

Grundlagen der Sprachsignalverarbeitung

1. Einführung

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

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 (weitere Quellenangaben)

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

Grundlagen der Sprachsignalverarbeitung

1. Einführung

1.1 Grundbeobachtungen am Sprachsignal

1.2 Bau des menschlichen Sprechtrakts (kurze Einführung)

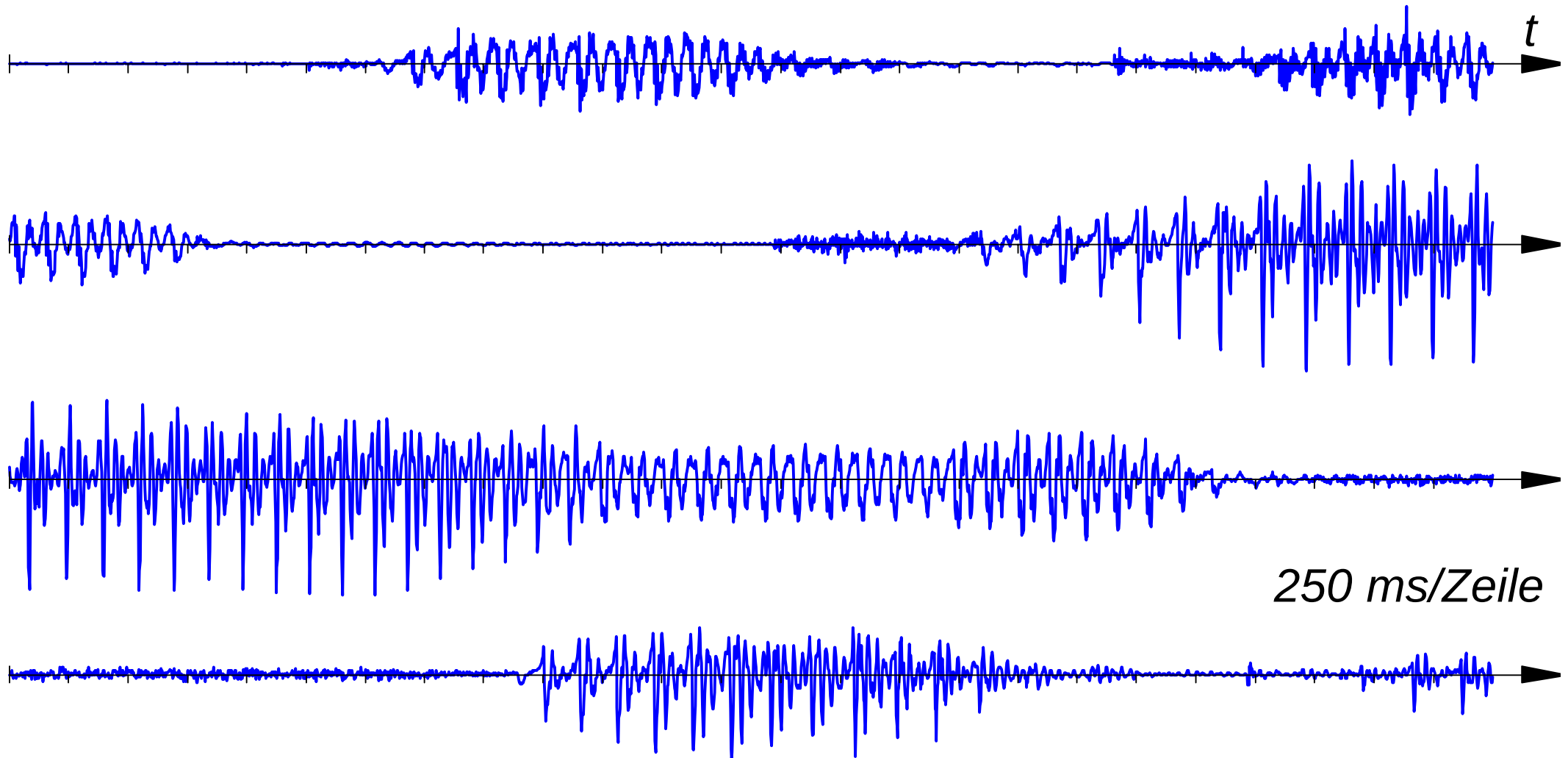
1.3 Signaldarstellung im Rechner

1.4 Der Informationsgehalt der Sprache

1.5 Die wesentlichen Aufgaben der Sprachsignalverarbeitung

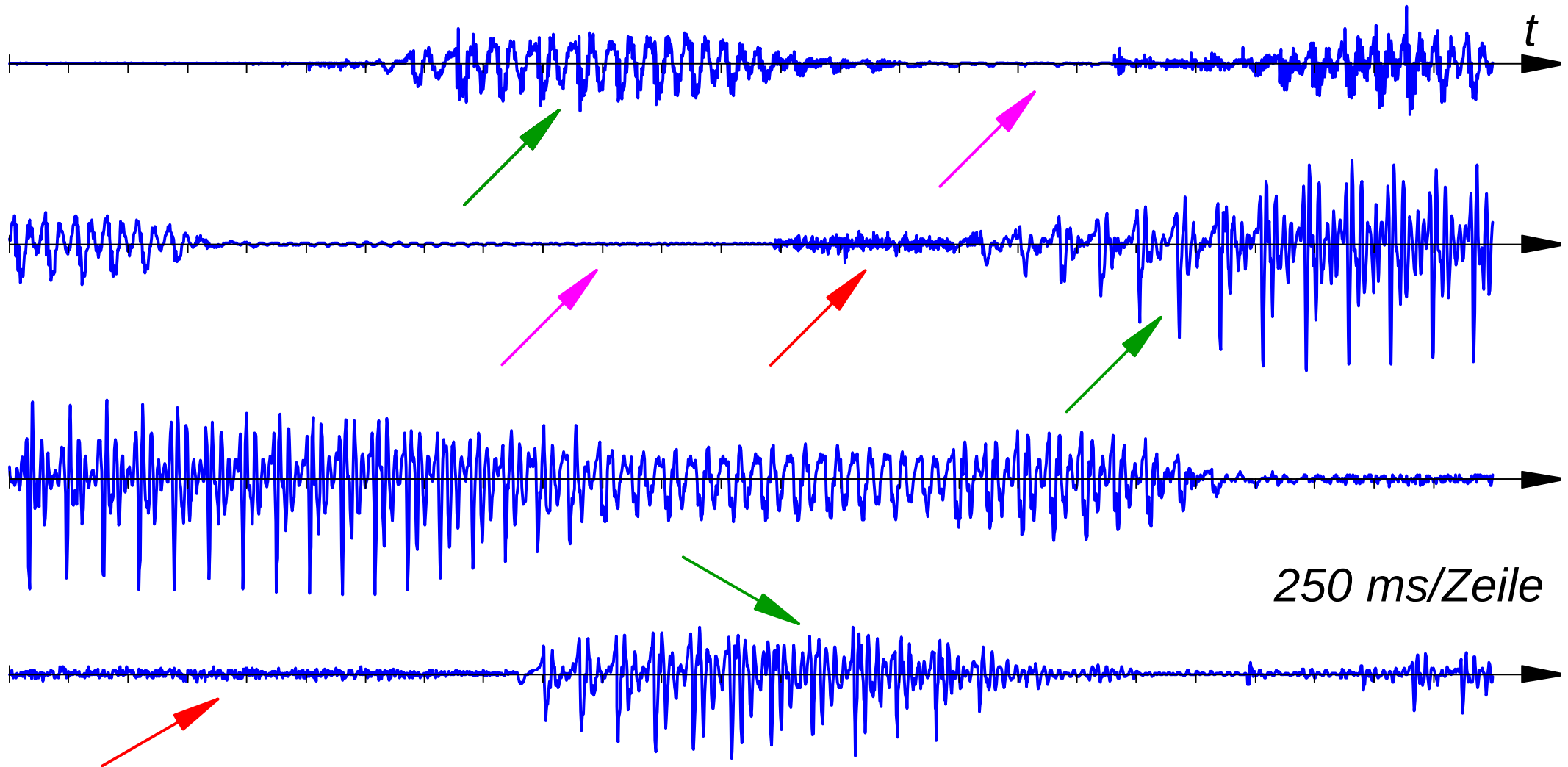
1.6 Phonetische Einheiten und die Sprachsignalverarbeitung

Grundbeobachtungen am Sprachsignal [1]



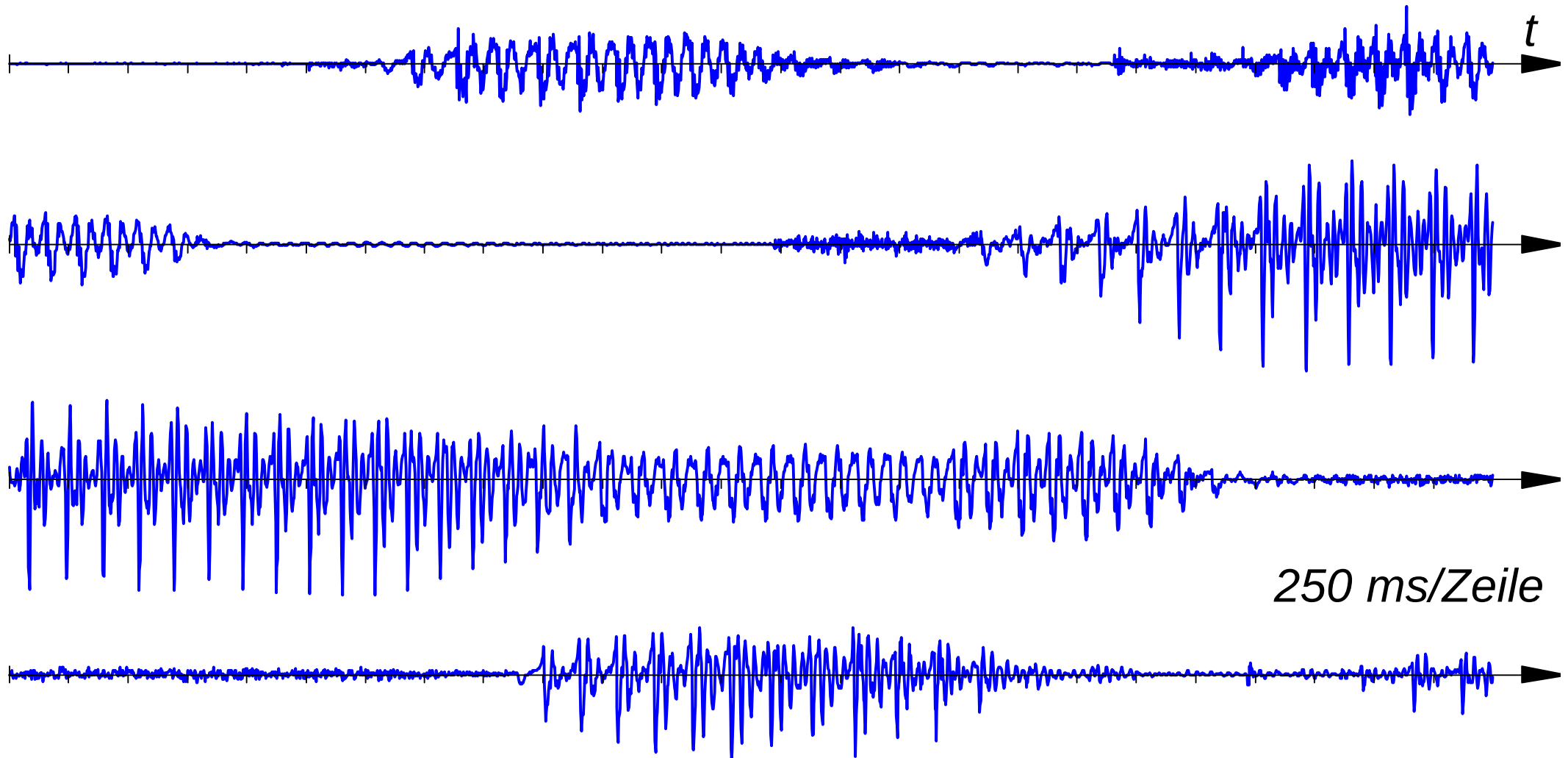
Das Sprachsignal ist **zeitveränderlich**. In der Art und Weise, wie die Zeitveränderung erfolgt, liegt die Information, die das Sprachsignal zum Kommunikationsträger für den Menschen macht.

Grundbeobachtungen am Sprachsignal [2]



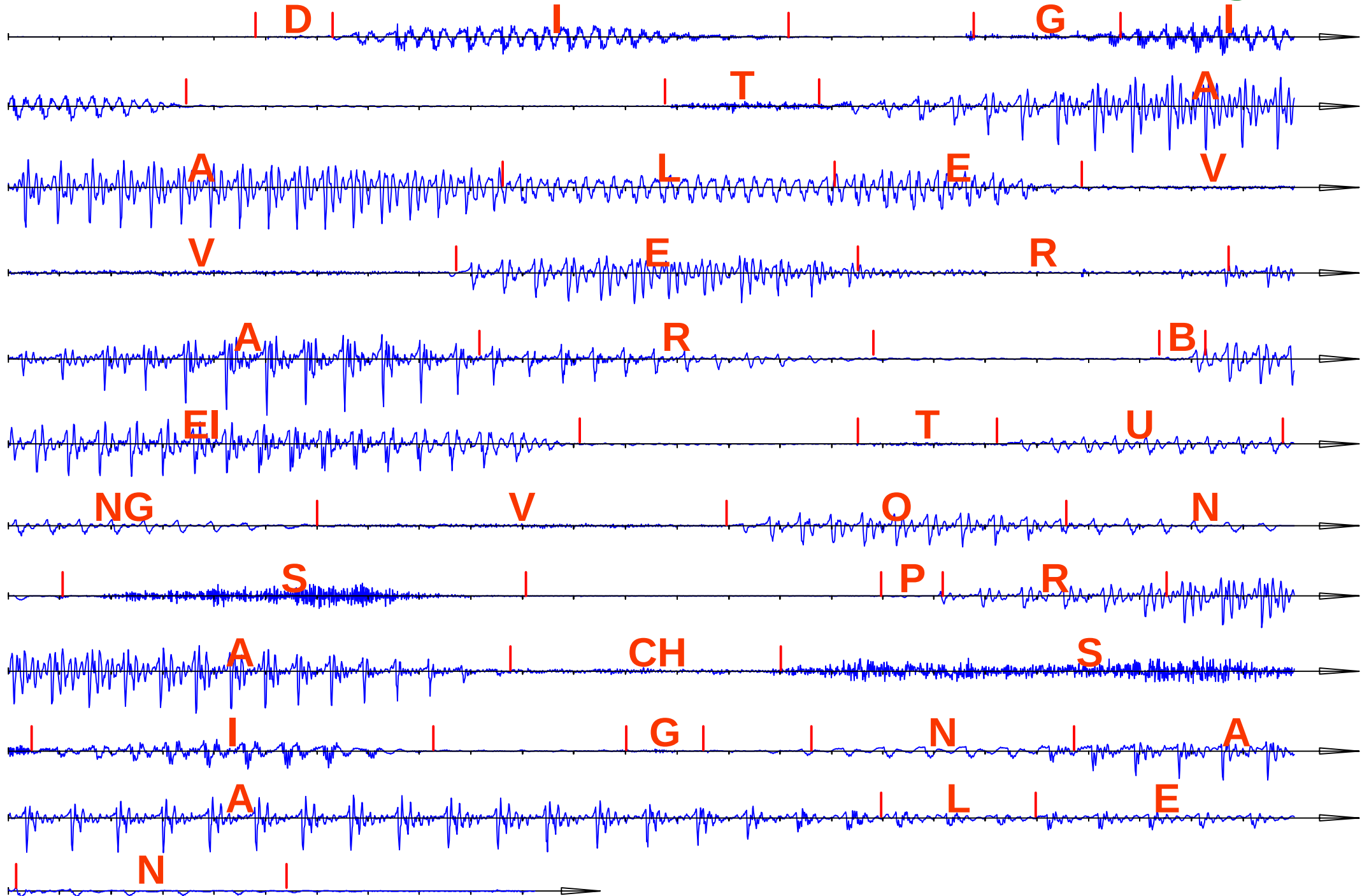
Das Sprachsignal ist in Abschnitten (quasi-)periodisch.
In anderen Abschnitten besitzt es die Struktur eines Rauschens.
Weiterhin existieren Pausen, die teilweise sehr kurz sind.

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.

Des Rätsels Lösung ...



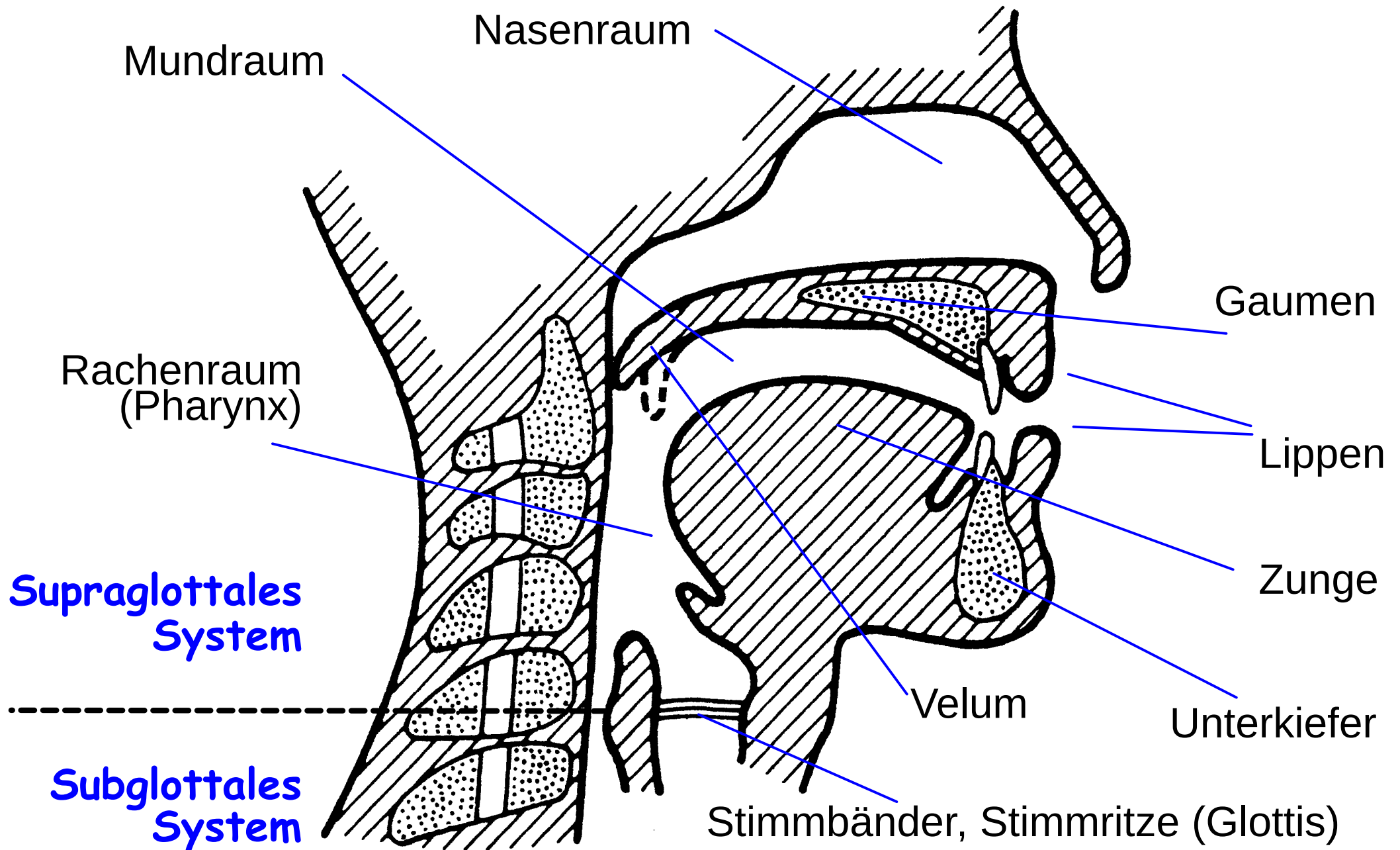
Grundlagen der Sprachsignalverarbeitung

1. Einführung

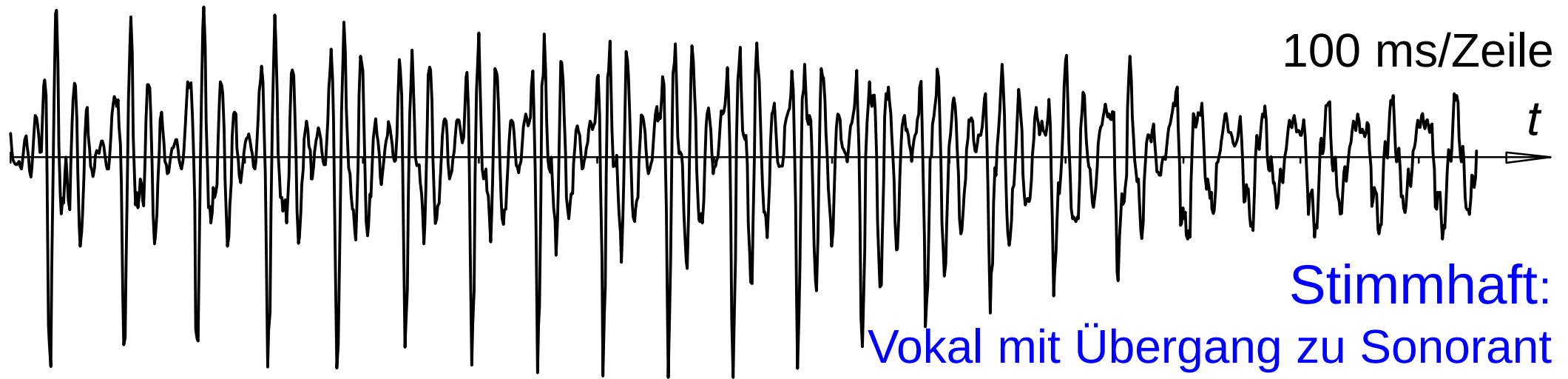
- 1.1 Grundbeobachtungen am Sprachsignal
- 1.2 **Bau des menschlichen Sprechtrakts (kurze Einführung)**
- 1.3 Signaldarstellung im Rechner
- 1.4 Der Informationsgehalt der Sprache
- 1.5 Die wesentlichen Aufgaben der Sprachsignalverarbeitung
- 1.6 Phonetische Einheiten und die Sprachsignalverarbeitung

Die menschlichen Sprechorgane

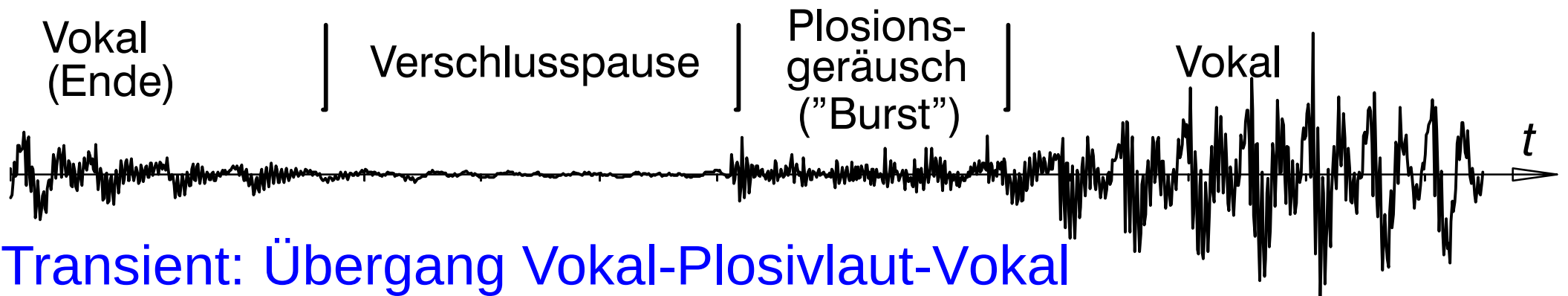
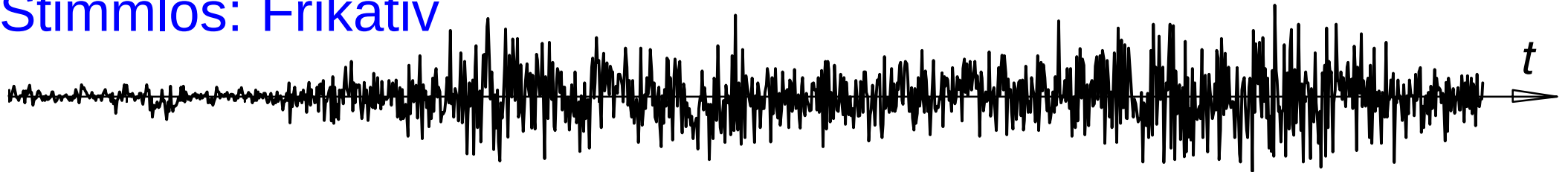
[aus: Grundlagen der Phonetik]



Sprachsignalbeispiele für die einzelnen Anregungsarten



Stimmlos: Frikativ



Grundlagen der Sprachsignalverarbeitung

1. Einführung

- 1.1 Grundbeobachtungen am Sprachsignal
- 1.2 Bau des menschlichen Sprechtrakts (kurze Einführung)
- 1.3 **Signalдарstellung 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 .

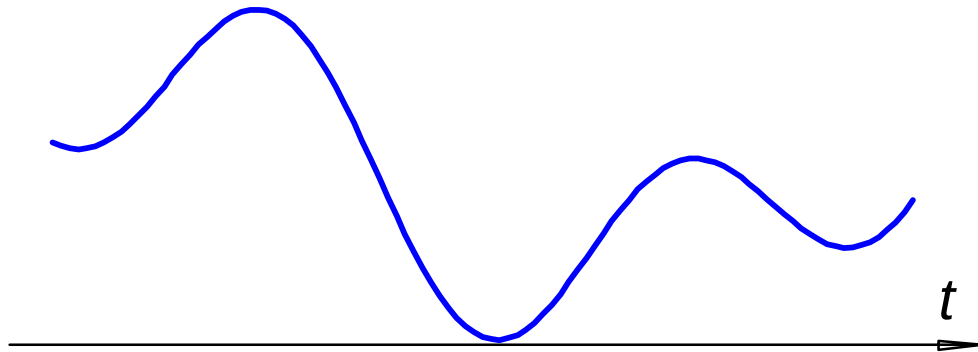
Signaldarstellung im Rechner

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

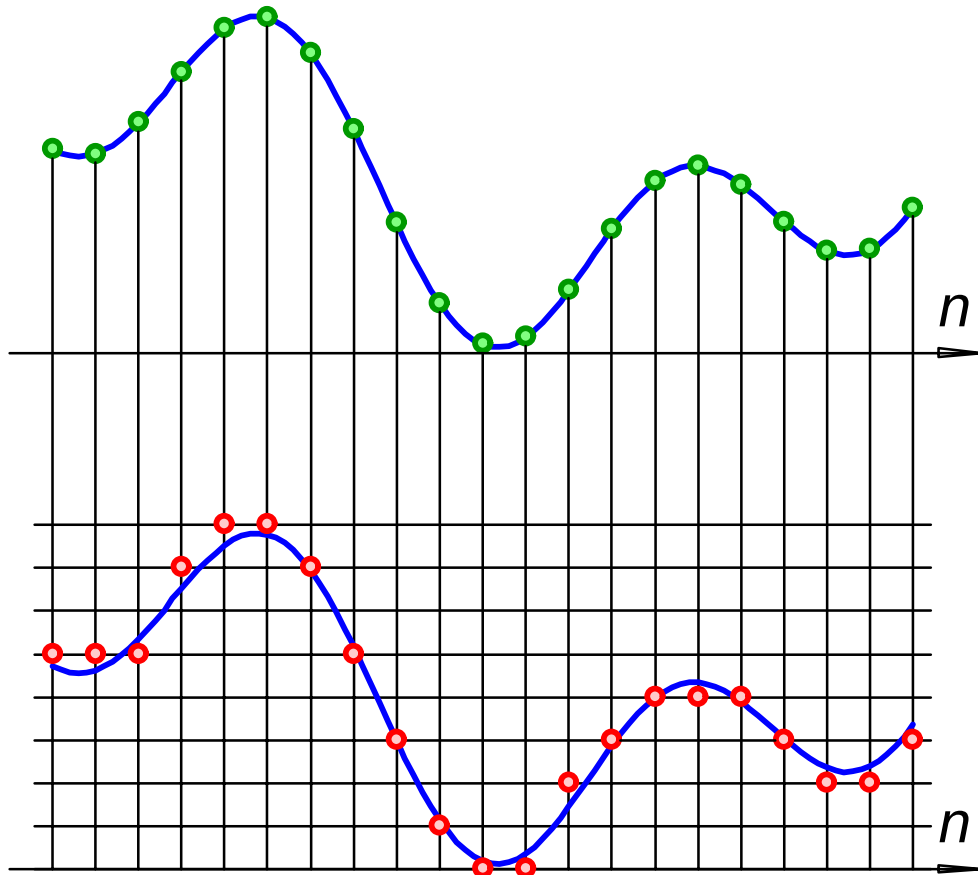
Die Umwandlung des Signals in die diskrete Darstellung bezeichnen wir als **Digitalisierung** oder **Quantisierung**. Sie läuft in 2 Schritten ab:

- 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]



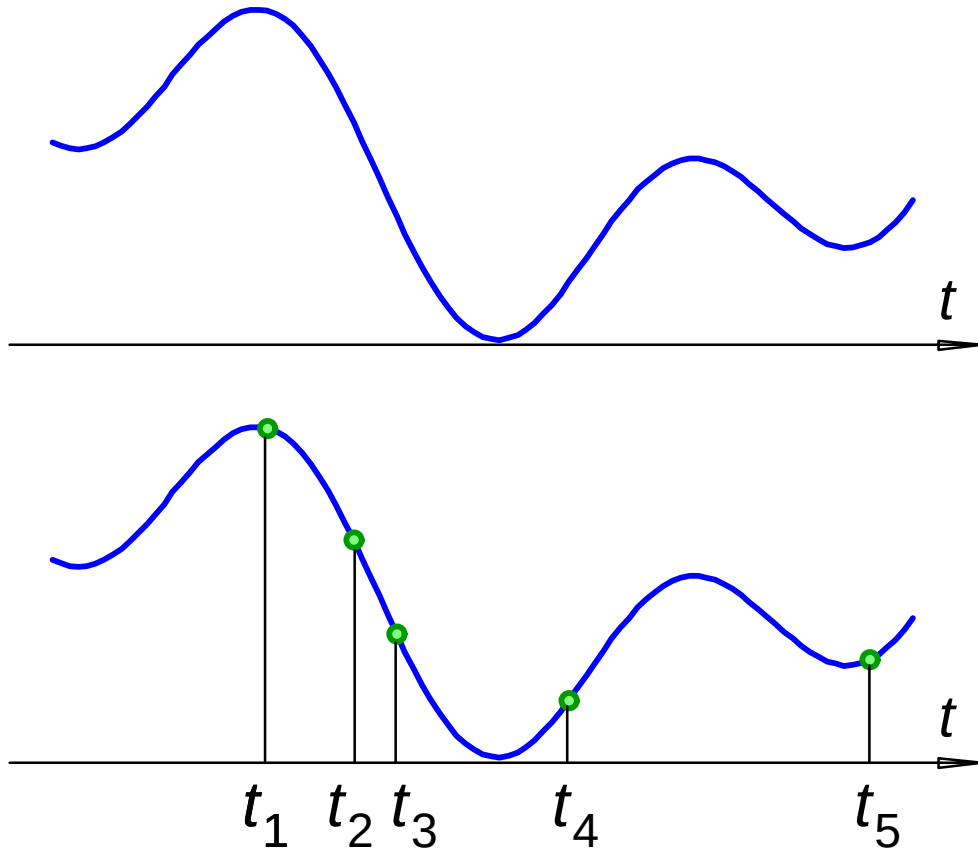
kontinuierlich ("analog")



abgetastet

digital
(abgetastet und quantisiert)

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



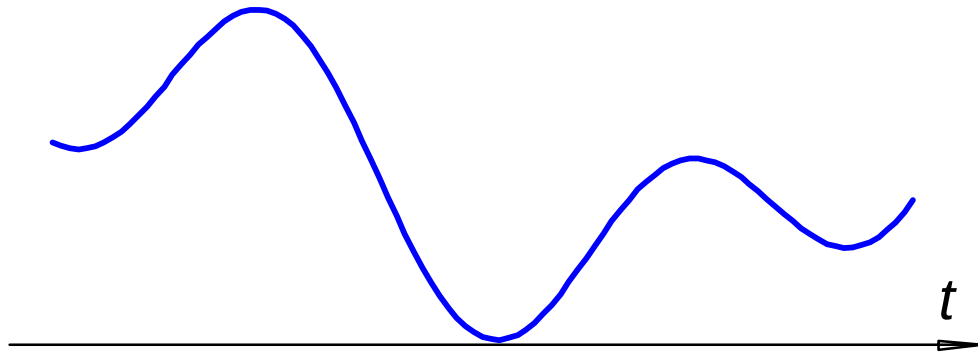
Darstellung als Zeitfunktion $s(t)$

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

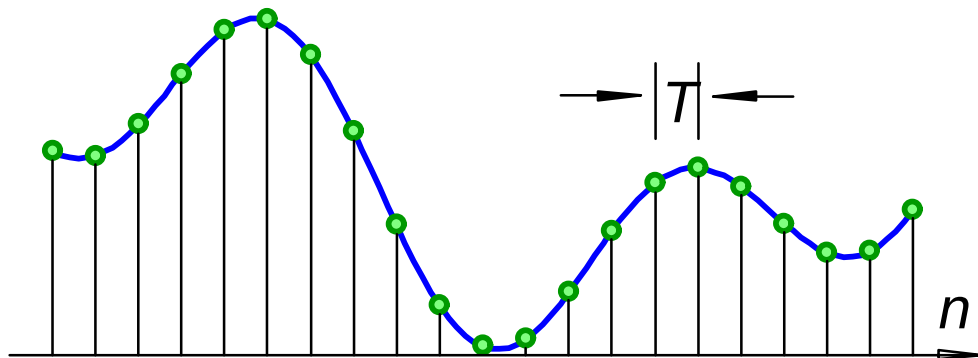
Quantisierung der unabhängigen Variablen ergibt **Abtastung** und zerlegt das Signal in eine **Folge von Abtastwerten**:

$$s(t) \rightarrow \{ \dots s(t_1), s(t_2), \dots \} \quad (1.2)$$

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:

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

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.

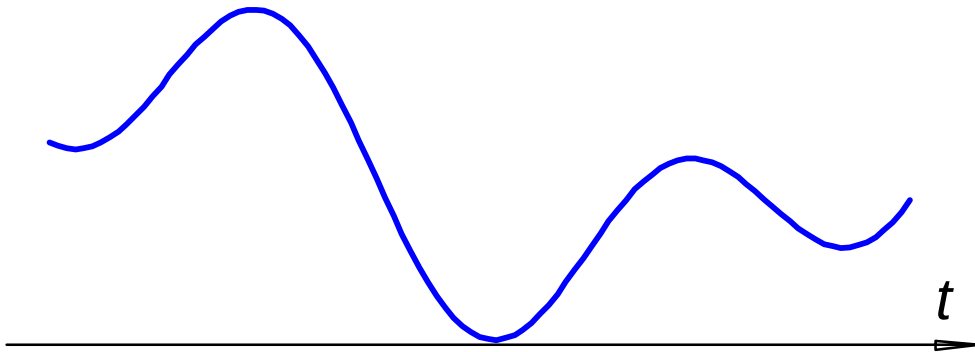
Missachtung dieser Regel führt zu groben Verzerrungen, die sich aber ausschließlich auf die Signalanteile auswirken, die das Abtasttheorem verletzen.

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

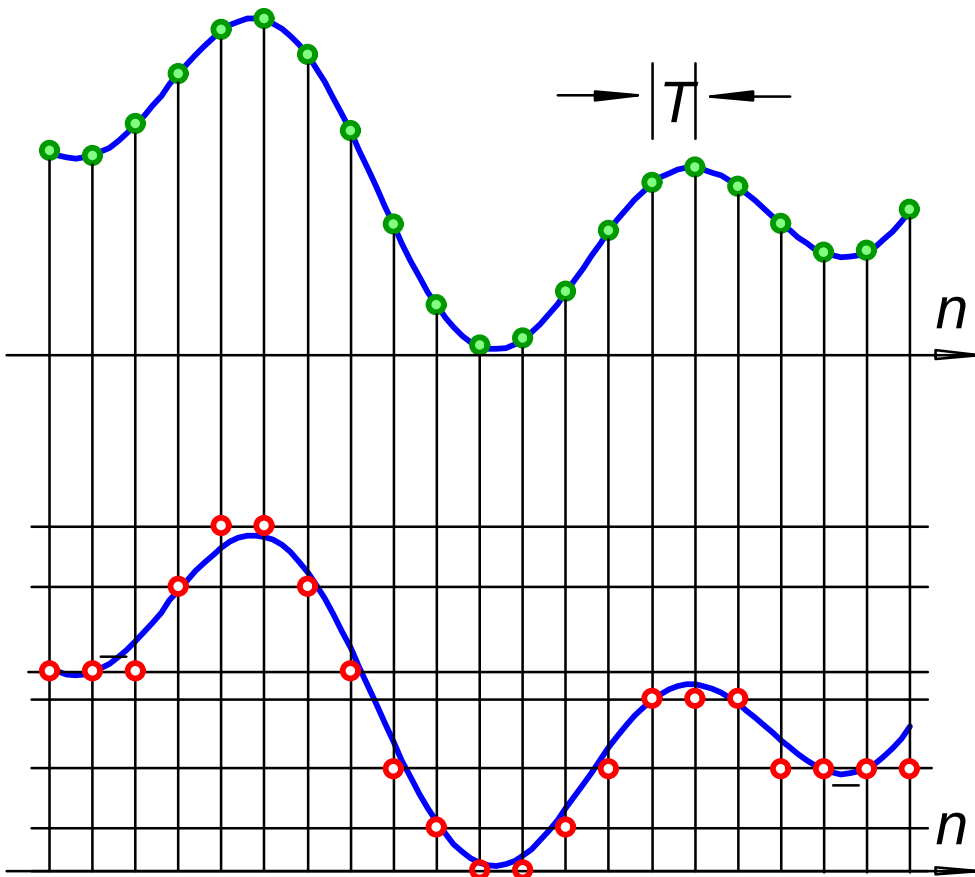
Gängige Werte der Abtastfrequenz für Audiosignale

Signal	Frequenzen bis [kHz]	Abtastfrequenz [kHz]
Sprachsignal guter Qualität	7.5	16
	10	22.05
Sprachsignal mindererer 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 $s(t)$

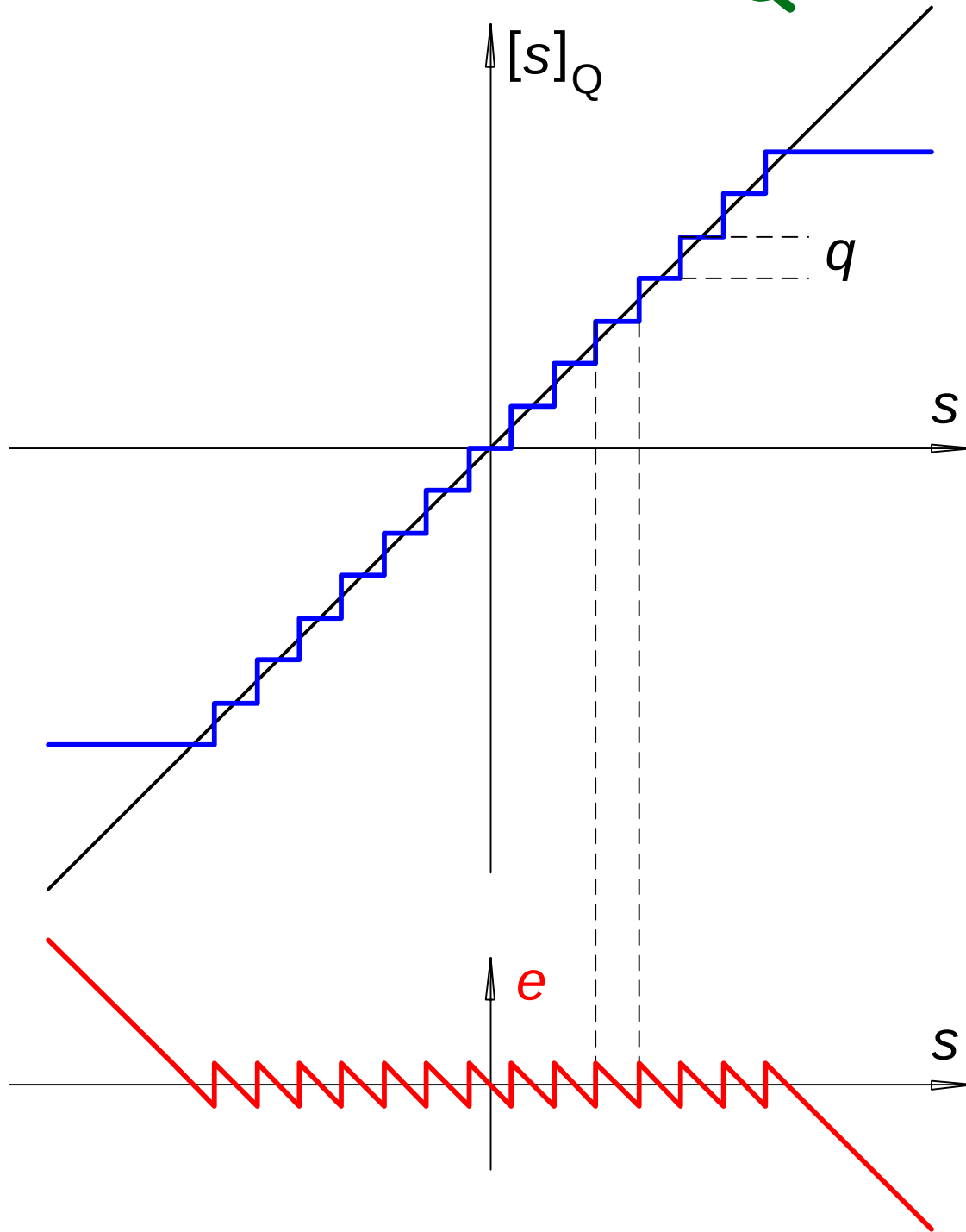


Darstellung als
abgetastete Zeitfunktion $s(n)$

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

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

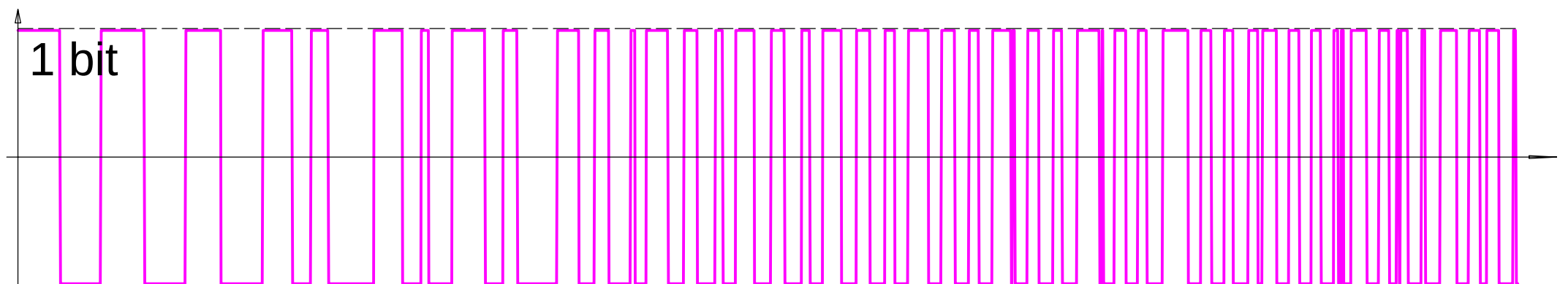
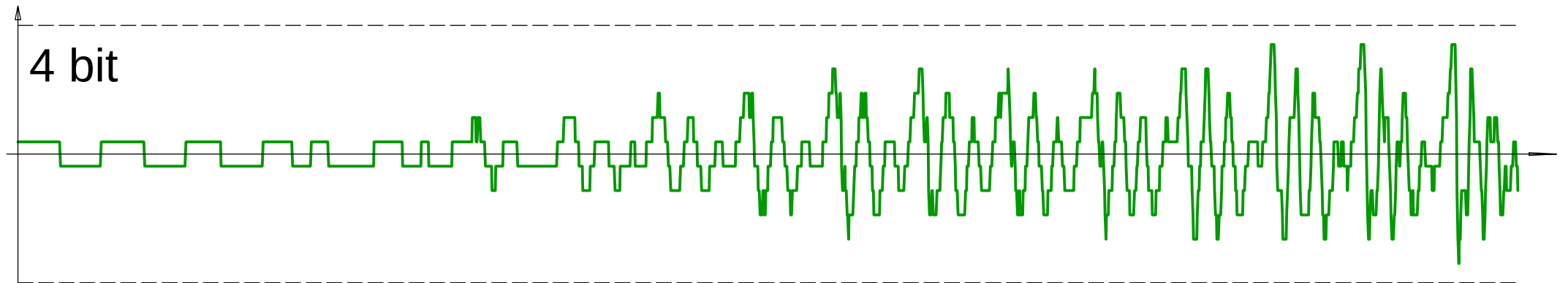
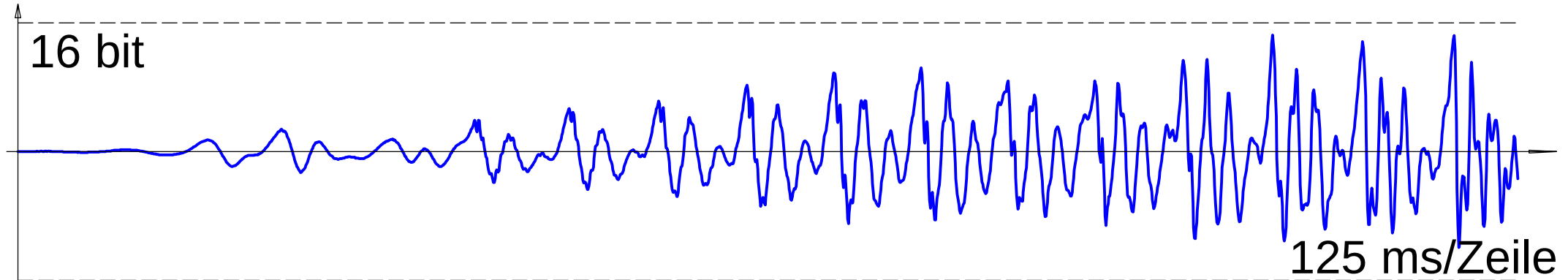
Quantisierung: Beispiel

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

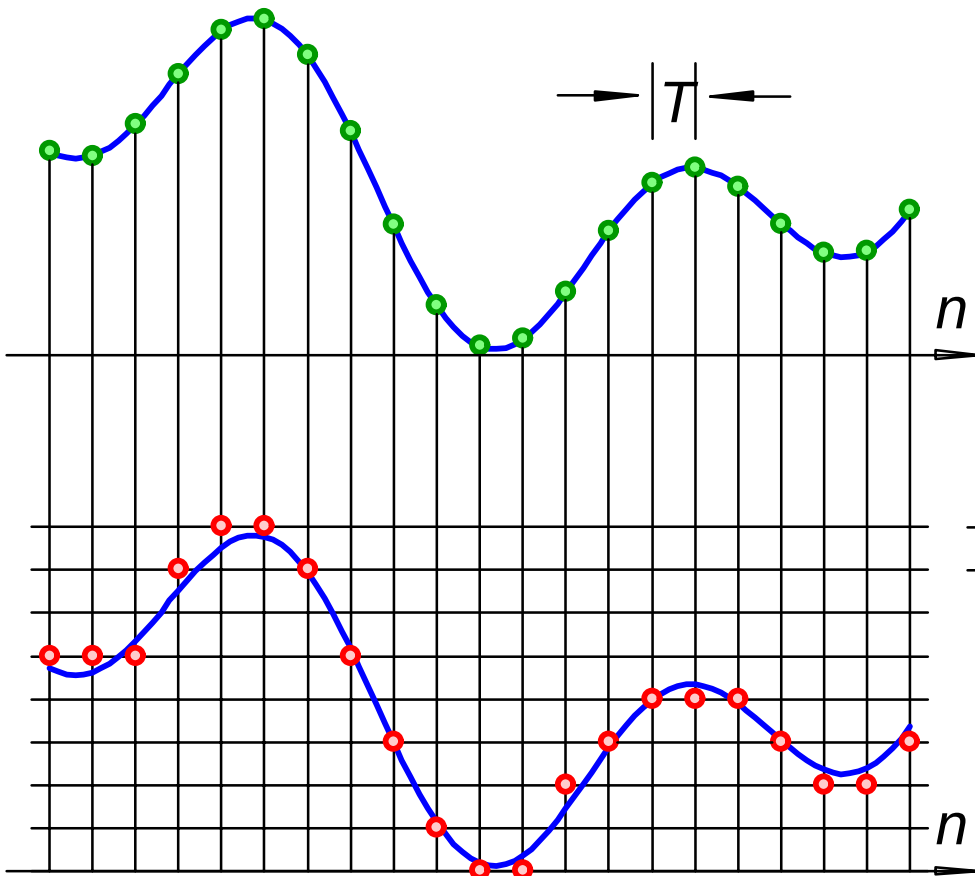
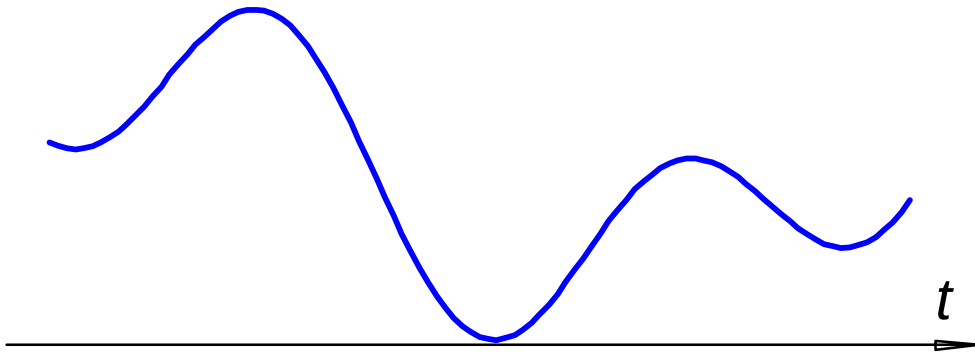
$$s(t) \rightarrow s(nT) \rightarrow s(n)$$

(bei gleichförmiger Abtastung)

Darstellung im Rechner
als Zahlenfolge

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

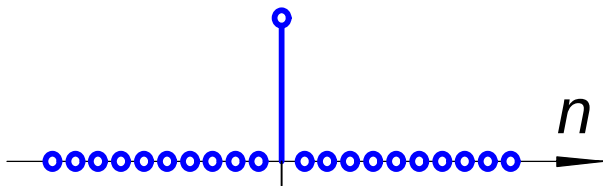
(bei gleichförmiger Quantisierung)



Einige spezielle Signale

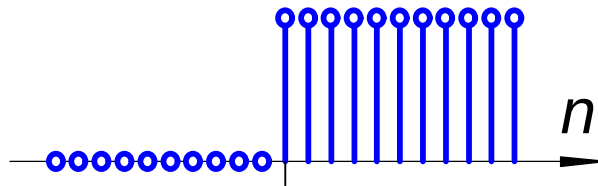
Einheitsimpuls

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



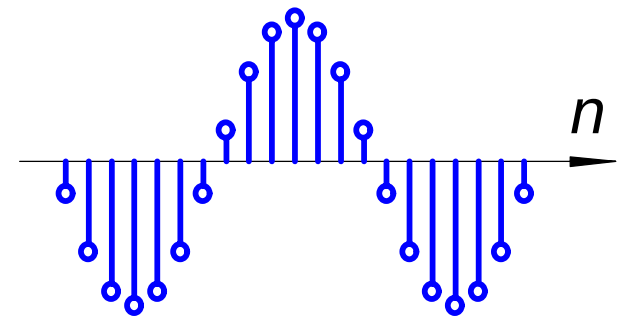
Einheitssprung

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



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

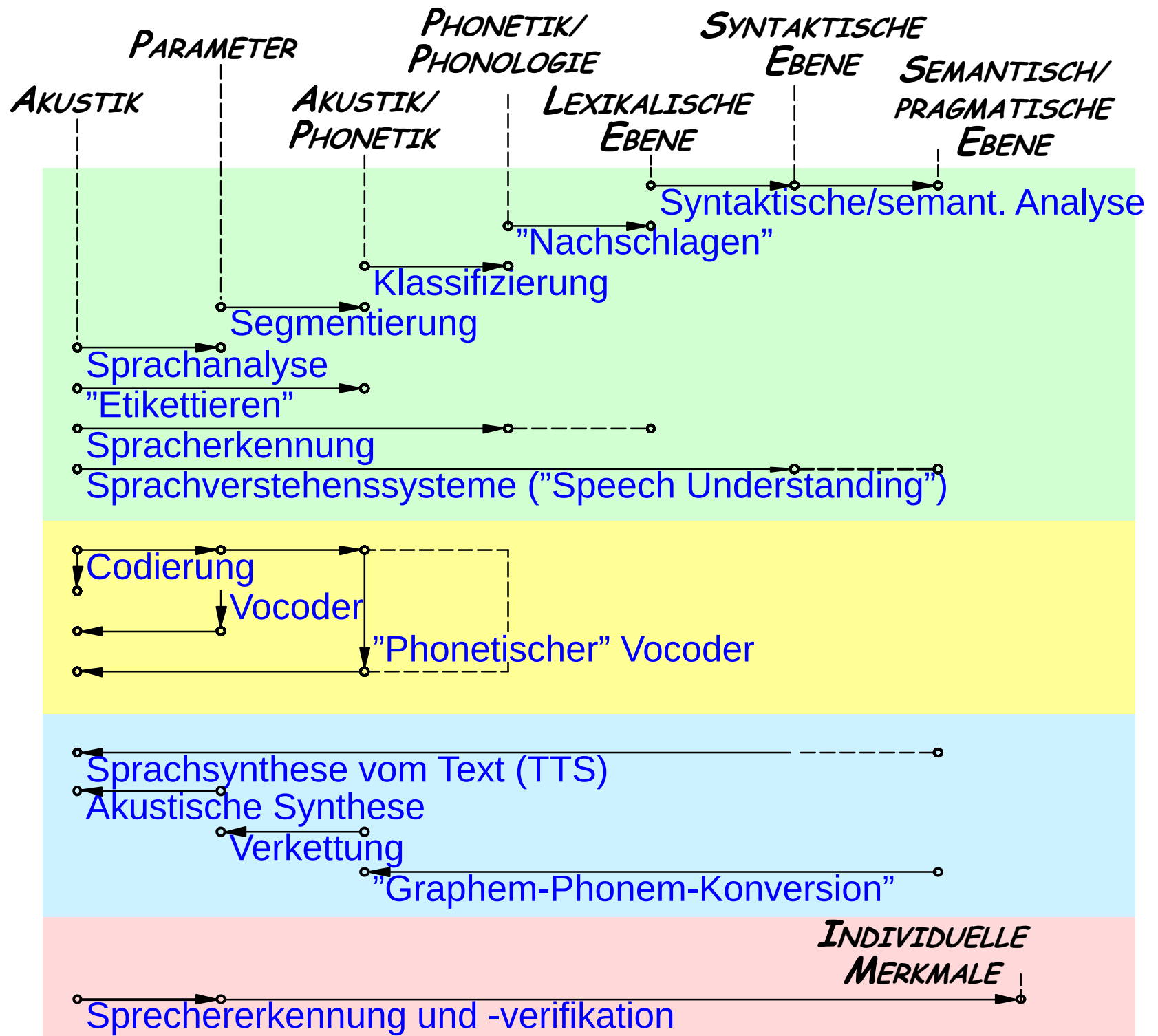
Inf.-Ebene	I.-Fluss (bit/s)	Repräsentation
Akustik	320000	HiFi-Sprachsignal (Abtastfrequenz 20 kHz, 16 bit/Abtastwert)
	64000	Sprachsignal in Telefon-PCM-Qualität (8 kHz, 8 bit/Abtastw.)
	2400	Untere Grenze für Repräsentation auf akustischer Ebene (mit aufwendigen Codierverfahren)
Parameter	4800	Parametrische 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



Grundlagen der Sprachsignalverarbeitung

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

Literaturempfehlung

Vorlesungsskripten – deutsch – bilingual deutsch/englisch

Allgemein zur Sprachsignalverarbeitung

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)

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

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

Speziell zu Kapitel 1

Fant, Gunnar (²1970, ¹1960): *Acoustic theory of speech production* (Mouton, 's-Gravenhage) **(F)**

Flanagan, James L. (²1972, ¹1965): *Speech analysis, synthesis, and perception* (Springer, Berlin) **(F)**

Schüßler, Hans-Wilhelm (³1993, ¹1984): *Netzwerke, Signale und Systeme. Band II – Theorie kontinuierlicher und diskreter Signale und Systeme* (Springer, Berlin) **(F)**

Zwicker, Eberhard / Zollner, Manfred (1984): *Elektroakustik*. Hochschultext (Springer, Berlin) **(F)**

(F) – Lektüre für Fortgeschrittene

Grundlagen der Sprachsignalverarbeitung

2. Kurzzeitanalyse; elementare Parameter des Sprachsignals

2.1 Einige Grundbegriffe der Systemtheorie

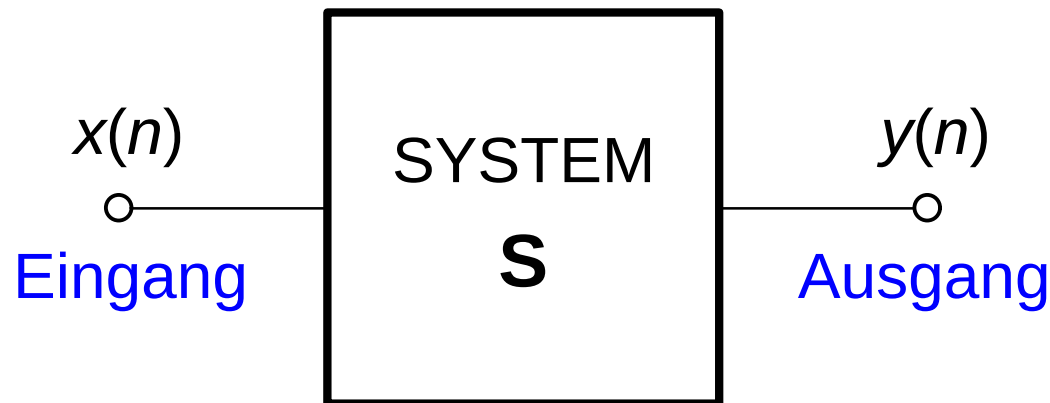
2.2 Das Prinzip der Kurzzeitanalyse

2.3 Elementare Sprachsignalparameter

2.4 Vereinfachte Modellierung des Sprechtrakts als lineares System

Zum Begriff des Systems

Unter einem **System** verstehen wir eine Vorrichtung, die nach einer bestimmten Rechenvorschrift **S** ein Signal x eindeutig in ein Signal y umwandelt. Hierbei wird x als **Eingangssignal**, y als **Ausgangssignal** bezeichnet. Die Signale x und y können Funktionen der gleichen unabhängigen Variablen sein; dies ist jedoch nicht unabdingbar. **S** wird gelegentlich als **Transformationsvorschrift** oder **Systemoperator** bezeichnet. Das Ausgangssignal y heißt auch **Systemantwort** bezüglich des Eingangssignals x .



Die *Systemtheorie* betrachtet und behandelt die Auswirkung der Rechenvorschrift **S** auf die Signale unabhängig von der tatsächlichen Realisierung des Systems.

Vier grundlegende Systemeigenschaften

Systeme der (Sprach-)Signalverarbeitung werden durch vier grundlegende Eigenschaften charakterisiert, die das Verhalten eines Systems entscheidend beeinflussen:

- **Linearität**
- **Zeitinvarianz**
- **Stabilität**
- **Kausalität**

Ein System kann jede dieser Eigenschaften besitzen oder auch nicht.

Systemeigenschaften: Linearität

Ein System ist **linear**, wenn es das **Superpositionsprinzip** erfüllt. Das heißt, wenn

$$y_1(n) = \mathbf{S}\{x_1(n)\} \quad \text{sowie} \quad y_2(n) = \mathbf{S}\{x_2(n)\}$$

dann ist das Superpositionsprinzip erfüllt, wenn gilt

$$k_1 y_1(n) + k_2 y_2(n) = \mathbf{S}\{k_1 x_1(n) + k_2 x_2(n)\}; \quad (2.3)$$

k_1 und k_2 sind hierbei Konstante. Legt man also an den Eingang eines Systems die (gewichtete) Summe zweier Ursachen an, so addieren sich am Ausgang die zugehörigen Wirkungen.

Systeme, die das Superpositionsprinzip nicht erfüllen, werden als **nichtlinear** bezeichnet.

Die Linearität ist eine sehr grundlegende Systemeigenschaft. Sie wird bei den meisten Berechnungen in der Sprachsignalverarbeitung vorausgesetzt.

Systemeigenschaften: Zeitinvarianz

Ein System ist **zeitinvariant**, wenn ein gegebenes Eingangssignal $x(n)$ zu allen Zeiten auf das gleiche Ausgangssignal $y(n)$ führt; d.h., wenn

$$y(n) = \mathbf{S}\{x(n)\} ,$$

dann gilt auch

$$y(n-k) = \mathbf{S}\{x(n-k)\} , \quad (2.4)$$

unabhängig von der (als beliebig, aber fest angenommenen) Verzögerung k .

Erfüllt das System diese Bedingung nicht, so wird es als **zeitveränderlich** oder **zeitvariant** bezeichnet.

Es ist üblich, diesen Begriff auch auf *Signale* auszudehnen. **Sprachsignale sind zeitvariant**, da sich ihre Eigenschaften mit der Zeit ändern.

Systemeigenschaften: Stabilität

Ein System ist (theoretisch) **stabil**, wenn es auf jedes wertemäßig beschränkte Eingangssignal $x(n)$ mit einem beschränkten Ausgangssignal $y(n)$ reagiert:

$$|x(n)| < M_x \rightarrow |y(n)| < M_y ; \quad M_x, M_y \text{ endlich .} \quad (2.5)$$

Diese Stabilitätsbedingung ist für wertkontinuierliche Systeme notwendig und hinreichend.

Quantisierungseffekte können in realen digitalen Systemen (also in wertdiskreter Realisierung) jedoch bewirken, dass ein System nicht stabil ist, obwohl die theoretische Stabilität gewährleistet ist. In solchen Fällen hilft der Begriff der **asymptotischen Stabilität** (Ljapunow-Stabilität). Sofern hier das Eingangssignal zu Null wird, strebt auch das Ausgangssignal gegen Null:

$$x(n) \equiv 0 \text{ für } n > n_1 \text{ ergibt } y(n) \rightarrow 0 \text{ für } n \rightarrow \infty . \quad (2.6)$$

Systemeigenschaften: Kausalität

Ein System ist **kausal**, wenn jeder Abtastwert des Ausgangssignals $y(n)$ nur von zeitlich zurückliegenden, höchstens aber gleichzeitigen Abtastwerten des Eingangssignals $x(n)$ abhängt:

$$y(n) = y [x(k), k = \dots (1) n] . \quad (2.7)$$

Anders ausgedrückt: Sind alle Abtastwerte eines Eingangssignals $x(n)$ Null für $n < n_0$, so kann auch das Ausgangssignal für $n < n_0$ keine von Null verschiedenen Abtastwerte besitzen.

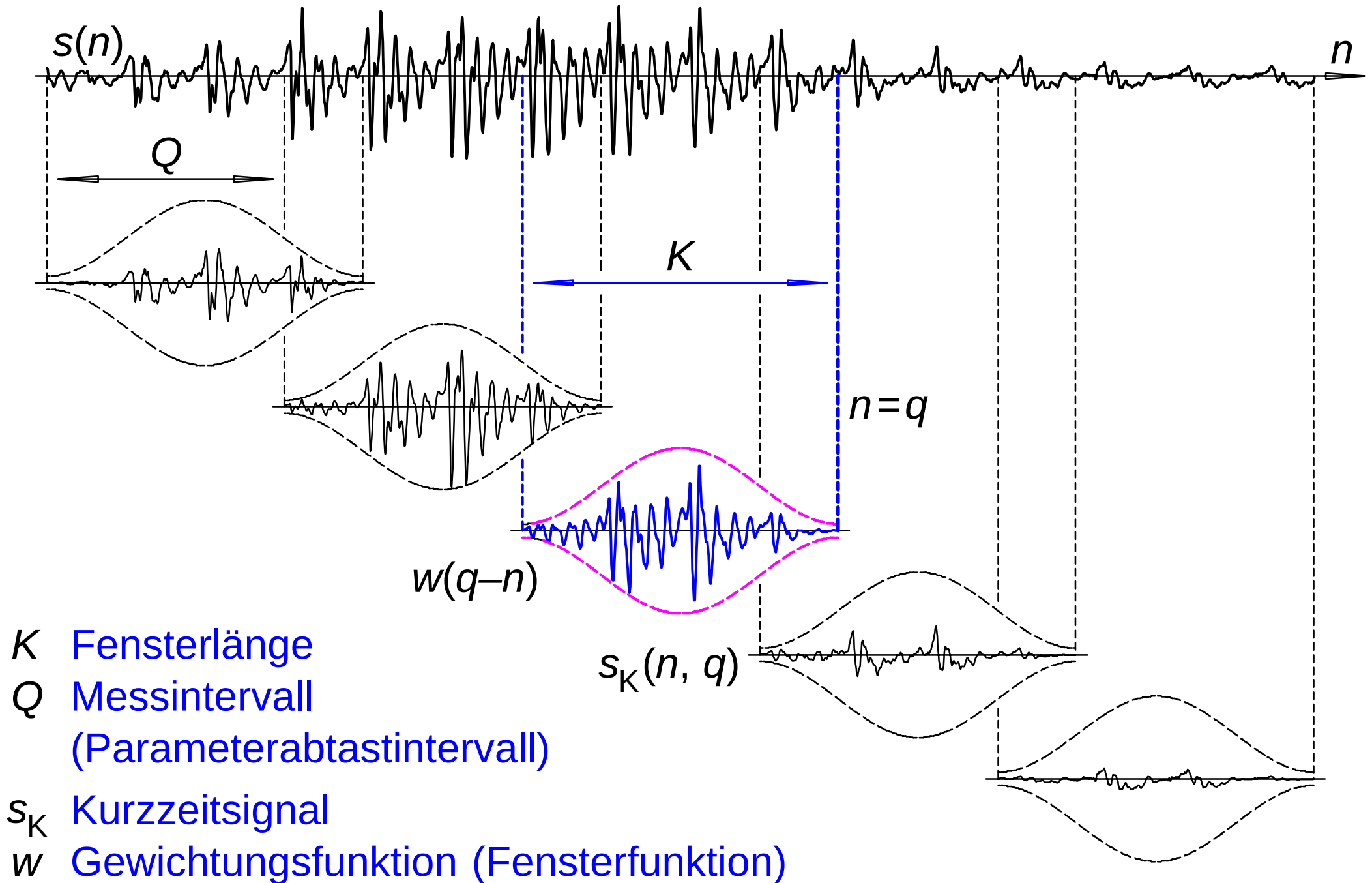
Alle in der Praxis realisierten Systeme sind kausal. Für theoretische Betrachtungen oder für Messungen an *gespeicherten*, zeitlich begrenzten Signalen ist es in bestimmten Fällen sinnvoll, das Kausalitätsprinzip (vorübergehend) zu verletzen. Dies ist dann möglich, wenn die Kausalität durch Hinzufügen einer endlichen Verzögerung wiederhergestellt werden kann.

Grundlagen der Sprachsignalverarbeitung

2. Kurzzeitanalyse; elementare Parameter des Sprachsignals

- 2.1 Einige Grundbegriffe der Systemtheorie
- 2.2 Das Prinzip der Kurzzeitanalyse**
- 2.3 Elementare Sprachsignalparameter
- 2.4 Vereinfachte Modellierung des Sprechtrakts als lineares System

Prinzip der Kurzzeitanalyse



Dimensionierung der Parameter für die Kurzzeitanalyse

Die Fensterlänge K ist so zu wählen, dass der zu bestimmende Sprachsignalparameter 1) zuverlässig genug gemessen werden kann, aber auch noch 2) mit hinreichender Genauigkeit als konstant angenommen werden darf.

Da diese beiden Bedingungen sich widersprechen, wird die Messung stets von einer gewissen Unschärfe begleitet sein.

In der Sprachsignalverarbeitung gebräuchliche Werte sind:

- für das **Messintervall** Q : 5 bis 25 ms
- für die **Fensterlänge** K : 20 bis 50 ms.

Ein Sonderfall ergibt sich, wenn die Fensterlänge K gleich der Dauer T_0 der Grundperiode wird. In diesem Fall lässt sich die natürliche Periodizität des Signals ausnützen. Die Methode hat den Vorteil, dass Verzerrungen durch das Fenster auf ein Minimum reduziert werden können (**grundperiodensynchrone** oder auch "**grundfrequenzsynchrone**" Analyse).

Drei Ansätze zur Kurzzeitanalyse

Das Kurzzeitsignal s_K ist zunächst nur *innerhalb* des Fensters definiert:

$$s_K(n, q) = \begin{cases} s(n) w(q-n); & n = q-K+1 \text{ (1) } q \\ \text{sonst (zunächst) undefiniert, d.h. beliebig.} \end{cases} \quad (2.9)$$

Für manche Anwendungen werden auch Signalwerte außerhalb des Fensters benötigt. Diese werden dann je nach Messverfahren gemäß einem der folgenden Ansätze festgelegt.

1) Das Signal s_K wird als **einmaliger Vorgang** angesehen. In diesem Fall werden die Signalwerte außerhalb des Fensters zu Null gesetzt (**"stationärer" Ansatz**).

2) Das Signal s_K wird als **periodischer Vorgang** angesehen. In diesem Fall wird es außerhalb des Fensters periodisch fortgesetzt; die Periode ist in der Regel gleich der Fensterlänge K [ist sie größer, so muss mit Nullwerten aufgefüllt werden] (**"synchroner" bzw. "periodischer" Ansatz**).

3) Das Signal s_K wird als **Ausschnitt aus einem zeitvarianten Vorgang** angesehen (**"nichtstationärer" Ansatz**). Werden Signalwerte außerhalb des Fensters gebraucht, so nimmt man die Originalwerte.

Grundlagen der Sprachsignalverarbeitung

2. Kurzzeitanalyse; elementare Parameter des Sprachsignals

2.1 Einige Grundbegriffe der Systemtheorie

2.2 Das Prinzip der Kurzzeitanalyse

2.3 Elementare Sprachsignalparameter

2.3.1 Physikalische Grundgrößen in der Sprachsignalverarbeitung

2.3.2 Amplitude, Pegel, Energie, Leistung des Sprachsignals in digitaler Darstellung

2.3.3 Grobmessung des Frequenzverhaltens

2.4 Vereinfachte Modellierung des Sprechtrakts als lineares System

Leistung, Energie und die digitale Darstellung [1]

In der Physik sind **Energie** (Arbeit) und **Leistung** vor allem aus Mechanik und Elektrizitätslehre wohlbekannt:

$$E = F \cdot s ; \quad [(\text{mechanische}) \text{ Arbeit gleich Kraft mal Weg}] \quad (2.10/11)$$

$$P = U \cdot I ; \quad [(\text{elektrische}) \text{ Leistung gleich Spannung mal Stromstärke}]$$

Energie und Leistung sind verbunden durch die ebenfalls wohlbekannte Beziehung

$$P = E/t ; \quad [\text{Leistung gleich Energie durch Zeit}] \quad (2.12a)$$

oder umgekehrt

$$E = P \cdot t ; \quad [\text{Energie gleich Leistung mal Zeit}] \quad (2.12b)$$

Leistung, Energie und die digitale Darstellung

[2]

Ein Grundgesetz der Elektrizitätslehre ist das **Ohm'sche Gesetz**, das die Größe des elektrischen **Widerstandes** definiert,

$$U/I = R ; \quad [\text{Widerstand gleich Spannung durch Stromstärke}] \quad (2.13)$$

Der Widerstand ist eine **Materialkonstante** und nicht abhängig von Strom oder Spannung.

Mit Hilfe des Widerstandes gelingt es, Energie bzw. Leistung durch *eine* der Größen U bzw. I auszudrücken; die Beziehung wird **quadratisch**:

$$P = U \cdot I = I^2 \cdot R = U^2/R . \quad (2.14)$$

Diese Beziehungen sind in dieser Form zunächst für den (einfachsten) Fall definiert, dass alle beteiligten Grundgrößen (Kraft, Spannung, Stromstärke und insbesondere auch die Leistung) *keine Funktionen der Zeit* sind.

Leistung, Energie und die digitale Darstellung [3]

Ist die elektrische Spannung eine Funktion der Zeit, so errechnet sich beispielsweise die in einem Verbraucher während des Zeitintervalls (t_2-t_1) umgesetzte elektrische Energie zu

$$E = \int_{t_1}^{t_2} u(t) i(t) dt . \quad (2.15a)$$

Der **Effektivwert** für $u(t)$ ist definiert als der Gleichspannungswert U_{eff} , der bei gleichem Widerstand R die gleiche (mittlere) Leistung ergibt wie $u(t)$; dies bedeutet insbesondere auch gleiche Energie im betrachteten Zeitintervall:

$$E = \int_{t_1}^{t_2} u(t) i(t) dt = U_{\text{eff}} \cdot I_{\text{eff}} \cdot (t_2-t_1) = I_{\text{eff}}^2 \cdot R \cdot (t_2-t_1) = \frac{U_{\text{eff}}^2}{R} \cdot (t_2-t_1) . \quad (2.15b)$$

Analog hierzu werden eine (Pseudo-)Energie, eine (Pseudo-)Leistung und ein (Pseudo-)Effektivwert für digitale Signale gebildet.

Grunddefinitionen der Akustik [1]

[siehe *Grundlagen der Phonetik*, Kap.1]

Akustische Signale breiten sich in Luft wie in jedem den Schall leitenden Medium in Form von *Schallwellen* aus.

An jeder Welle sind *zwei* physikalische Größen beteiligt; bei der Schallwelle sind dies der *Schalldruck* p und die *Schallschnelle* v .

Der *Schalldruck* p gibt an, *wie sich durch einen Schall am Ort der Messung der Luftdruck* (bzw. allgemein der Druck des Mediums, in dem sich eine Schallwelle bewegt) *ändert*. Gemessen wird er in *Pascal* (Pa). In der älteren Literatur ist auch noch die Einheit *Mikrobar* (μb) zu finden.

$$1 \text{ Pa} = 1 \text{ N/m}^2 = 10 \mu\text{b} .$$

Grunddefinitionen der Akustik [2]

[siehe *Grundlagen der Phonetik*, Kap.1]

Die *Schallschnelle* $\mathbf{v}$ gibt die *Geschwindigkeit* an, mit der sich ein *Luftteilchen infolge eines Schallereignisses bewegt*. Bei beliebiger Wellenausbreitung ist $\mathbf{v}$ ein Vektor; bei dem uns in der Sprachsignalverarbeitung ausschließlich interessierenden Fall der *ebenen Welle* (d.h., einer Welle, die sich nur in einer Richtung ausbreitet – im Unterschied etwa zu einer sich kugelförmig nach allen Richtungen ausbreitenden Welle) kann $\mathbf{v}$ als skalare Größe v dargestellt werden.

Nicht zu verwechseln ist die Schallschnelle mit der *Schallausbreitungsgeschwindigkeit* (oder kurz *Schallgeschwindigkeit*) c , mit der sich die *Schallwelle als Ganzes ausbreitet*. Während die Schallschnelle vom Schalldruck abhängt (und damit von der Schallintensität bzw. Lautstärke), ist die Schallgeschwindigkeit eine Materialkonstante; in Luft beträgt sie

$$c \approx 340 \text{ m/s} .$$

Grunddefinitionen der Akustik [3]

[siehe *Grundlagen der Phonetik*, Kap.1]

Entsprechend dem Ohm'schen Gesetz in der Elektrizitätslehre kann auch in der Akustik eine Widerstandsgröße gebildet werden; sie ist im allgemeinen Fall definiert als *Schallfeldimpedanz*

$$Z_0 := p / v \quad [\text{Einheit N s/m}^3 \text{ oder Pa s/m}] \quad (2.16)$$

und kennzeichnet die akustischen Eigenschaften des Übertragungsmediums. Für eine ebene Welle im Ausbreitungsmedium Luft gilt

$$Z_s = Z_0 = 414 \text{ N s/m}^3 = 414 \text{ Pa s/m} . \quad (2.16a)$$

Diese Größe wird *Schallkennimpedanz* genannt. Auch sie ist eine *Materialkonstante* und lässt sich für jedes Schallübertragungsmedium angeben.

Übertragungsmedien mit niedriger Schallkennimpedanz (beispielsweise Luft oder andere Gase) werden als *schallweich*, solche mit hoher Schallkennimpedanz (z.B. Flüssigkeiten oder feste Körper) als *schallhart* bezeichnet.

Pegelmaße: Schalldruckpegel

[siehe *Grundlagen der Phonetik*, Kap.1]

Die in der Akustik üblichen Werte für den Schalldruck variieren in weiten Grenzen (bis zu 7 Zehnerpotenzen in der Amplitude zwischen Ruheschwelle und Schmerzschwelle).

Daher wird der Schalldruck meist nicht direkt in Pa [Pascal] oder μPa , sondern als logarithmisches Verhältnismaß in der dimensionslosen Größe *Dezibel* (dB) anzugeben.

Definition für Leistungs- und Energiegrößen [$\lg \triangleq$ Zehnerlogarithmus]

$$L := 10 \lg (P/P_0) \quad [\text{dB}] \quad (2.17a)$$

Definition für Amplitudengrößen, wie den Schalldruck p , als

$$L := 20 \lg (p/p_0) \quad [\text{dB}] \quad (2.17b)$$

Diese Definitionen ergeben übereinstimmende Werte für Leistung und Amplitude.

Für den Schalldruck ist ein Wert $p_0 = 20 \mu\text{Pa}$ als **Bezugswert** international festgelegt und normiert. Dieser Wert liegt knapp unterhalb der Ruheschwelle bei 1000 Hz.

Grundlagen der Sprachsignalverarbeitung

2. Kurzzeitanalyse; elementare Parameter des Sprachsignals

2.1 Einige Grundbegriffe der Systemtheorie

2.2 Das Prinzip der Kurzzeitanalyse

2.3 Elementare Sprachsignalparameter

2.3.1 Physikalische Grundgrößen in der Sprachsignalverarbeitung

2.3.2 Amplitude, Pegel, Energie, Leistung des Sprachsignals in digitaler Darstellung

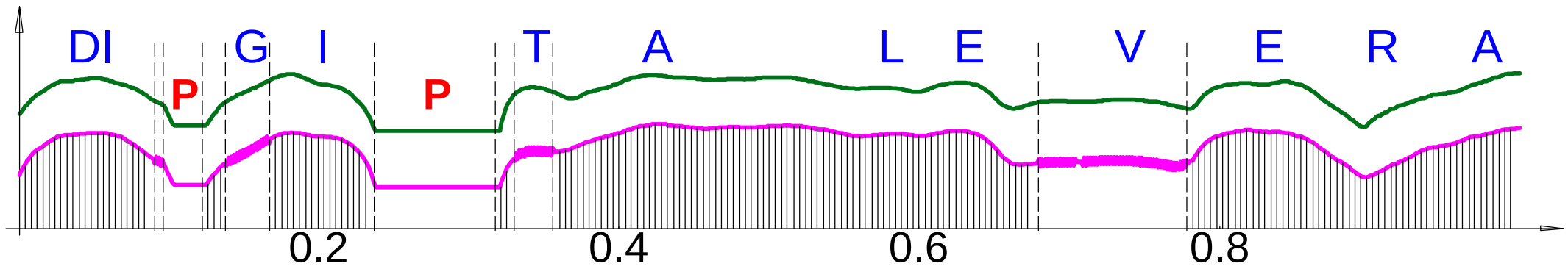
2.3.3 Grobmessung des Frequenzverhaltens

2.4 Vereinfachte Modellierung des Sprechtrakts als lineares System

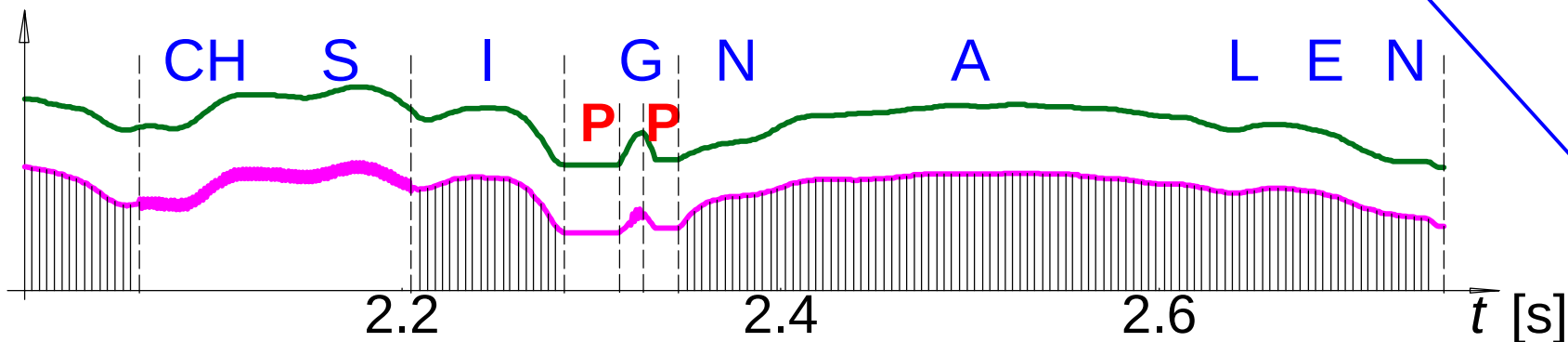
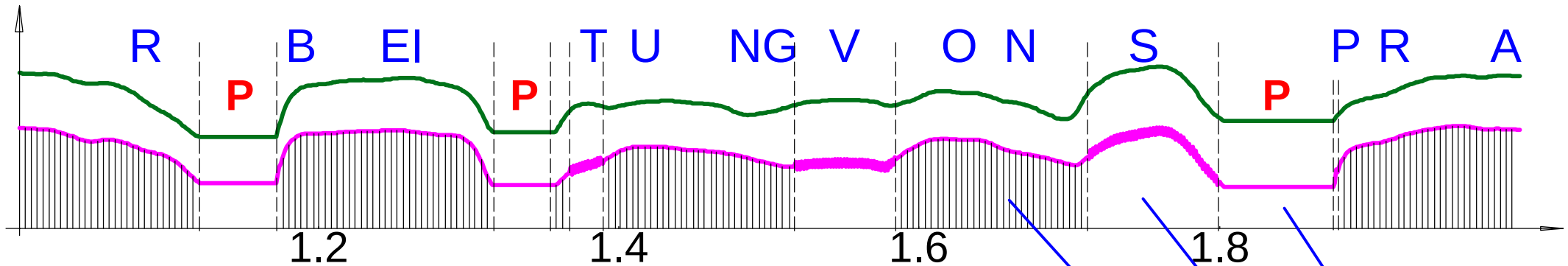
(Kurzzeit-)Amplitude, Pegel, Energie

Kurzzeitenergie	(2.22)	$U_K(q) := \sum_{n=q-K+1}^q s_K^2(n, q)$
Kurzzeitleistung	(2.21)	$P_K(q) := \frac{1}{K} \sum_{n=q-K+1}^q s_K^2(n, q)$
Kurzzeiteffektivwert	(2.20)	$u_K(q) := \sqrt{\frac{1}{K} \sum_{n=q-K+1}^q s_K^2(n, q)}$
Betragsmittelwert	(2.23)	$ \bar{s}_K(q) := \frac{1}{K} \sum_{n=q-K+1}^q s_K(n, q) $
Amplituden (Spitzenwerte)	(2.24a-d)	$A_b(q) := \max [s_K(n, q)], \quad n = q-K+1 \text{ (1) } q$ $A_+(q) := \max [s_K(n, q)], \quad A_-(q) := \min [s_K(n, q)]$ $A_{ss}(q) := \max [s_K(n, q)] - \min [s_K(n, q)]$

Kurzzeit-Amplituden-Verlauf: Beispiel



Betragsmittelwert (Signal) (differenziertes Signal) [logarithmische Darstellung]



• Pause
• Stimmlos
• Stimmhaft

(Kurzzeit-)Amplitude, Pegel, Energie

Ein extremes Beispiel

Die Maße für die Kurzzeitamplitude können je nach Signal erheblich divergieren. Extremes Beispiel:

Signal [Einzelimpuls im Fenster]

$$s_K(n, q) = \begin{cases} 100 & n=q \\ 0 & \text{sonst} \end{cases}; \quad K=100$$

Effektivwert

$$u_K(q) := \sqrt{\frac{1}{K} \sum_{n=q-K+1}^q s_K^2(n, q)} = \sqrt{\frac{1}{100} \cdot 100^2} = \sqrt{100} = 10$$

Betragsmittelwert

$$|\bar{s}_K(q)| := \frac{1}{K} \sum_{n=q-K+1}^q |s_K(n, q)| = \frac{1}{100} \cdot 100 = 1$$

Amplituden (Spitzenwerte)

$$A_b(q) = A_+(q) = A_{ss}(q) = 100$$

Beachte: Sprachsignale sind häufig der Impulsform angenähert!

Grobmessung des Frequenzverhaltens

Differenziertes Signal
(approximiert durch Filter 1. Grades)

$$(2.25) \quad s_D(n) \approx s(n) - s(n-1)$$

Quotient Amplitude
[Betragsmittelwert]
differenziertes Signal und Signal

$$(2.26) \quad Q_S := \frac{|\bar{s}_D(q)|}{2 |\bar{s}(q)|}$$

Relative
Nulldurchgangshäufigkeit

$$(2.27) \quad H_N(q) := \frac{1}{K} \sum_{n=q-K+1}^q x_S(n)$$

$$x_S(n) = \begin{cases} 0, & \text{sgn}[s(n)] = \text{sgn}[s(n-1)] \\ 1, & \text{sgn}[s(n)] \neq \text{sgn}[s(n-1)] \end{cases}$$

Grundlagen der Sprachsignalverarbeitung

2. Kurzzeitanalyse; elementare Parameter des Sprachsignals

- 2.1 Einige Grundbegriffe der Systemtheorie
- 2.2 Das Prinzip der Kurzzeitanalyse
- 2.3 Elementare Sprachsignalparameter
- 2.4 Vereinfachte Modellierung des Sprechtrakts als lineares System

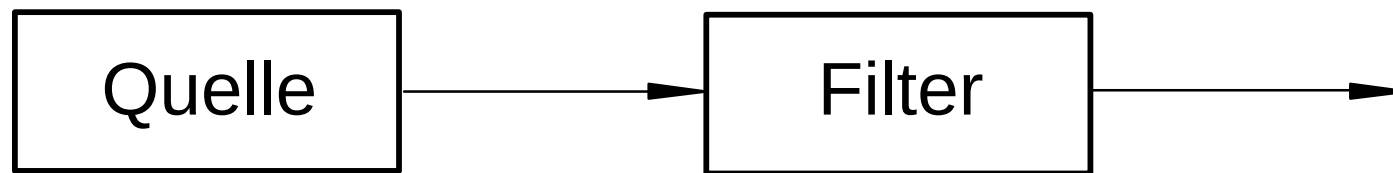
Quelle-Filter-Modell

[vgl. Grundlagen der Phonetik, Kap.1]

Im Prinzip erfolgt die Bildung des Sprachsignals in zwei Schritten:

- 1) Anregung ("Quelle")
- 2) Signalformung ("Filter")

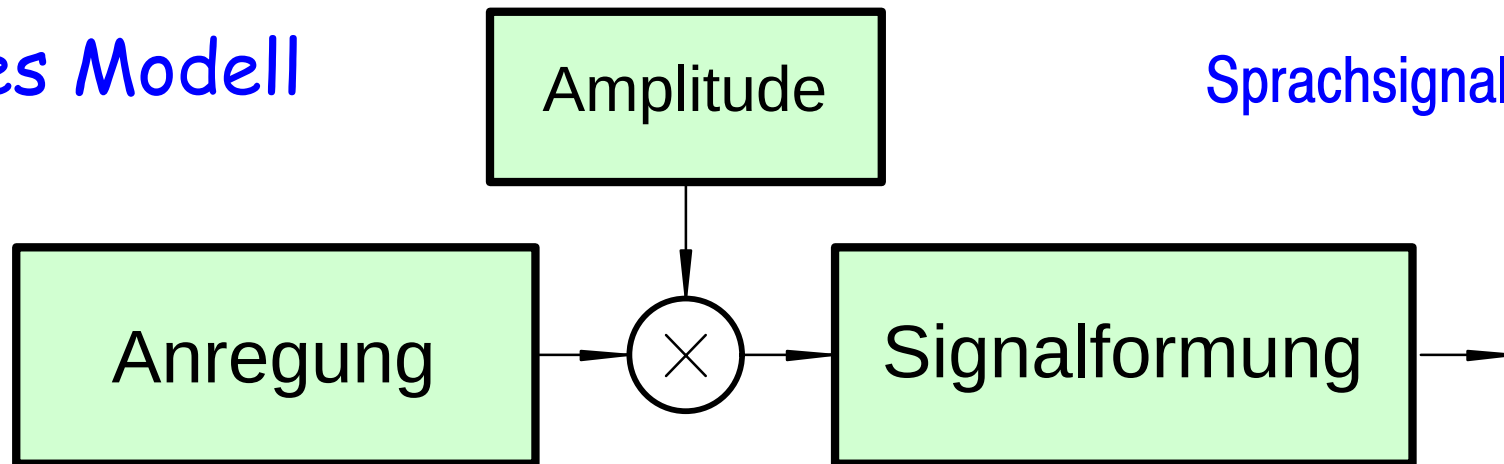
Diese beiden Komponenten können mit guter Näherung als unabhängig voneinander betrachtet werden.



Hier geht es darum, Modelle zu erhalten, welche die Komponenten quantitativ beschreiben.

Lineares Modell der Sprachproduktion Stimmhafter Pfad

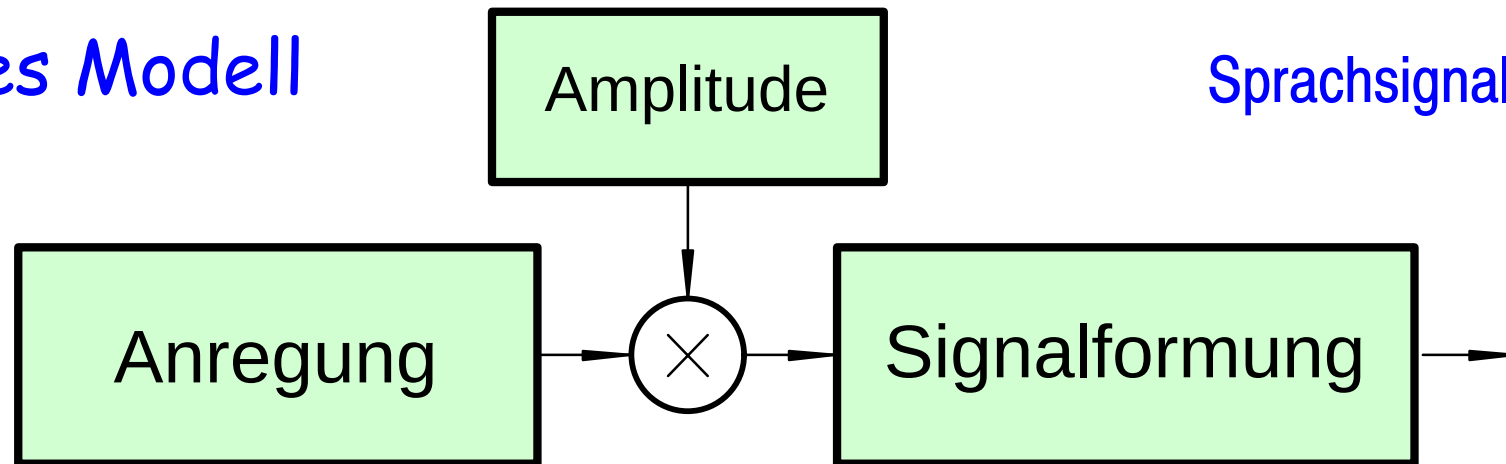
Einfachstes Modell



Lineares Modell der Sprachproduktion

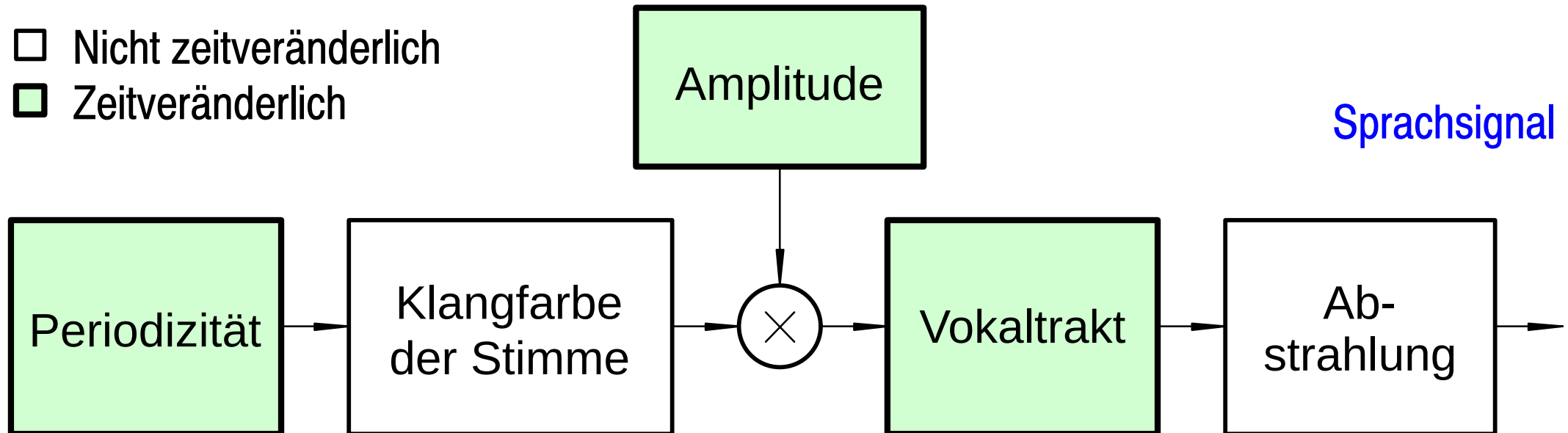
Stimmhafter Pfad

Einfachstes Modell

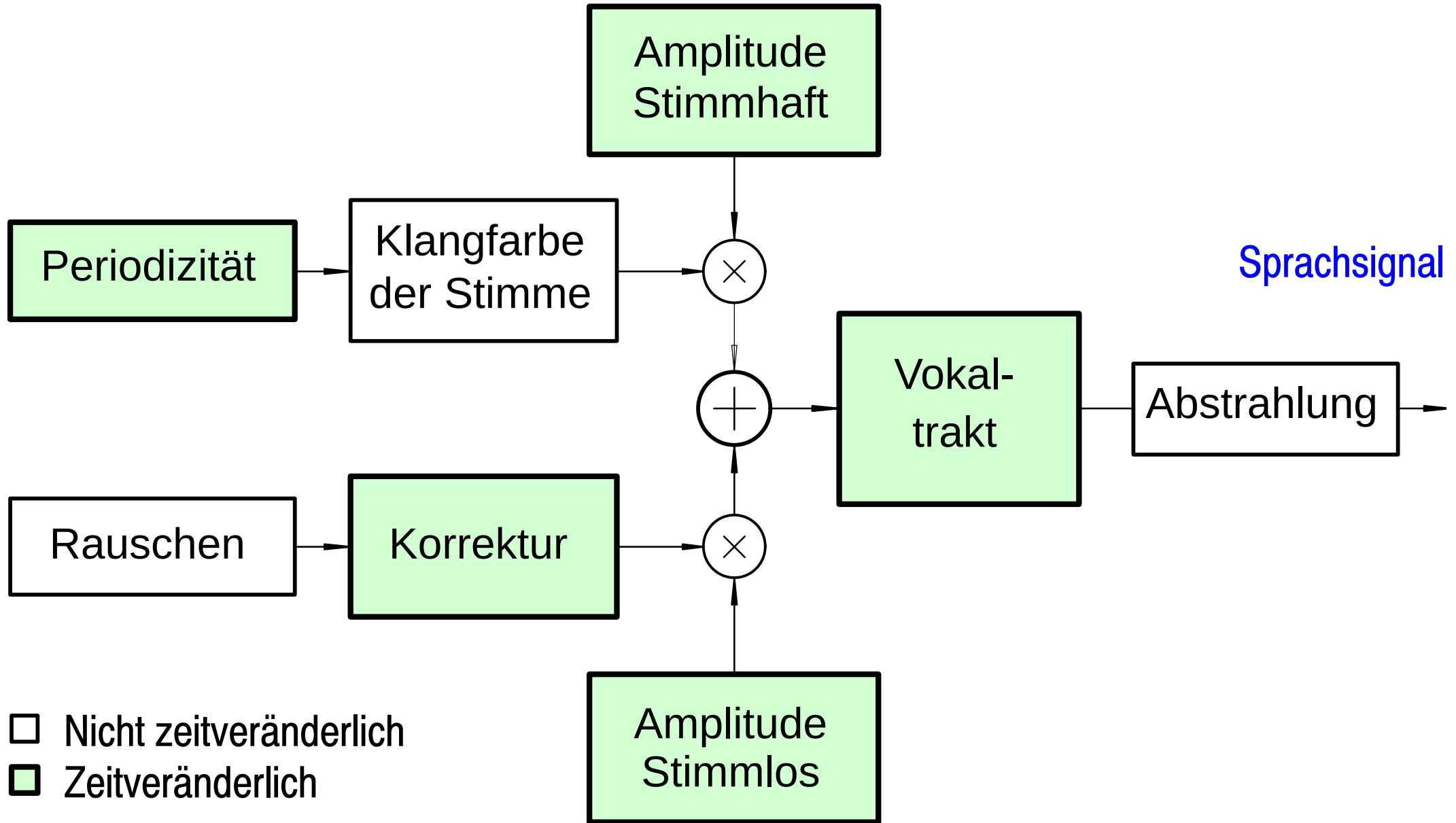


Detaillierteres Modell

- ☐ Nicht zeitveränderlich
- ☒ Zeitveränderlich



Lineares Modell der Sprachproduktion



Auf dieses Modell werden wir später in größerem Detail - vor allem quantitativ - zurückkommen.

Grundlagen der Sprachsignalverarbeitung

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

Literaturempfehlung

Vorlesungsskripten – deutsch – bilingual deutsch/englisch

Allgemein zur Sprachsignalverarbeitung

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)

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

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

Speziell zu Kapitel 3

Bronstein, I. N. / Semendjajew, K. A. (²⁵1991): *Taschenbuch der Mathematik* (Teubner, Stuttgart) [Ergänzungslektüre zur Mathematik]

Harris, Frederic J. (1978): "On the use of windows for harmonic analysis with the discrete Fourier transform." Proc. IEEE 66, 51...63 (F)

(F) – Lektüre für Fortgeschrittene

Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

3.1 Harmonische Analyse (diskreter) periodischer Signale

3.1.1 Sinoidalschwingung (Sinusschwingung) als Grundbaustein akustischer Signale

3.1.2 Orthogonalitätsbeziehung und Fourierkoeffizienten

3.1.3 Exkurs: Was benötigen wir an Mathematik in diesem Kapitel?

3.1.4 Übergang zur komplexen Exponentialfunktion. Die diskrete Fouriertransformation (DFT)

3.1.5 Einige besondere Zeitfunktionen und ihre Spektren

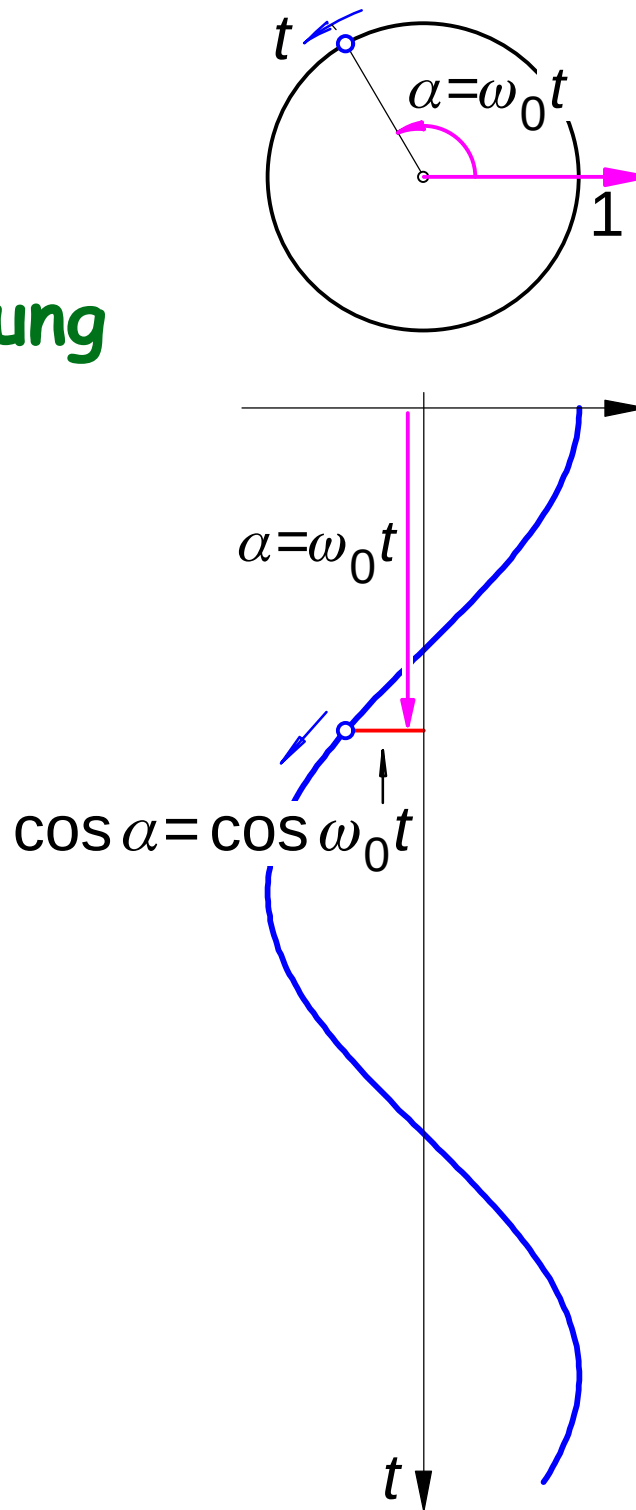
3.2 Nichtharmonische Spektralanalyse

3.3 Kurzzeitspektren von Sprachsignalen

3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

Die Sinoidalschwingung
(der Sinuston)
resultiert aus der
gleichförmigen
Kreisbewegung.

In der Akustik
wird ein Sinuston
als **reiner Ton**
bezeichnet.



Bewegt sich ein Punkt
gleichförmig auf dem
Umfang eines
Einheitskreises, so
führt seine
Projektion auf die
Abszisse des
zugehörigen
Koordinatensystems
eine
Kosinusschwingung
aus.

Sinuston als akustischer Elementarbaustein [1]

Der zeitliche Verlauf eines reinen Tons (Sinuston) ist durch eine **Kosinus-** oder **Sinusfunktion** beschrieben,

$$x_r(t) = x_{rg}(t) + x_{ru}(t) ; \quad \text{wobei}$$

$$x_g(t) = A_g \cos \omega_0 t ; \quad x_u(t) = A_u \sin \omega_0 t ; \quad (3.1)$$

$x_g(t)$ ist gerade, $x_u(t)$ ungerade; ω_0 ist die **Kreisfrequenz**; diese ist durch die Beziehung

$$\omega_0 = 2\pi f_0 = 2\pi / t_0 \quad (3.2)$$

mit der **Frequenz** bzw. der **Periode** der Schwingung verbunden.

Sinuston als akustischer Elementarbaustein [2]

Kosinus- und Sinusanteil lassen sich zusammenfassen,

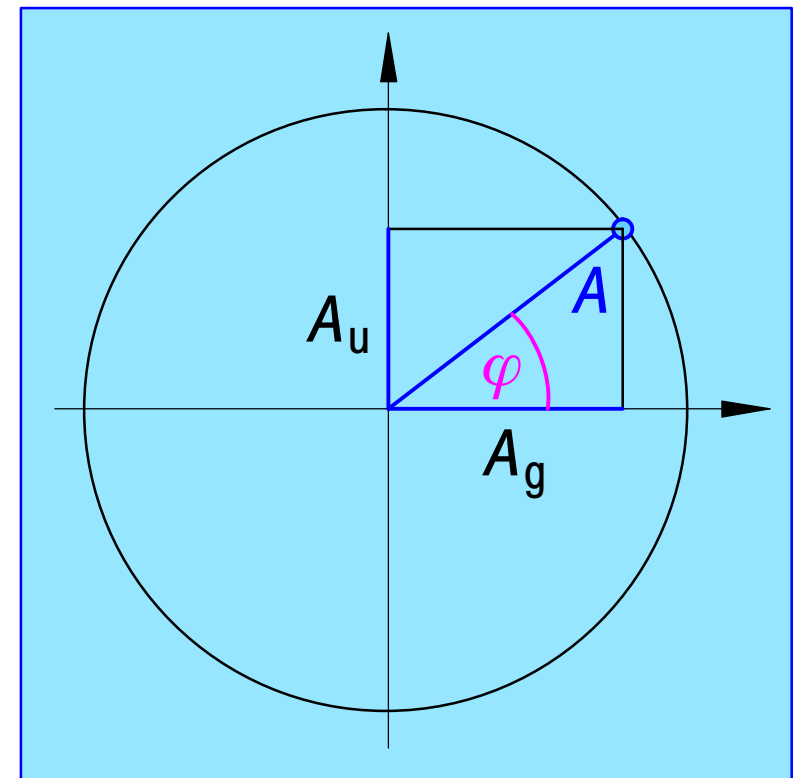
$$x_r(t) = x_{rg}(t) + x_{ru}(t) \quad (3.4)$$

$$= A_g \cos \omega_0 t + A_u \sin \omega_0 t = A \cos (\omega_0 t + \varphi)$$

$$A = \sqrt{A_g^2 + A_u^2} \quad \text{sowie} \quad \varphi = \arctan \frac{A_u}{A_g}. \quad (3.5a,b)$$

A ist die **Amplitude** der resultierenden Schwingung;
 φ deren **Phase**.

Durch Amplitude, Phase und Frequenz ist ein Sinuston vollständig bestimmt.



Der Ton als Summe von Sinustönen (Teilschwingungen)

In der Akustik verstehen wir unter einem **Ton** eine i.A. nichtsinusförmige periodische Schwingung $x_p(t)$ mit der Periode t_0 .

Eine solche Schwingung lässt sich durch Zerlegung in eine Reihe als Summe von (i.A. unendlich vielen) Sinoidalschwingungen exakt darstellen.

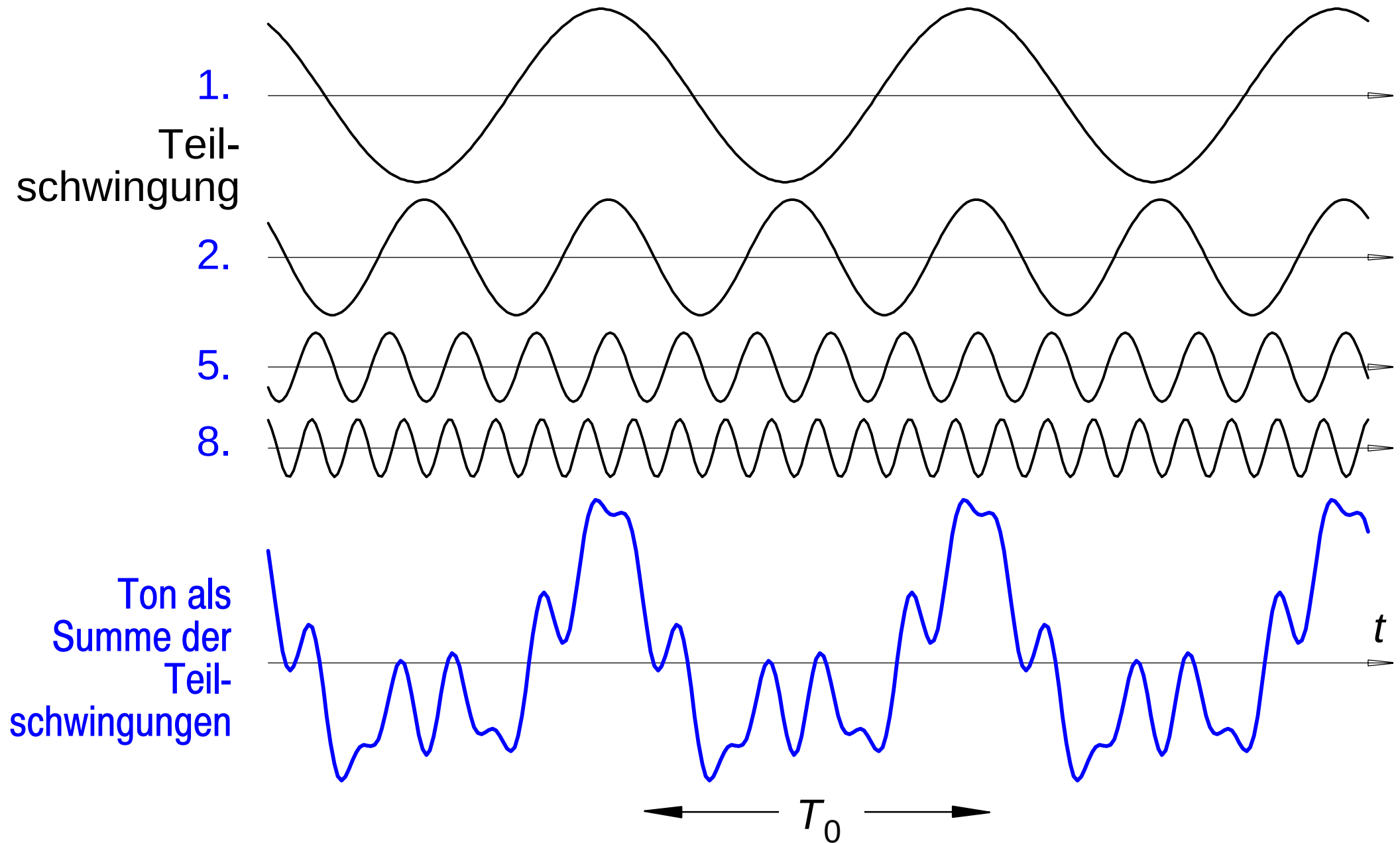
Die zugehörige Reihe heißt (nach einem ihrer Entdecker, dem Mathematiker JOSEPH B. FOURIER, benannt) **Fourierreihe**.

Die einzelnen Sinoidalschwingungen, die die Glieder der Reihe ausmachen, sind in ihren Frequenzen nicht beliebig wählbar. Ist t_0 die Periode der Gesamtschwingung, und gilt nach wie vor $f_0 = 1/t_0$, so können die Glieder der Reihe nur Frequenzen von

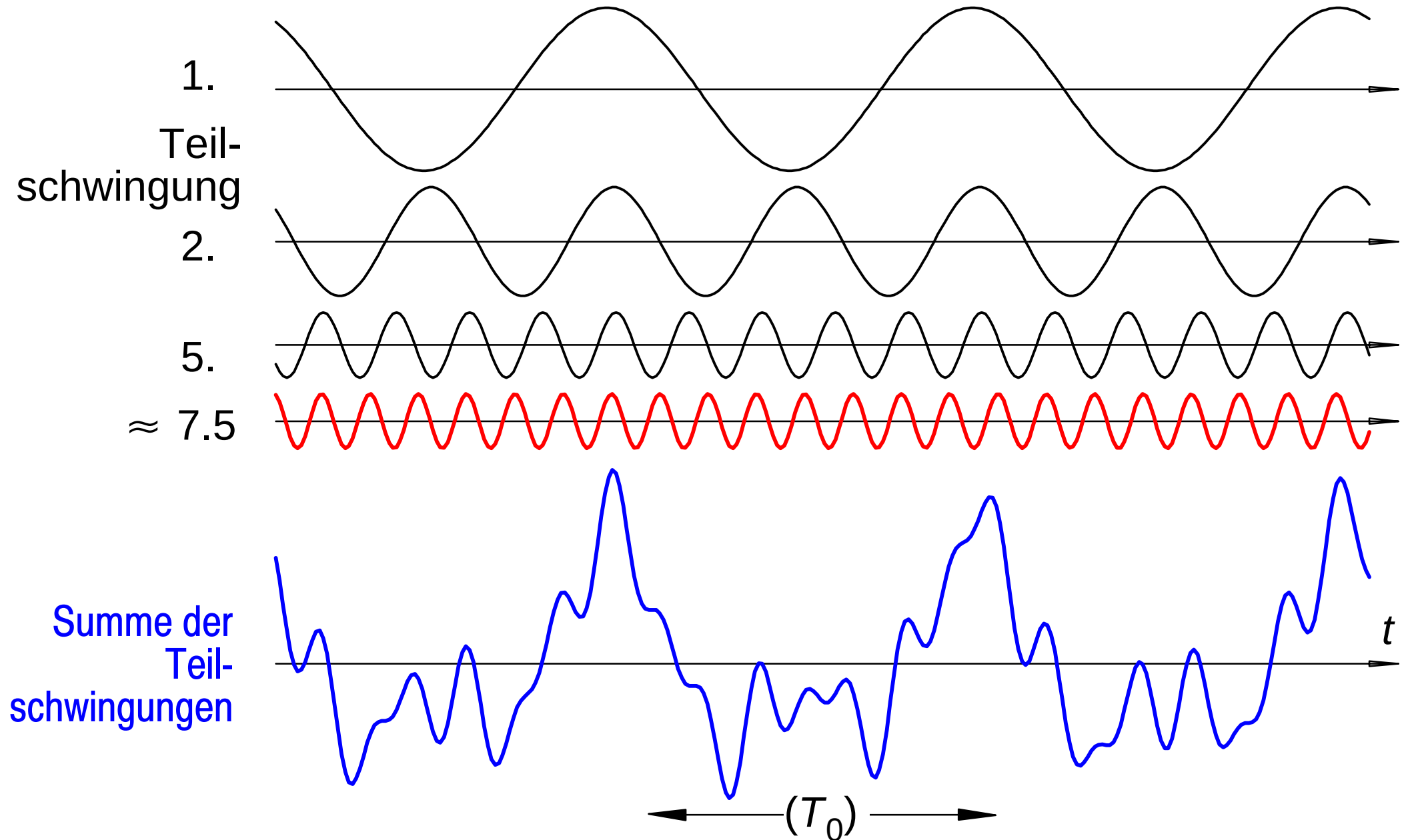
$$0, f_0, 2f_0, 3f_0, \dots mf_0 \dots, \quad m \text{ ganzzahlig} \quad (3.6)$$

annehmen. Anteile jeder anderen Frequenz würden die Bedingung verletzen, dass $x_p(t)$ mit der Periode t_0 periodisch ist.

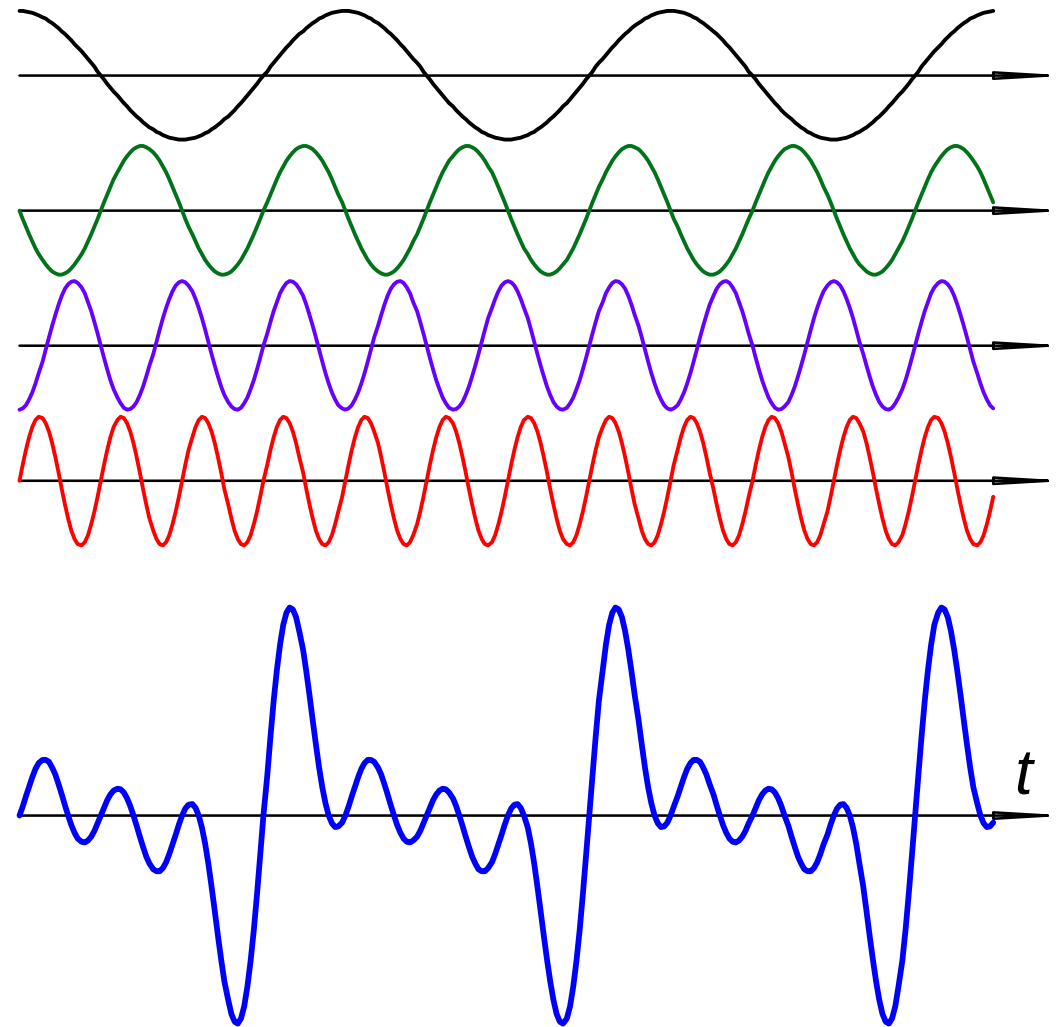
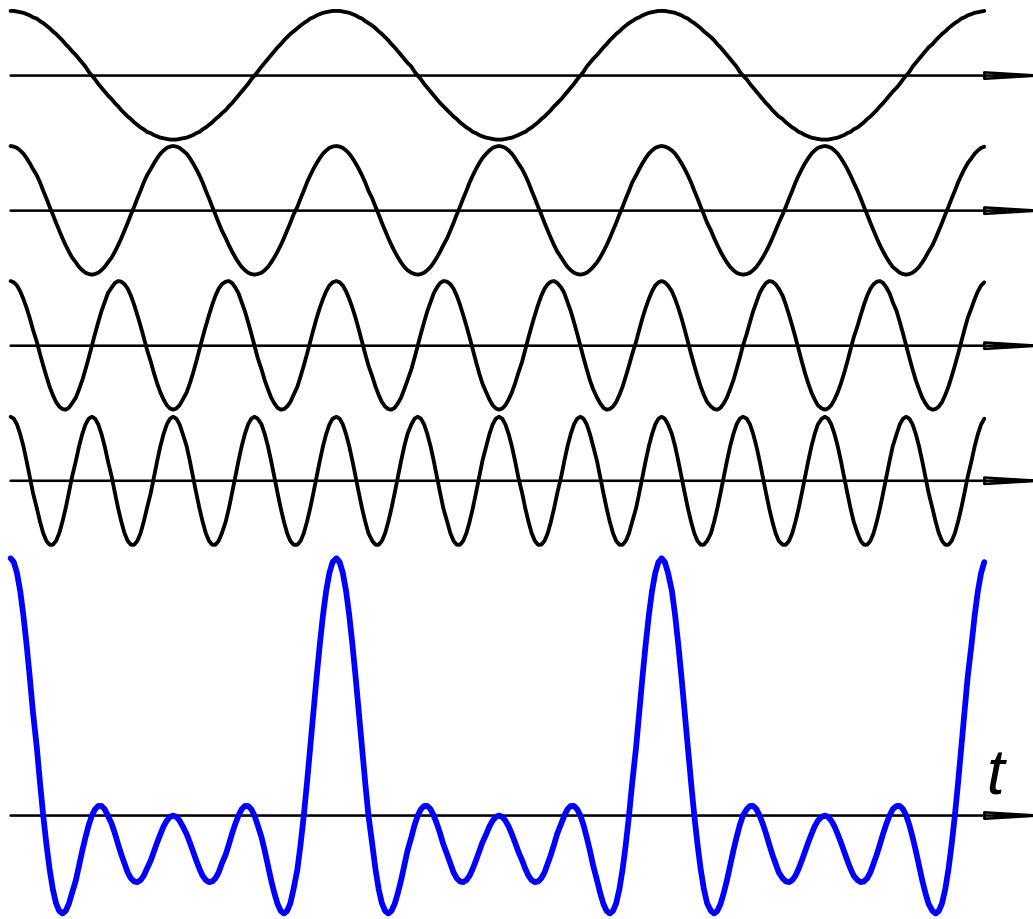
Der Ton als Summe von Sinustönen (Teilschwingungen)



Fällt auch nur eine Teilschwingung aus dem harmonischen Schema, so wird die Gesamtschwingung nicht mehr periodisch.



Phase und Kurvenform



Phasenverschiebungen einzelner Teilschwingungen ändern die Kurvenform der Gesamtschwingung. Dabei kann sich auch der Spitzenwert ändern. Die Energie ändert sich dagegen nicht, wenn nur die Phasen der Teilschwingungen geändert werden.

Der Ton als Summe von Sinustönen: Fourierreihe

Signal $x_P(t)$ als **Fourierreihe** dargestellt:

$$x_P(t) = \frac{a_0}{2} + a_1 \cos(\omega_0 t + \varphi_1) + a_2 \cos(2\omega_0 t + \varphi_2) + \dots + a_m \cos(m\omega_0 t + \varphi_m) + \dots \quad (3.7)$$

Jedes Glied der Reihe lässt sich in einen Kosinus- und einen Sinusanteil zerlegen; Kosinus- und Sinusanteil getrennt zusammengefasst,

$$\begin{aligned} x_{Pg}(t) &= \frac{a_0}{2} + a_{1g} \cos \omega_0 t + a_{2g} \cos 2\omega_0 t + \dots + a_{mg} \cos m\omega_0 t + \dots \\ x_{Pu}(t) &= a_{1u} \sin \omega_0 t + a_{2u} \sin 2\omega_0 t + \dots + a_{mu} \sin m\omega_0 t + \dots \end{aligned} \quad (3.8a,b)$$

Hierbei ist $x_{Pg}(t)$ der *gerade* Anteil von $x_P(t)$; $x_{Pu}(t)$ der *ungerade* Anteil.

Die Darstellung einer periodischen (Zeit-)Funktion mit Hilfe der Fourierreihe ist dann nützlich, wenn es gelingt, die Koeffizienten a_{ig} und a_{iu} bzw. a_i und φ_i aus dem Signal $x_P(t)$ zu gewinnen. Dies wird von der Fourierreihe geleistet.

Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

3.1 Harmonische Analyse (diskreter) periodischer Signale

3.1.1 Sinoidalschwingung (Sinusschwingung) als Grundbaustein akustischer Signale

3.1.2 Orthogonalitätsbeziehung und Fourierkoeffizienten

3.1.3 Exkurs: Was benötigen wir an Mathematik in diesem Kapitel?

3.1.4 Übergang zur komplexen Exponentialfunktion. Die diskrete Fouriertransformation (DFT)

3.1.5 Einige besondere Zeitfunktionen und ihre Spektren

3.2 Nichtharmonische Spektralanalyse

3.3 Kurzzeitspektren von Sprachsignalen

3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

Orthogonalitätsbeziehung der trigonometrischen Funktionen

Für analoge Signale:

$$\int_0^{2\pi} \cos kx \cos mx \, dx = \int_0^{2\pi} \sin kx \sin mx \, dx = \begin{cases} \pi & k = m \\ 0 & k \neq m \end{cases} \quad (3.9a,b)$$

$$\int_0^{2\pi} \cos kx \sin mx \, dx = 0 \quad (3.9c)$$

Für digitale Signale [Summenorthogonalität]:

$$\sum_{n=0}^{N-1} \cos \frac{k \cdot 2\pi n}{N} \cos \frac{m \cdot 2\pi n}{N} = \begin{cases} N/2 & k = m \\ 0 & k \neq m \end{cases} \quad (3.10a)$$

$$\sum_{n=0}^{N-1} \sin \frac{k \cdot 2\pi n}{N} \sin \frac{m \cdot 2\pi n}{N} = \begin{cases} N/2 & k = m \\ 0 & k \neq m \end{cases} \quad (3.10b)$$

$$\sum_{n=0}^{N-1} \cos \frac{k \cdot 2\pi n}{N} \sin \frac{m \cdot 2\pi n}{N} = 0 \quad (3.10c)$$

Herleitung der Orthogonalitätsbeziehung [1]

$$\int_0^{2\pi} \cos kx \cos mx \, dx = \begin{cases} \pi & k = m \\ 0 & k \neq m \end{cases} \quad (3.9a)$$

$$\cos kx \cos mx = \frac{1}{2} [\cos(k-m)x + \cos(k+m)x]$$

$$\begin{aligned} \int_0^{2\pi} \cos kx \cos mx \, dx &= \int_0^{2\pi} \left[\frac{1}{2} [\cos(k-m)x + \cos(k+m)x] \right] dx \\ &= \left[\frac{1}{2(k-m)} \sin(k-m)x + \frac{1}{2(k+m)} \sin(k+m)x \right] \bigg|_0^{2\pi} = 0 \end{aligned}$$

k≠m

$$\int_0^{2\pi} \cos kx \cos mx \, dx = \int_0^{2\pi} \left[\frac{1}{2} [1 + \cos 2kx] \right] dx = \left[\frac{x}{2} + \frac{1}{4k} \sin 2kx \right] \bigg|_0^{2\pi} = \pi$$

k=m

Herleitung der Orthogonalitätsbeziehung [2]

$$\int_0^{2\pi} \sin kx \sin mx \, dx = \begin{cases} \pi & k = m \\ 0 & k \neq m \end{cases} \quad (3.9b)$$

$$\sin kx \sin mx = \frac{1}{2} [\cos(k-m)x - \cos(k+m)x]$$

$$\begin{aligned} \int_0^{2\pi} \sin kx \sin mx \, dx &= \int_0^{2\pi} \left[\frac{1}{2} [\cos(k-m)x - \cos(k+m)x] \right] dx \\ &= \left[\frac{1}{2(k-m)} \sin(k-m)x - \frac{1}{2(k+m)} \sin(k+m)x \right] \bigg|_0^{2\pi} = 0 \end{aligned}$$

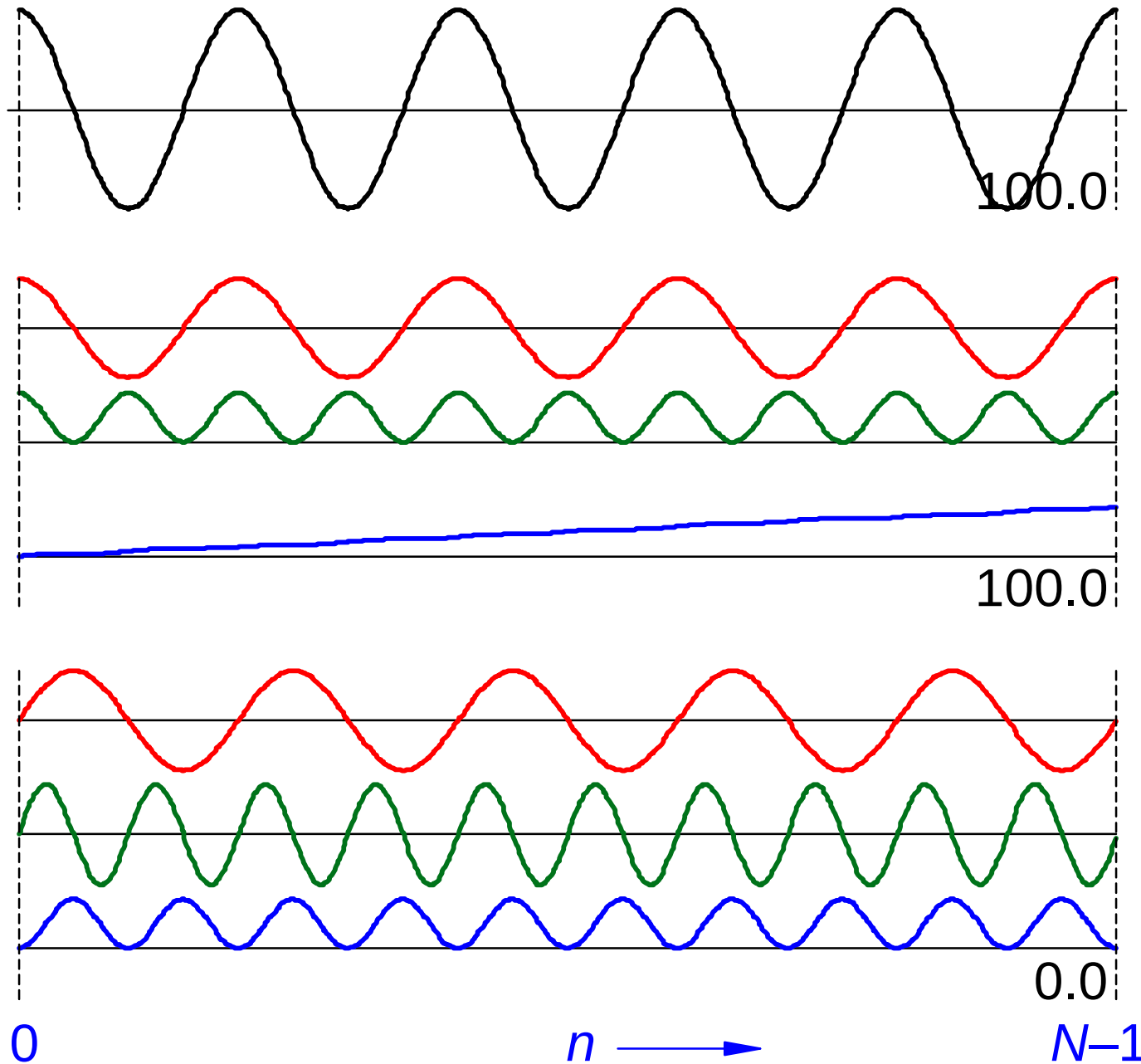
k≠m

$$\int_0^{2\pi} \sin kx \sin mx \, dx = \int_0^{2\pi} \left[\frac{1}{2} [1 - \cos 2kx] \right] dx = \left[\frac{x}{2} - \frac{1}{4k} \sin 2kx \right] \bigg|_0^{2\pi} = \pi$$

k=m

Orthogonalitätsbeziehung

Beispiel Kosinus $k=5$; $m=5$; $A=100$



$$x_P(n) = A \cos(k \cdot Q)$$
$$[Q = 2\pi n/N]$$

$$\cos(m \cdot Q)$$

$$x_P(n) \cos(m \cdot Q)$$

$$\frac{2}{N} \sum_{r=0}^n x_P(r) \cos(m \cdot Q)$$

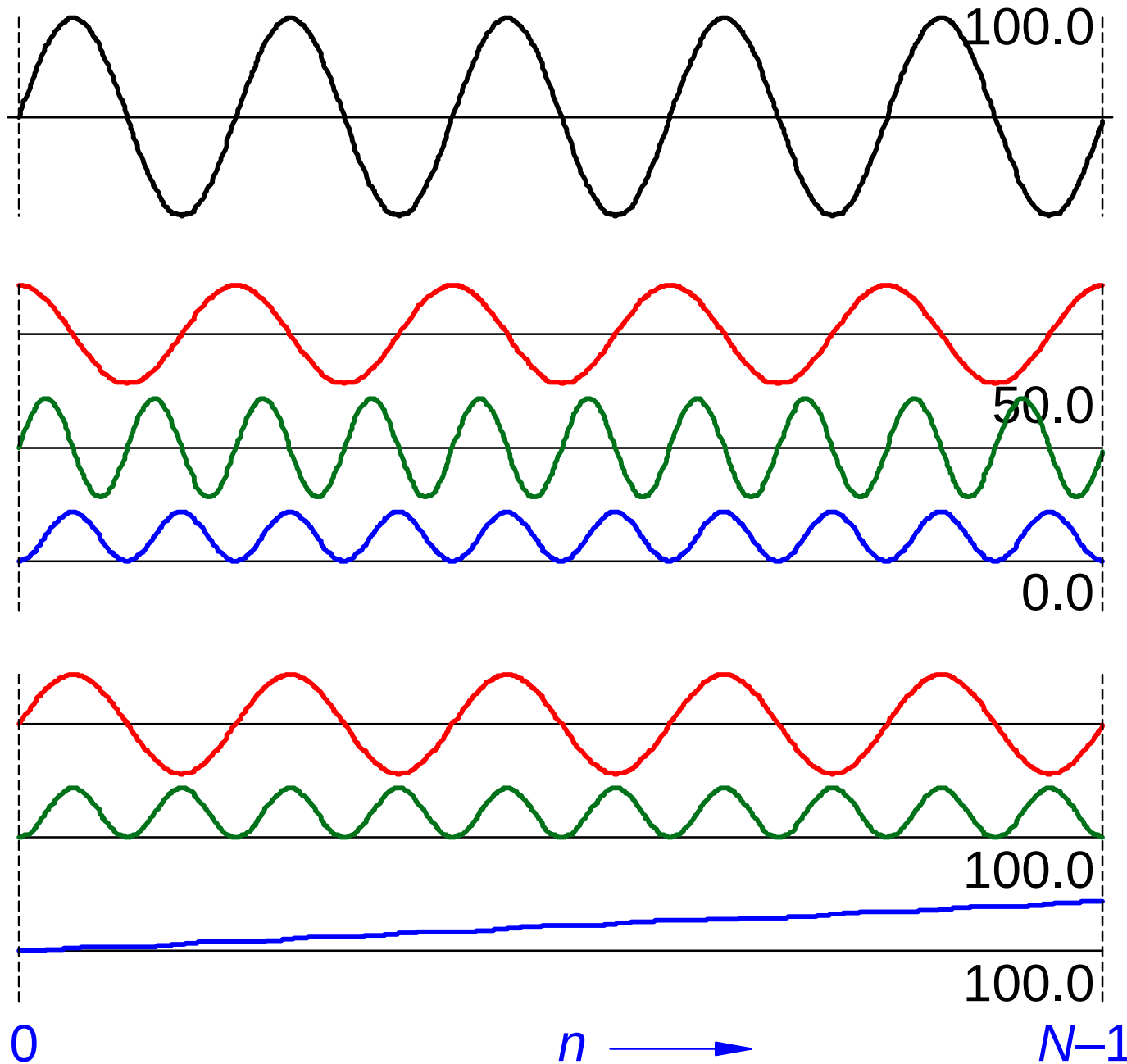
$$\sin(m \cdot Q)$$

$$x_P(n) \sin(m \cdot Q)$$

$$\frac{2}{N} \sum_{r=0}^n x_P(r) \sin(m \cdot Q)$$

Orthogonalitätsbeziehung

Beispiel Sinus $k=5$; $m=5$; $A=100$



$$x_P(n) = A \sin(k \cdot Q)$$
$$[Q = 2\pi n/N]$$

$$\cos(m \cdot Q)$$

$$x_P(n) \cos(m \cdot Q)$$

$$\frac{2}{N} \sum_{r=0}^n x_P(r) \cos(m \cdot Q)$$

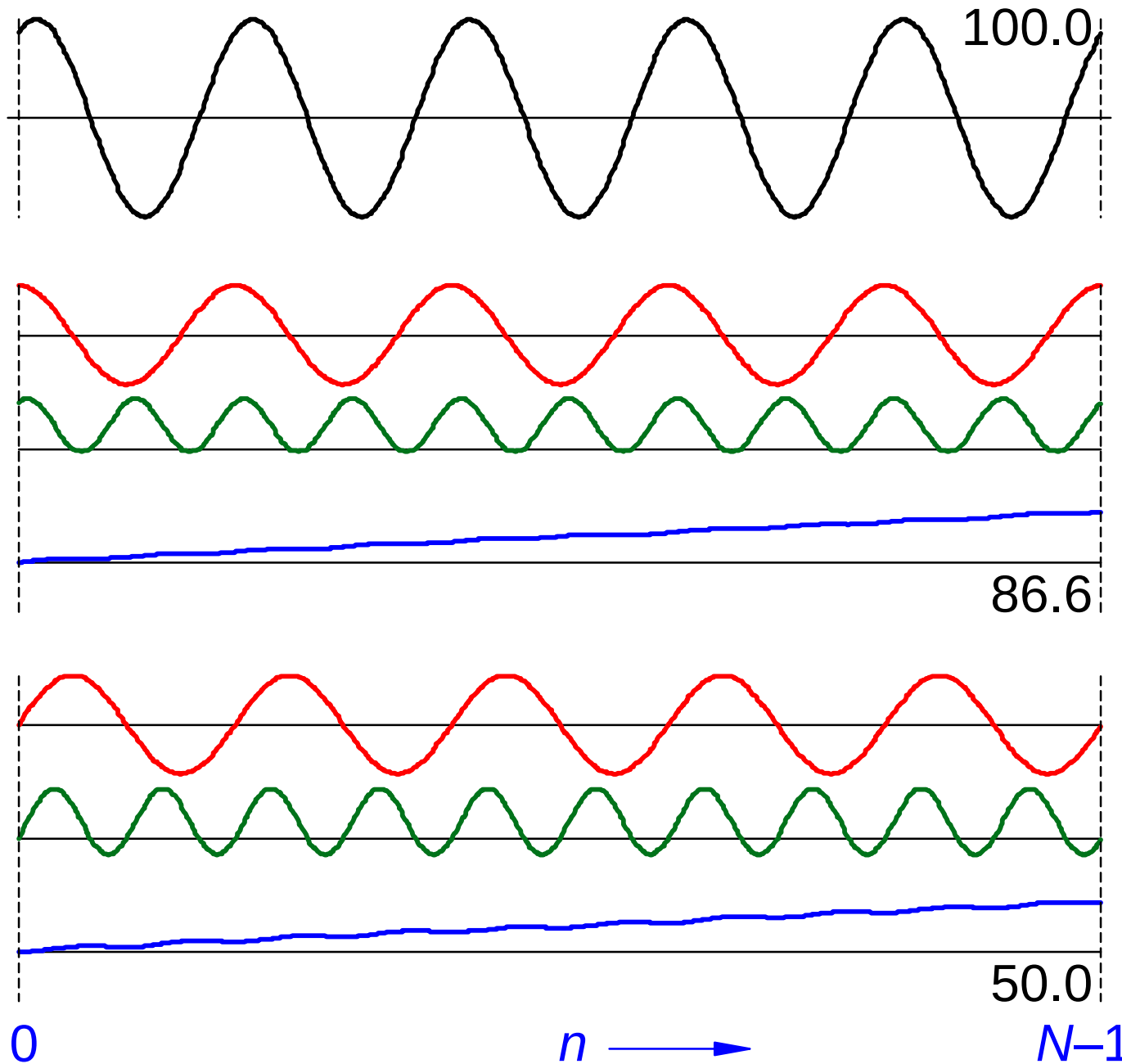
$$\sin(m \cdot Q)$$

$$x_P(n) \sin(m \cdot Q)$$

$$\frac{2}{N} \sum_{r=0}^n x_P(r) \sin(m \cdot Q)$$

Orthogonalitätsbeziehung

Beispiel Teilschwingung $k=5$; $m=5$; $A=100$; $\varphi=\pi/6$



$$x_P(n) = A \cos(k \cdot Q + \varphi)$$
$$[Q = 2\pi n/N]$$

$$\cos(m \cdot Q)$$

$$x_P(n) \cos(m \cdot Q)$$

$$\frac{2}{N} \sum_{r=0}^n x_P(r) \cos(m \cdot Q)$$

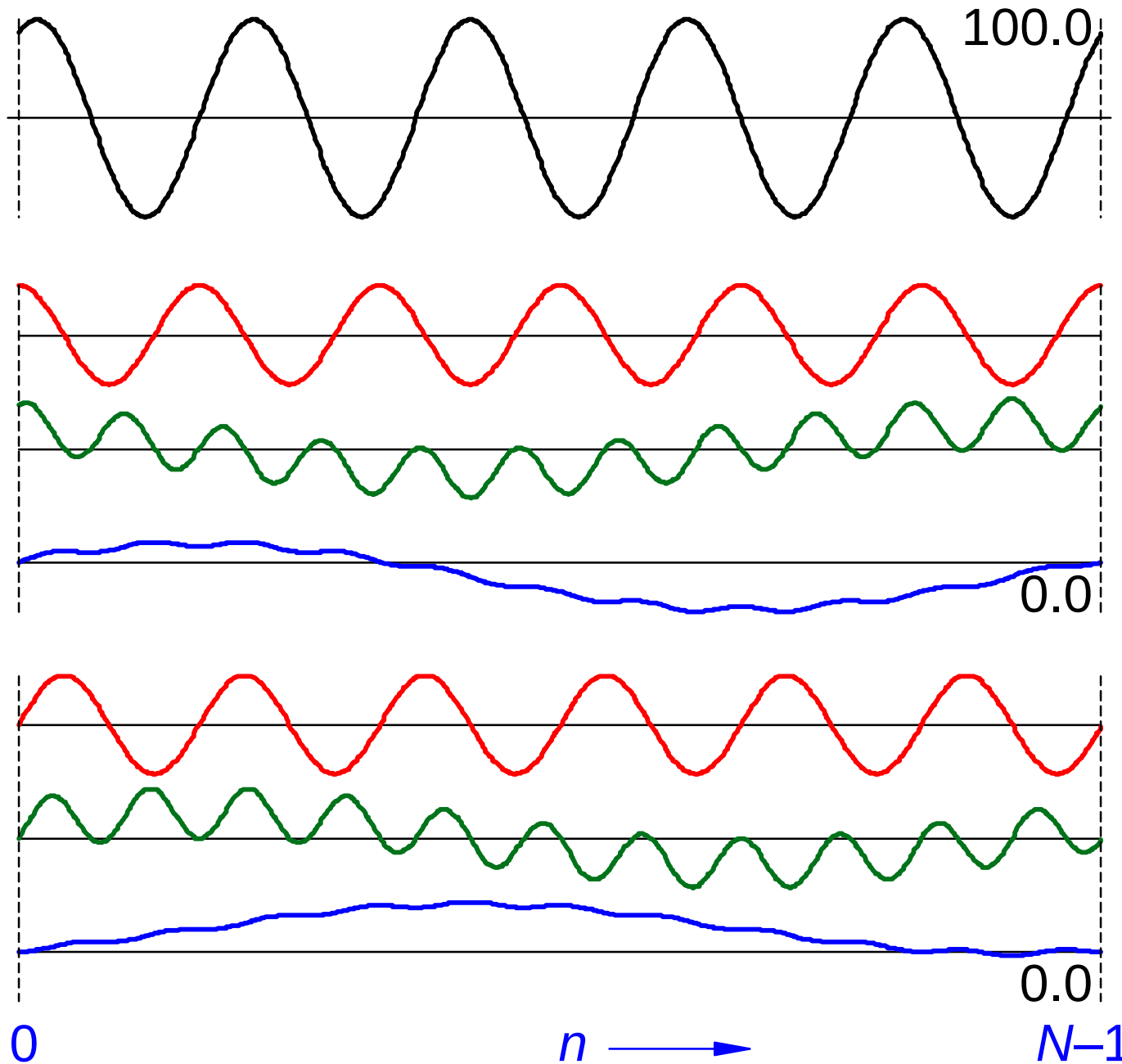
$$\sin(m \cdot Q)$$

$$x_P(n) \sin(m \cdot Q)$$

$$\frac{2}{N} \sum_{r=0}^n x_P(r) \sin(m \cdot Q)$$

Orthogonalitätsbeziehung

Beispiel Teilschwingung $k=5$; $m=6$; $A=100$; $\varphi=\pi/6$



$$x_P(n) = A \cos(k \cdot Q + \varphi)$$
$$[Q = 2\pi n/N]$$

$$\cos(m \cdot Q)$$

$$x_P(n) \cos(m \cdot Q)$$

$$\frac{2}{N} \sum_{r=0}^n x_P(r) \cos(m \cdot Q)$$

$$\sin(m \cdot Q)$$

$$x_P(n) \sin(m \cdot Q)$$

$$\frac{2}{N} \sum_{r=0}^n x_P(r) \sin(m \cdot Q)$$

Berechnung der Fourierkoeffizienten (analoger Fall)

Orthogonalitätsbedingung verallgemeinert:

$$\int_0^{t_0} \cos k\omega_0 t \cos m\omega_0 t dt = \int_0^{t_0} \sin k\omega_0 t \sin m\omega_0 t dt = \begin{cases} t_0/2 & k = m \\ 0 & k \neq m \end{cases}$$

Hieraus Berechnung der Fourierkoeffizienten **getrennt für den geraden und den ungeraden Anteil der Zeitfunktion:**

$$\begin{aligned} a_{mg} &= \frac{2}{t_0} \int_0^{t_0} x_P(t) \cos m\omega_0 t dt, \quad m = 1 (1) \dots & a_0 &= \frac{2}{t_0} \int_0^{t_0} x_P(t) dt \\ a_{mu} &= \frac{2}{t_0} \int_0^{t_0} x_P(t) \sin m\omega_0 t dt, \quad m = 1 (1) \dots & & \end{aligned} \quad (3.11a-c)$$

Der Ton als Summe von Sinustönen: Fourierreihe für diskrete Signale

Signal $x_P(n)$ als **Fourierreihe** dargestellt:

$$x_P(n) = \frac{a_0}{2} + a_1 \cos\left[\frac{2\pi n}{N} + \varphi_1\right] + a_2 \cos\left[\frac{2 \cdot 2\pi n}{N} + \varphi_2\right] \dots + a_{N/2} \cos\left[\frac{(N/2) \cdot 2\pi n}{N} + \varphi_{N/2}\right]$$

Jedes Glied der Reihe in geraden (Kosinus-) und ungeraden (Sinus-) Anteil zerlegen; Kosinus- und Sinusanteil getrennt zusammengefasst,

$$x_{Pg}(n) = \frac{a_0}{2} + a_{1g} \cos \frac{2\pi n}{N} + a_{2g} \cos \frac{2 \cdot 2\pi n}{N} + \dots + a_{N/2, g} \cos \frac{(N/2) \cdot 2\pi n}{N}$$
$$x_{Pu}(n) = a_{1u} \sin \frac{2\pi n}{N} + a_{2u} \sin \frac{2 \cdot 2\pi n}{N} + \dots + a_{N/2, u} \sin \frac{(N/2) \cdot 2\pi n}{N}$$

Wegen der Bandbegrenzung digitaler Signale auf die halbe Abtastfrequenz (Abtasttheorem!) ist die Fourierreihe hier endlich und auf den Bereich $\{0 \dots N/2\}$ begrenzt, wobei die ungeraden Anteile für 0 und $N/2$ verschwinden.

Berechnung der Fourierkoeffizienten (digitaler Fall)

Summenorthogonalität

$$\begin{aligned} \sum_{n=0}^{N-1} \cos \frac{k \cdot 2\pi n}{N} \cos \frac{m \cdot 2\pi n}{N} &= \begin{cases} N/2 & k = m \\ 0 & k \neq m \end{cases} \\ \sum_{n=0}^{N-1} \sin \frac{k \cdot 2\pi n}{N} \sin \frac{m \cdot 2\pi n}{N} &= \begin{cases} N/2 & k = m \\ 0 & k \neq m \end{cases} \end{aligned} \quad (3.10)$$

Hieraus Berechnung der Fourierkoeffizienten:

$$\begin{aligned} a_{mg} &= \frac{2}{N} \sum_{n=0}^{N-1} x(n) \cos \frac{m \cdot 2\pi n}{N}, \quad m = 0 (1) \frac{N}{2} \\ a_{mu} &= \frac{2}{N} \sum_{n=1}^{N-1} x(n) \sin \frac{m \cdot 2\pi n}{N}, \quad m = 1 (1) \frac{N}{2}-1; \quad a_{0u} = 0 \end{aligned} \quad (3.12)$$

Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

3.1 Harmonische Analyse (diskreter) periodischer Signale

3.1.1 Sinoidalschwingung (Sinusschwingung) als Grundbaustein akustischer Signale

3.1.2 Orthogonalitätsbeziehung und Fourierkoeffizienten

3.1.3 Exkurs: Was benötigen wir an Mathematik in diesem Kapitel?

3.1.4 Übergang zur komplexen Exponentialfunktion. Die diskrete Fouriertransformation (DFT)

3.1.5 Einige besondere Zeitfunktionen und ihre Spektren

3.2 Nichtharmonische Spektralanalyse

3.3 Kurzzeitspektren von Sprachsignalen

3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

Was benötigen wir hier an Mathematik?

Aus der Schulmathematik -

- Kenntnisse in Differentialrechnung (Regeln der Differentialrechnung, Ableitung transzendenter Funktionen)
- Grundkenntnisse in Integralrechnung (bestimmte Integrale reeller Funktionen)

Bei der Rechnung mit digitalen Signalen reduzieren sich die meisten Integrale auf Summen.

Über die Schulmathematik hinaus -

- Komplexe Zahlen
- Grundkenntnisse aus der Reihenrechnung
- Grundkenntnisse über die komplexe Exponentialfunktion

Auf dem Weg zu den komplexen Zahlen [1]

- Wir benötigen die Rechnung mit komplexen Zahlen, um Schwingungen mit Hilfe der komplexen Exponentialfunktion beschreiben zu können.
- Diese Beschreibungsweise ist, sobald man sie einmal kennt, die bei weitem eleganteste, universell einsetzbar und wird von Physikern und Ingenieuren ausnahmslos verwendet.

Wichtigstes Prinzip bei der Rechnung mit komplexen Zahlen ist das **Prinzip der Permanenz von Regeln und Operatoren** bei der Erweiterung des Zahlensystems. In anderen Worten: **bei der Erweiterung des Zahlensystems bleiben die Rechenregeln unverändert, und das bisherige Zahlensystem wird ein Sonderfall des erweiterten.**

Beim schrittweisen Übergang von den natürlichen Zahlen auf die komplexen Zahlen gilt das wie im Folgenden beschrieben.

Auf dem Weg zu den komplexen Zahlen [2]

Ausgangspunkt ist die Menge der natürlichen Zahlen:

$$\mathbb{N} = \{1, 2, 3, \dots\}$$

Mit natürlichen Zahlen lassen sich die Grundrechenarten **Addition** und **Multiplikation** sowie ihre Umkehrungen **Subtraktion** und **Division** definieren und ausführen.

Bei der Subtraktion natürlicher Zahlen können Ergebnisse herauskommen, die mit natürlichen Zahlen nicht mehr darzustellen sind.

- Wir erweitern um **Null und die negativen Zahlen** und erhalten die **Menge der ganzen Zahlen**; die natürlichen Zahlen sind eine Teilmenge hiervon.
- Die Rechenregeln gelten unverändert weiter; die Multiplikationsregel wird erweitert: minus mal minus gibt plus, Zahl mal null gibt null. Division durch Null ist nicht möglich.

Auf dem Weg zu den komplexen Zahlen [3]

Bei der Division natürlicher oder ganzer Zahlen können Ergebnisse herauskommen, die mit ganzen Zahlen nicht mehr darzustellen sind.

- Wir erweitern die Menge der ganzen Zahlen um die **gebrochenen Zahlen** und erhalten die **Menge der rationalen Zahlen**; die ganzen Zahlen sind eine Teilmenge hiervon.
- Die Rechenregeln gelten unverändert weiter.

Auf dem Weg zu den komplexen Zahlen [4]

Wir nehmen als neue Grundrechenart das **Potenzieren** hinzu:

$$(\text{Potenz}) = (\text{Basis})^{(\text{Exponent})}$$

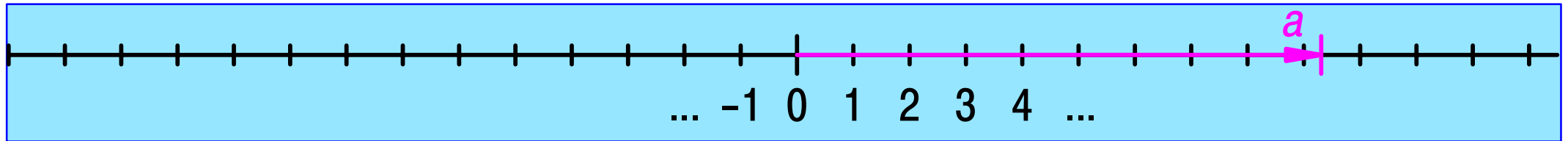
Ist die Basis rational und der Exponent ganzzahlig, so bleiben alle Potenzen im Rahmen der rationalen Zahlen.

Bei den beiden Umkehrungen **Radizierung** (Basis aus Potenz und Exponent) und **Logarithmierung** (Exponent aus Potenz und Basis) können Ergebnisse herauskommen, die nicht mit rationalen Zahlen darzustellen sind.

- Wir erweitern das Zahlensystem um die **irrationalen Zahlen** und erhalten die **Menge der reellen Zahlen**; die rationalen Zahlen sind eine Teilmenge hiervon.
- Die Rechenregeln gelten unverändert weiter; Radizieren und Logarithmieren werden (zumindest für positive Zahlen) möglich.
- Auch Potenzieren mit nicht ganzzahligen Exponenten wird möglich.

Auf dem Weg zu den komplexen Zahlen [5]

Mit den reellen Zahlen ist die Zahlengerade komplett. Jede Zahl ist geometrisch darstellbar als Strecke vom Nullpunkt der Zahlengeraden bis zu dem Punkt, der durch die Zahl markiert ist.



Die Quadratwurzel negativer Zahlen kann durch eine reelle Zahl nicht angegeben werden.

Grundproblem hierbei ist: welchen Wert besitzt $\sqrt{-1}$?

Das Problem wird gelöst, indem die Menge der reellen Zahlen um die **imaginären Zahlen** erweitert wird; wir erhalten damit die **Menge der komplexen Zahlen**. Die reellen Zahlen sind eine Teilmenge hiervon.

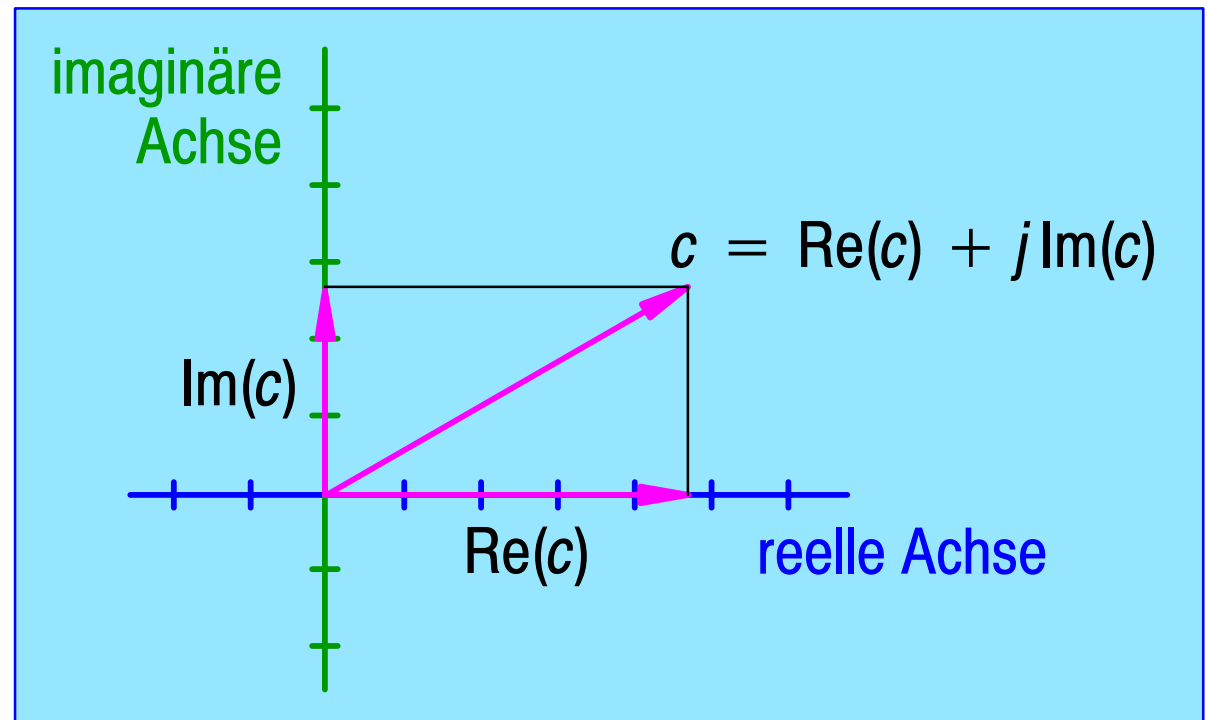
Auf dem Weg zu den komplexen Zahlen [6]

Aus der eindimensionalen Zahlengeraden wird die **zweidimensionale Zahlenebene**.

Sie wird beschrieben durch ein kartesisches Koordinatensystem:

- Abszisse: bisherige Zahlengerade der reellen Zahlen (*reelle Achse*)
- Ordinate: neue Zahlengerade der imaginären Zahlen (*imaginäre Achse*).

Jede komplexe Zahl c ist geometrisch darstellbar als Vektor vom Ursprung des Koordinatensystems bis zu dem Punkt in der Zahlenebene, der durch die Zahl markiert ist.



Auf dem Weg zu den komplexen Zahlen [7]

Dementsprechend werden die bisherigen Rechenregeln für komplexe Zahlen erweitert.

- Addition und Subtraktion folgen den Gesetzen der Vektorrechnung.
- Wichtig ist die alternative Darstellung im Polarkoordinatensystem durch Betrag und (Phasen-)Winkel.
- Multiplikation, Division und auch Potenzieren laufen in dieser Darstellung ab; z.B. bedeutet Multiplikation zweier komplexer Zahlen die Multiplikation der Beträge und die Addition der Winkel.
- Die Rechenoperationen für reelle Zahlen folgen genau den gleichen (erweiterten) Gesetzmäßigkeiten.

Im Rahmen der komplexen Zahlen lassen sich alle Probleme der Grundrechenarten und ihrer Umkehrungen lösen; einer nochmaligen Erweiterung des Zahlensystems bedarf es nicht mehr.

Reihenrechnung [1]

- Bei digitalen Signalen haben wir es häufig mit Summen zu tun (etwa über Signalwerte). Um solche Summen behandeln zu können, bedarf es der Rechnung mit Reihen.
- In der Mathematik ist die Reihenrechnung ein ganzer Wissenschaftszweig; wir brauchen davon nur einige wenige Grundkenntnisse. Diese sind im Folgenden beschrieben.

Definition: Eine Reihe ist eine Summe von Gliedern (Summanden). Jedes Glied (außer dem ersten) geht durch ein bestimmtes Bildungsgesetz aus dem vorangehenden Glied hervor.

Wir kennen endliche und unendliche Reihen; eine unendliche Reihe hat unendlich viele Glieder.

Beispiel: Unendliche geometrische Reihe

$$s = 1 + q + q^2 + q^3 + \dots + q^k + \dots = \sum_{n=0}^{\infty} q^n$$

Konvergenz: Eine unendliche Reihe ist konvergent, wenn die Summe einen endlichen Wert ergibt.

Wir müssen die Konvergenz im Einzelfall nachweisen.

Bei unserer geometrischen Reihe ist dies leicht möglich. Wir berechnen eine endliche geometrische Reihe nach der Summenformel

$$s_k = \sum_{n=0}^k q^n = \frac{1 - q^{k+1}}{1 - q}$$

Ist nun $|q| < 1$, so kann die Zahl k der Glieder gegen ∞ streben:

$$s = \sum_{n=0}^{\infty} q^n = \lim_{k \rightarrow \infty} \frac{1 - q^{k+1}}{1 - q} = \frac{1}{1 - q} ; \quad |q| < 1$$

Der Bereich, in dem eine Reihe konvergiert, wird **Konvergenzbereich** genannt.

Eine Reihe ist **absolut konvergent**, wenn sich jedes Glied durch seinen Betrag ersetzen lässt und die Reihe trotzdem konvergiert.

Unsere geometrische Reihe ist absolut konvergent für $|q| < 1$.

Merke: Konvergenz und absolute Konvergenz gelten auch im Komplexen!

Majorantenprinzip: Eine Reihe ist konvergent, wenn sich zu ihr eine andere konvergente Reihe (Majorante) finden lässt, deren sämtliche Glieder ab einem bestimmten Index (betragsmäßig) gleich oder größer sind als die entsprechenden Glieder der untersuchten Reihe.

Für viele Probleme in der Signalverarbeitung, die Reihen fordern, lässt sich eine konvergente geometrische Reihe als Majorante finden.

Komplexe Exponentialfunktion [1]

- Die **komplexe Exponentialfunktion** (also die Exponentialfunktion zur Basis e mit komplexem Exponenten) erlaubt eine einheitliche und elegante Darstellung ungedämpfter, exponentiell gedämpfter und exponentiell anklingender Schwingungen mit einer einzigen Funktion und führt zudem Kosinus und Sinus auf die Exponentialfunktion zurück.

Die Theorie der komplexen Exponentialfunktion ist kompliziert und sprengt den Rahmen dieser Vorlesung.

Kennen müssen wir

- die Eulersche Formel
- die Tatsache, dass die üblichen Gesetze der Potenzrechnung auch im Komplexen gelten.

Komplexe Exponentialfunktion [2]

Euler'sche Formel

$$\exp(jx) = e^{jx} = \cos x + j \sin x ; \quad j = \sqrt{-1} \quad (3.14)$$

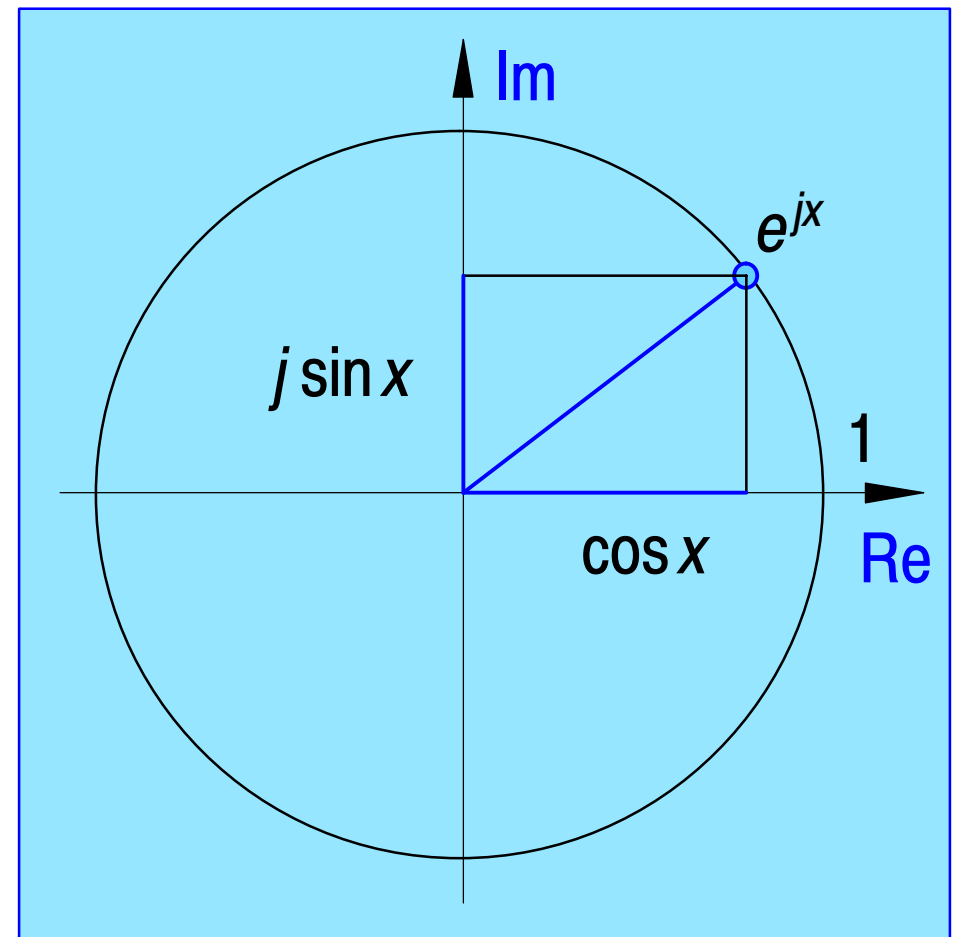
Der Wert e^{jx} markiert einen Punkt auf dem Einheitskreis (um den Ursprung) der komplexen Zahlenebene.

Rückrechnung in Kosinus bzw. Sinus:

$$\cos x = \frac{e^{jx} + e^{-jx}}{2} = \operatorname{Re}(e^{jx})$$

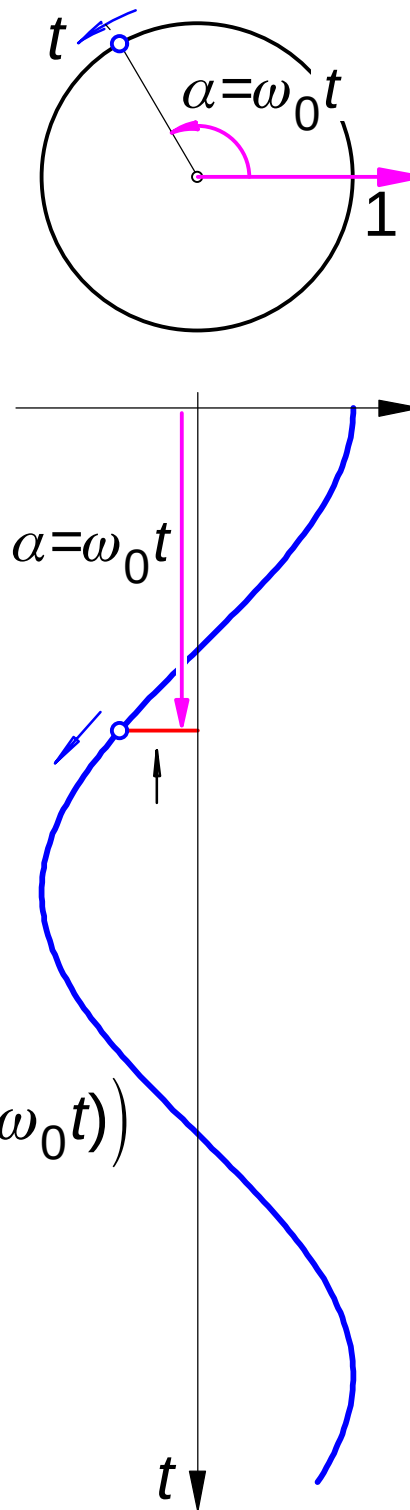
(3.15)

$$\sin x = \frac{e^{jx} - e^{-jx}}{2j} = \operatorname{Im}(e^{jx})$$



Die Sinoidalschwingung
(der Sinuston)
resultiert aus der
Rotation eines Zeigers
 $\exp(j\alpha) = \exp(j\omega_0 t)$ auf
dem Einheitskreis.

$$\begin{aligned}\cos \alpha &= \\ \cos \omega_0 t &= \\ \operatorname{Re} \left(\exp(j\omega_0 t) \right)\end{aligned}$$



Bewegt sich ein Punkt
 $\exp(j\alpha) = \exp(j\omega_0 t)$
gleichförmig auf dem
Umfang des
Einheitskreises, so
verläuft sein Realteil
in Form einer
Kosinusschwingung.

Komplexe Exponentialfunktion [3]

Nachweis der Euler'schen Formel

$$\exp(jx) = e^{jx} = \cos x + j \sin x ; \quad j = \sqrt{-1}$$

Der Nachweis erfolgt über den Ansatz der Taylorreihe. Hierbei sagen uns die Konvergenzbedingungen der Taylorreihe für die Exponentialfunktion und die trigonometrischen Funktionen, dass die vorgenommenen Operationen erlaubt sind. Nachstehend die Formeln; die Taylorentwicklung erfolgt für alle Funktionen $f(x)$ um den Punkt $x=0$ [MACLAURIN-Formel].

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(k)}(0)}{k!}x^k + \dots \quad (\text{im Konvergenzbereich})$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \dots ; \quad |x| < \infty$$

$$e^{jx} = 1 + jx - \frac{x^2}{2!} - j\frac{x^3}{3!} + \frac{x^4}{4!} + j\frac{x^5}{5!} + \dots ; \quad |x| < \infty$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} \mp \dots ; \quad \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} \mp \dots ; \quad |x| < \infty$$

Komplexe Exponentialfunktion [4]

Komplexes Argument

Es sei $c = a + jb$. Dann ist

$$e^c = e^{a+jb} = e^a \cdot e^{jb} = e^a \cdot (\cos b + j \sin b)$$

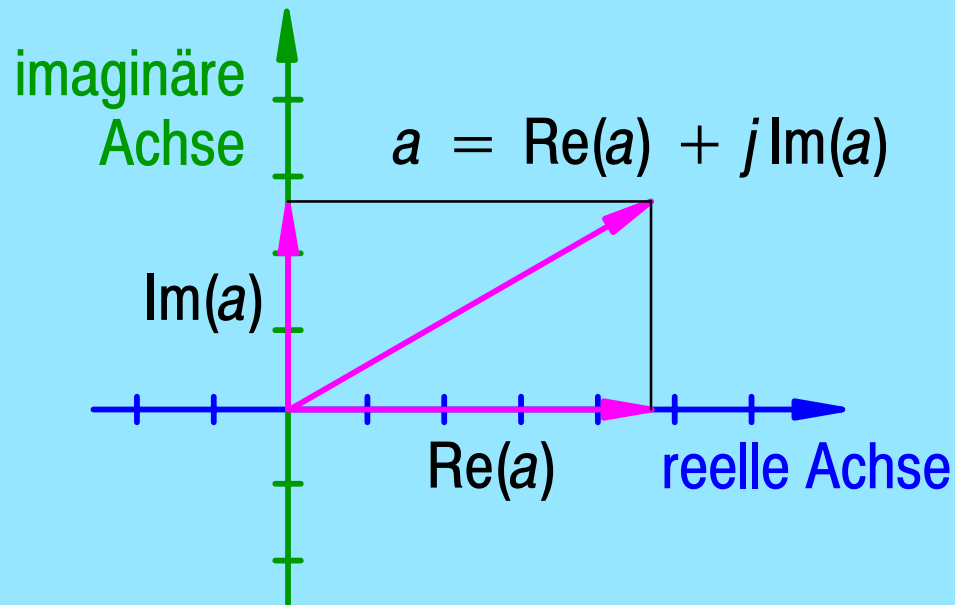
Der erste Term e^a hat ein reelles Argument und ist deshalb reell und stets positiv. Der zweite Term e^{jb} hat ein rein imaginäres Argument und ist nach der Eulerschen Formel zu behandeln; dementsprechend zerfällt er in einen reellen Kosinusanteil und einen imaginären Sinusanteil.

Ist c eine Funktion der Zeit, so können wir damit eine **exponentiell gedämpfte oder anklingende Schwingung** darstellen:

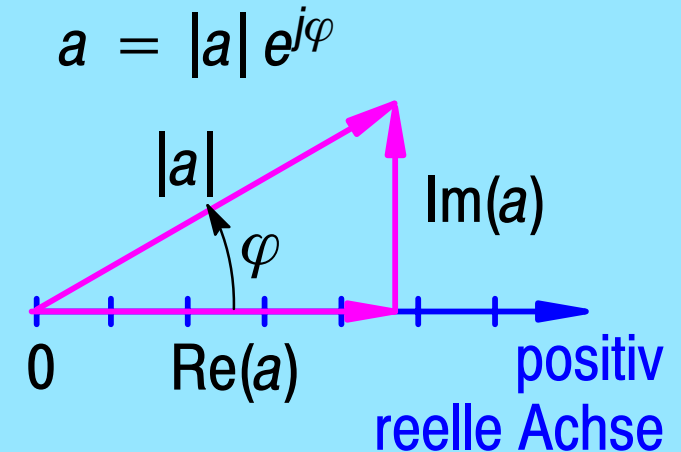
$$e^{(\alpha + j\omega)t} = e^{\alpha t} \cdot e^{j\omega t} = e^{\alpha t} \cdot (\cos \omega t + j \sin \omega t)$$

Darstellung komplexer Zahlen

Jede komplexe Zahl ist geometrisch darstellbar als Vektor vom Ursprung des Koordinatensystems bis zu dem Punkt in der Zahlenebene, der durch die Zahl markiert ist.



Darstellung in kartesischen Koordinaten



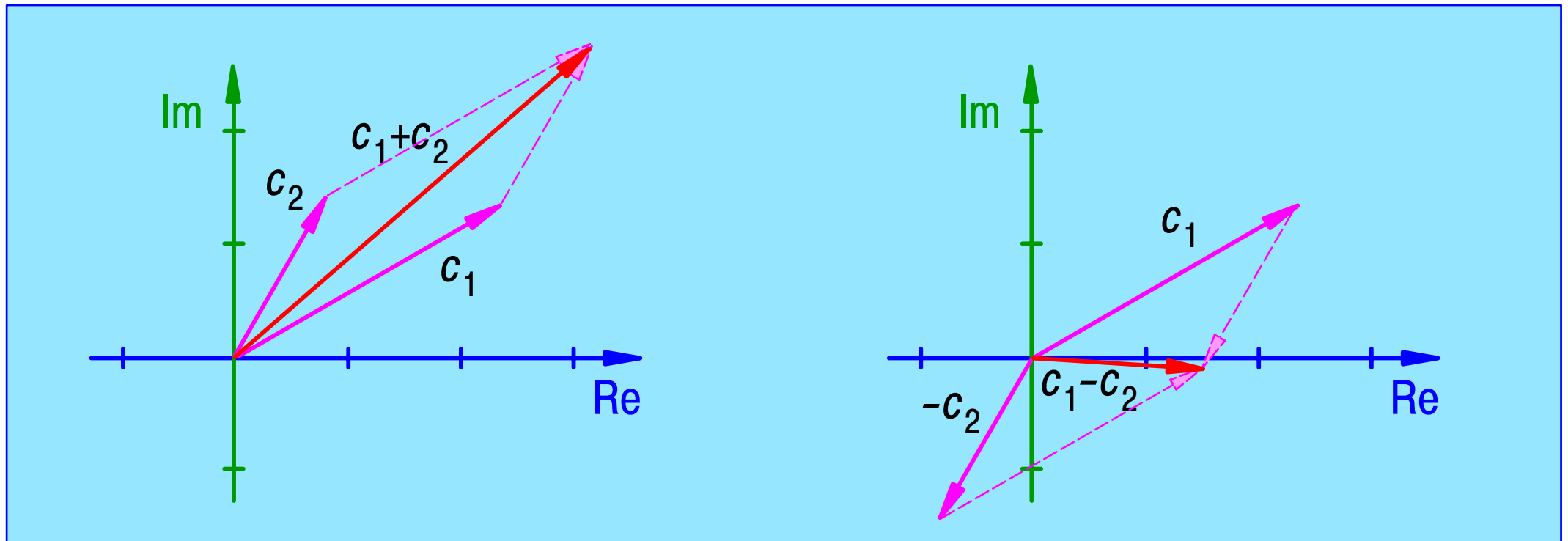
Darstellung in Polarkoordinaten

$$a = \text{Re}(a) + j \text{Im}(a) = |a| (\cos \varphi + j \sin \varphi) = |a| e^{j\varphi}$$

$$|a| = \sqrt{(\text{Re}(a))^2 + (\text{Im}(a))^2}$$

Addition komplexer Zahlen

Addition und Subtraktion komplexer Zahlen folgen den üblichen Regeln; in der Zahlenebene ergibt sich eine Vektoraddition.

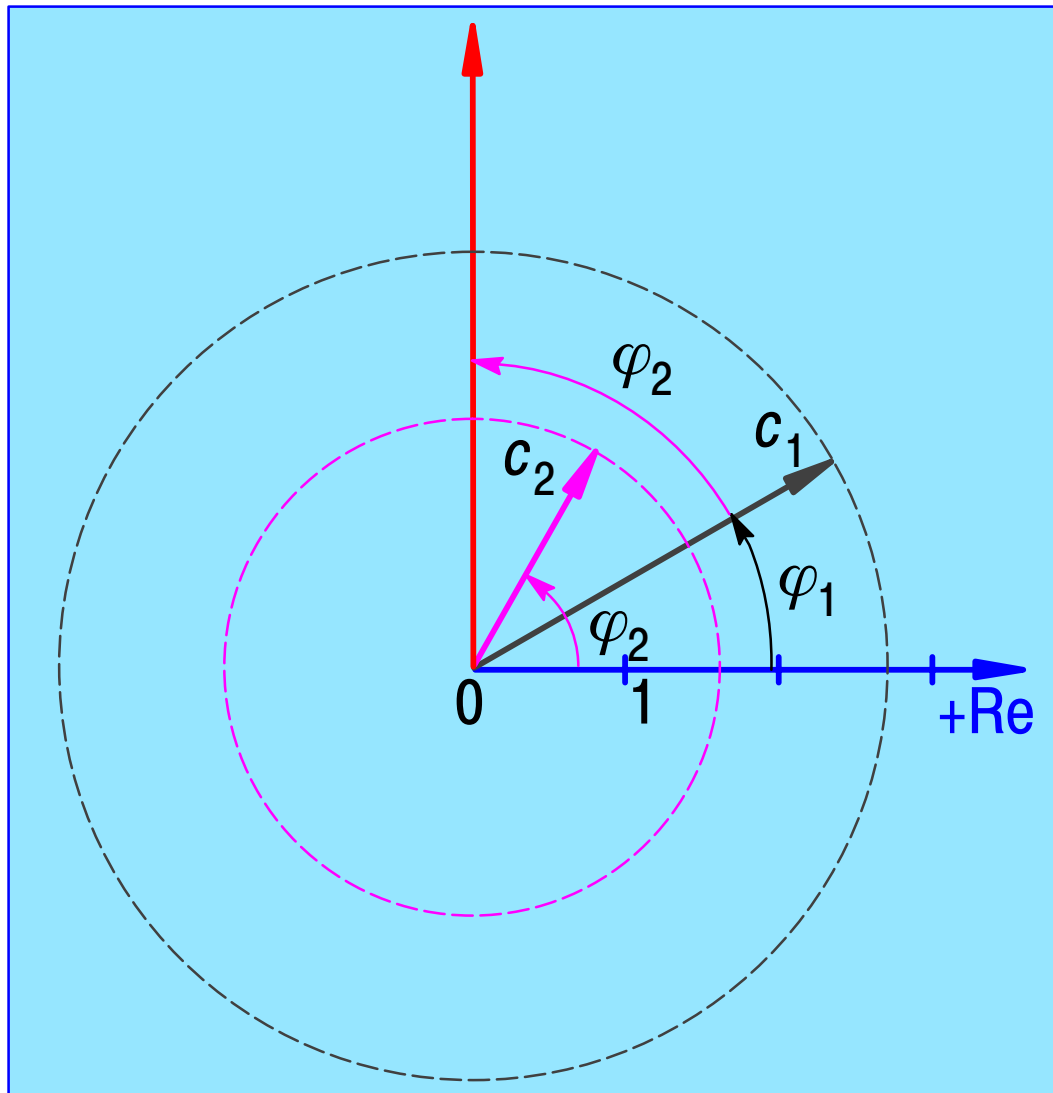


$$c_1 := a_1 + jb_1; \quad c_2 := a_2 + jb_2$$

$$c = c_1 \pm c_2 = (a_1 \pm a_2) + j(b_1 \pm b_2)$$

Multiplikation komplexer Zahlen

Das Produkt zweier komplexer Zahlen ergibt sich durch Multiplikation der Beträge und Addition der Winkel.



$$c_1 := a_1 + jb_1 = |a_1| e^{j\varphi_1}$$

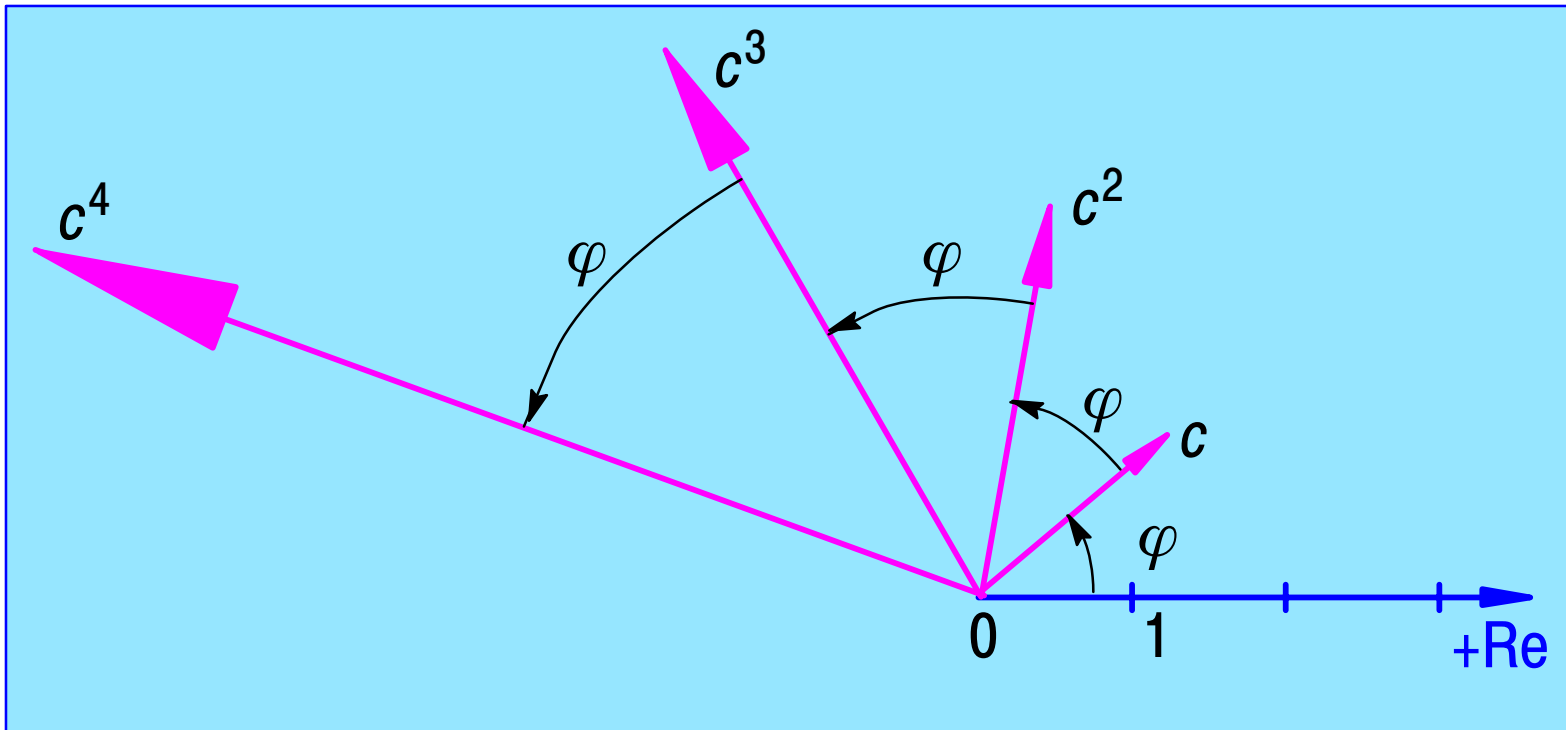
$$c_2 := a_2 + jb_2 = |a_2| e^{j\varphi_2}$$

$$\begin{aligned} c &= c_1 \cdot c_2 \\ &= |c_1| e^{j\varphi_1} \cdot |c_2| e^{j\varphi_2} \\ &= |c_1| |c_2| e^{j(\varphi_1 + \varphi_2)} \end{aligned}$$

Der Quotient zweier komplexer Zahlen ergibt sich entsprechend durch Division der Beträge und Subtraktion der Winkel.

Potenzen komplexer Zahlen

Die ganzzahlige Potenz einer komplexen Zahl ergibt sich durch Potenzieren des Betrags und Multiplizieren des Winkels mit dem Exponenten.

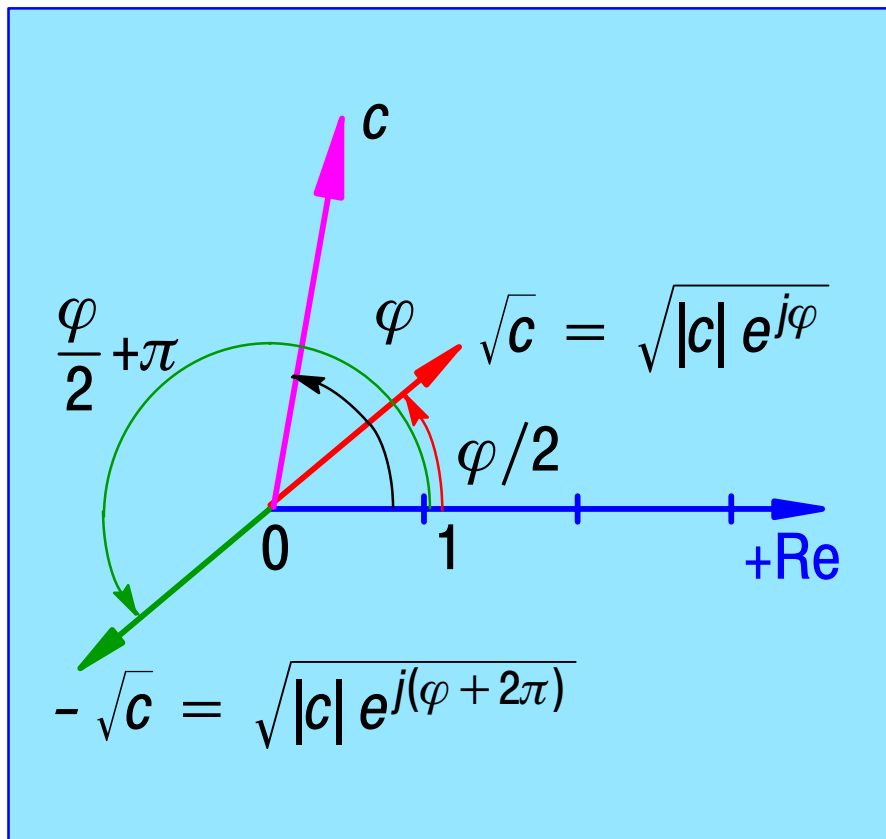


$$c := |c| e^{j\varphi}$$
$$c^k = |c|^k e^{k \cdot j\varphi}$$

Insbesondere gilt $j^2 = -1$; $j^3 = -j$; $j^4 = 1$.

Wurzeln aus komplexen Zahlen

Die k -te Wurzel aus einer komplexen Zahl ergibt sich durch Radizieren des Betrags und Dividieren des Winkels durch k ; wegen der Periodizität von $e^{j\varphi}$ mit 2π ergeben sich k verschiedene Lösungen; die Lösung für $m=0$ wird als *Hauptwert* bezeichnet.



$$\sqrt[k]{c} = c^{\left(\frac{1}{k}\right)} = \sqrt[k]{|c|} \cdot \exp\left[\frac{\varphi + m \cdot 2\pi}{k}\right]$$
$$m = 0, 1, \dots, k-1$$

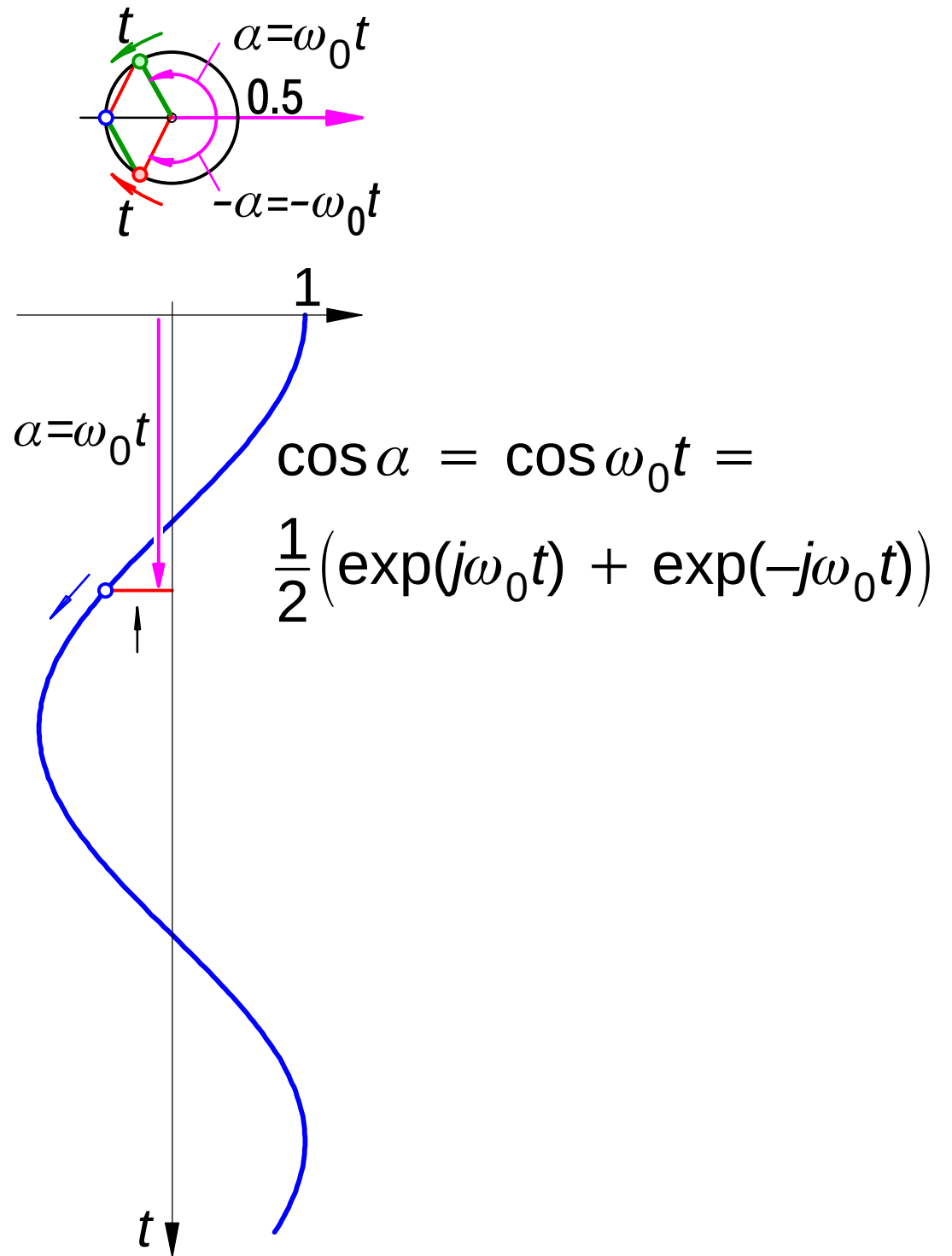
Insbesondere gilt

$$\sqrt{1} = \pm 1 ; \quad \sqrt{-1} = \pm j ;$$

$$\sqrt[3]{1} = \left\{ 1, \exp\left[\frac{2\pi j}{3}\right], \exp\left[\frac{4\pi j}{3}\right] \right\} ;$$

$$\sqrt[4]{1} = \{1, j, -1, -j\} .$$

Die Sinoidalschwingung
(der Sinuston)
resultiert aus der
Summe zweier
gegensinnig
rotierender Zeiger
der Länge 0,5.



Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

3.1 Harmonische Analyse (diskreter) periodischer Signale

3.1.1 Sinoidalschwingung (Sinusschwingung) als Grundbaustein akustischer Signale

3.1.2 Orthogonalitätsbeziehung und Fourierkoeffizienten

3.1.3 Exkurs: Was benötigen wir an Mathematik in diesem Kapitel?

3.1.4 Übergang zur komplexen Exponentialfunktion. Die diskrete Fouriertransformation (DFT)

3.1.5 Einige besondere Zeitfunktionen und ihre Spektren

3.2 Nichtharmonische Spektralanalyse

3.3 Kurzzeitspektren von Sprachsignalen

3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

Fourierreihe für diskrete Signale

(siehe Abschnitt 3.1.1/2)

Signal $x_P(n)$ als **Fourierreihe** dargestellt:

$$x_P(n) = \frac{a_0}{2} + a_1 \cos \left[\frac{2\pi n}{N} + \varphi_1 \right] + a_2 \cos \left[\frac{2 \cdot 2\pi n}{N} + \varphi_2 \right] \dots + a_{N/2} \cos \left[\frac{(N/2) \cdot 2\pi n}{N} + \varphi_{N/2} \right]$$

Darstellung nach geradem und ungeradem Anteil:

$$x_{Pg}(n) = \frac{a_0}{2} + a_{1g} \cos \frac{2\pi n}{N} + a_{2g} \cos \frac{2 \cdot 2\pi n}{N} + \dots + a_{N/2,g} \cos \frac{(N/2) \cdot 2\pi n}{N} \quad (3.13)$$

$$x_{Pu}(n) = a_{1u} \sin \frac{2\pi n}{N} + a_{2u} \sin \frac{2 \cdot 2\pi n}{N} + \dots + a_{N/2,u} \sin \frac{(N/2) \cdot 2\pi n}{N}$$

Koeffizienten:

$$a_{mg} = \frac{2}{N} \sum_{n=0}^{N-1} x(n) \cos \frac{m \cdot 2\pi n}{N}; \quad a_{mu} = \frac{2}{N} \sum_{n=1}^{N-1} x(n) \sin \frac{m \cdot 2\pi n}{N}, \quad m = 1 (1) \frac{N}{2}$$
$$a_{0u} = a_{N/2,u} = 0 \quad (3.12)$$

Fourierreihe für diskrete Signale

Ansatz über komplexe Exponentialfunktion [1]

$$\begin{aligned}a_{mg} \cos \frac{2\pi nm}{N} &= \frac{a_{mg}}{2} \cdot \left[\exp \frac{2\pi jnm}{N} + \exp \frac{-2\pi jnm}{N} \right] \\&= \frac{a_{mg}}{2} \cdot \left[\exp \frac{-2\pi jnm}{N} + \exp \frac{2\pi jnm}{N} \right] \\&= \frac{a_{mg}}{2} \cdot \left[\exp \frac{-2\pi jnm}{N} + \exp \frac{-2\pi jn \cdot (-m)}{N} \right]\end{aligned}$$

$$\begin{aligned}a_{mu} \sin \frac{2\pi nm}{N} &= \frac{a_{mu}}{2j} \cdot \left[\exp \frac{2\pi jnm}{N} - \exp \frac{-2\pi jnm}{N} \right] \\&= \frac{-j \cdot a_{mu}}{2} \cdot \left[\exp \frac{2\pi jnm}{N} - \exp \frac{-2\pi jnm}{N} \right] \\&= \frac{j \cdot a_{mu}}{2} \cdot \left[\exp \frac{-2\pi jnm}{N} - \exp \frac{2\pi jnm}{N} \right] \\&= \frac{j \cdot a_{mu}}{2} \cdot \left[\exp \frac{-2\pi jnm}{N} - \exp \frac{-2\pi jn \cdot (-m)}{N} \right]\end{aligned}$$

Diskrete Fouriertransformation

Wir spalten die Klammerausdrücke auf und definieren neue Koeffizienten

$$a_{mg} \cos \frac{2\pi nm}{N} = X_g(m) \exp \frac{-2\pi jnm}{N} + X_g(-m) \exp \frac{-2\pi jn \cdot (-m)}{N}, \quad X_g(m) = X_g(-m)$$

$$a_{mu} \sin \frac{2\pi nm}{N} = jX_u(m) \exp \frac{-2\pi jnm}{N} + jX_u(-m) \exp \frac{-2\pi jn \cdot (-m)}{N}, \quad X_u(-m) = -X_u(m)$$

$$\frac{a_0}{2} = X(0) = X_g(0)$$

$$X(m) := X_g(m) + jX_u(m)$$

Wegen der Periodizität der trigonometrischen Anteile mit N

$$\cos \frac{(m+kN) \cdot 2\pi n}{N} = \cos \frac{m \cdot 2\pi n}{N}, \quad \exp \frac{(m+kN) \cdot 2\pi jn}{N} = \exp \frac{m \cdot 2\pi jn}{N}, \quad k \text{ ganzzahlig}$$

kann der Index m von 0 bis $N-1$ laufen; dies ergibt die **diskrete Fouriertransformation**; entsprechend auch ihre Umkehrung

$$X(m) := \sum_{n=0}^{N-1} x(n) \exp \frac{-2\pi jnm}{N} ; \quad x(n) = \frac{1}{N} \sum_{m=0}^{N-1} X(m) \exp \frac{2\pi jnm}{N} \quad (3.20/21)$$

(Komplexes) Spektrum:

$$X(m) := \sum_{n=0}^{N-1} x(n) \exp \frac{-2\pi j n m}{N} ; \quad x(n) = \frac{1}{N} \sum_{m=0}^{N-1} X(m) \exp \frac{2\pi j n m}{N}$$

Amplituden- und Leistungsspektrum:

$$|X(m)| := \sqrt{X_g^2(m) + X_u^2(m)} ; \quad P(m) := X_g^2(m) + X_u^2(m) = |X(m)|^2$$

Logarithmiertes Amplituden- und Leistungsspektrum:

$$L(m) := 10 \lg P(m) = 20 \lg |X(m)| \quad (3.26)$$

Phasenspektrum:

$$\varphi(m) := \arctan \frac{X_u(m)}{X_g(m)} \quad (3.24)$$

Gerader und ungerader Anteil:

$$x(n) = x_g(n) + x_u(n) \rightarrow X(m) = X_g(m) + jX_u(m) \quad (3.22)$$

Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

3.1 Harmonische Analyse (diskreter) periodischer Signale

3.1.1 Sinoidalschwingung (Sinusschwingung) als Grundbaustein akustischer Signale

3.1.2 Orthogonalitätsbeziehung und Fourierkoeffizienten

3.1.3 Exkurs: Was benötigen wir an Mathematik in diesem Kapitel?

3.1.4 Übergang zur komplexen Exponentialfunktion. Die diskrete Fouriertransformation (DFT)

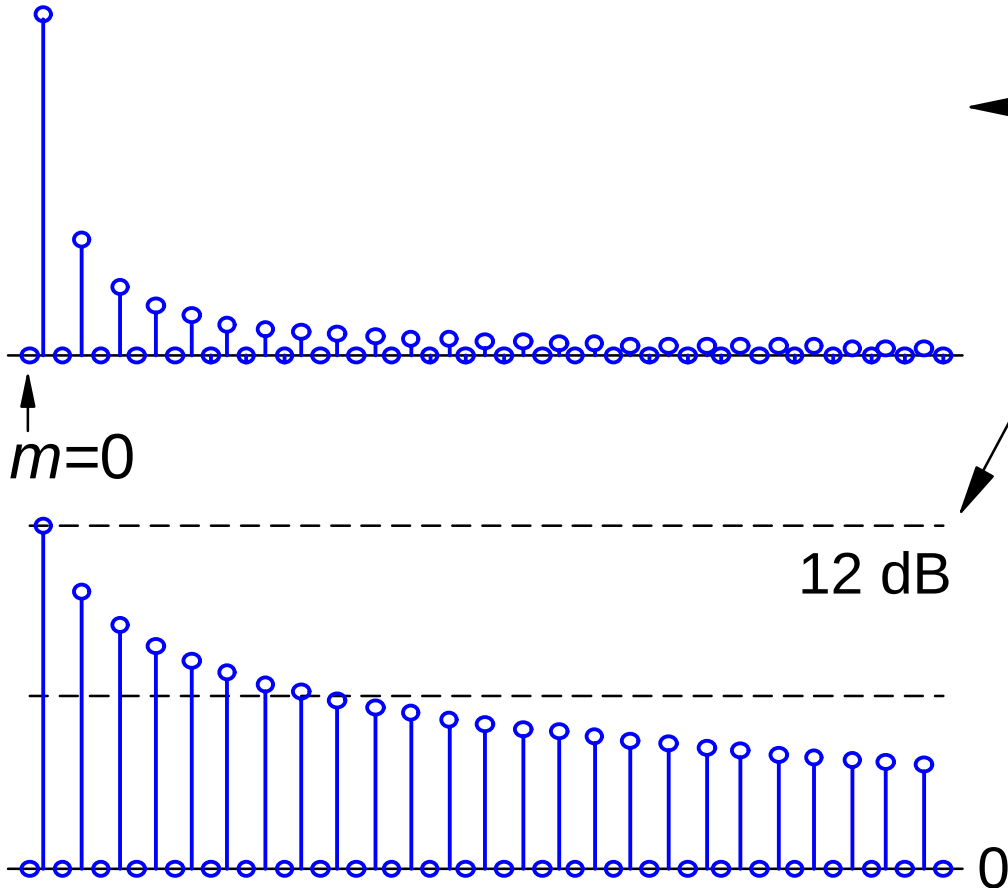
3.1.5 Einige besondere Zeitfunktionen und ihre Spektren

3.2 Nichtharmonische Spektralanalyse

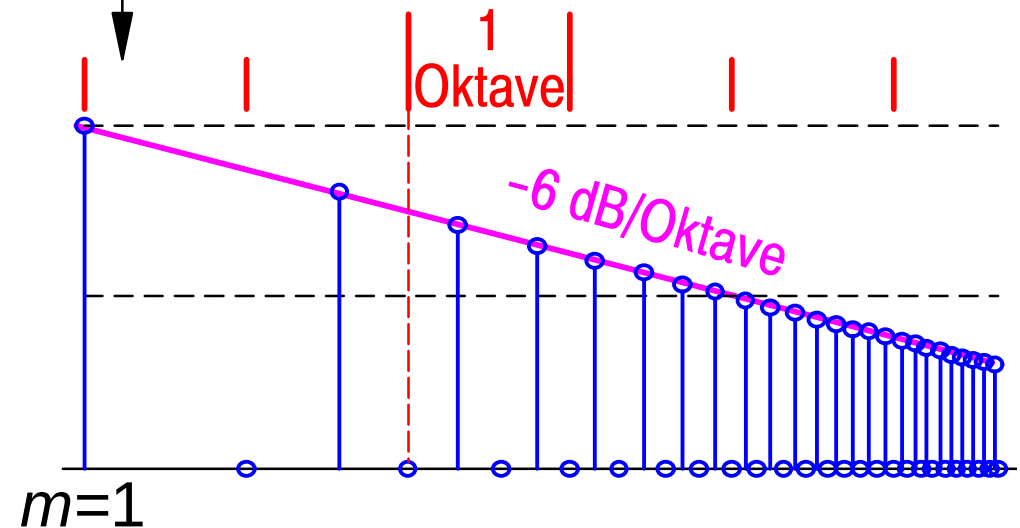
3.3 Kurzzeitspektren von Sprachsignalen

3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

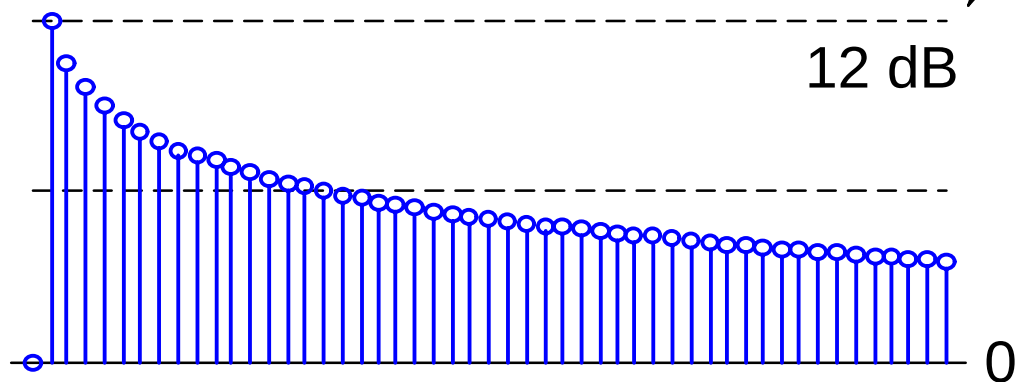
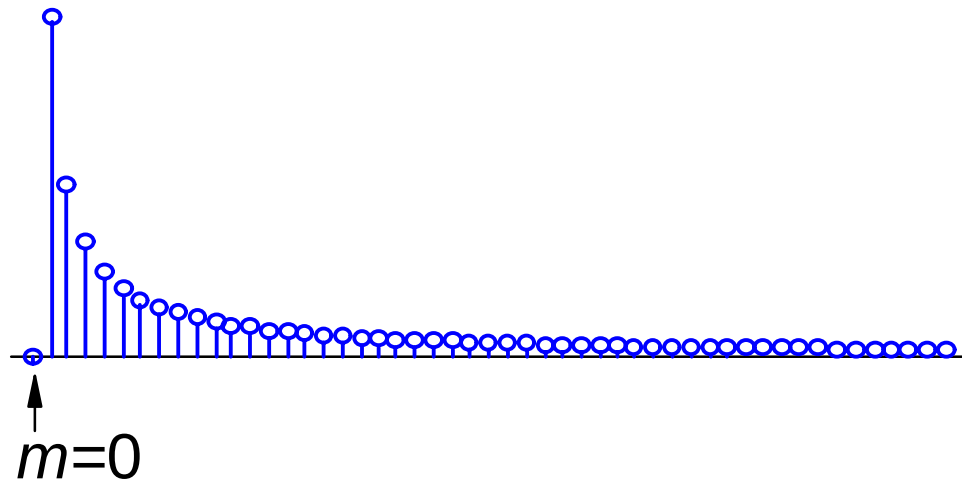
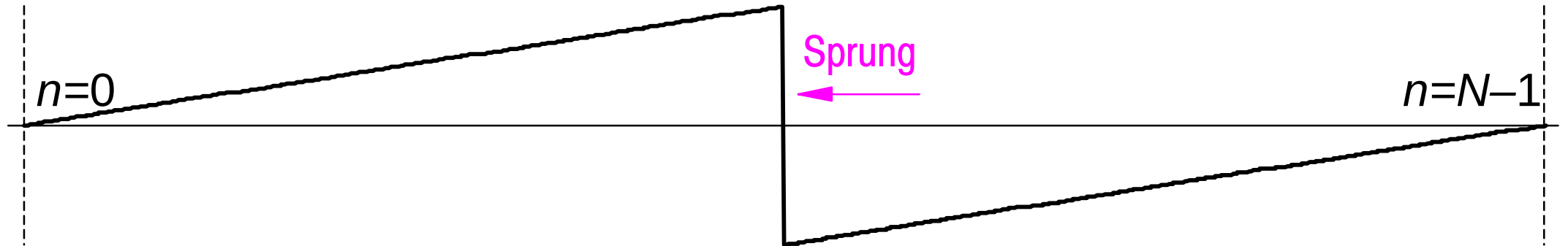
Zeitfunktionen und Spektren - Rechteck



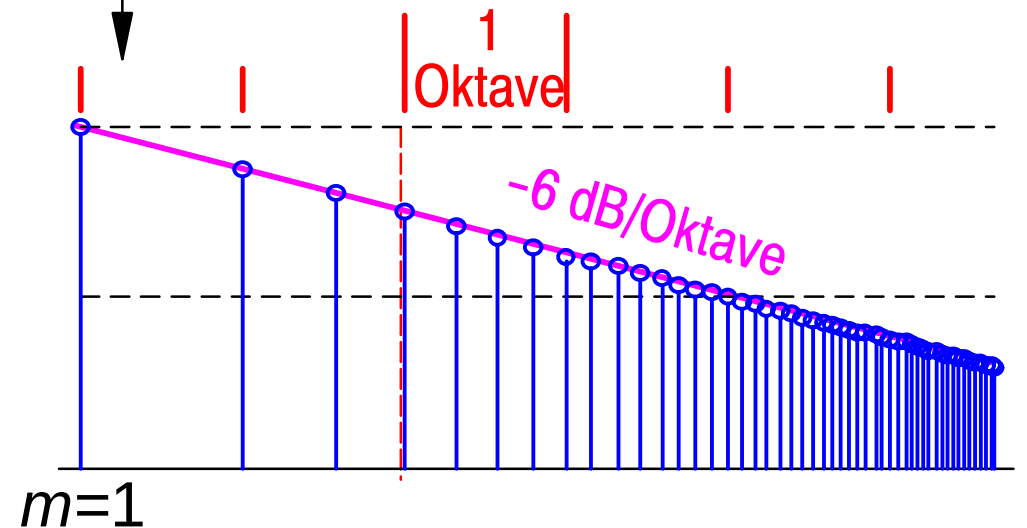
Frequenz: linear Amplitude: linear
Frequenz: linear Amplitude: logar.
Frequenz: logar. Amplitude: logar.



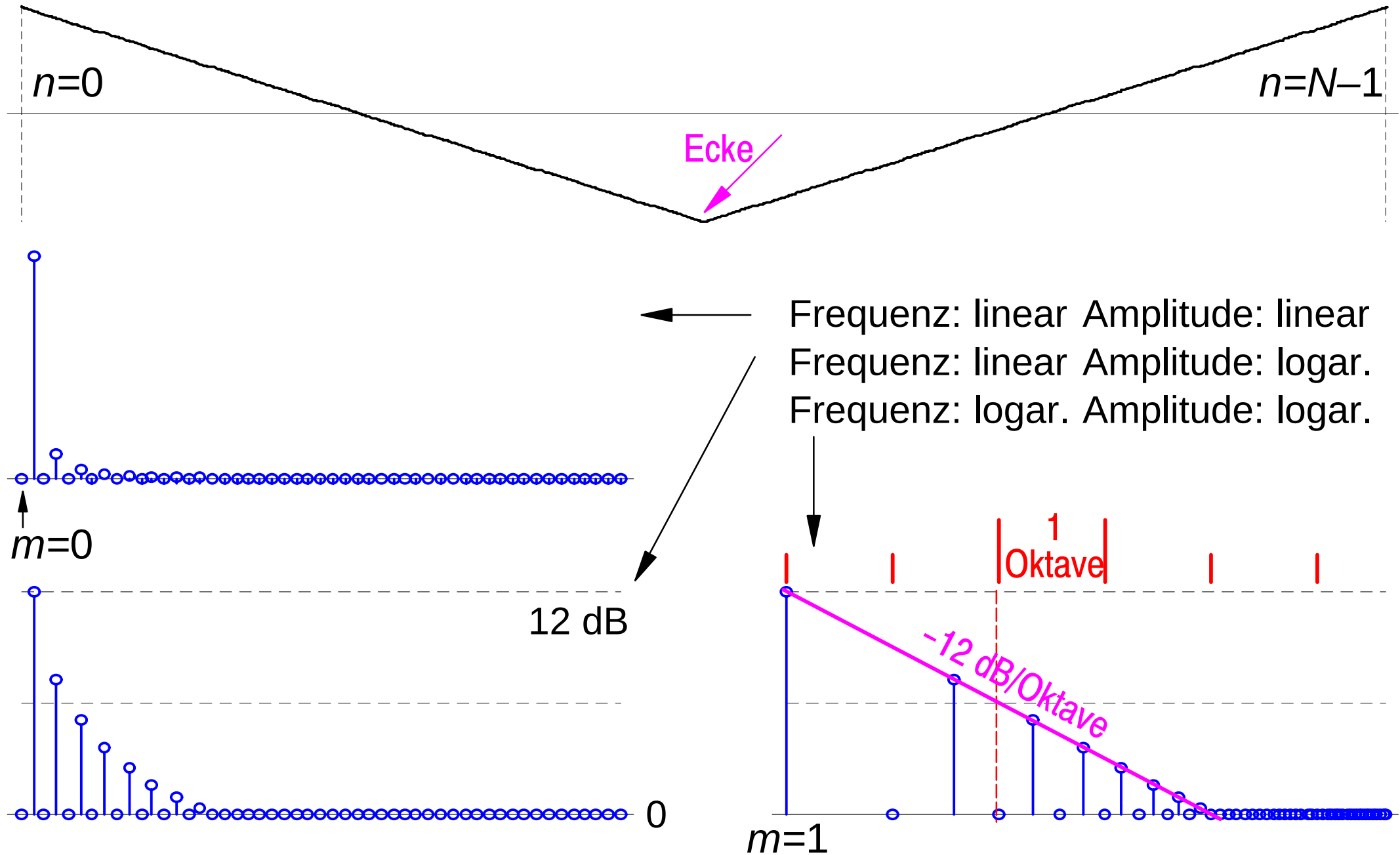
Zeitfunktionen und Spektren - Sägezahn



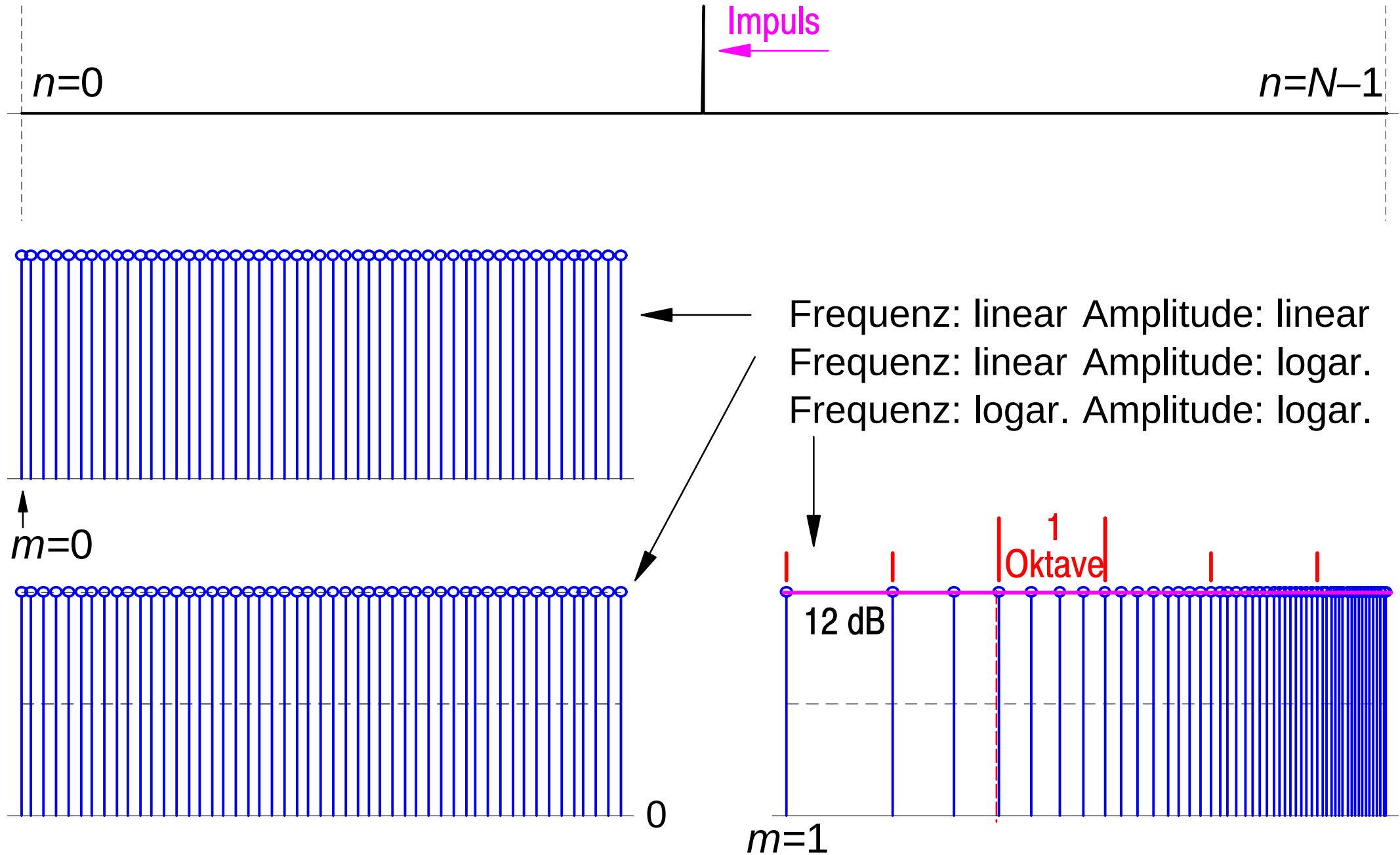
Frequency: linear Amplitude: linear
Frequency: linear Amplitude: logar.
Frequency: logar. Amplitude: logar.



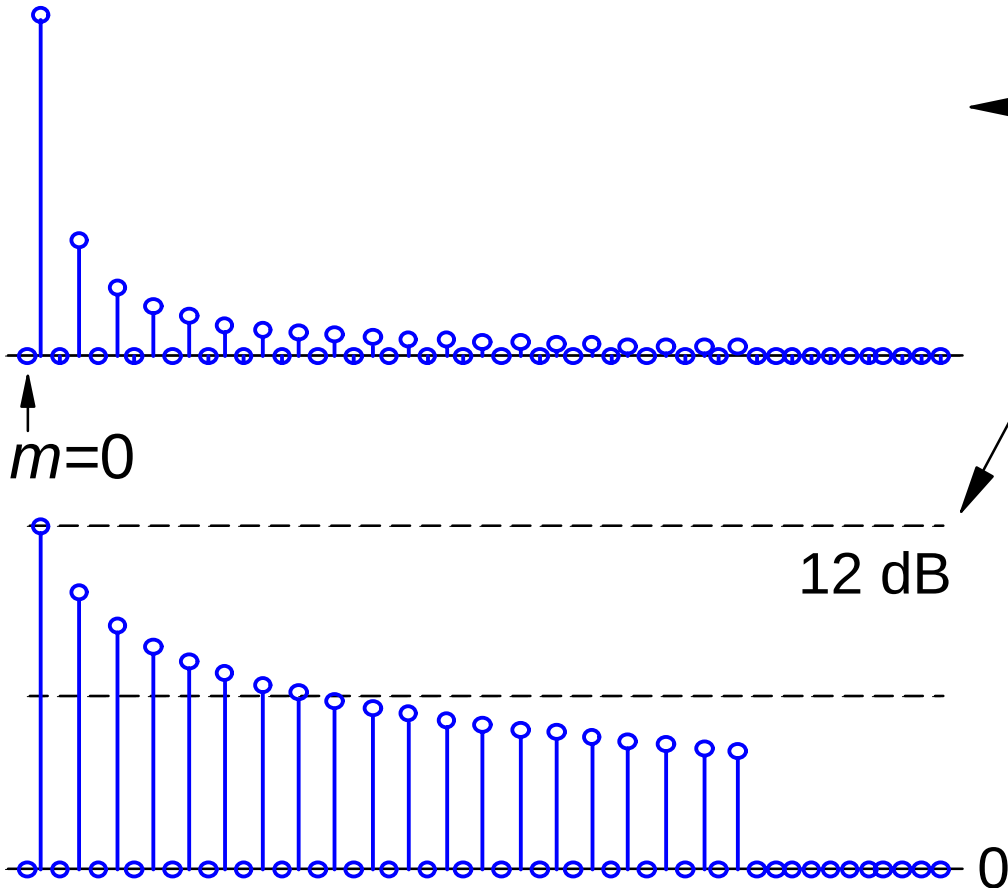
Zeitfunktionen und Spektren - Dreieck



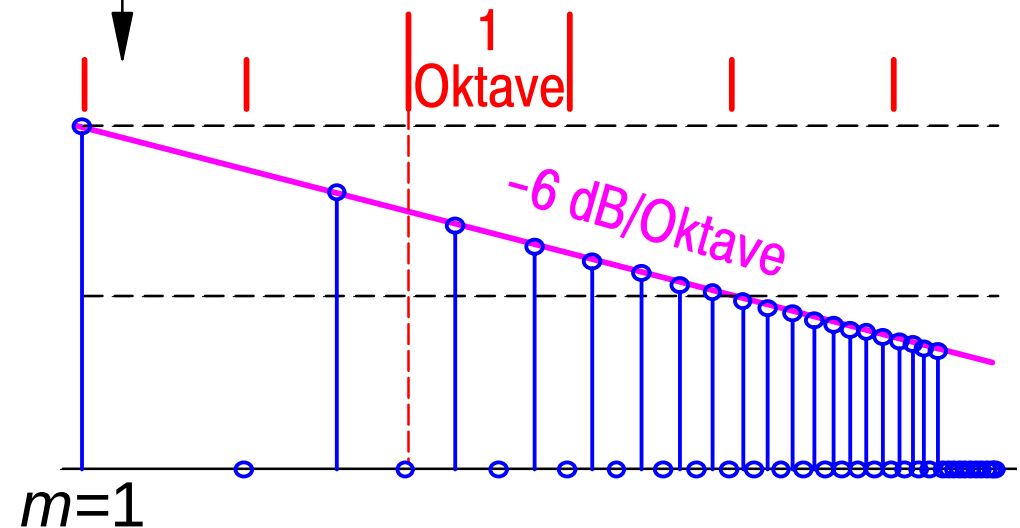
Zeitfunktionen und Spektren - Impuls



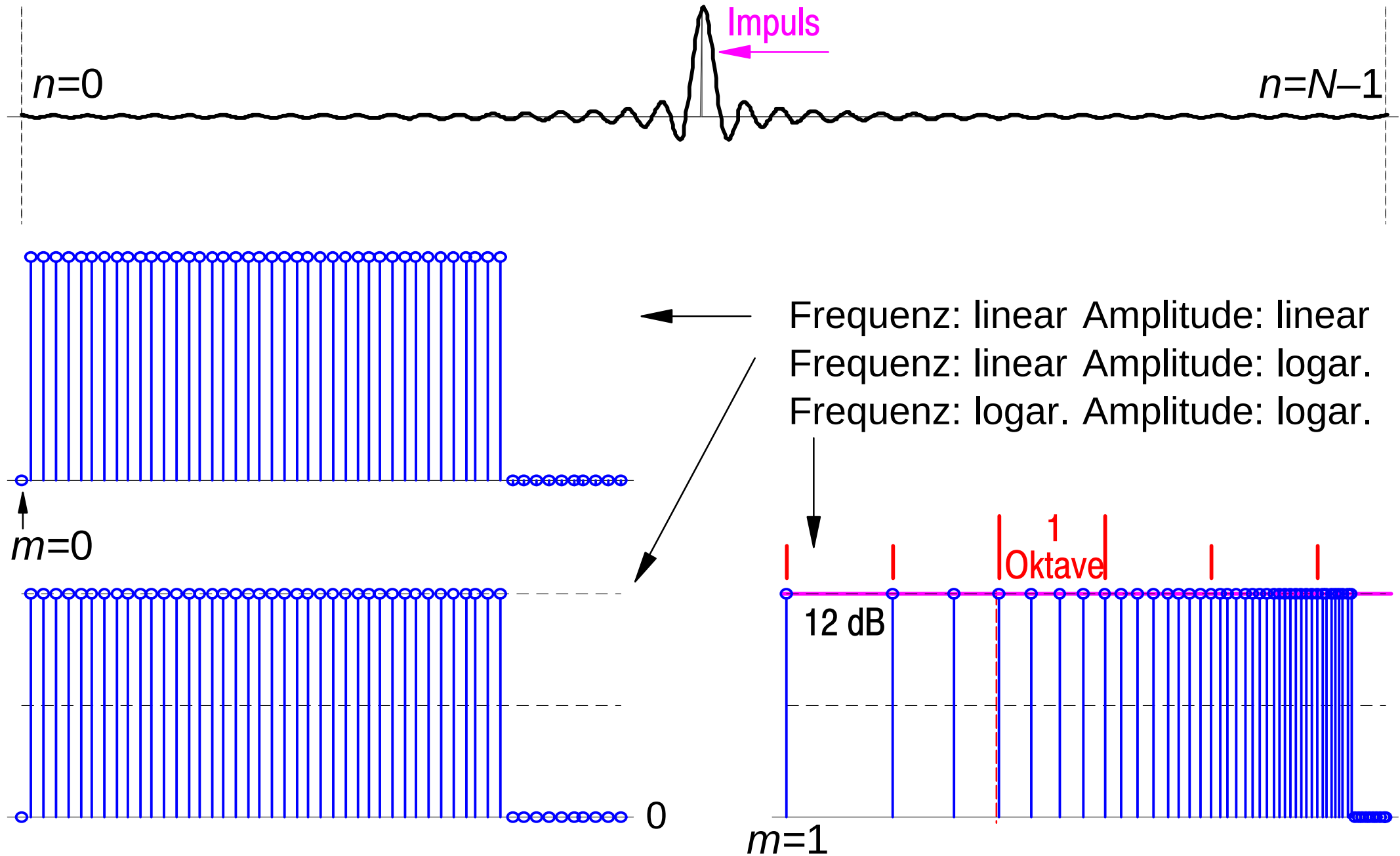
Zeitfunktionen und Spektren - Rechteck (erste 40 Teilschwingungen)



Frequenz: linear Amplitude: linear
Frequenz: linear Amplitude: logar.
Frequenz: logar. Amplitude: logar.



Zeitfunktionen und Spektren - Impuls (erste 40 Teilschwingungen)



Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

3.1 Harmonische Analyse (diskreter) periodischer Signale

3.2 Nichtharmonische Spektralanalyse

3.2.1 Spektraltransformation gedämpfter Schwingungen

3.2.2 Nichtharmonische Spektralanalyse

3.2.3 Gewichtung des Eingangssignals

3.3 Kurzzeitspektren von Sprachsignalen

3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

Nichtharmonische Spektralanalyse

Ausgangspunkt jeder Spektralanalyse mit der diskreten Fouriertransformation (DFT) ist die harmonische Analyse.

Für andere Signale - insbesondere Kurzzeit-Messintervalle - erzwingt die DFT die **periodische Fortsetzung** des Signals außerhalb des Messintervalls.

Dies kann zu Artefakten führen, wenn an den Rändern des Intervalls **Diskontinuitäten** (Sprünge, Ecken, ...) auftreten.

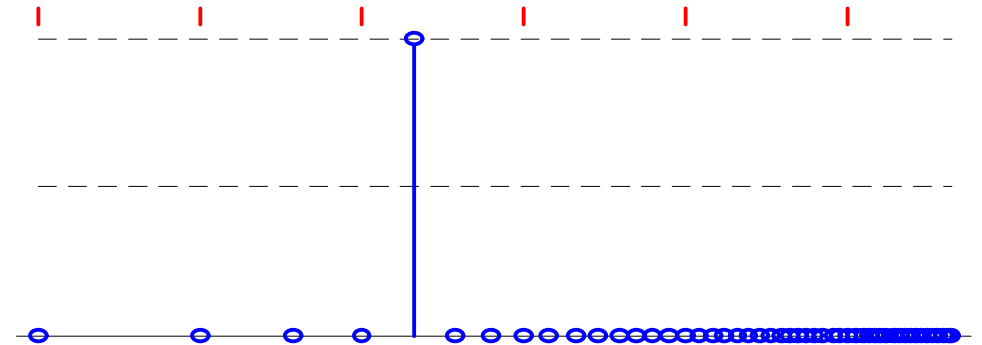
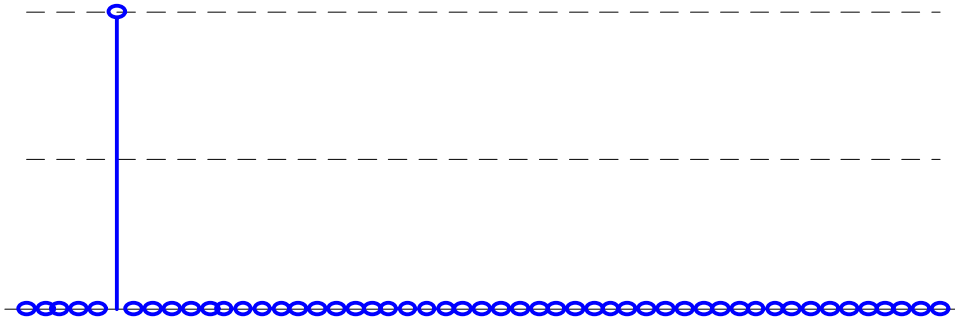
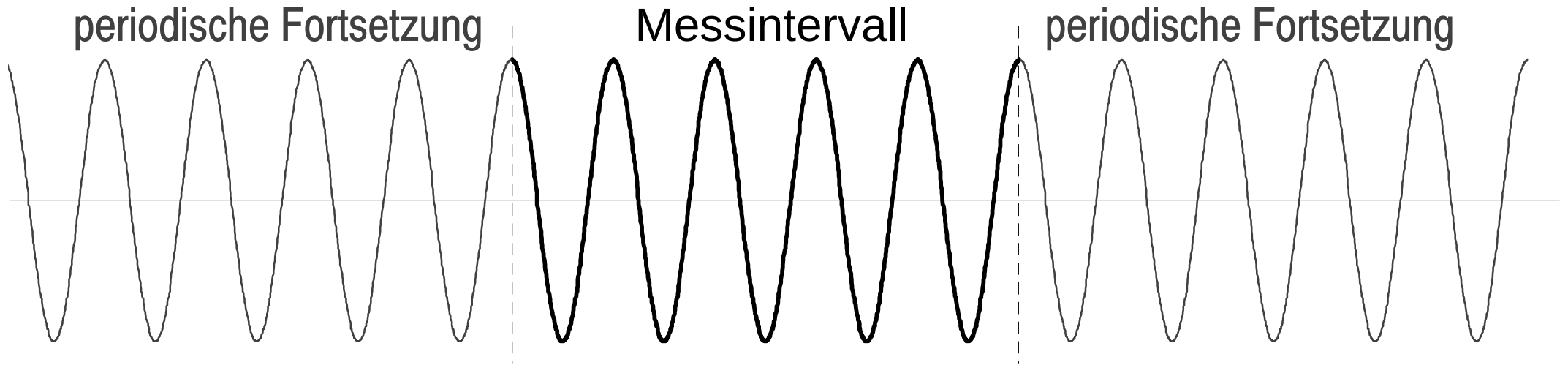
Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

- 3.1 Harmonische Analyse (diskreter) periodischer Signale
- 3.2 Nichtharmonische Spektralanalyse
 - 3.2.1 Spektraltransformation gedämpfter Schwingungen
 - 3.2.2 Keine ganze Zahl von Perioden im Intervall
 - 3.2.3 Gewichtung des Eingangssignals
- 3.3 Kurzzeitspektren von Sprachsignalen
- 3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

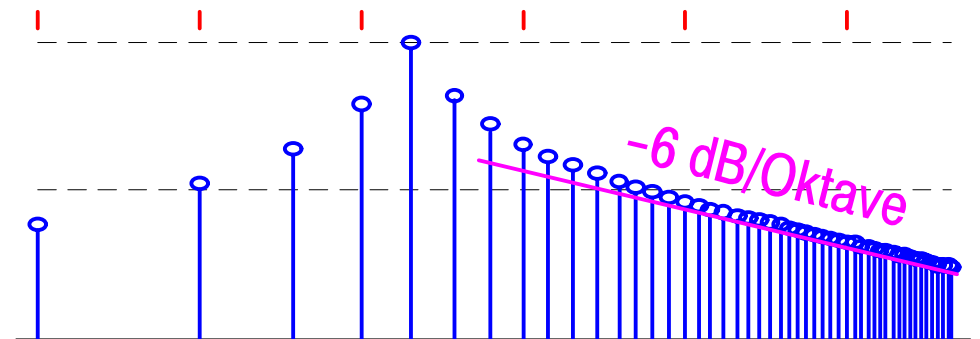
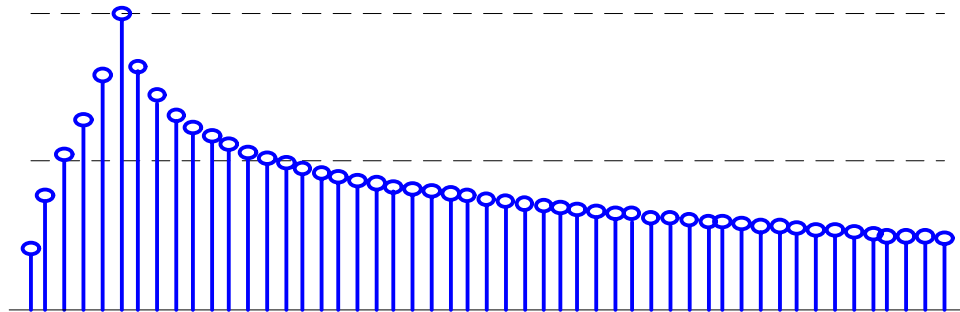
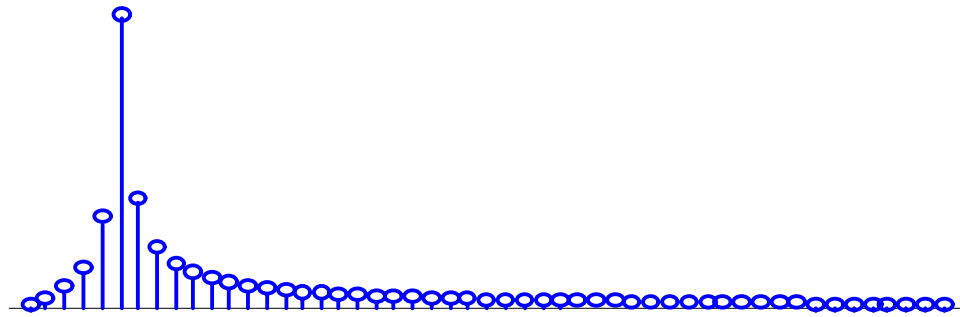
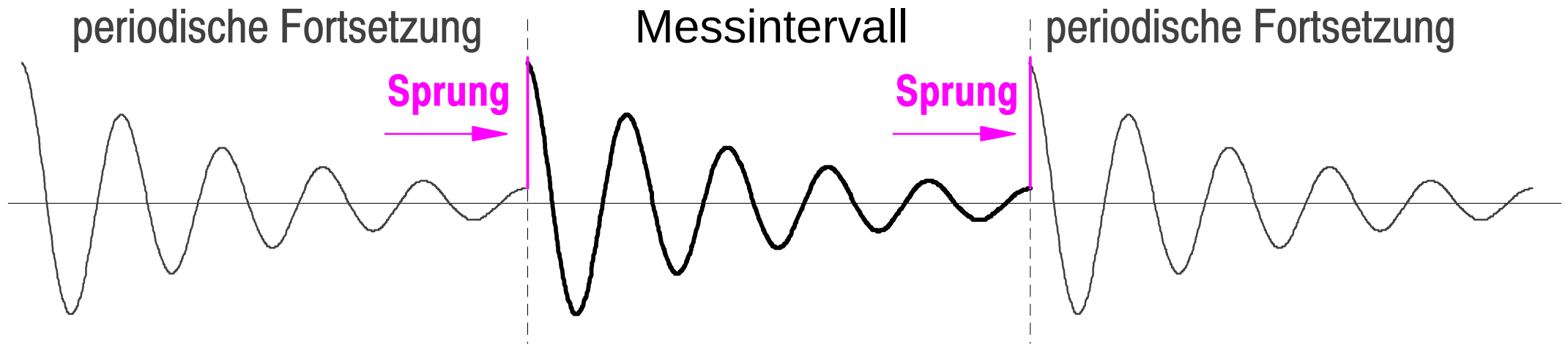
Ungedämpfte Kosinusschwingung

5 Perioden im Intervall



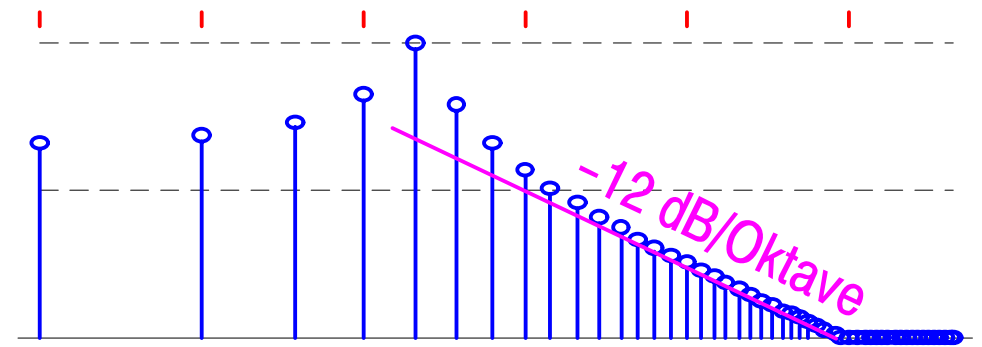
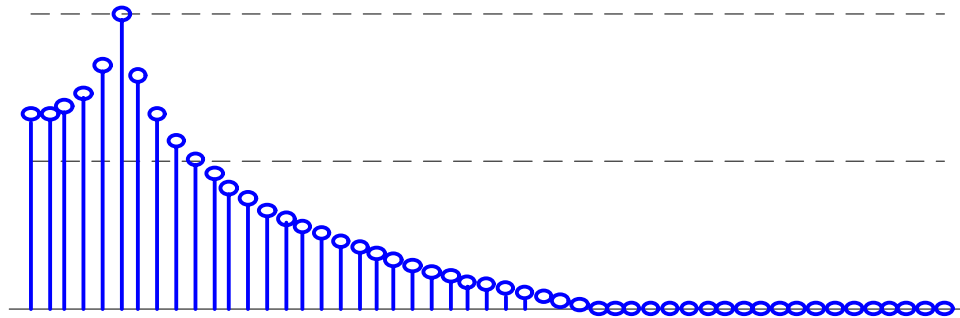
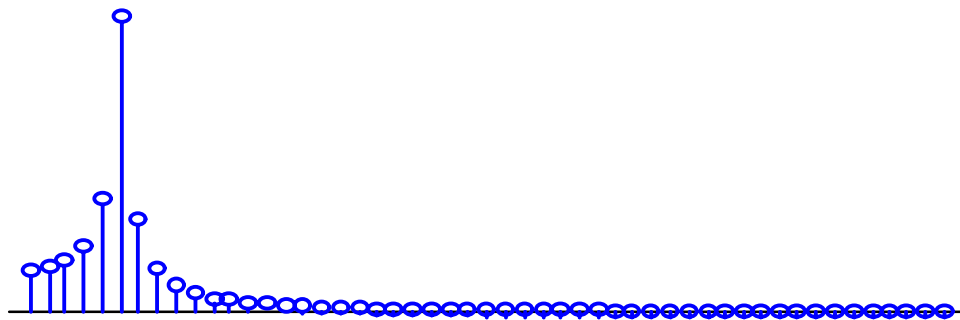
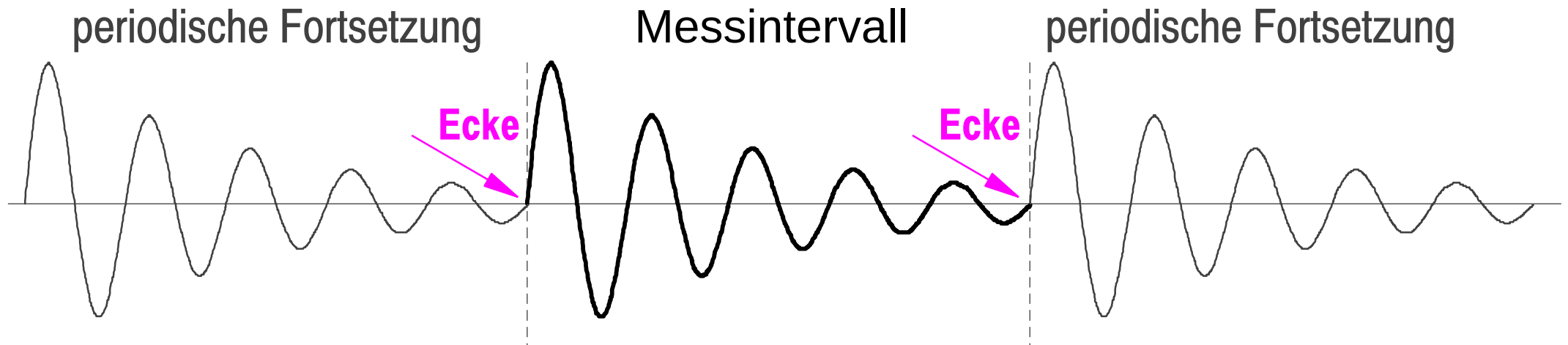
Gedämpfte Kosinusschwingung

5 Perioden im Intervall



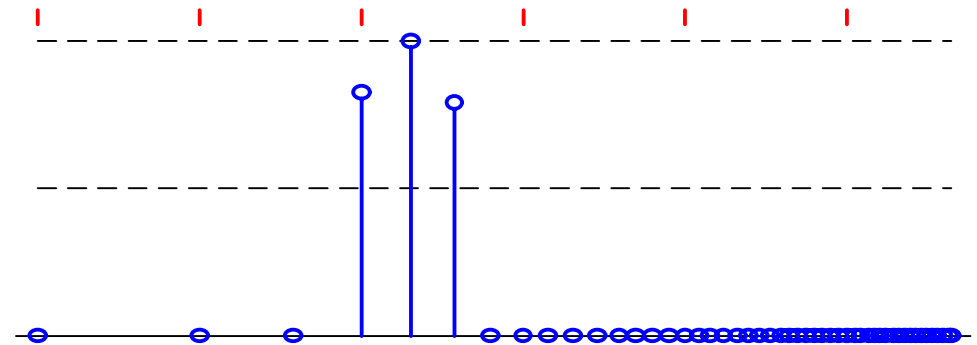
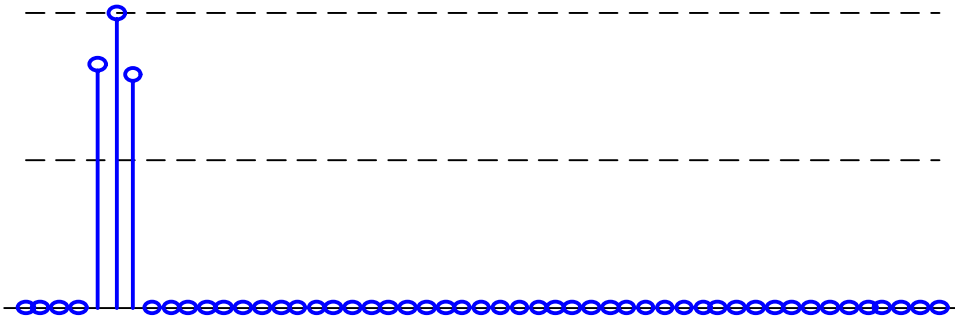
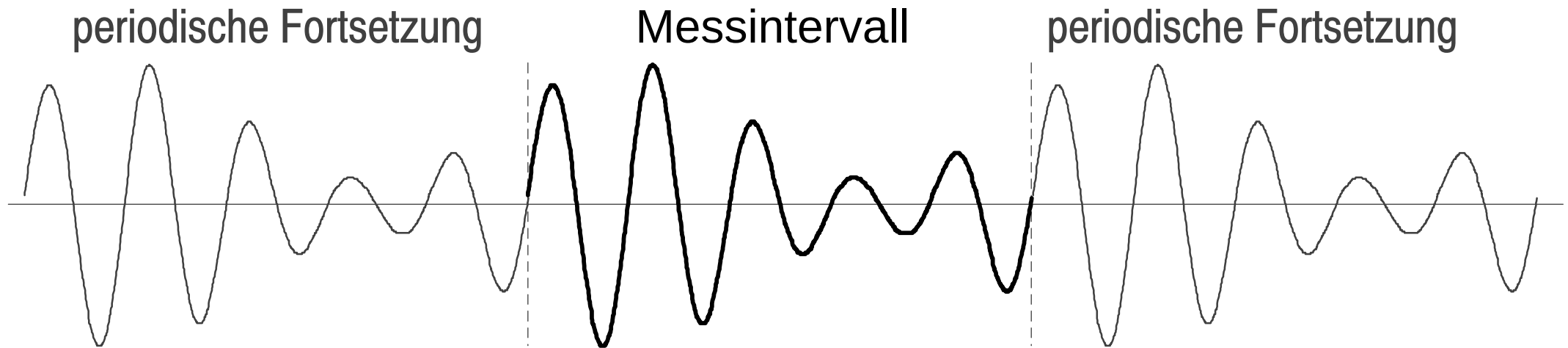
Gedämpfte Sinusschwingung

5 Perioden im Intervall



Gedämpfte Sinusschwingung

Approximation durch 3 Teilschwingungen

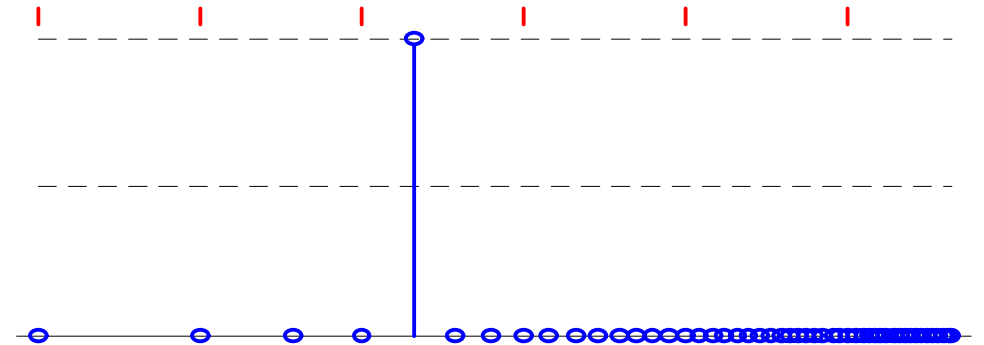
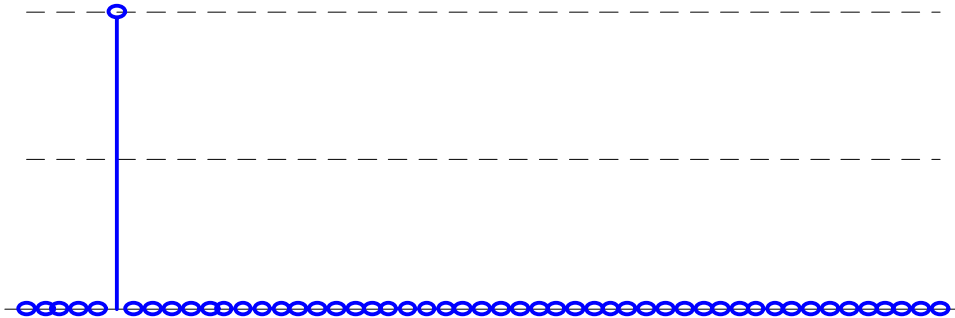
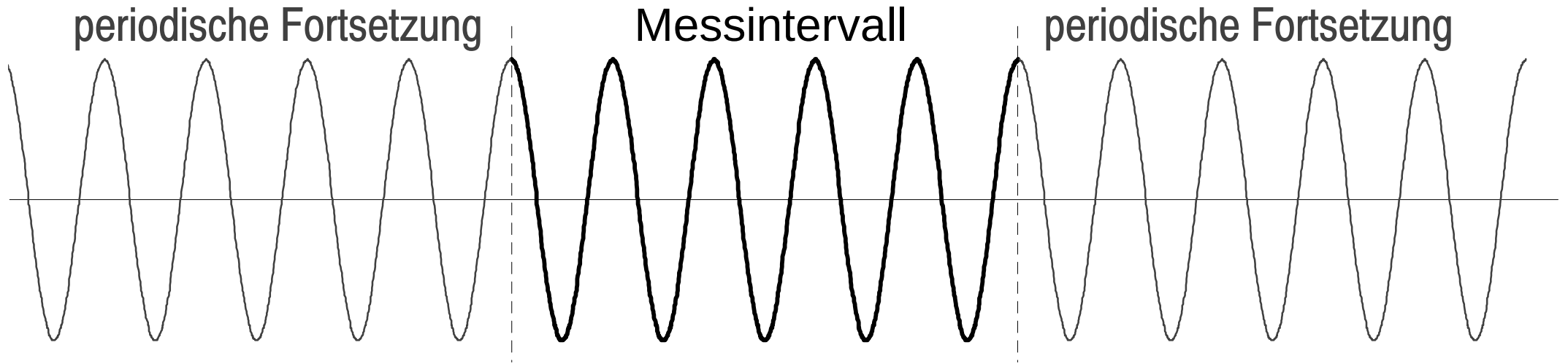


Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

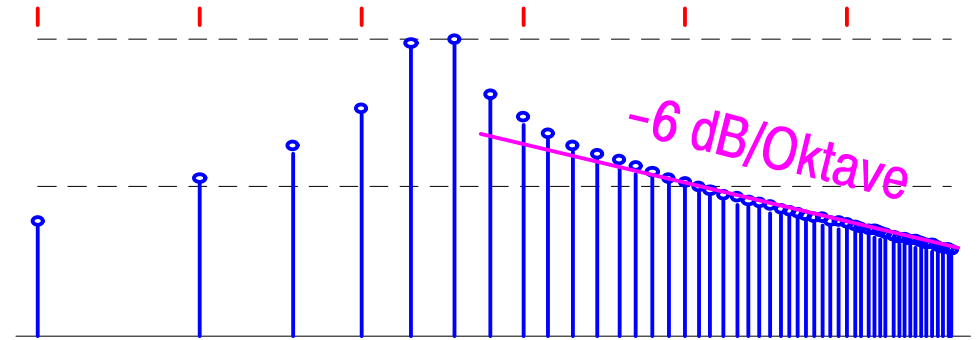
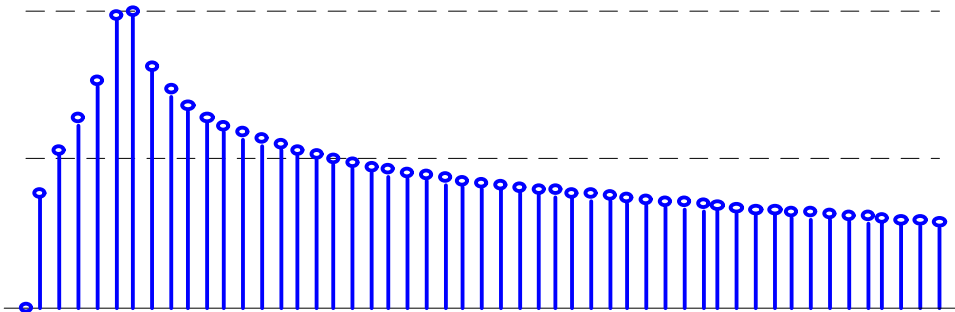
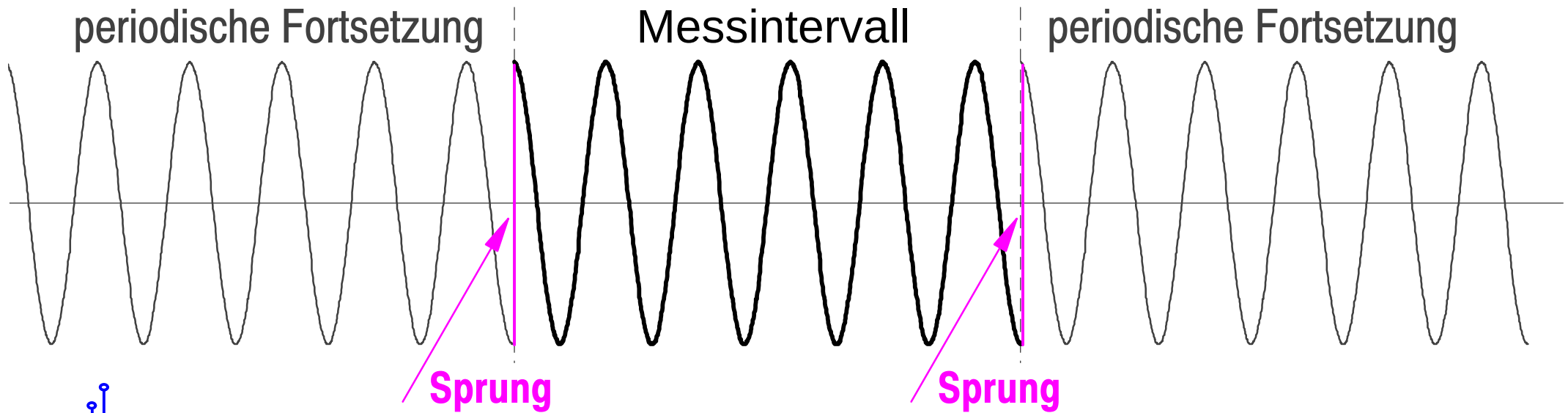
- 3.1 Harmonische Analyse (diskreter) periodischer Signale
- 3.2 Nichtharmonische Spektralanalyse
 - 3.2.1 Spektraltransformation gedämpfter Schwingungen
 - 3.2.2 Keine ganze Zahl von Perioden im Intervall
 - 3.2.3 Gewichtung des Eingangssignals
- 3.3 Kurzzeitspektren von Sprachsignalen
- 3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

Ungedämpfte Kosinusschwingung (vgl. Abschnitt 3.2.1)



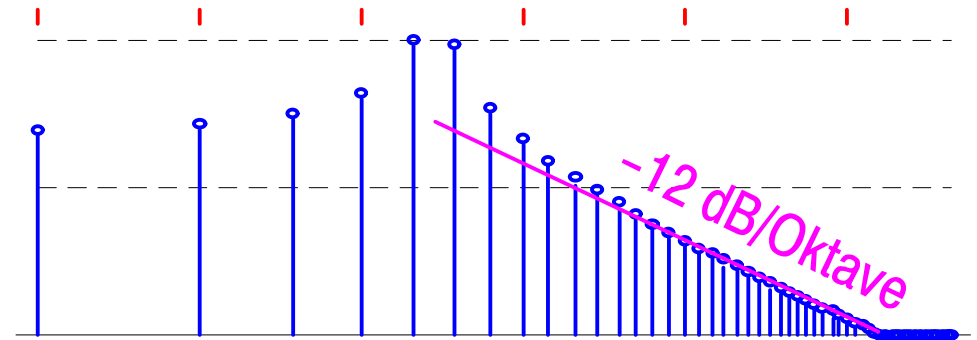
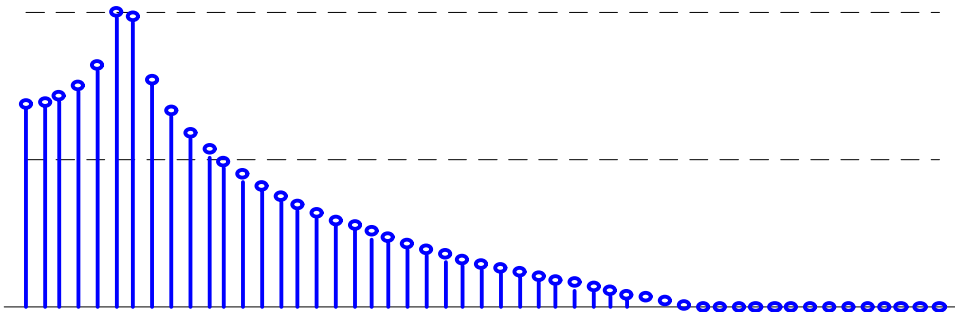
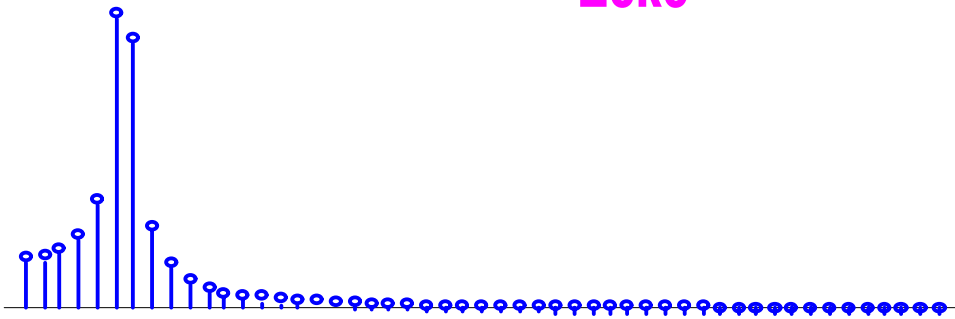
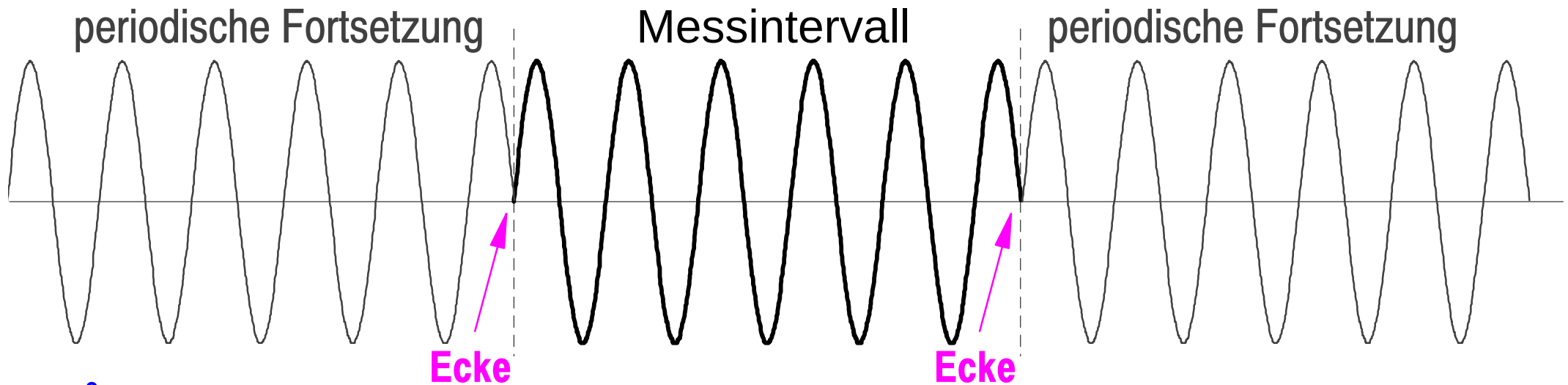
Ungedämpfte Kosinusschwingung

5,5 Perioden im Messintervall



Ungedämpfte Sinusschwingung

5,5 Perioden im Messintervall



Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

3.1 Harmonische Analyse (diskreter) periodischer Signale

3.2 Nichtharmonische Spektralanalyse

3.2.1 Spektraltransformation gedämpfter Schwingungen

3.2.2 Keine ganze Zahl von Perioden im Intervall

3.2.3 Gewichtung des Eingangssignals

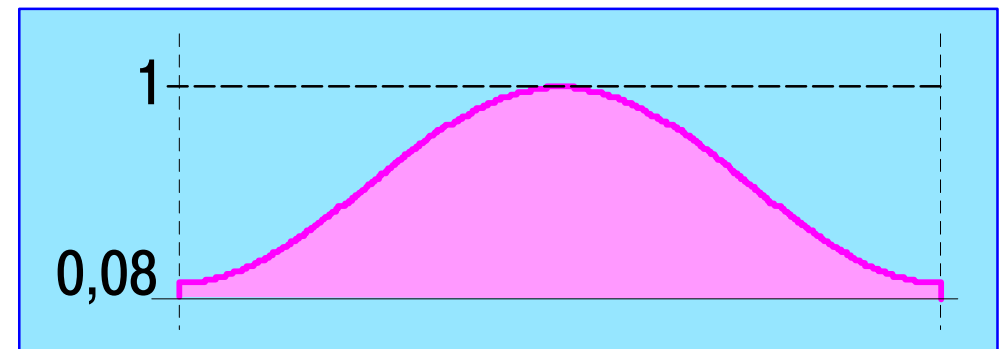
3.3 Kurzzeitspektren von Sprachsignalen

3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

Um die Auswirkung der Unstetigkeit eines beliebigen Eingangssignals an den Intervallgrenzen zu mildern, greift man in der Regel nach dem Mittel der **Gewichtung des Signals** vor Anwendung der DFT.

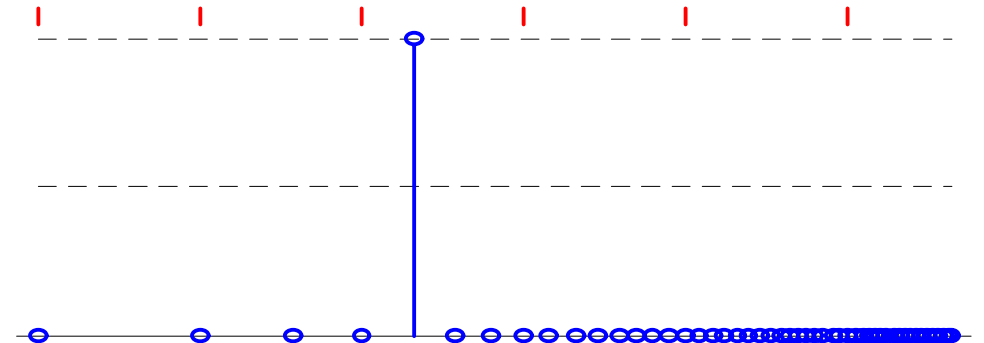
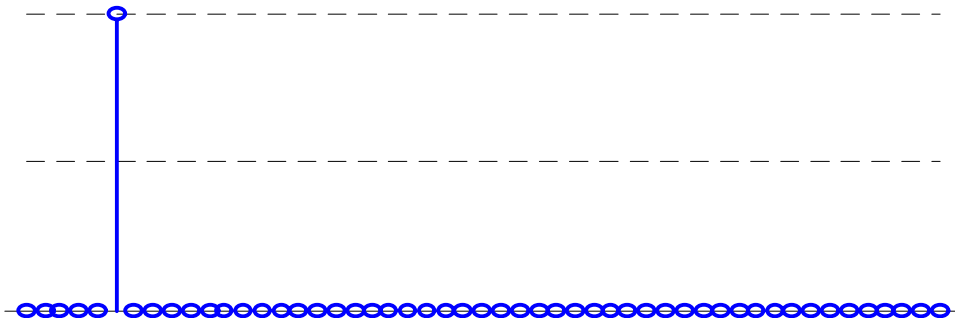
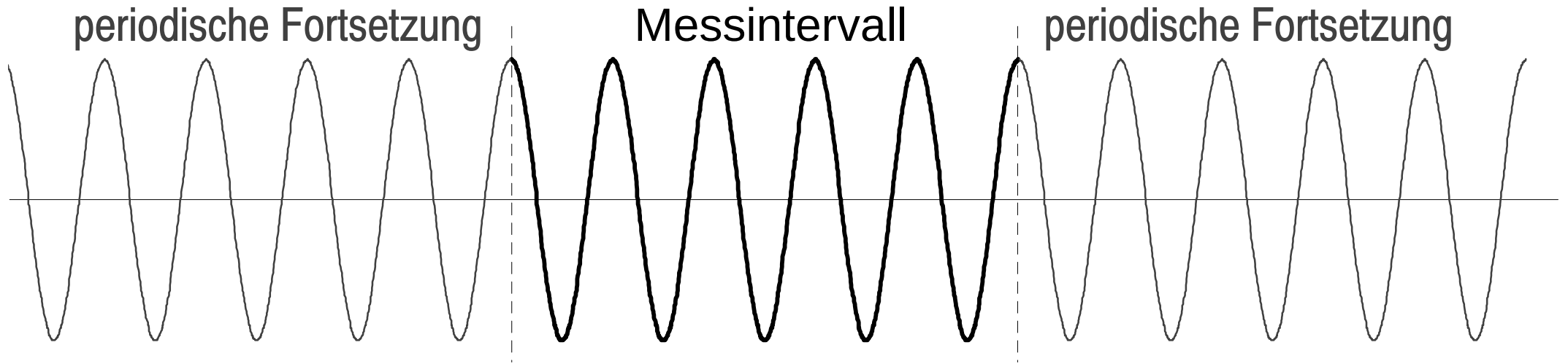
Hier wird das Signal über die gesamte Länge des Analyseintervalls mit einer **Gewichtungsfunktion** oder **Fensterfunktion** multipliziert, die in Intervallmitte den Wert 1 annimmt und nach den Intervallgrenzen hin sanft gegen Null oder einen kleinen, von Null verschiedenen Wert abfällt.

In der Sprachsignalverarbeitung sehr verbreitet ist das **Hamming-Fenster**:

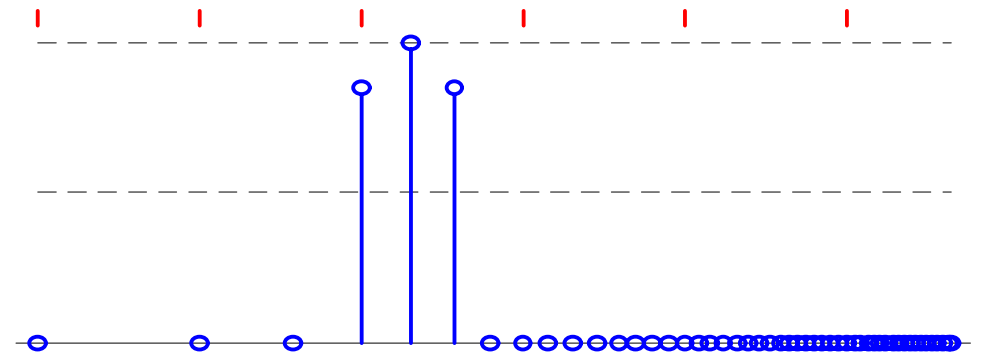
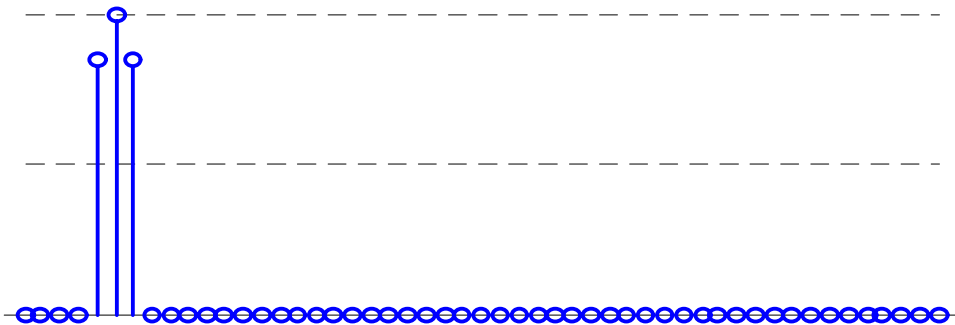
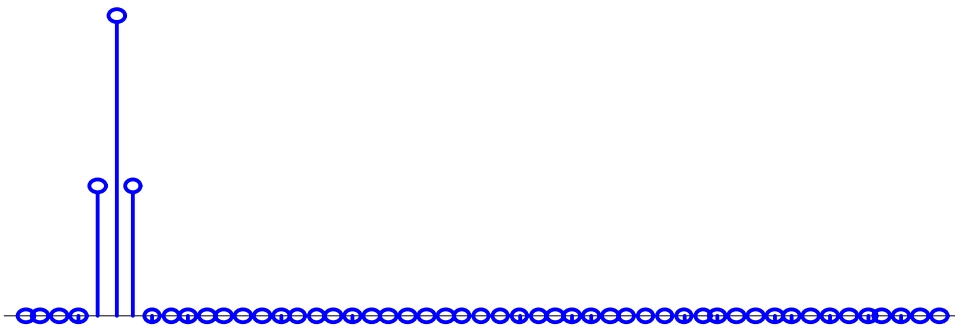
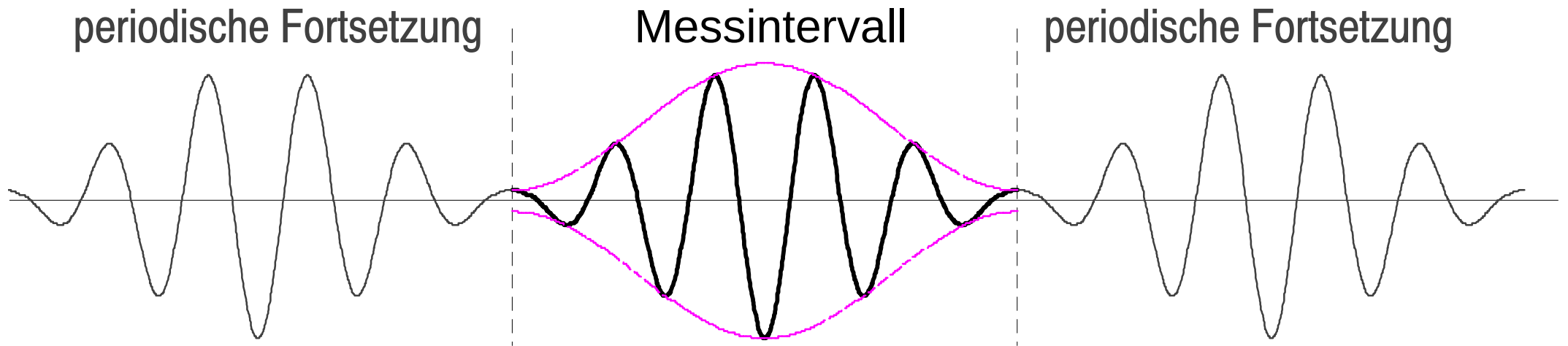


$$w_H(k) = 0,08 + 0,92 \cos^2 \left[\frac{\pi(k-K)}{N} \right] = 0,54 + 0,46 \cos \frac{2\pi(k-K)}{N} ; \quad K = \frac{N}{2} \quad (3.27)$$

Ungedämpfte Kosinusschwingung (vgl. Abschnitt 3.2.1)

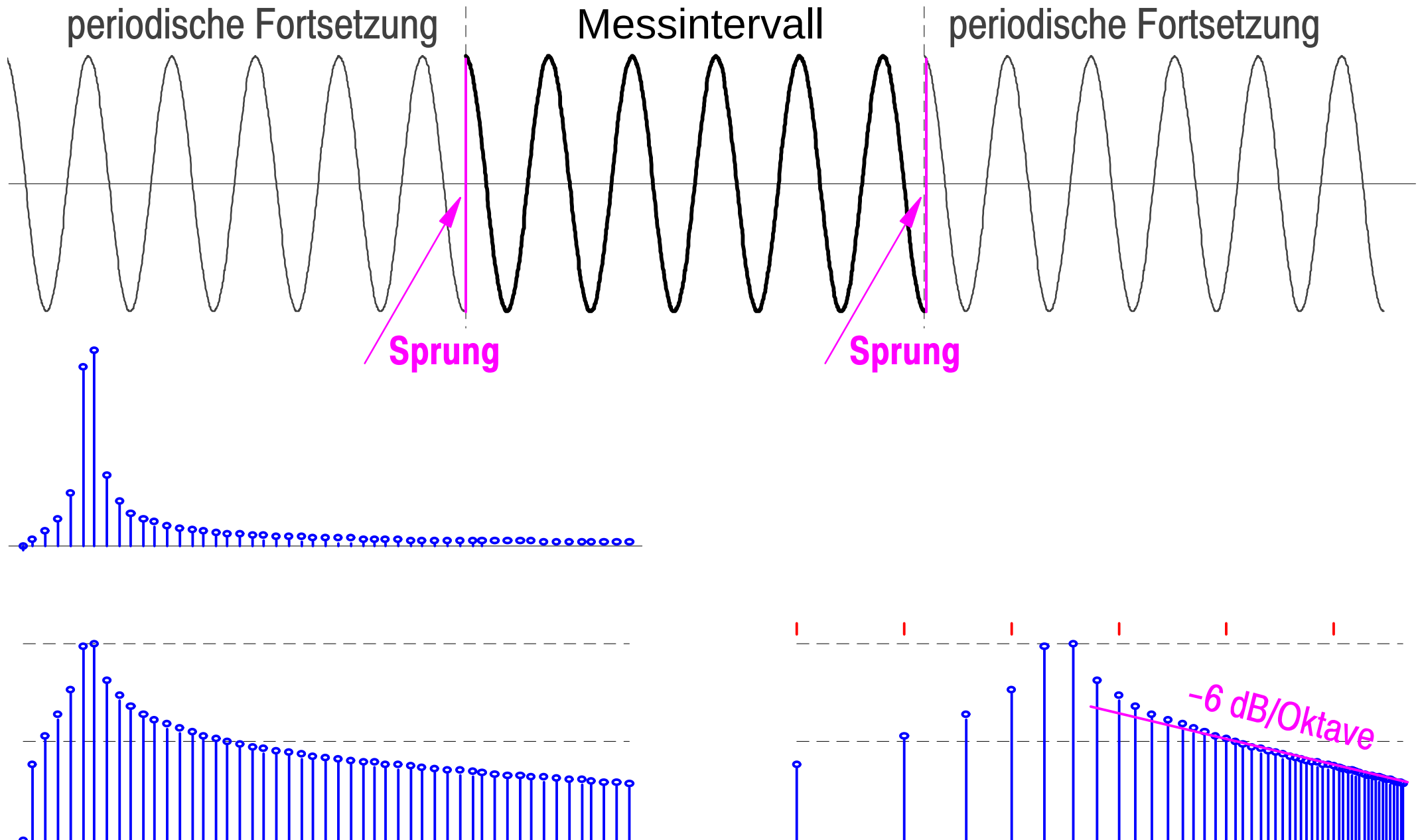


Ungedämpfte Kosinusschwingung gewichtet mit Hammingfenster



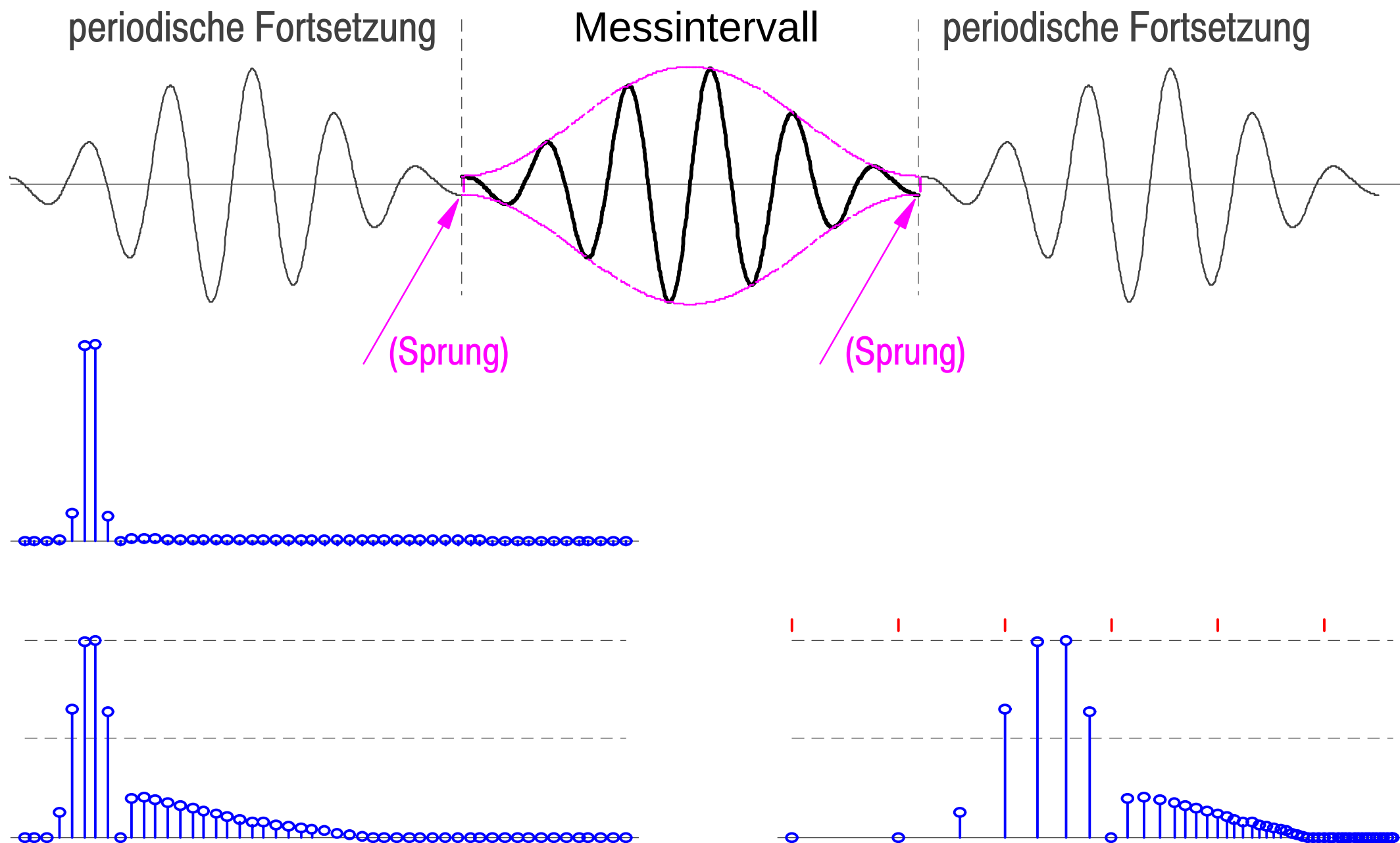
Ungedämpfte Kosinusschwingung

5,5 Perioden (vgl. Abschnitt 3.2.2)



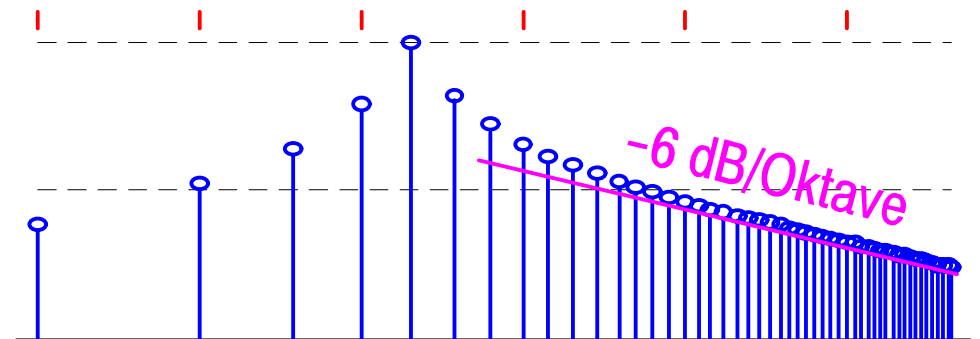
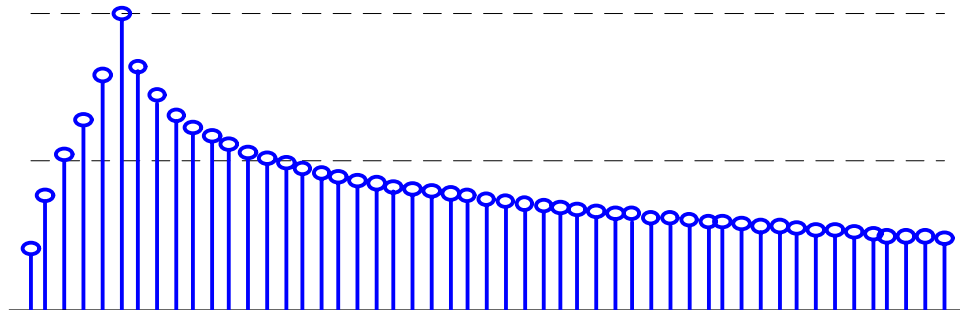
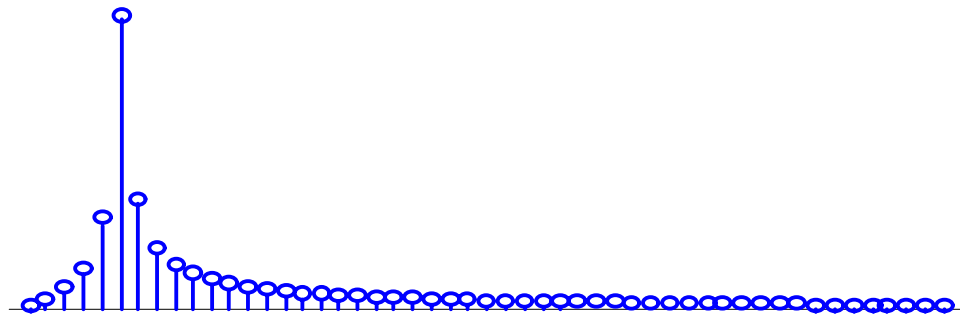
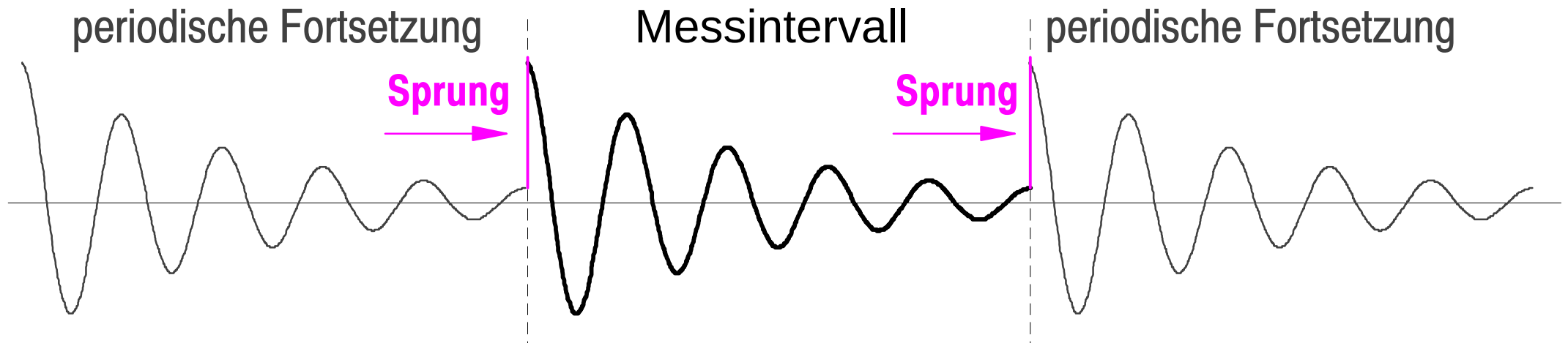
Ungedämpfte Kosinusschwingung

5,5 Perioden; gewichtet mit Hammingfenster



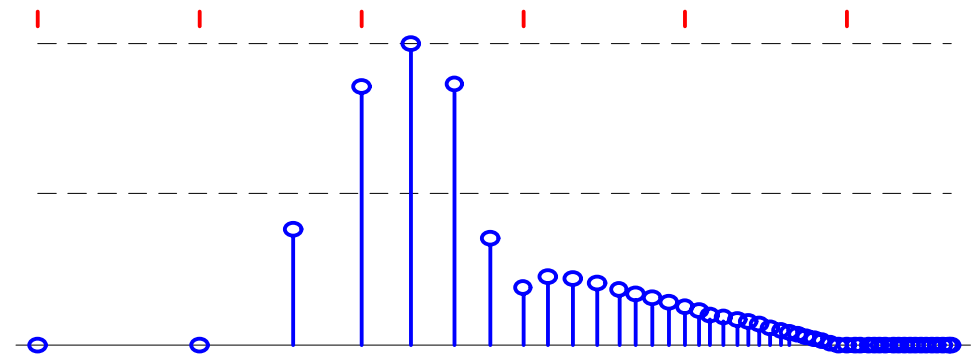
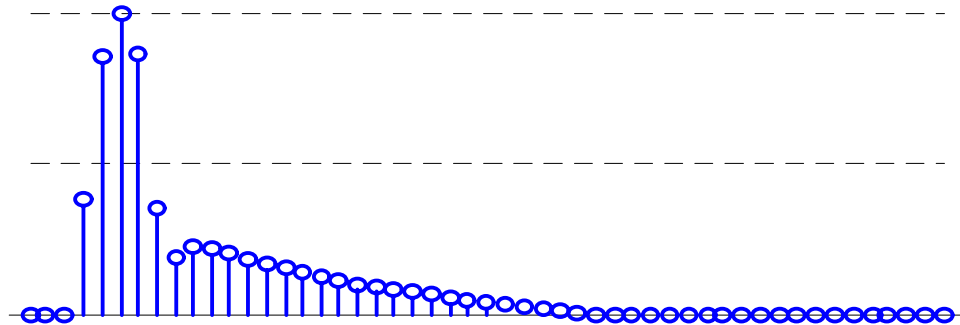
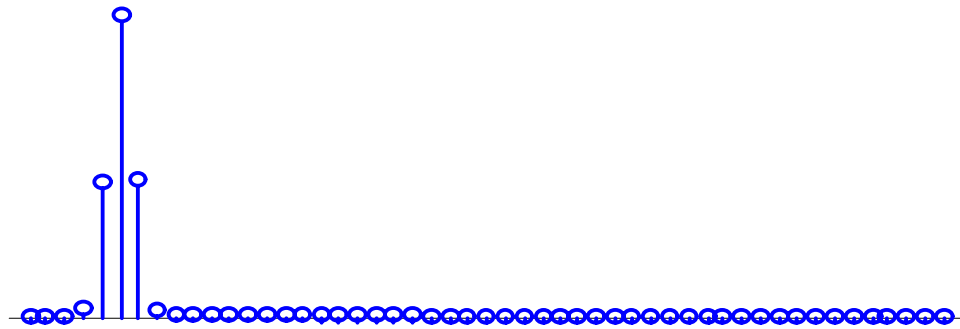
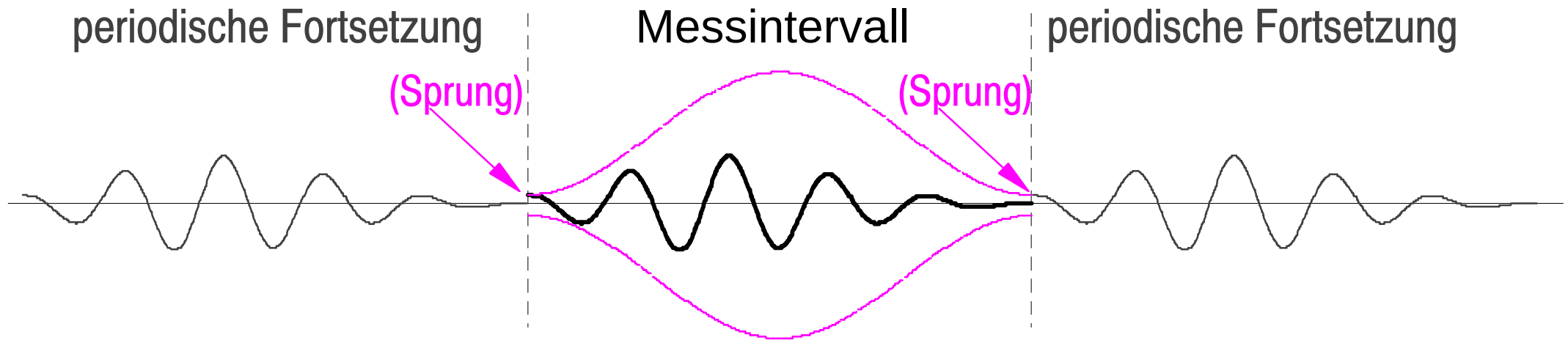
Gedämpfte Kosinusschwingung

5 Perioden (vgl. Abschnitt 3.2.1)



Gedämpfte Kosinusschwingung

5 Perioden; gewichtet mit Hammingfenster



Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

- 3.1 Harmonische Analyse (diskreter) periodischer Signale
- 3.2 Nichtharmonische Spektralanalyse
- 3.3 Kurzzeitspektren von Sprachsignalen**
- 3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung

Ansatz zur Kurzzeitanalyse

Das Kurzzeitsignal s_K ist zunächst nur *innerhalb* des Fensters definiert:

$$s_K(n, q) = \begin{cases} s(n) w(q-n) ; & n = q-K+1 \text{ (1) } q \\ \text{sonst (zunächst) undefiniert, d.h. beliebig.} & \end{cases} \quad (2.9)$$

Die Spektralanalyse erzwingt den **periodischen Ansatz**:

Das Signal s_K wird als **periodischer Vorgang** angesehen. In diesem Fall wird es außerhalb des Fensters periodisch fortgesetzt; die Periode ist gleich dem **Transformationsintervall N** .

Ist N größer als die Fensterlänge K , so muss mit Nullwerten aufgefüllt werden. Dann ist das Spektrum **überabgetastet**:

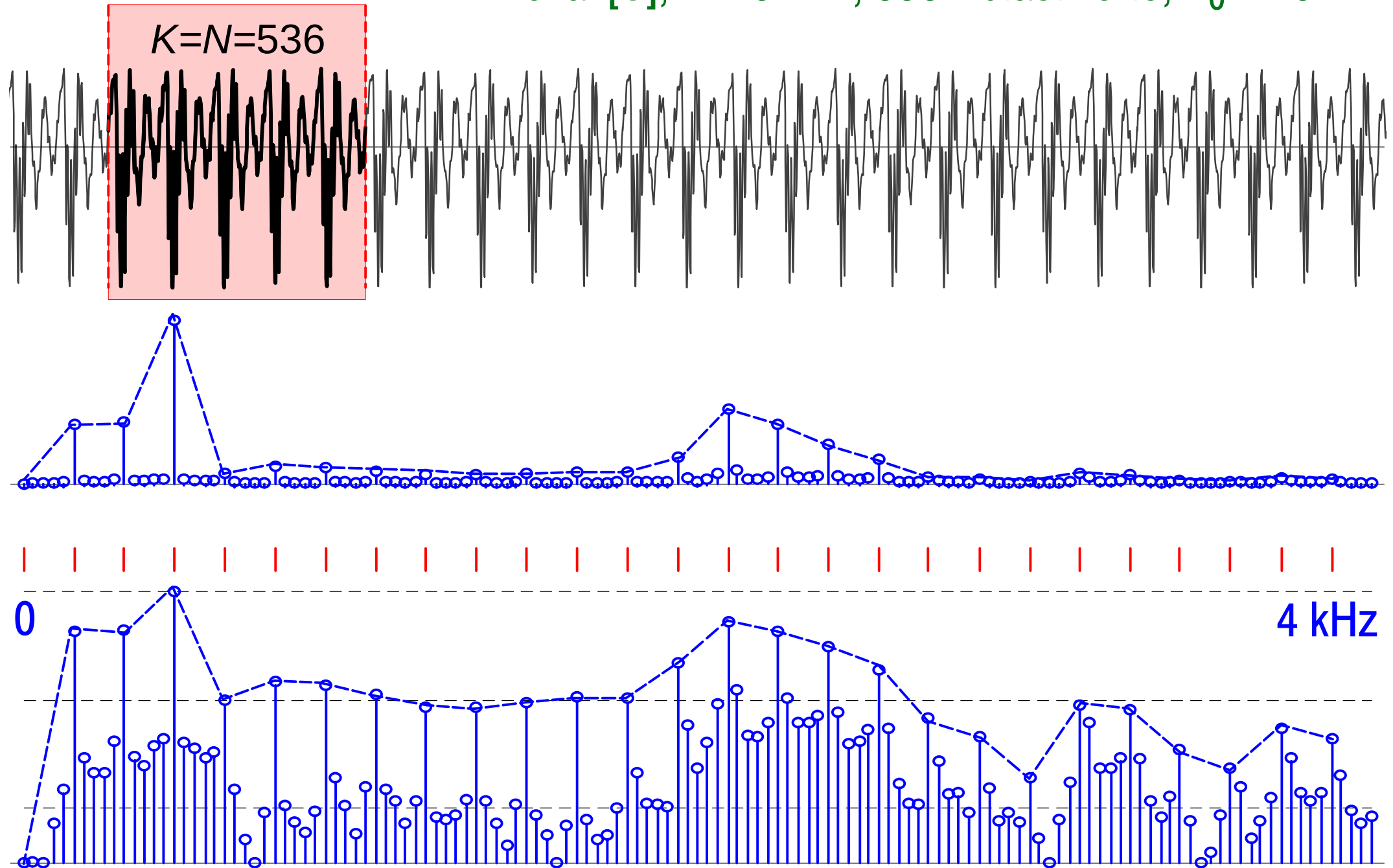
$$x_T(k) := \begin{cases} s_K(k, q) & k = 0 \text{ (1) } K-1 ; \text{ entspricht } n = q-K+1 \text{ (1) } q \\ 0 & k = K \text{ (1) } N-1 . \end{cases} \quad (3.28)$$

Spektrale Auflösung aus Abtastfrequenz F des Signals und N :

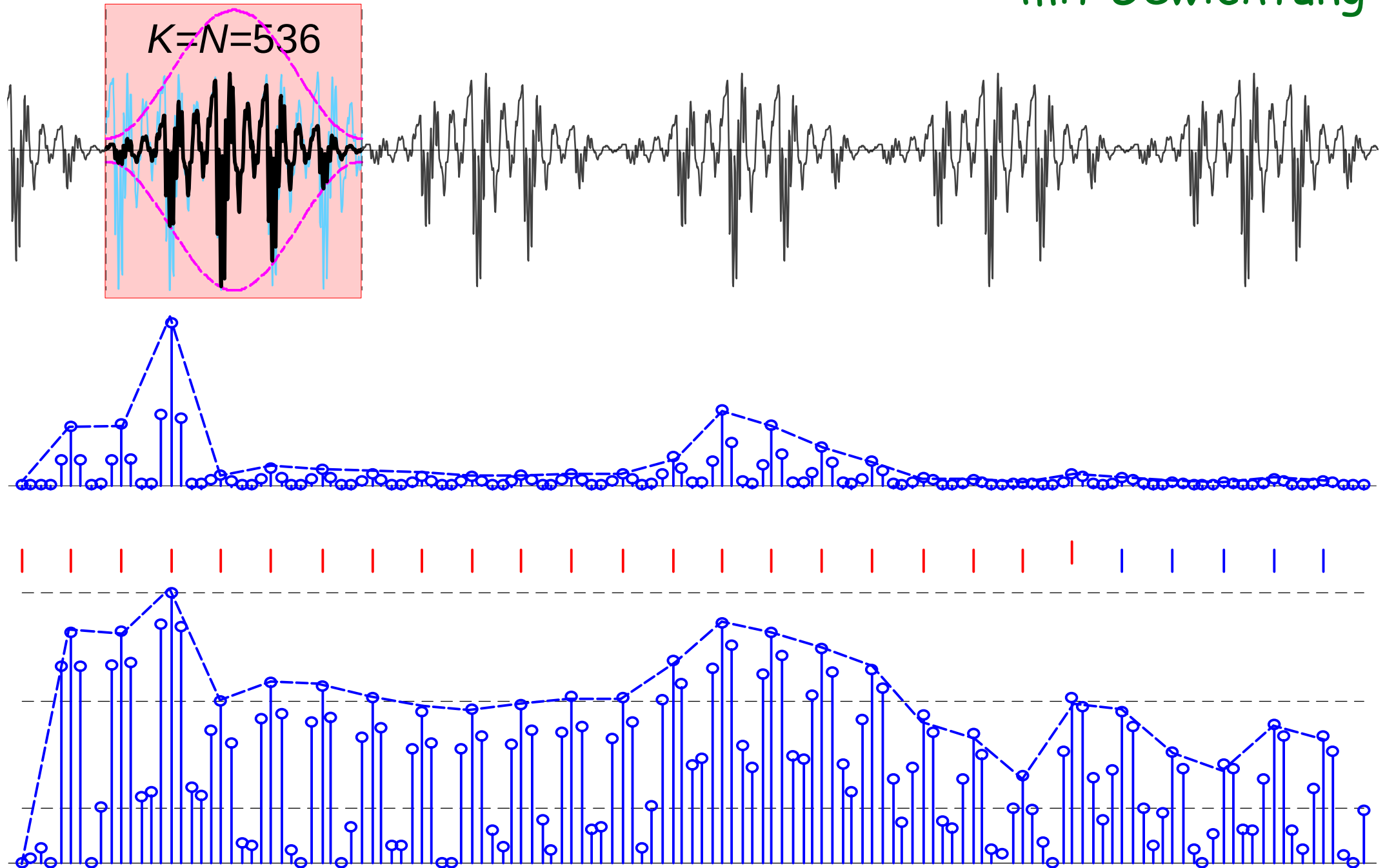
$$\Delta f = F / N \quad [\text{in Hz bzw. kHz}] \quad (3.29)$$

Harmonische Analyse (5 Perioden)

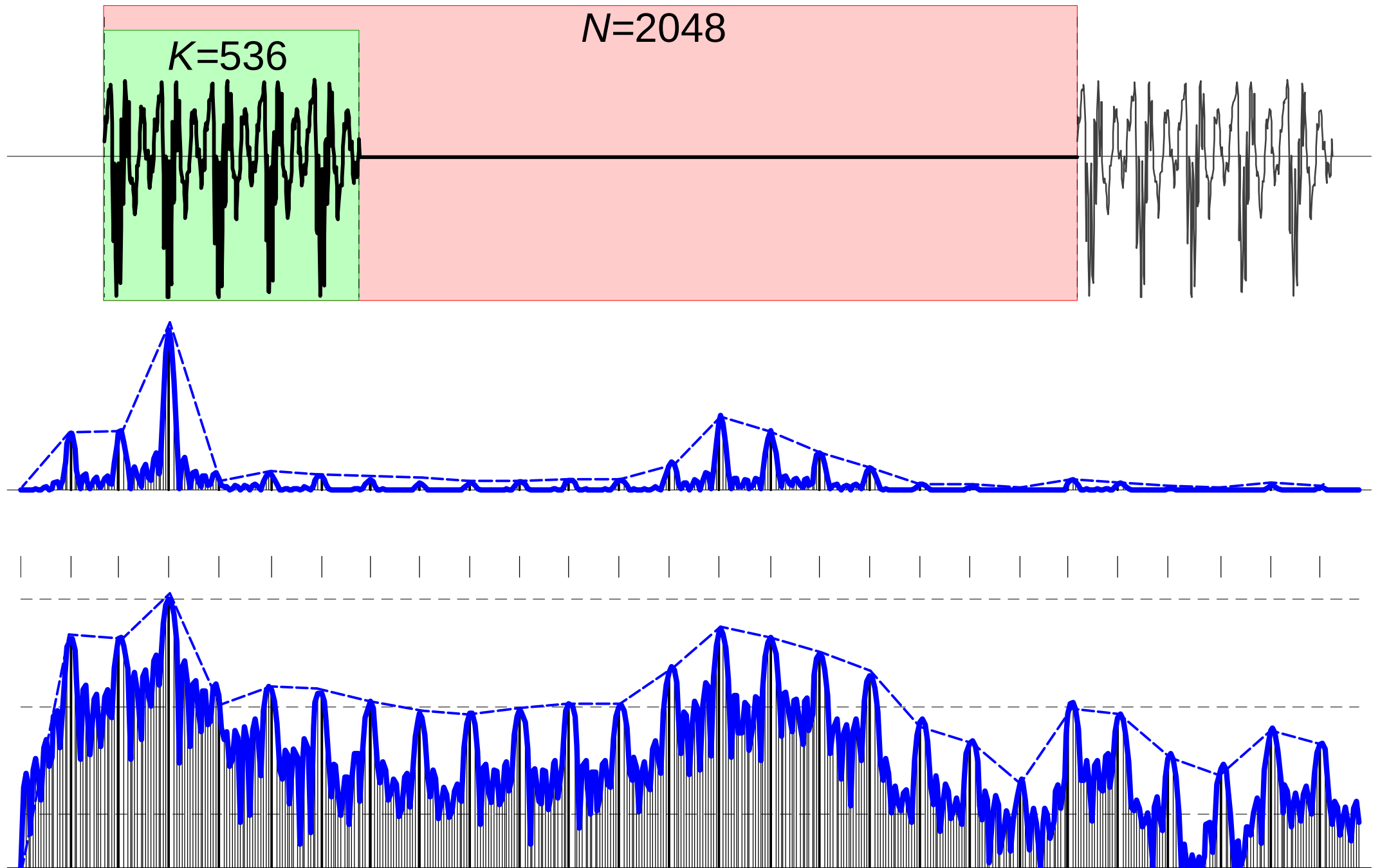
Vokal [ε], $F=16$ kHz, 536 Abtastwerte, $F_0=149$ Hz



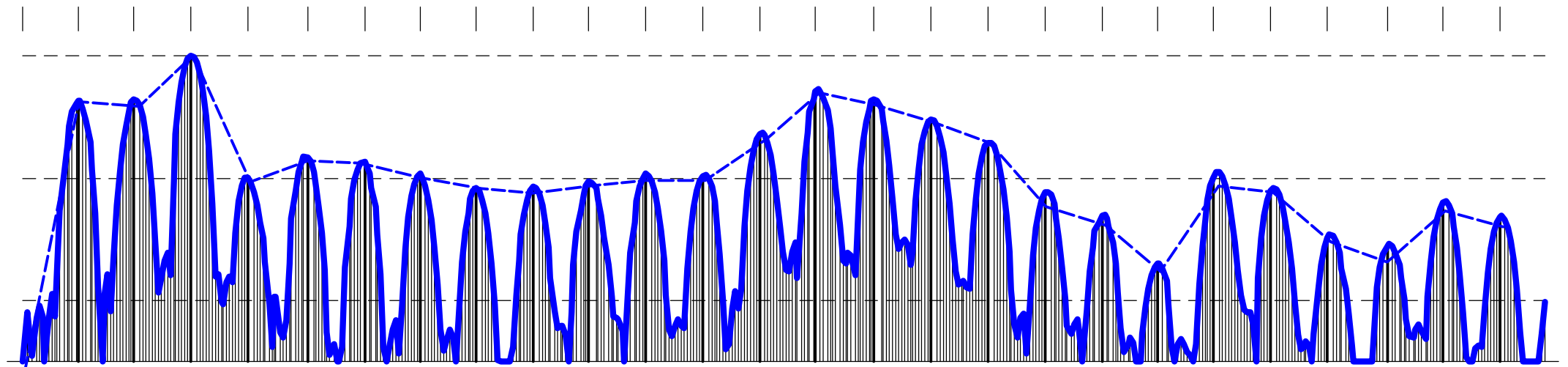
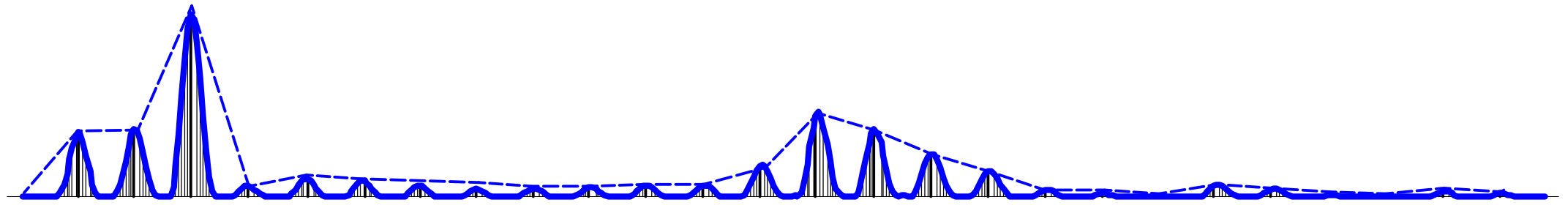
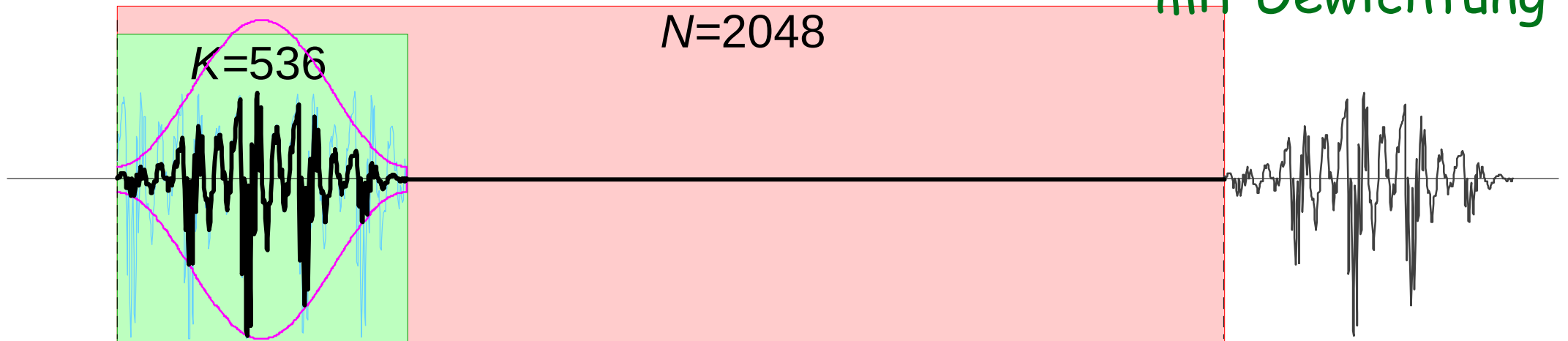
Harmonische Analyse (5 Perioden) mit Gewichtung



Nichtharmonische Analyse (5 Perioden)

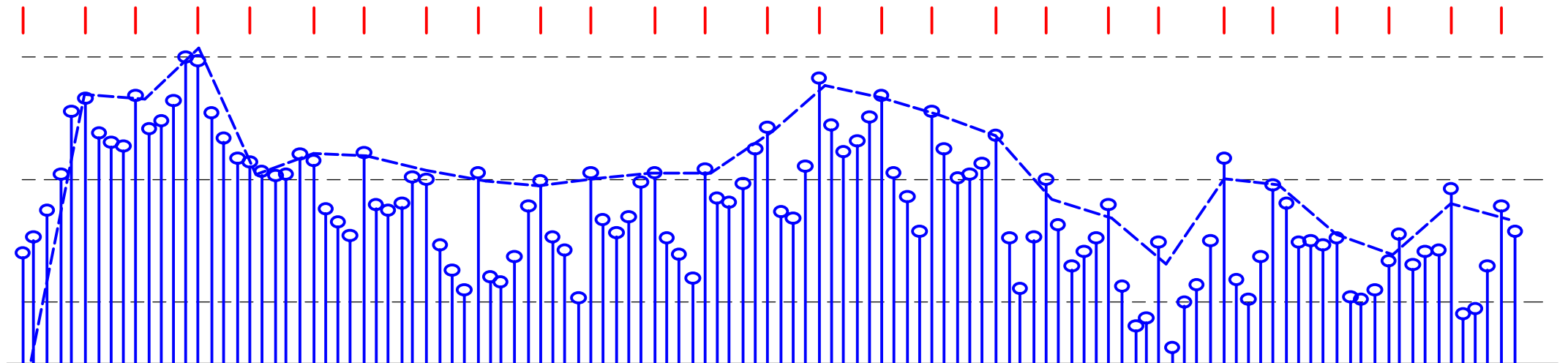
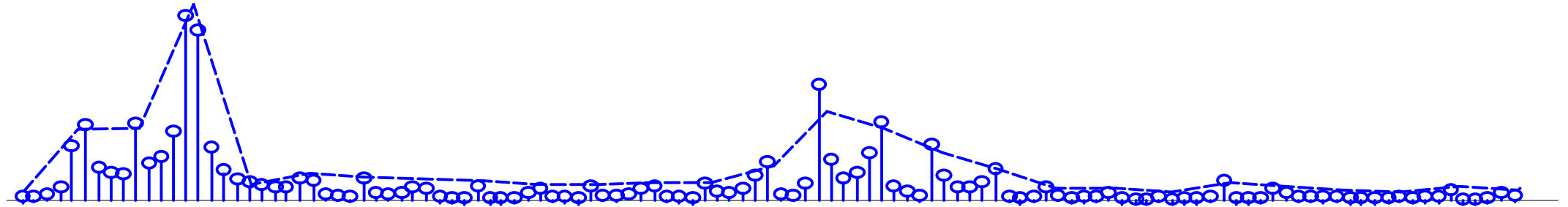
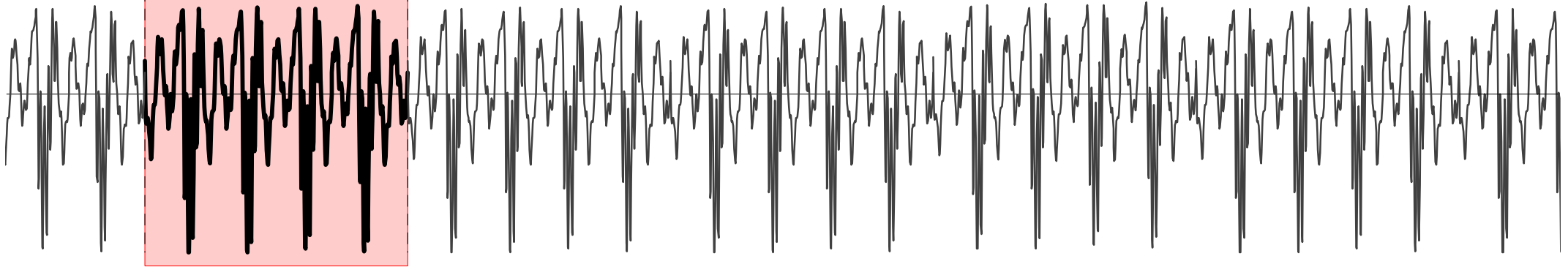


Nichtharmonische Analyse (5 Perioden) mit Gewichtung

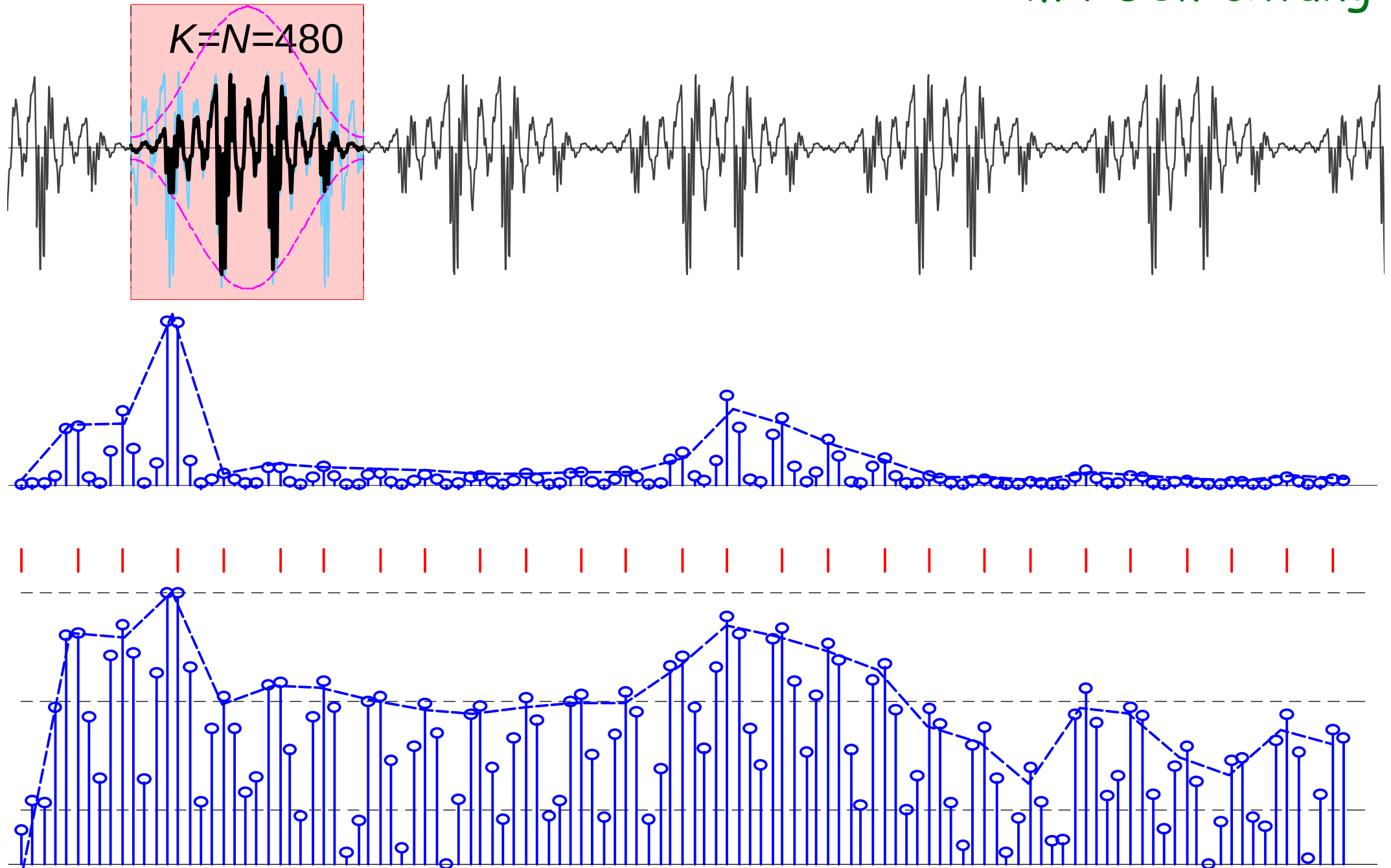


Harmonische Analyse (4,5 Perioden)

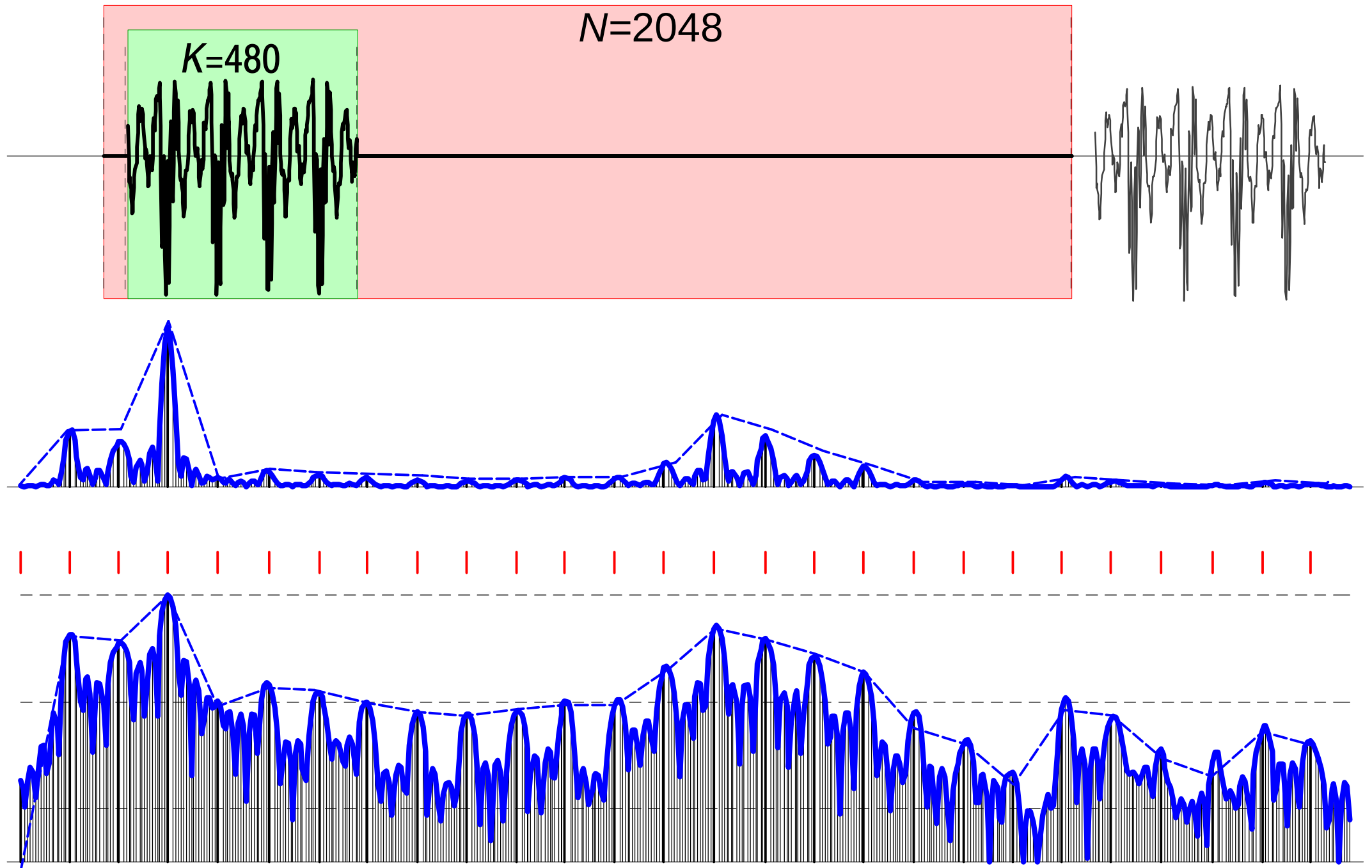
$K=N=480$



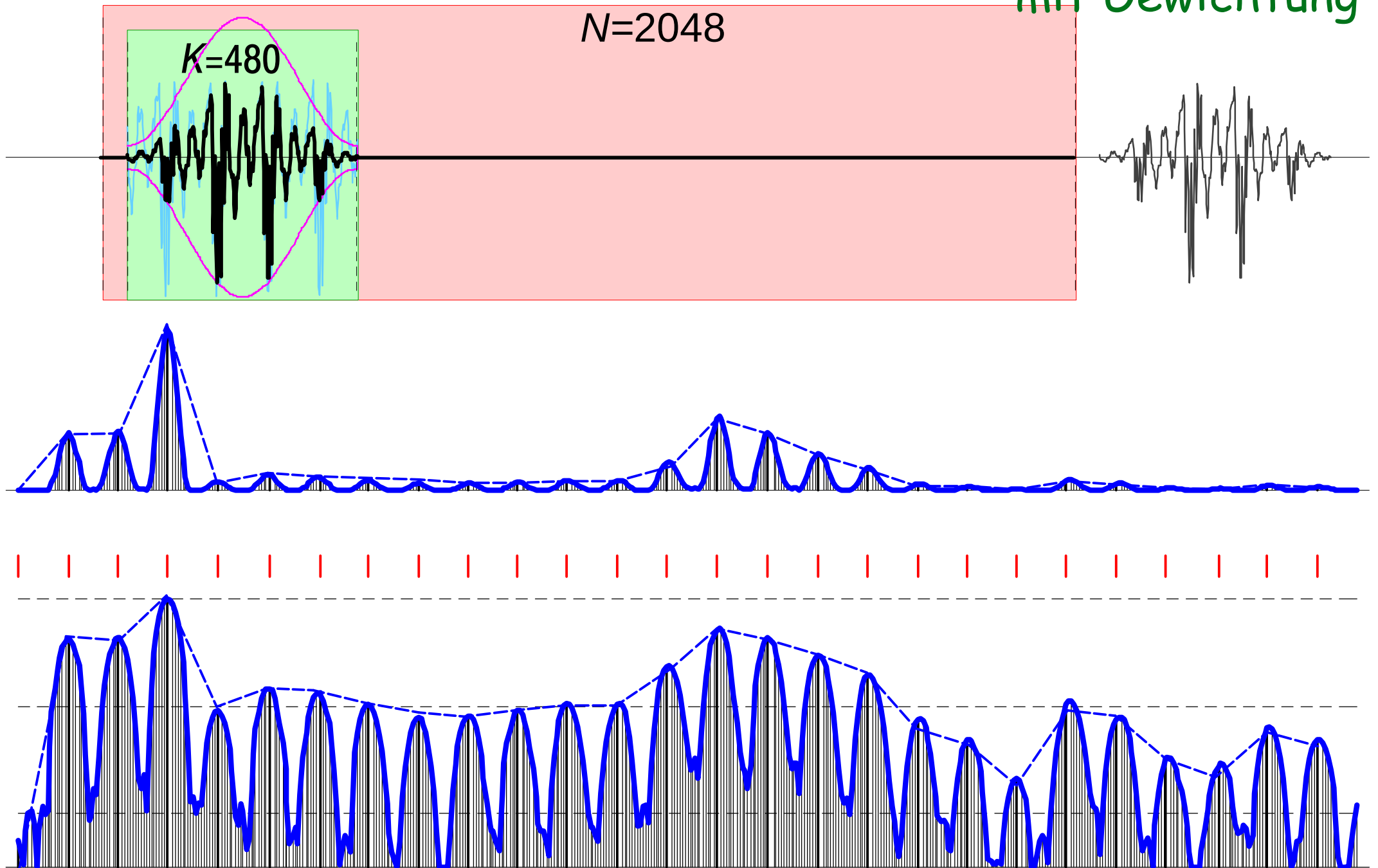
Harmonische Analyse (4,5 Perioden) mit Gewichtung



Nichtharmonische Analyse (4,5 Perioden)



Nichtharmonische Analyse (4,5 Perioden) mit Gewichtung

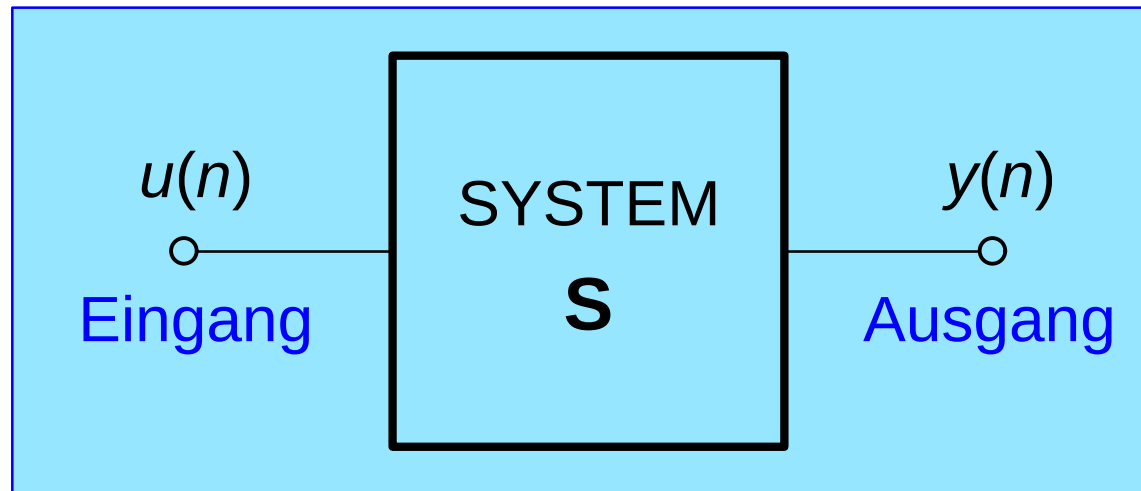


Grundlagen der Sprachsignalverarbeitung

3. Kurzzeitspektralanalyse

- 3.1 Harmonische Analyse (diskreter) periodischer Signale
- 3.2 Nichtharmonische Spektralanalyse
- 3.3 Kurzzeitspektren von Sprachsignalen
- 3.4 Das lineare Modell der Spracherzeugung in spektraler Darstellung**

Systemtheoretische Vorbetrachtung



Wir nehmen an, dass das System S kausal ist. Dann ist

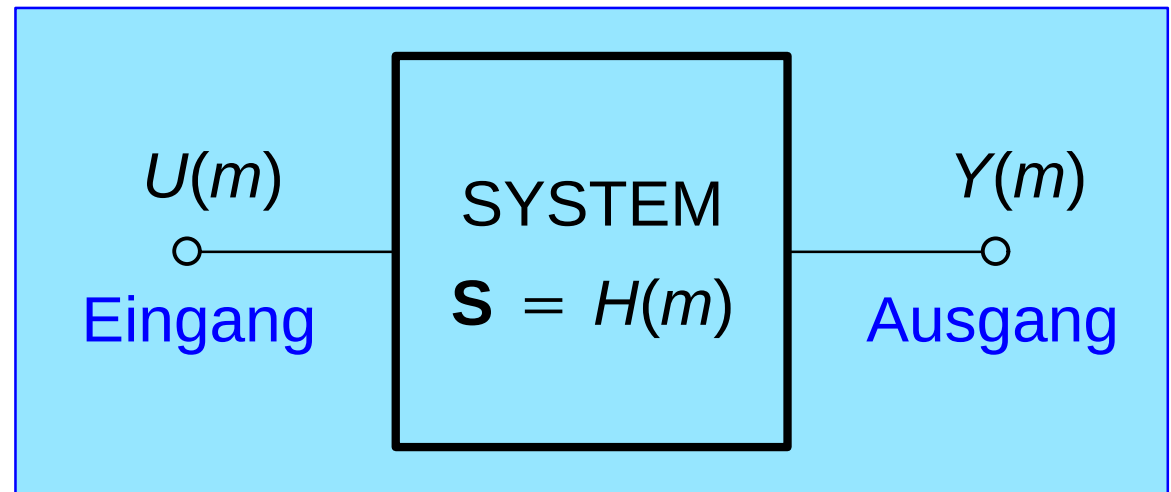
$$y(n) = S \{ u(k), k = \dots (-1) n \} \quad (2.7)$$

In dieser Form verbindet S das Ein- und das Ausgangssignal; genauer gesagt, wird für jeden Abtastzeitpunkt n der Abtastwert $y(n)$ des Ausgangssignals durch eine Verknüpfung der Gegenwart [repräsentiert durch den Abtastwert $u(n)$] und der gesamten Vergangenheit des Eingangssignals mit Hilfe der Rechenvorschrift S berechnet.

Übertragungsfunktion

Wir nehmen weiterhin an, dass das System **S** linear ist.

Ein lineares System verändert die Frequenz einer Sinusschwingung nicht.



Damit geht die Verknüpfung des Ausgangssignals $y(n)$ mit dem Eingangssignal $u(n)$ und seiner gesamten Vergangenheit über in die Verknüpfung *genau eines* Spektralwertes $U(m)$ des Eingangssignals mit dem zugehörigen Spektralwert $Y(m)$ des Ausgangssignals:

$$S \rightarrow H(m) := \frac{Y(m)}{U(m)} \quad \text{oder} \quad Y(m) = H(m) \cdot U(m) \quad (3.30)$$

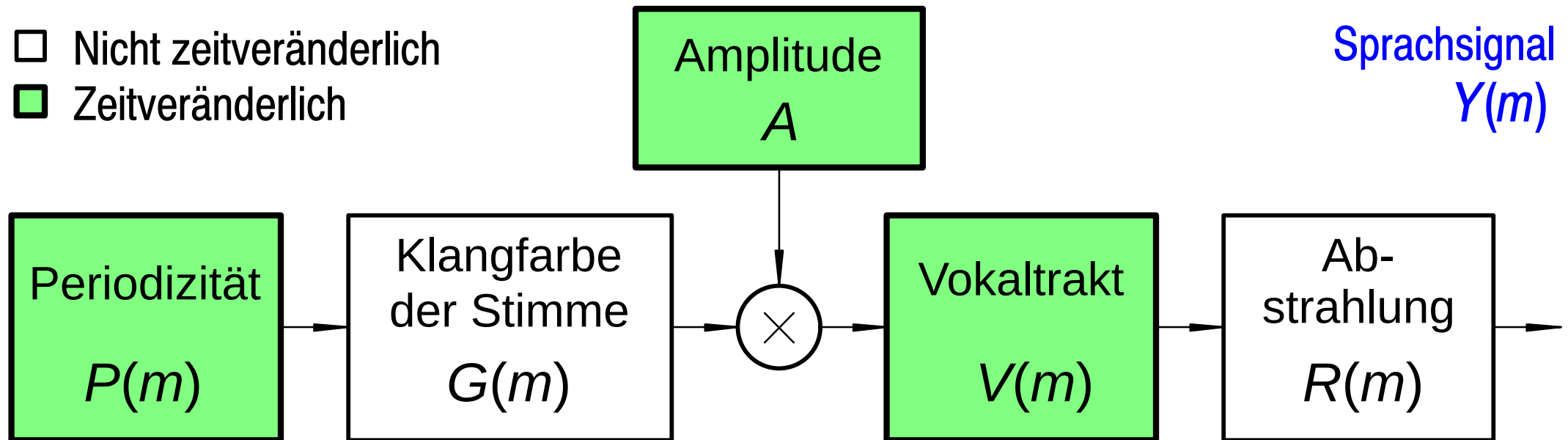
$H(m)$ wird als die **Übertragungsfunktion** des (linearen) Systems **S** bezeichnet.

Lineares Modell der Sprachproduktion

Stimmhafter Pfad

Spektrale Darstellung mit Übertragungsfunktion

- Nicht zeitveränderlich
- Zeitveränderlich



$$\begin{aligned} S(m) &= P(m) \cdot G(m) \cdot A \cdot V(m) \cdot R(m) \\ &= P(m) G(m) V(m) R(m) A \end{aligned}$$

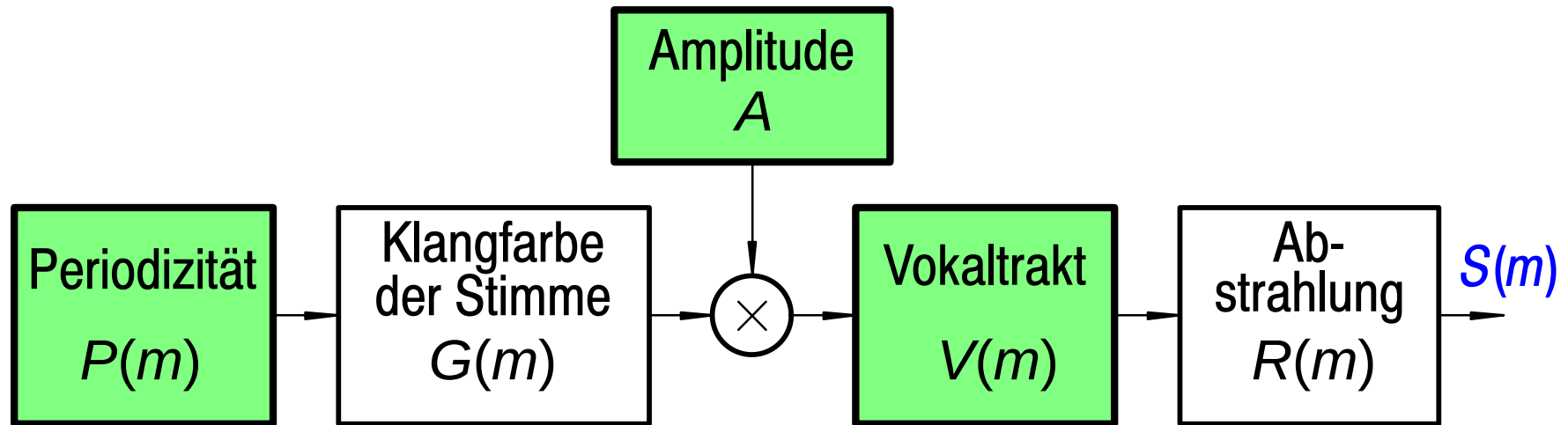
(3.31)

Grundlagen der Sprachsignalverarbeitung

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

Lineares Modell der Sprachproduktion

Stimmhafter Pfad

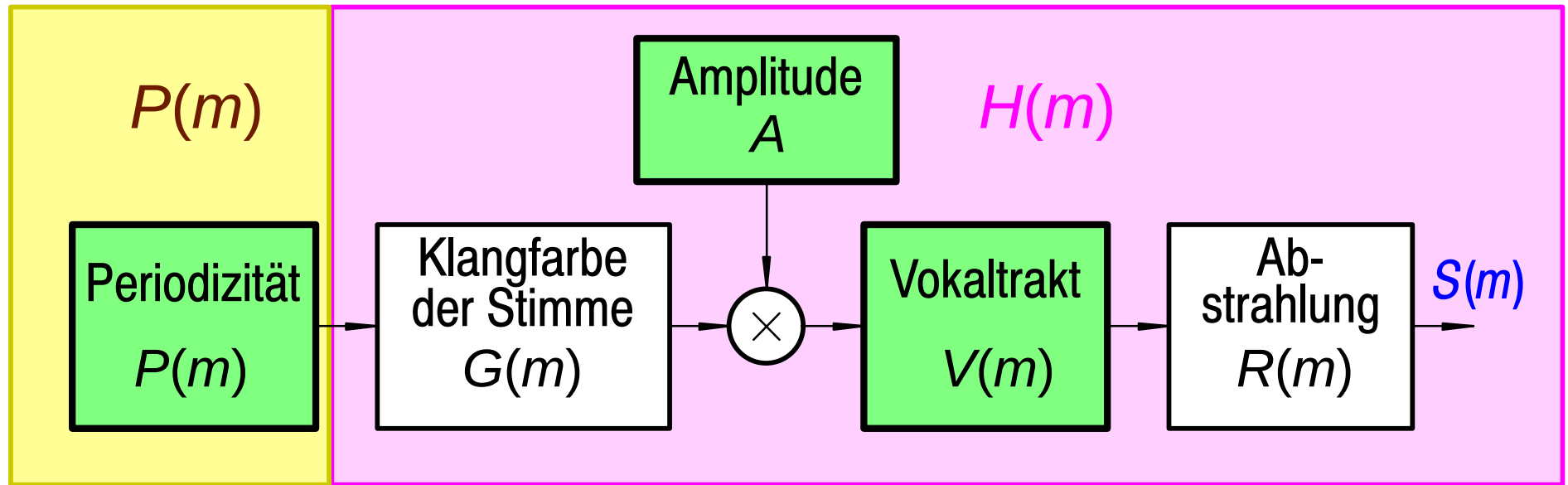


Eine der wichtigsten Aufgaben der Sprachsignalverarbeitung ist der Rückschluss vom Sprachsignal auf die Übertragungsfunktion der einzelnen Komponenten des Modells.

Dies ist eine Aufgabe der **Systemidentifikation**, also der Bestimmung einzelner Komponenten aus den Signalen.

Hierzu benötigt wird ein Verfahren, um die Komponenten quantitativ zu trennen.

Aufgaben der Sprachsignalanalyse

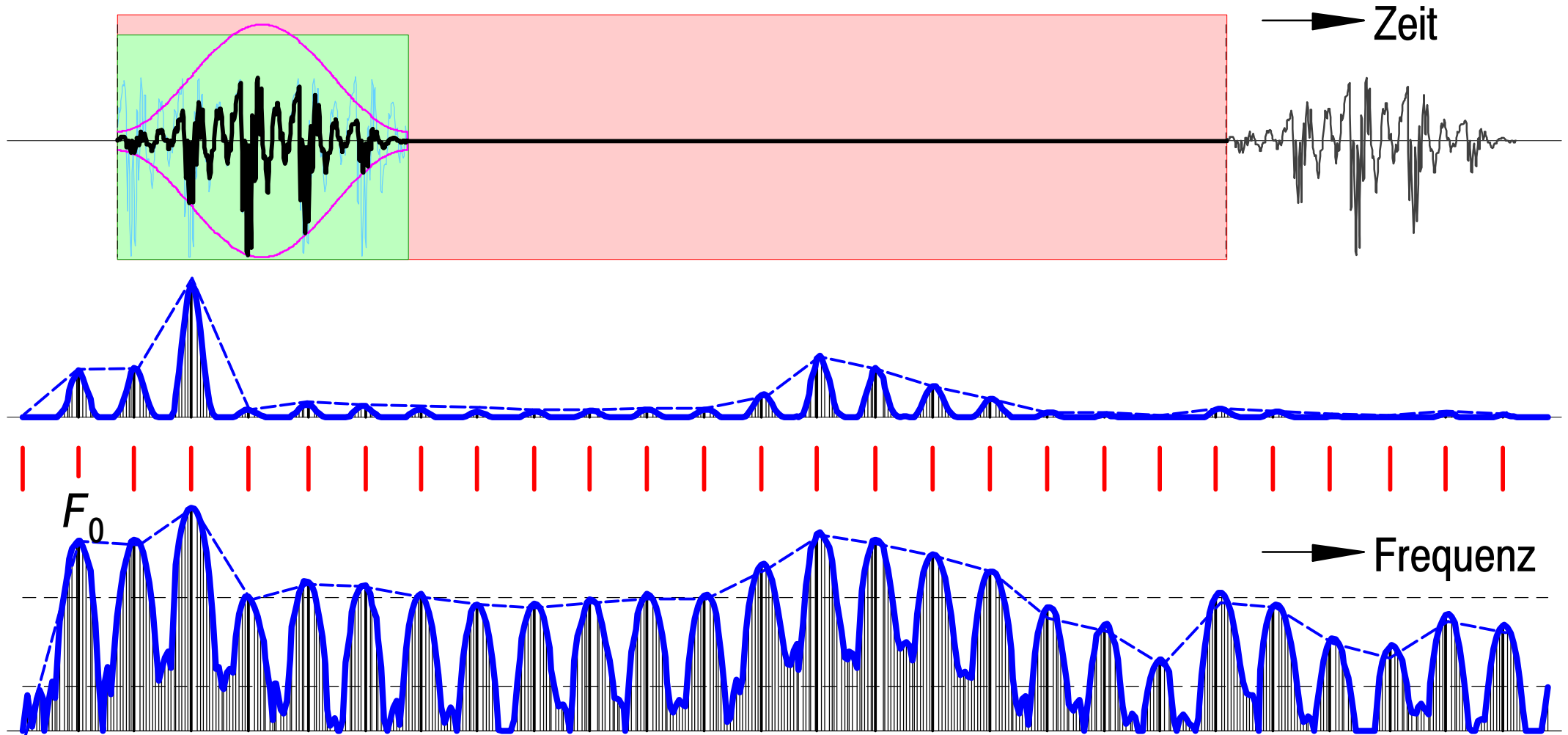


Vorrangige Aufgaben der Sprachanalyse:

- **Rückschluss auf $V(m)$:** Bestimmung der Vokaltraktparameter, Formantanalyse
- **Rückschluss auf $P(m)$:** Bestimmung der Sprachgrundfrequenz

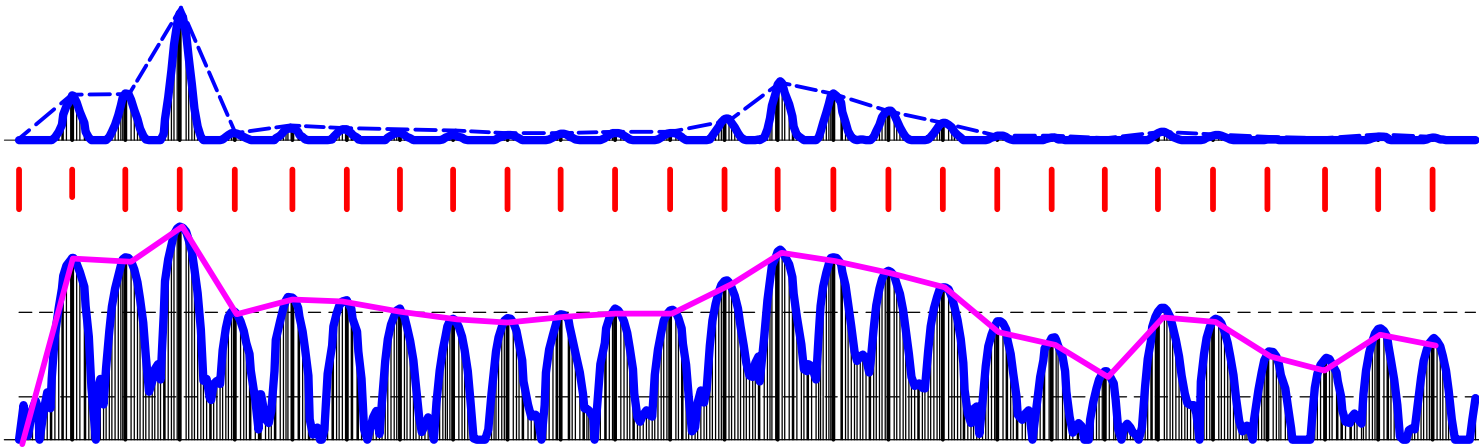
Besonders wichtig ist damit die Trennung der Komponente $P(m)$ vom Rest des Modells $H(m)$.

Spektrum und Modell



Die Periodizität äußert sich im Spektrum durch Maxima ("Welligkeiten") bei ganzzahligen Vielfachen von F_0 .

Grundannahme für spektrale Glättung

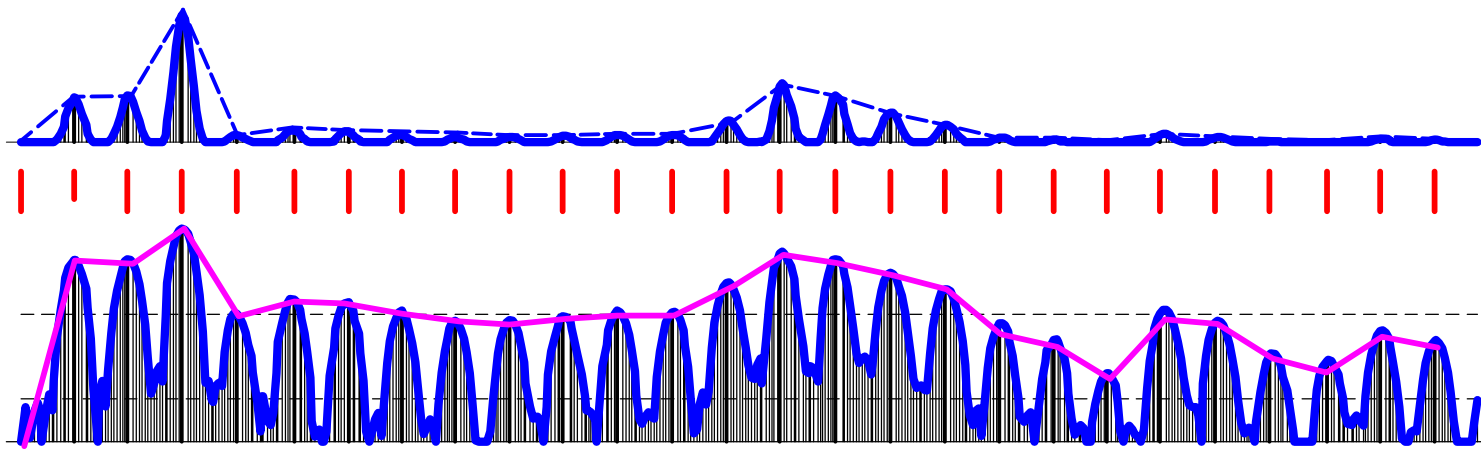


Spektrale Glättung entfernt die Welligkeiten aus dem Spektrum und erlaubt die Bestimmung der Parameter von $V(m)$, also der Formanten.

Grundannahme hierbei: Der Frequenzunterschied zwischen benachbarten Formanten ist groß gegen die Grundfrequenz F_0 .

Bei Frauen- und Kinderstimmen mit hoher Sprachgrundfrequenz ist diese Annahme oftmals nicht gegeben.

Drei grundsätzlich verschiedene Verfahren



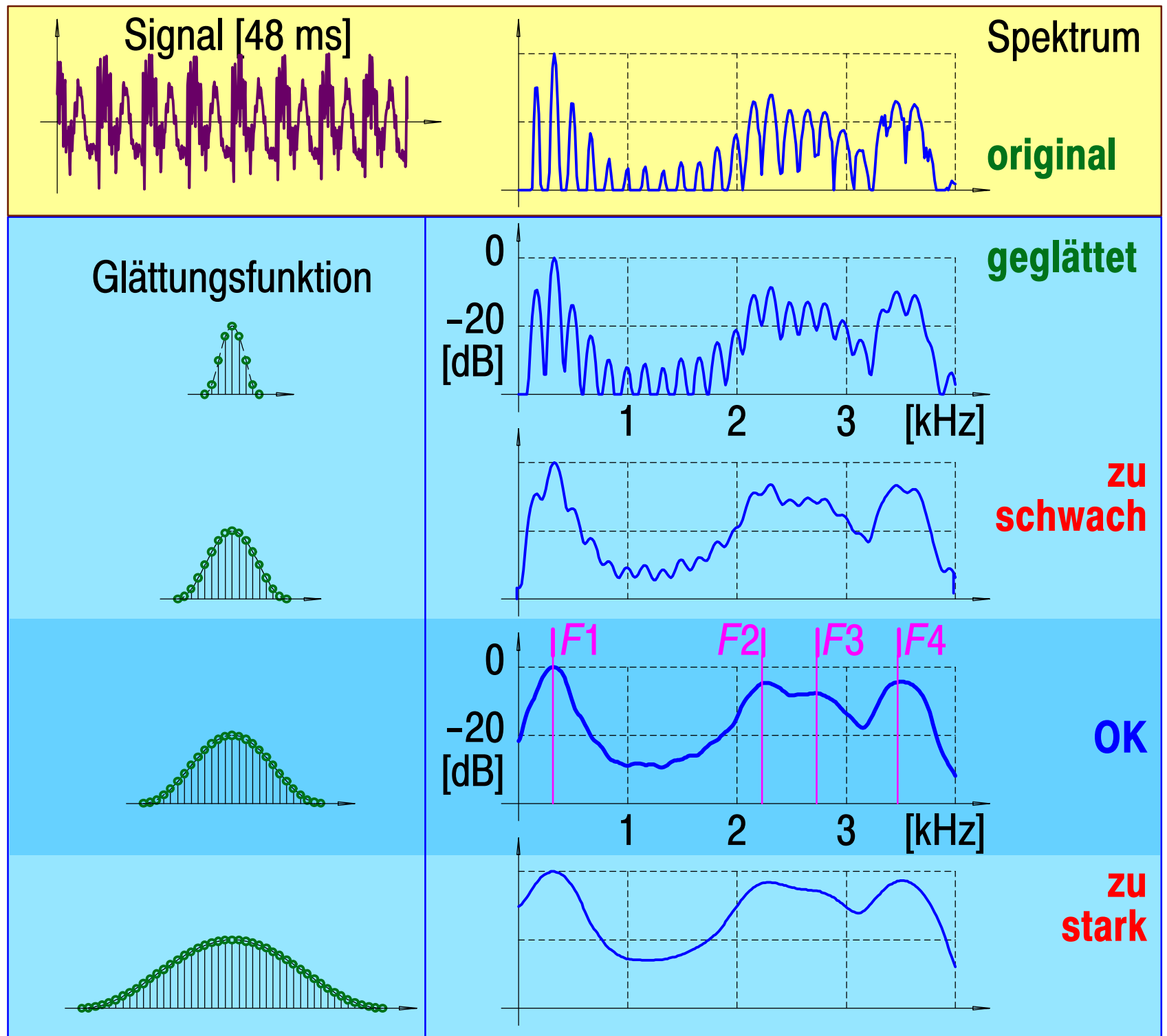
- Glättung durch **gewichtete Mittelwertbildung im Spektrum** (Abschnitt 4.1): Spektrales Tiefpassfilter; Wirkung der Periodizität wird entfernt.
- Glättung durch **Cepstrum** (Abschnitt 4.2): Trennung der Komponenten durch spektrale Filterung
- Beeinflussung des Spektrums durch **Veränderung der Fensterlänge K** bei der Spektralanalyse (Abschnitt 4.3)

Grundlagen der Sprachsignalverarbeitung

4. Vom Spektrum zum Spektrogramm

- 4.1 Glättung von Kurzzeitspektren durch gewichtete Mittelwertbildung
- 4.2 Glättung von Spektren mit Hilfe des Cepstrum-Verfahrens
- 4.3 Kombinierte Zeit-Frequenz-Darstellung. Spektrogramm und Lautgebirge

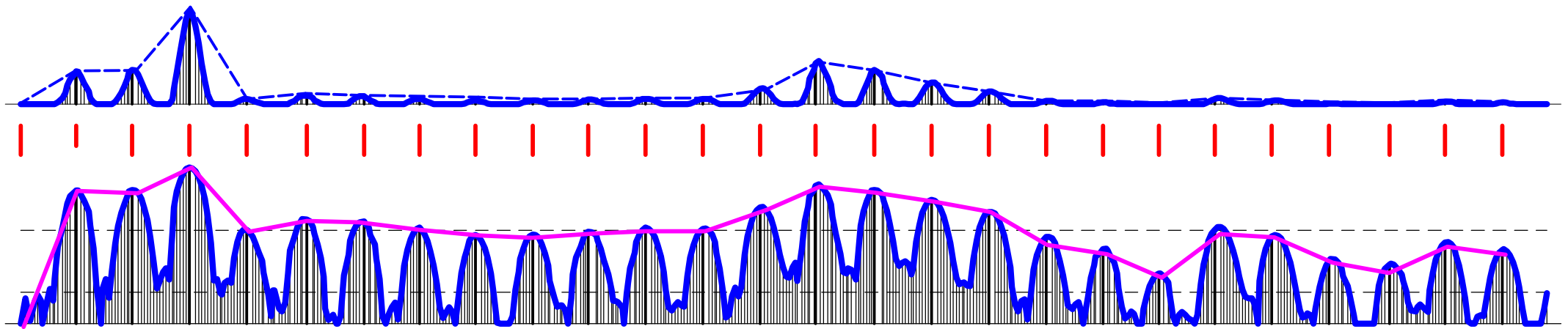
Glättung durch gewichtete Mittelung



Grundlagen der Sprachsignalverarbeitung

4. Vom Spektrum zum Spektrogramm

- 4.1 Glättung von Kurzzeitspektren durch gewichtete Mittelwertbildung
- 4.2 Glättung von Spektren mit Hilfe des Cepstrum-Verfahrens
- 4.3 Kombinierte Zeit-Frequenz-Darstellung. Spektrogramm und Lautgebirge

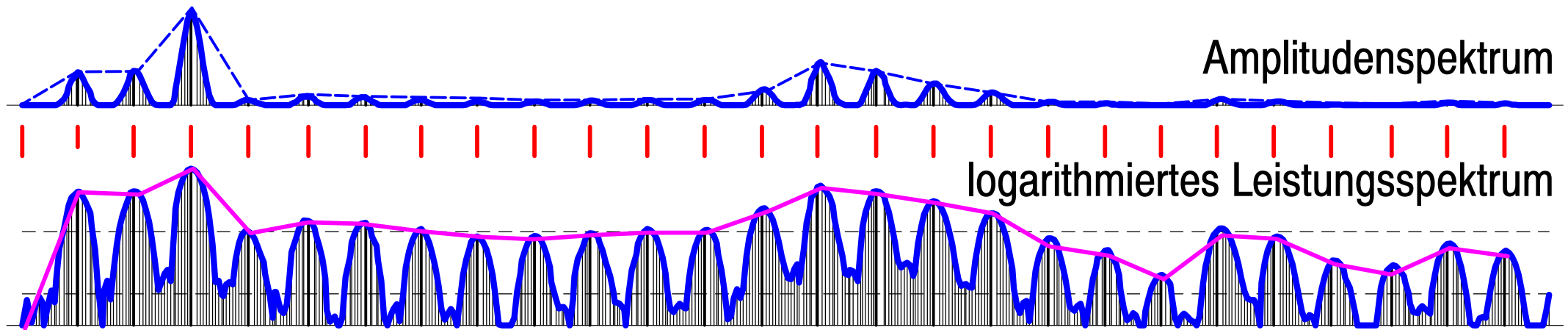


Betrachtet man das Spektrum als ein Signal (mit der Frequenz als unabhängiger Veränderlicher), so kann man es sich denken als Überlagerung einer kurzwelligen Schwingung mit der "Periode" F_0 und langwelliger Schwingungen, die den Verlauf der Übertragungsfunktion des Vokaltrakts darstellen.

Durch eine Art "Langpass"-Filterung können wir die kurzwellige Schwingung entfernen.

Hierzu berechnen wir ein "Spektrum des Spektrums" und sorgen dafür, dass die F_0 zugehörige Komponente verschwindet.

Grundidee [2]



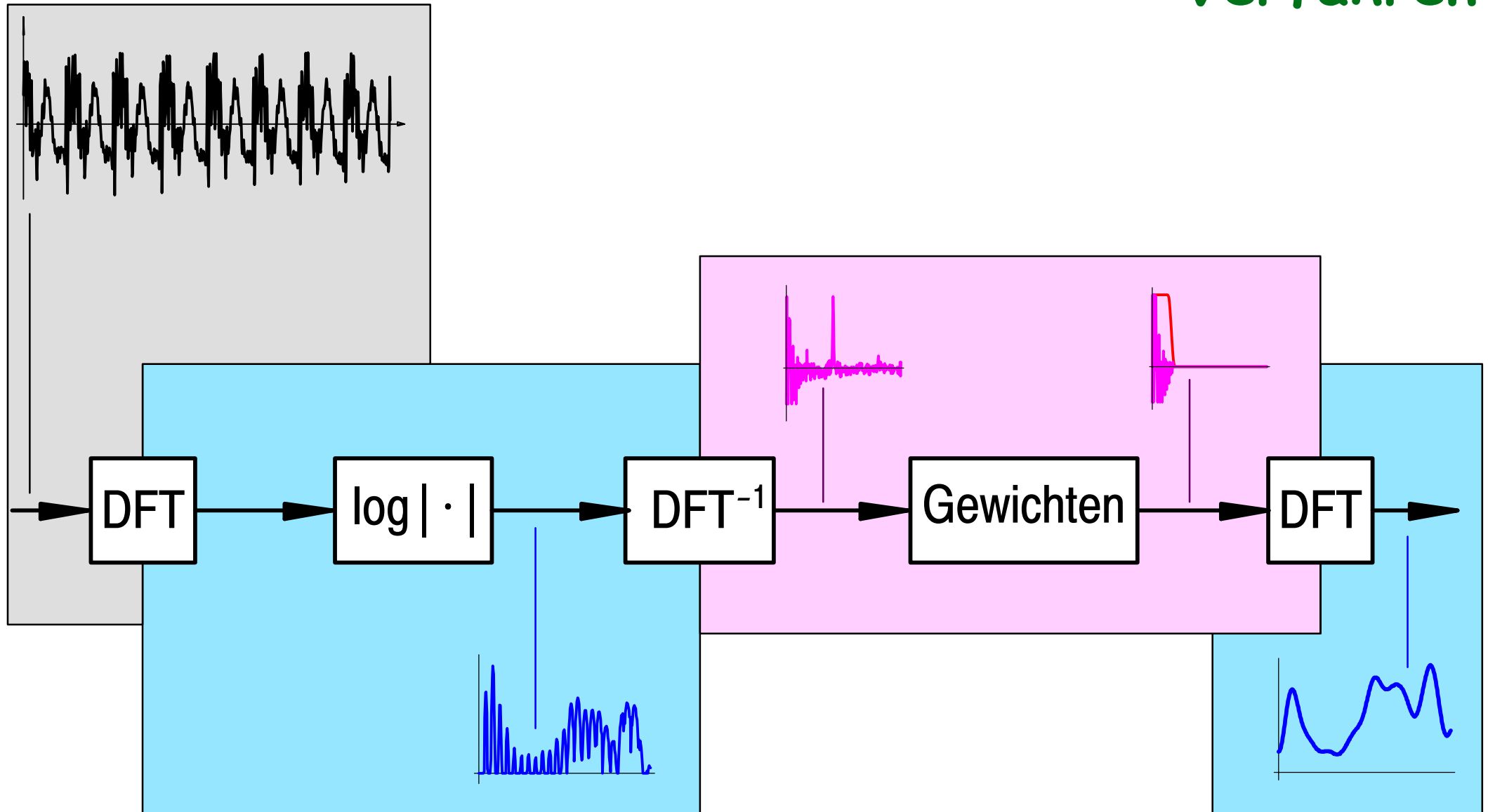
Im Amplitudenspektrum ist die Verknüpfung der Komponenten $P(m)$ und $V(m)$ multiplikativ.

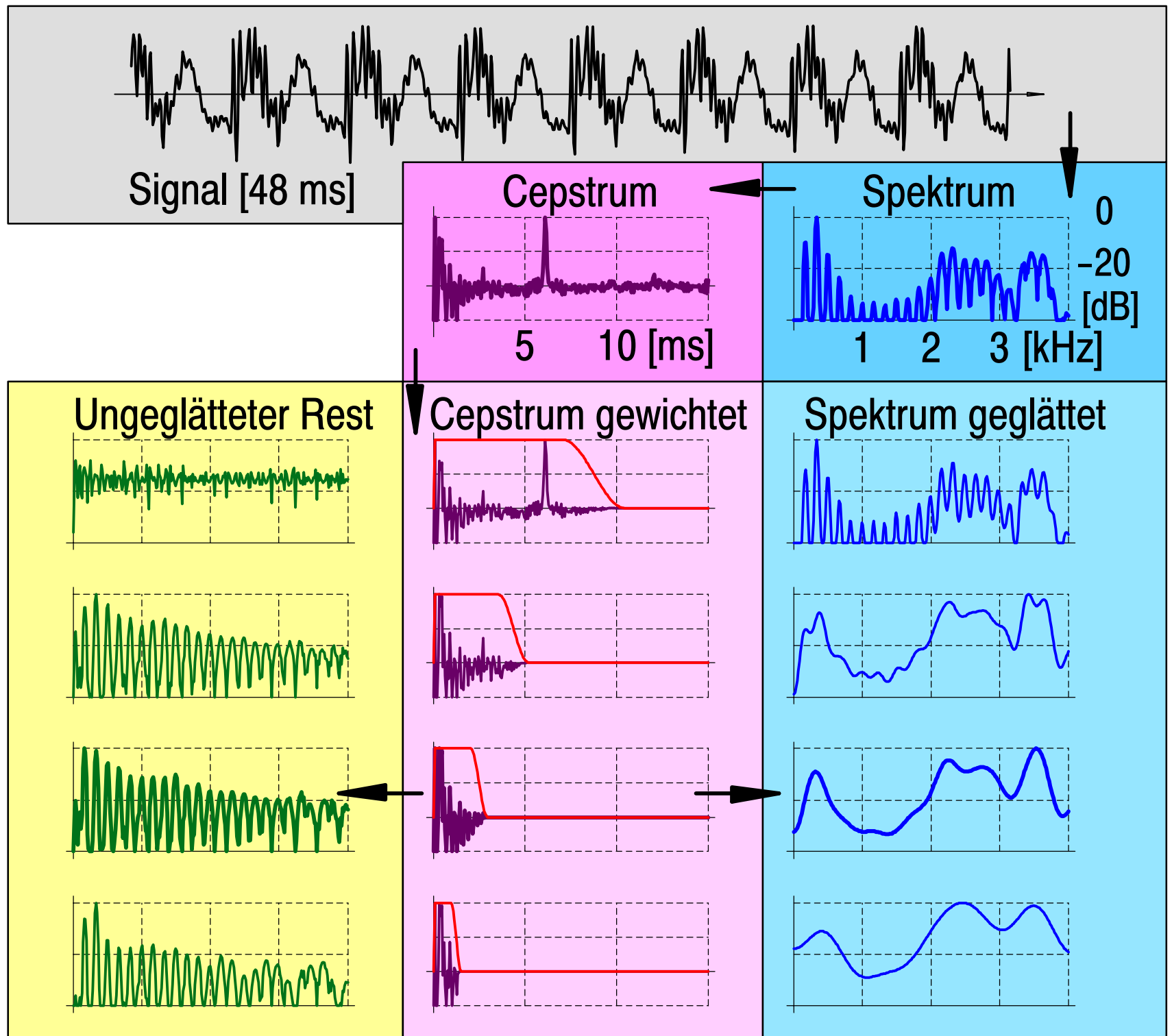
Die Verknüpfung wird additiv, wenn wir das Spektrum logarithmieren:

$$\log [P(m) \cdot V(m)] = \log P(m) + \log V(m)$$

Zur Trennung der Komponenten wird das logarithmierte Spektrum in den Zeitbereich zurücktransformiert. Dabei entsteht nicht mehr das ursprüngliche Signal.

Spektrale Glättung durch Cepstrum Verfahren

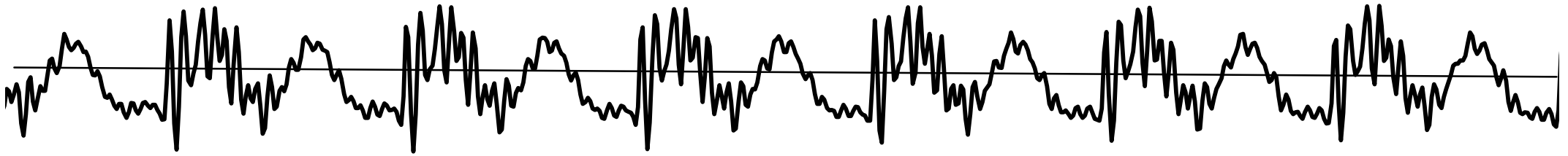




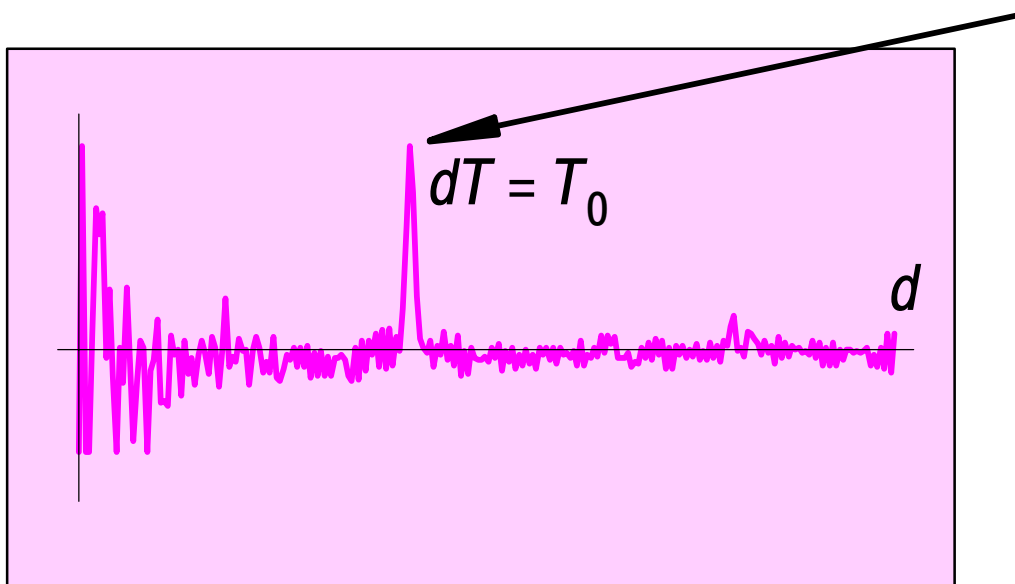
Spektrale
Glättung
durch
Cepstrum

Beispiel

Spitze bei $dT=T_0$



Im Sprachsignal treten die höchsten Amplituden zu beliebigen Zeitpunkten auf - wann immer die Glottis des Sprechers sich schließt.



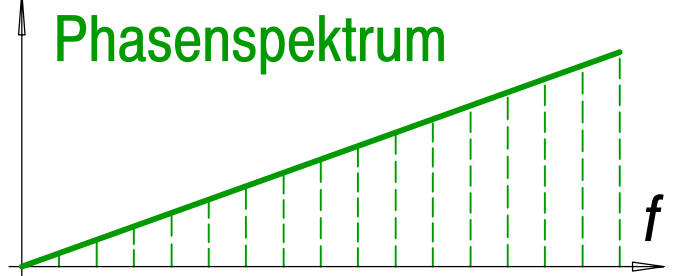
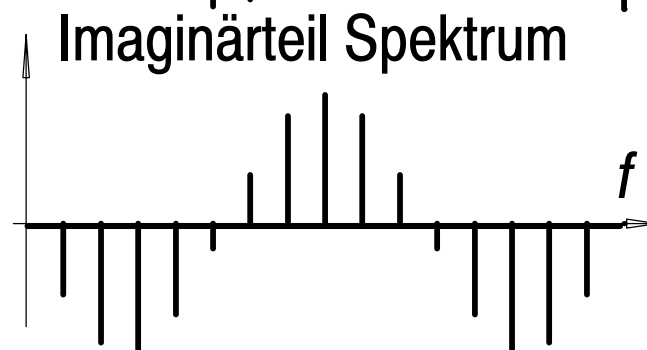
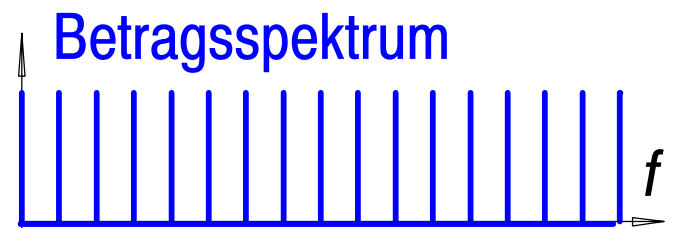
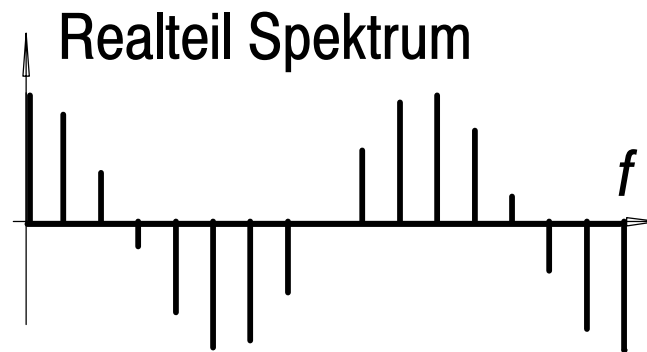
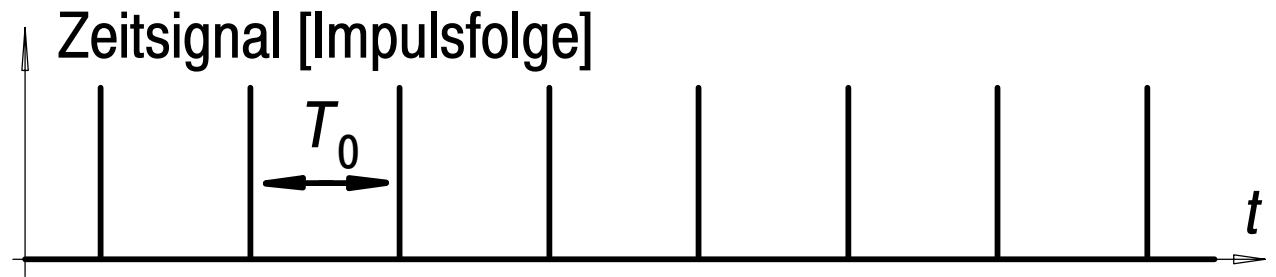
Warum erscheint im Cepstrum diese Spitze, die den Einfluss der Periodizität im Spektrum repräsentiert, stets bei $dT=T_0=1/F_0$?

Der komplexe Logarithmus trennt Betrag und Phase

$$\begin{aligned} \underbrace{\ln S(m)}_{\text{komplexes Spektrum}} &= \ln \left(\underbrace{|S(m)|}_{\text{Amplituden-Spektrum}} \underbrace{e^{j\varphi(m)}}_{\text{Phasen-Spektrum}} \right) = \ln |S(m)| + \ln \left(e^{j\varphi(m)} \right) \\ &= \underbrace{\ln |S(m)|}_{\text{Realteil [repräsentiert Amplituden-spektrum]}} + \underbrace{j\varphi(m)}_{\text{Imaginärteil [repräsentiert Phasen-spektrum]}} \end{aligned}$$

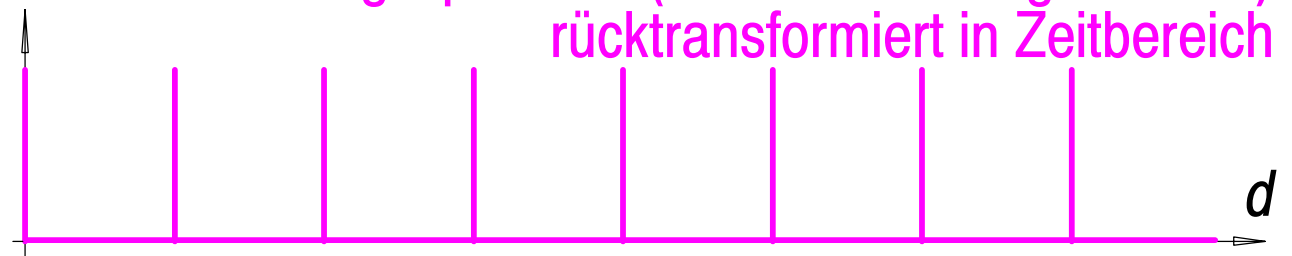
Wird nur das Amplitudenspektrum weiter bearbeitet, so heißt dies, dass die Phasen aller Teilschwingungen zu Null gesetzt werden.

Spektrum und Cepstrum einer Impulsfolge



Mit $\varphi=0$ liegt einer
der Impulse stets
bei $d=0$.

Betragsspektrum (oder dessen Logarithmus)
rücktransformiert in Zeitbereich



Grundlagen der Sprachsignalverarbeitung

4. Vom Spektrum zum Spektrogramm

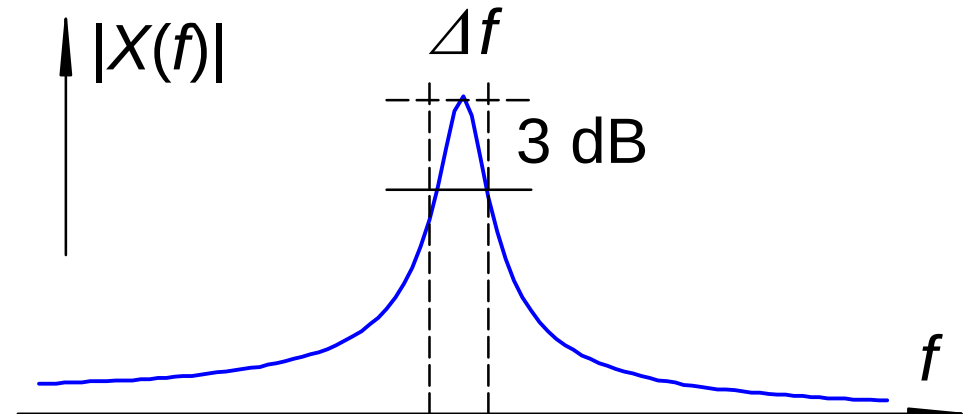
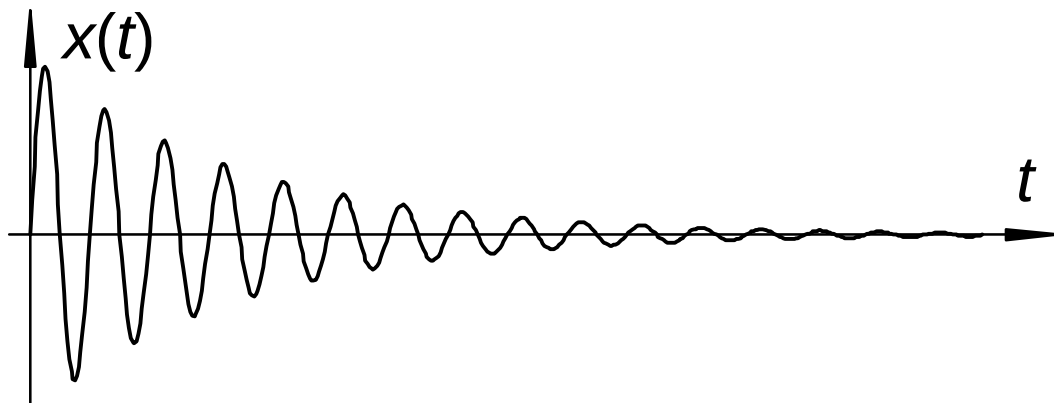
- 4.1 Glättung von Kurzzeitspektren durch gewichtete Mittelwertbildung
- 4.2 Glättung von Spektren mit Hilfe des Cepstrum-Verfahrens
- 4.3 Kombinierte Zeit-Frequenz-Darstellung. Spektrogramm und Lautgebirge
 - 4.3.1 Spektrale und zeitliche Auflösung oder das Zeit-Bandbreite-Gesetz der Nachrichtentechnik
 - 4.3.2 Visualisierungsformen

Dämpfung und Bandbreite [1]

Das spektrale Äquivalent zur Dämpfung ist die Bandbreite.

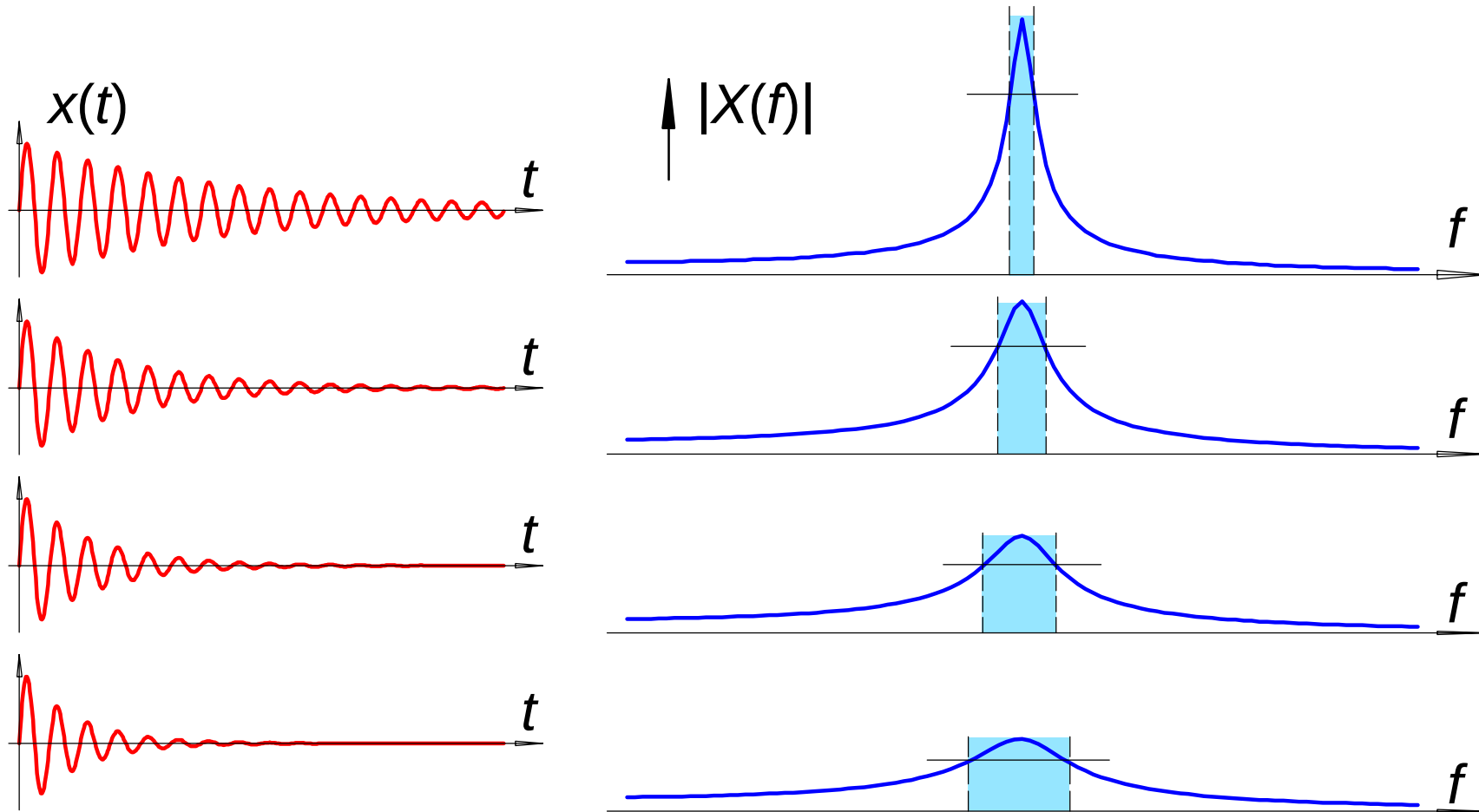
Die Bandbreite Δf ist im Frequenzbereich definiert als das Frequenzband, innerhalb dessen die spektrale Energie auf einen Wert von **50% des Maximalwertes** abgeklungen ist.

Dies entspricht einem Abfall von 3 dB und einem Rückgang der Amplitude auf $1/\sqrt{2} \approx 0,7071$.



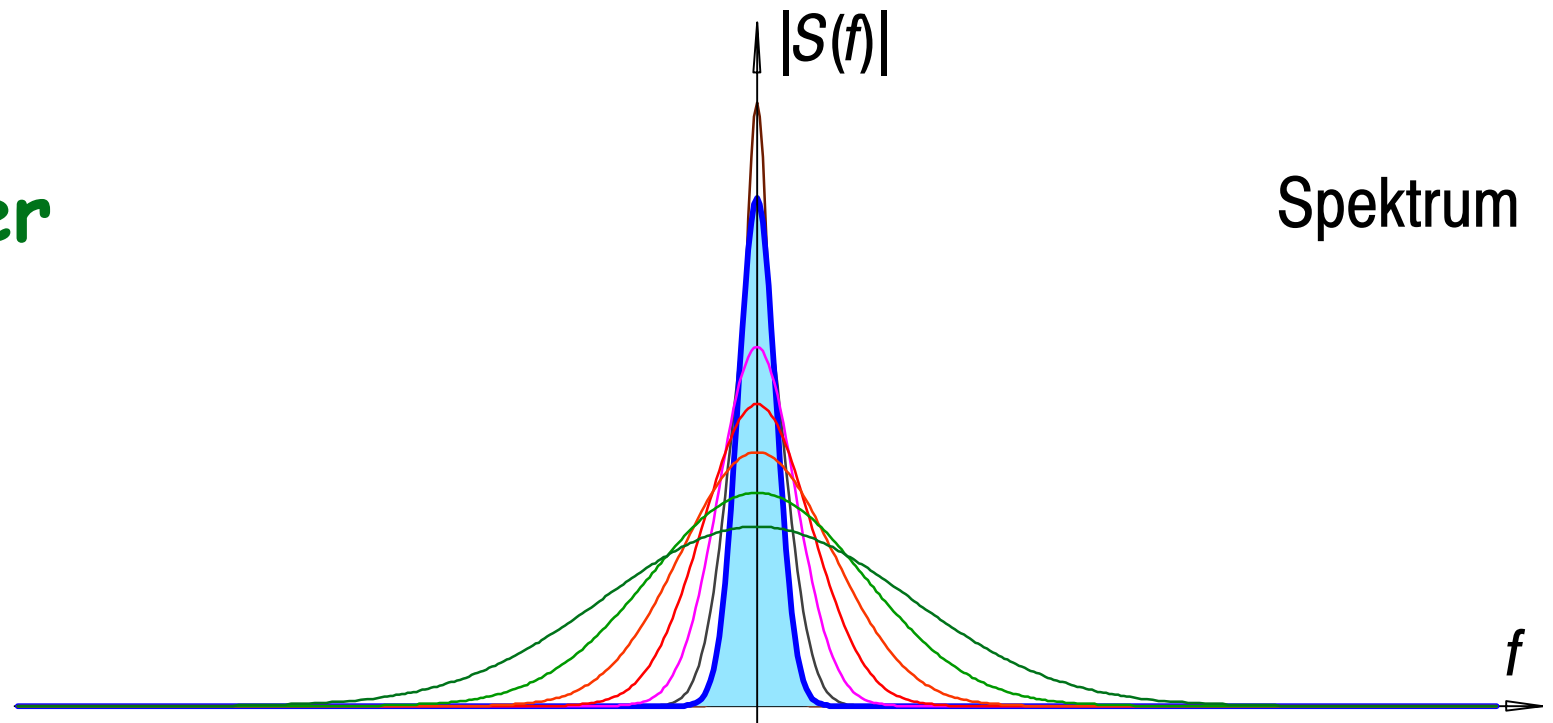
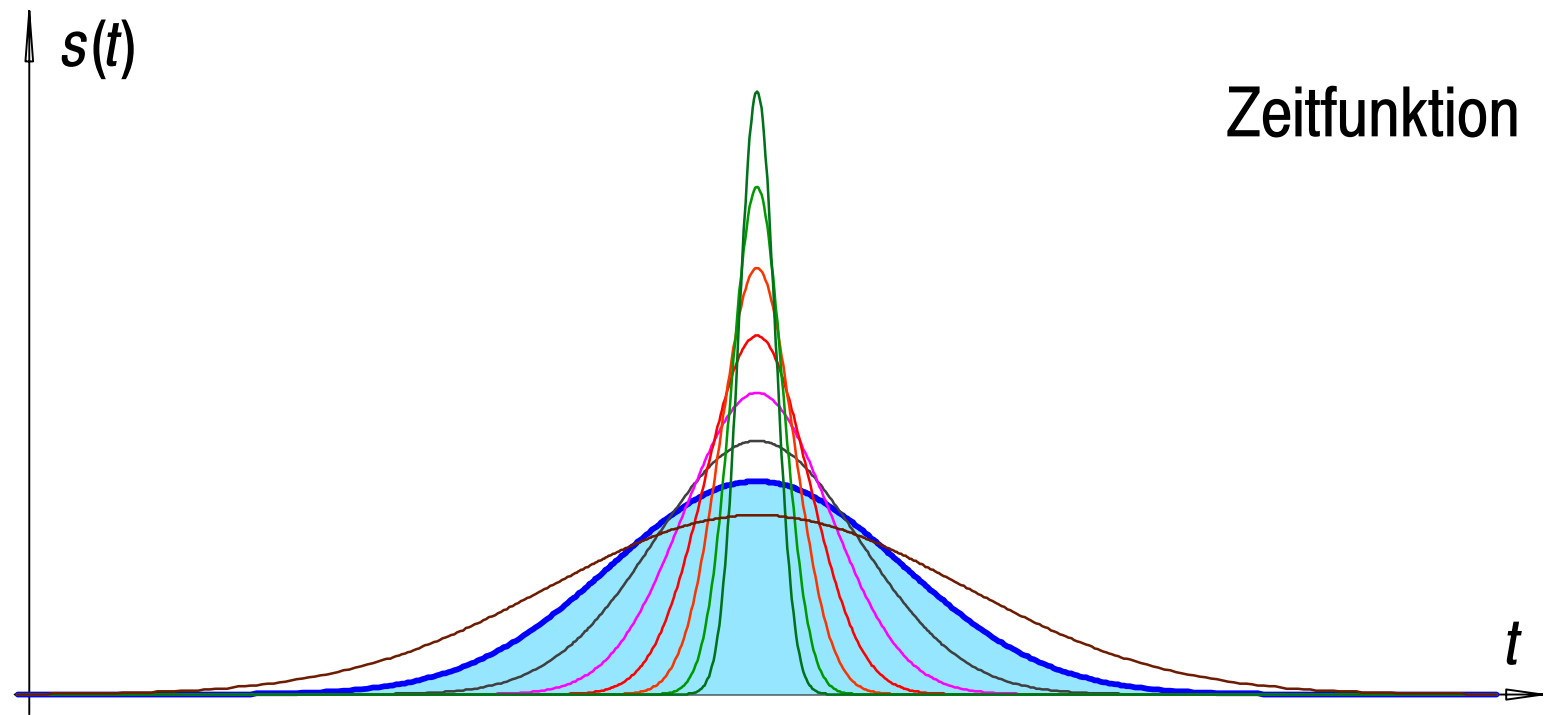
Dämpfung und Bandbreite [2]

Dämpfung und Bandbreite sind verbunden über das Zeit-Bandbreite-Gesetz der Nachrichtentechnik.



Je größer die Dämpfung einer Schwingung, desto höher die Bandbreite ihres Spektrums und umgekehrt.

Zeit-
Bandbreite-
Gesetz der
Nachrichten-
technik,
gezeigt an einer
Gauß'schen
Glockenkurve
(MARKO, 1956)



Zeit-Bandbreite-Gesetz der Nachrichtentechnik

Zeit mal Bandbreite ergibt eine Konstante.

Für die Gauß'sche Glockenkurve:

$$s(t) = \hat{s} \exp(-at^2)$$

$$S(f) = \hat{S} \exp(-\beta f^2)$$

$$\Delta t \cdot \Delta f \geq 1$$

Für eine Zeit-Frequenz-Darstellung von Sprachsignalen heißt dies: Die Information über die Periodizität ist entweder nur im Zeitbereich oder nur im Frequenzbereich sichtbar.

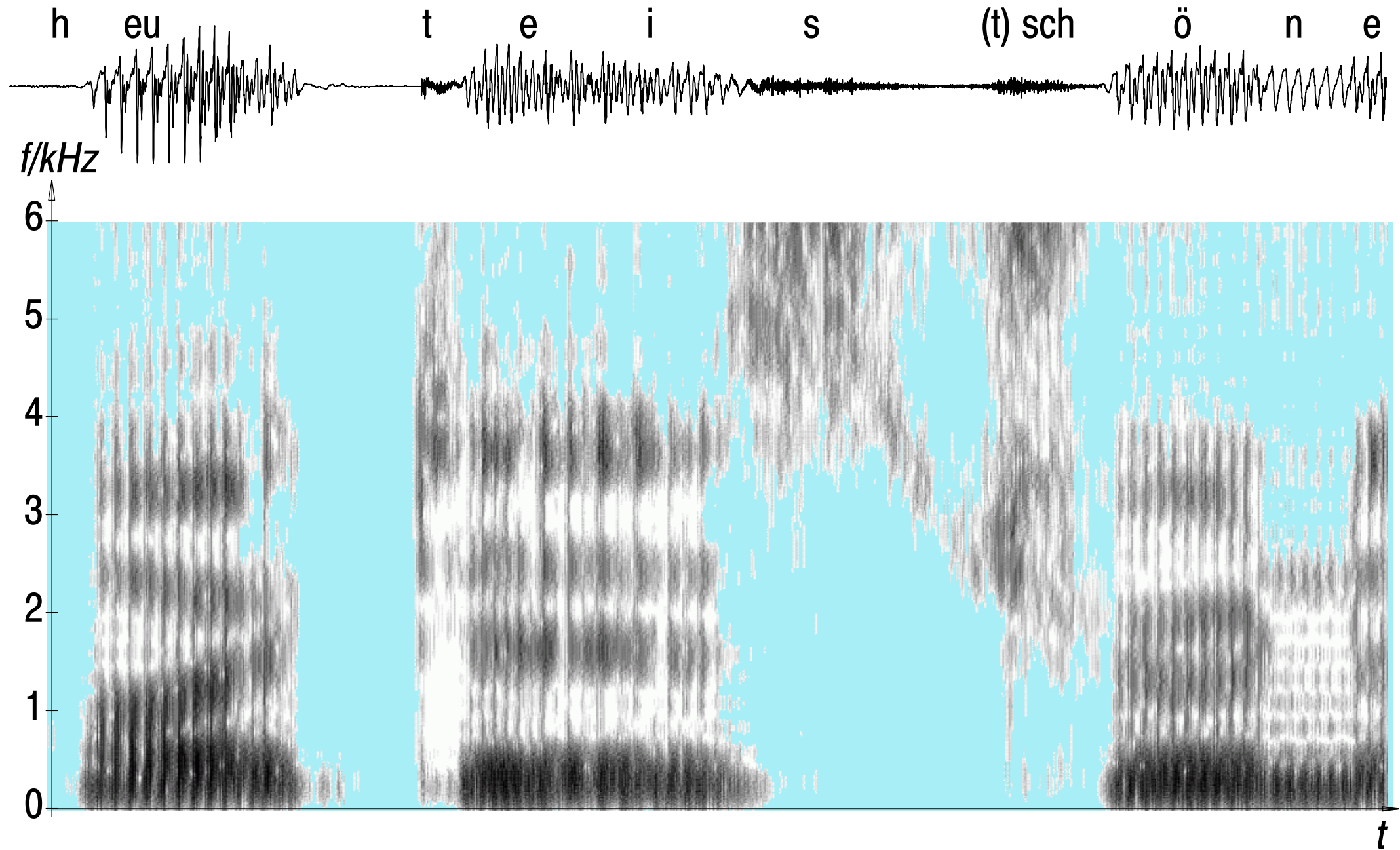
Wird ein Spektrogramm aus geglätteten Spektren zusammengesetzt, so erscheint die Information über die Periodizität nirgends.

Grundlagen der Sprachsignalverarbeitung

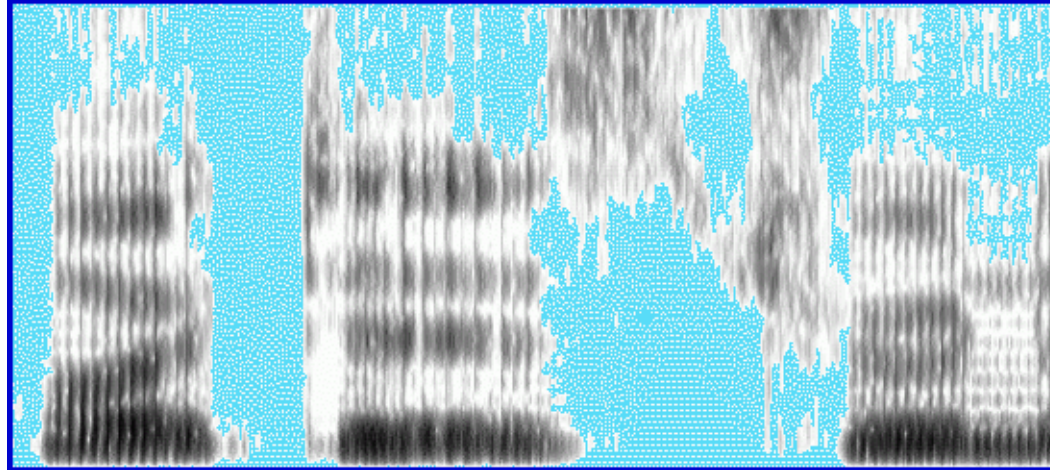
4. Vom Spektrum zum Spektrogramm

- 4.1 Glättung von Kurzzeitspektren durch gewichtete Mittelwertbildung
- 4.2 Glättung von Spektren mit Hilfe des Cepstrum-Verfahrens
- 4.3 Kombinierte Zeit-Frequenz-Darstellung. Spektrogramm und Lautgebirge
 - 4.3.1 Spektrale und zeitliche Auflösung oder das Zeit-Bandbreite-Gesetz der Nachrichtentechnik
 - 4.3.2 Visualisierungsformen

Breitbandspektrogramm



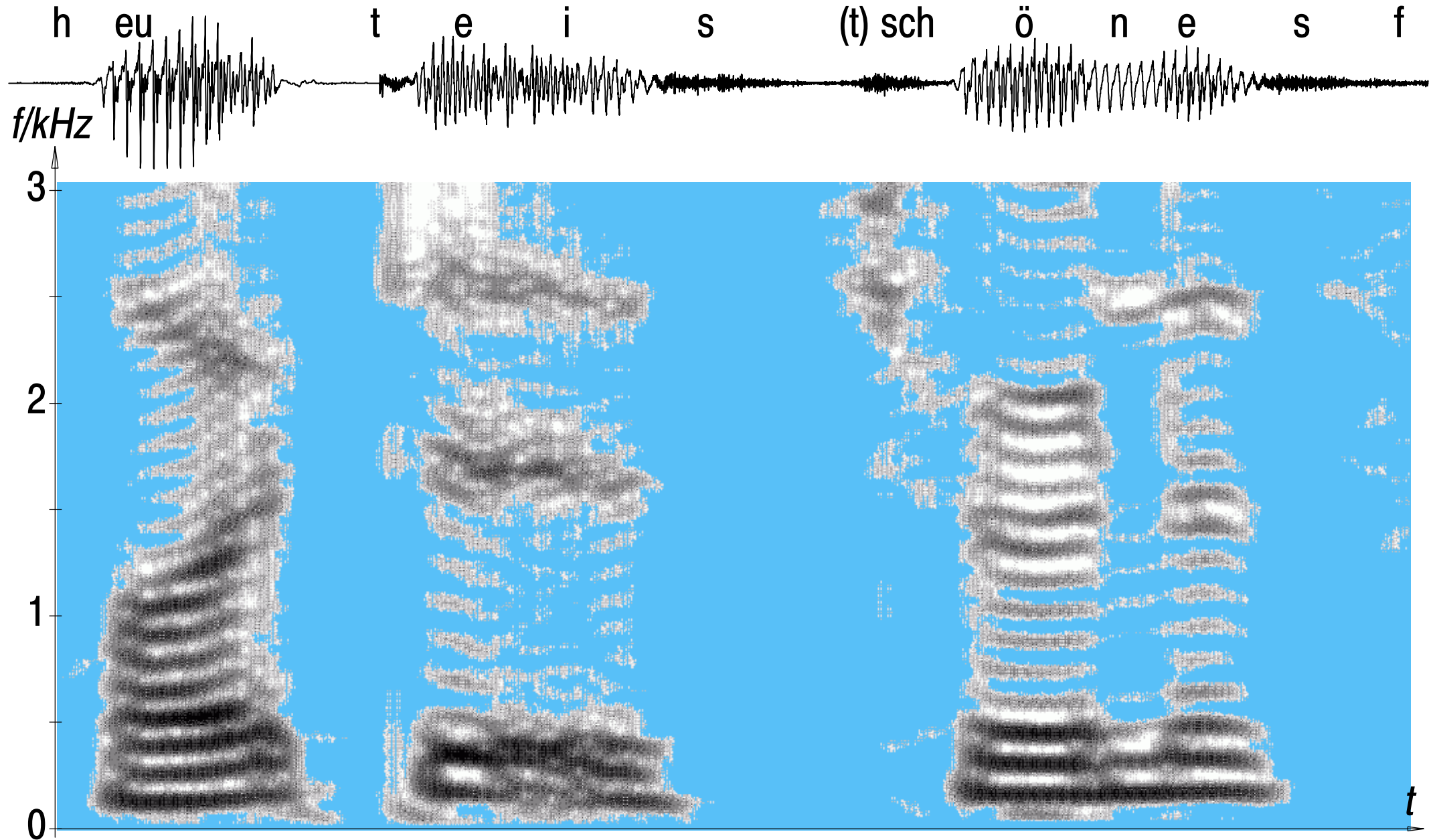
Das Breitbandspektrogramm ...



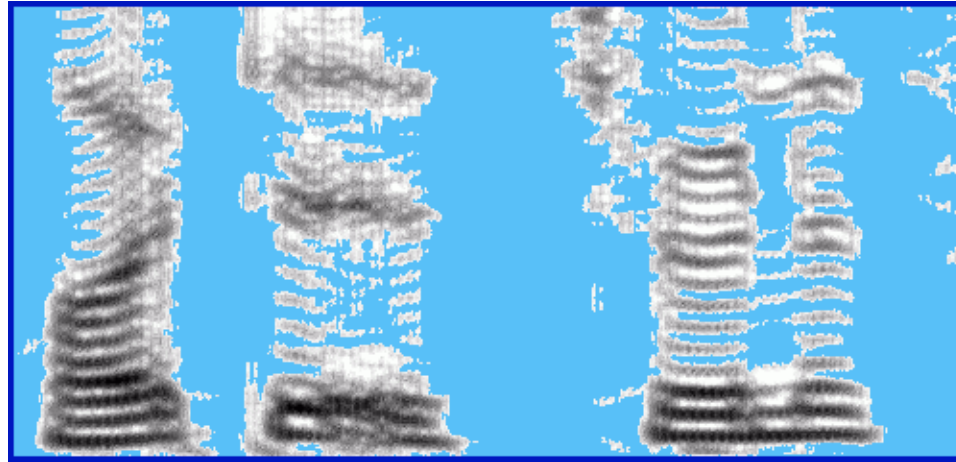
- hat ein kurzes Zeitfenster am Signaleingang
- kann deshalb *keine* Harmonischen in der spektralen Darstellung auflösen
- Da das Zeitfenster weniger als eine Grundperiode umfasst, erscheinen die Führungsamplituden am Beginn jeder Grundperiode als vertikale dunkle Linien (größere Amplituden).
- Die Information über die Grundfrequenz wird hier im Zeitbereich dargestellt.

Das Breitbandspektrum eignet sich zur Messung von Formanten und zur Visualisierung des Gesamtbildes einer sprachlichen Äußerung.

Schmalbandspektrogramm



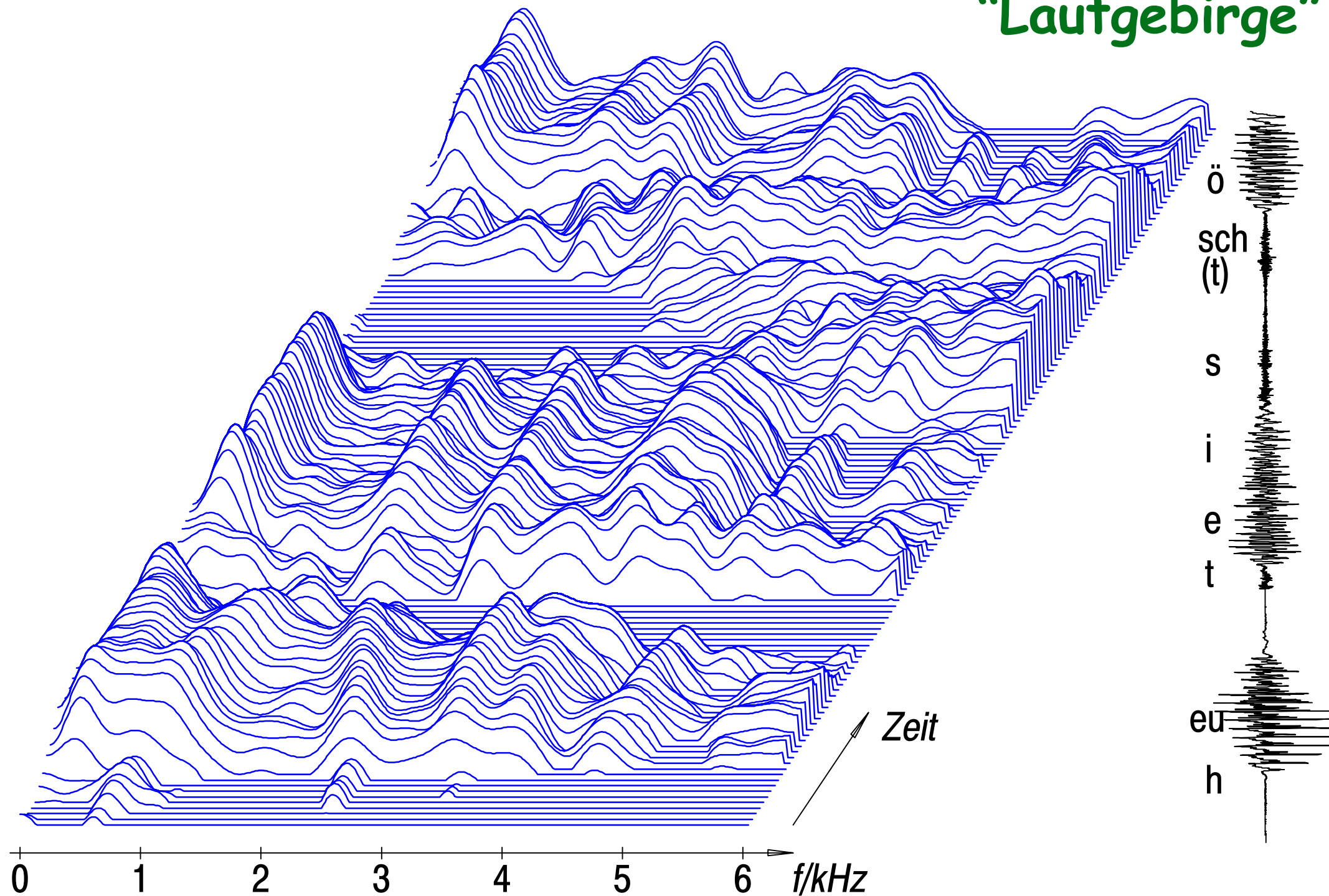
Das Schmalbandspektrogramm ...



- hat ein langes Zeitfenster am Signaleingang (in der Regel mehr als 2 Grundperioden)
- löst deshalb die Harmonischen in der spektralen Darstellung auf
- Die Information über die Grundfrequenz wird hier im Frequenzbereich dargestellt.
- Da das Zeitfenster mehr als eine Grundperiode umfasst, erscheint keinerlei Information über die Periodizität in der zeitlichen Darstellung.

Das Schmalbandspektrum eignet sich zur Messung und Visualisierung der Intonationskontur einer sprachlichen Äußerung.

“Lautgebirge”



Das Lautgebirge ...

- ist eine pseudo-dreidimensionale Amplitude-Zeit-Frequenz-Darstellung.
- Die einzelnen Spektren sind geglättet.
- Ein Lautgebirge enthält deshalb keine Information mehr über die Periodizität.

Das Lautgebirge eignet sich zur Visualisierung von Formantverläufen, da es einen sehr plastischen Eindruck vom Amplitudenverlauf vermittelt. Es hat sich jedoch gegenüber dem Spektrogramm nie richtig durchsetzen können.