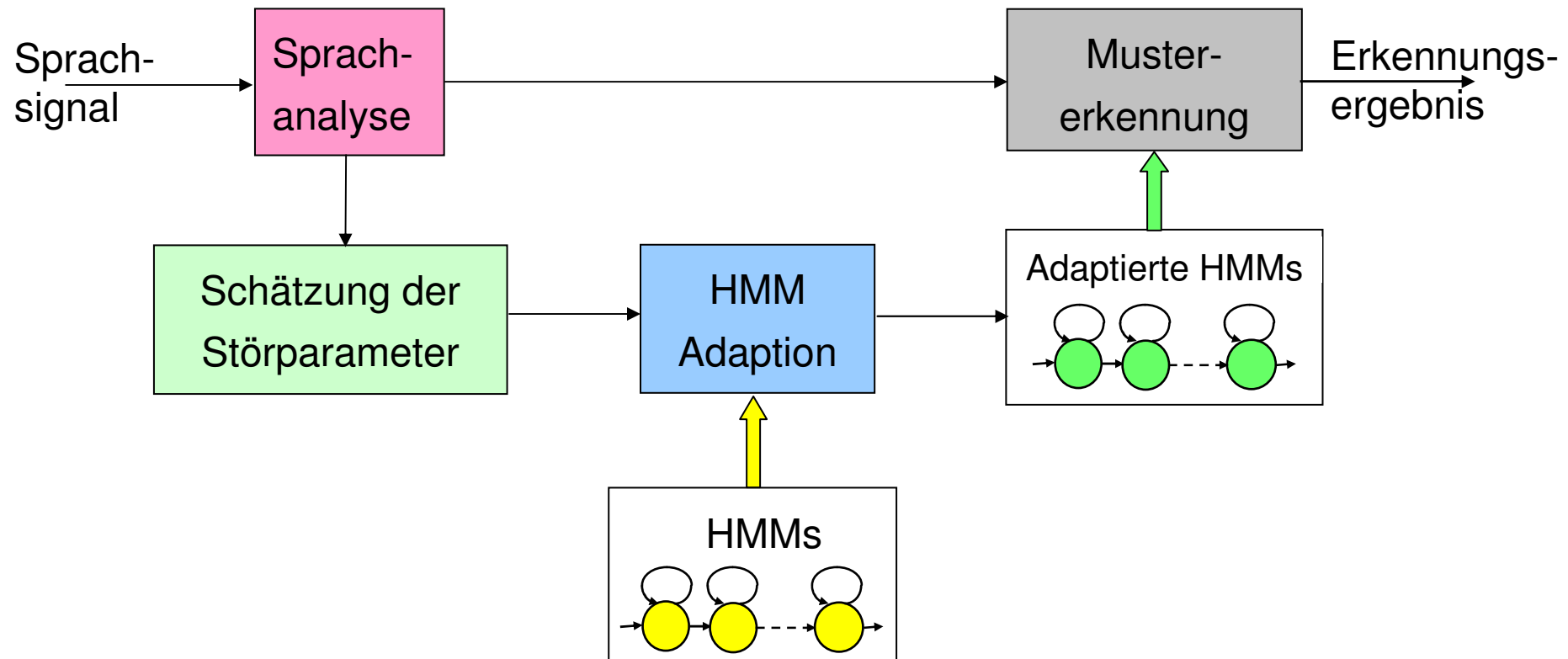
- Additive Hintergrundstörungen
- Unbekannte Frequenzgänge (z.B. Mikrofon, Telefonkanal)
- Hallige räumliche Umgebung
- Mobilfunkkanal (Sprachcodierung, Funkkanalstörungen)

Ansätze zur robusten Erkennung



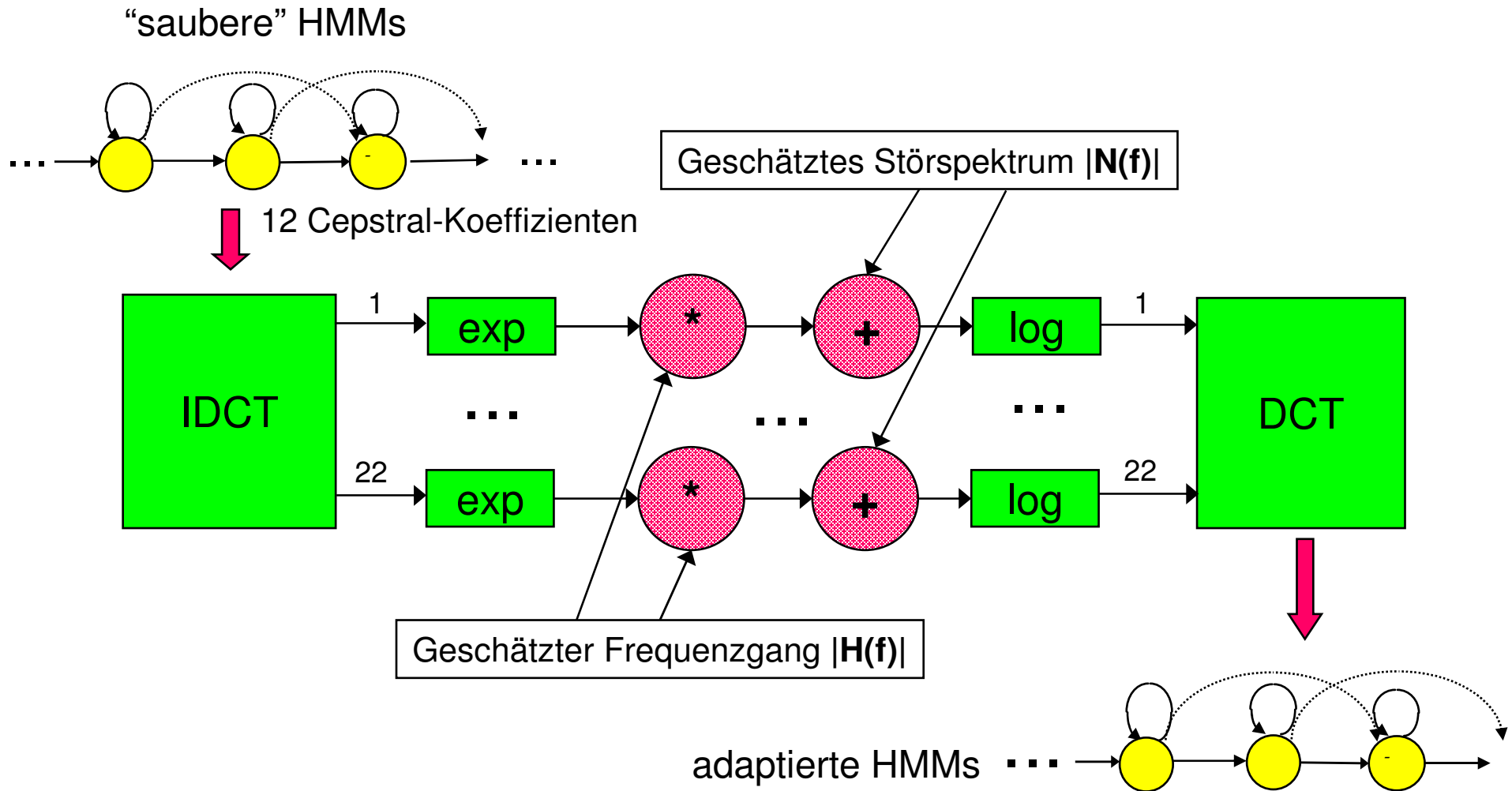
1. Extraktion robuster akustischer Merkmale bei der Sprachanalyse
2. Adaption der Referenzmuster auf die aktuelle Störumgebung

Adaption der Referenzmuster



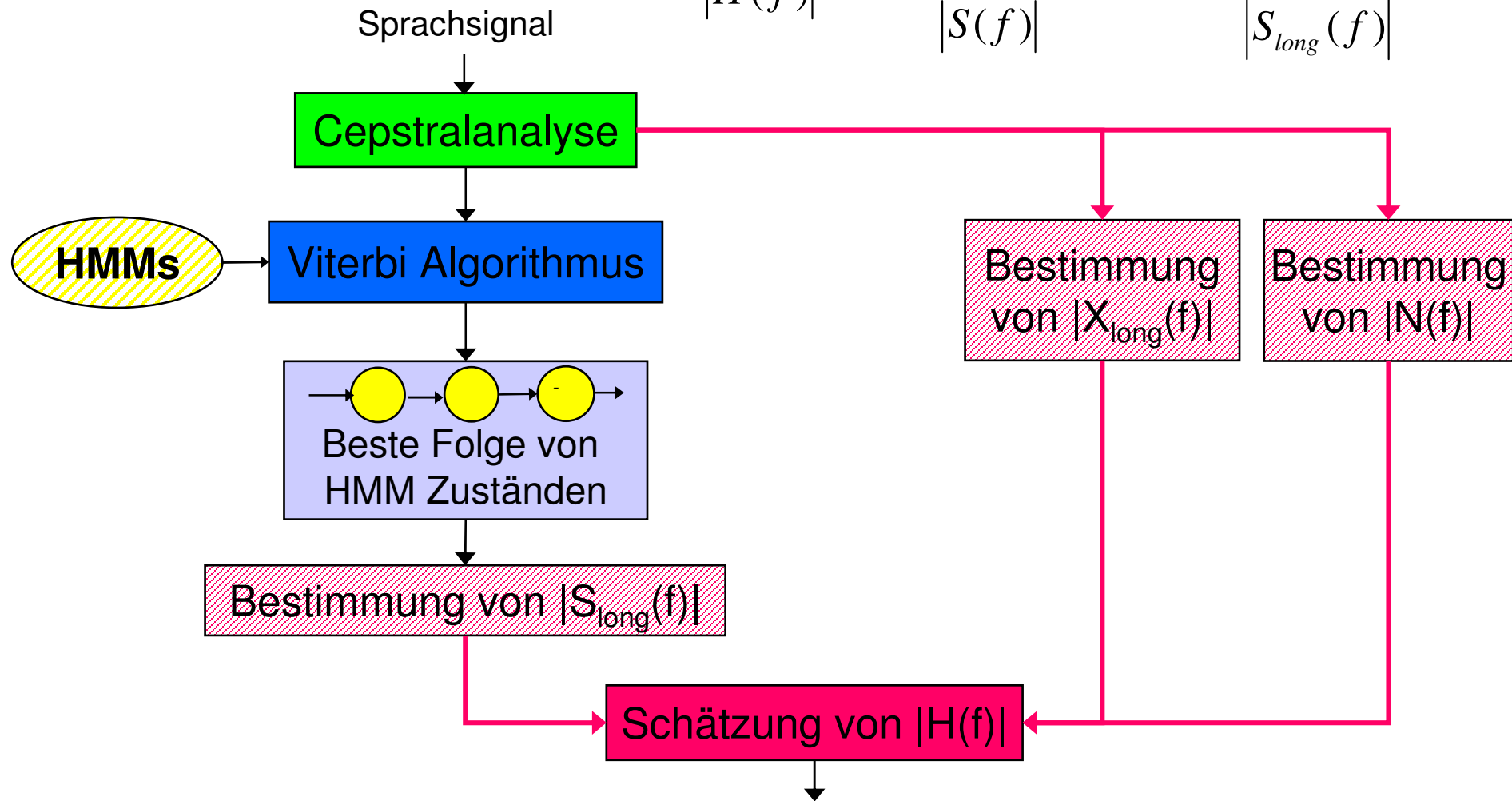
- Sprachanalyse: "Cepstral"analyse zur Bestimmung der spektralen Merkmale
- Referenzmuster: statistische Modellierung mit Hidden Markov Modellen (HMM)
- Bestimmung der HMMs mit dem Programmpaket HTK (Hidden Markov Model Toolkit)

Adaptionsschema (PMC, Young&Gales, Cambridge University)



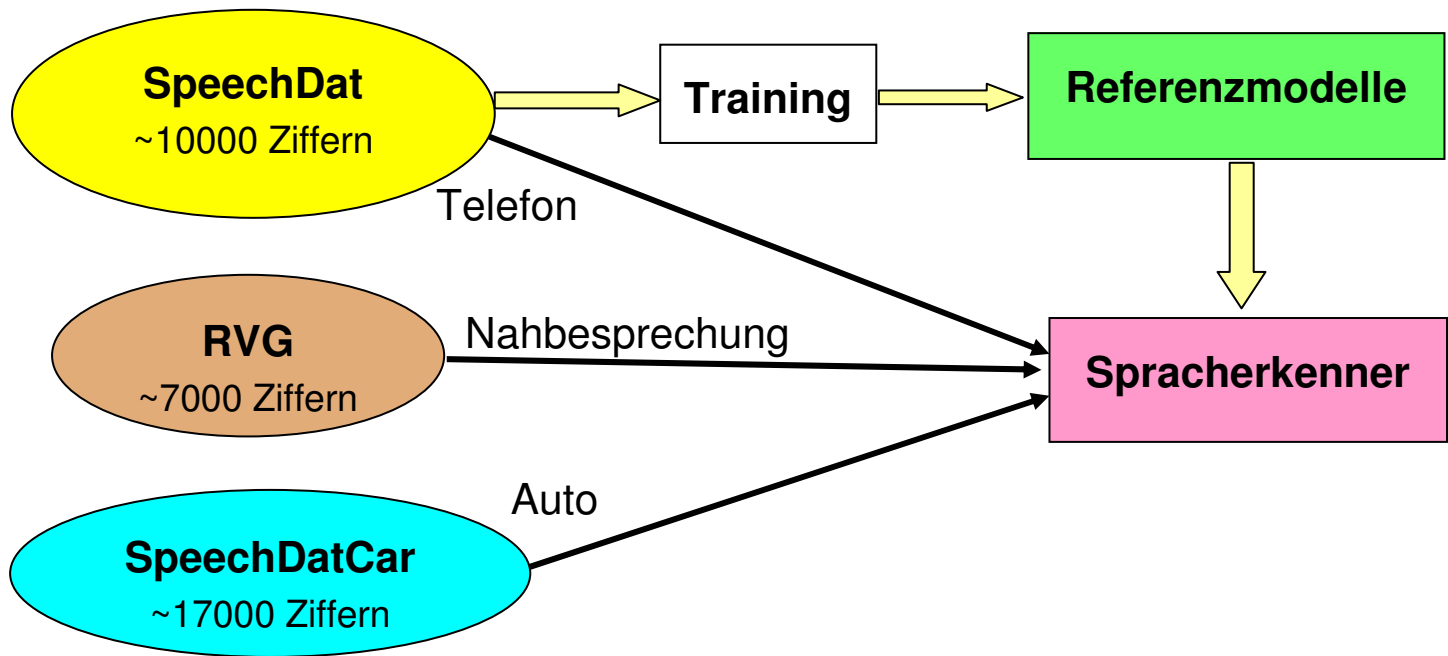
Schätzung der Übertragungsfunktion

$$|H(f)| = \frac{|X(f)| - |N(f)|}{|S(f)|} = \frac{|X_{long}(f)| - |N(f)|}{|S_{long}(f)|}$$



Erkennungsexperimente

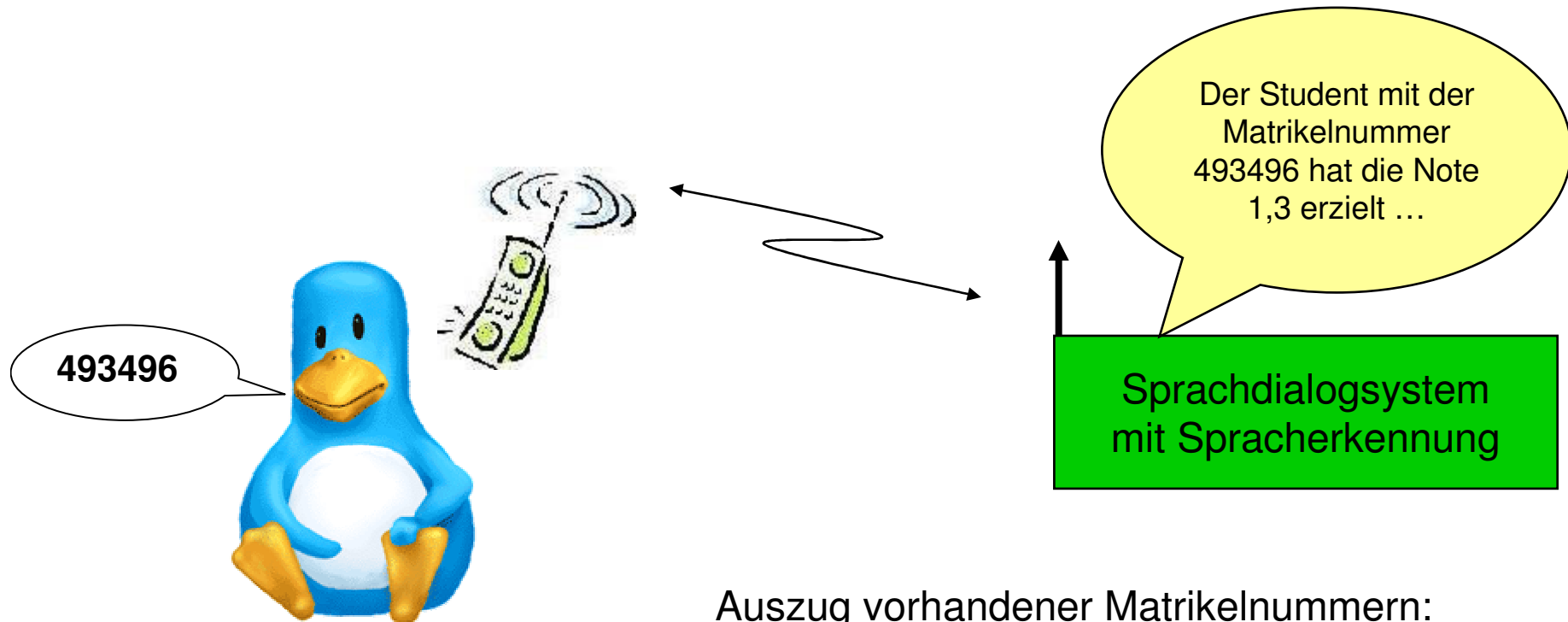
- Aufgabe: Erkennung deutscher Ziffern(ketten)



Worterkennungsraten

	SpeechDat	RVG	SpeechDatCar
39 Koeffizienten/Vektor → Standardisierte Merkmalsextraktion (ETSI-2)	96.11%	93.47 %	90.02 %
24 Koeffizienten/Vektor → Eigene Adaptionstechnik	95.88 %	93.88 %	87.21 %

Automatische Notenauskunft



Erkennung von
Ziffernketten

Auszug vorhandener Matrikelnummern:
356738 402828 448333 514670 514578
449698 470353 470239 515344 529704
470079 478062 507379 514647 537906
529704 ...

Tel.-Nr. 02151 643894

NRW- Verkehrsinfo



Ziffernketten, aber auch:
"A sieben-und-fünfzig"
und
DTMF-Erkennung

Tel.-Nr. 02151 643893

Resümee

Erfahrungen

- Spracherkennung funktioniert gut bei Vorhandensein stationärer Hintergrundstörungen
- Probleme im Fall instationärer Störungen (Sprache im Hintergrund)

Ausblick

- Bisher nur wortbasierte HMMs
- Inzwischen auch Ergänzung um phonembasierte Modelle möglich