



Seminararbeit aus SE Algorithmen in Akustik und
Computermusik 2 SS08:
Warping

Peter Sciri [peter-mario.sciri@student.kug.ac.at]
Alexander Wankhammer [alexander.wankhammer@student.kug.ac.at]

Graz, 1. Juli 2008

Betreuer: DI Franz Zotter [zotter@iem.at]

1 Einleitung

Das menschliche Gehör weist die Eigenschaft auf, Frequenzen einer nichtlinearen Skala entsprechend aufzulösen. Da jedoch den meisten heute verwendeten Systemen eine äquidistante Verteilung der Frequenzen im Frequenzraum zugrundeliegt, sieht man sich mit einer Fehlanpassung und einer daraus resultierenden Über- bzw Unterbestimmung in verschiedenen Bereichen konfrontiert.

Der Grund dafür liegt in der Verwendung der Einheitsverzögerung als fundamentaler Bestandteil aller Filterstrukturen, welche sich für alle Frequenzen gleich verhält und damit die lineare Frequenzverteilung determiniert.

Das Ziel dieser Arbeit soll nun sein, eine Methode zu erörtern, welche - nicht wie z.B. die Constant-Q Transformation durch Summenbildung ein Pseudo-Abbild aus dem linearen Raum auf einen exponentiellen Raum liefert - sondern durch Manipulation dieses grundlegenden Zusammenhangs zwischen Aufbau der Filterstruktur und Frequenzverteilung eine Reduktion dieser Fehlanpassung erreicht [7].

2 Auditorische Skalen

2.1 Cochlea

Die menschliche Frequenzwahrnehmung annähernd durch eine exponentielle Kurve der Basis 2 beschrieben werden.

Der Grund dafür liegt darin, wie sich Schallwellen im inneren des Gehörorgans verhalten. Nach der Impedanzanpassung durch das Mittelohr werden die Vibrationen am ovalen Fenster in Stehwellen im Inneren der Cochlea gewandelt. Dabei werden jeweils, je nach Frequenz, spezielle Ortsbereiche in Schwingung versetzt: es wird eine Frequenz - Ortstransformation durchgeführt.

Diese Beziehung ist gegeben durch

$$f = 165.4 (10^{0.06x} - 1) \quad (1)$$

bzw

$$x = \frac{1}{0.06} \log_{10} \left(\frac{f - 165.4}{165.4} \right) \quad (2)$$

Hier kann festgestellt werden, dass sich die Abstände der Schwingungsmaxima ähnlich der Tonheitsempfindung und damit exponentiell zur linearen Frequenzachse in Hertz verhalten.

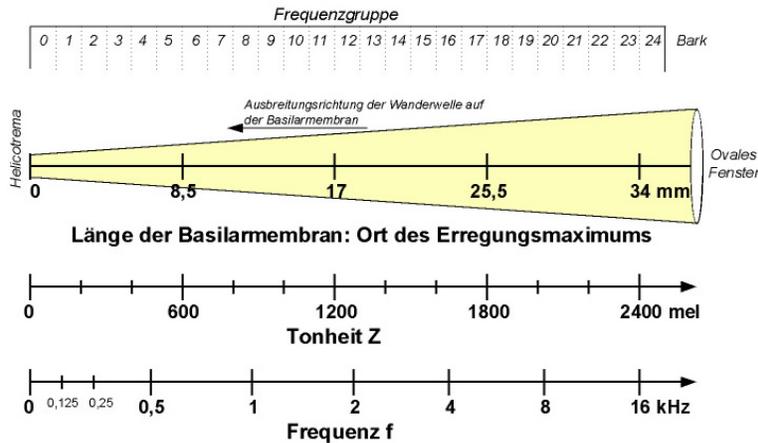


Abbildung 1: Orts-Frequenz Transformation an der Basilarmembran

2.2 Psychoakustische Skalen

Eine Weiterführung der nichtlinearen Frequenz-Orts-Beziehung der Cochlea stellen die diversen Skalen der Psychoakustik dar, welche durch empirische Messmethoden in der Vergangenheit ermittelt wurden — Beispiele dafür sind kritischer Bandbreiten (Bark), Mel und ERB ([12], [6], [9]) Die für unseren Zweck wichtigste ist jener der Einteilung in *Bark*:

$$1\text{bark} = 100\text{Mel} \quad (3)$$

was im Fall der Verdoppelung des *Mel*-Werts (Vgl. Abb. 1) sowohl einer doppelten Tonhöhenempfindung als auch einer ungefähr gleichen Fortschreitung entlang der Basilarmembran entspricht.

Hierbei werden sowohl Aspekte wie die Einteilung nach kritischen Bandbreiten, was der Modellierung des menschlichen Gehörs als Filterbank entspricht, als auch Verdeckungserscheinungen, Mithörschwellen und Frequenzunterscheidungsgrenzen miteinbezogen.

Bei der Einteilung in Bark-Bänder (kritische Bänder) wird die spektrale Energie innerhalb eines solchen Bandes betrachtet welche dann die Hörempfindung auslöst. Dieser Skala entsprechend kann festgestellt werden, dass das menschliche Ohr im tieffrequenten Bereich eine wesentlich bessere Unterscheidung von Frequenzen bzw. Bändern bewältigen kann als im höheren Frequenzbereich.

Dies stellt nun eine Diskrepanz zur linearen Frequenzverteilung konventioneller DSP Systeme dar: es werden zu viele Ressourcen für Hörbereiche

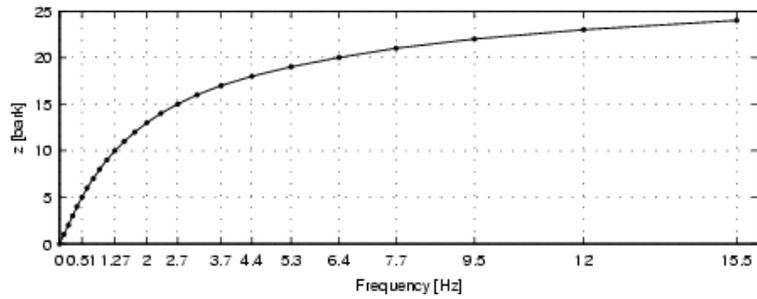


Abbildung 2: Barkskala

benötigt, die das Gehör kaum unterscheiden kann bzw. zu wenig Ressourcen für den eigentlich relevanten Frequenzbereich zu Verfügung gestellt.

Abhilfe für diese Problematik kann auf mehreren Wegen geschaffen werden:

Einen klassischen Ansatz, die spektrale Analyse auf das Auditorische Modell anzupassen, stellt die Constant-Q Transformation nach Judith Brown ([4]) dar. Hierbei wird jedoch eigentlich keine Rechenleistung gespart, da die aus der Analyse resultierenden, gleichverteilten Frequenz Bins (das Wort Bin deutet schon darauf hin, dass bereits hier mehrere 'kontinuierliche' Frequenzen in einen diskreten Wert zusammengefasst werden) in den Bark-Bändern entsprechenden Gruppen zusammengefasst werden.

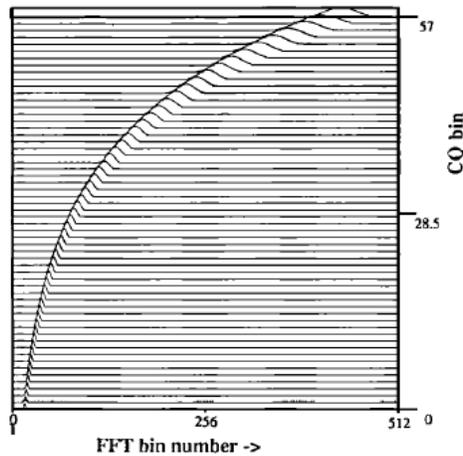


Abbildung 3: Filterbank der Constant-Q Transformation aus [4]

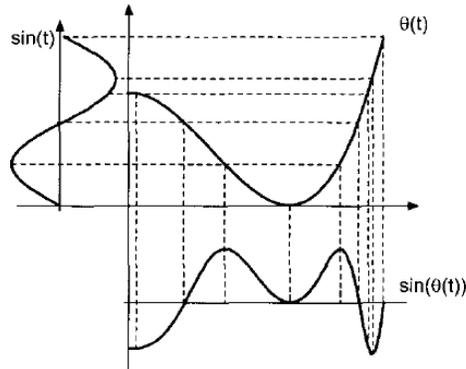


Abbildung 4: Warping eines Zeitsignals [13]

Die naheliegendste Lösung wäre also, die Frequenzachse ansich an die Bark Verteilung anzupassen - durch *warping*. Dies kann jedoch nicht ohne grundlegende Veränderungen der bekannten Filter- und Analysestrukturen geschehen und wird in den folgenden Kapiteln 3, 4 und 5 beschrieben. Eine spezielle Anwendung dieser Technik wird dann in Kapitel 6 beschrieben; außerdem wird aufgezeigt, wie groß die Recheneffizienz dieser Methode sein kann.

3 Warping

Der Begriff *warping* beschreibt im Allgemeinen eine nichtlineare Abbildung einer Domäne auf sich selbst. Dies geschieht z.B. in Verstärkerschaltungen, in denen der Arbeitspunkt falsch eingestellt ist oder durch Temperaturdrift verlassen wird.

Diese spezielle Form von *warping* ermöglicht jedoch keine nützliche Anwendung im Audio Bereich sondern ist eher eine unerwünschte Folge.

Wesentlich Interessanter wäre es jedoch wie schon im letzten Kapitel erwähnt, die Frequenzachse zu krümmen und dadurch eine inhärente Anpassung an das Auditorische Modell zu erreichen.

Ziel ist es auch, vorhandene Filterstrukturen weiterhin verwenden zu können. Da diese Strukturen (FIR, IIR,...) jedoch im wesentlichen nur aus Verzögerungsgliedern, Multiplikatoren und Summatoren bestehen, sind die Möglichkeiten der Einflussnahme sehr eingeschränkt.

Wie von Oppenheim in [3] und Schüssler in [10] beschrieben, bietet sich das Ersetzen der Einheitsverzögerungen durch Allpässe erster Ordnung als

zielführende Lösung an:

Es kann dadurch ein Filter im konventionellen Design (und vor allem auch mit konventionellen Methoden) entworfen werden, der einen zusätzlichen Freiheitsgrad in Form des Parameters λ aufweist, wodurch eine Verzerrung der Frequenzachse erreicht werden kann. Näheres dazu im folgenden Abschnitt.

4 Allpass Transformation

Zum Verständnis der folgenden mathematischen Beschreibung verwendeter Filterstrukturen, ist das Wissen über z-Transformation und zeitdiskrete Signalverarbeitung eine wichtige Voraussetzung [1].

Das Hauptmerkmal eines Allpasses 1. Ordnung ist sein konstanter Amplitudengang des Betrages 1. Dieser lässt sich aus der Übertragungsfunktion

$$H(z) = \frac{z^{-1} - \lambda^*}{1 - \lambda z^{-1}} \quad (4)$$

ableiten:

$$H(e^{j\omega}) = \frac{e^{-j\omega} - \lambda^*}{1 - \lambda e^{-j\omega}} = e^{-j\omega} \frac{1 - \lambda^* e^{j\omega}}{1 - \lambda e^{-j\omega}} \quad (5)$$

Wie von Oppenheim in [2] beschrieben, kann durch Gleichung 5 gezeigt werden, dass sich der konstante Phasengang des Allpasses 1. Ordnung aus der Eigenschaft ergibt, dass jedem Pol von $H(z)$ eine konjugiert reziproke Nullstelle gehört. Folglich gilt:

$$|H(e^{j\omega})| = |H(z)| = 1 \quad (6)$$

Da ein konstanter Amplitudengang die primäre Anforderung an ein z^{-1} Glied darstellt, kann es unter dieser Bedingung durch einen Allpass ersetzt werden. Weiters ist ersichtlich, dass für $\lambda = 0$ der Pol der Übertragungsfunktion aus der Gleichung fällt und somit nur $H(z) = z^{-1}$ bleibt.

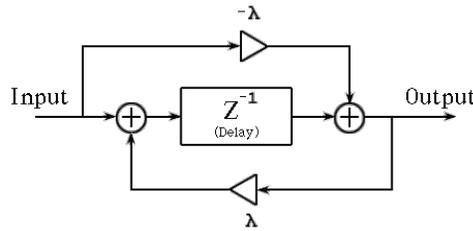


Abbildung 5: Blockdiagramm eines Allpasses 1. Ordnung

Im Unterschied zum konstanten Amplitudengang weist der Allpass 1. Ordnung einen Phasengang zwischen 0 und $-\pi$ auf, bei dem die Grenzfrequenz (entsprechend der gedachten -3dB Frequenz) nur vom Parameter λ der Übertragungsfunktion abhängt.

In mathematischer Form kann der Phasengang wie folgt beschrieben werden:

$$\arg \{H(e^{j\omega})\} = -\omega - 2 \arctan \left(\frac{\lambda \sin(\omega)}{1 - \lambda \cos(\omega)} \right) \quad (7)$$

bzw. für nicht reell Werte von λ in polarer Form $\lambda = r e^{j\theta}$

$$\arg \{H(e^{j\omega})\} = -\omega - 2 \arctan \left(\frac{r \sin(\omega - \theta)}{1 - r \cos(\omega - \theta)} \right) \quad (8)$$

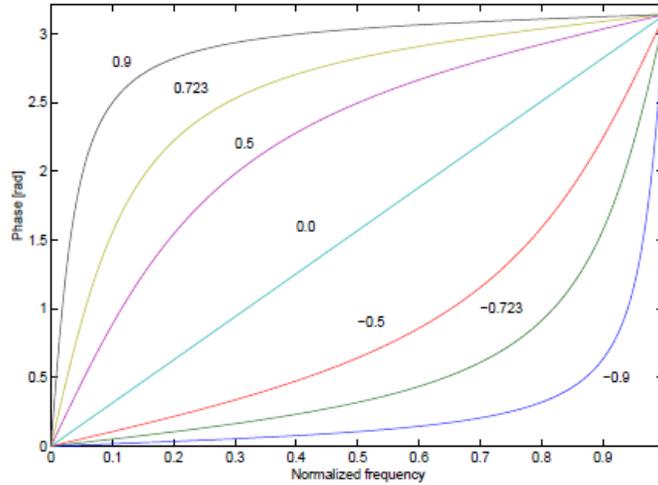


Abbildung 6: Phasengang eines Allpasses 1. Ordnung in Abh. von λ

Die Ableitung des Phasengangs nach der Frequenz ist als Gruppenlaufzeit (eng: *group-delay*) bekannt und ist eine weitere Möglichkeit, die Charakteristik eines Filters zu beschreiben.

$$gpd(\omega) = \frac{-\partial \arg \{H(e^{j\omega})\}}{\partial \omega} \quad (9)$$

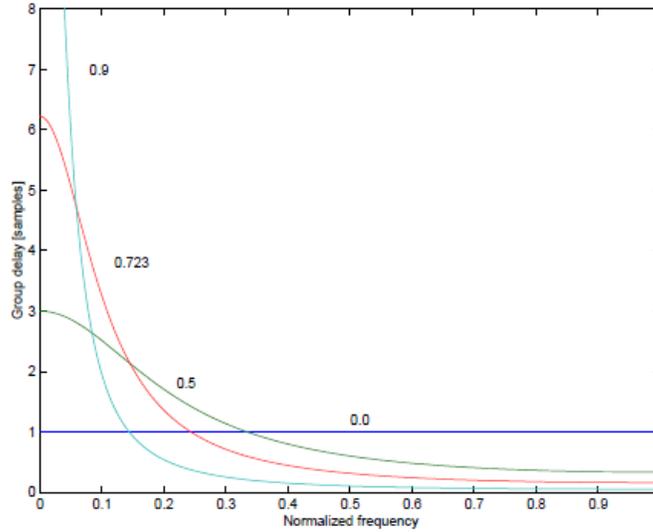


Abbildung 7: Gruppenlaufzeit eines Allpasses 1. Ordnung in Abh. von λ

Wird also $\lambda \neq 0$ gewählt, ist das System nicht mehr linearphasig und weist somit ein frequenzabhängiges Verzögerungsverhalten auf. Die innere Struktur der Signals zerfällt — die zeitliche Beziehung der Signalkomponenten zueinander entspricht nicht mehr jener des Eingangssignals — dies ermöglicht jedoch in weiterer Folge das Konstruieren eines dispersiven Mediums (siehe Kapitel 4.2).

4.1 Bilinear Transformation

Die z-Transformierte $S(z)$ einer Impulsantwort in der Zeitdomäne $s(n)$ sei gegeben durch

$$S(z) = \sum_{n=0}^{\infty} s(n)z^{-n} \quad (10)$$

Die mathematische Grundlage von *warping* stellt das sogenannte *bilineare Mapping* dar. Hierbei wird eine Domäne auf sich selbst durch einen bilinearen Zusammenhang abgebildet:

$$z^{-1} \iff \tilde{z}^{-1} \quad (11)$$

Das Mapping selbst ist wie folgt definiert:

$$\tilde{z}^{-1} = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \quad (12)$$

Folglich kann die Transformation als

$$z^{-1} \longrightarrow \tilde{z}^{-1} = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \quad (13)$$

und die Rücktransformation als

$$\tilde{z}^{-1} \longrightarrow z^{-1} = \frac{\tilde{z}^{-1} + \lambda}{1 + \lambda \tilde{z}^{-1}} \quad (14)$$

angeschrieben werden.

Dies entspricht einer konformen Abbildung von einer Einheitsscheibe auf eine andere. Somit kann man nun die Repräsentation

$$S(z) = \sum_{k=0}^{\infty} w(k) \tilde{z}^{-k} \quad (15)$$

anschreiben, wobei $w(k)$ den Samples der korrespondierenden verzerrten Impulsantwort entspricht.

Die Bedingungen zur Überführung eines Signals aus der z -Domäne in die \tilde{z} -Domäne bzw. vice-versa sind somit mathematisch befriedigt. Zur praktischen Umsetzung bedarf es nun einer Konstruktion, die es ermöglicht, zu k Zeitpunkten Werte für \tilde{z}^{-k} auszulesen.

Dies wird durch eine dispersive Verzögerungskette, wie im folgenden Kapitel 4.1 beschrieben, ermöglicht.

4.2 Dispersive Verzögerungskette

Eine dispersive (frequenzselektive) Verzögerungskette besteht im Grunde aus einer Reihenschaltung von Allpässen 1. Ordnung:

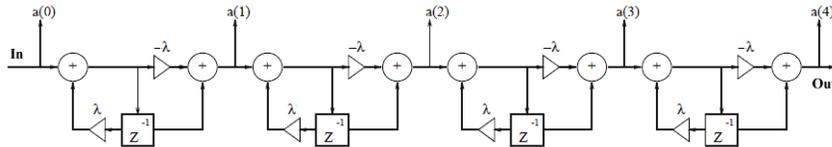


Abbildung 8: Allpasskette in Direktform II

Die in Darstellung 8 gewählte Implementierungsart in Direktform II stellt eine von mehreren Möglichkeiten dar.

Wird nun an den Eingang der Verzögerungskette ein Signal angelegt, so tritt es bei einer Verschaltung von m Allpässen mit dem m -fachen Group Delay behaftet wieder aus — die innere Struktur des Signals ist dabei verloren gegangen, da die Gruppenlaufzeit eine Funktion der Frequenz darstellt.

Während diese Eigenschaft von wenig praktischem Nutzen ist, kann im Inneren der Verzögerungskette folgendes beobachtet werden: durch die nicht-lineare Phase des Allpasses kommt es frequenzabhängig zu verschiedenen Ausbreitungsgeschwindigkeiten innerhalb der Allpasskette — der Vergleich mit einem dispersiven Medium wird also getroffen.

Die mathematische Erklärung für dieses Phänomen liefert die Phasenfunktion des Allpasses 1. Ordnung:

$$\phi(\omega) = \arg \{H(z)\} = -\omega - 2 \arctan \left(\frac{r \sin(\omega - \theta)}{1 - r \cos(\omega - \theta)} \right) \quad (16)$$

Für ein System mit stationärer Phase kann man zusätzlich definieren:

$$\varphi(n, \omega) = \phi_0 + (-\omega n) \quad (17)$$

bzw. die differentielle Definition der Frequenz als

$$\omega = \frac{\partial \varphi(n)}{\partial n} \quad (18)$$

Werden nun mehrere Allpassglieder zu einer Kette (dispersive Verzögerungskette) verknüpft, gelangt man zu einer neuen ortsdiskreten Dimension m , welche die Position entlang der Kette beschreibt. Da wir von $z = e^{j\omega}$ ausgehen und wir nur den Frequenzbereich berücksichtigen wollen, schreiben wir somit für die Funktion der Phase:

$$\tilde{\varphi}(m, n, \omega) = \phi_0 + (-\omega n) + m \arg \{H(\omega)\} \quad (19)$$

Betrachten wir nun wiederum die differentielle Form so ergibt sich aus den Gleichungen 18 und 19

$$\tilde{\omega}(m, n) = \frac{\partial \tilde{\varphi}(m, n)}{\partial m} = \arg \{H(\omega)\} \quad (20)$$

und somit gilt

$$\tilde{\omega} = \arg \{H(z)\} \quad (21)$$

Dies kann nun als mathematische Herleitung dafür betrachtet werden, dass die Phasenfunktion des Allpasses die *warping*-Kurve beschreibt.

Hierbei soll angemerkt werden, dass man korrekter Weise eigentlich von $\tilde{\omega}(\lambda)$ sprechen müsste, da ja der Parameter λ die Form des Arguments beschreibt. Da wir jedoch davon ausgehen, dass λ ein konstanter Wert ist, bleiben wir bei der Form $\tilde{\omega}$, ohne den Hinweis auf die Abhängigkeit von λ .

4.3 Möglichkeiten und Einschränkungen

Da die Phasenfunktion eines Allpasses nur einen Freiheitsgrad aufweist, sind die Möglichkeiten zur Gestaltung einer *warping*-Kurve relativ eingeschränkt. Smith und Abel haben in [11] eine analytische Beschreibung gefunden, um bei gegebener Sampling Frequenz eine möglichst exakte Anpassung an die Bark-Skala zu erreichen:

$$\lambda_{f_s} \approx 1.0674 \left(\frac{2}{\pi} \arctan \left(0.06583 \frac{f_s}{1000} \right) \right)^{1/2} - 0.1916 \quad (22)$$

Ein weiteres Charakteristikum der *warping*-Kurve ist ihre *turning-point-frequency*. Sie bestimmt jenen Punkt, über dem die Frequenzauflösung schlechter wird als im nicht verzerrten Bereich. Sie kann analytisch durch die Form

$$f_{tp} = \pm \frac{f_s}{2\pi} \arccos(\lambda) \quad (23)$$

beschrieben werden. Hier zeigt sich eine direkte Abhängigkeit der *turning-point-frequency* von der Samplerate.

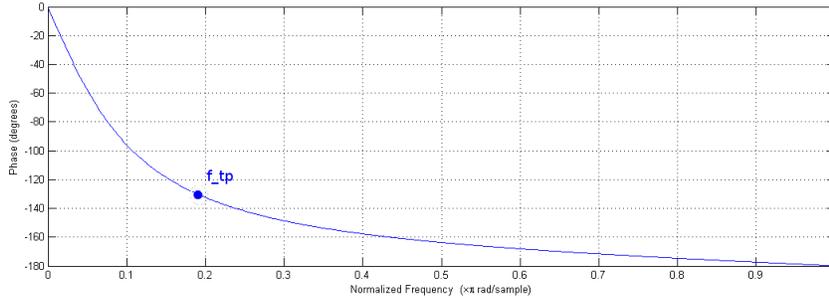


Abbildung 9: f_{tp} in der *warping*-Kurve

f_s [Hz]	f_{tp} [Hz]
44100	~ 5100
22050	~ 2500
11025	~ 1300

Tabelle 1: Beispiele für *turning-point* Frequenzen ($\lambda = 0.75$)

Wie schon aus dem bilinearen Zusammenhang zwischen Ursprungs- und verzerrter Domäne zu erwarten war, gibt es zwei Grundlegende Möglichkeiten um eine Signalanalyse in der \tilde{z} -Domäne durchzuführen:

1. **verzerrte Signale** → **Analyse mit normalen Methoden:**
z.B Warped FFT
2. **normale Signale** → **Analyse mit *warping* Methoden:**
Warped Filter

Methode 1 eignet sich insofern nicht für viele Anwendungen, da nach der Analyse/Bearbeitung ein verzerrtes Signal am Ausgang anliegt und dieses kaum praktisch brauchbar ist. Hier soll nun auch angemerkt werden, dass eine Rücktransformation in einem realen System nicht möglich ist, da sowohl Stabilität als auch Kausalität verloren gehen.

4.4 Warped FFT

Anhand dieses Beispiels wird ersichtlich, dass die Methode — die Fast-Fourier-Transform — ansich unverändert bleibt und das Signal verzerrt wird:

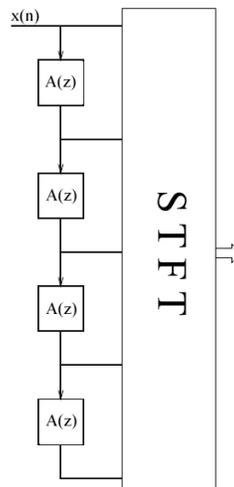


Abbildung 10: Warped FFT bzw. Warped STFT

Methode 2 hingegen eröffnet ein breites Feld an Anwendungsmöglichkeiten. Diese sollen in Kapitel 5 betrachtet werden.

5 Warped Filters

Die fundamentale Einheit aller klassischen Filterstrukturen ist das Einheitsverzögerungsglied. Es weist einen gleichmäßigen Amplitudengang und eine

lineare Phase auf. Dieses wird nun zum Behufe des Warpens durch Allpässe 1. Ordnung ersetzt. Die Implementierung der verschiedenen Filtergattungen kann jedoch nicht — wie folgend erörtert — immer *straight-forward* durchgeführt werden.

5.1 Warped FIR Filter

warping eines FIR Filters (in Folge WFIR genannt) ist sehr einfach — hier reicht das Ersetzen aller Einheitsverzögerungsglieder durch Allpässe aus:

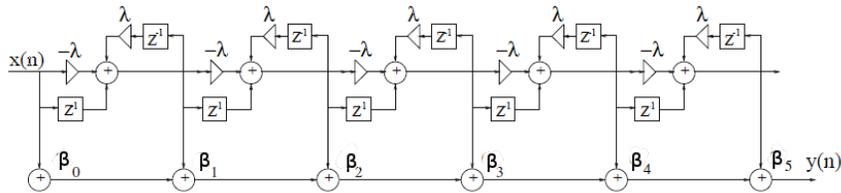


Abbildung 11: WFIR Netzwerk

Eigentlich wäre es nicht korrekt von einem WFIR Filter zu sprechen, da dieser auf Grund seiner Rückkopplungen in den einzelnen Allpässen eine unendliche Impulsantwort aufweist. Aufgrund der Ähnlichkeit bzw. der Abstammung der Filterstruktur wird dieser Terminus jedoch beibehalten.

Folgende nützliche Eigenschaft lässt sich ebenfalls erkennen: der Unterschied zwischen den von Schüssler in [10] vorgestellten Laguerre-Filtern und einem verzerrten FIR Filter ist marginal: es erfolgt nur eine zusätzliche Vorfilterung vor dem ersten Allpass durchgeführt, was — so Härmä in [7] — in der Praxis einen vernachlässigbaren Einfluss auf das Ergebnis hat. Der Vorteil der nahen Beziehung von verzerrten Filtern und Laguerre-Filter ist, dass alle Techniken, die für Laguerre-Filtern entwickelt wurden, auch für verzerrte Filter anwendbar sein müssten.

Zur Bestimmung der Filterkoeffizienten β gibt es entsprechenden den vorgegangenen Überlegungen zur Bilinear Transformation zwei Möglichkeiten.

Zum einen lässt sich die Impulsantwort mit gewöhnlichen Methoden bestimmen, unter der Voraussetzung, dass der Frequenzgang in der verzerrten Domäne vorliegt — z.B. der Bark-Domäne.

Zum anderen besteht die Möglichkeit, aus einer bekannten Impulsantwort eines konventionellen FIR Filters eine verzerrte Impulsantwort abzu-

leiten. Es gilt also

$$H(z) = \sum_{k=0}^{\infty} \tilde{h}(k) \tilde{z}^{-k} \quad (24)$$

und äquivalent

$$\tilde{H}(\tilde{z}) = \sum_{n=0}^{\infty} h(n) z^{-n} \quad (25)$$

In Worten formuliert bedeutet dies:

- verzerrter Filter mit normaler Impulsantwort \rightarrow Filterung in verzerrter Domäne
- verzerrter Filter mit verzerrter Impulsantwort \rightarrow Filterung in normaler Domäne

5.2 Warped IIR Filter

Im Unterschied zur Konversion eines FIR in einen WFIR Filter kann diese bei einem IIR Filter nicht *straight-forward* durchgeführt werden. Es ergibt sich hierbei vor allem ein Problem im Rückkopplungszweig: ein Pfad wird ohne Verzögerung rückgekoppelt (*delay-free-loops*).

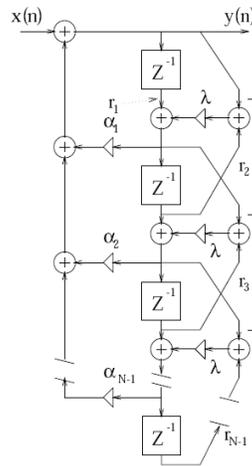


Abbildung 12: WIIR Netzwerk mit delay-free-loops

Dies führt zu einem mathematischen Problem — würde man eine Differenzgleichung in der Zeit-Domäne definieren so sähe diese wie folgt aus:

$$y[n] = x[n] + \alpha_1 d_1[y[n]] + \alpha_2 d_2[y[n]] \dots + \alpha_N d_N[y[n]] \quad (26)$$

Der Term d_k wird dabei definiert als *generalized shift operator*

$$d_k[x(n)] \equiv \underbrace{h(n) * h(n) * \dots * h(n)}_{k\text{-fache Faltung}} * x(n) \quad (27)$$

wobei $h(n)$ der Impulsantwort des Allpasses entspricht.

Die Berechnung der Faktoren α_0 bis α_N sei hier vorerst nicht weiter von Bedeutung.

Offensichtlich problematisch ist jedoch die Rekursion in der Differenzengleichung da wir annehmen

$$d_0[y(n)] = y(n) \quad (28)$$

wodurch $y[n]$ in eine direkte Abhängigkeit von sich selbst gerät (Vgl. Gl. 26). Dies kann entweder durch ein zweistufiges Verfahren oder durch Modifikation der Filterstruktur selbst beseitigt werden.

Eine zweistufige Berechnung wäre zwar im Ansatz einfacher, bedeutet jedoch einen doppelt so hohen Berechnungsaufwand, was diese Alternative weniger attraktiv erscheinen lässt. Die Modifikation erfolgt also durch eine Umwandlung der α s in σ s und einen gain-Faktor g . Der Algorithmus dazu wird von Härmä in [8] erläutert.

Der Term $\frac{1}{g}$ kann auch in [7] nachgelesen werden. Er wird dort als $\frac{1}{1-\chi}$ angeführt und als 'Rückkopplungsfaktor' bezeichnet (in Abbildung 13 gilt folglich die Entsprechung $g = 1 - \chi$).

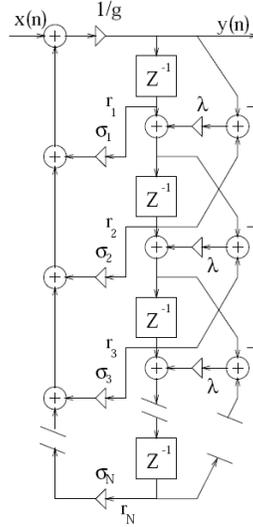


Abbildung 13: WIIR Netzwerk ohne delay-free-loops

Die Entwurfsmöglichkeiten für einen WIIR Filter entsprechen jenen des WFIR Filters.

6 Warped LPC

Linear Predictive Coding ist eine gängige Methode, um mit wenig rechen-technischem Aufwand eine verhältnismäßig gute Schätzung der spektralen Einhüllenden eines Signals zu erhalten. Sie wird vor allem in der Mobilfunk-technologie verwendet um dort eine Modellierung von Quelle und Filter durchzuführen.

Der Grundgedanke von LPC lautet: Vorhersage durch Linearkombinati-on von vergangenen Werten oder

$$x[n] = Gu[n] - \sum_{i=1}^k a_i x[n - i] \quad (29)$$

In der z-Domäne liefert dies den Analysefilter $A(z)$

$$A(z) = 1 + \sum_{k=1}^p \alpha_k z^{-k} \quad (30)$$

Der daraus resultierende, inverse Synthesefilter $H(z)$ kann angeschrieben werden als

$$H(z) = \frac{G}{1 + \sum_{k=1}^p \alpha_k z^{-k}} = \frac{G}{A(z)} \quad (31)$$

Der Summenterm kann dabei als Schätzer $\hat{x}[n]$ bezeichnet werden.

Wie schon zuvor bei den Filterstrukturen kann hier eine äquidistante Verteilung der Frequenzen über das gesamte Spektrum beobachtet werden. Aus den bekannten Gründen möchte man diese nun in der verzerrten Domäne betrachten. Man schreibt also

$$\hat{x}[n] = \sum_{k=1}^N \alpha_k d_k[x[n]] \quad (32)$$

Da es sich beim Prädiktor im Wesentlichen um einen FIR/WFIR Filter handelt kann man ihn durch konventionelle Adaptionsalgorithmen so tunen, dass der Fehler minimal wird. Dieser ist definiert als

$$e[n] = x[n] - \hat{x}[n] \quad (33)$$

Zur Minimierung bedient man sich des quadratischen Fehlers eines stochastischen Prozesses:

$$e = E \left\{ \left\| x[n] - \sum_{k=1}^N \alpha_k d_k[x[n]] \right\|^2 \right\} \quad (34)$$

Dies stellt wiederum ein klassisches MMSE Szenario dar dessen Optimierung durch Lösung der Wiener-Hopf Gleichung mittels Autokorrelationsmethode durchgeführt werden kann: aus der Autokorrelationsmatrix und der Signalwechsellistung (σ^2) können durch Lösung des Gleichungssystems die Koeffizienten des Analysefilters bestimmt werden. Ein dafür gängiger und effektiver Algorithmus wird von Levinson und Durbin in [5] beschrieben. Dieser basiert auf der Ausnutzung einer Rekursion innerhalb der Töplitzstruktur einer Autokorrelationsmatrix.

$$\mathbf{R}_{\mathbf{xx}} = \begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(N-1) & r(N-2) & \cdots & r(0) \end{bmatrix},$$

$$\mathbf{a} = \begin{bmatrix} a(0) \\ a(1) \\ \vdots \\ a(N-1) \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} P_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\mathbf{R}_{xx} \times \mathbf{a} = \mathbf{p} \quad (35)$$

$$\underbrace{\begin{bmatrix} r(0) & r(1) & \cdots & r(N-1) \\ r(1) & r(0) & \cdots & r(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(N-1) & r(N-2) & \cdots & r(0) \end{bmatrix}}_{\mathbf{R}_{xx}} \times \underbrace{\begin{bmatrix} a(0) \\ a(1) \\ \vdots \\ a(N-1) \end{bmatrix}}_{\mathbf{a}} = \underbrace{\begin{bmatrix} P_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{p}}$$

Voraussetzung dafür ist nun die Möglichkeit, die Autokorrelation eines Signals in der verzerrten Domäne zu betrachten.

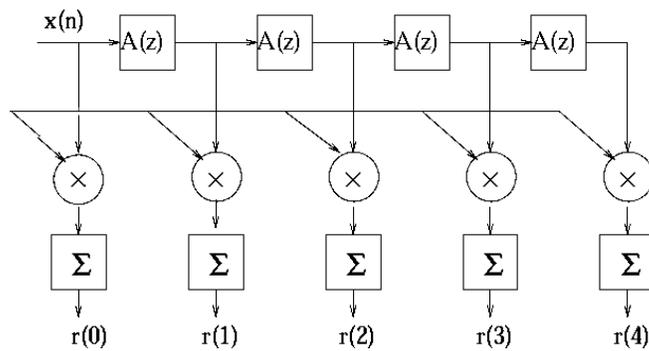


Abbildung 14: verzerrtes Autokorrelationsnetzwerk

Eine Autokorrelation N-ter Ordnung liefert somit N Werte der Korrelation und die Wechselleistung P_1 an der Stelle 0 (Entspricht σ^2). Wird nun der Levinson-Durbin Algorithmus darauf angewandt, so ergeben sich N-1

Werte, die direkt als Filterkoeffizienten des Synthesefilters eingesetzt werden können.

6.1 Ergebnisse

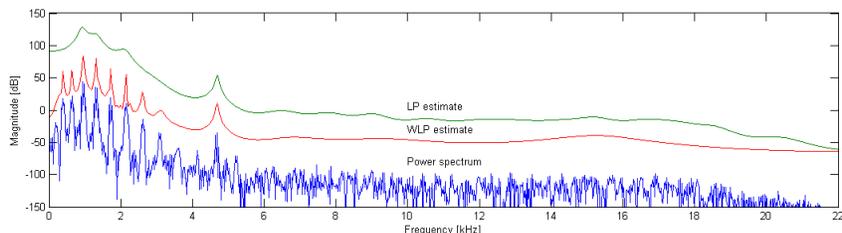


Abbildung 15: Vergleich WLPC/LPC

Wie aus Abbildung 15 sehr gut zu erkennen ist, zeigt der WLPC eine wesentlich bessere Auflösung im tieffrequenten Bereich und vernachlässigt 'uninteressante' Peaks im Bereich ab ca. 5kHz.

Zur Synthese eignen sich sowohl das Residuum als auch weißes Rauschen, dass schon bei sehr geringer Ordnung erstaunlich gute Ergebnisse liefert.

In der Praxis werden erfolgreich niedrige Abtastraten (8kHz bzw. 16 kHz in der Mobiltelefonie und Sprachverarbeitung) eingesetzt, um der hohen Auflösung des Gehörs bei tiefen Frequenzen gerecht zu werden). Für breitbandige Signalübertragung kann Frequency-Warping eine günstige Alternative dazu sein.

Eine notwendige Voraussetzung für effizientes *warping* ist daher eine verhältnismäßig hohe Samplingrate — die Praxis hat gezeigt, dass diese zumindest 22.05kHz betragen sollte (Vgl. Tabelle 1) um *warping* im auditorisch relevanten Bereich überhaupt durchführen zu können. Dies kann unter Umständen den Datenaufwand — zumindest während der Berechnung/Filterung — recht stark erhöhen.

Im Speziellen kann man jedoch feststellen, dass der WLPC ab Ordnung 10 bei Synthese mit weißem Rauschen sehr klar verständliche Sprache reproduzieren kann während der klassische LPC nur schwach moduliertes Rauschen am Ausgang liefert.

Problematisch kann jedoch auch eine Überbestimmung werden, da dann im auditorisch relevanten Bereich unter Umständen zu viele Pole detektiert werden, die dann das Ergebnis verfälschen.

7 Zusammenfassung

Die Anpassung der Frequenzachse an die menschliche Hörwahrnehmung bietet mit relativ geringen *computational costs* ein enormes Potential an Datenreduktion. Durch die Modifizierung der bekannten Filterstrukturen lassen sich herkömmlich Designmethoden (mit Vorbehalt) übernehmen und erstaunliche Ergebnisse erzielen.

Obwohl die Rechenleistung ständig steigt, bietet *warping* eine hervorragende Alternative zu Methoden wie QMF oder Constant-Q Transformation, da sie die Problematik der linearen Frequenzachse konventioneller DSP Systeme an der Wurzel ergreift und somit vor allem im mobilen Sektor, wo Rechen- und Speicherleistung noch in relativ beschränktem Ausmaß vorhanden sind, mit Sicherheit in vielen Sektoren Anwendung finden kann.

Literatur

- [1] John R. Buck Alan V. Oppenheim, Roland W. Schafer. *Zeitdiskrete Signalverarbeitung*. Pearson Studium, 2004.
- [2] John R. Buck Alan V. Oppenheim, Roland W. Schafer. *Zeitdiskrete Signalverarbeitung*, chapter 5.5 Allpässe. Pearson Studium, 2004.
- [3] C. Braccini and A. V. Oppenheim. Unequal bandwidth spectral analysis using digital frequency warping. *IEEE Transactions on Speech and Audio Processing*, ASSP-22:233–245, 1972.
- [4] J. Brown and M. Puckette. An efficient algorithm for calculation of a constant q transform. *J. Audio Engineering Society*, 92(5), November 1992.
- [5] J. Durbin. The fitting of time series models. *Rev. Inst. Int. Stat.*, 28(233-243), 1960.
- [6] H. Fastl E. Zwicker. *Psychoacoustics: Facts and Models*. Springer, Berlin, 1998.
- [7] Aki Härmä et al. Frequency-warped signal processing for audio application. *J. Audio Engineering Society*, 48(11), November 2000.
- [8] Aki Härmä. Implementation of recursive filters having delay free loops. *IEEE*, 1998.

- [9] R. D. Patterson. Auditory filter shapes derived with noised stimuli. *J. Acoust. Soc. Am.*, 59:640–654, 1976.
- [10] W. Schüssler. *Variable Digital Filters*, pages 524–525. Arch. Elek. Übertragung, 1970.
- [11] J. O. Smith and J. S. Abel. Bark and erb bilinear transform. *IEEE Transactions on Speech and Audio Processing*, 7:697–708, November 1999.
- [12] H. Traunmüller. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.*, 88:97–100, 1999.
- [13] Udo Zölzer. *DAFX*. Wiley and Sons, Feb. 2002.