

# **Dokumentation**

**zum**

**IT - Projekt**

## **Empirical Mode Decomposition**

**am Institut für  
Elektronische Musik (IEM)  
der Universität für Musik und darstellende  
Kunst Graz**

**Studienrichtung: Elektrotechnik - Toningenieur**

Verfasst von  
Christian Wallinger  
Matr.: 0230980  
Stdkz.: F750

Betreuer:  
- Dipl.-Ing. Dr. Alois Sontacchi  
- O.Univ. Prof. Dipl.-Ing. Mag. Dr. Robert Höldrich

**Inhalt:**

Vorwort.....	2
<b>1. Einführung in die Methode „Empirical Mode Decomposition“.....</b>	<b>4</b>
1.1. Literatur.....	4
1.2. Grundlagen.....	4
1.2.1. Motivation.....	4
1.2.2. EMD – Algorithmus.....	5
1.2.3. Hilbert – Huang – Spectrum.....	13
<b>2. Klanganalyse mittels EMD.....</b>	<b>15</b>
2.1. Zugrunde liegender Gedanke.....	15
2.2. Beschreibung einzelner Funktionsblöcke.....	18
2.2.1. Perzeptive Ermittlung der Cutoff – Frequenzen.....	18
2.2.1.1. Psychoakustische Grundlagen.....	20
2.2.2. Hochpassfilterung bei $f_{\text{cperzeptive}}$ .....	23
2.2.3. SSB – Modulation.....	26
2.2.3.1. Auflösungsfehler in Abhängigkeit von Samplingfrequenz des EMD Siebungsprozess.....	26
2.2.4. Blockfilterung.....	28
2.2.5. EMD – Siebungsprozess.....	29
<b>3. Ergebnisse.....</b>	<b>31</b>
3.1. Errechnete Intrinsische Mode Funktionen.....	31
3.2. Vergleich mit Gammatonfilterbank.....	40
Anhang.....	44

*Vorwort:*

In der Vergangenheit wurde für die Analyse von Daten bzw. Signalen eine Vielzahl von Methoden entwickelt. Einer großen Gruppe dieser Analysewerkzeuge liegt die Fouriertransformation zugrunde (z.B.: *Kurz – Zeit- Fouriertransformation, Wavelet – Transformation, usw...*). Allerdings sind für eine Fourier basierende Zerlegung an das zu untersuchende Signal zwei grundsätzliche Anforderungen gestellt: Das Signal muss über den betrachteten Zeitausschnitt *stationär* und *linear* sein.

Bei der Untersuchung reeller physikalischer Signale kann man diese beiden Eigenschaften im Allgemeinen nicht voraussetzen. Dadurch ergibt sich eine Einschränkung für die Detektierung von Frequenzanteilen, die nur über eine kurze Zeitspanne im Signal vorhanden sind. Diesem Nachteil Rechnung tragend, wurde von *N. E. Huang at al.* eine neue Analyseverfahren, genannt *Empirical Mode Decomposition (EMD)*, entwickelt.

Um ein Signal zu einem beliebigen Zeitpunkt in seine momentan vorhandenen Frequenzkomponenten zerlegen zu können, wird das Signal in sogenannte *Intrinsische Mode Funktionen (IMF)* zerlegt. Dies sind im Wesentlichen amplituden- und frequenzmodulierte Signale. Ihre Detektierung erfolgt über die Ermittlung des Mittelwerts, welcher aus den durch die lokalen Maxima und Minima interpolierten, das Signal einhüllende Kurvenverläufe gebildet wird.

Dabei ergibt sich aus Sicht der Frequenzanteile eine hierarchische Reihenfolge. D.h.: Jene IMF's welche anfänglich gefunden wurden, enthalten auch die höheren Frequenzkomponenten. Beendet wird dieser „*Siebungsprozess*“ des Signals, wenn eine IMF gefunden wird, welche eine monotone Funktion darstellt („Residuum“).

Aufgrund der Tatsache, dass die Zerlegung des Signals in seine IMF – Komponenten von den Eigenschaften des Signals selbst abhängig ist, kann keine *a priori* Analyse (vgl. im Gegensatz Fourieranalyse) erfolgen. Das Analyseverfahren adaptiert sich an das momentan zu untersuchende Signal. Aus den gefundenen IMFs wird in weiterer Folge über die Phase das *Hilbert – Huang - Spektrum (HHS)* errechnet. Dies repräsentiert das Signal in seinem Energie – Frequenzverlauf über die Zeit. Der Vorteil gegenüber den vergleichbaren Spektralanalysen liegt darin, dass keine „*Verwaschung*“ der Energie auf benachbarte Frequenzgebiete erfolgt. (*Leakage Effekt*)

Ziel dieses Projektes ist es, diese Methode soweit zu adaptieren, dass sie für die Analyse und Synthese von Klängen einsetzbar ist. Hierzu ist es notwendig, die Methode bezüglich der Wahl der Interpolationsmethoden für die Bildung der Einhüllenden, den Einfluss der Sampling – Frequenz sowie den Einfluss des Hilberttransformators für die Ermittlung des HHS zu untersuchen.

Weiterer Verbesserungsbedarf liegt in der Ermittlung der Randeffekte, welche durch die Interpolationsmethode am Beginn und am Ende des Signals entstehen.

Der Ablauf dieses Projektes gliederte sich grundsätzlich in drei Schritte:

- Literaturrecherche und –studie
- Einarbeitung in die bereits vorhandenen Matlab – Skripten, verfasst von Dr. Flandrin und Dr. Höldrich
- Adaptierung der Methode Empirical Mode Decomposition für die Analyse und Synthese von Klängen (insbesondere Gammelan – Instrumente)

# 1. Einführung in die Methode „Empirical Mode Decomposition (EMD)“

## 1.1. Literatur

Die EMD – Methode wurde von *Norden E. Huang at al.* Mitte der 1990er Jahre entwickelt und in der Arbeit „*The empirical mode decomposition and the Hilbert – Spectrum for nonlinear and non – stationary time series analysis*“ [1] veröffentlicht.

Ein in weiterer Folge, von *Patrick Flandrin at al.* verfasstes Manuskript „*On Empirical Mode Decomposition and its Algorithms*“ [2] wurde ebenfalls für diese Projektarbeit herangezogen.

## 1.2. Grundlagen [1]

### 1.2.1. Motivation

Daten- / Signalanalyse ist ein sehr wesentlicher und wichtiger Bestandteil (sozusagen die Basis) zur Bestimmung und Verifizierung von Modellparametern.

Hierzu ist die Transformation in den Frequenzbereich eine grundlegende Operation.

Historisch betrachtet, wurde erstmalig durch die *Fouriertransformation* eine generelle Spektralanalyse für die Bestimmung der Energie – Frequenzverteilung von Signalen ermöglicht.

Dieser Methode liegen allerdings zwei Forderungen zu Grunde:

*Die zu analysierenden Daten/Signal/System muss*

1) - **linear** und

2) –**stationär** sein,

ansonsten macht das ermittelte Spektrum wenig physikalischen Sinn.

D.h.: Die mittels Fourier detektierten Frequenzkomponenten müssen über den gesamten für die Transformation verwendeten Zeitraum im Signal enthalten sein.

( → Signalsynthesierung durch Fourierreihenoeffizienten.

Analyse:

$$X(f) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-j2\pi ft} dt$$

Synthese:

$$x(t) = \int_{-\infty}^{+\infty} X(f) \cdot e^{j2\pi ft} df \quad )$$

Grundsätzlich kann man aber davon ausgehen, dass die **zu untersuchenden physikalischen Signale weder stationär noch linear** sind!

Aus diesem Grunde wurden die bereits im Vorwort erwähnten Methoden, *Kurzzeit – Fourier – Transformation (STFT)*, *Wavelet – Transformation* usw., entwickelt.

Allerdings haben diese Methoden im betrachteten Analysefenster Forderungen nach Stationarität und Linearität inne, auch wenn diese von Methode zu Methode unterschiedlich sind.

Aus dieser Motivation heraus, ein Signal in seine tatsächlich vorhandenen physikalischen Frequenzkomponenten zu zerlegen, wurde von N. E. Huang *et al.* eine neue Methode, genannt *Empirical Mode Decomposition (EMD)*, entwickelt.

### 1.2.2. EMD - Algorithmus

Ausgangspunkt der Überlegungen ist der Wunsch, die momentan im Signal vorhandenen Frequenzkomponenten zu detektieren.

Dies kann im Allgemeinen durch die Definition der Frequenz über die Phase erfolgen:

mit  $x(t) = \sin(\theta(t))$  ( $\theta(t) = \omega t + \varphi$ ) ergibt sich:

$$\omega = \frac{d\theta(t)}{dt}$$

mit:  $\omega = 2\pi \cdot f$

f...Frequenz

$\theta(t)$ ...Phase

Für die Berechnung der Phase  $\theta(t)$  muss zunächst das reelle Eingangssignal  $x(t)$  mit Hilfe der Hilberttransformation in ein analytisches Signal  $z(t)$  transformiert werden.

$$y(t) = x(t) * \frac{1}{\pi \cdot t}$$

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(t')}{t - t'} dt'$$

Sodass man für

$$z(t) = x(t) + jy(t) \quad \text{bzw.} \quad y(t) = a(t)e^{j\theta(t)} \quad \text{erhält.}$$

Daraus lässt sich nun die Phase  $\theta(t)$  errechnen zu:

$$\theta(t) = \arctan\left(\frac{y(t)}{x(t)}\right)$$

Sowie der Betrag  $a(t)$ :

$$a(t) = (x(t)^2 + y(t)^2)^{1/2}$$

Nun stellt sich allerdings die Frage über die Bedeutsamkeit dieser ermittelten Momentanfrequenz.

Betrachtet man hierzu ein einfaches Sinussignal  $x(t) = \sin t$ , so erhält man durch Hilberttransformation  $y(t) = \cos t$  und als Lissajous - Figur einen Einheitskreis.

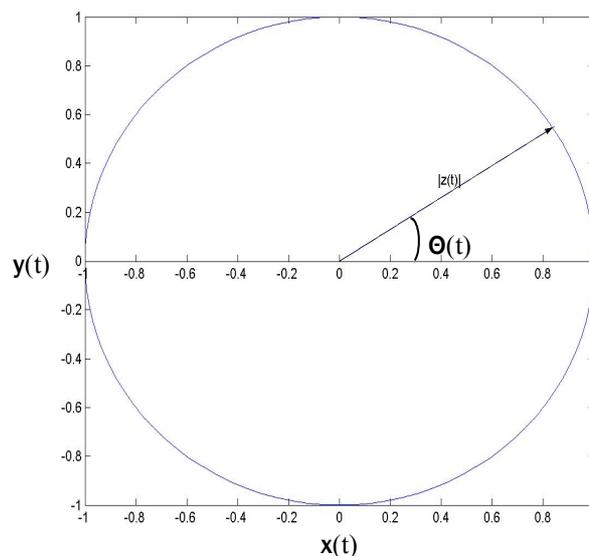


Abb.1: Lissajous - Figur

Der dazugehörige Phasengang ist ein Gerade (Abb.4 blauer Verlauf) und die Momentanfrequenz über die Zeit betrachtet ist konstant, wie zu erwarten war (Abb.2 erste-Schwingung und Abb.3 blaue Kennlinie).

Durch Hinzufügen eines Gleichanteiles  $\alpha$  zu  $x(t) = \alpha + \sin t$  ergibt sich in der x-y Ebene ebenfalls wieder ein Einheitskreis, allerdings wird der Mittelpunkt in Abhängigkeit des Gleichanteiles  $\alpha$  verschoben.

Dabei stellt sich heraus, dass mit größer werdendem  $\alpha$  der Mittelpunkt immer weiter aus dem Einheitskreis verschoben wird und die Momentanfrequenz über die Zeit betrachtet,

auch negative Wertigkeiten annimmt, die aber physikalisch nicht sinnvoll interpretierbar sind (siehe Abb.2 zweite und dritte Schwingung, sowie Abb.3 roter und grüner f – Verlauf).

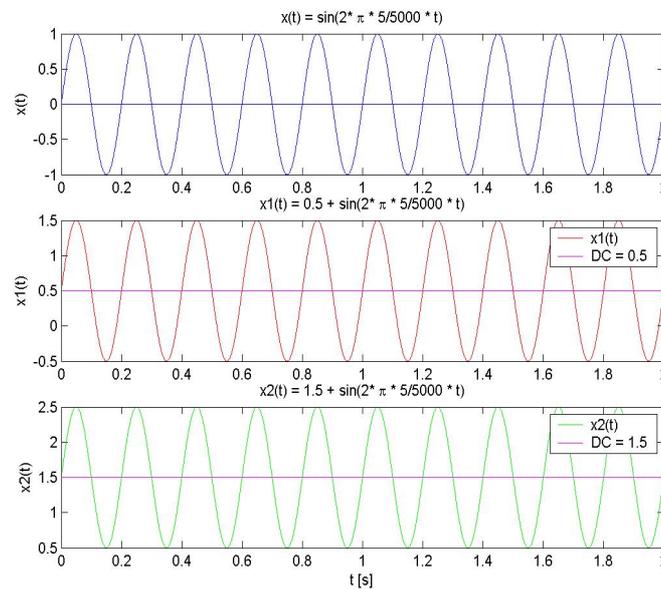


Abb.2: Signalverläufe mit unterschiedlichen DC – Anteilen

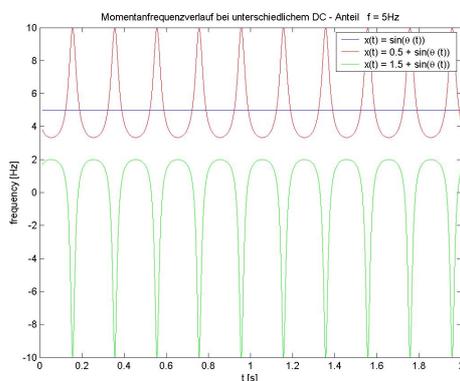


Abb.3: zugehörige f – Verläufe über die Zeit

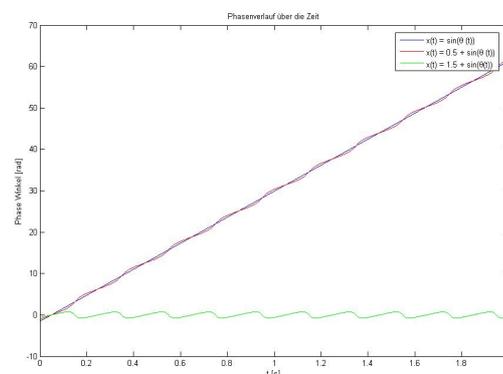


Abb.4: zugehöriger Phasenverlauf über die Zeit

Aus dieser Beobachtung heraus lässt sich nun die grundlegende Forderung nach einer Mittelwertfreiheit aufstellen, um bedeutsame Momentanfrequenzen zu erhalten.

Dies bedeutet, dass das Signal symmetrisch im Bezug auf die Abszisse sein muss, woraus sich im praktischen Sinne folgert, dass die Anzahl der Maxima und die Anzahl der Nullstellen gleich groß sein muss.

Für all jene Signale, die diese lokale Eigenschaft erfüllen, wird eine eigene Klasse von Funktionen definiert, welche als **Intrinsic Mode Function (IMF)** bezeichnet werden.

Zusammenfassend **müssen IMFs folgende zwei Bedingungen** erfüllen:

1) Die Anzahl der Extremwerte **und** die Anzahl der Nullstellen muss gleich groß sein, bzw. maximal um eins differieren.

2) Die IMF muss in jedem Punkt mittelwertfrei sein.

Für die Bestimmung dieser Klasse von Funktionen (IMF) müssen nun folgende Schritte durchgeführt werden ( $\rightarrow$  sifting process):

1. Lokalisierung aller Extremwerte ( Maxima und Minima) des Eingangssignals  $x(t)$
2. Interpolation aller Maxima respektive Minima. Daraus erhält man zwei das Eingangssignal einhüllende Kurvenverläufe ( $e_{\max}(t)$  und  $e_{\min}(t)$ )

$$3. \text{ Bildung des Mittelwerts } m(t) = \frac{e_{\min}(t) + e_{\max}(t)}{2} \quad (1)$$

4. Subtraktion des lokalen Mittelwerts (aus 3) vom ursprünglichen Signal:

$$h_1(t) = x(t) - m(t) \quad (2)$$

um lokale Mittelwertfreiheit zu erlangen.

5. Überprüfung, ob  $h_1(t)$  die beiden genannten Bedingungen einer IMF erfüllt.

**Werden die Kriterien nicht erfüllt, so müssen die Schritte 1 – 4 wiederholt werden.**

**Stellt sich allerdings heraus, dass  $h_1(t)$  eine IMF ist ( $\rightarrow h_1(t) = c_1$ ), so wird diese vom Originalsignal  $x(t)$  subtrahiert und man erhält ein Residuum  $r_1$ , welches nun als neues „Eingangssignal“ verwendet und sämtliche Schritte 1-5 wiederholt werden.**

(3)

$r_1 = x - c_1$ $r_1 - c_2 = r_3 \dots r_{n-1} - c_{n-1} = r_n$
---

Diese eben beschriebene (rekursive) Siebung des Signals (Zerlegung in IMF – Komponenten) erfolgt nun solange, bis man ein Residuum  $r_n$  erhält, welches eine monotone Funktion ist und somit daraus keine weitere Komponente ermitteln werden kann.

Durch aufsummieren aller IMF – Komponenten erhält man nun:

$$x(t) = \sum_{i=1}^n c_i + r_n \quad (4)$$

Illustration des Siebungsprozesses anhand eines Beispiels:

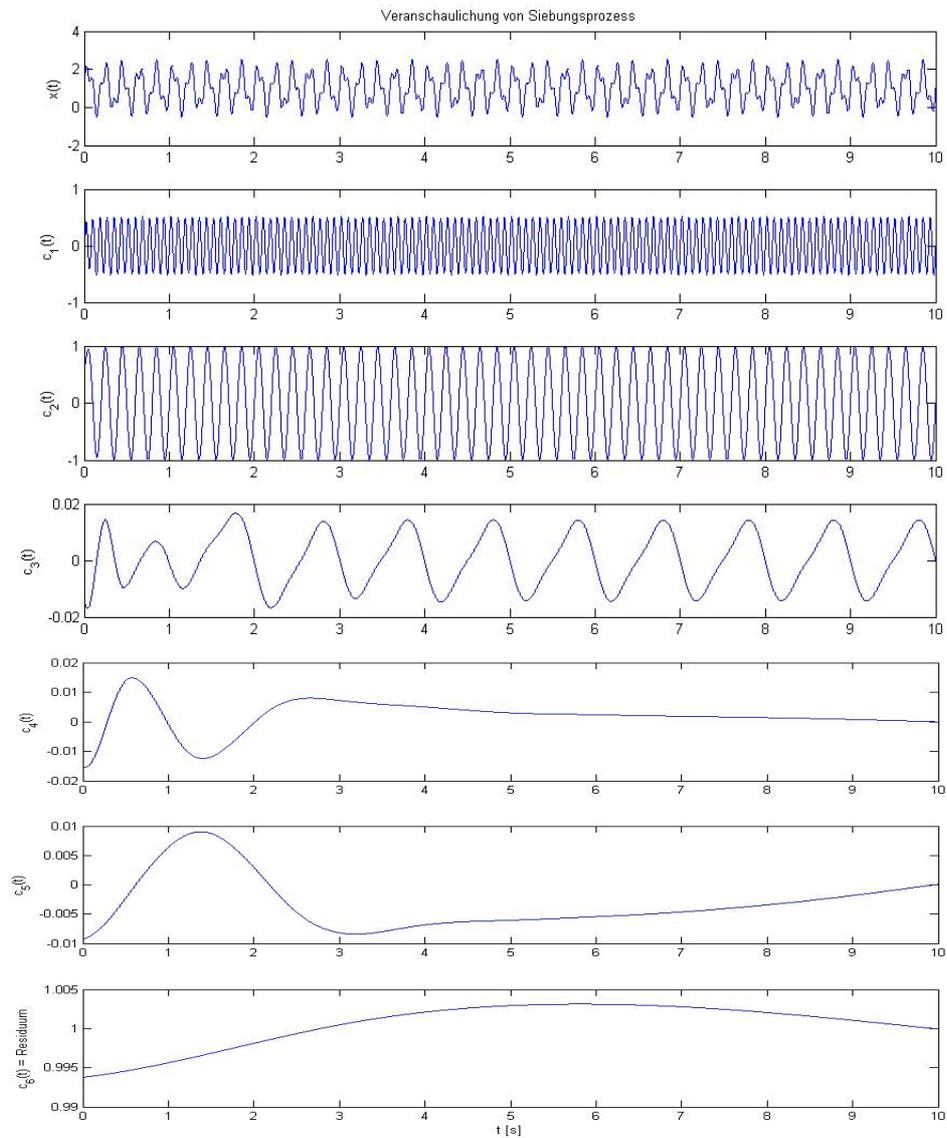


Abb.5: Zerlegung der Funktion  $x(t)$  in Intrinsic Mode Funktionen

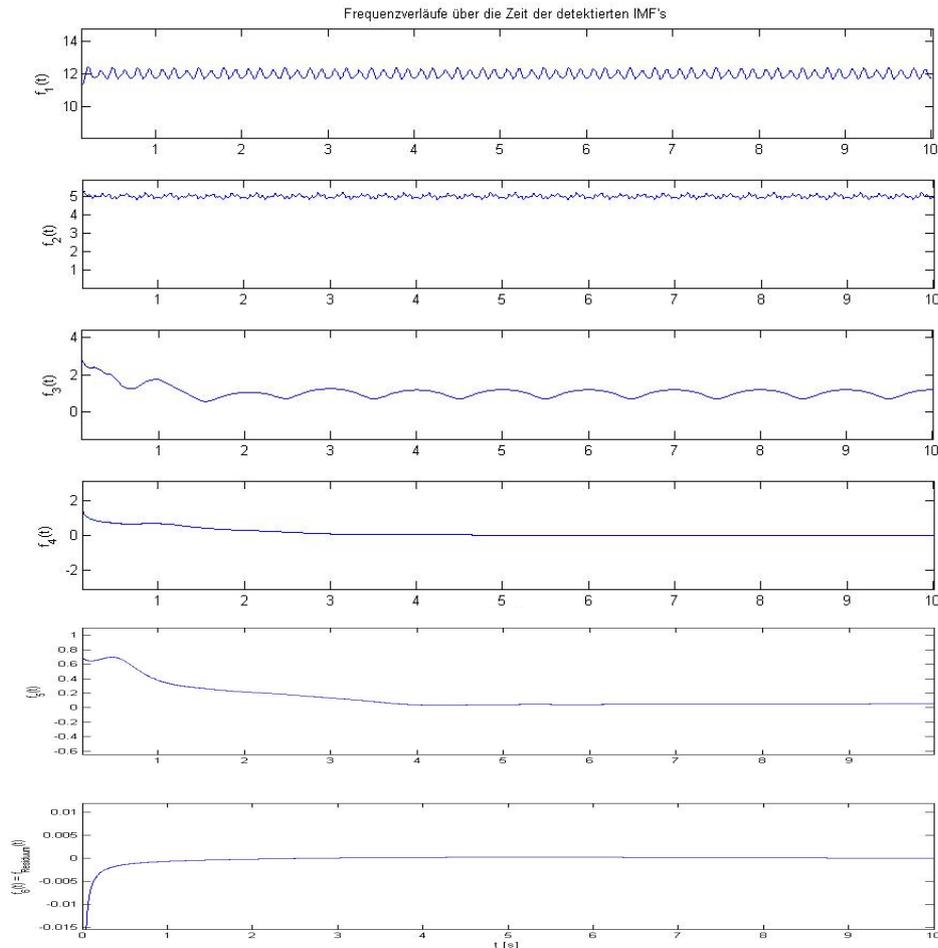


Abb.6: zugehörige Frequenzverläufe zu den detektierten IMF's

Wie in Abb.5 sehr deutlich gezeigt wird, wurde der in  $x(t)$  vorhandene Mittelwert bei der ersten gefunden IMF ( $c_1(t)$ ) bereits größtenteils herausgesiebt ( $\rightarrow$  es ist noch eine kleine Restwelligkeit vorhanden, allerdings  $\ll 1$ ).

Die in Abb.6 dargestellten Frequenzverläufe der gefunden IMFs zeigen eine eindeutig hierarchische Struktur. (Diese Eigenschaft wird in weiterer Folge in die Grundüberlegungen für den Entwurf eines Algorithmus zur Adaptierung dieser Methode zur Klanganalyse mit einfließen).

Dieses Beispiel zeigt sehr deutlich, dass es sich bei den Intrinsic Moden um ***amplituden- und/oder frequenzmodulierte Signale*** handelt.

Wie bereits in Schritt 5 beschrieben, muss überprüft werden, ob das neue mittelwertfreie Signal  $h(t)$  die Eigenschaften einer IMF besitzt oder nicht.

Allerdings muss an dieser Stelle die Frage nach einem Abbruchkriterium aufgeworfen werden. Denn es kann durchaus vorkommen, dass, obwohl die Anzahl der Extrema und die Anzahl der Nullstellen gleich groß ist, eine weitere Zerlegung durchgeführt wird. Dies führt zu einer Über – dekomposition und somit zum Verlust von Information. Das Ergebnis wäre ein rein frequenzmoduliertes Signal mit konstanter Amplitude.

Durch Begrenzung der Standardabweichung (SD) zweier aufeinander folgender Ergebnisse des sifting process wird ein solcher Informationsverlust vermieden.

$$\rightarrow SD = \sum_{t=0}^T \left[ \frac{|(h_{1(k-1)}(t) - h_{1k}(t))^2|}{h_{1(k-1)}^2(t)} \right] \quad (5)$$

$h_{1k}(t)$  entspricht dem Ergebnis der  $k$ -ten Iteration von Schritt 4

Typische Werte für SD liegen zwischen 0.2 und 0.3.

Eine weitere Möglichkeit eines Abbruchkriteriums wäre, die Anzahl der Iterationen festzulegen. Dabei hat sich gezeigt, dass 4 – 10 Iterationen zu einem brauchbaren Ergebnis führen [2].

Orthogonalität und Vollständigkeit der Siebung:

Betrachtet man diesen Punkt aus dem praktischen Blickwinkel, so ist die Vollständigkeit bzw. Orthogonalität des Siebungsprozesses mit Gleichung 4 eindeutig beschrieben:

→

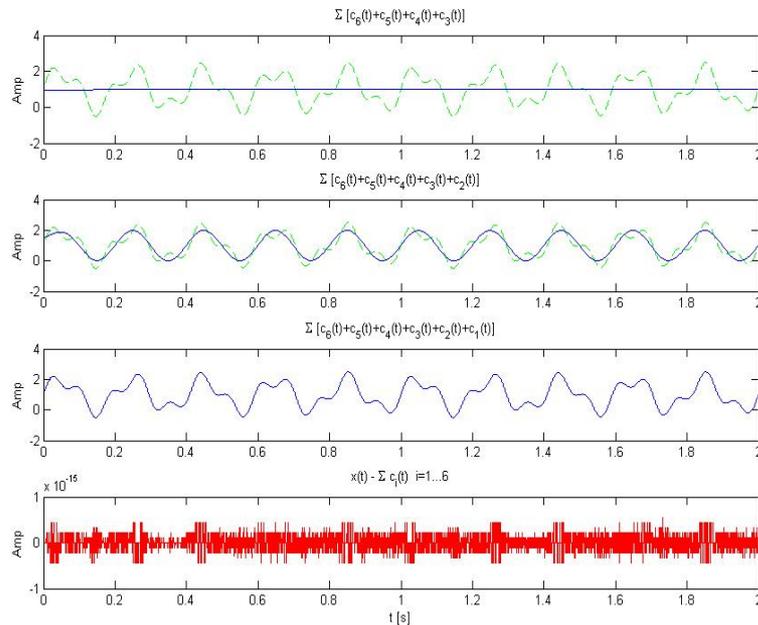


Abb.7: Rekonstruktion des Originalsignals aus den gefundenen IMFs

Abb.7 zeigt anhand des vorhergehenden Beispiels, dass das Originalsignal  $x(t)$  vollständig durch Anwendung von Gleichung 4 wieder hergestellt werden kann. Man beachte hierzu das unterste Diagramm in Abb.7 (Fehlerdarstellung aus  $x(t) - \sum c_i(t)$  mit  $i=1 \dots \#IMF + Resid.$ ). Vor allem die Dimension mit  $10^{-15}$  lässt eindeutig erkennen, dass es sich hier nur mehr um numerische Auflösungsfehler des Rechners handelt.

Um nun mathematische Orthogonalität der generierten IMFs zu finden (*a posteriori*) wird ein Orthogonalitätsindex **IO** eingeführt.

Diesen erhält man indem zunächst Gleichung 4 umgeändert wird in:

$$X(t) = \sum_{j=1}^{n+1} c_j(t) \quad (\text{dabei ist das Residuum bereits inkludiert}) \quad (6)$$

In weiterer Folge wird Gl. 6 quadriert, woraus sich

$$x^2(t) = \sum_{j=1}^{n+1} c_j^2(t) + 2 \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} c_j(t)c_k(t) \quad (7)$$

ableitet.

Wenn nun der Siebungsprozess orthogonal ist, bedeutet dies, dass der Kreuzkorrelationsterm

$2 \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} c_j(t) c_k(t) = 0$  ist, woraus sich nun der Index IO ergibt zu:

$$IO = \sum_{t=0}^T \left( \sum_{j=1}^{n+1} \sum_{k=1}^{n+1} c_j(t) c_k(t) / x^2(t) \right) \quad (8)$$

Für das gezeigte Beispiel erhält man einen, nach dieser Formel berechneten IO Index von  $IO = 0,005286$ .

Allerdings muss hier darauf geachtet werden, dass diese Definition, obwohl es anders scheint, keine globale Gültigkeit besitzt, sondern nur lokale Bedeutung erlangt, weil die Berechnung über den gesamten Zeitrahmen  $T$  erfolgt.

### 1.2.3. Hilbert – Huang – Spectrum

Wie bereits erwähnt und gezeigt, wird, um den Frequenz – Zeitverlauf zu erhalten, die Ableitung der Phase  $\theta(t)$  nach der Zeit  $t$  benötigt (siehe Seite 5).

Um aber aus einem reellen Signal eine Phase errechnen zu könne, wird das zugehörige imaginäre Signal benötigt.

Dies erhält man durch Hilberttransformation des reellen Signals.

Faltung im Zeitbereich des Signals  $x(t)$  mit Impulsantwort von Hilberttransformator:

$$y(t) = x(t) * \frac{1}{\pi \cdot t}$$

$$\rightarrow y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(t')}{t - t'} dt'$$

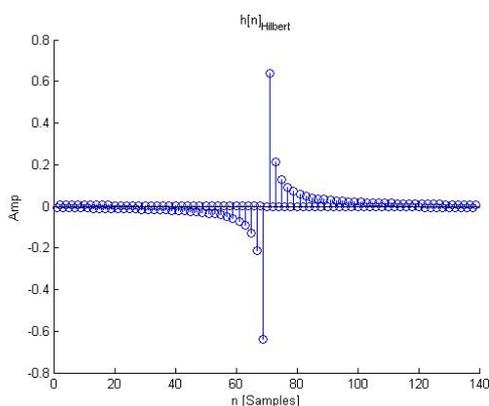


Abb.8a: Impulsantwort von Hilbert – Transformator

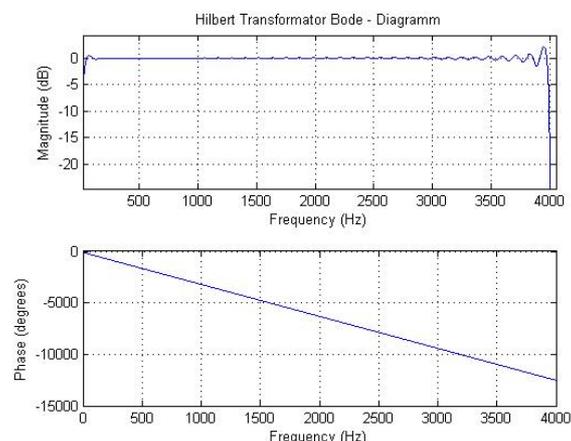


Abb.8b: Amplituden- und Frequenzgang von H.-Tr. bei Samplingfrequenz 8kHz

In weiterer Folge wird aus  $z(t) = x(t) + jy(t)$  die Phase  $\theta(t)$  errechnet und man erhält den Frequenzverlauf  $f(t)$  vom zugehörigen Signal  $x(t)$  als:

$$f(t) = \frac{1}{2 \cdot \pi} \cdot \frac{\partial \theta(t)}{\partial t} \quad (9)$$

Abbildung 6 zeigt hierzu ein Beispiel für den Frequenz – Zeitverlauf.

Wie bereits im Vorwort darauf hingewiesen wurde, liegt der generelle Vorteil dieser Methode (EMD mit Hilbert – Huang – Spektrum) im Vergleich zu anderen Spektralanalysen darin, dass keine „Verwaschung“ der Energie auf benachbarte Frequenzkomponenten erfolgt.

Als Beispiel sei hier das Spectrogramm von vorherigem Signal angeführt:

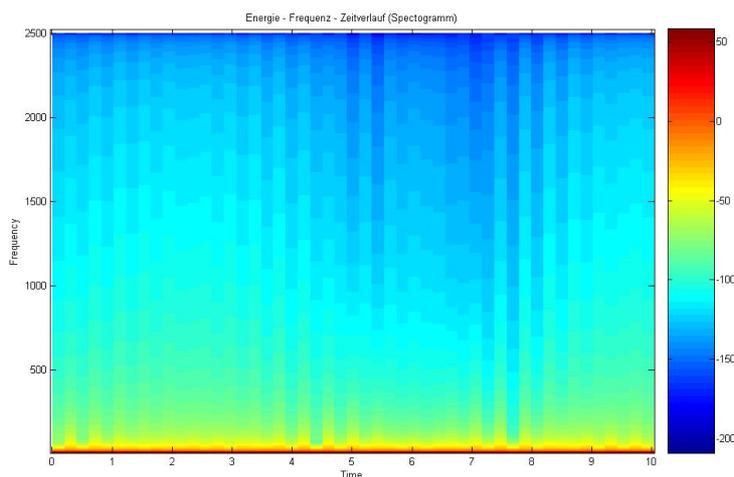


Abb.9a: Spectrogram

In Abb.9 ist der Energie – Zeit – Frequenz - Verlauf dargestellt. Die Energie ist farblich kodiert.

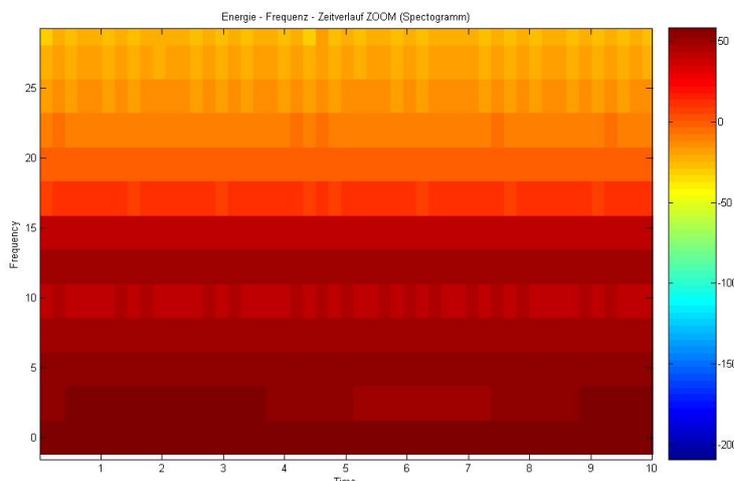


Abb.9b: Spectrogram – Zoom

Abb.9b zeigt eine gezoomte Version des Spectrograms. Sehr deutlich ist ersichtlich, dass keine eindeutige Detektion der  $f = 12\text{Hz}$   $f = 5\text{Hz}$  Komponenten und des DC-Anteiles von 1 erfolgt.

Im Vergleich siehe Abb.6  
 → eindeutige Identifizierung der f - Komponenten

## 2. Klanganalyse mittels EMD

### 2.1. Zugrunde liegender Gedanke

Wie in Kapitel 1 bereits erwähnt, ergibt sich für die detektierten Intrinsic Mode Funktionen aus Sicht der signalinhärenten Frequenzen eine hierarchische Reihenfolge. D.h. Jene IMF, welche als erste errechnet wird, beinhaltet auch die im Signal höchsten vorkommenden Frequenzen. Umgekehrt beinhalten letztere die tieffrequenten Anteile bzw. das Residuum, welches eine monotone Funktion darstellt. Diese zeigt den ungefähren „Trend“ – Verlauf des Signals.

Somit wäre ein möglicher Ansatz, dass man durch gezielt vorangehende Filterung das Signal in eine bestimmte Anzahl von Bändern zerlegt um dadurch eine „a priori“ Bestimmung über den Inhalt der in den einzelnen Bändern detektierten IMFs vorzugeben. Nachdem aber ein Siebungsprozess erst beendet ist, wenn das monotone Residuum ermittelt wurde, kann natürlich keine Vorabschätzung über die Anzahl der in den einzelnen Bändern gefundenen Mode Funktionen gemacht werden.

Ein Ausweg ist, vorab festzulegen, wie viele Moden pro Frequenzband herausgesiebt werden sollen und das erhaltene Residuum zum darunter liegenden Band hinzu zurechnen um keinen Informationsverlust zu generieren.

Allerdings muss an dieser Stelle die Definition des Residuums, oder „Trend“, geändert werden.

Es wurde schon mehrmals erwähnt, dass das Residuum des Siebungsprozesses eine monotone Funktion darstellt und somit auch keine Frequenzinformation enthält.

Wird aber eine derartige Verwendung des „Trends“, wie vorhergehend beschrieben, angestrebt, so muss sich zwingender Weise keine monotone Funktion einstellen.

Viel mehr wird das Residuum des Frequenzbandes mit den tieffrequentesten Anteilen auch die tiefsten im Signal vorkommenden Frequenzkomponenten enthalten.

Der hier soeben beschriebene Ansatz wurde im nachfolgend beschriebenen Algorithmus zur Analyse von Klängen realisiert.

Es hat sich, wie später noch ausführlicher erklärt, gezeigt, dass es sinnvoller ist, nicht das Residuum zum darunter liegenden Frequenzband zu addieren sondern die erhaltene Intrinsic Mode Funktion vom Filtereingangssignal zu subtrahieren (→kein Informationsverlust!) und das so neu erhaltene „tiefpass-gefilterte“ Signal für den weiteren Verlauf zu verwenden. Das Residuum findet somit keine weitere Verwendung mehr, außer

beim tieffrequentesten Band. Hier muss es natürlich aufgehoben und als quasi letzte IMF gehandhabt werden.

Somit ergibt sich folgende prinzipielle Struktur der Analyse:

- 1** Festlegen der zu untersuchenden Bänder
- 2** Hochpassfilterung des aktuellen, lokalen Signals
- 3** Finden der 1.IMF + Residuum
- 4** Subtraktion der gefundenen IMF von Filtereingangssignal (Pkt.1)
- 5** Überprüfung ob #der IMFs bzw. Bänder erreicht,  
wenn ja → Ende  
wenn nein → Wiederholung ab Schritt 2

Abbildung 10 zeigt das Flussdiagramm des zur Klanganalyse verwendeten Algorithmus. Auf die verwendete Einseitenbandmodulation wird nachfolgend genauer eingegangen.

Am Ende des gesamten Siebungsprozesses erhält man nun sämtliche Intrinsische Mode Funktion (Residuum bereits inkludiert) welche anschließend zur weiteren Berechnung des Hilbert – Huang – Spektrums verwendet werden.

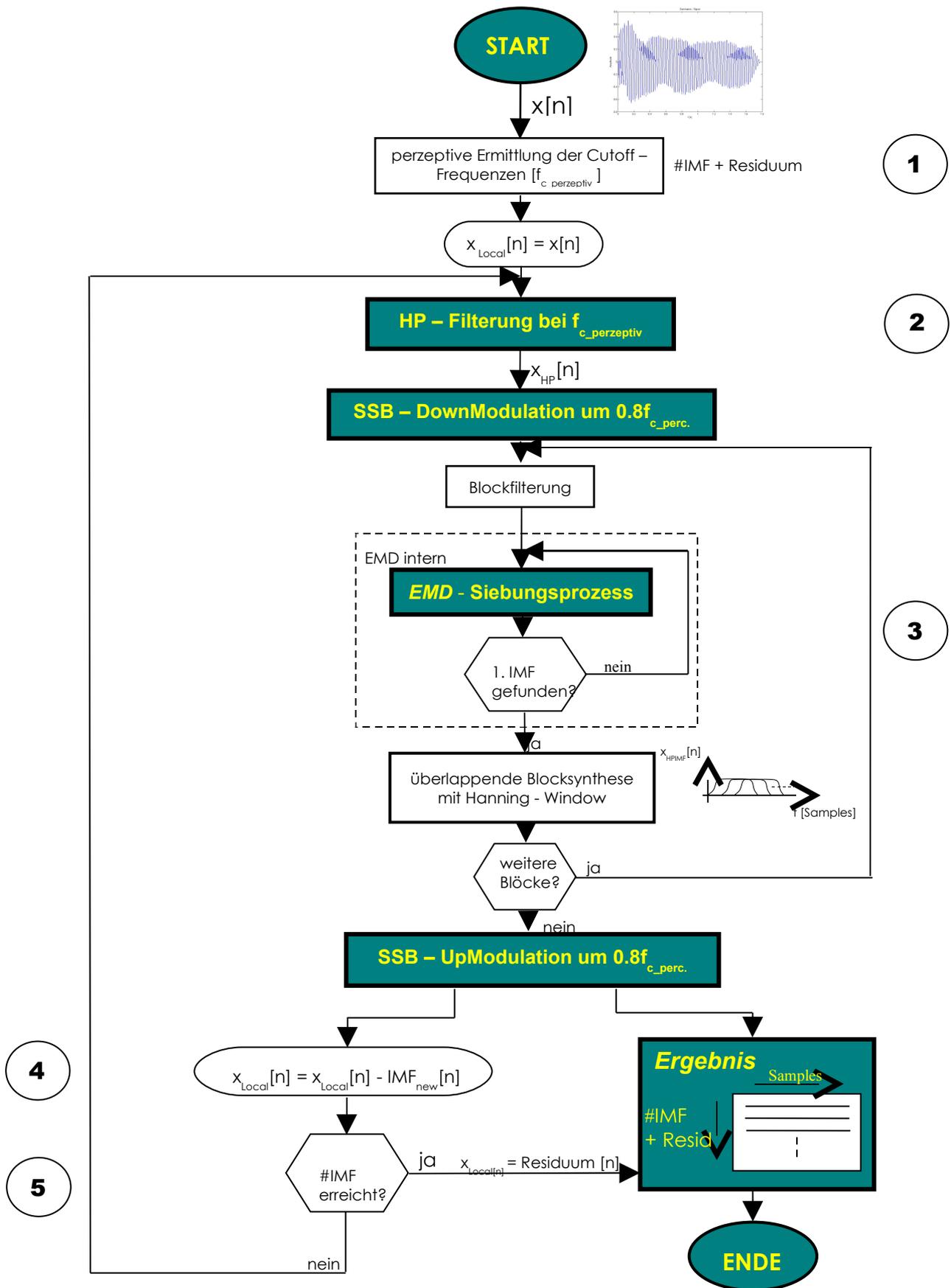


Abb.10: Flussdiagramm des Algorithmus zur Klanganalyse

## 2.2. Beschreibung einzelner Funktionsblöcke

### 2.2.1. Perzeptive Ermittlung der Cutoff – Frequenzen

Um nun eine Methode zu wählen, anhand der die für die Filterung notwendigen Grenzfrequenzen (-3dB) und somit auch die Anzahl der zu findenden Mode Funktionen ermittelt werden soll, ist es notwendig, hierzu das Ziel der Analyse festzulegen.

Es sollen mit Hilfe dieser Berechnungsvorschrift (Algorithmus) Frequenzverläufe über die Zeit ermittelt werden, die die im Signal prominent vorkommenden Spektral-

komponenten repräsentieren und in weiterer Folge als Eingangssignal von Oszillatoren

zur Resynthese der untersuchten Klänge dienen.

Bewusst werden die zu findenden IMFs als Repräsentanten von im Signalspektrum prominent vertretenen Spektralkomponenten bezeichnet, weil gerade bei dieser Art von Instrumenten die hier untersucht werden sollen (Gamelaninstrumente), kein notwendiger Weise harmonisches Spektrum vorhanden sein muss und trotzdem bestimmte Frequenzbereiche den Hauptanteil der Energie im Signal bestimmen und somit die „tonangebenden“ Komponenten sind.

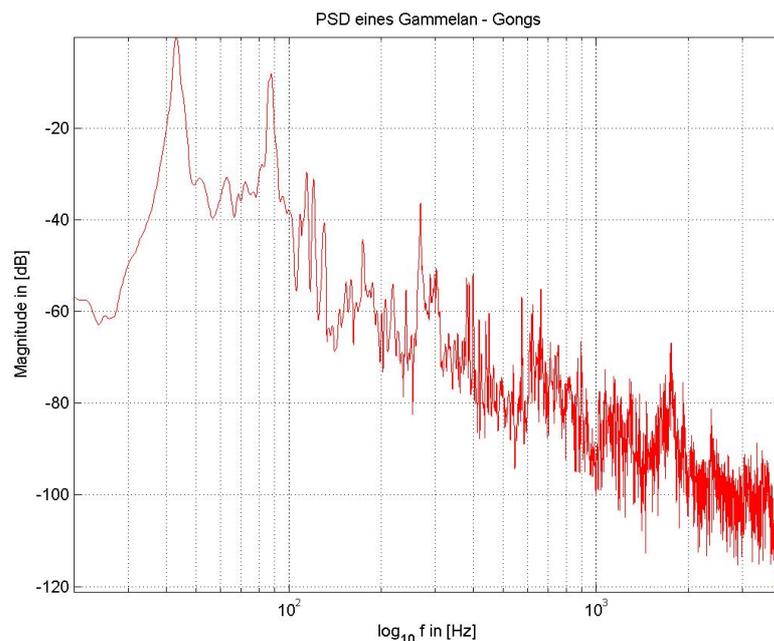


Abb.11: Leistungsdichtespektrum (PSD) eines Gamelan-Gongs

Abbildung 11 als Beispiel eines Leistungsdichtespektrums eines Gamelan – Gongs. (Hier sei darauf hingewiesen, dass diese Darstellung eine Mittelung über die Zeit darstellt!

D.h.: die in der PSD vertretenen Frequenzkomponenten müssen nicht unbedingt konstant im Spektrum über die Zeit vorhanden sein.)

Eine Möglichkeit besteht darin, dass innerhalb vorgegebener Grenzfrequenzen ( $f_{cu}$  untere- und  $f_{co}$  obere Grenzfrequenz) in Terzband- oder Oktavbandabständen die Filtergrenzfrequenzen gesetzt werden.

z.B.: Bestimmung der Filtergrenzfrequenzen in großen Terzbandabständen

(4 Halbtonschritte  $\rightarrow$  Oktave besteht aus 12 Halbtonschritten, wodurch sich ein

Faktor  $\sqrt[12]{2}$  pro Halbton ergibt) ab  $f_{co}$  bis  $f_{cu}$

mit:  $f_{cu} = 30\text{Hz}$   $f_{co} = 1\text{kHz}$  und  $k_{\Delta f} = 2^{-\frac{1}{3}}$

$\rightarrow$ ergibt sich die i-te Filtergrenzfrequenz zu

$$f_{ci} = f_{co} \cdot k_{\Delta f}^i$$

Wird die untere Grenzfrequenz  $f_{cu}$  erreicht, wird der Prozess abgebrochen und der restlich vorhandene Signalanteil als Residuum verwendet.

Bei dieser Art von Detektierung wird aber keinesfalls auf die, wie vorhin angeführt, im Signal prominent vorhandenen Spektralkomponenten eingegangen.

Nachdem pro Band eine IMF (Steuerfrequenz) gesiebt wird, muss bei der eben beschriebenen Möglichkeit mit mehr oder weniger starken Fluktuationen im ermittelten Frequenzverlauf gerechnet werden.

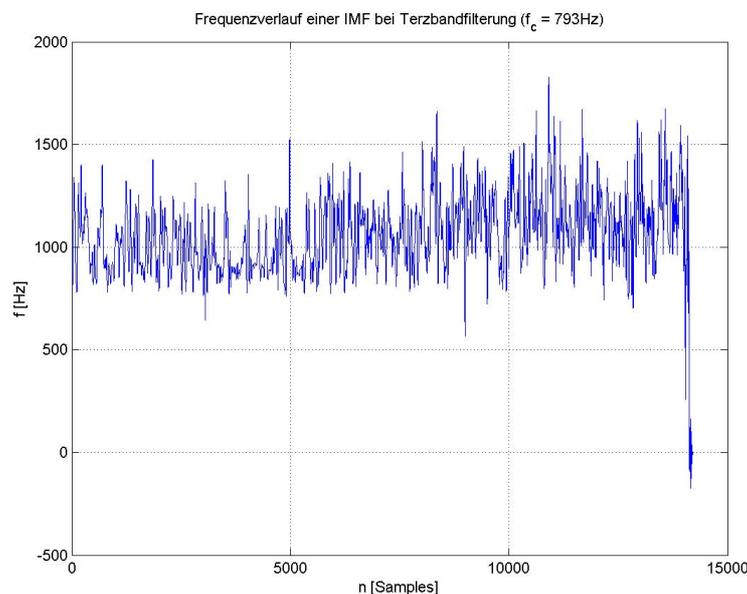


Abb.12: Frequenzverlauf einer IMF bei Terzbandfilterung

Die Bestimmung der Frequenzkomponenten, bei denen im Leistungsdichtespektrum ein energetisches Minimum auftritt und in weiterer Folge daraus die Filtergrenzfrequenzen resultieren, stellt eine weitere Option dar.

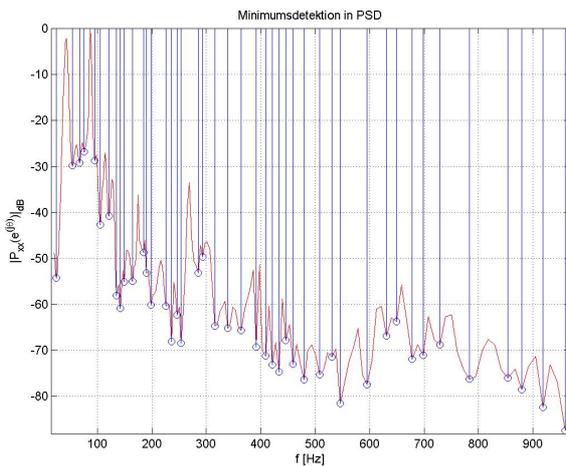


Abb.13a: Minima in PSD

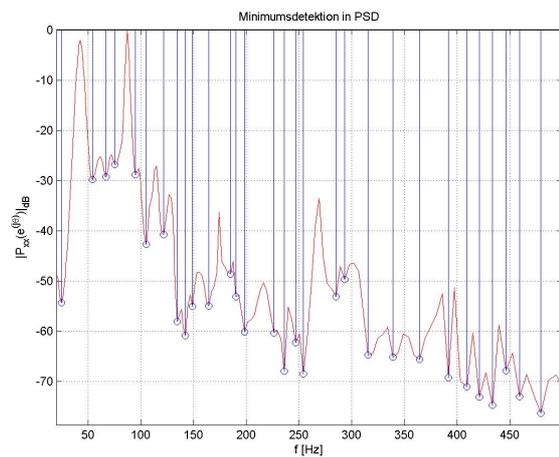


Abb.13b: Zoom zw. 10Hz und 500Hz

Allerdings stellt sich hier die Frage nach der Sinnhaftigkeit, wenn jedes detektierte Minimum als Filtergrenzfrequenz fungiert. Dieser Aspekt führt auf direktem Wege zu der Überlegung, welche der  $f$ -Komponenten, bei denen ein energetisches Maximum(!) vorherrscht, auch für die perzeptive Wahrnehmung des menschlichen Gehörs relevant sind.

### 2.2.1.1. Psychoakustische Grundlagen

In diesem Kapitel wird überblicksmäßig auf die dem Gehör innewohnende Eigenschaft der Maskierung eingegangen.

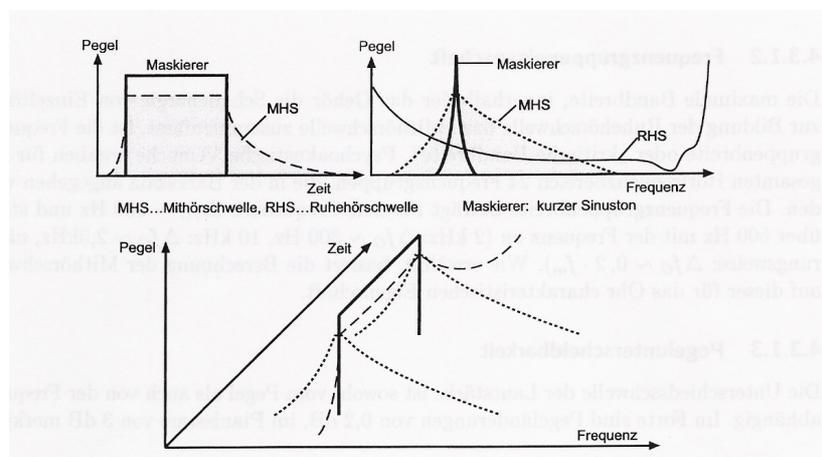


Abb.14: Verdeckungseffekt im Zeit- und Frequenzbereich [6]

Geht man von einem kurzen Sinuston aus, so werden benachbarte leisere Signalanteile sowohl im Zeitbereich als auch im Frequenzbereich verdeckt bzw. maskiert.

Im Zeitbereich spricht man hierbei von Vor- bzw. Nachverdeckung, im Frequenzbereich von Maskierung oder Simultanverdeckung.

In Abhängigkeit dieses Maskierers bildet sich eine so genannte Mithörschwelle aus. Alles, was unter dieser Schwelle liegt, kann nicht wahrgenommen werden.

Diese Eigenheit des menschlichen Sinnesorgans Ohr wird in Abbildung 14 nochmals dargestellt. Es ist deutlich zu erkennen, dass all jene Komponenten, die sich eben unter der Mithörschwelle (MHS) und über der Ruhehörschwelle (RHS) befinden, von dem tonal relevanten Signal verdeckt werden.

Allerdings ist diese Eigenschaft einerseits vom Pegel und der Frequenz des Maskierers, andererseits auch von der momentanen Verfassung des Menschen abhängig.

Diese Eigenschaft wird nun verwendet, um jene tonalen Komponenten aus dem Signalspektrum zu detektieren, die für die Analyse als Filtergrenzfrequenzen dienen. Hierzu wurde aus [4] der Teil II. EXTRACTION of TONAL COMPONENTS und Teil IIIA. EVALUATION of MASKING EFFECTS implementiert.

#### *ad. Extraction of Tonal Components:*

In diesem Abschnitt werden auf Basis des physikalischen Schalldruckes (Sound Pressure Level SPL) vom Signal all jene spektralen Komponenten eruiert, welche als Kandidaten für die im Signal prominent vorhandenen tonalen Anteile („Maskierer“) in Frage kommen.

Um den tatsächlichen Schalldruck vom zu analysierenden Signal zu erhalten, muss dieses zunächst mit einem Korrekturfaktor (zBsp: 11,99 für verwendeten Gamelangong) multipliziert werden (ergibt den Schalldruck in [Pa],) und dann auf den Bezugsschalldruck von  $20\mu\text{Pa}$  bezogen werden → Umrechnung in dB!

Als ersten Schritt werden dann die Schalldrücke ( $L_i$ ) der einzelnen Frequenzkomponenten (Ermittlung durch *N.Pkt - FFT*) mit den umliegenden benachbarten Komponenten verglichen und all jene herausgefiltert, die betragsmäßig größer sind.

$$L_{i-1} < L_i \geq L_{i+1}$$

In weiterer Folge, wenn ein Kandidat gefunden, wird getestet, ob der Schalldruck über einen bestimmten Schwellwert gegenüber den benachbarten Komponenten liegt. Ist dies der Fall, erfolgt eine endgültige Nominierung als Maskierer.

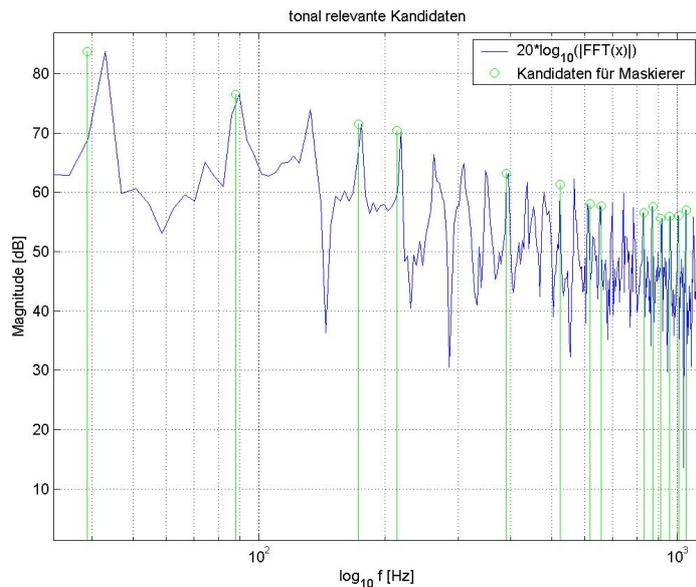
$$L_i - L_{i+j} \geq 8dB \quad j = -3, -2, +2, +3$$

Aufgrund dessen, dass die Berechnung der FFT nur eine begrenzte Frequenzauflösung

liefert (z.B.: bei  $f_s = 8\text{kHz}$  und 2048 Punkt – FFT ergibt dies ein  $(\Delta f = \frac{8000\text{Hz}}{2048\text{Pkt}} = 3.9\text{Hz})$ )

wird durch zusätzliche Interpolation zwischen den zwei unmittelbar benachbarten Frequenzkomponenten die „reale“ Frequenz des Maskierers ermittelt.

$$f_c = f_i + 0.46(\text{Hz} / \text{dB}) \cdot (L_{i+1} - L_{i-1})$$



In Abbildung 15 sind in grün all jene Kandidaten eingetragen, die als tonal relevante Komponente in Frage kommen. (Es wurden hier schon die interpolierten Frequenzwerte eingezeichnet)

Abb.15: Kandidaten für tonal relevante Komponenten

#### ad. Evaluation of Masking Effects

Die Evaluierung, welche der Kandidaten nun als tonal relevant gelten, wird durch die Differenz zwischen dem Schalldruck der individuellen Kandidaten und jenem Schalldruck, welcher aktuell die Maskierungsenergie präsentiert, errechnet.

$$LX_{\mu} = L_{\mu} - 10 \log_{10} \left[ \left( \sum_{v=1 \& v \neq \mu}^N 10^{L_{Ev}(f_v)/(20dB)} \right)^2 + I_{n\mu} + 10^{L_{TH}(f_{\mu})/(10dB)} \right] dB$$

mit  $LX_{\mu}$  als SPL – Differenz

$L_{Ev}(f_v)$  Anregungslevel

$L_{\mu}$  SPL des  $\mu$ -ten Kandidaten

$I_{n\mu}$  Rauschintensität (wurde nicht implementiert)

$L_{TH}$  Hörschwelle

Für weitere Details (Berechnung der einzelnen Schalldrücke) siehe [4].

In Abbildung 16 sind nun rot all jene Komponenten eingezeichnet, welche in weiterer Folge als Filtergrenzfrequenzen im Siebungsprozess dienen.

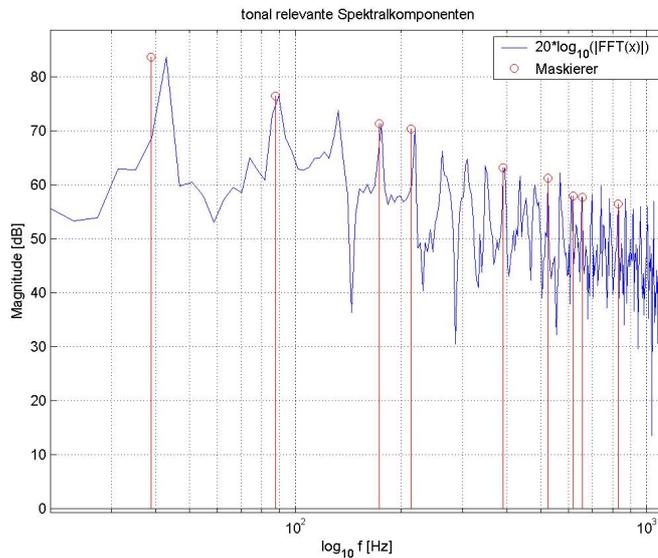


Abb.16: tonal relevante Komponenten

In diesem Beispiel liegen  $f_c$ -Komponenten bei

$f$ in [Hz]
39
88
173
213
391
524
618
657
831

Tabelle 1: Cutoff - Frequenzen

Vergleicht man die in diesem Punkt detektierten Frequenzen mit dem in Abb.11 dargestellten Leistungsdichtespektrum, so kann man durchaus eine Übereinstimmung feststellen. Betrachtet man aber zum Beispiel jenen Frequenzbereich um 250Hz (in Abb.11), so wurde diese Komponente nicht detektiert, weil genau in diesem Bereich, in dem die Schwellwertüberprüfung auf  $L_i - L_{i+j} \geq 8dB$   $j = -3, -2, +2, +3$  erfolgt, eine Aufweitung (breitere Frequenzkeule) stattfindet.

### 2.2.2. Hochpassfilterung bei $f_{c\_perzeptiv}$

Die Filterung des eingehenden Signals  $x_{Local}[n]$  mit einem Hochpassfilter erfolgt an denen im vorhergehenden Punkt eruierten Grenzfrequenzen.

Dabei sind an den Filter folgende Forderungen gestellt:

- keine Welligkeit im Durchlassbereich
- schmaler Übergangsbereich
- möglichst hohe Sperrdämpfung
- sowie eine Gruppenlaufzeit von  $\tau_{ph} = 0$

Aufgrund dessen, dass bei der Frequenz – Ortstransformation auf der Basilarmembran im Ohr ebenfalls eine kleine Zeitverzögerung zwischen den hohen- und tiefen Frequenzen erfolgt ( → Wanderwellenausbreitung mit endlicher Geschwindigkeit) ist eine Gruppenlaufzeit von exakt 0 nicht unbedingt erforderlich.

Der im Algorithmus implementierte Filter ist ein IIR – Butterworth – Filter 8.Ordnung.

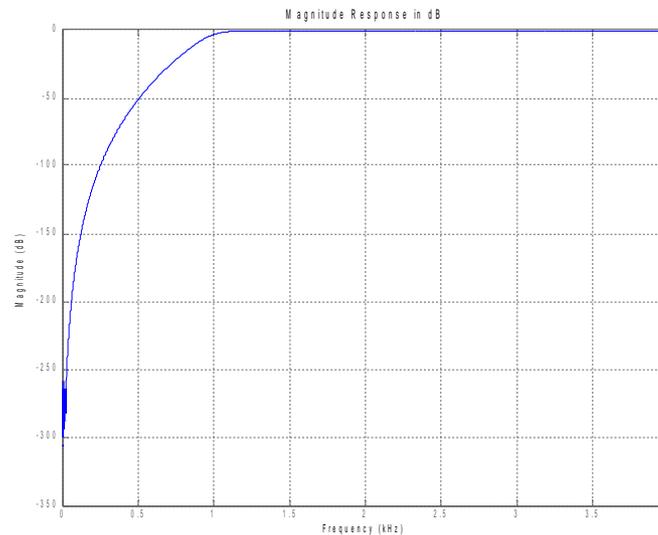


Abb.17: verwendeter HP-Filter bei  $f_c = 1\text{kHz}$

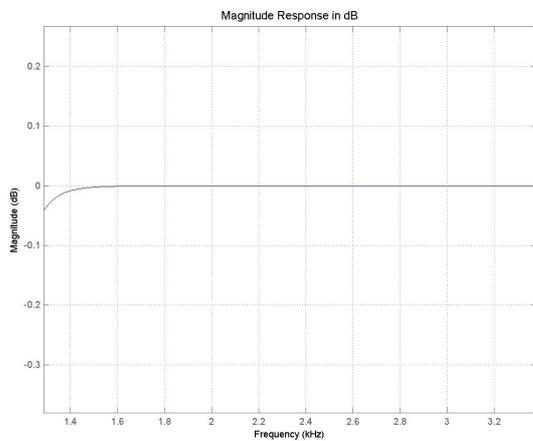


Abb.17a: Durchlassbereich von HP-Filter

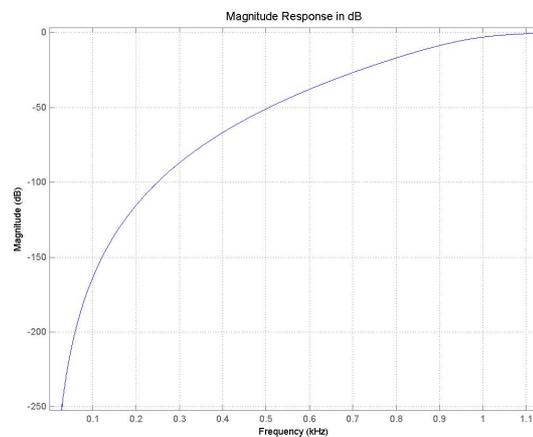


Abb.17b: Übergangsbereich von HP-Filter

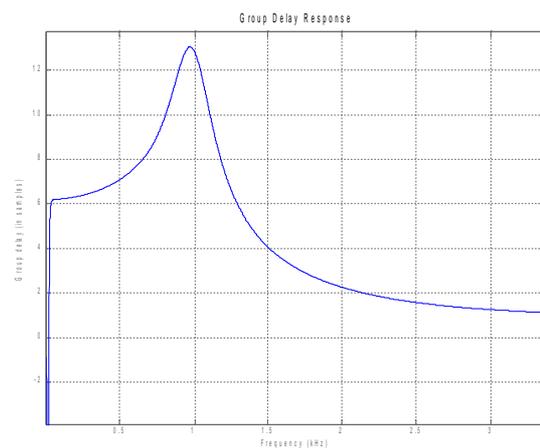
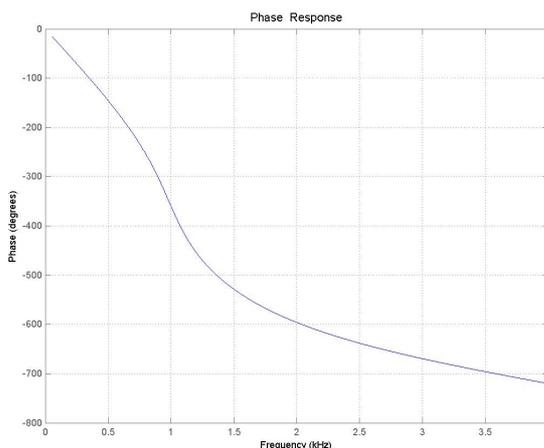


Abb.17c: Phasenverlauf von HP-Filter

Wie in Abb.17d zu sehen ist, hat der Filter bei der Grenzfrequenz eine Gruppenlaufzeit von 13 Samples. Dies entspricht bei einer Samplingfrequenz  $f_s = 8\text{kHz}$  einer Zeitverzögerung von

$$t = \text{Samples} \cdot \frac{1}{f_s} = 13 \cdot \frac{1}{8000} = 1.6\text{ms} \rightarrow \text{Dies ist ein Gruppenlaufzeit, die absolut im}$$

tolerierbaren Zeitbereich liegt.

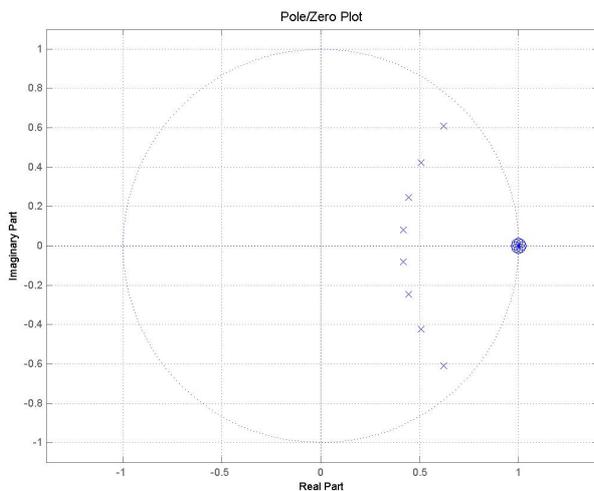


Abb.17e: PN-Plan von HP-Filter

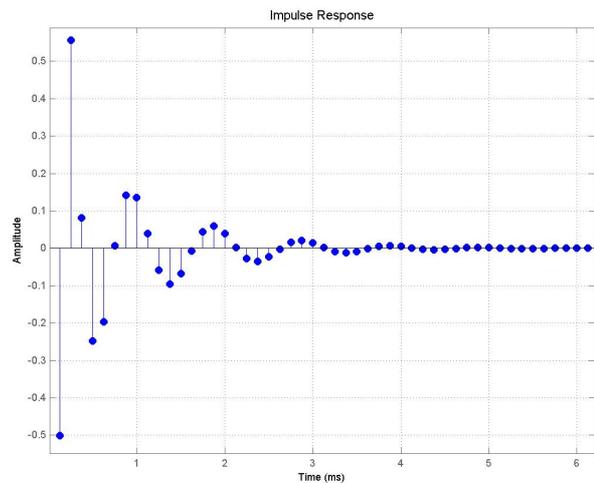


Abb.17f: Impulsantwort von HP-Filter

Um nun einen noch steileren Übergangsbereich zu bekommen, erfolgt eine 4-fache Filterung des Signals wodurch die resultierende Ordnung 32 beträgt.

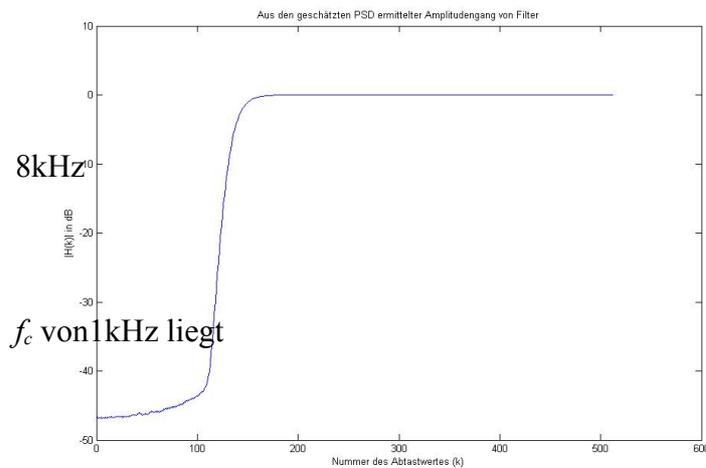


Abb.18: Amplitudenfrequenzgang von Filter mit 4-fach Filterung

Die Auflösung beträgt in

Abbildung 18 bei  $f_s =$   
und einer FFT-Länge von  
1024 Samples, pro  
Frequenzbin  $k = 7,81\text{Hz} \rightarrow$

bei  $k = 128$

Somit lässt sich der eindeutig steilere Übergangsbereich des HP-Filters erkennen.

Das aus diesem Prozess erhaltene hochpass – gefilterte Signal wird nun in weiterer Folge um den Faktor  $0.8 \cdot f_{c\_perzeptiv}$  mit Hilfe einer Einseitenband – Modulation (SingleSideBand-Modulation) Richtung DC moduliert.

### 2.2.3. SSB – Modulation

Das aus der Filterung erhaltene Signal wird, wie obig schon erwähnt, vor dem Siebungsprozess noch um den Faktor  $0.8 \cdot f_{c\_perzeptiv}$  Richtung DC verschoben.

Dies hat folgenden Grund:

Um eine Sinusschwingung eindeutig (ohne Informationsverlust) rekonstruieren zu können, muss das *Shannon* (bzw. *Nyquist*) – *Theorem* erfüllt sein:

$$f_s \geq 2 \cdot f_N \quad \text{mit } f_s \dots \text{Samplingfrequenz}$$

$$f_N \dots \text{Nyquist – Frequenz} \rightarrow \text{die höchste im Signal vorkommende Frequenzkomponente}$$

D.h: Die Abtastfrequenz muss mindestens doppelt so hoch sein, wie der höchste im Signal vorkommende Frequenzanteil.

Wendet man sich aber von dieser Bestimmung ab und betrachtet dies aus der Sicht der bei diesem Algorithmus verwendeten Generierung der Momentanfrequenzen, so werden eine minimale Anzahl an Datenpunkten der Schwingung benötigt, um daraus durch Ableitung der Phase die Frequenz zu erhalten.

Für die eindeutige Definition einer Oszillation (sinusoid) werden mindestens *vier* bzw. *fünf* Datenpunkte benötigt [1].

Diese Randbedingung führt somit zu einer Einschränkung des Basisfrequenzbereichs, der ohne Fehler detektiert wird.

#### 2.2.3.1. Auflösungsfehler in Abhängigkeit von Samplingfrequenz des EMD Siebungsprozesses

Um die Abhängigkeit des Auflösungsfehlers von der Samplingfrequenz zu bestimmen, wird die Eigenschaft der Mittelwertfreiheit ausgenutzt.

Der EMD – Algorithmus sollte, wenn man eine mittelwertfreie Sinusschwingung einspeist ( $\rightarrow$  #der Nullstellen = # der Maxima), diese sofort mit der ersten gefundenen IMF detektieren.  $\rightarrow$  Vergleicht man somit die 1. errechnete Mode Funktion mit dem

Originalsignal  $x_f[n]$ , so erhält man ein Maß für den Auflösungsfehler  $e(f)$  in Abhängigkeit von der Samplingfrequenz.

$$e(f)_{f_s} = \sqrt{\frac{\sum_n (x_f[n] - 1.IMF_f[n])^2}{\sum_n x_f[n]^2}}$$

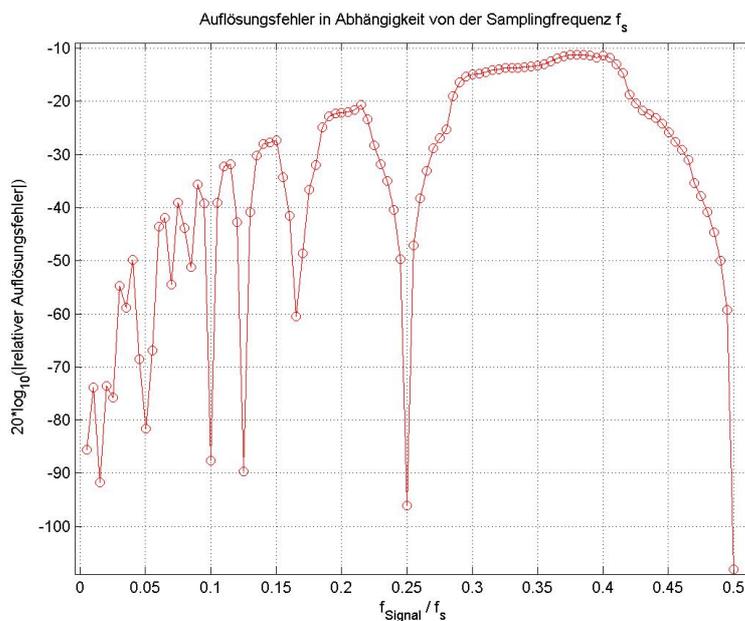


Abb.19.: Fehlermaß [dB] von verschiedenen Signalfrequenzen bei  $f_s = 8\text{kHz}$

Eindeutig lässt sich anhand der erhaltenen Kammfilterstruktur erkennen, dass der Auflösungsfehler sehr gering ist, wenn die Signalfrequenz ein ganz zahliger Teil der Samplingfrequenz ist.

Aber auch die vorhin beschriebene Mindestdatenmenge von 4 bis 5 Samples widerspiegelt sich im gemessenen Fehler → Bei  $\frac{1}{4} * f_s < \text{Signalfrequenz} < \frac{1}{2} * f_s$  tritt der größte Fehler auf.

→ Somit ist ersichtlich, dass die Auflösungsgenauigkeit erhöht werden kann, wenn man versucht, die höheren Frequenzanteile für die Siebung Richtung DC zu verschieben.

Die verwendete Einseitenbandmodulation (SSB) erfolgt nun folgendermaßen:

Bandbegrenzte Eingangssignal:  $x[n]$

→ Analytisches Signal:  $z[n] = x[n] + j * \text{imag}(\text{Hilbert}(x[n]))$

Modulation Richtun. DC:  $y[n] = \Re \left\{ z[n] \cdot e^{-j2\pi \cdot 0,8 f_c \text{_{perzeptiv}} \cdot n} \right\}$

$$y[n] = x[n] \cos(2\pi \cdot 0.8 f_{c\_perzeptiv} \cdot n) + \text{imag}(\text{hilbert}(x[n])) \sin(2\pi \cdot 0.8 f_{c\_perzeptiv} \cdot n)$$

#### 2.2.4. Blockfilterung

Bevor nun das aufbereitete Signal in den Siebungsprozess geschickt wird, erfolgt zuvor noch die Zerlegung in überlappende Blöcke (3fach), die dann einzeln gesiebt werden.

Nach dem Siebungsprozess werden diese dann wieder, gewichtet mit einem aufbereiteten *Hanning – Fenster* zusammengefügt und dementsprechend normiert.

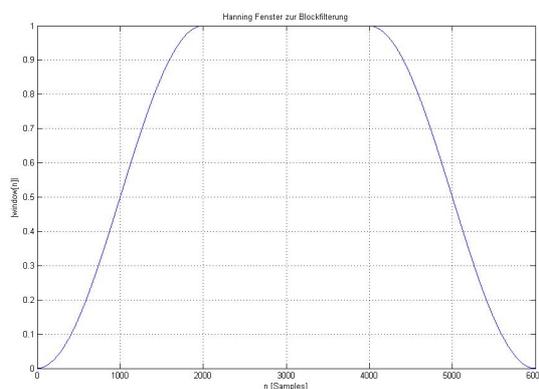


Abb.20: Aufbereitetes Hanningfenster zur Blockfilterung

Der Fehler der dabei gegenüber der nicht Blockverarbeitung gemacht wird, liegt bei ca.  $\sim -82.5\text{dB}$ . Die Berechnung erfolgte mit selbiger Formel, wie unter dem vorhergehenden Kapitel 2.2.3.1 verwendet. Dabei wurde eines ohne Blockverarbeitung und eines mit Blockverarbeitung gesiebt Sinussignal miteinander verglichen  
 →Dies ist durchaus ein Indiz dafür, dass der Algorithmus online fähig gemacht werden kann.

#### 2.2.5. EMD – Siebungsprozess

Der Algorithmus zur tatsächlichen Siebung (Findung der IMFs) wurde bereits in Kapitel 1 ausführlich beschrieben.

Als Grundlage für den EMD - ProgrammCode dienen die Matlab-Files von

Dr. Flandrin, welche in der „*Time Frequency Toolbox*“ zu Verfügung stehen [3].

Allerdings erfolgt der Abbruch der Mode Decomposition bereits nach der ersten gefundenen IMF, aufgrund der ebenfalls schon erläuterten Gründe.

Für die Berechnung einer IMF werden 20 Iterationen durchgeführt. (→ siehe Seite 11)

→ Nachdem der Siebungsprozess erfolgreich war, muss natürlich das erhaltene hochpassgefilterte, SSB- modulierte, gesiebte Signal wieder in seinen ursprünglichen Frequenzbereich demoduliert werden. Dies erfolgt auf umgekehrtem Wege zur Modulation Richtung DC (→ siehe Kapitel 2.2.3)

Wie am Beginn von Kapitel 2 schon erwähnt wurde, liegt die Grundidee für den Analyse – Algorithmus darin, dass man das hochpass - gefilterte Signal nach der Siebung vom originalen, für die Siebung aufbereiteten Signal, subtrahiert, um das neue Eingangssignal zu erhalten.

Ursprünglich bestand die Generierung des neuen Signals darin, dass schon vor der Siebung das gefilterte Signal vom Originalsignal abgezogen und das aus der EMD – Zerlegung gewonnene Residuum zu diesem wieder hinzu addiert wurde (um Informationsverlust zu vermeiden).

Durch die Detektion der Extremwerte bei der internen Mittelwertberechnung im EMD – Prozess werden aber auch all jene Komponenten die ursprünglich nicht so stark im Signal vorhanden waren, forciert (→ siehe auch [1]). Dies führt im vorliegenden Anwendungsfall dazu, dass der Frequenzbereich, der bei der Hochpassfilterung im Übergangs- und Sperrbereich gedämpft wird, trotzdem nach der Siebung energetisch stärker vorhanden ist, als durch den verwendeten Hochpassfilter vor der Siebung unterdrückt.

Somit hat das aus der Summenbildung der IMFs + Residuum resynthetisierte Signal in diesen Bereichen mehr Energie als das ursprüngliche und die unter Gleichung 4 definierte Vollständigkeit ist nicht mehr erfüllt.

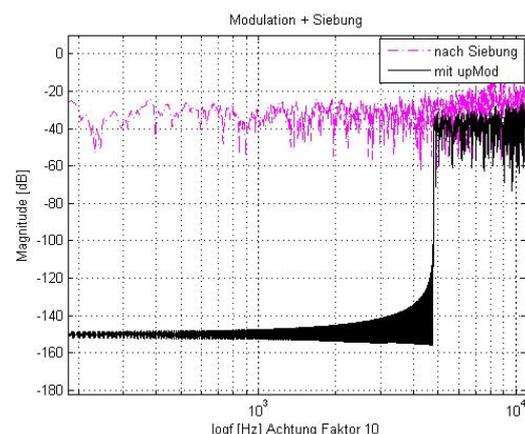
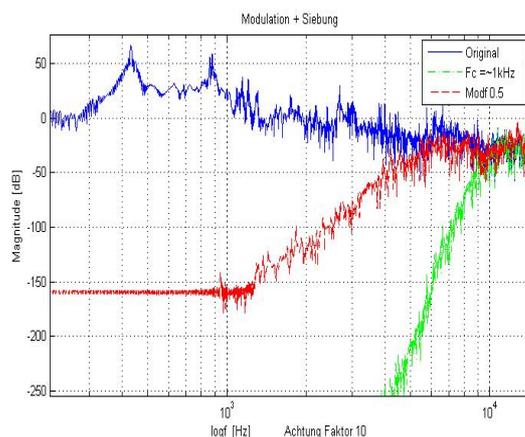


Abb.21a: Originalsignal (blau), HP- gefiltertes Signal (grün),  
Richtung DC um den Faktor  $0.5 \cdot f_{cutoff}$   
SSB-moduliertes Signal (rot)

Abb.21b: Signal nach der Siebung (violett)  
um den Faktor  $0.5 \cdot f_{cutoff}$   
SSB-upmoduliertes Signal  
(schwarz)

Abbildung 21a und 21b zeigen den soeben beschriebenen Vorgang. Im rechten Bild (21a) ist in blauer Farbe das ursprünglich, zu untersuchende Signal dargestellt. Grün zeigt das durch die Hochpassfilterung generierte Signal bei einer cutoff – Frequenz von 1kHz. Sehr gut lässt sich wiederum die Steilheit des verwendeten Filters erkennen. Danach erfolgt die Modulation Richtung niedrigerer Frequenzen um den Faktor  $0.5 \cdot f_{cutoff}$  (500Hz in diesem Bsp.). Dieses Signal wird nun gesiebt und man erkennt in linker Abbildung (21b) eindeutig die energetische Forcierung der tieffrequenten Komponenten. Das schwarz gefärbte Signal zeigt nun die endgültige, nach der Hochmodulation erhaltene IMF.

Die daraus erhaltenen Fehler zeigen nachfolgende Abbildungen:

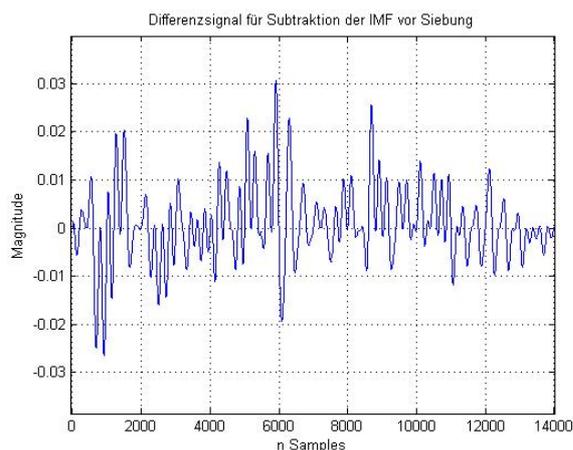


Abb.22a: Differenzsignal des aus den erhaltenen IMFs  
resynthetisierte Signal und dem Original,  
dem bei vor der Siebung  
erfolgter Subtraktion

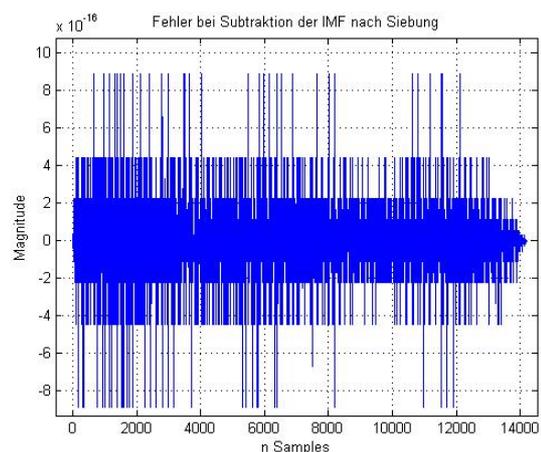


Abb.22b: Differenzsignal des aus den erhaltenen  
IMFs resynthetisierte Signal und  
Original, bei nach der Siebung  
erfolgter Subtraktion

Man beachte den Unterschied in den Größenordnungen der beiden Fehlersignale ( $10^{-2}$  zu  $10^{-16}$ ). Im Algorithmus selbst wird ein SSB – Modulationsfaktor von 0.8 verwendet. Eine Möglichkeit, die ebenfalls versucht wurde, bestand darin, dass man die Akzeptanz einer gefundenen IMF von einem Energiekriterium abhängig macht. D.h.: Es wurde die Energie im Spektralen Schwerpunkt der IMF mit der bei dieser Frequenz im Originalsignal vorhandenen Energie verglichen. Lag die Differenz in einem bestimmten Toleranzband, wurde die IMF akzeptiert, wenn nicht, wurde sie verworfen.

Dies führte allerdings zu keinen befriedigenden Ergebnissen, weil sich die Wahl des Toleranzbandes als sehr wage und kaum konstant über den gesamten Frequenzbereich herausstellte.

### 3. Ergebnisse

Sämtliche in diesem Kapitel dokumentierten Ergebnisse basieren auf dem Gamelan – Soundfile *'gong\_ageng\_1mAbstand\_0Grad\_8kHz\_kurz.wav'*. Dieses enthält einen impulsförmig angeregten Gong welcher ausklingt.

Somit lassen sich einige Erwartung an das Ergebnis stellen:

1. Aufgrund der impulsförmigen Anregung müssen auch höher - frequente Anteile detektiert werden.
2. Nachdem ein Ausklingvorgang stattfindet, ist zu erwarten, dass gegen Ende nur mehr die tieffrequenten, energiereichen Komponenten vorhanden sind und sämtliche höhere Spektralkomponenten schon frühzeitig ausklingen.
3. Allgemein muss ein Amplitudenabfall über die Zeit detektiert werden.

#### 3.1. Errechnete Intrinsic Mode Funktionen

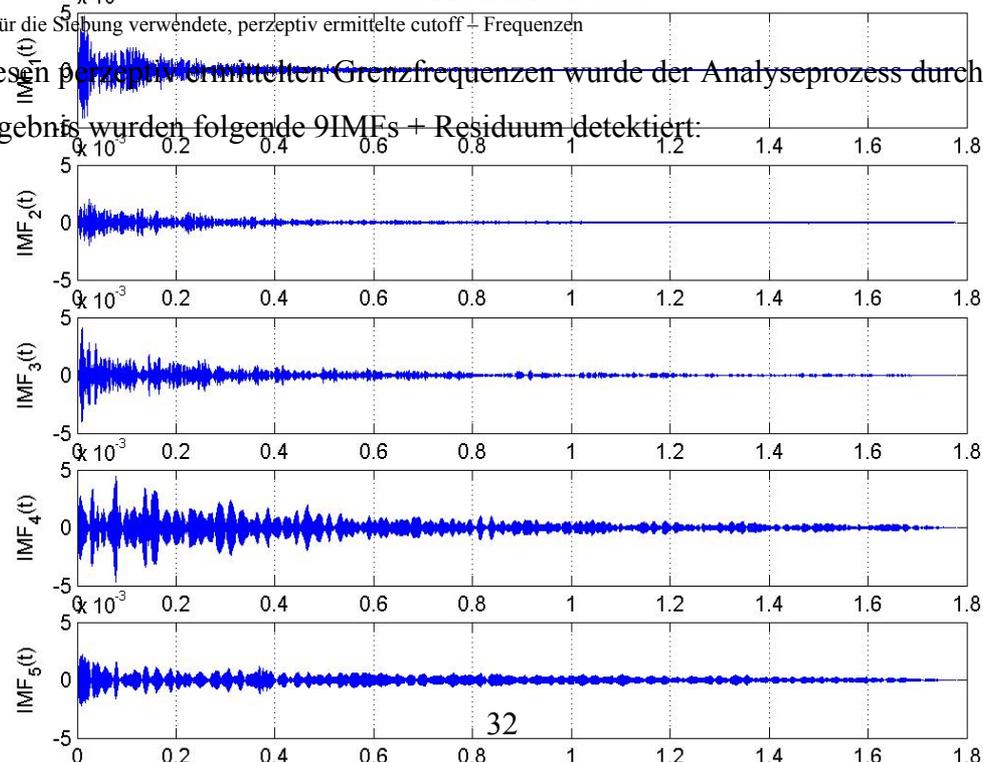
Die ermittelten cutoff – Frequenzen lauten:

Mode	IMF <sub>1</sub>	IMF <sub>2</sub>	IMF <sub>3</sub>	IMF <sub>4</sub>	IMF <sub>5</sub>	IMF <sub>6</sub>	IMF <sub>7</sub>	IMF <sub>8</sub>	IMF <sub>9</sub>
$f_c$ percept.	956	872	616	525	391	213	173	88	39

Tabelle 2: Für die Steigung verwendete, perzeptiv ermittelte cutoff – Frequenzen

Mit diesen perzeptiv ermittelten Grenzfrequenzen wurde der Analyseprozess durchgeführt.

Als Ergebnis wurden folgende 9IMFs + Residuum detektiert:



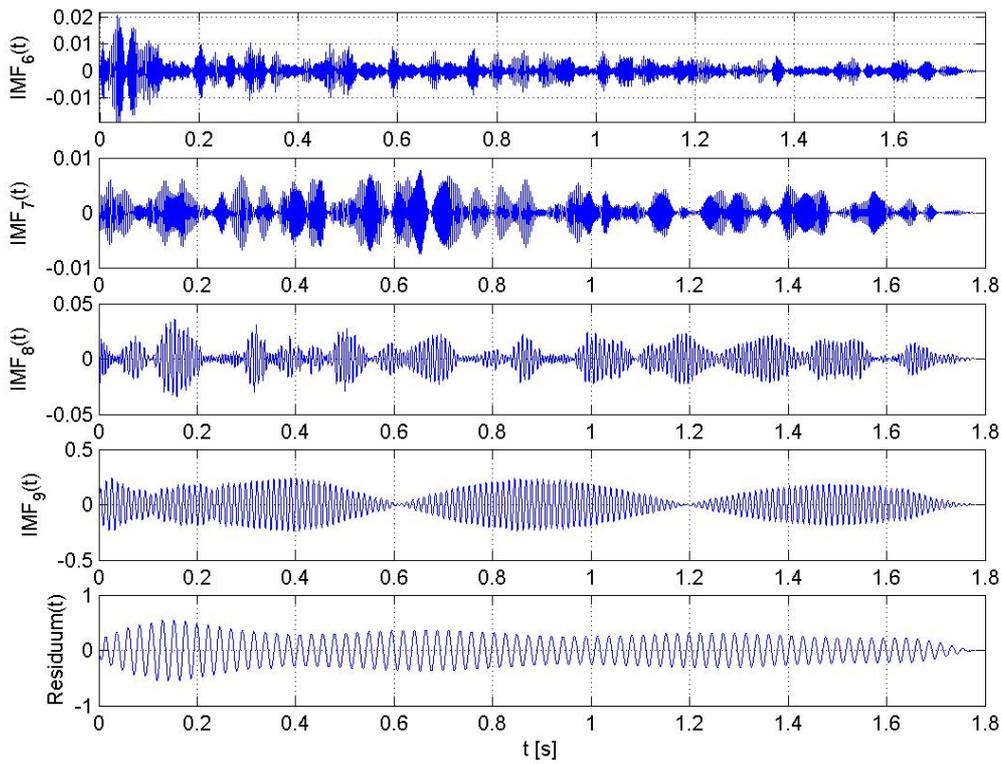
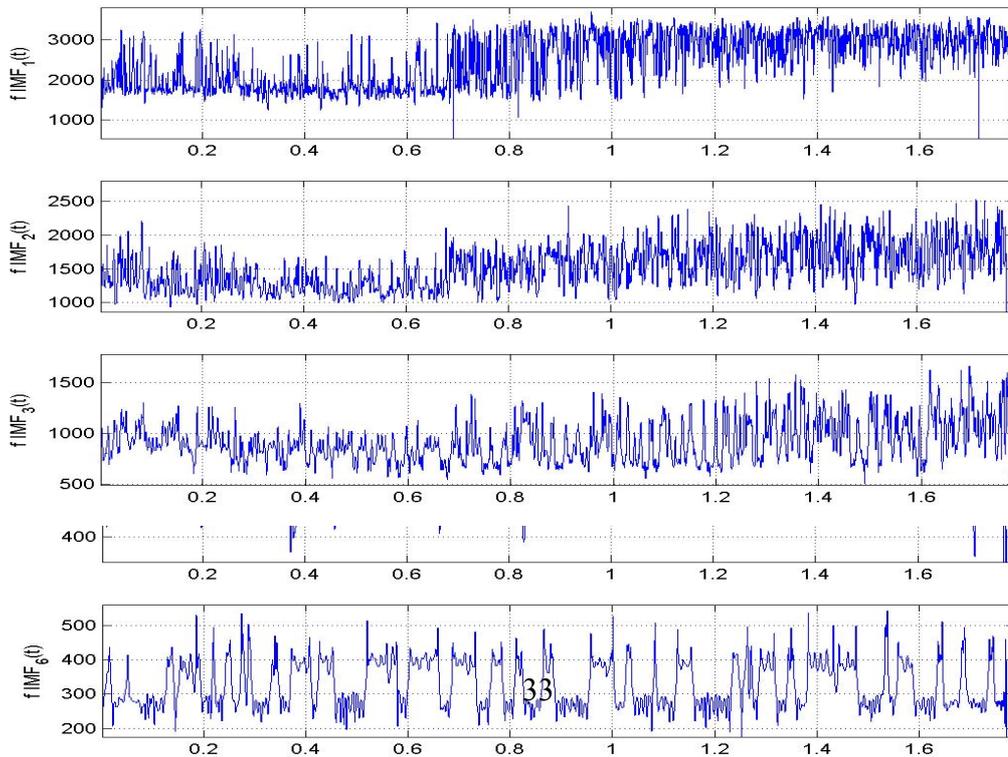


Abb.23: Detektierte IMF + Residuum

Deutlich ist in Abbildung 23 die frequenz - hierarchische Ordnung zu sehen. Diese wird zusätzlich durch den Frequenzinhalt untermauert, den nun jede dieser IMFs inkludiert.



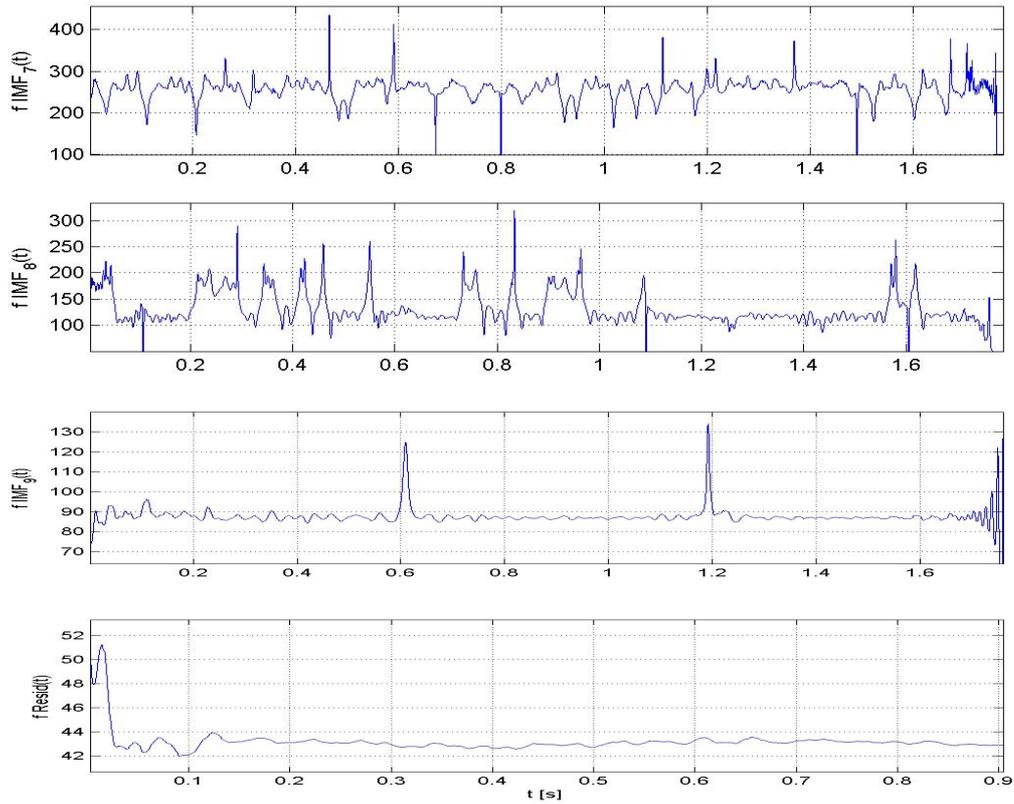
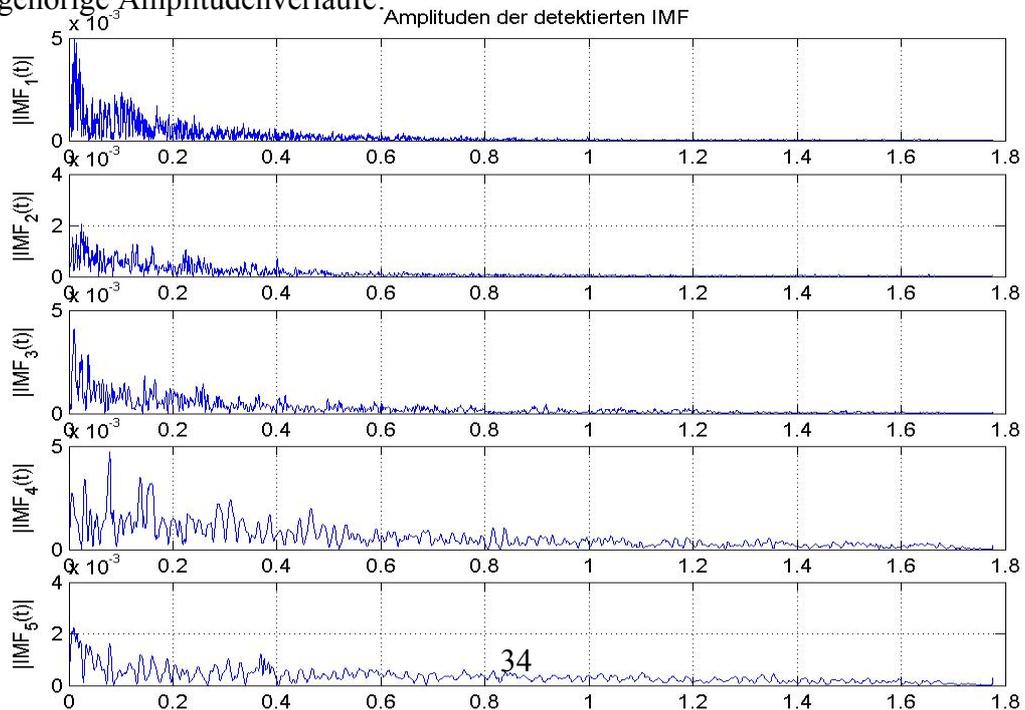


Abb.24: Frequenzverläufe der detektierten IMFs über die Zeit

→ zugehörige Amplitudenverläufe:



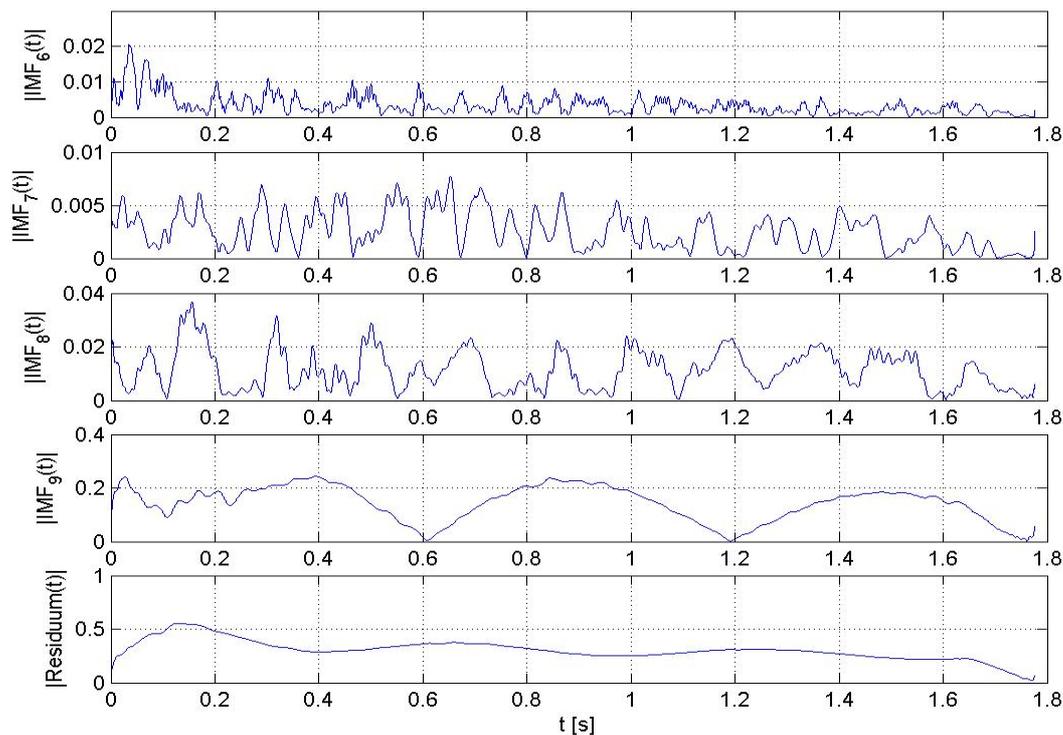


Abb.25: Amplitudenverläufe der gesiebten IMF über die Zeit

Wie zu erwarten war, sind die hochfrequenten Komponenten (IMF1 - 3) nur am „Beginn“ im Spektrum prominent vertreten und klingen innerhalb der ersten 0.6s angenähert nach einer e – Funktion ab.

Umgekehrt ist das Residuum, welches doch eine ziemlich konstante Frequenz bei ca.43Hz enthält und IMF9 bestimmend für den einhüllenden zeitlichen Verlauf des Signals.

Auch der Intrinsic Mode Nummer 9 löst eine Frequenz auf, welche um ca. 88Hz herum schwankt. Diese Schwankungen werden als „*inrawave frequency modulation*“ [1] bezeichnet. → Es kommen jene Frequenzen vor, innerhalb derer die Schwankung stattfindet. D.h.: im vorliegenden den Fall kommen die sehr eng bei einander liegenden Frequenzen von ca. 85Hz und 89Hz vor. (→ Schwebung)



Abb.26: FFT von Signal bis 100Hz

Wenn man eine FFT – Analyse verwendet, so werden die eben genannten Frequenzbereiche auch in diesem Sinne aufgelöst. Sehr deutlich sind die beiden  $f$  – Komponenten zwischen 85- und 90Hz erkennbar (Abb.26).

Allgemein lässt sich sehr gut in den vorhergehenden Abbildungen erkennen, dass es sich bei den Intrinsic Moden tatsächlich um amplituden- und frequenzmodulierte Funktionen handelt.

Beim Vergleich der cutoff – Frequenzen (*Tabelle2*) mit dem detektierten Frequenzinhalt der einzelnen IMFs lässt sich feststellen, dass genau die Komponenten die sich eigentlich in dem jeweiligen Band befinden müssten, erst bei der nächsten IMF detektiert werden (→ Vergleich bei Residuum und IMF9: die niederfrequenteste Grenzfrequenz liegt bei ~38Hz, also müsste die prominente 43Hz Komponente noch in der IMF9 zu finden sein.) Der Grund dafür liegt wiederum darin, dass der Algorithmus (EMD – Siebungsprozess) versucht, die im Band enthaltenen Anteile energetisch auszugleichen. Dadurch befinden sich, so wie hier passiert, die 85 und 88Hz Komponenten in Band 9 und die 44Hz im nächsten darunter liegendem Residuum.

Ein weiteres Problem stellt in diesem Zusammenhang der Hilberttransformator dar, welcher auf Fourier basiert und somit auch das Gibb'sche Phänomen eine wesentliche Rolle spielt, wenn eine größere Frequenzänderung vorliegt.

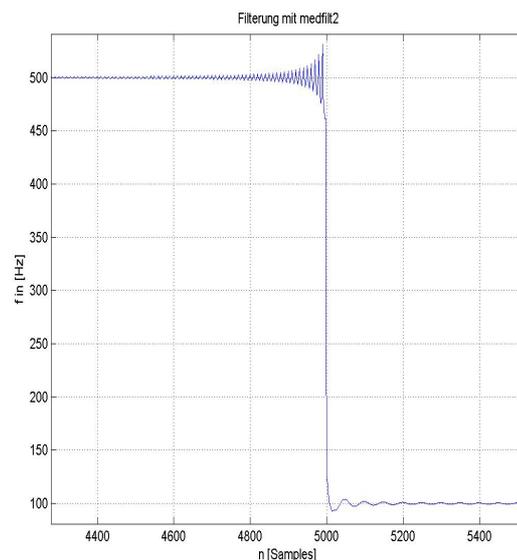
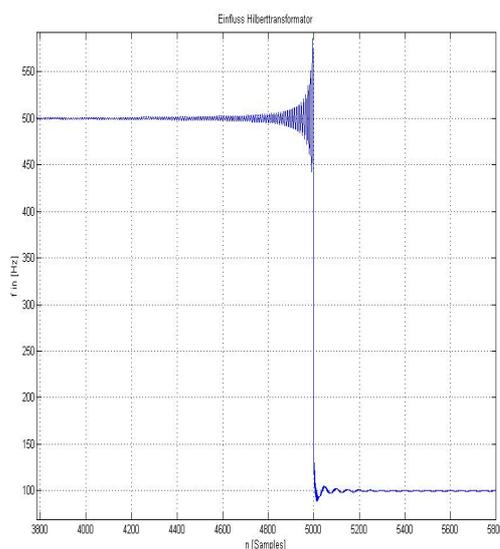
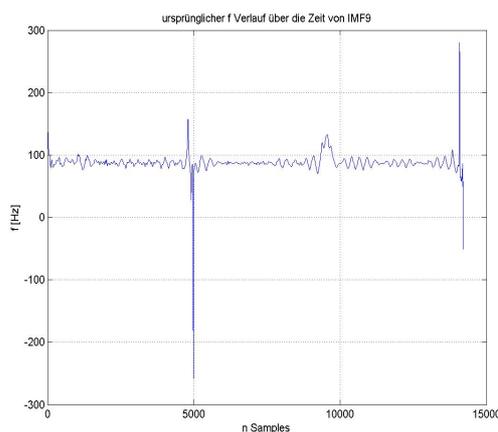
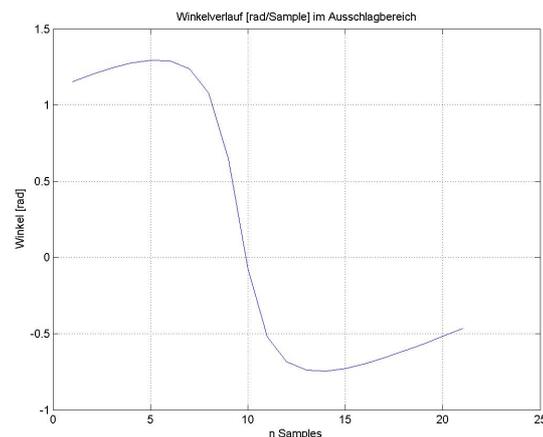


Abb.27a: Einfluss Hilberttransformator

Abb.27b: Begrenzung von Hilberteinfluß

In Abb. 28a/b ist jeweils der Frequenzverlauf eines Sinussignals, welches seine Frequenz ab Sample 5000 von 500Hz auf 100Hz ändert, dargestellt. Die linke Abbildung zeigt den Einfluß vom Hilberttransformator. Ganz deutlich ist das Aufschwingen vor dem  $f$  – Sprung zu erkennen (bis  $\sim 600$ Hz). Die rechte Abbildung zeigt jenes Ergebnis, welches erreicht wird, wenn das  $f$  – Signal noch mit einem 2dimensionalen Median Filter (*in Matlab: medfilt2*) bearbeitet wird. Dieser Filter funktioniert dermaßen, dass für jedes Sample der Medianwert, betrachtet in einer vordefinierten umliegenden Nachbarschaft, ausgegeben wird.

Dies führt in weiterer Folge auch zu einer Glättung der Frequenzverläufe, wodurch kein Problem mehr mit extremen, über kurzen Bereichen vorhandenen Signalausschlägen entsteht.

Abb.28a:  $f$  – Ausschläge ohne GlättungAbb.28b: Ausschnitt von Phasenverlauf bei  $f$  - Ausschlag

Die Ursache derartiger Ausschläge im  $f$  – Verlauf liegt darin, dass wenn sich der Phasenverlauf von einem Sample zum nächsten größer ändert (siehe Abb.28b), die Ableitung nach der Zeit für die Frequenz einen derartigen Sprung ergibt, wie er in Abb.28a bei Sample  $\sim 5000$  zu erkennen ist. In diesem Fall führt auch das *unwrap* der Phase nicht zum Ziel, weil diese Option nur  $\pi$  – Sprünge ausgleicht, die Phasenänderung die derartige  $f$  – Ausschläge verursacht, aber viel kleiner als  $\pi$  sein kann.  $\rightarrow$  Somit ist die Anwendung des verwendeten 2dimensionalen Medianfilters eine bessere Lösung.

Die Orthogonalität zwischen den aufeinander folgend gefunden IMFs lässt sich mit dem unter Gl.8 definierten Orthogonalitätsindex berechnen. (Seite 13).

Signl-	IMF	$\Sigma$ IMF									
--------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	--------------

vergleich	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	Original
$h$										
$IO$	0.011	0.043	0.283	0.243	0.137	0.12	0.038	0.08	0.006	1

Tabelle 3: Orthogonalität zwischen den aufeinanderfolgenden IMFs + Summe der IMF und dem Originalsignal

Deutlich ist zu erkennen, dass das Originalsignal aus den gefundenen IMFs wieder eindeutig rekonstruiert werden kann ( $\rightarrow IO = 1$ , entspricht im Prinzip dem Kreuzkorrelationstherm zwischen den beiden Signalen!).

Um die in Gl.4 (Seite 8) beschriebene Vollständigkeit auch bei dieser Zerlegung zu veranschaulichen, siehe folgende Aufsummierung:

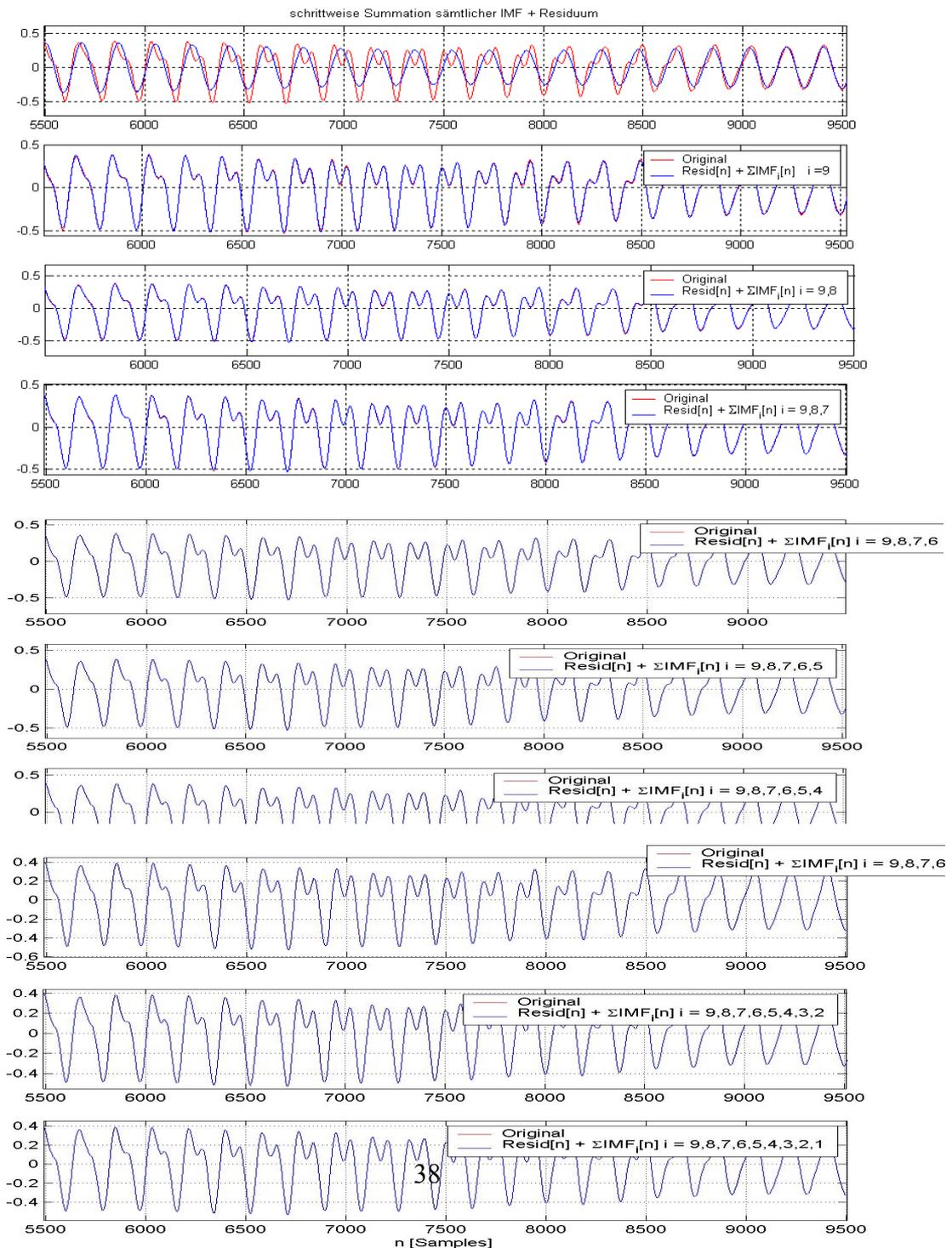


Abb.29: Nachvollzug der Vollständigkeit

Bei genauerer Betrachtung der Abbildung 29 lässt sich feststellen, dass in diesem Fall bereits das Residuum und IMF 9,8,7 eine beinahe Vollständigkeit des Signals gegenüber dem Originalsignal herbeiführen. Alle weiteren Beiträge liegen in einem Auflösungsbereich, welcher in dieser Darstellungsform nicht mehr aufgelöst wird. (Siehe auch Abb.25: Beiträge der hochfrequenten IMFs im Amplitudenbereich von  $10^{-3}$  )

Die erhaltene Differenz zwischen IMF – Summe und dem Originalsignal liegt wiederum im Bereich der numerischen Auflösung des Programms ( $10^{-15}$ ). Dieser Fehler ist bereits in Abbildung 22a dargestellt.

Vergleich der einzelnen Mode Funktionen mit dem Originalsignal im f – Bereich:

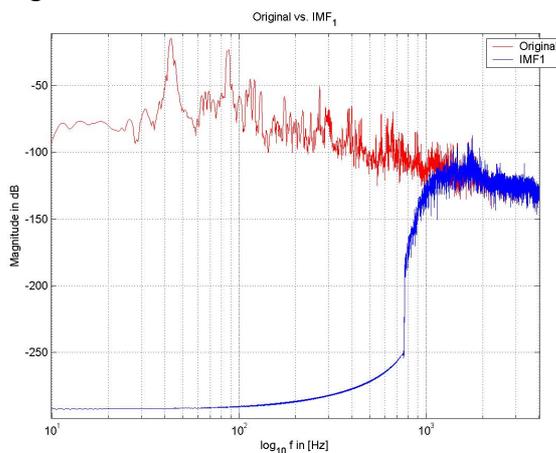


Abb.30a: Vergleich IMF1 mit Original

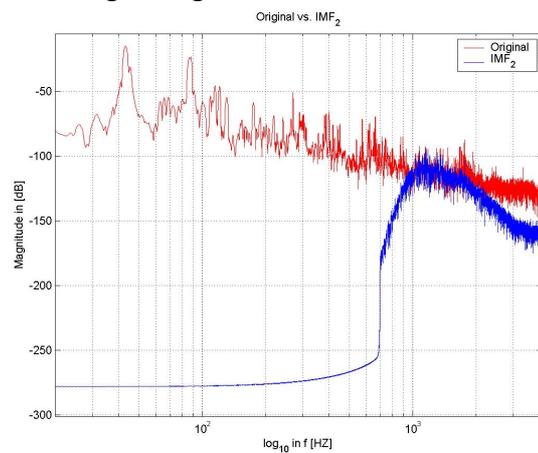


Abb.30b: Vergleich IMF2 mit Original

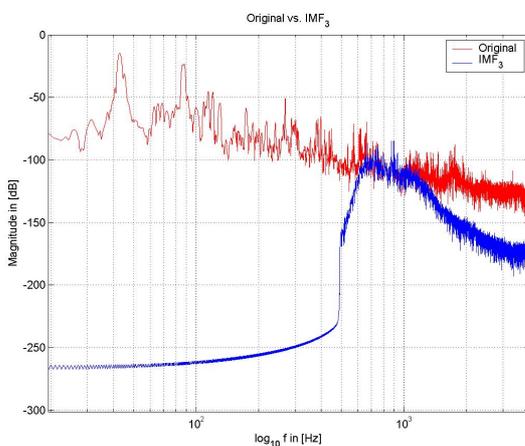


Abb.30c: Vergleich IMF3 mit Original

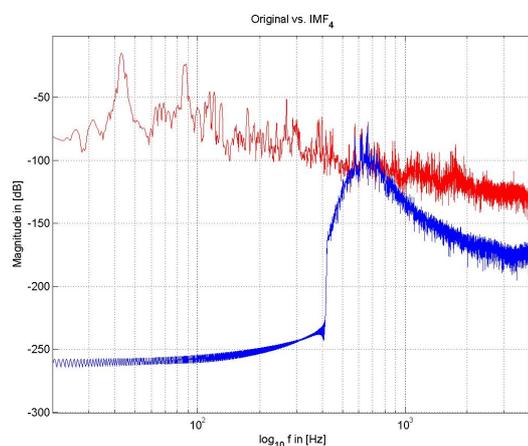


Abb.30d: Vergleich IMF4 mit Original

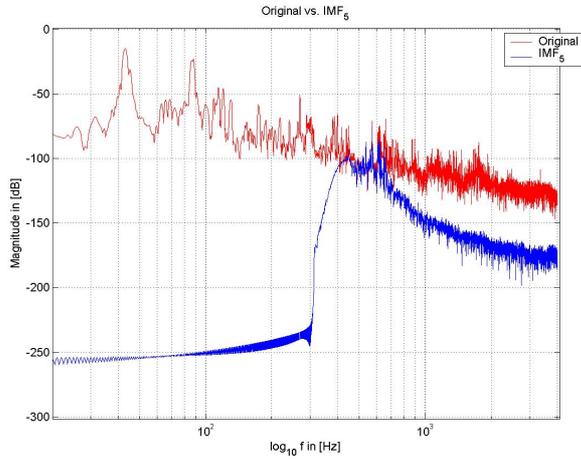


Abb.30e: Vergleich IMF5 mit Original

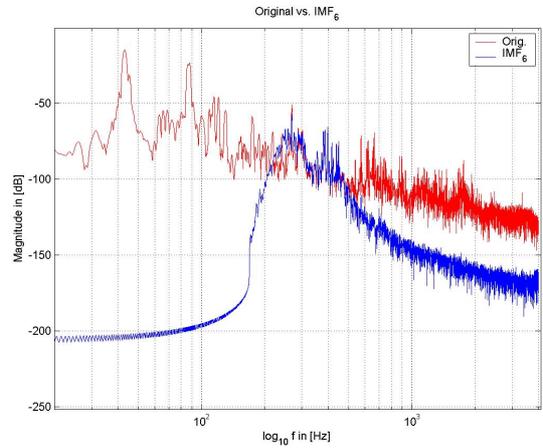


Abb. 30f: Vergleich IMF6 mit Original

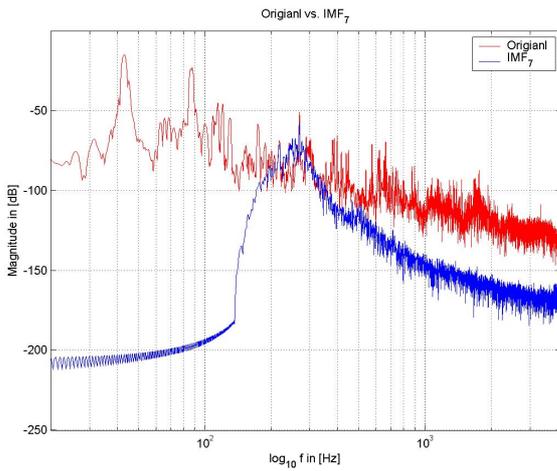


Abb.30g: Vergleich IMF7 mit Original

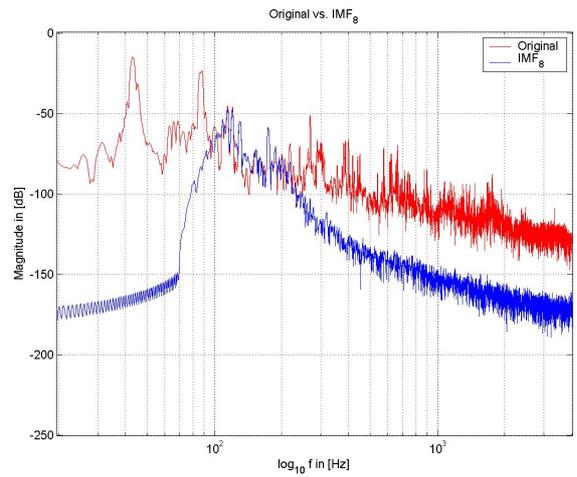


Abb.30h: Vergleich IMF8 mit Original

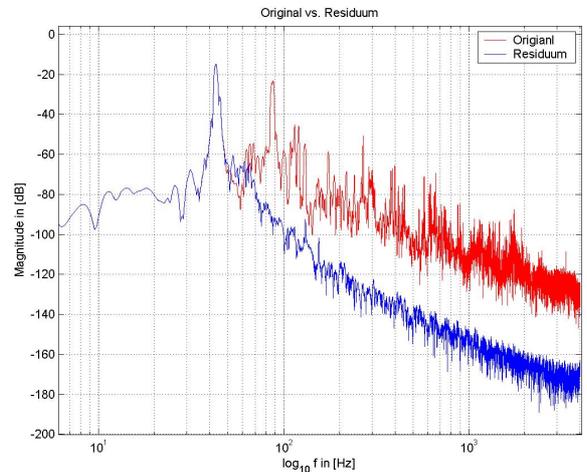
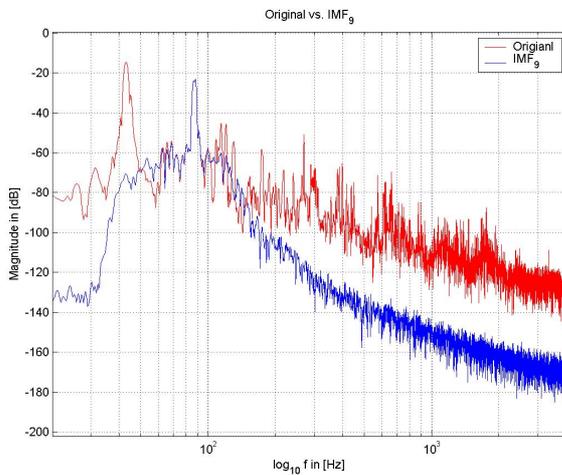


Abb.30i: Vergleich IMF9 mit Original  
Original

Abb.30j: Vergleich Resid. mit

Aus den erhaltenen Amplituden- und Frequenzsignalen wird nun wiederum eine Resynthese erstellt. Dabei dient jeder erhaltene Frequenzverlauf über die Zeit als Steuersignals eines Sinusoszillators sowie die dazugehörigen Amplituden.

### 3.2. Vergleich mit Gammatonfilterbank

Um nun einen objektiven Vergleich anstellen zu können, wie groß der Fehler des aus der Resynthese gewonnen Signals gegenüber dem Originalsignal in der perzeptiven Domäne ist, wurden beide Signale mit einer Gammatonfilterbank untersucht.

Die Gammatonfilterbank führt eine Filterung des Signals ähnlich der Basilarmembran im menschlichen Gehör durch. Dabei erfolgt eine Zerlegung in kritische Bandbreiten.

*kritische Bandbreiten:*

Die maximale Bandbreite, innerhalb der das Gehör die Schallenergie von Einzeltönen zur Bildung der Ruhehörschwelle bzw. Mithörschwelle zusammenfasst, ist die Frequenzgruppenbreite oder „kritische Bandbreite“. Diese werden in der Barkskala angegeben, wobei psychoakustische Versuche ergaben, dass der Mensch über den gesamten Hörfrequenzbereich 24 Frequenzgruppen besitzt.

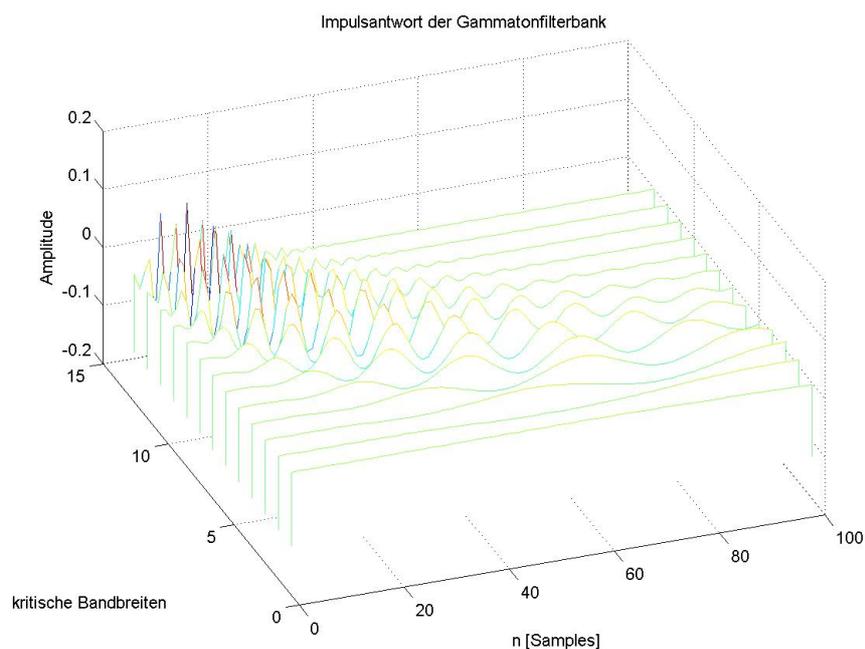


Abb.31: Impulsantwort der Gammatonfilterbank

In Abbildung 31 wird die Impulsantwort der verwendeten Gammatonfilterbank dargestellt. Es zeigt sich, dass die tieffrequenten Anteile (niedrige kritische Bandbreitenindex) zeitlich später aufgelöst werden als die hohen Frequenzen. (→ Auflösung hoher Frequenzen am Beginn der Basilarmembran und somit schnellere Auswertung im Gehirn!).

Bevor allerdings die eben beschriebene Filterung durchgeführt wird, wird noch mit einem A – Bewertungsfiler die Empfindlichkeit des Gehörs simuliert.

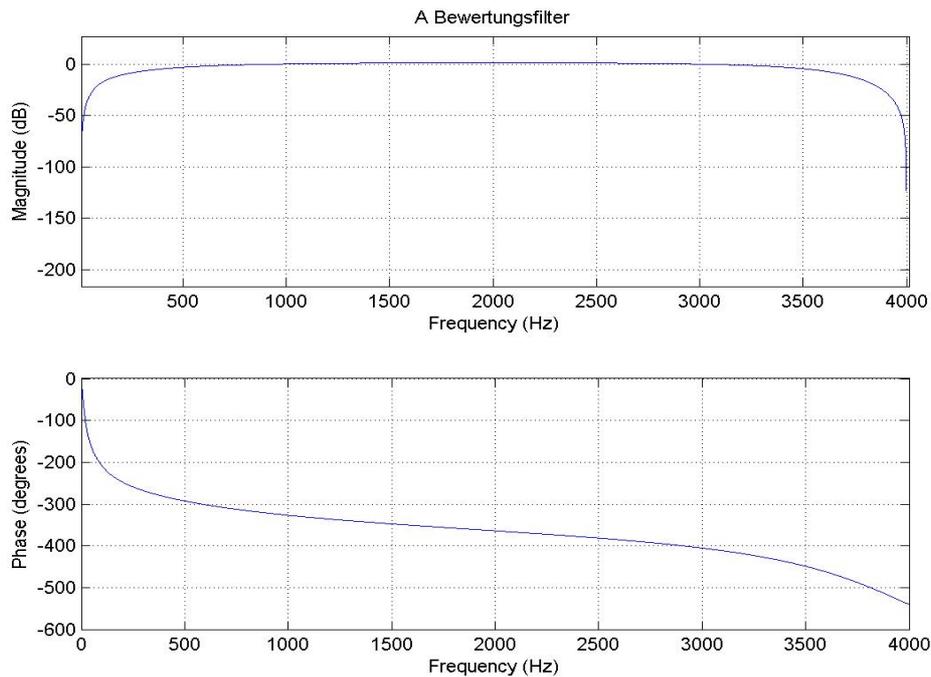


Abb.32: Bode – Diagramm von A – Bewertungsfiler mit  $f_s = 8\text{kHz}$

Als Vergleichskriterium zwischen den einzelnen kritischen Bandbreiten, welche bei dem Filterprozess der beiden Signale (Original und Resynthese) erhalten wurden, dient der Vergleich zwischen den Energien der beiden aktuell analysierten krit. Bändern.

Dabei wurden jeweils die Energien beider Ergebnisse detektiert, die Differenz gebildet, auf die Energie des krit. Bandes, welches vom Originalsignal errechnet wurde, bezogen, sowie daraus die Wurzel gezogen. Natürlich wird ein derartiges Fehlermaß in dB angegeben.

Es wurden 13 kritische Bandbreiten detektiert, bei einem Frequenzbereich von 25- bis 3000Hz.

Als Ergebnis erhält man nun für alle 13 Bänder einen Fehler (dB):

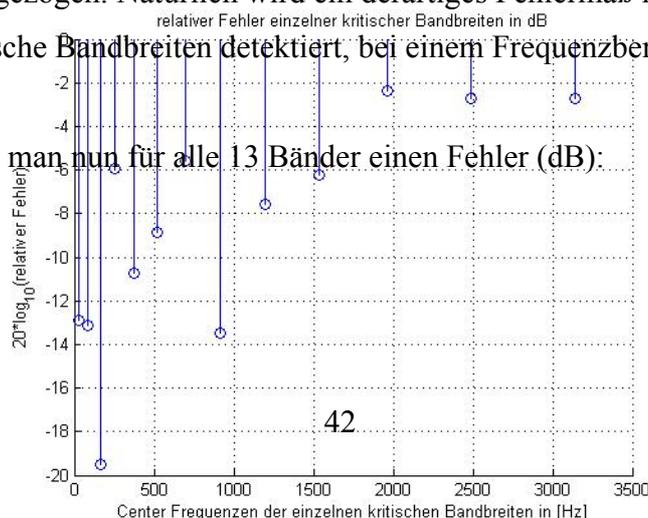


Abb.33: Fehler einzelner krit. Bandbreiten

Die zugehörigen Centerfrequenzen der einzelnen Bänder lauten wie folgt:

$f_{\text{Center}}$	25	86	161,8	255,8	372,5	517	696,5	920	1195,5	1538	1963	2490	3144
[Hz]													

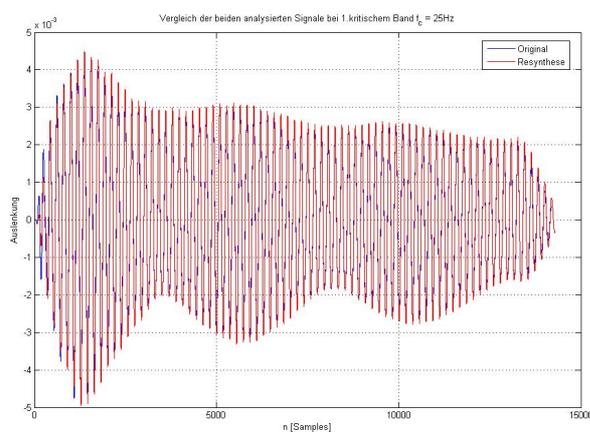
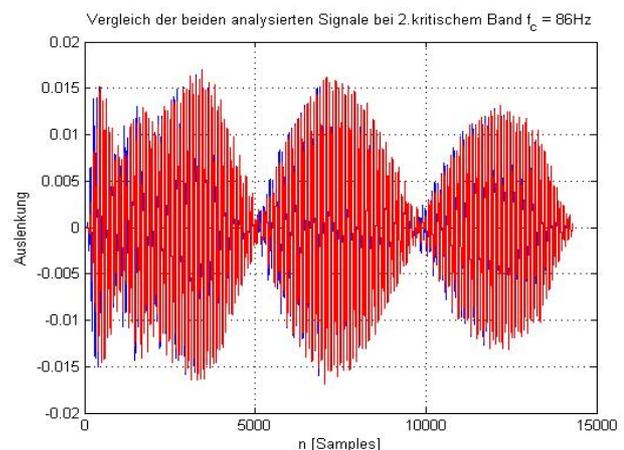
Tabelle 4: Center Frequenzen der krit. Bandbreiten

Abbildung 33 zeigt sehr deutlich, dass der energetische Fehler, bezogen auf die gesamte Zeitspanne des Signals, vor allem im tieffrequenten Bereich klein ist. Lediglich über

1500Hz liegt der Fehler bei  $\sim \frac{1}{\sqrt{2}}$  ( $\sim -3\text{dB}$ ). In diesem Bereich muss allerdings

berücksichtigt werden, dass der EMD inhärente Auflösungsfehler in diesem Bereich größer ist, als bei den tiefen Frequenzen ( $\rightarrow$  Auch wenn SSB-Modulation, denn bei erster cutoff-Frequenz von  $\sim 1\text{kHz}$  wird um  $800\text{Hz}$  gegen DC verschoben. Somit befinden sich die höchsten Komponenten bei ca.  $3\text{kHz}$  !! (siehe auch Abb.19 Seite 27) )

$\rightarrow$  Beispiel 2er kritischer Bandbreiten im Vergleich:

Abb.34a: Vergleich der krit.Bandbr.bei  $f_c = 25\text{Hz}$ Abb.354: Vergleich der krit.Bandbr.bei  $f_c = 86\text{Hz}$ 

Somit lässt sich durchaus schlussfolgern, dass das resynthetisierte Signal sehr ähnlich dem Originalsignal klingt.

Um dieses Kapitel noch zu beenden, folgt ein kurzer Ausflug in die Psychoakustik, um eine Vorstellung für eine Grenze zu bekommen, die beschreibt, wie groß nun die Unterschiede in der Lautstärke (bzw. Schallintensität) sein müssen, damit zwei Schallereignisse unterscheidbar werden [7]:

Zur Messung der Intensitätsauflösung, d.h. des kleinsten wahrnehmbaren Unterschieds in der Schallintensität, werden drei verschiedenen Methoden eingesetzt:

- Modulationsdetektion (bei niedriger Modulationsfrequenz)
- Detektion der Erhöhung des Pegels eines kontinuierlichen Hintergrundstimulus
- Intensitätsdiskrimination von gepulsten Stimuli („Welcher von 2 Stimuli war der lautere?“)

Die mittels dieser drei Methoden gewonnenen Ergebnisse stimmen relativ gut überein. Für Breitband und Bandpass-gefiltertes Rauschen trifft das Weber'sche Gesetz ( $DL/I = \text{konstant}$ ) [siehe Kap.1 [7] ] zu. Für Schallpegel von ca. 20-100 dB SL beträgt DL (in dB) ca. 0.5-1 dB und für schwellennahe Stimuli ist DL höher. Im Gegensatz dazu gilt das Weber'sche Gesetz nicht für Sinustöne:

Die Funktion von DL über L (in dB) ergibt eine Gerade mit einem Anstieg von 0.9 statt 1 (entsprechend dem Weber'schen Gesetz), d.h. die Intensitätsauflösung wird mit steigendem Pegel besser. Bei 20 dB SL beträgt DL ca. 1.5 dB, bei 40 dB SL ca. 0.7 dB und bei 80 dB SL ca. 0.3 dB (alle bei 1000 Hz). Diese Diskrepanz zu den Ergebnissen mit Rauschsignalen wird in der Literatur als „near miss to Weber's Law“ bezeichnet.

**Anhang:**            Verwendete Literatur und Internetlinks

- [1]    “*The empirical mode decomposition and the Hilbert spectrum for nonlinear and non – stationary time series analysis*” von Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shith, Quanan Zheng, Nai – Chyuan Yen, Chi Chao Tung and Henry H. Liu , eingereicht 3. Juni 1996; akzeptiert 4. November 1996,  
Seite 903 bis 995
- [2]    “*On empirical mode decomposition and its algorithms*” von Gabriel Rilling, Patrick Flandrin and Paulo Goncalves
- [3]    *TimeFrequency MatlabToolbox* von Flandrin Homepage  
<http://perso.ens-lyon.fr/patrick.flandrin/>
- [4]    “*Algorithm for extraction of pitch and pitch salience from complex tonal signals*” von Ernst Terhardt, Gerhard Stoll, Manfred Seewann, eingereicht 14.Jänner 1981; akzeptiert 29. September 1981, Seite 680 bis 682
- [5]    *BinauralToolbox* von Dr Michael A Akeroyd, MRC Institute of Hearing Research,  
<http://www.ihr.gla.ac.uk> bzw. [tp://www.ihr.mrc.ac.uk](http://www.ihr.mrc.ac.uk)
- [6]    „*Digitale Audiotechnik I*“ ,Ao. Univ.-Prof. Dr. Gerhard Graber, Institut für Breitbandkommunikation TU – Graz, Skriptum Vs.8 WS 2004/05
- [7]    „*Psychoakustik II Schwerpunkt: Experimentelle Audiologie*“, Bernhard Laback, Skriptum zur Vorlesung, Kapitel 4.4.3 Intensitätsauflösung, Seite 90

