

IV. Digitale Signalverarbeitung

Nachdem in den vorangegangenen Kapiteln die grundlegenden Eigenschaften der in der physikalischen Meßtechnik aufgenommenen Signale und ihrer Störungen behandelt wurden, sowie die Techniken, um diese Störungen zu minimieren, soll im folgenden die weitergehende Verarbeitung und Analyse dieser Signale behandelt werden, die fast ausschließlich digital erfolgt, d. h. unter Einsatz des Computers.

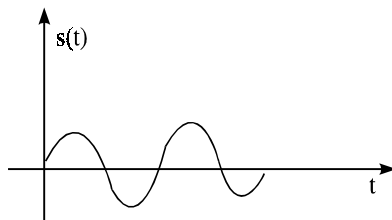
IV.1 Diskrete Spektralanalyse

Eine der häufigsten Anwendungen der digitalen Signalverarbeitung zur Auswertung von physikalischen Meßsignalen ist die Spektralanalyse von digitalisierten Signalen. Um die dabei auftretenden Probleme verstehen zu können, soll zunächst das Spektrum von analogen und digitalisierten Signalen betrachtet werden. Dabei wird für die digitale Signalverarbeitung der Fall angestrebt, daß sowohl das Zeitsignal als auch das Spektrum als diskrete Folgen im Rechner abgespeichert werden können. Um zu dieser Eigenschaft zu gelangen, betrachten wir zunächst aperiodische Signale, die entweder kontinuierlich oder diskret sein können. Für den diskreten Fall kann man das Signal auch als Produkt des kontinuierlichen Signals mit einem δ -Kamm auffassen, so daß sich das Spektrum dieses diskretisierten Signals als Faltung des Original-Spektrums mit der Fourier-Transformierten des δ -Kamms, nämlich dem δ -Kamm mit der inversen Periodizität schreiben läßt:

Aperiodische Signale

Kontinuierliche Signale

$$s(t), -\infty < t < +\infty$$

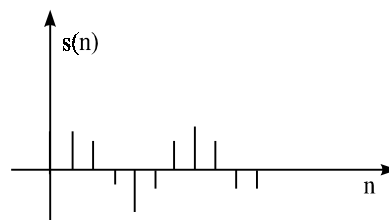


$$S(f) = \int_{-\infty}^{\infty} s(t) e^{-2\pi i f t} dt$$

Diskrete Signale

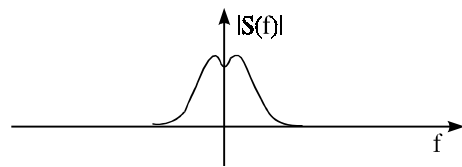
$$s(n), n \in \mathbb{Z}, s(n) = \begin{cases} s(t), t = n \\ 0 \text{ sonst} \end{cases}$$

(IV.1)

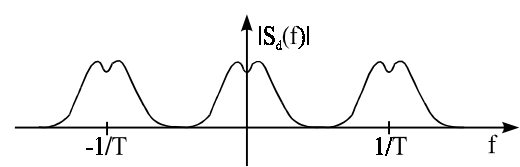


$$S_d(f) = S(f) * \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T}) = \sum_{n=-\infty}^{\infty} \frac{1}{T} S\left(f - \frac{n}{T}\right)$$

(IV.2)



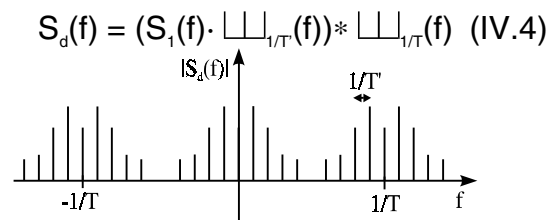
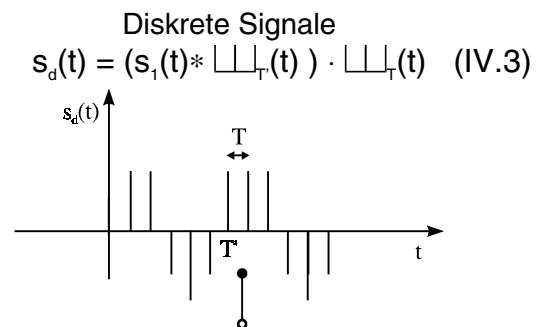
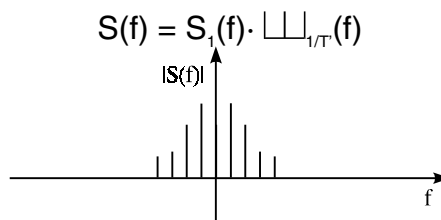
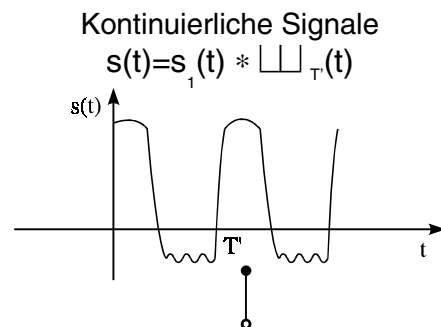
Spektrum aperiodisch, kontinuierlich



Spektrum periodisch, kontinuierlich

Für periodische Signale gilt: $s(t+nT') = s(t)$, wobei T' die Periodenlänge bezeichnet. Man kann ein periodisches Signal daher als die Faltung einer Periode dieses Signals mit einem δ -Kamm der Periodizität T' auffassen. Das Spektrum ist daher aperiodisch und diskret mit der Periodizität $1/T'$. Für das periodische diskrete Signal ergibt sich dagegen wiederum eine Multiplikation des periodischen Signals mit einem δ -Kamm der Periode T , so daß als Fouriertransformierte die Faltung resultiert:

Periodische Signale



Das Spektrum eines periodischen, diskreten Signals ist dabei wieder periodisch und diskret, wobei die Feinstruktur des Zeitsignals (d. h. die Abtastperiode) zur Großstruktur des Spektrums wird (d. h. zur Periodizität des Spektrums) während die Grobstruktur des Zeitsignals (d. h. die Periodizität des Zeitsignals) zur Feinstruktur des Spektrums wird (d. h. zur Linienstruktur des Spektrums). Dieser Fall ist für die digitale Signalverarbeitung aber besonders wichtig, weil sowohl das Zeitsignal als auch das Spektrum sich als diskrete Folge darstellen lassen. Die Beziehung zwischen **einer Periode** dieses diskreten, periodischen Zeitsignals und **einer Periode** des zugehörigen diskreten, periodischen Spektrums wird als **diskrete Fouriertransformation** bezeichnet. Für sie gilt:

$$s(n) \xrightarrow{\bullet} S(k) = \sum_{n=0}^{N-1} s(n) e^{-2\pi i \frac{nk}{N}}$$

$$S(k) \xrightarrow{\bullet} s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) e^{2\pi i \frac{nk}{N}} \quad (\text{IV.5})$$

Bei der diskreten Fouriertransformation wird also die Integration in eine Summation über eine Periode verwandelt. Da es sich auch hierbei um eine lineare, zeitinvariante Transformation handelt, sind die meisten Eigenschaften der Fouriertransformation auch für die diskrete Fouriertransformation die gleichen. Hin- und Rücktransformation der DFT unterscheiden sich nur durch das Vorzeichen des Exponenten. Die Normierung mit $1/N$ wird oft weggelassen. Ein wichtiger Unterschied zur kontinuierlichen Fouriertransformation ist jedoch die Faltung: Da die DFT nur als eine Periode des (unendlich langen, periodischen) Spektrums verstanden werden kann, gilt der Faltungssatz nicht für die exakte Faltung, sondern nur für die **zyklische Faltung**: Die diskrete Fourier-Rücktransformation des Produkts zweier diskreter Spektren ist nicht die Faltung der zugehörigen Zeitfunktionen, sondern der periodische fortgesetzten Zeitfunktionen, so daß sich u. U. das Ende der Periode in den Anfang des betrachteten Zeitbereichs hinein „verschmiert“.

Aufgrund der Definition der DFT benötigt die Berechnung eines Spektralwertes $S(k)$ n Multiplikationen und n Additionen, so daß der Rechenaufwand für eine vollständige diskrete Fouriertransformation proportional mit N^2 geht. Falls N eine Zweierpotenz ist, d. h. $N = 2^m$ läßt sich der Rechenaufwand durch die Anwendung der **Fast Fouriertransformation** (FFT) erheblich verringern. Dieses wird durch eine geschickte Zerlegung der für die Berechnung einer vollständigen DFT notwendigen Operationen erreicht, die sich in eine Abfolge von Elementar-Filterfunktionen mit zwei Eingängen und zwei Ausgängen (sogenannten Butterfly-Operationen) zerlegen lassen. Nachdem man $\log_2(N)$ Zwischenschritte berechnet hat, bei denen jeweils für jeden Wert eine weitere Butterfly-Operation durchgeführt wurde, können die Ausgangswerte in Bit-verkehrter Reihenfolge abgespeichert werden.

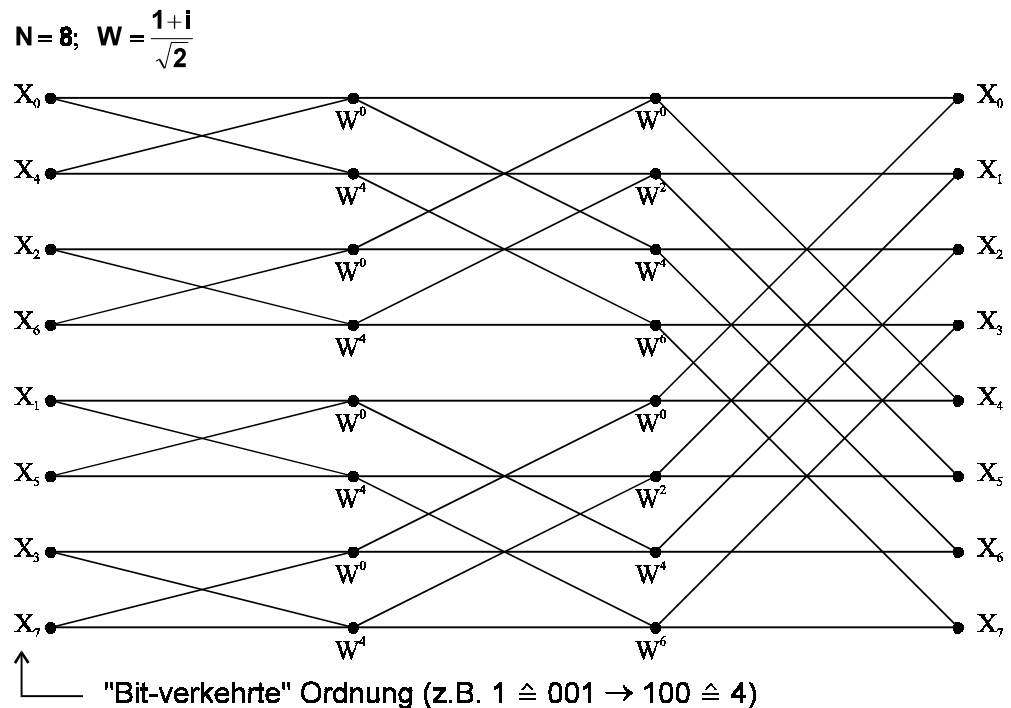


Abb. 4.1: Schema der FFT

In jedem dieser Teilschritte werden dabei N Multiplikationen und N Additionen durchgeführt, so daß der Gesamt-Rechenaufwand proportional zu $N \cdot \log_2 N$ wird. Da diese Größe mit zunehmenden N wesentlich geringer ansteigt als N^2 , ist gerade bei großen Signallängen eine Anwendung der FFT sinnvoll, so daß eine Spektralanalyse mit dem Computer fast immer eine FFT benutzt. Dies kann für Signale mit einer Anzahl von Samples, die nicht einer Zweierpotenz entspricht, z. B. durch Anfügen von Nullen erreicht werden. Außerdem existiert eine Zahl von zahlentheoretischen Fouriertransformationen, die für andere Primzahlen als für die 2 gelten, so daß auch bei Signallängen, die nicht einer Zweierpotenz, sondern einer günstigen Primzahlzerlegung entsprechen, effiziente Algorithmen für die Spektralanalyse eingesetzt werden können.

Im folgenden sollen zwei Probleme bei der diskreten Spektralanalyse näher behandelt werden, die von großer praktischer Bedeutung sind, und zwar die Fensterung und die zeitabhängige Spektralanalyse. Das Problem der **Fensterung** ergibt sich immer dann, wenn das Spektrum eines kontinuierlichen, aperiodischen Signals $s(t)$ mit Hilfe einer endlichen, diskreten Fouriertransformation geschätzt werden soll. Dazu wird das Signal mit einem Analyse-Zeitfenster zeitlich begrenzt und zudem äquidistant abgetastet:

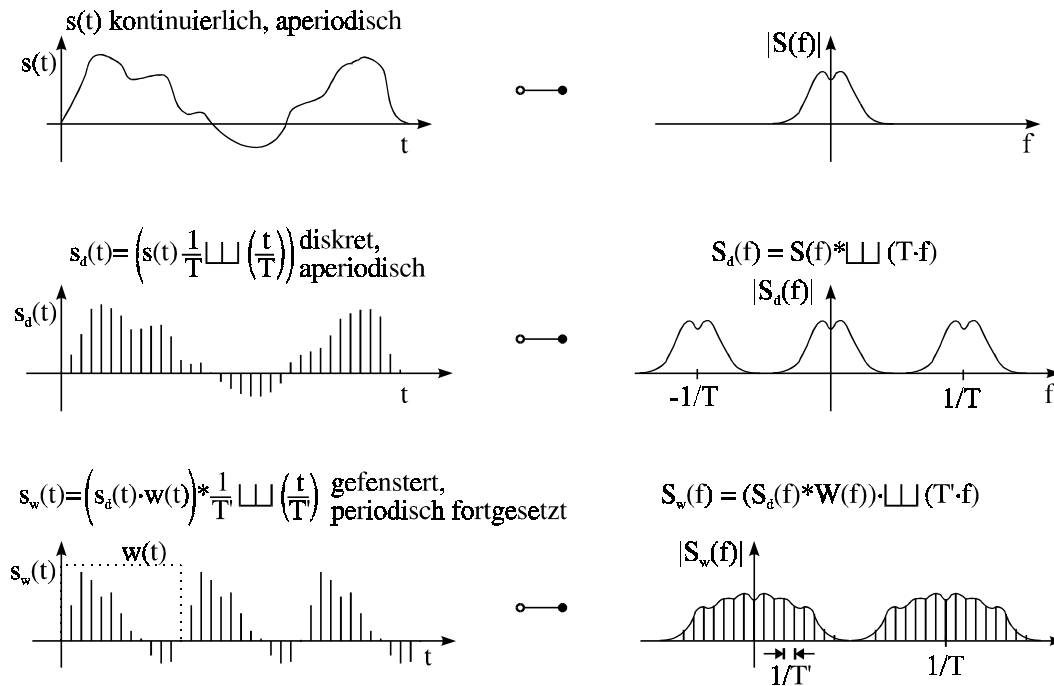


Abb. 4.2: Vom kontinuierlichen aperiodischen Signal zum diskreten, gefensterten Signal

Während das kontinuierliche, aperiodische Signal ein kontinuierliches, aperiodisches Spektrum $S(f)$ besitzt, mit der Zusatzannahme, daß $S(f)$ bandbegrenzt ist, besitzt das entsprechende abgetastete, aperiodische Signal $s_d(t)$ ein periodisches, kontinuierliches Spektrum. Bei der Fensterung des abgetasteten Signals, d.h. bei der Multiplikation mit dem Fenster $w(t)$, wird im Spektralbereich eine Faltung mit der Fouriertransformierten der Fensterfunktion $W(f)$ durchgeführt, so daß das resultierende Spektrum ein periodisches Linienspektrum ist, daß mit $W(f)$ gefaltet, also „verschmiert“ ist. Mit der diskreten Fouriertransformation (DFT) wird jeweils nur eine Periode dieses periodisch fortgesetzten, mit dem Fenster $w(t)$ abgeschnittenen und abgetasteten Signals $s_w(t)$ betrachtet. Das resultierende diskrete Spektrum ist demnach die diskrete Version des mit $W(f)$ verschmierten Originalspektrums. Um den dabei entstehenden Fehlerer möglichst klein zu halten, sollte ein Fenster $w(t)$ mit möglichst δ -förmiger Fouriertransformierten $W(f)$ gewählt werden. Dies kann durch ein möglichst langes Analyse-Fenster und durch eine optimierte Fensterfunktion approximativ erreicht werden. Am häufigsten werden die folgenden Fensterfunktionen gewählt:

$$\text{Hanning-Fenster:} \quad w_1(n) = 0,5 - 0,5 \cos \frac{2\pi n}{M-1} \quad (\text{IV.6})$$

$$\text{Hamming-Fenster:} \quad w_2(n) = 0,54 - 0,46 \cos \frac{2\pi n}{M-1} \quad (\text{IV.7})$$

$$\text{Blackman-Fenster: } w_3(n) = 0,42 - 0,5 \cos \frac{2\pi n}{M-1} + 0,08 \cos \frac{4\pi n}{M-1} \quad (\text{IV.8})$$

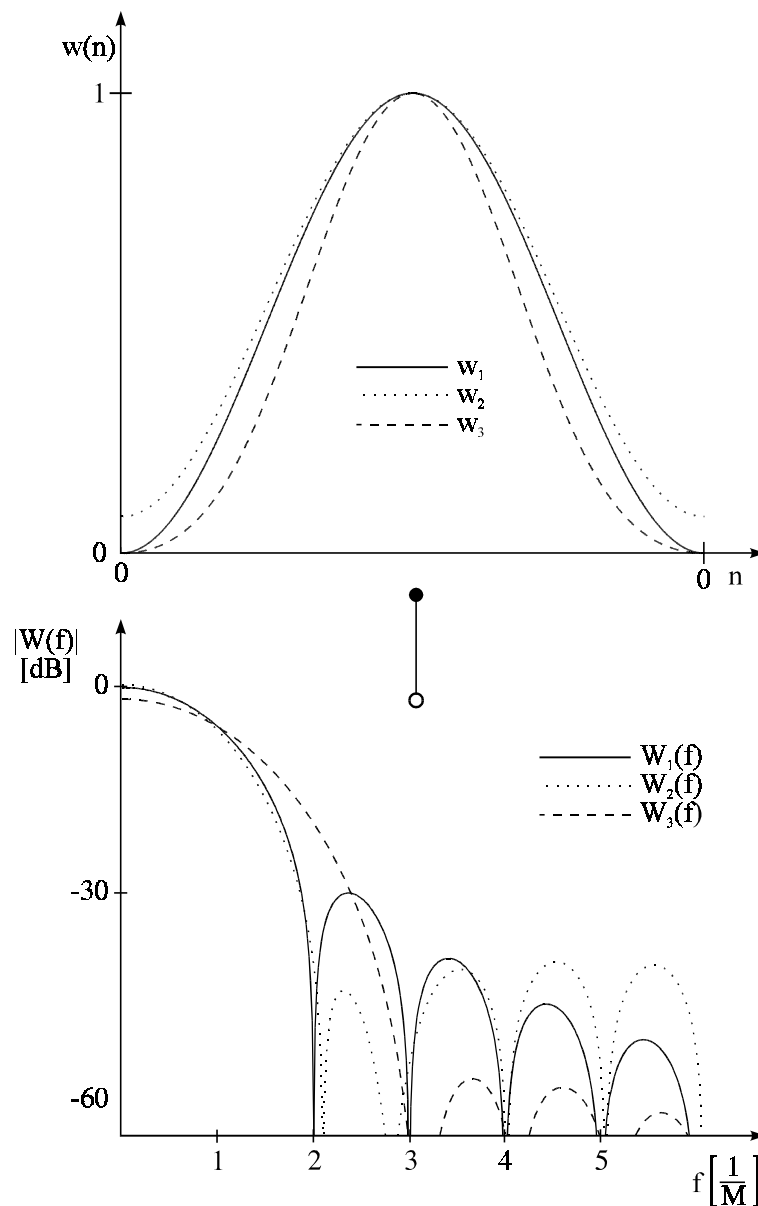


Abb.4.3: Die verschiedenen Fenster im Zeit- und Frequenzbereich

Um für die Spektralanalyse einen möglichst geringen Rechenaufwand durch Verwendung einer FFT zu benötigen, kann das Fenster mit Nullen „aufgefüllt“ werden, so daß die Gesamtfensterlänge gerade eine Zweierpotenz von Abtastwerten umfaßt. Durch dieses Hinzufügen von Nullen im Zeitbereich wird die Abtastung des verschmierten Spektrums im Frequenzbereich zwar verbessert, es wird jedoch keine zusätzliche Information gewonnen. Man kann daher eine Interpolation im Frequenzbereich dadurch berechnen, daß man eine entsprechend längere FFT wählt, für die man an das ursprüngliche Analyse-Fenster eine entsprechend größeren

Zahl von Nullen.

Das grundlegende Problem bei der zeitlichen Fensterung des zu analysierenden Zeitsignals ist die **Zeit-Frequenz-Unschärfe**:

Wenn eine hohe Frequenzauflösung gewünscht wird, muß $W(f)$ möglichst δ -förmig sein. Dies ist äquivalent damit, daß $w(t)$ möglichst lang ist, was wiederum mit einer geringen Zeitauflösung korrespondiert, d. h. das Spektrum wird über einen entsprechend langen Zeitausschnitt errechnet, so daß die auftretenden Spektralanteile nicht einem definierten Zeitpunkt zugeordnet werden können. Genau wie bereits in dem ersten Abschnitt besprochen, kann das Zeit-Bandbreitenprodukt aufgrund der Unschärfere-lation nicht unterschritten werden, so daß man sich entweder für eine hohe Zeitauflösung oder für eine hohe Frequenzauflösung entscheiden muß. Auch die im folgenden Abschnitt zu behandelnde Zeit-Frequenz-Analyse kann dieses prinzipielle Problem nicht lösen.

IV.2 Zeit-Frequenz-Analyse

Das Ziel einer gleichzeitig möglichst guten Zeit- und Frequenzauflösung wird mit der **diskreten Zeit-Frequenz-Analyse** angestrebt: Dabei soll der momentane „Frequenzgehalt“ eines Signals als Funktion der Zeit dargestellt werden. Es wird also eine Funktion $S(f, t)$ gesucht, die im diskreten Fall einem Raster von Werten in der Zeit-Frequenz-Ebene entspricht, d. h. es werden Funktionswerte nur für diskrete Zeitpunkt t_i und diskrete Frequenzen f_j berechnet. Die „klassische“ Zeit-Frequenz-Verteilungsfunktion ist die **Kurzzeit-Fourier-Analyse** (oder Spektrogramm) mit der allgemeinen Form für den analogen Fall

$$S(f, t) = \int_{-\infty}^{\infty} s(t'+t)w(t') \cdot e^{-2\pi i f t'} dt' \quad (\text{IV.9})$$

Dabei handelt es sich also um die Frequenzanalyse des um t zentrierten Signalausschnittes $s(t'+t)$, der mit dem Fenster $w(t')$ gewichtet wird.

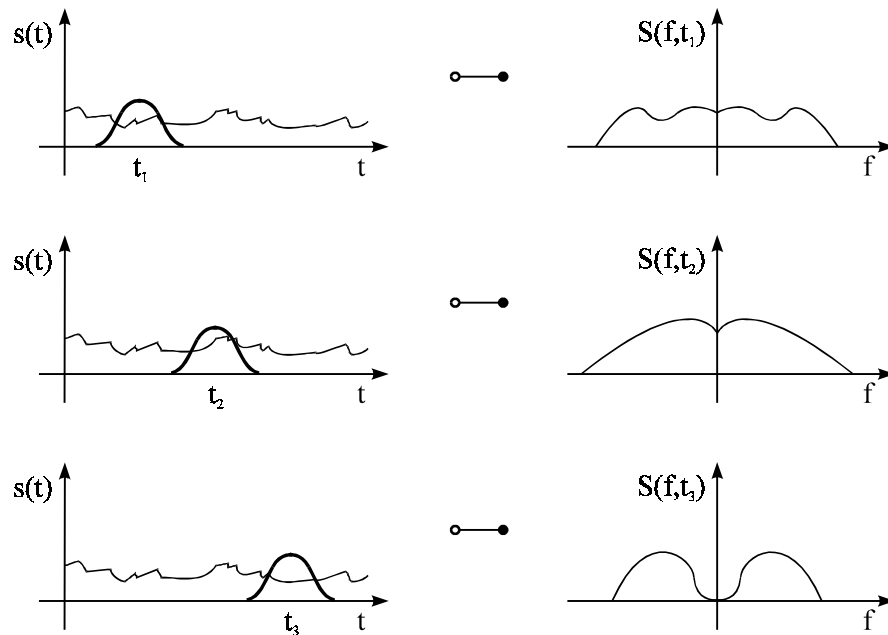


Abb.4.4: Zeitfrequenzanalyse

Im diskreten Fall wird ein Signalausschnitt zum Zeitpunkt t_m ausgewertet und die diskrete Fouriertransformation zu den Frequenzen f_k gebildet:

$$S(f_k, t_m) = \sum_{n=0}^{N-1} s(nT + t_m) w(n) e^{-j2\pi \frac{nk}{N}} \quad (\text{IV.10})$$

Bei dieser diskreten Kurzzeitfourieranalyse treten die folgenden Variablen auf:

- 1.) Die Abtastfrequenz $f_s = 1/T$, die bei der Diskretisierung von $s(t)$ in die diskrete Zahlenfolge $s(n)$ auftritt.
- 2.) Die Fensterlänge M (in Anzahl von Samples) der Fensterfunktion $w(n)$
- 3.) Länge N der diskreten Fouriertransformation (DFT)
- 4.) Abstand aufeinanderfolgender Analyse-Zeitpunkte t_m , d. h. $1/\text{„Framerate“ } F$.

Diese vier Variablen sollten so gewählt werden, daß das Originalsignal $s(t)$ aus dem diskreten Spektrogramm vollständig rekonstruiert werden kann. Es darf also bei der Berechnung des Spektrogramms bei entsprechender Wahl der Parameter kein Informationsverlust auftreten. Aus diesen For-

derungen resultieren die folgenden Abtasttheoreme:

1.) Abtasttheorem:

$$f_s = \frac{1}{T} \geq 2f_{\text{grenz}} \quad (\text{IV.11})$$

D. h. die höchste im Originalsignal vorkommende Frequenz f_{grenz} darf maximal die Hälfte der Abtastfrequenz betragen. Dieses ist das bereits behandelte Abtasttheorem der digitalen Signalverarbeitung.

2.) Die Länge des Analysefensters M ist umgekehrt proportional zur Zeitauflösung der Kurzzeitfrequenzanalyse und ist proportional zur Frequenzauflösung der Kurzzeitfrequenzanalyse, d. h. die Fensterlänge ist ein freier Parameter der Kurzzeitfrequenzanalyse, die je nach Anforderung entweder eine hohe Zeit- und niedrige Frequenzauflösung oder umgekehrt aufweist. Bei Sprachsignalen ist eine gängige Fensterlänge etwa 20 ms.

3.) $N \geq$ Fensterlänge M , d. h. die Länge der DFT muß mindestens so groß sein wie die Länge des Analysefensters. In der Regel wird die DFT sogar länger als das Analyse-Fenster gewählt, um einerseits eine bessere Abtastung des Spektrums zu erreichen und um andererseits mögliche Artefakte durch „zirkuläres Aliasing“ zu vermeiden, das bei der Durchführung von Modifikationen im Frequenzbereich und anschließender Rekonstruktion des Signals auftreten kann: Anhand der Kurzzeitspektralanalyse ist es nämlich möglich, gezielt zu bestimmten Zeitpunkten eine Änderung des Spektrums (Filterung) durchzuführen. Für das modifizierte Spektrum gilt dann:

$$S'(f_k, t_m) = H(f_n) \cdot S(f_k, t_m)$$
$$s'(n + t_m) = h(n) * (s(n + t_m) \cdot w(n)) \quad (\text{IV.12})$$

D. h., daß bei der Rücktransformation des im Spektralbereich veränderten Signals in den Zeitbereich eine Faltung des gefensterten Originalsignals mit der Impulsantwort $h(n)$ auftritt, so daß sich der Bereich des Signals, in dem keine Nullen auftreten, um die Länge der Impulsantwort $h(n)$ verlängert. Damit sich diese „Verschmierung“ des ursprünglich durch das Zeitfenster zeitlich begrenzten Signals nicht über die Ränder hinweg erstreckt (und damit ein „Überfließen“ vom Ende des DFT-Zeitsignals in den Anfang

dieses Signals passiert) muß also die Länge der DFT N genügend groß sein. Dieses **zweite Abtasttheorem** besagt also, daß

$$N \geq \text{Fensterlänge } M + \text{Länge von } h(n) \quad (\text{IV.13})$$

Bei der Hintransformation müssen demnach eine genügend große Anzahl von Nullen an das gefensterterte Signal angefügt werden, um die entsprechende DFT-Länge zu erreichen.

4.) Das dritte Abtasttheorem der Kurzzeitfrequenzanalyse beinhaltet eine Forderung für die Analyse-Frame-Rate. Die Kurzzeit-Fouriertransformierte $S(f, t)$ kann auch als Filterbank aufgefaßt werden, bei der in einer Reihe von Bandpaßkanälen jeweils ein Ausgangssignal erzeugt wird, das sich zeitlich nur sehr langsam ändert (bandbegrenzte Signale):

$$S(f, t) = \int_{-\infty}^{\infty} s(t'+t) w(t') e^{-2\pi i f t'} dt' = e^{-2\pi i f t} \cdot \int_{-\infty}^{\infty} s(t') \cdot e^{-2\pi i f t'} \cdot w(t'-t) dt' \quad (\text{IV.14})$$

$S(f_k, t_m)$ ist daher für ein festes f_k als Funktion von t_m eine Tiefpaßsignal. Die obere Grenzfrequenz dieses Signals f_{grenz} entspricht bei einem Hamming-Fenster $w(t)$ gerade der ersten Nullstelle der Übertragungsfunktion (bei $\frac{2}{M}$) und beträgt damit:

$$f_{\text{grenz}} \cong \frac{2}{M} \cdot f_s \quad (\text{IV.15})$$

Da ein derartiges Tiefpaßsignal mit mindestens der doppeltem vorkommenden Frequenz abgetastet werden muß, folgt daraus, daß die Frame-Rate (d. h. die Häufigkeit, mit der dieses Tiefpaßsignal abgetastet werden muß) gerade den Wert annehmen muß:

$$F = 4 \cdot \frac{f_s}{M} \quad (\text{IV.16})$$

Dieses entspricht einem Überlapp aufeinanderfolgender Analyse-Frames von $3/4$. Diese Forderung, daß zur vollständigen Repräsentation des Eingangssignals ein Überlapp der aufeinanderfolgenden Analyse Intervalle von $3/4$ auftreten sollte, wird als **drittes Abtasttheorem** bezeichnet. Falls ein größerer Fehler toleriert werden kann und als obere Grenzfrequenz des Bandpaß-Signals nicht die erste Nullstelle des Hamming-Fensters,

sondern bereits eine Abschwächung um ca. 10 dB ausreichend ist, genügt der halbe Überlapp aufeinanderfolgender Analyse-Frames.

Im folgenden sollen zwei prinzipielle Möglichkeiten zur Rekonstruktion des Nutzzeit-Fourier analysierten Signals diskutiert werden, die beide das Original-Signal perfekt rekonstruieren können, sich aber in ihrer Konzeption unterscheiden. Zur genaueren Abhandlung sei auf den Artikel von J. B. Allen (1977) verwiesen.

Der erste Ansatz ist die **Filterbank-Summation**: Dabei wird $S(f_k, t_m)$ für ein festes t_m als der Ausgang einer Filterbank aufgefaßt:

$$\begin{aligned}
 s(t_m) &= \frac{1}{N} \sum_{k=0}^{N-1} S(f_k, t_m) \cdot e^{\frac{2\pi i k m}{N}} \\
 &= \frac{1}{N} \sum_{n=0}^{N-1} s(n + t_m) \cdot w(n) \cdot \underbrace{\sum_{k=0}^{N-1} e^{\frac{2\pi i k(m-n)}{N}}}_{=\delta(n-m) \cdot N}
 \end{aligned} \quad (\text{IV.17})$$

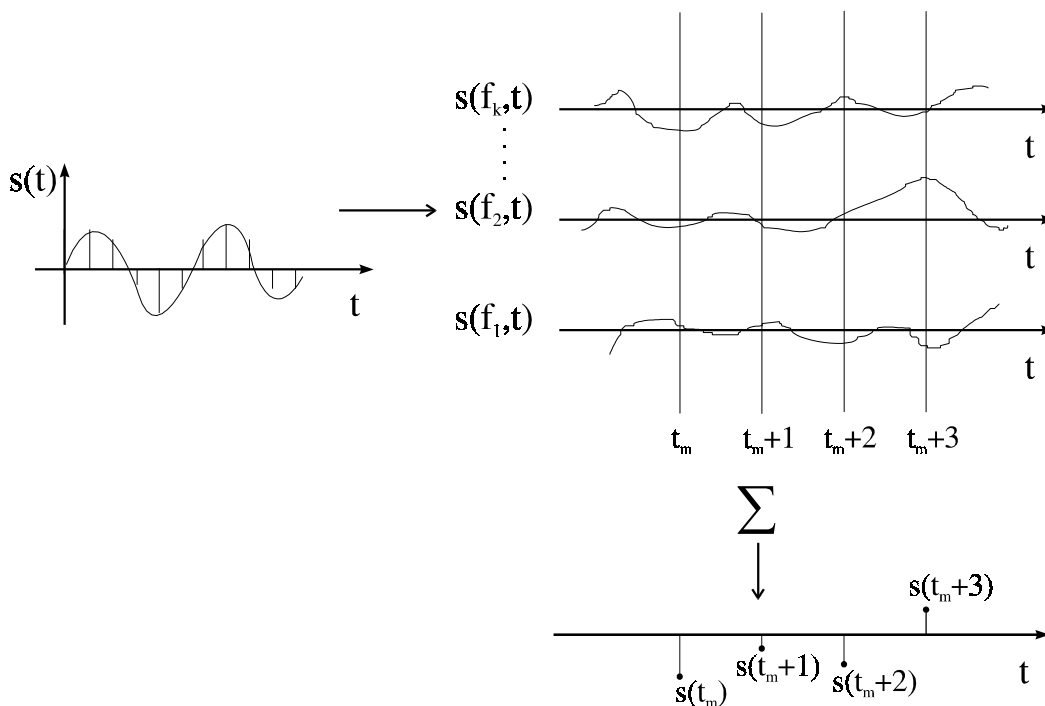


Abb.4.5: Filterbank-Summation

Der Vorteil dieser Filterbank-Summationstechnik, ist die Summation über wenige, u. U. nicht äquidistante Frequenzen, die zu einer vollständigen Rekonstruktion des Signals führen. Eine derartige nicht-äquidistante Aufteilung ist z. B. für die gehöradäquate Frequenzanalyse oder die Wavelet-

Analyse notwendig. Außerdem wirken sich spektrale Modifikationen sofort auf das rekonstruierte Signal aus, da die entsprechenden Bandpaßkanäle instantan abgeschwächt bzw. verstärkt werden. Zu den Nachteilen einer derartigen Technik gehört, daß für jeden rekonstruierten Wert $s(t_m)$ das gesamte Spektrum $S(f_k, t_m)$ vorliegen muß. Da das rekonstruierte Signal mit der Original-Abtastfrequenz wieder vorliegen soll, bedeutet dies, daß zu jedem Abtastzeitpunkt das gesamte Spektrum neu berechnet werden muß, so daß insgesamt der Rechenaufwand sehr hoch wird. Ein weiterer Nachteil ist, daß die zu der spektralen Modifikation $H(f_k, t_m)$ gehörende Impulsantwort $h(t_m)$ mit der Fensterfunktion gewichtet wird:

$$\hat{S}(f_k, t_m) = H(f_k, t_m) \cdot S(f_k, t_m) \quad \bullet \quad \hat{s}(t_m) = s(n) * [h(t_m) \cdot w(n)] \quad (\text{IV.18})$$

Dies bedeutet, daß nicht eine beliebige spektrale Modifikation durchgeführt werden kann, sondern nur eine spektrale Modifikation, deren zugehörige Impulsantwort in das Analyse-Fenster $w(n)$ „paßt“.

Die zweite Rekonstruktionsmethode ist das **Overlapp-Add-Verfahren** (OLA): Bei diesem Verfahren wird zu jedem Rekonstruktionszeitpunkt das gesamte Zeitfenster (und nicht nur ein Signalwert) aus der inversen DFT rekonstruiert. Die so entstandenen rekonstruierten Zeitfenster werden sukzessive überlappend aufaddiert:

$$\begin{aligned} s(n + t_m) \cdot w(n) &= \frac{1}{N} \sum_{k=0}^{N-1} S(f_k, t_m) e^{2\pi i \frac{n \cdot k}{N}} \\ s(n) &= \sum_{t_m} s(n + t_m) \cdot w(n) = \sum_{t_m} s(n) w(n - t_m) \\ &= s(n) \cdot \sum_{t_m} w(n - t_m) = s(n) \cdot \underset{\substack{\uparrow \\ \text{Gleichanteil} \\ \text{von } w(n)}}{W(0)} \end{aligned} \quad (\text{IV.19})$$

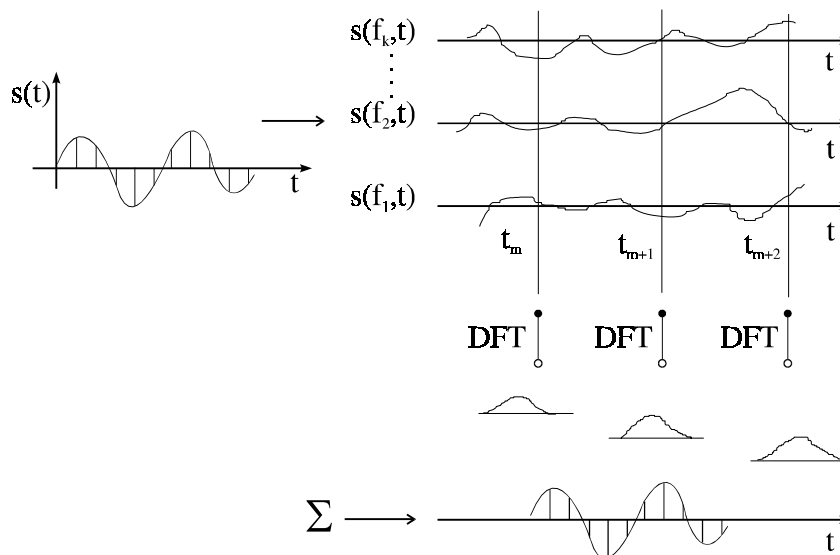


Abb.4.6: Overlapp-Add-Verfahren

Der Vorteil dieser Rekonstruktionsmethode ist, daß eine Zurücktransformation nur an wenigen Zeitpunkten t_m erforderlich ist, so daß der Rechenaufwand sinkt. Bei dem Überlapp aufeinanderfolgender Analyse-Frames von $\frac{3}{4}$ erhöht sich damit die Datenrate um den Faktor 4, während sie bei den oben diskutierten Filterbank-Summationsverfahren um den Faktor N (d. h. Länge der DFT) sich erhöhte. Ein weiterer Vorteil ist, daß keine Fensterung der zu einer spektralen Modifikation $H(f_k)$ gehörenden Impulsantwort $h(t)$ auftritt, sofern nicht die o. a. zirkulären Aliasing-Artefakte auftreten (d. h. die Länge der DFT ausreichend ist):

$$\begin{aligned}\hat{s}(n) &= \sum_{t_m} \frac{1}{N} \sum_{k=0}^{N-1} s(f_k, t_m) \cdot H(f_k) e^{2\pi i \frac{n \cdot k}{N}} \\ &= \sum_{t_m} w(n - t_m) \cdot [s(n) * h(n)] = [s(n) * h(n)] \cdot W(0)\end{aligned}\quad (\text{IV.20})$$

$\hat{s}(n)$ bezeichnet dabei das modifizierte Zeitsignal, das als eine exakte (zyklische) Faltung des Original-Signals mit der „gewünschten“ Impulsantwort resultiert.

Zu den Nachteilen dieser Rekonstruktionsmethode zählt, daß die exakt äquidistanten Frequenzen der inversen DFT für die Rekonstruktion erforderlich sind. Außerdem wirken sich zeitlich veränderliche spektrale Modifikationen nicht instantan auf das rekonstruierte Zeitsignal aus, sondern wirken sich auf mehrere Frames verteilt aus, so daß die spektralen Änderungen nur zeitlich tiefpaßgefiltert sich auf das rekonstruierte Zeitsignal auswirken. Wenn man an einer sofortigen Umsetzung der gewünschten

Filterfunktion in das rekonstruierte Zeitsignal interessiert ist, sollte man daher der Filterbank-Summationstechnik den Vorteil geben. Wenn andererseits eine geringe Datenrate bzw. ein geringerer Rechenaufwand im Vordergrund steht, ist das Overlap-Add-Verfahren im Vorteil. Dieses Verfahren wird auch in einer Reihe von Anwendungen vom digitalen Rundfunk über die Codierung von Sprache und Musiksignalen bis hin zur Realisierung moderner, „intelligenter“ digitaler Hörgeräte eingesetzt (vgl. Kollmeier, 1990).

IV. 3 Digitale Filter

Während wir im vorherigen Abschnitt die spektrale Analyse eines digitalisierten Zeitsignals und die mögliche Modifikation dieser Zeit-Frequenzdarstellung und ihrer Rekonstruktionsmöglichkeiten besprochen haben, wollen wir im folgenden die spektrale Modifikation eines beliebig langen Eingangssignals mit einem festen Filter behandeln, das digital realisiert wird. Ein derartiges digitales Filter operiert auf einer Folge $x(n)$ von diskreten Eingangswerten, wobei n sämtliche ganzen Zahlen überstreichen kann. Die Elementaroperationen eines digitalen Filters sind dabei:

- 1.) Multiplikation mit einer Konstanten $\alpha \in \mathbb{R}$: $x(n) \rightarrow \alpha \cdot x(n)$
- 2.) Die Addition zweier Signale: $x(n), y(n) \rightarrow \alpha x(n) + \beta y(n), \alpha, \beta \in \mathbb{R}$
- 3.) Die Zeitverzögerung des Signals um ein ganzzahliges Vielfaches der Abtastperiode: $x(n) \rightarrow x(n-1)$

Diese Elementar-Operationen IE sind sämtlich **linear**, d. h.

$$IE(\lambda x(n) + \mu y(n)) = \lambda IE(x(n)) + \mu IE(y(n)) \quad (\text{IV.21})$$

und **verschiebungsinvariant**, d. h.

$$y(n) = IE(x(n)) \Rightarrow y(n+k) = IE(x(n+k)) \quad (\text{IV.22})$$

Aus der beliebigen Kombination dieser Elementaroperationen läßt sich damit ein lineares, Zeit- bzw. Verschiebungs-invariantes System aufbauen, das wir als digitales Filter bezeichnen:

Wenn wir als Eingangsfolge die Werte $x(n)$ und als Ausgangsfolge die

Werte $y(n)$ bezeichnen, so wird das digitale Filter durch die sogenannte **Differenzengleichung** definiert:

$$y(n) = \sum_{k=0}^M \alpha_k x(n-k) + \sum_{k=1}^N \beta_k y(n-k) \quad (\text{IV.23})$$

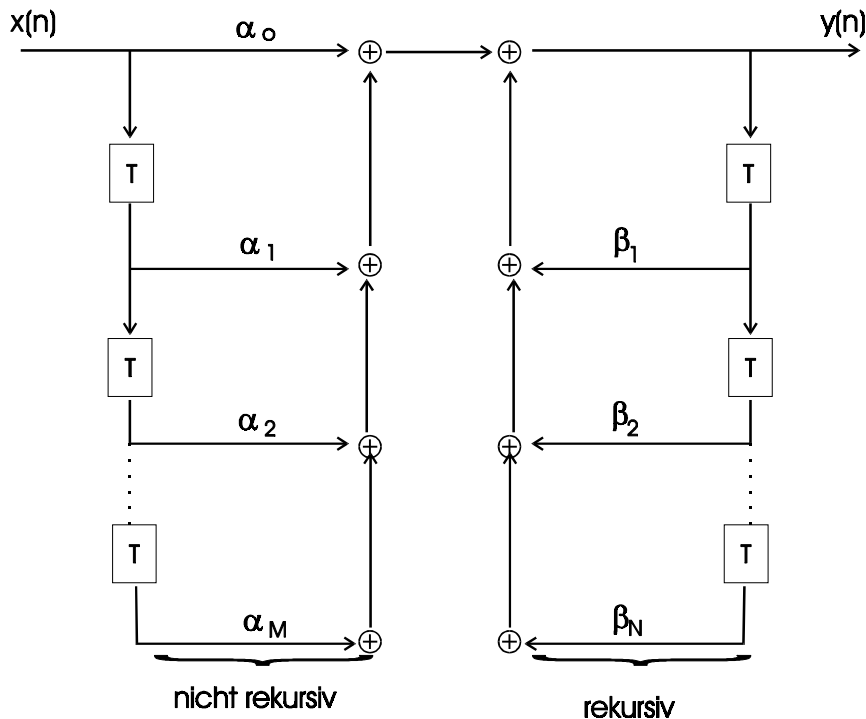


Abb. 4.7: Schema des digitalen Filters

Ein digitales Filter wird damit aus einem nichtrekursiven Anteil und einem rekursiven Anteil zusammengesetzt, wobei in letzterem der Ausgangswert wieder an den Eingang geführt wird und im ersteren Fall der Eingangswert nach Zeitverzögerung, Additionen und Multiplikationen an den Ausgang geführt wird. Um die Eigenschaften eines derartigen digitalen Filters näher beleuchten zu können, betrachten wir zunächst die Impulsantwort. Sie ist definiert als die Folge von Ausgangswerten, die sich bei Eingabe einer δ -Impulsfolge auf den Eingang ergibt.

Eingangssignal:
$$\delta(n) = \begin{cases} 1 & \text{für } n = 0 \\ 0 & \text{sonst} \end{cases}$$

$$\text{Ausgangssignal: } h(n) = \begin{cases} 0 & \text{für } n < 0 \\ \alpha_0 & \text{für } n = 0 \\ \alpha_n + \sum_{k=1}^{\min(n,N)} \beta_k h(n-k) & \text{für } n > 0 \end{cases} \quad (\text{IV.24})$$

Die Impulsantwort ist daher eindeutig durch die Koeffizienten α_k und β_k bestimmt, so daß durch diese Koeffizienten auch sämtliche Eigenschaften des Filters festgelegt werden.

Für den Fall, daß alle $\beta_k = 0$ sind, vereinfacht sich die Impulsantwort zu:

$$\beta_k = 0 \text{ für alle } k \Rightarrow h(n) = \alpha_n \text{ für } n \geq 0 \quad (\text{IV.25})$$

Dieser Fall einer endlich andauernden Impulsantwort wird als **Finite Impulse Response (FIR)-Filter** oder als **nichtrekursives Filter** bezeichnet.

Für den Fall, daß mindestens ein $\beta_k \neq 0$ ist, ergibt sich die Impulsantwort $h(n)$ zu

$$\beta_k \neq 0 \text{ für mind. ein } k \Rightarrow h(n) = \alpha_n + \sum_{k=1}^{\min(n,N)} \beta_k h(n-k), n \geq 0 \quad (\text{IV.26})$$

d. h. die Impulsantwort dauert unendlich an, so daß man dieses Filter als **Infinite Impulse Response (IIR)** oder **rekursives Filter** bezeichnet.

Als Ordnung des Filters bezeichnet man dabei die maximale Verzögerung im Filter, d. h. das Maximum aus M und N . Je höher diese Ordnung ist, d. h. je größer der Aufwand ist, den man für die Realisation des Filters betreiben muß, desto stärker ist die Auswirkung des Filters auf den Frequenzbereich.

Durch Fouriertransformation der Differenzengleichung (IV.23), bei der die Glieder einzeln fouriertransformiert werden können, ergibt sich aufgrund der Linearität und der Verschiebungseigenschaft der Fouriertransformation für die Fouriertransformierten $X(f)$ des Eingangssignals bzw. $Y(f)$ des Ausgangssignals die folgenden Beziehungen:

$$X(f) = \sum_{k=0}^M \alpha_k X(f) \cdot (e^{-2\pi i f T})^k + \sum_{k=1}^N \beta_k Y(f) \cdot (e^{-2\pi i f T})^k \quad (\text{IV.27})$$

Beim Zusammenfassen aller Glieder in X und in Y sowie bei Division durch X(f) (unter der Voraussetzung, daß $X(f) \neq 0$ für alle betrachteten Frequenzen) ergibt sich damit als Übertragungsfunktion H(f) des Filters

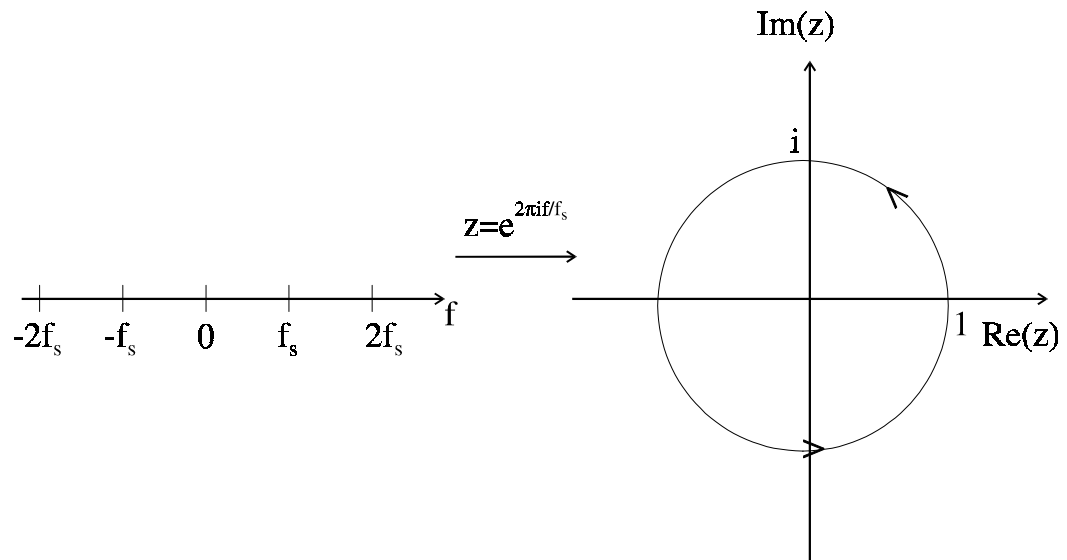
$$H(f) = \frac{Y(f)}{X(f)} = \frac{\sum_{k=0}^M \alpha_k (e^{-2\pi i f T})^k}{1 - \sum_{k=1}^N \beta_k (e^{-2\pi i f T})^k} \quad (\text{IV.28})$$

Diese Übertragungsfunktion H(f) hängt nur von $e^{2\pi i f T}$ ab, d. h., daß nur das Produkt $f \cdot T = f/f_s$ relevant ist. Sämtliche vorkommenden Frequenzen sind damit auf die Abtastfrequenz f_s normiert, so daß in der digitalen Signalverarbeitung die „normierte“ Frequenz von 0 bis 0,5 (d. h. halbe Abtastfrequenz läuft). Außerdem ist die Übertragungsfunktion periodisch mit f_s . Diese Eigenschaft folgt aus der diskreten Struktur der Impulsantwort $h(n)$, deren Fouriertransformierte gerade die Übertragungsfunktion sein muß. Um die Übertragungsfunktion H(f) nun in einfacherer Form schreiben zu können, führen wir eine Variablen-Transformation durch, die (im allgemeinen Fall) als **z-Transformation** bezeichnet wird:

$$z = e^{2\pi i f \cdot T} \quad (\text{IV. 29})$$

Für die Übertragungsfunktion ergibt sich dann:

$$H(f) = H(e^{2\pi i f T}) = H(z) = \frac{\sum_{k=0}^M \alpha_k z^{-k}}{1 - \sum_{k=1}^N \beta_k z^{-k}} \quad (\text{IV.30})$$

Abb. 4.8: Übertragungsfunktion, periodisch mit f_s

Wenn f sämtliche Werte der reellen Frequenzachse von $-\infty \rightarrow \infty$ durchläuft, durchläuft z zyklisch den komplexen Einheitskreis. Die Variablentransformation $z = e^{2\pi i f T}$ ist damit eine konforme (d. h. lokal winkelerhaltende) Abbildung, die die reelle Frequenzachse auf den (komplexen) Einheitskreis transformiert. Die Verallgemeinerung dieser Variablentransformation auf die gesamte komplexe Ebene stellt dann die (weiter unten behandelte) allgemeine z -Transformation dar.

Als Beispiel betrachten wir zunächst den Fall eines einfachen nichtrekursiven Filters zweiter Ordnung mit der Impulsantwort:

$$\alpha_0 = \frac{1}{2}, \alpha_1 = 1, \alpha_2 = \frac{1}{2}, \beta_1 = \beta_2 = 0 \quad (\text{IV.31})$$

Als Differenzengleichung ergibt sich:

$$y(n) = \frac{1}{2}x(n) + x(n-1) + \frac{1}{2}x(n-2) \quad (\text{IV.32})$$

Die Übertragungsfunktion als Funktion von f sowie als Funktion von z ergibt sich damit zu:

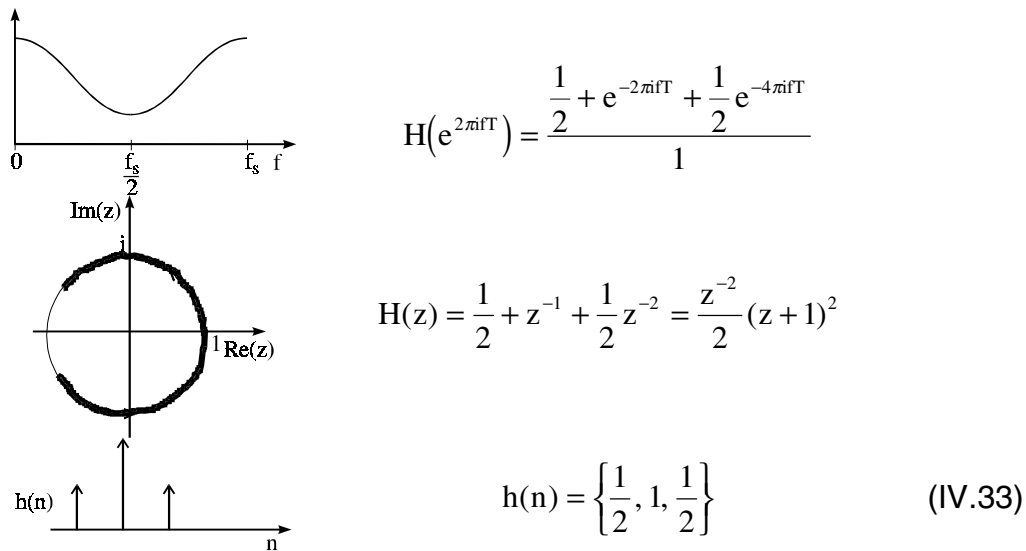


Abb. 4.9: Übertragungsfunktionen und Impulsantwort des nicht rekursiven Filters

Als zweites Beispiel behandeln wir ein rekursives Filter erster Ordnung mit den Filterkoeffizienten $\alpha_0 = 1$, $\alpha_1 = \alpha_2 = \alpha_k = 0$, $\beta_1 = 0,9$, $\beta_2 = \beta_k = 0$.

Als Impulsantwort errechnet sich:

$$h(n) = (0,9)^n, n \geq 0 \quad (\text{IV.34})$$

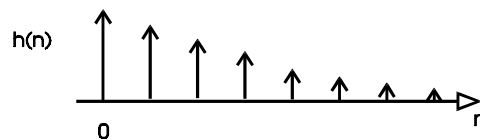
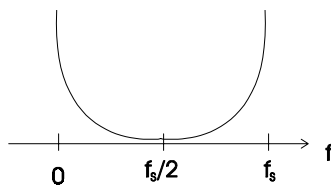
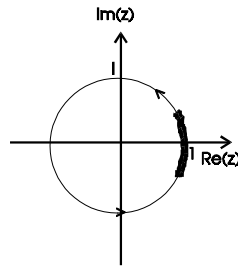


Abb. 4.10: Impulsantwort

Als Übertragungsfunktion $H(f)$ und Übertragungsfunktion $H(z)$ ergibt sich dann:



$$H(f) = \frac{1}{1 - 0,9 \cdot e^{-j2\pi fT}}$$



$$H(z) = \frac{1}{1 - 0,9 \cdot z^{-1}} \quad (\text{IV.35})$$

Abb. 4.11: Übertragungsfunktion von f und z

Anhand dieser Beispiele wird bereits deutlich, daß die z-Transformierte die einfachste und am direktesten aus der Impulsantwort folgende Beschreibung eines digitalen Filters ist. Diese leichte Handhabbarkeit ist insbesondere durch die diskrete Natur des behandelten linearen Systems bedingt. Für den allgemeinen Fall ergibt sich die folgende Definition der **z-Hintransformation**, die für ein beliebiges $z \in \mathcal{D}$ gilt, sofern $H(z)$ existiert:

$$h(n) \rightarrow H(z) = \sum_{n=-\infty}^{\infty} h(n)z^{-n} \quad (\text{IV.36})$$

Um zu zeigen, daß diese Definition konsistent ist mit der oben durchgeführten Fouriertransformation mit anschließender Variablensubstitution, drücken wir $h(n)$ wie folgt aus:

$$\begin{aligned} h(n) &= h(t) \cdot \text{III}\left(\frac{n}{T}\right) = \sum_{n=-\infty}^{\infty} h(t) \delta(t - nT) \\ H(f) &= \int_{-\infty}^{\infty} h(n) e^{-2\pi i f t} dt = \int_{-\infty}^{\infty} \left(\sum_{n=-\infty}^{\infty} h(t) \delta(t - nT) \right) \cdot e^{-2\pi i f t} dt \\ &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} h(t) \delta(t - nT) e^{-2\pi i f t} dt \\ &= \sum_{n=-\infty}^{\infty} h(n) e^{-2\pi i f T n} \\ &= \sum_{n=-\infty}^{\infty} h(n) z^{-n} \end{aligned} \quad (\text{IV.37})$$

Bei der **z-Rücktransformation** ergibt sich das Problem, daß die z-Transformierte auf einer gesamten Ebene (der **komplexen Ebene**) definiert ist (sofern sie in diesem Gebiet existiert), so daß für die Rücktransformation ein Linienintegral in der komplexen z-Ebene angegeben werden muß. Dieses Linienintegral kann z. B. als Integral über den gesamten

Einheitskreis, d. h. $|z| = 1$ ausgeführt werden:

$$H(z) \rightarrow h(n) = \frac{1}{2\pi i} \oint_{|z|=1} H(z) z^{n-1} dz \quad (\text{IV.38})$$

Um diese Definition plausibel zu machen, können wir uns zum einen überlegen, daß der Integrationsweg bei der inversen Fouriertransformation, der über sämtliche Frequenzen von $-\infty$ bis $+\infty$ geht, für den Fall der z-Rücktransformation über den gesamten Einheitskreis gehen müßte. Andererseits können wir uns auch die z-Rücktransformation für ein Signal $h(n)$ ansehen, deren z-Hintransformation wir kennen:

$$\begin{aligned} \frac{1}{2\pi i} \oint H(z) z^{n-1} dz &= \sum_{k=-\infty}^{\infty} \frac{1}{2\pi i} \oint h(k) z^{n-k-1} dz \\ &= \sum_{k=-\infty}^{\infty} \frac{1}{2\pi i} \cdot h(k) \cdot \delta(n-k) \cdot 2\pi i, \\ \text{da } \oint z^{n-k-1} dz &= \begin{cases} 0 & \text{für } n \neq k \\ 2\pi i & \text{für } n = k \end{cases} \end{aligned} \quad (\text{IV.39})$$

Dabei haben wir die Aussage des Cauchy'schen Integralsatzes bzw. Residuen-Satzes benutzt, daß das Integral über eine geschlossene Kurve gerade den Wert annimmt, den eine Reihenentwicklung von $H(z)$ bei dem Glied $1/z$ annimmt. Aus dieser Betrachtung folgt bereits die unmittelbare Anwendung der Sätze aus der Funktionentheorie für die digitale Signalverarbeitung. Weitere Eigenschaften der z-Transformationen finden sich bei Oppenheim/Schafer, (1975) Kapitel 2.

Die oben erhaltene Darstellung der Übertragungsfunktion als Polynom M-ten Grades im Zähler und Polynom N-ten Grades im Nenner läßt sich auch dazu ausnutzen, zu einer anschaulichen Interpretation der z-Transformierten zu gelangen. Nach dem Hauptsatz der Algebra lassen sich Zähler und Nenner-Polynom in Linearfaktoren zerlegen, wobei die Linearfaktoren im Zähler der Form $z - z_{ok}$ sind (z_{ok} bezeichnet die **Nullstellen** des Zählerpolynoms) bzw. der Form $z - z_{pi}$ (z_{pi} bezeichnen dabei die **Pole**, d.h. die Nullstellen des Nenner-Polynoms). Es ergibt sich damit die folgende Produktdarstellung (**Pol-Nullstellenzerlegung der Übertragungsfunktion**)

$$H(z) = \frac{\alpha_0 \prod_{k=1}^M (1 - z_{ok} z^{-1})}{\prod_{i=1}^N (1 - z_{pi} z^{-1})} \quad (\text{IV.40})$$

Für den Betrag von $H(z)$ ergibt sich damit aufgrund der Eigenschaft, daß der Betrag eines Produktes das Produkt der Beträge ist, die folgende Zerlegung:

$$|H(z)| = \frac{\prod \text{Abstände zwischen } z \text{ und den Nullstellen}}{\prod \text{Abstände zwischen } z \text{ und den Polen}} \quad (\text{IV.41})$$

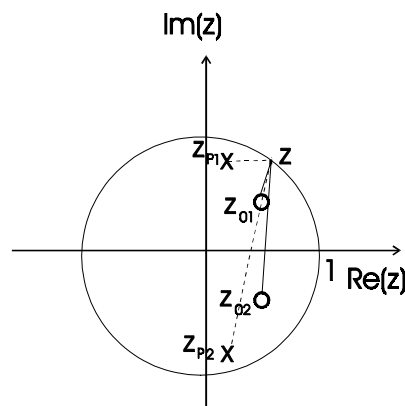


Abb. 4.12: Geometrische Interpretation des Betrags von $H(z)$

Als geometrische Interpretation dieser Formel ergibt sich der Betrag der Übertragungsfunktion bei einem festen, vorgegebenen Wert von z auf dem Einheitskreis als das Produkt der Abstände zwischen diesem Wert zu sämtlichen Nullstellen geteilt durch das Produkt aus den Abständen dieses Punktes zu sämtlichen Polen. Wenn eine Nullstelle dicht beim Einheitskreis liegt, wird der Betrag der Übertragungsfunktion sehr klein, während der Betrag sehr große Werte annimmt, wenn ein Pol sehr dicht an dem Einheitskreis liegt. Eine ähnliche Betrachtung läßt sich für die Phase der Übertragungsfunktion anstellen, die sich als die Summe über die Winkel der Nullstellenabstände minus der Summe über die Winkel der Polabstände interpretieren läßt. Mit dieser **Pol-Nullstellenzerlegung** der Übertragungsfunktion läßt sich daher aus der Verteilung der Pole und Nullstellen in der komplexen z -Ebene die Eigenschaften der Übertragungsfunktionen qualitativ abschätzen. Umgekehrt lassen sich durch Vorgabe von Polen und Nullstellen die Eigenschaften der Übertragungsfunktion gut vorhersagen, so daß für das **Filterdesign** zumeist die Lage der Pole und Nullstellen und nicht die Werte der Koeffizienten in der Differenzgleichung vorgegeben werden. Für die Vorgabe von Nullstellen z_{ok}

und Polen z_{pk} gelten allerdings die folgenden Einschränkungen:

- 1) Die Nullstellen z_{ok} bzw. die Pole z_{pi} dürfen entweder reell sein oder nur in konjugiert-komplexen Paaren auftreten. Diese Einschränkung folgt aus der Bedingung, daß die Koeffizienten α_k und β_k reell sein müssen, damit es sich um einen als reellen Filter realisierbares digitales Filter handelt.
- 2) $|z_{pk}| < 1$, d. h. Pole dürfen nur innerhalb des Einheitskreises liegen. Diese Bedingung ist notwendig, damit der Filter stabil ist. Wenn ein Pol außerhalb des Einheitskreises liegt, existiert zwar eine z-Transformierte, als z-Rücktransformation ergibt sich aber eine exponentiell ansteigende Impulsantwort, die damit einem nicht-stabilen Filter entspricht.

Als Beispiel dazu betrachten wir das obige rekursive Filter erster Ordnung mit der Übertragungsfunktion

$$H(z) = \frac{1}{1 - z_p z^{-1}}, \quad z_p \text{ reell, } z_p > 1$$

$$h(n) = z_p^n \rightarrow \infty \quad (\text{IV.42})$$

d. h. die Impulsantwort divergiert (instabiles System).

Als Pol-Nullstellendarstellung für die oben angeführten Beispiele eines nicht-rekursiven Filters zweiter Ordnung bzw. rekursiven Filters erster Ordnung ergeben sich folgende Zerlegungen:

a)
$$y(n) = \frac{1}{2}x(n) + x(n-1) + \frac{1}{2}x(n-2)$$

$$H(z) = \frac{1}{2} + z^{-1} + \frac{1}{2}z^{-2} = \frac{1}{2}((-1)z^{-1} - 1)((-1)z^{-1} - 1) \quad (\text{IV.43})$$

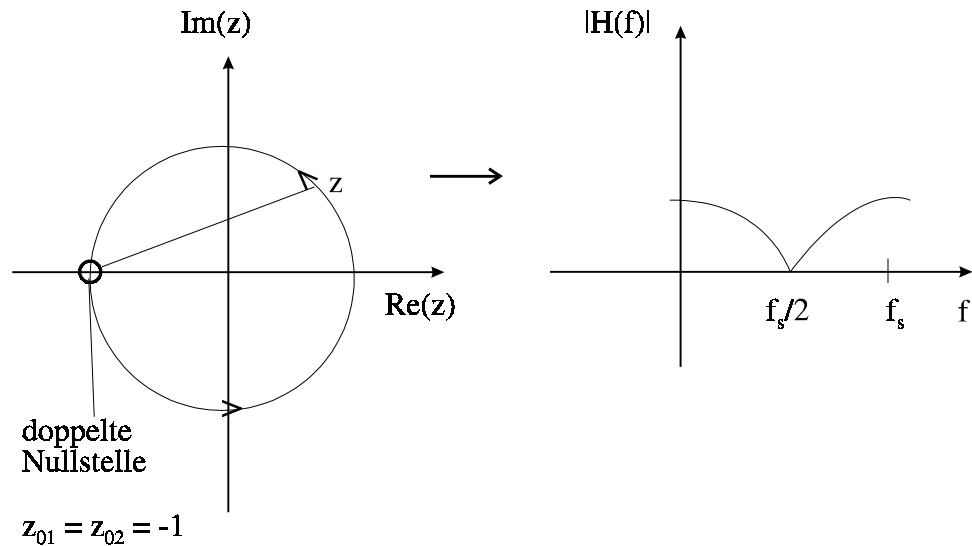


Abb. 4.13: Geometrische Interpretation und Übertragungsfunktion des 1. Beispiels

Dieses nichtrekursive Filter hat somit eine doppelte Nullstelle bei -1 und agiert damit als Tiefpaßfilter, bei dem die halbe Abtastfrequenz genau auf 0 abgeschwächt wird.

b)
$$y(n) = x(n) + 0,9 \cdot y(n-1)$$

$$H(z) = \frac{1}{1 - 0,9z^{-1}}, z_{p1} = 0,9 \quad (\text{IV.44})$$

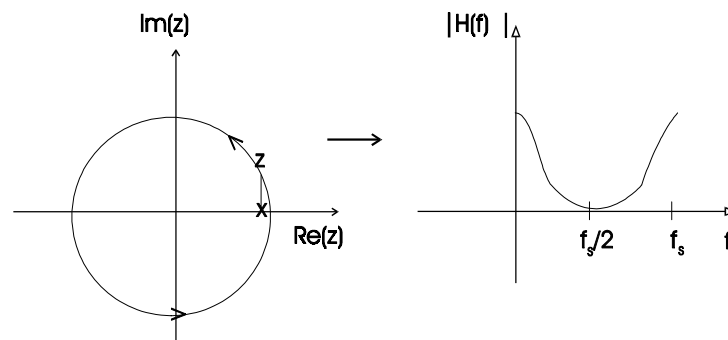


Abb. 4.14: 2. Beispiel

Dieses rekursive Filter erster Ordnung wirkt ebenfalls als Tiefpaßfilter, wobei die Übertragungsfunktion bei der Frequenz 0 um so spitzer und schärfer ausfällt, je dichter der Pol an den Einheitskreis rückt.

IV. 4 Entwurf („Design“) digitaler Filter

Während wir bisher nur den Fall vorgegebener Pole und Nullstellen und der daraus resultierenden Übertragungsfunktion bzw. Impulsantwort betrachtet haben, soll im folgenden das umgekehrte Problem betrachtet werden, in dem eine „gewünschte“ Übertragungsfunktion vorgegeben wird und die Lage der Nullstellen bzw. Pole so gewählt werden soll, daß ein digitales Filter realisiert wird, das diesen gewünschten Eigenschaften möglichst nahe kommt. Die Vorgabe eines entsprechenden Filters ist in der Praxis immer dann wichtig, wenn die digital abgespeicherten Daten auf eine bestimmte Art gefiltert werden sollen. Die gewünschte Filterfunktion wird dabei innerhalb von gewissen Toleranzgrenzen vorgegeben.

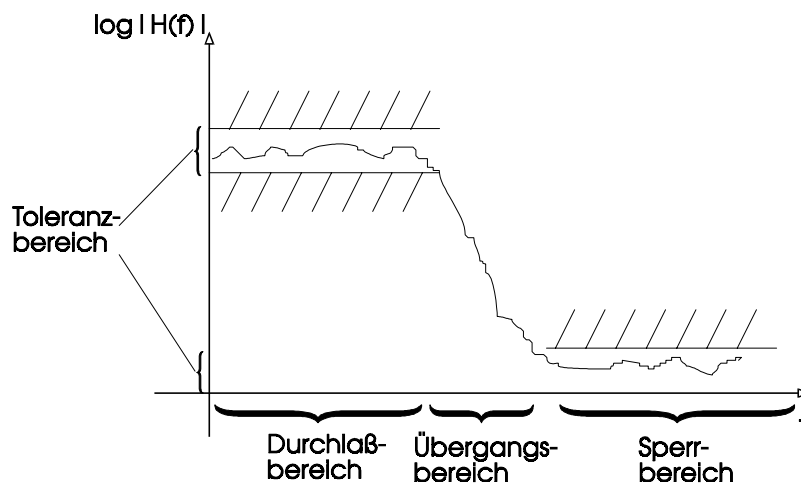


Abb. 4.15: Toleranzbereich der Übertragungsfunktion

Im allgemeinen wird der Betrag der Übertragungsfunktion im Durchlaßbereich und im Sperrbereich mit jeweils einer gewissen Toleranzgrenze vorgegeben, sowie der Frequenzbereich des Übergangs zwischen diesen beiden Bereichen. Die grundlegende Frage des Filterdesigns ist es nun, wieviele Nullstellen und wieviele Pole man wo hinlegt, um diese Filterfunktion zu erreichen. Dabei ist folgende Vorgehensweise angebracht:

1.) Soll es sich um ein **IIR** (Infinite Impulse Response) oder ein **FIR** (Finite Impulse Response) Filter handeln?

Ein IIR Filter hat die Vorteile eines geringen Implementierungsaufwandes bei hoher Filterwirkung und die direkte Übertragung „klassischer“ Filterdesigns (z. B. Butterworth, Tschebycheff, Bessel-Filter etc.). Dagegen besitzt ein IIR-Filter den Nachteil, gegebenenfalls numerisch instabil zu sein, insbesondere wenn die Pole zu dicht am Einheitskreis liegen. Aufgrund von Rechenungenauigkeiten bei den Filterkoeffizienten und den zu verarbeitenden digitalen Daten können instabile (d. h. exponentiell anwachsende) Impulsantworten auftreten oder sogenannte Grenzyklen entstehen (d. h.

periodisch durchlaufene Zustände des rekursiven Filters). Außerdem weisen IIR-Filter eine **Dispersion** auf, d. h. die einlaufenden Signale werden je nach ihrer Frequenz unterschiedlich lang durch den Filter verzögert. Dadurch können impulsartige Eingangssignale am Ausgang stark über den Zeitbereich verschmiert erscheinen. Obwohl mit einigen rekursiven Filtertypen (Bessel-Filtern) die Dispersion im Durchlaßbereich minimiert wurde, ist dieser Filtertyp nicht zu verwenden, wenn eine hohe Impulstreue (d. h. eine möglichst geringe Verfälschung der Zeitfunktion des Signals) gewünscht wird.

Ein **FIR-Filter** besitzt dagegen die Vorteile, numerisch stabil zu sein und bei symmetrischer Impulsantwort phasenlinear zu sein, d. h. es tritt keine Dispersion auf (Beweis siehe unten). Für diese Art von Filter sind analytische Lösungen mit geringem Designaufwand ausrechenbar (Windowing-Design mit Approximation „idealer“ Übertragungsfunktionen) und es existieren auch numerische Optimierungsalgorithmen für das direkte FIR-Filterdesign (z. B. Remez-Exchange-Algorithmus, vgl. Oppenheim/Schafer, 1975). Der wichtigste Nachteil von FIR-Filtern ist dagegen der relativ hohe Implementierungsaufwand, der für eine vorgegebene Filterwirkung auftritt. Im allgemeinen ist die Filterwirkung, die durch einen Pol erreicht wird, wesentlich stärker als die Filterwirkung, die durch eine Nullstelle bewirkt wird.

Um den oben angekündigten Beweis der **Phasenlinearität** eines FIR-Filters mit symmetrischer Impulsantwort zu erbringen, betrachten wir den Fall, daß

$$h(n) = h(N - n), \text{ für } 0 \leq n \leq \frac{N}{2} \quad (\text{IV.45})$$

Dabei gehen wir davon aus, daß N ungerade ist.

Als z-Transformierte ergibt sich davon:

$$\begin{aligned}
H(z) &= \sum_{n=0}^N h(n)z^{-n} = \sum_{n=0}^{N/2} h(n)[z^{-n} + z^{-(N-n)}] \\
&= z^{-\frac{N}{2}} \cdot \sum_{n=0}^{N/2} h(n) \left[z^{\frac{N}{2}-n} + z^{-\left(\frac{N}{2}-n\right)} \right] \\
&= z^{-\frac{N}{2}} \cdot \sum_{n=0}^{N/2} h(n) \cdot 2 \operatorname{Re} \left(z^{\frac{N}{2}-n} \right), \text{ da } h(n) \text{ reell ist} \\
&= z^{-\frac{N}{2}} \cdot |H(z)| \tag{IV.46}
\end{aligned}$$

(Für den Fall, daß N gerade ist, muß der Term mit $h\left(\frac{N}{2}\right)$ separat behandelt werden, an der prinzipiellen Rechnung ändert sich sonst nichts).

Die Übertragungsfunktion läßt sich demnach aufspalten in einen reinen reellen Faktor, der den Betrag der Übertragungsfunktion kennzeichnet und einen Phasenfaktor $z^{-N/2}$, der einer frequenzunabhängigen Verzögerung mit der halben Länge der Impulsantwort entspricht. Damit wäre der Beweis einer ausbleibenden Dispersion bei symmetrischer Impulsantwort erbracht.

2.) Nachdem der Typ des Filters festgelegt wurde, ist die **Ordnung des Filters** (d. h. die Anzahl der Nullstellen bzw. Pole) festzulegen. Generell gilt, daß die Approximation an eine „ideale“ Übertragungsfunktion umso besser wird, je höher die Ordnung des Filters ist. Da andererseits der Implementierungsaufwand mit zunehmender Ordnung zunimmt, muß in der Praxis immer ein Kompromiß zwischen beiden Anforderungen gefunden werden. Dieser Kompromiß hängt stark von den jeweiligen Randbedingungen ab (d. h. inwiefern der Implementierungsaufwand ein härteres Kriterium als die Güte der Übertragungsfunktion darstellt).

3.) Schließlich muß die Methode ausgewählt werden, mit der das gesuchte Filter entworfen werden soll. Dazu gibt es zum einen die Möglichkeit einer **numerischen, computerbasierten Optimierung** der Koeffizienten eines digitalen Filters, die vorwiegend für FIR-Filter durchgeführt wird, und andererseits die Übertragung eines analogen Filterdesigns in ein digitales Filter. Bei dieser letzteren Methode gibt es zum einen das **Impulsantwort-invariante Design**, bei der die Impulsantwort des analogen Filters in ein digitales Filter übernommen wird (dies ist vorwiegend beim Design von FIR-Filtern der Fall) oder ein **Übertragungsfunktion-invariantes Design**, bei dem die Übertragungsfunktion des analogen Filters (im Frequenzbereich von $0 \rightarrow \infty$) in die entsprechende Übertragungsfunktion des digitalen

Filters (von 0 bis zur halben Abtastfrequenz) übertragen wird. Diese Design-Methode wird vorwiegend bei IIR-Filtern angewandt.

Beim Impulsantwort-invarianten Design von FIR-Filtern, das auch als **Windowing-Design** bezeichnet werden kann, wird die „ideale“ Impulsantwort der „gewünschten“ analogen Übertragungsfunktion ausgerechnet. Falls diese „ideale“ Übertragungsfunktion beispielsweise eine rechteckförmige Tiefpaßcharakteristik darstellt, wäre die zugehörige Impulsantwort eine sinc-Funktion. Die Idee besteht nun darin, diese analoge, „ideale“ Impulsantwort zum einen zu diskretisieren und zum anderen mit einer Fensterfunktion zu multiplizieren, um aus der unendlichen Impulsantwort eine realisierbare, endliche Impulsantwort zu machen.

$$h(t) \rightarrow h(n) = h(t) \cdot \text{II}_T(t) \quad (\text{IV.47})$$

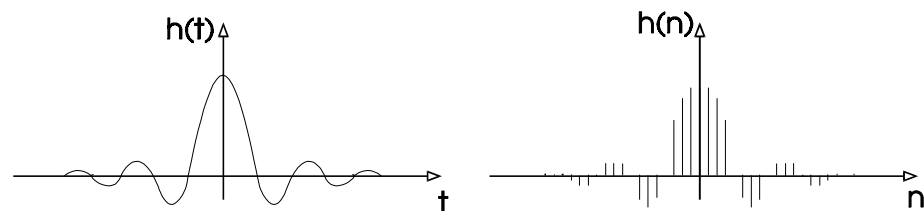


Abb. 4.16: Ideale Impulsantwort (z.B. sinc)

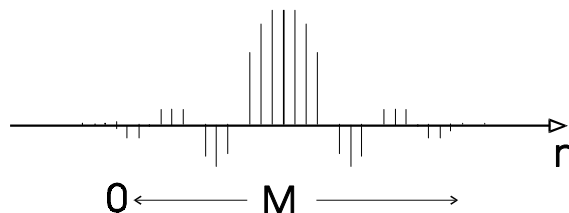


Abb. 4.17: Gefensterte Impulsantwort

Wenn in das vorgegebene Zeitfenster M Samples der diskretisierten Impulsantwort hineinpassen, läßt sich $h(t)$ durch ein nichtkursives FIR-Filter approximieren, für dessen Koeffizienten gelten:

$$\alpha_i = h(i) \text{ für } 0 \leq i \leq M \quad (\text{IV.48})$$

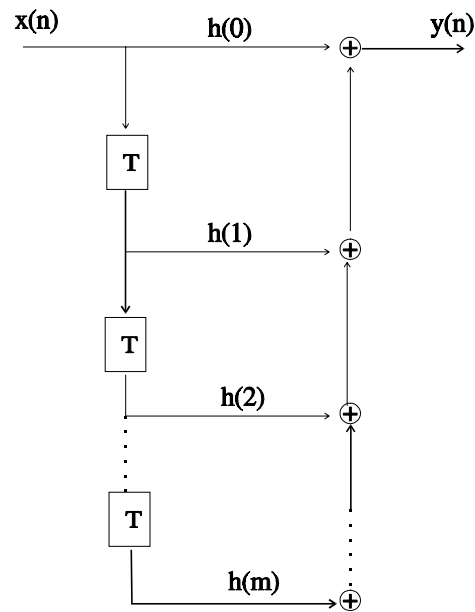


Abb. 4.18: Struktur des FIR-Filters

Um die resultierende Übertragungsfunktion des FIR-Filters zu berechnen, stellen wir zunächst die diskretisierte, abgeschnittene Impulsantwort $h_d(n)$ als Multiplikation mit einer Kammfunktion und einem Fenster sowie einer zeitlichen Verschiebung um $MT/2$ dar (d. h. Verschiebung um die halbe Fensterlänge):

$$h_d(n) = \left(h(t) \cdot \text{Kamm}(t) \cdot w\left(\frac{t}{MT}\right) \right) * \delta\left(t - \frac{MT}{2}\right)$$

$$H_d(f) = [(H(f) * \text{Kamm}(f \cdot T)) * W(f \cdot M \cdot T)] \cdot z^{-\frac{MT}{2}} \quad (\text{IV.49})$$

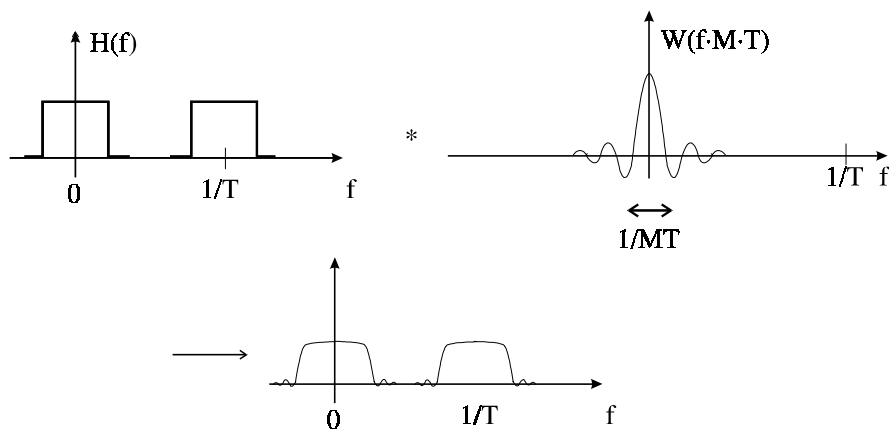


Abb. 4.19: Übertragungsfunktion, verschmiert mit FFT der Fensterfunktion

Die resultierende Übertragungsfunktion des digitalen Filters ist zum einen (als Konsequenz der zeitlichen Abtastung) eine periodische Überlagerung der ursprünglichen Übertragungsfunktion. Außerdem ist die „ideale“ Übertragungsfunktion mit der Fouriertransformierten des Fensters „verschmiert“, so daß als resultierende Übertragungsfunktion nicht notwendigerweise eine glatte, die ideale Übertragungsfunktion immer optimal approximierende Funktion herauskommt. Stattdessen können bei der Approximation Überraschungen und Interpolations-Produkte auftreten (z. B. durch das Gibb'sche Phänomen). Um diese Abweichungen zu vermeiden, sollte ein möglichst langes Fenster mit einer möglichst δ -förmigen Fouriertransformierten verwendet werden. Dazu bieten sich die bereits in der Frequenzanalyse vorgestellten Hamming- oder Hanning-Fenster an. Außerdem können die Vorgaben für die „ideale“ Übertragungsfunktion leicht modifiziert werden, um eine möglichst gute Approximation an den „gewünschten“ Verlauf der Übertragungsfunktion zu erhalten. Nähere Einzelheiten finden sich z. B. bei Oppenheim/Schafer (1975) oder bei Lacroix (1980).

Beim **Frequenzgangs-invarianten Filterdesign**, das für die Konstruktion von IIR-Filtern angewendet wird, besteht die grundlegende Idee darin, die auf der Laplace s -Ebene optimierte Lage von Polen und Nullstellen eines analogen digitalen Filters so auf die z -Ebene zu übertragen, daß die gesamte Frequenzachse der Laplace-Transformation eineindeutig auf den Einheitskreis der z -Transformation abgebildet wird. Dies kann so geschehen, daß bei tiefen Frequenzen annäherungsweise eine 1:1 Abbildung zwischen dem analogen und diskreten Fall auftritt, während bei hohen Frequenzen die Abbildung auf den Einheitskreis immer stärker komprimierend wirkt, so daß die Frequenzen $\pm \infty$ bei -1 auf dem Einheitskreis abgebildet werden.

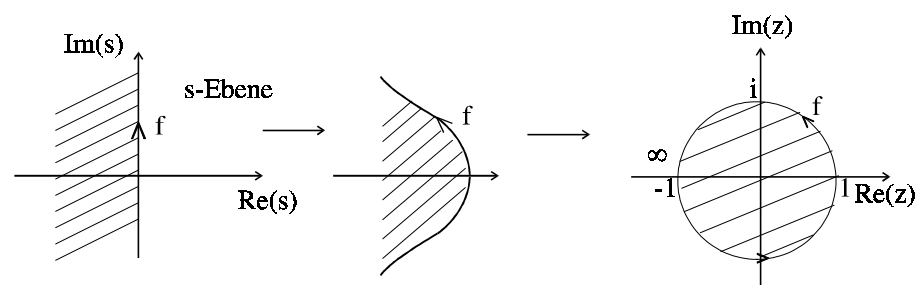


Abb. 4.20: Gesamte Frequenzachse, eineindeutig auf Einheitskreis abgebildet

Dabei wird die linke s -Halbebene in das Innere des Einheitskreises der z -Ebene abgebildet und die Imaginärteil-Achse der Laplace-Ebene wird auf den Rand des Einheitskreises in der z -Ebene transformiert. Die Punkte $\pm i \infty$ der Laplace-Ebene werden auf die Punkte $z = -1$ übertragen. Dem-

entsprechend wird der Frequenzgang eines analogen Filters in den Frequenzgang eines digitalen Filters so übertragen, daß die Übertragungsfunktion zu höheren Frequenzen stark „zusammengedrängt“ wird.

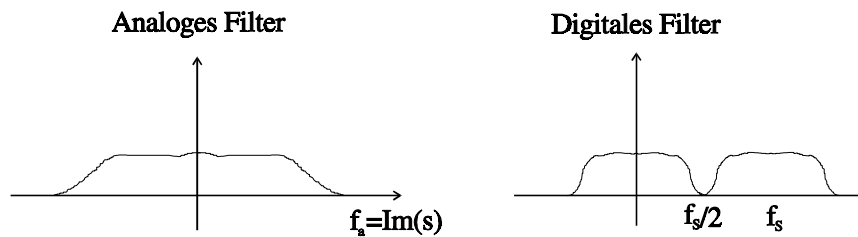


Abb. 4.21: Frequenzgang des analogen und digitalen Filters

Um eine derartige Abbildung zwischen der Laplace-s-Ebene und der z-Ebene der z-Transformation zu realisieren, bedient man sich der sogenannten bilinearen Transformation:

$$s = \frac{2}{T} \frac{1 - z^{-1}}{1 + z^{-1}} \Leftrightarrow z = \frac{1 + \left(\frac{T}{2}\right)s}{1 - \left(\frac{T}{2}\right)s} \quad (\text{IV.50})$$

Diese bilineare Transformation bewirkt eine Frequenz-Verzerrung (Frequency-Warping) bei der die Zuordnung zwischen der ursprünglichen Frequenz des analogen Filters f_a und der neuen Frequenz des digitalen Filters f_d , folgendermaßen realisiert ist:

$$\frac{2\pi T}{2} f_a = \tan\left(\frac{2\pi \cdot f_d}{2}\right) \quad (\text{IV.51})$$

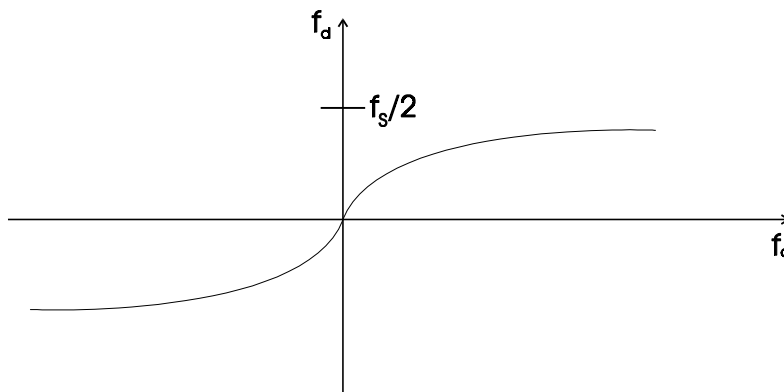


Abb. 4.22: Frequenzverzerrung („Warping“)

Für sehr kleine Werte von s ist diese Abbildung approximativ gleich 1.

Als Beispiel rechnen wir den Fall eines Butterworth-Tiefpaß-Filters zweiter Ordnung (Grenzfrequenz ω_0) durch.

$$\begin{aligned}
 H(s) &= \frac{1}{1 + \frac{\sqrt{2}}{\omega_0} s + \frac{1}{\omega_0^2} s^2}, \quad s \rightarrow \frac{2}{T} \frac{1 - z^{-1}}{1 + z^{-1}} \\
 \Rightarrow H(z) &= \frac{1}{1 + \frac{\sqrt{2}}{\omega_0} \frac{2}{T} \left(\frac{1 - z^{-1}}{1 + z^{-1}} \right) + \frac{1}{\omega_0^2} \frac{2}{T} \left(\frac{1 - z^{-1}}{1 + z^{-1}} \right)^2} \\
 &= \frac{(1 + z^{-1})^2}{\left(1 + \frac{2\sqrt{2}}{\omega_0 T} + \left(\frac{2}{\omega_0 T} \right)^2 \right) - \left(2 \left(\frac{2}{\omega_0 T} \right)^2 - 2 \right) z^{-1} - \left(\frac{2\sqrt{2}}{\omega_0 T} - \left(\frac{2}{\omega_0 T} \right)^2 - 1 \right) z^{-2}} \quad (\text{IV.52})
 \end{aligned}$$

Aus den zwei Polen des analogen Butterworth-Tiefpasses zweiter Ordnung werden daher zwei Pole und zwei Nullstellen (d. h. eine doppelte Nullstelle bei $z = -1$) für den digitalen Fall. Weiterhin muß man beachten, daß die ursprüngliche Grenzfrequenz des analogen Filters nicht mehr für den digitalen Filter gilt. Stattdessen gilt für die gewünschte Grenzfrequenz des digitalen Filters:

$$\begin{aligned}
 f_{d0} &= \frac{1}{8} f_s = \frac{1}{8 \cdot T} \\
 \Rightarrow \omega_0 &= 2\pi \cdot f_{a0} = \frac{2}{T} \tan \frac{2\pi T}{2} \cdot f_{d0} = 2\pi \cdot \frac{1}{T} \cdot \frac{1}{7.58} \quad (\text{IV.53})
 \end{aligned}$$

ω_0 liegt wegen der Frequenzverzerrung daher etwas höher als $2\pi \cdot f_{d0}$.

Als Übertragungsfunktion $H(z)$ ergibt sich damit:

$$H(z) = \frac{1}{10,24} \frac{(1 + z^{-1})^2}{1 - 0,9428z^{-1} + 0,333z^{-2}} \quad (\text{IV.54})$$

Die Lage der Polstellen $z_{p01/2}$ sowie die Koeffizienten α_0 bis α_2 errechnen sich demnach zu:

$$Z_{p01/2} = 0,4714 \pm 0,333i$$

$$\alpha_0 = 0,64 \cdot \frac{1}{16}, \quad \alpha_1 = 1,28 \cdot \frac{1}{16}, \quad \alpha_2 = 0,64 \cdot \frac{1}{16} \quad (\text{IV.55})$$

4.) Nachdem nun auf verschiedene Arten und Weisen digitale Filter aus analogen Filter-Design-Prinzipien berechnet wurden, soll noch erwähnt werden, daß digitale Filter aus digitalen Prototyp-Filtern errechnet werden können, indem spezielle Tiefpaß-Hochpaß- oder Tiefpaß-Bandpaß-Transformationen verwendet werden. Die grundlegende Idee ist dabei die Abbildung des Einheitskreises in der z -Ebene auf sich selbst, z. B. für den Fall einer Tiefpaß-Hochpaß-Transformation mit der Änderung $z \rightarrow -z$.

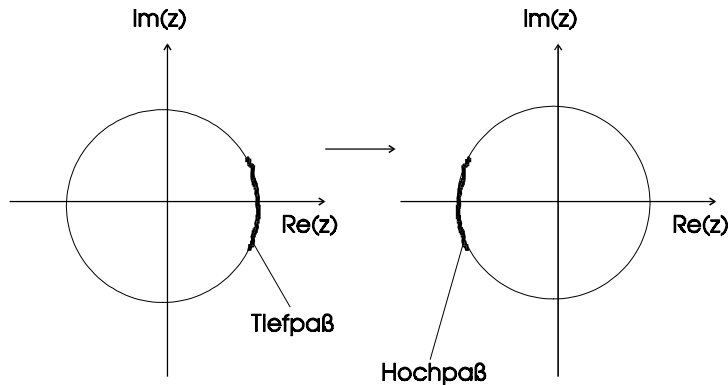


Abb. 4.23: Hoch- Tiefpaß-Transformation

D. h. aus der Tiefpaß-Übertragungsfunktion mit einem Maximum des Betrags in der Nähe von Null wird eine Hochpaß-Übertragungsfunktion, bei der das entsprechende Maximum bei $z = -1$ auftritt. Eine große Klasse von weiteren Transformationen, mit denen ein digitales Filter in ein anderes digitales Filter umgerechnet werden kann, findet sich in Tabelle 2.1.

Tabelle 4.1: Transformationen eines Tiefpaß-Digital-Filter-Prototyps mit der Grenzfrequenz θ_p

Filter-Typ	Transformation	associated design formulas
Tiefpaß	$z^{-1} = \frac{Z^{-1} - \alpha}{1 - \alpha Z^{-1}}$	$\alpha = \frac{\sin\left(\frac{\theta_p - \omega_p}{2}\right)}{\sin\left(\frac{\theta_p + \omega_p}{2}\right)}$ <p>ω_p = gewünschte Grenzfrequenz</p>
Hochpaß	$-\frac{Z^{-1} + \alpha}{1 + \alpha Z^{-1}}$	$\alpha = -\frac{\cos\left(\frac{\omega_p + \theta_p}{2}\right)}{\cos\left(\frac{\omega_p - \theta_p}{2}\right)}$ <p>ω_p = gewünschte Grenzfrequenz</p>
Bandpaß	$-\frac{Z^{-2} - \frac{2\alpha k}{k+1}Z^{-1} + \frac{k-1}{k+1}}{\frac{k-1}{k+1}Z^{-2} - \frac{2\alpha k}{k+1}Z^{-1} + 1}$	$\alpha = \frac{\cos\left(\frac{\omega_2 + \omega_1}{2}\right)}{\cos\left(\frac{\omega_2 - \omega_1}{2}\right)}$ $k = \cot\left(\frac{\omega_2 - \omega_1}{2}\right) \tan \frac{\theta_p}{2}$ <p>ω_2, ω_1 = gewünschte obere und untere Grenzfrequenz</p>
Bandsperre	$-\frac{Z^{-2} - \frac{2\alpha}{1+k}Z^{-1} + \frac{1-k}{1+k}}{\frac{1-k}{1+k}Z^{-2} - \frac{2\alpha}{1+k}Z^{-1} + 1}$	$\alpha = \frac{\cos\left(\frac{\omega_2 + \omega_1}{2}\right)}{\cos\left(\frac{\omega_2 - \omega_1}{2}\right)}$ $k = \tan\left(\frac{\omega_2 - \omega_1}{2}\right) \tan \frac{\theta_p}{2}$ <p>ω_2, ω_1 = gewünschte obere und untere Grenzfrequenz</p>

Literatur

- Allen, J. B. (1977): „Short term spectral analysis, synthesis, and modification by discrete Fourier transform,“ IEEE Trans. ASSP 25, 235-238.
- Alrutz, H.: „Über die Anwendung von Pseudoräuschsignalen zur Messung an linearen Übertragungssystemen“, Dissertation, Universität Göttingen, 1983.
- Claasen, T. A. C. M., Mecklenbräuker, W. F. G. (1980): The Wigner-Distribution- a Tool for Time-Frequency Signal Analysis, Part I-III, Philips Journal of Research 35, 217-250, 276-300, 372-389.
- Cohen, L. (1989): Time-Frequency Distributions-A Review; Proceedings of the IEEE 77, 941-981.
- Golomb, S.W.: Shift Register Sequences; Holden-Day 1967.
- Kollmeier, B.: Meßmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache; Habilitationsschrift, Universität Göttingen, 1990.
- Kunze, H. J.: Physikalische Meßmethoden; Teubner, 1986.
- Lacroix, A.: Digitale Filter: Eine Einführung in zeitdiskrete Signale und Systeme; Oldenbourg, 1980.
- Lüke, H. D.: Signalübertragung; Springer, 1989.
- Mildenberger, O.: System- und Signaltheorie, Vieweg, 1989.
- Müller, R.: Rauschen; (Halbleiter-Elektronik), Springer, 1990.
- Oppenheim, A. V. , Schafer, W.: Digital Signal Processing; Prentice Hall, 1975.
- Papoulis, A. : Probability, random variables and stochastic processes, McGraw Hill, 1965
- Papoulis, A: The Fourier Integral and its Applications; McGraw-Hill, 1962
- Press, W. Teukolsky, S., Vetterling, W. T., Flannery, B. P., Numerical Recipes in C-The Art of Scientific Computing; Cambridge University Press, 1992.
- Ronneberger, D.: Vorlesungsskript Spektral- und Korrelationsanalyse, III.Physikalisches Institut, Universität Göttingen, 1989.
- Sachs L.: Angewandte Statistik: Anwendung statistischer Methoden; Springer, 1990.
- Schroeder, M. R.: Fractals, Chaos, Power Laws; Freeman, 1991.
- Schroeder, M. R.: Number theory in science and communication; Springer, 1986.
- Stearns, S. D.: Digitale Verarbeitung analoger Signale, Oldenbourg, 1979.