PS Einführung in die Computerlinguistik

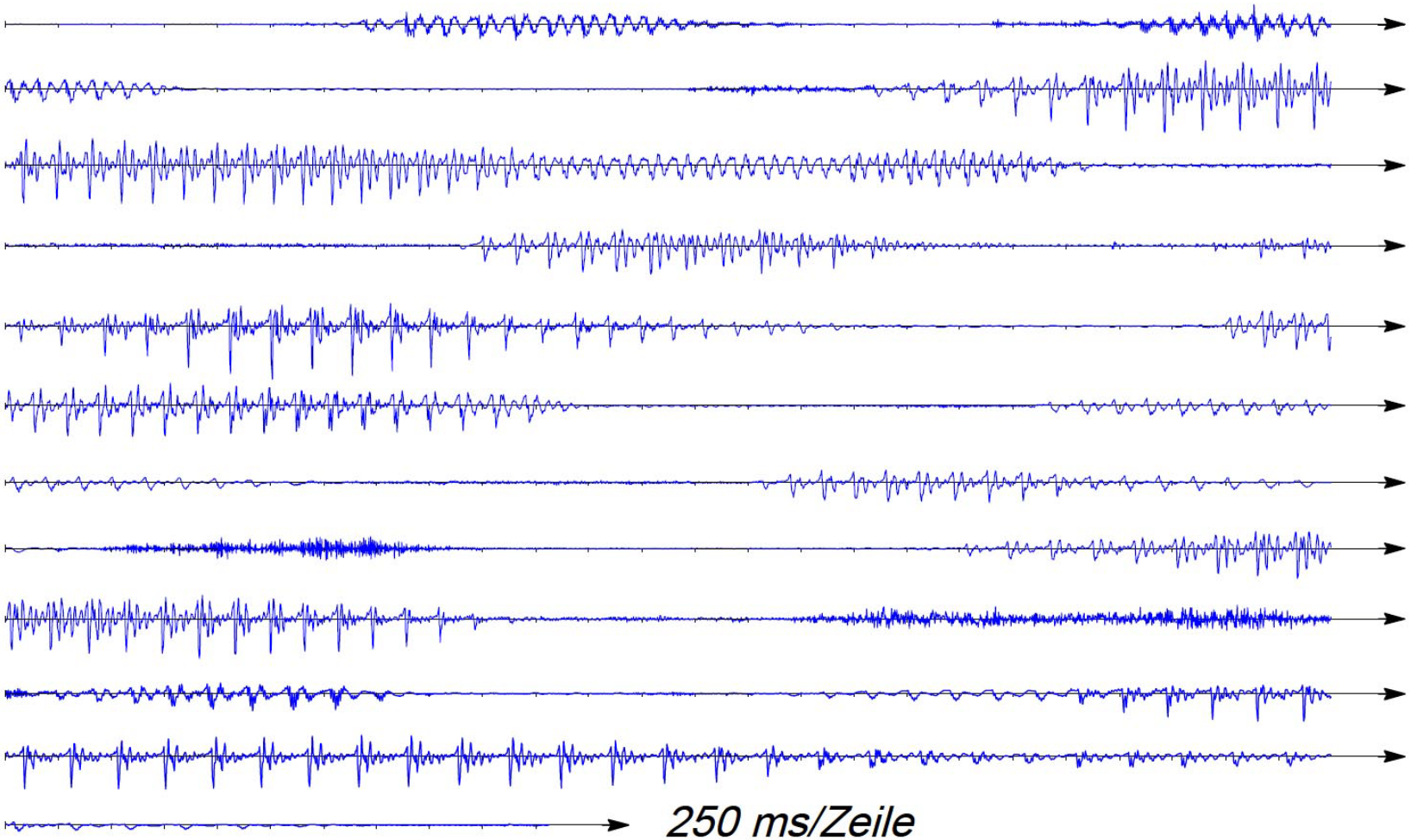
bzw.

SE aus Artificial Intelligence

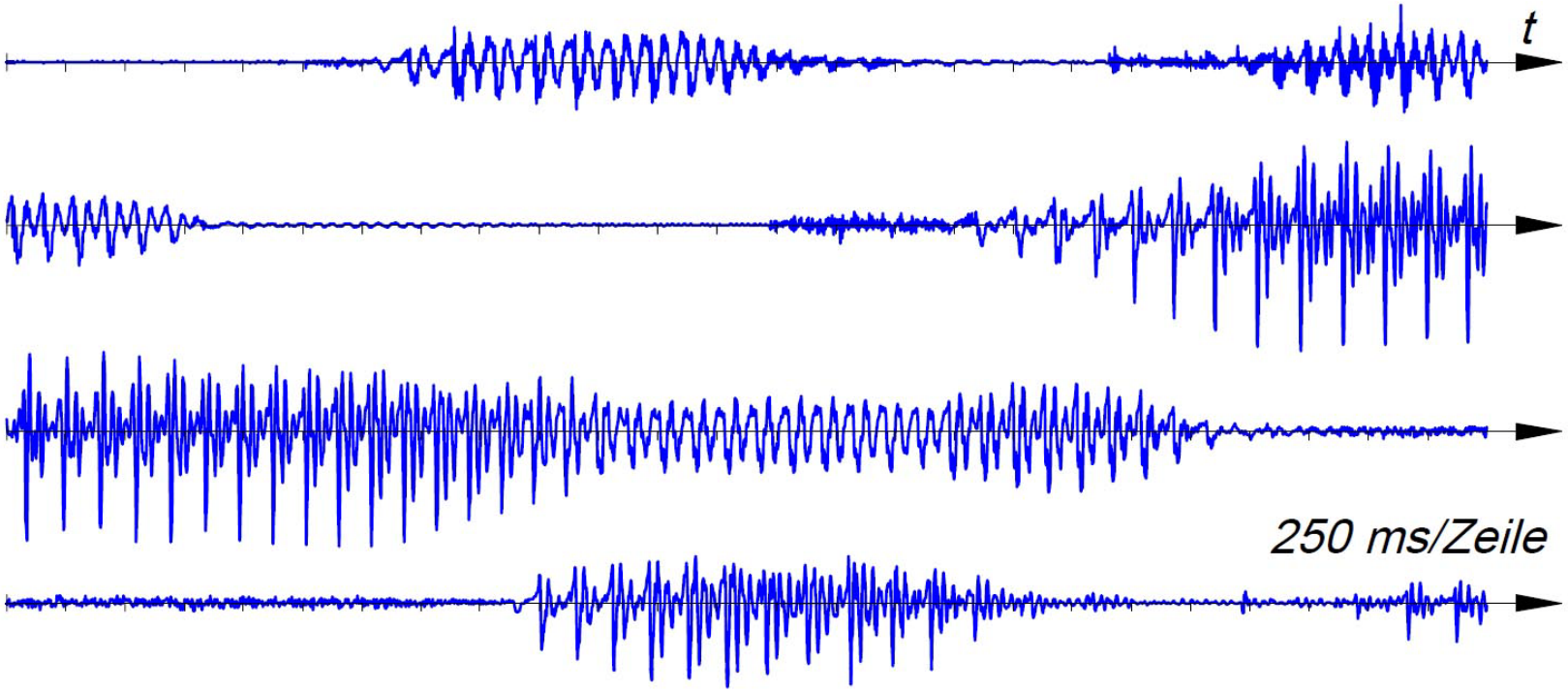
Digitale Verarbeitung von natürlicher Sprache

MMag. Gudrun Kellner

Ein Sprachsignal

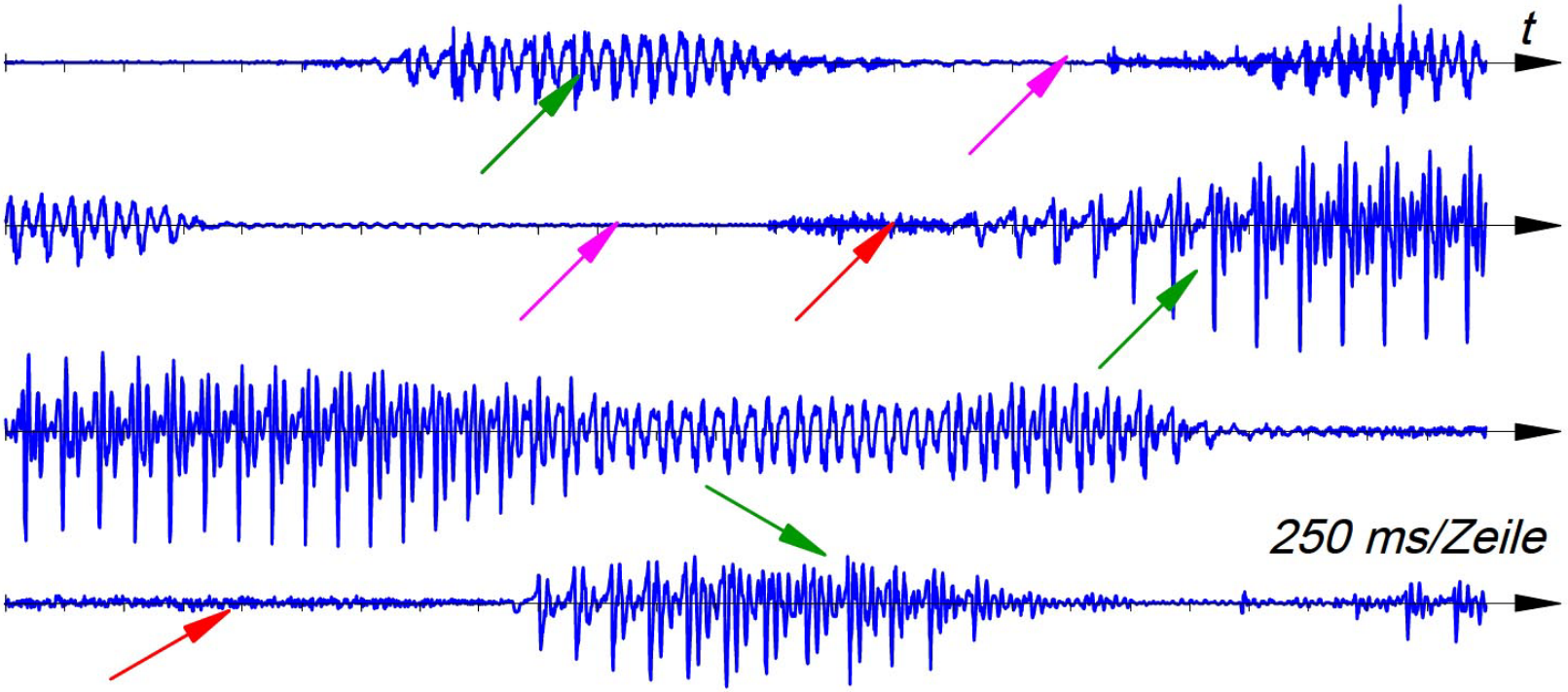


Beobachtungen am Sprachsignal 1



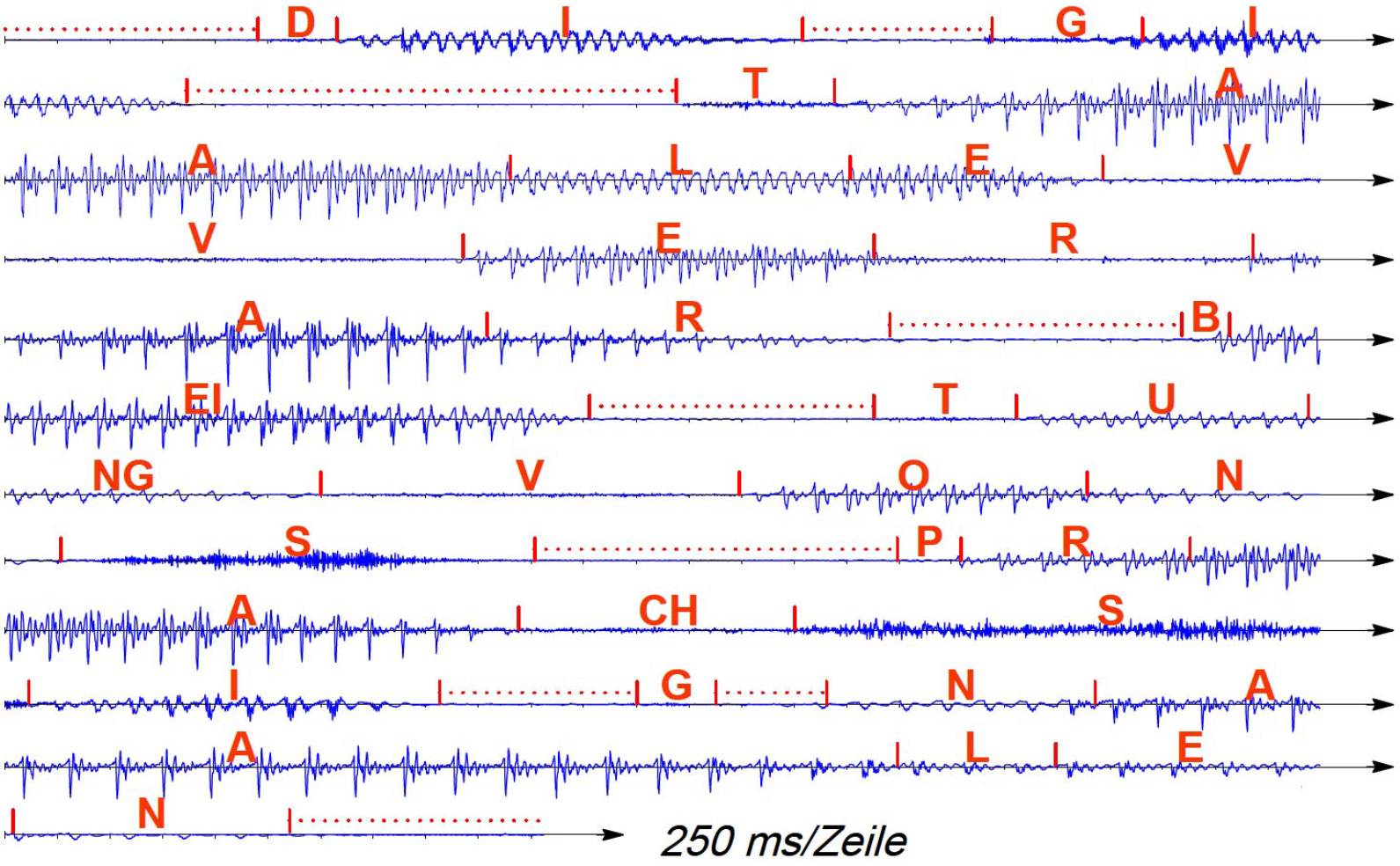
- Sprachsignale sind zeitveränderlich
- Sprachsignale stellen ein Kontinuum (Datenstrom) dar → Änderungen erfolgen fließend

Beobachtungen am Sprachsignal 2



- Sprachsignale sind in manchen Abschnitten periodisch, dazwischen können ein Rauschen oder auch Pausen auftreten

... des Rätsels Lösung

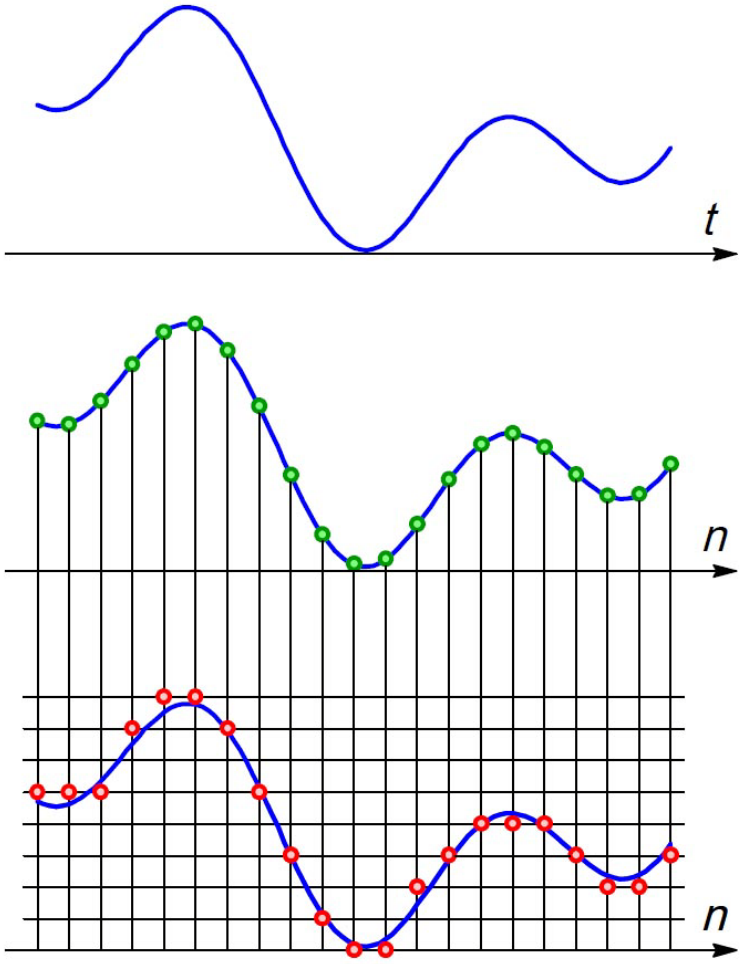


Signale

- Signal: Erscheinungsbild einer physikalischen Information
- Sprachsignale sind eindimensional
- Darstellung als Zeitfunktion:
  s(t) kann als Momentanwert, Augenblicksamplitude oder Wert des Signals zum Zeitpunkt t bezeichnet werden

- Digitalisierung: durch Abtastung und Quantifizierung wird das Signal in eine (am Computer speicherbare) Zahlenfolge umgewandelt
- Grundregel zur Qualitätssicherung: Die Abtastfrequenz muss mehr als das Doppelte der höchsten im Signal enthaltenen Frequenz betragen. (sonst: Verzerrungen)

Digitalisierung von Signalen



- kontinuierlich („analog“)
- abgetastet
- digital (abgetastet und quantisiert)

Einheiten von Sprache

- Satz – Satzteil – Wort – Silbe – Halbsilbe – Phonem
- Inventar an jew. Einheiten für die dt. Sprache:

Einheit	Inventar (dt.)
Satz	unbegrenzt
Wort	~ 500.000
Silbe	5000
Halbsilbe	2000
Phonem	< 50

- Ein Phonem kann in unterschiedlichen Graphemen ausgedrückt werden: /t/ = t (Tag) und dt (Stadt)

Variabilität von Sprache

- Jeder Mensch hat eine individuelle Sprechweise (abhängig von anatomischen Voraussetzungen und erlernter Technik).
- Sprachliche Äußerungen sind nicht exakt reproduzierbar (abhängig von z.B. Stimmung, Müdigkeit, Gesundheitszustand, Muskelspannung etc.)

- Aussprachevarianten zur Sprecherleichterung
  z.B. gibt es 10 Varianten für „sieben“ (nach Häufigkeit):
  si:bən, zibən, si:bn, si:bm, si:bɛn, zi:bm, zi:bn, zi:bɛn, zi:vən, zi:vn

- Koartikulation: der Verlauf eines Lautes wird an seinen aktuellen Kontext angepasst

Schwierigkeitsstufen

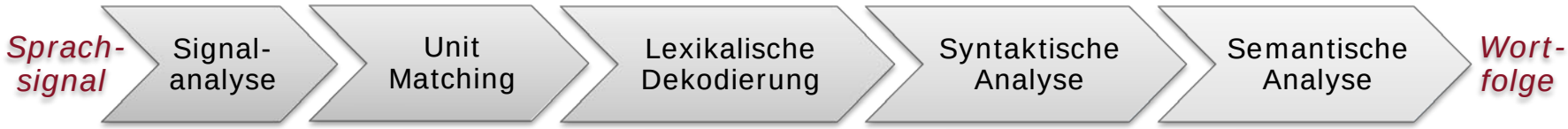
- abhängig von Sprechweise, Wortschatz und Benutzerkreis:

	einfach	schwierig
Sprechweise	einzelne Wörter	Sätze
Wortschatz	klein	groß
Benutzerkreis	sprecherabhängig	sprecherunabhängig

- akustische Umgebung:
  - Aufnahmequalität (Mikrophon vs. Handy-Verbindung)
  - Störgeräusche (Lärmumgebung, Verbindungsverluste)

- Anwendungen: vom „Command and Control“ (einzelne Wörter) bis zu Dialog- und Diktiersystemen (kontinuierliche Eingabe)

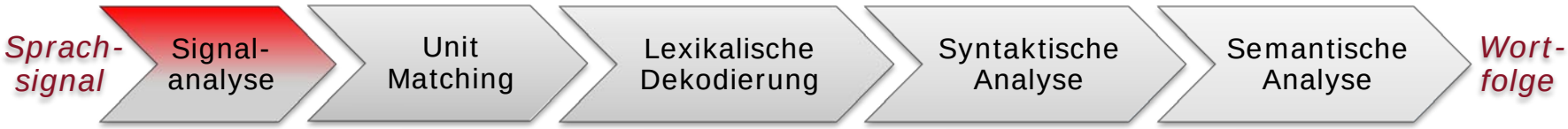
Spracherkennung



Drei große Einsatzfelder:

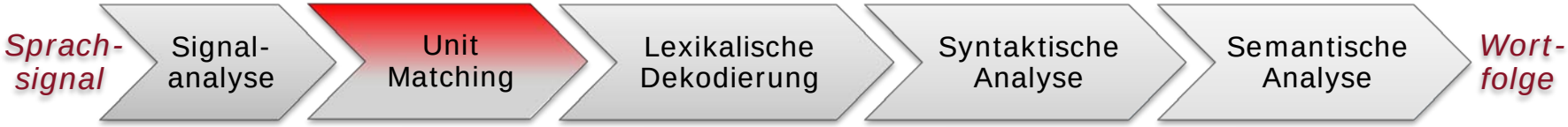
- Gerätesteuerung („Hands free“-Anwendungen):
  Vorteil: intuitive und komfortable Bedienung
  z.B.: Bedienung von Infotainmentkomponenten, Lagerverwaltungssysteme, Steuerung von Operationsmikroskopen
- Diktiersysteme:
  beste Ergebnisse bei Anpassung an den Sprecher
  auch mit Spezialvokabular möglich (Ärzte, Juristen etc.)
- Sprachdialogsysteme („Voice Portals“):
  z.B. Auskunfts-, Bestellungssysteme, autom. Telefonvermittlung

Signalanalyse



- Sprachdetektion: Erkennen von Beginn und Ende einer Spracheingabe (üblicherweise nach der Lautstärke)
- Wortgrenzenbestimmung: Schwierigkeiten durch stimmlose An- oder Auslaute mit geringer Energie, externe (zusätzliche) Lautquellen, Pausen innerhalb einer Äußerung, starke Hintergrundgeräusche
- Merkmalsextraktion: Blockbildung (15-20 ms/Block, Überlappung der Blöcke), Fourier-Transformation, Anpassung des Frequenzbereichs, Vektor-Quantisierung (Reduktion der Vektoren auf bestimmtes Musterset)

Unit Matching

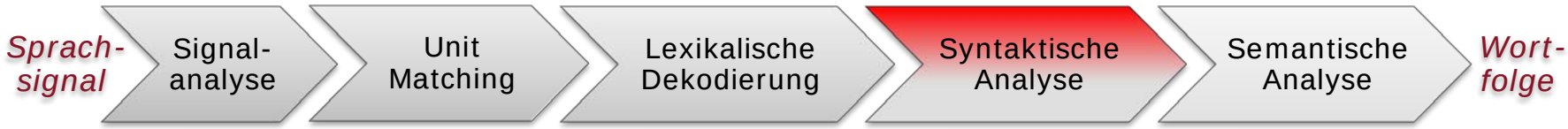


- Aufgabe: Erkennen von akustischen „Minimalblöcken“
- Vorgehen:
  - unbekannte Äußerungen werden mit allen bekannten Sprachmustern verglichen und die Unähnlichkeit berechnet (genau gleich: Wert 0)
  - Flexibilität bei Eigenschaften wie Längengestaltung, Aussprachevarianten, Koartikulation etc.
  - bei sehr nahen ersten Möglichkeiten wird entweder mit Wahrscheinlichkeit gearbeitet oder nachgefragt
- Verbesserung bei Einsatz von wahrscheinlichkeitsbasierten Methoden

Lexikalische Dekodierung

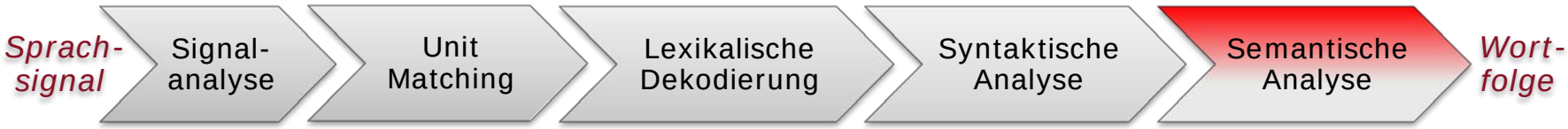
- Ziel: Zuordnung von Lautfolgen zu akustischen Wörterbucheinträgen (basierend auf den Units)
- kann auf unterschiedlichen Einheitsgrößen basieren:
  - ganze Wörter: nur geeignet bei beschränktem Wortschatz
  - Halbsilben oder Phoneme: Einsatz von Aussprachewörter-büchern (inkl. Aussprachevarianten)
- Herausforderungen: Eigennamen, Abkürzungen (USA, NATO), Fremdwörter/Wörter aus anderen Sprachen
- Verbesserung durch statistische Sprachmodelle

Syntaktische Analyse



- Ziel: Interpretationsmöglichkeiten reduzieren auf syntaktisch korrekte Aussagen
- Einsatz von unterschiedlichen Grammatikmodellen:
  - einseitig lineare Grammatiken (Chomsky-Hierarchie)
  - kontextfreie Grammatiken
  - Merkmalskategorien (z.B. Wortkategorie, Numerus, Genus)
- mögliche Wortkombinationen werden auf syntaktische Korrektheit geprüft

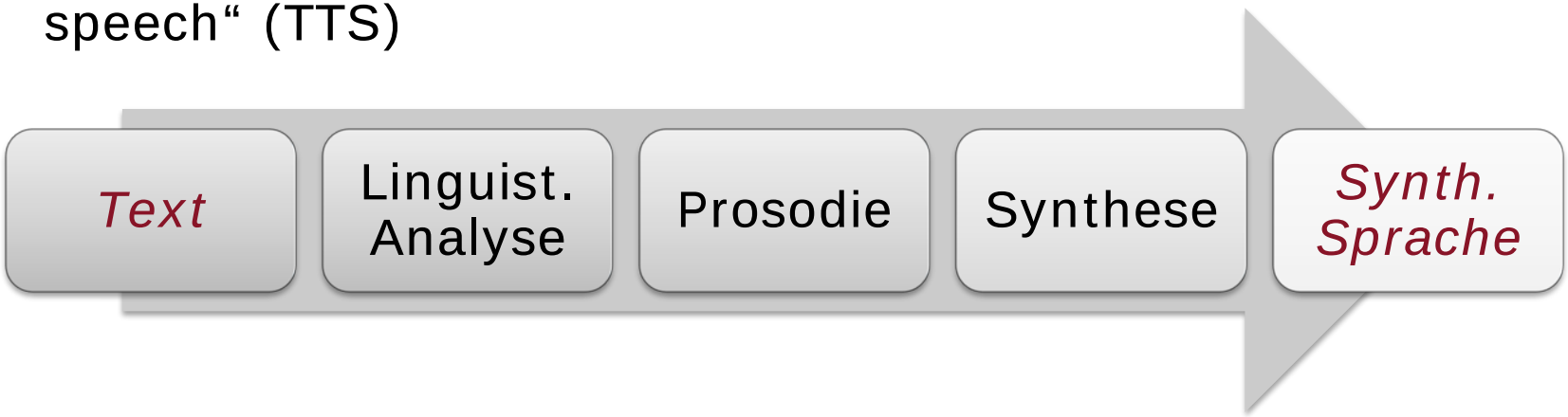
Semantische Analyse



- Ziel: Interpretationsmöglichkeiten reduzieren auf semantisch korrekte Aussagen
- Einsatz von unterschiedlichen Analysemethoden:
  - Merkmalsemantik (z.B. +/-belebt, +/-menschlich etc.)
  - Verbvalenzbestimmung (zum Erkennen von Zusammenhängen im Satz)
  - latente semantische Analyse (statistisches Verfahren)
- mögliche Wortkombinationen werden auf semantische Korrektheit geprüft

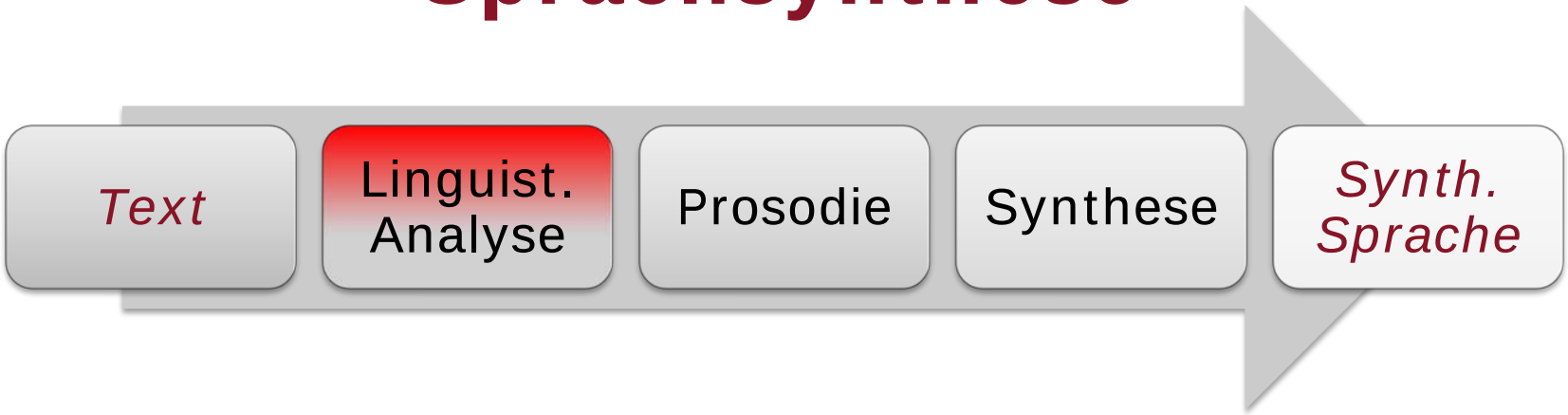
Sprachsynthese

- Automatische akustische Ausgabe von Texten: „text to speech“ (TTS)



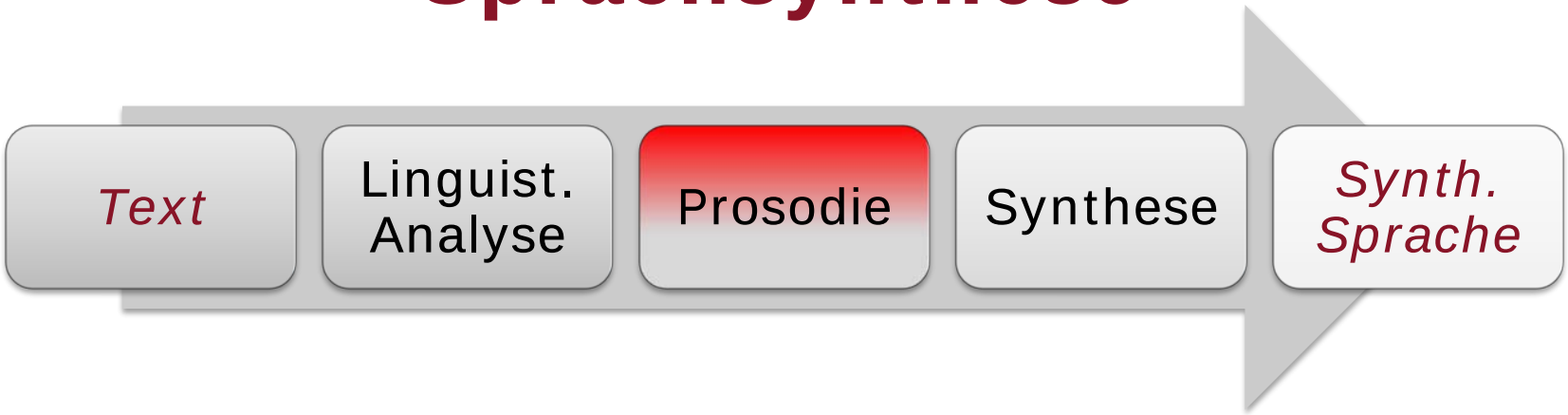
- Linguistische Analyse: lexikalisch, morphologisch, syntaktisch
- Prosodie: Bestimmung des Satzmodus, Festlegung von Silben-, Wort-, Phrasen- und Satzakzenten
- Synthese: Vollformenwörterbuch vs. Ausspracheregeln

Sprachsynthese



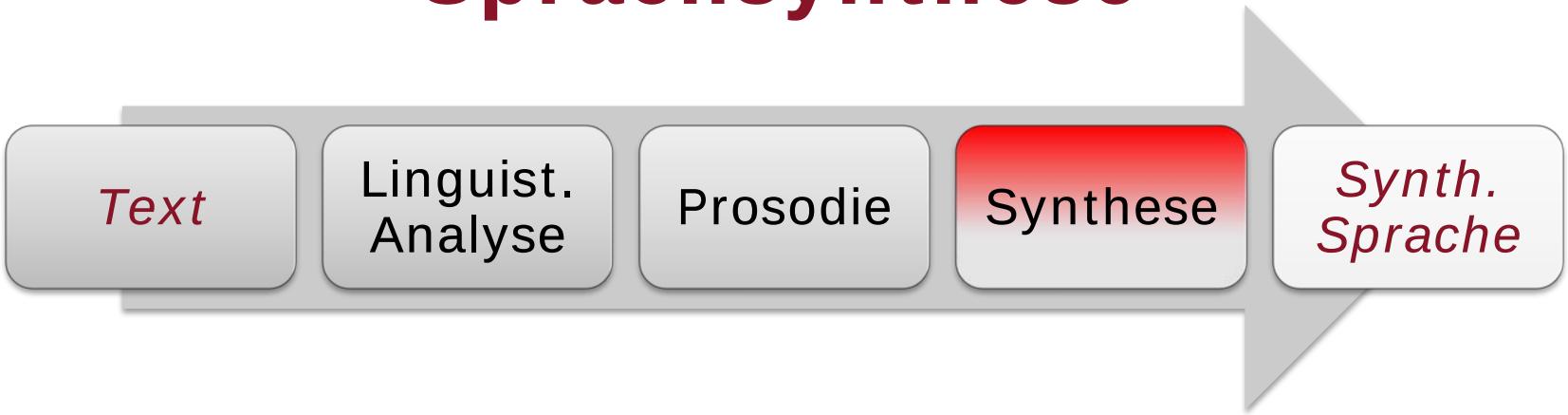
- Linguistische Analyse: lexikalisch, morphologisch, syntaktisch
- entsprechende Vorbereitung:
  - Punkte: Ende eines Satzes vs. Abkürzungsmarkierung vs. Komma
  - Zahlen: Entscheidung „normale Zahl“ oder Jahreszahl („eintausendneunhundertsechsundfünfzig“ vs. „neunzehnhundertsechsundfünfzig“)
  - Abkürzungen: Ermittlung der Aussprache („etc.“ zu „et cetera“)

Sprachsynthese

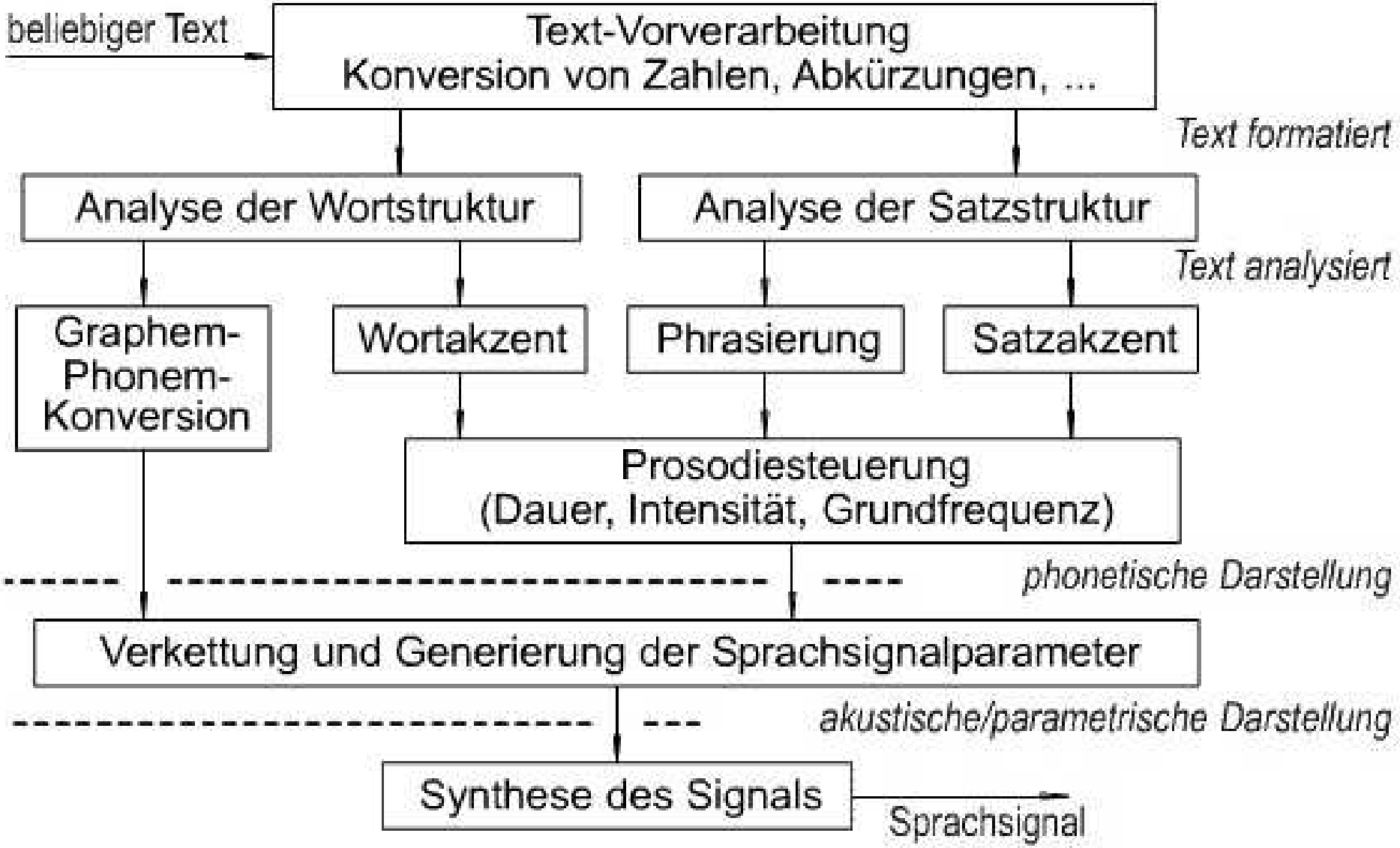


- Prosodie: Bestimmung des Satzmodus, Festlegung von Silben-, Wort-, Phrasen- und Satzakzenten
- Blickwinkel:
  - Artikulation: Vokale, Koartikulation
  - Wortprosodie: Töne, Wortakzent
  - Satzprosodie: Phrase, Satzebene, satzübergreifend
  - Para- und extralinguistische Parameter: Sprechstil, Emotion

Sprachsynthese



- Vollformenwörterbuch
  - Grundlage: mehrere Stunden Text vom selben Sprecher
  - Struktur des Korpus (Sätze, Wörter, Silben, Halbsilben)
- Ausspracheregeln
  - Umwandlung von (Buchstaben-)Schrift in Lautschrift
  - meist Kombination aus Regelwerk und Ausnahmelexikon


beliebiger Text
Text-Vorverarbeitung
Konversion von Zahlen, Abkürzungen, ...
Text formatiert
Analyse der Wortstruktur
Analyse der Satzstruktur
Text analysiert
Graphem-
Phonem-
Konversion
Wortakzent
Phrasierung
Satzakzent
Prosodiesteuerung
(Dauer, Intensität, Grundfrequenz)
phonetische Darstellung
Verkettung und Generierung der Sprachsignalparameter
akustische/parametrische Darstellung
Synthese des Signals
Sprachsignal

Dialogsysteme

- Kommunikation zwischen Benutzer und Maschine
  - Sprachliche Ein- und Ausgabe
  - Textbasiert oder akustisch

Ablauf:

- Interpretation natürlicher Sprache
- Aktionsplanung
  - Aufruf entsprechender Programme/Funktionen
  - Anfrage an internes Datenbank-System
- Ausgabeplanung
  - Formulieren der entsprechenden Antwort und/oder
  - Durchführen der gewünschten Aktion

Literatur

- Carstensen, Kai-Uwe et. al. (Hg.): Computerlinguistik und Sprachtechnologie. Eine Einführung. München: Elsevier, 2004.
- Euler, Stephen: Grundkurs Spracherkennung. Wiesbaden: Vieweg, 2006.
- Hess, Wolfgang: Grundlagen der Sprachsignalverarbeitung. http://www.ikp.uni-bonn.de/dt/lehre/materialien/grundl_ssv/gsv_1f_b.pdf
- Hess, Wolfgang: Systeme der akustischen Mensch-Maschine-Kommunikation. http://www.ikp.uni-bonn.de/dt/lehre/materialien/sammk/index.html

Referats- / Seminararbeitsthemen

- Referat (20 Min. + Diskussion 10 Min.)
- pro Referatsthema eine Theorie + bestehende oder vorstellbare Anwendungsszenarien präsentieren

- empfohlen: interdisziplinäre Teams (2-3 TeilnehmerInnen)

- Themenvorschläge (3 Vorschläge, ev. mit Präferenz) **bis Mi., 5.11.**, per e-Mail an **kellner@ec.tuwien.ac.at**