

Abschlussarbeit zum Toningenieurprojekt

Klangmodellierung von Gamelaninstrumenten basierend auf exponentiell ausschwingenden Sinuskomponenten

von

Margherita Jammer

Graz, am 8. November 2007

Betreuer: Univ.Ass. DI Dr. Alois Sontacchi

Mitbetreuer: DI Franz Zotter



UNIVERSITÄT
FÜR MUSIK UND
DARSTELLENDE KUNST
GRAZ - AUSTRIA



Zusammenfassung

Im vorliegenden Projekt werden Klänge von Gamelaninstrumente 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 gehörsrelevanten 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. Dass heißt, es erfolgt eine Berücksichtigung der Übertragungseigenschaft des Außen- und Mittelohres sowie die des Innenohrs durch Berechnung der simultanen Maskierung bzw. der zeitlich variablen Mithörschwelle (vgl. [3]).

In einem zweiten Schritt sollen mittels „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.

Inhaltsverzeichnis

1. Einleitung	7
1.1. Gamelan	7
2. FZ-ARMA Frequency Zooming Autoregressive Moving-Average	9
3. BZ-ARMA Bark Zooming Autoregressive Moving-Average	10
3.1. Erster Block - Psychoakustische Betrachtung	10
3.1.1. Gewichtung des Signals mit A-Filter Kurve	10
3.1.2. Maskierung	10
3.1.3. Zerlegung in Barkbänder und Nutzung der Maskierungseigenschaften des menschlichen Gehörs	13
3.1.4. Modulation und Downsampling	15
3.2. Zweiter Block - Analyse	15
3.2.1. LPC	16
4. Analyse-Programm	19
4.1. Das zu bearbeitende Material	19
4.2. Die Berechnungsschritte des Programms	20
4.2.1. Inputbegrenzung	20
4.2.2. Downsampling	20
4.2.3. Gewichtung mit A-Filter	21
4.2.4. Barkbandfilterung	21
4.2.5. Elimination maskierter Bänder und Signalanteile	21
5. Resynthese	23
5.1. Klangmodellierung mit Residuum	23
5.2. Klangmodellierung mit Decaying-Exponentials	23
5.3. Synthese-Filterbank	28
6. Analyse-Beispiele	30
6.1. Gong Suwukan	33
6.2. Bonang Barung	33
6.3. Kempul	37
6.4. Peking A	37

Literaturverzeichnis

41

Abbildungsverzeichnis

1.1. Bildausschnitt eines typischen Gamelanorchesters (Quelle: IME-Homepage)	7
3.1. Flussdiagramm Bark-Zooming ARMA	11
3.2. Schematische Darstellung der Simultanmaskierung	12
3.3. Zusammenhang zwischen echter Frequenz f und Barkfrequenz z	13
3.4. Bark Filterkurven	14
3.5. Schematische Darstellung der Maskierung von T30 und Einschwingzeit	15
3.6. Vergleich mehrerer Residuen eines Saron Demung Instrumentes	18
3.7. Vergleich mehrerer Residuen eines Slenthem Instrumentes	18
4.1. FFT-Spektren mehrerer Originaklangbeispiele	19
4.2. Analyse-Filterbank	20
4.3. Bark Filterkurven 2. (blau) und 4. (pink) Ordnung	21
4.4. Bark Filterkurven 4. Ordnung (blau) und 8. Ordnung (pink)	22
4.5. Blockdiagramm der LPC-Analyse	22
5.1. Synthese mit Residuum	23
5.2. Analyse mit Decaying Exponentials	24
5.3. Zur Synthese herangezogene Teiltöne eines Bonang Panerus mit zugehöriger Startamplitude und Ausklingzeit	25
5.4. Teiltonfrequenzen und zugehörige Ausklingzeiten eines Bonang Panerus	26
5.5. Zur Synthese herangezogene Teiltöne eines Bonang Barung mit zugehöriger Startamplitude und Ausklingzeit	26
5.6. Teiltonfrequenzen und zugehörige Ausklingzeiten eines Bonang Barung	27
5.7. Synthese-Filterbank	28
5.8. Blockdiagramm der Teiltongeneration	29
6.1. Vergleich eines Originalklages des Instruments Bonang Panerus mit seinem resynthetisierten Adäquat	30
6.2. Vergleich eines Originalklages des Instruments Gong Suwukan mit seinem resynthetisierten Adäquat	31
6.3. Vergleich eines Originalklages des Instruments Bonang Barung mit seinem resynthetisierten Adäquat	31

6.4. Vergleich eines Originalklanges des Instruments Kempul mit seinem re-synthetisierten Adäquat	32
6.5. Vergleich eines Originalklanges des Instruments Slenthem mit seinem re-synthetisierten Adäquat	32
6.6. Detektierte Teiltöne des Gong Suwukan mit zugehöriger Startamplitude und Ausklingzeit	34
6.7. Detektierte Teiltonfrequenzen des Gong Suwukan mit zugehörigen Ausklingzeiten	35
6.8. Detektierte Teiltöne des Bonang Barung mit zugehöriger Startamplitude und Ausklingzeit	36
6.9. Detektierte Teiltonfrequenzen des Bonang Barung mit zugehörigen Ausklingzeiten	36
6.10. Detektierte Teiltöne des Kempul mit zugehöriger Startamplitude und Ausklingzeit	38
6.11. Detektierte Teiltonfrequenzen des Kempul mit zugehörigen Ausklingzeiten	38
6.12. Vergleich der FFT-Analyse des Originalklanges Pl PekingA 62 mit seinem resynthetisierten Adäquat	39
6.13. Detektierte Teiltöne des Peking A mit zugehöriger Startamplitude und Ausklingzeit	40
6.14. Detektierte Teiltonfrequenzen des Peking A mit zugehörigen Ausklingzeiten	40

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 gehörredundanten Anteilen befreite Signanteile verschiedener Gamelaninstrumente zu analysieren, ihre Frequenzkomponenten zu extrahieren, und die zu den Frequenzen jeweils zugehörige Startphasen, Startamplituden und Ausklingkonstanten zu ermitteln. Diese vier Parameter Frequenz, Startphase, Startamplitude und Ausklingkonstante, ermöglichen letztendlich eine näherungsweise realistische Synthetisierung von Gamelan-klängen 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.

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

Mit $\omega = 2\pi f_m / f_s$ als Modulationsfrequenz und f_s als Samplefrequenz. In der z -Ebene bedeutet dies ein Verschieben der Pole (und Nullstellen) z_i um einen Winkel ω_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 wird weiter eingeschränkt und das Signal tiefpassgefiltert. Die rotierte Impulsantwort wird dann um einem 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.

Um auf die gehörspezifischen Aspekte eingehen zu können, wird in dieser Projektarbeit die FZ-ARMA Methode um ein psychoakustisches Selektionsmodell erweitert (siehe Kapitel 3).

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 Datenreduktion und Bandzerlegung, der zweite Block beinhaltet eine LPC-Analyse aus der in weiterer Folge die Pole und somit alle gewünschten Parameter hervorgehen. Weiters wird im zweiten Block zusätzlich ein Residuum des Signals berechnet, das bei der späteren Synthese mit dem resynthetisierten Klang überblendet werden könnte.

3.1. Erster Block - 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 Innenohrs durch Berechnung der simultanen Maskierung bzw. der zeitlich variablen Mit Hörschwelle (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])

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.

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 welchem Lautstärkenverhältnis sie

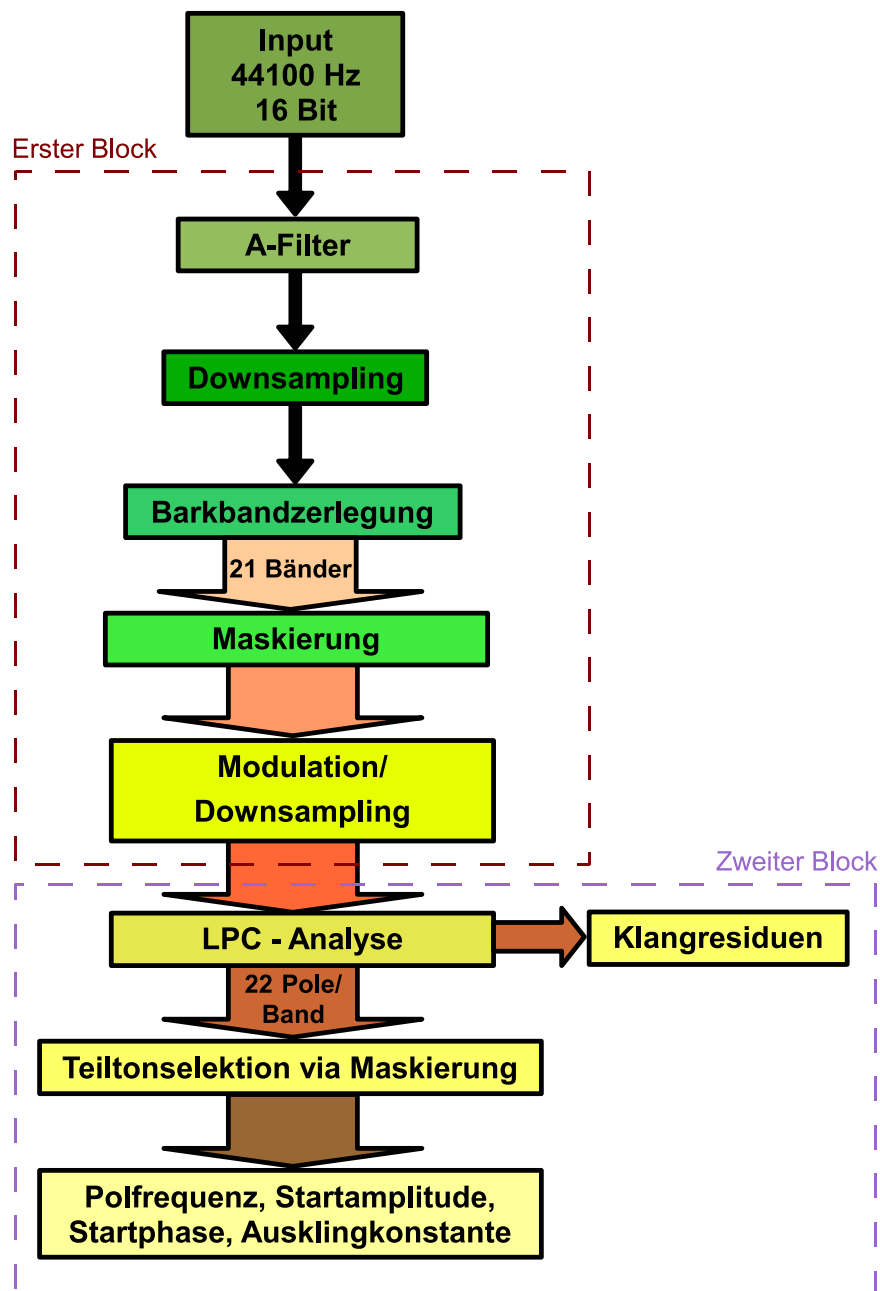


Abbildung 3.1.: Flussdiagramm Bark-Zooming ARMA

zu einander stehen. Begründet ist dieses Verhalten durch die angesteuerten Frequenzbereiche auf der Basilarmembran im Innenohr, die für benachbarte Frequenzen nahe bei einander liegen, wodurch zwei neben einander angeregte Schwingungen, ausgehend vom lautesten Ton, nur mit einer gewissen Dämpfung zu höheren und zu tieferen Frequenzen wahrgenommen werden. Zu höheren beträgt diese Dämpfung 24dB/Bark und zu tieferen Frequenzen 27dB/Bark.

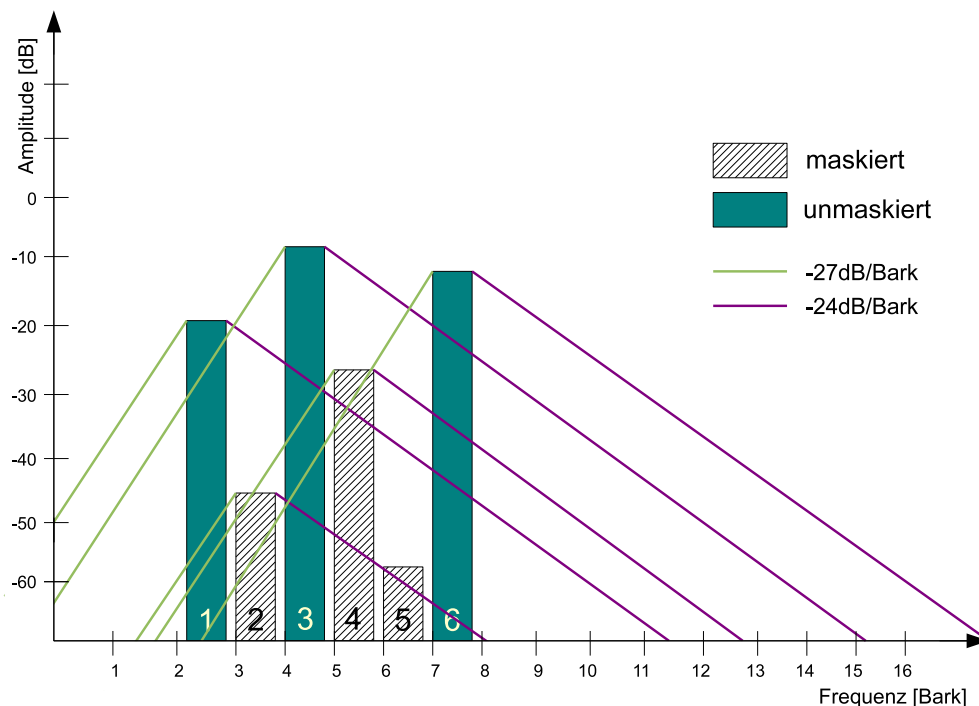


Abbildung 3.2.: Schematische Darstellung der Simultanmaskierung

Zusätzlich zu dieser Simultanmaskierung existiert noch die *zeitlich variable Mithörschwelle*. Sie beschreibt die Anpassung des Mittelohrs 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 so 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. Ein Bark ist 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. 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: 1,31Bark = 131Hz. 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.

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

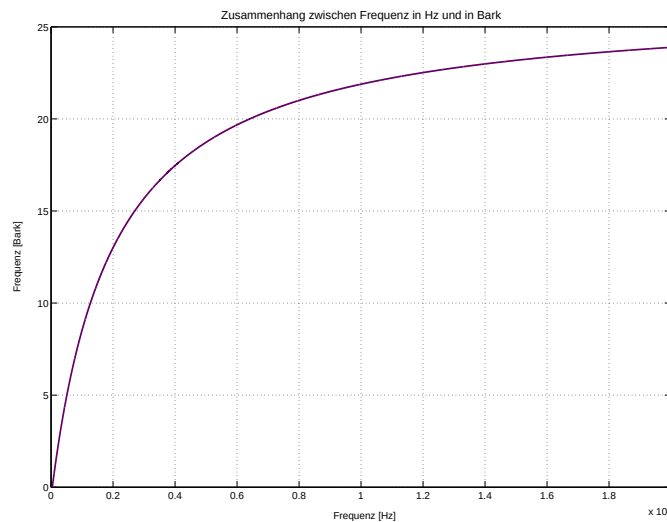


Abbildung 3.3.: Zusammenhang zwischen echter Frequenz f und Barkfrequenz z

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 echte untere Grenzfrequenz von 35Hz hatte. Ausgehend von diesen 35Hz wurden 21 Barkbänder errechnet.

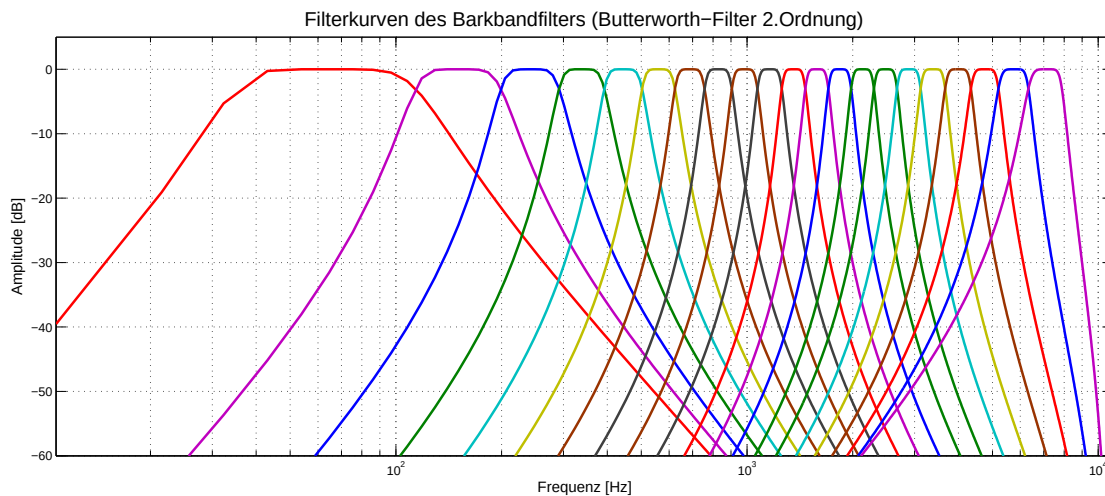


Abbildung 3.4.: Bark Filterkurven

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 weiß machen durch LPC (Residuum) in eine gehörgerechte Klangfarbe zu wandeln. Um die Maskierung anwenden zu können, wurde zu erst das Betragsquadrat der Einhüllenden jedes Bandes gebildet. Aus den Einhüllenden wurde für jedes Band eine Ausklingzeit T_{30} (die Zeit in der das Signal um 30dB abgefallen ist), sowie eine Einschwingzeit ermittelt.

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 gerechnet, 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 dem 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 3.5 stellt schematisch diesen Prozess der Maskierung über vorgegebene Zeitwerte dar. Band 1 ist während der gesamten Einschwingzeit und den Großteil der Ausklingzeit T_{30} 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.

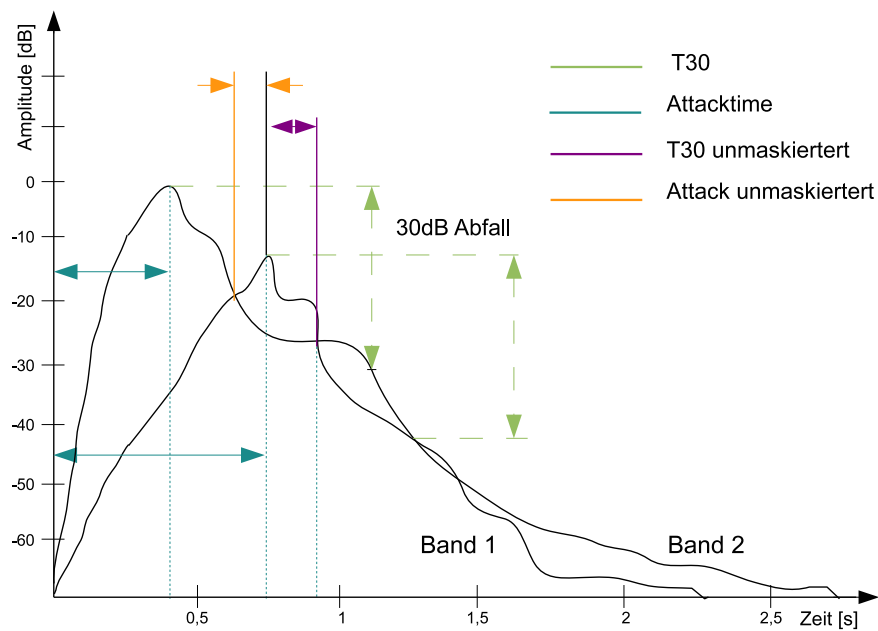


Abbildung 3.5.: Schematische Darstellung der Maskierung von T30 und Einschwingzeit

3.1.4. Modulation und Downsampling

Ausgehend von der Tatsache, dass jedes Barkband sich in einem anderen Frequenzbereich bewegt, war es nach Modulation eines jeden Bandes gegen Null möglich, ein Downsampling mit unterschiedlichen Faktoren vorzunehmen.

3.2. Zweiter Block - Analyse

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) einen 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:

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

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

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

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

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 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 3.6 und Abbildung 3.7 ist ersichtlich, dass die Residuen eine sehr hohe Teiltondichte besitzen und pro Instrument jeweils ähnliche Frequenzkomponenten aufweisen. Im Optimalfall wäre bei gleicher Anschlagtechnik eine absolute Übereinstimmung der einzelnen Residuen auf einem Instrument zu erwarten. Eine annähernde Übereinstimmung findet man in Abbildung 3.6 zwischen 600Hz und etwas über 1kHz.

3. BZ-ARMA Bark Zooming Autoregressive Moving-Average

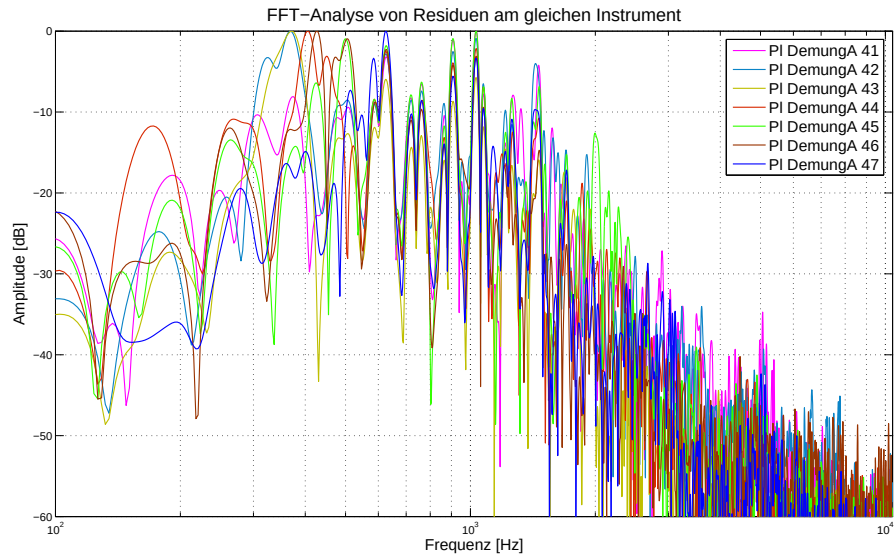


Abbildung 3.6.: Vergleich mehrerer Residuen eines Saron Demung Instrumentes

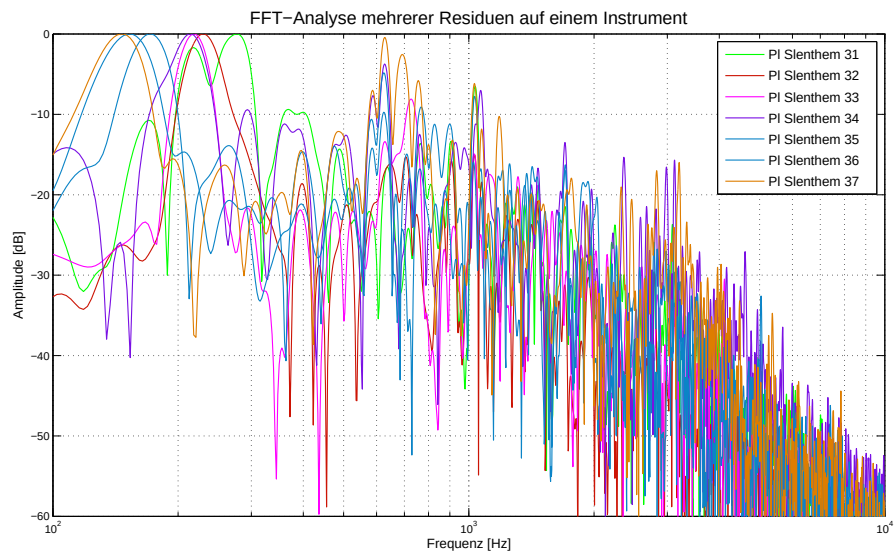


Abbildung 3.7.: Vergleich mehrerer Residuen eines Slenthem Instrumentes

4. Analyse-Programm

4.1. Das zu bearbeitende Material

Das Eingangssignal bestand zum Großteil aus Monosignalen. War es ein Stereosignal, so wurde der Einfachheit halber nur der linke Kanal in der Analyse berücksichtigt. Alle Soundfiles waren PCM Dateien mit 16Bit Wortbreite und einer Abtastrate von 44100Hz. Die Länge der Rohdaten bewegte sich zwischen 1 und 35 Sekunden.

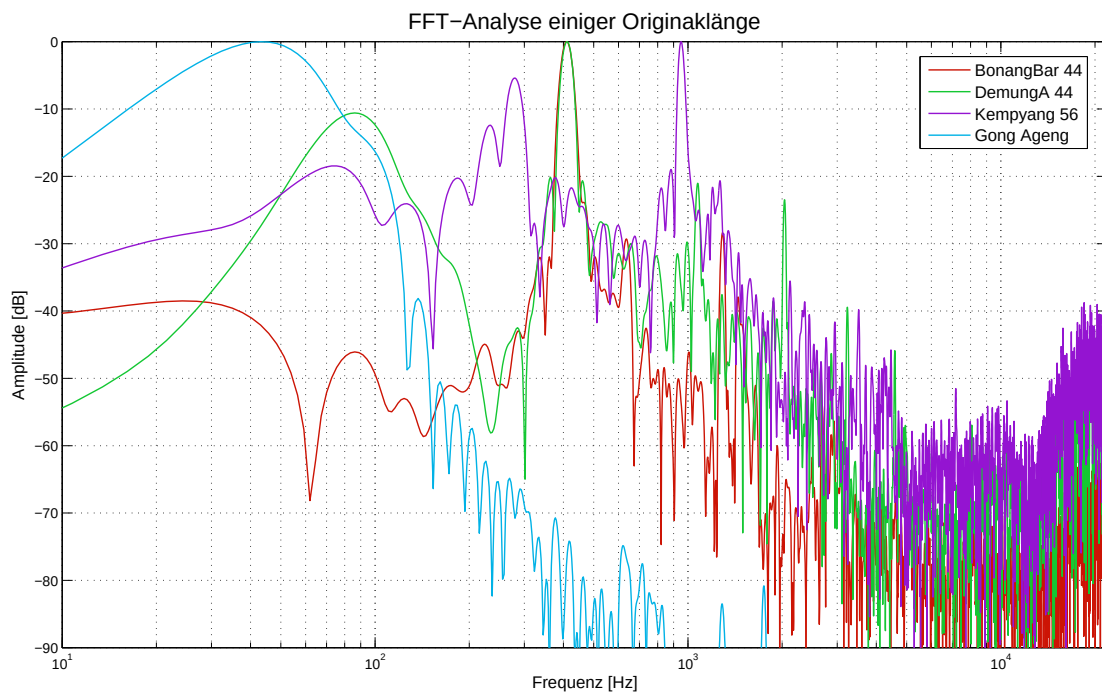


Abbildung 4.1.: FFT-Spektren mehrerer Originaklangbeispiele

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.

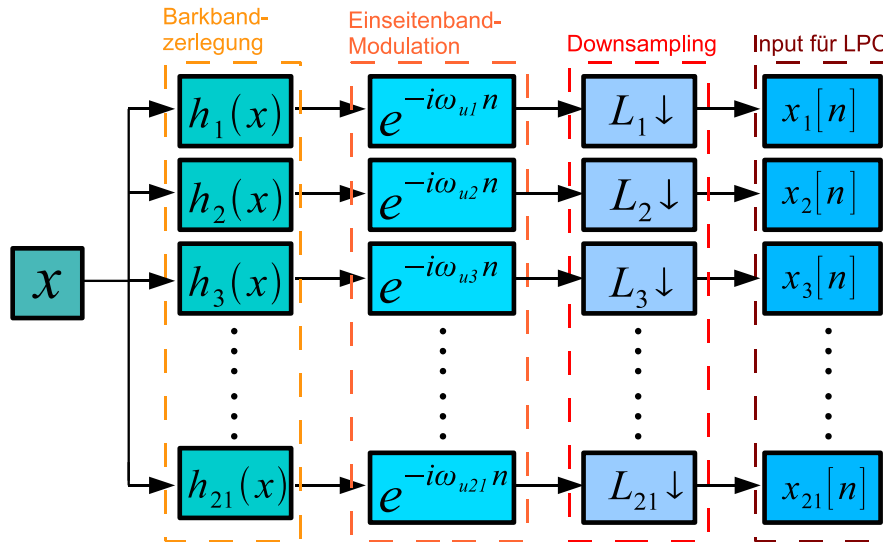


Abbildung 4.2.: Analyse-Filterbank

4.2.1. Inputbegrenzung

Um die Anzahl der zu verarbeitenden Samples zu senken, wurde das Eingangssignal auf maximal elf Sekunden begrenzt, was immer noch $44100 * 11 = 485100$ Samples ergibt. Jedoch wurde durch eben diese Begrenzung ein Berechnen der Klänge auf dem oben bereits erwähnten System möglich.

4.2.2. Downsampling

Da sich die für den Klangcharakter wesentlichen Frequenzanteile unter 8kHz befinden, konnte von einer Samplerate von 44100Hz auf 22050Hz downgesampelt werden, ohne wichtige Informationen für die Analyse zu verlieren. Vor dem Downsampling wurde das Signal selbstverständlich durch ein Antialiasing-Filter geschickt.

4.2.3. Gewichtung mit A-Filter

Wie bereits in 3.1.1 erklärt, wird in diesem Schritt das bereits um den Faktor 2 downgesamplete Eingangssignal, mit einer bewerteten A-Filterkurve an die für das menschliche Gehör übliche Frequenzgewichtung angepasst.

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 3.4), das um die Phase des gefilterten Signals stabil zu halten, einmal vorwärts und einmal rückwärts filtert, was die Flankensteilheit auf die eines Filters 4. Ordnung erhöht.

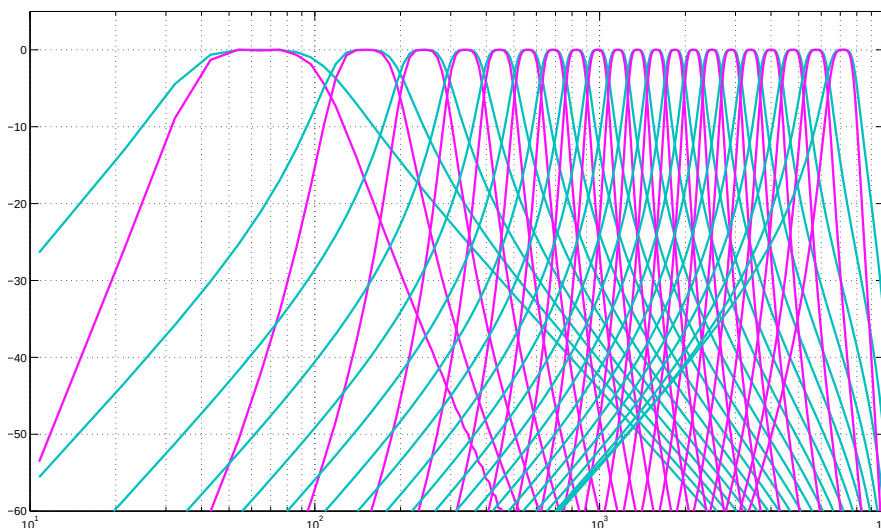


Abbildung 4.3.: Bark Filterkurven 2. (blau) und 4. (pink) Ordnung

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 einstellt (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 gehör-

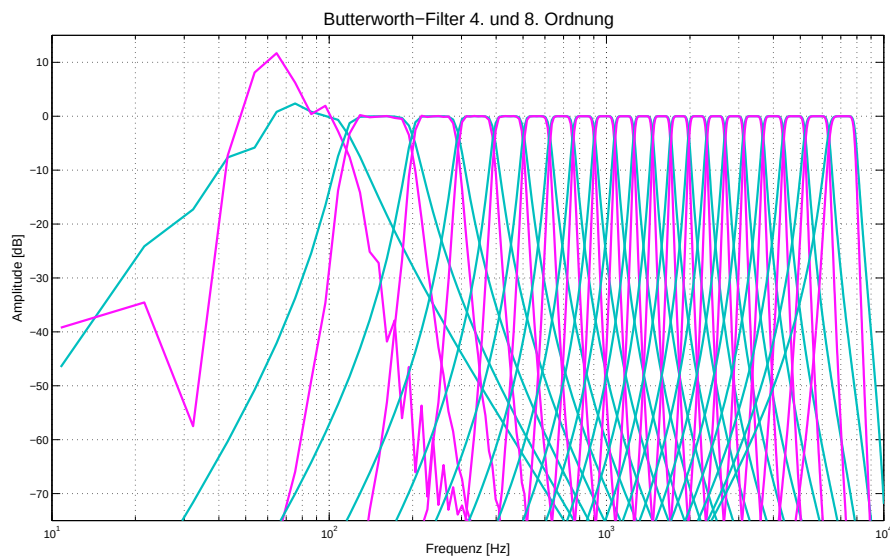


Abbildung 4.4.: Bark Filterkurven 4. Ordnung (blau) und 8. Ordnung (pink)

relevanten Anteile übrig bleiben. Aus diesen errechneten Anteilen wird anschließend eine LPC-Analyse (siehe Kapitel 3.2.1) mit 22 Polen pro eruiertem Barkband durchgeführt.

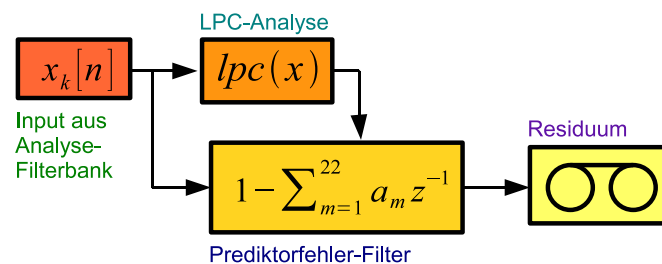


Abbildung 4.5.: Blockdiagramm der LPC-Analyse

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

5. Resynthese

5.1. Klangmodellierung mit Residuum

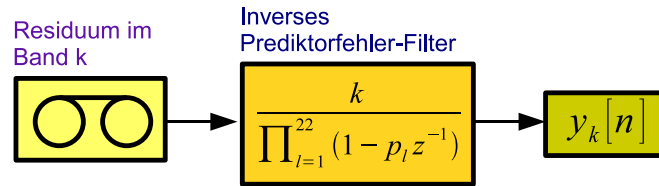


Abbildung 5.1.: Synthese mit Residuum

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 Θ_m aufweisen.

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

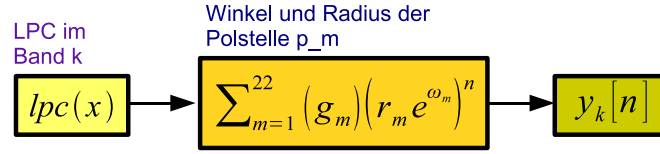


Abbildung 5.2.: Analyse mit Decaying Exponentials

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

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

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:

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

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

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

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

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 detektieren Teiltöne des Signals mit jeweiliger Startamplitude und Ausklingkonstante. Dabei liegt das Hauptaugenmerk bei den dreidimensionalen Plots auf den Verhältnissen der Teiltöne in

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

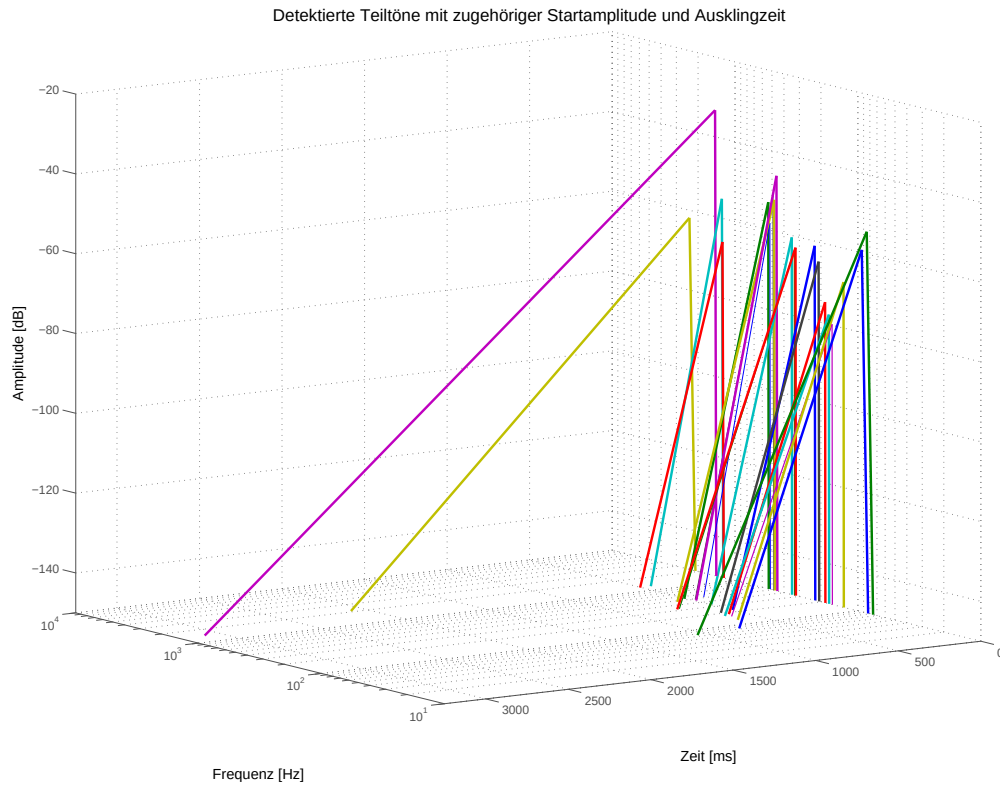


Abbildung 5.3.: Zur Synthese herangezogene Teiltöne eines Bonang Panerus mit zugehöriger Startamplitude und Ausklingzeit

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, sind 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

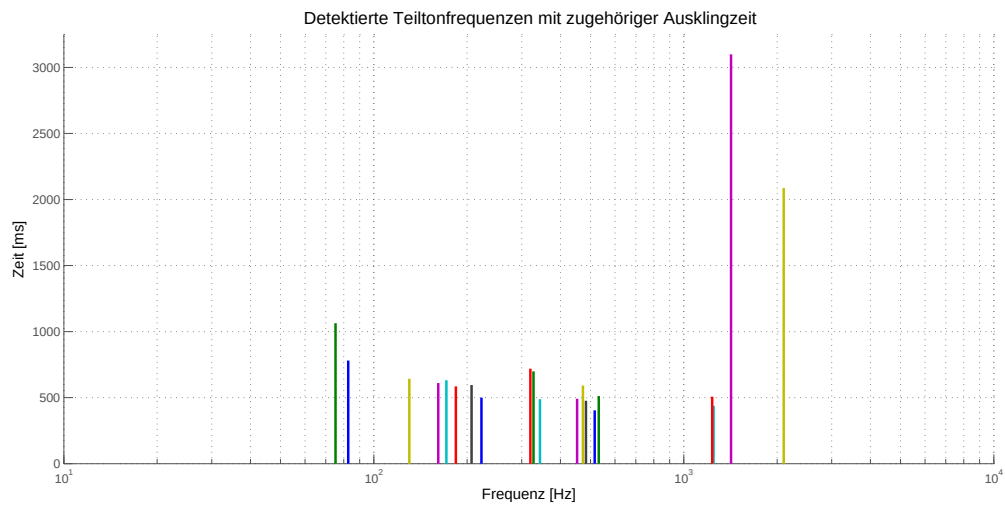


Abbildung 5.4.: Teiltonfrequenzen und zugehörige Ausklingzeiten eines Bonang Panerus

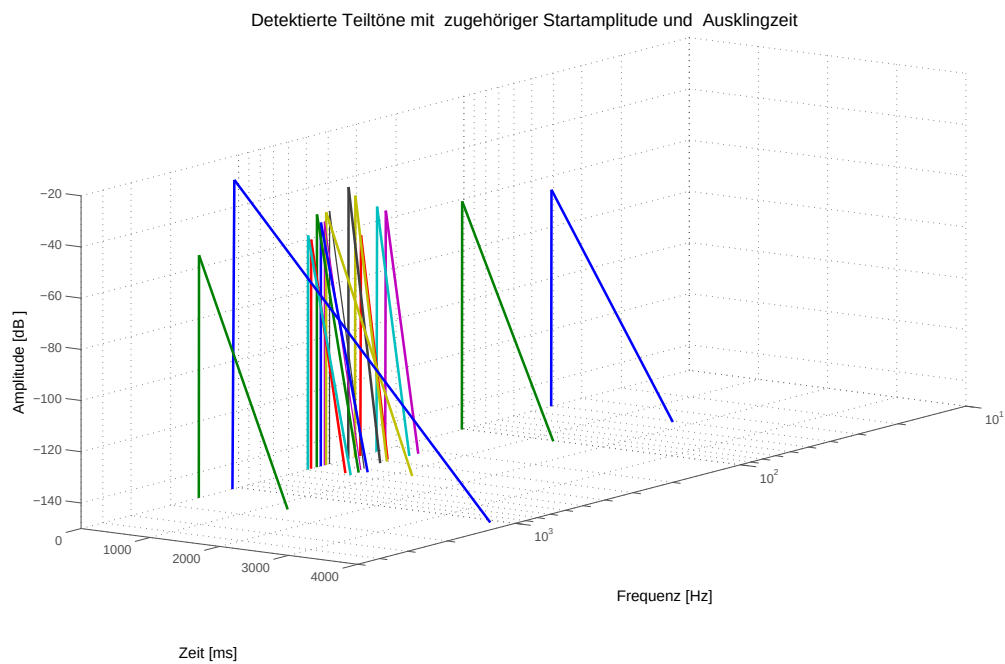


Abbildung 5.5.: Zur Synthese herangezogene Teiltöne eines Bonang Barung mit zugehöriger Startamplitude und Ausklingzeit

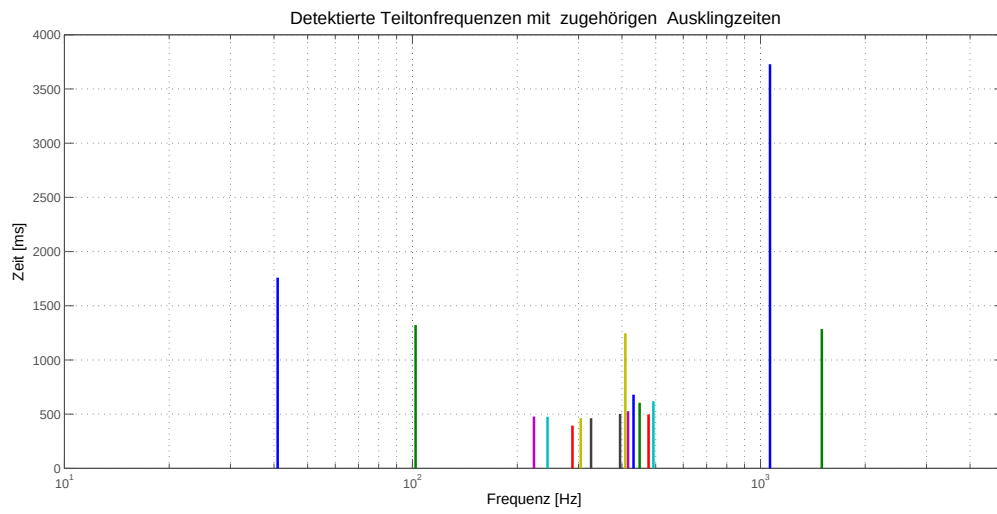


Abbildung 5.6.: Teiltonfrequenzen und zugehörige Ausklingzeiten eines Bonang Barung

Ohr dem Original sehr ähnlich empfunden.

5.3. Synthese-Filterbank

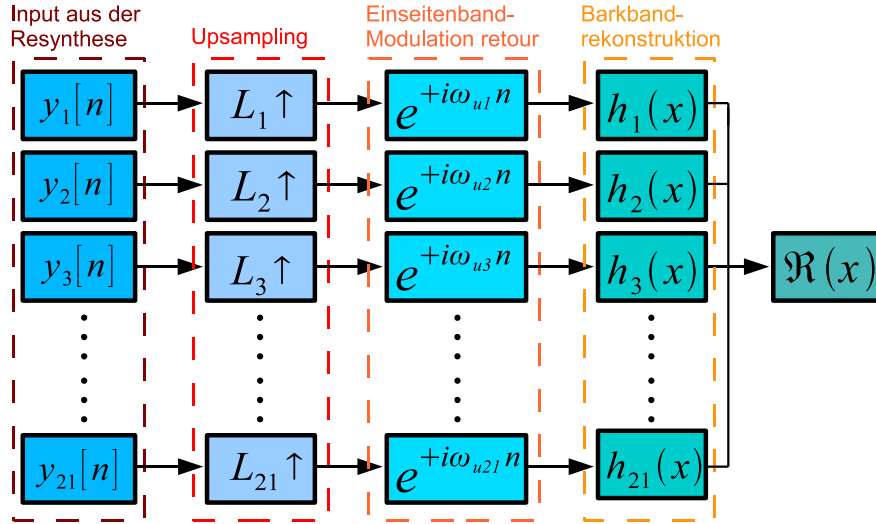


Abbildung 5.7.: Synthese-Filterbank

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 Herabmodulieren verwendeten, speziellen Modulationsfrequenz ω_{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. 3.4) gefiltert um wirklich nur die bandeigenen Anteile in das Syntheseergebnis mit einzubeziehen. 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} (Ausklingzeit), auf die Sinustöne aufgeprägt werden. Alle so generierten Frequenzkomponenten gemeinsam ergeben den fertige Klang.

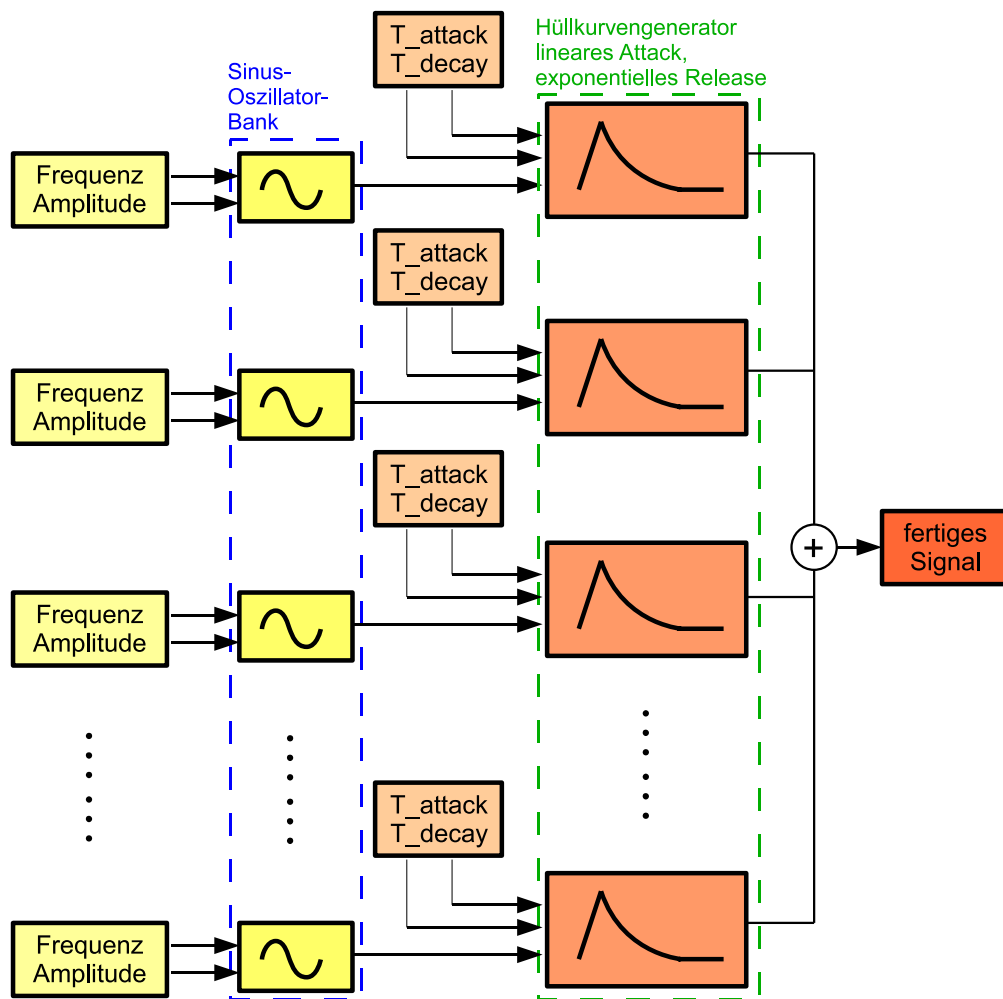


Abbildung 5.8.: Blockdiagramm der Teiltongeneration

6. Analyse-Beispiele

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

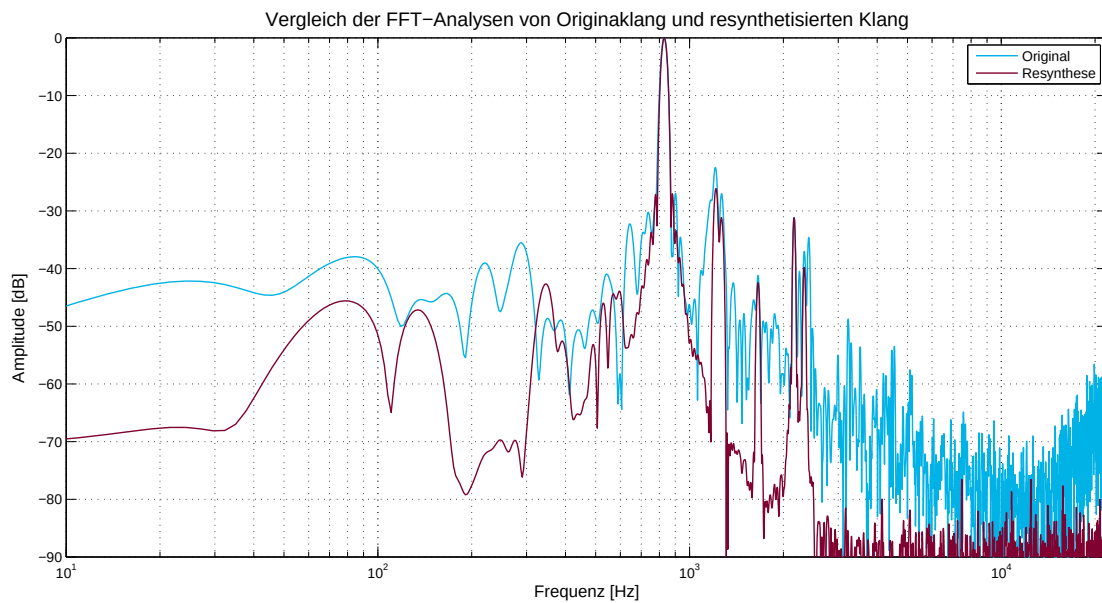


Abbildung 6.1.: Vergleich eines Originalklanges des Instruments Bonang Panerus mit seinem resynthetisierten Adäquat

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 des selben Instruments. Man kann recht gut erkennen, dass diese ausgewählten Beispiele eine relativ hohe Übereinstimmung zwischen den Frequenzkomponenten des Original- und Syntheseklanges aufweisen.

6. Analyse-Beispiele

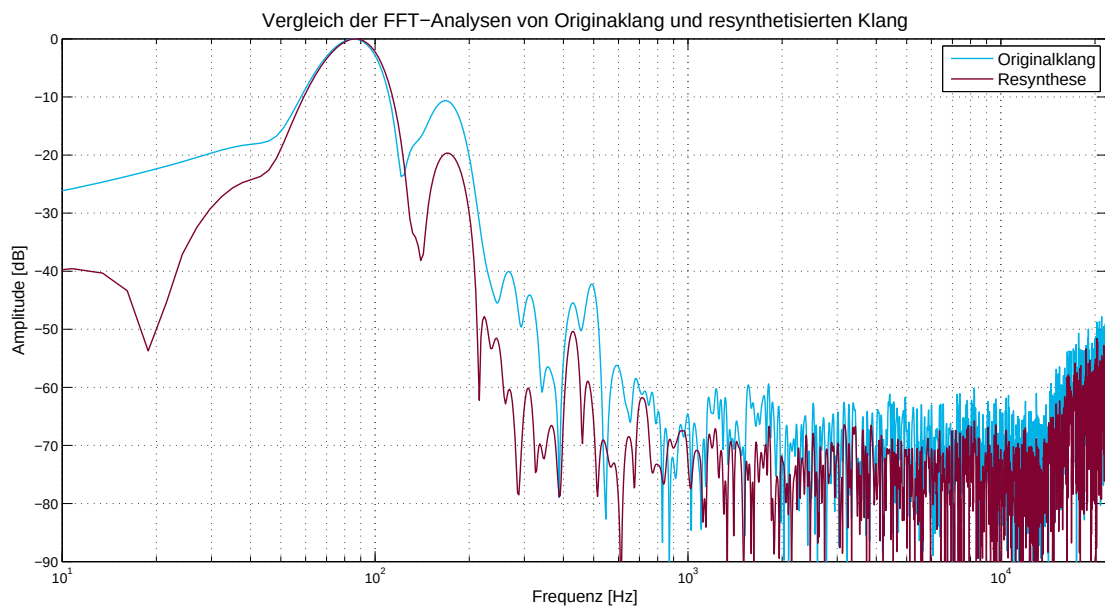


Abbildung 6.2.: Vergleich eines Originalklanges des Instruments Gong Suwukan mit seinem resynthetisierten Adäquat

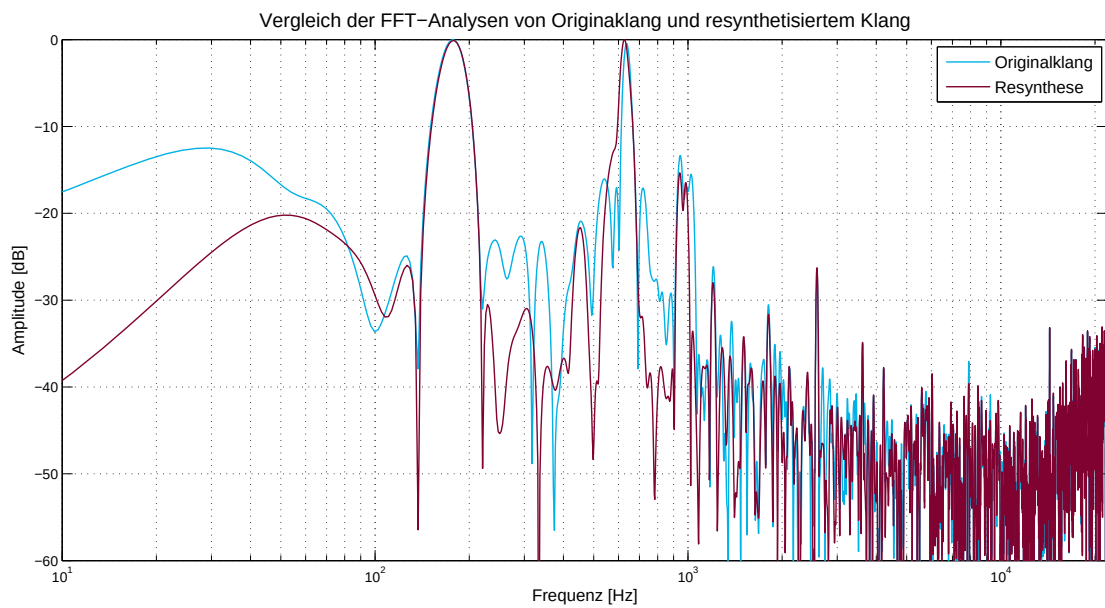


Abbildung 6.3.: Vergleich eines Originalklanges des Instruments Bonang Barung mit seinem resynthetisierten Adäquat

6. Analyse-Beispiele

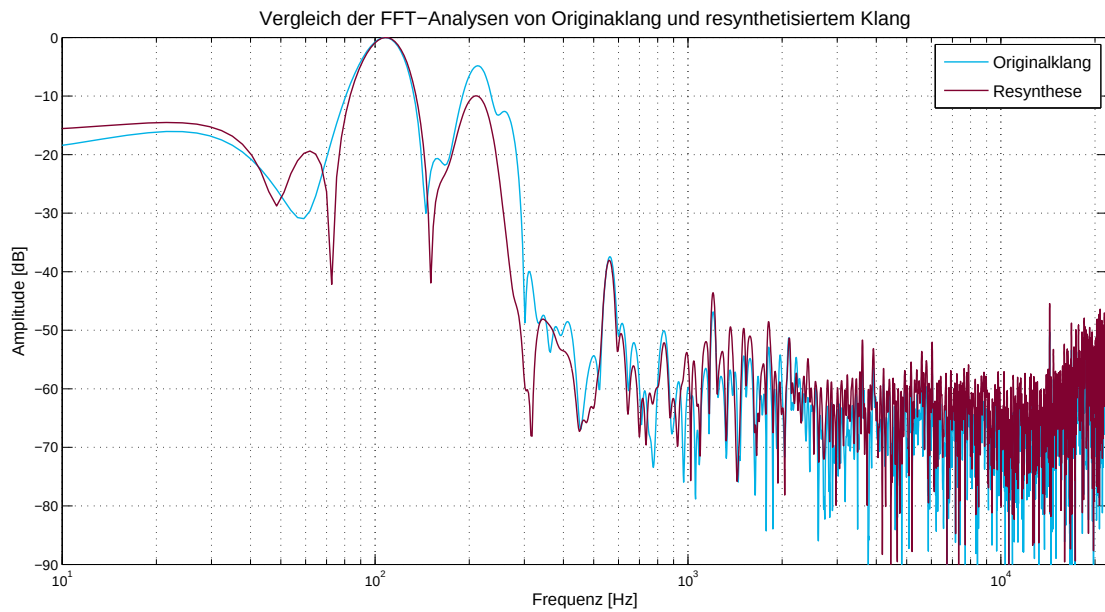


Abbildung 6.4.: Vergleich eines Originalklanges des Instruments Kempul mit seinem resynthetisierten Adäquat

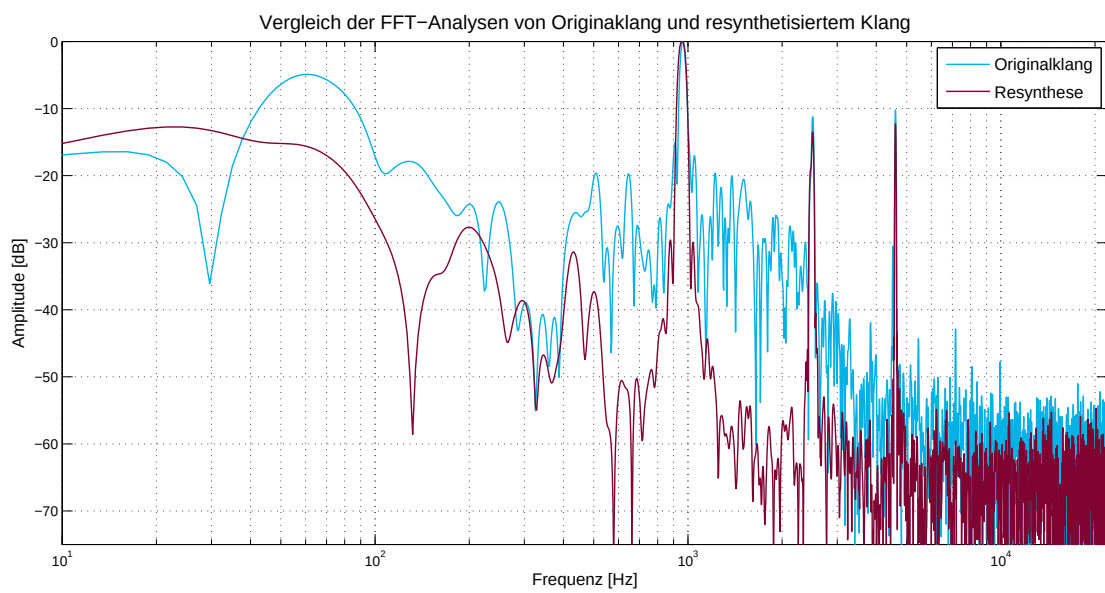


Abbildung 6.5.: Vergleich eines Originalklanges des Instruments Slenthem mit seinem resynthetisierten Adäquat

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 bei etwa 165Hz. Man sieht eine starke Übereinstimmung des resynthetisierten Klanges mit dem Original bei der Frequenzausprägung bei 82Hz. Die Übereinstimmung bei 165Hz ist nicht ganz so gut, 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. Abbildung 6.7 zeigt eben diese beiden starken Frequenzspitzen bei 82Hz und

Frequenz [Hz]	Amplitude [dB]	Phase [°]	Einschwingzeit [ms]	Ausklindauer [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
528.53	-92.5	-240	11	3597
514.39	-93.1	25	11	2475
739.42	-82.5	-38	9	2186
721.91	-77.7	-170	9	2995
697.64	-73.1	-71	9	2556

Tabelle 6.1.: Ermittelte Parameter des Audiofiles Pl GongSuw 22

108Hz mit ihren Ausklingkonstanten. Als am längsten ausklingende Komponenten heben sie sich deutlich von den restlichen als charaktergebende ab. 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 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.

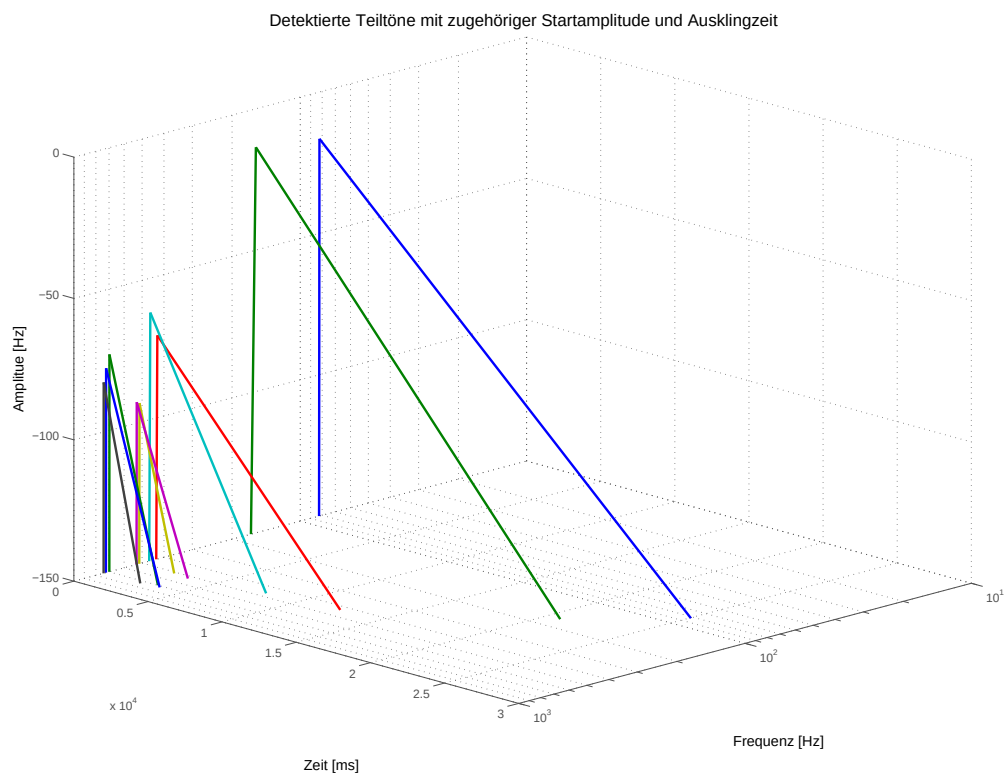


Abbildung 6.6.: Detektierte Teiltöne des Gong Suwukan mit zugehöriger Startamplitude und Ausklingzeit

6. Analyse-Beispiele

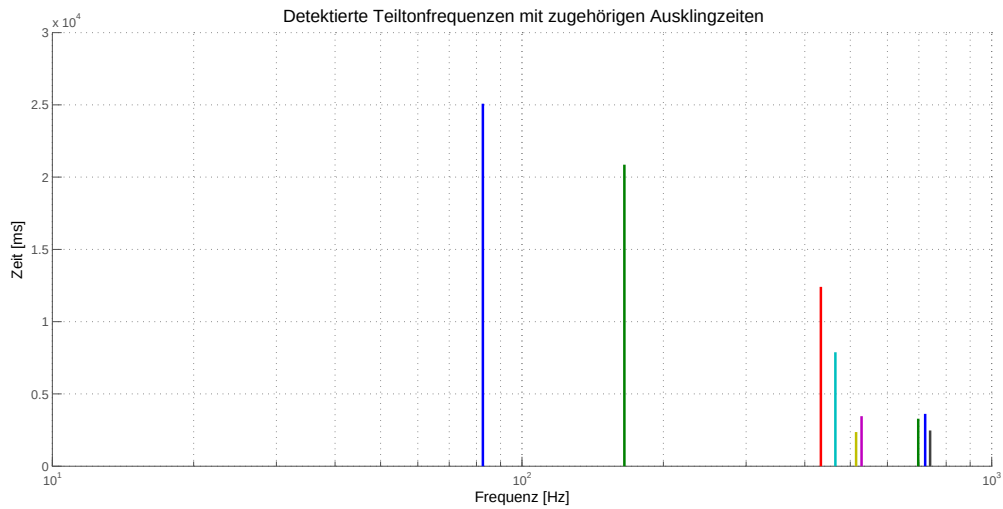


Abbildung 6.7.: Detektierte Teiltonfrequenzen des Gong Suwukan mit zugehörigen Ausklingzeiten

Frequenz [Hz]	Amplitude [dB]	Phase [°]	Einschwingzeit [ms]	Ausklingdauer [ms]
111.08	-87.1	7	6	664
43.99	-61.7	264	6	725
50.86	-68.3	161	6	755
74.70	-66.8	-81	6	739
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 Sl BonangBar 52

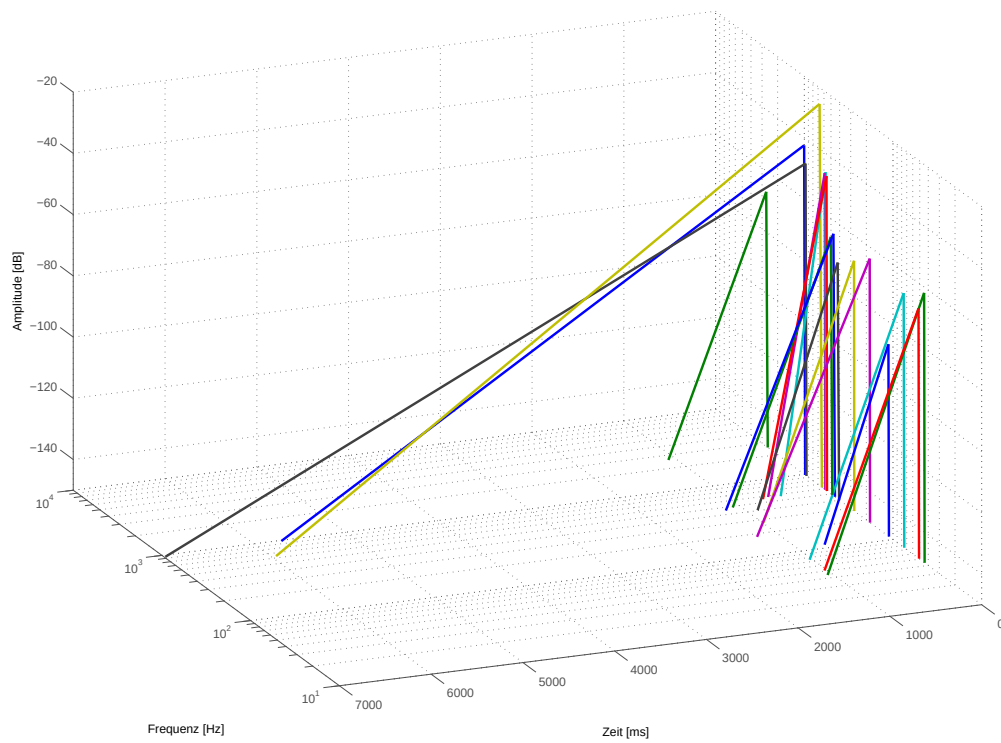


Abbildung 6.8.: Detektierte Teiltöne des Bonang Barung mit zugehöriger Startamplitude und Ausklingzeit

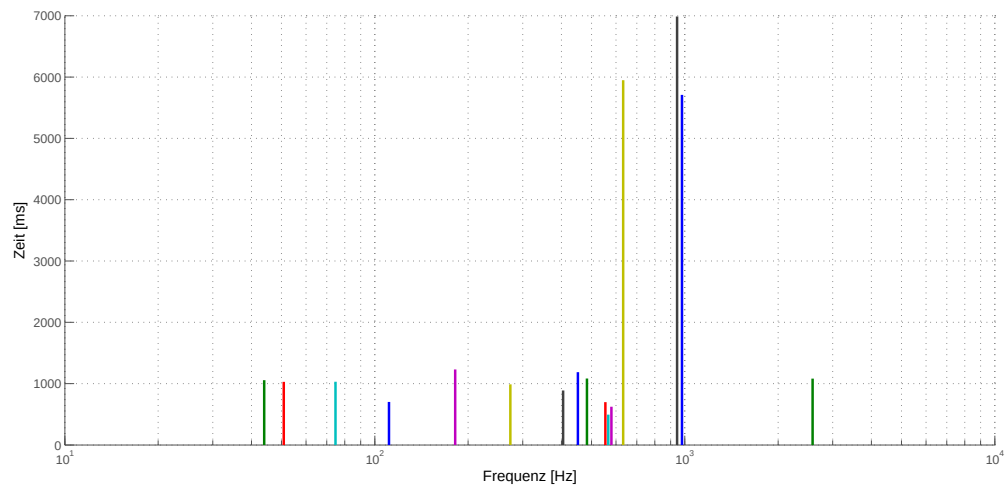


Abbildung 6.9.: Detektierte Teiltonfrequenzen des Bonang Barung mit zugehörigen Ausklingzeiten

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 Gong Suwukan, dessen Spektrum sich in einem ähnlichen Bereich bewegt - siehe Abbildung 6.2).

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
482.12	-83.7	-171	205	3135
417.70	-54.9	-47	205	2037
614.16	-62.5	97	91	3380
597.17	-69.9	-64	91	1815
556.87	-59.4	-155	91	2512
569.69	-63.0	73	91	3004
847.34	-71.4	5	3	3143

Tabelle 6.3.: Ermittelte Parameter des Audiofiles SI Kempul 25

6.4. Peking A

Das Analysebeispiel P1_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 der 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 enttarnt und somit von weiteren Rechenschritten ausgeschlossen.

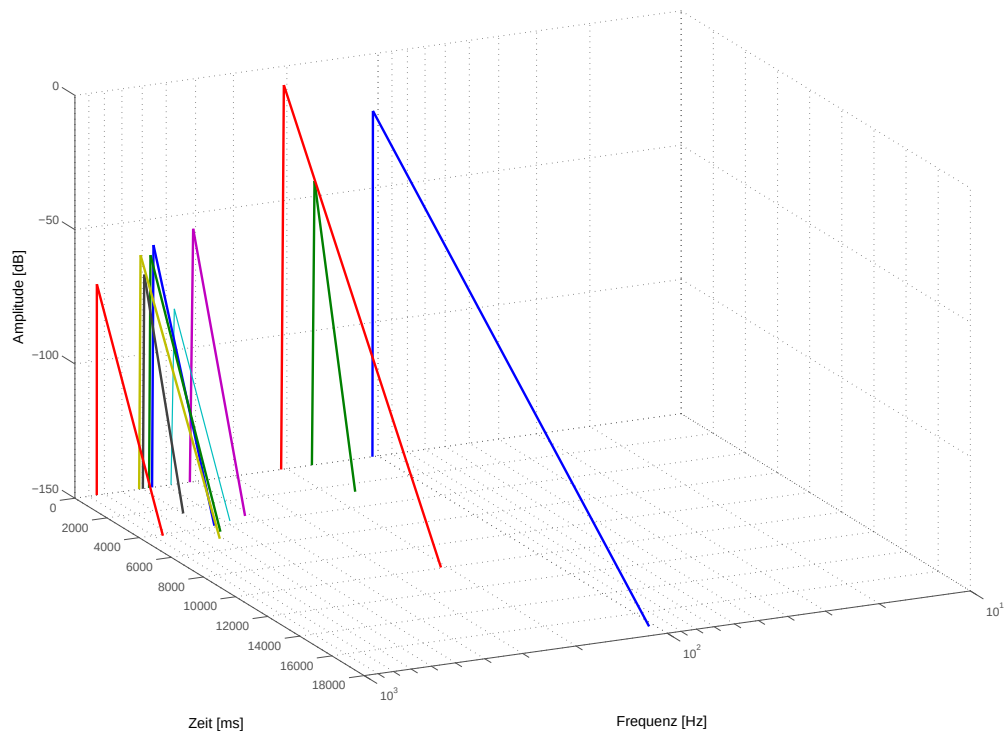


Abbildung 6.10.: Detektierte Teiltöne des Kempul mit zugehöriger Startamplitude und Ausklingzeit

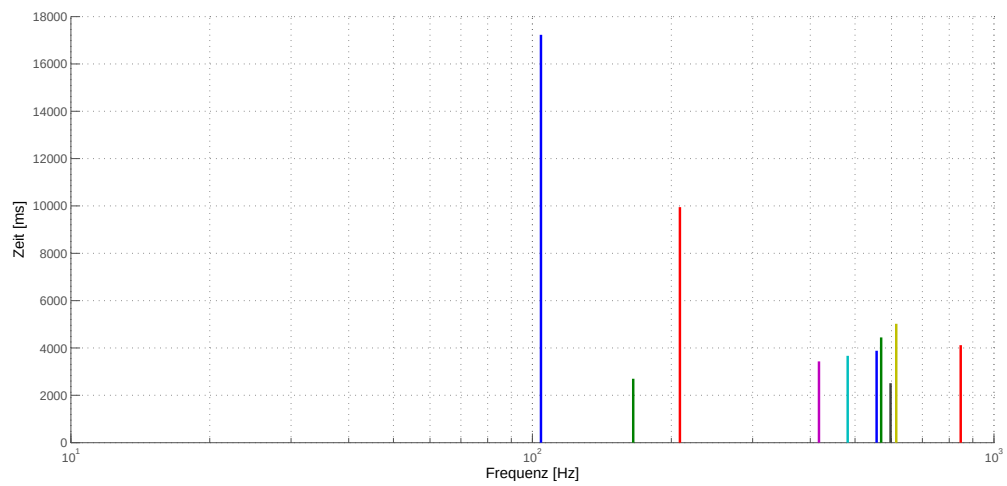


Abbildung 6.11.: Detektierte Teiltonfrequenzen des Kempul mit zugehörigen Ausklingzeiten

Frequenz [Hz]	Amplitude [dB]	Phase [°]	Einschwingzeit [ms]	Ausklingdauer [ms]
1321.34	-40.3	-93	4	4390
3327.83	-41.1	64	8	831

Tabelle 6.4.: Ermittelte Parameter des Audiofiles Pl PekingA 62

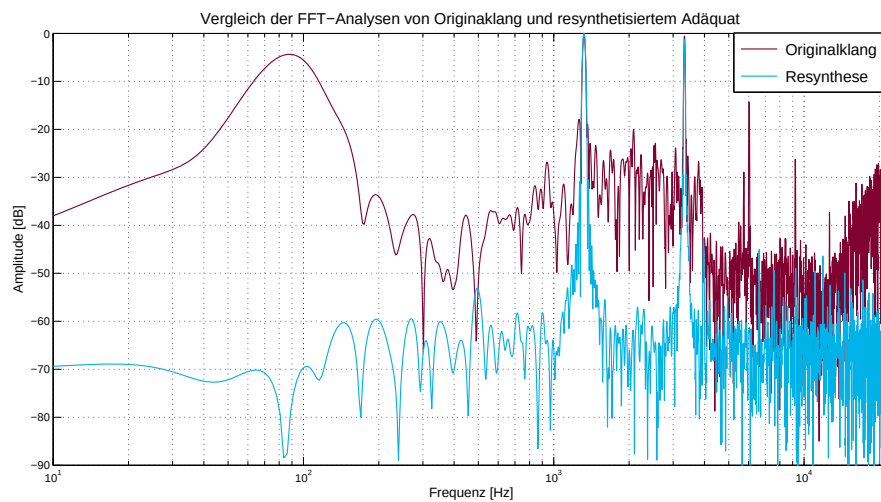


Abbildung 6.12.: Vergleich der FFT-Analyse des Originalklanges Pl PekingA 62 mit seinem resynthetisierten Adäquat

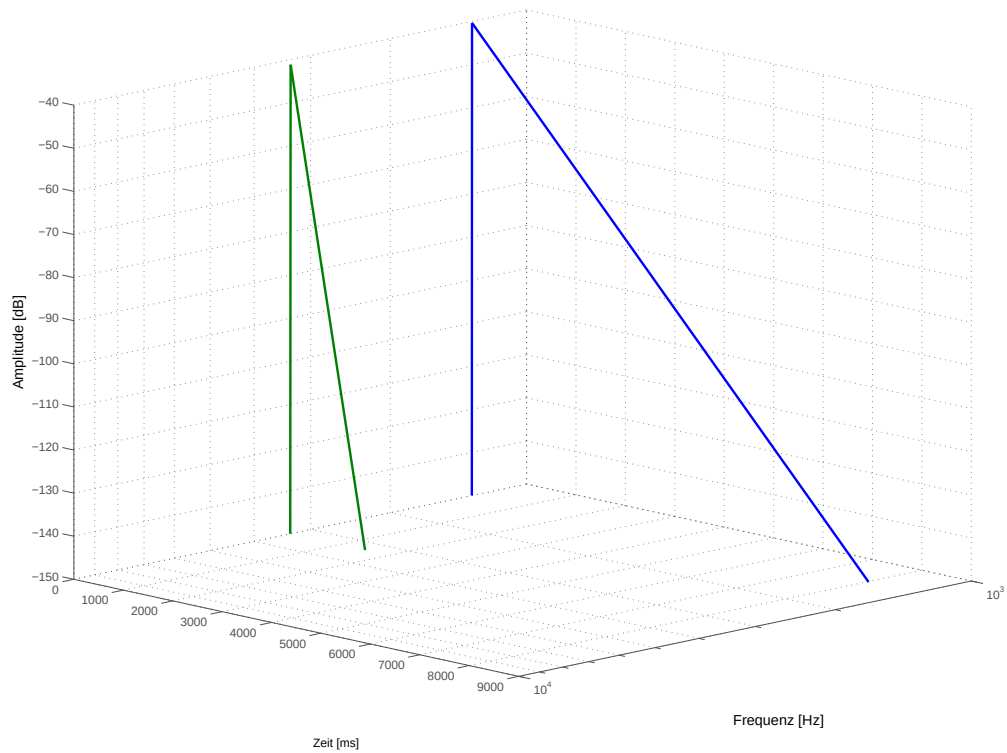


Abbildung 6.13.: Detektierte Teiltöne des Peking A mit zugehöriger Startamplitude und Ausklingzeit

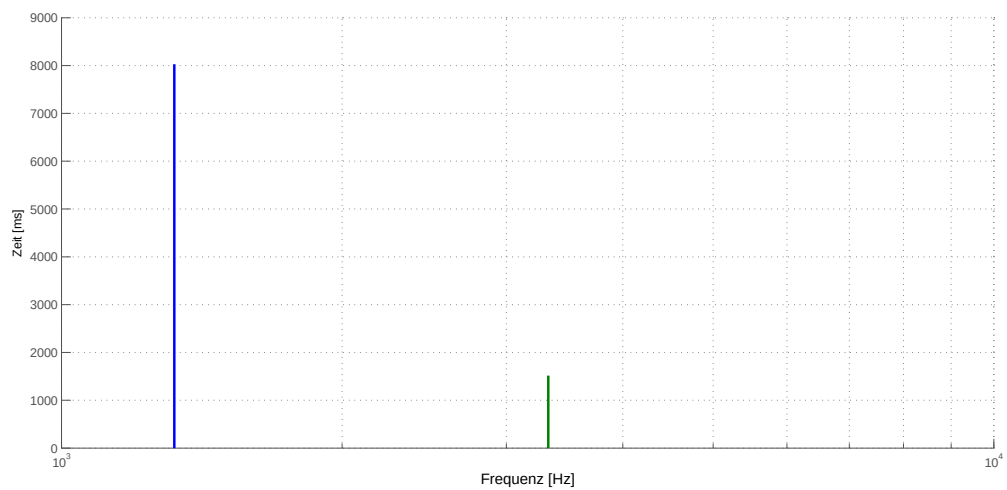


Abbildung 6.14.: Detektierte Teiltonfrequenzen des Peking A mit zugehörigen Ausklingzeiten

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, 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 Tonstudiotechnik“, 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