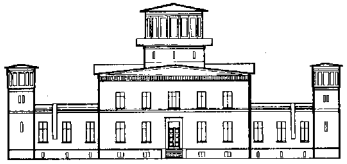
02.04.2002



# Grundlagen der Sprachsignalverarbeitung

Prof. Dr.-Ing. Wolfgang Hess

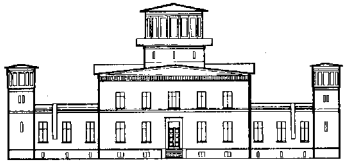
Institut für Kommunikationsforschung und Phonetik (IKP)
Universität Bonn
Poppelsdorfer Allee 47, 53115 Bonn

wgh@ikp.uni-bonn.de
http://www.ikp.uni-bonn.de

## Grundlagen der Sprachsignalverarbeitung

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

02.04.2002



# Fundamentals of Speech Signal Processing

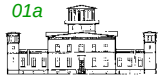
Prof. Dr.-Ing. Wolfgang Hess

Institute for Communication Research and Phonetics (IKP)
University of Bonn
Poppelsdorfer Allee 47, D-53115 Bonn, Germany

wgh@ikp.uni-bonn.de
http://www.ikp.uni-bonn.de

## Fundamentals of Speech Signal Processing

1. **Introduction**
2. Short-Time Analysis; Elemental Parameters
3. Short-Time Spectral Analysis
4. From Spectrum to Spectogram



# Literaturempfehlung

Vorlesungsskripten – deutsch – bilingual deutsch/englisch

Allgemein zur Sprachsignalverarbeitung

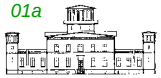
Rosen, Stuart / Howell, Peter (1991): *Signals and systems for speech and hearing* (Academic Press, London)
Rabiner, Lawrence R. / Schafer, Ronald W. (1978): *Digital processing of speech signals*. Prentice-Hall Signal Proc. Series, ed. by A.V.Oppenheim (Prentice Hall, Englewood Cliffs, NJ)
Vary, Peter / Heute, Ulrich / Hess, Wolfgang (1998): *Digitale Sprachsignalverarbeitung* (Teubner, Stuttgart) **(F)**

Speziell zu Kapitel 1

Meyer-Eppler, Werner (21969): 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





# Recommended Literature

Course material – German – bilingual German/English

Speech Signal Processing in General

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) **(A)**

For Chapter 1

Meyer-Eppler, Werner (21969): Grundlagen und Anwendungen der Informationstheorie. (2. Aufl. 1969; mit einer Erweiterung von G. Heike; 1. Aufl. 1959) (Springer, Berlin) **(A)**
Shannon, Claude E. (1949): "Communication in the presence of noise." Proc. IRE 37, 10-21 **(A)**

**(A)** – For advanced readers

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

Fundamentals of Speech Signal Processing

# 1. Introduction

1.1 **Basic Observations on the Speech Signal**

1.2 The Human Speech Organs (short Introduction)

1.3 Signal Representation in the Computer

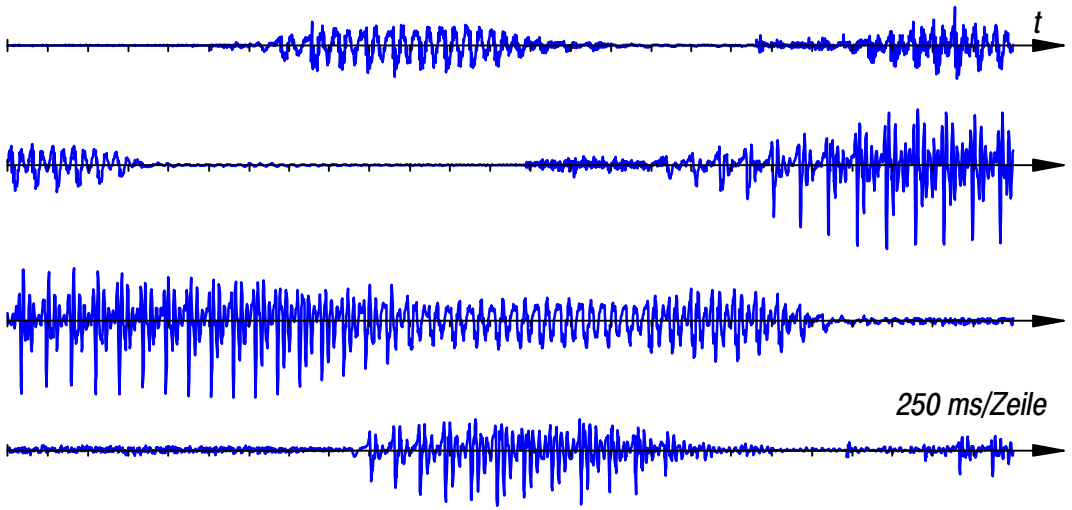
1.4 The Information Content of Speech

1.5 The Essential Tasks of Speech Signal Processing

1.6 Phonetic Units and Speech Signal Processing

004 gsv_1.1

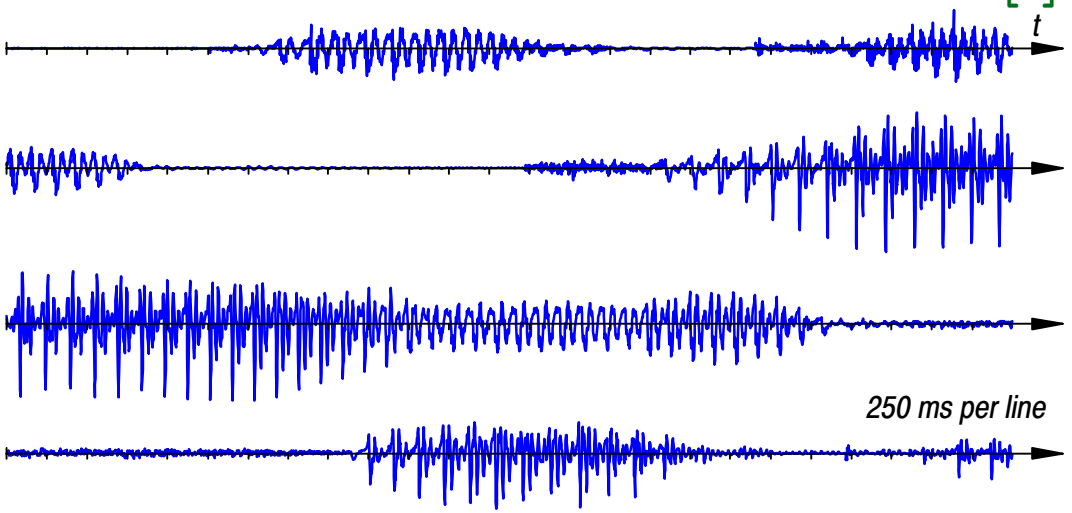
# Grundbeobachtungen am Sprachsignal [1]




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.

004 gsv_1.1

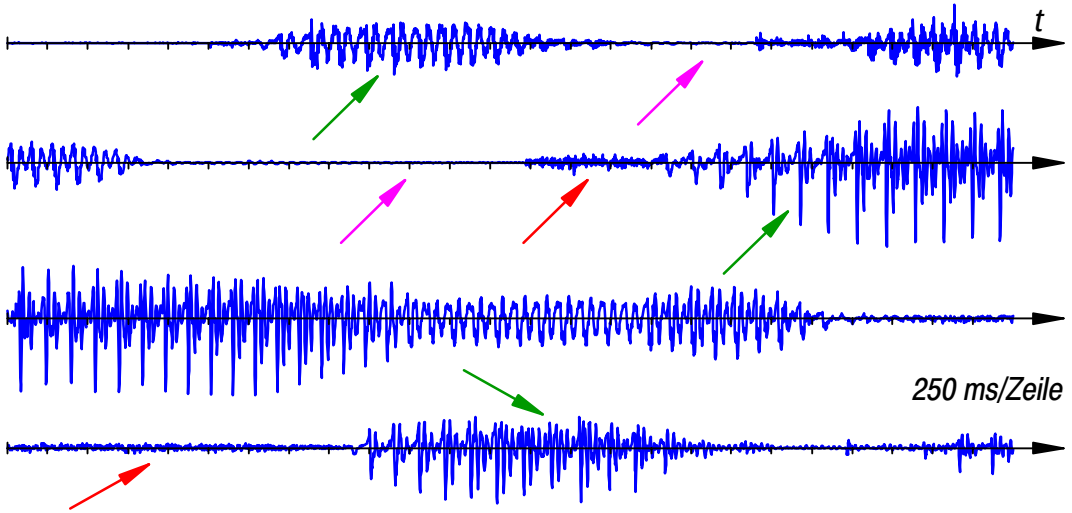
# Basic Observations on the Speech Signal [1]




The speech signal is time variant. In the way how the properties vary with time, the information is contained which makes speech signals to carriers of communication for human beings.

004 gsv_1.2

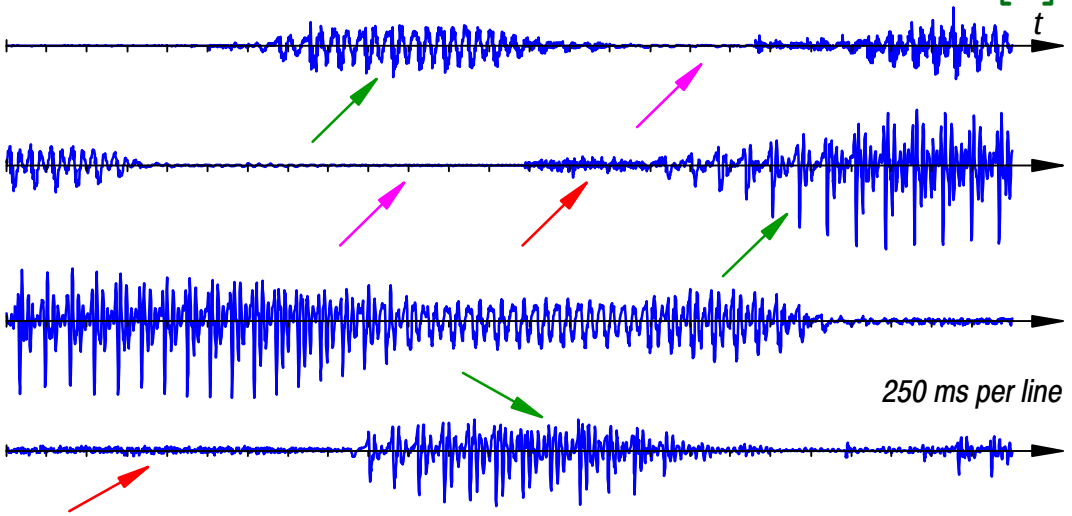
# Grundbeobachtungen am Sprachsignal [2]




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

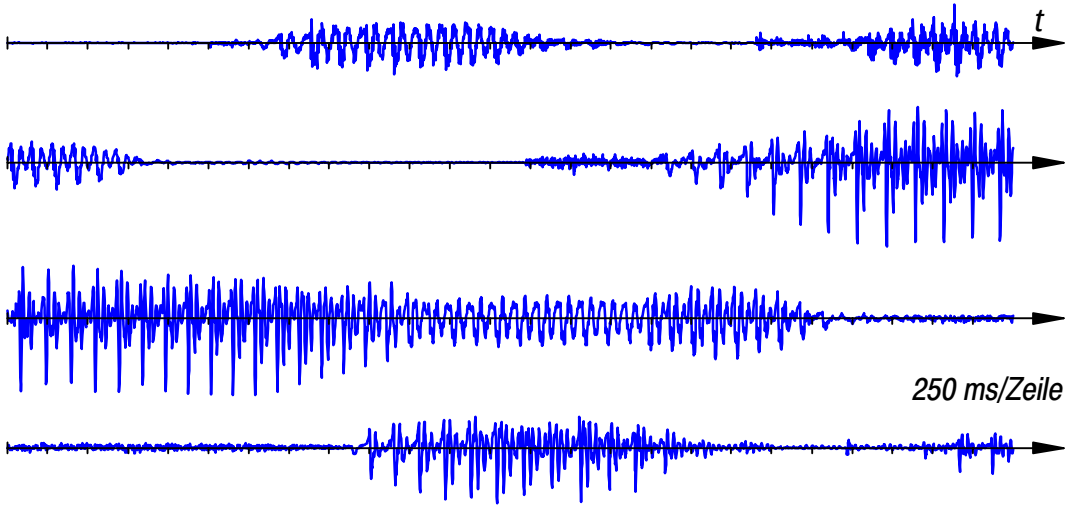
004 gsv_1.2

# Basic Observations on the Speech Signal [2]




The signal is (quasi-)periodic in individual segments.
In other segments it has the suructure of a noise.
Furthermore, there are pauses, some of which are very short.

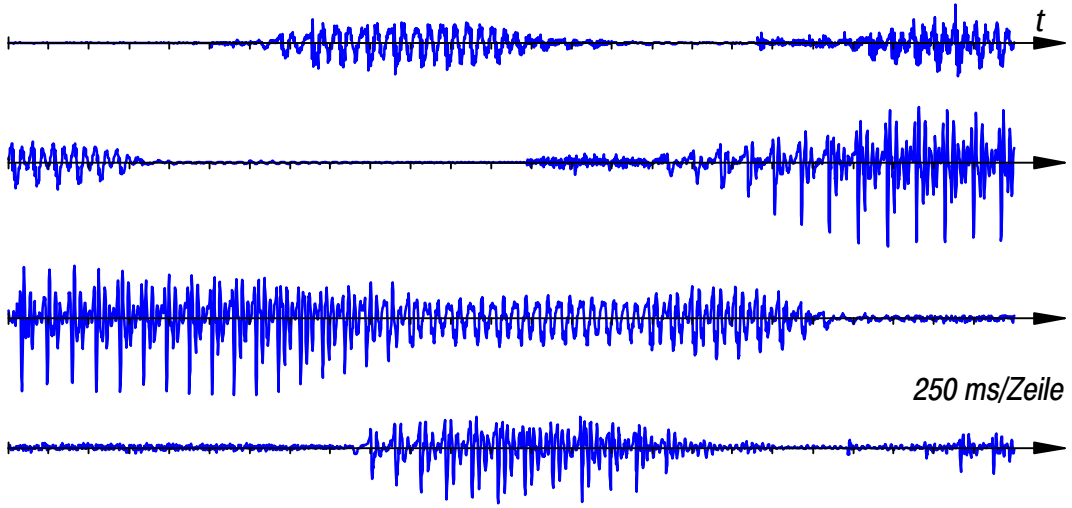
# Grundbeobachtungen am Sprachsignal [3]




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.



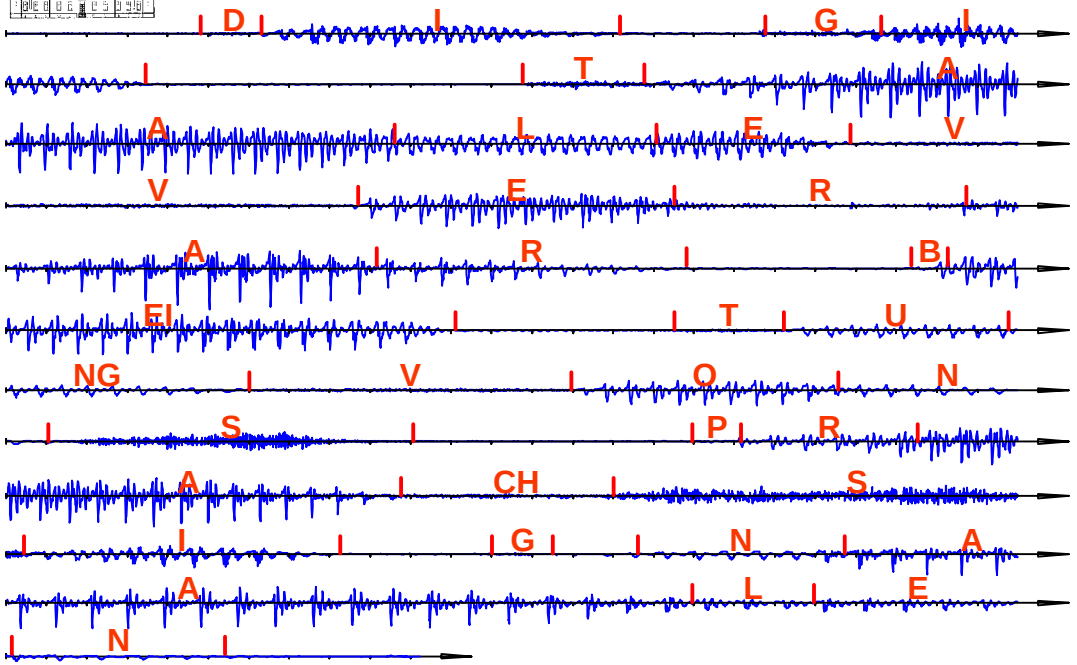
# Basic Observations on the Speech Signal [3]




Compared to the written form, where we have a chain, i.e., a sequence of symbols, the speech signal represents a continuum, a stream of data. Most signal structures change smoothly, not abruptly.

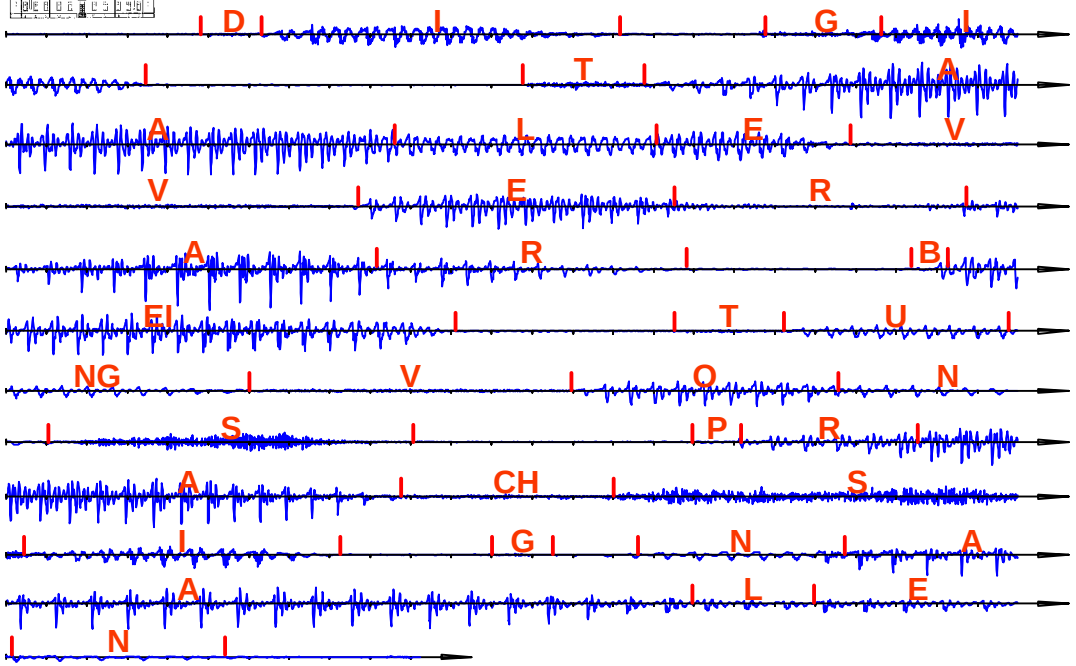


# Des Rätsels Lösung ...






# Solving the Puzzle ...



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

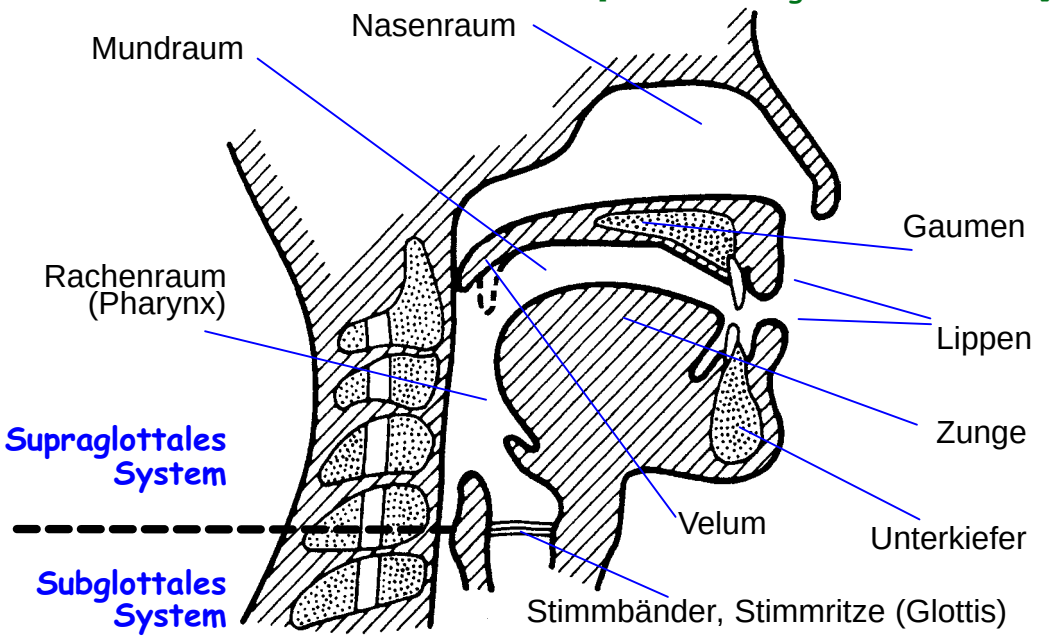
1.5 Die wesentlichen Aufgaben der Sprachsignalverarbeitung

1.6 Phonetische Einheiten und die Sprachsignalverarbeitung




## Die menschlichen Sprechorgane

[aus: Grundlagen der Phonetik]






Fundamentals of Speech Signal Processing

# 1. Introduction


1.1 Basic Observations on the Speech Signal

1.2 **The Human Speech Organs (short introduction)**

1.3 Signal Representation in the Computer

1.4 The Information Content of Speech

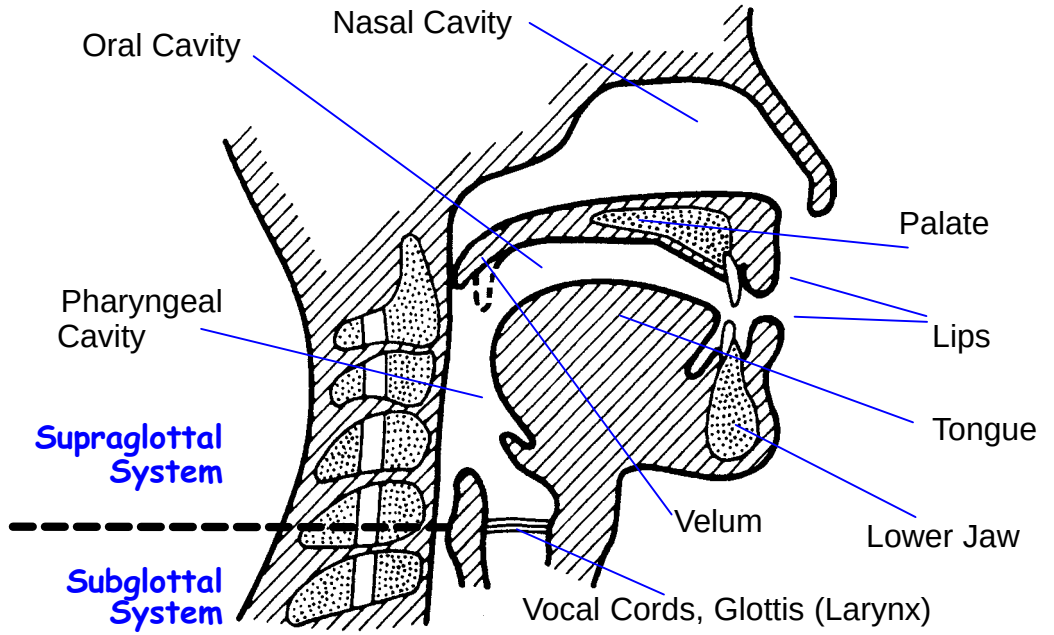
1.5 The Essential Tasks of Speech Signal Processing

1.6 Phonetic Units and Speech Signal Processing

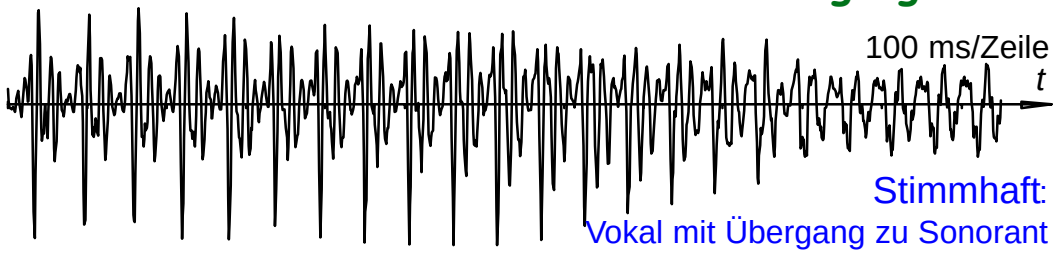



## The Human Speech Organs

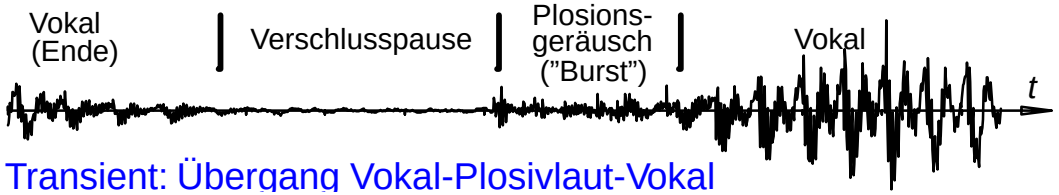
[from: Fundamentals of Phonetics]



## Sprachsignalbeispiele für die einzelnen Anregungsarten



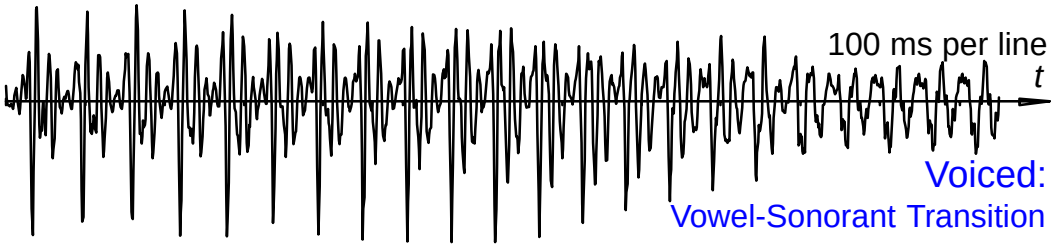

Stimmlos: Frikativ



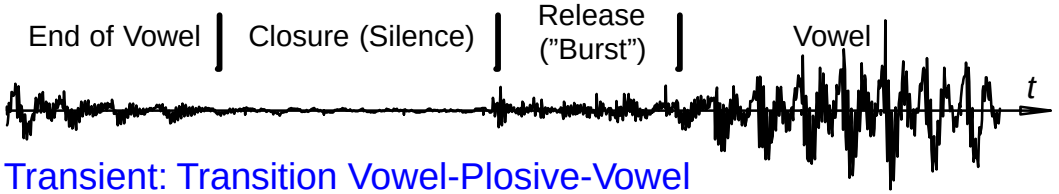

Transient: Übergang Vokal-Plosivlaut-Vokal



## Examples for the Individual Classes of Sounds




Voiceless: Fricative




Transient: Transition Vowel-Plosive-Vowel



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



Fundamentals of Speech Signal Processing

## 1. Introduction

1.1 Basic Observations on the Speech Signal

1.2 The Human Speech Organs (short introduction)

1.3 **Signal Representation in the Computer**

1.4 The Information Content of Speech

1.5 The Essential Tasks of Speech Signal Processing

1.6 Phonetic Units and Speech Signal Processing

## 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.



## Signals: Definition [1]

A signal is "the visible image of physical information" (Kunt, 1980). A signal can, for instance, be described through

- a mathematical function in the strict sense, and hence, an analytic correlation (in closed form), or
- a distributional law (e.g. for a stochastic signal). Here, the momentary value of the signal is unknown, since the distributional law describes only global signal properties, or
- empirically through a series of measurements.

In practice, most signals belong to this latter category.



## 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 Augenblicksamplitude, der Momentanwert oder einfach der Wert des Signals zu einem bestimmten Zeitpunkt $t$.



## Signals: Definition [2]

If a signal is a function of only one variable, it is called a one-dimensional signal, otherwise, it is called a multi-dimensional signal.

Examples for one-dimensional signals:

- speech signals
- any acoustic signal (music, noise, etc.)
- biological signals (electrocardiogram, electroencephalogram)
- physical signals (e.g., seismic waves)
- radar signals.

Examples for multi-dimensional signals:

- (two-dimensional) image signals

One-dimensional signals are usually encountered as time functions and designated $s(t)$; the value $s(t)$ is the instantaneous or momentary amplitude, the instantaneous value or simply the value of the signal at a given point $t$ in time.

# 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**)



# Signal Representation in the Computer

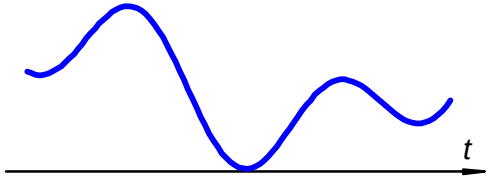
The representation of a digital signal in the computer, like anything else, is possible only as a sequence of numbers. All signals stored in a computer are thus discrete.

The conversion of a signal into the discrete representation is called digitalization or quantization. It is done in two steps:

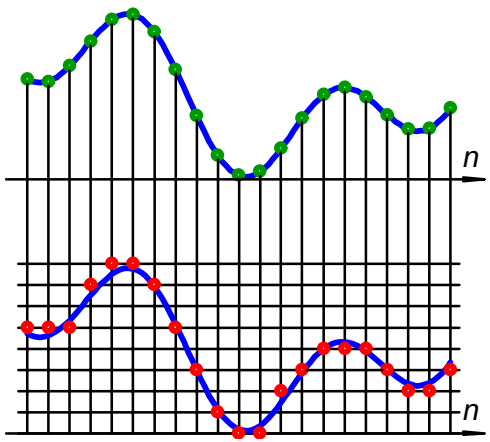
- Conversion of the independent variable into the discrete representation (**sampling**)
- Conversion of the dependent variable into the discrete representation (**quantization**)



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




kontinuierlich ("analog")

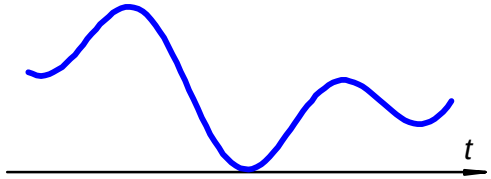



abgetastet

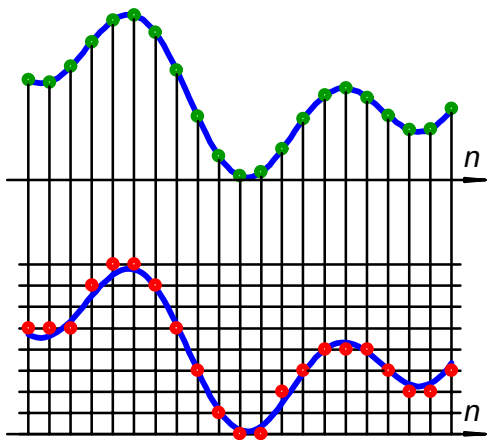
digital
(abgetastet und quantisiert)



# From Continuous ("Analog") to Digital Signals [1]



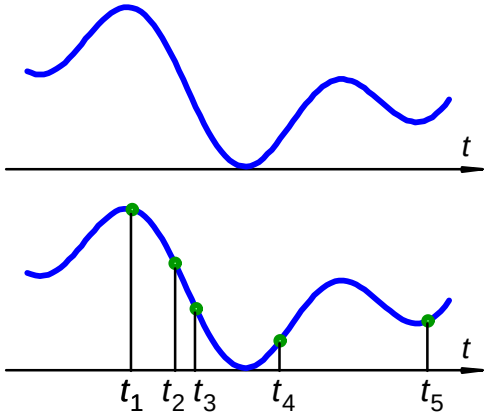

continuous ("analog")




sampled

digital
(sampled and quantized)

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




**Darstellung als Zeitfunktion** $s(t)$

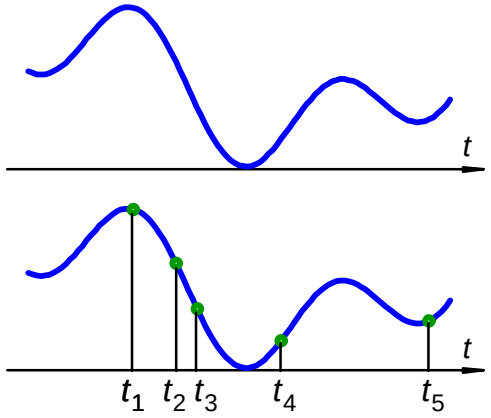
- allgemein: Augenblicksamplitude
- in Luft: Schalldruck
- in Mikrofon und Verstärker: elektrische Spannung

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

$$s(t) \rightarrow \{ \ldots s(t_1),\ s(t_2),\ \ldots \} \qquad (1.2)$$



# From Continuous ("Analog") to Digital Signals [2]




**Representation as time function** $s(t)$

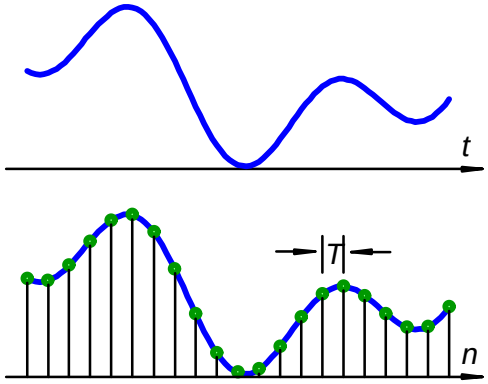
- general: instantaneous amplitude
- in air: sound pressure
- in microphone and amplifier: electric voltage

Quantizing the values of the independent variable results in *sampling* and splits the signal into a *sequence of samples*:

$$s(t) \rightarrow \{ \ldots s(t_1),\ s(t_2),\ \ldots \} \qquad (1.2)$$



# 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)

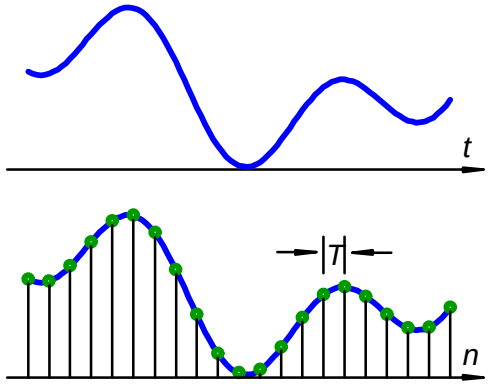
Vereinbarung: Die Abtastung erfolge *gleichförmig*, d.h., stets *in gleichen Intervallen der unabhängigen Variablen*:

$s(n) := s(nT)$ ; $n = -\infty \ldots -1,\ 0,\ 1 \ldots +\infty$ [ganzzahlig] (1.3)

$T$ : Abtastintervall



# From Continuous ("Analog") to Digital Signals [3]




Representation as time function $s(t)$

**Sampling**: representation as sampled time function

$s(t) \rightarrow s(nT) \rightarrow s(n)$
(*uniform* sampling)

Agreement: Sampling be *uniform*, i.e., always *in equal intervals of the independent variable*:

$s(n) := s(nT)$ ; $n = -\infty \ldots -1,\ 0,\ 1 \ldots +\infty$ [integer] (1.3)

$T$: sampling interval

## Abtastintervall und Abtastfrequenz

$s(n) := s(nT)$ ; $n = -\infty \ldots -1,\ 0,\ 1 \ldots +\infty$ [ganzzahlig]
$T$ : Abtastintervall (1.3)

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:

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



## Sampling Interval and Sampling Frequency

$s(n) := s(nT)$ ; $n = -\infty \ldots -1,\ 0,\ 1 \ldots +\infty$ [integer]
$T$: sampling interval (1.3)

The sampling interval $T$ is usually left out in the designation $s(n)$ for the signal.

Consequently, $n$ is now only indirectly a measurement of time; therefore, $n$ shall be designated as index oder address of the sample.

The reciprocal of the sampling interval is called sampling frequency or sampling rate, respectively:

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



## Abtasttheorem

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.



## Sampling Theorem

Which values must be assumed by the sampling interval $T$ or the sampling frequency $F$ for a given signal is expressed by the *sampling theorem*

> The sampling rate $F$ must be higher than twice the highest frequency contained in the signal

Violation of this rule leads to gross distortion, which however affects only those portions of the signal that violate the sampling theorem.

If the sampling theorem is fulfilled, the analog signal can be reconstructed from the digital representation without any errors - except for quantization effects.

# Gängige Werte der Abtastfrequenz für Audiosignale

| Signal | Frequenzen bis [kHz] | Abtastfrequenz [kHz] |
|---|---|---|
| Sprachsignal guter Qualität | 7.5<br>10 | 16<br>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 |

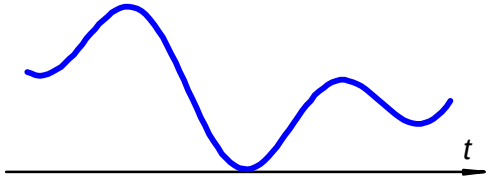


# Common Sampling Rates for Audio Signals

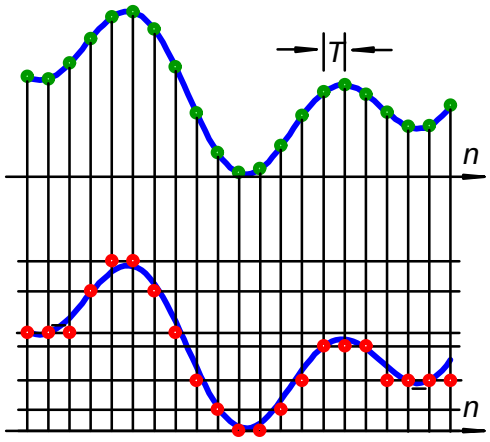
| Signal | Frequencies up to [kHz] | Sampling Rate [kHz] |
|---|---|---|
| Good-quality speech signal | 7.5<br>10 | 16<br>22.05 |
| Moderate-quality speech signal | 5 | 11.025 |
| Telephone quality | 3.4 | 8 |
| Hi-Fi music signal (minimum) | 15 | 32 |
| CD | 16 | 44.1 |
| other digital recording devices (DAT) | 16 | 48 |



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




Darstellung als Zeitfunktion $s(t)$



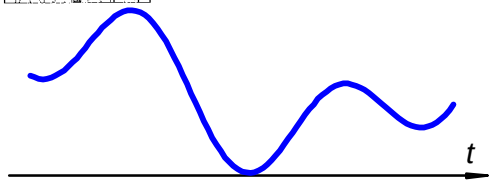

Darstellung als abgetastete Zeitfunktion $s(n)$

**Quantisierung:**
Darstellung im Rechner als Zahlenfolge

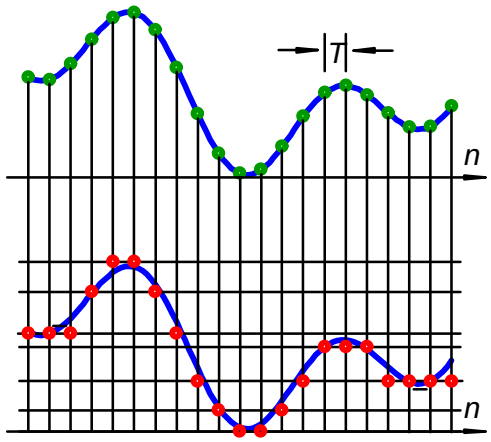
$s(n) \rightarrow [s(n)]_Q$



# From Continuous ("Analog") to Digital Signals [4]




Representation as time function $s(t)$



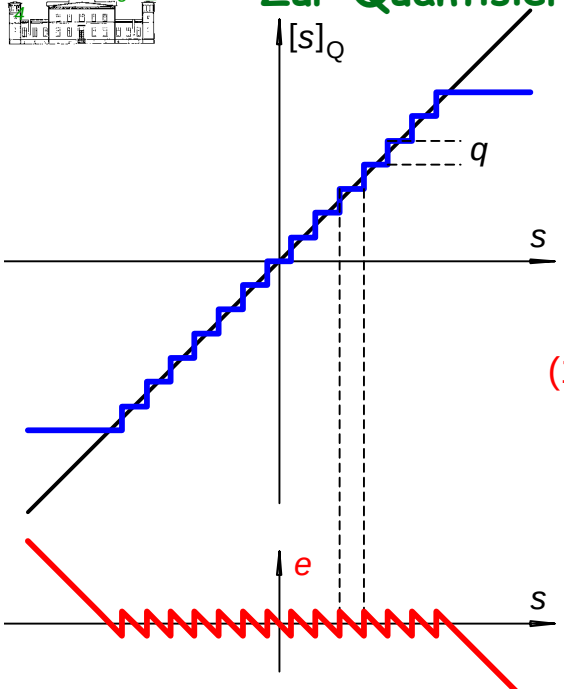

Representation as sampled time function $s(n)$

**Quantization:**
Representation in the computer as a sequence of numbers

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

00 gsv_1.17

## Zur Quantisierung digitaler Signale [1]




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

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

$k$ ganzzahlig

$s$ Eingangssignal

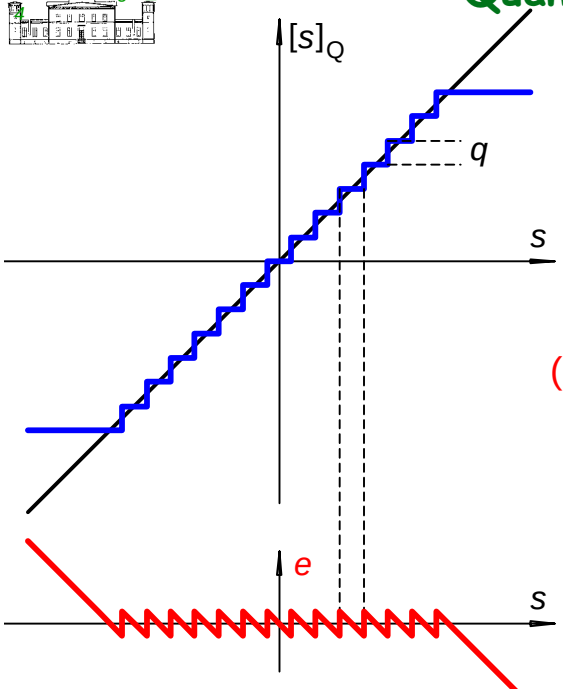
$[s]_Q$ Eingangssignal quantisiert

$q$ Quantisierungsstufe

$e$ Quantisierungsfehler

00 gsv_1.17

## Quantizing Digital Signals [1]




Agreement: The signal be quantized *uniformly*, i.e., always *in equal intervals of the dependent variable*:

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

$k$ integer

$s$ input signal

$[s]_Q$ input signal quantized

$q$ quantizing step

$e$ quantization error

00 gsv_1.18

## Zur Quantisierung digitaler Signale [2]

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<br>Musiksignale, CD | 16 | 65536 |
| Audiosignale im Studio | 20–24 | >200000 |
| Bildsignale ("*True Color*") | 24 | $2^{24}$ |

00 gsv_1.18

## Quantizing Digital Signals [2]

With respect to the quantizing step $q$ there is no strict rule comparable to the sampling theorem.

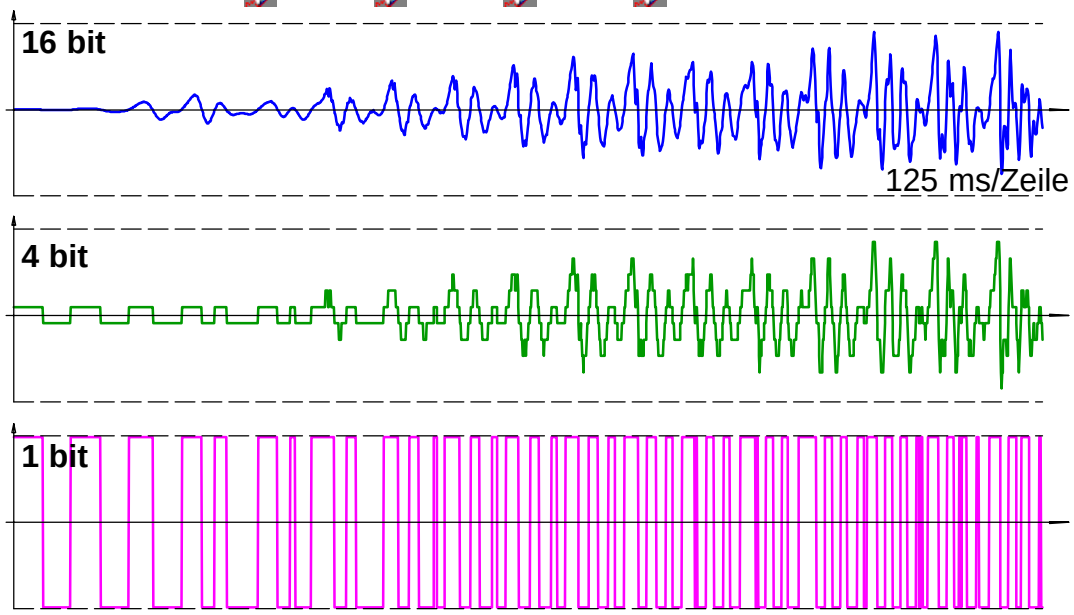
- If the quantizing step is too crude, this results in noise-like signal-correlated distortions known as quantization noise.

| Signal | Word length [bits] | Amplitude Levels |
|---|---|---|
| Telephone quality | 8 | $2^8$=256 |
| Good-quality speech signals<br>music signals, CD | 16 | 65536 |
| Audio signals in studio | 20–24 | >200000 |
| Image signals ("*True Color*") | 24 | $2^{24}$ |

01 gsv_1.19

# Quantisierung: Beispiel

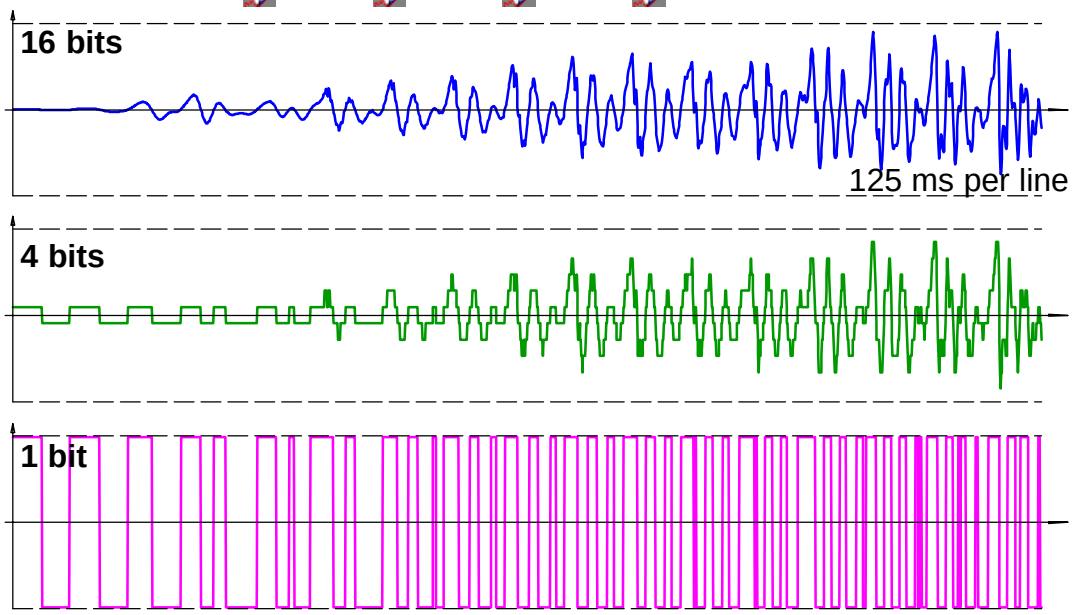
16 bit 8 bit 4 bit 1 bit




01 gsv_1.19

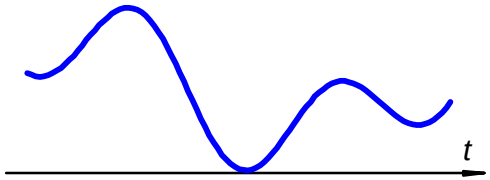
# Quantization: Example

16 bit 8 bit 4 bit 1 bit



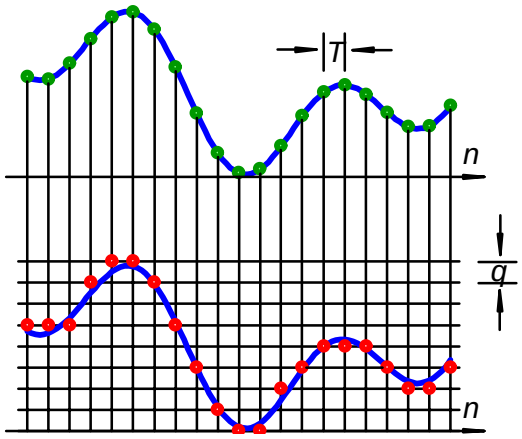

00 gsv_1.20

# 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)

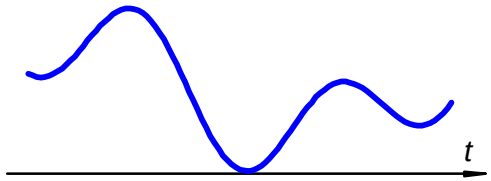
Darstellung im Rechner als Zahlenfolge

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

(bei *gleichförmiger* Quantisierung)

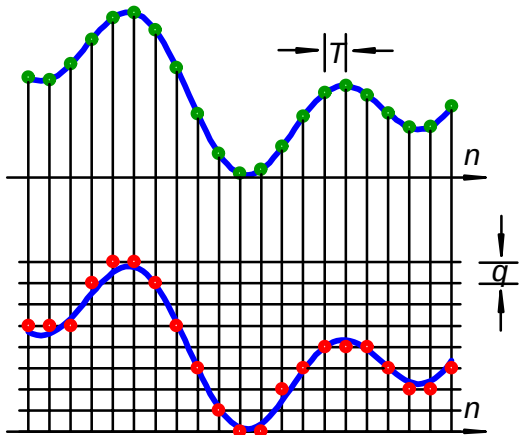
00 gsv_1.20

# From Continuous ("Analog") to Digital Signals [5]




Representation as time function $s(t)$

- general: instantaneous amplitude
- in air: sound pressure
- in microphone and amplifier: electric voltage




Representation as sampled time function

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

(*uniform* sampling)

Representation in the computer as a sequence of numbers

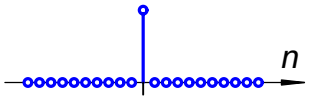
$s(n) \rightarrow [s(n)]_Q \rightarrow k(n) \cdot q \rightarrow k(n)$

(*uniform* quantization)

## Einige spezielle Signale

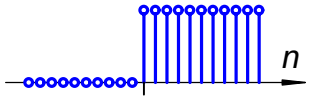
### Einheitsimpuls

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



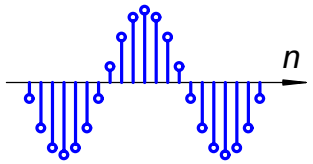

### Einheitssprung

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




### Sinusförmiges Signal

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

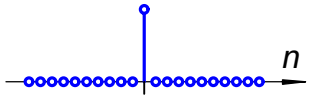





## Some Particular Signals

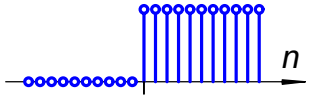
### Unit Impulse

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



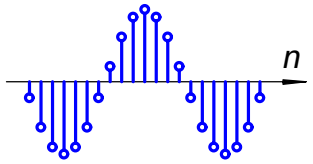

### Unit Step

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




### Sinusoid

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






## 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.



## Signal Representation Revisited

Digital signals are visually presented in a pseudo-analog way, i.e., as if they were analog.

In digital signal processing theory we can never ignore the sampling theorem and - with it - the fact that a digital signal is sampled; due to sampling there are different theoretical approaches for analog and digital signals; however, there are close relations between them.

On the other hand, quantization is largely ignored in theory and visualization except when a problem of quantization itself is dealt with.

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




Fundamentals of Speech Signal Processing

# 1. Introduction


1.1 Basic Observations on the Speech Signal

1.2 The Human Speech Organs (short introduction)

1.3 Signal Representation in the Computer

1.4 **The Information Content of Speech**

1.5 The Essential Tasks of Speech Signal Processing

1.6 Phonetic Units and Speech Signal Processing




## Einige sprachliche Einheiten

Laut (Phonem)
Silbe
Wortteil: Stamm, Präfix, Suffix (Morphem)
Wort
Wortgruppe (Satzglied, Phrase)
Satz
Satzgruppe
. . . .



## Some Units of Language

Sound (phoneme)
Syllable
Part of word: stem, prefix, suffix (morpheme)
Word
Group of words (constituent, phrase)
Sentence
Utterance
. . . .

## 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 |

## Language's Various Levels of Information

| Level | Representation |
|---|---|
| Acoustic | Speech signal |
| Parametric | Source and vocal-tract parameters |
| (Acoustic/)Phonetic | Realizations of phonemes (”phones”) or of other phonetic units |
| (Phonetics/)Phonology | Phonemes or corresponding elementary units (e.g., syllables) |
| Lexical level | Morphemes or words and their realizations |
| Syntactic level | Phrase and/or sentence |
| Semantic level | Contents and meaning of an utterance |
| Pragmatics | ”Environment” or ”world model” in which the utterance is embedded. |
| Prosody | Accenting, speech melody, duration (rhythm) and intensity of individual sounds, syllables etc. |
| Individual features | Voice timbre, dialect, intonation, speaker’s mood, wording, and speaking style |

## 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). |

## Information Flow in Different Representations

| Inf.Level | Flow (bits/s) | Representation |
|---|---|---|
| Acoustic | 320000 | Hi-Fi speech signal (sampling rate 20 kHz, 16 bits/sample) |
| | 64000 | Telephone PCM quality (sampling rate 8 kHz, 8 bits/sample) |
| | 2400 | Lower limit for representation on the acoustic level (with sophisticated coding technology) |
| Parametric | 4800 | Parametric representation in good quality or mixed (semiparametric) representation |
| | 2400 | Parametric representation with good quality |
| | 1200 | Parametric representation with moderate quality |
| | 400 | Lower limit for parametric representation with very crude quantization |
| Acoustic/phonetic | 100-500 | Mixed parametric/phonetic representation (e.g., “phonetic vocoder”) |
| Phonetics/Phonology | 60 | Sequence of phonemes (6 bits/phoneme; 10 phones/s) |
| Phonology and above | 10-20 | Information necessary for understanding the message when all linguistic and communicative constraints (syntax, semantics, pragmatics) are taken into account |

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




Fundamentals of Speech Signal Processing

# 1. Introduction


1.1 Basic Observations on the Speech Signal

1.2 The Human Speech Organs (short introduction)

1.3 Signal Representation in the Computer

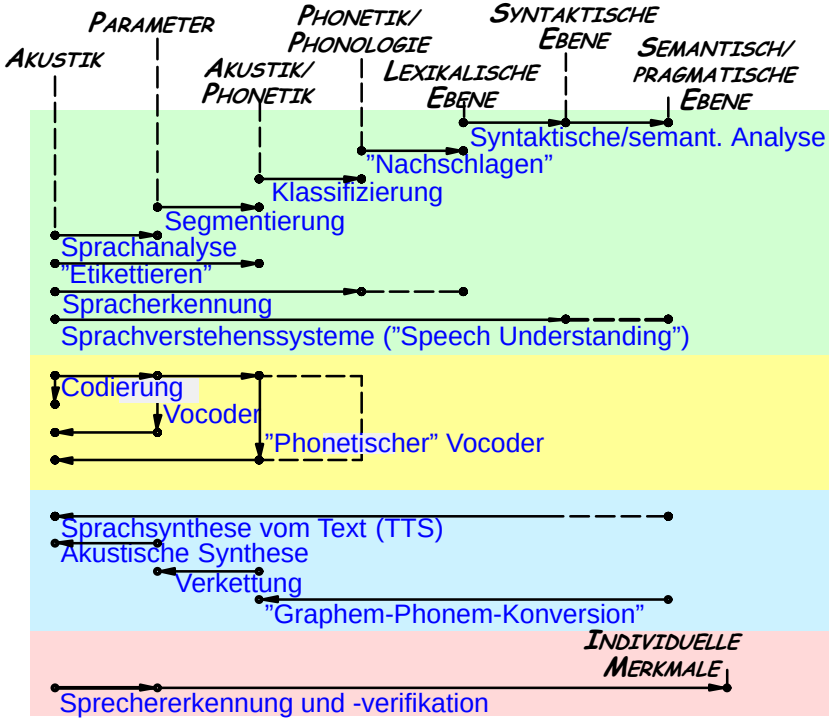
1.4 The Information Content of Speech

1.5 **The Essential Tasks of Speech Signal Processing**

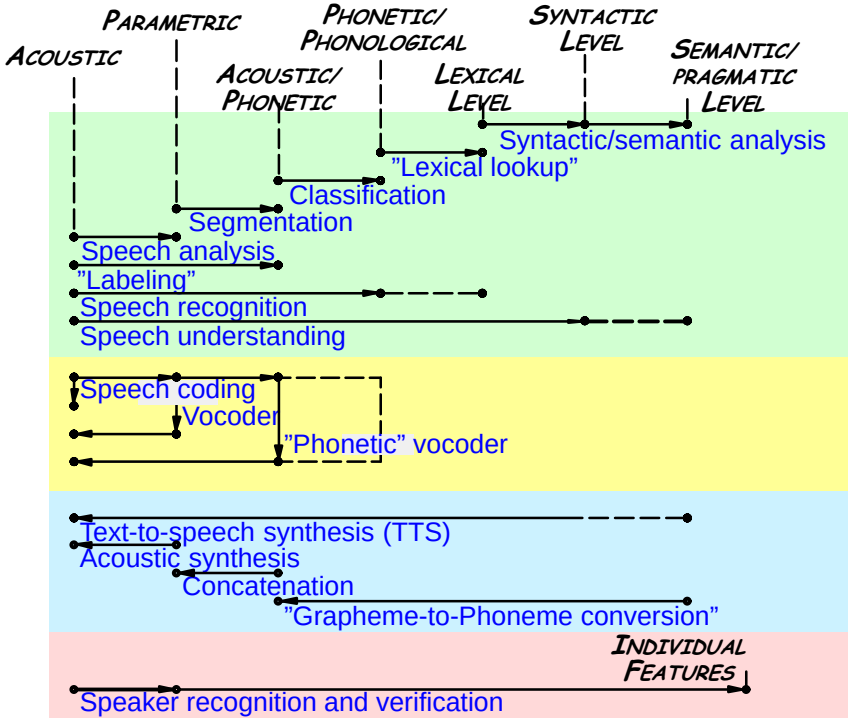
1.6 Phonetic Units and Speech Signal Processing




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






**The Essential Tasks of Speech (Signal) Processing**