

Abschlussarbeit zum Toningenieurprojekt

# Klangmodellierung von Gamelaninstrumenten basierend auf exponentiell ausschwingenden Sinuskomponenten

Margherita Jammer

Graz, am 7. Januar 2008

Betreuer: Univ.Ass. DI Dr. Alois Sontacchi

Mitbetreuer: DI Franz Zotter



**UNIVERSITÄT**  
FÜR MUSIK UND  
DARSTELLEND KUNST  
GRAZ - AUSTRIA



## Zusammenfassung

Im vorliegenden Projekt werden Klänge von Gamelaninstrumenten des lauten Stils durch ein einfaches parametrisierbares Signalmodell beschrieben. Die untersuchten Instrumente gehören zu den Idiophonen. Deren Klänge können durch die typische Signalform weitgehend abrupt einsetzender und exponentiell abklingender Sinusschwingungen beschrieben werden. Diese Schwingungsformen stellen die so genannten Eigenmoden bzw. Teiltöne des jeweiligen Klangkörpers dar. Da im Rahmen des Projektes „Virtual Gamelan Graz (VGG)“ (vgl. [1]) die Eigenschaften der einzelnen Teiltöne getrennt voneinander behandelt werden müssen, soll das Analyse- und Syntheseverfahren diese Anforderung erfüllen. Bei der Klangsynthese sollen jeweils die gehörrelevanten Signalkomponenten eines Klanges ermittelt werden, die in Folge durch eine reduzierte Anzahl von Parametern beschrieben werden. Dabei wird im ersten Schritt eruiert, aus welchen Frequenzkomponenten sich der Klang zusammensetzt. Die Klänge werden im Vorfeld so gefiltert, dass die Empfindlichkeit des Gehörs berücksichtigt wird und für das Gehör nicht relevante oder maskierte Komponenten eliminiert werden. Das heißt, es erfolgt eine Berücksichtigung der Übertragungseigenschaft des Außen- und Mittelohres sowie die des Innenohres durch Berechnung der simultanen Maskierung bzw. der zeitlich variablen Mithörschwelle (vgl. [3]).

In einem zweiten Schritt sollen mittels „Bark Zooming Autoregressive Moving Average (BZ-ARMA)“, einem Verfahren basierend auf „Frequency Zooming Autoregressive Moving Average (FZ-ARMA)“ (vgl. [2]), aus dem übrig bleibenden Spektrum die jeweilige Frequenz mit zugehöriger Startphase, Startamplitude und Ausklingkonstante ermittelt werden. Aus diesen vier Parametern soll später die Resynthesierung des Gamelanklanges erfolgen. Wie in der hier vorgelegten Arbeit später ersichtlich werden wird, muss zur befriedigenden Rekonstruktion der Audiodaten, ein zusätzlicher Parameter - die Einschwingzeit - eingeführt werden.

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>5</b>
1.1. Gamelan . . . . .	5
<b>2. FZ-ARMA Frequency Zooming Autoregressive Moving-Average</b>	<b>7</b>
<b>3. BZ-ARMA Bark Zooming Autoregressive Moving-Average</b>	<b>9</b>
3.1. Psychoakustische Betrachtung . . . . .	9
3.1.1. Gewichtung des Signals mit A-Filter Kurve . . . . .	9
3.1.2. Maskierung . . . . .	11
3.1.3. Zerlegung in Barkbänder und Nutzung der Maskierungseigenschaften des menschlichen Gehörs . . . . .	12
3.1.4. Modulation und Downsampling . . . . .	15
3.2. Signalanalyse . . . . .	16
3.2.1. LPC . . . . .	16
<b>4. Analyse-Programm</b>	<b>20</b>
4.1. Das zu bearbeitende Material . . . . .	20
4.2. Die Berechnungsschritte des Programms . . . . .	21
4.2.1. Inputbegrenzung . . . . .	21
4.2.2. Downsampling . . . . .	22
4.2.3. Gewichtung mit A-Filter . . . . .	22
4.2.4. Barkbandfilterung . . . . .	22
4.2.5. Elimination maskierter Bänder und Signalanteile . . . . .	22
<b>5. Resynthese</b>	<b>25</b>
5.1. Klangmodellierung mit Residuum . . . . .	25
5.2. Klangmodellierung mit Decaying-Exponentials . . . . .	25
5.3. Synthese-Filterbank . . . . .	30
<b>6. Analyse-Beispiele</b>	<b>32</b>
6.1. Gong Suwukan . . . . .	35
6.2. Bonang Barung . . . . .	35
6.3. Kempul . . . . .	37
6.4. Peking A . . . . .	39

<b>7. Test der Funktionalität des Analysprogrammes</b>	<b>44</b>
7.1. Programmtest . . . . .	44
7.1.1. Test mit einfachen Sinussignalen . . . . .	44
7.1.2. Test mit Sinuskombinationen . . . . .	46
7.2. Verbesserung der Analyseergebnisse durch gezielte Manipulation des Programmablaufes . . . . .	49
7.2.1. Gong Ageng 13 . . . . .	50
<b>Literaturverzeichnis</b>	<b>55</b>

# 1. Einleitung

## 1.1. Gamelan

Ein Gamelan ist ein südostasiatisches Ensemble, das vorwiegend aus gestimmten Idiophonen (Gongs, Gongspiele, Metallophone) und weiteren Instrumenten besteht. „In Zentraljava (Indonesien) wird mit solchen Orchestern die an lokalen Fürstenhöfen entstandene klassische Musik gespielt, die heute auch an örtlichen Musikhochschulen gelehrt wird.(...)



Abbildung 1.1.: Bildausschnitt eines typischen Gamelanorchesters (Quelle: IME-Homepage)

(...),„Virtual Gamelan Graz (VGG)“ ist ein Projekt, das im Hinblick auf musikethnologische Forschung neue Möglichkeiten bieten soll, welche auch den gegenwärtigen Stand der Entwicklung im Bereich der spektralen und spatialen Klangsynthese sowie der algorithmischen Verfahren zur Generierung von musikalischen Strukturen nutzen.(...) Es sollte also ein virtuelles Gamelan-Orchester erzeugt werden, welches selbständig Kompositionen der klassischen Musik Zentraljavas in verschiedenen Stimmungen und modifizierbaren Klangfarben generiert, und wodurch das Spiel auf unterschiedlichen Sets von Instrumenten simuliert werden kann.“ (vgl. [1] Zusammenfassung und Kapitel 1, Einleitung). Für die sinnvolle Synthetisierung der besagten Klänge ist es jedoch von absoluter

Notwendigkeit, die genaue spektrale Zusammensetzung eines jeden Klangs zu kennen.

Wie bereits in der Zusammenfassung erwähnt, war die Aufgabe für diese Projektarbeit, von für das Gehör irrelevanten Anteilen befreite Signale verschiedener Gamelaninstrumente zu analysieren, ihre Frequenzkomponenten zu extrahieren, und die zu den Frequenzen jeweils zugehörige Startphasen, Startamplituden und Ausklingkonstanten zu ermitteln. Jedoch stellte sich im Laufe der praktischen Umsetzung heraus, dass ein fünfter Parameter, der den Einschwingvorgang beschreibt, vonnöten war, um die Klänge gut rekonstruieren zu können. Die fünf Parameter - Frequenz, Startphase, Startamplitude, Einschwingzeit und Ausklingkonstante - ermöglichen letztendlich eine näherungsweise realistische Synthetisierung von Gamelanklängen für die weitere Nutzung in „Virtual Gamelan Graz“.

## 2. FZ-ARMA Frequency Zooming Autoregressive Moving-Average

Als Grundlage der vorliegenden Projektarbeit dient das FZ-ARMA (Frequency Zooming Autoregressive Moving-Average) Verfahren. Es wurde entwickelt um sehr nahe beieinander liegende Spektralkomponenten in breitbandigen Signalen aufzuspüren. Dabei wird grundsätzlich der für die Analyse interessante Teil des Signals gegen Null herab moduliert.

$$h(m) = e^{j\omega_m n} h(n) \quad (2.1)$$

Wobei  $\omega = 2\pi f_m / f_s$  die Modulationsfrequenz, und  $f_s$  die Samplefrequenz angibt. In der  $Z$ -Ebene bedeutet dies ein Verschieben der Pole (und Nullstellen)  $z_i$  um einen Winkel  $\omega_m$  im Uhrzeigersinn.

$$\omega_{i,rot} = \omega_i - \omega_m = \arg(z_i) - \omega_m \quad (2.2)$$

Die Polradien werden jedoch beibehalten. Die Bandbreite des Signals wird durch Tiefpassfilterung weiter eingeschränkt. Die rotierte Impulsantwort wird dann um einen Zoomingfaktor  $K_{zoom}$  downgesamlet. Die neue Samplingrate ergibt sich zu  $f_{s,zoom} = f_s / K_{zoom}$ . Das wiederum impliziert ein Mappen der Pole (und Nullstellen) auf eine neue  $z$ -Ebene, in der eine Skalierung mit  $z_{i,zoom} = z_i^{K_{zoom}}$  gilt und eine Reduktion der Polradien, wie in Grafik 2.1 dargestellt.

Um auf die gehörspezifischen Aspekte eingehen zu können, wird in dieser Projektarbeit die FZ-ARMA Methode um ein psychoakustisches Selektionsmodell erweitert worauf im folgenden Kapitel 3 näher eingegangen wird.

## 2. FZ-ARMA Frequency Zooming Autoregressive Moving-Average

---

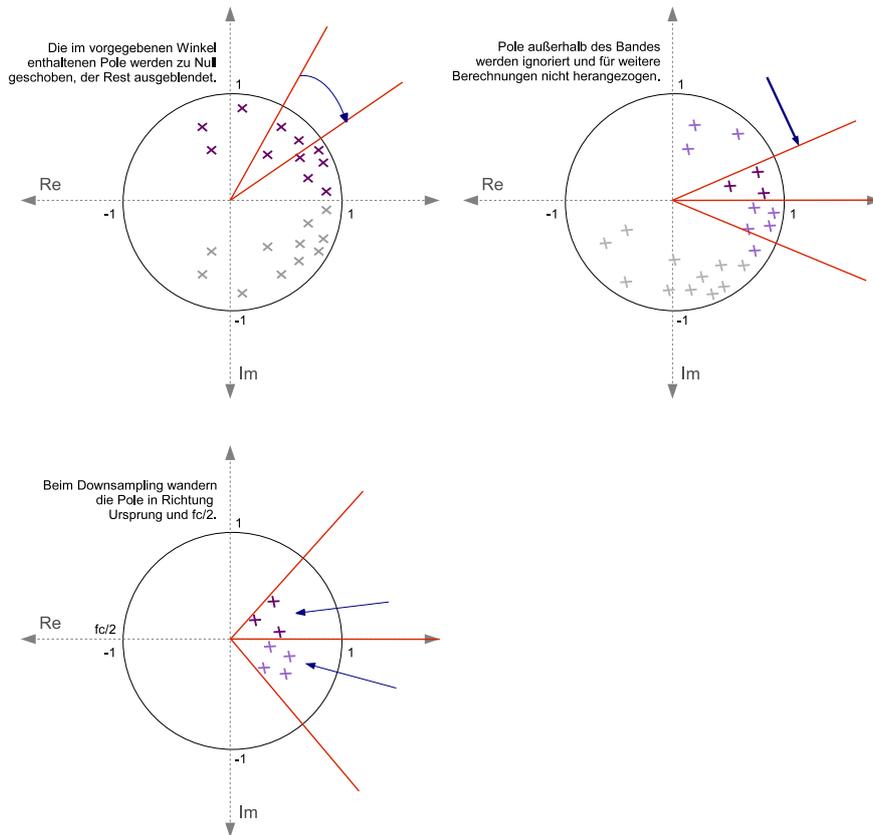


Abbildung 2.1.: Schritte der FZ-ARMA

## 3. BZ-ARMA Bark Zooming Autoregressive Moving-Average

In dem im Rahmen dieser Projektarbeit erstellten Programm, sollte die Datenreduktion durch psychoakustisch ausschließbare Signalanteile verbessert werden.

Grundsätzlich besteht Bark Zooming ARMA aus zwei wesentlichen Blöcken. Der erste beinhaltet eine psychoakustische Betrachtung und Bandzerlegung der Signale, was starke Datenreduktion ermöglicht. Im zweiten Block wird im Zuge der Signalanalyse eine LPC-Analyse durchgeführt, aus der in weiterer Folge die Pole und somit alle gewünschten Parameter hervorgehen.

Als wesentliche Neuerung des Verfahrens gilt die Verwendung eines, ebenfalls im zweiten Block gebildeten, Signalresiduums. Dieses könnte als innovative klangverbessernde Maßnahme eingesetzt werden, indem man es bei der späteren Synthese mit dem resynthetisierten Klang überblendet.

### 3.1. Psychoakustische Betrachtung

Die Hauptaufgabe dieses ersten Blocks beläuft sich auf das Miteinbeziehen der grundlegenden psychoakustischen Eigenheiten des menschlichen Gehörs. Es erfolgte also eine Berücksichtigung der Übertragungseigenschaft des Außen- und Mittelohres sowie die des Innenohres durch Berechnung der simultanen Maskierung bzw. der zeitlich variablen Mithörschwelle (vgl. [3]).

#### 3.1.1. Gewichtung des Signals mit A-Filter Kurve

Das menschliche Gehör ist, aufgrund anatomischer Gegebenheiten, in verschiedenen Frequenzbereichen verschieden empfindlich. Diese Frequenzabhängigkeit beim Hören manifestiert sich in den Phonkurven (auch Kurven gleicher Lautstärke nach DIN 45 360), die bei 20 bis 40 phon der Kurve A des bewerteten Schalldruckpegels entsprechen. (vgl. [4] und [6])

Um die zuvor erwähnte Abhängigkeit des Gehörs zu simulieren, und die vom Ohr zu stark gedämpften Spektralanteile aus den Berechnungen ausklammern zu können, wurde jedes Soundfile vor der Weiterverarbeitung mit einem A-Filter gewichtet.

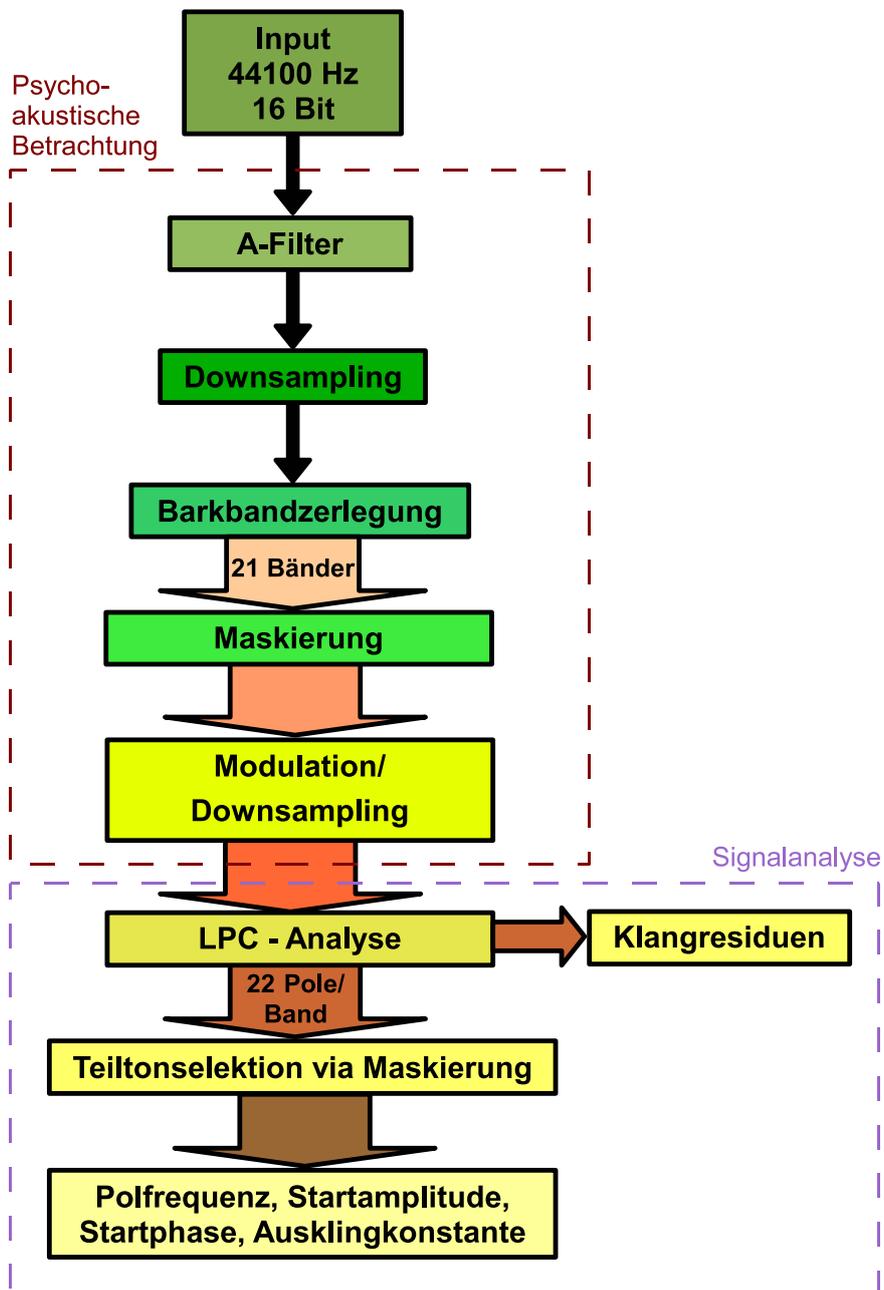


Abbildung 3.1.: Flussdiagramm Bark-Zooming ARMA

### 3.1.2. Maskierung

Eine weitere Eigenheit des menschlichen Gehörs ist die Eigenschaft, gleichzeitig auftretende Spektralkomponenten unter gewissen Umständen nicht richtig aufzulösen. Dabei kommt es immer darauf an, in welchem Abstand und in welchem Lautstärkenverhältnis sie zu einander stehen. Begründet ist dieses Verhalten durch die angesteuerten Frequenzbereiche auf der Basilarmembran im Innenohr, die für benachbarte Frequenzen nahe bei einander liegen. Es wird also jeweils ein gewisser Frequenzbereich vom Ohr als zusammengehörige Frequenzgruppe (Barkband) verarbeitet und somit als Einheit wahrgenommen, innerhalb der zusätzlich auftretende Komponenten als klangfärbend, aber kaum als lautheitssteigernd wirken. Unmittelbar benachbarte Frequenzgruppen wirken sich durch die enge örtliche Ansiedlung der Verarbeitungsbereiche auf der Basilarmembran stark aufeinander aus, indem der jeweils stärker angeregte Bereich mit den anderen Bereichen interferiert. Diese Eigenheit manifestiert sich als Dämpfung in der aktiven Wahrnehmung jener Gruppen (oder in diesen Gruppen vorkommender Töne). Zu höheren Frequenzen (oder Frequenzgruppen) beträgt diese Dämpfung 24dB/Bark und zu tieferen Frequenzen 27dB/Bark. In Kapitel 3.1.3 wird noch näher auf dieses Phänomen eingegangen.

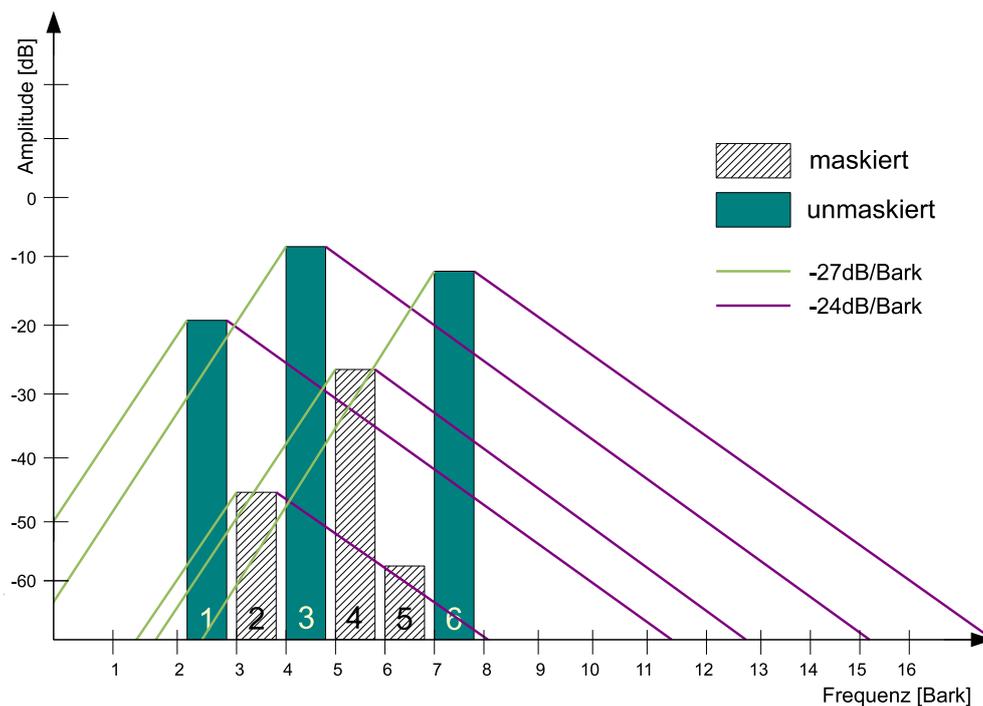


Abbildung 3.2.: Schematische Darstellung der Simultanmaskierung

Zusätzlich zu dieser Simultanmaskierung existiert noch die *zeitlich variable Mithör-*

*schwelle*. Sie beschreibt die Anpassung des Mittelohres an die Lautstärkenänderung der an das Trommelfell dringenden akustischen Signale. Normalerweise passiert diese so, dass die Druckübertragung durch die Gehörknöchelchen so abgeschwächt wird, dass der Schallpegel der Umgebung für den Hörer im angenehmen Bereich liegt. Passiert jedoch eine Anpassung durch einen lauten Impuls, so benötigt das Gehör zumindest 30 ms um sich wieder auf einen normalen Schallpegel einstellen zu können. Tatsächlich geht dieses Phänomen so weit, dass teilweise sogar Anteile eines Schallereignisses, die kurz vor dem lauten Impuls stattfinden nicht mehr aufgelöst werden (Vorverdeckung).

### 3.1.3. Zerlegung in Barkbänder und Nutzung der Maskierungseigenschaften des menschlichen Gehörs

Die Bark-Skala ist eine psychoakustische Skala, die sich für die Beschreibung von Lautheitswahrnehmung und spektraler Klangfarbenwahrnehmung anbietet. Die Skala ist definiert von 0,2 bis 25Bark. Sie ist verknüpft mit der Tonheit in Mel nach Eberhard Zwicker. Es gilt: 1Bark = 100Mel. Geht man von der Eigenschaft des Ohres aus, unterschiedliche

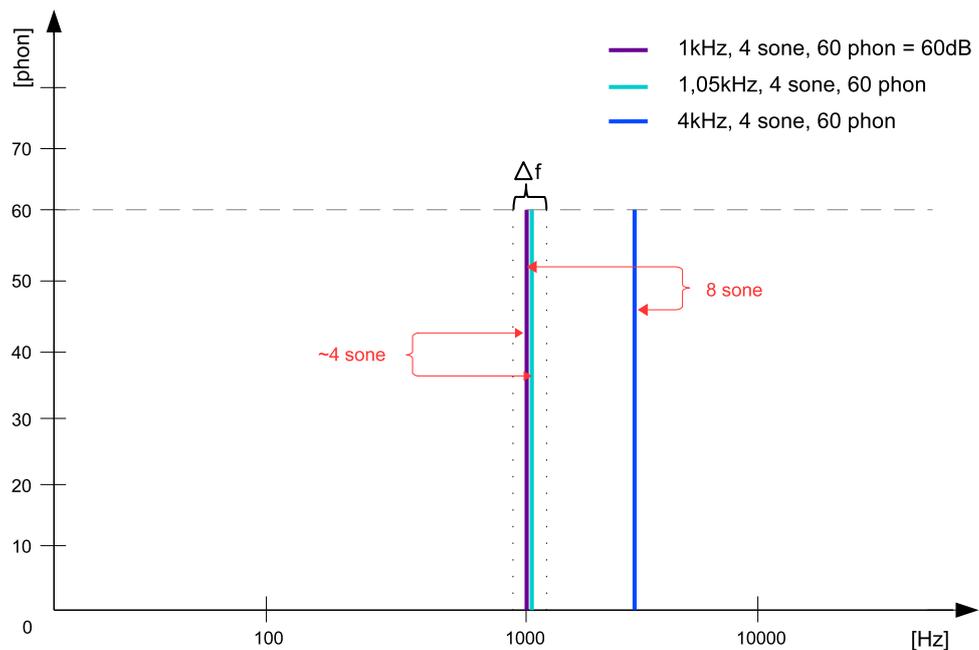


Abbildung 3.3.: Schematische Darstellung der Lautheitsempfindung

Frequenzbereiche unterschiedlich empfindlich wahrzunehmen, so gelangt man schnell zur subjektiven Lautstärkewahrnehmung - der Lautheit eines Tones beschrieben in sone. Als Referenzwert und Verbindung zwischen dB, phon und sone gilt hier ein Sinuston bei 1kHz

mit 40dB, das entspricht 40phon, was wiederum einem sone entspricht. Die Lautheitsskala ist nicht linear aufgebaut. So entsprechen 50phon 2sone, 60phon 4sone und 70phon 8sone etc. Hört man also gleichzeitig 2 Töne die jeweils 60phon, also je 4sone laut sind, so geht man davon aus, dass man sie gemeinsam als doppelt so laut - also 8sone - empfinden würde. Dies gilt jedoch nur solange die beiden Frequenzen nicht innerhalb der selben Frequenzgruppe liegen. Berücksichtigt man jetzt die gruppierte Verarbeitung des Gehörs, so ergibt sich für zwei Töne, die sich innerhalb einer Frequenzgruppe befinden, nicht die doppelte Lautheit (8sone), sondern die Gesamtlautheit erhöht sich marginal (vgl. [6]). Wie in Abbildung 3.3 dargestellt, trägt der Sinuston bei 1.05kHz nur als in die Frequenzgruppe um 1kHz integrierte Komponente, zur Gesamtlautheit bei. Der Pegel in dB erhöht sich bei diesem Beispiel von 60dB auf 63dB - was in Lautheit etwas über 4 sone bedeuten würde. Der 2.Ton beeinflusst also nicht die Lautstärke, sondern primär die Klangfarbe des Schallereignisses. Für den bei 4kHz dargestellten Ton gilt, bei gleichzeitigem Erklingen mit dem Ton bei 1kHz, eine Verdoppelung der Lautheit (also 2x4sone), da die beiden Töne nicht innerhalb der selben Frequenzgruppe liegen. Dies wiederum schafft die Verbindung zur Bark-Skala, die eben diese gruppierte Verarbeitung als Grundlage nimmt. Ein Bark ist also der Bereich, in dem das Ohr die eingehenden Frequenzen als Frequenzgruppe wahrnimmt, was in weiterer Folge eine gemeinsame Verarbeitung der Frequenzen innerhalb dieser Gruppe bedeutet (siehe Kapitel 3.1.2). Liegen Teile des Signals außerhalb der Frequenzgruppe, werden sie zu ihren eigenen Frequenzgruppen zugeordnet und dort verarbeitet. Im letzten Schritt werden alle Frequenzgruppen zu einem Gesamteindruck zusammengeführt. Normiert wird die Barkskala auf den musikalischen Ton C bei 131Hz. Es gilt:  $1,31\text{Bark} = 131\text{Hz}$ . Der Zusammenhang zwischen Frequenz und Bark ergibt sich unter 500Hz als nahezu linear, bei Frequenzen über 500Hz ist der Zusammenhang besser als ein Logarithmischer zu beschreiben. Formel 3.1 nach H. Traunmüller (vgl. [5]) zeigt den Zusammenhang zwischen Barkfrequenz und linearer Frequenz.

$$z = \frac{26.81}{1 + 1960/f} - 0.53 \quad (3.1)$$

Um eben diese Frequenzgruppen nutzen zu können, wurde in der Projektarbeit jedes Klangbeispiel in einzelne Barkbänder zerlegt. Das unterste Band wurde so gewählt, dass es eine untere Grenzfrequenz von 35Hz hatte. Beginnend bei 35Hz wurden 21 Barkbänder errechnet.

Die Bark-Skala wird hierbei genutzt um verdeckte Signalkomponenten zu eliminieren, und um in weiteren Schritten (siehe Kapitel 3.2.1) transiente Einschwingvorgänge nach dem Weißmachen durch inverse Filterung mit aus der LPC erhaltenen Koeffizienten (Residuum), in eine gehörgerechte Klangfarbe zu wandeln. Um die Maskierung anwenden zu können, wurde zuerst das Betragsquadrat der Einhüllenden jedes Bandes gebildet. Aus den Einhüllenden wurde für jedes Band eine Ausklingzeit T30 (die Zeit in der das Signal um 30dB abgefallen ist), sowie eine Einschwingzeit ermittelt.

### 3. BZ-ARMA Bark Zooming Autoregressive Moving-Average

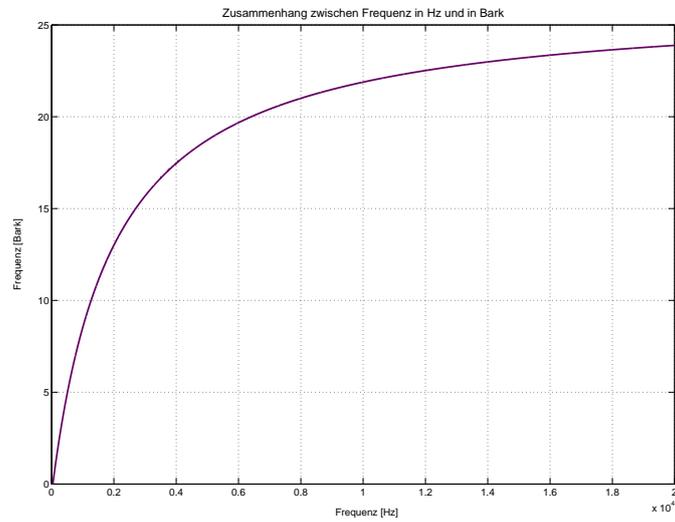


Abbildung 3.4.: Zusammenhang zwischen linearer Frequenz  $f$  und Barkfrequenz  $z$

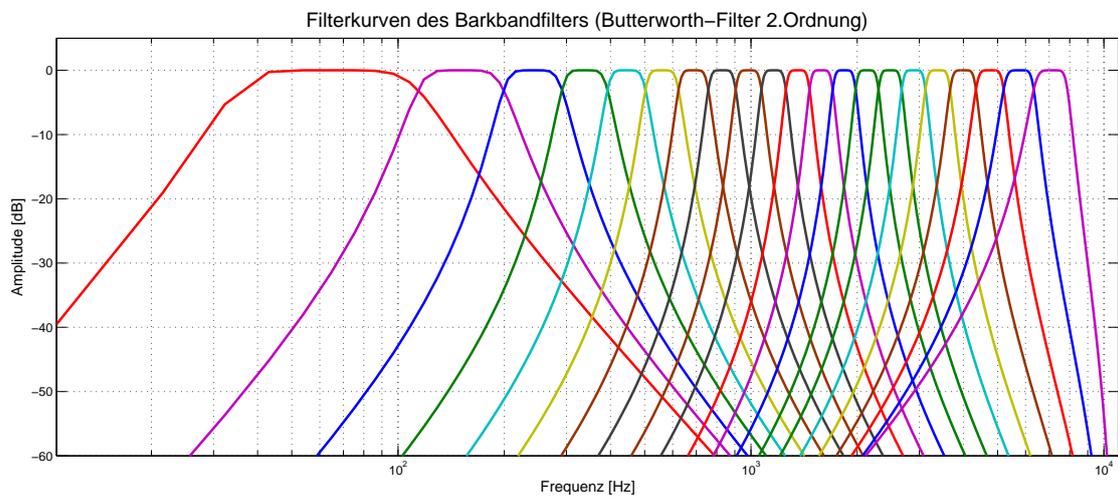


Abbildung 3.5.: Barkfilterkurven

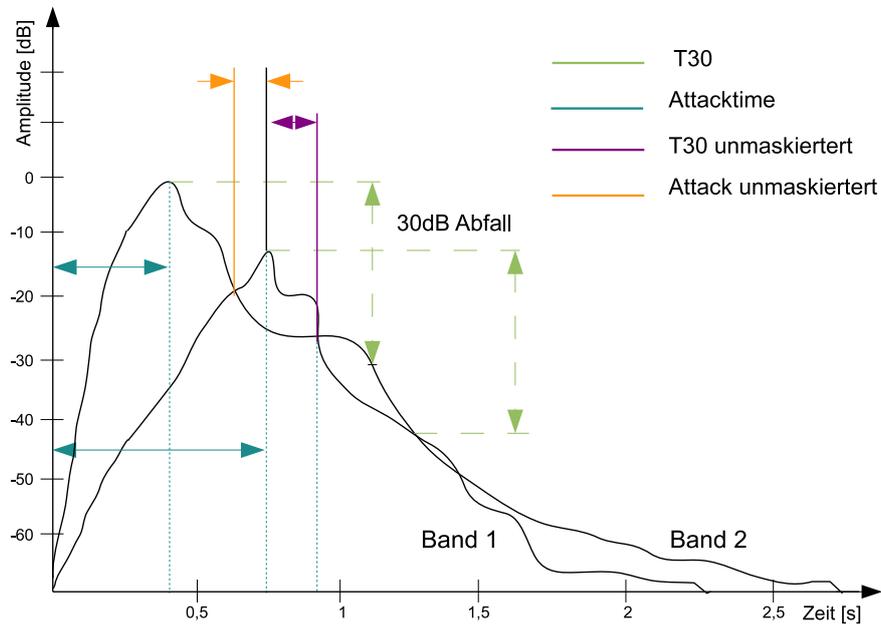


Abbildung 3.6.: Schematische Darstellung der Maskierung von T30 und Einschwingzeit

Ausgehend von den Einhüllenden wurden die durch Maskierungsflanken für das Gehör verdeckten Bänder eliminiert. Dabei wurde in diesem Fall die Maskierung ausgehend von den Barkbandkanten jeden halben Bandes errechnet, nachdem letztere mit einer Sinus-Halbwellen geglättet wurden. Wie in Abbildung 3.2 dargestellt, sind alle Bänder, die unter der Maskierungsschwelle liegen, für das menschliche Ohr maskiert.

Im nächsten Schritt werden alle Bänder bezüglich ihrer Ausklingzeit miteinander verglichen. Errechnet aus der längeren Ausklingzeit eines Bandes mit seinem Benachbarten ergibt sich eine Zeitspanne, in der das jeweilige Band unmaskiert ist, die über oder unter einem voreingestellten Wert für die Simultanmaskierung (0,2 Sekunden) liegen kann. Ist das Band länger maskiert als der vorgegebene Wert, so wird es allgemein als maskiert behandelt und verworfen. Abbildung 3.6 stellt schematisch diesen Prozess der Maskierung über vorgegebene Zeitwerte dar. Band 1 ist während der gesamten Einschwingzeit und den Großteil der Ausklingzeit T30 unmaskiert. Band 2 wird während der Einschwing- und der Ausklingzeit jeweils länger als 0,2 Sekunden maskiert und somit für alle Zeitpunkte als maskiert behandelt, unabhängig vom weiteren zeitlichen Verlauf.

### 3.1.4. Modulation und Downsampling

Basierend auf der Tatsache, dass jedes Barkband sich in einem anderen Frequenzbereich bewegt und somit andere Bandgrenzen aufweist, war es nach Modulation eines jeden

Bandes gegen Null möglich, ein Downsampling mit unterschiedlichen Faktoren, abhängig von der tatsächlichen Bandbreite, vorzunehmen.

## 3.2. Signalanalyse

Wir betrachten hier die Annahme, dass Instrumentalklänge, welche durch eine impuls-hafte Anregung zum Klingen gebracht werden (z.B. Idiophone), als Quelle-Filter-Modell beschrieben werden können. Dieser Ansatz ist für Klänge ineffizient, welche *frequency-glides* aufweisen - sich also Teiltöne des Klanges in ihrer Frequenz verstimmen. Signifikante Datenreduktion kann insbesondere dann erreicht werden, wenn der Instrumentalklang stationäre, ausklingende Teiltöne besitzt, welche durch eine äußerst kurze Anregung zum Klingen gebracht werden. An dieser Stelle wird angenommen, dass dieses Modell für die betrachteten Klänge idiophoner Gamelan-Instrumente gut geeignet ist.

### 3.2.1. LPC

Der Ansatz des Linear Predictive Coding besitzt eine Struktur, in welcher ein spektral ungefärbtes Signal (weiß) ein Resonanzfilter anregt. Davon ist nur das Ausgangssignal  $s[n]$  bekannt. Als erste Aufgabe gilt es nun, aus den vergangenen  $N - 1$  Samples von  $s[n]$  das aktuelle Sample zu schätzen. Der Fehler wird in diesem Ansatz der Schätzung beschrieben als:

$$e[n] = \underbrace{s[n]}_{\text{bekanntes Signal}} - \underbrace{\sum_{m=-N+1}^0 s[n+m-1]a_m}_{\text{Linearkombination unbekannter Koeffizienten } a_m} \quad (3.2)$$

Um daraus ein Least-Squares (Minimierung der Fehlerquadrate) Problem formen zu können, benötigt man:

$$|e[n]|^2 \rightarrow \min. \quad (3.3)$$

Da es mathematisch sinnvoll ist, darf davor der Ausdruck  $\sum_{m=-N+1}^0 s[n+m-1]a_m$  als Vektor geschrieben werden:

$$\sum_{m=-N+1}^0 s[n+m-1]a_m = \vec{s}^t[n-1]\vec{a}, \quad (3.4)$$

Wobei man von der Definition ausgeht:

$$\vec{s}[n] = \begin{pmatrix} s[n] \\ \vdots \\ s[n-N+1] \end{pmatrix}, \quad (3.5)$$

$$\frac{\partial}{\partial \vec{a}} |s[n] - \vec{s}^t[n-1]\vec{a}|^2 \stackrel{!}{=} 0 \quad (3.6)$$

$$2(s[n] - \vec{s}^t[n-1]\vec{a})\vec{a} = 0 \quad (3.7)$$

$$\vec{s}^t[n-1]\vec{a} = s[n] \quad (3.8)$$

$$\Rightarrow \vec{a}[n] = (\vec{s}^t[n-1]\vec{s}^t[n-1])^{-1}\vec{s}^t[n-1]s[n]. \quad (3.9)$$

Um die Koeffizienten zu finden, die im Mittel über einen längeren Zeitraum  $m = n + (-Na + 1, \dots, 0)$  gelten, wird erweitert auf:

$$\vec{e}[n] = \begin{pmatrix} e[n+0] \\ e[n-1] \\ \vdots \\ e[n-N_a+1] \end{pmatrix} \quad (3.10)$$

$$\vec{e}[n] = \begin{pmatrix} e[n+0] \\ e[n-1] \\ \vdots \\ e[n-N_a+1] \end{pmatrix} - \begin{pmatrix} \vec{s}^t[n-1] \\ \vec{s}^t[n-2] \\ \vdots \\ \vec{s}^t[n-N_a] \end{pmatrix} \vec{a} \quad (3.11)$$

$$\|e[n]\|^2 \xrightarrow{\Rightarrow} \min \vec{a}_{opt}[n] = (S[n-1]S^t[n-1])^{-1}S[n-1] \begin{pmatrix} s[n] \\ s[n-1] \\ \vdots \\ s[n-N_a+1] \end{pmatrix} \quad (3.12)$$

$$\approx R_{ss}^{-1}r_{ss} \quad (3.13)$$

Mit dem Prediktorfehlerfilter

$$r[n] = e[n] = s[n] - \sum_{m=-N+1}^0 a_m s[n+m-1] = \sum_{m=-N}^0 b_m s[n+m], \quad (3.14)$$

also den Koeffizienten  $b_0 = 1$  und  $b_n = -a_n$ , erhält man das Fehlersignal, das Residuum. Nimmt man das dazu inverse Filter, so ergibt sich ein Filtermodell, das die Resonanzen des Klangs nachbildet.

Innerhalb dieser Arbeit wurde das Linear Predictive Coding nicht mit dem vollständigen, sondern mit dem um den Anschlag verkürzten Signal durchgeführt. Dadurch sind

### 3. BZ-ARMA Bark Zooming Autoregressive Moving-Average

---

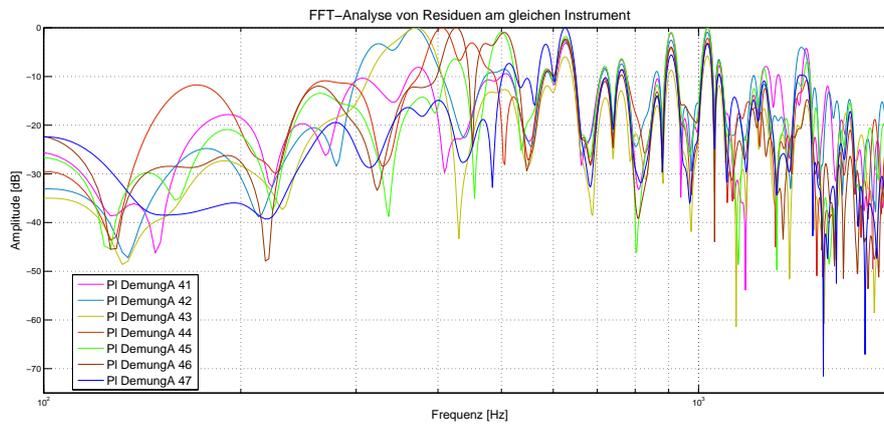


Abbildung 3.7.: Vergleich mehrerer Residuen eines Saron Demung Instrumentes

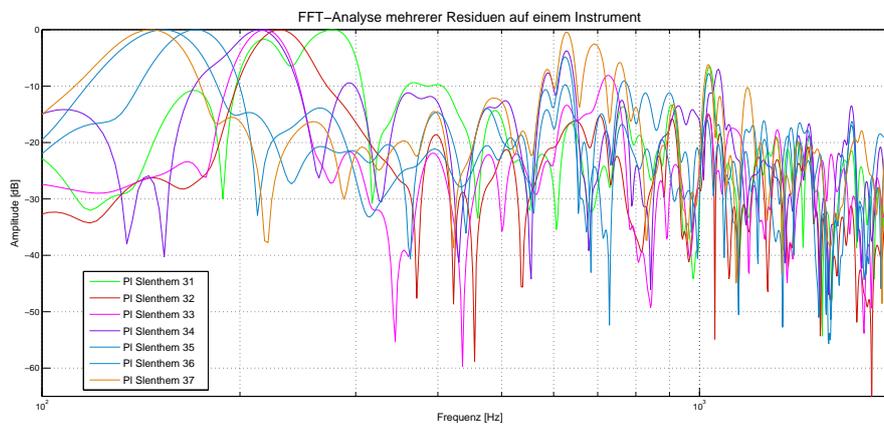


Abbildung 3.8.: Vergleich mehrerer Residuen eines Slenthem Instrumentes

die Residuen nicht weiß gefärbt, wie es normalerweise sein sollte. Sie spiegeln die Klangfarbe des Anschlags wieder, die sich aus den Residuen-Teilsignalen aller 21 Barkbänder zusammensetzt.

In Abbildung 3.7 und Abbildung 3.8 ist ersichtlich, dass die Residuen eine sehr hohe Teiltondichte besitzen und pro Instrument jeweils ähnliche Frequenzkomponenten aufweisen. Grundsätzlich ist bei beiden Abbildungen eine globale Struktur erkennbar, die sich aus ähnlichen Pegelbereichen bei unterschiedlichen Frequenzen, und das bei verschiedenen Instrumentparts, ergibt. Weiters ist jedoch eine innere, lokale Struktur vorhanden, deren Verlauf sich aus der Minima- und Maximaverteilung des Spektrums wesentlich macht. Im Optimalfall wäre bei gleicher Anschlagtechnik eine absolute Übereinstimmung der einzelnen Residuen auf einem Instrument zu erwarten. Eine annähernde Übereinstimmung findet man in Abbildung 3.7 zwischen 600Hz und etwas knapp 2kHz. Die weniger starke Übereinstimmung dieser Residuengruppe unter 500Hz ist jedoch unwesentlich, da der Pegel der globalen Struktur hier bereits um durchschnittlich 20dB unter dem Pegel des gut übereinstimmenden Bereiches liegt, der wiederum in einem Frequenzbereich liegt in dem das Ohr eine erhöhte Empfindlichkeit aufweist (siehe Kapitel 3.1.1). Auch die in Abbildung 3.8 dargestellte Residuengruppe zeigt eine schlechtere Kongruenz im tieferfrequenten Bereich, in dem auch das menschliche Gehör um einiges unempfindlicher ist. Die Kongruenz erhöht sich ab etwa 800Hz und hält im höheren Maße bis gegen 2kHz an - ein Bereich, der vom Ohr wiederum empfindlicher wahrgenommen wird (vgl. [6] Kapitel 8.1.).

## 4. Analyse-Programm

### 4.1. Das zu bearbeitende Material

Als Input dienten Monosignale mit 16Bit Wortbreite und einer Abtastrate von 44100Hz. Alle Soundfiles waren PCM-Codiert. Die Länge der Rohdaten bewegte sich zwischen einer Sekunde bei den kürzesten, impulshaften Trommelklängen und 35 Sekunden bei den längsten (primär gongartigen) Klängen.

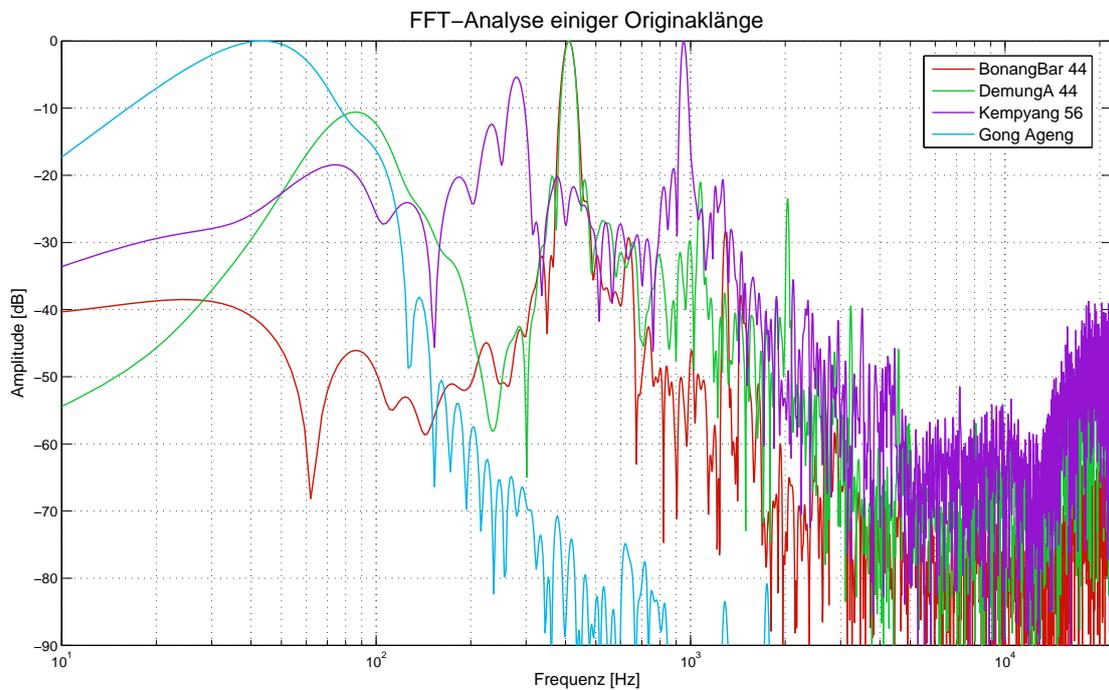


Abbildung 4.1.: FFT-Spektren mehrerer Originaklangbeispiele

In Abbildung 4.1 werden mehrere FFT-Spektren unterschiedlicher Gamelanklänge dargestellt, um einen Eindruck des Ausgangsmaterials zu bekommen.

## 4.2. Die Berechnungsschritte des Programms

Das Analyse-Programm ist darauf ausgelegt, eine möglichst genaue, aufwandsreduzierte Berechnung mit 22 Polen durchzuführen. Es wurden diverse Maßnahmen zur Datenreduktion getroffen, um den Berechnungsaufwand im Rahmen zu halten.

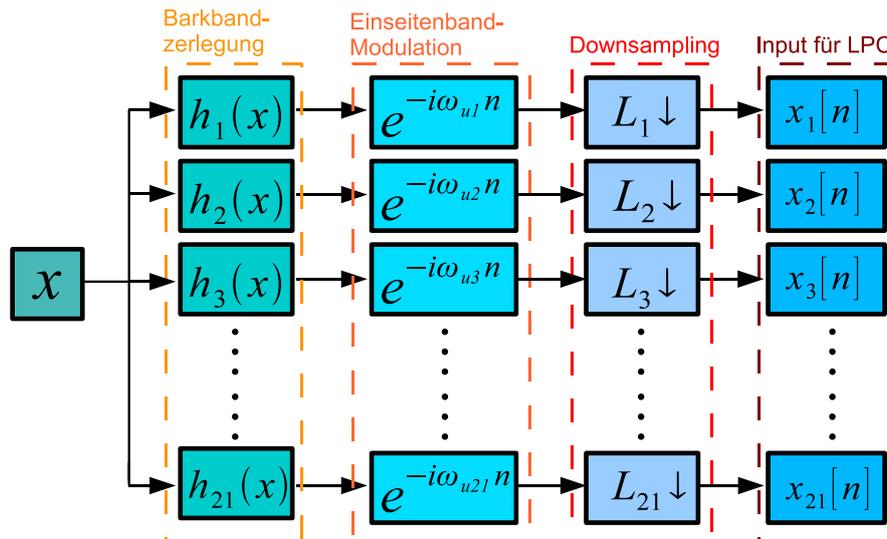


Abbildung 4.2.: Analyse-Filterbank

### 4.2.1. Inputbegrenzung

Um die Anzahl der zu verarbeitenden Samples zu senken, wurde das Eingangssignal auf maximal elf Sekunden begrenzt. Hierbei wurde von der Tatsache ausgegangen, dass es genügt, die Zeit zur Berechnung heranzuziehen, die das Signal benötigt, um 30dB unter den Maximalwert zu fallen ( $T_{30}$ ). Diese Zeit ist in Anbetracht der zu erwartenden, exponentiell verlaufenden Signalformen in jedem Fall genug, um aus ihr die Ausklingkonstante  $\tau$  zu bestimmen.

$$\tau = \frac{T_{30}}{3 \ln(10)} \quad (4.1)$$

Davon ausgehend, dass also nach dem ersten Maximum des Signals, aufgrund der Signalanregung mit einmaligen Anschlag, keine weiteren Maxima mehr zu erwarten sind, ist es kein Problem das Eingangssignal auf den gewünschten Wert von elf Sekunden zu reduzieren. Diese elf Sekunden sollten für alle erwarteten Eingangssignale genug Information enthalten um mehr als nur die  $T_{30}$  zu extrahieren. An Datenaufwand würde das

44100 \* 11 = 485100 Samples pro Input bedeuten, was den Aufwand der Berechnung im Vergleich zur Ausgangssituation erheblich verringert.

### 4.2.2. Downsampling

Da sich die für den Klangcharakter wesentlichen Frequenzanteile unter 8kHz befinden, konnte von einer Samplerate von 44100Hz auf 22050Hz downgesamlet werden, ohne wichtige Informationen für die Analyse zu verlieren. Vor dem Downsamplen wurde das Signal dem standardmäßigen Verfahren einer Antialiasing-Filterung unterzogen.

### 4.2.3. Gewichtung mit A-Filter

Wie schon in 3.1.1 erklärt, wird in diesem Schritt das bereits um den Faktor 2 downgesamlete Eingangssignal, mit einer bewerteten A-Filterkurve an die für das menschliche Gehör übliche Frequenzgewichtung angepasst. Dieser Bearbeitungsschritt setzt voraus, dass das fertig analysierte Signal bei der Resynthese wieder mit der selben, aber inversen bewerteten A-Filterkurve entzerrt wird, um die gehörspezifische Frequenzgewichtung wieder auszuklammern.

### 4.2.4. Barkbandfilterung

Aus den bereits in 3.1.3 erläuterten Gründen war es notwendig eine Barkfilterbank zu erstellen. Dies wurde in diesem Fall über ein Butterworth-Filter 2. Ordnung implementiert (siehe Abbildung 3.5) das, um die Phase des gefilterten Signals stabil zu halten, einmal vorwärts und einmal rückwärts filtert. Dadurch konnte die Flankensteilheit von der eines Filters 2. Ordnung auf die eines Filters 4. Ordnung erhöht werden.

Abbildung 4.3 stellt für alle errechneten Barkbänder die zugehörigen Filterkurven nach der Vorwärtsfilterung (blaue Kurven) und nach der Rückwärtsfilterung (pinke Kurven) dar. Die ursprüngliche Idee, die einzelnen Barkbänder noch schärfer heraus zu filtern, war nicht umsetzbar, da sich im tieffrequenten Bereich bereits bei 4. Ordnung des Filters eine gewisse Welligkeit einstellte (siehe Abbildung 4.4), die sich beim zweiten Durchlaufen des Filters natürlich um ein Vielfaches erhöht (bis zu 12dB).

### 4.2.5. Elimination maskierter Bänder und Signalanteile

Wie bereits in Kapitel 3.1.3 beschrieben, werden maskierte Barkbänder und Signalkomponenten so eliminiert, dass für die tatsächliche Detektion nur noch die wirklich hörrelevanten Anteile übrig bleiben. Aus diesen errechneten Anteilen wird anschließend eine LPC-Analyse (siehe Kapitel 3.2.1) mit 22 Polen pro eruiertem Barkband durchgeführt.

Dabei werden gleichzeitig die Koeffizienten für ein Prediktorfehlerfilter ermittelt

## 4. Analyse-Programm

---

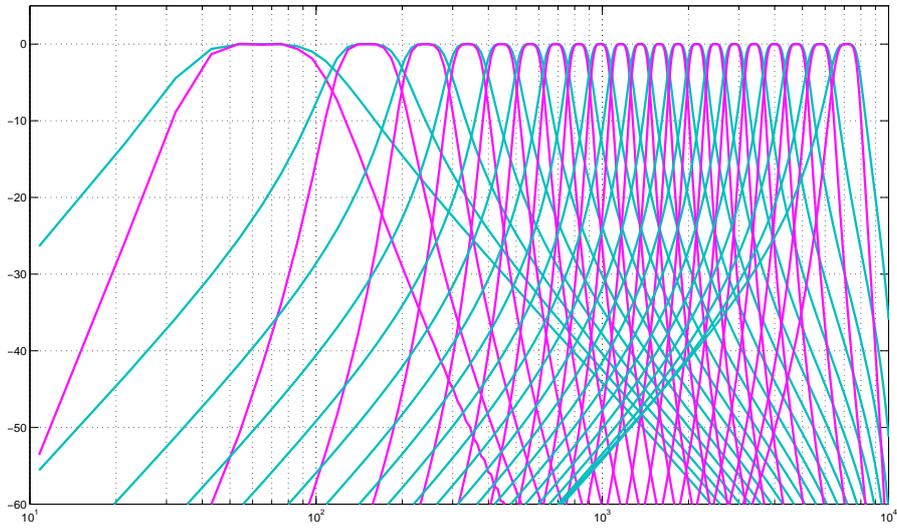


Abbildung 4.3.: Barkfilterkurven 2. (blau) und 4. (pink) Ordnung

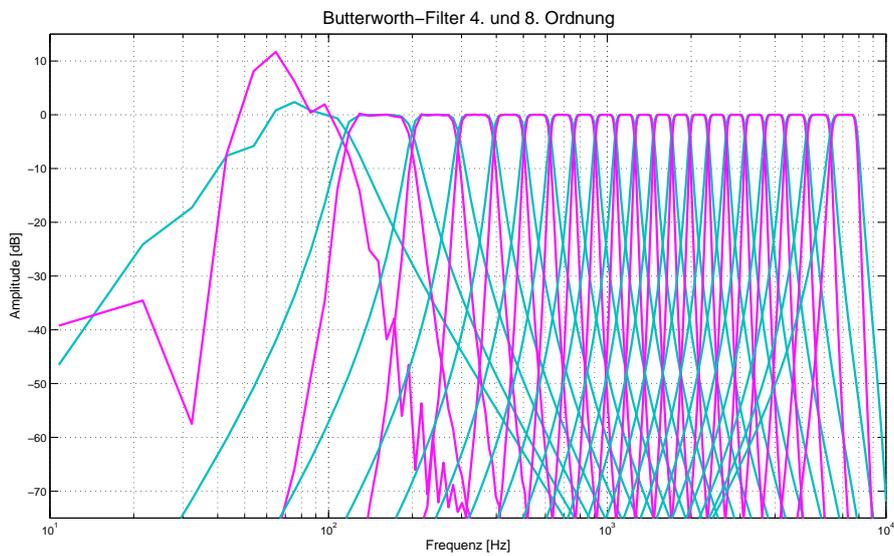


Abbildung 4.4.: Barkfilterkurven 4. (blau) und 8. (pink) Ordnung

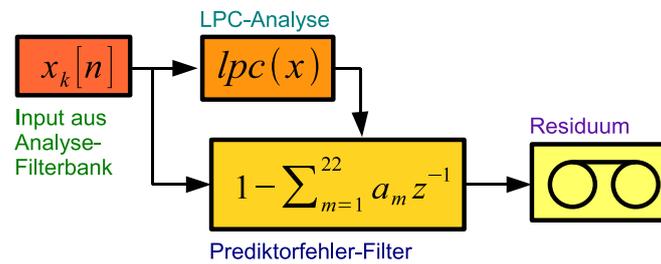


Abbildung 4.5.: Blockdiagramm der LPC-Analyse

Die innerhalb der LPC-Analyse errechneten Pole dienen als Grundlage für die Errechnung der Startamplituden, Startphasen, Einschwingzeiten und Ausklingkonstanten zu den zugehörigen Polfrequenzen.

## 5. Resynthese

### 5.1. Klangmodellierung mit Residuum

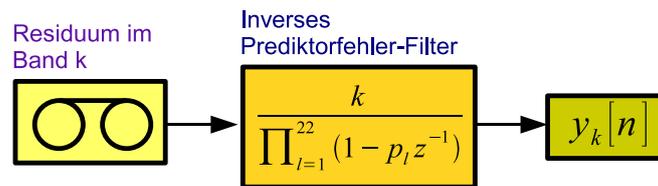


Abbildung 5.1.: Synthese mit Residuum

Ausgehend vom bereits erwähnten Residuum (Restsignal, Anregungssignal des Resonanzmodells) kann mit Hilfe eines inversen Prädiktorfehlerfilters das Originalsignal sehr einfach wieder berechnet werden. Dazu ist es nötig folgendes Resonanzfilter auf das Residuum anzuwenden:

$$H(z) = \frac{1}{1 - \sum_{m=0}^N (1 - a_m z^{-1})} = \frac{k}{\prod_{m=1}^N (1 - p_m) z^{-1}} \quad (5.1)$$

Wie oben bereits angedeutet, kann das Filtermodell als Verkettung mehrerer Resonanzfilter mit komplexwertiger Polstelle (dargestellt durch das Produkt im Nenner) umgesetzt werden.

Für die Residuen soll im Allgemeinen gelten, dass Sie für die gleiche Anschlagtechnik annähernd gleiche Frequenzen enthalten und somit nahezu gleich klingen.

### 5.2. Klangmodellierung mit Decaying-Exponentials

Das Filtermodell des inversen Prädiktorfehlerfilters aus dem LPC-Ansatz kann auch verwendet werden, um die zu den ermittelten Polstellen zugehörigen Resonanzfrequenzen und Ausklingkonstanten zu bestimmen. Dazu wird das Polstellenpolynom wieder in Linearfaktoren aufgespalten, so dass jene komplexwertigen Polstellen  $p_m = r_m e^{i\Theta_m}$  sichtbar werden, die Ausklingkonstanten  $r_m$  und Resonanzfrequenzen  $\Theta_m$  aufweisen.

$$H(z) = \frac{1}{1 - \sum_{m=0}^N (1 - a_m z^{-1})} = \frac{k}{\prod_{m=1}^N (1 - p_m) z^{-1}} \quad (5.2)$$

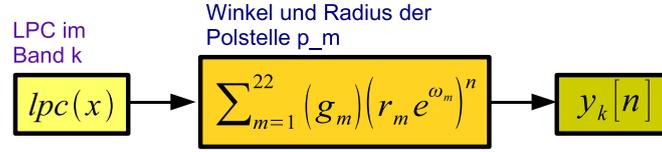


Abbildung 5.2.: Analyse mit Decaying Exponentials

Einer einzelnen Polstelle  $p_m$  kann dabei eine ausklingende Exponentialschwingung mit Startamplitude 1 und Startphase 0 bei  $n = 0$  zugeordnet werden:

$$u_m[n] = p_m^n = r_m^n e^{i\Theta_m^n} \quad (5.3)$$

Um Phase und Betrag mit den im Signal  $s[n]$  vorhandenen Anfangswerten abgleichen zu können, muss nun folgendes Gleichungssystem mit komplexwertigen Gewichten  $\vec{g}$  aufgestellt und für einen Zeitraum  $N_a$  gelöst werden:

$$\begin{pmatrix} s[0] \\ s[1] \\ \vdots \\ s[N_a - 1] \end{pmatrix} = \underbrace{\begin{pmatrix} u_1[0] & u_2[0] & \cdots & u_N[0] \\ u_1[1] & u_2[1] & \cdots & u_N[1] \\ \vdots & \vdots & \vdots & \vdots \\ u_1[N_a - 1] & u_2[N_a - 1] & \cdots & u_N[N_a - 1] \end{pmatrix}}_U \underbrace{\begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_N \end{pmatrix}}_{\vec{g}} \quad (5.4)$$

$$\Rightarrow \vec{g} = \text{pinv}(U) \begin{pmatrix} s[0] \\ s[1] \\ \vdots \\ s[N_a - 1] \end{pmatrix} \quad (5.5)$$

Die komplexwertigen Gewichte  $\vec{g}$  beinhalten sowohl Startphase, als auch Startamplitude. Das nachgebildete Signal ist hier - natürlich ähnlich wie oben - skalar für alle Zeitpunkte  $n \geq 0$  angeschrieben:

$$\vec{s}[n] = \sum_{m=1}^N (r_m e^{i\Theta_m})^n \quad (5.6)$$

Für dieses Syntheseverfahren erzeugt ein Sinusoszillator eine gegebene Frequenz mit Phase, linearer Hüllkurve für den Einschwingvorgang, Amplitudenwert am Maximum und einer exponentiellen Hüllkurve für den Ausklingvorgang, für jede detektierte unmaskierte Polfrequenz.

Die Abbildungen 5.3, 5.4, 5.5 und 5.6 zeigen die in der Analyse detektierten Teiltöne des Signals mit jeweiliger Startamplitude und Ausklingkonstante. Dabei liegt das Hauptaugenmerk bei den dreidimensionalen Plots auf den Verhältnissen der Teiltöne in

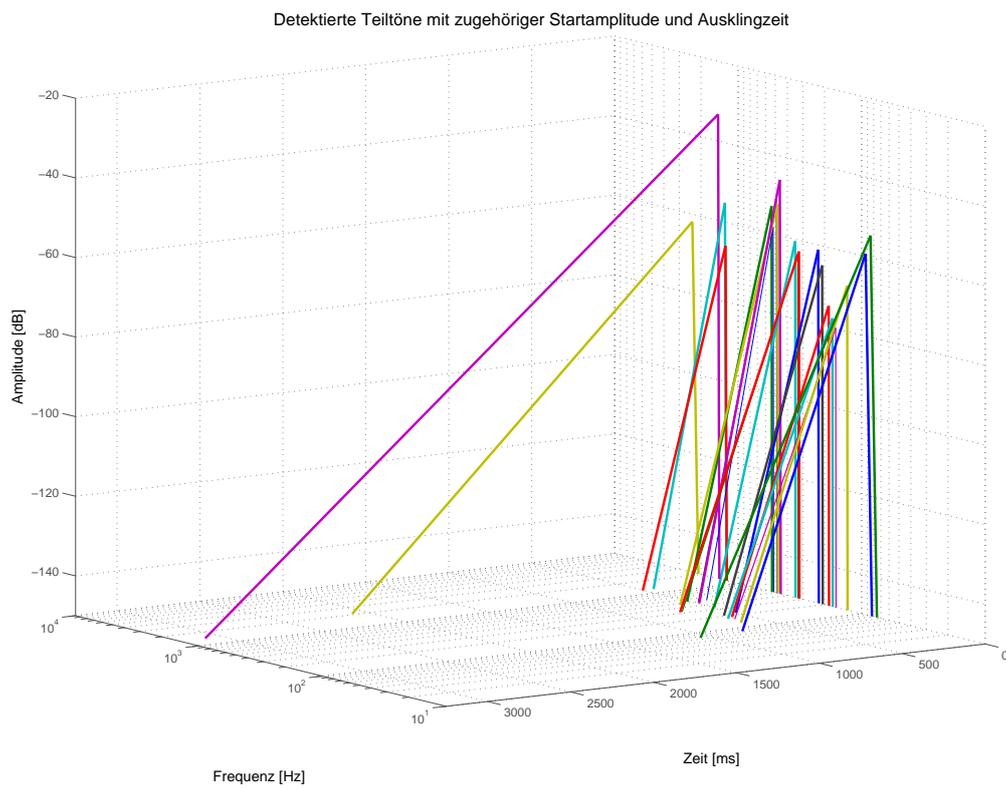


Abbildung 5.3.: Zur Synthese herangezogene Teiltöne eines Bonang Panerus mit zugehöriger Startamplitude und Ausklingzeit

## 5. Resynthese

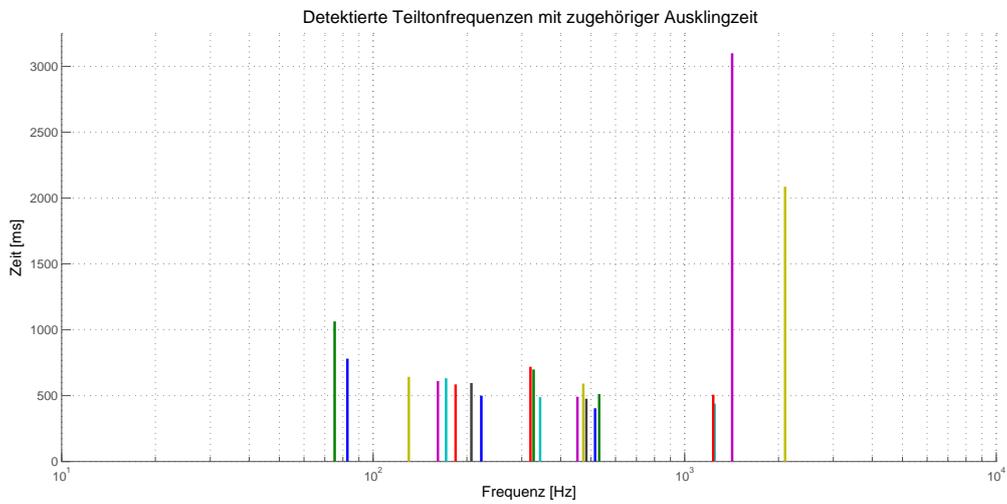


Abbildung 5.4.: Teiltonfrequenzen und zugehörige Ausklingzeiten eines Bonang Panerus

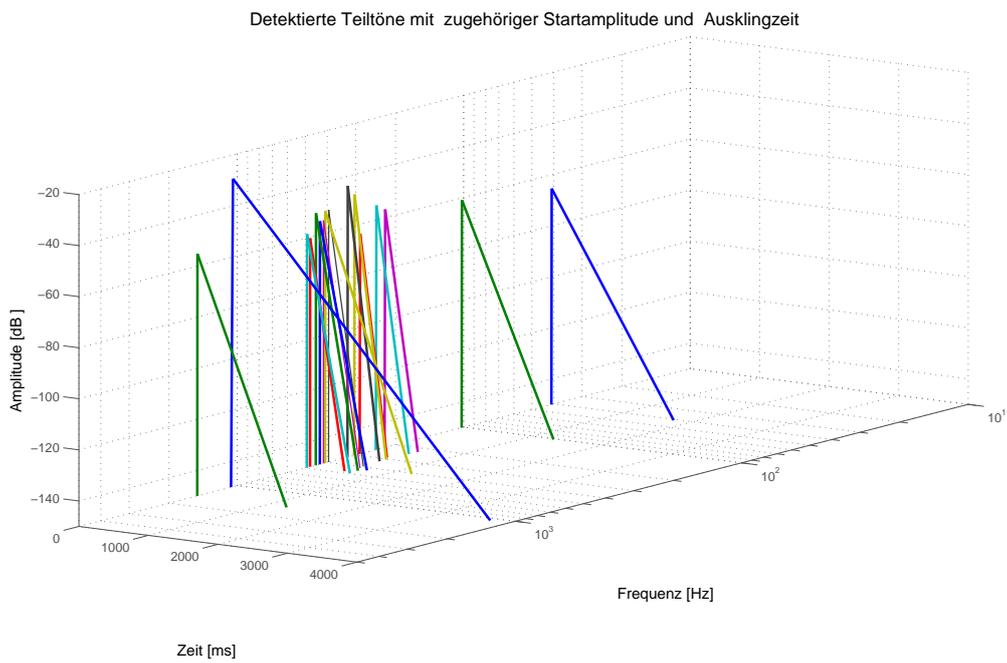


Abbildung 5.5.: Zur Synthese herangezogene Teiltöne eines Bonang Barung mit zugehöriger Startamplitude und Ausklingzeit

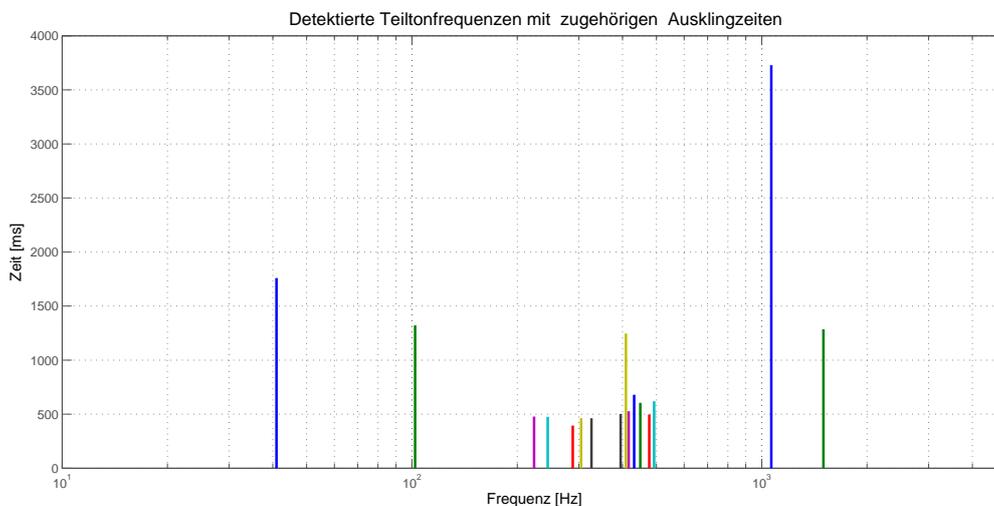


Abbildung 5.6.: Teiltonfrequenzen und zugehörige Ausklingzeiten eines Bonang Barung

Amplitude und Ausklingvorgang zueinander. Die zweidimensionalen Plots zeigen eine genauere Darstellung der Ausklingzeit bei den einzelnen detektierten Frequenzen.

Anhand der ausgewählten Bilder ist erkennbar, dass sich meist ein oder zwei stark ausgeprägte Komponenten als Hauptträger des Klanges (Hauptkomponenten) manifestieren, während sich viele kleinere Nebenkomponten mit kürzerer Ausklingzeit als Klangfarben gebende Elemente hinzugesellen. Die richtige Analyse jener Hauptkomponenten als Tonhöhenvorgabe bzw. Grundtonhöhe des resynthetisierten Klanges, ist enorm wichtig als erstes Erkennungsmerkmal beim akustischen Vergleich des Originalklanges mit seinem resynthetisierten Adäquat. Wie später in Kapitel 6 in den Grafiken 6.1, 6.2, 6.3, 6.4 und 6.5 auffällt, sind die FFTs des Originals und der Resynthese bei weitem nicht hundertprozentig übereinstimmend. Sieht man aber genauer hin, gibt es bei jedem Spektrenpaar eine starke Similarität bei den für den jeweiligen Klang charakteristischen Frequenzspitzen. Die Nebenkomponten des Klanges sind jedoch oft gar nicht, in gedämpfter Form, oder sogar leicht verschoben vorhanden. Trotzdem wird der resynthetisierte Klang vom Ohr dem Original sehr ähnlich empfunden.

## 5.3. Synthese-Filterbank

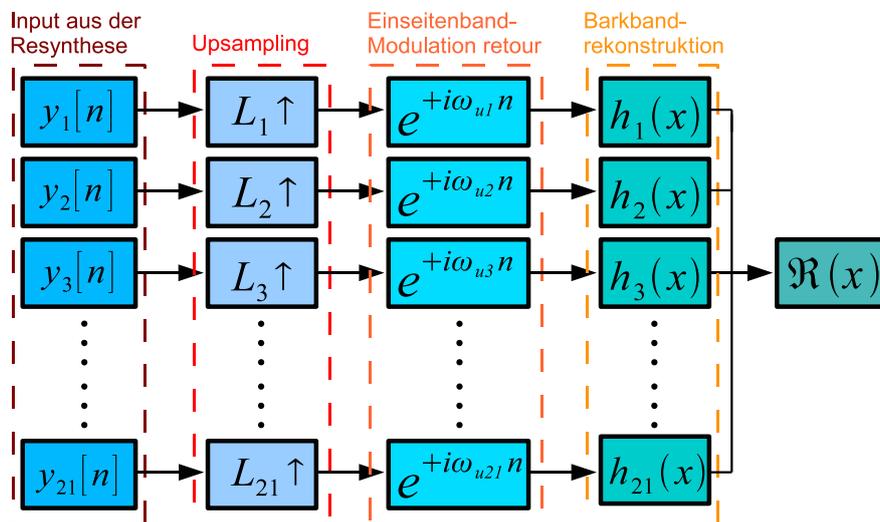


Abbildung 5.7.: Synthese-Filterbank

Um nach den, zuvor in Kapitel 5 beschriebenen, Resynthese-Verfahren jedes 21-Band Signal wieder in ein Vollband-Signal überzuführen, muss es alle in Abbildung 5.7 gezeigten Schritte durchlaufen. Die 21 separaten Eingangssignale der Synthese-Filterbank werden auf die ursprüngliche Samplingrate zurückgebracht. Dabei werden die selben Faktoren zum Upsamplen verwendet, die zuvor als Downsamplingfaktoren ermittelt wurden, weshalb jedes Band mit seinem individuellen Faktor upgesamplet wird. Im nächsten Schritt werden die auf die ursprüngliche Abtastrate gebrachten Signal-Teilbänder mit ihrer, schon beim Demodulieren verwendeten, speziellen Modulationsfrequenz  $\omega_{uk}$ , wieder aus dem Basisband durch Einseitenband-Modulation in ihr eigentliches Frequenzband zurück moduliert. Dort angekommen wird jedes Band nochmals mit dem Barkbandfilter (siehe Abb. 3.5) gefiltert um wirklich nur die bandeigenen Anteile in das Syntheseergebnis miteinzubeziehen. Anschließend werden die Realteile aller gefilterten Bänder bei der jeweiligen Originalfrequenz aufsummiert um das Synthese-Endergebnis zu bilden.

Die Teiltöne werden generiert (siehe Abbildung 5.8) indem eine Bank von Sinusoszillatoren mit den Parametern Frequenz und Amplitude gespeist werden, die entsprechende Sinustöne produzieren, deren Einhüllende im nächsten Block, vorgegeben durch die Parameter  $T\_attack$  (Einschwingzeit) und  $T\_decay$  (Ausklingszeit), auf die Sinustöne aufgeprägt werden. Alle so generierten Frequenzkomponenten gemeinsam ergeben den fertigen Klang.

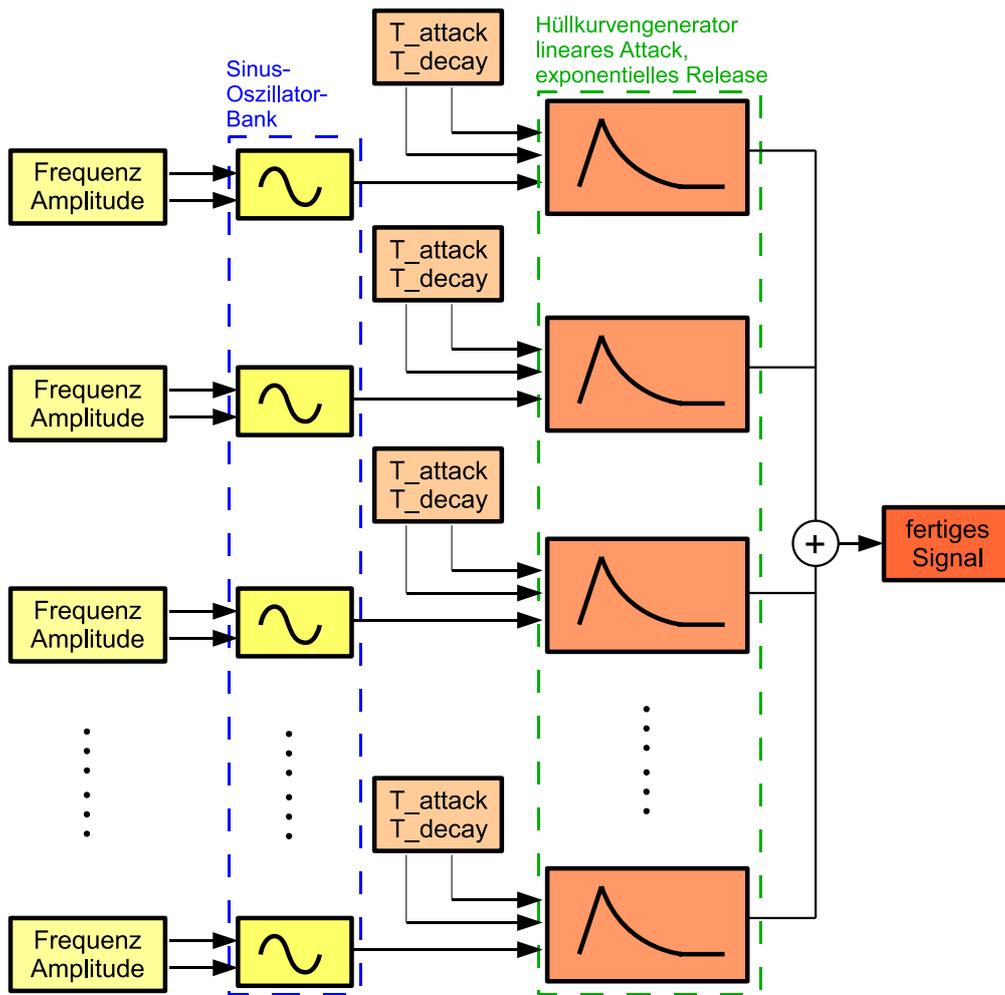


Abbildung 5.8.: Blockdiagramm der Teiltongeneration

## 6. Analyse-Beispiele

In diesem Kapitel werden einige Klanganalysen als Beispiel und Repräsentation der gesamten Arbeit vorgelegt.

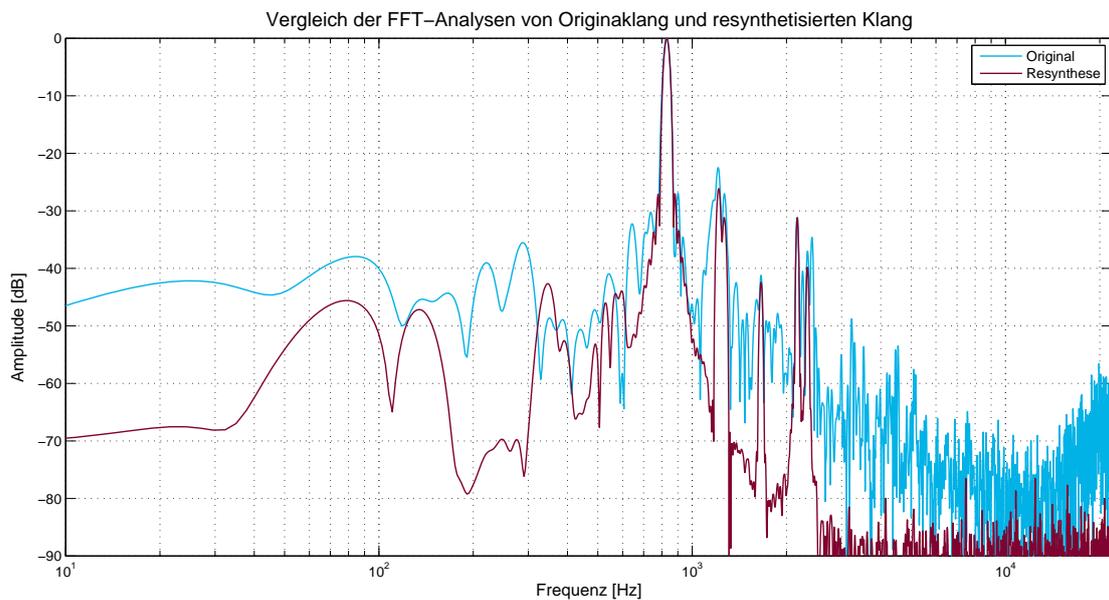


Abbildung 6.1.: Vergleich eines Originalklanges des Instruments Bonang Panerus mit seinem resynthetisierten Adäquat

Die Abbildungen 6.1, 6.2, 6.3, 6.4 und 6.5 zeigen jeweils die FFT des Originalklanges eines Instruments, verglichen mit der FFT des resynthetisierten Klanges desselben Instruments. Verwendet wurde hier eine  $2^{14}$  Punkte FFT. Es wurden jeweils 2048 Samples entnommen und mit einem Hanningfenster gewichtet, wonach Nullen angefügt wurden, bis die FFT-Länge von  $2^{14}$  erreicht wurde. Dieses Hanningfenster wurde, bis das Ende des Signal erreicht wurde, immer um 2048 Samples weiter verschoben, wonach immer auf die Länge der FFT Zeropadding durchgeführt wurde. Dadurch erhielt man mehrere genaue Kurzzeitspektren, die nach einer Mittelung eine sehr gut repräsentative Darstellung der im Klang enthaltenen Komponenten erreichten. Man kann gut erkennen, dass die, für diese Kapitel ausgewählten, Beispiele eine relativ hohe Übereinstimmung zwischen den Frequenzkomponenten des Original- und Syntheseklanges aufweisen.

## 6. Analyse-Beispiele

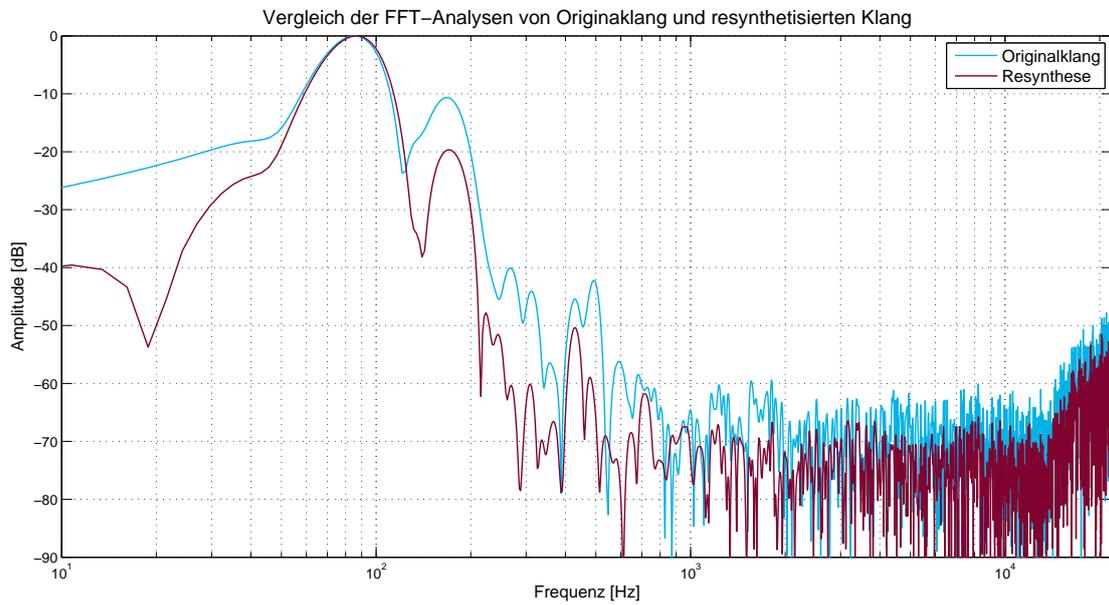


Abbildung 6.2.: Vergleich eines Originalklanges des Instruments Gong Suwukan mit seinem resynthetisierten Adäquat

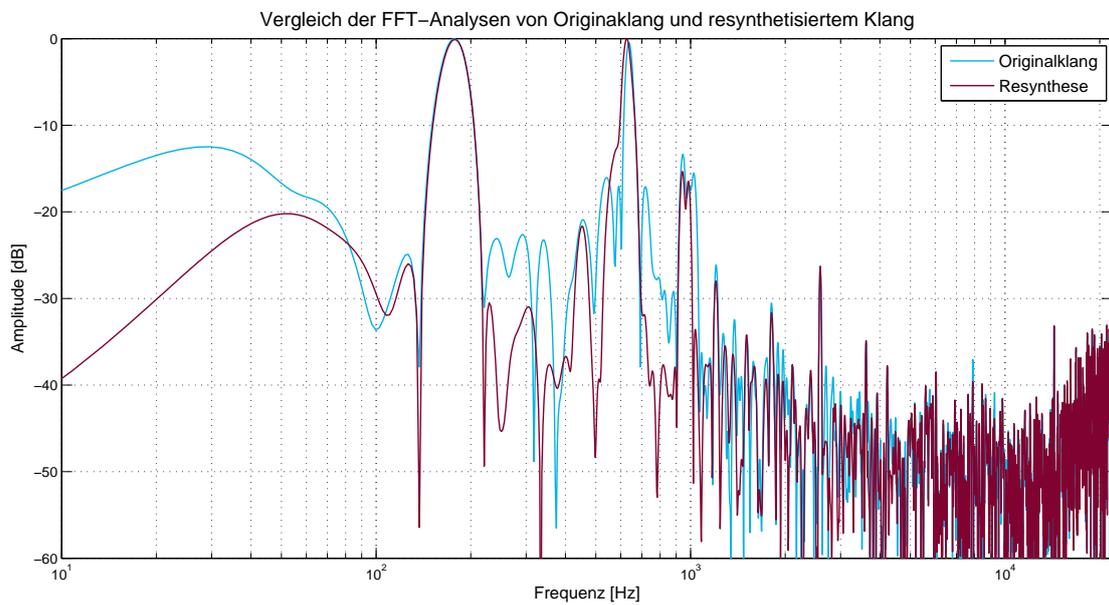


Abbildung 6.3.: Vergleich eines Originalklanges des Instruments Bonang Barung mit seinem resynthetisierten Adäquat

## 6. Analyse-Beispiele

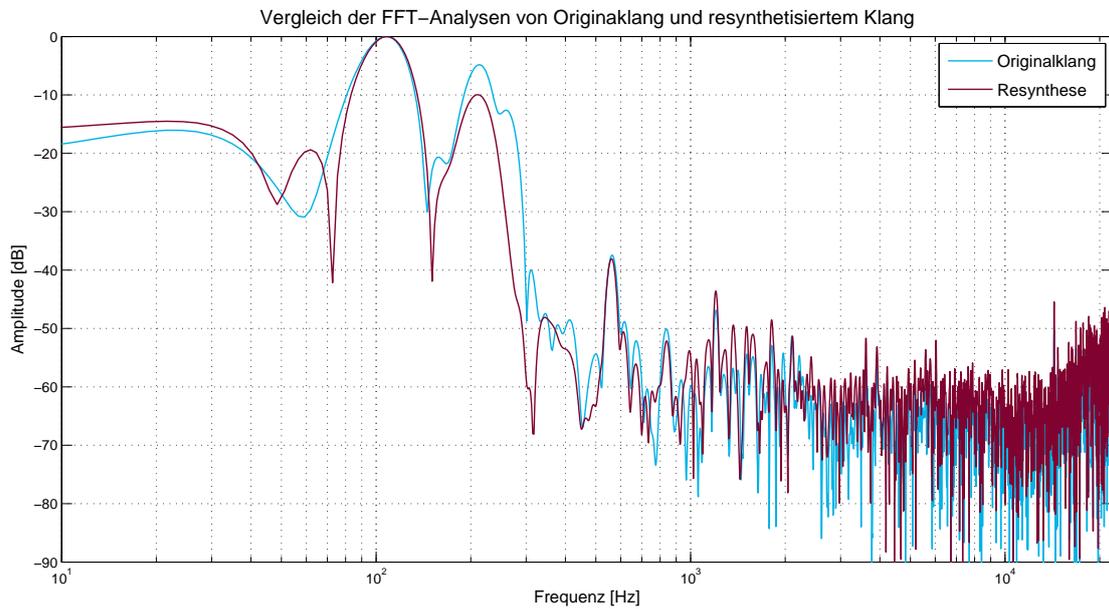


Abbildung 6.4.: Vergleich eines Originalklanges des Instruments Kempul mit seinem resynthetisierten Adäquat

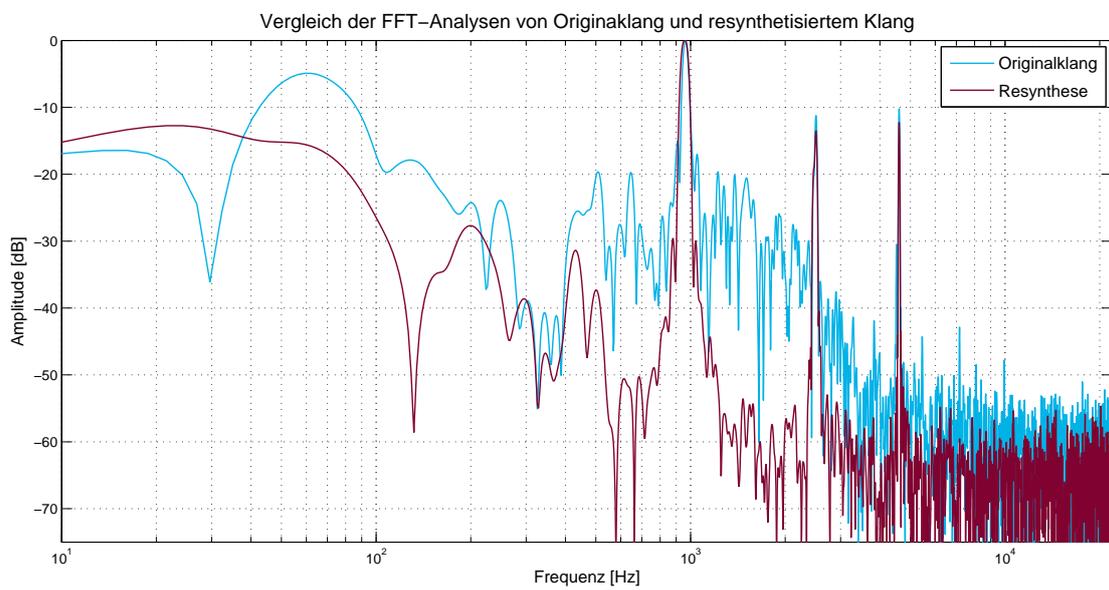


Abbildung 6.5.: Vergleich eines Originalklanges des Instruments Slenthem mit seinem resynthetisierten Adäquat

## 6.1. Gong Suwukan

Wie bereits in Abbildung 6.2 ersichtlich ist, hat der Gong Suwukan eine starke Ausprägung seines Spektrums bei ca. 82Hz und eine weitere kleinere Ausprägung bei etwa 165Hz. Man sieht eine starke Übereinstimmung des resynthetisierten Klanges mit dem Original bei der Frequenzausprägung bei 82Hz. Die Übereinstimmung bei 165Hz ist weniger exakt, der dortige synthetisierte Verlauf ist um fast 10dB gedämpft. Die verschiedenen kleineren Ausprägungen im höherfrequenten Bereich stimmen verhältnismäßig wenig überein. Trotzdem kommt die akustische Überprüfung auf ein täuschend ähnliches Ergebnis zum Originalklang. Tabelle 6.1 demonstriert die in der Analyse eruierten Parameter des Gong Suwukan.

Frequenz [Hz]	Amplitude [dB]	Phase [°]	Einschwingzeit [ms]	Ausklinddauer [ms]
82.57	-16.5	-342	27	11259
165.17	-12.7	17	336	8963
432.81	-70.7	-183	74	9327
464.85	-62.0	-11	74	5322
514.39	-93.1	25	11	2475
528.53	-92.5	-240	11	3597
697.64	-73.1	-71	9	2556
721.91	-77.7	-170	9	2995
739.42	-82.5	-38	9	2186

Tabelle 6.1.: Ermittelte Parameter des Audiofiles Pl GongSuw 22

Abbildung 6.7 zeigt eben diese beiden starken Frequenzspitzen bei 82Hz und 165Hz mit ihren Ausklindkonstanten. Als am längsten ausklingende Komponenten heben sie sich deutlich von den restlichen ab und sind somit charaktergebend. Genau genommen ergeben die analysierten 165Hz den ersten Oberton - also die Oktave - zu der Grundfrequenz bei 82Hz, wobei dieser erste Oberton zum Grundton ein wenig verstimmt ist, was wiederum die für den Gong charakteristische Schwebung im Klang erzeugt. Die für den Gong Suwukan wesentlichen und von der Analyse erkannten Teilfrequenzen, befinden sich durchgehend unterhalb von 1kHz. Als Gong ist der Gong Suwukan dadurch mit typischer, eher tieferer Grundfrequenz und wenigen, stark ausgeprägten Komponenten beschreibbar.

## 6.2. Bonang Barung

Das Audiofile Sl\_BonangBar\_52.wav ist ein gutes Beispiel für einen Klang, der trotz vieler detektierter Frequenzen, deutlich auf die, im Laufe der Berechnung vorgenommene, Elimination der maskierten Frequenzanteile hinweist. Abbildung 6.3 verdeutlicht dies

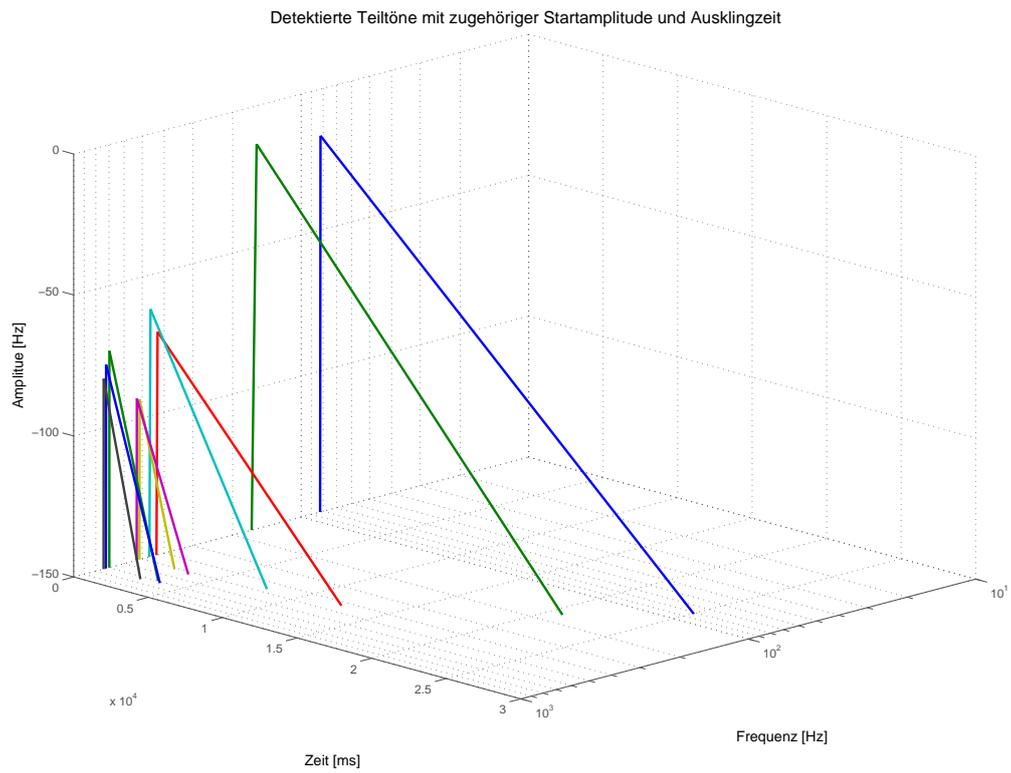


Abbildung 6.6.: Detektierte Teiltöne des Gong Suwukan mit zugehöriger Startamplitude und Ausklingzeit

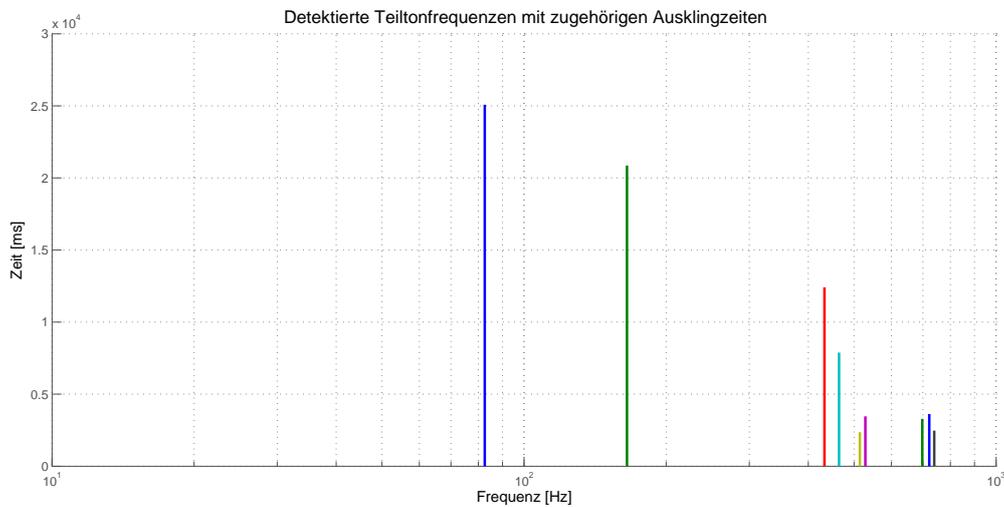


Abbildung 6.7.: Detektierte Teiltonfrequenzen des Gong Suwukan mit zugehörigen Ausklingzeiten

durch die immer wiederkehrenden, starken Frequenzeinbrüche, wo das Original teils stärkere Anteile aufweist. Diese Eigenheit ist auch in Abbildung 6.5 auffällig. Sieht man sich die Spitzen bei ca. 950Hz, ca. 2500Hz und ca. 4500Hz und die jeweiligen charakteristischen Abfälle zu höheren und tieferen Frequenzen an, fühlt man sich an Abbildung 3.2 mit den für die Simultanmaskierung üblichen Flanken erinnert.

Es sind zwei tiefe Frequenzen mit ca. 43Hz und 50.86Hz in diesem Klang enthalten. Zusätzlich erscheint noch eine Schwingung mit 111.08Hz, deren fünfter Oberton bei rund 554Hz im Spektrum enthalten ist. Jedoch liegt die musikalisch prägnanteste Frequenz bei 632.49Hz. Als Ton mit der stärksten Startamplitude ist er für die Wahrnehmung und Tonhöhenordnung maßgebend. Weiters wird seine Position als Grundton durch das Auftreten der Doppeloktave bei ca. 2583Hz gestärkt. Die ansonst auftretenden Frequenzen werden aus dem Anschlag stammend vermutet.

### 6.3. Kempul

Beim Kempul ist wesentlich und charaktergebend, dass sich alle zehn eruierten Teilfrequenzen zwischen 100Hz und 1kHz befinden. Die stärksten Ausprägungen findet man bei ca. 104Hz, ca. 165Hz und etwa 209Hz, wobei hier die 209Hz bereits wie der erste musikalische Oberton (Oktave) wirken. Das hier analysierte Instrument zählt zu den besten Ergebnissen aller Analysedurchläufe. Dies könnte zum einen daran liegen, dass bei der Analyse kaum Frequenzkomponenten eliminiert wurden (vgl. Abbildung 6.4), zum anderen ist der Klang in seinem Spektrum von vornherein eher begrenzt (wie auch der

## 6. Analyse-Beispiele

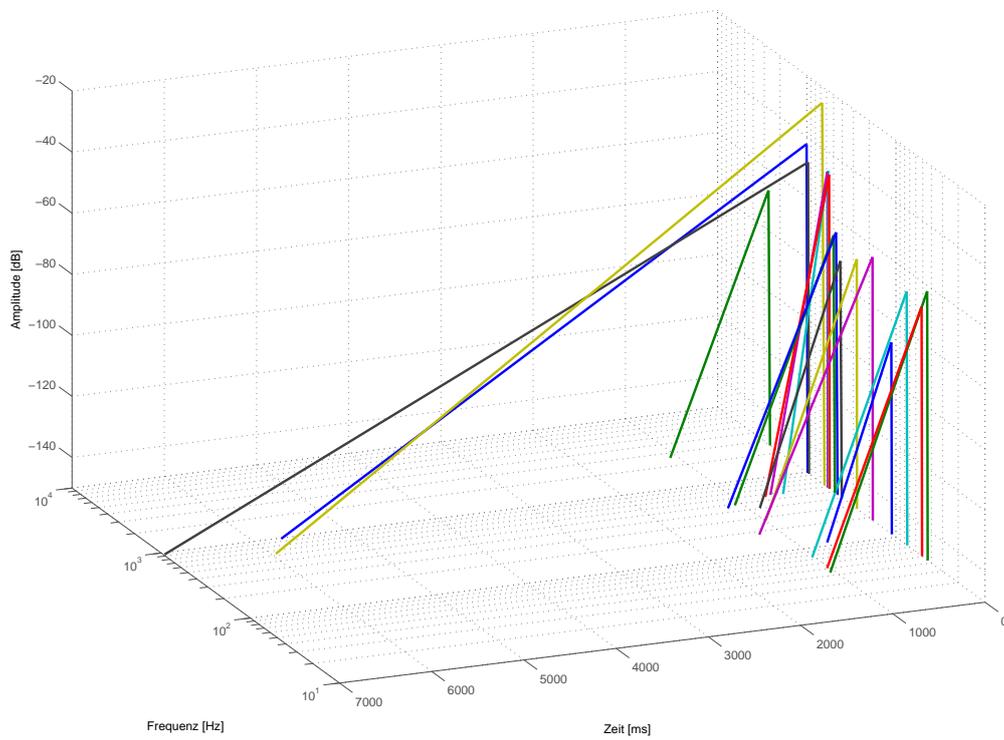


Abbildung 6.8.: Detektierte Teiltöne des Bonang Barung mit zugehöriger Startamplitude und Ausklingzeit

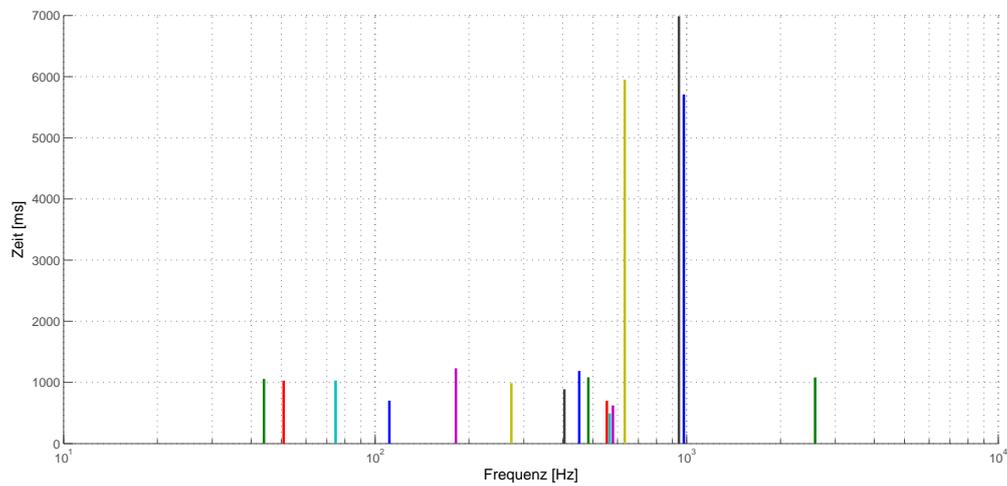


Abbildung 6.9.: Detektierte Teiltonfrequenzen des Bonang Barung mit zugehörigen Ausklingzeiten

Frequenz [Hz]	Amplitude [dB]	Phase [°]	Einschwingzeit [ms]	Ausklinddauer [ms]
43.99	-61.7	264	6	725
50.86	-68.3	161	6	755
74.70	-66.8	-81	6	739
111.08	-87.1	7	6	664
181.63	-63.9	-128	4	877
273.56	-68.3	-270	4	724
404.97	-72.4	63	16	675
452.07	-64.0	-220	16	818
483.35	-65.7	-24	16	760
554.48	-47.1	-191	6	404
566.52	-46.0	-217	6	282
579.59	-46.5	-234	6	358
632.49	-24.7	-120	25	2836
944.42	-47.9	-114	11	4100
978.89	-42.3	125	11	3174
2583.56	-66.4	-29	17	765

Tabelle 6.2.: Ermittelte Parameter des Audiofiles SI BonangBar 52

Gong Suwukan, dessen Spektrum sich in einem ähnlichen Bereich bewegt - siehe Abbildung 6.2). Hier ist auf den ersten Blick die Oktave zum Grundton mit 104Hz bei 208Hz erkennbar. Auch die Doppeloktave mit 417Hz ist leicht erkennbar. Dieses dreimalige Auftreten der Grundtonhöhe ergibt ein recht tiefgehendes Tonhöhenempfinden bei diesem Klang. Bei ca. 614Hz scheint der sechste Oberton auf, bei rund 847Hz der achte Oberton. Ein zweiter Grundton mit 165Hz tritt hier mit seinem dritten Oberton (Oktave plus Quinte) bei ca. 482Hz auf. Die anderen nicht in die Obertonreihe passenden Frequenzen könnten aus dem Anschlaggeräusch sowie aus Interferenzen bei den sich im Instrument ausbreitenden Moden entstehen.

## 6.4. Peking A

Das Analysebeispiel Pl\_PekingA\_62 steht stellvertretend für die Gruppe der Pekings, in der die meisten Klänge mit weniger als sechs Frequenzen für die Resynthesierung auskommen. Das hier dargelegte Beispiel benötigt sogar nur zwei Frequenzen. Mit akustischer Überprüfung ist der resynthetisierte Klang, verhältnismäßig für die Reduktion der ursprünglichen Frequenzkomponenten, immer noch sehr gut als jener Klang zu erkennen, der er eigentlich sein sollte. Hier ist wirklich nur noch eine Grundfrequenz zur Darstellung der Tonhöhe und eine zweite, farbgebende Frequenz enthalten. Alle anderen ursprünglichen Komponenten werden innerhalb der Berechnung als nicht lange genug unmaskiert

## 6. Analyse-Beispiele

Frequenz [Hz]	Amplitude [dB]	Phase [°]	Einschwingzeit [ms]	Ausklindauer [ms]
104.39	-21.1	-51	32	8006
165.38	-44.0	-134	161	1996
208.76	-6.3	-171	164	4086
417.70	-54.9	-47	205	2037
482.12	-83.7	-171	205	3135
556.87	-59.4	-155	91	2512
569.69	-63.0	73	91	3004
597.17	-69.9	-64	91	1815
614.16	-62.5	97	91	3380
847.34	-71.4	5	3	3143

Tabelle 6.3.: Ermittelte Parameter des Audiofiles SI Kempul 25

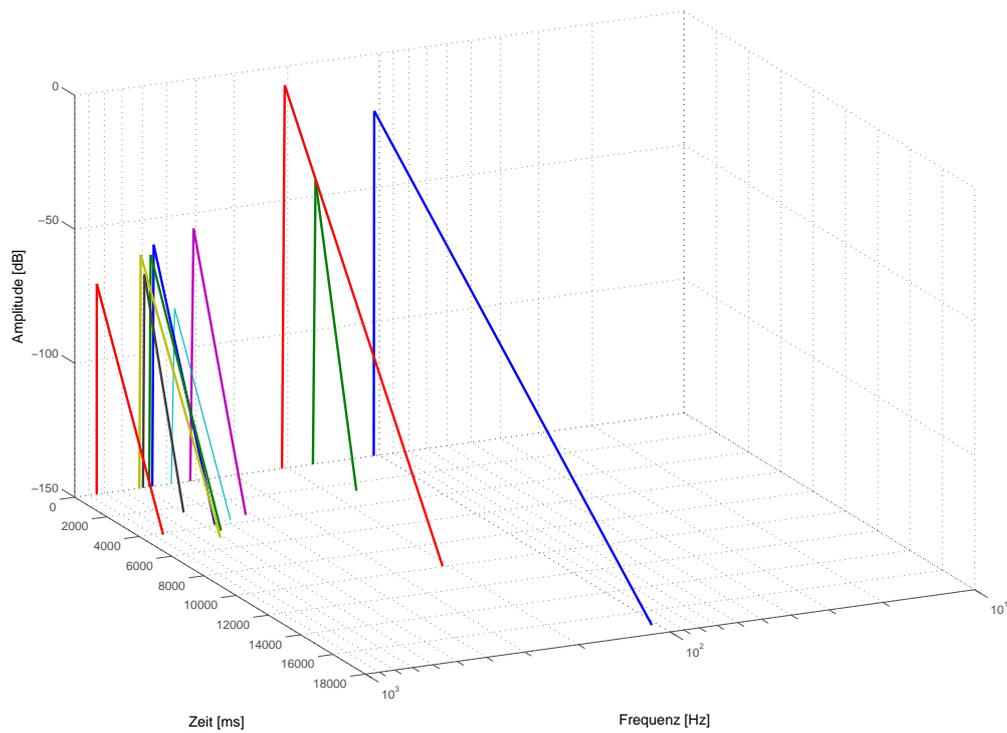


Abbildung 6.10.: Detektierte Teiltöne des Kempul mit zugehöriger Startamplitude und Ausklingzeit



## 6. Analyse-Beispiele

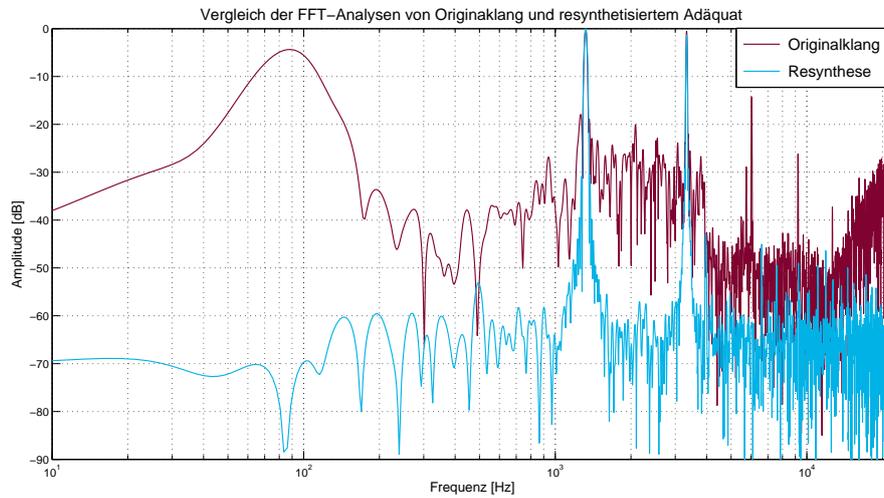


Abbildung 6.12.: Vergleich der FFT-Analyse des Originalklanges Pl PekingA 62 mit seinem resynthetisierten Adäquat

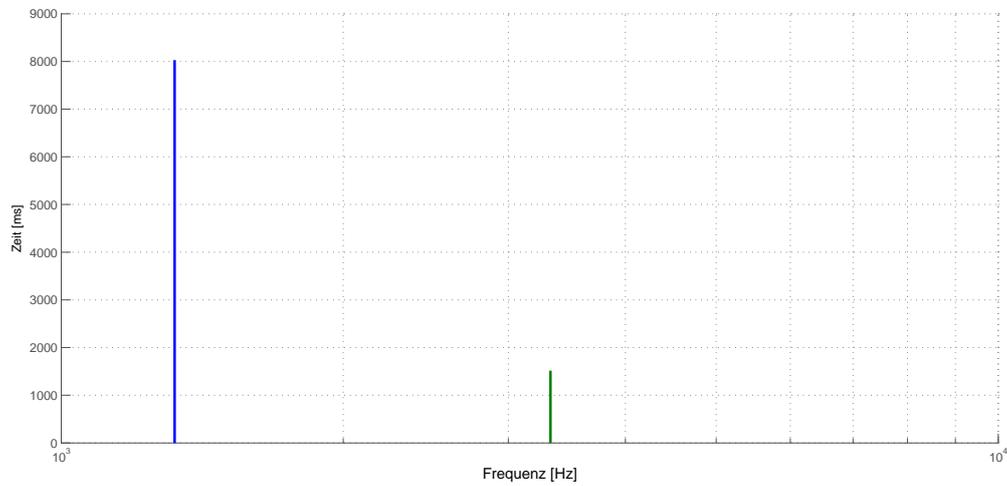


Abbildung 6.13.: Detektierte Teiltonfrequenzen des Peking A mit zugehörigen Ausklingzeiten

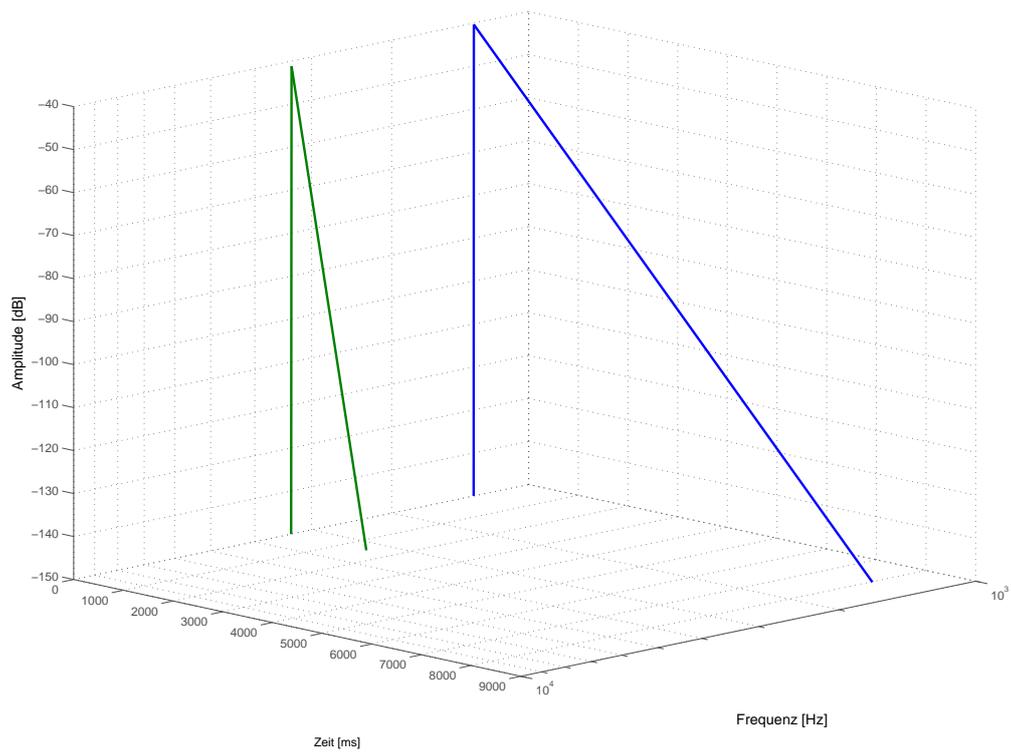


Abbildung 6.14.: Detektierte Teiltöne des Peking A mit zugehöriger Startamplitude und Ausklingzeit

## 7. Test der Funktionalität des Analyseprogrammes

Nicht alle Analyseergebnisse waren so zufriedenstellend wie die im vorhergehenden Kapitel 6 diskutierten. Um den Ursachen der teilweise unschöneren Ergebnisse auf den Grund zu gehen, wurde das Programm mit künstlichen, zuvor bekannten Signalen getestet.

### 7.1. Programmtest

Wie eben erwähnt waren nicht alle Ergebnisse zu voller Zufriedenheit. Zur Überprüfung der Grenzen des Programmes wurden künstlich generierte, bekannte Signale durch die Analyse geschickt und die Analyseergebnisse anschließend mit dem ursprünglichen Signal verglichen. Die synthetischen Testsignale waren Sinustöne oder Kombinationen von Sinustönen, die jeweils eine Hüllkurve mit linearem Einschwingvorgang und exponentiell abklingenden Ausschwingvorgang aufgeprägt bekamen.

#### 7.1.1. Test mit einfachen Sinussignalen

Begonnen wurde mit einem Test der Frequenzdetektion innerhalb der unterschiedlichen Bänder. Dabei wurde das Programm für den Test in seiner ursprünglichen Konfiguration belassen. Wurde als Testsignal ein einzelner Sinus herangezogen, dessen Amplitude, Einschwingzeit und Ausklingdauer vorgegeben waren, so wurde diese sehr genau an der richtigen Frequenz detektiert. Wie man Tabelle 7.1.1 entnehmen kann, wurde die Frequenz durchgehend sehr genau detektiert. Der Fehler bewegt sich hier für alle getesteten Frequenzen unter 1%, was eine exakte Frequenzdetektion des erarbeiteten Programmes verifiziert. Als Testfrequenzen wurden einige Frequenzen an den Bandgrenzen, und einige Frequenzen weiter innerhalb der Bänder herangezogen. Zum Beispiel liegen 36Hz am unteren Rand des ersten Bandes, 113Hz am oberen Rand, 80Hz liegen mitten im ersten Band.



Die blau markierten Zeilen stellen jeweils ein Testsignal, das gerade noch unter der oberen Bandkante liegt, dar. Auffällig bei diesen Frequenzen ist die fehlerhafte Detektion der Amplitude. Die Abweichung steigt hier mit der Höhe der Frequenz und bewegt sich im schlimmsten Fall im Bereich 6dB. Weiters lässt die Tabelle auf größere Ungenauigkeiten bei der Erkennung der Ausklingzeit schließen. Der Fehler bei diesem Parameter bewegt sich im Bereich von ca. 33%. Diese extrem hohe Abweichung ist nicht wünschenswert und sollte im Falle einer fortführenden Arbeit gesondert behandelt werden. Auch der Fehler bei der Detektion der Einschwingvorgänge liegt mit durchschnittlichen 15% in einem Bereich der noch Verbesserungen zulässt.

### 7.1.2. Test mit Sinuskombinationen

In dieser Testphase war es Ziel die Frequenzauflösung des Programmes zu erörtern. Dazu wurden Kombinationen zweier Sinustöne in unterschiedlichen Abständen zueinander und verschiedenen Frequenzbereichen als Testsignale verwendet. Hierbei war das tieffrequenteste Band das zu erst getestet. Es wurde eine Kombination zweier Sinussignale bei 36Hz und 37Hz als Input generiert. Die für dieses Signal benötigte Frequenzauflösung wurde mit den standardmäßig eingestellten Anzahl von 22 Polen nicht erreicht. Um in diesem Frequenzband, am untersten Rand des betrachteten Frequenzbereiches, zwei Sinustöne zu detektieren, die so nahe beieinanderliegen, war mindestens 24 Pole notwendig. Die beiden Töne konnten mit leichter Verschiebung zur gewünschten Frequenz detektiert werden. Diese Verschiebung bewegt sich im Bereich von einem Prozent und schrumpft bei höherer Polanzahl, bis er sich auf eine Frequenzverschiebung von 0.1Hz nach unten bei der tieferen Frequenz und 0.1 Hz nach oben bei der höheren Frequenz einpendelt (vgl. Tabellen 7.1.2 bis 7.1.2). Dies ist ein Bereich, der unterhalb der *Just Noticeable Difference* (JND) (vgl. [6]) liegt, wodurch der Fehler sogar in diesem niederfrequenten Bereich nicht ins Gewicht fällt. Die Just Noticeable Difference ist ein Phänomen, das dem Ohr eine gewisse Toleranz gegenüber der Tonhöhe einräumt. Dabei ist das Ohr Frequenzmodulationen gegenüber empfindlicher als statischen Tonhöhenänderungen. Bei jenen Verschiebungen im tieffrequenten Bereich nimmt das Ohr erst ab einer Änderung von ca. 10% eine tatsächliche neue Tonhöhe wahr. Denkt man dabei an die musikalische Bedeutung der Tonhöhe und Intervallabstände in Hertz bzw. Prozent, so erscheint eine Änderung von 10% im tieffrequenten Bereich sehr groß, um nicht zu sagen, musikalisch nicht mehr gleichwertig dem ursprünglichen Ton. Tatsächlich ist es aber so, dass der Ton in der Musik deshalb zu tief oder zu hoch wahrgenommen wird, da zusätzlich Obertöne mitschwingen, die sich im Verhältnis genauso verschieben. Dies ergibt ein leichteres Wahrnehmen der Tonhöhenänderung, da das Ohr zu hohen Frequenzen empfindlicher auf Verschiebungen wird.

Die Erkennung der Signalkomponenten war immer der Polanzahl entsprechend gut. Grundsätzlich lässt sich sagen, dass bei geringerer Polanzahl, die beiden Sinussignale immer näher zusammen wandern, bis sie schließlich als eine Frequenz, liegend nahe an

7. Test der Funktionalität des Analysesprogrammes

---

Polanzahl	Frequenz [Hz]		Detektion			
			Detektierte Frequenzen		Fehler [%]	
22	36	37	-	-	-	-
24	36	37	36.403	36.994	1.119	0.016
26	36	37	36.320	36.922	0.888	0.210
28	36	37	36.169	36.969	0.469	0.083
30	36	37	36.100	37.000	0.277	0.000
32	36	37	36.000	37.000	0.000	0.000
34	36	37	35.900	37.100	0.277	0.272
36	36	37	35.900	37.100	0.277	0.272
40	36	37	35.900	37.100	0.277	0.272
48	36	37	35.900	37.100	0.277	0.272
56	36	37	35.900	37.100	0.277	0.272

Tabelle 7.2.: Analyse von einem Testsignal bei 36Hz und 37Hz mit unterschiedlichen Polanzahlen - Vergleich von jeweils erwarteter und realer Frequenz

Polanzahl	Frequenz [Hz]		Detektion			
			Detektierte Frequenzen		Fehler [%]	
26	112	113	-	-	-	-
28	112	113	112.443	112.571	0.395	0.379
30	112	113	112.092	112.913	0.082	0.077
32	112	113	112.008	112.998	0.007	0.002
34	112	113	111.955	113.052	0.040	0.046
36	112	113	111.900	113.100	0.089	0.088
48	112	113	111.900	113.100	0.089	0.088
56	112	113	111.900	113.100	0.089	0.088

Tabelle 7.3.: Analyse von einem Testsignal bei 112Hz und 113Hz mit unterschiedlichen Polanzahlen - Vergleich von jeweils erwarteter und realer Frequenz

7. Test der Funktionalität des Analysesprogrammes

Polanzahl	Frequenz [Hz]		Detektion			
			Detektierte Frequenzen		Fehler [%]	
28	198	199	-	-	-	-
30	198	199	198.453	198.540	0.228	0.231
32	198	199	198.100	198.906	0.050	0.047
36	198	199	198.000	199.000	0.000	0.000
40	198	199	197.900	199.100	0.050	0.050
48	198	199	197.900	199.100	0.050	0.050
56	198	199	197.900	199.100	0.050	0.050

Tabelle 7.4.: Analyse von einem Testsignal bei 198Hz und 199Hz mit unterschiedlichen Polanzahlen - Vergleich von jeweils erwarteter und realer Frequenz

Polanzahl	Frequenz [Hz]		Detektion			
			Detektierte Frequenzen		Fehler [%]	
28	290	291	-	-	-	-
30	290	291	290.455	290.534	0.156	0.160
32	290	291	290.228	290.764	0.078	0.081
34	290	291	290.096	290.899	0.033	0.035
40	290	291	289.900	291.100	0.034	0.034
48	290	291	289.900	291.100	0.034	0.034
56	290	291	289.900	291.100	0.034	0.034

Tabelle 7.5.: Analyse von einem Testsignal bei 290Hz und 291Hz mit unterschiedlichen Polanzahlen - Vergleich von jeweils erwarteter und realer Frequenz

Polanzahl	Frequenz [Hz]		Detektion			
			Detektierte Frequenzen		Fehler [%]	
38	622	623	-	-	-	-
40	622	623	622.356	622.645	0.057	0.058
42	622	623	622.200	622.800	0.032	0.032
46	622	623	622.102	622.901	0.016	0.016
48	622	623	622.000	623.000	0.000	0.000
50	622	623	622.000	623.100	0.000	0.016
52	622	623	622.000	623.100	0.000	0.016
54	622	623	621.900	623.100	0.016	0.016
56	622	623	621.900	623.100	0.016	0.016

Tabelle 7.6.: Analyse von einem Testsignal bei 622Hz und 623Hz mit unterschiedlichen Polanzahlen - Vergleich von jeweils erwarteter und realer Frequenz

Polanzahl	Frequenz [Hz]		Detektion			
			Detektierte Frequenzen		Fehler [%]	
46	904	905	-	-	-	-
48	904	905	904.450	904.557	0.050	0.049
50	904	905	904.426	904.581	0.047	0.046
52	904	905	904.330	904.678	0.036	0.035
54	904	905	904.204	904.804	0.022	0.022
56	904	905	904.100	904.900	0.011	0.011

Tabelle 7.7.: Analyse von einem Testsignal bei 904Hz und 905Hz mit unterschiedlichen Polanzahlen - Vergleich von jeweils erwarteter und realer Frequenz

der unteren, im Testsignal enthaltenen Komponente aufscheinen. Wie bereits in Tabelle 7.1.2 sichtbar wird, ist eine Frequenzauflösung von 1Hz ab diesem Bereich mit sinnvoller Polanzahl kaum mehr möglich. Aus dem Verlauf des Fehlers über die Frequenz kann man schließen, dass mit steigender Frequenz für gleich gute Auflösung eine ungleich höhere Polanzahl nötig sein wird. Für Fälle bei denen im Bereich über 1kHz eine Auflösung von 1Hz benötigt würde wäre es wahrscheinlich sinnvoller gleich das originale Klangbeispiel zu verwenden, da der Rechenaufwand natürlich mit erhöhter Polanzahl steigt.

Für die im Rahmen dieser Arbeit behandelten Problemstellungen ist eine gute Frequenzauflösung im tieferfrequenten Bereich wünschenswert. Mit der Grundeinstellung von 22 Polen können im Bereich bis etwa 650Hz zwei Frequenzen im Abstand von 2Hz ohne weiteres detektiert werden. Zu höheren Frequenzen benötigt man dann bereits einen Frequenzabstand von 5Hz oder mehr, bis hin zur oberen Frequenzbereichsgrenze bei 7880Hz, wo man zumindest 20Hz Abstand um noch zwei separate Töne zu detektieren.

## 7.2. Verbesserung der Analyseergebnisse durch gezielte Manipulation des Programmablaufes

Ausgehend von oben erwähnten Grenzen des Programmes, wurde versucht, die Analyse durch gezieltes Eingreifen in den Programmablauf zu verbessern. Sieht man sich hier die beschränkte Frequenzauflösung der Analyse an, so muss man, um gewisse Klänge noch gut reproduzieren zu können, an der richtigen Stelle manipulieren. Für die meisten Klänge reicht eine Analyse mit 22 Polen, wie sie standardmäßig eingesetzt wurde. Jedoch wird im Falle von zu eng beieinander stehender Komponenten eine bessere Auflösung notwendig. Nachdem das psychoakustische Modell im Zuge der Analyse nur für das Ohr irrelevante Teile des Signals eliminiert, wird die Verfeinerung der Auflösung primär durch eine Verbesserung der LPC - also einer Erhöhung der Polanzahl - geschehen. Dieser Ansatz wurde an einigen Beispielen getestet, deren Klang es an akustischer Kongruenz zum

Original fehlte. Es wurden bei diesem Analyseverfahren durchgehend bessere Ergebnisse erzielt.

Will man im höherfrequenten Bereich eine so hohe Auflösung erreichen, so ist es von Nutzen die Polanzahl dementsprechend zu erhöhen. So wird im folgenden Kapitel 7.2.1 ersichtlich, dass bei 86Hz un 87Hz für eine Detektion zweier so knapp nebeneinander liegender Frequenzen 22 Pole nicht mehr genügen.

### 7.2.1. Gong Ageng 13

Sieht man sich die analysierten Komponenten des Gong Ageng an, so fehlen wesentliche Komponenten die bei ca. 85Hz liegen müssten. Diese Komponenten sind im Originalklang einfach zu hören. Sie erzeugen eine Schwebung die den eigentlichen Grundcharakter des Klanges ergibt. Abbildung 7.1 zeigt die mit der Standardeinstellung von 22 Polen detektierten Teiltöne. Erst mit 30 Polen wird eine der oben erwähnten Komponenten

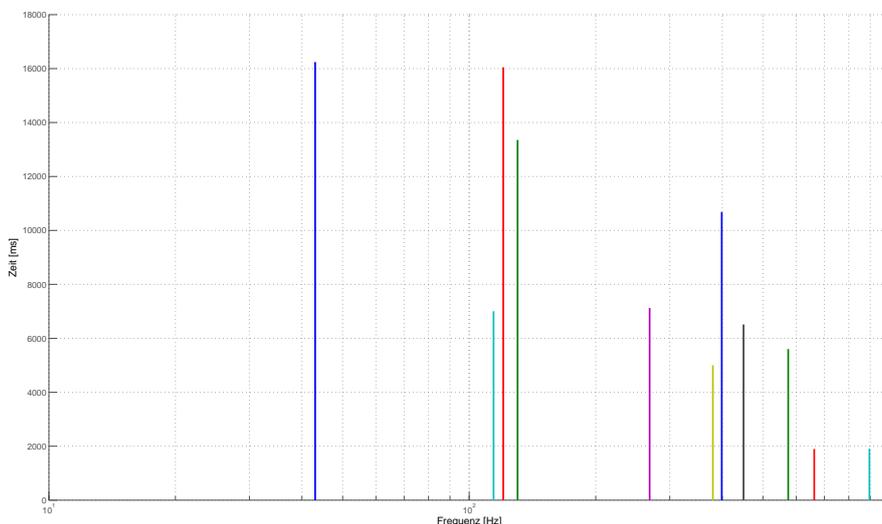


Abbildung 7.1.: Mit 22 Polen detektierte Teiltöne des Audiofiles Gong Ageng 13

detektiert. Abbildung 7.2 zeigt die in eben dieser Auflösung von 30 Polen detektierten Komponenten. Hier sieht man gut, dass allgemein mehr Teiltöne erfasst werden können. Trotz der bereits 30 Pole fehlt immer noch die zweite Komponente für die Schwebung, die dicht neben der bei rund 85Hz liegen müsste. erhöht man die Polanzahl der LPC weiter so erhöht sich auch die Anzahl der detektierten Komponenten. Mit 56 Polen wird die zweite Schwebungskomponente gut erkannt. Überprüft man den Klang nun akustisch so ist er eindeutig der selbe Gong Ageng an dessen Klang die Analyse durchgeführt wurde. Abbildung 7.3 lässt einen Einblick in die Komplexität des Gong Ageng zu. Nicht nur bei 86Hz und 87Hz stehen zwei Komponenten nahe aneinander, auch zu höheren Frequenzen

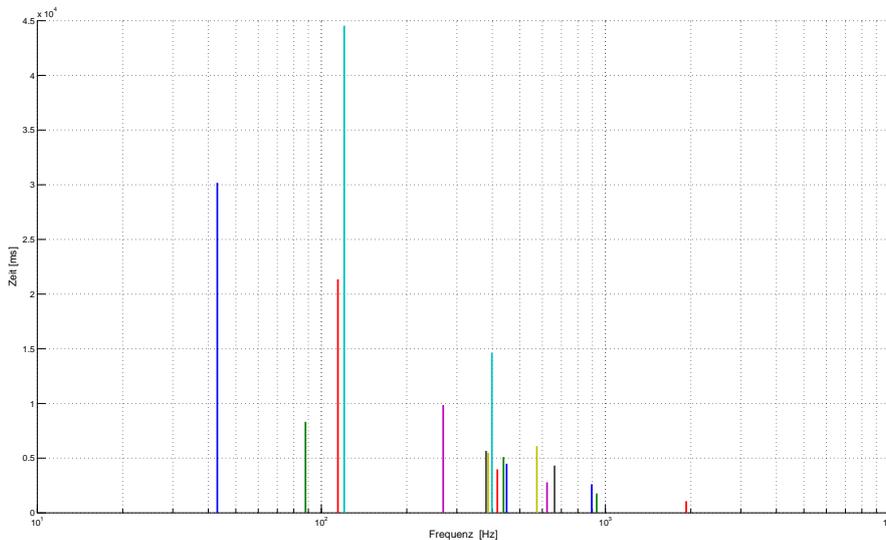


Abbildung 7.2.: Mit 30 Polen detektierte Teiltöne des Audiofiles Gong Ageng 13

gibt es eng beieinanderliegende Töne. Vermutlich würde dieses Anwachsen von Teiltönen auch bei jedem anderen Klang auftreten. Jedoch ist das manuelle Eingreifen in den Analyseprozess nur bei schlechten Analyse- und Syntheseergebnissen, wie zum Beispiel beim hier behandelten Gong Ageng, notwendig und sinnvoll, und man könnte gleich zu den ursprünglichen Originalklängen zurückkehren.

Abbildung 7.4 zeigt noch einmal vergleichend die Spektren des originalen Klanges, des mit 22 Polen analysierten und resynthetisierten Klanges und des Klanges aus einer Analyse mit 56 Polen. Zur Resynthese des mit 56 Polen analysierten Klanges wurden alle detektierten Frequenzen mit ihren zugehörigen Startamplituden, Startphasen, Einschwingvorgängen und Ausklingdauern herangezogen. Der so resynthetisierte Klang wurde anschließend mit einer FFT (siehe Kapitel 6) analysiert und mit dem Originalklang verglichen. Hier ist der deutliche Abfall des Klanges aus der Analyse mit 22 Polen zu erkennen, wo eigentlich die wichtigsten beiden Komponenten stehen sollten. Im Vergleich dazu ist die Kongruenz des Originals und des Klanges aus der Analyse mit 56 Polen extrem hoch. Im tieffrequenten, klangcharakterisierenden Bereich ist die Deckung der beiden Klänge nahezu perfekt.

## 7. Test der Funktionalität des Analysprogrammes

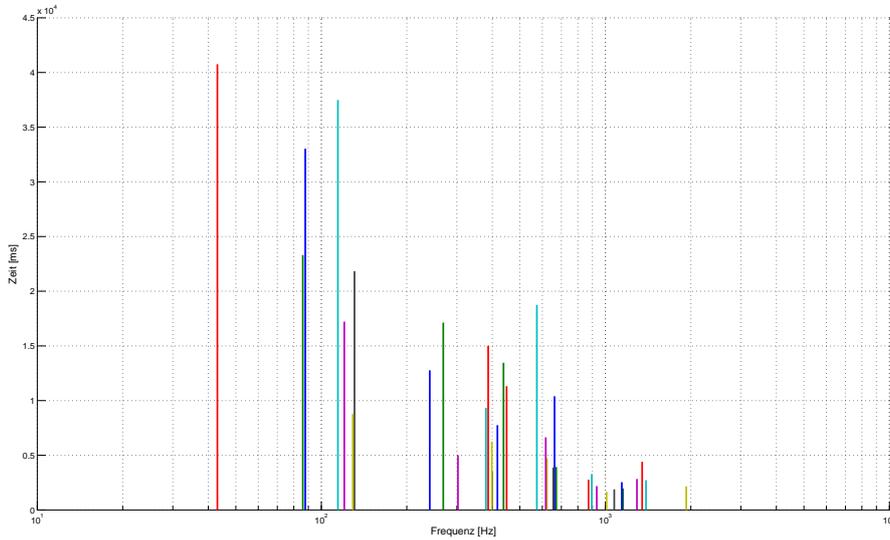


Abbildung 7.3.: Mit 56 Polen detektierte Teiltöne des Audiofiles Gong Ageng 13

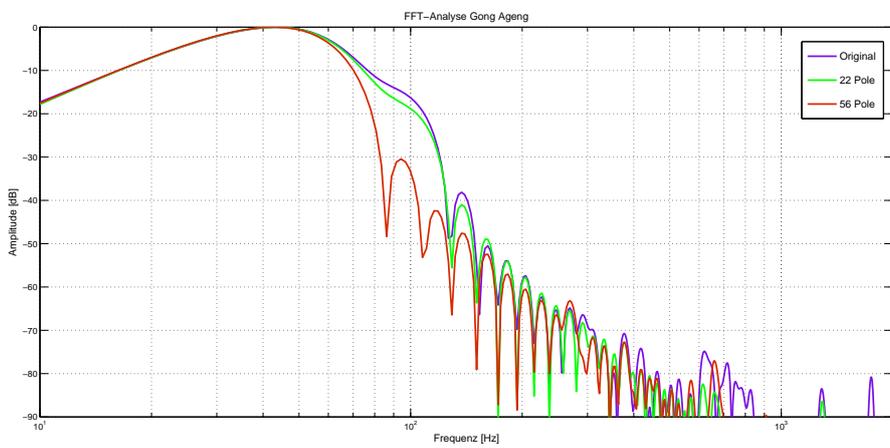


Abbildung 7.4.: FFT des originalen Gong Ageng, der Analyse mit 22 Polen und mit 56 Polen

# Abbildungsverzeichnis

1.1. Bildausschnitt eines typischen Gamelanorchesters (Quelle: IME-Homepage)	5
2.1. Schritte der FZ-ARMA	8
3.1. Flussdiagramm Bark-Zooming ARMA	10
3.2. Schematische Darstellung der Simultanmaskierung	11
3.3. Schematische Darstellung der Lautheitsempfindung	12
3.4. Zusammenhang zwischen linearer Frequenz $f$ und Barkfrequenz $z$	14
3.5. Barkfilterkurven	14
3.6. Schematische Darstellung der Maskierung von T30 und Einschwingzeit	15
3.7. Vergleich mehrerer Residuen eines Saron Demung Instrumentes	18
3.8. Vergleich mehrerer Residuen eines Slenthem Instrumentes	18
4.1. FFT-Spektren mehrerer Originalklangbeispiele	20
4.2. Analyse-Filterbank	21
4.3. Barkfilterkurven 2. (blau) und 4. (pink) Ordnung	23
4.4. Barkfilterkurven 4. (blau) und 8. (pink) Ordnung	23
4.5. Blockdiagramm der LPC-Analyse	24
5.1. Synthese mit Residuum	25
5.2. Analyse mit Decaying Exponentials	26
5.3. Zur Synthese herangezogene Teiltöne eines Bonang Panerus mit zugehöriger Startamplitude und Ausklingzeit	27
5.4. Teiltonfrequenzen und zugehörige Ausklingzeiten eines Bonang Panerus	28
5.5. Zur Synthese herangezogene Teiltöne eines Bonang Barung mit zugehöriger Startamplitude und Ausklingzeit	28
5.6. Teiltonfrequenzen und zugehörige Ausklingzeiten eines Bonang Barung	29
5.7. Synthese-Filterbank	30
5.8. Blockdiagramm der Teiltongeneration	31
6.1. Vergleich eines Originalklanges des Instruments Bonang Panerus mit seinem resynthetisierten Adäquat	32
6.2. Vergleich eines Originalklanges des Instruments Gong Suwukan mit seinem resynthetisierten Adäquat	33

6.3. Vergleich eines Originalklanges des Instruments Bonang Barung mit seinem resynthetisierten Adäquat . . . . .	33
6.4. Vergleich eines Originalklanges des Instruments Kempul mit seinem resynthetisierten Adäquat . . . . .	34
6.5. Vergleich eines Originalklanges des Instruments Slenthem mit seinem resynthetisierten Adäquat . . . . .	34
6.6. Detektierte Teiltöne des Gong Suwukan mit zugehöriger Startamplitude und Ausklingzeit . . . . .	36
6.7. Detektierte Teiltonfrequenzen des Gong Suwukan mit zugehörigen Ausklingzeiten . . . . .	37
6.8. Detektierte Teiltöne des Bonang Barung mit zugehöriger Startamplitude und Ausklingzeit . . . . .	38
6.9. Detektierte Teiltonfrequenzen des Bonang Barung mit zugehörigen Ausklingzeiten . . . . .	38
6.10. Detektierte Teiltöne des Kempul mit zugehöriger Startamplitude und Ausklingzeit . . . . .	40
6.11. Detektierte Teiltonfrequenzen des Kempul mit zugehörigen Ausklingzeiten	41
6.12. Vergleich der FFT-Analyse des Originalklanges Pl PekingA 62 mit seinem resynthetisierten Adäquat . . . . .	42
6.13. Detektierte Teiltonfrequenzen des Peking A mit zugehörigen Ausklingzeiten	42
6.14. Detektierte Teiltöne des Peking A mit zugehöriger Startamplitude und Ausklingzeit . . . . .	43
7.1. Mit 22 Polen detektierte Teiltöne des Audiofiles Gong Ageng 13 . . . . .	50
7.2. Mit 30 Polen detektierte Teiltöne des Audiofiles Gong Ageng 13 . . . . .	51
7.3. Mit 56 Polen detektierte Teiltöne des Audiofiles Gong Ageng 13 . . . . .	52
7.4. FFT des originalen Gong Ageng, der Analyse mit 22 Polen und mit 56 Polen	52

# Literaturverzeichnis

- [1] Alois Sontacchi, Franz Zotter, „Zwischenreport zum Projekt Virtual Gamelan Graz (VGG)“, IEM Report 35/06, [http://iem.at/projekte/publications/iem\\_report/report35\\_06/report35\\_06](http://iem.at/projekte/publications/iem_report/report35_06/report35_06), 2006.
- [2] Matti Karjalainen et al., „AR/ARMA Analysis and Modeling of Modes in Resonant and Reverberant Systems“, 112th AES Convention, Convention Paper 5590, May 10-13 Munich, Germany, 2002.
- [3] Frank Baumgarte, „Ein psychophysiologisches Gehörmodell zur Nachbildung von Wahrnehmungsschwellen für die Audiocodierung“, Dissertation, Fachbereich Elektrotechnik und Informationstechnik, Universität Hannover, 2000.
- [4] Michael Dickreiter, „Handbuch der Tonstudioteknik“, K.G.Saur Verlag München, Band I und II, 6. Verbesserte Auflage, 1997
- [5] Hartmut Traunmüller, „Analytical expressions for the tonotopic sensory scale“, J. Acoust. Soc. Am. 88: 97-100, 1990
- [6] Eberhard Zwicker und Hugo Fastl, „Psychoacoustics - Facts and Models“, Springer-Verlag Berlin, Second Updated Edition, 1999
- [7] Helmut Fleischer, „Schwingungsuntersuchung an Gongs“, Heft 1/01 der Reihe „Beiträge zur Vibro- und Psychoakustik“, Herausgeber: Helmut Fleischer und Hugo Fastl, 2001
- [8] Helmut Fleischer, „Hörversuche mit Glockenklängen“, Heft 1/07 der Reihe „Beiträge zur Vibro- und Psychoakustik“, Herausgeber: Helmut Fleischer und Hugo Fastl, 2007