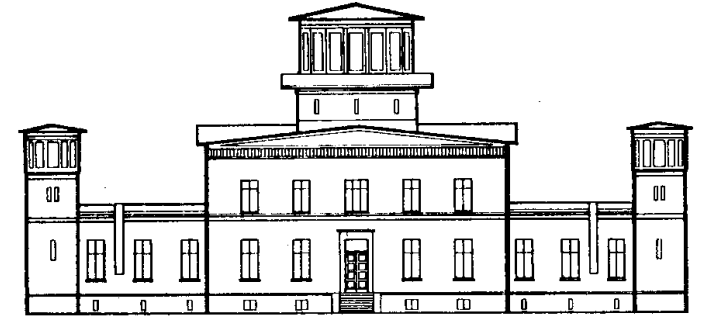


28.04.2003



# Grundlagen der Sprachsignalverarbeitung

Prof. Dr.-Ing. Wolfgang Hess

Institut für Kommunikationsforschung und Phonetik (IKP)

Universität Bonn

Poppelsdorfer Allee 47, 53115 Bonn

[wgh@ikp.uni-bonn.de](mailto:wgh@ikp.uni-bonn.de)

<http://www.ikp.uni-bonn.de>



# Grundlagen der Sprachsignalverarbeitung

1. Einführung
2. Kurzzeitanalyse; elementare Parameter
3. Kurzzeitspektralanalyse
4. Vom Spektrum zum Spektrogramm



Vorlesungsskripten – deutsch – bilingual deutsch/englisch

Allgemein zur Sprachsignalverarbeitung

Rosen, Stuart / Howell, Peter (1991): *Signals and systems for speech and hearing* (Academic Press, London)

Rabiner, Lawrence R. / Schafer, Ronald W. (1978): *Digital processing of speech signals*. Prentice-Hall Signal Proc. Series, ed. by A.V.Oppenheim (Prentice Hall, Englewood Cliffs, NJ)

Vary, Peter / Heute, Ulrich / Hess, Wolfgang (1998): *Digitale Sprachsignalverarbeitung* (Teubner, Stuttgart) **(F)**

Speziell zu Kapitel 1 (weitere Quellenangaben)

Meyer-Eppler, Werner (<sup>2</sup>1969): *Grundlagen und Anwendungen der Informationstheorie*. (2. Aufl. 1969; mit einer Erweiterung von G. Heike; 1. Aufl. 1959) (Springer, Berlin) **(F)**

Shannon, Claude E. (1949): "Communication in the presence of noise." Proc. IRE 37, 10-21 **(F)**

**(F)** – Lektüre für Fortgeschrittene



# Grundlagen der Sprachsignalverarbeitung

## 1. Einführung

### 1.1 Grundbeobachtungen am Sprachsignal

1.2 Bau des menschlichen Sprechtrakts (kurze Einführung)

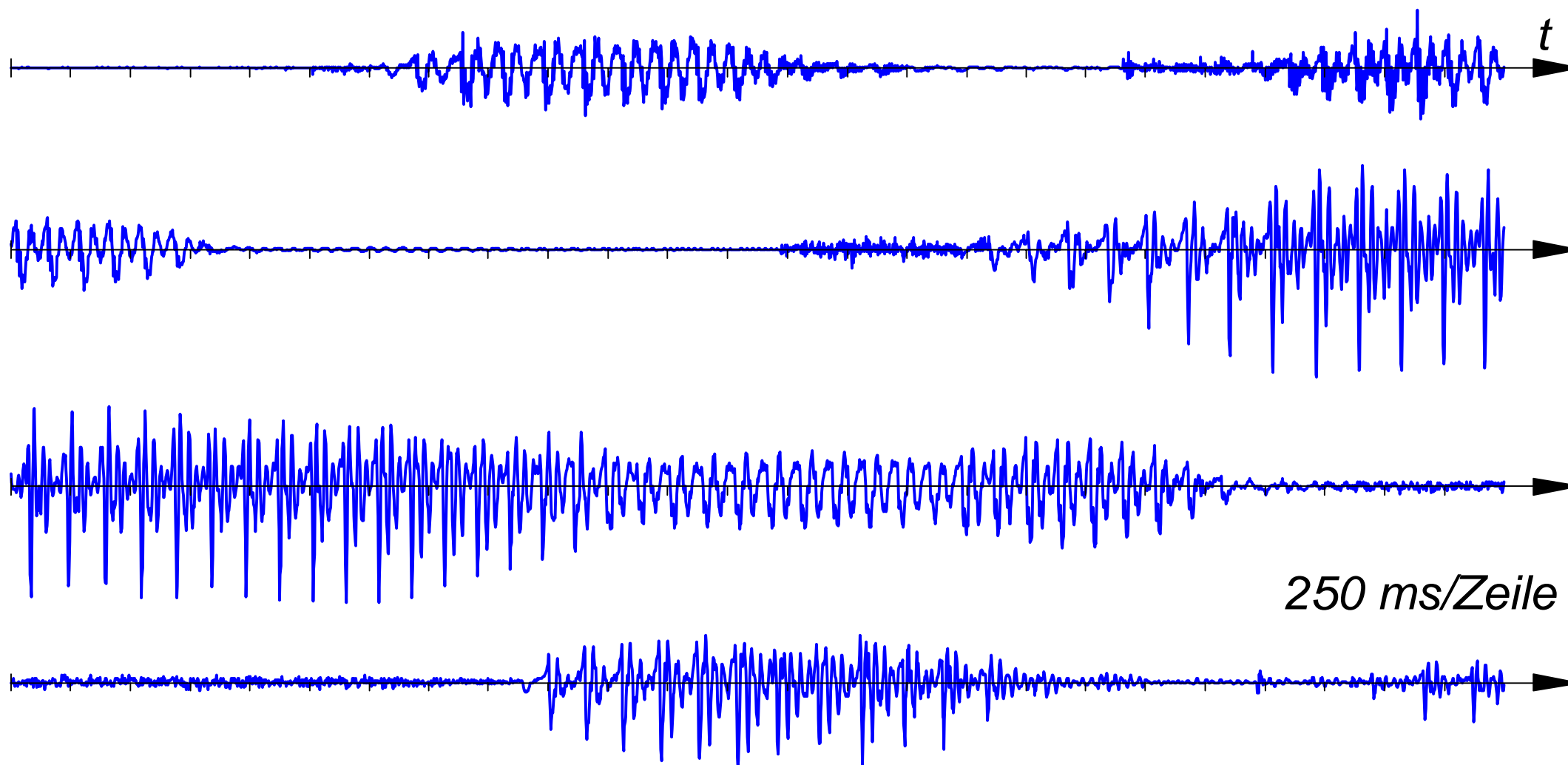
1.3 Signaldarstellung im Rechner

1.4 Der Informationsgehalt der Sprache

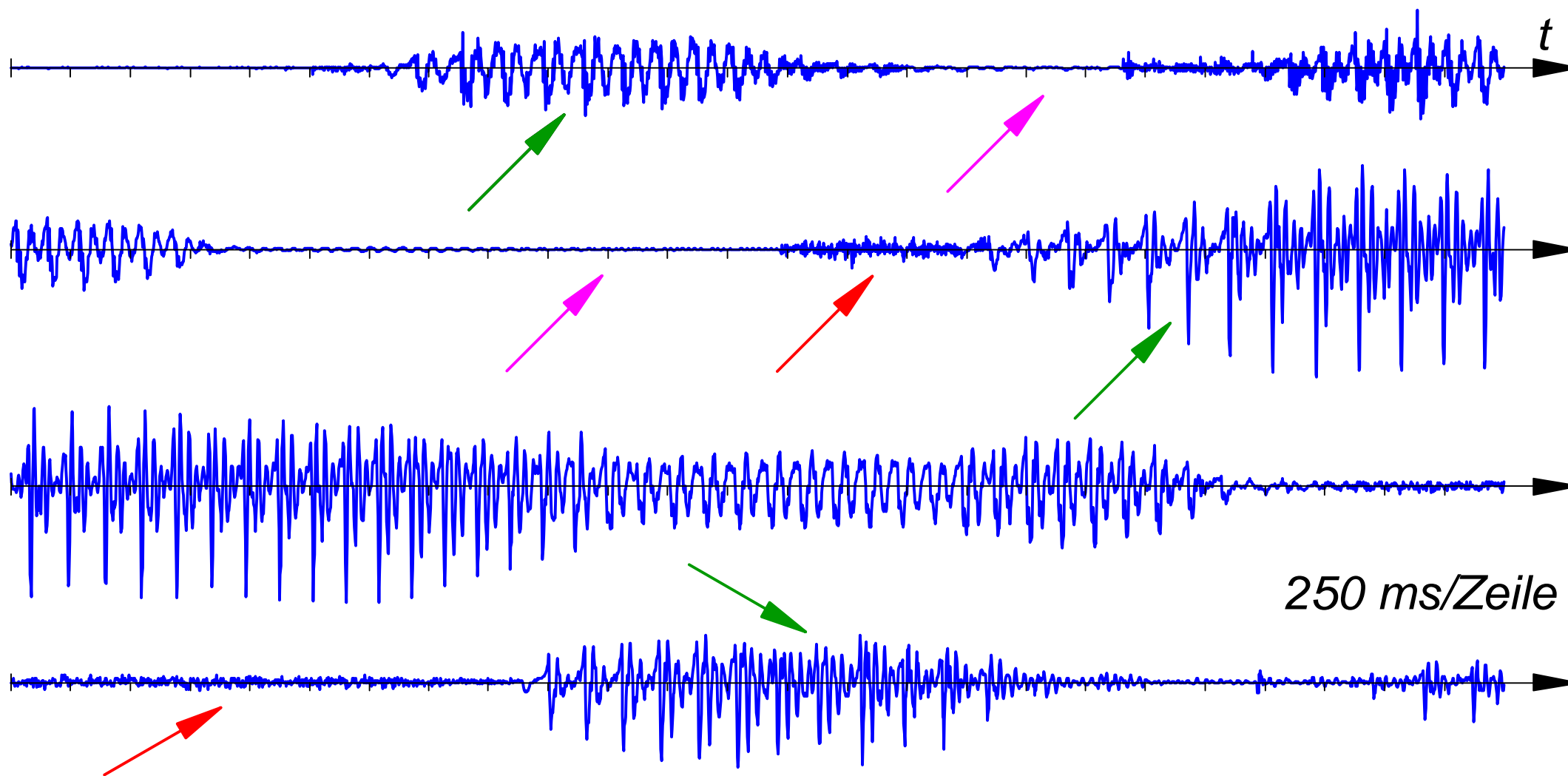
1.5 Die wesentlichen Aufgaben der Sprachsignalverarbeitung

1.6 Phonetische Einheiten und die Sprachsignalverarbeitung

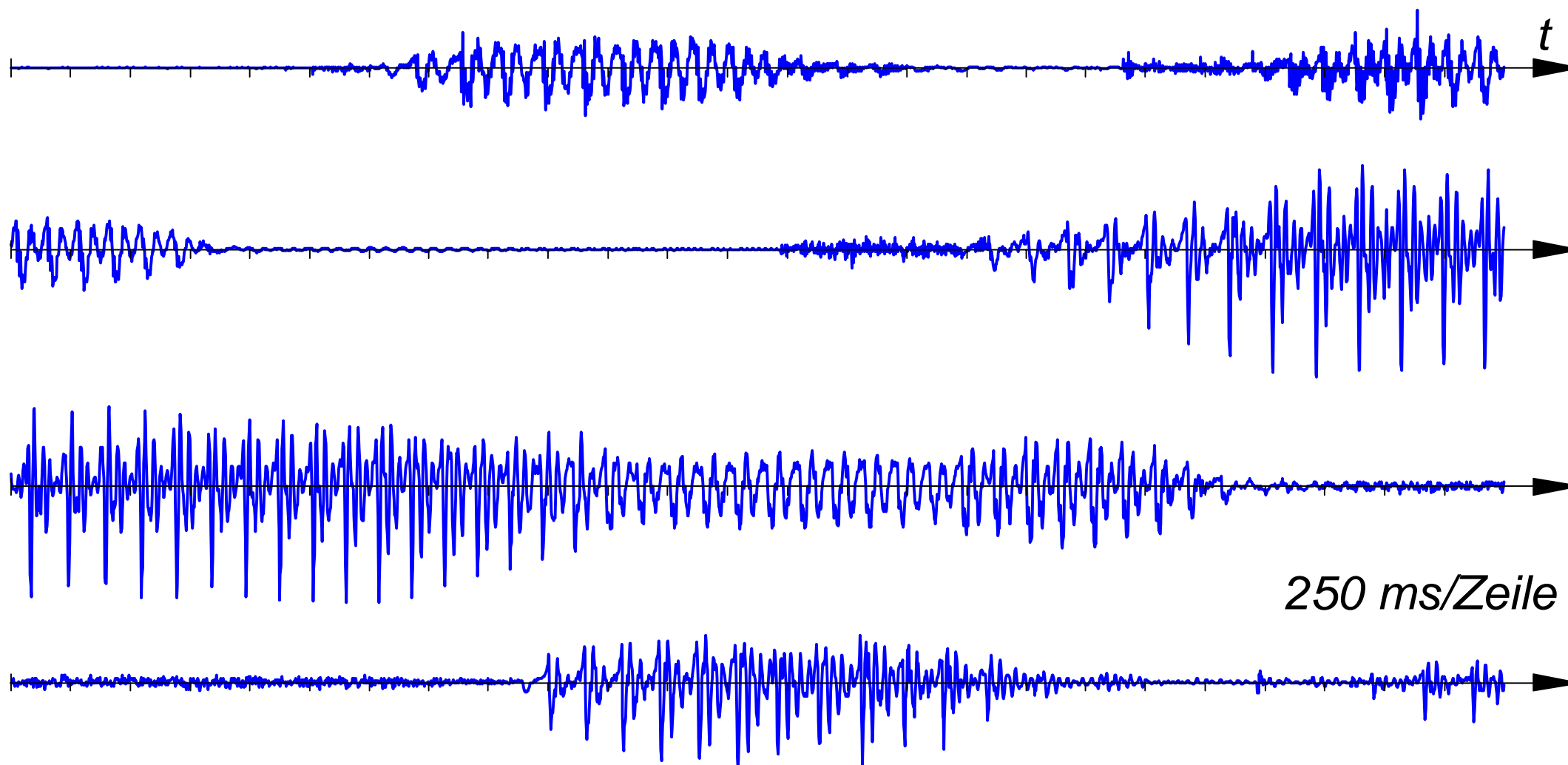
# Grundbeobachtungen am Sprachsignal [1]



Das Sprachsignal ist **zeitveränderlich**. In der Art und Weise, wie die Zeitveränderung erfolgt, liegt die Information, die das Sprachsignal zum Kommunikationsträger für den Menschen macht.



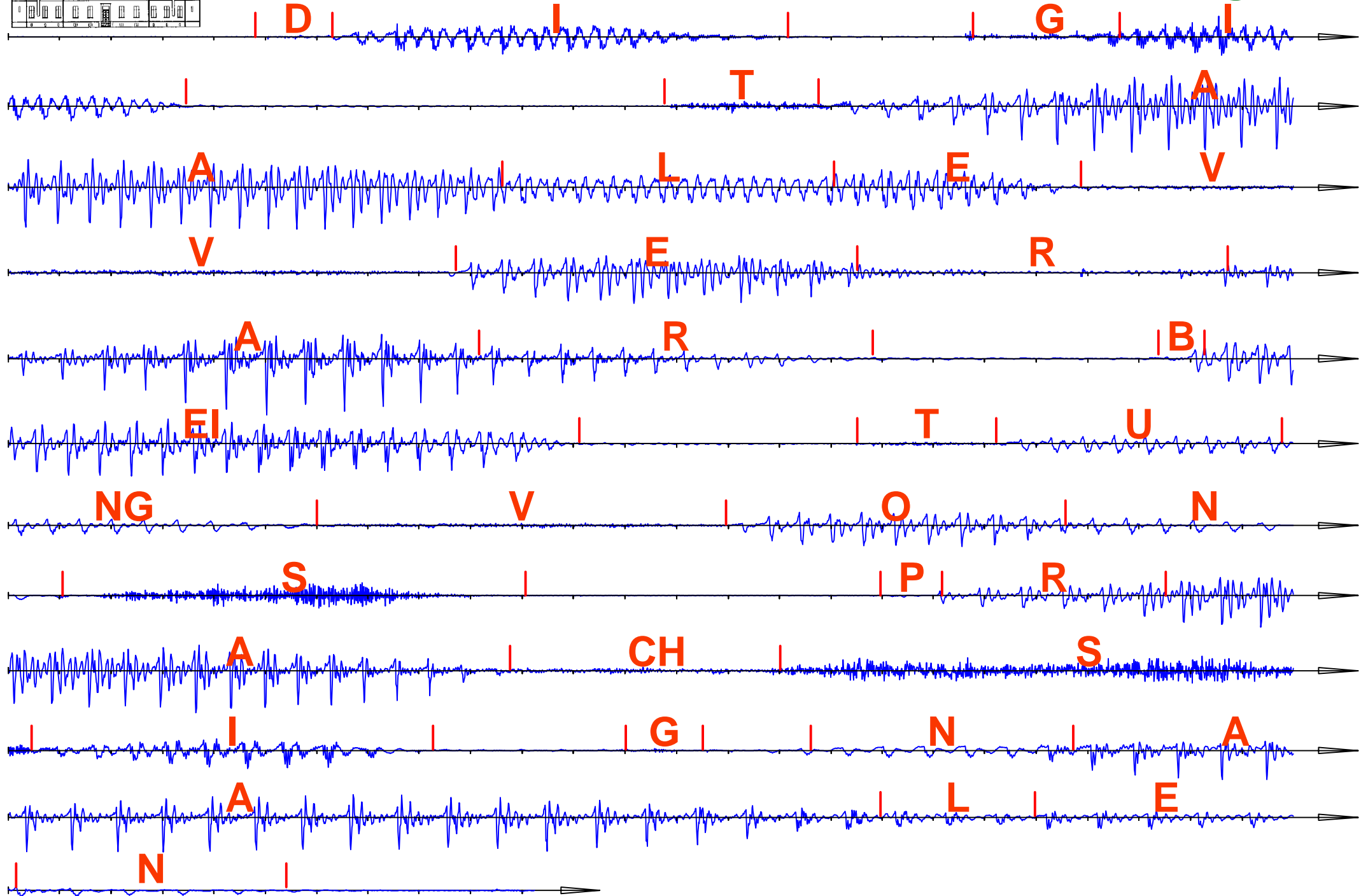
Das Sprachsignal ist in Abschnitten (quasi-)periodisch.  
In anderen Abschnitten besitzt es die Struktur eines Rauschens.  
Weiterhin existieren Pausen, die teilweise sehr kurz sind.



Im Vergleich zur Schriftform, wo wir eine Zeichenkette vorliegen haben, stellt das Sprachsignal ein **Kontinuum**, einen **Datenstrom** dar. Änderungen der Signalstruktur erfolgen zumeist fließend, nicht abrupt.



# Des Rätsels Lösung ...







# Grundlagen der Sprachsignalverarbeitung

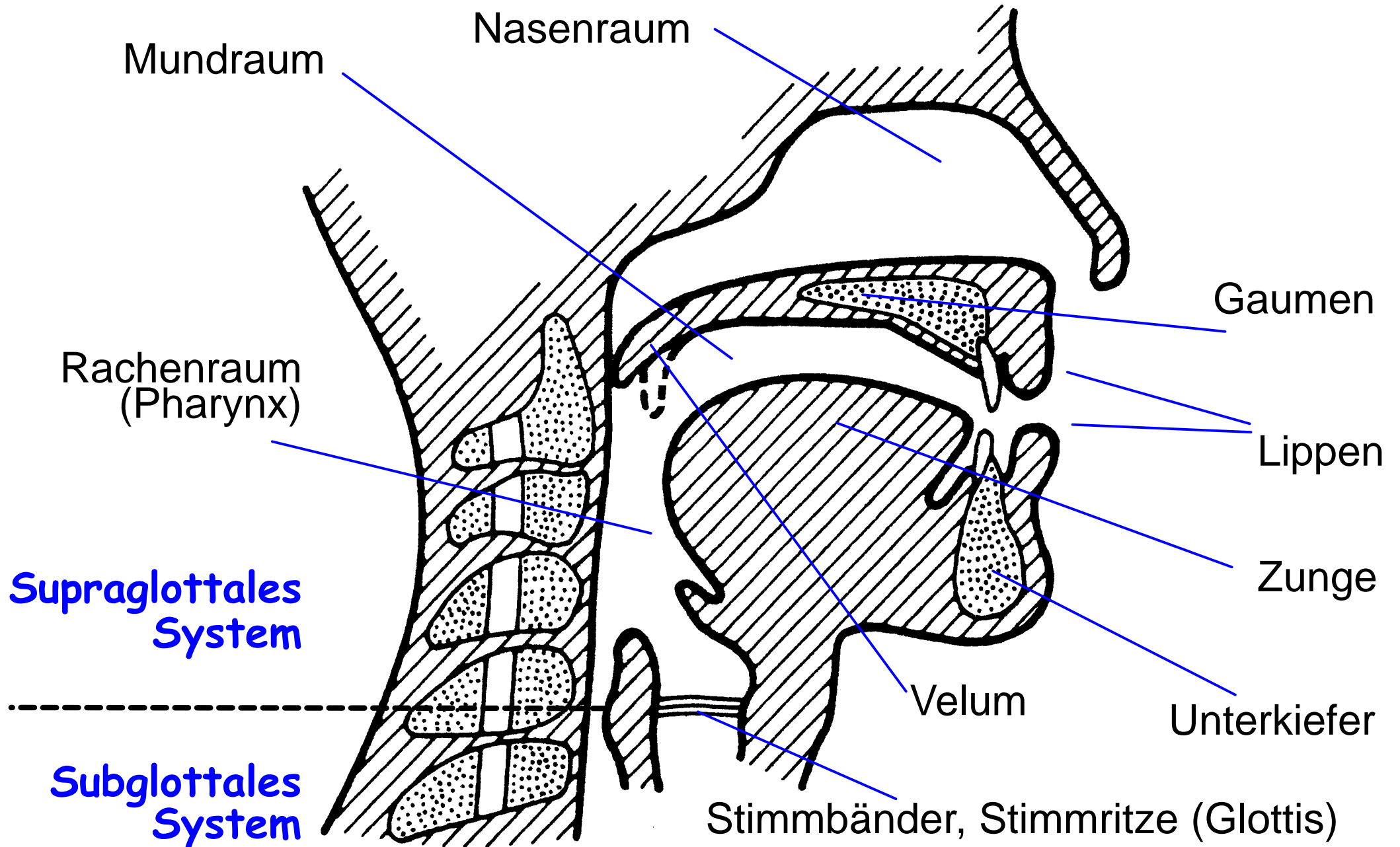
## 1. Einführung

- 1.1 Grundbeobachtungen am Sprachsignal
- 1.2 Bau des menschlichen Sprechtrakts (kurze Einführung)**
- 1.3 Signaldarstellung im Rechner
- 1.4 Der Informationsgehalt der Sprache
- 1.5 Die wesentlichen Aufgaben der Sprachsignalverarbeitung
- 1.6 Phonetische Einheiten und die Sprachsignalverarbeitung



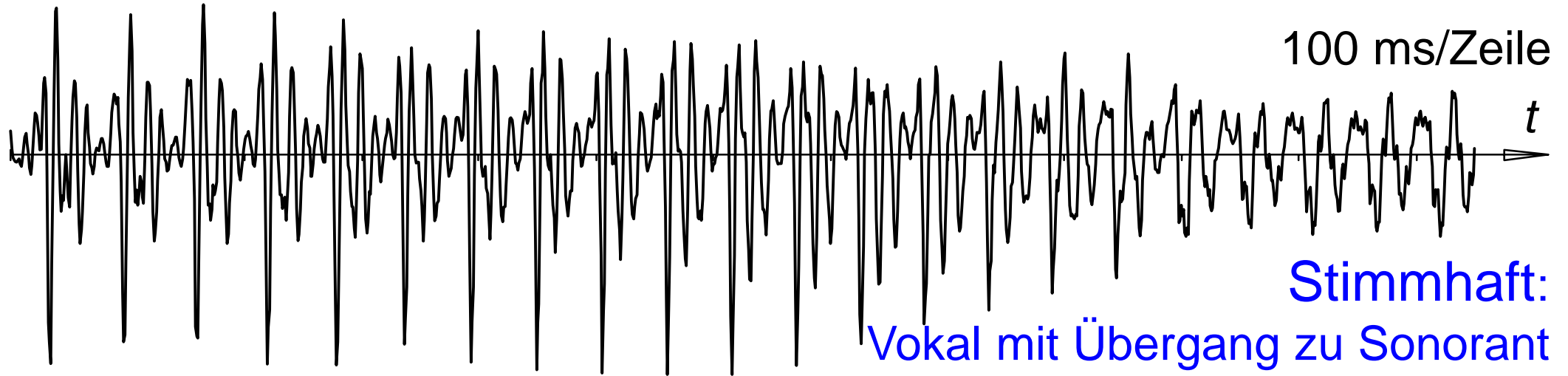
# Die menschlichen Sprechorgane

[aus: Grundlagen der Phonetik]

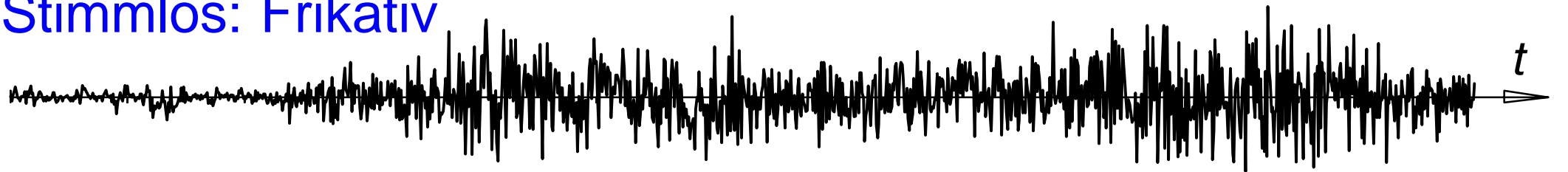




# Sprachsignalbeispiele für die einzelnen Anregungsarten



## Stimmlos: Frikativ

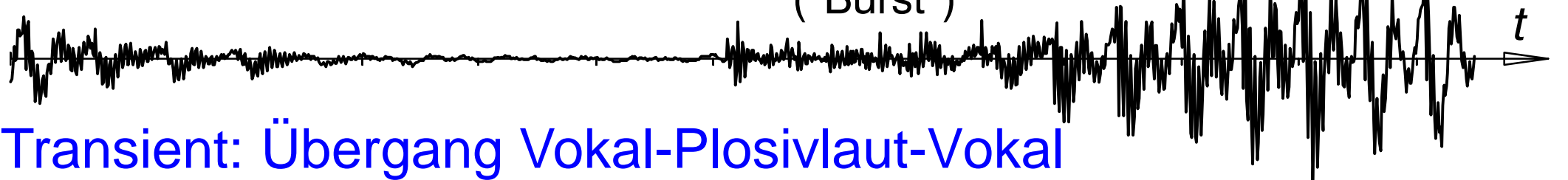


Vokal  
(Ende)

Verschlusspause

Plosions-  
geräusch  
("Burst")

Vokal



## Transient: Übergang Vokal-Plosivlaut-Vokal



# Grundlagen der Sprachsignalverarbeitung

## 1. Einführung

- 1.1 Grundbeobachtungen am Sprachsignal
- 1.2 Bau des menschlichen Sprechtrakts (kurze Einführung)
- 1.3 **Signaldarstellung im Rechner**
- 1.4 Der Informationsgehalt der Sprache
- 1.5 Die wesentlichen Aufgaben der Sprachsignalverarbeitung
- 1.6 Phonetische Einheiten und die Sprachsignalverarbeitung



# Signale: Definition [1]

Ein **Signal** ist "das Erscheinungsbild einer physikalischen Information" (Kunt, 1980). Ein Signal ist beispielsweise beschreibbar durch

- eine **mathematische Funktion** im strengen Sinn, also einen analytischen Zusammenhang (in geschlossener Form), oder
- ein **Verteilungsgesetz** (z.B. für ein stochastisches Signal). Hier ist der Augenblickswert des Signals nicht bekannt, da die Verteilungsfunktion nur globale Signaleigenschaften beschreibt, oder
- **empirisch** durch eine **Messreihe**.

Die meisten Signale in der Praxis gehören zu dieser letzten Kategorie.



# Signale: Definition [2]

Ist ein Signal Funktion nur einer Veränderlichen, so sprechen wir von einem **eindimensionalen Signal**, ansonsten von einem **mehrdimensionalen Signal**.

Beispiele für eindimensionale Signale:

- Sprachsignale
- akustische Signale aller Art (Musik, Geräusch)
- biologische Signale (Elektrokardiogramm, Elektroencephalogramm)
- physikalische Signale (z.B. seismische Wellen)
- Radarsignale.

Beispiel für mehrdimensionaler Signale:

- (zweidimensionale) Bildsignale

Eindimensionale Signale sind meist **Zeitfunktionen** ("Vorgänge"): sie seien als  $s(t)$  bezeichnet;  $s(t)$  ist die **Augenblicks-amplitude**, der **Momentanwert** oder einfach der **Wert** des Signals zu einem bestimmten Zeitpunkt  $t$ .



# Signaldarstellung im Rechner

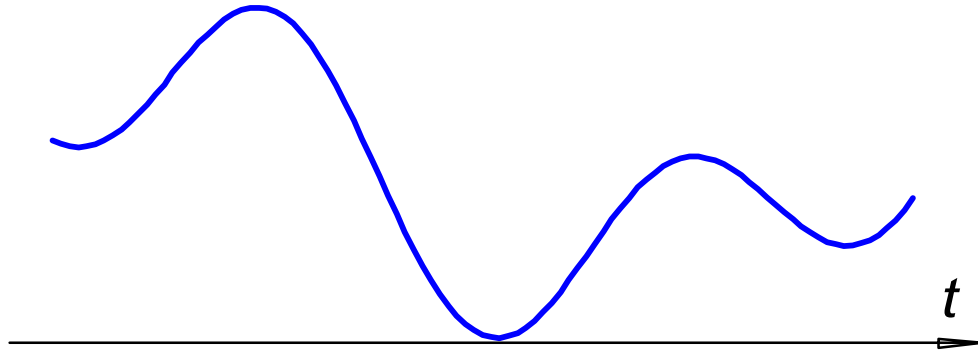
Signale können im Rechner - wie alles andere auch - nur als **Zahlenfolgen** gespeichert werden. Im Rechner wird ein Signal also **in diskreter Form** gespeichert.

Die Umwandlung des Signals in die diskrete Darstellung bezeichnen wir als **Digitalisierung** oder **Quantisierung**. Sie läuft in 2 Schritten ab:

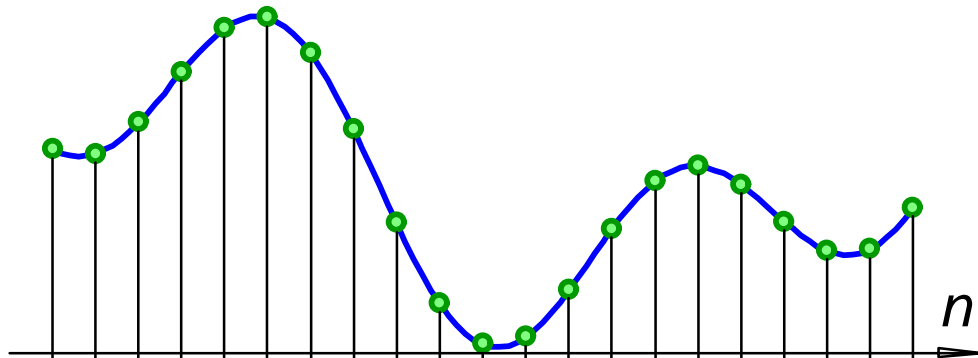
- Umwandlung der **unabhängigen Veränderlichen** in die diskrete Darstellung (**Abtastung**)
- Umwandlung der **abhängigen Veränderlichen** in die diskrete Darstellung (**Quantisierung**)



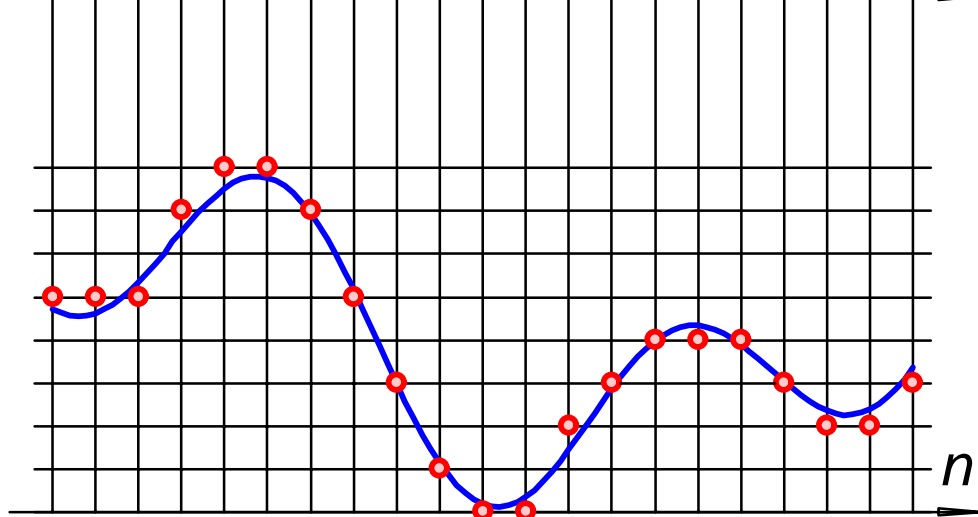
# Vom kontinuierlichen ("analogen") zum digitalen Signal [1]



kontinuierlich ("analog")



abgetastet

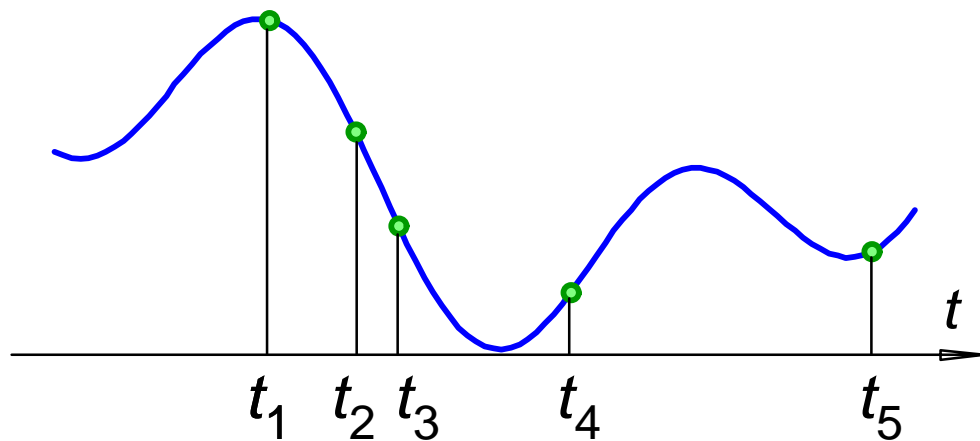
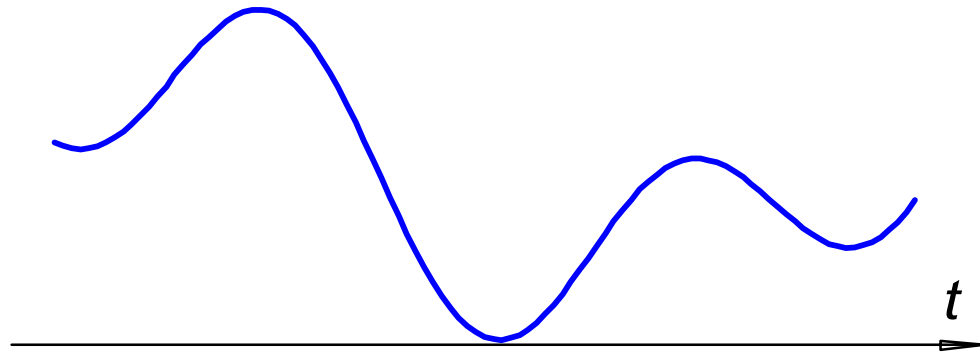


digital  
(abgetastet und quantisiert)





# Vom kontinuierlichen ("analogen") zum digitalen Signal [2]



## Darstellung als Zeitfunktion $s(t)$

- allgemein: Augenblicksamplitude
- in Luft: Schalldruck
- in Mikrofon und Verstärker: elektrische Spannung

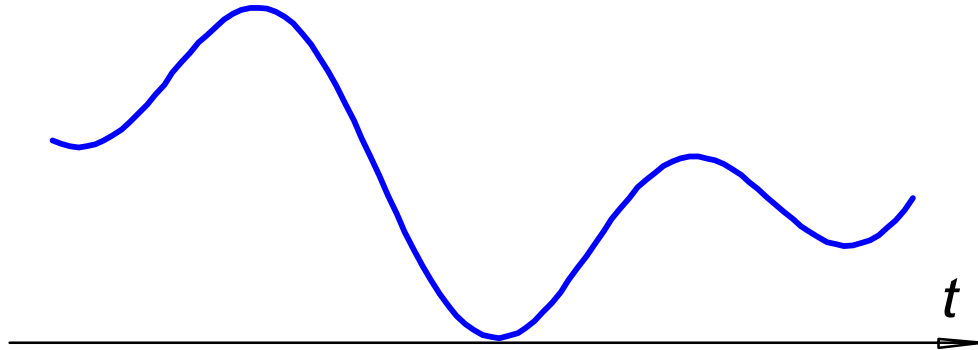
Quantisierung der unabhängigen Variablen ergibt **Abtastung** und zerlegt das Signal in eine **Folge von Abtastwerten**:

$$s(t) \rightarrow \{ \dots s(t_1), s(t_2), \dots \}$$

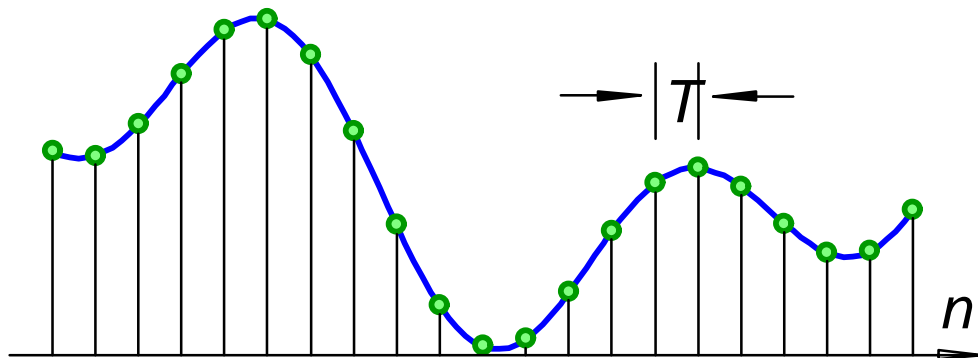
(1.2)



# Vom kontinuierlichen ("analogen") zum digitalen Signal [3]



Darstellung als Zeitfunktion  $s(t)$



**Abtastung:** Darstellung als  
abgetastete Zeitfunktion

$$s(t) \rightarrow s(nT) \rightarrow s(n)$$

(bei *gleichförmiger* Abtastung)

Vereinbarung: Die Abtastung erfolge *gleichförmig*, d.h., stets *in gleichen Intervallen der unabhängigen Variablen*:

$$s(n) := s(nT) ; \quad n = -\infty \dots -1, 0, 1 \dots +\infty \text{ [ganzzahlig]}$$

$T$ : Abtastintervall

(1.3)



# Abtastintervall und Abtastfrequenz

$$s(n) := s(nT) ; \quad n = -\infty \dots -1, 0, 1 \dots +\infty \text{ [ganzzahlig]} \quad (1.3)$$

$T$ : Abtastintervall

Das **Abtastintervall**  $T$  wird in der Bezeichnung  $s(n)$  für das Signal meist weggelassen.

Dementsprechend ist  $n$  nur noch indirekt ein Maß für die Zeit; daher sei  $n$  als **Index** oder **Adresse** des Abtastwertes (**Messwertadresse**) bezeichnet.

Der Kehrwert des Abtastintervalls wird als **Abtastfrequenz** bzw. **Abtastrate** bezeichnet:

$$F := \frac{1}{T} \quad (1.4)$$



Welche Werte das Abtastintervall  $T$  bzw. die Abtastfrequenz  $F$  annehmen müssen, sagt uns das **Abtasttheorem**:

Die Abtastfrequenz  $F$  muss mehr als das Doppelte der höchsten im Signal enthaltenen Frequenz betragen.

Missachtung dieser Regel führt zu groben Verzerrungen, die sich aber ausschließlich auf die Signalanteile auswirken, die das Abtasttheorem verletzen.

Ist das Abtasttheorem eingehalten, so lässt sich das analoge Signal - bis auf Quantisierungseffekte - fehlerfrei aus dem digitalen Signal rückwandeln.

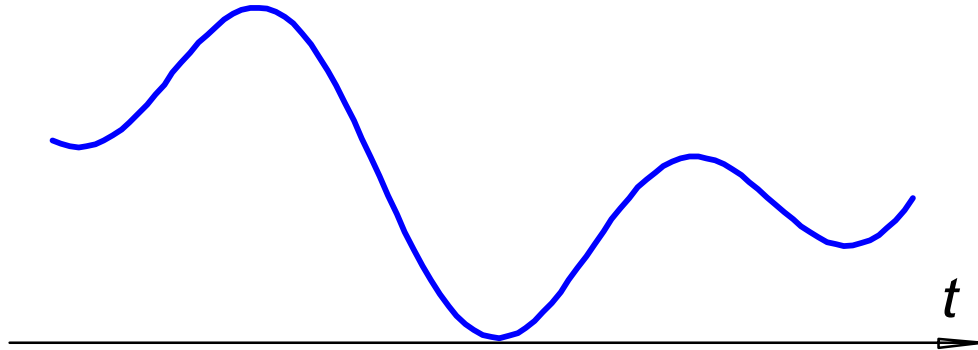


# Gängige Werte der Abtastfrequenz für Audiosignale

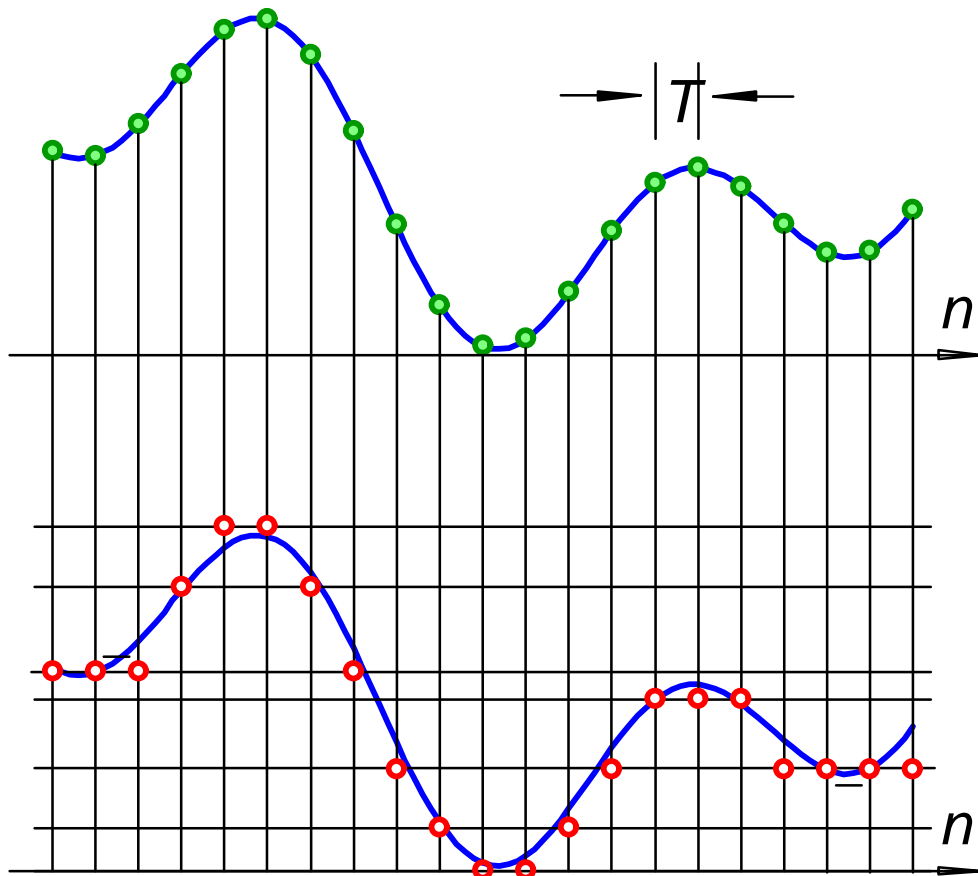
Signal	Frequenzen bis [kHz]	Abtastfrequenz [kHz]
Sprachsignal guter Qualität	7.5	16
	10	22.05
Sprachsignal minderer Qualität	5	11.025
Sprachsignal Telefonqualität	3.4	8
Musiksignal HiFi mindestens	15	32
CD	16	44.1
sonstige digitale Tonaufzeichnung (DAT)	16	48



# Vom kontinuierlichen ("analogen") zum digitalen Signal [4]



Darstellung als Zeitfunktion  $s(t)$



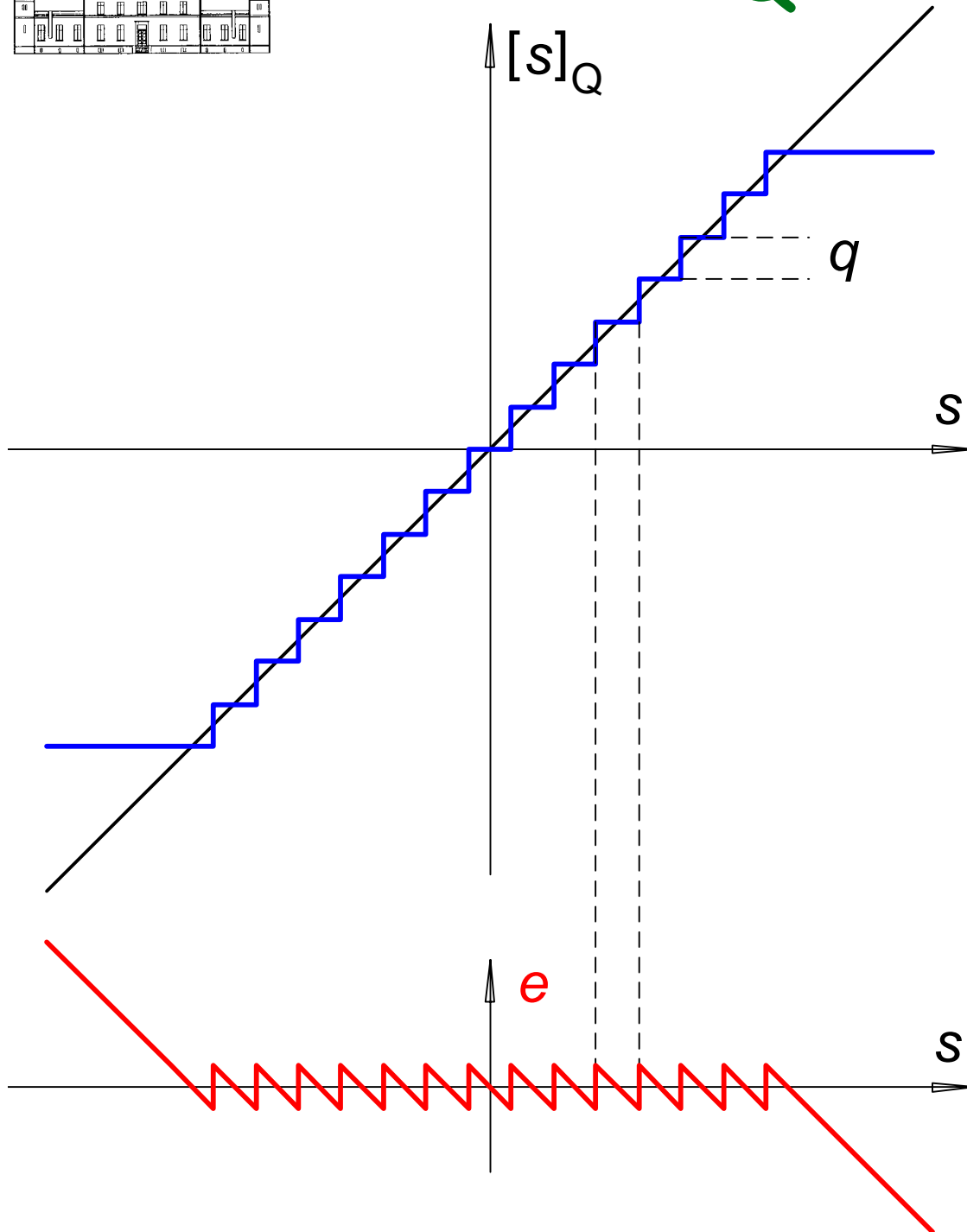
Darstellung als  
abgetastete Zeitfunktion  $s(n)$

**Quantisierung:**  
Darstellung im Rechner  
als Zahlenfolge

$$s(n) \rightarrow [s(n)]_Q$$



# Zur Quantisierung digitaler Signale



Vereinbarung: Die Quantisierung erfolge *gleichförmig*, d.h., stets *in gleichen Intervallen* der abhängigen Variablen:

$$(1.5) \quad s(n) \rightarrow [s(n)]_Q = k(n) \cdot q$$

$k$  ganzzahlig

$s$  Eingangssignal

$[s]_Q$  Eingangssignal  
quantisiert

$q$  Quantisierungsstufe

$e$  Quantisierungsfehler

Bei der Quantisierung gibt es hinsichtlich der Größe der Quantisierungsstufe  $q$  keine verbindliche, dem Abtasttheorem vergleichbare "Spielregel".

- Eine zu grobe Wahl der Quantisierungsstufe führt zu geräuschähnlichen Verzerrungen, die als **Quantisierungsrauschen** bekannt sind.

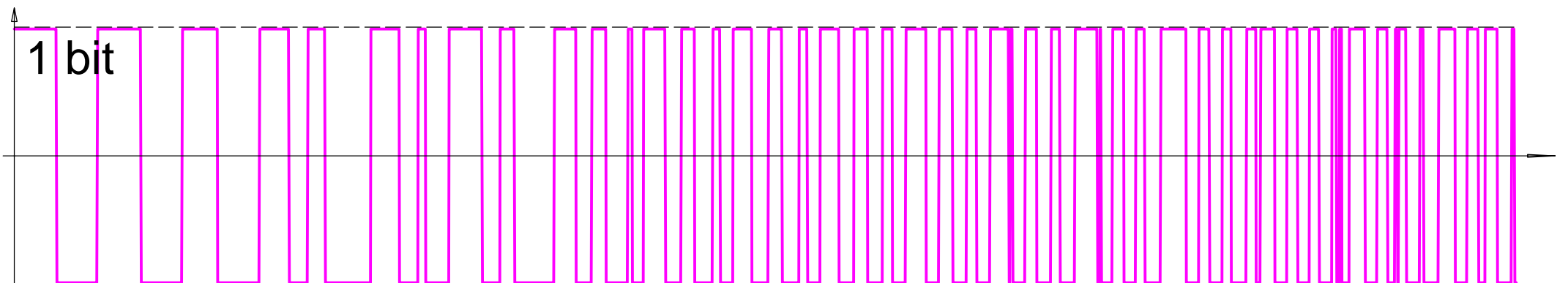
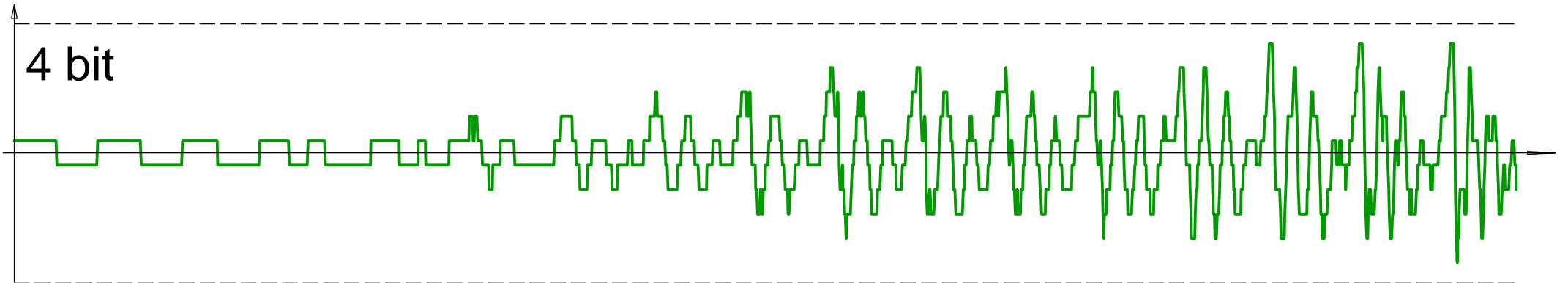
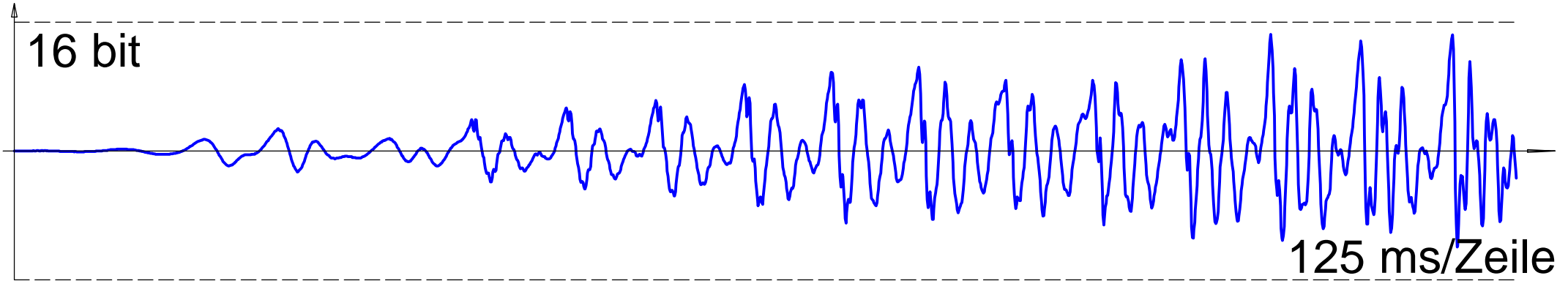
Signal	Wortlänge [bit]	Stufenzahl
Sprachsignal Telefonqualität	8	$2^8=256$
Sprachsignal guter Qualität Musiksignale, CD	16	65536
Audiosignale im Studio	20–24	>200000
Bildsignale (" <i>True Color</i> ")	24	$2^{24}$





# Quantisierung: Beispiel

16 bit  8 bit  4 bit  1 bit 





# Vom kontinuierlichen ("analogen") zum digitalen Signal [5]

## Darstellung als Zeitfunktion $s(t)$

- allgemein: Augenblicksamplitude
- in Luft: Schalldruck
- in Mikrofon und Verstärker: elektrische Spannung

## Darstellung als abgetastete Zeitfunktion

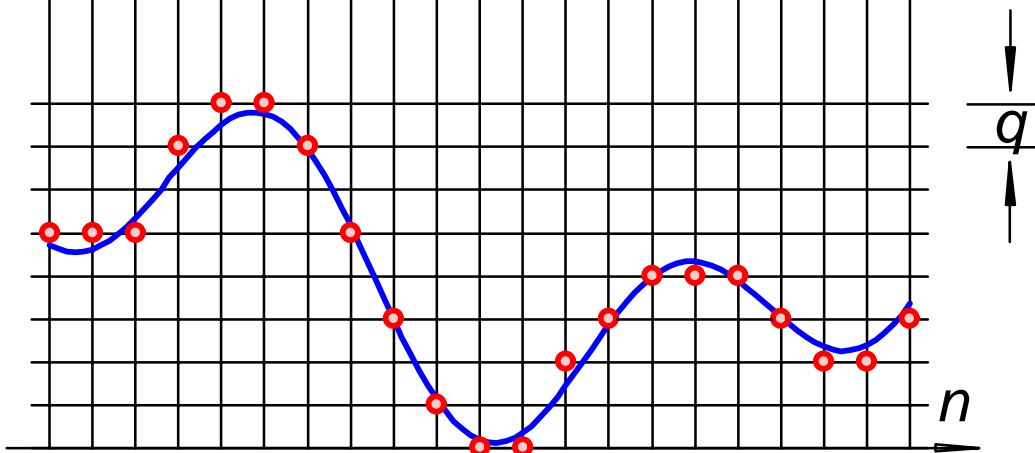
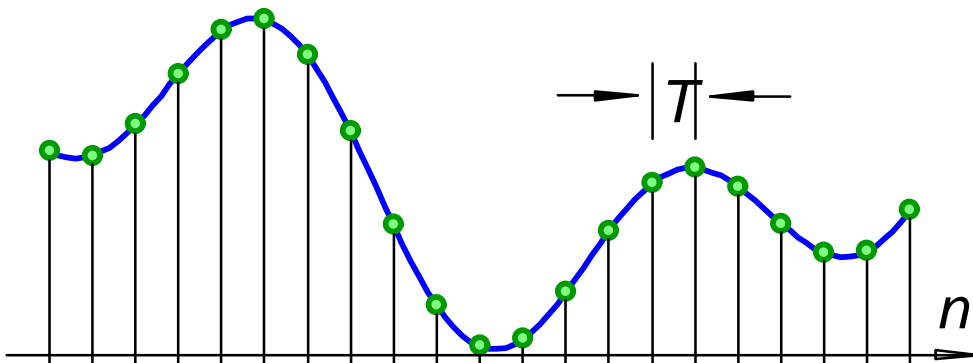
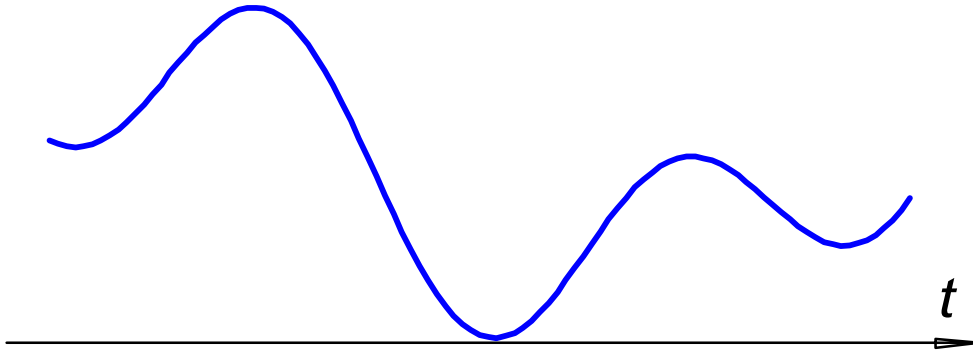
$$s(t) \rightarrow s(nT) \rightarrow s(n)$$

(bei gleichförmiger Abtastung)

## Darstellung im Rechner als Zahlenfolge

$$s(n) \rightarrow [s(n)]_Q \rightarrow k(n) \cdot q \rightarrow k(n)$$

(bei gleichförmiger Quantisierung)

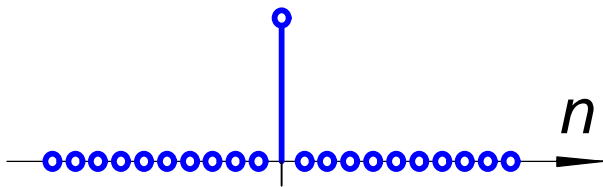




# Einige spezielle Signale

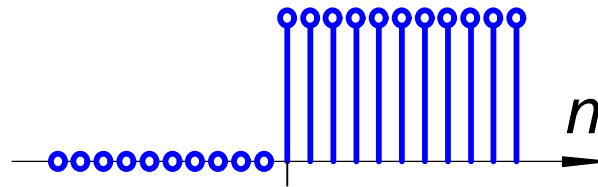
## Einheitsimpuls

$$\delta(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$



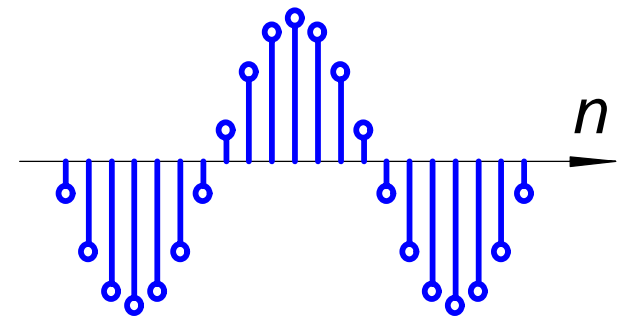
## Einheitssprung

$$s(n) = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases}$$



## Sinusförmiges Signal

$$x(n) = \cos(\Omega_0 n + \varphi_0)$$



# Nochmals zur Signaldarstellung

Die visuelle Darstellung digitaler Signale erfolgt in der Regel so, als seien sie analog.

Bei der theoretischen Betrachtung digitaler Signale kann das Abtasttheorem und damit die Tatsache der Abtastung nie vernachlässigt werden; die Abtastung führt dazu, dass für analoge und digitale Signale verschiedene Kalküle existieren, zwischen denen aber enge Beziehungen bestehen.

Demgegenüber wird die Quantisierung in Theorie und Visualisierung meist vernachlässigt, es sei denn, es geht um die Effekte der Quantisierung selbst.



# Grundlagen der Sprachsignalverarbeitung

## 1. Einführung

- 1.1 Grundbeobachtungen am Sprachsignal
- 1.2 Bau des menschlichen Sprechtrakts (kurze Einführung)
- 1.3 Signaldarstellung im Rechner
- 1.4 Der Informationsgehalt der Sprache**
- 1.5 Die wesentlichen Aufgaben der Sprachsignalverarbeitung
- 1.6 Phonetische Einheiten und die Sprachsignalverarbeitung

# Einige sprachliche Einheiten

Laut (Phonem)

Silbe

Wortteil: Stamm, Präfix, Suffix (Morphem)

Wort

Wortgruppe (Satzglied, Phrase)

Satz

Satzgruppe

. . . .



# Die verschiedenen Informationsebenen der Sprache

Informationsebene	Repräsentation
Akustik	Sprachsignal
Parameter	Quellen- und Vokaltraktparameter
(Akustik/)Phonetik	Realisationen von Phonemen ("Phone") oder anderer phonetischer Einheiten
(Phonetik/)Phonologie	Phoneme oder entsprechende elementare Einheiten (z.B. Silben)
Lexikalische Ebene	Morpheme oder Wörter bzw. deren Realisierung
Syntaktische Ebene	Wortgruppe und/oder Satz
Semantische Ebene	Inhalt und Bedeutung einer Aussage
Pragmatik	"Umfeld" oder "Weltmodell", in das die Aussage eingebettet ist.
Prosodie	Betonung, Sprechmelodie, Dauer (Rhythmus) und Intensität einzelner Laute, Silben etc.
Individuelle Merkmale	Klangfarbe der Stimme, Dialekt, Tonfall, Stimmung, Wortwahl und Sprechstil



# Informationsfluss in verschiedenen Repräsentationen

Inf.-Ebene	I.-Fluss (bit/s)	Repräsentation
Akustik	320000	HiFi-Sprachsignal (Abtastfrequenz 20 kHz, 16 bit/Abtastwert)
	64000	Sprachsignal in Telefon-PCM-Qualität (8 kHz, 8 bit/Abtastw.)
	2400	Untere Grenze für Repräsentation auf akustischer Ebene (mit aufwendigen Codierverfahren)
Parameter	4800	Parametrische Darstellung in guter Qualität oder gemischte Darstellung akustisch-parametrisch
	2400	Parametrische Darstellung in guter Qualität
	1200	Parametrische Darstellung in mäßiger Qualität
	400	Untergrenze parametrische Darstellung bei sehr grober Quantisierung
Akustik/Phonetik	100-500	Gemischte parametrisch-phonetische Darstellung (z.B. "Phonetischer Vocoder")
Phonetik/Phonologie	60	Darstellung als Phonemfolge (6 bit/Phonem; 10 Phone/s)
Höhere	10-20	Enthaltene Information, die zum Verstehen der Nachricht notwendig ist, unter Einbeziehung aller Einschränkungen linguistischer Art (Syntax, Semantik, Pragmatik).

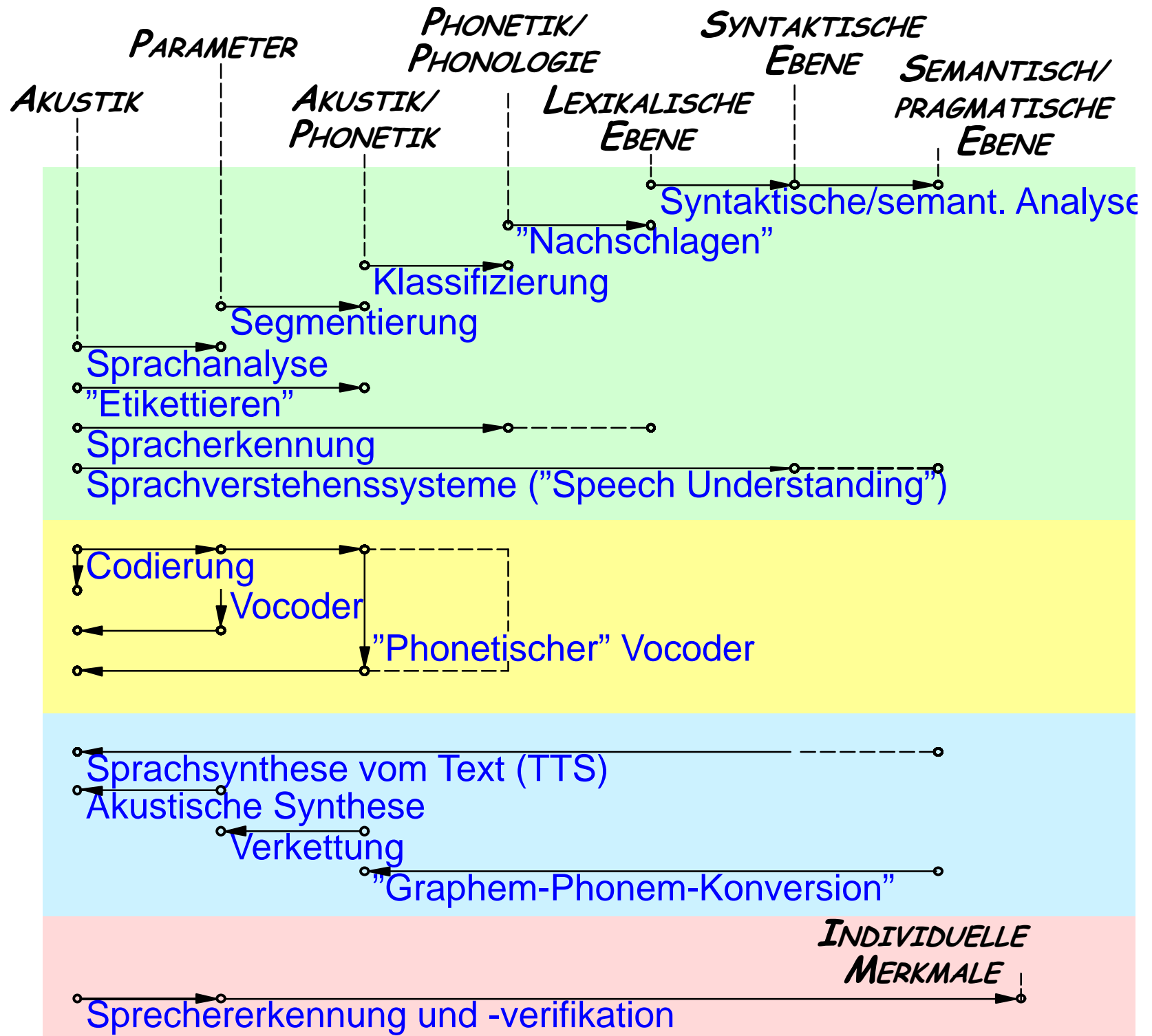




# Grundlagen der Sprachsignalverarbeitung

## 1. Einführung

- 1.1 Grundbeobachtungen am Sprachsignal
- 1.2 Bau des menschlichen Sprechtrakts (kurze Einführung)
- 1.3 Signaldarstellung im Rechner
- 1.4 Der Informationsgehalt der Sprache
- 1.5 **Die wesentlichen Aufgaben der Sprachsignalverarbeitung**
- 1.6 Phonetische Einheiten und die Sprachsignalverarbeitung



# Die wichtigsten Aufgaben der Sprach-(signal-)verarbeitung