

Implementierung einer Ausfallsverschleierung mittels spektraler Signalmodellierung

Projektarbeit

von

Sebastian Rieck

Betreuer:

DI Franz Zotter

Graz, Oktober 2009



Universität für Musik und darstellende
Kunst Graz



Institut für Elektronische Musik und
Akustik

Inhaltsverzeichnis

1	Einleitung	1
2	Spectral Modeling Synthesis (SMS) plus Noise	2
3	Der Analyse/Synthese Algorithmus	3
3.1	prinzipielle Struktur	3
3.2	Fensterung	5
3.3	Peakdetektion	5
3.3.1	Lokalisation der Peaks	5
3.3.2	Amplitudenkorrektur	7
3.3.3	Frequenzkorrektur	7
3.3.4	Phasenkorrektur	8
3.3.5	Schwellwert	9
3.4	Spurenbildung	11
3.4.1	Parameter einer Spur	11
3.4.2	Funktionsweise	11
3.4.3	Minimale Lebenszeit	13
3.5	Deterministischer Teil	15
3.5.1	Synthese der Partialtöne	15
3.5.2	Phaseninterpolation	16
3.6	Stochastischer Teil	18
3.6.1	Filterbankparameter	18
3.6.2	Filterbankimplementierung	18
3.6.3	Analyse und Synthese	20
4	Ausfallsverschleierung	22
5	Programmbedienung	25
5.1	Parameter	25
5.2	Ausgabe	27
6	Evaluierung	28
6.1	Analyse/Synthese	28
6.1.1	Qualität der Korrekturverfahren	28
6.1.2	zeitliche Fehler	31
6.1.3	Schwellwerte	33
6.1.4	Spurenbildung	33
6.1.5	Residuumfilterbank	33
6.2	Qualität der Ausfallsverschleierung	34
6.3	Berechnungsaufwand und Signalverzögerung	36
7	Mögliche Verbesserungen	37
8	Fazit	38
9	Literaturverzeichnis	39

Kurzfassung

Ziel dieses Projekts ist die Erstellung eines Modells auf Grundlage des Spectral Modeling Synthesis (SMS) plus Noise Verfahrens, mit dessen Hilfe die Fortführung eines zeitdiskreten Signals ermöglicht wird. Es wird davon ausgegangen, dass ein Signal bis zu einem bestimmten Zeitpunkt vorhanden ist und anschließend ein Ausfall stattfindet. Die Analyse des letzten gültigen Signalsegments erfolgt über die Berechnung der Kurzzeit-Fouriertransformation (KZFT), wobei das Signal im Zeitbereich zunächst gefenstert (Hanning) und anschließend mit der diskreten Fouriertransformation (DFT) transformiert wird. Als Signalparameter erhält man daraus das Amplituden- und Phasenspektrum, welche spektral modelliert werden, d.h. aus Sinustönen und Rauschen nachgebildet werden. Die Frequenzauflösung dieser Modellierung ist durch die Länge der DFT gegeben und kann durch geeignete Korrekturverfahren verbessert werden (zB. parabolic interpolation). Durch Beibehalten des Amplitudenspektrums und kontinuierlicher Fortsetzung der Phase wird das Signal im Falle eines Ausfalls fortgesetzt. Dies kann als Fortführung des Signals im Frequenzbereich verstanden werden. Die eigentliche Resynthese des Signals erfolgt jedoch im Zeitbereich.

1 Einleitung

Als weitverbreitetes Basissystem in der Verarbeitung von digitalen Audiosignalen wird eine Einheit, bestehend aus Analyse und Resynthese einer zeitdiskreten Signalfolge, verwendet. In der Analysestufe werden aus einem Signal Merkmale, d.h. Parameter, extrahiert. Diese werden, z.B. durch Modifikation der Zeit- und/oder Frequenzstruktur (siehe Phasenvokoder), entsprechend einer gewünschten Zielanwendung verändert. Komplementär zur Analyse erzeugt die Resynthese aus den modifizierten Parametern wiederum ein Zeitsignal.

2 Spectral Modeling Synthesis (SMS) plus Noise

Dieses additive Synthesemodell geht davon aus das sich ein Signal $x[n]$ in einen deterministischen Teil $d[n]$, bestehend aus M Sinustönen, und einen stochastischen Teil $s[n]$, bestehend aus Rauschen bzw. Geräuschanteilen, zerlegen lässt.

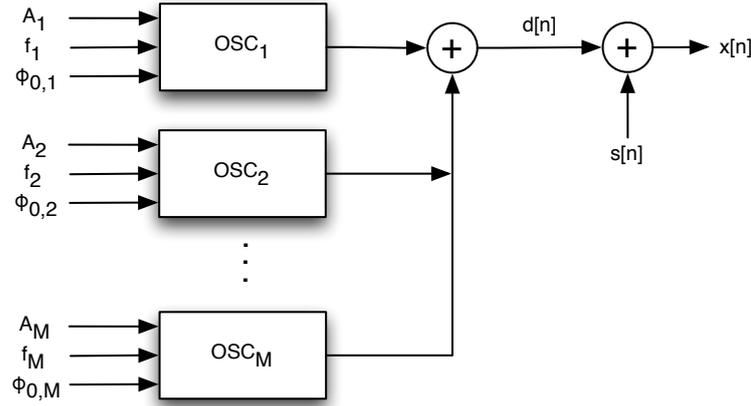


Abbildung 1: SMS Synthese Modell

Der deterministische Teil lässt sich im zeitdiskreten Fall beschreiben durch

$$d[n] = \sum_{i=1}^M A_i[n] \cos(\phi_i[n]) \quad (1)$$

wobei $A_i[n]$, $\phi_i[n]$ die Momentanwerte der Amplitude und Phase des i -ten Sinustons darstellen. Die Momentanphase ist gegeben durch

$$\phi_i[n] = \phi_{0,i} + \sum_{n'=0}^{n-1} \frac{f_i[n']}{f_s} 2\pi \quad , \quad n > 0 \quad (2)$$

mit f_s der Abtastfrequenz und f_i der Momentanfrequenz. Der stochastische Teil, auch bezeichnet als Residuum, ergibt sich aus der Differenz

$$s[n] = x[n] - d[n] \quad (3)$$

und stellt somit jenen Teil des Signals dar, der sich nicht durch schmalbandige Komponenten beschreiben lässt. Das SMS Modell geht des Weiteren davon aus das sich die Amplitude und Frequenz jedes Sinustones zeitlich gesehen nur langsam ändert. Diese Stationarität stellt damit die Voraussetzung dar das Signal blockweise zu verarbeiten. Typische Blöcklängen sind 1024 ($\hat{=}$ 23ms@44,1 kHz). Eine genauere Beschreibung lässt sich in [Ser89] bzw. [Ser97] finden.

3 Der Analyse/Synthese Algorithmus

3.1 prinzipielle Struktur

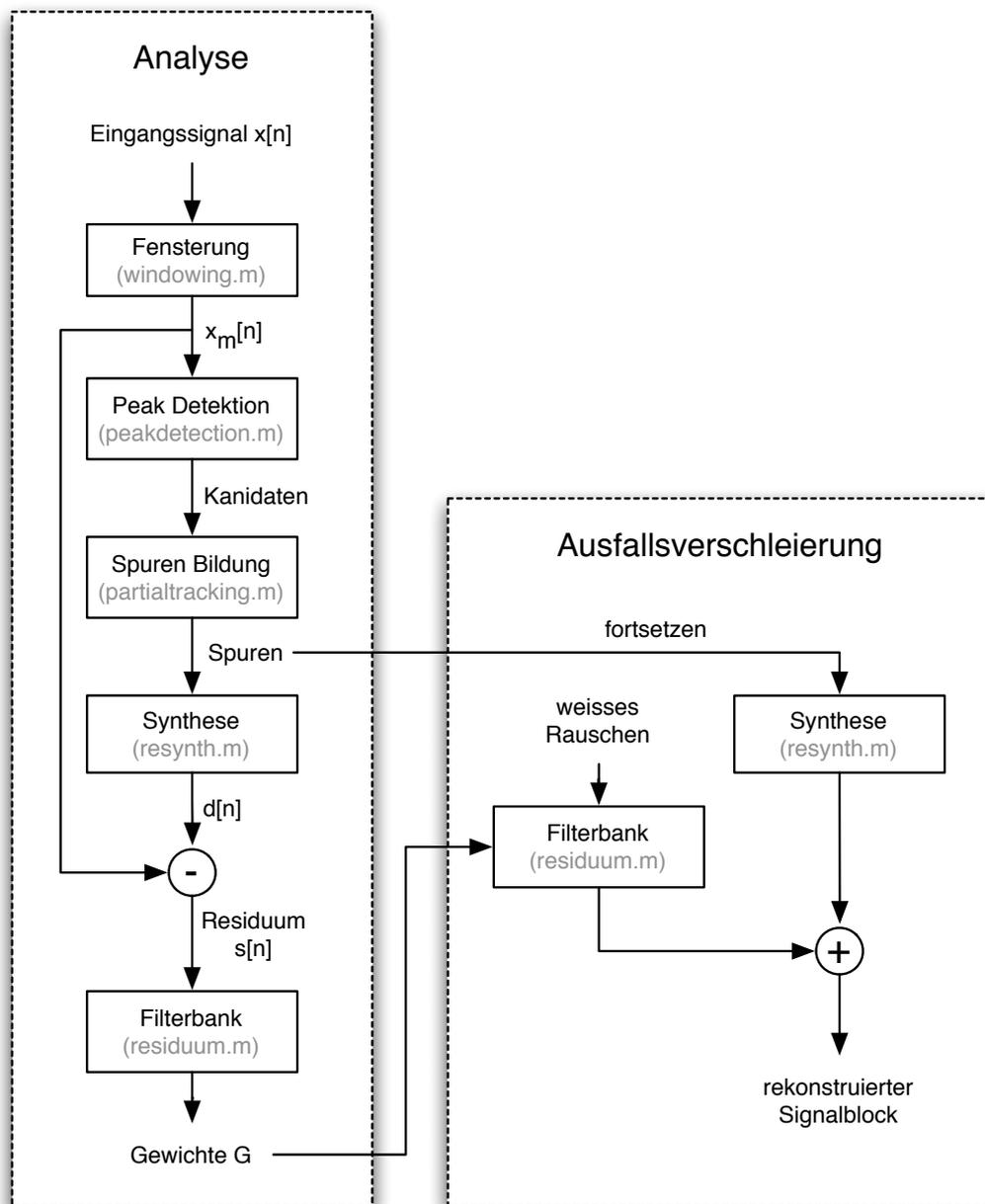
Das SMS + Noise Modell (Abbildung 1) sowie die 2002 verfasste Diplomarbeit von Cornelia Falch "*MATLAB®-Tool für die Parameterberechnung zur Additiven Klangsynthese*" [Fal02] stellt die Grundlage des hier verwendeten Systems dar. Die prinzipiellen Schritte zur Bestimmung der Parameter (A, f, ϕ_0) der Partialtöne sind:

1. Fensterung des einkanaligen Eingangssignals $x[n]$
2. Peakdetektion
3. Amplituden-, Frequenz- und Phasenkorrektur der gefundenen Peaks
4. Zusammenfassen der Peaks zu Spuren

Jede Spur stellt dann einen Partialton dar. In weiterer Folge wird das Residuum bestimmt und modelliert:

5. Synthese der Partialtöne (Spuren) $d[n]$
6. Bestimmung des Residuums $s[n] = x[n] - d[n]$
7. Analyse des Residuums:
Filterung durch Filterbank, Bestimmung der Gewichte $g \hat{=}$ Leistung in jedem Band
8. Synthese des Residuums:
weisses Rauschen wird gefiltert durch dieselbe Filterbank und skaliert mit den Gewichten g

Dieser Ablauf stellt die grundsätzliche Funktionsweise des Analyse/Synthese Systems dar. Bezogen auf die Problematik der Ausfallsverschleierung, werden die Parameter der Partialtöne (A, f, ϕ_0) und die berechneten Gewichte g der Residuumfilterbank benutzt, um im Falle eines Ausfalls die resultierende parametrische Darstellung des Signals fortzusetzen und damit den Ausfall zu füllen. Abbildung 2 zeigt diesen schematischen Ablauf. Eine genauere Beschreibung der einzelnen Stufen, bzw. deren Implementierung, folgt in den kommenden Abschnitten.

Abbildung 2: Prinzipielle Struktur des Algorithmus (*dropcon.m*)

3.2 Fensterung

Das Eingangssignal $x[n]$ wird überlappend (Hopsize hs) ausgeschnitten und mit einem Hann-Fenster $w[n]$ (siehe Abbildung 4) gewichtet. Das resultierende Signal wird im Weiteren als Frame bezeichnet.

$$x'_m[n] = x[n + (m - 1) \cdot hs] \quad (4)$$

$$x_m[n] = x'_m[n] \cdot w[n] \quad (5)$$

mit $0 \leq n \leq \text{Fensterlänge} - 1$

wobei $x_m[n]$ das gefensterte Signal des m -ten Frames darstellt (siehe Abbildung 3). Falls die im Signal vorhandenen Abtastwerte nicht für ein vollständiges Fenster ausreichen, wird das Fenster mit Nullen aufgefüllt.

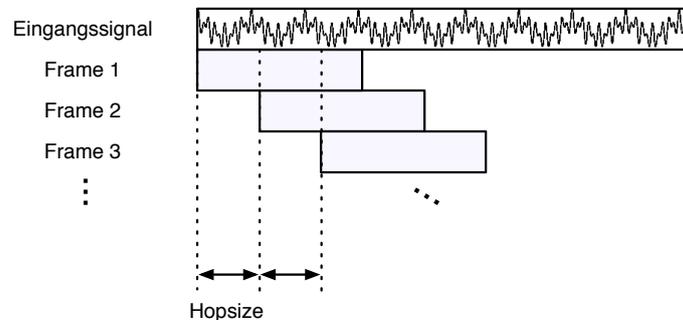


Abbildung 3: Fensterung des Eingangssignals

3.3 Peakdetektion

3.3.1 Lokalisation der Peaks

Der aus der Fensterung resultierende Signalblock $x_m[n]$ der Länge N wird mit einer DFT zerlegt und das Betragsspektrum $|X[k]|$ berechnet.

$$|X_m[k]| = 20 \log_{10} \left(\frac{1}{\sum w[n]} \left| \sum_{n=0}^{N-1} x_m[n] e^{-2\pi i k n / N} \right| \right) \quad (6)$$

Zur Detektion spektraler Spitzen (Peaks) wird die zeitdiskrete Ableitung $|\tilde{X}_m[k]|$ gebildet

$$|\tilde{X}_m[k]| = \sum_{k=1}^{\frac{N}{2}-2} |X_m[k+1]| - |X_m[k]| \quad (7)$$

und der Wechsel von positiven auf negative Werte lokalisiert. Dies entspricht dem Auffinden der lokalen Maxima im Betragsspektrum.

Bei Erfüllen der Bedingung

$$|\tilde{X}_m[k]| > 0 \quad \text{und} \quad |\tilde{X}_m[k+1]| < 0 \quad (8)$$

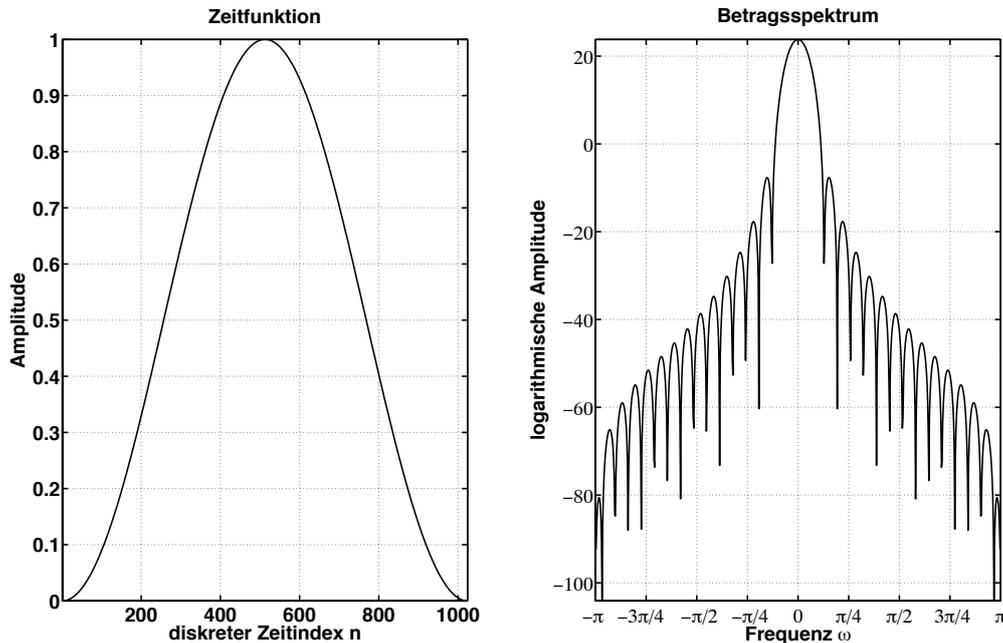


Abbildung 4: Hann Fenster, (links) Funktion im Zeitbereich, (rechts) Betragsspektrum

markiert die Stelle k einen Extremwert und damit einen gefundenen Peak. Die dazugehörige Amplitude $|X_m[k]|$ wird im nächsten Schritt korrigiert (siehe 3.3.2), um feststellen zu können ob sie über einem definierten Schwellwert liegt, ab welchem das Verfahren die spektrale Spitze als Teiltonkanidat erkennt. Durch diese Korrektur wird garantiert dass der Schwellwert die tatsächlichen Amplitudenwerte der Peaks mit einbezieht.

Grundsätzlich stellt sich die Frage, wie groß der minimale Abstand zweier Peaks in ihrer Frequenz sein darf, um gerade noch erkannt zu werden. Damit zwei eng beieinander liegende Frequenzen f_1, f_2 gerade noch erkannt werden, müssen sie im Betragsspektrum als separate Peaks erkennbar sein. Dies ist nur der Fall, wenn die korrespondierenden Bins k_1 und k_2 mindestens einen Bin Abstand besitzen. Das heißt also:

$$\Delta k_{min} = k_2 - k_1 = 2 \quad (9)$$

Daraus folgt für den minimalen Frequenzabstand Δf_{min} :

$$\Delta f_{min} = f_2 - f_1 = 2 \frac{f_s}{N} \quad (10)$$

mit N der Fensterlänge und der Samplingfrequenz f_s . Beispielhaft ergibt sich für die hier verwendete Fensterlänge von $N = 1024$ und einer angenommenen Abtastfrequenz $f_s = 22050 \text{ Hz}$: $\Delta f_{min} = 43,066 \text{ Hz}$

Zu beachten bleibt, dass das gebildete Spektrum $|X_m[k]|$ nur Stützstellen des kontinuierlichen Spektrums $|X(\omega)|$ an den Stellen $k \cdot f_s/N$ repräsentiert. Es kann daher weder garantiert werden, dass alle vorkommenden Frequenzen nur auf diesen Stützstellen liegen, noch dass die Länge des Fensters immer einem ganzzahligen Vielfachen der Signalperiode

entspricht. Um das tatsächliche Spektrum besser abzubilden, muss zumindest eine Korrektur der gemessenen Werte für Amplitude, Frequenz und Phase vorgenommen werden. Die verwendeten Verfahren werden im Folgenden näher erläutert.

3.3.2 Amplitudenkorrektur

Es wurde das Verfahren der parabolischen Interpolation ausgewählt. Es nutzt die Eigenschaft, dass ein lokales Maximum im Betragsspektrum, zusammen mit den links und rechts benachbarten Bins, annähernd eine Parabel beschreibt. Diese Eigenschaft ist eine direkte Folge der Fensterung eines Signals und der darauf folgenden FFT. Wird ein Signal im Zeitbereich mit einer Fensterfunktion multipliziert, so entspricht dies im Frequenzbereich der Faltung der beiden Transformierten. Da die Transformierte der Fensterfunktion in ihrem Mittelpunkt eine parabelähnliche Form besitzt (siehe Abbildung 4), bildet sich diese bei Faltung mit dem Signalspektrum eines Sinustones verschoben ab.

Zusammen mit den benachbarten Amplituden eines lokalen Maximums, berechnet die parabolische Interpolation das neue Maximum des korrigierten Amplitudenwerts (siehe Abbildung 5 bzw. [Ser89]).

$$\text{Mit } \begin{array}{lll} x_1 = f_1 = (k-1) & f_s/N & \\ x_2 = f_2 = k & f_s/N & \\ x_3 = f_3 = (k+1) & f_s/N & \end{array} \quad \text{und} \quad \begin{array}{l} y_1 = 20 \log_{10} |X[k-1]| \\ y_2 = 20 \log_{10} |X[k]| \\ y_3 = 20 \log_{10} |X[k+1]| \end{array}$$

sowie den Parametern α und γ

$$\alpha = \frac{y_2 - y_1}{(x_2 - \gamma)^2 - (x_1 - \gamma)^2} \quad (11)$$

$$\gamma = \frac{1}{2} \cdot \frac{(y_3 - y_1)(x_2 - x_1)(x_2 + x_1) - (y_2 - y_1)(x_3 - x_1)(x_3 + x_1)}{(y_3 - y_1)(x_2 - x_1) - (y_2 - y_1)(x_3 - x_1)} \quad (12)$$

aus der allgemeinen Funktionsgleichung einer Parabel

$$y(x) = \alpha(x - \gamma)^2 + \beta \quad (13)$$

kann das neue Maximum $X'_{mag}[k]$ berechnet werden:

$$X'_{mag}[k] = y_2 - \alpha(x_2 - \gamma)^2; \quad (14)$$

3.3.3 Frequenzkorrektur

Die Frequenzkorrektur wurde implementiert nach dem von Cathrine Desainte in "*High Precision Fourier Analysis of Sounds using Signal Derivatives*" [DCM00] vorgeschlagenen Verfahren. Die dahinter liegende Idee ist, dass die zeitliche Ableitung einer komplexen Exponentialschwingung wieder jene Schwingung ergibt, welche nun jedoch mit $i\omega$ skaliert ist. Die Division der Beträge ergibt folglich die Frequenz der Exponentialschwingung.

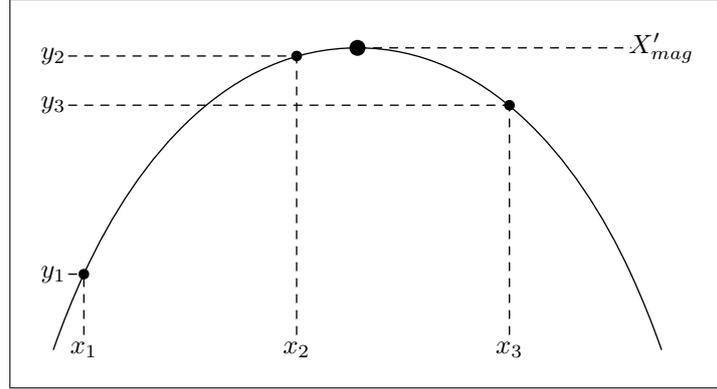


Abbildung 5: Parametrierung der Parabel zur Amplitudenkorrektur

Im allgemeinen, kontinuierlichen Fall lässt sich die korrigierte Frequenz f'_p aus der gemessenen Frequenz f_p des Partialtons folgendermaßen berechnen:

$$f'_p = \frac{1}{2\pi} \cdot \frac{|FT^{i+1}(f_p)|}{|FT^i(f_p)|}$$

bzw. im hier verwendeten diskreten Fall:

$$f'_p = \frac{1}{2\pi} \cdot \frac{|DFT^1[m_p]|}{|DFT^0[m_p]|} \quad (15)$$

mit $m_p = N \cdot f_p / f_s$ und N der Fensterlänge. Weiters bezeichnet DFT^1 die diskrete Fourier Transformation DFT der ersten Ableitung des Zeitsignals.

$$DFT^1 = DFT \left\{ \left(\sum_{k=0}^{N-1} x[k+1] - x[k] \right) w[k] \right\} \quad (16)$$

Eine wichtige Eigenschaft der Gleichung (15) ist, dass sich der Einfluss des Fensters $w[n]$ auf die Spektren durch die Division kompensiert, sofern beide Male das selbe verwendet wurde. Dies wurde in der Implementierung berücksichtigt.

3.3.4 Phasenkorrektur

Durch das Analysefenster entsteht eine Phasenmodulation in Abhängigkeit des Abstandes der tatsächlichen Frequenz f'_p zur Frequenz f_{bin} des nächstgelegenen Bins.

$$DFT'[x_m] = DFT[x_m] \cdot e^{-\frac{2\pi i(f'_p - f_{bin})}{f_s} \frac{(N-1)}{2}} \quad (17)$$

Der Term $e^{-j2\pi(N-1)/2}$ in Gleichung (17) entspricht einem zeitlichen Versatz des Fenster um eine halbe Länge und stellt dadurch den Bezug des gemessenen Phasenwertes zum Anfang des Fensters dar. Von der korrigierten Transformation DFT' wird abschließend das Phasenspektrum gebildet und ausgelesen. Die korrigierte Phase ϕ'_0 ist dann:

$$\phi'_0(f'_p) = \arg(DFT'[f_{bin} \cdot N / f_s]) \quad (18)$$

Zur Gegenprobe: Sind die Frequenzen f'_p und f_{bin} identisch, d.h. die Frequenz des Partialtons ein ganzzahliges Vielfaches der Frequenzauflösung der DFT ist, fällt der Exponentialterm in Gleichung (17) weg und die Phase des Partialtons kann ohne Verfälschung direkt ausgelesen werden.

3.3.5 Schwellwert

Der Schwellwert bezeichnet jene Amplitude ab der ein Peak im Betragsspektrum als Partialtonkanidat betrachtet wird. Es wurden hier unterschiedliche Varianten implementiert:

1. **fixer globaler Schwellwert GS [dB]**
Ist Signal- und Frameunabhängig
2. **fixer lokaler Schwellwert LS [dB]**
Ist definiert als das Verhältnis zum lautesten Peak eines Frames
3. **dynamischer Schwellwert DS**
Der Schwellwert wird für jedes Frame neu berechnet

Der lokale Schwellwert stellt eine deutliche Verbesserung zum starren Konzept des fixen globalen Schwellwerts dar, nachdem Letzterer meist nur ungenügend geschätzt werden kann. Des Weiteren führt die Verwendung des lokalen Schwellwerts dazu, dass auch in leisen Passagen Partialtöne erkannt werden, welche sonst durch den höheren GS nicht berücksichtigt werden. Eine weitere Möglichkeit stellt der dynamische Schwellwert dar. Es sei hier erwähnt, dass es sich bei dem dynamischen Schwellwert nicht um einen konstanten Wert, sondern um ein Verfahren handelt, den Schwellwert für das aktuelle Frame zu berechnen. Um jenen Schwellwert zu bestimmen, werden folgende Schritte durchgeführt:

1. Vom aktuellen gefensterten Signalblock wird das logarithmische Betragsspektrum gebildet
2. dieses wird wiederrum gefenstert, also in Frequenzblöcken weiterverarbeitet
3. für jeden Frequenzblock wird der Quantilwert bestimmt
4. alle Quantile Werte werden arithmetisch gemittelt
5. dies stellt dann jenen Schwellwert dar, welcher in der Peakdetektion für den aktuellen Frame verwendet wird

Quantile haben ihre ursprüngliche Verwendung in der Statistik. Sie sind Punkte einer nach Größe der Einzelwerten sortierten statistischen Verteilung, wobei die Ordnung p jenen Wert beschreibt, unterhalb dessen ein vorgegebener Anteil p aller Fälle der Verteilung liegt. Im hier vorliegenden Fall wurden die Parameter dieses Verfahrens experimentell¹ ermittelt und folgendermaßen festgelegt:

Fensterlänge	=	200 Abtastzeitpunkte
Hopsize	=	100 Abtastzeitpunkte
p	=	$0.8 \hat{=} 80\%$

¹Näheres in Abschnitt 6.1.3

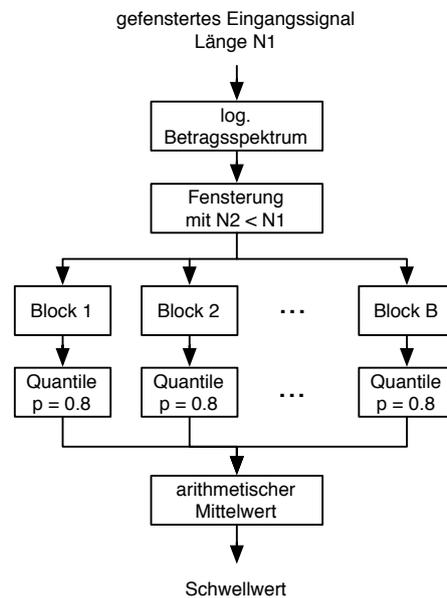


Abbildung 6: dynamischer Schwellwert, Berechnung des aktuellen Schwellwerts

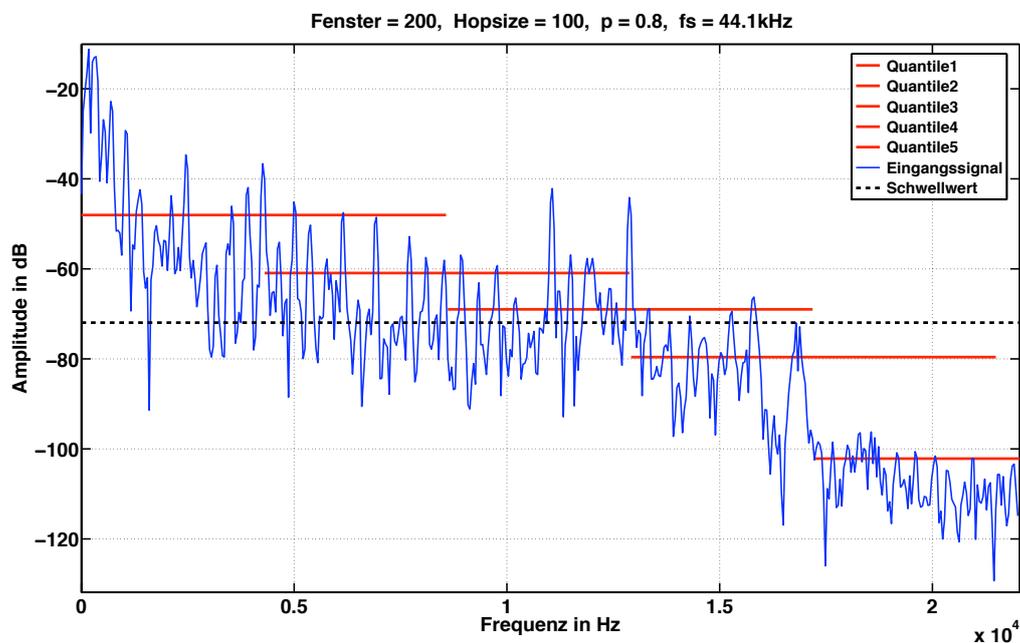


Abbildung 7: dynamischer Schwellwert, am Beispiel des Betragsspektrums einer zeitlich gefensterten akustischen Gitarre, punktierte Linie repräsentiert den resultierenden Schwellwert

Die Amplitudenwerte in jedem Frequenzblock werden demnach der Größe nach sortiert und jener Wert bestimmt der 80% des resultierenden Wertebereichs beinhaltet.

Abbildung 6 zeigt den Ablauf zur Bestimmung des aktuellen Schwellwerts, Abbildung 7 den resultierenden Schwellwert am Beispiel einer zeitlich gefensterten akustischen Gitarre. Letztere zeigt, dass Peaks, deren Frequenz größer als ca. 13 kHz ist, vernachlässigt werden. In der Praxis hat sich gezeigt, dass die dadurch vernachlässigten Partialtöne die empfundene Qualität der gesamten Ausfallsverschleierung in nur sehr geringem Maße beeinflusst. Es wurde deshalb auf einen frequenzabhängigen Schwellwert verzichtet. Es sei außerdem angemerkt dass der Schwellwert grundsätzlich auf -96dB, was der Dynamik eines mit 16bit quantisierten Signals entspricht, beschränkt wird.

Nachdem nun alle Partialtonkanidaten bestimmt und deren Parametern für Amplitude A , Frequenz f und Phase ϕ_0 korrigiert wurden, werden Sie abschließend in einer Matrix gespeichert und an die Spurenbildung weitergegeben.

3.4 Spurenbildung

Mit den Parametern der Partialton-Kanidaten (A, f, ϕ_0) werden jetzt Partialtonspuren gebildet um benachbarte Analysezeitpunkte verbinden zu können. Dies geschieht, um einen zeitlich kontinuierlichen Verlauf der analysierten Parameter zu erzeugen.

3.4.1 Parameter einer Spur

Grundsätzlich werden alle Spuren mit einem numerischen Namen versehen und besitzen außer den Werten für Amplitude A , Frequenz f und Phasenverschiebung ϕ_0 , ein Synthese-Flag (Tabelle 1), sowie einen MLT(Minimum LifeTime)-Zähler welcher angibt wieviele Frames die Spur bereits vorhanden ist.

Synthese-Flag	Bedeutung
1 (<i>start</i>)	neue Spur beginnt
2 (<i>continue</i>)	Spur wird fortgesetzt
3 (<i>stop</i>)	Spur wird beendet
4 (<i>hold</i>)	Spur wird gehalten und nicht resynthesiert

Tabelle 1: Syntheseflag und dessen Bedeutung

Das Flag *hold* und der sog. MLT-Zähler kommen aus der Überlegung der *Minimalen Lebenszeit* welche in Abschnitt 3.4.3 erläutert wird.

3.4.2 Funktionsweise

Wir bezeichnen mit S_m^i die i-te Spur im aktuellen Frame m und entsprechend S_{m-1}^i die Spur des vergangenen Frames. Selbiges gilt für die aktuell detektierten Peaks P_m^i .

Prinzipiell lassen sich folgende Fälle unterscheiden:

a) keine Peaks P_m detektiert, keine Spuren S_{m-1} vorhanden

Wurden im aktuellen Frame m keine Peaks detektiert und im vergangenen Frame keine Spuren gebildet, wird die Spurenbildung übersprungen.

b) keine Peaks P_m detektiert, aber Spuren S_{m-1} vorhanden

Wurden im aktuellen Frame m keine Peaks detektiert, werden alle vorhandenen Spuren S_{m-1} beendet.

c) Peaks P_m detektiert, aber keine Spuren S_{m-1} vorhanden

Die Peaks P_m^i , welche nach ihrer Frequenz aufsteigend sortiert vorliegen, werden der Reihe nach abgearbeitet. Am Beispiel einer neuen Spur wird der Ablauf erklärt.

Es wird im ersten Schritt eine neue Spur S_m^1 erzeugt, welche die Parameter $(A_1, f_1, \phi_{0,1})$ von P_m^1 übertragen bekommt. Danach wird aus der Frequenz f_1 ein Bereich Δf_1 definiert in welchem auftretende Peaks der gleichen Spur zugeordnet werden.

$$\Delta f_1 = \Delta_f \cdot f_1 \quad \text{mit} \quad 0.01 \leq \Delta_f \leq 0.1$$

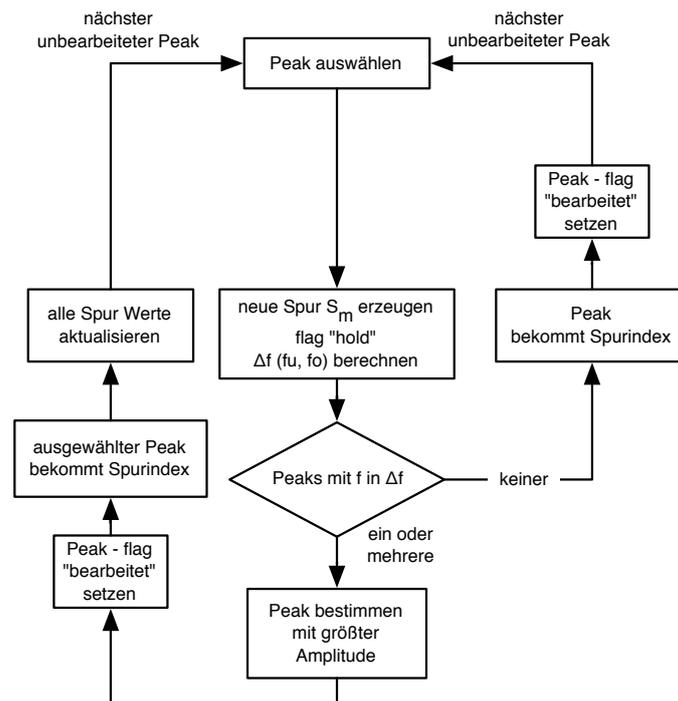


Abbildung 8: Fall b) Bildung neuer Spuren

Der Wertebereich von Δ_f ist prinzipiell frei wählbar, wird hier aber auf sinnvolle Weise beschränkt¹. Es hat sich in der Praxis $\Delta_f = 0.01$ ($\hat{=}$ 17 Cent) als bester Wert herausgestellt.

Treten Peaks mit Frequenzen im Bereich von $f_1 \pm \Delta f_1$ auf, wird in weiterer Folge der Lauteste bestimmt und die Spur S_m^1 bekommt die analysierten Parameter jenes Peaks übertragen. Alle betrachteten Peaks werden für die weitere Bildung neuer Spuren nicht

¹Siehe Abschnitt 6.1.4 für die Wahl von Δ_f und ihre Auswirkungen

mehr berücksichtigt. Abbildung 8 zeigt den hier erläuterten Ablauf.

d) Peaks P_m detektiert und Spuren S_{m-1} vorhanden

Dies ist der allgemeine Fall. Unter der Voraussetzung das im aktuellen Analysezeitfenster überhaupt ein Signal vorhanden ist und mindestens ein Frame bereits analysiert wurde, sind Peaks detektiert und auch Spuren bereits gebildet worden.

Es wird hier, analog zu c), für jede Spur mit der Frequenz f_{sp} ein Bereich Δf_{sp} berechnet:

$$\Delta f_{sp} = \Delta_f \cdot f_{sp} \quad \text{mit} \quad 0.01 \leq \Delta_f \leq 0.1$$

Im nächsten Schritt werden alle berechneten Bereiche $f_{sp} \pm \Delta f_{sp}$ nacheinander betrachtet und geprüft ob sie sich mit dem Bereich der nächsten Spur überschneiden. Ist dies der Fall so wird jene Spur mit der geringeren Amplitude beendet und für den weiteren Ablauf nicht mehr berücksichtigt. In den Frequenzbereichen der verbliebenen Spuren wird nun nach detektierten Peaks mit ihrer Frequenz f_{peak} gesucht.

$$f_{sp} - \Delta f_{sp} \leq f_{peak} \leq f_{sp} + \Delta f_{sp} \quad (19)$$

Es ergeben sich 4 zu unterscheidende Fälle:

- 1) kein f_{peak} liegt im erlaubten Intervall Δf_{sp}
- 2) ein f_{peak} liegt in einem Δf_{sp}
- 3) mehrere f_{peak} liegen in einem Δf_{sp}
- 4) ein f_{peak} liegt in mehreren Δf_{sp}

Im Fall 1) wird die Spur in dessen Δf_{sp} kein Peak gefunden wurde beendet. Im Fall 2) bekommt die Spur die Parameter des lokalisierten Peaks übertragen und wird fortgesetzt. Im 3ten Fall wird der lauteste Peak der Spur zugeordnet, die wiederum die Parameter des ausgewählten Peaks bekommt und fortgesetzt wird. Im Fall 4) wird der Peak jener Spur zugeordnet, dessen Frequenz f_{sp} die geringste Abweichung zur Frequenz f_{peak} des Peaks aufweist.

3.4.3 Minimale Lebenszeit

Das Konzept der minimalen Lebenszeit gibt an wie lange, zeitlich gesehen, ein Partialtonkanidat auftreten muss, bevor er resynthetisiert wird. Diese Länge wurde hier in Vielfachen eines Frames realisiert. Aufgrund des hier erläuterten Verfahrens werden grundsätzlich alle Spuren die neu erzeugt werden mit dem Flag *hold* versehen, welches angibt das die Spur nicht resynthetisiert wird. Am Ende der Spurenbildung wird bei allen die eben jenes Flag besitzen überprüft ob sie das Kriterium der Lebenszeit erfüllen. Gegebenenfalls wird das Flag auf *start* umgetragen und die Spur damit in der Resynthese der Partialtöne berücksichtigt.

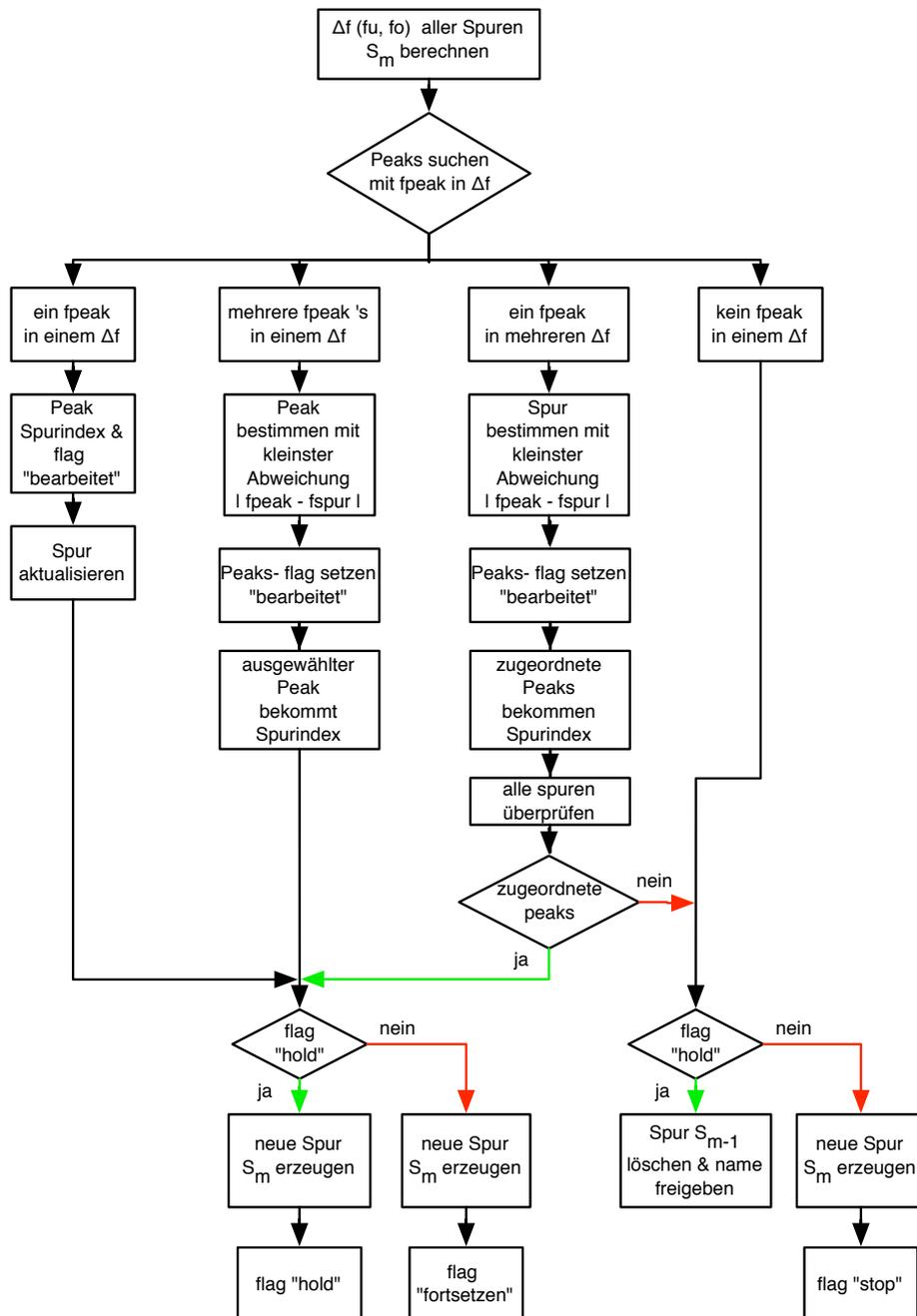


Abbildung 9: Zuordnung der Peaks zu Spuren

3.5 Deterministischer Teil

Die Analysedaten wurden durch einen Signalauschnitt (Frame) der Länge N gewonnen. Die darauffolgenden Daten werden nach einem zeitlichen Versatz des Fensters um eine Hopsizelänge hs berechnet (Abbildung 3, S.5). Folglich muss die Synthese nur für eine Hopsizelänge durchgeführt werden. Es wird hier nicht das oft verwendete Verfahren „Overlap&Add“ angewendet, sondern der erzeugte Verlauf zeitdiskreter Samples direkt aneinander gefügt. Damit dies möglich ist, muss garantiert werden, dass sich der Phasenverlauf des entstehenden Signals kontinuierlich fortsetzt. Zusammen mit der Tatsache, dass sich eine Spur in Amplitude, Frequenz und Phase von Frame zu Frame ändern darf, erfordert es eine geeignete Interpolation aller Parameter. Während die Amplitudenwerte linear interpoliert werden können, muss für die Interpolation der Phasen- und Frequenzwerte eine gemeinsame Funktion entwickelt werden, welche in Abschnitt 3.5.2 beschrieben wird (siehe hierzu [MQ86]).

3.5.1 Synthese der Partialtöne

Zur Bildung des deterministischen Teils, betrachten wir hier den aktuellen Frame m mit seinen Spuren S_m . Es bleibt zu beachten, dass sich alle aktuellen Analysedaten immer auf den Beginn des Frames m beziehen. Allgemein lässt sich die Synthese der Partialtöne durch Gleichung (1) und (2) auf Seite 2 beschreiben.

Abhängig vom Synthese-Flag einer Spur (siehe Tabelle 1, Seite 11) wird die Synthese folgendermaßen durchgeführt:

start:

Da sich die Spurparameter immer auf den Anfang des Frames beziehen muss eine neu beginnende Spur mit ihrer momentanen Amplitude A_m bereits in Frame $m - 1$ eingeblendet werden. Dies geschieht durch Nullsetzen von A_{m-1} und einer linearen Interpolation mit ihrem Endwert bei A_m . Des Weiteren wurde eine Momentanphase ϕ_m bestimmt, mit Hilfe dessen die Phase ϕ_{m-1} berechnet werden kann:

$$\phi_{m-1} = \phi_m - 2\pi \frac{f_m}{f_s} hs$$

mit der aktuellen Spurfrequenz f_m und Hopsizelänge hs .

continue:

Da sich eine Spur in Ihren Parametern von Frame zu Frame ändern darf werden die Werte für Amplitude, Frequenz und Phase zwischen den Frames $m - 1$ und m interpoliert. Die Amplituden A_{m-1} und A_m werden linear interpoliert, während Frequenz und Phase eine gemeinsame kubische Interpolation erfahren (siehe 3.5.2). Das Ergebnis aller Interpolationen stellt das synthetisierte Signal des Frames $m - 1$ dar.

stop:

Ein Partialton der im Frame m nicht mehr existiert und deshalb das Flag *stop* bekommen hat, wird in Frame $m - 1$ ausgeblendet.

Aufgrund der Eigenschaft das beginnende Spuren in $m - 1$ eingebledet, endende Spuren in $m - 1$ ausgbledet und Interpolationen im Falle von *continue* immer von $m - 1$ auf m erfolgen, ist der Signalblock der synthetisiert wird nur der von Frame $m - 1$.

3.5.2 Phaseninterpolation

Um zwischen den Frequenzen $\omega_0 = 2\pi f_{m-1}$ und $\omega_T = 2\pi f_m$, sowie den Momentanphasen $\phi_0 = \phi_{m-1}$ und $\phi_T = \phi_m$ zu interpolieren wird eine kubische Funktion der Form

$$\phi(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 \quad (20)$$

verwendet (siehe [MQ86]). Die zeitliche Ableitung der Phase entspricht der Momentanfrequenz $\omega(t)$ und ist folglich:

$$\omega(t) = \frac{d\phi(t)}{dt} = a_1 + 2a_2 t + 3a_3 t^2 \quad (21)$$

Durch Einsetzten der Randbedingungen

$$\begin{aligned} \phi(0) = \phi_0 = a_0 & \quad \text{und} \quad \phi(T) = \phi_T + 2\pi M \\ \omega(0) = \omega_0 = a_1 & \quad \text{und} \quad \omega(T) = \omega_T \end{aligned} \quad (22)$$

in (20) bzw. (21) erhält man

$$\omega(T) = \omega_0 + 2a_2 T + 3a_3 T^2 = \omega_T \quad (23)$$

$$\Rightarrow a_2 + \frac{3}{2} a_3 T = \frac{\omega_T - \omega_0}{2T} \quad (24)$$

$$\phi(T) = \phi_0 + \omega_0 T + a_2 T^2 + a_3 T^3 = \phi_T + 2\pi M \quad (25)$$

$$\Rightarrow a_2 + a_3 T = \frac{\phi_T + 2\pi M - \phi_0 - \omega_0 T}{T^2} \quad (26)$$

Zur vollständigen Beschreibung von (20) muss nun noch a_2 , a_3 und M bestimmt werden. Die beiden a Koeffizienten lassen sich durch Verwendung von (24) und (26) bestimmen und ergeben sich dann mit den Vereinfachungen

$$\Delta\omega = \omega_T - \omega_0 \quad \text{und} \quad \Delta\phi(M) = \phi_T - \phi_0 + 2\pi M$$

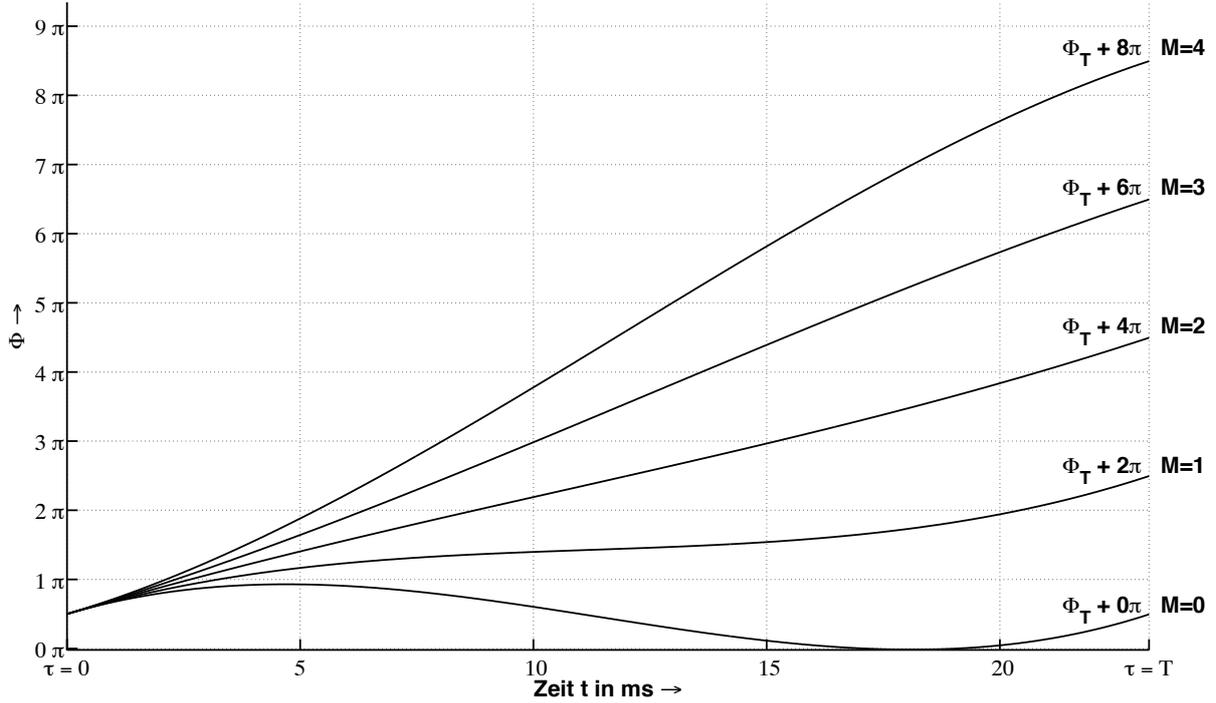
zu

$$a_2 = \frac{3\Delta\phi(M) - 3\omega_0 T - \Delta\omega}{T^2} \quad (27)$$

$$a_3 = \frac{\Delta\omega T - 2\Delta\phi(M) + 2\omega_0 T}{T^3} \quad (28)$$

Grundsätzlich ist die Interpolationsfunktion (20) damit ausreichend bestimmt. Die Startwerte ϕ_0 und ω_0 , sowie die Endwerte ϕ_T und ω_T sind bereits Teil der Lösung. Es soll jedoch auch erreicht werden das der Verlauf der Funktion möglichst glatt sein soll (Abbildung 10). Diese Bedingung lässt sich formulieren als eine möglichst kleine Frequenzänderung:

$$\int_0^T \left(\frac{d\omega(t)}{dt} \right)^2 = \int_0^T (2a_2 + 6a_3 t)^2 dt \rightarrow \text{Minimum} \quad (29)$$

Abbildung 10: Funktionsschaar der Interpolation für verschiedene Werte für M

Gesucht ist also jenes M welches den Ausdruck (29) minimal werden lässt.

$$\frac{d}{dM} \int_0^T (4a_2^2 + 24a_2a_3t + 36a_3^2t^2) dt \stackrel{!}{=} 0 \quad (30)$$

Die Lösung von (30) ergibt sich dann zu:

$$M = \text{round} \left[\frac{1}{2\pi} \left(\frac{T\Delta\omega}{2} - \phi_T + \phi_0 + \omega_0 T \right) \right] \quad (31)$$

Da $\cos(\phi_T) = \cos(\phi_T + 2\pi M)$ nur für ganzzahlige Werte von M gilt wird in (31) gerundet. Damit ist nun die Interpolationsfunktion (20) vollständig beschrieben. Abschließend fehlt nun noch die Umsetzung vom zeitkontinuierlichen auf den zeitdiskreten Fall:

$$0 < t < T \quad \rightarrow \quad 0 < n < hs \quad (32)$$

mit $n = t \cdot f_s$ und der Hopsize hs .

3.6 Stochastischer Teil

Der stochastische Signalanteil besteht aus Geräuschanteilen, welche nicht schmalbandig sind bzw. eine kurze zeitliche Ausdehnung besitzen. Er berechnet sich aus der zeitlichen Differenz (Gleichung (3), S. 2) zwischen Originalsignal und den synthetisierten Partiatönen (deterministischer Signalanteil). Ein stochastisches Signal lässt sich vollständig durch sein Leistungsdichtespektrum beschreiben. Es ist außerdem weder notwendig die Momentanphase noch die exakten Amplitudenwerte des Spektrums für jeden Signalblock abzubilden.

Das hier verwendete Verfahren zur Analyse bzw. Synthese des Residuums stellt eine Kombination aus der Constant-Q Transformation und einem klassischen Filterbank Ansatz dar. Ersteres weil die Verteilung der Mittenfrequenzen der Filter und deren Bandbreiten nach der Idee der CQ Transformation geschieht. Die Transformation selbst wird jedoch nicht durchgeführt. Die Idee war eine Filterbank zu entwerfen welche der menschlichen Hörwahrnehmung Rechnung trägt. Der geringe Berechnungsaufwand, sowie die Nähe zur Bark Skala, gaben den Ausschlag für diese Implementierung.

3.6.1 Filterbankparameter

Durch die festgelegte Anzahl b an Filtern pro Oktave und der tiefsten(ersten) Mittenfrequenz f_0 ergibt sich die Mittenfrequenz f_k des k -ten Filters durch:

$$f_k = f_0 2^{\frac{k}{b}} \quad \text{mit } k = 0, \dots, K-1 \quad (33)$$

Die Bandbreiten Δf_k lassen sich berechnen durch:

$$\Delta f_k = f_k (2^{1/b} - 1) \quad (34)$$

Es soll weiters garantiert werden dass die obere Grenzfrequenz f_g des letzten Filters $K-1$ unter $f_s/2$ liegt.

$$\begin{aligned} f_{g,K-1} &= f_{K-1} + \frac{1}{2} \Delta f_{K-1} \\ &= f_0 2^{\frac{K-1}{b}} + \frac{1}{2} f_0 2^{\frac{K-1}{b}} (2^{1/b} - 1) \\ &< f_s/2 \end{aligned} \quad (35)$$

Löst man diese Gleichung nach K so ergibt sich daraus für die gesamte Anzahl an Filtern:

$$K < \text{ld} \left(\frac{f_s}{2f_0 + f_0(2^{1/b} - 1)} \right) b + 1 \quad (36)$$

Es wird also die Anzahl an Filtern pro Oktave b sowie die Mittenfrequenz des ersten Filters f_0 festgelegt und damit die gesamte Anzahl an Filtern K berechnet.

3.6.2 Filterbankimplementierung

Die einzelnen Bandpässe, welche in Matlab als Butterworth Filter 5ter Ordnung realisiert wurden, können bei tiefen Mittenfrequenzen und kleinen Bandbreiten, Instabilitäten aufweisen. Dies ist ein Problem der Matlab eigenen Funktion *butter()*, welche die Filterkoeffizienten berechnet. Zur Umgehung dieses Problems wurden die erzeugten Butterworth-Filter in Kaskaden, bestehend aus 5 Filtern 2ter Ordnung, zerlegt (Abbildung 11). Die

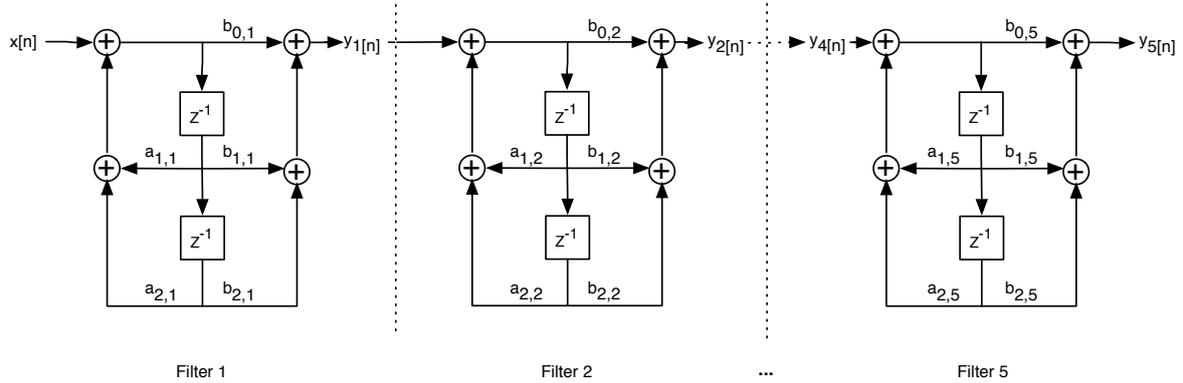


Abbildung 11: Zerlegung eines Bandpasses 5ter Ordnung in 5 Filter 2ter Ordnung

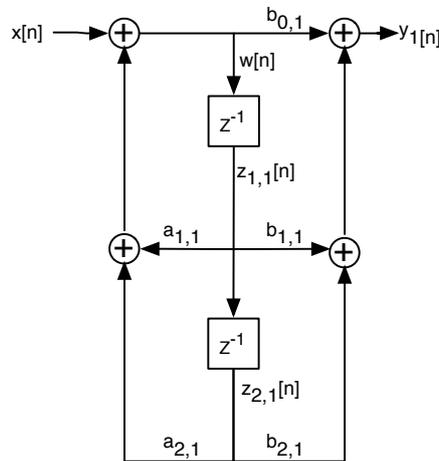


Abbildung 12: BSP: Erstes Filter einer Kaskade im Detail

Umrechnung der Filterkoeffizienten wurde mit der Matlab eigenen Funktion *zp2sos()* vorgenommen.

Am Beispiel eines Filters (Abbildung 12) sei die Funktionsweise zur Bestimmung der Ausgangsfolge beschrieben:

Mit Hilfe der Zwischengröße $w[n]$

$$w[n] = x[n] + a_{1,1}z_{1,1}[n] + a_{2,1}z_{2,1}[n] \tag{37}$$

ergibt sich die Ausgangsfolge zu

$$y_1[n] = b_{0,1}w[n] + b_{1,1}z_{1,1}[n] + b_{2,1}z_{2,1}[n] \tag{38}$$

$y_1[n]$ stellt in weitere Folge den Eingang des nächsten Filters dar.

Da alle Signale in Blockverarbeitung abgearbeitet werden, muss garantiert werden dass die Ausgänge der Filter stets einen kontinuierlichen Verlauf haben. Die inneren Zustände $z_1[n]$ und $z_2[n]$ aller 5 Filter einer Kaskade, sowie aller k Bandpässe, müssen am Ende

eines Frames gespeichert werden, um damit alle Filter am Beginn des nächsten Frame zu initialisieren. Dies gilt sowohl für die Analyse- als auch für die Synthese-Filterbank. Erst dadurch ist sichergestellt, dass alle im Eingangssignal enthaltenen Ein- und Auschwingvorgänge korrekt reproduziert werden und keine Unstetigkeiten auftreten.

3.6.3 Analyse und Synthese

Wie in Abbildung 13 zu sehen, wird das gefensterte Eingangssignal $x_m[n]$ durch die k Bandpässe in Teilbänder unterteilt. In jedem Band wird anschließend der RMS Wert bestimmt.

$$RMS = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (x_{m,k}[n])^2} = g_m(k) \quad (39)$$

mit N der Fensterlänge. In der Synthese (Abbildung 14) wird nun die gleiche Filterbank mit weißem Rauschen angeregt und die Ausgänge der Bandbässe mit den den Gewichten $g_m(k)$ multipliziert. Abschließend werden alle Signale aufsummiert und mit der Varianz σ_{noise}^2 von weißem Rauschen skaliert.

Die Varianz von weißem Rauschen lässt sich berechnen durch

$$\sigma_{noise}^2 = \frac{\Delta^2}{12} = \frac{(\max(x_{noise}) - \min(x_{noise}))^2}{12} = \frac{(1 - (-1))^2}{12} = \frac{1}{3} \quad (40)$$

Um einen kontinuierlichen Verlauf der einzelnen Gewichte zu garantieren werden die analysierten Werte des vergangenen Frames $g_{m-1}(k)$ und die des aktuellen $g_m(k)$ für die Verwendung in der Synthese linear interpoliert.

(41)

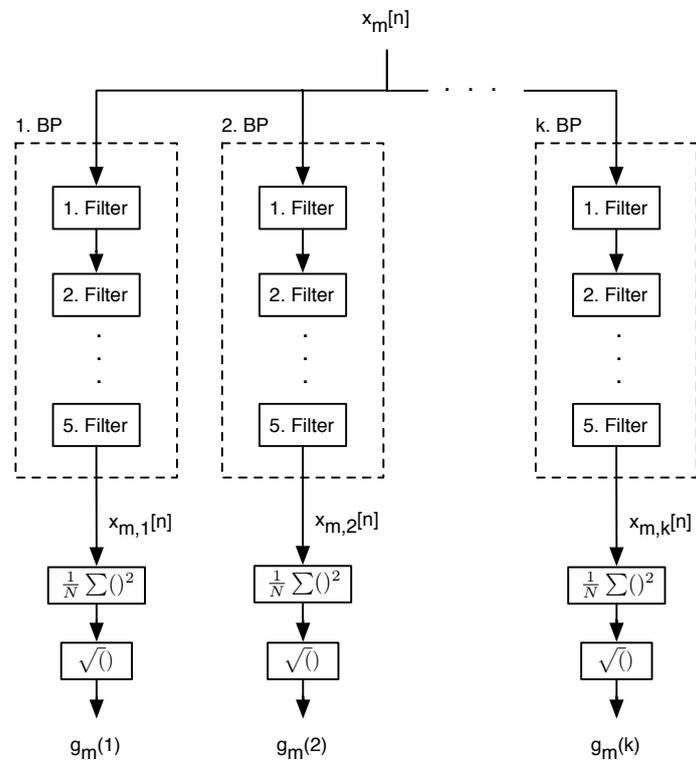


Abbildung 13: Residuumanalyse und Bestimmung der amplitudenbezogenen Gewichte $g_m(k)$ aus gefensterter Signalblock $x_m[n]$

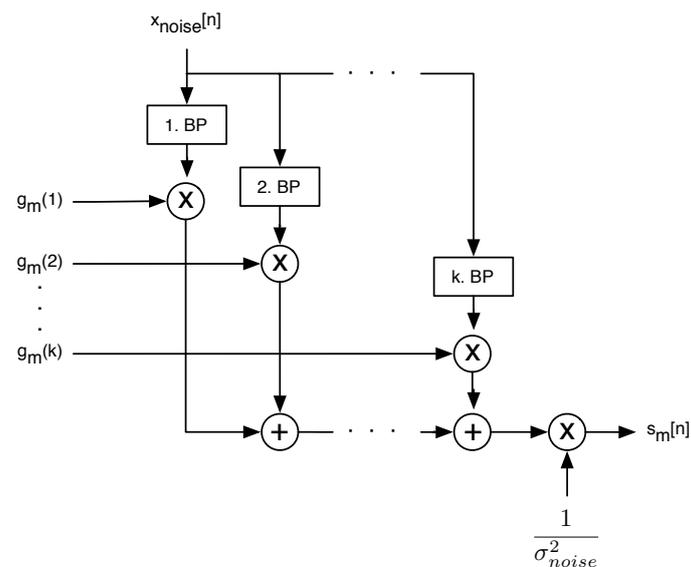


Abbildung 14: Synthese des stochastischen Teils durch gefiltertes weißes Rauschen und Gewichtung mit $g_m(k)$

4 Ausfallsverschleierung

Da die Detektion eines Ausfalls nicht Teil dieses Projektes ist, kann der Zeitpunkt eines Ausfalls als bekannt angenommen werden. Um eine Verschleierung durchzuführen, wird folgendermaßen vorgegangen.

Durch die in Abschnitt 3 beschriebene Analyse sind folgende Parameter des Eingangssignals bekannt:

a) **Partialtöne (Spuren):**

- 1) Amplitude A
- 2) Frequenz f
- 3) Phase ϕ

b) **Residuum**

- 4) Spektrale Gewichte g

Findet nun im aktuellen Frame m ein Ausfall statt, so werden die gewonnenen Parameter des vorangegangenen Frames $m - 1$ dazu genutzt, jene parametrische Darstellung des Signals fortzusetzen. Zur genaueren Beschreibung betrachten wir die beiden Hauptkomponenten des synthetisierten Signals getrennt.

a) **Partialtöne / Spuren**

Tritt im aktuellen Frame m ein Ausfall auf, so werden alle im Frame $m - 1$ gestartete Spuren, sowie bereits laufende Spuren fortgesetzt. Die Nullphase $\phi_{0,m}^i$ der i -ten Spur wird nicht neu detektiert sondern aus der im letzten Frame gemessenen Phase $\phi_{0,m-1}^i$ und Frequenz f_{m-1}^i berechnet:

$$\phi_{0,m}^i = \phi_{0,m-1}^i + 2\pi \frac{f_{m-1}^i}{f_s} h s \quad (42)$$

Die Amplitude A^i und Frequenz f^i wird für die Dauer des Ausfalls konstant gehalten. Wenn in einem vollständigen Fenster kein Ausfall mehr vorkommt, werden die Analysedaten wieder genutzt und es wird eine Hopsize nach dem normalen, in Abschnitt 3.5.1 beschriebenen, Verfahren synthetisiert. Abbildung 15 zeigt den analysierten Verlauf der Partialtöne, eines Klavierauszugs der Pathétique von Ludwig v. Beethoven. Deutlich zu erkennen ist das Beibehalten des Amplitudenspektrums während der Ausfallsverschleierung.

b) **Residuum**

Analog zu a) werden auch hier die gewonnenen Parameter genutzt, um im Falle eines Ausfalls die parametrische Darstellung des Residuums fortzusetzen. Wenn ein Ausfall in Frame m stattfindet, muss bereits in Frame $m - 1$ die Analyse des Residuums stattgefunden haben, wodurch die Gewichte g_k aller k Filter der Filterbank ermittelt wurden. Dazu ist es weiters notwendig das auch die Synthese des deterministischen Teils bereits in Frame $m - 1$ durchgeführt wurde.

Die Analyse des deterministischen Teils (Peak Detektion, Spurenbildung) wird in der Implementierung der Ausfallsverschleierung immer durchgeführt, und zwar unabhängig vom Auftreten eines Ausfalls. Die Synthese der Partialtöne und die Modellierung des Residuums geschieht jedoch nur ab einem Frame vor dem Ausfall bis inklusive des ersten Frames nach dem Ausfall (Abbildung 16).

Das in dieser Arbeit beschriebene Analyse/Synthese Verfahren kann eine perzeptiv fehlerfreie Nachbildung des Originalsignals erzeugen, wenn alle signifikanten Signalanteile korrekt detektiert und synthetisiert werden. Da es aber in der Praxis leicht zu Fehlern kommen kann, z.B. durch einen nicht optimal signalangepassten Schwellwert oder Fehler bei der Parameterkorrektur der Partialtöne (Abweichung zwischen erkannten und tatsächlich auftretenden Partialtönen), wurde die Verwendung von Cross-Fades implementiert. Dadurch besteht die Möglichkeit vor, beziehungsweise nach dem Ausfall das Originalsignal und das synthetisierte Signal zu überblenden (siehe Abbildung 16). Die hierfür zuständigen Parameter sind *drop_prefade* und *drop_postfade* (siehe Abschnitt 5.1, 6.2). Die Beeinflussung dieser Parameter auf die Ausfallsverschleierung lässt sich in Abbildung 17 erkennen.

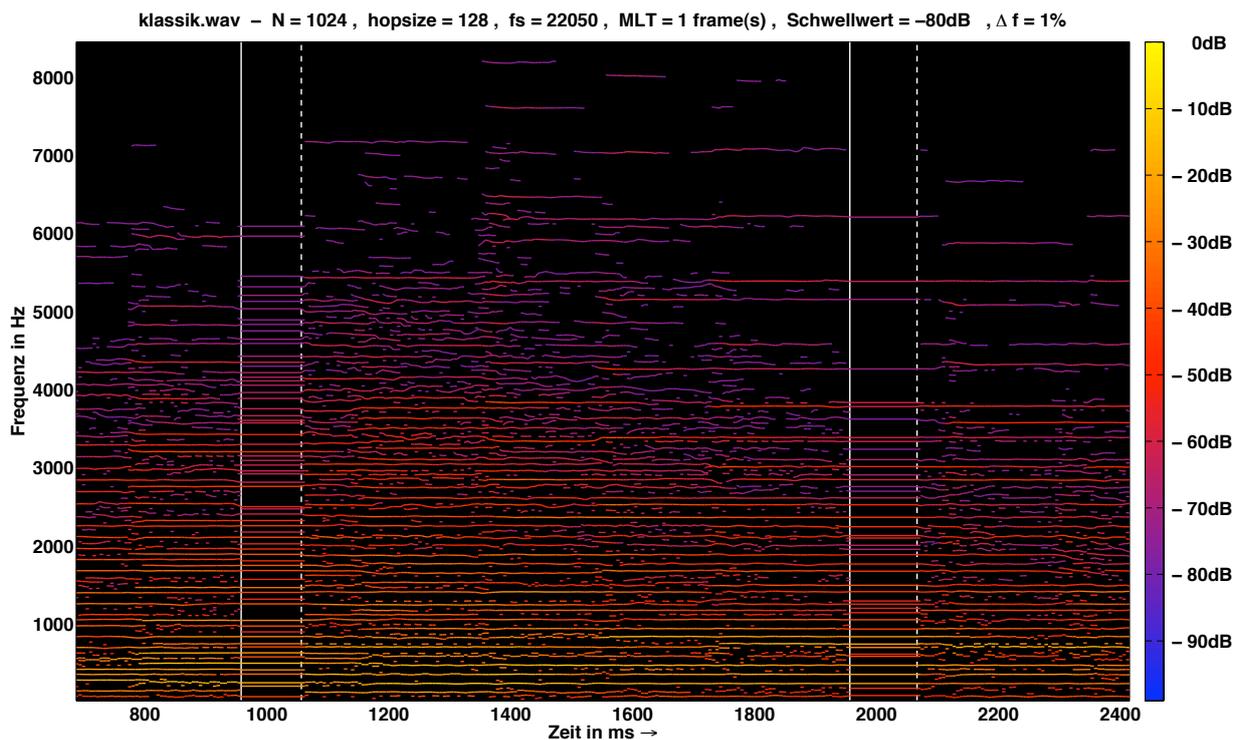


Abbildung 15: Partialtonverlauf eines Klavierauszugs der „Pathetique von Ludwig v. Beethoven“, durchgezogene Linie markiert den Beginn und strichlierte Linie das Ende der Ausfallsverschleierung

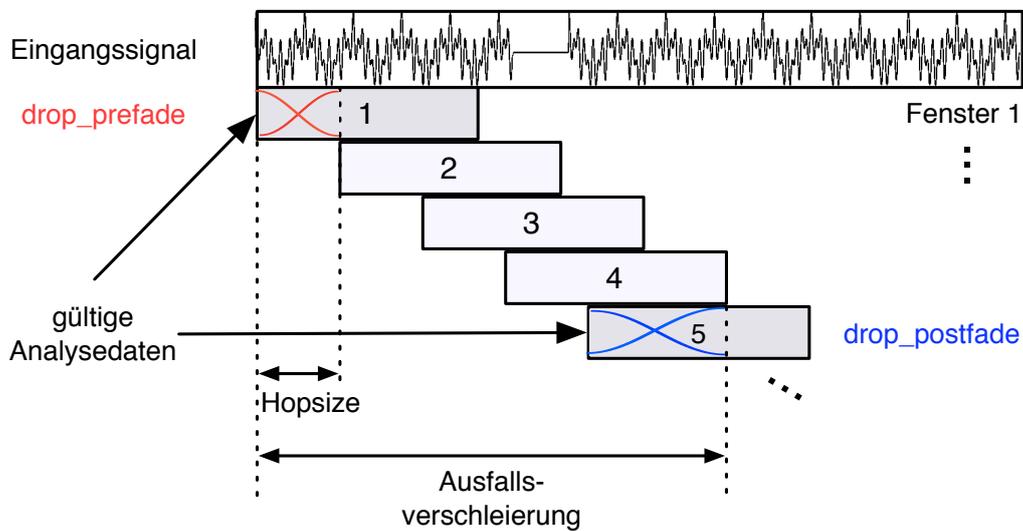


Abbildung 16: Ausfallsverschleierung, (Fenster 1) beginnende Ausfallsverschleierung, (Fenster 5) endende Ausfallsverschleierung

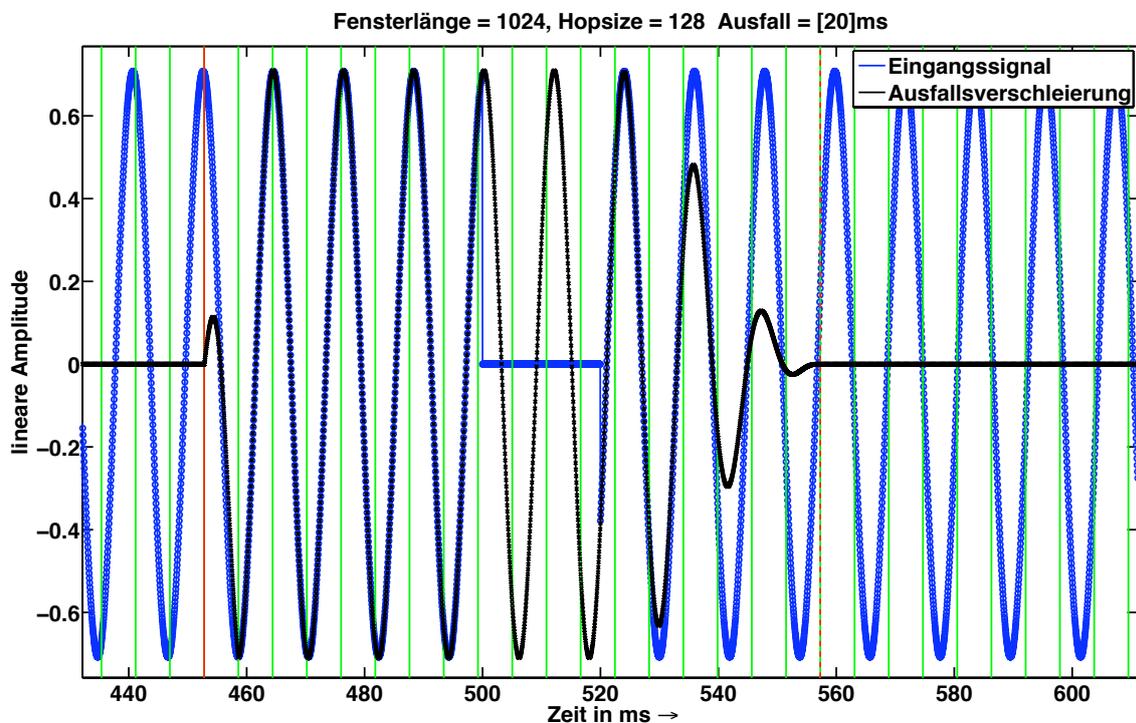


Abbildung 17: Ausfallsverschleierung am Beispiel eines Sinustons, durchgezogene rote Linie markiert den Beginn und strichlierte rote Linie das Ende der Ausfallsverschleierung

5 Programmbedienung

5.1 Parameter

Auflistung der im Hauptprogramm *dropcon.m* einstellbaren Parameter:

Parameter:	signal
Funktion:	Wahl des Eingangssignals
mögliche Werte:	showdialog öffnet nach Programmstart eine Dialogbox zur Auswahl einer spezifischen Datei
Parameter:	dropout_pos
Funktion:	Position des Ausfalls in ms
mögliche Werte:	Kann als Zeilenvektor auch mehrere Positionen beschreiben
Parameter:	dropout_len
Funktion:	Länge des Ausfalls in ms
mögliche Werte:	Kann als Zeilenvektor auch mehrere Längen beschreiben
Parameter:	drop_prefade (siehe Abschnitt 4)
Funktion:	Einblendung des rekonstruierten Signals vor beginnendem Ausfall
mögliche Werte:	TRUE / FALSE
Parameter:	drop_postfade (siehe Abschnitt 4)
Funktion:	zeitliche Länge in ms der Ausblendung des rekonstruierten Signals nach Wiedereinsetzen des Original Signals
mögliche Werte:	ganzzahlige Werte ≥ 0 ; $0 \hat{=}$ Keine Ausblendung
Parameter:	win_len
Funktion:	Länge des Fensters der Blockverarbeitung
mögliche Werte:	ganzzahlige Werte, Zweierpotenzen sinnvolle Werte: 2048, 1024
Parameter:	hopsiz
Funktion:	Hopsiz der Blockverarbeitung
mögliche Werte:	ganzzahlige Werte; muss kleiner als Fensterlänge sein, sinnvolle Werte: win_len/8, win_len/4, win_len/2
Parameter:	thresh (siehe Abschnitt 3.3.5)
Funktion:	Fixer globaler Schwellwert in dB der Peak Detektion
mögliche Werte:	ganzzahlige negative Werte
Parameter:	thresh_local (siehe Abschnitt 3.3.5)
Funktion:	Fixer lokaler Schwellwert in dB der Peak Detektion
mögliche Werte:	ganzzahlige negative Werte
Parameter:	dyn_threshold (siehe Abschnitt 3.3.5)
Funktion:	frameweise Berechnung des Schwellwerts zur Peak Detektion
mögliche Werte:	TRUE / FALSE

Parameter:	delta_f (siehe Abschnitt 3.4.2, 6.1.4)
Funktion:	Zuordnung der Peaks zu Spuren im Bereich von $\pm\Delta_f$ einer Spur
mögliche Werte:	jede Prozentangabe ≥ 1
Parameter:	mlt (siehe Abschnitt 3.4.3)
Funktion:	Minimale Lebenszeit einer Spur
mögliche Werte:	Anzahl der Frames
Parameter:	b (siehe Abschnitt 3.6.1)
Funktion:	Anzahl an Filtern pro Oktave der Residuumfilterbank
mögliche Werte:	ganzzahlige Werte ≥ 1
Parameter:	f0 (siehe Abschnitt 3.6.1)
Funktion:	Mittenfrequenz des ersten (tiefsten) Filters der Residuumfilterbank
mögliche Werte:	Angabe in Hz, ganzzahlige Werte
Parameter:	progress
Funktion:	Fortschrittsanzeige im Command Window
mögliche Werte:	TRUE / FALSE
Parameter:	plots_all
Funktion:	gibt an ob Plots ausgegeben werden sollen
mögliche Werte:	TRUE / FALSE, Wenn TRUE: keine Ausfälle \rightarrow Betragsspektren von Eingangssignal und Synthese dyn_threshold=TRUE \rightarrow Schwellwertverlauf wird angezeigt
Parameter:	plot_time
Funktion:	Plot Amplitude über Zeit der Signale im Zeitbereich
mögliche Werte:	TRUE / FALSE
Parameter:	plot_partial
Funktion:	Plot der Partialtöne über Zeit und Amplitude
mögliche Werte:	TRUE / FALSE
Parameter:	make_newdir
Funktion:	Erzeugt für jeden Durchlauf ein eigenes Verzeichnis
mögliche Werte:	TRUE / FALSE
Parameter:	verbose
Funktion:	zusätzliche Informationen und Warnungen werden bei Auftauchen im Command Window ausgegeben
mögliche Werte:	TRUE / FALSE

5.2 Ausgabe

Abseits der Ausgabe im Command Window und der grafischen Ausgabe (Zeitsignale, Partialtonverlauf, Schwellwertverlauf,...), wird für jeden Durchlauf ein Bericht *report.txt* im Ausgabeverzeichnis *output/* erzeugt, der alle gewählten Parameter sowie zusätzlich vom Programm berechnete Daten zusammenfasst. Ein beispielhafter Bericht ist in Abbildung 18 zu sehen. Da der Algorithmus im Falle keiner Ausfälle im gesamten Signal das Eingangssignal vollständig analysiert und synthetisiert unterscheidet sich die Anzahl der gespeicherten Signale. Dies wird in Tabelle 2 mit Resynthese bezeichnet.

Dateiname	Ausfallsverschleierung	Resynthese
1_original.wav	Eingangssignal	
2_deterministic.wav	deterministischer Teil	
3_residuum.wav	-	Residuum
4_stochastic.wav	stochastischer Teil	
5_output.wav	rekonstruierte Signalböcke	resynthetisiertes Eingangssignal
6_restored.wav	rekonstruiertes Eingangssignal	-

Tabelle 2: Auflistung der gespeicherten Signale

```

dropout concealment
-----
TESTSIGNAL
Signal      : testsignals/flute.wav
length     : 132349 samples (3.0011 s)
fs         : 44100 Hz
-----
DROPOUT(s)
dropout 1   : 10ms (441 samples) @ 400ms
dropout 2   : 20ms (882 samples) @ 1200ms
drop_prefade : true
drop_postfade : 30 ms
-----
WINDOWING
win_len    : 1024 samples (23.22 ms)
hopsize    : 128 samples (2.9025 ms)
-----
PEAK-DETECTION
dyn_threshold : true
thresh_local  : -30 dB
-----
PARTIAL TRACKING
delta_f     : 1 %
mlt        : 1 frame(s)
-----
STOCHASTIC SYNTHESIS
b          : 3 filter per octave
f0        : 200 Hz
-----
SUMMARY
dropout    : 3456 samples (78.3673ms) effectivly replaced
dropout fade : 1408 samples (31.9274ms) effectiv crossfade @ end of each dropout
-----
OUTPUT
directory  : output/18_06_2009--09_55_54/
-----
REMARKS
-----

```

Abbildung 18: Beispielhafter Bericht *report.txt* nach Programmdurchlauf

6 Evaluierung

Ziel dieses Projektes war es einen Ausfall, im besten Falle, nicht hörbar zu verschleiern. Die erreichte Qualität lässt sich demnach am besten durch Hörbeispiele beurteilen. Um dennoch eine qualitative Beschreibung zu geben betrachten wir die maßgeblichen Funktionen und Faktoren des hier vorgestellten Algorithmus.

6.1 Analyse/Synthese

Ein kritischer Faktor des gesamten Analyse/Synthese Systems liegt im Finden der signifikanten Signalanteile. Eine schlechte Erkennung führt zu einem deterministischen Teil welcher zu wenig Partialtöne enthält und dadurch wird das Residuum Sinuskomponenten enthalten, welche in weiterer Folge durch die Residuumsfilterbank analysiert und mit weißem Rauschen nachgebildet werden. Abseits des kleinst möglichen Frequenzabstands zweier Peaks $\Delta f_{min} = 2f_s/N$ (Siehe Abschnitt 3.3.1, S.6), ist die Qualität der Parameterkorrektur detektierter Peaks von entscheidender Bedeutung.

Es stellt sich also die Frage, wie genau die Verfahren zur Korrektur der Amplitude, Frequenz und Phase eines Partialtons sind.

6.1.1 Qualität der Korrekturverfahren

Abbildung 19 zeigt den entstehenden Fehler der parabolischen Amplitudenkorrektur. Er wurde mit einem Sinuston fixer Amplitude ($=-6\text{dB}$) und zwischen 2 Frequenzbins ermittelt.

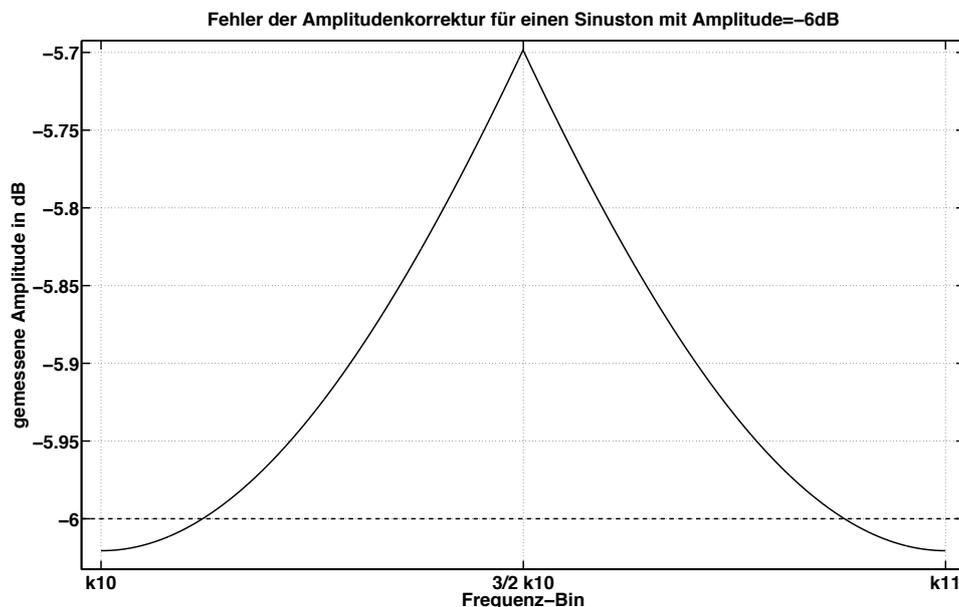


Abbildung 19: Fehler der Amplitudenkorrektur für einen Sinuston mit Amplitude = - 6 dB zwischen den Frequenzbins k_{10} und k_{11} , $f_s=44,1\text{kHz}$

Da der gemessene Fehler nur vom Abstand der tatsächlichen Frequenz zum nächstgelegenen

Bin abhängt, ist er über den gesamten Frequenzbereich gleich, bzw. ließe sich die Messung für beliebige aufeinander folgende Bins mit gleichem Ergebnis durchführen. Der maximale Fehler von 0,3 dB wird exakt zwischen zwei Bins erreicht und stellt einen schwer wahrnehmbaren Fehler dar.

Der Fehler der Frequenzkorrektur, zu sehen in Abbildung 20, wurde mit einem Sinuston mit ansteigender Frequenz und fixer Amplitude -10dB ermittelt. Die tiefste Testfrequenz entspricht dem ersten Frequenzbin ($\sim 43\text{Hz}$) und die strichlierte Linie markiert den zweiten Frequenzbin ($\sim 86\text{Hz}$). Der relevanteste Fehler im Bezug auf den jeweiligen Frequenzbereich stellt die erste Erhöhung dar und liegt mit 3,5 Hz für ca. 65 Hz bei fast einer kleinen Sekunde ($\sim 4.3\text{ Hz}$). Das die Position dieses Fehlers in der Mitte zwischen erstem und zweiten Frequenzbin liegt daran liegen das allgemein die Information (Betrag, Phase) eines Sinustons dessen Frequenz genau zwischen zwei Bins liegt, am unschärfsten ist und sich daher nur schwer korrigieren lässt (vergleiche Abbildung 19). Dies lässt sich auch bei höheren Frequenzen betrachten, nur ist der Fehler in Bezug auf die auftretende Frequenz wesentlich kleiner, was ihn weniger bedeutend macht. Dass der entstehende Fehler zu tiefen ($<100\text{Hz}$) und zu hohen ($>10\text{kHz}$) Frequenzen zunimmt, begründet sich in der Tatsache dass, das Verfahren von Desainte [DCM00] auf der diskreten Ableitung im Zeitbereich beruht. Zu tiefen Frequenzen hin entsteht ein immer kleiner werdender Amplitudenunterschied zwischen zwei aufeinander folgenden Abtastzeitpunkten und die Korrektur wird demnach immer ungenauer. Zu hohen Frequenzen hin tendieren benachbarte Amplitudenwerte zum Extremfall $+1/-1/+1 \dots$ bei welchem die Ableitung zu Null wird.

Es wurde zum Vergleich auch die Frequenzkorrektur mit Hilfe der parabolischen Interpolation getestet. Sie lieferte eine, über den gesamten Frequenzbereich, relativ konstante Abweichung zur Testfrequenz, jedoch mit größerem Betrag und wurde daher bei der Implementierung nicht verwendet.

Der Fehler in der Erkennung der korrekten Phase (Abbildung 22) eines Sinustons stellt, aufgrund der Bildung der zeitlichen Differenz zwischen Originalsignal und deterministischen Teil zur Berechnung des Residuums, den kritischsten Fehler dar. Selbst wenn die Amplitude und Frequenz eines Sinustons exakt, die Phase jedoch nur ungenau bestimmt wurde, ergibt sich im Residuum durch die zeitliche Differenz ein Sinuston mit gleicher Frequenz und Amplitude, nur mit anderem Phasenbelag. Dieser würde in weiterer Folge durch die Filterbankanalyse und Synthese zu einem starken Rauschen führen. Abseites dessen führt ein zu großer Phasenfehler in der Synthese zu Unstetigkeiten in der Phasenfurthersetzung eines Partialtons. Dies wird durch die in Abschnitt 3.5.2 beschriebene Phaseninterpolation verhindert. Der Fehler der Frequenzkorrektur (Abbildung 20) und jener der Phasenkorrektur (Abbildung 21) stehen in direktem Zusammenhang, da zur Korrektur der Phase der korrigierte Frequenzwert herangezogen wird. Abbildung 22 zeigt, dass in jenem Frequenzbereich (ca 200 Hz - 2 kHz) in welchem die Frequenzkorrektur genaue Werte liefert, die Phasenkorrektur eine hohe Genauigkeit besitzt.

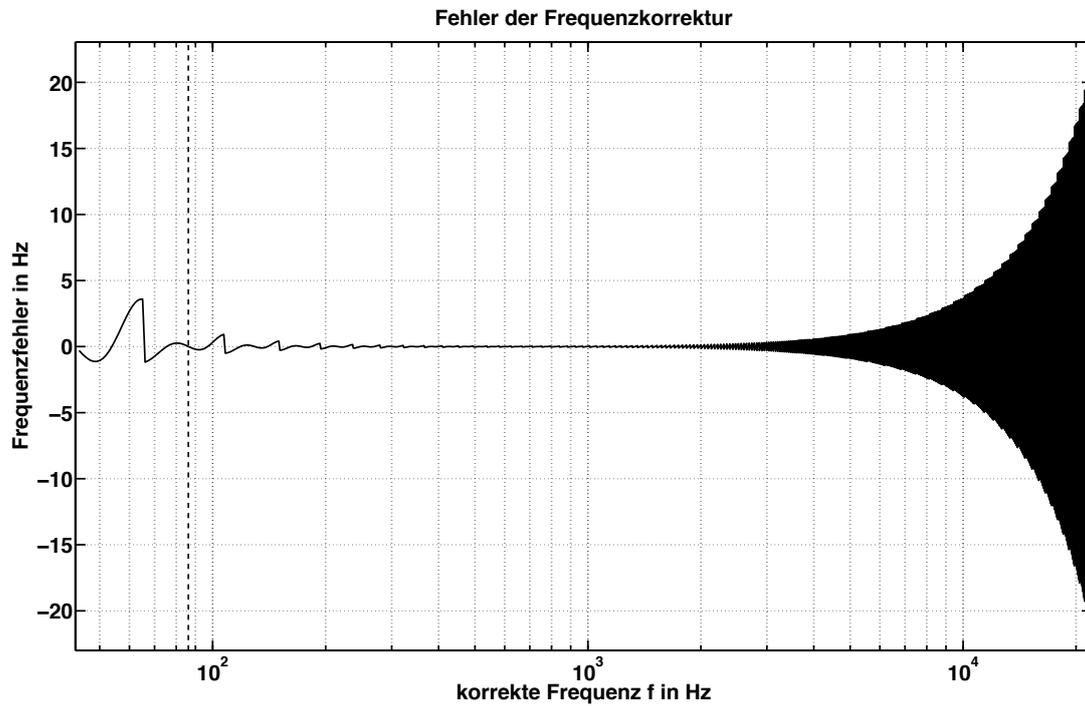


Abbildung 20: Fehler der Frequenzkorrektur, Sinuston mit Amplitude=-10dB, $f_s=44,1$ kHz, strichlierte Linie markiert dem 2ten Frequenzbin ~ 86 Hz

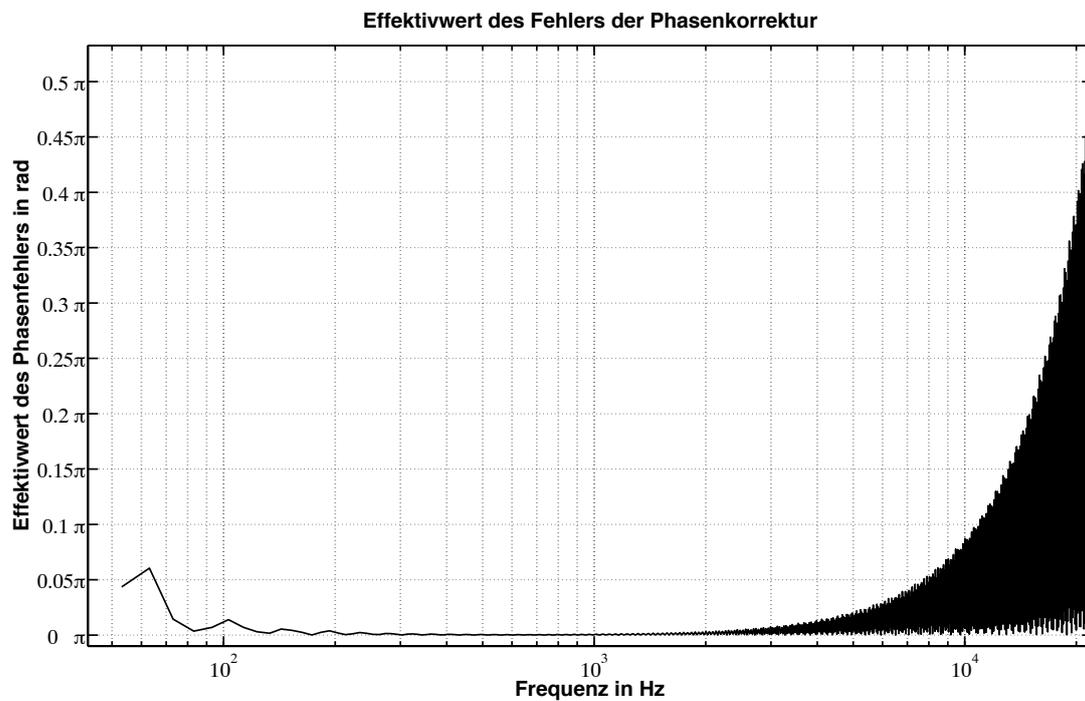


Abbildung 21: Effektivwert des Fehlers der Phasenkorrektur, $f_s=44,1$ kHz

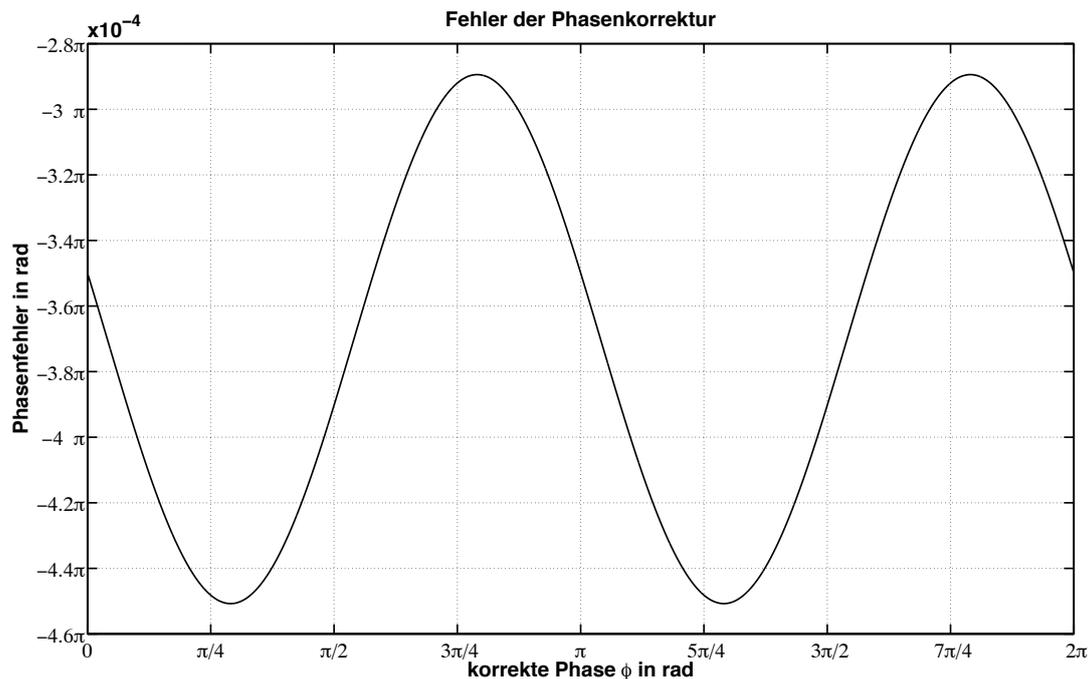


Abbildung 22: Fehler der Phasenkorrektur, $f_s=44,1\text{kHz}$, gemessen mit Sinuston mit $f=1\text{kHz}$

6.1.2 zeitliche Fehler

Wir haben bis jetzt die entstehenden Fehler im Frequenzbereich betrachtet. Ein entscheidendes Kriterium für die Qualität des Systems ist außerdem das zeitliche Verhalten. Grundsätzlich erzeugt die verwendete Fensterlänge eine zeitliche Unschärfe. Beispielfähig ergibt sich diese bei der hier verwendeten Fensterlänge von 1024 Samples zu $\sim 23\text{ms}@44,1\text{kHz}$. Zeitliche Abläufe die kürzer sind können nicht aufgelöst werden, was z.B. für Transienten der Fall ist. In Bezug auf die Problematik der Ausfallsverschleierung ist jedoch das Nichterkennen von Transienten eher ein Vorteil. Könnten Transienten erkannt werden und würde weiters ein Ausfall direkt nach Auftreten eines Solchen stattfinden, würde sein geräuschhaftes Spektrum fortgesetzt und dadurch als sehr störend wahrgenommen werden.

Zeitliche Fehler entstehen auch durch die Eigenschaften und Vorgaben des Analyse bzw. Synthese Systems. Die Abbildungen 23 und 24 zeigen das Eingangssignal, die deterministische Synthese sowie das entstehende Residuum im Zeitbereich. Die schwarzen Linien markieren jeweils eine Hopsizelänge und damit jenen Bereich der pro Frame synthetisiert wird. Deutlich zu sehen ist die Vorgabe der Spurenbildung, dass eine beginnende Spur (Partialton) bereits ein Frame früher eingeblendet wird (Abbildung 23). Dies verursacht einen Fehler von einer Hopsizelänge. Für die hier verwendete Hopsizelänge von 128 Samples, ergibt sich ein Fehler von $2,9\text{ms}@44,1\text{kHz}$. Abbildung 24 zeigt das Verhalten des deterministischen Teils bei einem ausklingenden Sinuston. Gegen Ende füllt der abklingende Ton immer weniger ein Analysefenster. Folglich ist die gemessene Amplitude geringer und die Synthese klingt bereits früher ab.

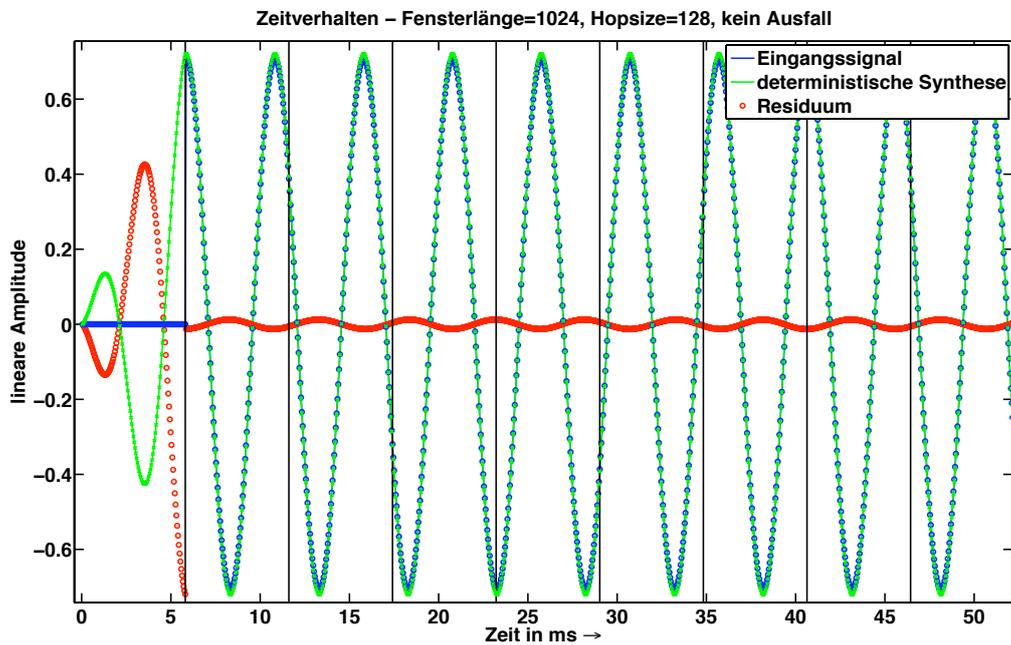


Abbildung 23: Zeitverhalten der Synthese, Beginn eines Sinustons, vertikale Linien markieren jeweils eine Hopsize

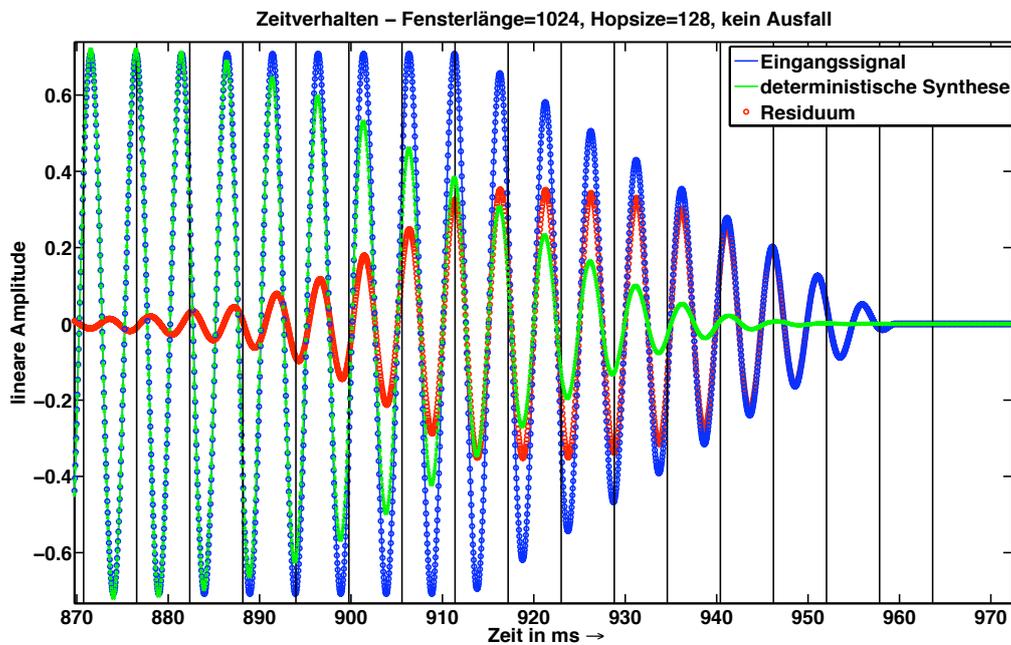


Abbildung 24: Zeitverhalten der Synthese, Ende eines Sinustons, vertikale Linien markieren jeweils eine Hopsize

6.1.3 Schwellwerte

Der dynamische Schwellwert (Abschnitt 3.3.5) berechnet aus dem logarithmischen Betragsspektrum des aktuellen Analysezeitfensters, den Amplitudenschwellwert für die Peak Detektion. Im besten Fall sollte sich ein Schwellwert ergeben welcher direkt über dem Pegel der geräuschhaften Signalanteile liegt. Der hierfür optimale Wert für die Ordnung p zur Bestimmung der Quantile, steht im direkten Zusammenhang mit der verwendeten Fensterlänge, Hopsizelänge und dem betrachteten Signalspektrum. Letzteres stellt den wichtigsten Faktor dar. Grundsätzlich gilt, je dichter ein Signalspektrum, desto kleiner muss p gewählt werden. Die hier verwendete Wert $p = 0.8$ stellt, zusammen mit einer Fensterlänge 200 und einer Hopsizelänge von 100 Abtastwerten, einen guten Kompromiss dar. Das Signalspektrum bei z.B Gitarre, Klavier, Holzbläser, Querflöte und auch Sprache wird gut geschätzt, wohingegen sehr dichtere Spektren einen erhöhten Schwellwert zur Folge haben.

Der lokale Schwellwert wurde empirisch mit -30 dB bestimmt. Wenn ein fixer globaler Schwellwert verwendet wird, stellt der lokale Schwellwert eine deutliche Verbesserung dar, denn gerade bei leisen Passagen eines Signals führt er dennoch zu einer positiven Peakdetektion. Auch bei einem zu hoch berechneten dynamischen Schwellwert (z.B im Falle nicht optimal an das vorhandene Audiomaterial angepasster Fensterlänge, Hopsizelänge und Ordnung p) kommt dieser positive Effekt zum Tragen.

6.1.4 Spurenbildung

Abseits zeitlicher Fehler (6.1.2) kann es zu spektralen "Fehlern" kommen, nämlich dann, wenn der gebildete deterministische Teil mit jenem des Eingangssignals nicht übereinstimmt. Amplituden-, Frequenz- und Phasenfehler der Peak Detektion wurden bereits betrachtet (6.1.1). Eine weitere Fehlerquelle stellt der erzeugte zeitliche Verlauf eines Partialtons also die Zusammenfassung der gefundenen Peaks zu Spuren dar. Der hierfür maßgebliche Faktor ist Δ_f . Wie bereits in Abschnitt 3.4.2 zu sehen, gibt er jenen Frequenzbereich an, in welchem ein detektierter Peak als möglicher Kandidat der Spur betrachtet wird. Wird nun Δ_f groß gewählt, darf sich also die Frequenz einer Spur stark von Frame zu Frame ändern, was zu einer Art singendem Effekt führt. Außerdem wird das Spektrum stark vereinfacht, nachdem in einem sehr breiten Frequenzbereich aus sehr vielen Peaks ein einziger ausgewählt und der Spur zugeordnet wird. Es hat sich deshalb in der Praxis ein Wert von $\Delta_f = 1\%$ als beste Wahl herausgestellt.

Weiters sei hier erwähnt das das Konzept der minimalen Lebenszeit (3.4.3) implementiert wurde, sich aber als wenig sinnvoll heraus gestellt hat. Spuren welche nur eine kurze Lebenszeit besitzen, entstehen hauptsächlich bei Anschlägen und Einschwingvorgängen. Gibt man also eine minimale Lebenszeit von mehreren Frames an, werden diese nicht berücksichtigt. Es wurde daher die minimale Lebenszeit auf 1 Frame gesetzt, im Grunde also nicht verwendet.

6.1.5 Residuumfilterbank

Um die Qualität der Filterbank zu beurteilen, betrachten wir eine vollständige Analyse und Synthese eines Querflötentons. Abbildung 25 zeigt das entstehende Residuum nach

Abzug des deterministischen Teils vom Eingangssignal. Es wurden $b = 3$ Filter pro Oktave und $f_0 = 100\text{Hz}$ verwendet. Mit $f_s = 44,1\text{kHz}$ ergibt sich nach Gleichung (36) (S. 18) eine gesamte Anzahl von 23 Filtern. Man kann erkennen dass das zu analysierende Residuum unterschätzt wird was aber, auch wenn es keiner idealen Rekonstruktion entspricht, auch einen Vorteil in sich birgt. Sollte das zu analysierende Residuum schmalbandige Sinuskomponenten enthalten (was in der Praxis nicht selten vorkommt, da alle Verfahren zur Partialtondetektion und Korrektur der Parameter, nur eine beschränkte Geauigkeit besitzen), welche in einem solchen Fall auch meist lauter als die geräuschhaften Signalanteile sind, so führt dies nicht nicht zu einem überdeutlichen Rauschen. Es gilt zudem dass der deterministische Signalanteil einen deutlichen größeren Beitrag zur wahrgenommenen Qualität der Ausfallsverschleierung liefert.

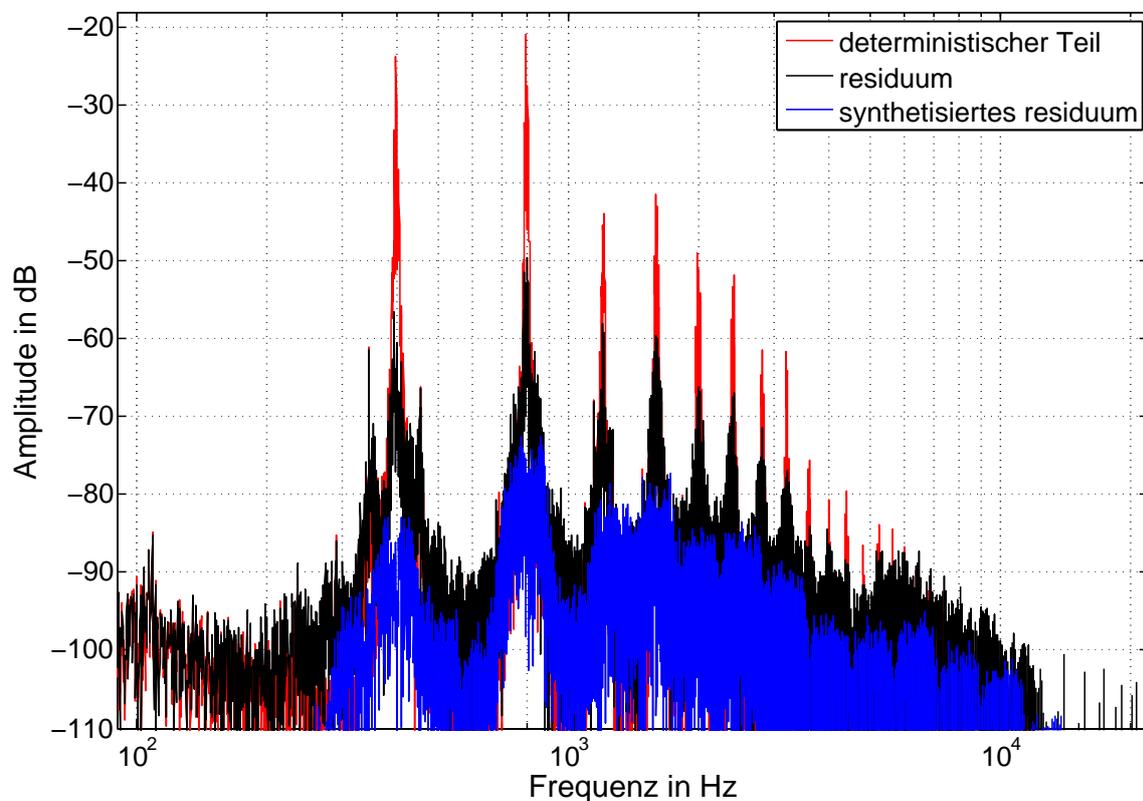


Abbildung 25: Betragsspektrum des deterministischen Teils, des Residuums und des synthetisierten Residuums

6.2 Qualität der Ausfallsverschleierung

Die bislang betrachteten Fehler stellen jene eines Systems dar, welches ein Eingangssignal vollständig analysiert und synthetisiert. Für die Ausfallsverschleierung ist es jedoch nicht notwendig, eine perfekte Signalrekonstruktion zu erreichen, da Verdeckungseffekte und die sehr kurze Ausfallszeit eine wahrnehmbare Verschleierung verringert.

Die Qualität der Ausfallsverschleierung hängt grundsätzlich von jener der Analyse des deterministischen und stochastischen Teils ab. Wird das Spektrum des Eingangssignals

korrekt erkannt und dadurch parametrisiert, so lässt sich ein Ausfall durch direktes einsetzen der Synthese ohne hörbaren Übergang füllen. Bei sehr dichten Spektren können jedoch, durch die beschränkte Frequenzauflösung, nicht alle Partialtöne erkannt werden. Die Folge sind Unstetigkeiten in der Phasenfortsetzung, welche als "Knackser" hörbar werden. Diese können durch eine Überblendung zwischen Eingangs- und rekonstruiertem Signal verhindert werden (*drop-prefade* und *drop-postfade*, Abschnitt 4, Abbildung 17).

Des Weiteren ist die Länge der Ausfälle entscheidend. Da der hier vorgestellte Algorithmus keine Prädiktion beinhaltet entstehen zwangsläufig hörbare Verschleierungen wenn sich der musikalische Kontext während eines Ausfalls ändert. Aber nicht nur spektrale Änderung während eines Ausfalls, sondern auch das spektrale Verhältnis des Eingangssignals von vor zu nach dem Ausfall spielt eine wichtige Rolle. Grundsätzlich werden alle Partialtöne bei einem Ausfall in Amplitude und Frequenz konstant gehalten. Setzt nun das Eingangssignal wieder ein werden dementsprechend Partialtöne beendet, neue erzeugt oder in ihrer Frequenz moduliert um das neue Spektrum zu erzeugen. Besitzt nun das Eingangssignal nach dem Ausfall eine zu große spektrale Differenz zu jenem vor dem Ausfall, entstehen dadurch sehr kurze, schnelle Sweeps welche als störend wahrnehmbar sind. Wird der Parameter *drop-postfade* verwendet, wird diese Art der Störung, durch konstant halten der Partialtöne in ihrer Frequenz und Anzahl bei gleichzeitigem Überblenden von rekonstruiertem Signal und Eingangssignal, verhindert. Es sei erwähnt, dass die zeitliche Länge der Überblendung in ganzzahlige Vielfache einer Hopsizes umgerechnet wird. Für eine Abtastrate von $f_s = 44,1\text{kHz}$ ergeben sich daraus Schrittweiten von 2,9ms. In der Praxis hat sich eine Schrittweite als bester Wert herausgestellt.

Was die Anzahl der zu synthetisierenden Samples N_{synth} betrifft ist zu beachten, dass nicht nur die Anzahl der ausgefallenen Abtastwerte $N_{ausfall}$ erzeugt werden muss, sondern

$$N_{synth} \sim \text{ceil} \left(\frac{N_{ausfall}}{hopsizes} \right) hopsizes + N_{Fenster} \quad (43)$$

Der Grund hierfür liegt in der Gegebenheit, dass nur ein Signalblock in welchem kein Ausfall stattfindet gültige Analysedaten liefert (siehe Abbildung 16, S. 24), also alle Abtastwerte zwischen jenen zwei Blöcken synthetisiert werden müssen. Es sei angemerkt dass Gleichung (43) deshalb nur eine ungefähre Abschätzung darstellt, da nicht nur die Länge eines Ausfalls, sondern auch dessen Position in einem Fenster eine entscheidende Rolle spielt, der gesamte Zusammenhang sich aber mathematisch schwer beschreiben lässt.

Ein Beispiel:

Bei einem Eingangssignal mit $f_s = 44,1\text{kHz}$ und einem Ausfall von 1 Sample, welches mit einer Fensterlänge von 1024 und einer Hopsizes von 128 blockverarbeitet wurde ergibt sich die gesamte Anzahl an erzeugten Samples zu $128 + 1024 = 1152 \hat{=} 26,1\text{ms}$ (zuzüglich der optionalen Überblendung).

Diese Tatsache beschränkt deutlich die maximale Länge eines Ausfalls, denn bei längeren Ausfällen von zB. 30ms müssen bereits ca. 52 ms erzeugt werden und die Wahrscheinlichkeit eine solche Verschleierung nicht mehr wahrzunehmen sinkt erheblich (siehe oben, Länge eines Ausfalls).

6.3 Berechnungsaufwand und Signalverzögerung

Der genaue Berechnungsaufwand lässt sich nicht allgemeingültig berechnen, da er von sehr vielen Faktoren abhängt und dadurch sehr unterschiedlich ausfällt. Qualitativ gesehen teilt sich der gesamte Berechnungsaufwand im Durchschnitt wie in Tabelle 3 zu sehen auf.

Funktion	% des Gesamtaufwands
Spuren Bildung	48
Peak Detektion	29
stochastischer Teil	8
Hauptprogramm	8
dynamischer Schwellwert	6
Partialtonsynthese	0.5
diverse Funktionen	0.5

Tabelle 3: qualitativer Anteil einzelner Funktionen am gesamten Berechnungsaufwand

Zu erkennen ist, dass die aufwendigste Funktion die Spurenbildung darstellt. Der hierfür maßgebliche Faktor ist die spektrale Dichte des Eingangssignals und dessen zeitliches Verhalten. Ein sehr dichtes Spektrum hat zur Folge dass viele Spuren gebildet werden müssen. Ändert es sich zudem noch zeitlich gesehen sehr schnell, so reduziert sich die Lebensdauer der Spuren und es müssen sehr oft Spuren beendet und neu gebildet werden. Dies führt insgesamt zu einer großen Anzahl an Spuren, die bearbeitet werden müssen. Durch die vielen inhärenten logischen Überprüfungen der Spurenbildung, führt dies zu einem hohen Berechnungsaufwand. Wie bereits erwähnt schwankt die in Tabelle 3 zu sehende Verteilung sehr stark, nicht zuletzt mit der Anzahl und Länge der im Eingangssignal vorkommenden Ausfälle.

Die aus dem Algorithmus resultierende Signalverzögerung ist bedingt durch den notwendigen "look-ahead", da bekannt sein muss ob im nächsten Block ein Ausfall stattfindet um das Residuum und daraus die Gewichte der Residuumfilterbank bestimmen zu können. Wie in Abbildung 26 zu sehen, ergibt sich also eine Verzögerung t_{verz} von

$$t_{verz} = \frac{N_{fenster} + N_{hopsizel}}{f_s} \quad (44)$$

Bei Verwendung der Standardwerte $N_{fenster} = 1024$, $N_{hopsizel} = 128$ und einer Samplefrequenz von $f_s = 44,1\text{kHz}$ ergibt sich nach (44) eine Verzögerung von 26,1 ms.

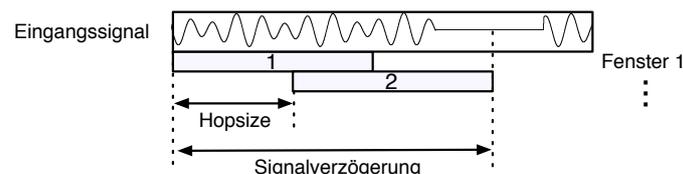


Abbildung 26: Signalverzögerung

7 Mögliche Verbesserungen

Um die Qualität des hier vorgestellten Verfahrens noch zu steigern bieten sich, in Bezug auf die Teilbereiche des Algorithmus, unterschiedliche Möglichkeiten.

Grundsätzlich liegt die größte Beschränkung, wie bei jedem auf der KZFT basierenden Verfahren, in der Zeit-/Frequenzauflösung. Verbesserungen können mittels „Time Frequency Reassignment“ ([HW01], [FH01]) erreicht werden. *Frequency Reassignment* basiert auf der Idee das Spektrogramm in der Art anzupassen, dass die spektrale Energie in jedem Frequenzbin nicht der Mittenfrequenz des Bins sondern jener Frequenz, die seinem spektralen Schwerpunkt entspricht, zugeordnet wird. Analoges gilt für *Time Reassignment*. Dies verbessert deutlich, die durch die Fensterung entstehende spektrale/zeitliche Unschärfe, insbesondere für quasi periodische Signale.

Die Entkopplung der Phasenerkennung und Korrektur von der Frequenz ließe sich durch Verwendung nullphasiger Fenster (siehe [Ser89], [Har78]) erreichen. Eine ungerade Anzahl an Abtatszeitpunkten (Fensterlänge), sowie einen zeitlichen Versatz um eine Halbe Fensterlänge (fftshift) lässt eine symmetrische Impulsantwort entstehen, welche dem entstehenden Spektrum keinen zusätzlichen Phasenbelag aufprägt.

Die Abhängigkeit der zeitlichen Position eines Ausfalls und der daraus resultierenden Qualität der Verschleierung, könnte durch Verfahren aus dem Bereich „*Music Information Retrieval*“ verringert werden. Informationen über den musikalischen Kontext könnten für eine zeitliche und spektrale Prädizierung genutzt werden. Dies würde auch die maximale Länge eines nicht wahrnehmbaren Ausfalls deutlich erhöhen.

Abseits dieser speziellen Verfahren, bietet sich ganz Allgemein eine adaptive Anpassung der Algorithmusparameter auf das Eingangssignal an. Dies stellt, gemessen am Aufwand, wahrscheinlich die einfachste und effektivste Möglichkeit dar.

8 Fazit

Ziel dieses Projektes war es ein Algorithmus auf Grundlage des SMS plus Noise Modells zu erstellen mit dessen Hilfe der Ausfall eines zeitdiskreten Signals verschleiert werden kann. Die Fortführung des Signals sollte weiters möglichst nicht hörbar sein.

Grundsätzlich hängt die Qualität der Verschleierung vom Eingangssignal und von der zeitlichen Position der Ausfälle ab. Anhand der Beispiele lässt sich jedoch erkennen, dass Ausfälle bis $\sim 10\text{ms}$ sehr gut und bis $\sim 35\text{ms}$ gut verschleiern lassen. Außerdem bleibt festzuhalten, dass selbst wenn die Verschleierung hörbar wird, sie weitaus weniger störend ist als die durch einen Ausfall entstehenden „Knack-Geräusche“. Über eine Ausfalllänge von 35ms hinaus nimmt man nur dann die Verschleierung nicht oder nur kaum wahr, wenn das Spektrum des Eingangssignals möglichst wenige klar voneinander getrennte spektrale Spitzen besitzt, wie z.B bei einer Querflöte. Außerdem sollte sich das Spektrum zeitlich gesehen nur langsam ändern, damit die erzeugte „Stationarität“ des Spektrums den musikalischen Kontext nicht verletzt. Ist dies der Fall so sind auch Ausfälle bis 50ms verschleierbar.

Abschließend lässt sich sagen dass, das hier vorgestellte Verfahren einen deutlichen Qualitätsgewinn darstellt. Zudem eignet sich dieses Verfahren durch seine relativ geringe Signalverzögerung von $26,1\text{ms}@44,1\text{kHz}$ und dem moderaten Berechnungsaufwand auch zur Implementierung in einer „echtzeitfähigen“ Sprache wie z.B. C/C++.

Als Anmerkung sei erwähnt, dass eine Auswertung mittels Hörversuche interessant gewesen wäre, aufgrund des zeitlichen Rahmens zukünftigen Arbeiten überlassen wird.

9 Literaturverzeichnis

- [DCM00] M. Desainte-Catherine and S. Marchand, “high precision fourier analysis of sounds using signal derivatives’,” *Journal of the Audio Engineering Society*, vol. Vol. 48, no. No. 7-8, pp. p 654–667, 2000.
- [Fal02] C. Falch, “matlab®-tool für die parameterberechnung zur additiven klangsynthese’,” *Diplomarbeit, IEM Institut für Elektronische Musik und Akustik*, 2002.
- [FH01] K. Fitz and L. Haken, “on the use of time-frequency reassignment in additive sound modeling’,” *AES Journals*, 2001.
- [Har78] F. J. Harris, “on then use of windows for harmonic analysis with the discrete fourier transform’,” *Proceedings of the IEEE*, vol. 60, no. 1, 1978.
- [HW01] S. W. Hainsworth and P. J. Wolfe, “time-frequency reassignment for music analysis’,” *Proc. International Computer Music Conference*, 2001.
- [MQ86] R. J. McAulay and T. F. Quatieri, “speech analysis/synthesis based on a sonusoidal representation’,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. Vol. ASSP 34, no. No. 4, pp. P. 749 – 751, 1986.
- [Ser89] X. Serra, “a system for sound analysis/transformation/synthesis based on a deterministic plus stochastic decomposition’,” *Ph.D. Thesis Stanford University*, 1989.
- [Ser97] —, “musical sound modeling with sinusoids plus noise’,” *Musical Signal Processing*, 1997.