

## III Verarbeitung und Analyse akustischer Signale

### III.1 Aufnahme- und Wiedergabetechnik: Bestandteile der Übertragungskette

Die Aufnahme, Analyse, Verarbeitung und Wiedergabe akustischer Signale erfordert eine Reihe von technischen Elementen und Techniken, deren detaillierte Behandlung für jedes Einzelelement den Rahmen der vorliegenden Ausarbeitung bei weitem sprengen würden. Es soll daher zunächst in groben Zügen erleutert werden, welche Prinzipien bei den einzelnen Schritten von Bedeutung sind und was bei den einzelnen technischen Elementen an Besonderheiten zu beachten ist.

Die Übertragungskette der Aufnahme- und Wiedergabetechnik kann aufgeteilt werden in den **Schallaufnehmer**, den **Übertragungskanal bzw. Speicher** des Schallsignals und Elemente zur **Schallwiedergabe** (vgl. Tabelle 3.1). Bei den Schallaufnehmern ist das Mikrofon das praktisch wichtigste Element. In einer praktischen Aufnahme-Situation werden zu meist mehrere Mikrophone in unterschiedlicher Anordnung verwendet, die eine gewisse Richtcharakteristik jeweils aufweisen. Für einfache Stereo-Aufnahmen genügen beispielsweise zwei Mikrophone in einem gewissen Abstand voneinander, deren Richtcharakteristik sich räumlich etwas unterscheidet (z. B. Super-Nieren-Charakteristik, die jeweils etwas nach rechts bzw. links von der Mitte geschwenkt wird). Durch eine derartige Anordnung werden im Stereo-Signal sowohl Intensitäts- als auch Laufzeit (bzw. Phasen-) Unterschiede codiert, die beim Abhören der Aufnahme (z. B. über Kopfhörer) einen räumlichen Eindruck vermitteln. Um ein nachträgliches Abmischen beispielsweise einer Musikaufnahme zu ermöglichen, werden dagegen mehrere Mikrophone verwendet, die jeweils dicht vor den auszunehmenden Schallquellen angeordnet sind und deren Richtcharakteristik eine maximale Übertragung der „gewünschten“ Schallquelle mit größtmöglicher Dämpfung „unerwünschter“ Richtungen ermöglichen sollte. Je nach Verstärkung eines Mikrophon-Ausgangs-Signals für den linken oder rechten Stereokanal kann die räumliche Position dieser Schallquelle in der Aufnahme eingestellt werden, wobei allerdings keine Phasenunterschiede, sondern nur Intensitätsunterschiede zwischen den beiden Stereokanälen auftreten („Intensitäts-Stereophonie“). Die Verringerung des Abstandes zwischen Mikrophon und Schallquelle dient auch zur Verringerung des Einflusses des Grundrauschens der Mikrophone, das selbst bei den **rauschärmsten** erhältlichen Mikrophenen noch über dem Pegel der Ruhehörschwelle bei Normalhörenden liegt. Dieses Rauschen wird bei der Wiedergabe unter die Ruhehörschwelle von Normalhörenden erst gedrückt, wenn eine Abschwächung des aufgenommenen akusti-

schen Signals erfolgt, die aufgrund des geringen Mikrofon-Abstandes zur Schallquelle oder aufgrund der Vielzahl der aufgenommenen Schallquelle notwendig wird. Die Weiterleitung des Mikrofon-Signals an die nachfolgende Schallübertragung oder Speicherungs-Einheit kann entweder drahtgebunden erfolgen oder drahtlos, d. h. durch eine Infrarot- oder Hochfrequenz-Sende-Empfangs-Anordnung.

Um eine möglichst realistische und naturgetreue Schallwiedergabe zu ermöglichen, wird anstelle von Mikrofonen oft ein Kunstkopf eingesetzt, bei dem eine Schulter-, Kopf-, und Außenohr-Nachbildung verwendet wird, bei der anstelle der Gehörgänge Mikrophone eingesetzt sind. Bei Kopfhörerwiedergabe der mit diesen Mikrofonen an einem bestimmten Ort aufgenommenen Schallsignale wird der Eindruck erweckt, als befände man sich an dem Aufnahmeort, an dem der Kunstkopf sich befunden hat. Am originalgetreuesten und am wenigsten richtungsverzerrend wirkt die Aufnahme dann, wenn sie mit den individuellen, eigenen Ohren der abhörenden Person aufgenommen wurde. Dies kann z. B. durch die Verwendung von Sondenmikrofonen im eigenen Gehörgang bei der Schallaufnahme erfolgen oder durch „virtuelle“ Akustik, bei der die Außenohr-Übertragungsfunktion für jede Versuchsperson individuell für jede in Frage kommende Schalleinfallrichtung ausgemessen wird. Zur Simulation des Schalleinfalls aus einer bestimmten Richtung wird dann im Computer das entsprechende Schallsignal mit der gemessenen, individuellen Außenohr-Impulsantwort gefaltet, so daß die Kunstkopftechnik damit simuliert wird.

Zur Aufnahme von akustischen Phänomenen, die nicht von dem Medium Luft, sondern von anderen Medien übertragen werden (z. B. Wasser, andere Flüssigkeiten oder Festkörper) gibt es eine Reihe von spezialisierten Schallaufnehmern (z. B. Hydrophone, Körperschallaufnehmer), auf die hier nicht näher eingegangen werden soll.

Tabelle 3.1: Mögliche Bestandteile einer Schallübertragungskette

Schallaufnehmer:	Übertragungskanal/Speicher:	Schallwiedergabe:
Mikrophon	Impedanzwandler/Verstärker	Lautsprecher
<ul style="list-style-type: none"> <li>• Anzahl</li> <li>• Anordnung</li> <li>• Richtcharakteristik</li> <li>• Empfindlichkeit</li> <li>• Rauschen</li> <li>• drahtgebunden/drahtlos</li> </ul>	<ul style="list-style-type: none"> <li>Mischpult (analog/digital)</li> <li>Tonband (analog/digital)</li> <li>Equalizer</li> <li>(End-)Verstärker</li> <li>Signalleitung</li> <li>Signalübertragung (IR, HF)</li> </ul>	<ul style="list-style-type: none"> <li>• Bauform</li> <li>• Frequenzgang</li> <li>• Leistung</li> <li>• Raumakustik</li> </ul>
Kunstkopf		Kopfhörer
<ul style="list-style-type: none"> <li>• realer Kopf</li> <li>• nachgebildeter Kopf</li> <li>• simulierter Kopf</li> </ul>		<ul style="list-style-type: none"> <li>• Bauform</li> <li>• Frequenz-Gang</li> <li>• Phasengang</li> <li>• max.Pegel</li> </ul>
		Schwingungsgeber

Die mittlere Spalte von Tabelle 3.1 enthält einige wichtige Beispiele von Übertragungs-Elementen von Schallsignalen bzw. von Speicherelementen. Für Aufnahmen mit einem Audio- oder Meßmikrophon ist meistens die Verwendung eines Impedanz-Wandlers und Vorverstärkers notwendig, die für eine möglichst störungsfreie Übertragung des vom Mikrophon erzeugten, sehr kleinen elektrischen Signals an den eigentlichen Verstärker bzw. das Speichermedium notwendig sind. Während beispielsweise die meisten dynamischen Mikrophone eine kleine Ausgangsimpedanz von wenigen 100 Ohm aufweisen (d. h. nur geringe Spannungen, dafür aber reaktiv hohe Ströme „liefern“ können), haben die meisten Eingangsstufen von Leistungsverstärkern eine relativ hohe Eingangsimpedanz (d. h. bei Anlegen einer gewissen Wechselspannung fließt nur ein kleiner Wechselstrom). Da auf der Kabelverbindung zwischen Mikrophon und Endverstärker von einer gewissen Störeinstreuung mit konstanter Leistung ausgegangen wird, muß nach Möglichkeit möglichst viel von der in elektrische Energie umgewandelten Schallenergie von Mikrophon zum (End-) Verstärker übertragen werden. Diese Leistungsübertragung ist optimal, wenn die Impedanz des Mikrophons mit der Eingangs-Impedanz des Verstärkers übereinstimmt, was beispielsweise durch die Verwendung eines Übertragers (Transformator zur Impedanzanpassung) oder durch einen Vorverstärker bewerkstelligt werden kann, der im Mikrophon bzw. in unmittelbarer Nähe des Mikrophons angebracht wird und eine Impedanzwandlung mit gleichzeitiger Vorverstärkung der Mikrophon-Signale erreicht. In der Studio-Technik werden die Mikrophon-Signale dann einem Mischpult zugeführt, bei dem der Pegel, der Klang (durch Anheben/Absenken bestimmter Frequenz-Bereiche), die Kanalzuordnung (linker bzw. rechter Stereokanal) und weitere Kontrollparameter (z. B. Übertragung für die Bühnen-Monitor-Anlage) eingestellt werden kann. Neben der konventionellen Analog-Technik setzt sich zunehmend auch die Digital-Technik für den Studio-Bereich durch. Dies betrifft auch die Ab-

speicherung von Schall-Signalen, die beispielsweise mit Tonbandgeräten (analoge Technik) bzw. digitalen Bandgeräten (z. B. DAT-Tape) nahezu erfolgen kann. Mit Hilfe eines **Equalizers** kann der Gesamt-Frequenzgang der Aufnahme beeinflußt werden und anschließend einem Endverstärker zugeführt werden, der die Leistung des elektrischen Eingangssignals soweit verstärkt, daß das Signal über Lautsprecher wieder hörbar gemacht werden kann. Am Ausgang des Endverstärkers befindet sich über eine Signalleitung verbunden beispielsweise ein Lautsprecher. Ebenso wie bei den Eingangssignalen des Mischpults (z. B. bei drahtlosen Mikrophon-Anlagen) kann die Übertragung des Ausgangssignals auch drahtlos (über Infrarot oder Hochfrequenz-Verbindungsstrecke) erfolgen. Diese Übertragungsart weist einige praktische Vorteile auf (z. B. Verringerung des Verlegeaufwandes von Leitungen, Beweglichkeit von Musikern auf der Bühne).

Bei der **Schallwiedergabe** ist das häufigste Grundelement der Lautsprecher, der in verschiedenen Bauformen erhältlich ist und in unterschiedliche Gehäuse eingebaut werden kann. Das Ziel ist dabei ein möglichst ausgewogener Frequenzgang, d. h. eine möglichst glatte Übertragungsfunktion vom elektrischen Signal zum akustischen Signal in einiger Entfernung des Lautsprechers. Neben dem Frequenzgang des Lautsprechers ist auch die maximal abgebbare Leistung ein wichtiger Parameter, bei der zum einen zwischen der Eingangsleistung (d. h. maximal verkräfteter elektrischer Antrieb des Lautsprechers) und der akustischen Leistung (d. h. effektiv abgestrahlte akustische Signal-Leistung) unterschieden werden muß. Gebräuchliche HiFi-Lautsprecher weisen einen Wirkungsgrad von nur wenigen Prozent auf. Obwohl die Übertragungsfunktion einer Lautsprecher-Box im reflektionsarmen Raum durchaus einen glatten Verlauf aufweisen kann, ist für den Betrieb derselben Lautsprecher-Box in einem geschlossenen, mit reflektierenden Wänden versehenen Raum mit einer deutlichen Veränderung der Übertragungsfunktion zu rechnen: Beispielsweise werden durch die Anregung von Raumresonanzen ausgeprägte Minima und Maxima in der Übertragungsfunktion bewirkt, so daß bei der Beurteilung der Qualität von Lautsprecher-Boxen neben den reinen akustischen Eigenschaften auch die Interaktion zwischen der Box und dem Raum zu beachten ist.

Dieses Problem tritt nicht bei Kopfhörer-Wiedergabe auf, wobei sich die einzelnen Kopfhörer ebenfalls in ihrer Bauform (z. B. circumoral, d. h. das gesamte Ohr umschließend, supraaural, geschlossen / offen) und in ihrem Frequenz- und Phasengang unterscheiden. Für den Frequenzgang von Kopfhörern wird eine bestimmte Vorgabe (z. B. Freifeld entzerrt oder Diffusfeld entzerrt) verlangt, d. h. die elektroakustische Übertragung des Kopfhörers soll die Situation nachbilden, bei der im Freifeld die hörende

Person das Schallsignal von vorne (freifeldentzerrt) bzw. statistischerweise von allen Richtungen (Diffusfeld-Entzerrung) angeboten bekommt. Ebenfalls ist der maximal mit einem Kopfhörer erreichbare Pegel von Bedeutung, einerseits weil eine Begrenzung notwendig erscheint, (z. B. bei Walkman-Kopfhörern, die bei sehr hohen Darbietungspegeln zu irreversiblen Gehörschäden führen können) oder andererseits bei Audiometrie-Kopfhörern, bei denen zu Testzwecken auch bei hochgradig schwerhörigen Versuchspersonen noch genügend akustische Energie übertragen werden muß.

Auf die Diskussion von Schwingungsgebern für die Anregung von Schallwellen auf anderen Materialien (z. B. Festkörpern oder Flüssigkeiten) soll hier verzichtet werden.

Die meisten der von den bisher betrachteten Komponenten der akustischen Übertragungskette durchgeführten Operationen sind lineare, zeitinvariante Filterungs- bzw. Verstärkungsoperationen. Ohne auf die genaue Filterwirkung bei den einzelnen Techniken näher einzugehen, sollen zwei Prototypen von Filtern stellvertretend diskutiert werden, und zwar ein analoges und zum anderen ein digitales Filter.

Ein analoges Filter besteht aus einem Netzwerk aus passiven Bauelementen wie Widerständen, Kondensatoren und Induktivitäten, sowie aktiven Elementen, die durch Operationsverstärker erzeugt werden, in deren Gegenkopplungszweig sich wiederum passive Bauelemente befinden. Ein Beispiel eines derartigen analogen Filters ist in untenstehender Figur abgebildet.

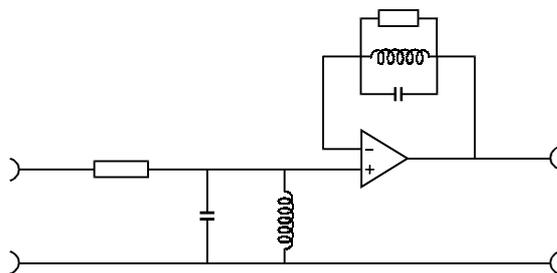


Abbildung 3.1: Schaltbild eines analogen Filters (schematisch)

Zur Berechnung der Wirkungsweise des Filters und der Dimensionierung der Bauelemente sei auf die einschlägige Literatur (lineare Systemtheorie, Laplace-Transformation, Design analoger Filter) verwiesen. Der Vorteil analoger Filter ist ihr relativ geringer Implementierungsaufwand und ihrer Einsetzbarkeit auch bei sehr hohen Frequenzen. Der Vorteil digitaler Filter

dagegen ist die Einsetzbarkeit auch bei sehr niedrigen Frequenzen, die hohe erzielbare Genauigkeit und Stabilität der Filter- Übertragungsfunktion und die Möglichkeit, dispersionsfreie Filter (d. h. linearphasige Filter) zu konstruieren. Der hohe Implementierungsaufwand wird mit zunehmendem Fortschritt der Mikroelektronik allerdings immer weniger ausschlaggebend, so daß zum Beispiel in jedem Kompakt-Disk-Spieler bereits komplexe digitale Filter eingebaut sind. Der prinzipielle Aufbau eines digitalen Filters, das auf zeit- und wertediskreten Signalen (abgetasteten Signal-Samples) operiert, ist im folgenden aufgeführt:

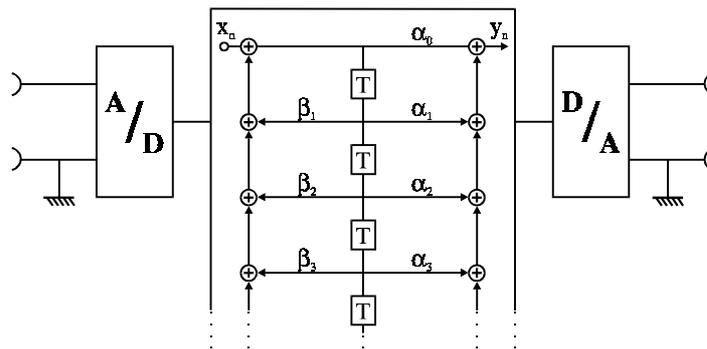


Abbildung 3.2: Schema eines digitalen Filters (schematisch)

Nach Analog-Digital-Wandlung, bei der das zeit- und wertekontinuierliche elektrische Eingangssignal bei einer festen Abtastfrequenz in ein zeit- und wertediskretes, im Computer darstellbares Signal umgewandelt wird, besteht das eigentliche Filter aus einem rekursiven und einem nichtkursiven Teil. Beiden Teilen ist gemeinsam, daß die vorhergehenden Werte gespeichert werden und für die Berechnung der nachfolgenden Ausgangswerte verwendet werden. Beim rekursiven Anteil (links in der Abbildung) wird das Ausgangssignal des Filters nach Zeitverzögerung wieder mit den Koeffizienten  $\beta_1$ ,  $\beta_2$ , bis  $\beta_M$  multipliziert, aufaddiert und wieder auf das Eingangssignal gegeben, wobei  $M$  die Ordnung des Filters bezeichnet. Diese rekursive Berechnung des Ausgangssignals führt zu unendlich langen Impulsantworten, so daß sie auch als Infinite Impulse Response (IIR)-Filter bezeichnet werden. Sie haben den Vorteil einer hohen Filterschärfe aber den Nachteil, bei bestimmten Parameter-Kombinationen instabil zu werden und unter Umständen Probleme bei Quantisierungs-Fehlern zu bereiten. Unproblematischer, aber mit einem höheren Aufwand versehen, sind dagegen die nichtrekursiven oder Finite Impulse Response (FIR)-Filter, bei denen das zeitverzögerte Eingangssignal nach Multiplikation mit dem Koeffizienten  $\alpha_1$   $\alpha_M$  aufaddiert und auf den Ausgang des Filters gegeben werden. Dieser Ausgang wird dann durch D/A-Wandlung wieder in ein zeit- und wertekontinuierliches Signal verwandelt. Zur genaueren Analyse und zum Design digitaler Filter sei auf die einschlägige Literatur über di-

gitale Signalverarbeitung und das Skript zur Vorlesung „Physikalische Meßtechnik und digitale Signalverarbeitung“ verwiesen.

Während in der akustischen Aufnahme- und Wiedergabetechnik die oben prinzipiell erläuterte Filterung von elektrischen Signalen an mehreren Stellen des Übertragungsweges benötigt wird (z. B. zum Abschneiden unerwünschter, hoher Frequenzen, die nur Rauschanteile enthalten oder vom Ausgleichen der elektroakustischen Übertragungsfunktion mit einem Equalizer) werden in der akustischen Meßtechnik Filter vorwiegend für die **Analyse** von akustischen Signalen verwendet. Dieses ist z. B. für die frequenzabhängige Schallpegel-Bestimmung notwendig, bei der das Mikrofon-Signal erst über ein Bandpaßfilter an das Pegelmeßgerät angeschlossen wird, so daß nur die in dem Durchlaßbereich des Bandpaßfilters fallenden Frequenzen zu dem gemessenen Schallpegel beitragen und eine grobe Approximation des Frequenz-Spektrums des zu messenden Signals möglich ist, in den eine gesamte Bank von Bandpaßfiltern verwendet wird. Das Prinzip einer derartigen **Filterbank**, die in der akustischen Meßtechnik zumeist in Oktavschritten oder in Terzschritten erfolgt, ist in folgender Abbildung aufgeführt:

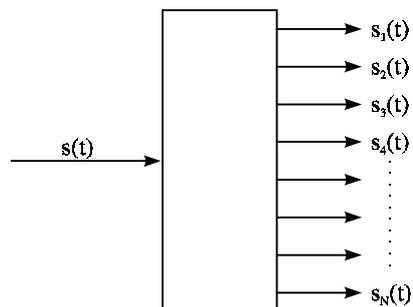


Abbildung 3.3: Prinzipieller Aufbau einer Filterbank

Das Eingangssignal wird in  $N$  verschiedene Ausgangssignale aufgespalten, die unterschiedlichen Bandpaßcharakteristiken entsprechen. Bei Oktav-Filterbänken sind die Grenzfrequenzen im Verhältnis einer Oktave angeordnet (d. h. 1:2), während bei Terz-Filterbänken, das Abstandsverhältnis 4:5 beträgt. Die für die akustische Meßtechnik normierten Grenzfrequenzen von Oktav- bzw. Terzbändern sind in der folgenden Tabelle aufgeführt:

Tabelle 3.2: Mittenfrequenzen  $f_m$  von Terz- und Oktavbändern mit den Frequenzgrenzen  $f_u$  und  $f_o$ .

Oktave			Terz		
$f_u$ (Hz)	$f_o$ (Hz)	$f_m$ (Hz)	$f_u$ (Hz)	$f_o$ (Hz)	$f_m$ (Hz)
11	22	16	14,1	17,8	16
			17,8	22,4	20
			22,4	28,2	25
22	44	31,5	28,2	35,5	31,5
			35,5	44,7	40
			44,7	56,2	50
44	88	63	56,2	70,7	63
			70,7	89,1	80
			89,1	112	100
88	177	125	112	141	125
			141	178	160
			178	224	200
177	355	250	224	282	250
			282	355	315
			355	447	400
355	710	500	447	562	500
			562	708	630
			708	891	800
710	1420	1000	891	1122	1000
			1122	1413	1250
			1413	1778	1600
1420	2840	2000	1778	2239	2000
			2239	2818	2500
			2818	3548	3150
2840	5680	4000	3548	4467	4000
			4467	5623	5000
			5623	7079	6300
5680	11360	8000	7079	8913	8000
			8913	11220	10000
			11220	14130	12500
11360	22720	16000	14130	17780	16000
			17780	22390	20000

Weil diese Oktav- bzw. Terzanalyse praktisch von großer Bedeutung ist, seien im folgenden einige der prinzipiellen Eigenschaften dieser Art von Frequenzanalyse aufgeführt:

- Jedes der Frequenzbänder hat eine konstante **relative** Bandbreite  $\Delta f/f$ , d. h. die Frequenzbänder sind nicht äquidistant, sondern äquidistant auf einer logarithmischen Frequenz-Skala.
- Für sämtliche Ausgangskanäle ergibt sich approximativ die gleiche Ausgangsleistung, wenn als Eingangssignal rosa Rauschen anliegt, d. h. die spektrale Leistungsdichte  $S^2(f) \sim 1/f$ . Dieses rosa Rauschen oder  $1/f$ -Rauschen findet sich sehr häufig bei natürlichen Rauschvorgängen. Für die meisten natürlichen akustischen Signale (z. B. Umweltgeräusche und

Sprache) findet sich ebenfalls approximativ ein  $1/f$  Spektrum. Durch die entsprechende Wahl der Filterbank ist so sichergestellt, daß im höchsten Frequenzband ungefähr dieselbe Energie wie im niedrigsten Frequenzband erfaßt wird, so daß sämtliche Frequenzbänder für die Beurteilung des Schalls etwa die gleiche Wichtigkeit haben.

- Die konstante relative Bandbreite entspricht ungefähr der Frequenzanalyse im Gehör. Bei der im Innenohr stattfindenden Frequenz-Ortstransformation werden die höchsten Frequenzen zu Beginn der Basilarmembran im Innenohr und die niedrigsten Frequenzen am Ende (Apex bzw. Helicotrema) abgebildet, wobei approximativ der Ort auf der Basilarmembran dem Logarithmus der Frequenz entspricht. Diese logarithmische Abbildung gilt allerdings erst oberhalb von etwa 500 Hz bis etwa 20 kHz. Unterhalb von 500 Hz befindet sich dagegen eine ungefähr lineare Frequenzabbildung.

- Die Oktav- bzw. Terzanalyse entspricht der in letzter Zeit häufig diskutierten Wavelet-Analyse, bei der ein Zeitsignal mit selbstähnlichen Analysefunktionen (den Wavelets, zu Deutsch etwa Wellenpaketchen) analysiert wird. Für den allgemeinen Fall einer Wavelet-Analyse wird die Wavelet-Analyse-Funktion  $h(t)$  definiert, die z. B. die Form der Impulsantwort eines Bandpaßfilters aufweist:

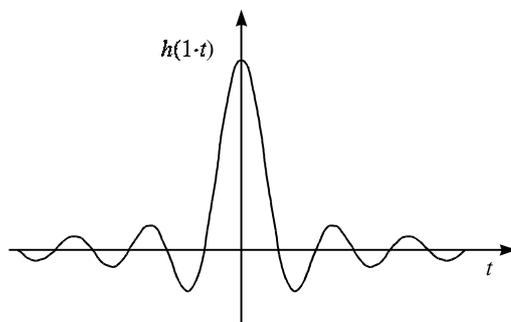


Abbildung 3.4: Beispiel für eine Wavelet-Analyse-Funktion  $h(t)$

Durch Dehnung bzw. Stauchung dieses Wavelets können nun Filter-Impulsantworten generiert werden, die selbstähnlich zu der ursprünglichen Impulsantwort sind und zu Filtern führen, die genau die gleiche relative Bandbreite haben wie der ursprüngliche Bandpaßfilter. Wenn wir also einen Zeit-Stauchungs-Faktor  $a$  mit dem zugehörigen gestauchten Wavelet  $h(a \cdot t)$  definieren (siehe Abbildung 3.5), können wir die Wavelet-Analyse-Funktion als Ausgangssignal des Bandpaßfilters mit dem Frequenz-Stauchungsparameter  $a$  betrachten:

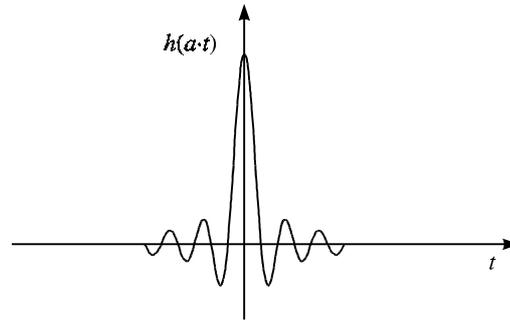


Abbildung 3.5: Mit dem Parameter  $a$  gestauchte Wavelet-Analyse-Funktion aus Abb. 3.4

$$\Rightarrow S_a(t) = \int_{-\infty}^{+\infty} s(t-t') \cdot h(at') dt = s(t) * h(at) \quad (\text{III.1})$$

Somit stellt sich die Wavelet-Analyse als eine Zeit-Frequenz-Verteilungsfunktion mit konstanter relativer Bandbreite dar, die in guter Näherung der Analyse im menschlichen Ohr entspricht. Eine ähnliche Analogie lässt sich auch für die zweidimensionale Wavelet-Analyse und der Transformation finden, die im visuellen System des Menschen durchgeführt wird.

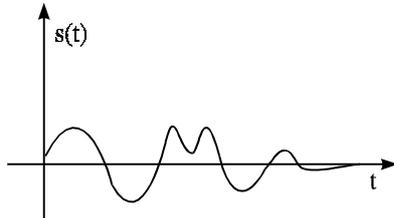
### III.2 Diskrete Spektralanalyse

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  $\delta$ -Kamm auffassen, so daß sich das Spektrum dieses diskretisierten Signals als Faltung des Original-Spektrums mit der Fourier-Transformierten des  $\delta$ -Kamms, nämlich dem  $\delta$ -Kamm mit der inversen Periodizität schreiben läßt:

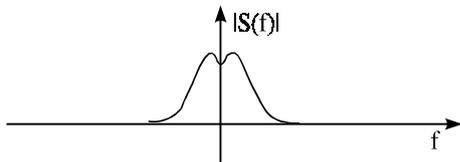
*Aperiodische Signale*

Kontinuierliche Signale

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



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

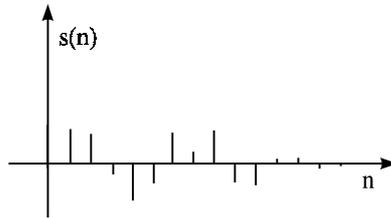


Spektrum aperiodisch, kontinuierlich

Diskrete Signale

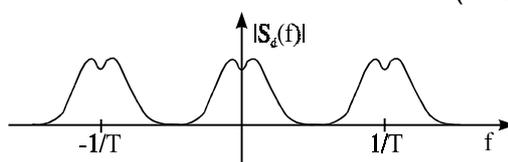
$$s(n), n \in \mathbb{Z}, s(n) = \begin{cases} s(t), & t = n \\ 0 & \text{sonst} \end{cases}$$

(III.2)



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

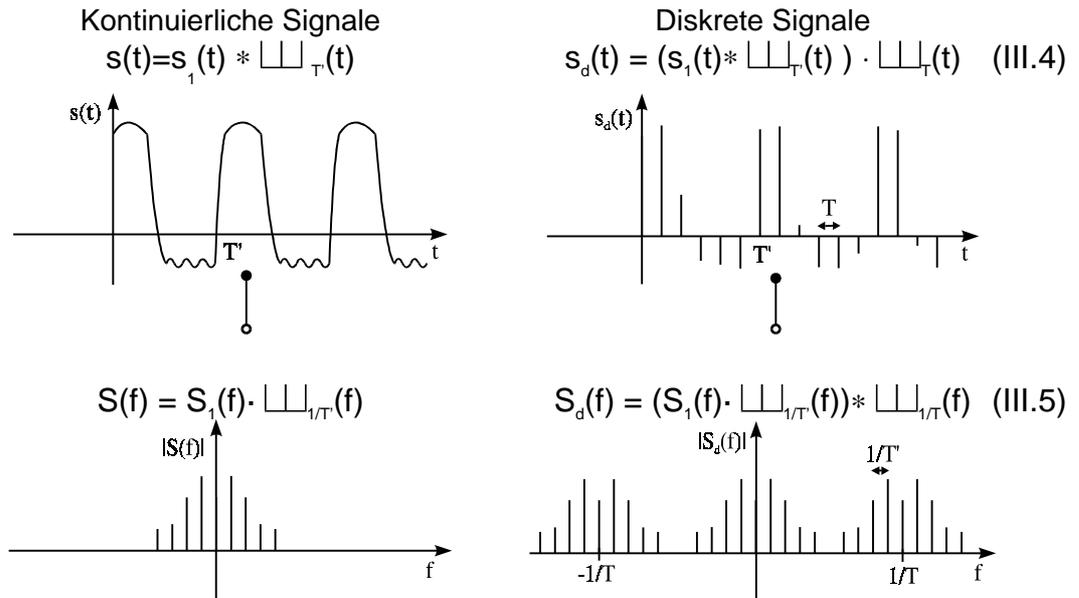
(III.3)



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  $\delta$ -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  $\delta$ -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:

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

$$S(k) \longleftrightarrow s(n) = \frac{1}{N} \sum_{K=0}^{N-1} S(k) e^{2\pi i \frac{nk}{N}} \quad (\text{III.6})$$

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 Pe-

riode 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.

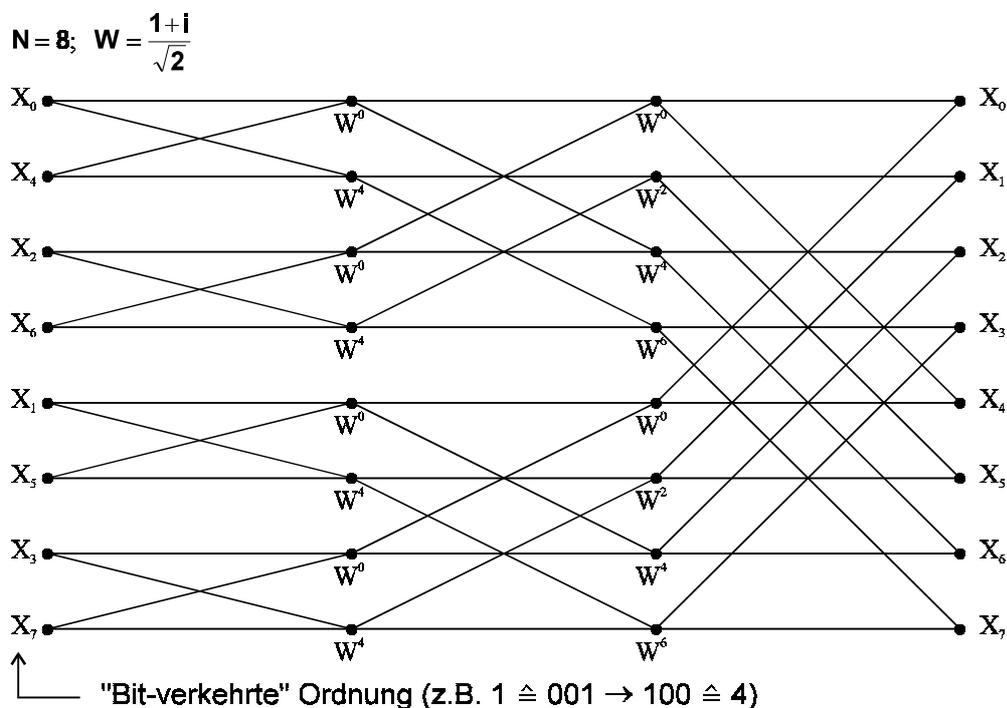


Abbildung 3.6: Schema der FFT

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:

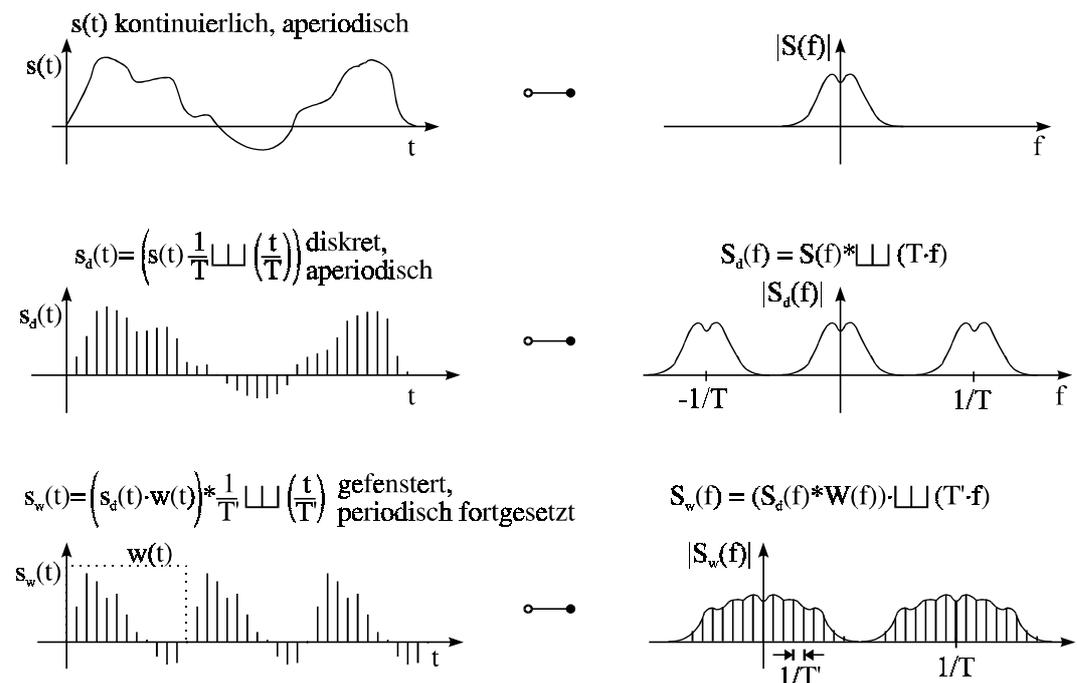


Abbildung 3.7: Vom kontinuierlichen aperiodischen Signal zum diskreten, gefensterten Signal

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  $S_d(f)$  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 Orginalspektrums. Um den dabei entstehenden Fehler möglichst klein zu halten, sollte ein Fenster  $w(t)$  mit möglichst  $\delta$ -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:

$$\text{Hanning-Fenster: } w_1(n) = 0,5 - 0,5 \cos \frac{2\pi n}{M-1} \quad (\text{III.7})$$

$$\text{Hamming-Fenster: } w_2(n) = 0,54 - 0,46 \cos \frac{2\pi n}{M-1} \quad (\text{III.8})$$

$$\text{Blackman-Fenster: } w_3(n) = 0,42 - 0,5 \cos \frac{2\pi n}{M-1} + 0,08 \cos \frac{4\pi n}{M-1} \quad (\text{III.9})$$

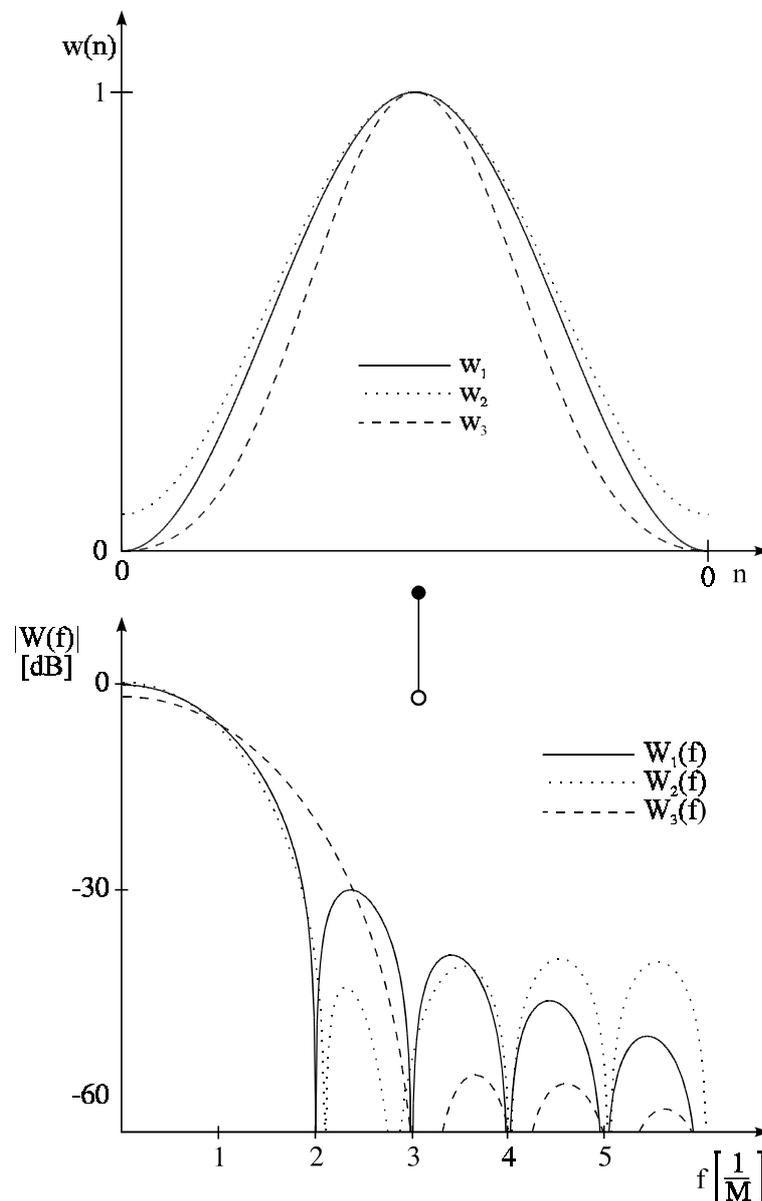


Abbildung 3.8: Die verschiedenen Fenster im Zeit- und Frequenzbereich

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 anhängt.

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  $\delta$ -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 Unschärferelation 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.

### III.3 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

$$S(f, t) = \int_{-\infty}^{\infty} s(t'+t)w(t') \cdot e^{-2\pi i f t'} dt' \quad (\text{III.10})$$

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

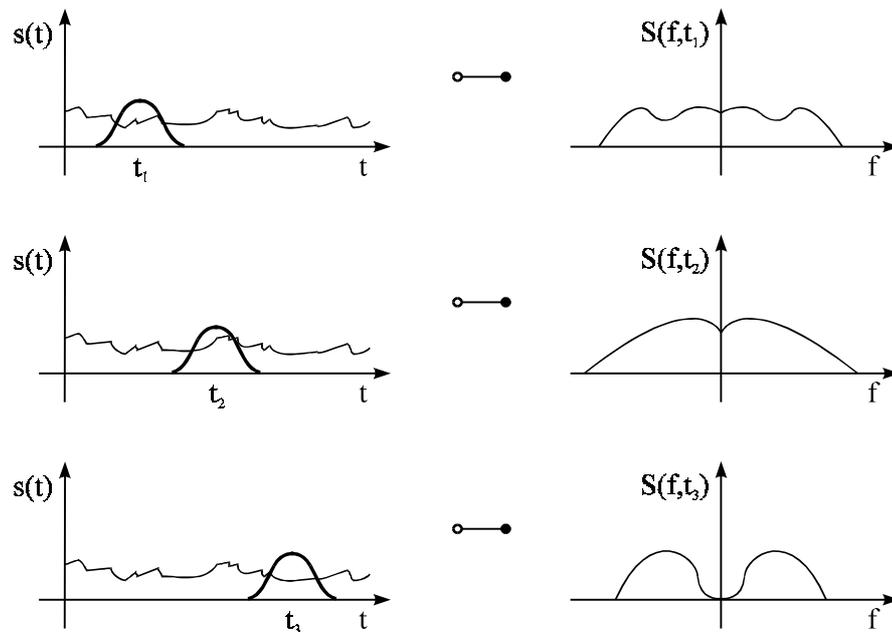


Abbildung 3.9: Zeitfrequenzanalyse

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

$$S(f_k, t_m) = \sum_{n=0}^{N-1} s(nT + t_m) w(n) e^{-2\pi \frac{nk}{N}} \quad (\text{III.11})$$

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/„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 Forderungen resultieren die folgenden Abtasttheoreme:

1.) Abtasttheorem:

$$f_s = \frac{1}{T} \geq 2f_{\text{grenz}} \quad (\text{III.12})$$

D. h. die höchste im Originalsignal vorkommende Frequenz  $f_{\text{grenz}}$  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:

$$S'(f_k, t_m) = H(f_n) \cdot S(f_k, t_m)$$

$$\downarrow$$

$$s'(n + t_m) = h(n) * (s(n + t_m) \cdot w(n)) \quad (\text{III.13})$$

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

$$N \geq \text{Fensterlänge } M + \text{Länge von } h(n) \quad (\text{III.14})$$

Bei der Hintransformation müssen demnach eine genügend große Anzahl von Nullen an das gefensterete 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):

$$S(f,t) = \int_{-\infty}^{\infty} s(t'+t)w(t')e^{-2\pi ift'} dt' = e^{2\pi ift} \cdot \int_{-\infty}^{\infty} s(t') \cdot e^{-2\pi ift'} \cdot w(t'-t)dt' \quad (\text{III.15})$$

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

$$f_{\text{grenz}} \cong \frac{2}{M} \cdot f_s \quad (\text{III.16})$$

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:

$$F = 4 \cdot \frac{f_s}{M} \quad (\text{III.17})$$

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, stellt ein **drittes Abtasttheorem** dar. Falls ein größerer Fehler toleriert werden kann und als obere Grenzfrequenz des Bandpaß-Signals nicht die erste Nullstelle des Hamming-Fensters, son-

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

