

Sprachverarbeitung I / 10 HS 2016

Einführung in die Spracherkennung

Buch: Kapitel 11.1 bis 11.6

Beat Pfister



Übersicht

- 4 Nachmittage:
- Einführung in die Spracherkennung
 - Merkmalsextraktion
 - Spracherkennung mit Mustervergleich
 - Spracherkennung mit statistischen Modellen

Am Schluss werden Sie in der Lage sein,
einen eigenen Spracherkenner zu programmieren!

Programm heute

- Vorlesung:
- Warum Spracherkennung?
 - Probleme der Spracherkennung
 - Messen der Erkennungsleistung
 - Anwendungen

Übung: (keine)

Warum Spracherkennung?

Lautsprache ist ein **mächtiges**, **natürliches** und **effizientes** Kommunikationsmittel des Menschen

- mächtig → es lässt sich alles beschreiben
- natürlich → wird am häufigsten verwendet
- effizient → schneller als beispielsweise Schreiben
(ca. 150 – 300 Wörter / Minute)

→ Lautsprache ist attraktives Mittel für Mensch-Maschine-Kommunikation

Ziel der Spracherkennung

Sprachsignal \Rightarrow Was wurde gesagt? \Rightarrow Reaktion

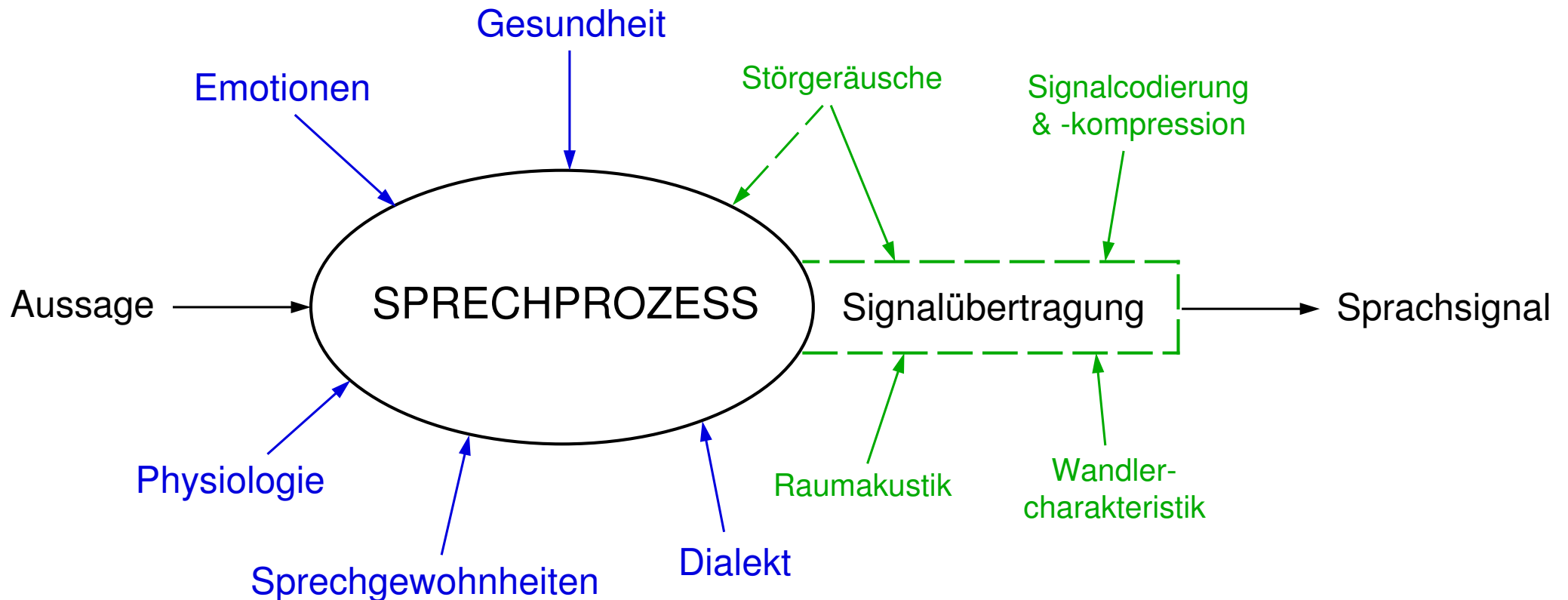
Probleme bei der Spracherkennung

- Diverse Einflüsse auf das Sprachsignal
- Variation der Dauer, der Lautstärke etc.
- Koartikulation
- Aussprachevarianten und Allophone
- Fehlende Wortgrenzen
- Nicht zum Vokabular des Erkenners gehörende Wörter

Probleme bei der Spracherkennung

- Diverse Einflüsse auf das Sprachsignal
- Variation der Dauer, der Lautstärke etc.
- Koartikulation
- Aussprachevarianten und Allophone
- Fehlende Wortgrenzen
- Nicht zum Vokabular des Erkenners gehörende Wörter

Einflüsse auf das Sprachsignal

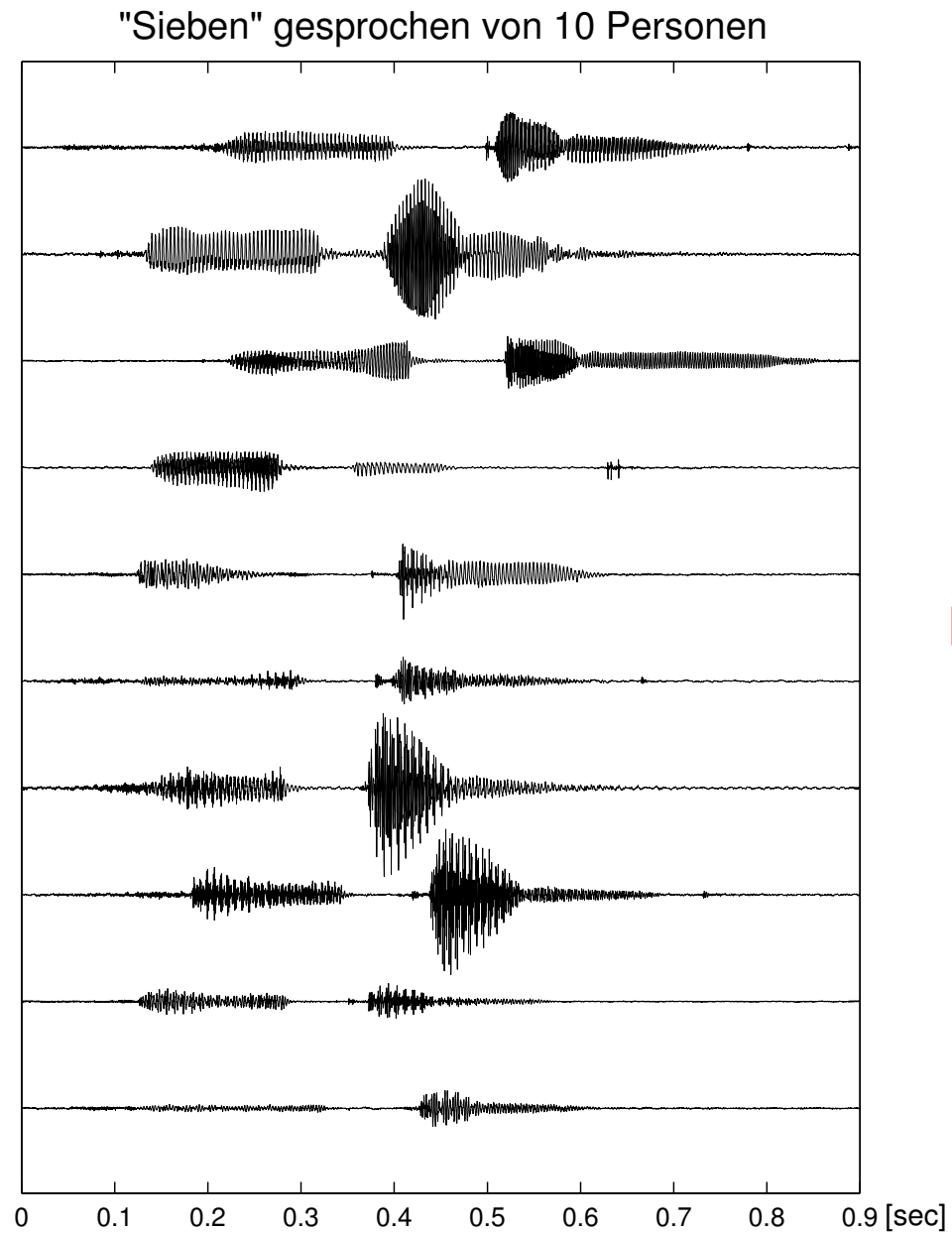


Probleme bei der Spracherkennung

- Diverse Einflüsse auf das Sprachsignal
- Variation der Dauer, der Lautstärke etc.
- Koartikulation
- Aussprachevarianten und Allophone
- Fehlende Wortgrenzen
- Nicht zum Vokabular des Erkenners gehörende Wörter

Variabilität

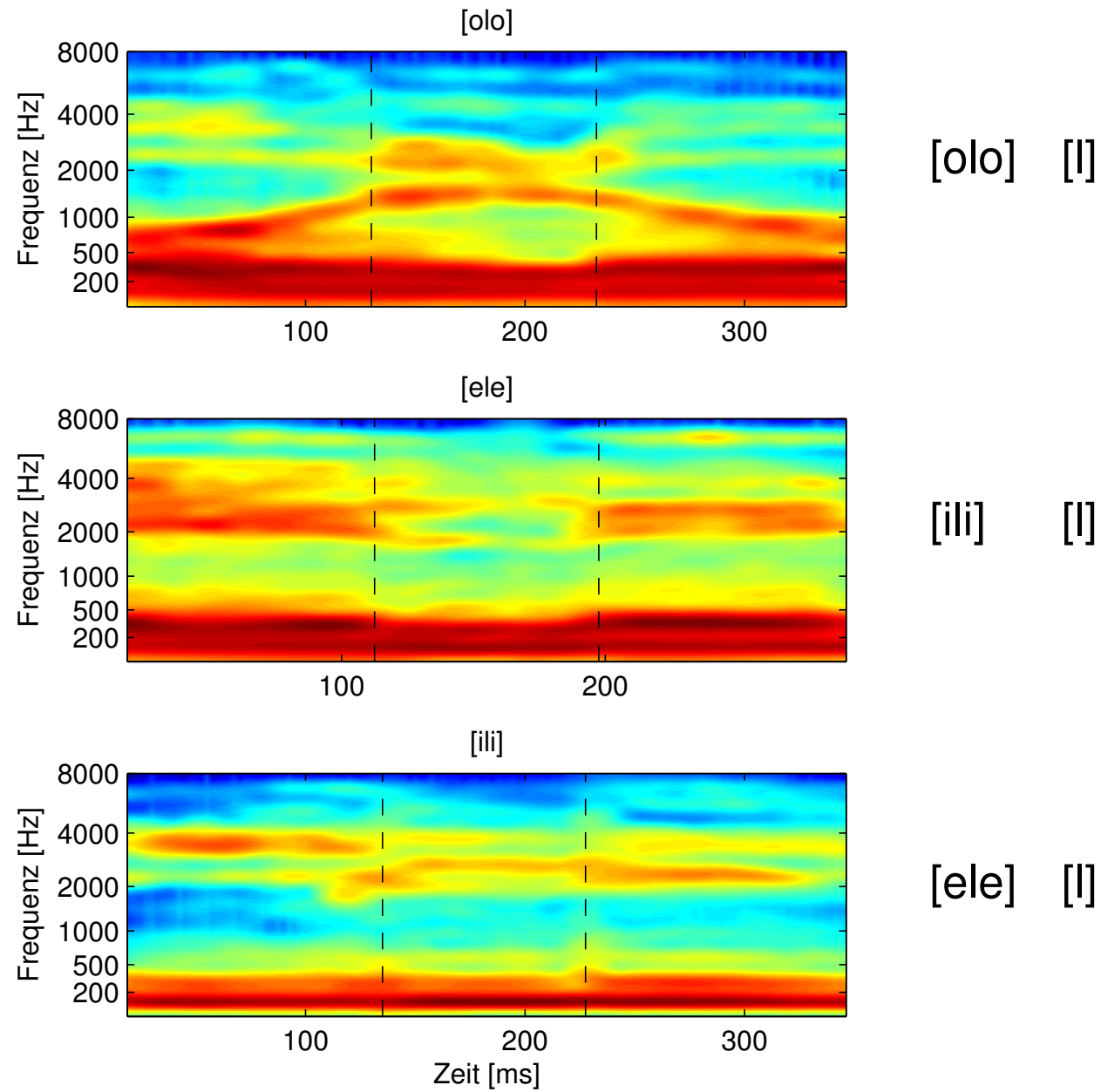
Dauer,
Lautstärke,
Stimme
etc.



Probleme bei der Spracherkennung

- Diverse Einflüsse auf das Sprachsignal
- Variation der Dauer, der Lautstärke etc.
- Koartikulation
- Aussprachevarianten und Allophone
- Fehlende Wortgrenzen
- Nicht zum Vokabular des Erkenners gehörende Wörter

Koartikulation



Probleme bei der Spracherkennung

- Diverse Einflüsse auf das Sprachsignal
- Variation der Dauer, der Lautstärke etc.
- Koartikulation
- Aussprachevarianten und Allophone
- Fehlende Wortgrenzen
- Nicht zum Vokabular des Erkenners gehörende Wörter

Aussprachevarianten

fünzig	[fʏnfʦɪç] → [fʏnfʦɪk]	(regionale Variante)
fünf	[fʏnf] → [fʏmf]	(lautliche Anpassung)
haben	[habŋ] → [habən]	(Einfügung / Auslassung)
guten Morgen	[gu:tən mɔrgən] → [gʊn mɔən]	(Verschleifung)

Allophone

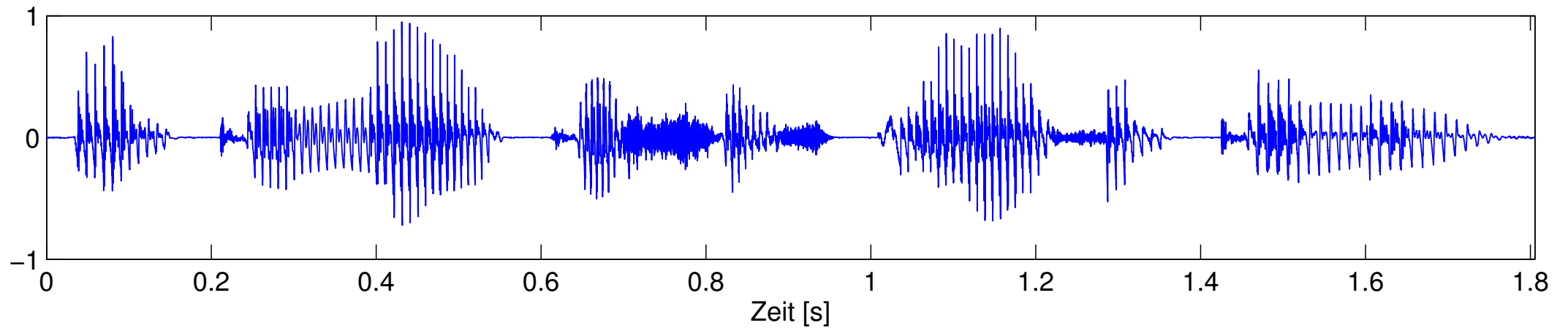
Mehrschlägiges Zungenspitzen-R	[r]	Vordergaumen-ch	[ç]
Einschlägiges Zungenspitzen-R	[r]	Hintergaumen-ch	[x]
Zäpfchen-R	[R]		
Reibe-R	[ʁ]		

Probleme bei der Spracherkennung

- Diverse Einflüsse auf das Sprachsignal
- Variation der Dauer, der Lautstärke etc.
- Koartikulation
- Aussprachevarianten und Allophone
- **Fehlende Wortgrenzen**
- Nicht zum Vokabular des Erkenners gehörende Wörter

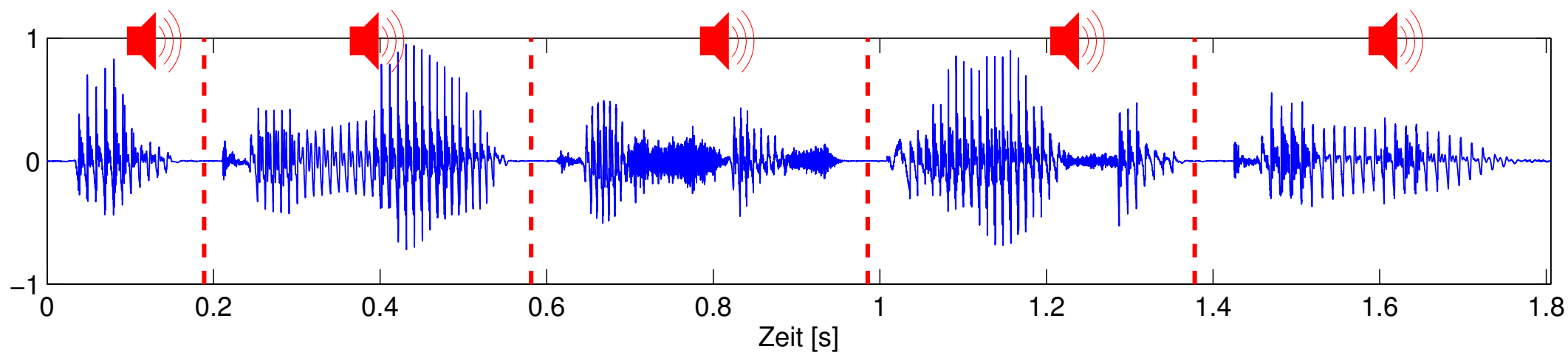
Wortgrenzen

Wo sind in diesem Sprachsignal die Wortgrenzen?



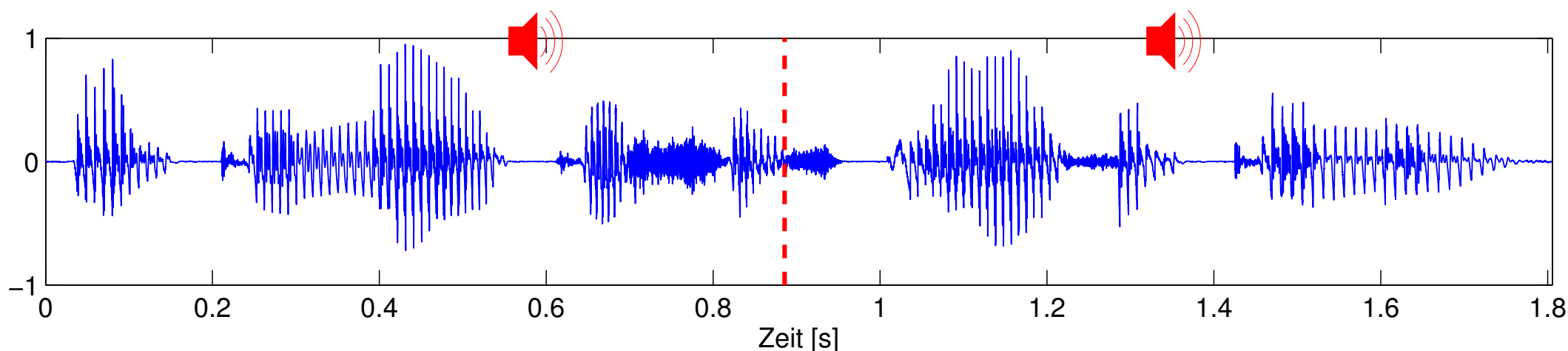
Wortgrenzen

Wo sind in diesem Sprachsignal die Wortgrenzen?



Wortgrenzen

Wo sind in diesem Sprachsignal die Wortgrenzen?



Wortgrenzen sind im Sprachsignal nicht sichtbar

—> **Wieviele und welche Wörter gesagt wurden muss gleichzeitig entschieden werden!**

Wie wirken sich fehlende Wortgrenzen aus?

Die Buchstabensequenz

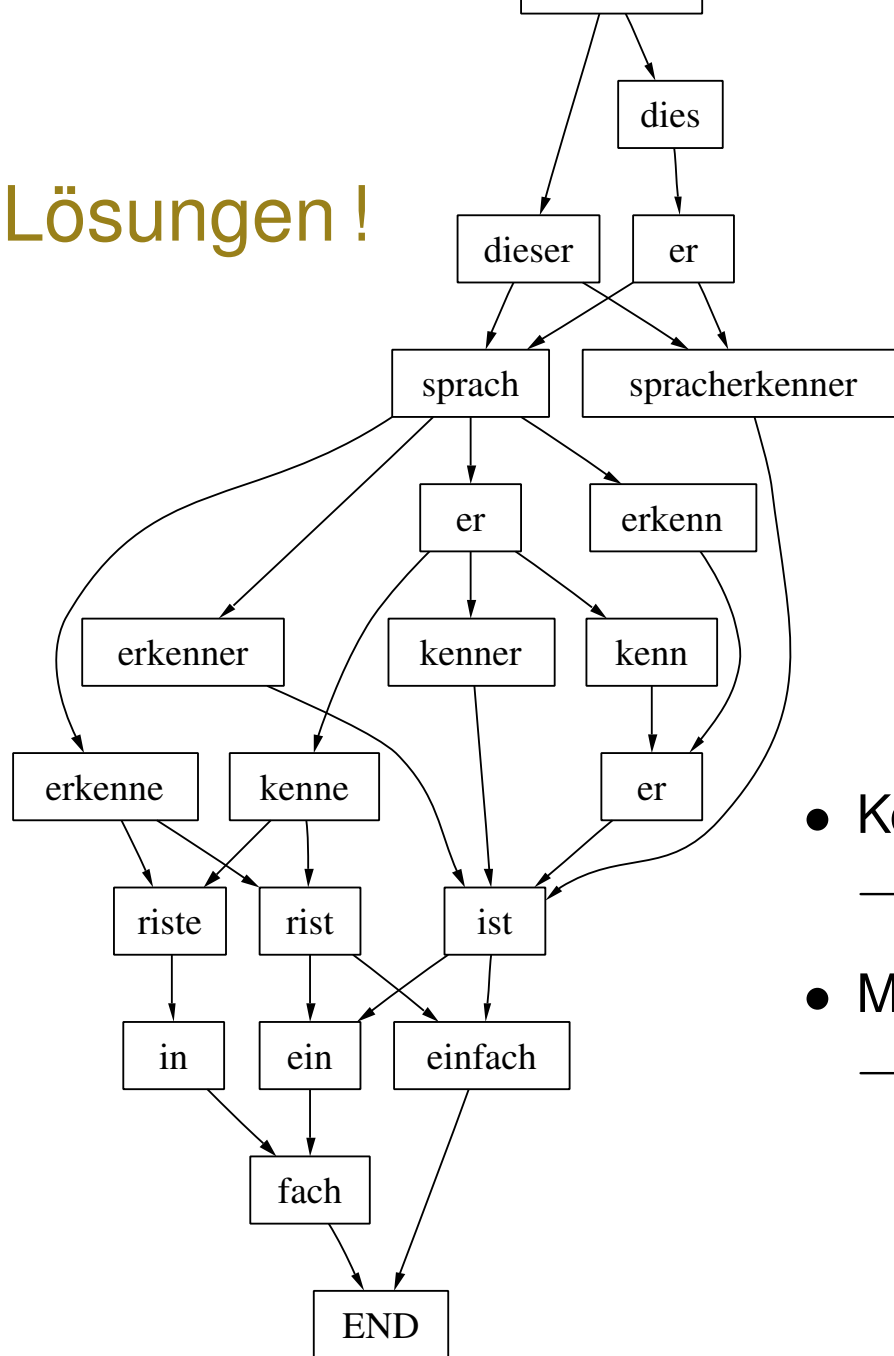
dieserspracherkenneristefach

soll in alle möglichen Folgen von deutschen Wörtern zerlegt werden, also z.B.

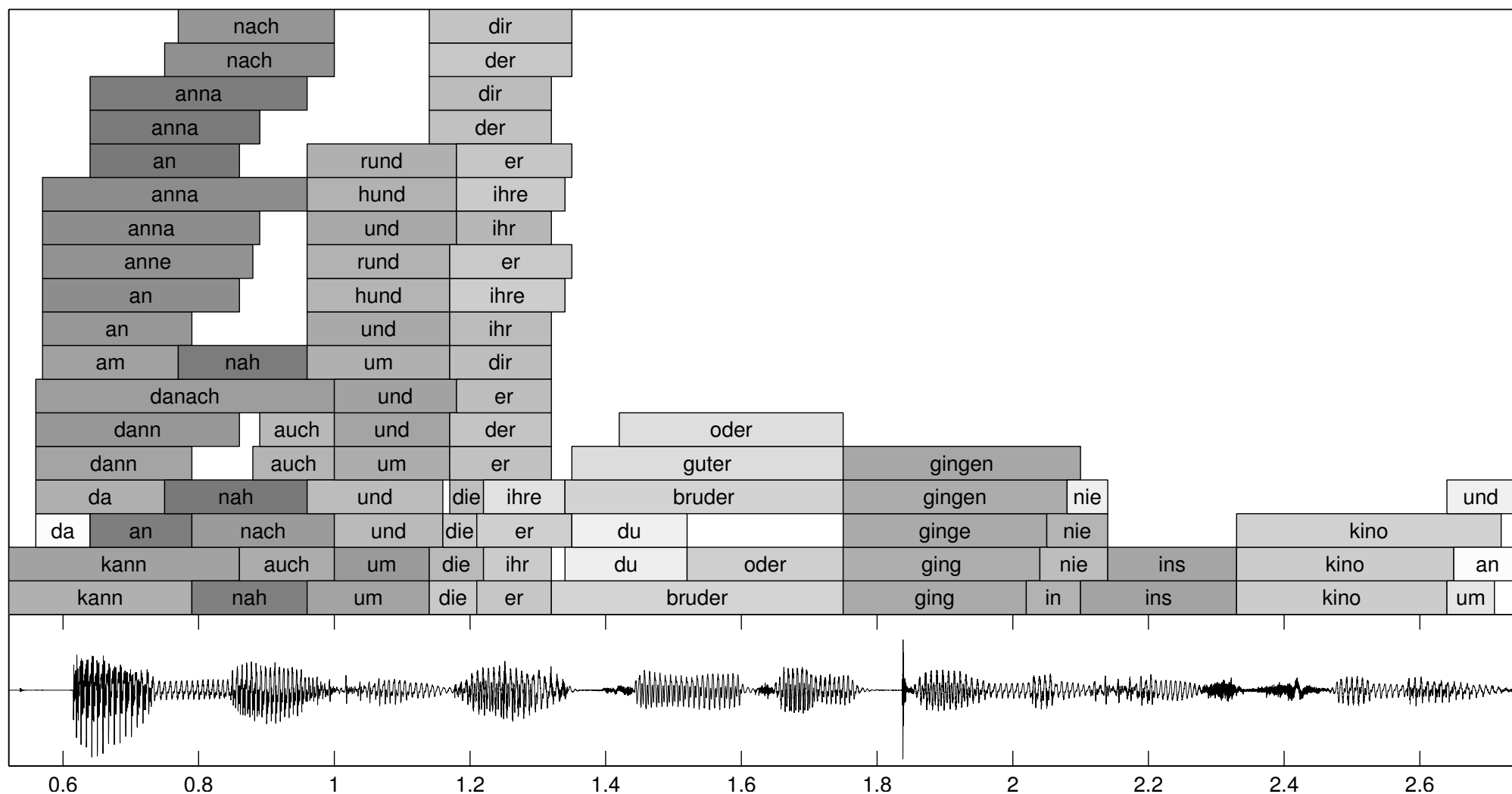
dieser spracherkenner ist efach

Wieviele Lösungen gibt es?

32 Lösungen !



- Keine Wortgrenzen
→ Starker Anstieg der Anzahl Lösungen
- Mehr Lösungen
→ höhere Fehlerwahrscheinlichkeit des Spracherkenners



Probleme bei der Spracherkennung

- Diverse Einflüsse auf das Sprachsignal
- Variation der Dauer, der Lautstärke etc.
- Koartikulation
- Aussprachevarianten und Allophone
- Fehlende Wortgrenzen
- Nicht zum Vokabular des Erkenners gehörende Wörter

Für den Spracherkenner unbekannte Wörter

Ein unbekanntes Wort kann in einem Mensch-Maschinen-Dialog zu grossen Verständigungsschwierigkeiten führen!

Beispiel: *Reisedialogsystem der Deutschen Bahn*



Problem: Es gibt zwar eine Ortschaft “**Rossau**” in Deutschland,
aber keinen Bahnhof mit diesem Namen.
Deshalb wird ein ähnlicher Ortsname wie “**Goslar**” erkannt.

Probleme bei der Spracherkennung

- Diverse Einflüsse auf das Sprachsignal
- Variation der Dauer, der Lautstärke etc.
- Koartikulation
- Aussprachevarianten und Allophone
- Fehlende Wortgrenzen
- Nicht zum Vokabular des Erkenners gehörende Wörter

Hauptproblem der Spracherkennung

Hohe Variabilität des Sprachsignals

- in zeitlicher Hinsicht
- in spektraler Hinsicht

—→ Es gibt keine Merkmale, anhand denen die sprachlichen Einheiten (d.h. die Laute) eindeutig unterschieden werden können !

Ansätze zur Spracherkennung

a) Sprachmustervergleich

- ein Muster für jedes Wort des Vokabulars bereitstellen
- ★ zu erkennende Äusserung wird mit allen Mustern verglichen
→ Wort mit dem ähnlichsten Muster gilt als erkannt

b) Statistischer Ansatz der Spracherkennung

- statistische Beschreibung für jedes Wort bereitstellen
- ★ zu erkennende Äusserung wird an den Statistiken getestet
→ Wort, dessen Statistik am besten passt, gilt als erkannt

- Training des Erkenners
- ★ Einsatz des Erkenners

Spracherkenner

Den idealen Spracherkenner, der für jeden Anwendungsfall einsetzbar ist (d.h. mit akzeptabler Fehlerrate), den gibt es nicht.

—→ Spracherkenner werden für spezielle Aufgaben ausgelegt, um eine genügend kleine Fehlerrate zu erreichen.

Einteilung von Spracherkennungssystemen

Spracherkenner können aufgrund der folgenden Kriterien eingeteilt werden:

- Zu erkennende Äusserungen
- Sprecherabhängigkeit
- Vokabulargrösse

Einteilung von Spracherkennungssystemen

Spracherkenner können aufgrund der folgenden Kriterien eingeteilt werden:

- Zu erkennende Äusserungen
- Sprecherabhängigkeit
- Vokabulargrösse

Kriterium: Zu erkennende Äusserungen

Systemklasse	Verarbeitbare Äusserungen
Einzelworterkenner	Einzelne Wörter oder kurze Kommandos isoliert gesprochen
Keyword-Spotter	Einzelne Wörter oder kurze Kommandos in einer sonst beliebigen Äusserung
Verbundworterkenner	Sequenz von fliessend gesprochenen Wörtern aus kleinem Vokabular (z.B. Telefonnummern)
Kontinuierlicher Spracherkenner	Ganze, fliessend gesprochene Sätze

Einteilung von Spracherkennungssystemen

Spracherkenner können aufgrund der folgenden Kriterien eingeteilt werden:

- Zu erkennende Äusserungen
- Sprecherabhängigkeit
- Vokabulargrösse

Kriterium: Sprecherabhängigkeit

- sprecherabhängig (nur mit Sprachsignalen des Benutzers trainiert)
- sprecherunabhängig (mit sehr vielen Stimmen trainiert)
- sprecheradaptiert (wird vor oder beim Einsatz an den Benutzer angepasst)

Einteilung von Spracherkennungssystemen

Spracherkenner können aufgrund der folgenden Kriterien eingeteilt werden:

- Zu erkennende Äusserungen
- Sprecherabhängigkeit
- **Vokabulargrösse**

Kriterium: Vokabulargrösse

Vokabulargrösse:

klein	< 100	Wörter
mittel	100 – 1'000	Wörter
gross	1'000 – 10'000	Wörter
sehr gross	> 10'000	Wörter

Heute existieren Systeme mit einem Vokabular > 500'000 Wörter.

Messung der Erkennungsleistung: Wortfehlerrate

Gesagt wurde: “Guten Morgen, es ist Viertel vor sechs”

Erkannt wurde: “Guten morgens ist vierte vor sie sechs”

Guten Morgen es ist Viertel vor ### sechs

Guten **morgens** **##** ist **vierte** vor **sie** sechs

Anzahl Wortfehler = **#Ersetzungen** + **#Auslassungen** + **#Einfügungen**

(Anzahl der Wortfehler entspricht Editierdistanz)

$$\text{Wortfehlerrate} = \frac{\text{Anzahl Wortfehler}}{\text{Anzahl zu erkennende Wörter}}$$

Angabe der Erkennungsleistung

Wortfehlerrate: Prozentualer Anteil der nicht erkannten Wörter

Worterkennungsrate: Prozentualer Anteil der erkannten Wörter

→ $\text{Erkennungsrate} = 100\% - \text{Fehlerrate}$

Anwendungen der Spracherkennung

- Eyes-/Hands-Busy-Anwendungen
 - Steuerung von Telefon, Stereoanlage oder Klimaanlage im Auto
 - Bedienung von Geräten im Operationssaal oder im Dunkelraum
 - Bedienungshilfen für Behinderte
 - Eingabe von medizinischen Befunden (während Untersuchung)
 - Bedienung von Miniaturgeräten (zu klein für Tastatur)
- Telefonbasierte Anwendungen
 - Automatische Auskunftssysteme (Wetter, Telefonnummern, Fahrplan)
 - Automatische Telefondienste (Teleshopping, Telebanking)
 - Intelligente Anrufweiterleitung (Call Preselection)
 - Intelligenter Telefonbeantworter (Lokal- und Fernbedienung)
- Diktiersysteme

Bedingungen für erfolgreichen Einsatz

- Erkennungsrate über einer bestimmten Schwelle (WER > 95 %)
- Systemantwortzeit kurz genug
- Echter Nutzen für Benutzer (SE an sich ist kein Nutzen)
- Benutzer muss sich wohl fühlen (z.B. immer wissen wie weiter)

>>>

Thema der nächsten Lektion:

Merkmalsextraktion

Zur Übersicht der Vorlesung *Sprachverarbeitung I* >>>

Bürgertelefon der Stadt Augsburg

© 1999 Randy Glasbergen.
www.glasbergen.com



Bürgertelefon der Stadt Augsburg

© 1999 Randy Glasbergen.
www.glasbergen.com



**“...If you’d like to hear all of your options again,
press 49. If you’ve forgotten why you called
in the first place, press 50.”**

<<<

