

# **Detektion und Unterdrückung von Rückkopplungen in Lautsprecher-Mikrofon Anordnungen**

Diplomarbeit

durchgeführt von

**Henning Petersen**

Institut für Elektronische Musik und Akustik  
der Universität für Musik und darstellende Kunst, Graz

Leiter: o. Univ.- Prof. Mag. DI. Dr. Robert Höldrich

Begutachter: o. Univ.- Prof. Mag. DI. Dr. Robert Höldrich

Betreuer: DI Markus Noisternig

Graz, im Januar 2006

## Zusammenfassung

Diese Arbeit zeigt, wie Rückkoppelfrequenzen in Lautsprecher-Mikrofon Anordnungen unterdrückt werden können. Es werden sowohl traditionelle Methoden zur Stabilisierung elektroakustischer Schleifen, wie Frequenzschieber oder *pitch shifter*, als auch verschiedene Detektionsmethoden für Rückkoppelfrequenzen, die eine Bedämpfung dieser mittels Kerbfiltern ermöglichen, vorgestellt. Die unterschiedlichen Detektionsmethoden werden getestet und hinsichtlich Detektionsgeschwindigkeit und Effizienz vergleichend gegenübergestellt. Darüber hinaus erfolgt der Entwurf einer fixen und adaptiven Filterbank, sowie deren Ansteuerheuristik. Die Einbindung in das Gesamtsystem erlaubt eine effiziente Unterdrückung von Rückkoppelfrequenzen und führt zu einer Erhöhung der Koppel- bzw. Stabilitätsgrenze. Das vorgestellte System ist echtzeitfähig und wird in Pure Data<sup>1</sup> implementiert. Abschließend wird das implementierte Gesamtsystem getestet und bewertet.

## Abstract

This work shows how howling frequencies in loudspeaker microphone setups can be suppressed. Besides traditional stabilization methods for electroacoustic loops such as frequency shifters or pitch shifters, methods for detecting and suppressing howling using notch filter techniques are demonstrated. The different detection methods are tested and compared to each other concerning their detection speed and efficiency. Furthermore the design of a fixed and adaptive filter bank and its filter tracking is sketched. The integration in a total system allows efficient howling suppression and an increase of the maximum gain of the loudspeaker microphone setup. The presented system is a real-time implementation programmed in Pure Data. In conclusion, the entire implemented system is tested and evaluated.

---

1) Pure Data von Miller Puckette ist eine grafisch orientierte Computermusik-Programmierungsumgebung zur Echtzeitsignalverarbeitung.

## Danksagung

An dieser Stelle möchte ich all jenen danken, die durch ihre fachliche und persönliche Unterstützung zum Gelingen dieser Diplomarbeit beigetragen haben.

Speziell möchte ich mich bei Herrn o. Univ.- Prof. Mag. DI. Dr. Robert Höldrich, Herrn DI Markus Noisternig und Herrn Thomas Musil für die kompetente Betreuung meiner Diplomarbeit und die zahlreichen wissenschaftlichen Ratschläge bedanken, welche stets zur Verbesserung der Arbeit beigetragen haben.

Besonderer Dank gebührt meinen Eltern, die mir dieses Studium durch ihre großzügige Unterstützung ermöglicht haben.

Außerdem danke ich meiner Freundin Nora für ihre Geduld und ihre moralische Unterstützung.

# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einleitung</b>  | <b>6</b>  |
| 1.1      | Gliederung der Arbeit .....  | 6         |
| 1.2      | Zielsetzung .....  | 7         |
| <b>2</b> | <b>Traditionelle Methoden zur Stabilisierung elektroakustischer Schleifen</b>          | <b>8</b>  |
| 2.1      | Adaptive Line Enhancement .....  | 8         |
| 2.1.1    | FIR Struktur .....   | 9         |
| 2.1.2    | IIR Struktur .....   | 11        |
| 2.1.3    | Vergleich der beiden Strukturen .....  | 13        |
| 2.2      | Adaptiver Störunterdrücker als Kerbfilter nach Widrow/Stearns .....                    | 14        |
| 2.3      | Frequenzschieber .....   | 17        |
| 2.3.1    | Hilbert Filter .....   | 18        |
| 2.4      | Pitch Shifter .....  | 21        |
| <b>3</b> | <b>Präzise Fourier-Analyse mittels Signalableitungen</b>                               | <b>23</b> |
| 3.1      | FT <sup>n</sup> : Fourier Transformation unter Verwendung von Signal Ableitungen ..... | 23        |
| 3.2      | DFT <sup>1</sup> : Die Diskrete FT <sup>1</sup> .....                                  | 25        |
| 3.3      | Der Analyse Algorithmus .....  | 27        |
| 3.3.1    | Beispiel für ein frequenzmoduliertes Signal .....                                      | 28        |
| <b>4</b> | <b>Detektion von Rückkoppelfrequenzen</b>  | <b>29</b> |
| 4.1      | Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins .....             | 29        |
| 4.1.1    | Glättung .....   | 30        |
| 4.1.2    | Zeitkonstanten bzw. Verzögerung .....  | 32        |
| 4.1.3    | Untersuchung der benachbarten Frequenz-Bins .....                                      | 34        |
| 4.1.4    | Fensterung .....   | 36        |
| 4.1.5    | Fehldetektionen .....  | 37        |
| 4.1.6    | „Drop-out“ Blöcke .....  | 39        |
| 4.1.7    | Detektion von mehreren Rückkoppelfrequenzen im Testsignal .....                        | 40        |
| 4.1.8    | Einfluss der Steuerparameter auf die Detektionszeit .....                              | 43        |
| 4.2      | Spektrale Flachheit (SFM-Methode) .....  | 48        |
| 4.2.1    | Glättung des Spektrums .....   | 52        |
| 4.2.2    | Nullsetzen der größten drei Frequenz-Bins - Berechnung von P <sub>Sub</sub> .....      | 54        |
| 4.3      | Vergleich der beiden Detektionsmethoden .....  | 55        |

---

|          |   |           |
|----------|---|-----------|
| <b>5</b> | <b>Filter- und Filterbankentwurf</b>                                      | <b>62</b> |
| 5.1      | Parametrisches Kerbfilter .....   | 62        |
| 5.2      | Berechnung der Half Transition Bandbreite .....                           | 66        |
| 5.3      | Filterbankentwurf .....   | 69        |
| 5.3.1    | Terzfilterbank .....  | 69        |
| 5.3.2    | Parametrische adaptive Filterbank .....                                   | 72        |
| <b>6</b> | <b>Gesamtsystem</b>   | <b>74</b> |
| 6.1      | Tracking .....  | 74        |
| 6.2      | Nachklingen der Koppelfrequenzen .....                                    | 75        |
| 6.3      | Glättung und Spitzenwert-Detektion .....                                  | 76        |
| 6.4      | Totzeit und adaptive Dämpfung .....                                       | 77        |
| 6.5      | Zusammenfassung .....   | 78        |
| <b>7</b> | <b>Evaluierung des Gesamtsystems</b>                                      | <b>79</b> |
| 7.1      | Erhöhung der Stabilitätsgrenze .....                                      | 80        |
| 7.2      | Bewertung der Sprach- und Klangqualität im Bereich der Koppelgrenze ..... | 81        |
| <b>8</b> | <b>Ausblick</b>   | <b>83</b> |
| 8.1      | Langzeitheuristik .....   | 83        |
| 8.2      | Reduzierung der Ausgangsverstärkung .....                                 | 83        |
| <b>9</b> | <b>Literaturverzeichnis</b>   | <b>84</b> |

---

# 1 Einleitung

Rückkopplungsfrequenzen entstehen durch den Rückkopplungspfad bei Lautsprecher-Mikrofon Anordnungen. Im ungünstigsten Fall kann eine derartige Rückkopplung zu einer Instabilität des Gesamtsystems und einem damit verbundenen lauten Heulen (engl. howling) führen. Dies geschieht, wenn ein ausreichend hoher Anteil des Lautsprechersignals vom Mikrofon wieder aufgenommen wird und in der Rückkopplungsschleife zirkuliert (siehe Abbildung 1-1). Wenn die Lautsprecher-Mikrofon Anordnung mit einer Frequenz gespeist wird, deren Schleifenverstärkung  $g$  größer als eins ist, dann wird das Signal bei jedem Schleifendurchgang verstärkt, und es entstehen Rückkopplungsfrequenzen. In der Praxis hängt das Rückkoppelverhalten von der Nichtlinearität des elektronischen Systems ab. Es existieren viele Ansätze die das Auftreten solcher Koppel frequenzen verhindern. Einige davon werden in dieser Arbeit vorgestellt und implementiert. Für die Simulationen und Darstellungen wird Matlab verwendet, die Echtzeitimplementation erfolgt in Pure Data.

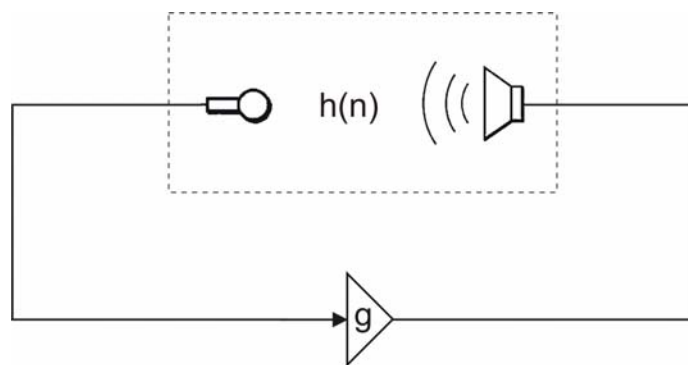


Abbildung 1-1 akustische Rückkopplungsschleife

## 1.1 Gliederung der Arbeit

Zunächst werden einige Methoden zur Stabilisierung elektroakustischer Schleifen vorgestellt. Im Detail werden der *adaptive line enhancer*, der eigentlich zu den Geräuschunterdrückungsfiltern gehört, und ein adaptives Filter zur Störgeräuschunterdrückung, das von B. Widrow et al. entwickelt wurde, präsentiert. Außerdem wird gezeigt, wie ein *Frequenzschieber* bzw. ein *pitch shifter* eingesetzt werden kann, um die Koppelgrenze der Lautsprecher-Mikrofon Anordnung zu erhöhen.

In Kapitel 3 wird eine Methode erläutert, mit der es möglich ist, die nahezu exakte Frequenz einer Rückkopplung durch eine spezielle Variante der Kurzzeit-Fourier-Analyse zu ermitteln. Diese Frequenz bietet bei Einsatz von Kerbfiltern zur Unterdrückung von Koppelfrequenzen die Möglichkeit der exakten Positionierung dieser Filter, und somit eine optimale Dämpfung des gewünschten Frequenzbereichs bei maximaler Filtergüte.

Kapitel 4 stellt zwei Methoden zur Detektion von Rückkoppelfrequenzen vor, die der Ansteuerung der Kerbfilterbank dienen um mehrere Koppelfrequenzen zu unterdrücken. Beide Methoden werden hinsichtlich ihrer Geschwindigkeit und ihrer Effizienz getestet und miteinander verglichen.

Kapitel 5 beschäftigt sich mit dem Entwurf von Kerbfiltern und Filterbänken. In Kapitel 6 wird die Ansteuerung einer adaptiven Filterbank mit begrenzter Anzahl von Filtern beschrieben, die es ermöglicht, das gesamte Frequenzband abzudecken.

Das implementierte Gesamtsystem wird in Kapitel 7 hinsichtlich seiner Effizienz getestet und evaluiert. Kapitel 8 bietet einen kurzen Ausblick und Vorschläge zur weiteren Verbesserung des Gesamtsystems

## 1.2 Zielsetzung

Das Ziel dieser Arbeit ist der Entwurf eines Gesamtsystems zur automatisierten Detektion und Unterdrückung störender Rückkoppelfrequenzen mit Hilfe adaptiver Filterbänke. Die Koppel- bzw. Stabilitätsgrenze einer Lautsprecher-Mikrofonanordnung soll nach oben verschoben, und somit ein Lautheitsgewinn erzielt werden. Dabei soll die Signalqualität so wenig wie möglich beeinträchtigt werden. Das System soll in ein am IEM entwickeltes Gesamtsystem eingebettet werden, in dem AEC (Acoustic Echo Cancellation) vorgeschaltet ist. Die Algorithmen zur AEC werden in dieser Arbeit nicht betrachtet. Mit dem vorgestellten System sollen die verbleibenden Rückkoppelfrequenzen (residual feedback) unterdrückt werden, die vom AEC-System nicht ausgelöscht worden sind, da der Lautsprecher-Mikrofonpfad meist nicht exakt nachgebildet werden kann. Dies soll ohne Schätzung des Übertragungspfades geschehen. Das System soll echtzeitfähig sein und in Pure Data implementiert werden.

## 2 Traditionelle Methoden zur Stabilisierung elektroakustischer Schleifen

### 2.1 Adaptive Line Enhancement

*Adaptive Line Enhancer* dienen ursprünglich dem Zweck, Sinussignale, die durch breitbandiges Rauschen gestört sind, anzuheben. Vertauscht man die Rollen von Rauschen und Sinussignal, so können *Line Enhancement* Algorithmen auch zur Unterdrückung sinusförmiger (schmalbandiger) Störungen unbekannter Frequenz eingesetzt werden [4].

Das im folgenden beschriebene Verfahren ist somit den Algorithmen der adaptiven Geräuschunterdrückung zuzuordnen. Das Konzept ist in Abbildung 2-1 zu sehen.

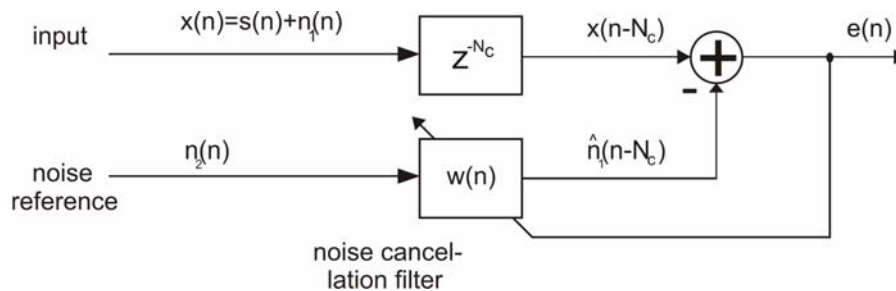


Abbildung 2-1 Schema eines adaptiven Geräuschunterdrückungs-Filter

Ein Signal  $s(n)$  wird vom Sensor aufgenommen. Neben diesem Nutzsignal nimmt der Sensor zusätzlich Hintergrundrauschen  $n_1(n)$  auf. Ein zweiter Sensor empfängt das Rauschen  $n_2(n)$ , das als Referenzsignal für das *Geräuschunterdrückungs-Filter* dient. Voraussetzung für dieses Konzept ist, dass  $n_1(n)$  linear mit  $n_2(n)$  korreliert, jedoch weder  $n_1(n)$  noch  $n_2(n)$  mit  $s(n)$  korrelieren.

Falls die Korrelation zwischen  $n_1(n)$  und  $n_2(n)$  auch akausale Anteile enthält, muss das Signal im Primärpfad um  $N_c$  Abtastwerte verzögert werden. Dieser Fall tritt dann auf, wenn sich der

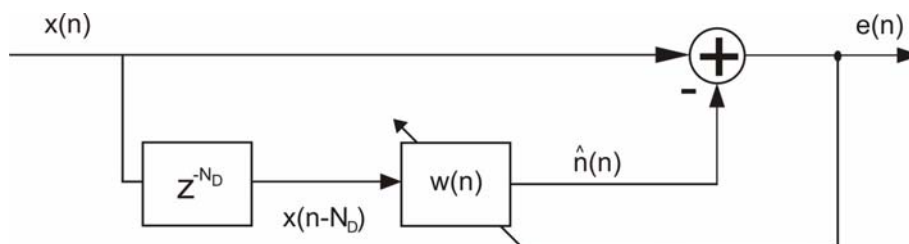


Referenzsensor näher an der Störquelle befindet als der Primärsensor. Das Filter  $w(n)$  wird nun so adaptiert, dass sein Ausgang  $\hat{n}_1(n - N_c)$  möglichst genau  $n_1(n - N_c)$  nachbildet. Durch Subtraktion dieses Ausgangssignals vom verzögerten Primärsignal wird Geräuschreduktion erreicht.

Falls kein Referenzsignal für das *Geräuschunterdrückungs-Filter* zur Verfügung steht, so muss dieses entweder künstlich generiert, oder aus dem Primärsignal gewonnen werden. Unter der Voraussetzung einer periodischen Störung kann eine verzögerte Version des Primärsignals als Referenzsignal dienen. Diese Verzögerung muss so eingestellt werden, dass die Signalkomponenten im Primärpfad nicht mehr mit jenen im Referenzpfad korrelieren. Aufgrund der Periodizität des Störsignals bleibt die Korrelation der Störkomponenten in Primär- und Referenzpfad jedoch bestehen. Die Einsatzmöglichkeiten für dieses Modell sind Beschallungssysteme, Hörgeräte und andere Systeme in denen geschlossene akustische Schleifen existieren können. In [4] wird für die zu wählende Verzögerung ein Wert von 1-2 ms vorgeschlagen um die Kurzzeitkorrelation von Sprache ausreichend zu verringern. Bei einer Abtastrate von 44,1 kHz entspricht dies 44-88 Samples. Bei der in der Sprachsignalverarbeitung üblichen Abtastrate von 8 kHz würde dies 8-16 Samples entsprechen. Aufgrund der periodischen Anteile von Sprachsignalen sollte das Gedächtnis des Filters die Periodendauer der Grundfrequenz des Sprachsignals nicht übersteigen. Andernfalls würden auch periodische Anteile des Sprachsignals unterdrückt werden. Bei einer Abtastrate von 8 kHz dürfte das Filter nicht mehr als 40-60 Koeffizienten enthalten.

### 2.1.1 FIR Struktur

Abbildung 2-2 zeigt das Grundprinzip eines *adaptive line enhancers* mit FIR-Struktur.



**Abbildung 2-2** adaptive line enhancement Filter mit FIR Struktur

In den meisten Anwendungen wird zur Adaptierung des Filters ein normierter LMS-Algorithmus (NLMS-Algorithmus) mit einer Schrittweite von  $\mu \in \{0.01, 0.00001\}$  verwendet. Diese kleinen Schrittweiten werden deshalb gewählt, weil der Adaptionsprozess relativ langsam erfolgen soll, damit nur die periodischen Anteile des Signals unterdrückt werden. Anstelle des NLMS-Algorithmus können natürlich auch andere Adaptionsalgorithmen wie zum Beispiel der RLS-Algorithmus verwendet werden, der aber für diese Anwendung aufgrund seiner hohen Rechenintensität und des schnellen Adaptionsprozesses nicht sinnvoll wäre.

Mit dem Eingangsvektor

$$\mathbf{x}(n - N_D) = [x(n - N_D), x(n - N_D - 1), \dots, x(n - N_D - N + 1)]^T \quad [2.1]$$

und dem Vektor der Filterkoeffizienten

$$\mathbf{w}(n) = [w_0(n), w_1(n), \dots, w_{N-1}(n)]^T \quad [2.2]$$

ergibt sich der Ausgang des adaptiven Filters zu:

$$\hat{n}(n) = \mathbf{w}(n)^T \mathbf{x}(n - N_D) = \mathbf{x}^T(n - N_D) \mathbf{w}(n). \quad [2.3]$$

Durch Subtraktion des geschätzten Störsignals  $\hat{n}(n)$  vom primären Eingangssignal  $x(n)$  wird das Ausgangssignal

$$e(n) = x(n) - \hat{n}(n) \quad [2.4]$$

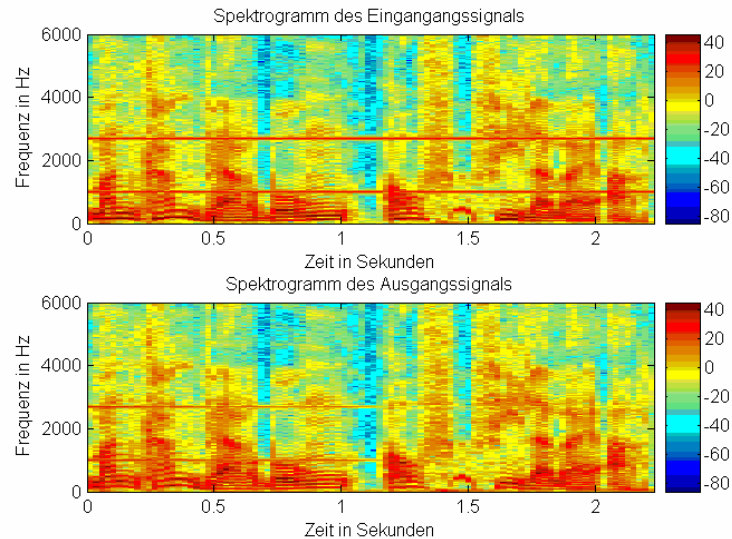
erzeugt, das gleichzeitig als Fehlersignal für den NLMS Adaptionprozess dient:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \frac{\mathbf{x}(n - N_D) \cdot e(n)}{\|\mathbf{x}(n - N_D)\|^2 + \beta}. \quad [2.5]$$

$\beta \ll 1$  dient zu Regularisierung und verhindert, dass der Nenner in [2.5] Null werden kann.

Aufgrund der periodischen Komponenten des Sprachsignals versucht das Filter auch Teile des Sprachsignals zu unterdrücken. Durch die Wahl einer kleinen Schrittweite  $\mu$  kann dieses Verhalten verhindert werden und es werden nur periodische Störungen, die für längere Zeit präsent sind eliminiert. Andererseits führt eine kleine Schrittweite zu einer langsamen Konvergenz, so dass plötzlich auftretende periodische Störungen erst nach sehr langer Zeit abgeschwächt werden. Aus diesen Gründen gilt es einen guten Kompromiss für die Schrittweite zu finden.

Abbildung 2-3 zeigt ein Beispiel für *adaptive line enhancement*. Als Eingangssignal dient hier ein Sprachsignal, das von zwei Sinussignalen mit den Frequenzen  $f_1=1000\text{Hz}$  und  $f_2=2700\text{Hz}$  gestört ist (siehe Abbildung 2-3 oben). Es ist deutlich zu sehen wie die beiden Störfrequenzen nach kurzer Adaptionzeit (ca. 1 Sekunde) unterdrückt werden. Siehe Abbildung 2-3 unten.

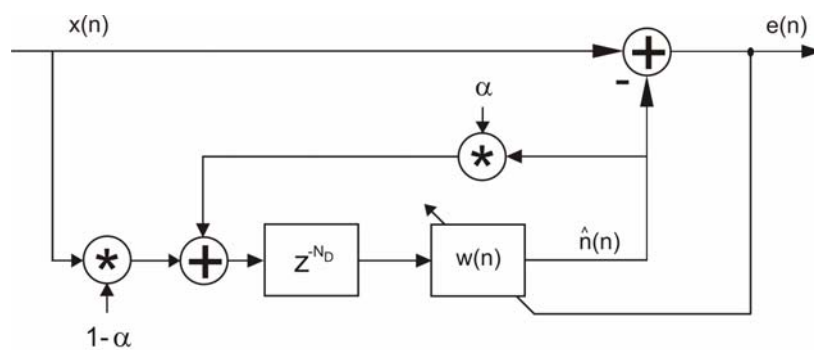


**Abbildung 2-3** Zeit-Frequenz-Analyse des Eingangs- bzw. Ausgangssignals des *adaptiven line enhancers*

Die Abtastfrequenz  $f_s$  beträgt für dieses Beispiel 44,1 kHz. Das Eingangssignal wird für den Referenzpfad um  $N_D=75$  Samples verzögert, die Anzahl der Filterkoeffizienten ist  $N=250$  und die Schrittweite für die Adaption wird auf  $\mu=0,0014$  eingestellt.

## 2.1.2 IIR Struktur

Mit der IIR-Struktur des *adaptive line enhancers* ist es möglich sehr schmale Kerben für das adaptive Filter zu erreichen. Ein Nachteil dieser Struktur ist, dass das Filter instabil werden kann, wenn der Adaptionprozess die Pole des Filters aus dem Einheitskreis der  $z$ -Domäne drängt. Abbildung 2-4 zeigt das Schema des *adaptiven line enhancers* mit IIR-Struktur.



**Abbildung 2-4** adaptives line enhancement Filter mit IIR Struktur

Um die Vorteile dieser Struktur zu verdeutlichen wird die Simulation wie unter 2.1.1 wiederholt.  $\alpha$  hat hierbei den Wert 0,7. Auch hier wird der NLMS-Algorithmus als Adaptionsalgorithmus für die Filterkoeffizienten  $w_i(n)$  verwendet. Außerdem wird folgende einfache *leakage-Kontrolle* eingeführt um zu vermeiden, dass das Filter instabil wird:

$$\mathbf{w}(n+1) = \lambda(n)\mathbf{w}(n) + \mu \frac{\tilde{\mathbf{x}}(n-N_D) \cdot e(n)}{\left\| \tilde{\mathbf{x}}(n-N_D) \right\|^2 + \beta}. \quad [2.6]$$

Der Anregungsvektor  $\tilde{\mathbf{x}}(n-N_D)$  ist nun eine Mischung aus dem Primärsignal  $x(n)$  und dem geschätzten Störsignal  $\hat{n}(n)$ :

$$\tilde{\mathbf{x}}(n-N_D) = [\tilde{x}(n-N_D), \tilde{x}(n-N_D-1), \dots, \tilde{x}(n-N_D-N+1)]^T$$

mit  $\tilde{x}(n) = (1-\alpha)x(n) + \alpha\hat{n}(n)$ .

Wenn die Kurzzeit-Energie des Fehlersignals  $\overline{|e(n)|^2}$  jene des Eingangssignals  $\overline{|x(n)|^2}$  um 3dB übersteigt, wird die Leakage-Konstante  $\lambda(n)$  auf 0,95 gesetzt, andernfalls hat sie den Wert 1:

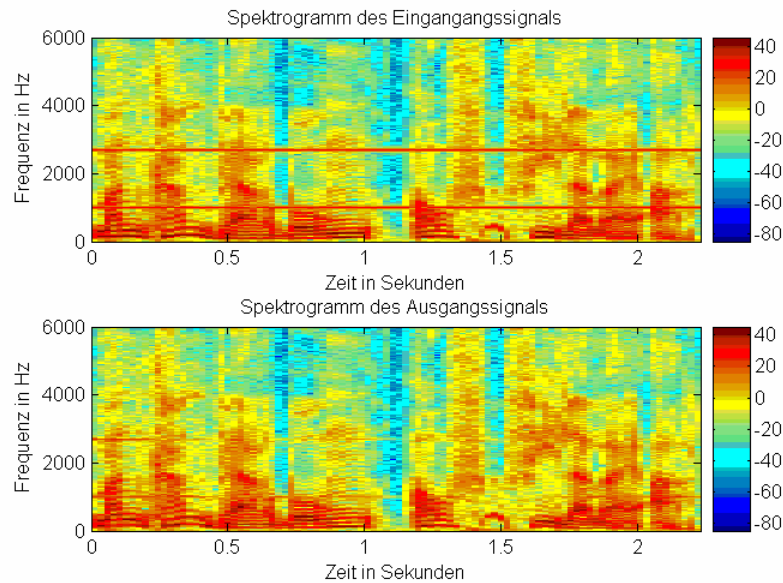
$$\lambda(n) = \begin{cases} 0,95, & \text{für } \overline{|e(n)|^2} > 2\overline{|x(n)|^2} \\ 1, & \text{sonst} \end{cases} \quad [2.7]$$

Diese Kurzzeit-Energien werden unter Zuhilfenahme eines IIR-Filters erster Ordnung folgendermaßen geschätzt:

$$\overline{|x(n)|^2} = 0,999 \cdot \overline{|x(n-1)|^2} + 0,001 \cdot |x(n)|^2, \quad [2.8]$$

$$\overline{|e(n)|^2} = 0,999 \cdot \overline{|e(n-1)|^2} + 0,001 \cdot |e(n)|^2 \quad [2.9]$$

Das Simulationsergebnis ist in Abbildung 2-5 zu sehen.



**Abbildung 2-5** Zeit-Frequenz-Analyse des Eingangs- bzw. Ausgangssignals des *adaptiven line enhancers* mit IIR-Struktur

Auch hier beträgt die Abtastrate  $f_s = 44,1$  kHz. Die Verzögerung beträgt  $N_D = 75$  Samples, die Koeffizientenanzahl beträgt  $N = 250$  und die Schrittweite für die Adaption ist  $\mu = 0,0014$  eingestellt.

### 2.1.3 Vergleich der beiden Strukturen

Das Simulationsergebnis zeigt, dass die beiden Störfrequenzen im Falle der IIR-Struktur etwas besser unterdrückt werden. Dies liegt an der tieferen und schmäleren Kerbe, die die IIR-Filterstruktur aufweist.

Allerdings hat die IIR-Struktur auch Nachteile. Bei den beiden präsentierten Beispielen handelt es sich nicht um rückgekoppelte Systeme. Das heißt, der Ausgang des line-enhancement-Filters wird nicht an den Eingang zurückgeführt. Wäre dies der Fall, wie zum Beispiel bei einer Lautsprecher-Mikrofon Anordnung, dann würde die Energie der periodischen Störkomponente durch die Filterung mit niedrigerer Intensität in die Kostenfunktion des adaptiven Filters eingehen. Dies würde die Kerbtiefe des Filters reduzieren und in weiterer Folge wieder zu einem höheren periodischen Störanteil im Signal führen. Die Tiefe der Kerbe würde also bei einem rückgekoppelten System oszillieren.

Eine Beschreibung der Funktion des Filters innerhalb des Gesamtsystems folgt in Kapitel 6.

## 2.2 Adaptiver Störunterdrücker als Kerbfilter nach Widrow/Stearns

Steht eine Rückkopplfrequenz durch eine Detektion, wie sie beispielsweise in Kapitel 3 beschrieben wird, zur Verfügung, so besteht die Möglichkeit ein Kerbfilter nach Widrow und Stearns zu implementieren [13]. Die Bandbreite dieses Filters lässt sich einfach kontrollieren und das Filter ist in der Lage dem Referenzeingangssignal zu folgen. Außerdem ist der geringe Rechenaufwand von nur zwei Koeffizienten ein großer Vorteil. Leider ist die Dämpfung des Filters nicht einstellbar, sondern geht immer gegen unendlich. Das Filter könnte sich aber, wie auch der *adaptive line enhancer* (siehe Kapitel 2.1), wegen seiner hohen Geschwindigkeit und eben dieser hohen Dämpfung zum Beispiel als letzte Instanz in einem Gesamtsystem, in dem vorher eine Filterung mittels einer Filterbank durchgeführt wurde als „Notfilter“ eignen. In Abbildung 2-5 ist das *single frequency adaptive notch filter* nach Widrow/Stearns zu sehen.

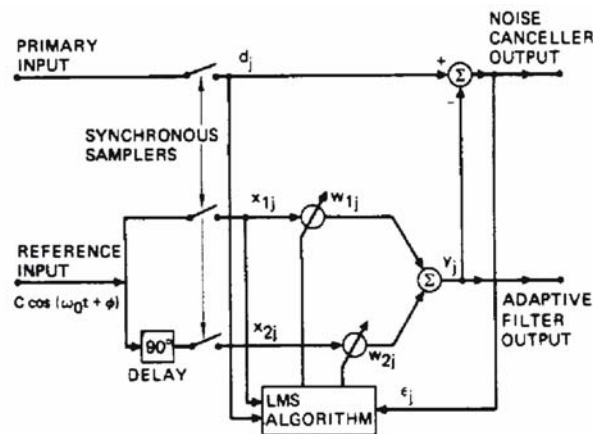


Abbildung 2-5 Single-frequency adaptive notch filter [13]

Das Referenzsignal  $C \cdot \cos(\Omega_0 t + \Phi)$ , ein reiner Kosinus, stellt das Eingangssignal  $x_{1k}$ , und nach einer  $90^\circ$  Phasenverschiebung  $x_{2k} = C \cdot \sin(\Omega_0 t + \Phi)$  dar. Die Gewichte  $w_{1k}$  und  $w_{2k}$  werden nach dem LMS Algorithmus folgendermaßen aktualisiert:

$$\begin{aligned} w_{1,k+1} &= w_{1k} + 2\mu\epsilon_k x_{1k} \\ w_{2,k+1} &= w_{2k} + 2\mu\epsilon_k x_{2k} \end{aligned} \quad [2.10]$$

wobei  $\epsilon_k$  das Fehlersignal darstellt.

Mit  $\omega_0 = 2\pi f_0 T$ , ergibt sich für die Referenz-Eingangssignale:

$$\begin{aligned} x_{1k} &= C \cos(k\omega_0 + \Phi) \\ x_{2k} &= C \sin(k\omega_0 + \Phi) \end{aligned} \quad [2.11]$$

In [13] wird gezeigt, dass die Übertragungsfunktion  $H(z)$  des Filters durch

$$H(z) = \frac{z^2 - 2z \cos \omega_0 + 1}{z^2 - 2(1 - \mu C^2)z \cos \omega_0 + 1 - 2\mu C^2} \quad [2.12]$$

gegeben ist.

Gleichung [2.12] zeigt, dass es sich bei diesem Filter um ein Kerbfilter bei der Frequenz  $\omega_0$  handelt. Die Nullstellen der Übertragungsfunktion befinden sich in der  $z$ -Ebene bei

$$z = e^{\pm j\omega_0} \quad [2.13]$$

und somit exakt auf dem Einheitskreis bei den Winkeln  $\pm\omega_0$  rad. Die Pole befinden sich bei

$$z = (1 - \mu C^2) \cos \omega_0 \pm j[(1 - 2\mu C^2) - (1 - \mu C^2)^2 \cos^2 \omega_0]^{1/2}. \quad [2.14]$$

Sie liegen innerhalb des Einheitskreises mit einem radialen Abstand von  $(1 - 2\mu C^2)^{1/2} \approx 1 - \mu C^2$  zum Mittelpunkt und bei den Winkeln

$$\pm \cos^{-1}[(1 - \mu C^2) \cdot (1 - 2\mu C^2)^{-1/2} \cos \omega_0]. \quad [2.15]$$

Für eine langsame Adaption, das heißt für kleine Werte von  $\mu C^2$ , hängen diese Winkel vom Faktor

$$\frac{1 - \mu C^2}{(1 - 2\mu C^2)^{1/2}} = \left( \frac{1 - 2\mu C^2 + \mu^2 C^4}{1 - 2\mu C^2} \right)^{1/2} = (1 + \mu^2 C^4 + \dots)^{1/2} = 1 + \mu^2 C^4 + \dots$$

ab, der beinahe 1 ist. Das bedeutet, dass die Winkel der Polstellen nahezu identisch mit jenen der Nullstellen sind. Die Positionen der Pol- und Nullstellen bzw. die Punkte der halben Leistung sind in Abbildung 2-6 zu sehen.

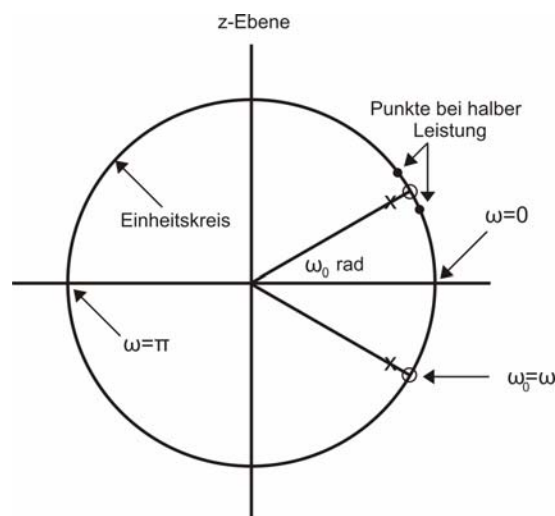


Abbildung 2-6 Pol-Nullstellendiagramm Single-frequency adaptive notch filter [13]

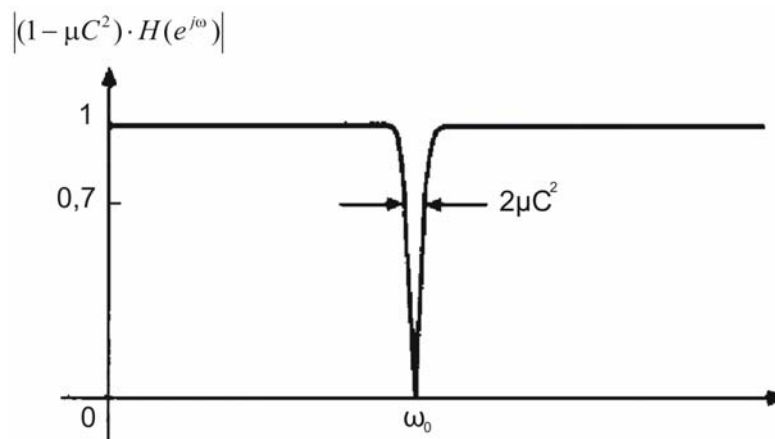
Da die Nullstellen auf dem Einheitskreis liegen, ist die Tiefe der Kerbe des Filters bei  $\omega = \omega_0$   $-\infty$  dB. Die Breite der Kerbe wird durch die Nähe der Pole zu den Nullstellen bestimmt. Die Polstellen haben in etwa den Abstand  $\mu C^2$  zu den Nullstellen. Die Bogenlänge entlang des Einheitskreises zwischen den Punkten der halben Leistung stellt die Bandbreite des Filters dar und ergibt sich zu:

$$bw = 2\mu C^2 [\text{rad}] = \frac{\mu C^2}{\pi T} [\text{Hz}] \quad [2.16]$$

Die Güte  $Q$  des Filters als Maß für die Schärfe der Kerbe ergibt sich aus dem Verhältnis von Mittenfrequenz zu Bandbreite:

$$Q_{-3dB} = \frac{\text{Mittenfrequenz}}{\text{Bandbreite}} = \frac{\omega_0}{2\mu C^2} \quad [2.17]$$

Abbildung 2-7 zeigt den Betrag der Übertragungsfunktion des *single-frequency adaptive notch filter*.



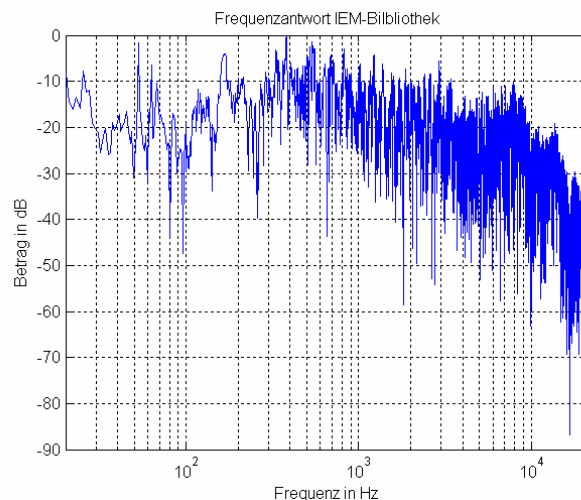
**Abbildung 2-7** Betrag der Übertragungsfunktion des single-frequency adaptive notch filter [13]

Wenn das Referenzsignal eine reine Kosinusschwingung ist, dann verhält sich das Filter wie ein stabiles Kerbfilter. Die Tiefe der Kerbe ist im allgemeinen ausgeprägter als die eines fixen Kerbfilters, da durch den Adaptionprozess die korrekte Phasenbeziehung für die Unterdrückung beibehalten wird. Dies gilt auch, wenn sich die Referenzfrequenz langsam ändert.



## 2.3 Frequenzschieber

Frequenzschieber werden benutzt um akustische Schleifen zu stabilisieren. Rückkopplungssysteme sind stabil, solange die Schleifenverstärkung für alle Frequenzen kleiner als 1 ist. Unter der Annahme eines linearen Verstärkers werden Rückkopplfrequenzen zuerst an der Stelle auftreten, an der die Übertragungsfunktion ihr Maximum hat. Für jene Frequenzen, die sich am Empfänger konstruktiv überlagern, also jene Frequenzen die sich gleichphasig am Mikrofon addieren, werden höhere Schleifenverstärkungen auftreten als für andere. Der Ansatz nützt die Übertragungscharakteristik von großen Räumen aus. Misst man die Übertragungsfunktion vom Lautsprecher zum Mikrofon, so werden spektrale Täler und Hügel wie in Abbildung 2-9 auftreten. Der spektrale Bereich der für die Koppelfrequenzen verantwortlich ist, wird durch den Frequenzschieber nach und nach in ein spektrales Tal geschoben bis die Stabilitätsbedingung wieder erfüllt ist. Die Abbildungen 2-8 und 2-9 zeigen die Frequenzantwort der IEM-Bibliothek<sup>1</sup>. Die spektralen Hügel und Täler sind deutlich zu erkennen. Die lokalen Maxima und Minima haben einen durchschnittlichen Abstand von 10Hz [4]. Ihr typischer Amplitudenunterschied liegt etwa bei 10dB, teilweise sogar bei bis zu 30 dB.



**Abbildung 2-8** Frequenzantwort der IEM-Bibliothek

1) Die Messung der Impulsantwort erfolgte mit der „swept sine Methode“ nach Angelo Farina [3].

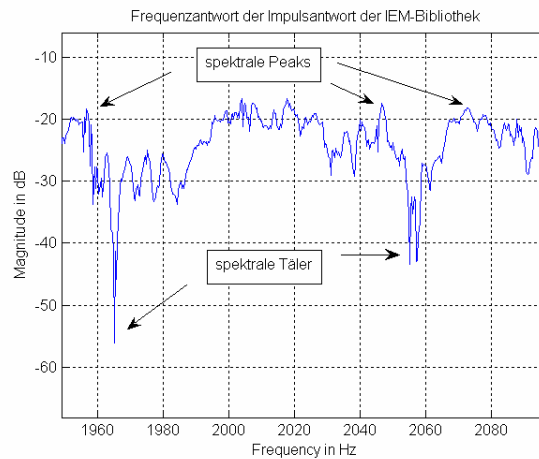


Abbildung 2-9 spektrale Hügel und Täler in einer Frequenzantwort

### 2.3.1 Hilbert Filter

Dieser Abschnitt zeigt, wie ein Frequenzschieber implementiert werden kann [4].

Gegeben sei ein bandbegrenztetes Signal. Das Spektrum wird zunächst in ein unteres und ein oberes Seitenband aufgeteilt:

$$S(e^{j\Omega}) = S_L(e^{j\Omega}) + S_U(e^{j\Omega}) \quad [2.18]$$

Bei einer Verschiebung in Richtung höherer Frequenzen wird das obere Seitenband in Richtung  $+\pi$ , das untere Seitenband in Richtung  $-\pi$  verschoben. Die Verschiebungs-Frequenz sei  $\Omega_0$ . Zunächst wird das Signal mit einer Kosinusfunktion multipliziert, was zu zwei verschobenen Versionen des Originalspektrums im Frequenzbereich führt:

$$s_{\cos}(n) = s(n) \cdot \cos(n\Omega_0) \quad [2.19]$$

$$S_{\cos}(e^{j\Omega}) = \frac{1}{2} S(e^{j(\Omega+\Omega_0)}) + \frac{1}{2} S(e^{j(\Omega-\Omega_0)}) \quad [2.20]$$

Teilt man das Spektrum wie in [2.18] wieder in ein unteres und ein oberes Seitenband auf, so ergibt sich:

$$S_{\cos}(e^{j\Omega}) = \frac{1}{2} [S_L(e^{j(\Omega-\Omega_0)}) + S_U(e^{j(\Omega+\Omega_0)}) + S_L(e^{j(\Omega+\Omega_0)}) + S_U(e^{j(\Omega-\Omega_0)})] \quad [2.21]$$

Die beiden letzten Terme aus Gleichung [2.21] sind in die entgegengesetzte Richtung der beiden ersten und somit in die falsche Richtung verschoben. Um diese zu eliminieren wird ein zweites Trägersignal generiert. Es wird diesmal eine Sinusfunktion als Modulationsfunktion verwendet. Vor der Modulation wird das Signal hilbertgefiltert:

$$s_{Hil}(n) = s(n) * h_{Hil,n} \quad [2.22]$$

Das Hilbert-Filter verschiebt alle negativen Frequenzen des Eingangsspektrums um  $\pi/2$  und alle positiven Frequenzen um  $-\pi/2$ :

$$S_{Hil}(e^{j\Omega}) = -j \operatorname{sgn}(\Omega) S(e^{j\Omega}) \quad [2.23]$$

die Übertragungsfunktion von  $h_{Hil,i}$  lautet also:

$$H_{Hil}(e^{j\Omega}) = -j \operatorname{sgn}(\Omega) \quad [2.24]$$

durch Multiplikation des Hilbert gefilterten Signals mit der Sinusfunktion

$$s_{\sin}(n) = s_{Hil}(n) \sin(n\Omega_0) \quad [2.25]$$

erhält man im Frequenzbereich:

$$S_{\sin}(e^{j\Omega}) = j \frac{1}{2} S_{Hil}(e^{j(\Omega+\Omega_0)}) - j \frac{1}{2} S_{Hil}(e^{j(\Omega-\Omega_0)}) \quad [2.26]$$

Einsetzen von [2.23] und [2.18] ergibt:

$$S_{\sin}(e^{j\Omega}) = \frac{1}{2} (S_L(e^{j(\Omega-\Omega_0)}) + S_U(e^{j(\Omega+\Omega_0)}) - S_L(e^{j(\Omega+\Omega_0)}) - S_U(e^{j(\Omega-\Omega_0)})) \quad [2.27]$$

Durch Addition der beiden modulierten Signale, bleibt nur das obere Seitenband, dass in Richtung  $+\pi$  verschoben wurde, und das untere Seitenband, dass in Richtung  $-\pi$  verschoben wurde, übrig.

$$U(e^{j\Omega}) = S_{\cos}(e^{j\Omega}) + S_{\sin}(e^{j\Omega}) = S_L(e^{j(\Omega+\Omega_0)}) + S_{\sin}(e^{j(\Omega-\Omega_0)}) \quad [2.28]$$

Abbildung 2-10 zeigt das Blockschaltbild des Frequenzschiebers.

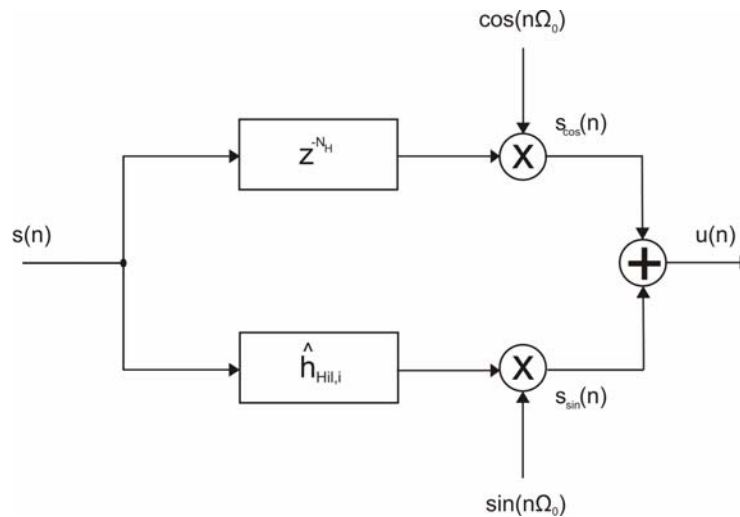


Abbildung 2-10 Frequenzschieber

Die Impulsantwort des Hilbert-Filters kann nun noch berechnet werden. Die inverse Fourier-Transformation von [2.24] ergibt nach kurzer Berechnung [4] die diskrete Lösung:

$$h_{Hil,i} = \begin{cases} 0, & \text{für } i \text{ ungerade} \\ \frac{2}{i\pi}, & \text{sonst} \end{cases} \quad [2.29]$$

Diese Impulsantwort ist unendlich lang und nicht kausal. Sie muss daher zunächst mit einem geeignetem Fenster (z.B. Hamming-Fenster um  $i=0$ ) zwischen  $-N_H$  und  $N_H$  ausgeschnitten und um  $N_H$  verschoben werden.

$$\hat{h}_{Hil,i} = h_{Hil,i-N_H} h_{Win,i-N_H} \quad [2.30]$$

Aufgrund dieser Verschiebung muss auch der kosinusmodulierte Signalpfad in Abbildung 2-10 um  $N_H$  Samples verzögert werden.

Eine Möglichkeit für die Realisierung einer  $90^\circ$  Phasendrehung ist der Einsatz besonderer Allpass Strukturen. Solche Anordnungen können zwar nicht einzelne Signale alleine um  $90^\circ$  in der Phase drehen, dafür aber zwei Signale so bearbeiten, dass zwischen ihnen eine  $90^\circ$  Phasendifferenz auftritt. Zur Dimensionierung sucht man sich eine Mittenfrequenz, um welche herum die beiden Allpässe der Ordnung  $M$  eine genaue  $90^\circ$  Phasendifferenz zueinander haben sollen. Der Phasenwinkel  $M \cdot 90^\circ$  ist die Hälfte der gesamten erreichbaren Allpass Phasendrehung. Von dieser Phasendrehung ausgehend, setzt man einen Allpass auf  $M \cdot 90^\circ + 45^\circ$ , den anderen auf  $M \cdot 90^\circ - 45^\circ$  bei der festgelegten Mittenfrequenz. Dadurch erreicht man für diese spezielle Frequenz eine exakte  $90^\circ$  Phasendifferenz, für benachbarte Frequenzen eine angenäherte  $90^\circ$  Phase. Diese Dimensionierung kann nun auch für mehrere

Frequenzen durch Verkettung solcher Allpass Strukturen vorgenommen werden, um eine breitere Wirkung zu erzielen.

In [4] wird die Erhöhung der Stabilitätsgrenzen durch die Verwendung eines Frequenzschiebers für die Verschiebungs-Frequenzen 6, 9 und 12Hz in drei verschiedenen Orten präsentiert. Es handelt sich um einen Vorlesungssaal, die Eingangshalle der Universität Darmstadt und einen Hallraum. Die Orte weisen also eine sehr unterschiedlich lange Nachhallzeit auf. Bei Anregung mit weißem Rauschen sind im Hallraum Verstärkungen bis zu 11dB möglich, ohne dass Rückkoppelfrequenzen auftreten. Da eine Anregung mit weißem Rauschen nichts über die Sprachverständlichkeit aussagt, werden die gleichen Versuche mit Sprachsignalen wiederholt. Zunächst wird die Verstärkung des Systems ohne Frequenzschieber so lange erhöht, bis die ersten Qualitätsminderungen am Signal auftreten. Dies geschieht knapp unter der Stabilitätsgrenze. Danach wird der Frequenzschieber dazugeschaltet und der gleiche Versuch wiederholt. Bei dieser Versuchsanordnung ist die maximale Erhöhung der Verstärkung weitaus geringer. In dem Vorlesungssaal kann die Stabilitätsgrenze etwa um 1 - 2dB herauf gesetzt werden, in der Eingangshalle sind Erhöhungen bis zu 4 dB möglich. Die größten Verbesserungen können im Hallraum erzielt werden. Hier sind Erhöhungen von 5 - 6dB möglich. Ab einer Verschiebungs-Frequenz von 12 Hz sind sehr störende Artefakte im Sprachsignal zu hören.

Grundsätzlich lässt sich sagen, dass ein Frequenzschieber zwar die Stabilitätsgrenze nach oben verschiebt, er aber bei deutlicher Erhöhung der Schleifenverstärkung immer zu relativ starken Einbußen in der Qualität des Signals führt. Innerhalb des in dieser Arbeit erstellten Gesamtsystems existiert ein Frequenzschieber. Eine Evaluierung dieses Frequenzschiebers befindet sich in Kapitel 7.

## 2.4 Pitch Shifter

Ein Nachteil des Frequenzschiebers ist, dass durch seinen Einsatz die Obertonstruktur des Signals zerstört wird. Bei Einsatz eines *pitch shifters*, bei dem die Frequenzen multiplikativ verschoben werden, ist dies nicht der Fall. Im Rahmen dieser Arbeit wird eine sehr einfache Variante eines Pitch Shifters eingesetzt.

Das Signal wird in eine delayline geschrieben, und aus dieser mit modulierter Geschwindigkeit ausgelesen. Dies entspricht praktisch einer Modulation der Abtastrate. Durch diese Modulation werden die Koppelfrequenzen ähnlich wie beim Frequenzschieber in spektrale Täler verschoben (siehe Kapitel 2.3). Der Unterschied besteht darin, dass hier die Obertonstruktur erhalten bleibt. Auch der *pitch shifter* wird in Kapitel 7 getestet und evaluiert.

Es existieren natürlich weitaus bessere *pitch shifting* Algorithmen, die aber aufgrund ihrer Komplexität und ihrer hohen Rechenintensität hier nicht behandelt werden. Genaueres findet man in [15].

Im Versuch wurde festgestellt, dass sich durch Einsatz eines *pitch shifters* anstelle eines Frequenzschiebers keine Verbesserung der Signalqualität einstellt. Daher wurde die Implementierung aufwändiger Algorithmen nicht weiterverfolgt.

## 3 Präzise Fourier-Analyse mittels Signalableitungen

Die Fourier-Transformation transformiert ein Zeitsignal in seine spektrale Darstellung. Sie liefert ein spektrales Bild des gesamten Klages und mittelt das Signal in ein einziges Spektrum. Dieses Spektrum stimmt aber nur für stationäre Signale mit unserer Wahrnehmung überein. Da sich die meisten Signale mit der Zeit verändern, müssen Analyse-Algorithmen zeitvariante Ergebnisse liefern. Solche Algorithmen wiederholen dieselben Prozeduren für kurze aufeinanderfolgende Abschnitte des Signals. Diese Art der Verarbeitung heißt Kurzzeit-Analyse. Der bekannteste Vertreter der Kurzzeit-Analyse ist die Kurzzeit Fourier-Analyse, die eine zeitabhängige Version der Fourier-Transformation liefert.

$$X[m] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \cdot e^{-j\frac{2\pi}{N}nm} \quad [3.1]$$

Für jeden Block  $x$ , bestehend aus  $N$  Samples wird die Fourier-Transformierte berechnet.

Sei  $S$  das Leistungsspektrum des Signals  $s$ , und  $S(f)$  der Wert bei der Frequenz  $f$ . Die diskrete Fourier-Transformation würde ein Spektrum zwischen 0Hz und der halben Abtastfrequenz  $f_s/2$  liefern, und zwar in Schritten von  $f_s/N$  Hz.

Die Frequenzgenauigkeit ist also indirekt proportional zur Blocklänge  $N$ . Eine gute Frequenzauflösung würde eine hohe Blocklänge  $N$  erfordern, was aber zu einer schlechten Zeitauflösung führen würde. Bei der Kurzzeit Fourier-Analyse muss also ein Kompromiss zwischen Zeitauflösung und Frequenzauflösung gefunden werden. Je besser die Zeitauflösung, desto schlechter die Frequenzauflösung und umgekehrt. In diesem Kapitel soll eine Methode vorgestellt werden, die die Fourier-Analyse deutlich verbessert, indem sie diesen Konflikt minimiert.

### 3.1 FT<sup>n</sup>: Fourier Transformation unter Verwendung von Signal Ableitungen

Für die folgende Herleitung wird unter Bezug auf [2] von einem spektralen Modell ausgegangen, dass aus einer deterministischen Komponente  $d(t)$  und einer stochastischen Komponente  $s(t)$  besteht:

$$q(t) = d(t) + s(t)$$

Der deterministische Teil ist die Summe aus sinusförmigen Oszillatoren, deren Frequenzen und Amplituden sich langsam mit der Zeit verändern. Ein solch sich langsam verändernder Oszillator wird *Teilton* oder *Partial* genannt:

$$d(t) = \sum_{p=1}^P \text{osc}(f_p(t), a_p(t))$$

wobei  $P$  die Anzahl der Teiltöne und

$$\text{osc}(f_p(t), a_p(t)) = a_p(t) \cdot \cos(\Phi_p(t))$$

mit

$$\frac{d\Phi_p}{dt} = 2\pi f_p(t) \text{ d.h. } \Phi_p(t) = \Phi_p(0) + 2\pi \int_0^t f_p(u) \cdot du \text{ ist.}$$

$f_p$ ,  $a_p$  und  $\Phi_p$  stellen die Frequenz, Amplitude und Phase des  $p$ -ten Teiltöns dar.

Psychoakustische Tests haben gezeigt, dass das Gehör nicht sensitiv auf die Phase im Ursprung reagiert,  $\Phi_p(0)$  also beliebig wählbar ist und somit zu Null gesetzt werden kann [2]. Im Folgenden soll auch der stochastische Anteil von  $q(t)$ , der das Rauschen darstellt, vernachlässigt werden und nur der deterministische Teil betrachtet werden. Es ergibt sich also folgendes Signalmodell:

$$q(t) = \sum_{p=1}^P \text{osc}(f_p(t), a_p(t)) \quad [3.2]$$

Gegeben sei nun ein einzelner Oszillator:

$$o_p(t) = \text{osc}(f_p(t), a_p(t)) = a_p(t) \cdot \cos(\Phi_p(t))$$

Die erste Ableitung ist dann:

$$\frac{do_p}{dt}(t) = \frac{d}{dt}(a_p(t) \cdot \cos(\Phi_p(t))) = a_p(t) \cdot \frac{d}{dt}(\cos(\Phi_p(t))) + \frac{da_p}{dt} \cdot \cos(\Phi_p(t))$$

wenn  $a_p$  sich nur langsam mit der Zeit ändert, dann kann  $\frac{da_p}{dt} = 0$  gesetzt werden und es ergibt sich:

$$\frac{do_p}{dt} = a_p(t) \cdot \frac{d}{dt}(\cos(\Phi_p(t))) = -a_p(t) \frac{d\Phi_p}{dt}(t) \cdot \sin(\Phi_p(t))$$



mit  $\frac{d\Phi_p}{dt} = 2\pi f_p(t)$  folgt:

$$\frac{do_p}{dt} = -a_p(t)(2\pi f_p(t)) \cdot \sin(\Phi_p(t)) = 2\pi \cdot a_p(t) f_p(t) \cdot \cos\left(\Phi_p(t) - \frac{\pi}{2}\right) \quad [3.3]$$

Durch Beweisführung mittels Induktion kann gezeigt werden, dass allgemein gilt:

$$\frac{d^k o_p}{dt^k} = a_p(t)(2\pi f_p(t))^k \cdot \cos\left(\Phi_p(t) + \left(-k \cdot \frac{\pi}{2}\right)\right)$$

Da die Ableitung eine lineare Operation ist, lässt sich unter Berücksichtigung von Gleichung [3.2] obige Bedingung wie folgt formulieren:

$$\frac{d^k q_p}{dt^k} = \sum_{p=1}^P a_p(t)(2\pi f_p(t))^k \cdot \cos\left(\Phi_p(t) + \left(-k \cdot \frac{\pi}{2}\right)\right) \quad [3.4]$$

Sei  $FT^k$  die Fourier-Transformation der  $k$ -ten Ableitung des Signals  $\frac{d^k q}{dt^k}$  ( $k > 0$ ), und  $FT^k(f)$  ihre Amplitude bei der Frequenz  $f$ . Eine Schlussfolgerung aus Gleichung [3.4] ist, dass für jeden Teilton  $p$ , ein Maximum für  $FT^i$  an der Stelle  $f_p$  existieren muss (angenommen  $a_p > 0$  und  $f_p > 0$ ) und:

$$f_p = \frac{1}{2\pi} \cdot \frac{FT^{i+1}(f_p)}{FT^i(f_p)} \quad [3.5]$$

## 3.2 DFT<sup>1</sup>: Die Diskrete FT<sup>1</sup>

Für die  $FT^n$ -Transformation benötigt man die ersten  $n$  Ableitungen. Im Folgenden soll gezeigt werden wie sich diese aus dem digitalen Signal berechnen lassen.

Die erste Ableitung  $\frac{dq}{dt}$ , auch als  $q'$  bezeichnet, ist mathematisch definiert durch:

$$q'(t) = \lim_{\varepsilon \rightarrow 0} \frac{q(t+\varepsilon) - q(t)}{\varepsilon}$$

Die linke und rechte Ableitungen für  $\varepsilon$  negativ bzw. positiv sind gegeben durch:

$$q'_-(t) = \lim_{\varepsilon \rightarrow 0_-} \frac{q(t+\varepsilon) - q(t)}{\varepsilon} = \lim_{\eta \rightarrow 0_+} \frac{q(t) - q(t-\eta)}{\eta} \quad (\eta = -\varepsilon)$$

$$q_+'(t) = \lim_{\varepsilon \rightarrow 0_+} \frac{q(t + \varepsilon) - q(t)}{\varepsilon}$$

Linke und rechte Ableitung sind also gleich, d.h.:

$$q_-'(t) = q_+'(t) = q'(t)$$

In der Praxis ist das Audiosignal  $q$  ein diskretes Signal  $q[i]$ , das  $q\left(i \cdot \frac{1}{f_s}\right)$  repräsentiert, und der kleinste  $|\varepsilon|$  für die Berechnung der Ableitung ist die Samplingperiode  $\frac{1}{f_s}$ . Es werden also folgende Näherungen angenommen:

$$q_-'[i] = (q[i] - q[i-1]) \cdot f_s \quad \text{und} \quad q_+'[i] = (q[i+1] - q[i]) \cdot f_s$$

Jetzt ist  $q_-'(t) \neq q_+'(t)$ , aber:

$$q_+'[i] = (q[i+1] - q[i]) \cdot f_s = q_-'[i+1]$$

$$q_-'[i] = (q[i] - q[i-1]) \cdot f_s = q_+'[i-1]$$

Die beiden Ableitungen stellen also bis auf eine Verschiebung um ein Sample die gleichen diskreten Funktionen dar. Diese Verschiebung soll vernachlässigt werden und die Näherung der linken Ableitung als Näherung für die eigentliche Ableitung verwendet werden. ( $q[i]=0$  für  $i < 0$ ):

$$q'[i] = f_s (q[i] - q[i-1]) \tag{3.6}$$

Gleichung [3.6] stellt ein digitales Hochpassfilter mit linearer Phase dar. Seine Differenzgleichung lautet:

$$y[n] = f_s x[n] - f_s x[n-1]$$

Die Übertragungsfunktion ist also:

$$H(z) = f_s (1 - z^{-1})$$

mit der Verstärkung:

$$|H(e^{j\omega})| = f_s \cdot \sqrt{2(1 - \cos(\omega))} \quad \text{mit} \quad \omega = \frac{2\pi f}{f_s}$$

Die Ableitung ist eine lineare Operation, die auch als Filterungsoperation mit der Verstärkung  $2\pi f_p$  angesehen werden kann (s. Gleichung [3.3]). Die theoretische Verstärkung ist also  $f_s \omega$ , die sich deutlich von der praktischen Verstärkung unterscheidet. Dieser Unterschied kann durch Multiplikation des Leistungsspektrums der Signalableitung, mit einem Skalierungsfaktor  $F$  ausgeglichen werden.

$$F(\omega) = \frac{\omega}{\sqrt{2(1 - \cos(\omega))}} \quad \text{mit} \quad \omega = \frac{2\pi f}{f_s} \quad [3.7]$$

Die diskrete Form von Gleichung [3.5] für  $i=0$  lautet:

$$f_p = \frac{1}{2\pi} \cdot \frac{DFT^1[m_p]}{DFT^0[m_p]} \quad [3.8]$$

$m_p$  ist der zugehörige Index zu  $f_p$ , der das Maximum der  $DFT^0$  angibt. Für jedes Maximum  $m$  der  $DFT^0$  kann also mit Gleichung [3.8] die exakte Frequenz  $f_p$  berechnet werden. Da nur die erste Ableitung von  $a$  benutzt wird, heißt diese Methode in der Literatur *first order derivative Fourier analysis* [2].

### 3.3 Der Analyse Algorithmus

Im folgenden wird in sieben kurzen Schritten der Algorithmus zur Berechnung der exakten Frequenz  $f_p$  mit der  $DFT^1$  Methode beschrieben.

1. Fensterung des Eingangssignals  $q$
2. Berechnung der  $DFT^0$  also der gewöhnlichen diskreten Fourier-Transformation von  $q$
3. Berechnung von  $q'$  unter Verwendung von Gleichung [3.6]
4. Fensterung von  $q'$  mit der gleichen Fensterfunktion (z.B. Hann-Fenster)
5. Berechnung der  $DFT^1$  also der diskreten Fourier-Transformation von  $q'$
6. Skalierung des Leistungsspektrums der  $DFT^1$  mit dem Skalierungsfaktor  $F$  unter Verwendung von Gleichung [3.7]
7. Berechnung der genauen Frequenz für jedes lokale Maximum der  $DFT^0$  unter Verwendung von Gleichung [3.8]

Natürlich müssen die Werte der  $DFT^1$  nur für die lokalen Maxima der  $DFT^0$  berechnet werden, was den Rechenaufwand deutlich verringert.

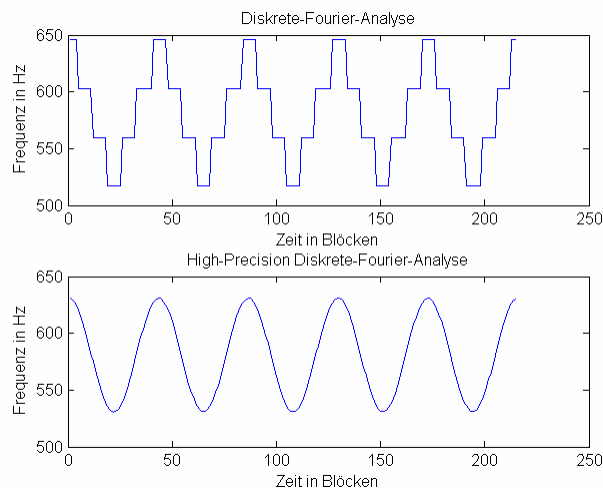
### 3.3.1 Beispiel für ein frequenzmoduliertes Signal

Als Testsignal für die  $DFT^1$  dient ein frequenzmoduliertes Signal:

$$x(t) = A_1 \sin(\omega_1 t + A_2 \sin(\omega_2 t)) \quad \text{mit}$$

$$A_1=1, \quad A_2=50, \quad \omega_1=2\pi f_1, \quad \omega_2=2\pi f_2, \quad f_1=581\text{Hz}, \quad f_2=50\text{Hz}$$

Abbildung 3-1 zeigt das Testergebnis.



**Abbildung 3-1**  $DFT$  (oben) und  $DFT^1$  (unten) für ein frequenzmoduliertes Testsignal

Es ist deutlich zu erkennen, dass die normale Kurzzeit-Fourier-Transformation ein sehr viel ungenaueres Ergebnis liefert als die  $DFT^1$ -Methode. Die  $DFT^0$  liefert für das Frequenzmaximum nur 4 unterschiedliche Werte, und zwar:

$$f_{\max} = \frac{f_s}{N} \cdot k = \frac{44100}{1024} \cdot k, \quad k \in [12,13,14,15]$$

Dagegen liefert die  $DFT^1$ -Methode einen nahezu kontinuierlichen Verlauf der Frequenzmaxima des frequenzmodulierten Eingangssignals (siehe Abbildung 3-1 unten).

Mit dieser präzisen Fourier-Analyse ist es also möglich, die nahezu exakte Frequenz einer Rückkopplfrequenz zu ermitteln. Danach könnte man beispielweise mit Filtern nach Widrow/Stearns (siehe Kapitel 2.2) sehr schmalbandige Filterungen des Signals vornehmen.

Allerdings wird in Kapitel 6 gezeigt, dass es nicht unbedingt notwendig ist, die exakte Frequenz einer Störfrequenz zu kennen. Durch geschickte Wahl der Güte eines Kerbfilters ist es sogar möglich mehrere Koppelfrequenzen mit einem Filter zu unterdrücken. Speziell für den Fall einer Filterbank mit begrenzter Anzahl von Filtern ist die Frequenzauflösung einer FFT mit 1024 oder 2048 Punkten ausreichend genau.

## 4 Detektion von Rückkoppelfrequenzen

In diesem Kapitel werden zwei Methoden vorgestellt, die es ermöglichen, auftretende Rückkoppelfrequenzen innerhalb breitbandiger Sprachsignale zu erkennen. Die Methoden werden getestet, miteinander verglichen und schließlich innerhalb des Gesamtsystems eingesetzt.

### 4.1 Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins

Bei dieser Detektionsmethode wird blockweise das Leistungsspektrum berechnet und anschließend für jeden Frequenz-Bin überprüft, ob die zugehörige Leistung der Amplitude gegenüber ihren vergangenen Werten streng monoton ansteigt. Übersteigt die Dauer des Anstiegs einen zeitlichen Schwellwert  $S_z$  (z.B.  $S_z=8$  Blöcke, was bei einer FFT mit  $N=1024$  und einer Abtastrate von 44,1 kHz etwa 184ms entspricht), so wird eine Koppelfrequenz detektiert. Als Nebenbedingung wird überprüft, ob es sich bei diesem Frequenz-Bin um ein lokales Maximum handelt. Ansonsten würden breitbandige Signale mit kontinuierlich ansteigender Amplitude als Rückkopplung erkannt werden. Für diese Überprüfung wird ein Algorithmus zur Spitzenwert-Detektion implementiert, welcher in Kapitel 6.3 genauer beschrieben wird.

Es wird zunächst ein Testsignal erzeugt, das aus einem Sinuston bei  $f=560\text{Hz}$  mit linear ansteigender Amplitude und überlagertem gaußverteilten weißen Rauschen mit einem SNR von 10 dB besteht.

Die Signallänge  $L$  beträgt fünf Sekunden bei einer Abtastrate von 44,1kHz. Die Blocklänge für die FFT beträgt 1024 Samples. Daraus ergibt sich eine Anfangsverzögerung von ca. 23ms, wobei zur Analyse keine Blocküberlappung und Rechteckfensterung verwendet wird. Der Sinuston startet bei  $L/4$ , d.h. im 54. Analyseframe. Die Amplitude des Sinustons steigt kontinuierlich und linear von 0 auf 0.6 mit 0.00001 pro Sample an. Ab dem 113. Block darf keine Koppelfrequenz mehr detektiert werden.

Eine Rückkoppelfrequenz soll von der Detektion dann erkannt werden, wenn die Leistung  $P_{ss}$  eines Frequenz-Bins in z.B. 7 aufeinanderfolgenden Blöcken streng monoton ansteigt ( $7*23\text{ms}=161\text{ms}$ ). Dieser zeitliche Schwellwert muss empirisch ermittelt werden. Es muss gewährleistet sein, dass die Bedingung eines solchen Anstiegs bei einem ungestörtem Sprachsignal nicht erfüllt ist, da sonst Fehldetektionen entstehen.  $P_{ss}$  berechnet sich aus:

$$P_{ss} = |X(n)|^2 = X(n) \cdot X^*(n)$$

$X(n)$  ... Amplitude eines Frequenz - Bins

Daraus ergibt sich durch die zusätzliche Anfangsverzögerung eine minimale Verzögerung  $T_{\min}$  von:

$$T_{\min} = (S_z + 1) \cdot \frac{N}{fs}$$

$S_z$  ... zeitlicher Schwellwert

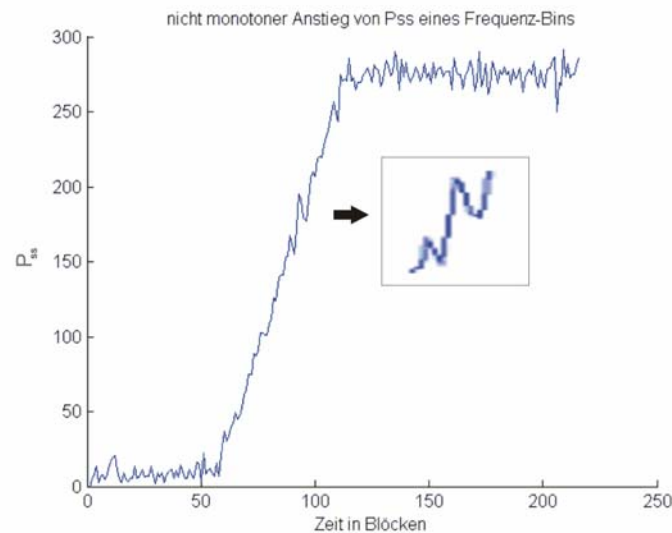
$N$  ... Blocklänge der FFT

$fs$  ... Abtastrate

Je kürzer man diesen zeitlichen Schwellwert wählt, desto schneller werden Koppelfrequenzen detektiert. Allerdings ergeben sich, wie oben erwähnt, Fehldetektionen, wenn man diesen Schwellwert zu klein wählt, weil ein kürzerer streng monotoner Anstieg der Leistung eines Frequenz-Bins auch zufällig auftreten kann. Ein langer streng monotoner Anstieg hingegen tritt nur bei einer Rückkoppelfrequenz auf. Bei der Wahl des zeitlichen Schwellwerts gilt es also zwischen der Latenzzeit und der Häufigkeit der Fehldetektionen abzuwägen.

## 4.1.1 Glättung

Der SNR spielt für das Zusammenwirken von Glättung und Monotonie – unter Annahme eines Sinustons als Eingangssignal – eine wesentliche Rolle. Auch die Qualität des Störers wirkt auf diese Abhängigkeit ein. Durch die Verrauschung des Signals ist die Bedingung des streng monotonen Anstiegs von  $P_{ss}$  trotz kontinuierlichem Anstieg der Leistung nicht mehr erfüllt (s. Abbildung 4-1).



**Abbildung 4-1** nicht streng monotoner Anstieg von  $P_{ss}$  des Frequenz-Bins 13 @ FFT-Blocklänge=1024 →  $f=559\text{Hz}$

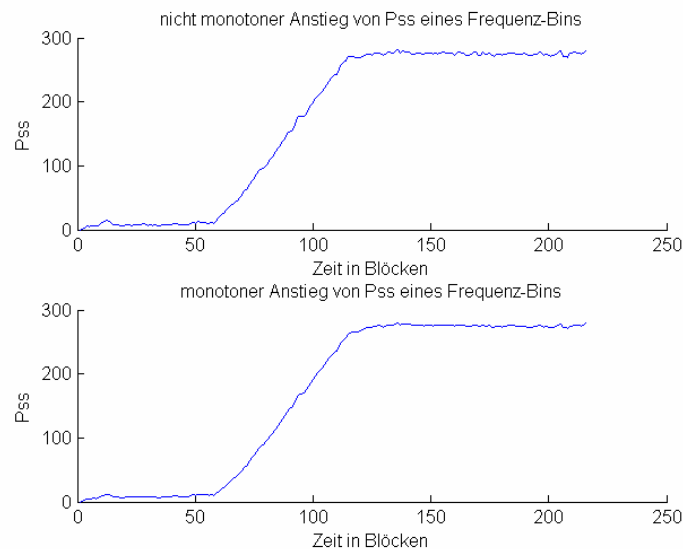
Die Detektion kann also die auftretende Koppelfrequenz aufgrund der Verrauschung nicht erkennen. Es ist daher notwendig eine zeitliche Glättung des Spektrums einzuführen. Diese wird folgendermaßen vorgenommen:

$$|X_j(n)|^2 = \alpha \cdot |X_j(n)|^2 + (1 - \alpha) \cdot |X_j(n-1)|^2 \quad \forall j = 0 \dots N-1$$

$N$ ... Länge der FFT bzw. Blockgröße

$X_j$  ist die Amplitude des  $j$ -ten Frequenz-Bins,  $\alpha$  der Vergessensfaktor und  $n$  der aktuelle Block.

Führt man diese Glättung für obiges Beispiel mit  $\alpha=0.3$  bzw.  $\alpha=0.2$  ein ergibt sich folgendes Verhalten für  $P_{ss}$ .



**Abbildung 4-2** oben: geglätteter ( $\alpha=0.3$ ) nicht streng monotoner Anstieg unten: ( $\alpha=0.2$ ) monotoner Anstieg von  $P_{ss}$  des Frequenz-Bins 13 @ FFT-Blocklänge=1024  $\rightarrow f=559\text{Hz}$

Man erkennt im oberen Teil der Abbildung zwar eine deutliche Verbesserung, also einen glatteren Anstieg von  $P_{ss}$ , allerdings verläuft der Anstieg des Betrags von  $P_{ss}$  noch immer nicht *streng monoton*.

Bei einer Glättung mit  $\alpha=0.2$  hingegen ist die Voraussetzung für die Detektion, also der *streng monotone* Anstieg gewährleistet.

## 4.1.2 Zeitkonstanten bzw. Verzögerung

Die Zeitkonstante  $\tau$  gibt diejenige Zeit an, die das System benötigt, um bei einem Einheitssprung als Eingangssignal, am Ausgang 63% der Maximalamplitude zu erreichen.

Sie berechnet sich bei Glättung zu:

$$\tau = -\frac{1}{\ln(1-\alpha)} \cdot \frac{1}{fs^*}$$

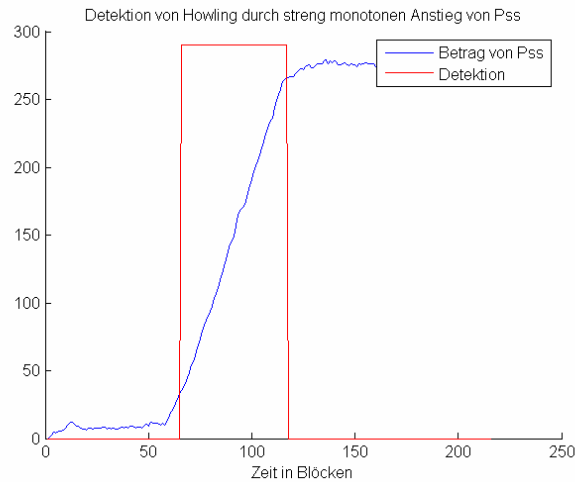
Für die Abtastrate  $fs^*$  muss hierbei berücksichtigt werden, dass die FFT blockweise berechnet wird, und  $fs^*$  somit nicht der Abtastrate des Audiosignals entspricht sondern mit der Blocklänge  $N$  der FFT unterabgetastet ist:

$$fs^* = \frac{fs}{N}$$

Die Zeitkonstante  $\tau$  für  $\alpha=0.2$  bei einer Blocklänge von  $N=1024$  ergibt sich beispielsweise zu:



$$\tau = -\frac{1}{\ln(1-0.2)} \cdot \frac{1024}{44100} \approx 0,1s$$



**Abbildung 4-3** Detektion einer Rückkoppelfrequenz durch monotonen Anstieg von  $P_{ss}$  des Frequenz-Bins 13 @ FFT-Blocklänge=1024  $\rightarrow$   $f=559\text{Hz}$

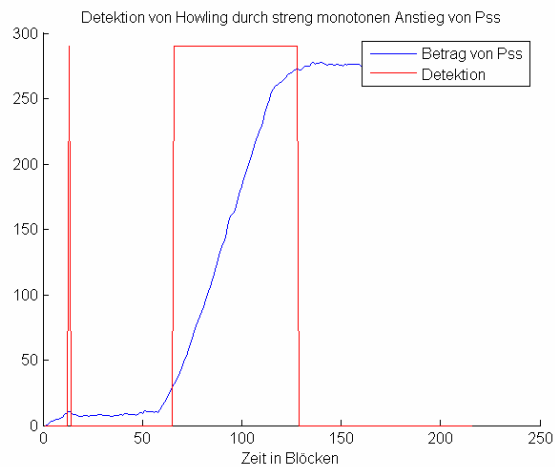
In Abbildung 4-3 ist zu sehen wie die Detektion erfolgt. Der monotone Anstieg von  $P_{ss}$  beginnt im 58. Block. Sieben Blöcke später, also ab dem 65. Block wird eine Koppelfrequenz erkannt. Diese Zeitverzögerung würde also, unter Berücksichtigung der Anfangsverzögerung, oben beschriebenen 184ms entsprechen. Der Sinuston beginnt wie vorher beschrieben im 54. Block.

Da die Koppelfrequenz erst im 65. Block erkannt wird, ergibt dies insgesamt eine Verzögerung von 11 Blöcken. Dies entspricht ca. 253ms. Da der Sinuston ab dem 113. Block eine konstante Amplitude aufweist, die Detektion aber erst im 117. Block auf Null zurückspringt, ergibt sich auch hier eine Verzögerung von 4 Blöcken.

Diese Zeitverluste sind auf die Glättung des Spektrums zurückzuführen, die aber notwendig ist, damit bei verrauschten oder amplitudenmodulierten Signalen wie etwa Gesang auch eine Detektion möglich ist. Wird obiger Versuch mit einem nicht verrauschten Sinus und ohne Glättung des Spektrums, also  $\alpha=1$ , durchgeführt, ergibt sich die erwartete Verzögerung von 7 Blöcken. Der anschwellende Sinuston beginnt im 54. Block, die Detektion erkennt ihn mit dem 61. Block.

Verringert man  $\alpha$  noch weiter auf 0.15, so ergibt sich wegen der langsameren Aktualisierung zu Beginn eine Fehldetektion (s. Abbildung 4-4). Dies liegt daran, dass die beschriebene Glättung wie ein Tiefpassfilter wirkt und somit der anfänglich relativ schnelle Anstieg auf mehrere Blöcke verteilt wird, was zu einem streng monotonem Anstieg von  $P_{ss}$  führt.

Der Sinuston wird genauso wie oben im 65. Block erkannt.

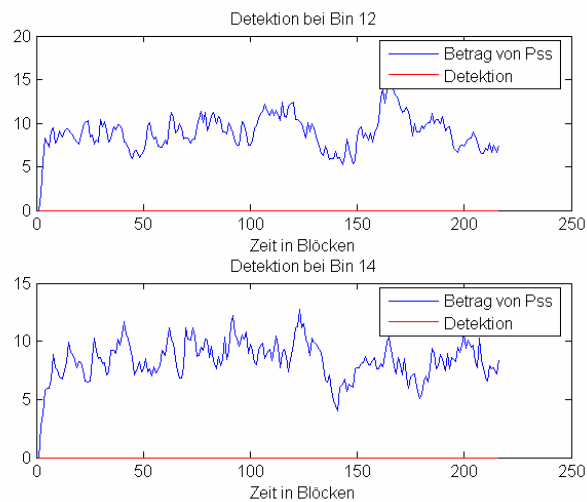


**Abbildung 4-4** Fehldetektion durch zu langsamen Update bei  $\alpha=0.15$ , Frequenz-Bin 13 @ FFT-Blocklänge=1024  $\rightarrow f=559\text{Hz}$

Außerdem wird das Ende des Sinustons erst im 128. Block erkannt, also 11 Blöcke=253ms später als für den Fall  $\alpha=0.2$

### 4.1.3 Untersuchung der benachbarten Frequenz-Bins

Die benachbarten Frequenz-Bins sind durch den Anstieg des Sinustons bei 560Hz nicht beeinflusst. Abbildung 4-5 zeigt den Verlauf der beiden benachbarten Frequenz-Bins bei  $f_{\text{Bin1}}=517\text{Hz}$  und  $f_{\text{Bin2}}=603\text{Hz}$  und die zugehörige Detektion, die auf Null bleibt, weil keine Rückkopplfrequenz detektiert wird. Dies ist auch leicht ersichtlich, da 560Hz genau auf einem Bin der FFT liegt und bei einer Rechteckfensterung die benachbarten Bins genau bei den Nullstellen der Sinc-Funktion im Frequenzbereich zu liegen kommen.



**Abbildung 4-5** keine Fehldetektion bei Frequenz-Bin 14  $\rightarrow f_{\text{Bin1}}=603\text{Hz}$  und Frequenz-Bin 12  $\rightarrow f_{\text{Bin2}}=517\text{Hz}$

#### 4.1.3.1 Detektion von Rückkoppelfrequenzen zwischen zwei FFT-Bins

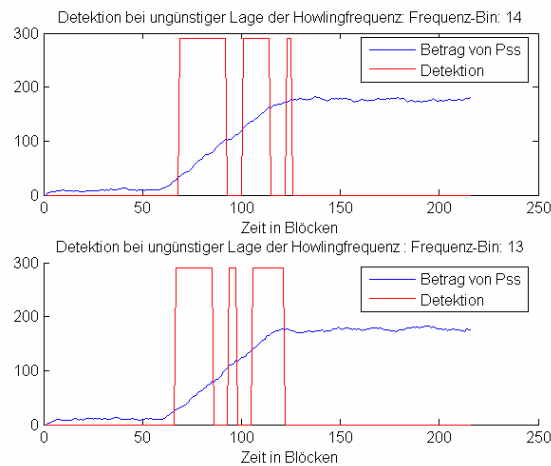
Es wird zunächst der ungünstigste Fall, nämlich der Fall, bei dem die Rückkoppelfrequenz genau in der Mitte zwischen zwei Frequenz-Bins der FFT zu liegen kommt, untersucht. Dazu wird als Testsignal wie oben, ein verrauschter Sinuston (SNR=10dB) mit linear ansteigender Amplitude und der Frequenz

$$f_{\text{sin}} = \frac{f_s}{N} \cdot 13,5 = \frac{44100\text{Hz}}{1024} \cdot 13,5 \cong 581,4\text{Hz}$$

$N$ : Länge der FFT

generiert. Das Ergebnis dieses Tests ist in Abbildung 4-6 zu sehen. Man erkennt, dass sowohl beim 13. als auch beim 14. Frequenz-Bin, also sowohl bei  $f_1=559\text{Hz}$  als auch bei  $f_2=630\text{Hz}$  Koppelfrequenzen detektiert werden. Diese Detektion könnte bei einer nachgeschalteten Filterbank entweder ein Filter in der Mitte der beiden detektierten Frequenzen ansteuern, oder die Güte eines der beiden Filter müsste entsprechend verkleinert werden. Letzteres wird bei jenem Filter passieren, der von der Detektion früher gesetzt wurde, also dessen Frequenz früher erkannt wurde. Jedenfalls findet trotz der sehr ungünstigen Lage der Rückkoppelfrequenz eine Detektion statt, die bei entsprechender Handhabung für die Problemstellung verwendet werden, und eine nachgeschaltete Filterbank ansteuern kann.

Die Koppelfrequenz würde hier allerdings aufgrund dieser ungünstigen Lage erst im 66. bzw. 68. Block erkannt werden, was einer zusätzlichen Verzögerung von 23ms bzw. 69ms entspräche.



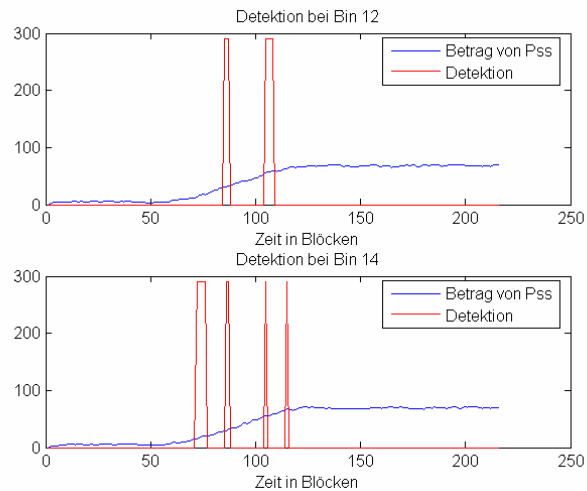
**Abbildung 4-6** Detektionsverhalten bei den Frequenz-Bins 13 und 14 für ein Testsignal mit  $f=581,4\text{Hz}$

Obiger Fall stellt den „worst case“, also die ungünstigste Lage einer auftretenden Koppelfrequenz dar. Je näher die Rückkopplfrequenz sich an einem Bin der FFT befindet, desto besser funktioniert die Detektion.

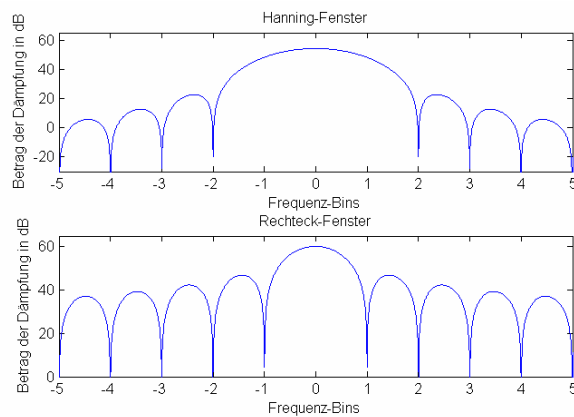
#### 4.1.4 Fensterung

Obige Untersuchung ist mit einer Rechteckfensterung durchgeführt worden. Führt man beispielsweise eine Hanningfensterung ein, so liegen die benachbarten Frequenz-Bins noch innerhalb der 4 Bin breiten Hauptkeule des Hanningfensters und ihre Leistungen folgen dem monotonen Anstieg des relevanten Frequenz-Bins. Dies führt zu Fehldetektionen, die in Abbildung 4-7a zu sehen sind. In dieser Hinsicht kommt es also weniger auf die Nebenkeulendämpfung des verwendeten Fensters als auf die Breite der Hauptkeule an (siehe Abbildung 4-7b). Allerdings wird für das Gesamtsystem eine weitere Detektionsmethode benötigt, die Koppelfrequenzen mit einer Schleifenverstärkung knapp unter 1 detektiert (siehe Kapitel 6.2 und 6.3). Für diese Detektionsmethode erweist sich eine Hanningfensterung als vorteilhaft. Da aufgrund des Rechenaufwands nicht zwei Fouriertransformationen durchgeführt werden sollen, und durch einen Tracking-Algorithmus (siehe Kapitel 6.1) die Fehldetektionen, die aufgrund der Hanningfensterung entstehen nicht ins Gewicht fallen, wird als Analysefenster das Hanningfenster verwendet.

Die Analyse muss aufgrund der Hanningfensterung überlappend durchgeführt werden. Würde man nicht überlappend analysieren, dann würden Ereignisse, die am Rand des Analysefensters auftreten geringer gewichtet werden als Ereignisse, die in der Mitte des Fensters auftreten. Besonders schlecht wäre dies dann, wenn es sich um periodische Ereignisse handelte, die immer am Rand des Fensters auftreten.



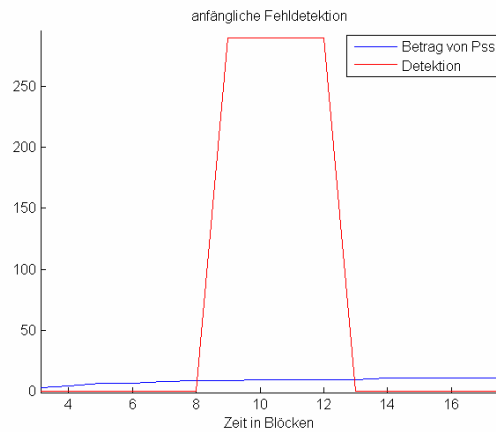
**Abbildung 4-7a** Fehldetektion durch Hanningfensterung bei Frequenz-Bin 14  $\rightarrow f_{\text{Bin1}}=603\text{Hz}$  und Frequenz-Bin 12  $\rightarrow f_{\text{Bin2}}=517\text{Hz}$



**Abbildung 4-7b** Hauptkeule und erste Nebenkeulen des Hanning- bzw. Rechteckfensters

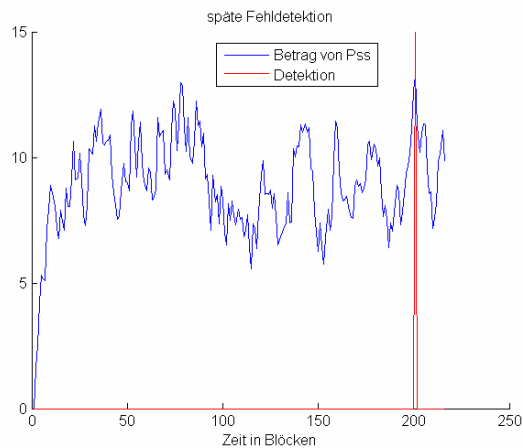
### 4.1.5 Fehldetektionen

Neben den Fehldetektionen durch die Fensterung entstehen auch noch andere Fehldetektionen. Hauptsächlich treten diese zu Beginn des Signals auf, wenn durch die Trägheit, die durch die Glättung entsteht, die Spektren zu langsam aktualisiert werden. Für den Frequenz-Bin 60 @ FFT-Blocklänge=1024  $\rightarrow f_{\text{Bin}}=2584\text{Hz}$ , den Fall  $\alpha=0,2$  und obiges Testsignal ist in Abbildung 4-8 eine solche Fehldetektion zu sehen.



**Abbildung 4-8** Fehldetektion, Frequenz-Bin 60 @ FFT-Blocklänge=1024  $\rightarrow f_{\text{Bin}}=2584\text{Hz}$ ,  $\alpha=0,2$

Es treten allerdings auch zu späteren Zeitpunkten kurze Fehldetektionen auf. Sehr kurze Fehldetektionen kann man leicht verhindern, in dem man den zeitlichen Schwellwert beispielsweise von 7 auf 8 Blöcke erhöht. Man handelt sich dadurch allerdings bei den wahren Detektionen eine zusätzliche Verzögerung von einer Blocklänge ein. Bei der Wahl des zeitlichen Schwellwerts gilt es genau zu beachten, mit welcher Art von Signal man es zu tun hat: Die Kurzzeitstationarität des Signals spielt dabei eine wesentliche Rolle. Eine *späte* Fehldetektion ist in Abbildung 4-9 für den Frequenz-Bin 17  $\rightarrow f_{\text{Bin}}=732\text{Hz}$  zu sehen.



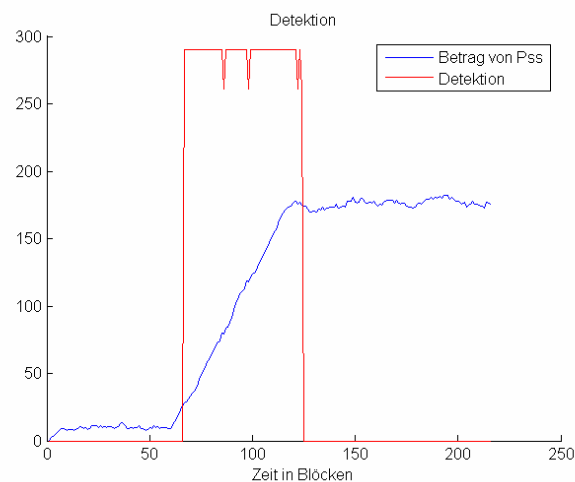
**Abbildung 4-9** späte Fehldetektion, Frequenz-Bin 17 @ FFT-Blocklänge=1024  $\rightarrow f_{\text{Bin}}=732\text{Hz}$ ,  $\alpha=0,2$

## 4.1.6 „Drop-out“ Blöcke

Im unteren Teil der Abbildung 4-6 ist zu erkennen, dass sich der monotone Anstieg von  $P_{ss}$  zwischenzeitlich zwei Mal nur für einen Block („drop-out“ Block) nicht fortsetzt. Danach dauert es jeweils wieder 8 Blöcke bis eine Rückkopplfrequenz erkannt wird, da die Bedingung für eine Detektion ein Anstieg über 8 Blöcke ist. Um dieses Verhalten zu verbessern, wird eine weitere Bedingung für die Detektion eingeführt:

Ist bereits eine Koppelfrequenz erkannt, und setzt sich der monotone Anstieg nur für einen einzigen Block nicht fort, so soll für diesen Block trotzdem weiterhin Rückkopplung detektiert werden.

Abbildung 4-10 zeigt das Resultat einer solchen Bedingung. Die „drop-out“ Blöcke sind durch die Einschnitte in der Detektionslinie leicht zu erkennen. Der Anstieg von  $P_{ss}$  wird nun durchgehend erkannt. Selbstverständlich zögert diese Veränderung das Detektionsende um mindestens einen Block hinaus.



**Abbildung 4-10** Einführung von „drop-out“ Blöcke

## 4.1.7 Detektion von mehreren Rückkoppelfrequenzen im Testsignal

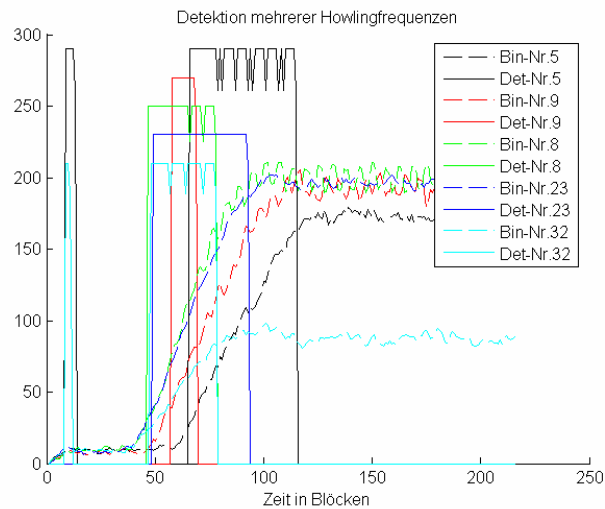
Für diesen Test wird ein Testsignal generiert, das aus 5 Sinustönen mit linear ansteigender Amplitude besteht. Diese sind wiederum von weißem, gaußverteilterm Rauschen mit einem SNR von 10 dB überlagert. Die Frequenzen der Testtöne und deren zugehörige Bin-Nummer, sowie ihre ideale und reale Detektionszeit sind Tabelle 5.1 zu entnehmen. Man sieht, dass die Zeiten bis zur Detektion zwischen 11 und 16 Blöcken liegen, also zwischen 255ms und 371ms.

Die Signallänge  $L$  beträgt wiederum 5 Sekunden bei einer Abtastfrequenz von  $f_s=44100\text{Hz}$  und einer Analyse-Blocklänge von  $N_{\text{FFT}}=1024$ .

|                                  | Sinus1 | Sinus2 | Sinus3 | Sinus4 | Sinus5 |
|----------------------------------|--------|--------|--------|--------|--------|
| Frequenz [Hz]                    | 210    | 340    | 380    | 1000   | 1360   |
| Relevante Bin Nr.                | 5      | 8      | 9      | 23     | 32     |
| Startblock bzw. ideale Detektion | 54     | 36     | 44     | 36     | 31     |
| reale Detektion in Block         | 65     | 46     | 57     | 48     | 47     |

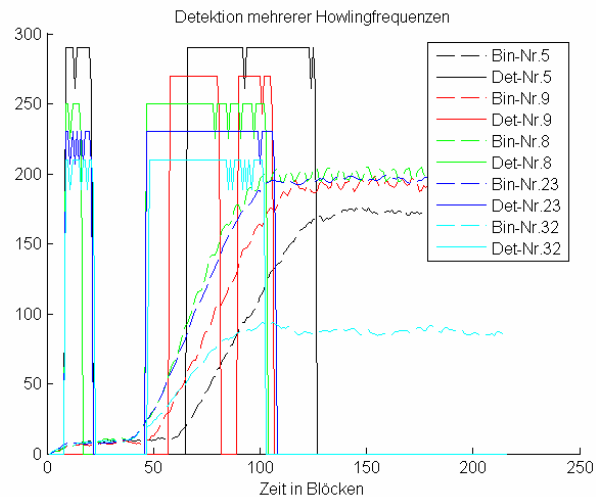
**Tabelle 4-1** Testtöne für die Detektion mehrerer Rückkoppelfrequenzen





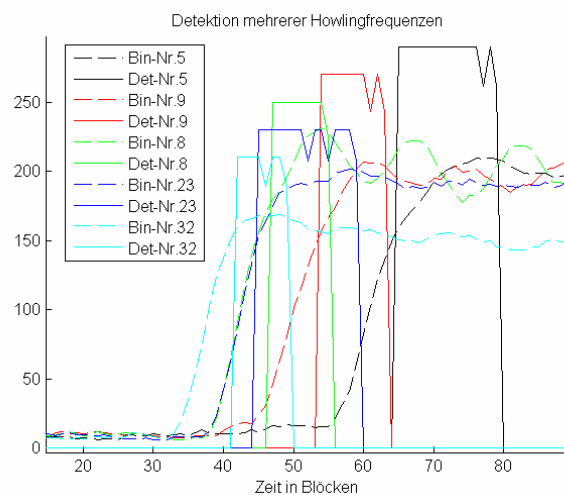
**Abbildung 4-11** Detektion mehrerer Rückkopplfrequenzen,  $\alpha=0.2$

Man sieht in Abbildung 4-11 Fehldetektionen am Anfang des Signals. Dieser Effekt wurde weiter oben beschrieben und ist auf die Glättung zurückzuführen. Außerdem sind die „drop-out“ Blöcke zum Beispiel bei der schwarzen Detektionslinie für die Frequenz  $f=215,3$  sehr schön zu erkennen. Leider wird der Anstieg der Amplituden nicht bei allen Testtönen während des gesamten Zeitraums des Amplitudenanstiegs detektiert. Erhöht man die Glättung auf  $\alpha=0.1$  (siehe Abbildung 4-12), so verbessert sich zwar dieses Verhalten, jedoch treten am Anfang umso mehr Fehldetektionen auf. Durch die Einführung einer Initialisierungsphase des Systems von zum Beispiel 1 Sekunde, in der keine Filter gesetzt werden können, stellen diese Fehldetektionen allerdings kein Problem dar.



**Abbildung 4-12** Detektion mehrerer Rückkopplfrequenzen höhere, Glättung  $\alpha=0.1$

Signale mit schneller ansteigenden Amplituden werden besser detektiert. Dazu wird ein Testsignal wie oben generiert, mit dem Unterschied, dass die Amplituden der Sinustöne 10 mal so schnell ansteigen. Das Ergebnis ist aus Abbildung 4-13 entnehmen.



**Abbildung 4-13** Detektion mehrerer Rückkopplfrequenzen,  $\alpha=0.2$ , zeitl. Schwellwert: 8 Blöcke, 10 mal schnellerer Amplitudenanstieg

Die realen Detektionszeiten nähern sich hier den idealen Detektionszeiten an und schwanken nun nur noch zwischen 9 und 11 Blöcken, das heißt zwischen 209ms und 255ms. Die Detektion für den Frequenz-Bin 23, die innerhalb von 7 Blöcken stattfindet, muss ein durch das Rauschen verursachter zufälliger Ausreißer sein, da wegen des zeitlichen Schwellwerts mindestens 8 Blöcke für die Detektion nötig sind. Mindestens der erste Anstieg, der zu den

erforderlichen acht Anstiegen gehört um eine Koppelfrequenz zu detektieren, muss also zufällig durch das Rauschen hervorgerufen worden sein.

## 4.1.8 Einfluss der Steuerparameter auf die Detektionszeit

### 4.1.8.1 Einfluss der Glättung $\alpha$ auf die Detektionszeit

Abbildung 4-14 zeigt die Abhängigkeit der Detektionszeit von  $\alpha$  für zwei Sinustöne mit unterschiedlich schnell ansteigenden Amplituden. Dabei stellt die untere, grüne Linie Testton1 mit einer 10 mal schneller ansteigenden Amplitude als Testton2 (obere blaue Linie) dar. Für Glättungen, die Werte von  $\alpha < 0.05$  aufweisen, ist die Zeitkonstante

$$\tau = -\frac{1}{\ln(1-\alpha)} \cdot \frac{1}{fs^*} = -\frac{1}{\ln(1-0.05)} \cdot \frac{1024}{44100} > 0.45,$$

was zu lange für eine schnelle Detektion von anschwellenden Störfrequenzen ist. Diese Fälle werden hier nicht weiter betrachtet. Für  $\alpha > 0,45$  ist die Glättung so niedrig, dass für den Testton1 gar keine Detektion mehr stattfindet. Dasselbe gilt für Testton2 bei  $\alpha > 0,8$ . Die Abhängigkeit der Detektionszeit von  $\alpha$  ist hier insgesamt sehr gering, vor allem innerhalb des relevanten Bereichs von  $0.05 < \alpha < 0.3$ .

Man sieht in Abbildung 4-14 auch deutlich, dass eine schneller ansteigende Amplitude zu einer besseren, das heißt früheren Detektion führt.

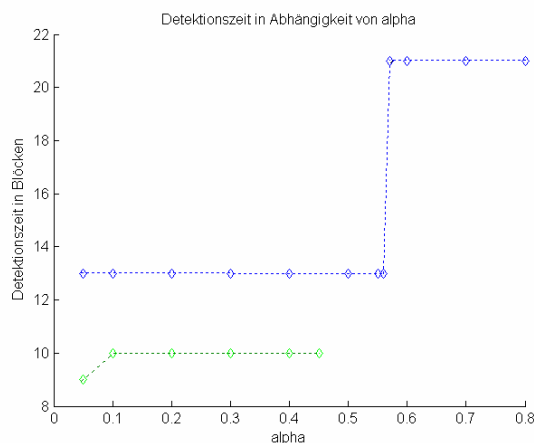


Abbildung 4-14 Detektionszeit in Abhängigkeit von  $\alpha$

### 4.1.8.2 Einfluss des zeitlichen Schwellwerts auf die Detektionszeit

Die Auswirkung des zeitlichen Schwellwerts auf die Detektionszeit ist offensichtlich. Bei der Erhöhung des Schwellwerts um einen Block (von  $n$  auf  $n+1$ ), wird sich auch die Detektionszeit um einen Block erhöhen, außer es tritt der ungünstige Fall auf, dass genau in diesem zusätzlichen Block ein zufälliger Ausreißer, d.h. eine zufällige Unterbrechung des monotonen Anstiegs von  $P_{ss}$ , stattfindet. In diesem Fall würde, vorausgesetzt der monotone Anstieg setzt sich anschließend fort, die Detektion weitere  $n+1$  Blöcke später stattfinden, also nach  $2n+2$  Blöcken. Dieser Fehler könnte bei einer zu geringen Glättung auftreten.

### 4.1.8.3 Einfluss des zeitlichen Verhaltens von auftretenden Rückkoppelfrequenzen auf die Detektionszeit

Interessanter ist die Untersuchung des zeitlichen Verhaltens von auftretenden Rückkoppelfrequenzen, da dies ja offenbar Einfluss auf die Detektionszeit hat. Dazu muss zunächst grundsätzlich geklärt werden, wie sich  $P_{ss}$  eines Frequenz-Bins verhält, wenn Koppelfrequenzen auftreten.

Dazu werden, wie in Abbildung 4-15 zu sehen, realistische Rückkoppelfrequenzen mit Hilfe einer realen Impulsantwort erzeugt, die in der IEM-Bibliothek nach der Methode von Angelo Farina [3] aufgenommen wurde.  $h(n)$  ist hierbei die aufgenommene Impulsantwort,  $g$  ist die Schleifenverstärkung. Als Eingangssignal dient weißes Rauschen.

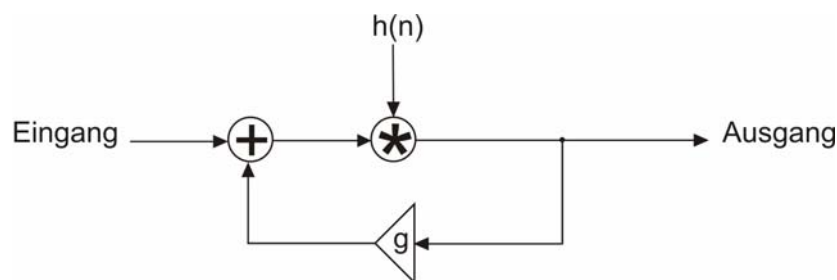
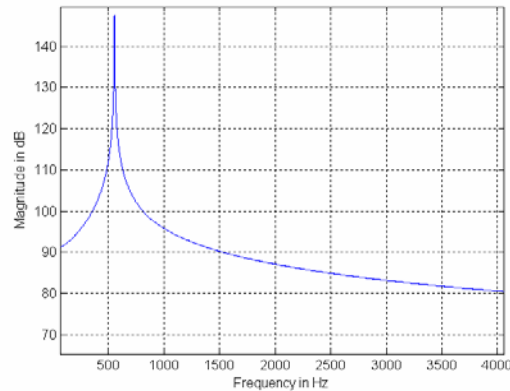


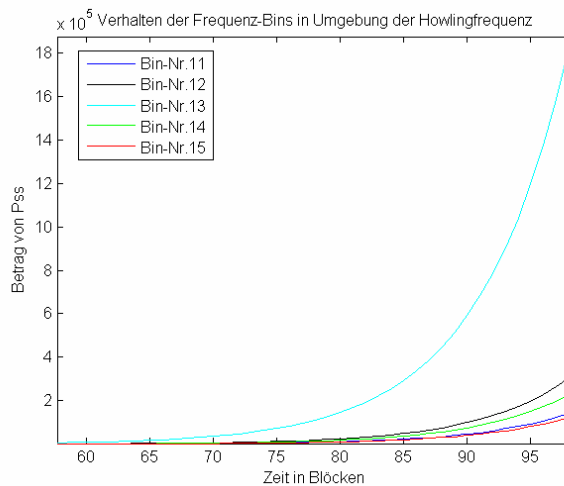
Abbildung 4-15 Erzeugung von Rückkoppelfrequenzen

Die Koppelfrequenz liegt, wie in Abbildung 4-16 zu sehen, etwa bei 554Hz.



**Abbildung 4-16** FFT des Ausgangssignals aus Abb.16

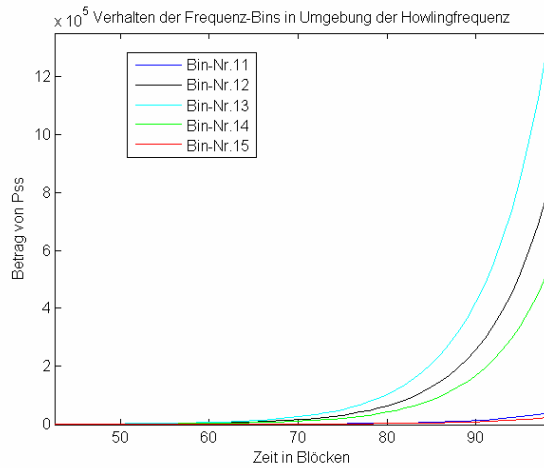
Bei einer Blockgröße von  $N=1024$  wird im Folgenden der 13. Frequenz-Bin und seine unmittelbare Umgebung genauer untersucht.



**Abbildung 4-17a** Verhalten der Frequenz-Bins in unmittelbarer Umgebung der Koppelfrequenz, Koppelfrequenz ca. 554Hz

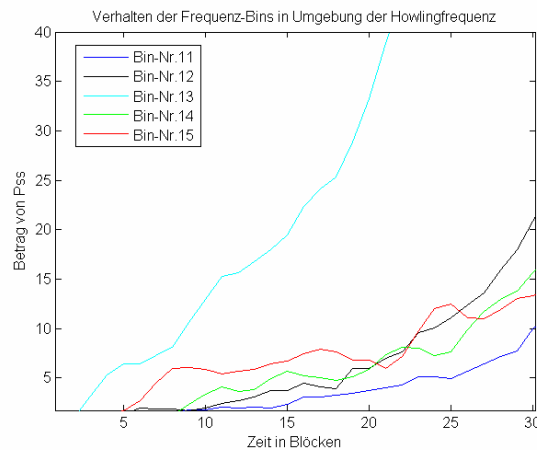
Der exponentielle Anstieg von  $P_{ss}$  ist bei allen Bins deutlich zu erkennen. Der Frequenz-Bin 13 ist etwa auf den 10-fachen Wert des zweithöchsten Bins angestiegen.

Zum Vergleich ist in Abbildung 5-17b die selbe Untersuchung mit einer Hanning-Fensterung zu sehen. Die beiden benachbarten Bins steigen deutlich stärker an als bei einer Rechteck-Fensterung, was eine eindeutige Detektion schwieriger macht.



**Abbildung 4-17b** Verhalten der Frequenz-Bins in unmittelbarer Umgebung der Koppelfrequenz, Koppelfrequenz ca. 554Hz, mit Hanning-Fensterung

Die Vergrößerung von Abbildung 4-17a zeigt das Verhalten der Leistungen  $P_{ss}$  der Bins während der Anfangsphase, die ja für die Detektion wichtig ist. Die Bedingung des streng monotonen Anstiegs von  $P_{ss}$  ist nur für den Frequenz-Bin 13 erfüllt, also wird nur hier eine Rückkopplfrequenz erkannt.



**Abbildung 4-18** Verhalten der Frequenz-Bins in unmittelbarer Umgebung der Koppelfrequenz

Wie wirkt sich nun hier eine Glättung des Spektrums auf die Detektionszeit aus? Dazu wird in Abbildung 4-19 der Steuerparameter  $\alpha$  der Detektionszeit gegenübergestellt.

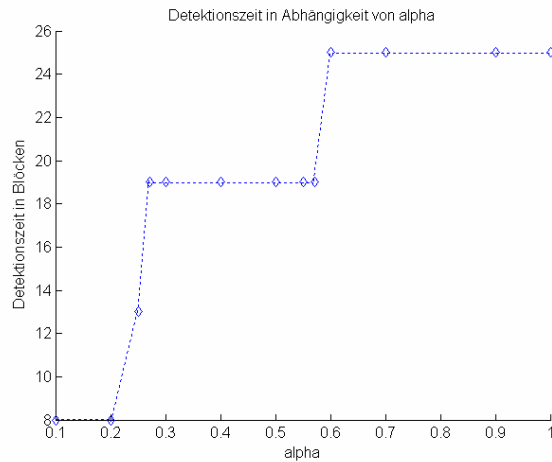


Abbildung 4-19 Detektionszeit in Abhängigkeit von  $\alpha$

Je kleiner  $\alpha$  ist, also je größer die Glättung, desto kürzer ist die Detektionszeit, weil durch eine höhere Glättung die Detektionsbedingung der Monotonie eher eingehalten wird.

#### 4.1.8.4 Kurze Fehldetektionen

Abbildung 4-20 zeigt das Verhalten der Detektion für ein gewöhnliches nicht rückgekoppeltes Sprachsignal. Es sind einige Fehldetektionen von sehr kurzer Dauer (1 Block) zu sehen. Durch eine Erhöhung des zeitlichen Schwellwerts von 7 auf 8 Blöcke können diese verhindert werden.

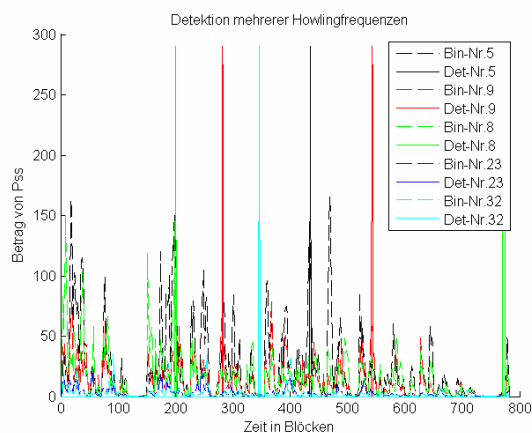


Abbildung 4-20 Einige kurze Fehldetektionen, die durch Erhöhung des zeitl. Schwellwerts vermieden werden können

## 4.2 Spektrale Flachheit (SFM-Methode)

Diese Methode folgt einem Patent von Kawamura et. al. [6]. Bei dieser Detektionsmethode wird blockweise die Leistung eines Frequenz-Bins ins Verhältnis zur Durchschnittsleistung des jeweiligen Blockes gesetzt. Übersteigt dieses Verhältnis einen absoluten Schwellwert  $S_a$ , dann gilt diese Frequenz als *Koppelkandidat*. Wird dieser Schwellwert länger als ein zeitlicher Schwellwert  $S_z$  überschritten, dann wird eine Rückkopplfrequenz detektiert.

Es wird hierfür zunächst blockweise (z.B. Blocklänge: 1024 Samples) das *normierte Leistungsdichtespektrum*  $P_s$  bis zur halben Abtastfrequenz berechnet:

$$P_s = \frac{X_j \cdot X_j^*}{L} = \frac{X_j \cdot X_j^*}{1024},$$

$$\text{mit } X(k) = \sum_{n=0}^{N-1} x[n] \cdot e^{-2\pi j \frac{nk}{N}}$$

$X^*$  ... konjugiert komplexe von  $X$  und

$L$  ... Blocklänge

Danach wird mit

$$P_{Ges} = \sum_{j=1}^N P_{s_j}, N=512$$

die Gesamtleistung des Blockes bis zur halben Abtastfrequenz ermittelt und ein spezieller Durchschnitt der Leistung ohne die drei höchsten Amplitudenwerte berechnet.

$$P_{AV} = \frac{P_{Ges} - (P_s(f_1) + P_s(f_2) + P_s(f_3))}{N-3}$$

mit  $P_s(f_1) \geq P_s(f_2) \geq P_s(f_3) \geq P_s(f_n) \quad \forall n \in [0 \dots N-1] / f_1, f_2, f_3$ ,

Dieser Durchschnitt liegt also etwas niedriger als der Gesamtdurchschnitt des Blocks und führt zu einem größerem Abstand zwischen Durchschnitts- und Maximalwert innerhalb eines Blocks.

Schließlich wird mit

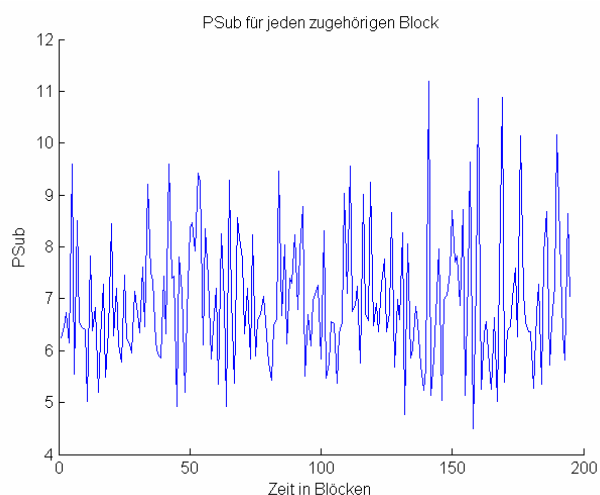


$$P_{Sub} = \frac{P_{max}}{P_{AV}} = \frac{P_s(f_1)}{P_{AV}}$$

derjenige Wert ermittelt, der mit einem Pegelschwellwert verglichen wird. Wird dieser Schwellwert von  $P_{Sub}$  überschritten, so gilt die Frequenz  $f_1$  als *Koppelfrequenzkandidat*. Übersteigt  $P_{Sub}$  diesen Schwellwert länger als eine bestimmte vorgegebene Zeit, so wird eine Rückkopplfrequenz detektiert. Die Überprüfung muss für jeden Frequenz-Bin durchgeführt werden.

Zunächst soll ermittelt werden wie hoch der absolute Schwellwert sein muss, der eine Frequenz als *Koppelfrequenzkandidat* bestimmt.

Dazu wird als Testsignal weißes gaußverteiltes Rauschen erzeugt und  $P_{Sub}$  für jeden Block in Abbildung 4-21 dargestellt.

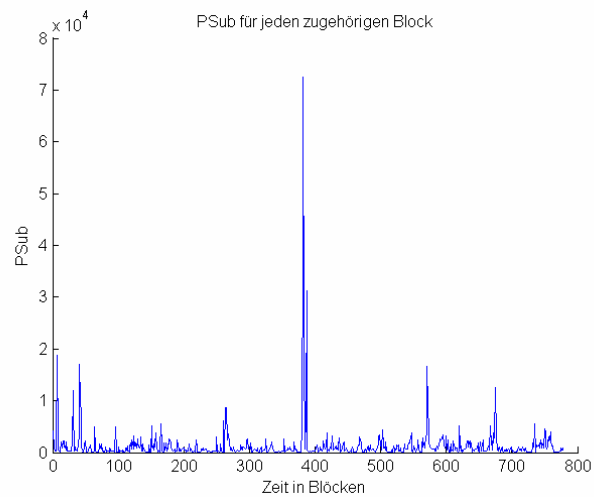


**Abbildung 4-21**  $P_{Sub}$  für weißes Rauschen

Man sieht hier, dass der Wert 12 nie überschritten wird. Das bedeutet, das Verhältnis zwischen Maximalwert und dem speziellen Durchschnittswert ist nie größer als 12. Für breitbandiges gaußverteiltes Rauschen war auch zu erwarten, dass der Maximalwert nicht sehr hoch sein wird. Da für eine Applikation von nichtstationären Störungen ausgegangen wird, muss in weiterer Folge ein Schätzer für das Rauschen eingeführt werden, der ständig adaptiert wird.

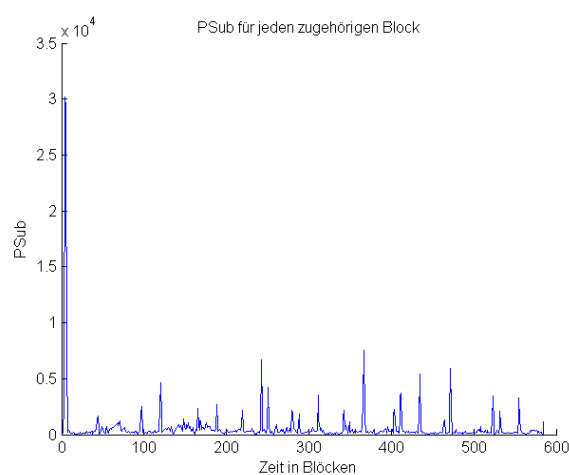
Für die folgenden Tests, Auswertungen und den Vergleich mit der Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins wird zur besseren Darstellung von stationären Verhältnissen ausgegangen.

Der gleiche Versuch wird nun für ein Sprachsignal durchgeführt. Abbildung 4-22 zeigt das Ergebnis.



**Abbildung 4-22**  $P_{\text{Sub}}$  für ein Sprachsignal

Für diese Art von Signal ergeben sich sehr viel höhere Werte. Allerdings treten sehr hohe Werte nur innerhalb sehr kurzer Zeit auf.



**Abbildung 4-23**  $P_{\text{Sub}}$  für ein Musiksignal

Abbildung 4-23 zeigt  $P_{\text{Sub}}$  für ein Musiksignal. Auch hier werden sehr hohe Spitzenwerte für  $P_{\text{Sub}}$  erreicht, die aber auch nur von kurzer Dauer sind.

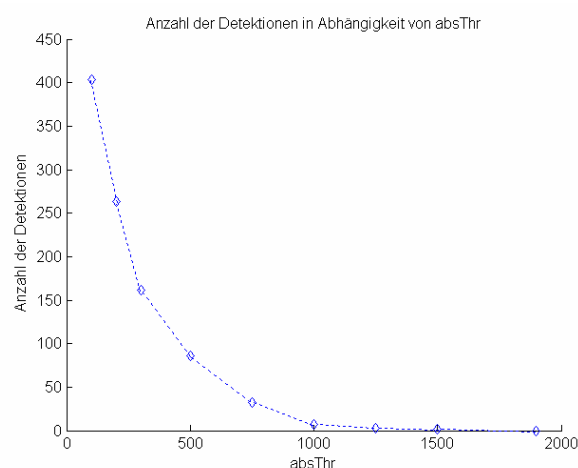
Die frei wählbaren Parameter ergeben sich zu:

- 1) zeitlicher Schwellwert ( $S_z$ )
- 2) absoluter Schwellwert ( $S_a$ )

Der zeitliche Schwellwert wirkt sich natürlich direkt auf die minimale Detektionszeit aus. Für die ersten Tests wird er auf 7 Blöcke gesetzt. Somit ist ein Vergleich zur Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins möglich. Die minimale Detektionszeit beträgt 8 Blöcke.

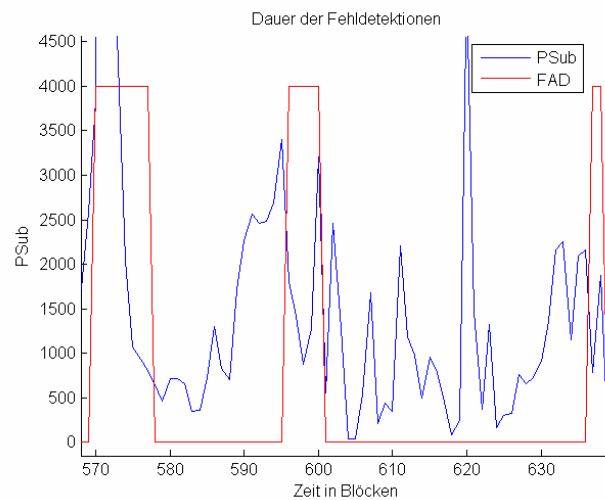
Ein Kriterium für die Wahl des Parameters  $S_a$  sind die Fehldetektionen. Bei einem reinen Sprachsignal als Testsignal sollten keine Fehldetektionen auftreten, da wegen der kurzen Quasistationarität, von etwa 10-30ms [7] von Sprache, keine sehr lang anhaltenden, schmalbandigen Anteile vorhanden sind.

$S_a$  wird nun auf Werte zwischen 100 und 1900 gesetzt und die Anzahl der Blöcke gezählt, in denen Rückkoppelfrequenzen detektiert werden. Als Testsignal dient wiederum ein Sprachsignal. Das Ergebnis ist in Abbildung 4-24 zu sehen. Die Fehldetektionen nehmen mit dem Anstieg von  $S_a$  deutlich ab.



**Abbildung 4-24** Anzahl der Detektionen in Abhängigkeit von  $S_a$ ,  $S_z=7$  Blöcke

Man kann auch zeigen, dass eine vertretbare Erhöhung des zeitlichen Schwellwerts keine Verbesserung des Detektionsverhaltens bringen würde, da die meisten Fehledetektionen deutlich länger als 2 bis 3 Blöcke andauern. Man müsste  $S_z$  um 8 Blöcke erhöhen um alle Fehledetektionen zu eliminieren. Das würde zu einer nicht vertretbaren Mindestlatenzzeit von 16 Blöcken also 368ms führen. Abbildung 4-25 verdeutlicht dies anhand von  $P_{\text{Sub}}$  aufgetragen über der Zeit in Blöcken und der zugehörigen Detektion.



**Abbildung 4-25**  $P_{\text{Sub}}$  aufgetragen über der Zeit und zugehörige Detektion,  $S_a=750$ ,  $S_z=7$  Blöcke

## 4.2.1 Glättung des Spektrums

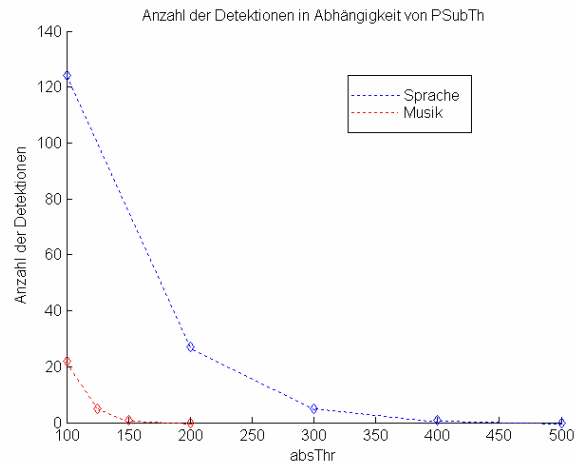
Es wird nun für die SFM-Detektion die gleiche spektrale Glättung wie bei der Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins eingeführt, also

$$|X_j(n)|^2 = \alpha \cdot |X_j(n)|^2 + (1 - \alpha) \cdot |X_j(n-1)|^2 \quad \forall j = 0 \dots N-1$$

$N$ ... Länge der FFT bzw. Blockgröße

$X_j$  ist also der  $j$ -te Frequenz-Bin,  $\alpha$  der Vergessensfaktor und  $n$  der aktuelle Block.

Mit  $S_z=7$  und  $\alpha=0.2$  werden wieder die Fehledetektionen in Abhängigkeit von  $S_a$  gezählt (siehe Abb. 4-26). Als Testsignale dienen ein Sprach- und ein Musiksignal.

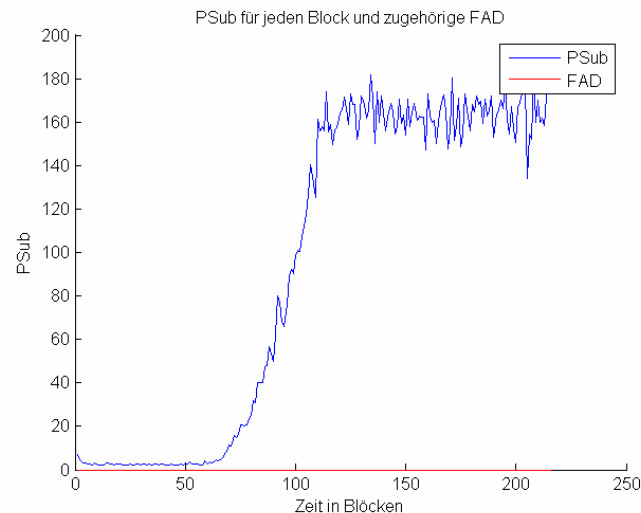


**Abbildung 4-26** Sprach- und Musiksignal als Eingangssignal und zugehörige Detektion,  $S_z=7$ ,  $\alpha=0.2$ ; Signallänge  $l=10$  Sekunden

Die Fehldetektionen nehmen auch hier deutlich mit der Erhöhung von  $S_a$  ab. In Anbetracht des Ergebnisses für das Sprachsignal muss man  $S_a$  mindestens auf 200 setzen, eigentlich sogar auf 300. Bei  $S_a=300$  sind es nur noch zwei, bei  $S_a=500$  treten keine Fehldetektionen mehr auf.

Für ein Musiksignal muss dieser Schwellwert nicht ganz so hoch gewählt werden. Aber Abbildung 4-26 zeigt, dass ein Schwellwert von 100 auch hier zu Fehldetektionen führt. Bei einem Schwellwert von 200 treten für dieses Musiksignal keine Fehldetektionen mehr auf.

Auch für diese Methode wird ein Sinuston mit kontinuierlich ansteigender Amplitude als Testsignal verwendet und mit weißem, gaußverteilterm Rauschen überlagert. Der Pegel-Schwellwert wird auf 200 gesetzt und das Testsignal wird auf Koppelfrequenzen überprüft. Es ergibt sich für die SFM-Methode keine Detektion, weil  $P_{Sub}$  nie den Wert 200 übersteigt (siehe Abbildung 4-27).



**Abbildung 4-27** Sinuston mit kontinuierlich ansteigender Amplitude als Testsignal,  $P_{\text{Sub}}$  für jeden Block,  $S_a=200$ ,  $S_z=7$ ,  $\alpha=0.2$

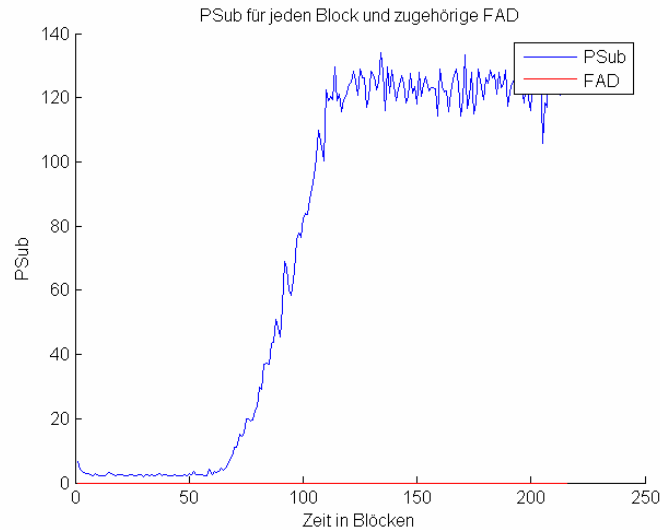
## 4.2.2 Nullsetzen der größten drei Frequenz-Bins - Berechnung von $P_{\text{Sub}}$

Obiger Test wird erneut durchgeführt, mit dem Unterschied, dass diesmal die drei höchsten Frequenz-Bins zur Berechnung von  $P_{\text{Sub}}$  nicht Null gesetzt werden.

Dies wird bei der Durchschnittsbildung insofern berücksichtigt, als dass hier durch die Anzahl *aller* Frequenz-Bins dividiert wird.

$$P_{AV} = \frac{P_{Ges}}{N}$$

$$P_{Sub} = \frac{P_{\max}}{P_{AV}} = \frac{P_s(f_1)}{P_{AV}}$$



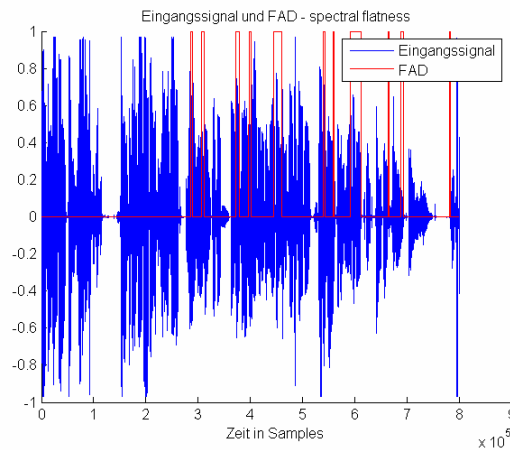
**Abbildung 4-28**  $P_{\text{Sub}}$  für jeden Block,  $S_a=200$ ,  $S_z=7$ ,  $\alpha=0.2$ , kein Nullsetzen der höchsten drei Frequenz-Bins für die Berechnung von  $P_{\text{Sub}}$

Man sieht in Abbildung 4-28, dass  $P_{\text{Sub}}$  deutlich niedrigere Werte als in Abbildung 4-27 aufweist. Das Nullsetzen der Frequenz-Bins macht die Detektion also sichtbar empfindlicher. Wäre bei diesem Test ein absoluter Schwellwert von etwa 110 nötig damit eine Detektion möglich wäre, so dürfte dieser unter obiger Berechnungsmethode um etwa 40 höher, also bei ca. 150 liegen. Der Patentvorschlag bringt also eine Verbesserung.

### 4.3 Vergleich der beiden Detektionsmethoden

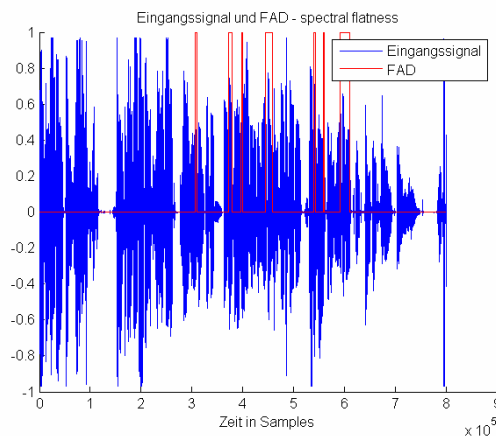
Es stellt sich trotz obiger Verbesserung die Frage, ob die SFM-Detektionsmethode empfindlich genug ist.

Die Erhöhung des zeitlichen Schwellwert innerhalb eines akzeptablen Bereiches bringt wie schon oben erwähnt nicht die nötige Verbesserung. Es ergeben sich trotzdem zu viele Fehldetektionen. Abbildung 4-29 verdeutlicht dies noch mal.



**Abbildung 4-29** Sprachsignal als Eingangssignal und zugehörige Detektion,  $S_a=100$ ,  $S_z=10$ ,  $\alpha=0.2$

Auch eine höhere Glättung  $\alpha=0,1$  bringt nicht die erforderliche Genauigkeit (siehe Abbildung 4-30)



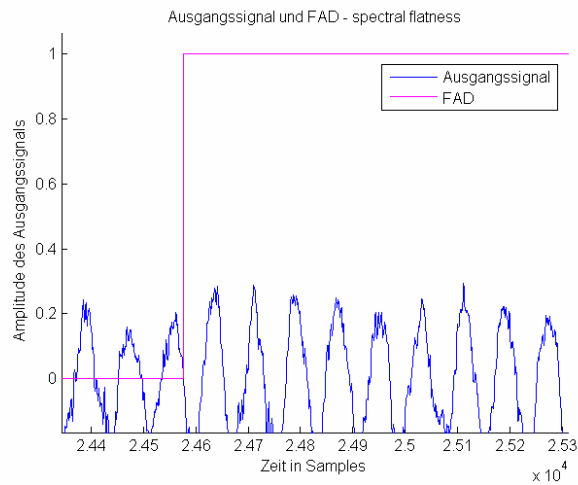
**Abbildung 4-30** Sprachsignal als Eingangssignal und zugehörige Detektion,  $S_a=100$ ,  $S_z=10$ ,  $\alpha=0.1$

Die Frage ist nun noch, ob der verwendete, in der Amplitude ansteigende Sinuston ein realistisches Testsignal ist, oder ob ein absoluter Schwellwert von 300 ein ausreichend empfindliches Kriterium für auftretende Koppelfrequenzen darstellt?

Dazu werden wie schon bei der Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins realistische Koppelfrequenzen durch Rückkopplung erzeugt. Das Blockschaltbild zur Rückkopplungserzeugung ist in Abbildung 4-15 zu sehen.

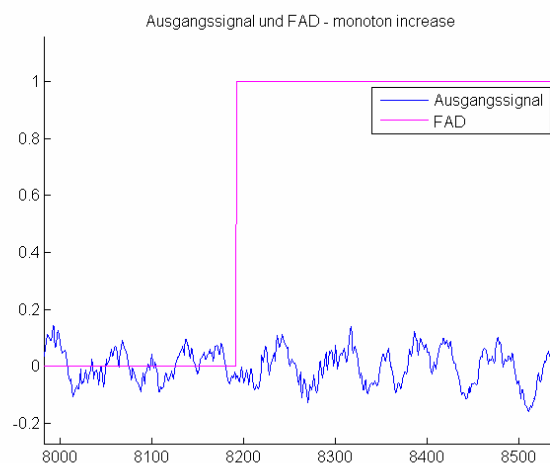


Man sieht in Abbildung 4-31, dass eine Rückkopplfrequenz erkannt wird, und zwar nach 24580 Samples.



**Abbildung 4-31** weißes Rauschen als Eingangssignal und zugehörige Detektion,  $S_a=300$ ,  $S_z=7$ ,  $\alpha=0.2$ , Detailausschnitt

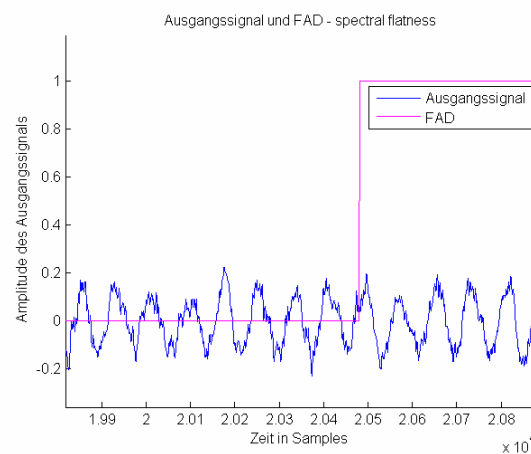
Zum Vergleich wird mit dem gleichen Eingangssignal und der gleichen Verstärkung der Test für die Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins durchgeführt.



**Abbildung 4-32** weißes Rauschen als Eingangssignal und zugehörige Detektion, zeitl. Schwellwert: 8 Blöcke,  $\alpha=0.2$ , Detailausschnitt

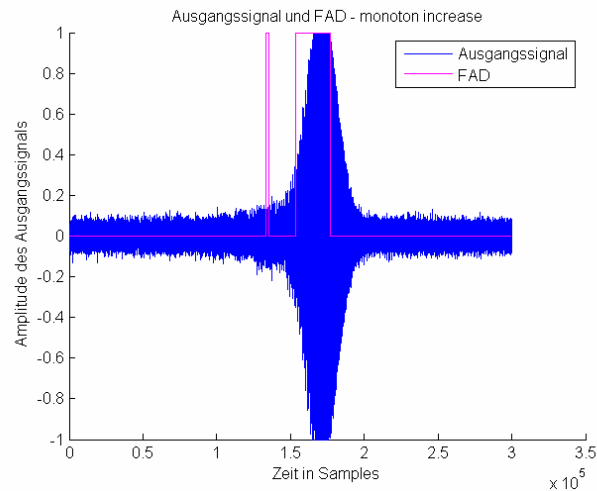
Die Koppelfrequenz wird hier schon nach etwa 8200 Sample erkannt. Das bedeutet, die Detektion findet bei einer Abtastfrequenz von 44,1kHz etwa 0,37 Sekunden früher statt. Dieses Ergebnis spricht eindeutig für die Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins.

Selbst bei Verringerung des absoluten Schwellwerts auf  $S_a=100$ , was ja zu den oben angesprochenen Fehldetektionen führt, liegt die Detektionszeit noch deutlich höher als bei der Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins, nämlich bei ungefähr 20500 Samples. Das entspricht bei  $f_s=44,1\text{kHz}$  etwa 465ms (siehe Abbildung 4-33).

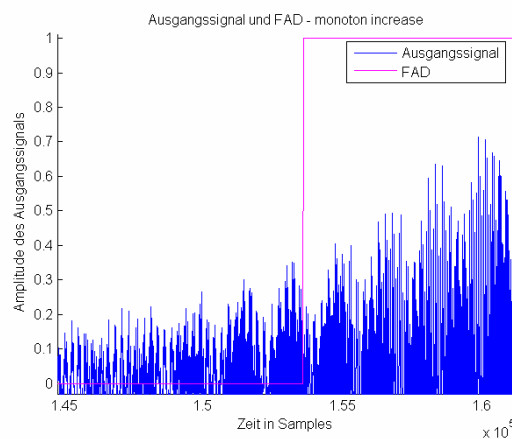


**Abbildung 4-33** weißes Rauschen als Eingangssignal und zugehörige Detektion,  $S_a=300$ ,  $S_z=7$ ,  $\alpha=0,2$ ,  
Detailausschnitt

Für einen weiteren Vergleichstest wird erneut eine realistische Rückkopplfrequenz erzeugt, die anschwillt und wieder abfällt. Dies wird über eine veränderliche Schleifenverstärkung  $g$  erreicht. Die Koppelfrequenz liegt bei 714 Hz.



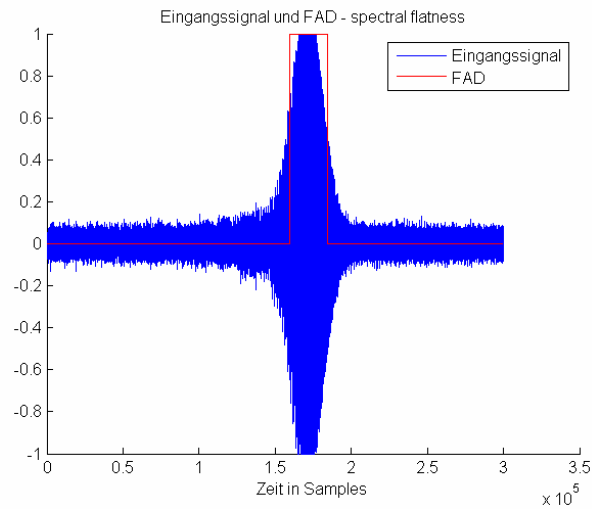
**Abbildung 4-34** Testsignal mit Koppelfrequenz und zugehöriger Detektion, zeitl. Schwellwert: 8 Blöcke,  $\alpha=0.2$



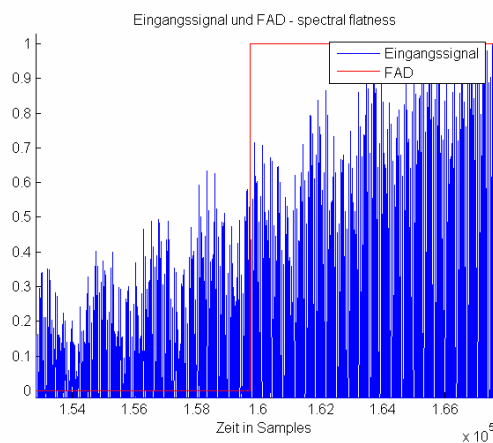
**Abbildung 4-35** Testsignal mit Koppelfrequenz und zugehörige Detektion, zeitl. Schwellwert: 8 Blöcke,  $\alpha=0.2$ ,  
Detailausschnitt

Es ist zunächst eine kurze Fehldetektion zu sehen. Die Koppelfrequenz wird bei der Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins nach etwa 153500 Samples erkannt. Die Abbildungen 4-34 und 4-35 zeigen dies.

Das Ergebnis für die *spectral flatness measure* Methode ist den Abbildungen 4-36 und 4-37 zu entnehmen.

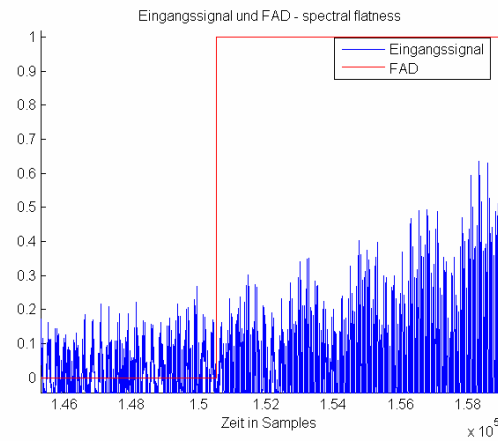


**Abbildung 4-36** Testsignal mit Koppelfrequenz und zugehörige Detektion,  $S_a=300$ ,  $S_z=7$ ,  $\alpha=0.2$



**Abbildung 4-37** Testsignal mit Koppelfrequenz und zugehörige Detektion,  $S_a=300$ ,  $S_z=7$ ,  $\alpha=0.2$

Die Rückkopplfrequenz wird hier etwa nach 160000 Samples detektiert, also etwas später als oben.



**Abbildung 4-38** Testsignal mit Koppelfrequenz und zugehörige Detektion,  $S_a=100$ ,  $S_z=7$ ,  $\alpha=0.2$

Verringert man  $S_a$  auf 100, so verkürzt sich die Detektionszeit und die Koppelfrequenz wird schon etwa nach 150500 Samples erkannt. (s. Abbildung 4-38)

Die Methode der streng monoton ansteigenden Leistung eines Frequenz-Bins liefert insgesamt eindeutig bessere Ergebnisse und wird deshalb auch im Gesamtsystem verwendet.

## 5 Filter- und Filterbankentwurf

### 5.1 Parametrisches Kerbfilter

Für die Implementierung der Filterbank in Matlab bzw. Pure Data (PD) werden Peak Filter zweiter Ordnung verwendet. Der Entwurf der Filterbank wird in Kapitel 6-3 näher erläutert.

Die Koeffizienten für die Filter werden aus folgender Übertragungsfunktion ermittelt [15]:

$$H(z) = 1 + \frac{H_0}{2} [1 - A_2(z)] , \quad [5.1]$$

$$\text{mit } H_0 = V_0 - 1, \quad V_0 = 10^{\frac{G}{20}}$$

$V_0$ ... Verstärkung,  $G$ ... Verstärkung in dB

und zugehöriger analoger Allpass-Übertragungsfunktion zweiter Ordnung

$$A_2(P) = \frac{1 - \frac{P}{Q} + P^2}{1 + \frac{P}{Q} + P^2}, \quad [5.2]$$

$$\text{mit } P = \frac{j\omega}{\omega_g} = j \frac{f}{f_g} = j\Omega \quad f_g \dots \text{ Grenzfrequenz}$$

wobei obiges Filter mittels Bilineartransformation mit  $P = L \cdot \frac{1 - z^{-1}}{1 + z^{-1}}$  und  $L = \cot\left(\frac{\pi \cdot f_0}{f_s}\right)$  [11]

in den zeitdiskreten Bereich überführt werden kann. Die Güte des Filters berechnet sich über die Bandbreite bei  $-3\text{dB}$ .

Das Blockschaltbild in Abbildung 5-1 zeigt die Struktur eines Peak Filters 2. Ordnung. Es handelt sich um ein Filter mit rückgekoppelter Allpass-Struktur.

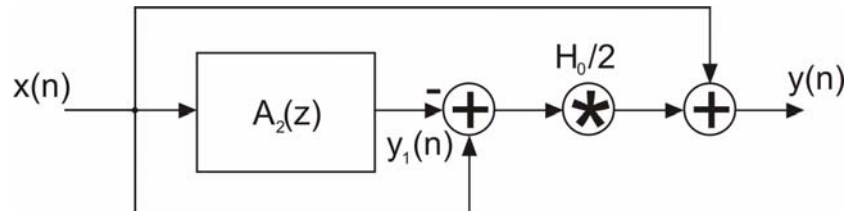


Abbildung 5-1 Peak Filter 2.Ordnung

Einsetzen in [5.2] ergibt:

$$A_2(z) = \frac{1 - \frac{L(1-z^{-1})}{Q(1+z^{-1})} + \frac{L^2(1-z^{-1})^2}{(1+z^{-1})^2}}{1 + \frac{L(1-z^{-1})}{Q(1+z^{-1})} + \frac{L^2(1-z^{-1})^2}{(1+z^{-1})^2}} = \frac{\underline{Q}(1+z^{-1})^2 - L(1-z^{-1})(1+z^{-1}) + \underline{Q}L^2(1-z^{-1})^2}{\underline{Q}(1+z^{-1})^2 + L(1-z^{-1})(1+z^{-1}) + \underline{Q}L^2(1-z^{-1})^2}$$

durch Kürzen und Zusammenfassen der Terme mit gleicher Potenz von  $z$  erhält man:

$$A_2(z) = \frac{z^{-2}(\underline{Q}(1+L^2)+L) + z^{-1}(2\underline{Q}(1-L^2)) + \underline{Q}(1+L^2) - L}{z^{-2}(\underline{Q}(1+L^2)-L) + z^{-1}(2\underline{Q}(1-L^2)) + \underline{Q}(1+L^2) + L} = \frac{Z(z)}{N(z)}$$

Einsetzen in [5.1] ergibt:

$$H(z) = 1 + \frac{H_0}{2} - \frac{H_0}{2} \cdot A_2(z) = \frac{2 \cdot N(z) + H_0 \cdot N(z) - H_0 \cdot Z(z)}{2 \cdot N(z)}$$

durch Ausmultiplizieren und Zusammenfassen der Terme mit gleicher Potenz von  $z$  ergibt sich:

$$H(z) = \frac{b_3 \cdot z^{-2} + b_2 \cdot z^{-1} + b_1}{a_3 \cdot z^{-2} + a_2 \cdot z^{-1} + a_1}, \text{ mit}$$

$$a_1 = 2Q + 2QL^2 + 2L$$

$$a_2 = 4Q - 4QL^2$$

$$a_3 = 2Q - 2L + 2QL^2$$

$$b_1 = 2Q + 2QL^2 + 2L + 2H_0L$$

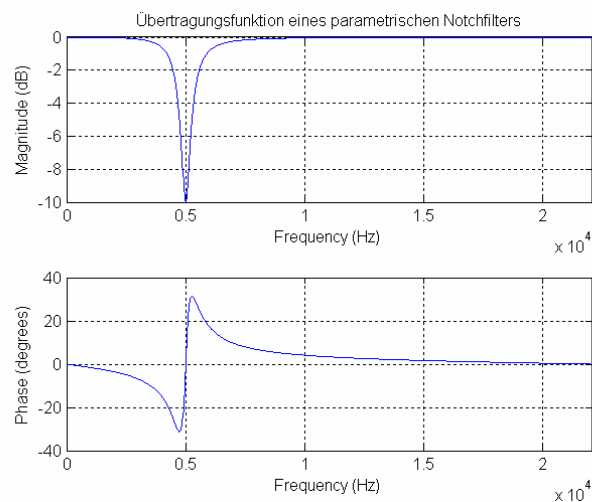
$$b_2 = 4Q - 4QL^2$$

$$b_3 = 2Q - 2L + 2QL^2 - 2H_0L$$

Die Differenzen-Gleichung des Filters lautet also:

$$y[n] = \frac{b_1}{a_1} \cdot x[n] + \frac{b_2}{a_1} \cdot x[n-1] + \frac{b_3}{a_1} \cdot x[n-2] - \frac{a_2}{a_1} \cdot y[n-1] - \frac{a_3}{a_1} \cdot y[n-2] \quad [5.3]$$

Die Übertragungsfunktion des Filters für den Fall  $Q=5$  und  $G=10\text{dB}$  ist in Abbildung 5-2 zu sehen.



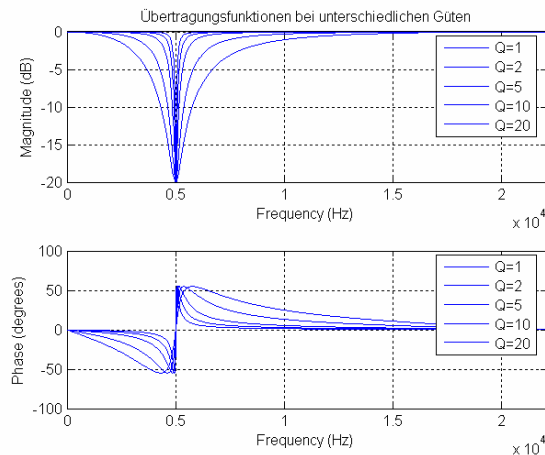
**Abbildung 5-2** Übertragungsfunktion  $Q=5$ ,  $G=10\text{dB}$

Durch Variation von  $Q$  lässt sich die Bandbreite  $bw$  der Kerbe einfach einstellen.

$$Q = \frac{f_g}{bw} \quad \rightarrow \quad bw = \frac{f_g}{Q}$$

Abbildung 5-3 zeigt dies.





**Abbildung 5-3** Übertragungsfunktionen mit unterschiedlich breiten Kerben durch Variation von  $Q$ ,  $G=20$

In [15] ergibt sich aus Gleichung [5.1]:

$$H(z) = 1 + \frac{H_0}{2} [1 - A_2(z)],$$

mit dem Allpassfilter 2.Ordnung:

$$A_2(z) = \frac{-a_c + (d - da_c)z^{-1} + z^{-2}}{1 + (d - da_c)z^{-1} - a_c z^{-2}} \quad [5.4]$$

ein Peak Filter 2.Ordnung.

Dabei ist

$$d = -\cos\left(\frac{2\pi f_c}{f_s}\right)$$

$$V_0 = H(e^{j2\pi f_c / f_s}) = 10^{-\frac{G}{20}}$$

$$H_0 = V_0 - 1$$

Die Bandbreite  $f_b$  wird dabei durch den Parameter  $a_c$  bestimmt und ergibt sich aus:

$$a_c = \frac{\tan\left(\frac{\pi \cdot f_b}{f_s}\right) - V_0}{\tan\left(\frac{\pi \cdot f_b}{f_s}\right) + V_0}$$

Daraus lassen sich die Koeffizienten für ein Biquad-Filter mit einer Differenzgleichung wie Gleichung [5.3] folgendermaßen berechnen:

$$a_1 = 1 \quad a_2 = \frac{2(K^2 - 1)}{1 + \frac{V_0}{Q}K + K^2} \quad a_3 = \frac{1 - \frac{V_0}{Q}K + K^2}{1 + \frac{V_0}{Q}K + K^2}$$

$$b_1 = \frac{1 + \frac{1}{Q}K + K^2}{1 + \frac{V_0}{Q}K + K^2} \quad b_2 = \frac{2(K^2 - 1)}{1 + \frac{V_0}{Q}K + K^2} \quad b_3 = \frac{1 - \frac{1}{Q}K + K^2}{1 + \frac{V_0}{Q}K + K^2}$$

mit

$$K = \tan\left(\frac{\pi \cdot f_c}{f_s}\right)$$

$Q$  berechnet sich hier aus der Bandbreite, die über einen 3dB-Abfall in der Übertragungsfunktion definiert ist.

Da eine Definition der Güte über die Bandbreite bei -3dB unpraktisch ist und in manchen Fällen dazu führt, dass diese Bandbreite nicht mehr definiert ist, wird im folgenden die Übertragungsfunktion eines parametrischen Bandpasses mit einer geschickteren Güte-Definition hergeleitet.

## 5.2 Berechnung der Half Transition Bandbreite

$$Q_{\text{half-Transition}} = \frac{f_g}{bw_{\text{half-Transition}}},$$

wobei  $bw_{\text{half-Transition}}$  hier die Bandbreite bei der halben Verstärkung bzw. Dämpfung ist.

$$H(P) = 1 + \frac{H_0}{2} \left( 1 - \frac{1 - \frac{P}{Q} + P^2}{1 + \frac{P}{Q} + P^2} \right) = 1 + \frac{H_0}{2} \left( \frac{1 + \frac{P}{Q} + P^2 - 1 + \frac{P}{Q} - P^2}{1 + \frac{P}{Q} + P^2} \right) = \frac{1 + \frac{P}{Q} + P^2 + H_0 \frac{P}{Q}}{1 + \frac{P}{Q} + P^2}$$

$$= \frac{1 + (H_0 + 1) \frac{P}{Q} + P^2}{1 + \frac{P}{Q} + P^2},$$

mit  $P = j \frac{f}{f_g} = j\Omega$  ergibt sich:

$$H(j\Omega) = \frac{1 + j \frac{1}{Q} \Omega (1 + H_0) - \Omega^2}{1 + j \frac{1}{Q} \Omega - \Omega^2}.$$

Da an der Grenzfrequenz

$$H(j\Omega)|_{\Omega=1} = \frac{1 + j \frac{(1+H_0)}{Q} - 1}{1 + j \frac{1}{Q} - 1} = 1 + H_0$$

gilt, folgt für die Frequenzen bei der halben Verstärkung:

$$|H(j\Omega)|_{\frac{1}{2}} = \sqrt{1+H_0} = \frac{\sqrt{(1-\Omega^2)^2 + \frac{\Omega^2}{Q^2} (1+H_0)^2}}{\sqrt{(1-\Omega^2)^2 + \frac{\Omega^2}{Q^2}}} \rightarrow$$

$$\sqrt{(1-\Omega^2)^2 \cdot (1+H_0) + \frac{\Omega^2}{Q^2} \cdot (1+H_0)} = \sqrt{(1-\Omega^2)^2 + \frac{\Omega^2}{Q^2} \cdot (1+H_0)^2}$$

$$(1-\Omega^2)^2 \cdot H_0 + \frac{\Omega^2}{Q^2} (1+H_0 - 1 - 2H_0 - H_0^2) = 0$$

$$(1-\Omega^2)^2 \cdot H_0 - \frac{\Omega^2}{Q^2} \cdot H_0 \cdot (1+H_0) = 0 \quad | \quad \frac{1}{H_0}$$

$H_0=0$  ist die triviale Lösung und wird nicht weiter betrachtet

$$\left[ (1-\Omega^2) + \frac{\Omega}{Q} \sqrt{1+H_0} \right] \cdot \left[ (1-\Omega^2) - \frac{\Omega}{Q} \sqrt{1+H_0} \right] = 0 \rightarrow$$

$$\Omega_{1,2}^2 - \frac{\Omega_{1,2}}{Q} \cdot \sqrt{1+H_0} - 1 = 0$$

$$\Omega_{3,4}^2 + \frac{\Omega_{3,4}}{Q} \cdot \sqrt{1+H_0} - 1 = 0$$

$$\Omega_1 = \frac{\sqrt{1+H_0}}{2Q} + \sqrt{\frac{1+H_0}{4Q^2} + 1} > 0$$

[5.5]

$$\Omega_2 = \frac{\sqrt{1+H_0}}{2Q} - \sqrt{\frac{1+H_0}{4Q^2} + 1} < 0 \quad \rightarrow \quad \text{keine Lösung}$$

$$\Omega_3 = -\frac{\sqrt{1+H_0}}{2Q} + \sqrt{\frac{1+H_0}{4Q^2} + 1} > 0 \quad [5.6]$$

$$\Omega_4 = -\frac{\sqrt{1+H_0}}{2Q} - \sqrt{\frac{1+H_0}{4Q^2} + 1} < 0 \quad \rightarrow \quad \text{keine Lösung}$$

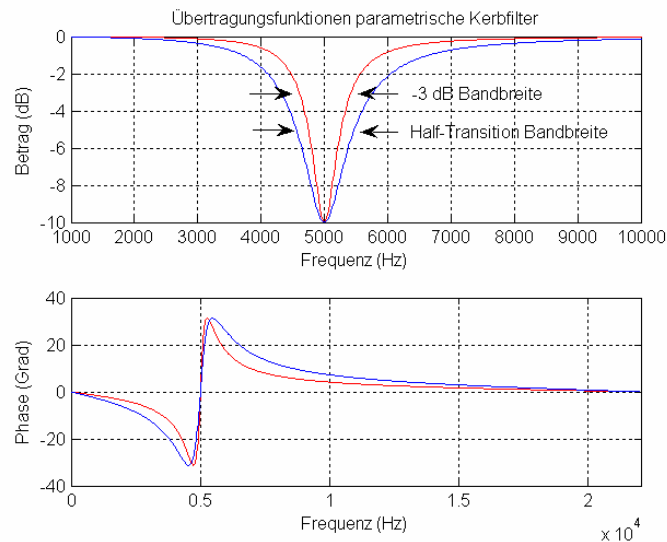
$$\frac{bw}{f_0} = \Omega_1 - \Omega_3 = \frac{\sqrt{1+H_0}}{Q}$$

$$\frac{Q}{\sqrt{1+H_0}} = \frac{f_g}{bw} \quad \rightarrow$$

$$Q_{\text{half-Transition}} = \frac{Q_{BP}}{\sqrt{V_0}} \quad [5.7]$$

$$H(j\Omega) = \frac{1 + j \frac{1}{Q_{\text{half-Transition}}} \Omega \cdot \sqrt{V_0} - \Omega^2}{1 + j \frac{1}{Q_{\text{half-Transition}}} \Omega \cdot \frac{1}{\sqrt{V_0}} - \Omega^2}$$

Abbildung 5-4 zeigt die Übertragungsfunktionen für die beiden unterschiedlichen Güte­defini­tionen.



**Abbildung 5-4** Übertragungsfunktionen parametrischer Kerbfilter mit  $-3\text{dB}$ -Güte und Half-Transition-Güte

## 5.3 Filterbankentwurf

Im folgenden Abschnitt werden verschiedene Möglichkeiten eines Filterbankentwurfs vorgestellt und miteinander verglichen.

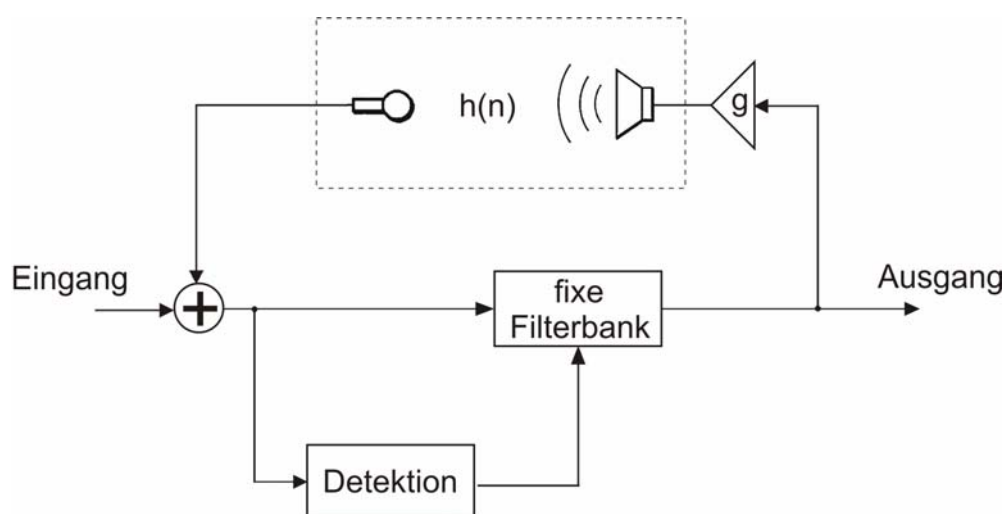
### 5.3.1 Terzfilterbank

Für die Eliminierung der Rückkopplfrequenzen soll eine Filterbank entworfen werden. Zunächst wird hierzu ein Terzband-Equalizer verwendet. Dazu wird das Spektrum von 25 Hz bis 12,5 kHz in Terzbänder aufgeteilt. Die Mittenfrequenzen einer Terzfilterbank sind Tabelle 5-1 zu entnehmen.

| Nr. | Mittenfrequenz [Hz] | Nr. | Mittenfrequenz [Hz] | Nr. | Mittenfrequenz [Hz] | Nr. | Mittenfrequenz [Hz] |
|-----|---------------------|-----|---------------------|-----|---------------------|-----|---------------------|
| 1   | 25                  | 8   | 125                 | 15  | 630                 | 22  | 3150                |
| 2   | 31.5                | 9   | 160                 | 16  | 800                 | 23  | 4000                |
| 3   | 40                  | 10  | 200                 | 17  | 1000                | 24  | 5000                |
| 4   | 50                  | 11  | 250                 | 18  | 1250                | 25  | 6300                |
| 5   | 63                  | 12  | 315                 | 19  | 1600                | 26  | 8000                |
| 6   | 80                  | 13  | 400                 | 20  | 2000                | 27  | 10000               |
| 7   | 100                 | 14  | 500                 | 21  | 2500                | 28  | 12500               |

**Tabelle 5-1** Mittenfrequenzen einer Terzfilterbank

Für jedes Terzband wird nun ein Kerbfilter (vgl. Kapitel 5.1) mit fester Mittenfrequenz bereitgestellt. Die Filter werden in Serie geschaltet, so dass das Ausgangssignal des ersten Filters als Eingangssignal des zweiten dient usw. Die Güte  $Q$  für die Filter muss dabei so gewählt werden, dass eine Filterung an jeder Stelle des Frequenzspektrums möglich ist. Die Dämpfung soll adaptiv angepasst werden. Das Blockschaltbild ist in Abbildung 5-5 zu sehen.



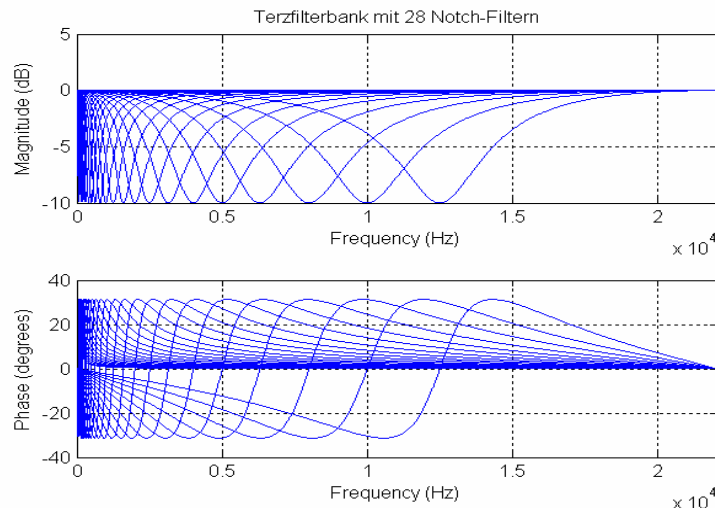
**Abbildung 5-5** Gesamtsystem mit fixer Terzfilterbank

Wählt man beispielsweise  $Q=1$  und  $G=10\text{dB}$ , so ergeben sich die Übertragungsfunktionen wie in Abbildung 5-6.

Aus der Güte  $Q$  und der Mittenfrequenz  $f_g$  lässt sich mit:

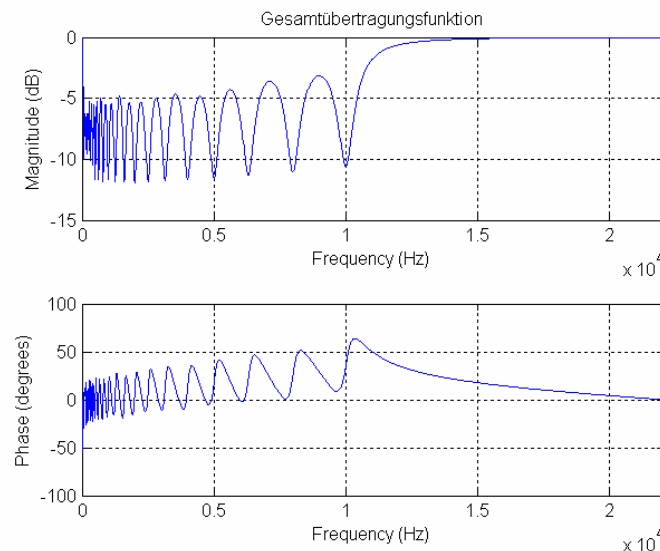
$$bw = \frac{f_g}{Q}$$

die Bandbreite  $bw$  für jedes Filter ermitteln.



**Abbildung 5-6** Übertragungsfunktionen der 28 Kerbfiler einer Terzfilterbank,  $G=10\text{dB}$ ,  $Q=1$

Leider sind die Kerben der Filter nicht besonders schmal, so dass das Eingangssignal, vor allem bei hohen Frequenzen, nicht nur direkt an der Stelle der auftretenden Rückkoppelfrequenz beeinflusst bzw. gedämpft wird, sondern auch in seiner unmittelbaren Umgebung. Wählt man aber eine höhere Güte  $Q=5$  und verringert somit die Bandbreite des Filters, so ist es nicht mehr möglich an allen Frequenzstellen Koppelfrequenzen zu eliminieren. Abbildung 5-7 verdeutlicht dies. Mit einer Filterbank mit fixen Filtern ist es schwierig, Frequenzen, die zwischen zwei Mittenfrequenzen der Filterbank liegen gut zu dämpfen. Man muss dann entweder einen der beiden Filter mit einer sehr hohen Dämpfung versehen, oder beide Filter einsetzen, um diese Koppelfrequenz zu unterdrücken. Beides bedeutet einen weitaus größeren Eingriff in das Signal, als eine Filterung mit einem adaptivem Kerbfiler, der seine Kerbe genau auf die zu unterdrückende Koppelfrequenz setzt.



**Abbildung 5-7** Gesamtübertragungsfunktion der 28 Kerbfiler einer Terzfilterbank,  $G=10\text{dB}$ ,  $Q=5$

Eine solche Terzfilterbank wurde in *Pure Data* zu Testzwecken implementiert aber aus Gründen der hohen Rechenintensität und ihrer Ineffizienz wieder verworfen.

### 5.3.2 Parametrische adaptive Filterbank

Da der Rechenaufwand in Anbetracht der Effizienz für obige fixe Filterbank insbesondere für eine Echtzeitanwendung zu hoch ist, wird im folgenden eine Filterbank entworfen, die mit sieben parametrischen Kerbfiltern auskommt. Die Mittenfrequenzen der Filter sollen sich also adaptiv verändern, je nach dem an welcher Stelle im Frequenzband eine Rückkopplfrequenz auftritt.

Die Dämpfung der Filter ist adaptiv veränderbar, je nach dem wie stark eine Koppelfrequenz auftritt bzw. wie viele Koppelfrequenzen innerhalb eines kleinen Frequenzbandes auftreten. Die Bandbreite, die bei der halben Dämpfung definiert ist (siehe Kapitel 5.2), wird für alle Filter konstant auf eine große Terz (Frequenzverhältnis 5:4) eingestellt.

$$bw = f_o - f_u$$

$$\frac{f_o}{f_u} = \frac{f_g + bw/2}{f_g - bw/2} = \frac{5}{4}$$

$$4 \cdot \left[ f_g + \frac{bw}{2} \right] = 5 \cdot \left[ f_g - \frac{bw}{2} \right]$$

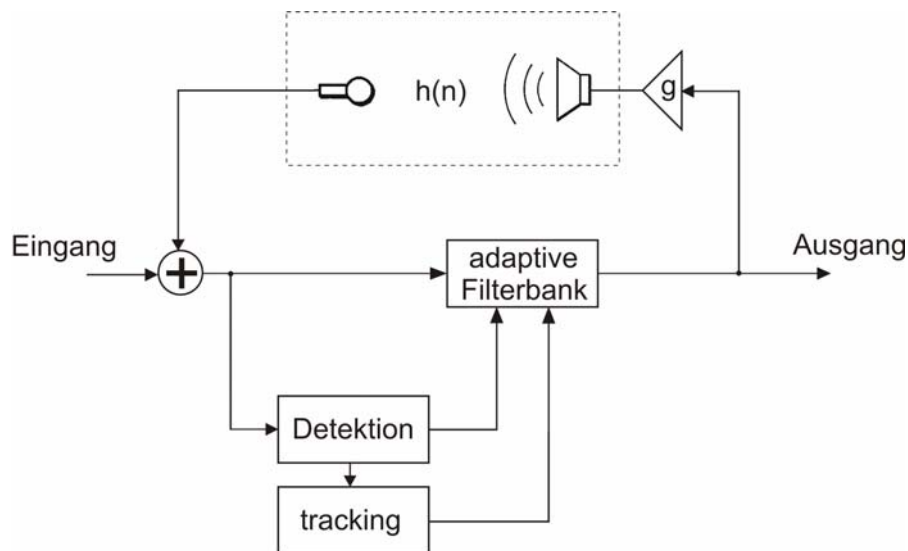
$$9bw = 2f_g$$



$$bw = \frac{2}{9} f_g$$

$$Q_{half-Transition} = \frac{f_g}{bw} = \frac{f_g}{\frac{2}{9} f_g} = 4,5$$

Durch einen speziellen Tracking-Algorithmus (siehe Kapitel 6.1) lassen sich bei günstiger Lage der Koppelfrequenzen mit den sieben zur Verfügung stehenden Filtern mehr als sieben Frequenzen unterdrücken, da ein Filter dessen Frequenz durch ein anderes Filter mitübernommen wird, freigegeben wird und eine achte Koppelfrequenz eliminieren kann. Abbildung 5-8 zeigt das Blockschaltbild.



**Abbildung 5-8** Gesamtsystem mit einer adaptiver Filterbank

## 6 Gesamtsystem

Im folgenden Kapitel wird beschrieben wie eine parametrische adaptive Filterbank angesteuert werden kann um effektiv Koppelfrequenzen zu unterdrücken.

### 6.1 Tracking

Für die in Kapitel 5.3.2 beschriebene Filterbank soll nun ein Tracking-Algorithmus entworfen werden. Ähnlich wie beim „partial tracking“, bei dem Teiltöne eines Klanges verfolgt werden, soll dieser Algorithmus die auftretenden Rückkopplfrequenzen verfolgen und geschickt einem der sieben Filter zuweisen. Die Anforderungen an den Algorithmus werden im folgenden beschrieben:

1. Mindestdauer einer Filterung.

Wird eine Rückkopplfrequenz detektiert, so muss eine gewisse Mindestdauer (z.B. 5 Sekunden) eingeführt werden, während der das Filter gesetzt ist, damit die Koppelfrequenz mit Sicherheit unterdrückt wird.

2. Attack- und Release-Zeit für die Filter

Die Filter werden mit einem *attack* und einem *release* für die Dämpfung versehen, um Signalverzerrungen zu vermeiden. Die *attack*-Zeit muss möglichst kurz (200ms) gewählt werden damit eine schnelle Unterdrückung der Koppelfrequenz gewährleistet ist. Nach diesem *attack* wird die Dämpfung mit einem *release* von der maximalen Dämpfung (5dB) auf 0dB abgesenkt. Die *release*-Zeit ergibt sich aus der Differenz von Mindestdauer und *attack*-Zeit, also zu 4,8 Sekunden.

3. Einflussbereich eines Filters

Der Einflussbereich  $B$  eines gesetzten Filters, in dem er andere Frequenzen mit unterdrücken kann, soll  $\pm 10\%$  seiner Mittenfrequenz  $f_m$  sein.

4. Übernehmen von Frequenzen

Koppelfrequenzen, die bereits aufgetreten sind, sollen von Filtern übernommen werden, die vorher auch diese Frequenzen unterdrückt haben.

Ist ein Filter bereits bei einer Frequenz  $f_m$  gesetzt, und wird während dieser Zeit die gleiche Koppelfrequenz detektiert, dann soll kein neues Filter gesetzt werden, sondern seine Dämpfung mit einem neuerlichen *attack* bis zur Maximaldämpfung angehoben werden. Wird eine Frequenz detektiert, die innerhalb des Einflussbereiches  $B$  des Filters liegt, so wird seine maximale Dämpfung erhöht und somit seine Güte verringert. Danach wird die Dämpfung wie oben mit einem *attack* auf die maximale Dämpfung angehoben.

## 5. Priorität der Koppelfrequenzen

Werden mehr als sieben Frequenzen detektiert, die alle einen eigenen Einflussbereich  $B$  haben, so werden die sieben Frequenzen unterdrückt, die am „gefährlichsten“ für die Stabilität des Systems sind. Das bedeutet es werden diejenigen Frequenzen unterdrückt, deren Amplituden am schnellsten ansteigen.

## 6.2 Nachklingen der Koppelfrequenzen

Mit dem System wie es in Abbildung 5-8 zu sehen ist, lassen sich zwar anschwellige Koppelfrequenzen sehr gut detektieren und unterdrücken, Frequenzen deren Schleifenverstärkung knapp unter 1 liegt, klingen aber sehr lange nach. Sie werden zwar nicht bei jedem Schleifendurchlauf verstärkt, werden aber nur sehr langsam abgeschwächt, so dass sie sehr lange in der akustischen Schleife zirkulieren. Dies führt zu einer starken Beeinträchtigung der Signalqualität. Abbildung 6-1 oben zeigt das Spektrogramm für ein Sprachsignal, dass durch solche nachklingenden Koppelfrequenzen gestört ist.

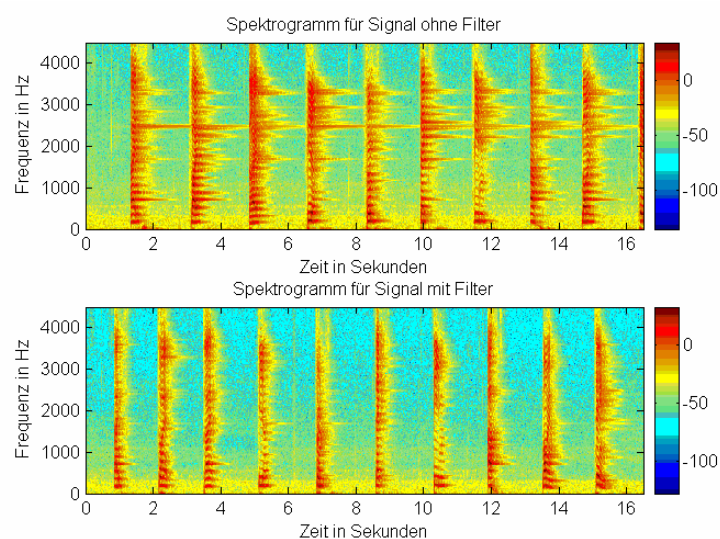


Abbildung 6-1 oben: Nachklingen der Koppelfrequenzen; unten: gefiltertes Signal

Um diese Koppelfrequenzen zu eliminieren, muss eine andere Detektionsmethode eingeführt werden, da die Leistung dieser Störfrequenzen nicht monoton ansteigt, sondern sehr langsam abfällt.

### 6.3 Glättung und Spitzenwert-Detektion

Für die Detektion dieser nachklingenden Frequenzen wird das Kurzzeitspektrum wiederum einer zeitlichen Mittelung unterzogen. Die Glättung ist diesmal aber deutlich höher. Der Vergessensfaktor  $\alpha$  beträgt nun den Wert 0,02. Somit ergibt sich für eine Blocklänge der FFT von  $N=2048$  beispielsweise eine Zeitkonstante von:

$$\tau = -\frac{1}{\ln(1-\alpha)} \cdot \frac{1}{fs^*} = -\frac{1}{\ln(0.98)} \cdot \frac{2048}{44100} \approx 2,3s$$

Es werden nun mit Hilfe einer Spitzenwert-Detektion diejenigen Frequenzen aus dem gemittelten Spektrum detektiert, die für das störende Nachklingen verantwortlich sind. Da die nachklingenden Koppelfrequenzen nur nach energiereichen Anregungen, wie zum Beispiel Sprache auftreten, wird die Aktualisierung des Spektrums zusätzlich mit der im Spektrum enthaltenen Energie  $P$  und einer Regularisierung, die eine Division durch Null verhindert, gewichtet.

$$Spektrum(neu) = \alpha \cdot Spektrum(alt) + (1-\alpha) \cdot \frac{FFT(neu)}{P + \beta}$$

$P$ : enthaltene Energie im Spektrum

$\alpha$ : Vergessensfaktor

$0 < \beta \ll 1$ : Regularisierung

Mit dieser Art der Aktualisierung wird vermieden, dass die relativ leise nachklingenden und energiearmen Störfrequenzen durch energiereiche Sprache im gemittelten Spektrum verdeckt werden.

Für die Spitzenwert-Detektion erweist sich eine Hanning-Fensterung aufgrund ihrer höheren Nebenkeulendämpfung als günstiger im Vergleich zu einer Rechteck-Fensterung.

Sei  $X[n]$  der zu untersuchende Bin,  $X[n-1]$ ,  $X[n-2]$  ... usw. seine linken Nachbarn,  $X[n+1]$ ,  $X[n+2]$  ... usw. seine rechten Nachbarn und  $P[n]$  seine Energie. Die Spitzenwert-Detektion detektiert aufgrund der 4 Bin breiten Hauptkeule des Hanning-Fensters und der daraus resultierenden Peakbreite  $Pb$  dann ein Maximum im gemittelten Spektrum, wenn gilt:

$$P[n]-P[n-1] > 8\text{dB und } P[n]-P[n+1] > 8\text{dB} \quad Pb=1$$

oder

$$P[n]-P[n-2] \text{ und } P[n-1]-P[n-2] > 8\text{dB und } P[n]-P[n+1] \text{ und } P[n-1]-P[n+1] > 8\text{dB} \quad Pb=2$$

oder

$$P[n]-P[n-1] \text{ und } P[n+1]-P[n-1] > 8\text{dB und } P[n]-P[n+2] \text{ und } P[n+1]-P[n+2] > 8\text{dB} \quad Pb=2$$

... usw.

Obige Bedingungen müssen bis zur Peakbreite  $Pb=4$  durchvariiert werden. Die benötigte Differenz zweier Bins von 8dB für eine Detektion ist im Hinblick auf Fehldetektionen empirisch ermittelt worden.

Es wurde ebenfalls versucht die nachklingenden Koppelfrequenzen mit dem in Kapitel 2.1 vorgestellten *adaptiven line enhancer* zu eliminieren. Zwar unterdrückt der *line enhancer* die Störfrequenzen sehr gut, jedoch unterdrückt er auch große Teile des Sprachsignals, so dass es zu Signalverzerrungen und einer zu starken Qualitätsminderung des Sprachsignals kommt. Der *adaptive line enhancer* ist für diese Anwendung nur bedingt geeignet.

## 6.4 Totzeit und adaptive Dämpfung

Für das Gesamtsystem werden zur Unterdrückung dieser Störfrequenzen 7 weitere Filter bereitgestellt. Grundsätzlich werden diese Filter wie in Kapitel 6.1 beschrieben getrackt. Bei einer ersten Detektion von einer Koppelfrequenz  $f_{Koppel}$  wird die Dämpfung eines Filters zunächst mit einem *attack* (200ms) auf einen Maximalwert von 3dB gesetzt. Aufgrund der langsamen Aktualisierung des Kurzzeitspektrums wird eine „Totzeit“, in der alle weiteren Detektionen von  $f_{Koppel}$  ignoriert werden, von 2 Sekunden für jedes Filter eingeführt. Diese Totzeit entspricht in etwa der Zeitkonstanten  $\tau=2,3$  Sekunden und wurde messtechnisch ermittelt. Dazu wurden mehrere Sinustöne mit unterschiedlicher Amplitude und Frequenz 10 Sekunden lang in das Gesamtsystem eingespeist und jene Zeit gemessen, die die Detektion nach Beendigung des jeweiligen Tons weiterhin eine Rückkoppelfrequenz detektiert. Während dieser Totzeit wird die Dämpfung aufgrund des *release* langsam zurückgesetzt. Die *release*-Zeit beträgt hier 9,8 Sekunden. Wird  $f_{Koppel}$  nach dieser Totzeit, aber noch während der *release*-Zeit erneut detektiert, so wird der Maximalwert der Dämpfung bei jeder weiteren Detektion, die nicht während der Totzeit stattfindet, um 1dB erhöht. Verstreicht die *release*-Zeit, so fällt auch der Maximalwert der Dämpfung wieder auf 3 dB zurück. Die Dämpfung wird somit adaptiv von den Detektionen gesteuert.

Das Ergebnis einer solchen Detektion bzw. Filterung ist in Abbildung 6-1 unten sehr gut zu sehen.

## 6.5 Zusammenfassung

Das Gesamtsystem besteht also aus zwei Filterbänken mit jeweils sieben Filtern. Dabei unterdrückt die erste Filterbank jene Rückkopplfrequenzen die schnell auftreten und durch ihre Schleifenverstärkungen, die größer als 1 sind, auch in der Lautstärke ansteigen. Wegen der hohen Geschwindigkeit mit der solche Frequenzen auftreten, muss die zeitliche Glättung des Spektrums für die Detektion niedrig sein. Der Vergessensfaktor  $\alpha$  beträgt bei der Echtzeitimplementation in PD den Wert 0,1. Die Zeitkonstante  $\tau$  ergibt sich bei einer Blocklänge für die FFT von  $N=2048$  zu:

$$\tau = -\frac{1}{\ln(1-\alpha)} \cdot \frac{1}{fs^*} = -\frac{1}{\ln(0,9)} \cdot \frac{2048}{44100} \approx 0,44s$$

Die zweite Filterbank unterdrückt jene Frequenzen, deren Schleifenverstärkung knapp unter 1 liegt. Da diese Frequenzen nicht sehr laut im Vergleich zum Hintergrundrauschen und vor allem im Vergleich zu energiereicher Sprache sind, muss hier erstens eine längere zeitliche Mittelung des Spektrums vorgenommen, und zweitens die Aktualisierung des Spektrums mit der im Signal enthaltenen Energie gewichtet werden (siehe Kapitel 6.3). Der Wert der Zeitkonstante beträgt hier  $\tau=2,3$  Sekunden. Das Gesamtsystem mit zwei adaptiven Filterbänken ist in Abbildung 6-2 zu sehen.

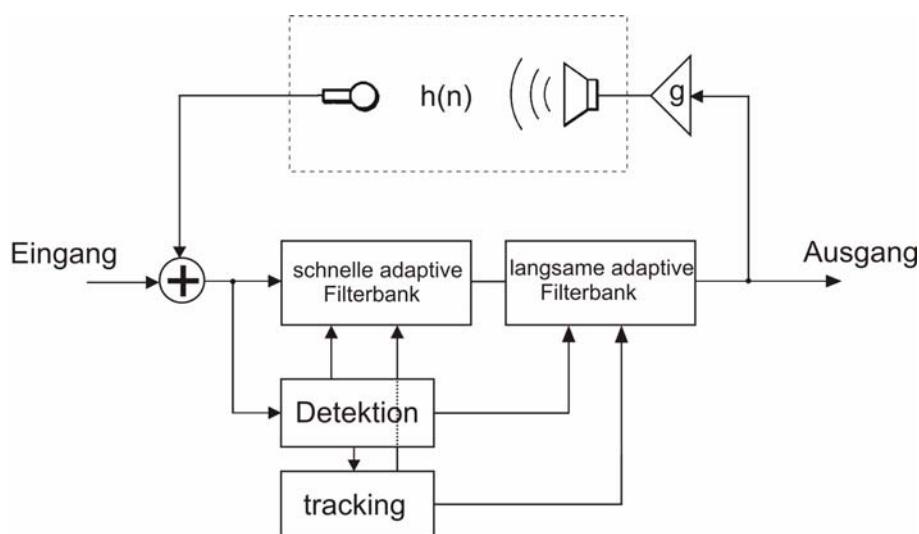


Abbildung 6-2 Gesamtsystem mit zwei adaptiven Filterbänken

## 7 Evaluierung des Gesamtsystems

Durch den Einsatz des vorgestellten Systems verbessert sich das Übertragungsverhalten kurz unterhalb, und besonders oberhalb der Koppelgrenze der Lautsprecher-Mikrofon Anordnung. Auf einige Punkte wird in diesem Kapitel im Detail eingegangen. Dazu wurden einige Messungen im Experimentalstudio des IEMs durchgeführt. Der Messaufbau ist Abbildung 7-1 zu entnehmen.

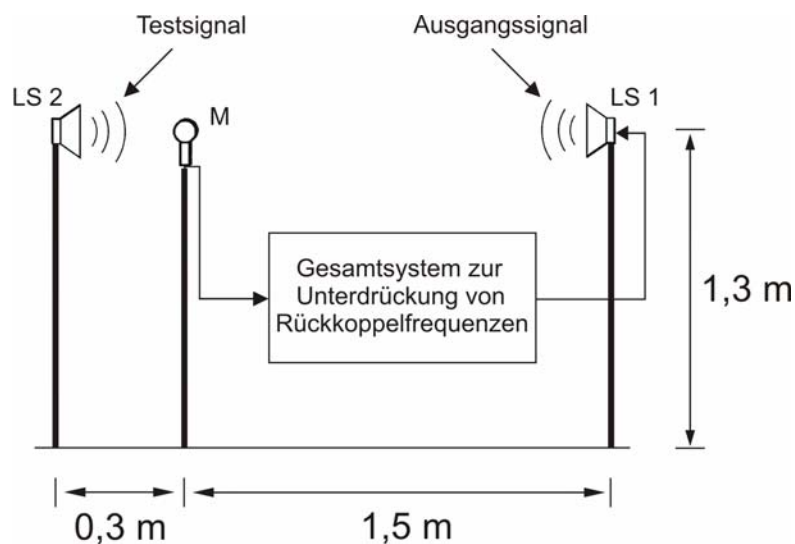
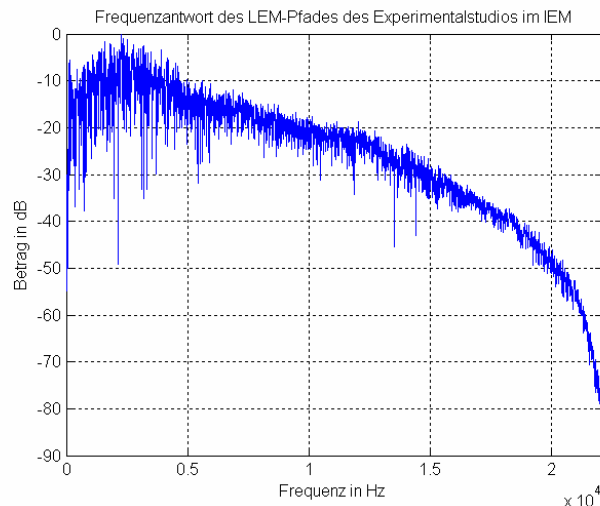


Abbildung 7-1 Messaufbau zur Evaluierung des Gesamtsystems

In Abbildung 7-2 ist die Frequenzantwort des LEM-Pfades (loudspeaker enclosure microphone) für obigen Messaufbau dargestellt. Die zugehörige Impulsantwort wurde wiederum nach [3] gemessen.



**Abbildung 3-2** Frequenzantwort des LEM-Pfades des Messaufbaus des Experimentalstudio im IEM

Bei Erhöhung der Ausgangsverstärkung ist zu erwarten, dass die ersten Koppelfrequenzen an den Stellen der größten Überhöhung der Frequenzantwort auftreten. Also wie in Abbildung 3-2 zu sehen etwa im Bereich von 2500Hz. Dies bestätigt sich auch bei den Messungen.

Mit einem Lautsprecher LS 2 werden nun Testsignale in das rückgekoppelte System eingespeist. Diese Signale werden von einem Mikrofon M aufgenommen und über einen Lautsprecher LS 1 abgespielt. Das Mikrofon M nimmt somit das Testsignal von LS 2 und den rückgekoppelten Teil des Testsignals von LS 1 auf. Über die regelbare Ausgangsverstärkung können nun verschiedene Szenarien mit unterschiedlichen Schleifenverstärkungen getestet werden. Die drei verschiedenen Testsignale sind zwei generisch erzeugte, genormte ITU-Signale (Norm: ITU-T Rec. P.501, AMFM und MSMP), und ein männlicher Sprecher. Die Testsignale werden in 30 cm Entfernung vom Mikrofon abgespielt und haben, ohne rückgekoppelten Anteil, am Mikrofon einen Schalldruckpegel von 74 dB(A). Der Abstand vom Mikrofon zum Lautsprecher LS 1 beträgt 1,5 m, die Lautsprecher und das Mikrofon befinden sich vom Boden aus in einer Höhe von 1,3 m.

## 7.1 Erhöhung der Stabilitätsgrenze

Eine wichtige Verbesserung ist die Erhöhung der Stabilitätsgrenze der Lautsprecher-Mikrofon Anordnung durch das Gesamtsystem. Zunächst wird die Ausgangsverstärkung des Systems ohne die Filterbank so lange erhöht, bis die Stabilitätsgrenze erreicht ist, das heißt, bis durch die zu hohe Schleifenverstärkung ein lautes Heulen auftritt. Dies geschieht zuerst ohne ein Testsignal von Lautsprecher LS 2. Danach wird der Versuch mit der zwischengeschalteten Filterbank wiederholt. Die Differenz der beiden Ausgangsverstärkungen beträgt für diese Versuchsanordnung 11 dB.

Als nächstes wird der Versuch mit den drei Testsignalen aus Lautsprecher LS 2 wiederholt. Hier ist das Ergebnis deutlich geringer. Für das MSMP-Signal und das Sprachsignal ergibt



sich eine Differenz von 6 dB, für das AMFM-Signal beträgt die Differenz 7 dB, allerdings entstehen bei diesem Signal sehr viele Fehldetektionen.

Der Frequenzschieber erreicht bei einer Verschiebungs-Frequenz von 4 Hz und dem Sprachsignal als Testsignal ebenfalls eine Erhöhung der Stabilitätsgrenze von 6 dB, allerdings bei deutlich höheren Qualitätseinbußen des Signals (siehe auch Kapitel 7.2). Der pitch shifter erhöht die Stabilitätsgrenze bei einer Verschiebungsfrequenz von 4 Hz und einem Frequenzhub von 30 Cent um 5 dB, hier sind die entstehenden Artefakte am unangenehmsten und die Signalqualität am niedrigsten.

Speist man das Sprachsignal in das Gesamtsystem bestehend aus der Filterbank und dem Frequenzschieber ein, so lässt sich die Stabilitätsgrenze um 10 dB nach oben drücken, allerdings unter enormen Einbußen der Sprach- und Klangqualität.

## 7.2 Bewertung der Sprach- und Klangqualität im Bereich der Koppelgrenze

Für die Bewertung der Sprach und Klangqualität, muss die Signalqualität auch unterhalb der Stabilitätsgrenze betrachtet werden. Auch bei dieser Bewertung dient der männliche Sprecher mit einem Schalldruckpegel von 74 dB(A) am Mikrophon als Testsignal. Etwa 4-5 dB unterhalb der Stabilitätsgrenze beginnen bei obigem Messaufbau die Qualitätsminderungen am Signal. Sie sind zunächst kaum hörbar und nahezu nicht störend. Nähert man sich auf etwa 2-3 dB an die Stabilitätsgrenze an, so erreicht die adaptive Filterbank eine deutliche Verbesserung der Signalqualität. Während ohne den Einsatz der Filter Koppelfrequenzen auftreten, die durch ihr langes Nachklingen die Sprach- und im speziellen die Klangqualität einschränken, entstehen bei Zwischenschaltung der Filterbank zwar hörbare, aber kaum bis gar nicht störende Artefakte. Nähert man sich der Stabilitätsgrenze noch weiter bis auf 1 dB, so werden die Vorzüge der Filterbank noch deutlicher. Das Signal wird nun ohne Filterung von sehr langen, sehr störenden, nachklingenden Koppelfrequenzen gestört. Schaltet man die adaptive Filterbank dazwischen, so entstehen zwar ebenfalls leicht störende Artefakte, allerdings ist die Signalqualität deutlich besser als ohne Filterung.

Erhöht man die Ausgangsverstärkung noch weiter, dann ist die Lautsprecher-Mikrofon Anordnung, wie in Kapitel 7.2 beschrieben, nicht mehr stabil. Eine Bewertung der Signalqualität kann nun nur noch mit zwischengeschalteter Filterbank erfolgen.

Direkt an und 1 dB oberhalb der Stabilitätsgrenze sind mit der Filterbank nur leichte Einbußen der Signalqualität bemerkbar. 2-3 dB oberhalb der Grenze sind die entstehenden Artefakte störend für den Hörer. Zwischen 4 und 6 dB oberhalb der Grenze entstehen trotz der Filterung immer wieder längere nachklingende Koppelfrequenzen, da schon mehr Rückkoppelfrequenzen detektiert werden als Filter zur Verfügung stehen. Die Filterbank unterdrückt in diesem Fall nur die prominentesten Koppelfrequenzen, was zu sehr störenden Verzerrungen und zu einer starken Qualitätsminderung des Signals führt. Etwa bei 7 dB oberhalb der Stabilitätsgrenze stehen nicht mehr genug Filter zur Verfügung um die

Schleifenverstärkung für alle Frequenzen kleiner als 1 zu halten. Das Signal schwingt bei einer Frequenz auf und es kommt zu einem lauten Heulen.

Der Frequenzschieber schiebt die Stabilitätsgrenze, wie oben bereits erwähnt, ebenfalls um 6 dB nach oben. Allerdings sind hier die auftretenden Artefakte störender als bei der adaptiven Filterbank. 3 dB oberhalb der Stabilitätsgrenze sind die Qualitätsminderungen des Signals sehr störend. Der pitch shifter erzeugt noch früher unangenehme Artefakte und schafft auch nur eine Erhöhung der Stabilitätsgrenze von 5 dB.

## 8 Ausblick

In diesem Kapitel werden kurz zwei Punkte genannt, die das bestehende System noch robuster gegenüber Koppelfrequenzen und somit noch effizienter gestalten könnten.

### 8.1 Langzeitheuristik

Geht man davon aus, dass die Umgebungsverhältnisse einigermaßen stationär sind, das heißt das die Übertragungsfunktion vom Mikrofon zum Lautsprecher und das Hintergrundrauschen einigermaßen konstant bleiben, dann würde die Einführung einer Langzeitheuristik eine Verbesserung für das Gesamtsystem bedeuten. In einer Konferenzsituation könnten diese Verhältnisse beispielsweise gegeben sein. Bei konstanten Umgebungsverhältnissen und einer zu hohen Schleifenverstärkung  $g$  werden immer wieder die selben Störfrequenzen auftreten. Versieht man das System mit einem Langzeitgedächtnis, in dem häufig auftretende Rückkopplfrequenzen gespeichert werden, so bestünde die Möglichkeit, ein Filter für eine längere Zeit mit einer konstanten Dämpfung zu setzen. Dies würde das ständige Auftreten dieser Frequenzen verhindern und eine deutliche Verbesserung der Signalqualität bedeuten.

### 8.2 Reduzierung der Ausgangsverstärkung

Eine weitere Möglichkeit zur Verbesserung des Systems ist die Reduzierung der Gesamtausgangsverstärkung. Ist das System durch ständiges Auftreten von sehr vielen Koppelfrequenzen völlig ausgelastet, das heißt es existieren sehr viele Frequenzen deren Schleifenverstärkung  $g$  über oder knapp unter 1 liegen, dann wird die Signalqualität, selbst wenn die Filterung eine völlige Zerstörung des Signals und ein lautes Heulen vermeidet, sehr unter diesen Umständen leiden. Es wäre zu überlegen, ob bei einer sehr hohen Schleifenverstärkung für sehr viele Frequenzen, eine Reduzierung der Gesamtausgangsverstärkung nicht erstens einfacher, und zweitens auch sinnvoller im Hinblick auf die Signalqualität wäre. Man würde auf diese Weise zwar keinen Lautheitsgewinn erzielen, allerdings würde die Qualität des Signals deutlich verbessert werden. Gerade an der Koppelgrenze führt eine Verringerung der Ausgangsverstärkung von 1 – 2 dB zu einen immensen Rückgang der Koppelfrequenzanzahl.

## 9 Literaturverzeichnis

- [1] Dattorro J.: *Effect Design Part 1: Reverberator and Other Filters*, J Audio Eng Soc, Vol. 45, No. 9, September 1997
- [2] Desainte-Catherine Myriam, Marchand Sylvain: *High Precision Fourier Analysis of Sounds using Signal Derivatives* May 1, 1998
- [3] Farina Angelo: *Simultaneous measurement of impulse response and distortion with a swept-sine technique*, AES 108<sup>th</sup> convention, Paris, February 2000
- [4] Hänslers E., Schmidt G.: *Acoustic echo and noise control – a practical approach*, John Wiley & Sons, Ltd West Sussex, England, 2004
- [5] Jianqiang Wei et al.: *A new algorithm for howling detection*, IEEE 2003
- [6] Kawamura, Akihisa et. al.: *Sound amplifying apparatus with automatic howl-suppressing function* United States Patent, Patent Number: 5,442,712 August 1995
- [7] Kolokolov A. S.: *Preprocessing and Segmentation of the Speech Signal in the Frequency Domain for Speech Recognition*, Russian Academy of Science, Moscow, Russia June 20, 2002
- [8] Laakso T.: *Design and Implementation of Efficient IIR Notch Filter with Quantization Error Feedback*, Transactions on instrumentation and measurement, Vol. 43, No. 3, June 1994
- [9] Leotwassana W.: *Adaptive howling canceller using adaptive IIR notch filter: Simulation and implementation*, IEEE Int. Conf. Neural Networks & Signal Processing, Nanjing, China December 2003
- [10] McAuley R.: *Speech Analysis/Synthesis Based on a Sinusoidal Representation*, IEEE Transactions on signal processing, Vol. Assp-34, No. 4, August 1986
- [11] Tietze U., Schenk Ch.: *Halbleiter-Schaltungstechnik*, 11. Auflage Springer-Verlag, 1999
- [12] Tsugio I.: *Apparatus detecting howling by decay profile of impulse response in sound system*, United States Patent, Patent Number: 6,442,280 August 2002
- [13] Widrow B. et al., *Adaptive Noise Canceling: Principles and Applications*, IEEE, December 1975

[14] Yong Ching Lim: *A piloted adaptive notch filter*, IEEE Transactions on signal processing, Vol. 53, No. 4, April 2005

[15] Zölzer Udo: *DAFX Digital Audio Effects*, John Wiley & Sons, Ltd West Sussex , England, 2002