PASCAL KAAP

SOUND DESIGN OF MODAL SYNTHESIS MODELS

SOUND DESIGN OF MODAL SYNTHESIS MODELS

AUTOMATIC CALIBRATION OF MODAL SYNTHESIS MODELS FOR SOUND SYNTHESIS

MASTERARBEIT

VERFASST UND VORGELEGT VON

PASCAL KAAP

[redacted]

AM 09.02.14

STUDIENGANG: PHYSIKALISCHE INGENIEURWISSENSCHAFT (MASTER)

Fakultät I - Geisteswissenschaften
Institut für Sprache und Kommunikation
TU Berlin

BETREUER:

Prof. Dr. Weinzierl
Dr. Edgar Berdahl

Pascal Kaap: *Sound Design of Modal Synthesis Models*, Februar 2014

ABSTRACT

This master's thesis is a study in sound synthesis based on the physical properties of musical instruments, which is known as „physical modeling". The focus of the model generation is on modal synthesis of the examined objects. In prior work, the modeling quality was limited due to the low computational power of consumer devices, so that necessary model simplifications were unavoidable for real-time applications, which usually had a negative effect on sound-quality. The ever-increasing computing capabilities allow for more accurate real-time modeling and as a result, more complex sound structures. The main part of this work consists of the programming of sound analysis scripts and the implementation of physical models for the sound synthesis software *Synth-A-Modeler*. With the help of analysis scripts in *MATLAB*, new physical models are calibrated. The calibrations are primarily based on sound examples of real musical instruments. In addition, calibration of analytically derived model parameters is carried out. To ensure the synthesis is as efficient as possible, the models are trimmed according to psychoacoustic methods. With the resulting physical models, 120 sounds were created, which can be synthesized in real-time. While linear sound bodys could be re-synthesized accurately, it was not possible to develop satisfactory models for non-linear objects. This work is limited to the simulation of structural vibrations of the modeled objects. The simulation of the radiation is not part of this work.

ZUSAMMENFASSUNG

Die vorliegende Master-Arbeit befasst sich mit der Klangsynthese auf Grundlage physikalischer Eigenschaften von Musikinstrumenten, der sogenannten „Physikalischen Modellierung". Der Fokus der Modellgenerierung liegt auf der Modalsynthese der zu untersuchenden Objekte. In früheren Arbeiten war die Modellierung aufgrund der niedrigen Rechenleistung von Endverbrauchergeräten deutlich beschränkt, so dass notwendige Modellvereinfachungen unumgänglich waren, was sich meist negativ auf die Klangqualität auswirkte. Die immer weiter steigenden Rechenkapazitäten lassen exaktere Modellierungen und im Ergebnis komplexere Klangkonstruktionen zu. Der Hauptbestandteil dieser Arbeit besteht in der Programmierung von Klanganalyseskripten und der Implementierung physikalischer Modelle für die Sound-Synthese Software *Synth-A-Modeler*. Mit Hilfe der in *MATLAB* angefertigten Analyse-Skripten, werden neue physikalische Modellsimulationen kalibriert. Grundlage der Kalibrierung bilden in erster Linie Klangbeispiele von realen Musikinstrumenten. Weiterhin werden Kalibrierungen anhand von analytisch hergeleiteten Modellparametern vorgenommen. Um die Synthese so effizient wie möglich zu gestalten, werden die Modelle anhand von psychoakustischen Methoden bearbeitet. Mit Hilfe der physikalischen Modelle wurden insgesamt 120 Klangsynthesen erstellt, welche in Echtzeit spielbar sind. Lineare Klangkörper konnten sehr gut synthetisiert werden, während für nicht lineare Objekte keine zufriedenstellenden Modelle

entwickelt werden konnten. Diese Arbeit ist beschränkt auf die Simulation der Bewegungsgesetze der modellierten Objekte. Die Simulation der Abstrahlung ist nicht Bestandteil dieser Arbeit.

INHALTSVERZEICHNIS

i EINLEITUNG 1

1 EINLEITUNG 3

ii GRUNDLAGEN 7

2 GRUNDLAGEN 9

2.1 Physikalische Grundlagen . . . 9

2.1.1 Impulsanregung . . . 9

2.1.2 Resonatoren . . . 9

2.1.3 Lineare Schwinger mit einem Freiheitsgrad . . . 10

2.1.4 Lineare Schwinger mit N Freiheitsgraden . . . 12

2.1.5 Nichtlineare Schwinger . . . 14

2.1.6 Dämpfung schwingender Strukturen . . . 15

2.2 Mathematische Grundlagen . . . 18

2.2.1 Fourier Transformation . . . 18

2.2.2 Schnelle Fourier Transformation . . . 18

2.2.3 Kurzzeit Fourier Transformation . . . 18

2.2.4 Spektrale Analyse von Signalen . . . 19

2.2.5 Energy Decay Relief . . . 21

2.2.6 Quadratische Peak Interpolation . . . 23

2.2.7 Generieren eines minimalphasigen Signals . . . 24

2.3 Psychoakustische Grundlagen . . . 25

2.3.1 Gehörrichtige Frequenzskala . . . 25

2.3.2 Signalleistung und Schalldruck im Band . . . 26

2.3.3 Absolute Hörschwelle . . . 26

2.3.4 Mithörschwelle . . . 26

iii KLANGMODELL 29

3 KLANGMODELL 31

iv METHODEN 35

4 METHODEN 37

4.1 Analyse der Klangbeispiele . . . 37

4.2 Gesamtschau der Signalanalyse . . . 37

4.3 Signalvorbearbeitung . . . 39

4.4 Spektralanalyse . . . 39

4.5 Modenauswahl . . . 41

4.5.1 Stutzen der Moden . . . 42

4.6 Bestimmung der modalen Eigenschaften . . . 45

4.6.1 Präzisieren der modalen Frequenz . . . 45

4.6.2 Bestimmung der Abklingzeit . . . 46

4.6.3 Bestimmung der Amplitude . . . 49

4.6.4 Analyse und Synthese von Schwebungen . . . 52

4.7 Signalnachbearbeitung . . . 54

4.8 Analyse und Synthese nichtlinearer Modelle . . . 55

4.9 Bestimmung des frequenzabhängigen Dämpfungsfaktors 57
4.10 Erstellen von analytischen Modellen 57

V ERGEBNISSE UND AUSWERTUNG 59
5 ERGEBNISSE UND AUSWERTUNG 61
5.1 Spektrale Analyse 61
5.2 Modendetektierung 62
5.3 Stutzen nicht hörbarer Moden 64
5.4 Hüllkurvenannäherung 64
5.5 Resynthese von Schwebungen 67
5.6 Analyse von nichtlinearen Modellen 69
5.7 Synthese von Klangsignalen 71
5.7.1 Sinussignal 72
5.7.2 Weißes Rauschen 74
5.7.3 Tom 77
5.7.4 Becken 78
5.7.5 Bassdrum 81
5.7.6 Snaredrum 83
5.7.7 Clave 85
5.7.8 Glocke 87
5.7.9 Klaviersaite 89
5.7.10 Marimba 91
5.7.11 Akustikgitarre 92
5.7.12 Kantilan 94

VI FAZIT UND AUSBLICK 97
6 DISKUSSION UND AUSBLICK 99
6.1 Diskussion 99
6.2 Ausblick 101

VII ANHANG 103
A ANHANG 105
A.1 Daten CD 105
A.2 Einteilung in gehörrichtige Frequenzgruppen 106
A.3 Berechnungsvorschriften analytischer Resonanzfrequenzen . . . 106
A.3.1 Resonanzfrequenzen Stab 106
A.3.2 Resonanzfrequenzen vorgespannte Saite 107
A.3.3 Resonanzfrequenzen vorgespannte rechteckige Platte . . . 107
A.3.4 Resonanzfrequenzen runde Platte 108
A.3.5 Resonanzfrequenz elliptische Platte 108
A.3.6 Resonanzfrequenz Holzstrukturen 108
A.3.7 Resonanzfrequenz Zylinderschale 109
A.3.8 Resonanzfrequenz Rundmembran 109
A.4 User interface guide for Mode-A-Lizer 110
A.4.1 Audiopreprocessing 110
A.4.2 Spectral Analysis 112
A.4.3 Mode Detection 112
A.4.4 Mode Properties Detection 114
A.4.5 Mode Trimming 116

A.4.6 Supplementary notes 118

LITERATURVERZEICHNIS 119

ABBILDUNGSVERZEICHNIS

Abbildung 1 Perkussives Audiosignal 5
Abbildung 2 Einmassenschwinger 10
Abbildung 3 Komplexes E-Modul [29] 16
Abbildung 4 Rayleigh-Dämpfung [27] 17
Abbildung 5 Overlapping 21
Abbildung 6 Quadratische Peak Interpolation [37] 23
Abbildung 7 Absolute Hörschwelle 27
Abbildung 8 Maskierung benachbarter Frequenzgruppen [41] 28
Abbildung 9 Schalldruckfeld einer atmenden Kugel [43] 32
Abbildung 10 Signalvergleich einer Mode 33
Abbildung 11 Weg des Signals bis zur Analyse 37
Abbildung 12 Schematische Darstellung der Signalanalyse 38
Abbildung 13 Spektralanalyse 39
Abbildung 14 Mit Nullen aufgefülltes Signal 40
Abbildung 15 Modevorauswahl 41
Abbildung 16 Absolute Hörschwelle 43
Abbildung 17 Lokale und globale Mithörschwelle 44
Abbildung 18 Bestimmung der Abklingzeit 46
Abbildung 19 Regressionsanalyse: erster Fall 48
Abbildung 20 Regressionsanalyse: zweiter Fall 48
Abbildung 21 Bestimmung der Amplitude 50
Abbildung 22 Amplitude und Wendepunkt 51
Abbildung 23 Postprocessing 54
Abbildung 24 Signalverlauf und normierter Frequenzverlauf 56
Abbildung 25 Detektierung der amplitudenabhängigen Modenfrequenz 56
Abbildung 26 Frequenzabhängiger Dämpfungsfaktor 57
Abbildung 27 Frequenzverlauf eine Sinussignals mit zeitabhängiger Frequenz $f(t)$ 63
Abbildung 28 Modendetektierung: konstanter und nicht konstanter Frequenzverlauf 63
Abbildung 29 Korrekte Hüllkurvenannäherung 1 65
Abbildung 30 Korrekte Hüllkurvenannäherung 2 65
Abbildung 31 Ungenaue Hüllkurvenannäherung 1 66
Abbildung 32 Ungenaue Hüllkurvenannäherung 2 66
Abbildung 33 Ungenaue Hüllkurvenannäherung 3 67
Abbildung 34 Ungenaue Hüllkurvenannäherung 4 67
Abbildung 35 Korrekte Synthese einer Schwebung 68
Abbildung 36 Ungenaue Synthese einer Schwebung 68
Abbildung 37 Zusammenhang zwischen der Modenfrequenz und der Amplitude einer Tom 69
Abbildung 38 Zusammenhang zwischen der Modenfrequenz und der Amplitude einer Hightom 70
Abbildung 39 Zusammenhang zwischen der Modenfrequenz und der Amplitude eines Becken 70

Abbildung 40 Zusammenhang zwischen der Modenfrequenz und der Amplitude eines Kantilan . . . 71
Abbildung 41 EDR Sinus . . . 72
Abbildung 42 Hüllkurvenannäherung bei reinem Sinus . . . 73
Abbildung 43 Signalvergleich Sinus . . . 73
Abbildung 44 EDR weißes Rauschen . . . 74
Abbildung 45 Hüllkurvenannäherung bei Tom . . . 75
Abbildung 46 Signalvergleich weißes Rauschen . . . 75
Abbildung 47 Signalvergleich weißes Rauschen mit Phasenbezug . . . 76
Abbildung 48 EDR Tom . . . 77
Abbildung 49 Hüllkurvenannäherung bei Tom . . . 77
Abbildung 50 Signalvergleich Tom . . . 78
Abbildung 51 EDR Ride Becken . . . 79
Abbildung 52 EDR Hi-Hat . . . 79
Abbildung 53 Hüllkurvenannäherung bei Becken . . . 80
Abbildung 54 Signalvergleich Becken . . . 80
Abbildung 55 EDR Bassdrum . . . 81
Abbildung 56 Hüllkurvenannäherung bei Bassdrum . . . 82
Abbildung 57 Signalvergleich Bassdrum . . . 82
Abbildung 58 EDR Snaredrum . . . 83
Abbildung 59 Hüllkurvenannäherung einer Snaredrum . . . 84
Abbildung 60 Signalvergleich Snaredrum . . . 84
Abbildung 61 EDR Clave . . . 85
Abbildung 62 Hüllkurvenannäherung Clave . . . 86
Abbildung 63 Signalvergleich Clave . . . 86
Abbildung 64 EDR Glocke . . . 87
Abbildung 65 Hüllkurvenannäherung bei Glocke . . . 88
Abbildung 66 Signalvergleich Glocke . . . 88
Abbildung 67 EDR Klavier . . . 89
Abbildung 68 Hüllkurvenannäherung bei Glocke . . . 90
Abbildung 69 Signalvergleich Klavier . . . 90
Abbildung 70 EDR Marimba . . . 91
Abbildung 71 Hüllkurvenannäherung bei Marimba . . . 91
Abbildung 72 Signalvergleich Marimba . . . 92
Abbildung 73 EDR Akustikgitarre . . . 92
Abbildung 74 Hüllkurvenannäherung bei Akustik Gittare . . . 93
Abbildung 75 Signalvergleich Akustikgitarre . . . 93
Abbildung 76 EDR Kantilan . . . 94
Abbildung 77 Hüllkurvenannäherung Kantilan . . . 95
Abbildung 78 Signalvergleich Kantilan . . . 95
Abbildung 79 Tab Overview . . . 110
Abbildung 80 Tab Audiopreprocessing . . . 110
Abbildung 81 UI Controls Audiopreprocessing . . . 111
Abbildung 82 Slider Audiopreprocessing . . . 111
Abbildung 83 Tab Spectral Analysis . . . 112
Abbildung 84 UI Controls Spectral Analysis . . . 113
Abbildung 85 Tab Mode Detection . . . 113
Abbildung 86 Mode Detection UI Controls . . . 114
Abbildung 87 Tab Mode Properties Detection . . . 115

Abbildung 88 Mode Properties Detection UI Controls 116
Abbildung 89 Tab Postprocessing 116
Abbildung 90 Postprocessing Axes 117
Abbildung 91 Postprocessing UI Controls 118

TABELLENVERZEICHNIS

Tabelle 1 Nichtlineare Klangbeispiele 69
Tabelle 2 Kennwerte Sinussignal 74
Tabelle 3 Kennwerte weißes Rauschen 75
Tabelle 4 Kennwerte Tom 78
Tabelle 5 Kennwerte Becken 81
Tabelle 6 Kennwerte Bassdrum 82
Tabelle 7 Kennwerte Snaredrum 85
Tabelle 8 Kennwerte Clave 87
Tabelle 9 Kennwerte Glocke 89
Tabelle 10 Kennwerte Glocke 90
Tabelle 11 Kennwerte Marimba 92
Tabelle 12 Kennwerte Akustikgitarre 94
Tabelle 13 Kennwerte Kantilan 96
Tabelle 14 Zusammenfassung Resynthese 100
Tabelle 15 Einteilung in gehörrichtige Frequenzgruppen [41] 106
Tabelle 16 Resonanzfrequenzen eingespannte Rundplatte [29] . . . 108
Tabelle 17 Resonanzfrequenzen Rundplatte mit freien Enden[29] . . 108
Tabelle 18 Resonanzfrequenzen einfach gelagerte Rundplatte [29] . 108
Tabelle 19 Rundmembran [29] 109

LISTE DER ALGORITHMEN

1 Berechnung von H_{abs} 40
2 Auswahl der Moden ω_k 42
3 Entfernen von Moden unterhalb der absoluten Hörschwelle 44
4 Entfernen von Moden unterhalb der globalen Maskierungsschwelle 45
5 Berechnung der Abklingzeit τ 49
6 Berechnung der Amplitude A 52
7 Analyse von Schwebungen 53

SYMBOLVERZEICHNIS

a	m	Radius
A	1	Amplitude
b	m	Radius Nebenachse
$\mathbf{B}$	$\frac{kg}{s}$	Dämpfungsmatrix
c	$\frac{1}{s}$	Dämpfungskonstante Feder
c_L	$\frac{m}{s}$	Longitudinalwellengeschwindigkeit
c_y	$\frac{m}{s}$ Richtung	Longitudinalwellengeschwindigkeit y
c_x	$\frac{m}{s}$ Richtung	Longitudinalwellengeschwindigkeit x
c	$\frac{1}{s}$	Dämpfung
E	$\frac{kg}{s^2 m}$	Elastizitätsmodul
EDR	1	diskretes EDR
f	$\frac{1}{s}$	Frequenz
f_H	$\frac{1}{s}$	Frequenz der Einhüllenden
f_R	$\frac{1}{s}$	Frequenz der Überlagerungsschwingung
f_S	$\frac{1}{s}$	Abtastfrequenz
f_0	$\frac{1}{s}$	Resonanzfrequenz
F	$\frac{kg \cdot m}{s^2}$	Kraft
h	m	Höhe
H	1	diskrete Kurzzeitfouriertransformierte
$\mathbf{I}$	1	Einheitsmatrix
k	$\frac{kg}{s^2}$	Federsteife
K	m	Trägheitsradius
$\mathbf{K}$	$\frac{kg}{s^2}$	Steifigkeitsmatrix
L	m	Länge
L_x	m	Länge in x Richtung
L_y	m	Länge in y Richtung

m	kg	Masse
$\mathbf{M}$	kg	Massenmatrix
q	1	Normierte modale Koordinaten
r	m	Radius Hauptachse
r_0	$\frac{kgm}{s^2}$	Erregerfunktion
S	m^2	Fläche
t	s	Zeit
T	$\frac{kg}{s^2 m}$	Spannung (Tension)
T_f	s	Schwingungsdauer
$\mathbf{U}$	1	Eigenvektor
v	$\frac{m}{s}$	Geschwindigkeit
x	m	Auslenkung
z	m	Abstand neutrale Faser
α	$\frac{1}{s^2}$	Erste Rayleighdämpfungskonstante
β	1	Zweite *Rayleigh*dämpfungskonstante
δ	1	Dirac Impuls, Delta-Distribution
Δ	1	Differenz
Δ_s	1	Differenz der Signalenergien
ϵ	1	Dehnung
θ	1	Phase
$\boldsymbol{\lambda}$	1	Diagonalmatrix Eigenwerte
μ	$\frac{kg}{m^4}$	Flächen spezifische Dichte
ν	$\frac{kg}{m^5}$	*POISSON* Konstante isotropes Material
ν_{xy}	$\frac{kg}{m^3}$	*POISSON* Konstante xy Richtung
ν_{yx}	$\frac{kg}{m^3}$	*POISSON* Konstante yx Richtung
ξ	$\frac{1}{s}$	*Rayleigh*dämpfung
ρ	$\frac{kg}{m^3}$	Dichte
ρ_a	$\frac{kg}{m^3}$	Dichte Luft
$\rho_{O,A}$	1	Korrelation Original - Additives Signal
$\rho_{O,S}$	1	Korrelation Original - Synthetisiertes Signal

σ	$\frac{kg}{m^4}$	Längen spezifische Dichte
ϕ	1	Tangentiale Koordinatenrichtung
ϕ_n	1	Eigenschaften einer Mode n
ω_k	$\frac{1}{s}$	Samplenummer der Mode k
ω_d	$\frac{1}{s}$	Eigenkreisfrequenz gedämpftes System
ω_0	$\frac{1}{s}$	Eigenkreisfrequenz
ω	$\frac{1}{s}$	Kreisfrequenz
$\vec{\nabla}$	$\left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right)$	Nablaoperator

ABKÜRZUNGSVERZEICHNIS

AIFF Audio Interchange File Format

DFT Discrete Fourier Transformation

DSP Digital Signal Processing

DTFT Discrete Time Fourier Transformation

EDR Energy Decay Relief

FEM Finite Element Method

FFT Fast Fourier Transformation

FM Frequency Modulation

FS Fourier Series

MDOF Multi Degree Of Freedom

PM Physical Modelling

STFT Shorttime Fast Fourier Transformation

SAM Synth A Modeler

SDOF Single Degree Of Freedom

SFM Spectral Flatness Measure

Teil I

EINLEITUNG

1 EINLEITUNG

Die Beschreibung physikalischer Vorgänge von Klängen beziehungsweise von Musik und deren Entstehung ist seit je her ein wichtiges und weit verbreitetes Forschungsthema. Bereits die Griechen, insbesondere Euklid und Archytas, versuchten die Eigenschaften der Musik physikalisch-mathematisch zu beschreiben, hierbei wurde zum Beispiel der Frage nach dem Ursprung des Klanges und der mathematischen Beschreibung von Harmonie nachgegangen.
Eine umfassende Abhandlung der jüngeren Geschichte über die Physik der Schallentstehung und des Klanges stammt von Rayleigh [1].
So scheint es naheliegend, die Synthese von Klängen auf Grundlage der physikalischen Vorgänge in einem Musikinstrument zu realisieren. Die Beschreibung der physikalischen Gesetze dient außerdem dem Verständnis der untersuchten Instrumente und kann helfen Bauweisen auf Grundlage der physikalischen Eigenschaften zu verändern. Die Simulation von Parametermodifikationen, wie es zum Beispiel in der Automobilindustrie üblich ist, kann den Entwicklungsprozess beschleunigen und vergünstigen.

Gegenüber klassischen Syntheseformen wie zum Beispiel der FM- oder der additiven Synthese hat die Physikalische Modellierung den entscheidenden Vorteil, dass die Klangmodellierung direkt mit Hilfe physikalischer Parameter, wie zum Beispiel der Länge einer Saite oder dem Durchmesser einer Membran erfolgen kann. Einer der Hauptnachteile besteht in dem großen Rechenaufwand. Die Physikalische Modellierung wird meistens mittels einer der folgenden Synthesemethoden realisiert: Die hier verwendete Modalsynthese, die rein numerische Lösung der partiellen Differentialbewegungsgleichungen, die Beschreibung der physikalischen Bauteile als digitale Filter und die Beschreibung durch Masse-Feder-Netzwerke. Für eindimensionale Instrumente wie zum Beispiel die Saiten einer Gitarre, eignet sich auch die Wellenleitersynthese.
Die Kombination verschiedener Methoden ist oft von Vorteil. Die Modalsynthese bietet sich besonders bei kompakten Strukturen mit einer komplexen Geometrie an, da die modalen Eigenschaften nicht nur mit Hilfe analytischer Modelle berechnet sondern auch gut mit anderen Methoden wie zum Beispiel der finite Elemente Methode (FEM) angenähert werden können. In Verbindung mit der Beschreibung durch Masse-Feder-Netzwerke können auch zusammengesetzte Bauteile gut beschrieben werden.
Die Simulation eines linearen Musikinstrumentes umfasst im Wesentlichen drei physikalische Vorgänge: Die Anregung von Körperschall in einem Objekt, die Bewegungen des Objekts, und schließlich die Abstrahlung, also die Umwandlung von Körperschall in Luftschall, wobei letzteres in dieser Arbeit nicht behandelt wird. Um Klänge von hoher Güte zu erhalten, müssen modale Modelle ausgearbeitet werden, welche die relevanten Kernkomponenten der komplexen physikalischen Vorgänge möglichst genau abbilden. Grundlage der Modellierung ist die sinnvolle Parametrisierung des Modells. Um diese Simulationen durchführen zu können, ist es meist unzureichend, die in der Regel nicht li-

nearen physikalischen Vorgänge auf lineare Problemstellungen zu reduzieren. Andererseits soll dem Endnutzer (Musiker) sowohl eine Echtzeitsimulation, als auch eine intuitive Klanggestaltung ermöglicht werden. Dies führt zu zwei der Hauptproblematiken der PM-Synthese: Die Echtzeitsimulation komplexer physikalischer Vorgänge und die intuitive Klangmodellierung und Klangsteuerung.

Die Physikalische Modellierung im Bereich der Akustik entstand mit dem Aufkommen des digitalen Zeitalters [2]. Als einer der Pioniere in der Simulation in Echtzeit kann Julius O. Smith genannt werden, der insbesondere im Bereich der physikalischen Signalanalyse und Verarbeitung bedeutende Vorarbeiten geleistet hat [3] [4] [5] . Die Modalsynthese hat ihren Ursprung in der Untersuchung von komplexen linearen- [6] [7] und nicht linearen [8] schwingungsfähigen Mehrteilsystemen im Bereich des Maschinen- und Anlagenbaus. Später fand sie den Weg in die Akustik.
Für die in dieser Arbeit verwendete Modalsynthese zur Klangerzeugung, wird der Klang über die modalen Eigenschaften eines Objektes definiert. Diese setzen sich im Wesentlich aus der Frequenz f der Mode, der über die Dämpfung definierten Abklingzeit τ und der Amplitude A zusammen. Ein Bestandteil dieser Arbeit besteht in der Ermittlung dieser Parameter mit Hilfe von Klangaufnahmen impulsartig angeregter linearer Musikinstrumente.
Die analytische Modalsynthese wird typischerweise beschränkt auf lineare Musikinstrumente. Zum Beispiel die Saite einer Gittare [9] [10] [11], die Stäbe eines Kantilan [12], oder die Membran einer Trommel [13] [14]. Die Dämpfungscharakteristiken müssen dabei weiterhin aus Messungen realer Instrumente ermittelt oder per Hand eingegeben werden. Die Erstellung nicht lineare Modelle ist außerdem mit einem hohen Rechenaufwand verbunden. Für frei gestaltete Geometrien ist eine analystische Lösung nicht möglich und es müssen numerische Methoden zur Modellparametrisierung genutzt werden [15] [16] [17]. Dies gibt dem Nutzer die Möglichkeit der freien Geometriegestaltung. Doch eignen sich diese Methoden nicht für zusammengesetzte Bauteile. Ein weiteres Problem besteht darin, dass dem Anwender keine Möglichkeit der intuitiven Klanggestaltung gegeben wird. Außerdem führen manche Ansätze zu einer falschen Vereinfachung und Parametrisierung des Modells, insbesondere ohne vorherige Messungen wichtiger Materialeigenschaften, was sich negativ auf den erzeugten Klang auswirkt.
Eine dieser Masterarbeit vorangegangene, vergleichbare Forschungsarbeiten, versuchten einen aufgenommen Klang eines geometrischen Körpers auf Körper anderer Geometrie und gleichen Materials zu übertragen und diese in Echtzeit in Computerspielen zu synthetisieren [18]. Ren untersuchte daraufhin die Audiowahrnehmung Geometrie-Invarianter [19] Materialeigenschaften und entwickelte ein erweitertes Modell zur Übertragung von Materialeigenschaften gemessener Objektgeometrien auf neu gestaltete Geometrien [20]. Die Untersuchungen zeigten, dass es in einigen Fällen möglich ist, Materialeigenschaften wie die Dämpfung auf Grundlage von Messungen zu bestimmen und auf eine andere Geometrie zu übertragen. Das Hauptinteresse dieser Untersuchung lag jedoch nicht auf der Echtzeitsynthese.

Daher wird in dieser Masterarbeit die Realisierbarkeit der Echtzeitresynthe-

se mit Hilfe modaler Eigenschaften aufgenommener Instrumente untersucht. Grundlage der Untersuchung sind dabei Audiodateien von perkussiven Klängen. Ein perkussiver Klang zeichnet sich durch eine kurze Anstiegszeit (engl: attack time) aus, die im Bereich unter 100ms liegt und wird im mathematisch einfachsten Fall durch einen Impuls angeregt. In Abbildung 1 ist ein solches Audiosignal Abgebildet. Der entstehende Klang eines linearen Modells kann durch folgende Gleichung (1) mathematisch Beschrieben werden:

$$s(t) = \sum_{n=1}^{N} A_n e^{-t/\tau_n} \sin(2\pi f_n t) \tag{1}$$

Es handelt sich also um die lineare Überlagerung von N Moden, welche exponentiell Ausklingen.

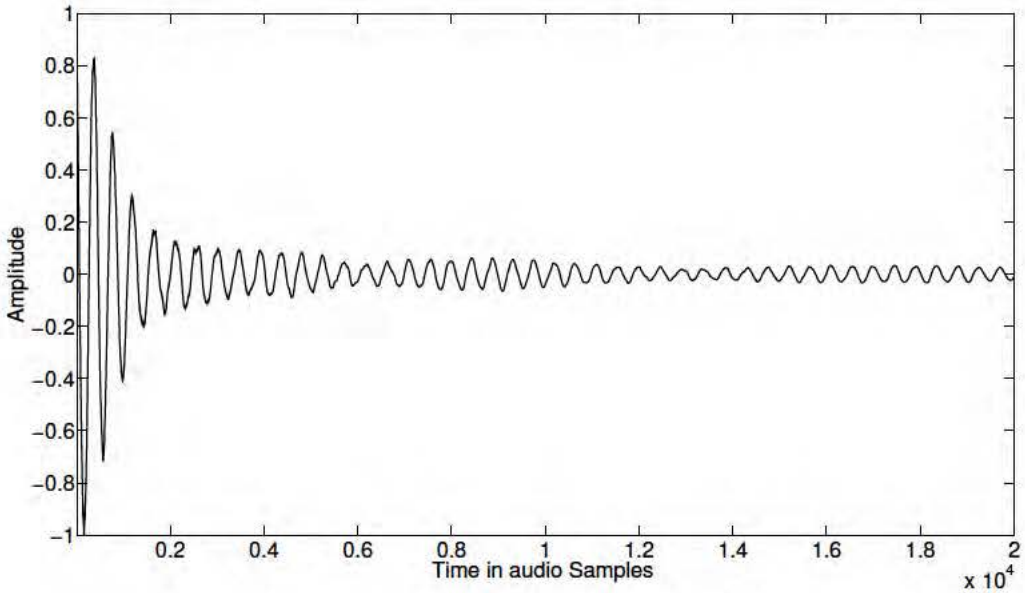


Abbildung 1: Perkussives Audiosignal

Die in dieser Arbeit verwendeten Audiosignale stammen zum großen Teil aus der Internetresource `www.freesounds.org` und sind lizenzfrei. Weiter wurden eigene Klänge aufgenommen. Ein Audiosignal einer nichtlineare Drum wurde von [1] heruntergeladen. Diese Audiodatei, zeichnet sich durch eine gute Klangqualität aus, weiter können die Ergebnisse der Synthese mit den Ergebnissen von Avanzini [21] verglichen werden.
Die Analyse der Klangbeispiele beinhaltet psychoakustische Methoden, um vom Menschen nicht wahrnehmbare Klangkomponenten aus den Modellsimulationen zu entfernen. Diese Bestandteile werden also in der Synthese nicht berücksichtigt.
Die ermittelten modalen Parameter, werden anschließend in ein *Synth-A-Modeler* Synthesemodell [22] integriert. Mit Hilfe der Digitalen Signal Verarbeitung (DSP) ermöglicht *Synth-A-Modeler* die physikalische Klangmodellierung in Echtzeit.
Zusätzlich zu den Parametrisierungen durch die Analyse von Klangdateien, werden lineare Modelle in *Synth-A-Modeler* mit Hilfe von analytischen Lösungen für einfache Geometrieren berechnet. Die Abklingzeiten und die Amplituden werden durch ein vereinfachtes Modell angenähert und können durch

1 `http://smc.dei.unipd.it/examples/membranes/`

den Anwender moduliert werden. Zusätzlich werden Materialparameter aus den Klangsignalen berechnet und in die analytischen Modellsimulationen integriert.
Die Ergebnisse der Modellierung werden anschließend ausgewertet und die Güte der Klänge sowohl mittels numerischer Kennwerte als auch informell bewertet.

Teil II

GRUNDLAGEN

2 GRUNDLAGEN

2.1 PHYSIKALISCHE GRUNDLAGEN

Im folgendem Abschnitt werden die Grundlagen der Schwingungen einfacher Körper erläutert.
Der Hauptbestandteil dieser Arbeit liegt in der Analyse von Klangbeispielen perkussiver Instrumente. Diese werden idealisiert als solide Körper angesehen und durch einen Impuls angeregt. Sie sind schwach gedämpft, so dass ihre modalen Amplituden exponentiell abnehmen. Da die Abstrahlung kein Bestandteil dieser Arbeit ist, wird angenommen, dass es sich bei den abstrahlenden Objekten um atmende Kugeln mit frequenzunabhängigem Abstrahlgrad handelt. Auf die angenommenen Vereinfachungen und Idealisierungen wird in Kapitel 3 näher eingegangen

2.1.1 *Impulsanregung*

Die impulsförmige Erregerfunktion ist allgemein definiert als [23]:

$$F(t) \quad r_0 \cdot \delta(t). \tag{2}$$

Die Anregung erfolgt im einfachen Fall über die impulsförmige Kraftanregung:

$$F(t) \quad \frac{F_0}{k} \cdot \delta(t). \tag{3}$$

Die *Dirac* Funktion δ ist dabei folgendermaßen definiert:

$$\delta(t) \begin{cases} 0 \text{ für } t < -\Delta t \\ / \; 0 \text{ für } t \leqslant -\Delta t \leqslant \Delta t \; ; \lim\limits_{\Delta t \to 0} \int\limits_{-\Delta t}^{\Delta t} \delta(t) dt \quad 1 \\ 0 \text{ für } \Delta t < t. \end{cases} \tag{4}$$

Es handelt sich also um eine infinitisimal kurze Anregung mit einer infinitismal großer Amplitude. Im Frequenzbereich zeichnet sie sich durch eine konstante Amplitude über das gesamte Frequenzspektrum aus:

$$F(\delta) \quad \int\limits_{-\infty}^{\infty} \delta(t) e^{j\omega t} dt \quad 1. \tag{5}$$

Es findet also eine freuquenzunabhängige Anregung stattfindet.

2.1.2 *Resonatoren*

Ein Resonator ist ein schwingfähiges System mit N Freiheitsgraden (MDOF), welches bei Anregung mit einer bestimmten Frequenz anfängt zu schwingen

und diese Schwingung ohne Dämpfung unendlich weiterführen würde. Diese Frequenzen f_n werden Eigenfrequenz oder Moden genannt. Bei dem im Folgenden beschriebenen Einmassenschwinger existiert nur eine solche Mode erster Ordnung. In realen Objekten mit N Freiheitsgraden kommen auch Moden höherer Ordnung vor: $f_{n,m,\ldots}; \quad n, m, \cdots \in [0, 1, 2, \ldots, N]$.
Die Klangerzeugung musikalischer Instrumente basiert grundsätzlich auf zwei Arten von Resonatoren. Zum einer dem mechanischen Resonator, als Klangerzeuger (z.B Gitarrensaite) und dem akustischen Resonator, als Resonanzkörper (z.B Gitarrenkörper). Für das Beispiel der Gitarre ist die Saite der Klangerreger, welcher definierte und modulierbare Eigenfrequenzen besitzt und der Gitarrenkörper der Klangverstärker, welcher die Resonanzfrequenzen der Saite verstärkt. Die Schwingungen der Saite werden dabei hauptsächlich über die Lagerungen der Saite auf den Gitarrenkörper übertragen.

2.1.3 *Lineare Schwinger mit einem Freiheitsgrad*

Der Einmassenschwinger ist eine spezielle Form des Resonators und dient hier der Anschaulichkeit des Prinzips eines Resonators, da die Auswirkungen einer Anregung relativ leicht berechnet und die Ergebnisse auf andere Schwingsysteme übertragen werden können.

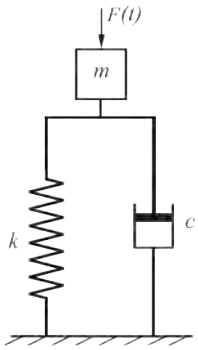


Abbildung 2: Einmassenschwinger

Der in Abbildung 2 gezeigte Einmassenschwinger besteht aus einer Feder mit Federsteife k, einem Dämpfer mit der Dämpfungskonstante c und einer Masse m. Er wird mit der Kraft $F(t)$ angeregt.
Die Resonanfrequenz f_0 des ungedämpften Schwingers ist proportional zur Wurzel des Verhältnisses von Federsteife zur Masse.

$$f_0 \quad \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{6}$$

In der Regel sind Schwinger gedämpft. Die allgemeine Differentialgleichung für den frei Schwingenden gedämpften linearen Einmassenschwinger lautet [24]:

$$m\frac{\partial^2 x}{\partial t^2} + 2d\frac{\partial x}{\partial t} + kx \quad 0 \quad ; d \quad \frac{c}{2m}. \tag{7}$$

Der Parameter d beschreibt hier die von der Masse m und der Dämpfungskonstante c abhängige Dämpfung. Die Dämpfung ist das Inverse der Abklingzeit

$\tau \quad 1/d$.
Nach Einsetzen von Gleichung (6) und Division durch die Masse m erhält man folgende homogene Differentialgleichung zweiter Ordnung:

$$\frac{\partial^2 x}{\partial t^2} + 2d\frac{\partial x}{\partial t} + \omega_0^2 x \quad 0. \tag{8}$$

Für den Fall einer freien schwach gedämpften Schwingung $d < \omega_0$ mit den Anfangsbedingungen $x(0) \quad x_0$ und $\dot{x}(0) \quad v_0$ erhält man folgende Lösung für die Auslenkung $x(t)$ [24]:

$$x(t) \quad Ae^{-dt}\sin(\omega_d t + \theta) \quad ; \omega_d \quad \sqrt{\omega_0^2 - d^2} \tag{9}$$

Also eine exponentiell abklingende Amplitude Ae^{dt}. mit der Amplitude A:

$$A \quad \frac{1}{\omega}\sqrt{x_0^2\omega_d^2 + (v_0 + dx_0)^2}, \tag{10}$$

der Phase

$$\theta \quad \arctan\left(\frac{\omega_d x_0}{v_0 + dx_0}\right), \tag{11}$$

und der Schwingperiode $T \quad 2\pi/\omega_d$. Wie in Gleichung (9) angemerkt, wurde reduziert sich die Resonanzfrequenz eines gedämpften Schwingers ω durch die Dämpfung d gegenüber der Resonanzfrequenz ω_0 des ungedämpften Schwingers.
Findet eine erzwungene Anregung statt, lautet die allgemeine Differentialgleichung für ein solches fremd erregtes System mit einem Freiheitsgrad [25] :

$$m\frac{\partial^2 x}{\partial t^2} + c\frac{\partial x}{\partial t} + kx \quad m\omega_0^2 r(t), \tag{12}$$

beziehungsweise für eine Krafterregung, normiert und mit vereinfachter Schreibweise:

$$\ddot{x} + 2d\dot{x} + \omega_0^2 x \quad \omega_0^2 r(t) \quad \frac{F(t)}{m} \quad \omega_0^2\frac{F(t)}{k} \tag{13}$$

Wobei $F(t)$ der Erregerfunktion entspricht. Im Grenzfall eines unendlichen kurzen Impulses, sind die Terme $2d\dot{x}$ und $\omega_0^2 x$ gegenüber $\ddot{x}$ vernachlässigbar klein [23], da sich die Masse zum Zeitpunkt der Anregung nur ein infinitesimal kleines Stück x aus der Ruhelage $x(0) \quad 0$ bewegt hat und somit auch eine, nur infinitesimal kleine Geschwindigkeit $\dot{x}$ besitzt. Die Beschleunigung $\ddot{x}$ ist jedoch sehr groß. Somit gilt für den Zeitpunkt der Anregung von $-\Delta t$ bis Δt :

$$\ddot{x} \quad \omega_0^2 r_0 \cdot \delta(t) \tag{14}$$

Durch Integration über die Impulsdauer erhält man die Geschwindigkeit nach dem Impuls:

$$\dot{x} \quad \lim_{\Delta t \to 0} \int_{-\Delta t}^{\Delta t} \ddot{x} \quad \omega_0^2 r_0 \quad \omega_0^2\frac{F_0}{k} \quad \frac{F_0}{m} \tag{15}$$

Nach dem Zeitpunkt der Anregung $t > \Delta t$, wirkt keine äußere Kraft mehr auf die Masse. Somit gilt für die Bewegung der Masse Gleichung 9. Diese Lösung

muss nun mit Hilfe der Anfangsbedingungen $x(0) = 0$ und $\dot{x} = F_0/m$ angepasst werden. Aus der ersten Anfangsbedingung $x(0) = 0$ folgt:

$$\begin{aligned} &x(0) = 0 = Ae^{-d\cdot 0}\sin(\omega_d \cdot 0 + \theta)\omega_0 \\ &\Rightarrow \theta = 0. \end{aligned} \tag{16}$$

Aus der zweiten Anfangsbedingung $\dot{x}(0) = \frac{F_0}{m}$ folgt:

$$\begin{aligned} &\dot{x}(0) = \frac{F_0}{m} = -A\beta e^{-d\cdot 0}\sin(\omega_d \cdot 0) + A\omega_d e^{-d\cdot 0}\cos(\omega_d \cdot 0) \\ &\Rightarrow A = \frac{F_0}{m\omega} \end{aligned} \tag{17}$$

Daraus folgt für die Bewegung für ein durch einen Kraftimpuls erregtes System mit einem Freiheitsgrad:

$$x(t) = \frac{F_0}{m\omega_d} e^{-dt}\sin(\omega_d t) \tag{18}$$

Insbesondere der Fakt, dass die Initialphase $\theta = 0$ ist, spielt für die Analyse und Synthese eine wichtige Rolle, da die Phasenformation bei der späteren Betrachtung vernachlässigt wird und angenommen wird, dass die untersuchten Klänge durch einen Impuls angeregt werden. Lineare Schwinger besitzen bestimmte Eigenschaften bezüglich ihres Übertragungsverhaltens. Die folgenden Eigenschaften charakterisieren einen linearen Schwinger, beziehungsweise ein lineares System, mit der Systemgleichung G, welche das Ausgangsverhalten charakterisiert:

- Homogenität:

$$G(cF(t)) = cG(F(t)) \quad ; \quad c \in \mathbb{R} \tag{19}$$

- Superposition:

$$G(F_1(t) + F_2(t)) = G(F_1(t)) + G(F_2(t)) \tag{20}$$

Die Impulsantwort dieses linearen Systems im Zeitbereich lautet auf Grund der Homogenität:

$$h(t) = \frac{1}{m\omega_d} e^{-dt}\sin(\omega_d t) \tag{21}$$

2.1.4 *Lineare Schwinger mit* N *Freiheitsgraden*

Für lineare gedämpfte Schwinger mit N Freiheitsgraden gilt folgende Bewegungsgleichung [26]:

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{B}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{F}(\mathbf{t}). \tag{22}$$

Mit der Massenmatrix $\mathbf{M}(N \times N)$, der Dämpfungsmatrix $\mathbf{B}(N \times N)$ und der Steifigkeitsmatrix $\mathbf{K}(N \times N)$. Mit Hilfe der Rayleigh-Dämpfung [27] (siehe Kapitel 2.1.6) kann die Dämpfungsmatrix $\mathbf{B}$ umgeschrieben werden:

$$\mathbf{B} = \alpha\mathbf{M} + \beta\mathbf{K} \tag{23}$$

Die Dämpfungsmatrix $\mathbf{B}$ ist nun ebenfalls diagonal. Die Rayleigh-Dämpfung kann also als mathematischer Trick angesehen werden. Es gilt dann folgendes Eigenwertproblem zu lösen:

$$\mathbf{K}\Phi \quad \lambda\mathbf{M}\Phi \Leftrightarrow (\mathbf{K} - \lambda\mathbf{M})\,\Phi \quad 0, \tag{24}$$

wobei Φ die Eigenvektoren und λ die Diagonalmatrix mit den Eigenwerten repräsentieren. Der Lösungsweg für das System wird hier nicht näher erläutert, er ist in [27] detailliert beschrieben. Das entkoppelte System, nach Lösung des Eigenwertproblems, lautet:

$$\ddot{\mathbf{q}} + (\alpha\mathbf{I} + \beta\lambda)\dot{\mathbf{q}} + \lambda\mathbf{q} \quad \Phi^T\mathbf{F}(\mathbf{t}). \tag{25}$$

Hierbei ist $\mathbf{I}$ die Einheitsmatrix. Der Wert $\mathbf{q}$ beschreibt die modalen Koordinaten oder Einheitskoordinaten. Es gilt:

$$\mathbf{x} \quad \sum_{n=1}^{N} u_n q_n \quad \mathbf{U}\mathbf{q}. \tag{26}$$

Die N Gleichungen (25) können nun wie die Gleichungen des Einmassenschwingers gelöst werden, da sie linear unabhängig sind. Mit den Lösungen für q_n für jede Mode n:

$$q_n \quad A_n e^{-d_n t} \sin(2\pi f_n t + \theta_n) \tag{27}$$

und der frequenzabhängigen Dämpfungskonstante $d_n \quad 1/\tau_n$:

$$d_n \quad \frac{\alpha}{2}\frac{1}{\omega_n} + \frac{\beta}{2}\omega_n. \tag{28}$$

Gleichung (27) entspricht also der Lösung des Einmassenschwingers (18). Die Gesamtlösung ergibt sich für ein System mit N Freiheitsgraden zu:

$$\mathbf{q} \quad \sum_{n=1}^{N} A_n e^{-d_n t} \sin(2\pi f_n t + \theta_n). \tag{29}$$

Ein gesamtes System kann also unter Vernachlässigung von θ_n durch drei Parameter pro Mode n beschrieben werden:

$$\varphi_n \quad (f_n, d_n, A_n). \tag{30}$$

Die Konstanten f_n und d_n sind von der Geometrie und dem Material abhängig. A_n ist zudem von der Art der Anregung abhängig. Im Gegensatz zum Schwinger mit nur einem Freiheitsgrad besteht die Lösung bei einem System mit N Freiheitsgraden also auch aus N unabhängigen Schwingungsformen, welche sich überlagern.
Vorherige Untersuchungen [19] [20] haben gezeigt, dass die materialabhängigen Parameter α und β für eine bekannte Struktur auch auf andere Strukturen gleichen Materials übertragen werden können.

2.1.5 *Nichtlineare Schwinger*

In der Regel sind reale Schwinger nicht linear, lassen sich teilweise aber aufgrund der kleinen Auslenkungen gut linear annähern. Werden die Verformungen des schwingenden Körpers größer, so gilt zum Beispiel beim Einmassenschwinger das Newtonsche $F \quad k \cdot x$ Gesetz für die Federkennlinie nicht mehr. Es treten bei Systemänderungen nichtlineare Faktoren auf. Am Beispiel der runden Membran soll dies verdeutlicht werden. Die lineare partielle Differentialgleichung für die vertikale Auslenkung der Membranoberfläche $z(r, \varphi, t)$ einer runden Membran lautet [13]:

$$\begin{aligned} D\nabla^4 z(r,\varphi,t) + \sigma\frac{\partial^2 z(r,\varphi,t)}{\partial t^2} - T_0\nabla^2 z(r,\varphi,t) \\ + d_1\frac{\partial z(r,\varphi,t)}{\partial t} + d_3\frac{\partial \nabla^2 z(r,\varphi,t)}{\partial t} \quad f^{(ext)}. \end{aligned} \tag{31}$$

Dabei ist σ die flächenspezifische Dichte, T_0 die Spannung der ruhenden Oberfläche, d_1 der frequenzunabhängige Verlustfaktor, d_3 der frequenzabhängige Verlustfaktor, E das Elastizitätsmodul, ν die *POISSON*-Zahl und R der Radius der Membran. Die linke Seite $f^{(ext)}$ der Gleichung 31 beschreibt die Anregung der Membran. Für eine nicht lineare runde Membran gilt folgende partielle Differentialgleichung [21]:

$$\begin{aligned} D\nabla^4 z(r,\varphi,t) + \sigma\frac{\partial^2 z(r,\varphi,t)}{\partial t^2} - [T_0 + T_{NL}(z(r,\varphi,t))]\nabla^2 z(r,\varphi,t) \\ + d_1\frac{\partial z(r,\varphi,t)}{\partial t} + d_3\frac{\partial \nabla^2 z(r,\varphi,t)}{\partial t} \quad f^{(ext)} \end{aligned} \tag{32}$$

Der Unterschied von Gleichung (32) zu (31) liegt im Term $T_{NL}(z)$, welcher die zusätzliche Spannung der Membranoberfläche in Abhängigkeit der Auslenkung z beschreibt. Dabei gilt für die, durch die Auslenkung hervorgerufene Spannung:

$$\begin{aligned} T_{NL}(z) \quad \frac{Eh}{2\pi R^2(1-\nu^2)} \\ \int_0^R \int_0^{2\pi} \left[\left(\frac{\partial z}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial z}{\partial \varphi}\right)^2 \right] r\,d\varphi\,dr \end{aligned} \tag{33}$$

Das Doppelintegral kann dabei als die Differenz der ruhenden Membranoberfläche A_0 und der, durch die Auslenkungen z verformte Membranoberfläche $A(z)$, angesehen werden. Die Lösungen der beiden Gleichungen werden hier nicht erläutert, da sie in dieser Arbeit keine Verwendung finden, sie sind in [13] zu finden. Es soll lediglich verdeutlicht werden, dass die Auslenkung z eines Objektes dazu führt, dass sich die Spannungen auf diesem Objekt erhöhen und es dadurch zu einer Veränderung der Eigenfrequenzen kommt. In der Regel führt eine höhere Spannung zu einer Vergrößerung der Eigenfrequenzen.
Weiter sind die Übertragungseigenschaften eines nichtlinearen Systems im Gegensatz zu einem linearen System auch von der Größe der Anregung abhängig. Eine Impulsantwort kann nicht angegeben werden. Das Superpositionsprinzip (20) und die Homogenität (19) sind für nichtlineare Systeme nicht gültig.

2.1.6 *Dämpfung schwingender Strukturen*

Für die Erstellung von modalen Modellen bedarf es der Bestimmung der Abklingkonstante τ, welche über die Dämpfung d der Schwingungen eines Objekts ermittelt werden kann $\tau = 1/d$. Die Dämpfung akustisch strahlender Objekte ist auf die Umwandlung der in der Schwingung enthaltenen Energie in eine andere Energieform zurückzuführen [28].
Die in der mechanischen Schwingung enthaltene Energie wandelt sich hauptsächlich in Wärme-, Strahlungs- und Schwingungsenergie anderer Systeme um. Bei der Umwandlung in Wärmeenergie wird unterschieden zwischen der materialabhängigen inneren Dämpfung und der Dämpfung durch die Luft.
Im folgenden werden die Dämpfungen für den einfachen Fall einer schwingende Saite erklärt [29]. Dieses Modell wird für die Berechnung der Abklingkonstante τ einer analytisch erstellten modalen Saite und vereinfacht auch für andere Modelle verwendet. Die Dämpfung durch die Luft ergibt sich aus dem Luftwiderstand der bewegten Saite. Für einen Zylinder der Länge L, des Radius r, welcher sich mit der Geschwindigkeit v und der Frequenz f bewegt ist die durch den Luftwiderstand hervorgerufene Kraft

$$F_r = 2\pi^2 \rho_a f v r^2 L \left(\frac{\sqrt{2}}{M} + \frac{1}{2M^2} \right), \tag{34}$$

mit $M = (r/2)\sqrt{2\pi f/\nu_a}$ und der Luftdichte $\rho_= 1,2 kg/m^3$. Daraus ergibt sich die Konstante für die Abklingzeit zu:

$$\tau_1 = \frac{\rho}{2\pi\rho_a f} \left(\frac{2M^2}{2\sqrt{2}M + 1} \right). \tag{35}$$

Die interne Dämpfung ist eine materialabhängige Konstante. Das komplexe E-Modul für isotrope Materialien, also für Materialien bei denen die physikalischen Eigenschaften unabhängig von der Richtung sind, lautet [29]:

$$E = E_1 + jE_2. \tag{36}$$

Für eine sinusartige Verfomung mit $\epsilon = \Delta l/l_0$ und $\sigma_n = E_n \epsilon$ die daraus resultierende Spannung eines Materials gilt:

$$\begin{aligned} \epsilon(t) &= \epsilon e^{j\omega t} = \epsilon(\cos(\omega t) + j\sin(\omega t)) \\ \sigma_1(t) &= E_1 \epsilon \cos(\omega t) \\ \sigma_2(t) &= -E_2 \epsilon \sin(\omega t) = E_2 \epsilon \cos(\omega t + \pi/2). \end{aligned} \tag{37}$$

Der Realteil des komplexen E-Moduls schwingt also in Phase mit der Auslenkung, der Imaginärteil schwingt in Phase mit der Geschwindigkeit und führt daher zur Abgabe von Energie aus dem System heraus.
Nach weiterer Rechnung [29] kann aus dieser Annahme das komplexe E-Modul in Abhängigkeit von der Kreisfrequenz ω und der Relaxationskonstante τ [30] berechnet werden:

$$E(\omega) = E_1(\omega) + jE_2(\omega) = \left[E_\infty - \frac{E_\infty - E_0}{1 + \omega^2\tau^2} \right] + j \left[\frac{(E_\infty - E_0)\omega\tau}{1 + \omega^2\tau^2} \right]. \tag{38}$$

Die Konstanten E_0 und E_∞ beschreiben hierbei das reale E-Modul bei sehr niedrigen und sehr hohen Frequenzen. In Abbildung 3 sind die Zusammenhänge zwischen Frequenz und E-Modul normiert dargestellt.

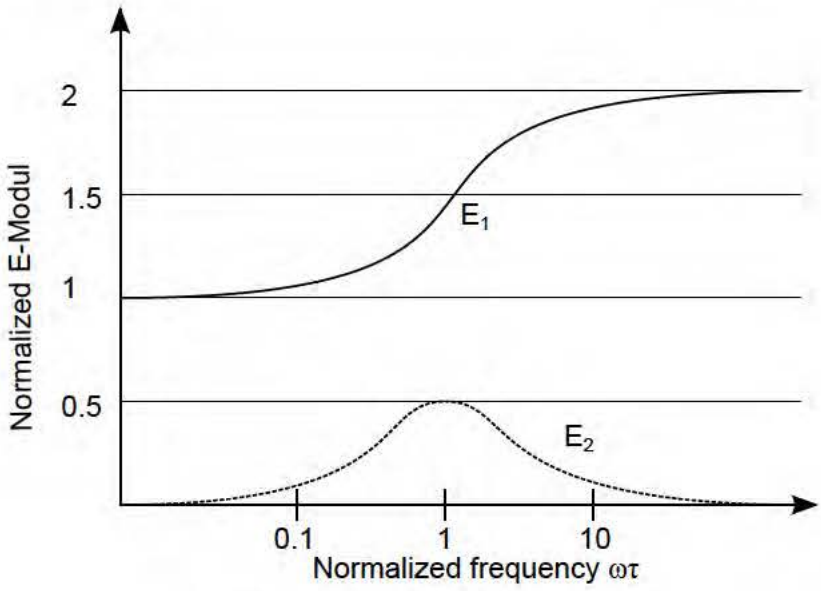


Abbildung 3: Komplexes E-Modul [29]

Bei bekanntem realem E-Modul E_1, und bekannter Realaxationskonstante [30] τ und den Annahmen $E_\infty = 2E_1$ und $E_0 = E_1$, kann das oben beschriebene frequenzabhängige Verhalten nützlich sein. Für die frequenzabhängige Abklingkonstante gilt also nach Gleichung (38):

$$\tau_2(\omega) = \frac{1}{\pi f}\frac{E_1(\omega)}{E_2(\omega)} = \frac{1}{\pi f}\left(-\frac{1}{\omega\tau} + \frac{E_\infty(1+\omega^2\tau^2)}{(E_\infty - E_0)\omega\tau}\right) = \frac{2}{\omega^2\tau} + 2\tau. \tag{39}$$

Eine dritte Dämpfung ergibt sich durch die Lagerung [31]. Über dieses wird Energie von der Saite an andere Systeme abgegeben. Der Realteil der Admittanz einer Saite ist definiert als:

$$G = \frac{v_n}{F_n \alpha_G}. \tag{40}$$

Mit der unbekannten Konstante α_G, der vertikal wirkenden Kraft F_n und der ebenfalls vertikalen Schnelle v_n. Die Dämpfung resultiert aus der in der Saite gespeicherten Energie. Die Abklingzeit mit der längenspezifischen Dichte σ ergibt sich zu:

$$\tau_3 = (8\sigma L f^2 G)^{-1}. \tag{41}$$

Aus der Kombination von τ_1, τ_2 und τ_3 kann nun die Abklingkonstante für eine schwingende Saite τ berechnet werden:

$$\tau = \frac{1}{\frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}}. \tag{42}$$

Um die Dämpfung einer unbekannten Struktur bei bekannten Materialparametern α und β grob zu berechnen eignet sich die Rayleigh-Dämpfung [27] [20]. In dieser Arbeit wird dieses einfache Modell zur Bestimmung der Dämpfung in Abhängigkeit der Frequenz angewendet. Die Rayleigh-Dämpfung verbindet zwei Dämpfungstheorien miteinander: Die massenproportionale Dämpfung $\xi = \alpha/(2\omega)$ und die steifigkeitsproportionale Dämpfung $\xi = \beta\omega/2$:

$$\xi = \frac{\alpha}{2}\frac{1}{\omega} + \frac{\beta}{2}\omega. \tag{43}$$

In Abbildung 4 ist die Rayleigh-Dämpfung gegen die Frequenz aufgetragen. Außerdem wurden die massen- und steifigkeitsproportionale Dämpfung eingezeichnet.

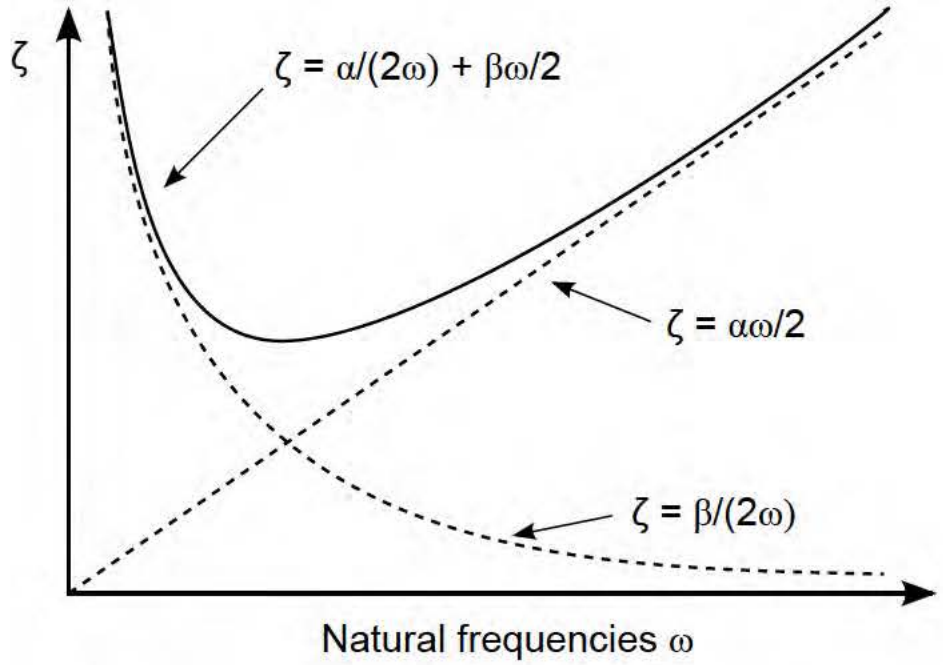


Abbildung 4: Rayleigh-Dämpfung [27]

Für anisotrope Materialien wie Holz, reicht die einfache Beschreibung der inneren Dämpfung nicht aus. Kann die Verformung aufgrund von Krafteinwirkung eines isotropen Materials durch drei unabhängige Parameter: Dem Schermodul G, dem E-Modul E und der *POISSON*-Zahl ν beschrieben werden, so müssen für ein orthotropes Material neun voneinander unabhängige Konstanten herangezogen werden. Jeweils drei für die drei Raumrichtungen. Mit diesem Wissen über die Struktur des verwendeten Holzes kann aber auch hier ein Dämpfungsmodell entwickelt werden [32]. Eine ausführliche Abhandlung und tabellarische Übersichten der wichtigsten Parameter sind in Woodhouse [33] zu finden.

Eine weitere Größe, welche in das oben genannte Dämpfungsmodell für eine Saite nicht einfließt ist die Dämpfung durch Abstrahlung. Der Strahlungsverlustfaktor [28] ist definiert als:

$$\eta = \frac{W_S}{2\pi W_k} \tag{44}$$

Der Strahlungsverlustfaktor beschreibt das Verhältnis von der abgestrahlten Energie pro Schwingungsperiode W_S, zu reversibler Körperschallenergie W_K. Da die Bestimmung dieser Größe nur über Messung erfolgen kann, ist die Anwendung hier problematisch. In exakteren Modellen sollte aber auch dieser Parameter berücksichtigt werden.

2.2 MATHEMATISCHE GRUNDLAGEN

2.2.1 *Fourier Transformation*

Die Mathematische Grundlage zur Spektralen Untersuchung von Signalen bildet die *FOURIER*-Transformation (FT). Mit Hilfe dieser Transformation können kontinuierliche Zeitsignale beliebiger Form in ein kontinuierliches Spektrum zerlegt werden.

$$F(\omega) \quad \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \quad ; \ \omega \in (-\infty, \infty). \tag{45}$$

Für diskrete Signale gilt folgende Gleichung für die Berechnung der diskreten Zeit *FOURIER*-Transformation (DTFT) [34]:

$$X(k) \quad \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega_k n} \quad ; \ \omega \in (-\pi, \pi). \tag{46}$$

Ein endliches periodisches Signal mit der Periode P kann mit Hilfe der *FOURIER*-Reihe (FS) in den Spektralbereich überführt werden:

$$F(\omega) \quad \frac{1}{2\pi}\int_{0}^{P} f(t)e^{-j\omega_k t}dt \quad ; \ k \quad -\infty, \ldots, +\infty. \tag{47}$$

Für periodische diskrete Signale gilt analog zur *FOURIER*-Reihe die diskrete *FOURIER*-Transformation (DFT):

$$X(k) \quad \sum_{n=0}^{N-1} x(n)e^{-j\omega_k n} \quad ; \ k \quad 0, 1 \ldots, N-1 \tag{48}$$

2.2.2 *Schnelle Fourier Transformation*

Die Schnelle Fourier Transformation (FFT) dient zur effizienten Transformation eines diskreten Zeitsignals $x(n)$ in den diskreten Spektralbereich $X(k)$ [34]. Die Grundidee ist die Aufteilung des Eingangsvektors $x(n)$ in einen Vektor mit geraden Indizes und einen Vektor mit ungeraden Indizes, welche unabhängig voneinander transformiert werden. Somit kann die diskrete Fourier Transformation in zwei Schritten aufgeteilt werden. Die FFT wird mit Hilfe der DFT berechnet, daher ist es notwendig, dass die Länge des Eingangsvektors einer Zweierpotenz entspricht.

2.2.3 *Kurzzeit Fourier Transformation*

Der Unterschied der diskreten Kurzzeit Fourier Transformation zur diskreten Fourier Transformation besteht in der Unterteilung des Signals s in kleine Segmente $s_W \subset s$, die unabhängig voneinander transformiert werden. Dazu wird das Fenster w mit der Fensterlänge W eingeführt. Ein Analysesegment hat al-

so die Länge W, die Gesamtlänge des Signals beträgt N. Die diskrete Kurzzeit Fourier Transformation (STFT)ist definiert als:

$$H(m,k) \quad \sum_{n=infty}^{\infty} s(n) \cdot w(n - m \cdot O_{step}) \cdot e^{-2\pi kn} \tag{49}$$

Hierbei entspricht f_S der Abtastfrequenz, O_{step} der Sprungweite und m dem zugeordneten Zeitschritt des Signals. Die Fensterung ist notwendig, da die DFT ein sich periodisch fortsetzendes Signal zur Transformation benötigt. Mit Hilfe der Kurzzeit Fourier Transformation ist es möglich ein Signal in Abhängigkeit von der Zeit in seine spektralen Bestandteile zu zerlegen. Die Fensterlänge W für die FFT muss $W \quad 2^k \; ; k \in \mathbb{Z}$ betragen. Ist dies nicht der Fall so muss der gefensterte Zeitraum mit Nullen aufgefüllt werden, bis $W \quad 2^k \; ; k \in \mathbb{Z}$ erfüllt ist. Das Auffüllen mit Nullen (Zero Padding) kann außerdem dazu genutzt werden, die Auflösung im Spektralbereich zu erhöhen, wenn hinter das gefensterte Signal der Länge W, $W_{zeroPad} \quad W \cdot k; \quad k/2 \in \mathbb{Z}$ Nullen hinzugefügt werden. Dies dient nur dazu, die Auflösung im Spektralbereich zu erhöhen, es erhöht nicht der Informationsgehalt, da für diesen Bereich keine Frequenzinformationen vorliegen. Die FFT-Länge des Fensters beträgt nach dem Auffüllen mit Nullen $W_{FFT} \quad W + W_{zeroPad}$.

2.2.4 *Spektrale Analyse von Signalen*

Die Spektralanalyse bildet das Kernstück der Signalverarbeitung für die spätere Parametrisierung des Modells. Da ein endliches und diskretes Signal nicht analytisch durch eine Fouriertransformation in den Spektralbereich transformiert werden kann, ist eine beliebig genaue zeitliche Auflösung bei einer beliebig genauen spektralen Auflösung nicht möglich. Es muss also ein Konsens gefunden werden. Die spektrale Auflösung ist begrenzt durch die Länge des Analysesegments beziehungsweise der Fensterlänge W und der Abtastrate f_S des Signals. Die zeitliche Auflösung ist durch die Anzahl der analysierten Segmente definiert.
Nach dem *NYQUIST-SHANNON*-Abtasthheorem ist die höchste Spektralkomponente durch die Abtastfrequenz f_S begrenzt:

$$f_S > 2 \cdot f_{max} \Rightarrow f_{max} \quad \frac{f_S}{2} \tag{50}$$

Somit können mit einer bei Audiosignalen üblichen Abtastfrequenz von f_S 44.1kHz alle Frequenzen unterhalb der psychoakustischen Hörschwelle von 20.05kHz erfasst werden.
Die unterste Spektralkomponente ist limitiert durch die die Länge des Analysesegments der Länge W und der Abtastfrequenz f_S. Die Periodendauer T einer Frequenz muss kleiner sein als der Analysezeitraum. Daraus folgt für die untere Spektralkomponente:

$$W > \frac{f_S}{f_{min}} \Leftrightarrow f_{min} > \frac{f_S}{W} \tag{51}$$

Somit kann bei einer Fensterlänge von $W \quad 4096$ Samples eine minimale Spektralkomponente von 12Hz erfasst werden. Dies liegt unterhalb der psychoakustischen unteren Hörschwelle von 20Hz. Somit kann signaltechnisch mit einer

Abtastrate von $f_S = 44,1\text{kHz}$ und einer Fensterlänge von $W = 4096$ der Bereich des menschlichen Hörens abgedeckt werden.
Die spektrale Auflösung eines Signals ist bestimmt durch die Abtastfrequenz und die Länge des gewählten Fensters. Wie oben erwähnt, kann mit einer Abtastfrequenz von $f_S = 44,1\text{kHz}$ eine Spektralkomponente von maximal $f_{max} = 22,05\text{kHz}$ bestimmt werden. Die spektrale Auflösung ist abhängig von der Anzahl der Samples W pro Fenster und beträgt effektiv $W/2$ aufgrund der Limitierung der oberen Grenzfrequenz f_{max}. So ergibt sich der Frequenzvektor f_k zu:

$$f_k = k\frac{f_S}{2(W/2-1)}; \quad k = 0,1,\ldots,W/2-1. \tag{52}$$

Wird eine Kurzzeit Fourier Transformation durchgeführt, erhält man ebenfalls einen Zeitvektor t_n, der folgende Eigenschaften hat: Die zeitliche Auflösung also die Länge L_t des Vektors t_n ist abhängig von dem Verhältnis der gewählten Segmentlänge W und der Gesamtlänge des Signals N, also $L_t = N/W$.
Die Zuordnung zu Zeitwerten ergibt sich aus der Länge des Signals M und der Abtastfrequenz f_s mit Hilfe der zeitlichen Länge des Signals $t_{max} = M/f_S$ und der Segmentgröße W.
Sinnvoller Weise liegt der zugeordnete Zeitwert t_n, bei der Anwendung symmetrischer Fenster in der Mitte eines analysierten Segments. Daher gilt für den minimalen Zeitwert $t_{min} = W/(2f_S)$ und für den maximalen Zeitwert $t_{max} = (2N-W)/(2f_s)$. Der Zeitvektor t_n ergibt sich folgendermaßen:

$$t_n = \frac{W}{2f_s} + \frac{N-W}{f_s(N/W-1)}n; \quad n = 0,1,\ldots,N/W-1. \tag{53}$$

Für ein Signal $s(n)$ mit einer Länge $N = 32768$ und einer Fensterlänge von $W = 8192$ sind also nur vier Zeitwerte vorhanden. Die zeitliche Auflösung kann erhöht werden indem man die Fenster überlappt (engl. „Overlapping“). Die Analysegebiete werden also nicht unabhängig von einander in den Spektralbereich transformiert, sondern enthalten Informationen von benachbarten Analysegebieten.
Abbildung 5 zeigt ein Beispiel für die Überlappung von Fenstern. In $a)$ wurden die Fenster nicht überlappt, die Analysegebiete sind unabhängig voneinander. In Beispiel $b)$ dagegen, wird 50% des Fensters m von dem Fenster $m+1$ überlappt. In Beispiel $c)$ dagegen, wird 75% des Fensters m von dem Fenster $m+1$ überlappt. Die Wahl der Überlappungsgröße $O \in [0,1[$ ist frei wählbar und unterliegt bei einer beliebig hohen Abtastrate f_S keiner mathematischen Beschränkung.

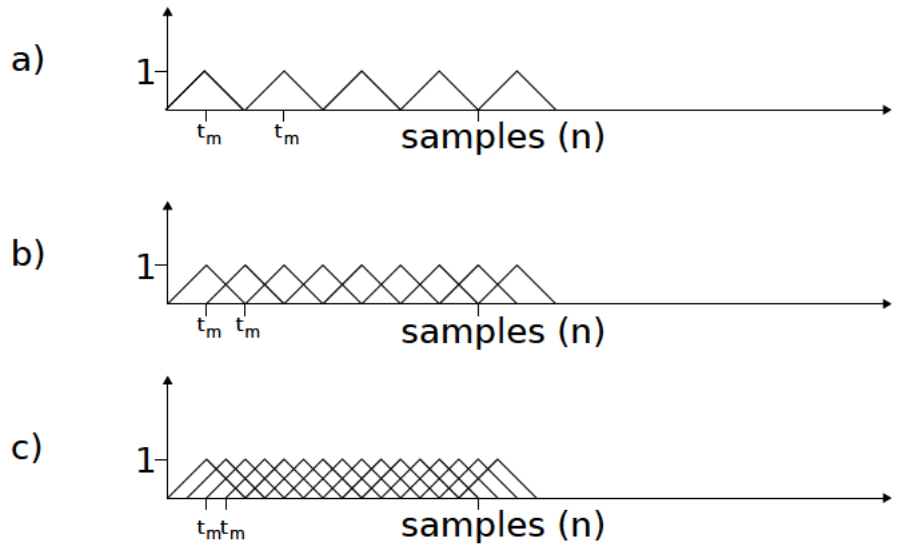


Abbildung 5: Overlapping
Verschiedene Overlapping Größen: a) kein overlapping b) overlapping 50% c) overlapping 75%

Mit Hilfe der Überlappung kann die zeitliche Auflösung also erhöht werden. Der Zeitvektor t_n ist nun ebenfalls von der Überlappungsgröße O abhängig. Da für die minimale und maximale Zeitwerte weiterhin $t_{min} \quad W/(2f_S)$ und $t_{max} \quad (2N - W)/(2f_S)$ gilt, ist es sinnvoll die Überlappung so zu wählen, dass gilt:

$$O \quad \frac{2^p - 1}{2^p}; \quad p \quad 0, 1, 2, \ldots, \log_2(W) - 1. \tag{54}$$

Für die Schrittweite O_{step}, die festlegt wie viele Samples das Segments $m + 1$ hinter Segment m liegt gilt also:

$$O_{step} \quad \frac{W}{2^p}; \quad p \quad 0, 1, 2, \ldots, \log_2(W) - 1. \tag{55}$$

Diese Größe sollte so gewählt sein, dass das letzte Segment, das selbe ist wie in der Analyse ohne Overlapping. Der Zeitvektor t_n ergibt sich nun zu:

$$t_n \quad \frac{W}{2f_s} + \frac{N - W}{f_s(O_{step}(N/W) - 1)} n; \quad n \quad 0, 1, \ldots, N \cdot O_{step}/W - 1. \tag{56}$$

2.2.5 *Energy Decay Relief*

Das Energy Decay Relief für eine bestimmte Frequenz f zum Zeitpunt t ist definiert als die in der Frequenz zum Zeitpunkt t vorhandene Energie. Damit die Energie für die vorhandenen Frequenzen einen endlichen Wert ergeben, wird vorraus gesetzt, dass es sich um Signale gedämpfter System handelt, deren Energie monoton fällt.
Das Energy Decay Relief bietet sich an um Signale zu untersuchen, welche Rauschbelastet sind. Insbesondere, wenn das Rauschen R eine Funktion der Frequenz f aber nicht der Zeit t ist. Grundlage des Energy Decay Reliefs bildet die Umwandlung eines Signals in ihre Frequenzbestandteile zu jedem Zeit-

punkt t mit Hilfe der Kurzzeitfouriertransformation. Das Energy Decay Relief ist definiert als [35]:

$$\hat{z}(t,\omega) \quad \int_t^\infty |F(t,\omega)|^2 \, dt. \tag{57}$$

Grundlage des diskreten Energie Decay Reliefs bildet die diskrete Kurzzeitfouriertransformierte $H(m,k)$ mit dem diskreten Frequenzvektor k und dem diskreten Zeitvektor m. Das Energy Decay Relief für das Zeitfenster n mit M Gesamtfenstern entspricht dann:

$$EDR(t_n, f_k)\hat{} \sum_{m=n}^{M} |H(m,k)|^2 \, . \tag{58}$$

Wie in Kapitel 2.1 gezeigt, hat die Amplitude eines harmonischen schwach gedämpften Oszillator für eine Frequenz f_0 im spektralbereich einen ideal exponentiell abfallenden Verlauf:

$$A(t) \quad \hat{A} e^{-t/\hat{\tau}}. \tag{59}$$

Für die spätere Modellbildung wird das EDR für die Bestimmung der Abklingzeiten τ herangezogen. Daher ist es notwendig die vom EDR erhaltenen Daten zurück zu transformieren. Folgende Rechnung zeigt, dass die Amplitude A und die Abklingkonstante τ aus dem Energy Decay Relief berechnet werden kann, wenn angenommen wird das im Signal kein Rauschen vorhanden ist:

$$\begin{aligned}
\hat{z}_H(t) \quad & \int_t^\infty \left(\hat{A} e^{-t/\hat{\tau}}\right)^2 dt \\
& \hat{A}^2 \int_t^\infty e^{-2t/\hat{\tau}} dt \\
& \hat{A}^2 \left[-\frac{\hat{\tau}}{2} e^{-2t/\hat{\tau}}\right]_t^\infty \\
& \hat{A}^2 \left[0 - \frac{\hat{\tau}}{2} e^{-2t/\hat{\tau}}\right] \\
& \hat{A}^2 \frac{\hat{\tau}}{2} e^{-2t/\hat{\tau}}.
\end{aligned} \tag{60}$$

Daher gilt für Rücktransformation der Amplitude:

$$A \quad \frac{2}{\tau}\sqrt{\hat{A}} \tag{61}$$

und die Abklingkonstante, die man vom EDR schätzen könnte, um die tatsächlichen Werte im Zeitbereich zu gewinnen:

$$\tau \quad \frac{\hat{\tau}}{2}. \tag{62}$$

Ist ein Signal mit einem zeitlich konstanten Rauschen R behaftet so ergibt sich folgender Rechenweg:

$$\begin{aligned}\hat{z}_H(t) &= \int_t^\infty \left(\hat{A}e^{-t/\hat{\tau}} + R\right)^2 dt \\ &= \int_t^\infty \hat{A}^2 e^{-2t/\hat{\tau}} + 2\hat{A}e^{-t/\hat{\tau}}R + R^2 dt \\ &= \left[-\hat{A}^2\frac{\hat{\tau}}{2}e^{-2t/\hat{\tau}} - 2\hat{A}\hat{\tau}e^{-t/\hat{\tau}} + R^2 t\right]_t^\infty \\ &= -\hat{A}^2\frac{\hat{\tau}}{2}e^{-2t/\hat{\tau}} - 2\hat{A}\hat{\tau}e^{-t/\hat{\tau}} + \left[R^2 t\right]_t^\infty .\end{aligned} \tag{63}$$

Es ist somit nicht möglich die Amplitude und die Abklingkonstante aus einem rauschbehaftetetn Signal analytisch zu berechnen. Daher wird in der Analyse der Abklingzeiten τ in dieser Arbeit vereinfacht davon ausgegangen, dass das Signal quasi rauschfrei ist.

2.2.6 *Quadratische Peak Interpolation*

Mit Hilfe der quadratischen Peak Interpolation kann die Genauigkeit, der in einer diskreten Frequenzanalyse ermittelten Amplituden und Frequenzen erhöht werden. Dabei wird ausgenutzt, dass die abfallenden Flanken rechts und links eines Peaks im interpolierten Amplitudengang theoretisch quasi-parabolisch abfallen. Abbildung 6 zeigt den Zusammenhang zwischen ermittelter Amplitude, den benachbarten Punkten und der neuen Amplitude [36].

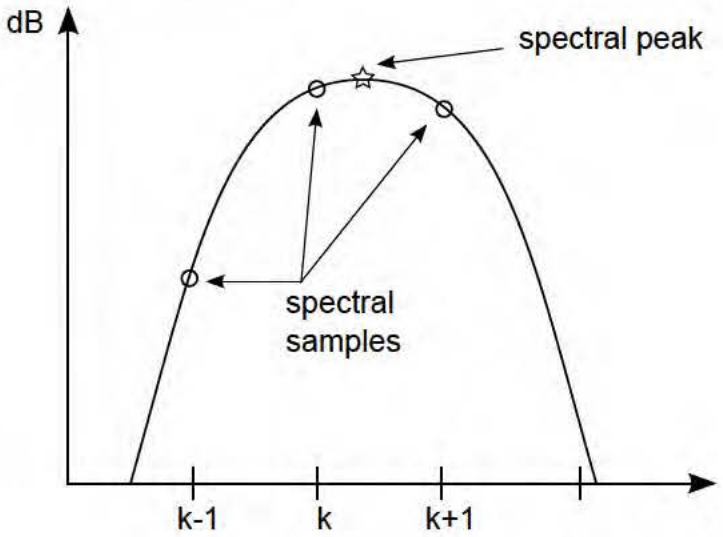


Abbildung 6: Quadratische Peak Interpolation [37]

Die generelle Form einer quadratischen Kurve lautet:

$$y(x) \hat{=} a(x-p)^2 + b. \tag{64}$$

Wobei, p die Position des globalen Extremum angibt und b der Amplitude in dB entspricht. Die Krümmung entspricht $2a$ und ist abhängig von der ge-

wählten Fensterfunktion. Nun kann mit Hilfe der benachbarten Samples des detektierten Peaks $y(k)$

$$\begin{aligned} y(k-1) &\quad \alpha \\ y(k) &\quad \beta \\ y(k+1) &\quad \gamma \end{aligned} \tag{65}$$

und der Annahme, dass es sich um eine quadratische Kurve handelt, folgendes Gleichungssystem aufgestellt werden:

$$\begin{aligned} \alpha &\quad ap^2 + 2ap + a + b \\ \beta &\quad ap^2 + b \\ \gamma &\quad ap^2 - 2ap + a + b. \end{aligned} \tag{66}$$

Welches aufgelöst, nach den bekannten Werten α, β und γ folgende Gleichungen für p

$$p \quad \frac{1}{2}\frac{\alpha-\beta}{\alpha-2\beta+\gamma} \tag{67}$$

und $y(p)$ ergibt:

$$y(p) \quad \beta - \frac{1}{4}(\alpha-\gamma)p. \tag{68}$$

Die interpolierte Frequenz f_{int} kann nun folgendermaßen mit Hilfe der Abtastfrequenz F_S und der Segmentlänge N berechnet werden:

$$f_{int} \quad \frac{(k+p)f_S}{W_{FFT}}\text{Hz}. \tag{69}$$

2.2.7 *Generieren eines minimalphasigen Signals*

Da in dem hier verwendeten physikalischen Modell keine Phaseninformationen verwendet werden, ist es notwendig, dass die zu analysierenden Signale möglichst minimalphasig sind also die darin enthaltenden spektralen Bestandteile, jeweils keine großen Phasenverschiebung zueinander enthalten.
Mit Hilfe der folgenden Schritte kann ein diskretes Signal minimalphasig gemacht werden. Der in dieser Arbeit verwendete Code wurde bereitgestellt von Julius O. Smith [38]. Im ersten Schritt wir der Logarithmus jedes Zeitschrittes t berechnet:

$$s_{log}(n) \quad \log(s(n)). \tag{70}$$

Anschließend wird die inverse Fourier Transformation des Signals berechnet:

$$s_{log}(m) \quad \frac{1}{N}\sum_{j=0}^{N-1} s_{log}(n)e^{2\pi i\frac{jm}{N}}dt. \tag{71}$$

Anschließend werden Werte, welche eine Phase θ / 0 besitzen minimalphasig gemacht. In dem die linke Seite von $S_{log}(m)$ mit dem komplex konjugierten

der Rechten Saite von $S_{log}(m)$ addiert wird, und anschließend die rechte Seite mit Nullen aufgefüllt wird. Dabei ist N die Gesamtlänge des Signals.

$$\begin{aligned} k_{min}(m) &\quad \left(S_{log}(m) + (S_{log}(N-m+1))^*\right) \cdot \alpha; \\ \alpha &\quad \begin{cases} 1; & \text{wenn} \quad m \quad 1, \ldots N/2 \\ 0; & \text{wenn} \quad m \quad N/2, \ldots N. \end{cases} \end{aligned} \tag{72}$$

Anschließend wird die e Funktion auf jeden Wert angewandt:

$$s_{min}(m) \quad e^{k_{min}(m)}. \tag{73}$$

Um dann das Signal durch eine Fourier Transformation wieder in den Zeitbereich zu überführen

$$s_{min}(n) \quad \sum_{j=0}^{N-1} s_{min}(m) e^{-2\pi i \frac{jn}{N}} dt. \tag{74}$$

Mit diesem Verfahren ist es möglich die spektralen Bestandteile eines Signals s minimalphasig zu machen.

2.3 PSYCHOAKUSTISCHE GRUNDLAGEN

Die hier vorgestellten psychoakustischen Modelle dienen zur späteren Daten- und der damit verbundenen Rechenreduktion der späteren Soundsynthese. Von Interesse ist hier die absolute Hörschwelle [39], also die Schwelle bei der ein Mensch einen Ton gerade noch wahrnimmt und sogenannte Maskierungsphänomene [40]. Dies sind akustische Phänomene welche Auftreten, wenn ein Lauter spektraler Bestandteil eines Tons einer bestimmten Frequenz f_1 einen zweiten leiseren spektralen Bestandteil der Frequenz f_2 maskiert, also dazu führt das der spektrale Bestandteil der Frequenz f_2 vom menschlichen Gehör nicht mehr wahrgenommen werden kann.

2.3.1 *Gehörrichtige Frequenzskala*

Für die Zusammenfassung einzelner Frequenzen zu Frequenzgruppen ist es notwendig diese Gruppen gehörrichtig anzuordnen. Hierzu muss die lineare Frequenzskala in eine gehörrichtige Frequenzskala transformiert werden, die sogennante *BARK*-Skala [41], welche die Tonheit z angibt:

$$\frac{z}{\text{Bark}} \quad 13 \arctan\left(0.76 \frac{f}{\text{kHz}}\right) + 3.5 \arctan\left(\frac{f}{7.5\text{kHz}}\right)^2. \tag{75}$$

Mit Hilfe der *BARK*-Skala, kann die Frequenzskala in Gruppen folgender Breite eingeteilt werden:

$$\frac{\Delta f_G}{\text{Hz}} \quad 25 + 75\left(1 + 1,4\left(\frac{f}{\text{kHz}}\right)^2\right)^{0,69}. \tag{76}$$

Die tabellarische Einteilung der Frequenzgruppen ist im Anhang A.2 zu finden.

2.3.2 *Signalleistung und Schalldruck im Band*

Auf Grundlage des Kurzzeitspektrums $X(k)$ kann das Leistungsdichtespektrum für jeden Wert k geschätzt werden:

$$\begin{aligned} S_P(e^{j\Omega}) \quad & S_P(e^{j\frac{2\pi k}{N}}) \quad X_R^2(e^{j\frac{2\pi k}{N}}) + X_I^2(e^{j\frac{2\pi k}{N}}) \\ & S_P(k) \quad X(k)_R^2 + X(k)_I^2 \quad ; 0 \leqslant k \leqslant N-1. \end{aligned} \tag{77}$$

Die Signalleistung in der Frequenzgruppe i kann dann folgendermaßen berechnet werden:

$$S_P(i) \quad \sum_{\Omega=\Omega_{ui}}^{\Omega_{oi}} S_P(e^{j\Omega}). \tag{78}$$

Für den Schalldruck in der Frequenzgruppe i gilt dann:

$$L_S(i) \quad 10 \cdot \log_{10} S_P(i). \tag{79}$$

2.3.3 *Absolute Hörschwelle*

Die absolute Hörschwelle (Ruhehörschwelle) gibt für jede Frequenz einen Schalldruckpegel an, für den das Gehör einen Sinuston gerade noch wahrnimmt [41].

$$\frac{L_{T_q}}{dB} \quad 3.64\left(\frac{f}{kHz}\right)^{-0.8} - 6.5\exp\left(-0.6\left(\frac{f}{kHz}-3.3\right)^2\right) + 10^{-3}\left(\frac{f}{kHz}\right)^4. \tag{80}$$

Unterhalb dieser Schwelle werden keine Signale mehr wahrgenommen. Bei einer Bewertung der Lautstärke eines Signals in Phon, also Kurven gleich empfundener Lautstärke, bildet die absolute Hörschwelle die Isophone mit dem Wert 0. Die Einteilung der Lautstärken einzelner Frequenzen in die Einheit Phon kann dabei helfen Lautstärken einzelner Frequenzen gehörrichtig zu bewerten.

2.3.4 *Mithörschwelle*

Die Mithörschwelle gibt an inwieweit ein Maskierungssignal benachbarte Töne maskiert, also überdeckt, so dass sie vom menschlichen Gehör nicht mehr wahrgenommen werden können. Bevor die Mithörschwelle der einzelnen Bänder ermittelt werden kann, muss der Abstand zwischen dem Pegel des Maskierungssignals und der Mithörschwelle bestimmt werden. Hierfür kann nach [41] als Näherungswert folgendes angenommen werden:

$$O(i) \quad \alpha(14.5+i) + (1-\alpha)a_v. \tag{81}$$

Der Koeffizient α bezeichnet den Tonalitätsindex und a_v den Maskierungsindex. Der Maskierungsindex wird vereinfachend zu $a_v \quad 5.5$ gesetzt. Der Abstand zwischen dem Pegel des Maskierungssignals und der Mithörschwelle ist abhängig vom Charakter des Signalabschnitts. Handelt es sich um einen tonalen oder einen rauschartigen Charakter so ergibt sich der Abstand zu $O(i) \quad 5.5$

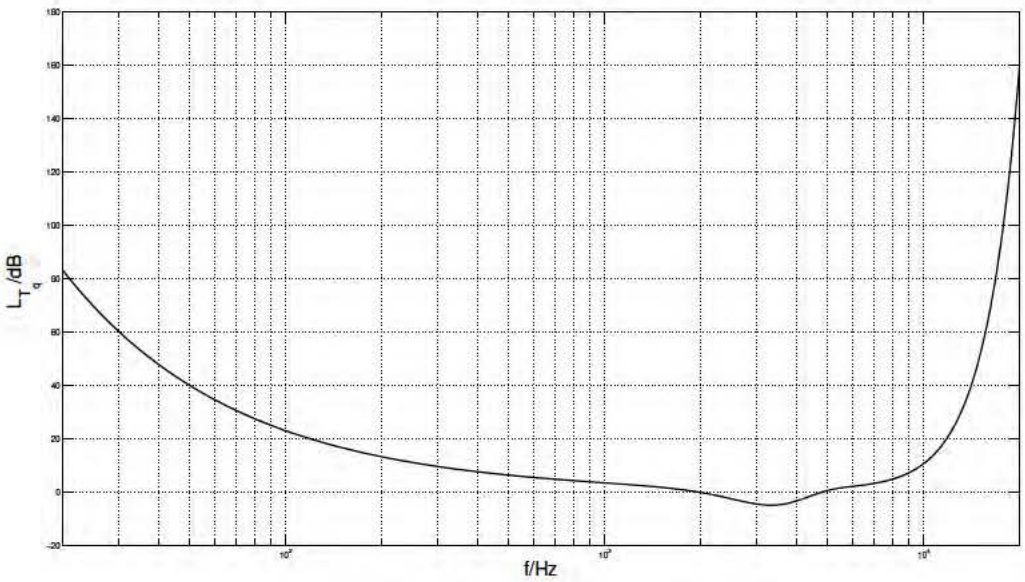


Abbildung 7: Absolute Hörschwelle

beziehungsweise $O(i) = 14.5 + i$ für die Frequenzgruppe i. Den Charakter kann man mithilfe der Spektralen Verteilung (SFM Spectral Flatness Measure) für eine Frequenzgruppe i bestimmen:

$$SFM = 10\log_{10}\left(\frac{\left[\prod_{k=1}^{\frac{N}{2}} S_p(e^{j\frac{2\pi k}{N}})\right]^{1/(\frac{N}{2}+1)}}{\frac{1}{N/2+1}\sum_{k=1}^{\frac{N}{2}} S_p(e^{j\frac{2\pi k}{N}})} i\right). \tag{82}$$

Der gewonnene SFM-Wert wird mit dem eines Sinussignals verglichen ($SFM_{max} = -60dB$) und daraus der Tonalitätsindex bestimmt:

$$\alpha = MIN\left(\frac{SFM}{SFM_{max}}, 1\right). \tag{83}$$

Anschließend kann die Mithörschwelle $T(i)$ folgendermaßen berechnet werden:

$$T(i) = 10^{[L_S(i)-O(i)]/10}. \tag{84}$$

Die Maskierung benachbarter Gruppen erfolgt nun mit Hilfe zweier Kurven. Hin zu tiefen Frequenzen fällt Kurve mit einem konstanten Wert mit 27dB pro Bark:

$$S_1 = 27dB/Bark. \tag{85}$$

Die obere Flanke ist von der Gruppenmittelfrequenz f_{mi} und des Gruppen Schallleistungspegels $L_S(i)$ abhängig:

$$S_2 = 24 + 0.23\left(\frac{f_{m_i}}{kHz}\right)^{-1} - 0.2\frac{L_S(i)}{dB}dB/Bark. \tag{86}$$

Benachbarte Frequenzen welche unterhalb dieser Flanken liegen, werden vom menschlichen Gehör nicht wahrgenommen, und können vernachlässigt werden.

Abbildung 8 zeigt dies grafisch. Die benachbarten Frequenzbänder von f_i, f_{i+1} und f_{i-1} sind vom menschlichen Gehör nicht wahrnehmbar, falls ihr Pegel unterhalb der linken S_1 beziehungsweise rechten Flanke S_2, der Mithörschwelle von Band i liegen.

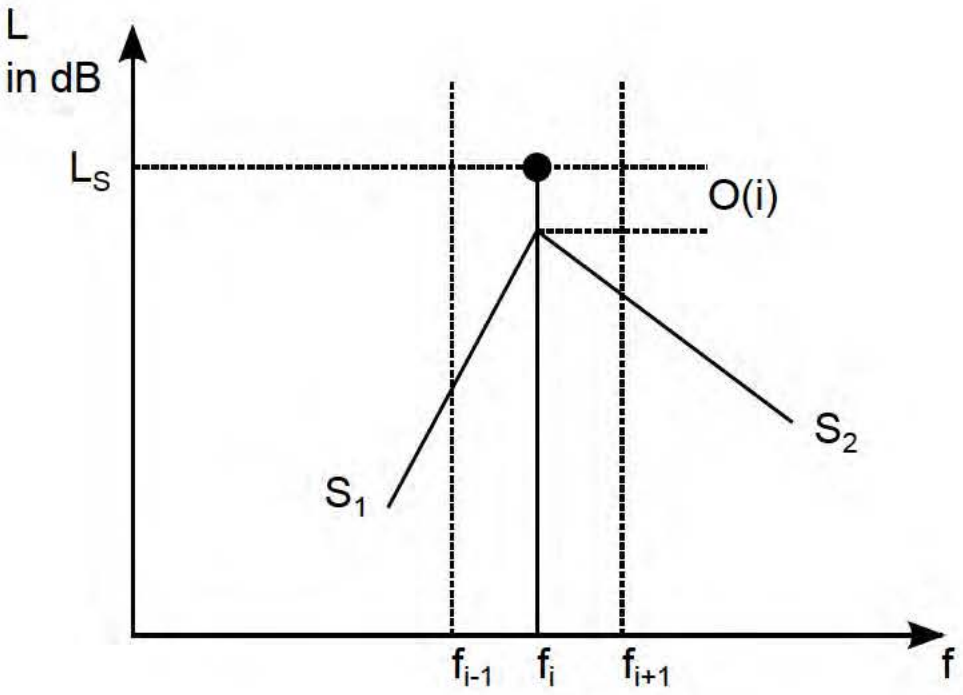


Abbildung 8: Maskierung benachbarter Frequenzgruppen [41]

Teil III

KLANGMODELL

3 KLANGMODELL

In diesem Kapitel wird das in dieser Arbeit verwendete Klangmodell physikalisch-mathematisch definiert. Da keine physikalischen Informationen über die untersuchten Klänge zur Verfügung stehen, muss ein gedankliches Modell erstellt werden, welches so in der Natur nicht vorkommt, aber notwendige Vereinfachungen erklärt.
Generell kann der Signalverlauf des Klanges eines Musikinstrumentes wie folgt mathematische beschrieben werden [42]:

$$s(t) \quad \sum_{n=1}^{N(t)} A_n(t)\sin(\theta_n(t)) + R(t). \tag{87}$$

Dabei beschreibt folgender Term die zeitabhängige Phase eines Spektralbestandteils n:

$$\theta_n(t) \quad 2\pi \int_0^t f_n(\xi)d\xi + \theta_{n_0}. \tag{88}$$

Wobei t die Zeit, $A_n(t)$ die Amplitude, n der Index der Mode, $N(t)$ die zeitabhängige Gesamtnummer der Moden, $\theta_n(t)$ die Phase der Mode n zur Zeit t, $f_n(t)$ die Frequenz der Mode n zur Zeit t, θ_{n_0} die Phase der Mode n zum Zeitpunkt $t \quad 0$ und $R(t)$ den zeitabhängigen Rauschanteil darstellt. Der Klang eines Instrumentes ergibt sich aus der Überlagerung der einzelnen Spektralanteile.
Ist der Rauschanteil $R(t)$ vernachlässigbar klein, so kann 87 zu

$$s(t) \quad \sum_{n=1}^{N(t)} A_n(t)\sin(2\pi \int_0^t f_n(\xi)d\xi + \theta_{n_0}). \tag{89}$$

vereinfacht werden.

In dieser Arbeit werden perkussive Klänge untersucht. Perkussive Instrumente zeichnen sich durch ein sehr kurzes Einschwingverhalten aus, so dass sich die Anregung im Idealfall durch eine Impulsanregung annähern lassen.
Klänge, welche diesem Kriterium genügen, sind Schlaginstrumente und Klaviere. Gezupfte Saiteninstrumente lassen sich weniger durch einen impulsartig angeregten Schwinger, als viel mehr durch einen freien Schwinger mit der Startbedingung $x \;/\; 0$ annähern.
Der Fokus dieser Arbeit liegt auf der Analyse und der Synthese linear schwingender Instrumente, bei denen $f_n(t) \quad f_n$ gilt und die Anzahl der Moden $N(t) \quad N$ konstant sind. Somit vereinfacht sich (89) weiter:

$$s(t) \quad \sum_{n=1}^{N} A_n(t)\sin(2\pi f_n t + \theta_{n_0}). \tag{90}$$

Das in dieser Arbeit verwendete Klangmodell basiert auf der Schwingung eines idealen gedämpften periodischen Mehrmassenschwingers. Für die Amplitude $A_n(t)$ gilt somit:

$$A_n(t) = A_n e^{-t/\tau_n}. \tag{91}$$

Da weiter vereinfachend davon ausgegangen wird, dass die Klänge impulsartig angeregt werden, fällt der Phasenterm in Gleichung (90) weg. Außerdem ist in der Synthese des Klanges von in *Synth-A-Modeler* keine Integration der Phase vorgesehen. Somit gilt für den Signalverlauf des in dieser Arbeit betrachteten Klangmodells:

$$s(t) = \sum_{n=1}^{N} A_n e^{-t/\tau_n} \sin(2\pi f_n t). \tag{92}$$

Der gesamte Signalverlauf lässt sich also durch drei Parameter pro Mode beschreiben. Der Modenfrequenz f_n, der Abklingkonstante τ_n und einer Amplitude A_n.
Im eigentlichen Sinne werden nicht die Moden bestimmt, sondern Frequenzen, dessen Energien sich vom Rauschsignal abheben. Die Abstrahlung wird in diesem Modell nicht berücksichtigt. Ebenso wird davon ausgegangen, dass es sich bei dem strahlenden Objekt um eine atmende Kugel [43] handelt, welche ungerichtet abstrahlt. Es gilt für die Auslenkung $z(t, \varphi, \vartheta)$auf der Kugeloberfläche:

$$\frac{\partial z}{\partial \varphi} = \frac{\partial z}{\partial \vartheta} = 0 \tag{93}$$

Abbildung 9 zeigt das Schallfeld einer atmenden Kugel. Die Abstrahlung ist hier unabhängig von der Richtung.

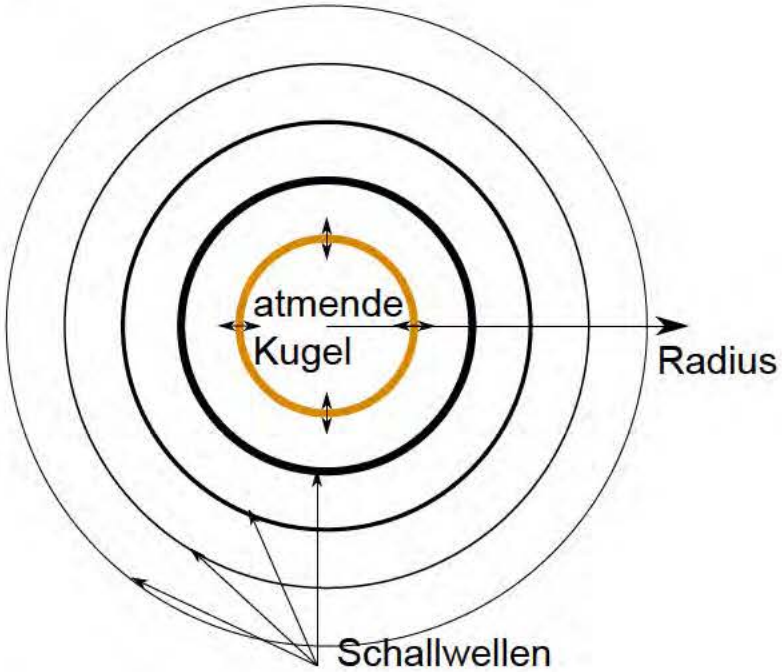


Abbildung 9: Schalldruckfeld einer atmenden Kugel [43]

Das idealisierte Schwingungsverhalten einer Mode n mit der Frequenz f_n zum Zeitpunkt t hat folgende Form:

$$s_n(t) = A_n e^{-t/\tau_n} \sin(2\pi f_n t) \tag{94}$$

Dies bedeutet, dass die Schwingung einer Mode zum Zeitpunkt $t = \pi/2 - \phi_n$ ihr Maximum erreicht und die Amplitude anschließend exponentiell abfällt. In Abbilung 10 ist der Verlauf einer solchen Mode zu sehen. Dabei ist $A_n e^{-t/\tau_n}$ die Einhüllende.

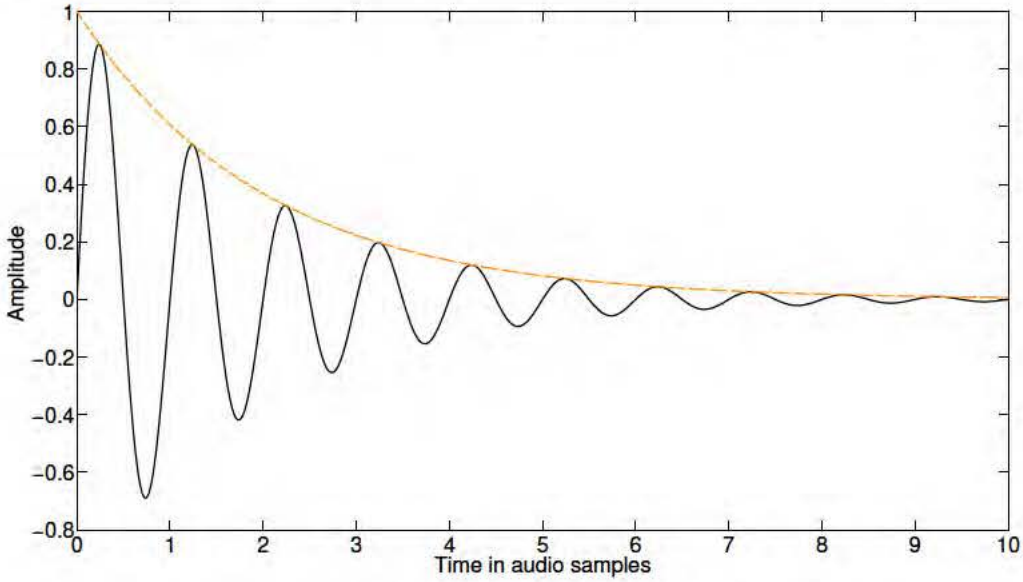


Abbildung 10: Signalvergleich einer Mode
— Einhüllende, — $s_n(t)$

Nichtlineare Schwinger können in der Regel nicht durch das verwendete Klangmodell beschrieben werden. Für die Parametrisierung einiger nicht linearer Signalverläufe kann $f_n(\xi)$ in Gleichung (95) durch eine geeignete Funktion dargestellt werden.

$$\theta_n(t) = 2\pi \int_0^t f_n(\xi) d\xi + \theta_{n_0} \tag{95}$$

Mit dieser Beschreibung können aber nur Signale mit einem nicht chaotischen Verhalten beschrieben werden. Wie in Kapitel 2.1.5 gezeigt, sind die Frequenzen der Moden abhängig von der Auslenkung z, beziehungsweise der dadurch hervorgerufenen Spannung $T(z)$. Im Fall der atmenden Kugel also vom zeitabhängigen Radius $r(t)$. Da in den zu untersuchenden Signaldateien keinerlei Informationen über die Geometrie der Instrumente vorhanden sind, kann nur die Amplitude $s(n)$ als Indikator für die Auslenkung z dienen. Somit muss ein Zusammenhang zwischen der Signalamplitude $|s(n)|$ und der veränderlichen Modenfrequenz $f_n(n)$ hergestellt werden. Im Falle der atmenden Kugel mit Ruheradius $R = z(0)$, ist die Auslenkung $z(t)$ konstant über die gesamte Kugeloberfläche. Somit ist die Änderung der Frequenz der Moden einfach eine unbekannte Funktion der Signalamplitude $s(n)$:

$$f_n(n) = F(|s(n)|) \tag{96}$$

Ob eine Annäherung durch die, in dieser Arbeit angewandten Methoden möglich ist, wird in Kapitel **??** und 5.6 behandelt und diskutiert.

Teil IV

METHODEN

4 METHODEN

4.1 ANALYSE DER KLANGBEISPIELE

Im Folgenden werden die Analyse- und Resynthesemethoden der Klangbeispiele erläutert.

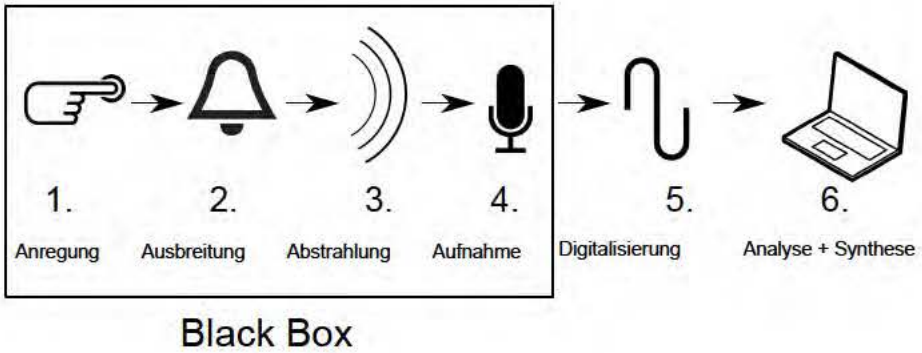


Abbildung 11: Weg des Signals bis zur Analyse

In Abbildung 11 ist der Signalweg des Klanges von der Anregung bis zur Analyse und Synthese schematisch dargestellt. Die zur Analyse bereitstehenden Daten bestehen allein im digitalisierten Signal (5). Es stehen keine Informationen über die Form der Anregung (1), die Ausbreitung im Objekt (2), über die Abstrahlung (3) und über die Art der Aufnahme (4) zur Verfügung. Diese Punkte sind kein Bestandteile der Analyse. Vereinfachend wird angenommen, dass es sich um eine Impulsanregung handelt. Die Ausbreitung, Abstrahlung und Aufnahme werden vernachlässigt. Dies bedeutet, dass in dem in dieser Arbeit verwendeten Modell, die Aufnahme keinen Einfluss auf die empfangene Schalldruckwelle hat und die Schalldruckwelle frequenzunabhängig von dem schwingenden Objekt angeregt wird. Mathematisch kann man diese Annahmen wie folgt ausdrücken:

$$s_a(t) = \sum_{n=-\infty}^{\infty} s(n\frac{1}{F_s})\delta(t - n\frac{1}{F_s}) = a \cdot m(t) = b \cdot p(t) = c \cdot x(t) \tag{97}$$

Dabei bezeichnen a, b und d Proportionalitätsfaktoren die unbekannt sind. Die Funktion $s_a(t)$ beschreibt das ideal abgetastete Signal [44], $s(t)$ beschreibt die Amplituden des kontinuierlichen Signals nach der Mikrophonaufnahme, $m(t)$ beschreibt die Auslenkungen der Mikrophonmembran, $p(t)$ den Schalldruck und $x(t)$ die Auslenkung der Objektoberfläche.

4.2 GESAMTSCHAU DER SIGNALANALYSE

In Abbildung 12 ist eine schematische Darstellung der Signalanalyse zu sehen. Ausgangspunkt der Analyse bildet eine WAVE oder AIFF Datei mit dem zu

analysierenden Klang. Im zweiten Schritt wird eine Vorbearbeitung der Audiodatei durchgeführt. Anschließend werden die spektralen Bestandteile des Signals in Abhängigkeit von der Zeit $H(m, k)$ mit Hilfe der STFT ermittelt. Mit Hilfe von $H(m, k)$, kann das Energy Decay Relief $EDR(t_n, f_k)$ berechnet werden. Im nächsten Schritt werden mögliche Eigenmoden f ausgewählt. Daraufhin wird die Eigenschaft der detektierten Moden hinsichtlich psychoakustischer Methoden untersucht und diese, wenn möglich, entfernt. Mit Hilfe der ausgewählten und gestutzten Moden f_s und des Energy Decay Reliefs können nun die Abklingkonstanten τ berechnet werden. Die Amplitude A wird mit Hilfe des Spektrogramms $H(m, k)$ und der im vorherigen Schritt berechneten Abklingkonstanten τ ermittelt. Anschließend werden die mit Hilfe des Klangmodells erstellten Amplitudenverläufe jeder Mode einer weiteren psychoakustischen Untersuchung unterzogen und wenn möglich wiederum gestutzt. In einem letzten Arbeitsschritt kann der Nutzer abhängig von ihrer Bedeutung für den Klang weitere Moden stutzen Das berechnete Klangmodell wird dann als *Synth-A-Modeler*-Datei (.mdl) gespeichert, so dass der Musiker diesen modellierten Klang nutzen kann. Die Analyseparameter und Methoden wurden

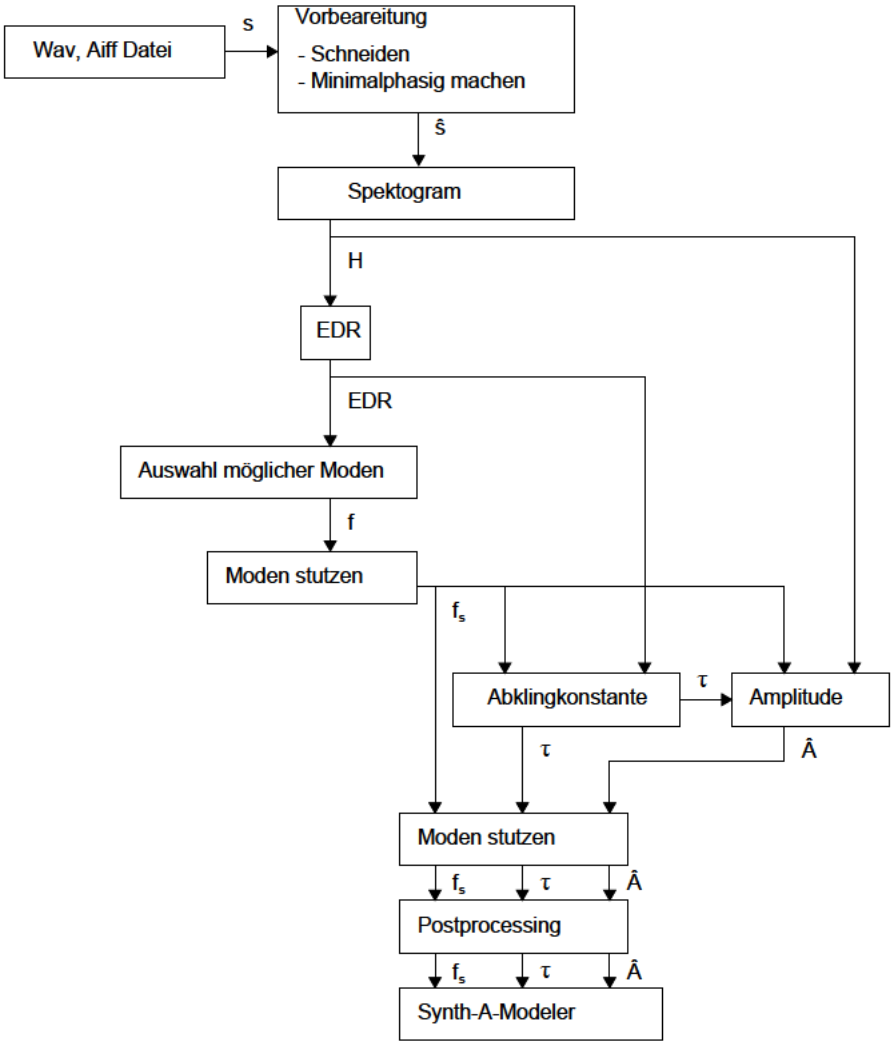


Abbildung 12: Schematische Darstellung der Signalanalyse

auf der Grundlage vieler Versuche optimiert, um die psychoakustische Qualität der synthetisierten Klänge zu verbessern.

4.3 SIGNALVORBEARBEITUNG

Die Signalvorbearbeitung umfasst zwei Schritte: Zum einen wird das Signal minimalphasig gemacht. Dies bedeutet, dass die Energie einer jeden Frequenz an den Anfang des Signals verschoben wird und so die Phasenverschiebungen minimiert werden. Der zweite Schritt besteht im Kürzen des Signals. Dieser Schritt ist aus zwei Gründen wichtig: Die Auswertung eines langen Signals ist sehr rechenintensiv, was die Analyseprozedur verlängern würde. Weiter ist die Auswertung des gesamten Signals nicht notwendig, da für die Parametrisierung des Modells neben der Frequenz f nur die Abklingkonstante τ und die Amplitude A angenähert werden muss. Die Amplitude ist nach dem hier verwendeten Modell am Anfang des Signals zu finden. Die Abklingkonstante kann bei einem idealen Signal, also ein Signal mit exponentiell abklingenden Spektralanteilen, schon aus den ersten beiden Zeitschritten des Spektrograms $H(1, f_k)$ und $H(2, f_k)$ berechnet werden. Um die Genauigkeit der Analyse zu erhöhen und unter Berücksichtigung realer Signale wird die Analysedauer des Signals $s(n)$ standardmäßig auf einen Wert von $N = 32768$ Audiosamples gesetzt.

4.4 SPEKTRALANALYSE

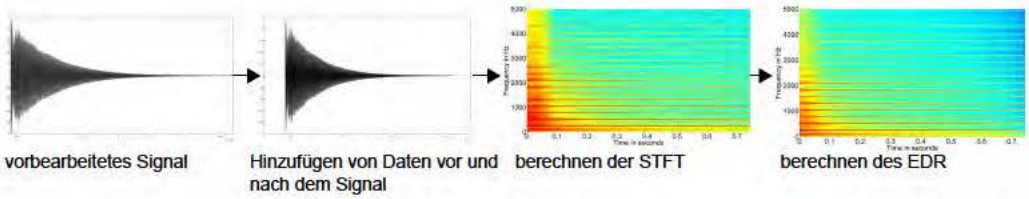


Abbildung 13: Spektralanalyse

Die Spektralanalyse beinhaltet zwei Berechnungen. Der erste Schritt besteht in der Berechnung der zeitabhängigen spektralen Verteilung mit Hilfe der Kurzzeit Fourier Transformation. Basierend auf diesen Daten wird das Energy Decay Relief berechnet.

Um eine möglichst originalgetreue Resynthese zu realisieren, ist es wichtig, dass die Daten im spektral- und Zeitbereich für jeden Zeitschritt übereinstimmen, im Spektralbereich also auch der Wert $t = 0$ und $t = t_{max}$ des Originalsignals vorhanden sind. Aus diesem Grund werden vor dem Analysesignal $W/2$ Nullen hinzugefügt. Ist die Länge des original Signals vor dem Kürzen N_{orig} kleiner als $N_{orig} < N + W/2$, so werden auch Nullen hinter das Signal angefügt. Ansonsten wird der Bereich hinter dem Signal mit $W/2$ Daten aus dem Originalsignal aufgefüllt. Die Anzahl der hinzugefügten Nullen hängt von der Fenstergröße ab und beträgt $W/2$. Die Auffüllung des Anfangs- und endes ist notwendig, um die Mitte des ersten beziehungsweise letzten Fensters jeweils zum Zeitschritt t_1 und t_{max} ansetzen zu können (Siehe Abbildung 14). Somit

können die Zeitdaten des Spektrogramms mit den Originaldaten verglichen werden.

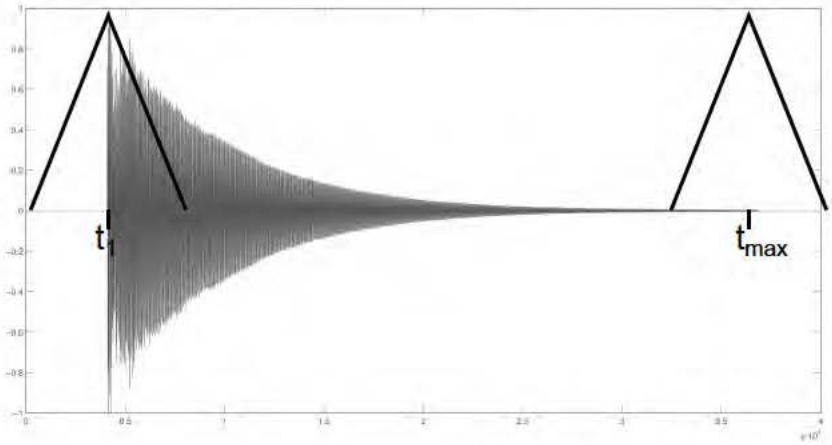


Abbildung 14: Mit Nullen aufgefülltes Signal

Mit Hilfe des verlängerten Signals $\hat{s}$ wird nun eine Kurzeit Fourier Transformation durchgeführt. Um möglichst viele Zeitschritte im Spektralbereich zu erhalten, wird die Kurzeit Fourier Transformation mit sich überlagernden Fenstern berechnet. Außerdem werden zu jedem gefensterten Analysesegment noch W Nullen hinzugefügt, so dass die effektive FFT Größe $W_{FFT} = 2W$ beträgt. Anschließend wird aus den Daten des Spektrogramms das Energy Decay Relief berechnet. Dieses dient als Grundlage zur Modendetektierung und Abklingkonstantenberechnung. Die Daten des Spektrogramms werden dann mit den vorher berechneten Abklingkonstanten zur Berechnung der Amplituden verwendet. Der Pseudocode 1 zeigt die Berechnung des Spektrogramms.

```
input  : ŝ(m),w(n),O
output : H_abs
begin
    step = 1 ;
    stepSize = W · O ;
    k = 1;
    f_i = {1, ..., W/2} ;
    while step < M − W/2 do
        m_i = {step, ..., step + W − 1} ;
        ŝ_m_i = ŝ(m_i) · w(n) ;
        H = fft(ŝ_m_i);
        H_abs(m, k) = abs(H(f_i)) · 2;
        step = step + stepsize ;
        k = k + 1;
    end
end
```

Algorithmus 1 : Berechnung von H_{abs}

4.5 MODENAUSWAHL

Die Modenauswahl erfolgt über die im Signal enthaltenen Energien einer jeden Frequenz f_k. Im ersten Schritt werden alle möglichen Kandidaten für Moden gesucht. Im einem zweite Schritt werden die irrelevanten Moden aussortiert. Mit Hilfe des EDR wird jeder Frequenz f_k eine Restenergie im Zeitfenster t_n zugeordnet. Zum Zeitpunkt $t_n = 1$ erhält man also die gesamte im Signal enthaltene Energie E pro Frequenz f_k. Mit Hilfe dieser Informationen werden die Frequenzen ausgewählt, die sich vom Rauschpegel abheben. Abbildung 15 zeigt ein EDR zum Zeitpunkt $t_n = 0$ und die ausgewählten Moden.

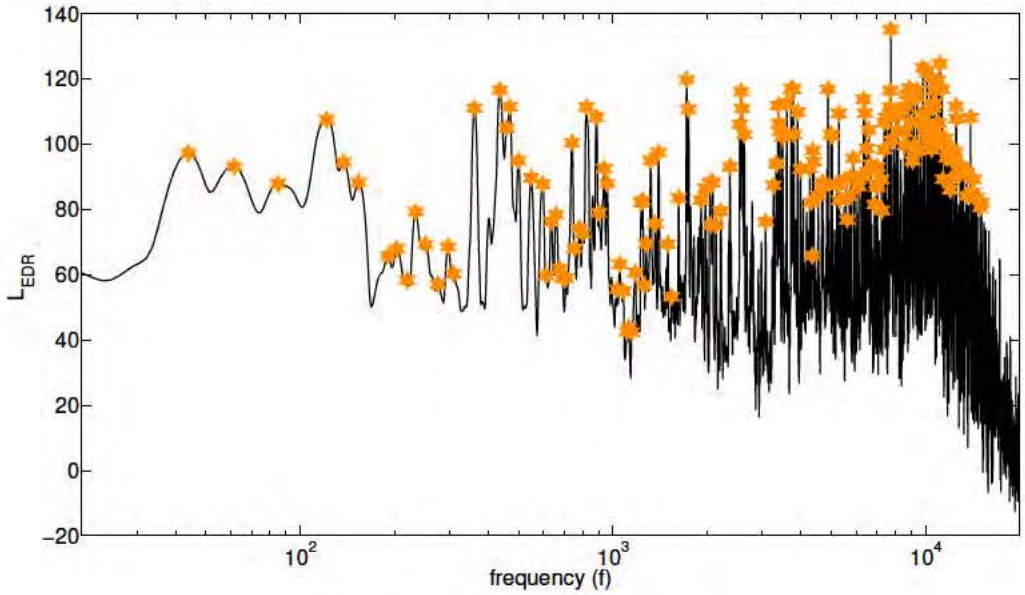


Abbildung 15: Modevorauswahl
$-$: $EDR(t_1, f_k)$; ★ : detektierte Moden

Die Auswahl der Moden erfolgt über die Spitzen im EDR im Zeitschritt t_1. Ein Peak bei einer Frequenz f_k zeichnet sich dadurch aus, dass die benachbarten Punkte im EDR einen kleineren Wert haben, also:

$$EDR(t_1, f_{k-1}) < EDR(t_1, f_k) > EDR(t_1, f_{k+1}) \tag{98}$$

Um Rauschbestandteile nicht zu berücksichtigen, wurde dabei eine Schwelle von $\gamma_t = 0.1$ gesetzt. Dies bedeutet, dass gefundene Peaks unter diese Schwelle nicht berücksichtigt werden. Die Schwelle wurde auf Grundlage der absoluten Hörschwelle bestimmt. Ein Mensch kann einen Ton egal welcher Frequenz, mit einem Schalldruckpegel von $L_p = -20\text{dB}$ nicht wahrnehmen. Dies bedeutet für die Schwelle γ, wenn der Schalldruckpegel aus den Daten der Fourier Transformierten H folgendermaßen berechnet wird:

$$-20\text{dB} > L_p(m,k) \geqslant 10\log_{10}\left(H(m,k)^2\right) \quad \Rightarrow H(m,k) \leqslant 0.1 \tag{99}$$

und die Moden mit Hilfe des EDRs mit $N > 10$ berechnet werden:

$$EDR(t_1, f_k) = \sum_{m=1}^{M} |H(m,k)|^2 . \tag{100}$$

Da $M \geqslant 10$ ist, also das Spektrogramm über mindestens 10 Zeitschritte verfügt, sind Werte $EDR(t_1, f_k) \leqslant 0.1$ umgerechnet in einen Schalldruckpegel für den Menschen nicht mehr wahrnehmbar.
Weiter wird hier ausgenutzt, dass der Mensch Frequenzen unterhalb von f 20Hz und oberhalb von 20kHz nicht wahrnehmen kann. Moden unterhalb und oberhalb dieser Schwelle werden ebenfalls nicht berücksichtigt.
Da die berechneten Peaks nicht zwingend mit den Resonanzen des Signals übereinstimmen, werden im nächsten Schritt weitere detektierte Moden aussortiert, welche vom Menschen nicht wahrgenommen werden. Dies dient der Reduzierung der Moden, und damit des Rechenaufwandes bei der Synthese, als auch der Entfernung von Frequenzen, die keine modalen Bestandteile des Signals sind.
Der Pseudocode 2 zeigt den Algorithmus zur Detektierung der Moden ω_k.

```
input  : EDR(1, f_k), γ
output : ω_k
begin
    i   1;
    for k ∈ [2, N − 1] do
        if EDR(1, f_k) > EDR(1, f_{k−1})   &&   EDR(1, f_k) >
        EDR(1, f_{k+1})   &&   EDR(1, f_k) > γ then
            ω_k   i;
            i   i + 1;
        else
        end
    end
end
```

Algorithmus 2 : Auswahl der Moden ω_k

4.5.1 *Stutzen der Moden*

Das Stutzen oder Entfernen der Moden erfolgt nach zwei psychoakustischen Methoden. Diese unterscheiden sich von dem in Kapitel 2.3 vorgestellten Modell. Sinn der Stutzung von Moden in der Audiokodierung ist die Reduzierung von Daten aufgrund der Wahl von gröberen Quantisierungsstufen in Spektralbereichen, welche maskiert werden. In dieser Arbeit jedoch sollen einzelne Moden gestutzt werden, um die Anzahl der Gesamtmoden zu reduzieren und so den Rechenaufwand der Synthese zu minimieren, ohne dass dadurch hörbare Veränderungen im Klang auftreten. Aus diesem Grund wird die Methodik aus Kapitel 2.3 angepasst. Eine erste Auswahl der zu stutzenden Moden wird mit Hilfe der absoluten Hörschwelle getroffen. Hierzu werden die Ergebnisse der STFT nach [41], in einen Schalldruckpegel umgewandelt:

$$L_s(m,k) \quad 10 \cdot \log_{10}\left(\frac{H_{Abs}(m,k)}{\gamma}\right)^2 \tag{101}$$

Anders als in Kapitel 2.3 vorgestellt, wird der Schalldruckpegel nicht für eine Frequenzgruppe berechnet, sondern für jede detektierte Mode. Der Korrektur-

faktor γ ist abhängig von der Fensterung [45]. Der Wert $\gamma = 0.42$ gilt für das *Blackman-Harris* Fenster. Für ein Rechteckfenster gilt $\gamma = 1$. Allgemein gilt für den Skalierungsfaktor γ der gefensterten Kurzeit Fourier Transformation:

$$\gamma = \frac{\sum_{n=1}^{W} w(n)}{W} \tag{102}$$

In Gleichung (102) beschreibt W die Fensterlänge und $w(n)$ das Fenster. Nun wird jeder Zeitschritt m von $L_s(m,k)$ mit dem absoluten Hörschwellenpegel $L_{T_{abs}}(k)$ verglichen. Abbildung 16 zeigt den absoluten Hörschwellenpegel — und den Pegel eines Signalabschnitts. Alle Moden dessen Pegel $L_S(m = 1, k)$ unterhalb der absoluten Hörschwelle liegen, sind vom menschlichen Gehör nicht wahrnehmbar.

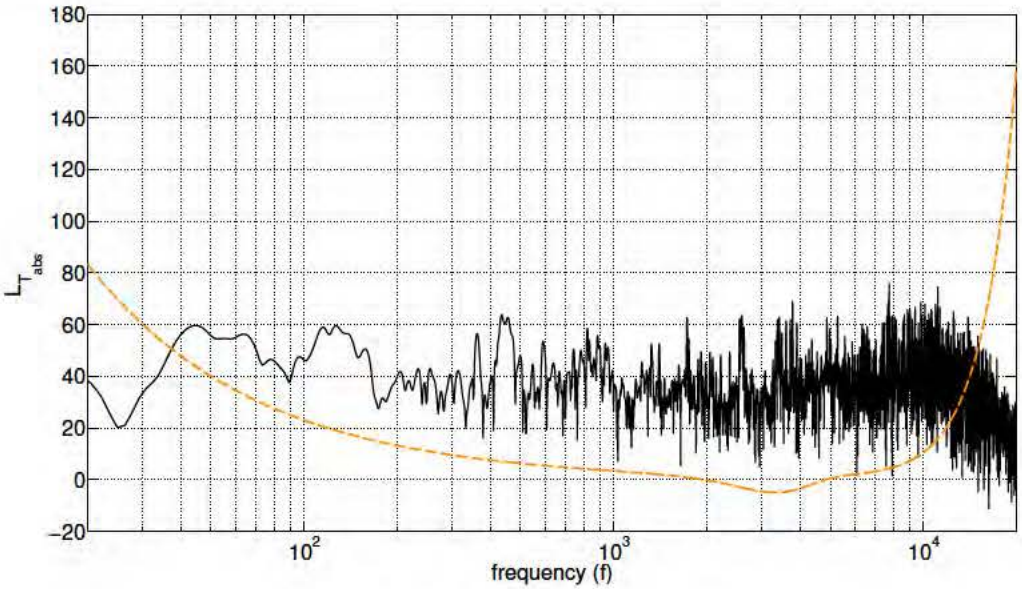


Abbildung 16: Absolute Hörschwelle
$-$: $L_S(m = 1, k)$; — : $L_{T_{abs}}(k)$

Eine Mode ω_k wird gestutzt wenn der Schalldruckpegel $L_s(m, \omega_k)$ in jedem Zeitschritt m unterhalb des zugehörigen absoluten Hörschwellenpegels $L_{T_{abs}}(\omega_k)$ liegt, also wenn für eine Mode gilt:

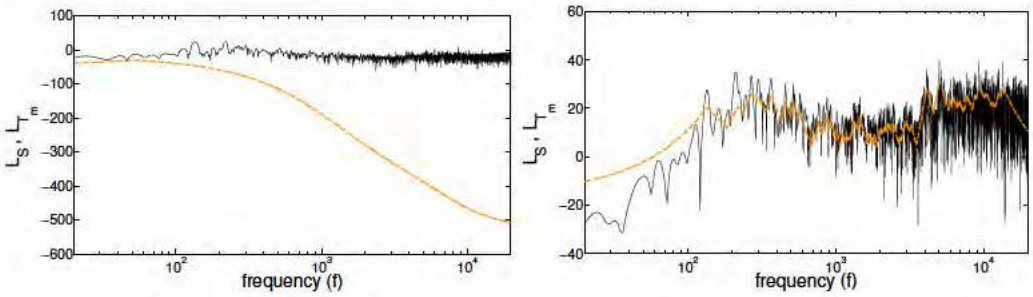
$$L_s(m, \omega_k) < L_{T_{abs}}(\omega_k) \quad \forall\, m \tag{103}$$

Da dieses Verfahren sehr große Rechenkapazitäten benötigt, wird jedoch praktisch nur jeder dritte Zeitschritt überprüft. Für exponentiell abfallende Amplituden und konstante Frequenzverläufe der Moden ist die Überprüfung eines jeden dritten Zeitschrittes ausreichend.

```
input  : H_abs(m,k), ω_k, γ
output : ω_k
begin
    L_s(m,k) = 10 · log10 ( H_Abs(m,k) / γ )^2 ;
    L_Tabs(k) ← absoluteHearingThreshold;
    foreach ω_k do
    inx = find(L_s(m, ω_k)) > L_Tabs(ω_k)  in  m) ;
    if inx == ∅ then
        delete(ω_k);
    end
    ;
end
```

Algorithmus 3 : Entfernen von Moden unterhalb der absoluten Hörschwelle

In einem weiteren Schritt werden Moden entfernt, welche unterhalb der globalen Mithörschwelle liegen.



a) Lokale Mithörschwelle
$-: L_S(1,k)$; —— : $L_{T_m}^{\omega_k}(1,k)$

b) Globale Mithörschwelle
$-: L_S(1,k)$; —— : $L_{T_{mglob}}(1,k)$

Abbildung 17: Lokale und globale Mithörschwelle

Dazu wird, ähnlich wie im ersten Schritt, für jeden Zeitschritt die globale Mithörschwelle (siehe Abbildung: 17 b)) aus den lokalen Mithörschwellen jeder Mode (siehe Abbildung: 17 a)) berechnet. Die globale Mithörschwelle wird aus den Maxima der einzelnen lokalen modalen Mithörschwellen $L_{T_m}^{\omega_k}(m,k)$ für jede Frequenz berechnet. Zum Beispiel für den ersten Zeitschritt $m = 1$:

$$L_{T_{mglob}}(1,k) = MAX(L_{T_m}^{\omega_k}(1,k)) \quad \forall\, \omega_k \tag{104}$$

Anschließend werden diejenigen Moden ω_k gestutzt, welche für alle Zeitschritte unterhalb des Mithörschwellenpegels des jeweiligen Zeitschrittes m liegen.

$$L_s(m,\omega_k) < L_{T_{mglob}}(m,\omega_k) \quad \forall\, m \tag{105}$$

```
input  : H_abs(m,k), ω_k
output : ω_k
begin
    L_s(m,k)   10 · log_10(H_Abs(m,k)^2) + γ;
    foreach m do
    L_T_mglob(m,k) ← findMax(L_T_m(m,k)) ;
    foreach ω_k do
    inx   find(L_s(m,ω_k)) > L_T_mglob(m,ω_k)   in   m) ;
    if inx   ∅ then
        delete(ω_k);
    end
    ;
end
```

Algorithmus 4 : Entfernen von Moden unterhalb der globalen Maskierungsschwelle

4.6 BESTIMMUNG DER MODALEN EIGENSCHAFTEN

Nach dem hier verwendeten physikalischen Klangmodell:

$$y(t) \quad \sum_{n=0}^{\infty} A_n e^{-t/\tau_n} \sin(2\pi f_n t + \theta_n), \tag{106}$$

bestehen die Eigenschaften φ_n einer jeden nicht schon gestutzten Mode aus drei Werten:

$$\varphi_n \quad (f_n, d_n, A_n) \tag{107}$$

Zum einen der Amplitude A_n und der Abklingkonstante τ_n. Nachdem die Frequnzen f_n der Moden bestimmt wurden müssen noch diese Eigenschaften ermittelt werden. Die Phasenverschiebung θ_n wird in erster Näherung vernachlässigt.

4.6.1 *Präzisieren der modalen Frequenz*

Die mit Hilfe der Daten der Kurzzeitfouriertransformation und der Modendetektierung ermittelten Modenfrequenzen unterliegen in ihrer Genauigkeit der spektralen Auflösung der Kurzzeitfouriertransformation. Um die Genauigkeit zu erhöhen, kann mit Hilfe der quadratischen Peak Interpolation die Frequenz der Mode genauer berechnet werden. Diese genauere Bestimmung erfolgt mit den Daten des EDR. Genau betrachtet muss im Gegensatz zu der in Kapitel 2.2.6 vorgestellten quadratischen Peak Interpolation, für das Spektrogramm eine biquadratische Peak Interpolationm, durchgeführt werden.
Die quadratische Peak Interpolation ist nur definiert für die Fouriertransformierte eines Zeitschrittes, da dort quasiparabole Kurven vorliegen. Für eine Folge von K Zeitschritten und N Moden, müssten also $K \times N$ quadratische Peak Interpolationen durchgeführt werden. Ebenso müssten für jeden Zeitschritt die lokalen Maxima beziehungsweise Moden berechnet werden, damit die quadratische Interpolation durchführbar ist. In dieser Arbeit wurden die Moden zur

Simplifikation mit Hilfe der Informationen des des ersten Zeitschrittes t_1 des Energy-Decay-Relief ermittelt. Somit muss auch die quadratische Peak Interpolation auf Grundlage dieser Daten durchgeführt werden. Um quasiparabole Kurvenverläufe EDR_{para} zu erhalten, werden vor der Quadratischen Peak Interpolation die Daten des EDR umgerechnet:

$$EDR_{para}(t_n, f_k) \quad \sqrt{EDR(t_n, f_k)}. \tag{108}$$

Anschließend wird mit den Daten von $EDR_{para}(t_n, f_k)$ und den vorher ermittelten Moden ω_k die quadratische Peak Interpolation durchgeführt.

4.6.2 *Bestimmung der Abklingzeit*

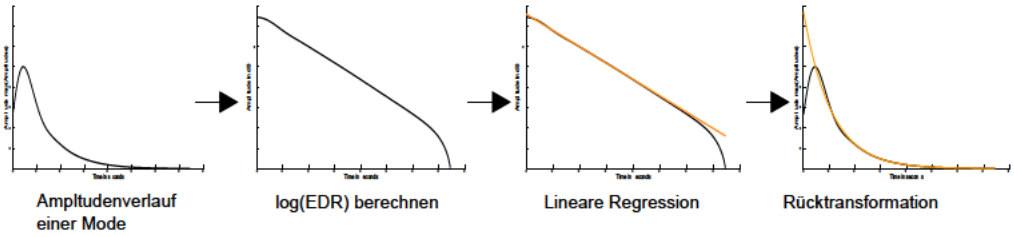


Abbildung 18: Bestimmung der Abklingzeit
— : exakter Verlauf, — — : angenäherter Verlauf

Die Bestimmung der Abklingzeit τ wird auf Grundlage der Daten des EDR realisiert und mit Hilfe zweier Regressionsanalysen durchgeführt. Für jede Mode ω_k liegt im Zeitbereich t_n ein Amplitudenwert vor. Für eine exponetiell abklingende Amplitude einer Mode ω_k gilt für die Hüllkurve des EDR nach Gleichung 60 folgender Verlauf:

$$\hat{z}(t_n, f_k) \quad \hat{A}^2 \frac{\hat{\tau}}{2} e^{-2t/\hat{\tau}}. \tag{109}$$

Logarithmiert man $\hat{z}(t_n, f_k)$, so erhält man einen linearen Verlauf und es ist möglich eine lineare Regression durchzuführen, um die Parameter β_0 und β_1 zu bestimmen:

$$\begin{aligned} \hat{z}(t_n, f_k(m))_{log} \quad & \log(\hat{A}^2 \frac{\hat{\tau}}{2} e^{-2t/\hat{\tau}}) \\ & \underbrace{\log(\hat{A}^2 \frac{\hat{\tau}}{2})}_{\beta_0} - \underbrace{\frac{2}{\hat{\tau}}}_{\beta_1} t. \end{aligned} \tag{110}$$

Die lineare Regression [46] ist definiert als Minimierungsproblem:

$$\min_{\beta_0, \beta_1} Q(\beta_0, \beta_1) \quad ; \quad Q(\beta_0, \beta_1) \quad \sum_{i=1}^{k} \hat{\varepsilon}_i^2 \quad \sum_{t_n=1}^{k} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{111}$$

und bestimmt die Konstanten β_0 und β_1 der linearen Gleichung (112) bei minimierten Q.

$$Y(m) \quad \beta_0 + \beta_1 m + \varepsilon_m \tag{112}$$

Grund für die mehrstufige Regression ist, dass die Amplitudennverläufe der Moden der analysierten Signale in der Regel keinen exponentiellen Verlauf annehmen, so dass eine einfache lineare Regressionsanalyse der logarithmierten Daten teilweise zu Abklingzeiten mit psychoakustisch schlechten Ergebnissen führen würde.

Im ersten Schritt wird eine lineare Regression vom Zeitschritt $t_n \quad \alpha$ bis zum Zeitschritt $t_n \quad \Omega_1$ durchgeführt. Der Wert α ergibt sich aus der Überlappungsgröße O, beziehungsweise aus der Schrittweite der Überlappung O_{step}. Und begründete sich durch die vor dem Signal hinzugefügten Nullen, welche die Amplitudenwerte der Fouiertransformation verfälschen. Der Startwert α ist also der erste Werte im spektral Bereich, welcher im Zeitbereich vollständig durch Signalwerte abgedeckt ist. Damit ergibt sich α zu:

$$\alpha \quad \frac{W}{O_{step}}. \tag{113}$$

Der Endwert t_{Ω_1} des ersten Regressionsbereich ist abhängig von der Wurzel von $Q(\beta_0, \beta_1)$, also von der Standardabweichung der linearen Regression:

$$\varepsilon_R \quad \sqrt{Q} \quad \sqrt{\frac{1}{M}\sum_{m=1}^{M}(Y(m) - \log(EDR(m, \omega_k)))^2} \quad \sqrt{\frac{1}{M}\sum_{m=1}^{N}(\epsilon_m)^2} \tag{114}$$

Das Ende des Bereichs der ersten linearen Regression ist dann erreicht, wenn der Fehler ϵ_R einen bestimmten Wert überschreitet. Der Wert ϵ_R wurde anhand von Versuchen auf folgenden Wert gesetzt.

$$\varepsilon_R < \frac{10}{\log(f)} \tag{115}$$

Unter Berücksichtigung, dass die Abklingzeit nach dem Dämpfungsmodell aus Kapitel 2.1.6 in der Regel mit steigender Frequenz abnimmt und somit schneller in den Rauschpegel übergeht, sinkt die Fehlerschranke mit steigender Frequenz, da gilt:

$$\omega_k > \omega_{k-1} \Rightarrow \tau_k < \tau_{k-1} \quad \text{und} \quad \lim_{k\to\infty} \tau_k \quad 0. \tag{116}$$

Für sehr große ω_k, würde also der Rauschboden bei einer zu großen Fehlerschranke ε_R fälschlicherweise als Amplitudenverlauf der Mode ω_k detektiert werden. Um dieses Problem zu vermeiden, ist die Fehlerschranke frequenzabhängig.
Der zweite Regressionsbereich liegt zwischen t_{Ω_1} und t_{Ω_2}, wobei Ω_2 nach dem gleichen Kriterium wie Ω_1 ermittelt wird.

Anschließend werden aus den berechneten Regressionsmodellen die benötigten Parameter ermittelt.
Hierbei werden zwei Fälle unterschieden. Im ersten Fall ist die Steigung der ersten Regression (*) größer als die die Steigung der zweiten Regression (——). Siehe hierzu Abbildung 19). Also wird die Steigung der zweiten Regression —— als Steigungskoeffizient ausgewählt .

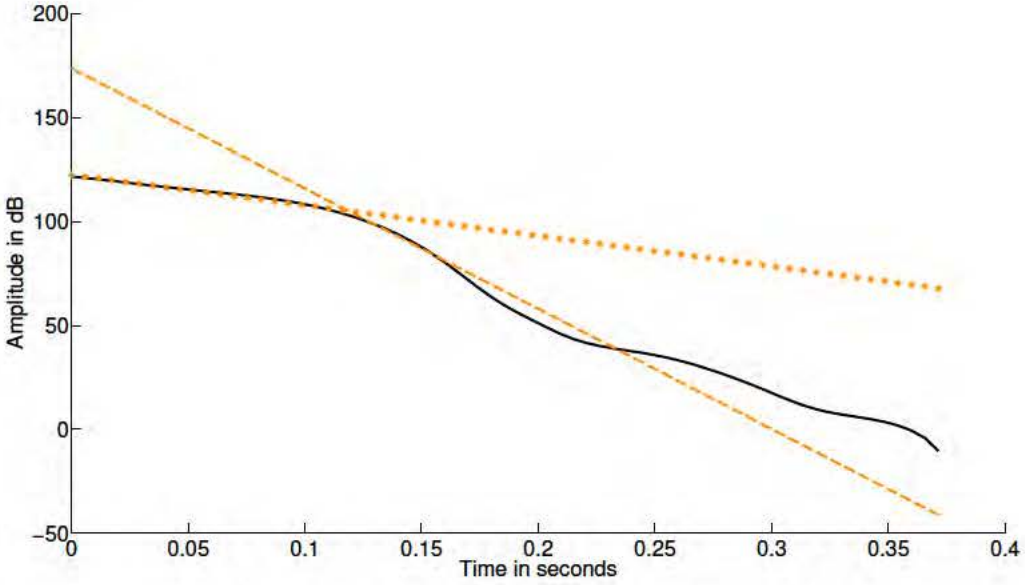


Abbildung 19: Regressionsanalyse: erster Fall
∗ : erste Regression; — : zweite Regression; — : gewählte Regression

Im zweiten Fall ist die Steigung der ersten Regression (—) kleiner als die die Steigung der zweiten Regression (∗). Siehe hierzu Abbildung 20). Also wird die Steigung der ersten Regression als Steigungskoeffizient ausgewählt. Der Steigungskoeffizient β_1 wird also einfach aus dem Minimum der beiden durch die linearen Regressionen angenäherten Steigungskoeffizienten ausgewählt: $\beta_1 = \min(\beta_{11}, \beta_{12})$

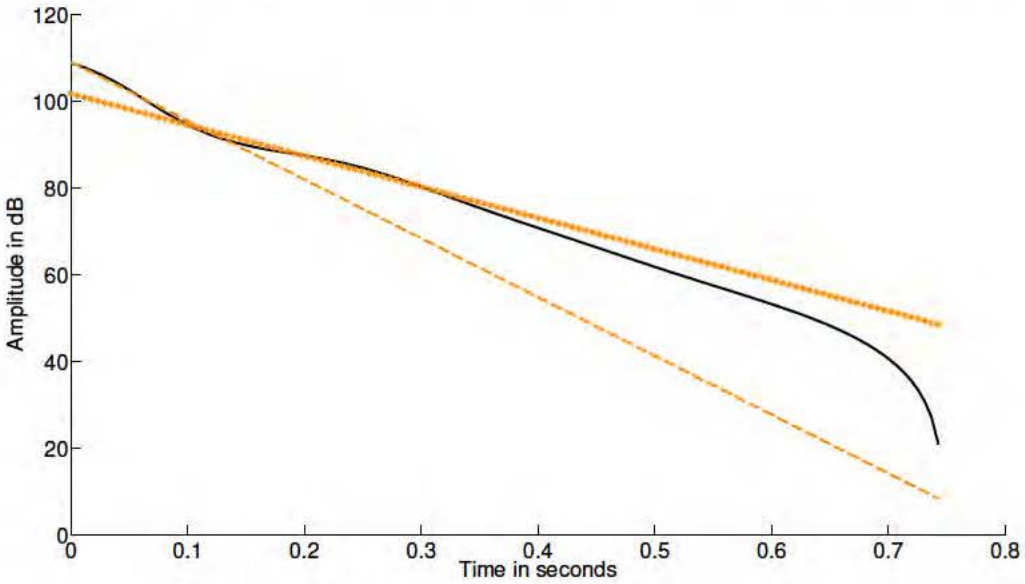


Abbildung 20: Regressionsanalyse: zweiter Fall
— : erste Regression; ∗ : zweite Regression; — : gewählte Regression

Nach erfolgter linearer Regression erhält man für jede Mode ω_k nun die Koeffizienten β_0 und β_1 der linearen Gleichung (112). Um nun die Abklingkonstante

τ zu erhalten, muss die erhaltene rücktransformierte Steigung $\beta_e \quad e^{\beta_1}$ nach Gleichung (60) nun einfach durch den Faktor Zwei dividiert werden.

$$\tau \quad \frac{\hat{\tau}}{2} \quad \frac{\beta_e}{2} \tag{117}$$

```
input  : EDR_log(t_n, f_k), ω_k, t_max, ε_R
output : τ_k
begin
    x_i  (t_max/(N−1))(n − 1) where n ∈ [1, N] ;
    foreach ω_k do
    k  1 ;
    y_i  EDR_log(t_n, ω_k) where t_n ∈ [1, N] ;
    while √(min_{β0,β1} Q(β01, β11)) < ε_R  &&  k < N do
        Find min_{β0,β1} Q(β01, β11);
        k  k + 1 ;
    end
    while √(min_{β2,β3} Q(β02, β12)) < ε_R  &&  k < N do
        Find min_{β2,β3} Q(β02, β12) ;
        k  k + 1 ;
    end
    τ_k  (1/2) min(β11, β12) ;
end
```

Algorithmus 5 : Berechnung der Abklingzeit τ

4.6.3 *Bestimmung der Amplitude*

Die Bestimmung der Amplitude kann auf zwei verschiedene Weisen durchgeführt werden. Zum einen kann nach Gleichung (60) aus den Daten der im vorherigen Kapitel beschriebenen linearen Regression ebenfalls die Amplitude A berechnet werden:

$$A \quad 2\tau\sqrt{\hat{A}} \quad 2\tau\sqrt{e^{\beta_0}}. \tag{118}$$

Die Bestimmung der Amplitude mit Hilfe der linearen Regression hat den großen Nachteil, dass sich die Fehler ϵ_R, welche bei der linearen Regression entstehen, fortpflanzen und auf die Genauigkeit der Amplitude A übertragen. Der Fehler kann dabei zwei Ursachen haben: Zum einen ist die lineare Regression selbst fehlerbehaftet und zum anderen wird der Fehler ϵ_R umso größer je weiter der untersuchte Klang von dem in dieser Arbeit verwendeten Klangmodell abweicht.

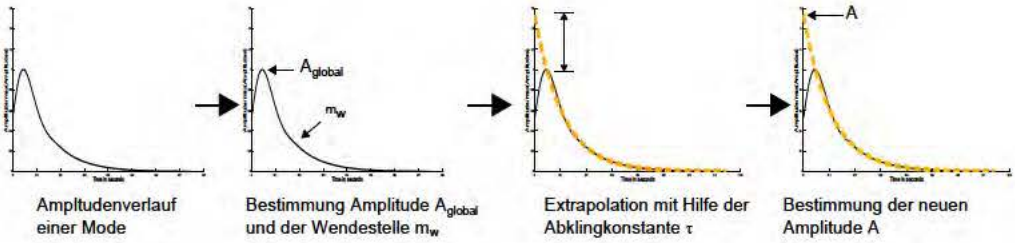


Abbildung 21: Bestimmung der Amplitude
– : exakter Verlauf, — : angenäherter Verlauf

Daher wird in dieser Arbeit eine andere Methode zur Amplitudenbestimmung verwendet. Die mit der Kurzzeitfouriertransformation gewonnenen Daten $H(m, k)$ beinhalten für jede detektierte Mode ω_k den Amplitudenverlauf in Abhängigkeit von der Zeit. Somit kann für jede Mode auch die globale Amplitude A_{global} ermittelt werden.
Das in dieser Arbeit verwendete Modell berücksichtigt keine Phaseninformationen, somit befindet sich die Amplitude A_{global} im Spektralbereich theoretisch immer im ersten Zeitschritt. Aus zwei Gründen ist dies in der Regel nicht der Fall: Zum einen benötigen die Moden eine unbekannte Zeit t_{ex}, um das Maximum zu erreichen. Die Zeit t_{ex} ist abhängig von der Art der Anregung und von dem untersuchten Objekt selbst. Zum anderen werden aufgrund der vor dem Signal hinzugefügten Nullen, die Werte zu Beginn des Signal fehlerhaft berechnet.
Um die Amplitude A des Klangmodells zu bestimmen, wird daher eine Extrapolation mit Hilfe der globalen Amplitude A_{global} bei $m = m_A$, dessen zugehöriger Zeitwert t_A und der Abklingkonstante τ durchgeführt. Da die Daten im Bereich von $H(m, k) \forall m < W/(2 \cdot O_{step})$ nicht exakt sind, kann eine globale Amplitude A_{global}, die für perkussive Instrumente in diesem Bereich liegen sollte, nicht als einziges Kriterium für die Bestimmung der Amplitude A hinreichend sein.
Daher wird im Bereich $m > m_A$ hinter der globalen Amplitude A_{global} der Wendepunkt als Parameter hinzugezogen. Da der Bereich nach der Globalen Amplitude eine negative Steigung aufweist wird also der Punkt gesucht in dem der Differentialquotient zweiter Ordnung sein Vorzeichen von negativ zu positiv wechselt. Abbildung 22 verdeutlicht dies. Das analysierte Signal ist ein exponentiell abfallender Sinuston.
Die detektierte globale Amplitude A_{global} liegt zeitlich kurz vor dem Wendepunkt. Würde nun die globale Amplitude als Referenzwert für die exponentielle Regression herangezogen werden, wäre das Ergebnis falsch. Da ein exponentiell abklingender Verlauf keinen Wendepunkt besitzt, muss der Wendepunkt $m = m_W$ als Referenzwert herangezogen werden. Die gesuchte Amplitude A kann so, im Falle eines exponentiell abklingenden Amplitudenverlaufs, korrekt ermittelt werden.

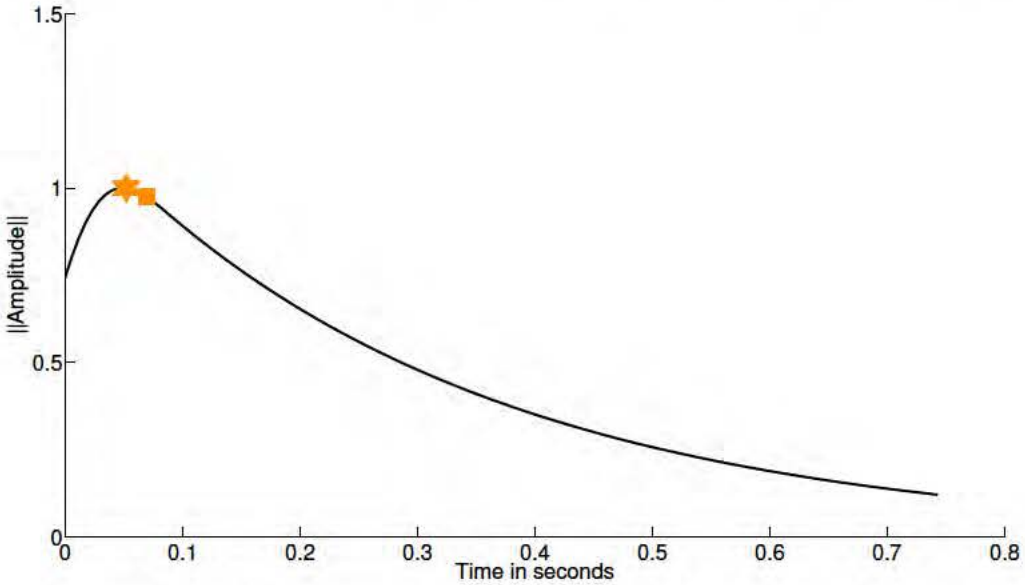


Abbildung 22: Amplitude und Wendepunkt
■ : Wendepunkt, ★ : Amplitude , — : Amplitudenverlauf

Mit Hilfe dieser Informationen wird der diskrete Differenzenquotient zweiter Ordnung im Bereich von $m = m_A$ bis $m = M$ für jede Moden ω_k berechnet, wobei m_A den Punkt der globalen Amplitude im Verlauf darstellt:

$$\begin{aligned}\vartheta(m,\omega_k)^2 &= \frac{\Delta^2 H(m,\omega_k)}{\Delta m^2} \\ &= \frac{H(m+1,\omega_k)+H(m-1,\omega_k)-2H(m,\omega_k)}{1^2} \qquad n \in [m_A-1, M]\end{aligned} \tag{119}$$

Nun wird nach dem ersten Wert $\vartheta(m,\omega_k)^2 > 0$ gesucht. Also nach der Wendestelle m_W, so dass die Amplitude A folgendermaßen berechnet werden kann:

$$A = H(m_W,\omega_k)\cdot e^{(m_W\tau_n)} \tag{120}$$

Da in dieser Arbeit perkussive Klänge analysiert werden, sollte die globale Amplitude A_{global} im Idealfall bei $t < 100ms$ liegen. Für den Sonderfall, dass die Amplitude A_{global} nicht in den ersten hundert Millisekunden des Signals liegt, wird die Amplitude folgendermaßen bestimmt:

$$A = A_{global}. \tag{121}$$

Dies berücksichtigt auch Signale, welche nicht dem hier verwendeten Klangmodell entsprechen, also ihre Amplitude A nicht zu Beginn des Signals besitzen.

```
input  : H_abs(m,k), ω_k, τ_k
output : A_k
begin
    x_i   t_max/(N−1) (n − 1) where n ∈ [1, N] ;
    foreach ω_k do
    inx   index(max(H_abs(m, omega_k))) ;
    if makeTimeValue(inx) < 0.1s then
        dH2   Δ²H_abs(m,omega_k)/Δn² ;
        inx2   find(dH2 > 0 in dH2(inx : end));
        A_k   H_abs(inx2, omega_k) · e^(makeTimeValue(inx2)τ_k) ;
    else
        A_k   H_abs(inx, omega_k);
    end
    ;
end
```

Algorithmus 6 : Berechnung der Amplitude A

4.6.4 *Analyse und Synthese von Schwebungen*

Schwebungen entstehen durch benachbarte Freuquenzen mit einer ähnlichen Frequenz und Amplitude [47]. Schwebungen welche aufgrund der zu niedrigen spektralen Auflösung nicht erfasst werden, zeichnen sich durch eine schwankende Amplitude im zeitlichen Verlauf aus . Mit der hier vorgestellten Methode können diese Schwebungen synthetisiert werden.
Schwebungen können vom verwendeten Klangmodell erfasst werden, wenn die Schwankungen regelmäßig auftreten, also durch eine abfallende Sinus Kurve der Form $y(t) \quad e^{-\tau t}\sin(2\pi f_{beat} t)$ beschrieben werden kann. Dazu werden die lokalen Maxima von $y(t)$ der Amplitudenverläufe jeder Mode ω_k untersucht. Ist die Standardabweichung der zeitlichen Abstände der lokalen Maxima unterhalb eines gewissen Schwellenwertes, so wird die Mode als „schwebend“ erkannt. Die Frequenz dieser Schwebung kann aus den Abständen der lokalen Maxima Δt berechnet werden:

$$f_{beat} \quad \frac{1}{\Delta t}. \tag{122}$$

Nun wird die Vereinfachung getroffen, dass die beiden die Schwebung der Frequenz f_S verursachenden Frequenzen f_1 und f_2 einen gleichen Abstand Δf zur Überlagerungsschwingung f_R haben und außerdem dieselbe Amplitude A_1 und A_2 haben. Somit kann mit Hilfe der Information der Frequenz, der Überlagerungsschwingung

$$f_R \quad \frac{f_1 + f_2}{2} \tag{123}$$

und der Frequenz der Einhüllenden

$$f_H \quad \frac{f_{beat}}{2} \quad \frac{|f_1 - f_2|}{2} \tag{124}$$

die Schwebung mit Hilfe der Frequenzen f_1 und f_2 synthetisiert werden:

$$\begin{aligned} f_1 &\quad f_r - \frac{f_H}{2} \\ f_2 &\quad f_r + \frac{f_H}{2} \\ &. \end{aligned} \tag{125}$$

Da angenommen wurde, dass die Frequenzen f_1 und f_2 die gleiche Amplitude besitzen, können diese einfach aus der Amplitude der Ausgangskurve berechnet werden:

$$A_1 \quad A_2 \quad \frac{A_R}{2}. \tag{126}$$

Die Abklingkonstante τ wird aufgrund der Linearität für die synthetisierten Kurven beibehalten. Zusammenfassend kann man den Vorgang für die Resynthese von Schwebung mit gegebenen f_R, A_R und f_H mathematisch beschreiben:

$$\begin{aligned} Y_R(t) \quad & A_R \cdot \underbrace{\sin(2\pi f_R t)}_{\text{Überlagerung}} \cdot \underbrace{\cos(2\pi f_S t)}_{\text{Einhüllende}} \\ & A_R \sin(2\pi \frac{f_1 + f_2}{2} t) \cos(2\pi \frac{|f_1 - f_2|}{2} t) \\ & A_1 \sin(2\pi f_1 t) + A_2 \sin(2\pi f_2 t) \end{aligned} \tag{127}$$

Der folgende Code beschreibt vereinfacht die Detektierung und Synthese von Schwebungen.

```
input  : H_abs(m,k), ω_k
output : ω_k_beat
begin
    foreach ω_k do
    MaxVec    findMax(H_abs(:, ω_k));
    ΔMaxVec    MaxVec(2 : end) − MaxVec(1 : end − 1);
    if standardDeviation(ΔMaxVec) < threshhold then
        timePeriode    mean(MaxVec) ;
        ω_k_beat    ω_k ;
        beatFrequency    makeFrequency(ω_k) − timePeriode/2;
        Frequency(ω_k)    makeFrequency(ω_k) + timePeriode/2;
        A_k(ω_k)    A_k(ω_k)/2;
        A_k(ω_k_beat)    A_k(ω_k);
        τ(ω_k_beat)    τ(ω_k);
    else
    end
    ;
end
```

Algorithmus 7 : Analyse von Schwebungen

4.7 SIGNALNACHBEARBEITUNG

Nach der Analyse werden nochmals Moden nach dem gleichen psychoakustischen Prinzip wie in Kapitel 4.5.1 entfernt. Dies ist möglich, da die realen Amplitudenverläufe nicht den angenäherten entsprechen und so die Möglichkeit besteht, dass die Synthese Moden enthält, welche vom menschlichen Gehör nicht wahrgenommen werden.

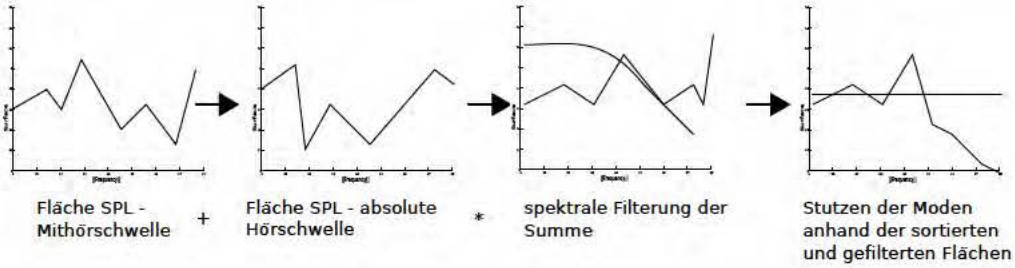


Abbildung 23: Postprocessing

Im Anschluss des Analyseprozesses hat der Nutzer die Möglichkeit die Anzahl der detektierten Moden manuell weiter zu reduzieren. Die Reduktion der Moden wird mit Hilfe der psychoakustischen Modelle durchgeführt, welche auch zum Stutzen der Moden verwendet wurden. Sowohl mit Hilfe des absoluten Hörschwellpegels $L_{T_{abs}}(f)$ als auch des vorher berechneten Mithörschwellenpegels $L_{T_m}(t,f)$. Die Bewertung des Verhältnisses eines Amplitudenpegels zum absoluten Hörschwellenpegel und zum Mithörschwellenpegel erfolgt hierbei nicht für einen Zeitpunkt t, sondern über die gesamte Länge des Signals t_{ges}. Dies bedeutet, dass zur Bewertung nicht der Abstand des Amplitudenlpegels zum Hörschwellenpegel beziehungsweise Mithörschwellenpegel als Bewertungsmaß herangezogen wird, sondern die Fläche zwischen Amplitdenpegelverlauf und des absoluten Hörschwellenpegelverlaufs beziehungsweise Mithörschwellenpgelverlaufs. Dazu wird im ersten Schritt der Mithörschwellenpgel $L_{T_m}(t,f)$ für die Schalldruckpegelverlaufs $L_s(t,f)$, berechnet. Weiter wird der Flächeninhalt zwischen dem Amplitudenpegelverlauf und dem Mithörschwellenpegels berechnet:

$$A_{AS}(f) = \int_{t=0}^{\infty} L_s(t,f) - L_{T_m}(t,f)\,dt. \tag{128}$$

Im nächsten Schritt wird der Flächeninhalt zwischen dem Amplitudenpegelverlauf und dem Pegelverlauf der absoluten Hörschwelle berechnet:

$$A_{AL}(f) = \int_{t=0}^{\infty} L_s(t,f) - L_{T_{abs}}(f)\,dt. \tag{129}$$

Diese beiden Flächen werden im Anschluss zu einer Gesamtfläche addiert:

$$A_{sum}(f) = A_{AL}(f)(1-x) + A_{AS}(f)x; \quad x \in [0;1]. \tag{130}$$

Wobei der Parameter x durch den Nutzer frei einstellbar ist.
Somit hat man für jede Frequenz einen Einzahlwert, welcher einen Zusammenhang zwischen den drei interessierenden Größen $L_{T_{abs}}(f)$,$L_{T_m}(t,f)$ und $L_s(t,f)$

herstellt.
Der Anwender hat die Möglichkeit eine maximale Anzahl von Moden N_{max} zu wählen. Die anschließende Entfernung der Moden kann nun über eine Größensortierung von $A_{sum}(f)$ erfolgen. Je Größer der Wert A_{sum} ist, desto wichtiger ist die Betreffende Mode der Frequenz f für den Klang. Dies Bedeutet, das alle Moden entfernt werden, welche nach der Größensortierung einen Index $i > N_{max}$ haben. Um dem Nutzer die Möglichkeit zu geben, diese Sortierung frequenzabhängig zu bewerten, hat dieser die Möglichkeit mit Hilfe einer grafisch modifizierbaren Filterkurve, einzelne Frequenzen zu betonen oder weniger abzuschwächen, so dass die Fläche $A_{sum}(f)$ noch mit einem filterabhängigen Paramter $Q_f \in [0.01; 1]$ multipliziert wird, um so eine filterabhängige Fläche $A_{sum}^{Q_f}(f) \quad A_{sum}(f) \cdot Q_f$ zu ergeben. Der Filter besteht aus einem Hochpass- einem Tiefpass- und zwei Peakfiltern. Der Filter ermöglicht es tiefe und hohe Frequenzen, sowie mittlere Frequenzen zu betonen oder abzuschwächen.

4.8 ANALYSE UND SYNTHESE NICHTLINEARER MODELLE

Um den nichtlinearen Schwingvorgang zu analysieren, wird der Frequenzverlauf der Mode mit dem global höchsten Peak analysiert. Dazu wird in einem ersten Schritt der größte Wert des Spektrogramms $max(H(k, m))$ ermittelt. Im nächsten Zeitschritt m wird in den fünf benachbarten Werten nach dem größten Wert gesucht. Dieser Schritt wird so lange fortgeführt, bis das Ende der Zeitachse erreicht ist. Mit Hilfe dieser Methode kann der Frequenzverlauf $f(m)$ einer Mode ermittelt werden. Für den Bereich $f(m) < O/2$, also in dem Bereich in dem die Werte der STFT vom restlichen Auswertungsbereich abweichen, können keine Werte ermittelt werden. Für den Bereich $f(m) > O/2$ müssen nun die Amplituden $s(n)$ des Signals mit dem Verlauf der Modalfrequenz $f(m)$ in Verbindung gebracht werden.
In Abbildung 24 ist der Signalverlauf und der detektierte Frequenzverlauf der energiereichsten Mode aufgetragen. Die detektierte Frequenz bei $max(H(k, m))$ wurde dabei auf den Wert Eins gesetzt. Die Frequenzveränderung wurde normiert.

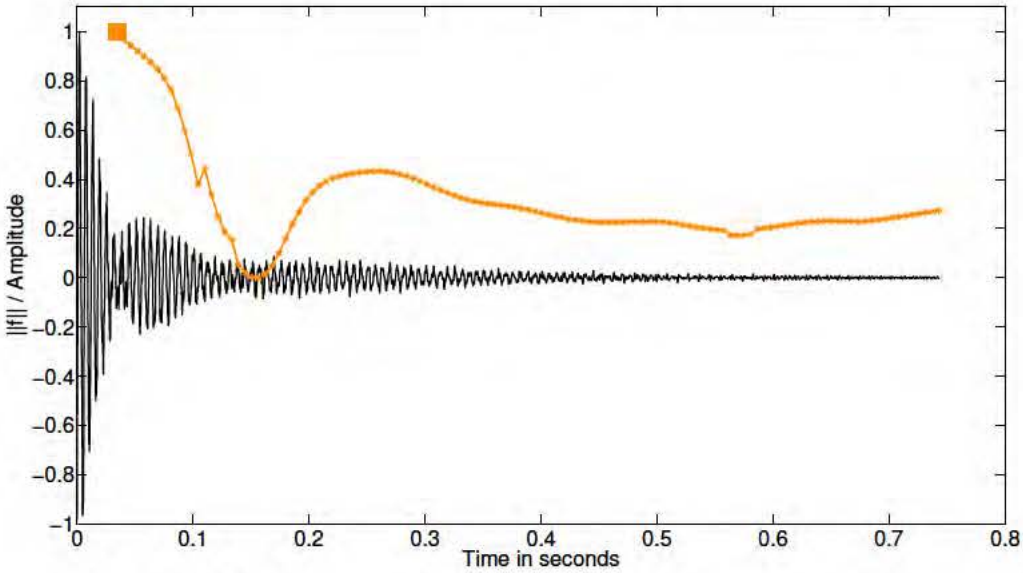
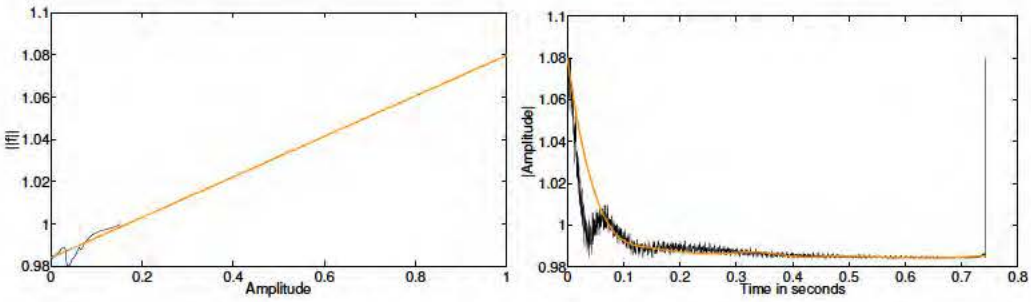


Abbildung 24: Signalverlauf und normierter Frequenzverlauf
■ : detektieres maximum, ★ : Frequenzverlauf , — : Signal

Nun werden die erhaltenen Frequenzen zum Zeitpunkt t (Abb. 24 ★) mit dem jeweiligen mittleren Signalwert $s_{mean}(t_n)$ zum Zeitpunkt t verknüpft. Der mittlere Signalwert bedeutet hier der gewichtete Mittelwert. Die Gewichtung erfolgt aufgrund der Fensterung der Kurzzeit Foruier Transformation. Die gewichtete Mittelung wird folgendermaßen berechnet:

$$s_{mean}(t_n) = \frac{\sum_{n=1}^{N} \sqrt{(s(n+t_n-N/2))^2 w(n)}}{\sum_{n=1}^{N} w(n)} \tag{131}$$

Im nächsten Schritt werden die ermittelten mittleren Signalamplituden s_{mean} gegen die dazugehörigen normierten Frequenzen aufgetragen und eine lineare Regression durchgeführt (siehe Abbildung 25 a)).



a) Zuordnung Amplitude -> Frequenz b) Berechnung der geglätteten Amplitude

Abbildung 25: Detektierung der amplitudenabhängigen Modenfrequenz

Die erhaltene Abhängigkeit der Frequenz von der Amplitude kann auf den geglätteten Amplitudenverlauf (siehe Abbildung 25 b)) des Originalsignals übertragen werden, und somit der Frequenzverlauf angenähert werden.

4.9 BESTIMMUNG DES FREQUENZABHÄNGIGEN DÄMPFUNGSFAKTORS

Die Bestimmung des Dämpfungsfaktors $d = 1/\tau$ in Abhängigkeit von der Frequenz wird mit Hilfe einer linearen Regression durchgeführt. Dabei werden nur solche ermittelten Dämpfungsfaktoren berücksichtigt, welche vertrauenswürdig ermittelt wurden. Als vertrauenswürdig gelten dabei Dämpfungsfaktoren, welche folgendes Kriterium erfüllen: Die bei der Bestimmung der Abklingzeit durchgeführte lineare Regression muss mindestens über die Hälfte des Analysezeitraums durchgeführt worden sein. Die lineare Regression wurde, um Ausreißer nicht zu stark in das Modell einfließen zu lassen mit der L1-Reggression durchgeführt [48]. Der verwendete Code kann unter [1] lizenzfrei heruntergeladen werden.

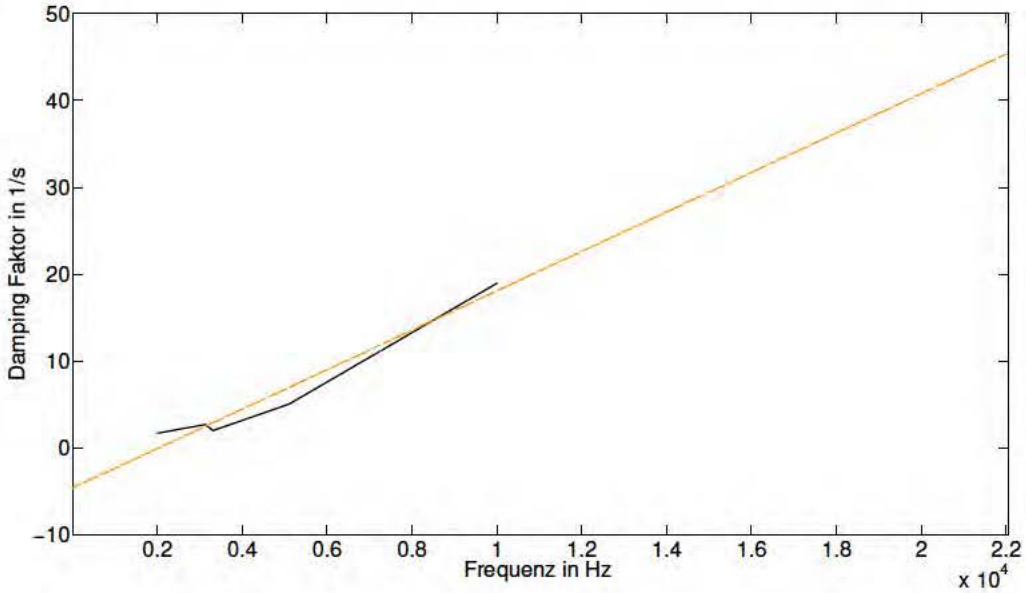


Abbildung 26: Frequenzabhängiger Dämpfungsfaktor
—— : angenäherter Verlauf; — : gemessener Verlauf

Dieses Modell entspricht nicht dem in Kapitel 2.1.6 vorgestellten Modell, da die notwendigen Informationen zur Erstellung komplexerer Dämpfungsmodelle fehlen. In Abbildung 26 ist gut zu erkennen, dass die erste detektierte Mode bei $f \approx 2$kHz liegt. Es liegen also keine Informationen über die Dämpfung für Frequenzen $f < 2$kHz vor. Dieses lineare Dämpfungsmodell stellt somit eine starke Vereinfachung da.

4.10 ERSTELLEN VON ANALYTISCHEN MODELLEN

Die Erstellung von eigenen Modellen erfolgt auf Grundlage von analytischen Modellen zur Berechnung von Resonanzfrequenzen. Die modalen Modelle wurden dabei hauptsächlich aus den Büchern Vibration of Plates and Shells [49]und The Physics of Musical Instruments [29] entnommen. Die verwendeten Resonanzmodelle sind in Kapitel A.3 aufgelistet. Da keine Information über die Amplituden vorliegen, wurde dieser Parameter frei wählbar gestaltet. Der Nutzer kann diese Parameter also frequenzabhängig einstellen. Die Abklingkon-

1 http://www.di.ens.fr/~mschmidt/Software/L1General.html

stanten wurden mit Hilfe des Dämpfungsmodells aus Kapitel 2.1.6 gestaltet. Weiter wurden einige der ermittelten frequenzabhängigen Dämpfungsfaktoren integriert. So wurde beispielhaft eine rechteckige Platte simuliert und mit den ermittelten Dämpfungsfaktoren einer Clave, eines Weinglases und einer Metallplatte kombiniert. Der Dämpfungsfaktor kann ebenfalls vom Nutzer modifiziert werden.

Teil V

ERGEBNISSE UND AUSWERTUNG

5 ERGEBNISSE UND AUSWERTUNG

5.1 SPEKTRALE ANALYSE

Die Aufgabe bei der spektralen Analyse bestand im Wesentlichen aus drei Punkten: Die Wahl eines geeigneten Fensters w für die Kurzzeitfouriertransformation, die Wahl der Länge des Fensters W und die Größe des Overlappings O. Die Hauptproblematik jeder Kurzzeit-Fourier-Transformation liegt in der Wahl der Fensterlänge. Die Fensterlänge definiert das Verhältnis aus zeitlicher und spektraler Auflösung ζ_f. Bei einer Fensterlänge $W = 8192$ und einer Abtastfrequenz von $f_S = 44.1\text{kHz}$ liegt die spektrale Auflösung im Bereich von $f = 0 - 22.05\text{kHz}$ bei $\xi = \frac{W}{2} + 1 = 4096$. Es können also maximal $N_{modes} = 2048$ Moden gefunden werden.
Bei einer Signallänge von $N = 32768$ können ohne Overlapping mit einer Fensterlänge von $W = 8192$ maximal vier Zeitschritte berechnet werden. Die zeitliche Auflösung liegt also bei $\zeta_t = 4$. Mit Hilfe überlappender Fenster konnte die zeitliche Auflösung erhöht werden. Eine Überlappung von 31/32 hat hierbei die besten Resultate erzielt. Die Rechenzeit bei diesem Wert ist vertretbar und die zeitliche Auflösung ist für die Bestimmung der Abklingzeit und der Amplitude hoch genug. Zusätzliche wurde noch die Auflösung im Spektralbereich mit Hilfe des Zero-Paddings erhöht. Zudem konnten mit Hilfe der quadratischen Peak Interpolation genauere Frequenzwerte für die detektierten Moden berechnet werden.
Als Fensterfunkion eignet sich laut Literatur das *Blackman-Harris* Fenster am besten [45], da es benachbarte Frequenzen mit unterschiedlichen Amplituden optimal detektiert. Für tonale Klänge trifft dies auch zu. Für Klänge mit einem rauschartigen Charakter eignete sich jedoch das *Hann* Fenster besser, da mehr Moden erkannt werden.
Bei sehr vielen Moden ($N_{modes} > 500$) sollte die Fensterlänge W groß genug sein, um die Moden identifizieren zu können. Die geeignete Fensterlänge liegt hier bei $W = 8192$. Für weniger Moden ($N_{modes} < 500$) genügt eine Fensterlänge von $W = 2048$.

Ein weiteres Problem war die Analyse des Datei- beziehungsweise Signalanfangs. Mit Hilfe der gewöhnlichen Kurzzeit-Fourier-Transformation über mehrere Zeitschritte konnte der Anfang des Signals nicht hinreichend genau aufgelöst werden. So befindet sich der Peak eines gefensterten Signalabschnitts mit einer Länge von $W = 8192$ und bei einer Abtastrate von $f_S = 44.1\text{kHz}$ bei $t = 0.1857\text{s}$. Im Zeitraum von $t < 0.1857\text{s}$ ist das Signal also nur unzureichend analysierbar. Daraus entsteht ein sogenannter Delay [50]. Daher wurden verschieden Methoden getestet, um diesen Signalabschnitt hinreichend analysieren zu können.
Eine Möglichkeit, den Delay zu verringern, liegt in der Anwendung variabler Fensterbreiten. Zu Beginn des Signals werden kleiner Fenster genutzt, beginnend mit W_{start}, zum Beispiel $W_{start} = W/2$ und dann größer werdenden

Fenstern bis zum erreichen der gewählten Fenstergröße W. Um eine konsistente Auflösung im Spektralbereich zu erhalten, werden dazu die restlichen Samples mit Nullen aufgefüllt. Der Test dieses Verfahrens zeigte jedoch, dass auf diese Weise keine ausreichende Analyse des Anfangszeitraums erreicht werden konnte.
Eine weiter Möglichkeit bestand in der Wahl asymmetrischer Fenster [51] [52] [53] [54]. Mit diesem Verfahren konnte der Delay zwar reduziert werden, aber nicht in zufriedenstellendem Ausmaß. Im Klang konnten immer noch große Unterschiede festgestellt werden. Des Weiteren waren die Fenstereigenschaften bezüglich der Seitenkeulen schlechter als bei symmetrischen Fenstern.
Daher wurde das in Kapitel 4.4 vorgestellte Verfahren verwendet. Dies führte dazu, dass die Fenstereigenschaften eines Rechteckfensters, bezüglich der Transformation in den Spektralbereich, für den Bereich $t < 0.1857s$ vorlagen, dennoch konnte der Signalanfang hinreichend analysiert werden. Aufgrund der in dieser Arbeit verwendeten Verfahren zur Bestimmung der Abklingkonstante τ und der Amplitude A spielten die fehlerhaften Amplituden im Anfangsbereich keine Rolle.

5.2 MODENDETEKTIERUNG

Die Detektierung der Moden mit Hilfe des EDR ist sehr effizient und schnell, wenn die Eigenfrequenzen konstant sind. Der verwendete Algorithmus detektiert teilweise zu viele Moden insbesondere bei rauscharmen Signalen aufgrund der Nebenkeulen, welche fälschlicherweise als Moden detektiert werden.
Außerdem werden auch Signalbestandteile als Moden detektiert, welche zwar eine Energie besitzen, die sich vom Rauschen abhebt, aber im eigentlichen Sinne keine Moden sind, sondern von der Anregung stammen. Für den Klang haben sie insbesondere zum Zeitpunkt der Anregung aber eine wichtige Rolle. Wird der Klang einer Glocke nur durch die Resonanzfrequenzen, Abklingzeiten und Amplituden beschrieben, unterscheidet er sich deutlich von dem Klang einer Glocke, die durch einen Klöppel angeregt wird [29]. Die Anregung selber spielt für den Klang also eine wesentliche Rolle. Somit wurden diese Bestandteile in dem Modell belassen. Ren [20] beschrieb diese Anteile als Residuen. Diese Bestandteile klingen meist sehr schnell ab.
Für nichtlineare Schwinger eignet sich die angewandte Detektierung nicht. Da ein nicht konstanter Frequenzverlauf $f(t)$ dazu führt, dass sich die Energie einer Mode nicht bei einer konstanten Frequenz befindet, sondern auf einen von der Schwankungsstärke von $f(t)$ abhängigen Bereich verteilt. Abbildung 27 zeigt den Frequenzverlauf einer Mode. Das analysierte Signal war ein reiner Sinus, dessen Frequenz exponentiell mit der Zeit abfällt.

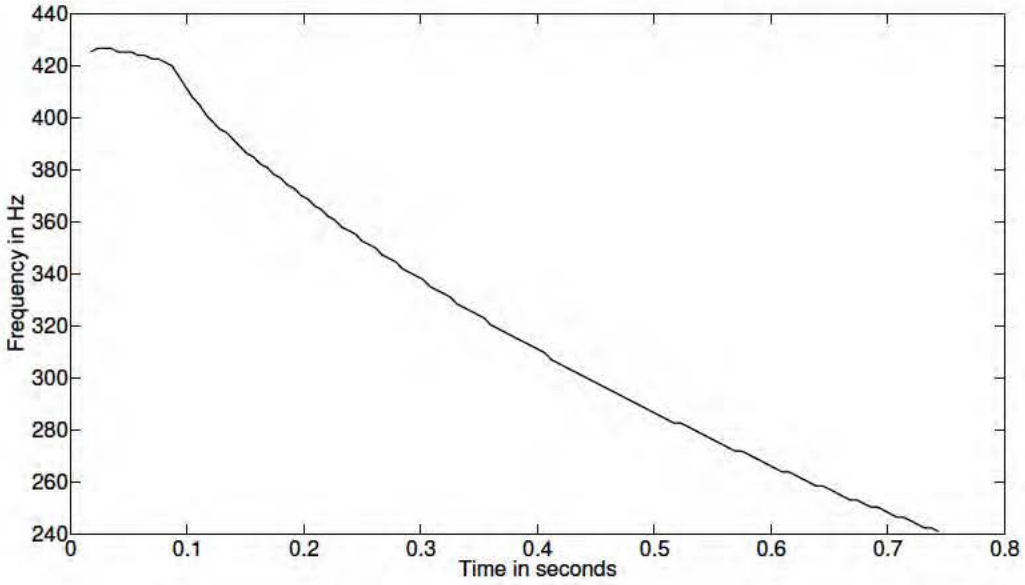


Abbildung 27: Frequenzverlauf eine Sinussignals mit zeitabhängiger Frequenz $f(t)$

Abbildung 28 zeigt die Amplituden eines EDR zum Zeitschritt $t_n = 1$ eines Signals, welches additiv aus einem Sinus mit konstanter Frequenz $f = 150\text{Hz}$ und einem Sinus mit exponentiell fallender Frequenz $f(t) = 440\text{Hz} * e^{-t}$ zusammengesetzt wurde. Zusätzlich wurden die detektierten Moden markiert. Im Vergleich ist der Unterschied der beiden Peaks gut zu erkennen. Bei konstanter Frequenz zeichnet sich das EDR durch schnell abfallende Flanken links und rechts des Peaks aus. Bei absteigender Frequenz fällt die Flanke rechts des Peaks steil ab. Links des Peaks jedoch fällt die Flanke nur aufgrund der fallenden Amplitude im Zeitbereich ab. Der Peak ist also sehr breit, wodurch eine eindeutige Bestimmung der Frequenz f, beziehungsweise des Frequenzverlaufs $f(t)$ mit dieser Methode nicht möglich ist.

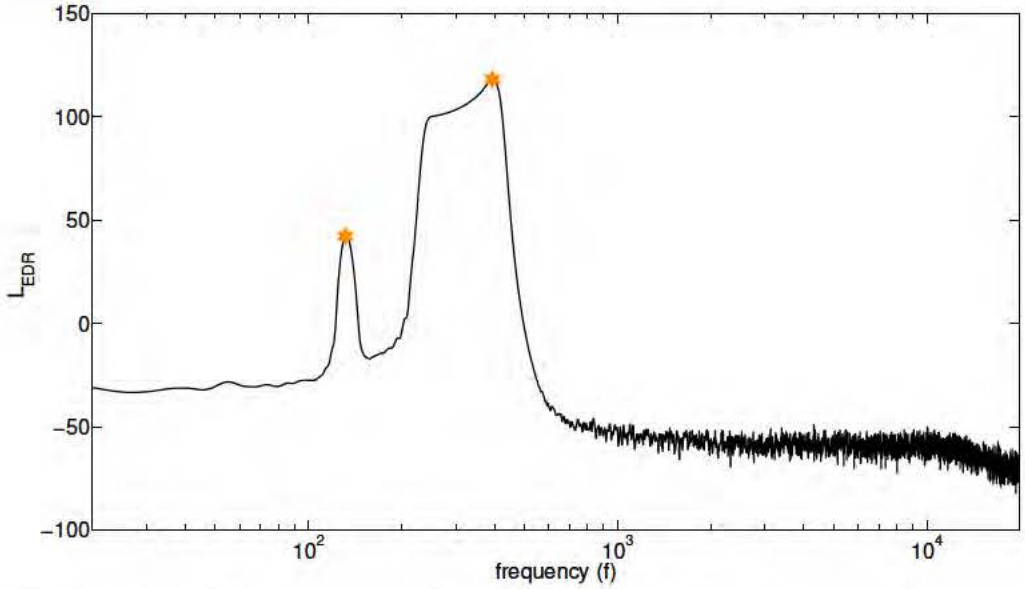


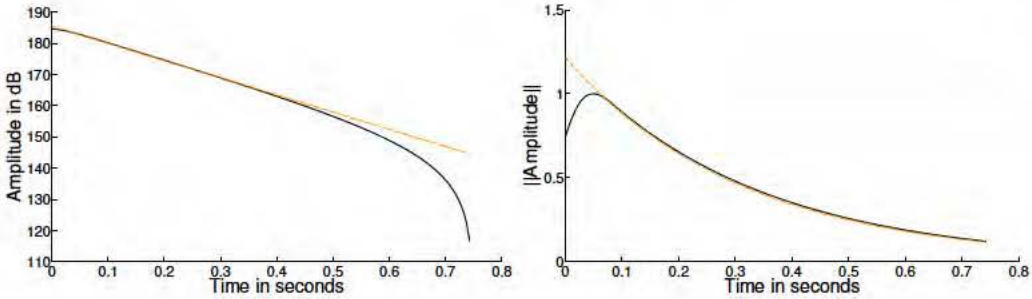
Abbildung 28: Modendetektierung: konstanter und nicht konstanter Frequenzverlauf
★ detektierte Moden, — EDR

5.3 STUTZEN NICHT HÖRBARER MODEN

Die Notwendigkeit der Entfernung nicht hörbarer spektraler Anteile liegt in der Reduzierung des Rechenaufwands bei der Synthese des Klanges mit *Synth-A-Modeler*. Die spektrale Maskierung eignet sich besonders für die Entfernung falsch erkannter Moden bei tonalen Klängen, da hier aufgrund der Seitenkeulen zu viele Moden erkannt werden.
Bei Signalen mit einem hohen Rauschanteil insbesondere im hohen Frequenzspektrum werden teilweise zu viele Moden gestutzt, so dass hohe Frequenzanteile in der späteren Resynthese fehlen.
Einer der Gründe für das Stutzen zu vieler Moden ist, dass aus Gründen der Rechenzeit nicht jeder Zeitschritt bezüglich der Wahrnehmung spektraler Anteile untersucht wurde, sondern nur jeder dritte. Somit wird nur ein drittel aller Zeitschritte untersucht. Insbesondere bei stark schwankenden Amplituden kann dies dazu führen, dass zu viele Moden gestutzt werden.
Ein weiter Grund könnte in der Umwandlung der Werte des Spektrogramms $H(k, m)$ in Schalldruckpegel $L_S(k, m)$ liegen. So werden diese Werte nicht wie in der Audiokodierung [41] üblich für eine Frequenzgruppe i, sondern für jede detektierte Mode ω_k berechnet.
Weiter kann es bei nichtlinearen Signalen aufgrund der durchgeführten Modendetektierung vorkommen, dass eine Mode ω_k bei einer bestimmten Frequenz f detektiert wird. Wegen der Zeitabhängigkeit der modalen Frequenz $f(t)$ bei nichtlinearen Schwingern werden die Maskierungseigenschaften falsch bewertet, da der Algorithmus die Maskierungseigenschaften nur für konstante Frequenzen f ausgehend von der detektierten Frequenz, exakt berechnet. Das führt dazu, dass die falschen Amplituden zur Ermittlung der Maskierungsschwelle genutzt werden. Diese fallen in der Regel niedriger aus, so dass die Maskierungsschwelle zu niedrig ist.

5.4 HÜLLKURVENANNÄHERUNG

Um eine Analyse der Klangdaten zu ermöglichen, werden (im Folgenden/in diesem Abschnitt/zur Auswertung) durch einen Dirac-Impuls angeregte Objekte untersucht. Außerdem wird weiter angenommen, dass deren modale Amplituden exponentiell abfallen. Durch die angewendete Kurzzeit-Fourier-Analyse mit Hilfe der Fensterung und des Overlapping werden die Kurven auch bei einem idealen exponentiellen Verlauf dahingehend verändert, dass die Zeitbereiche zu Beginn und am Ende der Analyse unterrepräsentiert sind, beziehungsweise kleinere Werte erhalten als wirklich vorhanden. Um die Analyse zu vereinfachen werden die ermittelten Hüllkurven logarithmiert, so dass eine lineare Regression durchgeführt werden kann.
In den folgenden Abbildungen sind Hüllkurven verschiedener Frequenzen und Klangdateien dargestellt. Abbildung 29 zeigt eine korrekte Erkennung einer Hüllkurve. Die durchgezogene schwarze Linie zeigt den tatsächlichen Verlauf. Die gestrichelte graue Linie zeigt die angenäherte Kurve.

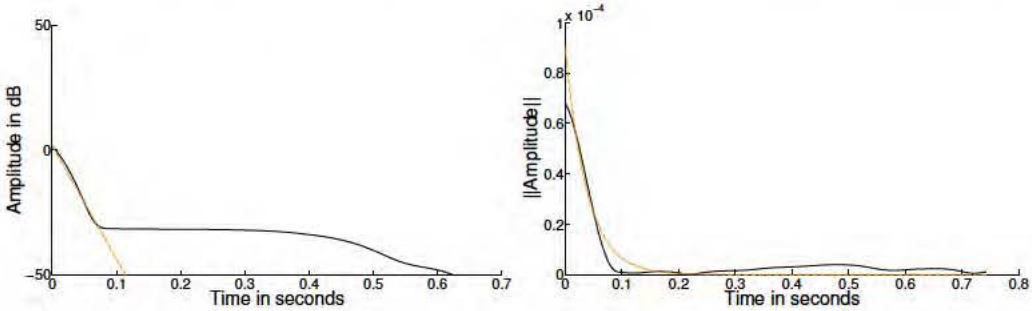


a) Logarithmische Darstellung EDR
— angenäherter Verlauf, — Verlauf

b) Lineare Darstellung H
— angenäherter Verlauf, — Verlauf

Abbildung 29: Korrekte Hüllkurvenannäherung 1

Es ist gut zu erkennen, dass die Regression im Bereich von $t = 0.05s$ bis $t = 0.2s$ sehr genau ist. In den Bereichen $t = 0 - 0.05s$ und $t = 0.4 - 0.8s$ ist die Erkennung ungenau. Dies liegt an der oben beschriebenen Fensterung und aufgrund des Auffüllens des Signals mit Nullen vor und nach den eigentlichen Signaldaten. Der Bereich der aufgefüllten Nullen wird bei der linearen Regression jedoch nicht beachtet.
Abbildung 34 zeigt eine weitere korrekte Hüllkurvenannäherung. Der spektrale Anteil besitzt Amplituden bis $t < 0.1s$, danach fällt er in den Rauschboden ab. in der linearen Darstellung ist erkennbar, dass die Amplituden nicht exponentiell abfallen. Somit ist eine exakte Annäherung mit einem exponentiellen Modell nicht möglich.

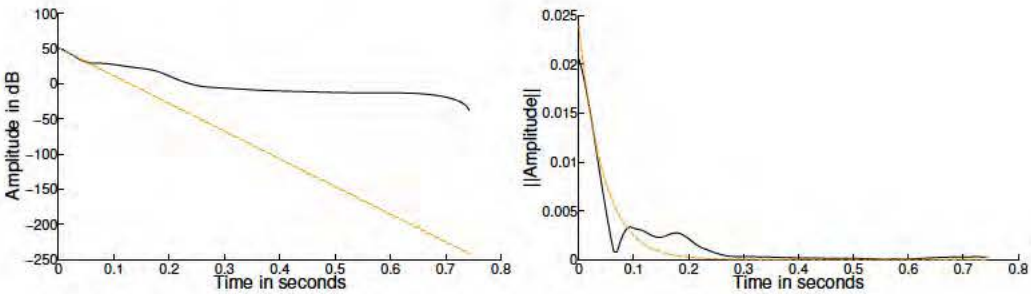


a) Logarithmische Darstellung EDR
— angenäherter Verlauf, — Verlauf

b) Lineare Darstellung H
— angenäherter Verlauf, — Verlauf

Abbildung 30: Korrekte Hüllkurvenannäherung 2

Abbildungen 31 und 32 zeigen weniger genaue Hüllkurvendetektierung. Der Algorithmus zur Hüllkurvendetektierung führt zwei lineare Regressionen zur Bestimmung der Abklingkonstante durch. Gewählt wird dann die Abklingkonstante mit dem kleineren Wert. Dies führt zwar bei nicht exponentiell abklingenden Frequenzamplituden dazu, dass die Abklingzeit als zu kurz bestimmt wird, im späteren Klang hat eine zu klein ermittelte Abklingzeit jedoch weniger Einfluss auf den Klangunterschied des synthetisierten Klangs zum Original, als eine zu groß ermittelte Abklingzeit.

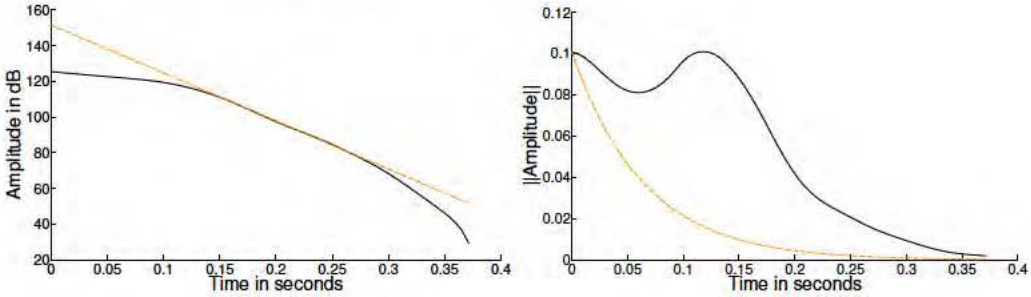


a) Logarithmische Darstellung EDR
— angenäherter Verlauf, — Verlauf

b) Lineare Darstellung H
— angenäherter Verlauf, — Verlauf

Abbildung 31: Ungenaue Hüllkurvenannäherung 1

In Abbildung 32 rechts ist ebenfalls eine falsche Ausklingzeit und Amplitude ermittelt worden. Die falsch ermittelte Abklingzeit hat den selben Grund wie im vorherigen Beispiel. Die falsch ermittelte Amplitude hängt damit zusammen, dass zur Ermittlung der Amplitude nur dann eine exponetielle Extrapolation durchgeführt wird, wenn das globale Amplitudenmaximum innerhalb der ersten hundert Millisekunden des Amplitudenverlauf liegt. Ist dies nicht der Fall, wird die globale Amplitude als Modellamplitude verwendet. Diese Methode wird gewählt, da die exponentielle Extrapolation, ausgehend von globalen Amplituden, welche nicht am Anfang des Amplitudenverlaufs liegen, sehr hohe Modellamplituden erzeugen würde.

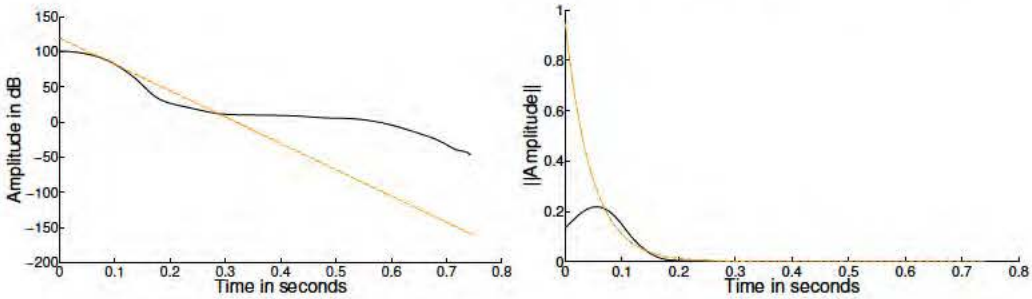


a) Logarithmische Darstellung EDR
— angenäherter Verlauf, — Verlauf

b) Lineare Darstellung H
— angenäherter Verlauf, — Verlauf

Abbildung 32: Ungenaue Hüllkurvenannäherung 2

Abbildung 33 zeigt eine weitere falsch ermittelte Hüllkurve. Der Amplitudenverlauf der logarithmierten Kurve zeigt im Bereich $t < 0.2s$ keinen linearen, sondern einen progressiv fallenden Verlauf, so dass der Algorithmus zwei sehr kurze lineare Regressionen durchführt und dann den kleineren der beiden ermittelten Steigungskoeffizienten wählt. Das führt dazu, dass die Modellamplitude zu hoch und die modellierte Abklingzeit falsch angenähert wird.

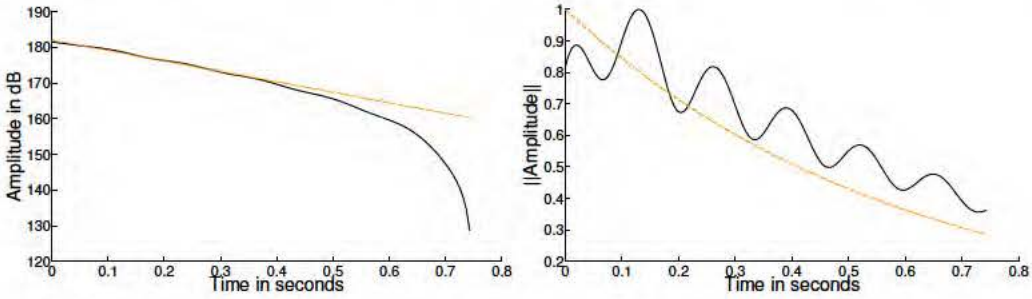


a) Logarithmische Darstellung EDR
— angenäherter Verlauf, — Verlauf

b) Lineare Darstellung H
— angenäherter Verlauf, — Verlauf

Abbildung 33: Ungenaue Hüllkurvenannäherung 3

In der letzten Abbildung 34 ist ein Beispiel für eine deutlich ausgeprägte Schwebung gezeigt. Die lineare Regression wird exakt durchgeführt, doch die Schwebungen werden durch das exponentielle Modell nicht abgebildet. Außerdem wird die exponentielle Extrapolation zur Ermittlung der Modellamplitude nicht durchgeführt, da die globale Amplitdue der Messkurve bei $t > 0.1s$ liegt, so dass der angenäherte Kurvenverlauf etwas niedriger als der gemessene ist. Wird die Schwebung von zwei spektralen Bestandteilen verursacht, welche ähnliche Frequenzen und Amplituden haben, so wird die Schwebung automatisch erfasst, wenn die spektrale Auflösung groß genug ist. Wird die Schwebung nicht automatisch erfasst, so kann dieser Schritt manuell vom Nutzer durchgeführt werden.



a) Logarithmische Darstellung EDR
— angenäherter Verlauf, — Verlauf

b) Lineare Darstellung H
— angenäherter Verlauf, — Verlauf

Abbildung 34: Ungenaue Hüllkurvenannäherung 4

5.5 RESYNTHESE VON SCHWEBUNGEN

Der Nutzer hat optional die Möglichkeit, Schwebungen, welche nicht durch die spektrale Auflösung erfasst werden, manuell zu analysieren und synthetisieren zu lassen. Abbildung 35 zeigt eine sehr genaue Detektierung der Schwebung. Der Verlauf des Originalsignals wird bis auf die ersten 1000 Samples sowohl hinsichtlich der Phase als auch des Amplitudengangs sehr genau wiedergegeben.

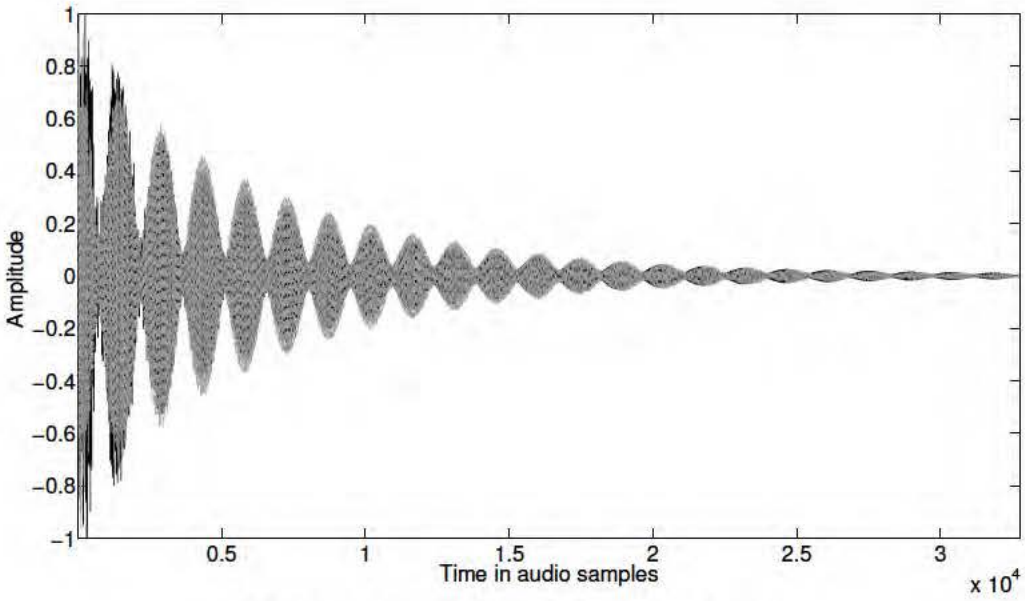


Abbildung 35: Korrekte Synthese einer Schwebung
— Synthesesignal, – Originalsignal

Durch das verwendete einfache Modell zur Synthese von Schwebungen können folgende Effekte auftreten: Zum einen kann die Phase der Schwebung nicht exakt resynthetisiert werden. Dies hängt damit zusammen, das die Phasen der einzelnen Moden nicht betrachtet werden. Den Effekt sieht man in Abbildung 36. Die angenäherte Schwebung ist phasenverschoben gegenüber der Schwebung des Originalsignals. Zum anderen wird die Amplitude nur unzureichend angenähert, so dass sie in diesem Fall, wie in Abbildungen 36 zu sehen, zu klein ist. Außerdem wird bei schwankender Schwebefrequenz f_{beat} der Mittelwert aus den detektierten Abständen Δ_{max} der Maxima zur Berechnung von f_{beat} verwendet, so dass der tatsächliche Verlauf von dem synthetisierten abweichen kann.

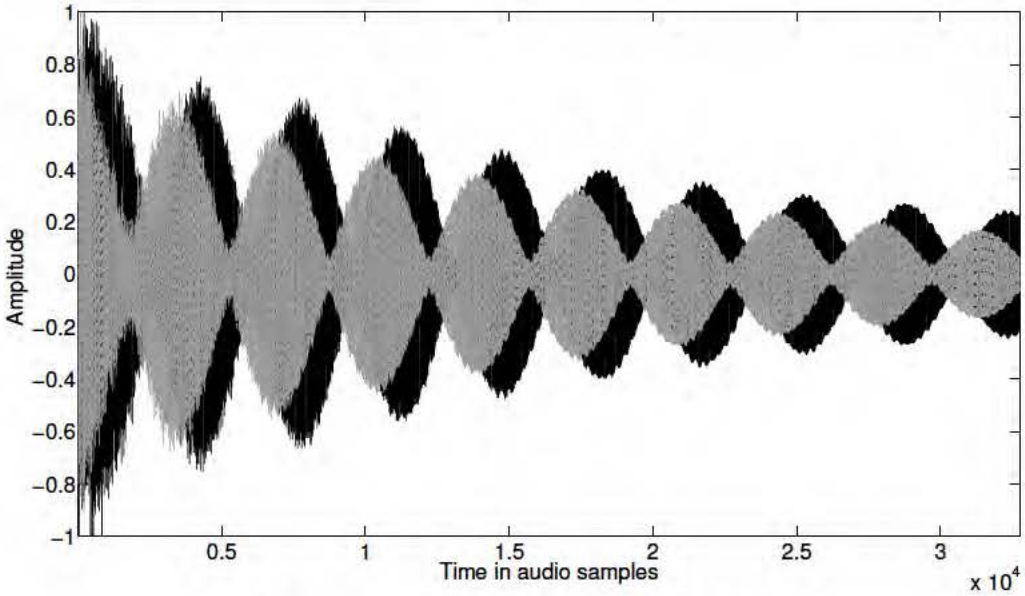


Abbildung 36: Ungenaue Synthese einer Schwebung
— Synthesesignal, – Originalsignal

Insgesamt ist der Höreindruck mit synthetisierten Schwebungen aber sehr gut. Bei Klängen mit stark ausgeprägter Schwebung bei konstanter Frequenz sollte die Analyse der Schwebung durchgeführt werden, um ein gutes Klangergebnis zu erhalten. Physikalisch gesehen wird eine weitere Mode, welche im Originalklang vorhanden war, aber durch die zu geringe spektrale Auflösung nicht wiedergegeben werden konnte, durch die Synthese der Schwebung reproduziert.

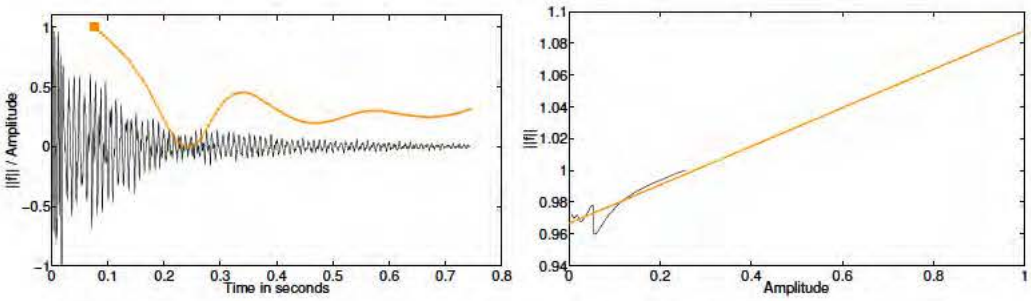
5.6 ANALYSE VON NICHTLINEAREN MODELLEN

Mit der vorgestellten Methode zur Resynthese von nichtlinearen Schwingungssignalen konnten keine zufriedenstellenden Ergebnisse erreicht werden. Es konnte kein ausreichender Zusammenhang zwischen Amplitude des Signal $s(n)$ und aktueller Modenfrequenz $f(n)$ hergestellt werden. Somit konnte die Resynthese nicht gut realisiert werden. Im Folgenden wird dies anhand einer Auswahl einiger Beispiele dargestellt. Die Ergebnisse können unter folgenden Dateinamen angehört werden:

NAME	DATEINAME
Tom	102802_mhc_acoustic – tom6
Hightom	184536_altemark_tom2
Becken	154713_mistorcaveman_cymbal_6_1_2
Kantilan	24621_anamorphosis_gmb – kantilan – 10

Tabelle 1: Nichtlineare Klangbeispiele

Abbildung a) zeigt den Signal- und den genormten Frequenzverlauf der energiereichsten Mode einer Tom. Es ist gut zu erkennen, dass die Frequenz mit der Zeit tendenziell sinkt aber auch schwankt. Die lineare Regression zwischen der Amplitude im Signal und der vorherrschenden genormten Frequenz ist in Abbildung 37 b) zu sehen. Die Regression nähert den Verlauf scheinbar gut an.



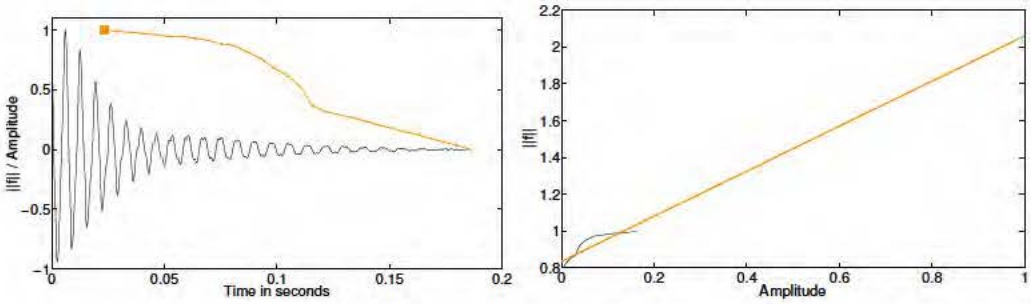
a) Signalverlauf und Frequenzverlauf
— Frequenzverlauf, – Signal

b) Amplitude und Frequenz
— Näherungskurve, – Messkurve

Abbildung 37: Zusammenhang zwischen der Modenfrequenz und der Amplitude einer Tom

Die Frequenz von der energiereichsten Mode einer Tom wiederum sinkt ebenfalls mit fallender Amplitude (siehe Abbildung 38 a)). Jedoch ist der Zusam-

menhang zwischen Amplitude und Frequenz nichtlinear, so dass die lineare Regression den Verlauf nicht gut widerspiegelt.

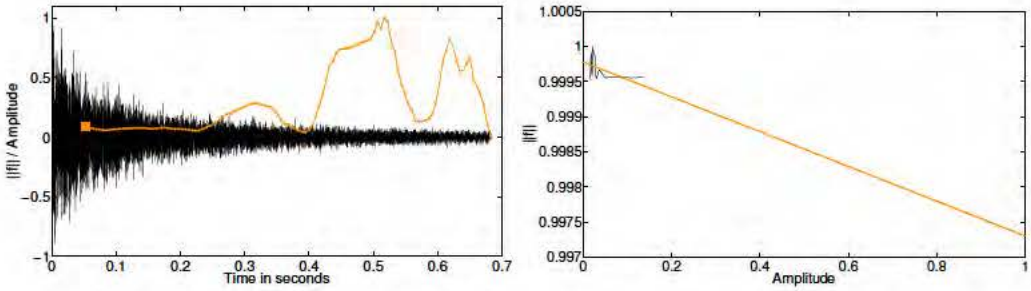


a) Signalverlauf und Frequenzverlauf
— Frequenzverlauf, — Signal

b) Amplitude und Frequenz
— Näherungskurve, — Messkurve

Abbildung 38: Zusammenhang zwischen der Modenfrequenz und der Amplitude einer Hightom

Der Verlauf der Frequenz eines Beckens scheint dagegen einen komplett anderen Verlauf anzunehmen. Abbildung 39 zeigt, dass hier die Frequenz mit fallender Amplitude zunimmt.

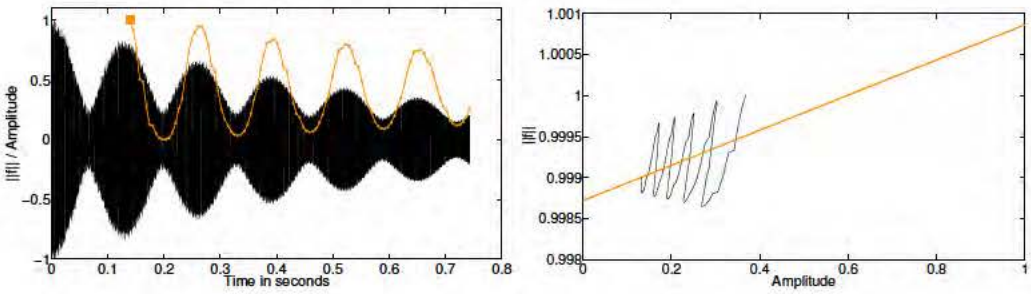


a) Signalverlauf und Frequenzverlauf
— Frequenzverlauf, — Signal

b) Amplitude und Frequenz
— Näherungskurve, — Messkurve

Abbildung 39: Zusammenhang zwischen der Modenfrequenz und der Amplitude eines Becken

Das letzte Beispiel in Abbildung 40 zeigt wiederum ein anderes Verhalten. Die Frequenz steigt mit steigender Amplitude und dies mit ähnlicher Phase und Periode. Abbildung 40 b) zeigt jedoch, dass die Höhe der Amplitude dabei scheinbar keine große Rolle spielt, so dass die lineare Regression den Verlauf nicht widerspiegeln kann.



a) Signalverlauf und Frequenzverlauf
— Frequenzverlauf, — Signal

b) Amplitude und Frequenz
— Näherungskurve, — Messkurve

Abbildung 40: Zusammenhang zwischen der Modenfrequenz und der Amplitude eines Kantilan

Die vier oben genannten Beispiele zeigen, dass eine einfache Parametrisierung der Frequenzverläufe im allgemeinen nicht möglich ist. Es wurden nur die energiereichsten Moden betrachtet. Der Verlauf energieärmerer Moden wurde nicht betrachtet. Aufgrund der Komplexität der Frequenzverläufe konnte auf diese Weise keine ausreichende Parametrisierung der Frequenzverläufe realisiert werden. In Kapitel 6.2 werden Vorschläge unterbreitet, um die Parametrisierung in Zukunft realisieren zu können.

5.7 SYNTHESE VON KLANGSIGNALEN

Im Folgenden werden die Analyseergebnisse anhand zwei exponentiell abfallender Signale erläutert. Das erste Referenzsignal ist ein reiner Sinuston, das zweite Signal ein weißes Rauschen. Anschließend werden die Ergebnisse der Resynthese ausgewählter Instrumente vorgestellt.

Die Analyse der Instrumente wurde mit einer Fensterlänge von $W = 8192$ durchgeführt. Als Fensterfunktion wurde das Hann-Fenster verwendet. Weiter wurde eine Überlappung von $O = 31/32$ gewählt. Nicht hörbarer Moden wurden bei der Analyse entfernt. Die Ergebnisse werden unter Einbeziehung der Phaseninformationen ausgewertet.
Neben der verbalen Beschreibung der Ergebnisse wird mit Hilfe des Korrelationskoeffizienten von dem Originalsignal und dem additiv zusammengesetzten Signal $\rho_{O,A}$, sowie dem Originalsignal und dem synthetisierten Signal $\rho_{O,S}$ auch ein numerischer Kennwert zur Beurteilung der Güte der Synthese herangezogen. Dabei wurde das additive Signal aus der Summe der Amplitudenverläufe der detektierten Moden berechnet. Das synthetisierte Signal wurde aus den Summen der parametrisierten Amplitudenverläufe der detektieren Moden berechnet. Als weitere numerische Kennwerte werden die Anzahl der Moden nach Stutzen N_{modes} die genormte Energie der Signaldifferenz $\Delta_s = \Delta_E/N$, mit der Signallänge in Samples N, zwischen dem Originalsignal s und dem synthetisierten Signal s_{syn} als Bewertungsmaß herangezogen. Δ_s ist also die gemittelte Energiedifferenz pro Sample. Für Δ_E gilt laut Aurich [55]:

$$\Delta_E = \|s_{syn}(n) - s(n)\|_2^2 = \sum_{n=1}^{M} |s_{syn}(n) - s(n)|^2. \tag{132}$$

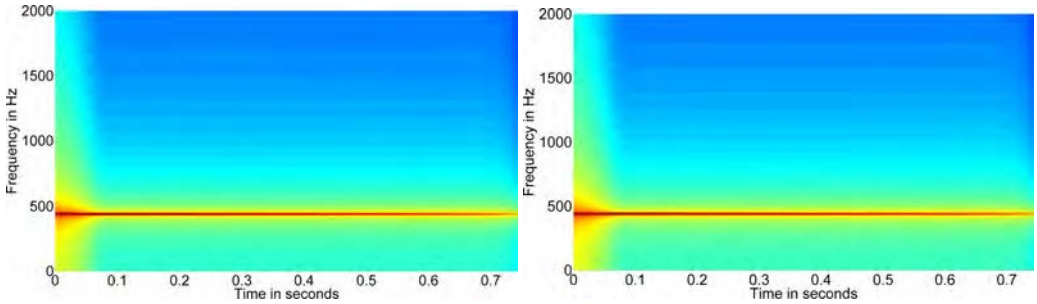
Weiter wird die Abweichung der detektierten ersten Modenfrequenz ϕ_0 und des Frequenzverlaufes f_m in Form der Standardabweichung σ_s betrachtet, welche wie folgt berechnet wird:

$$\sigma_s \quad \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\phi_0 - f_m)^2}. \tag{133}$$

Bei den numerischen Kennwerten sind nicht die Absolutwerte ausschlaggebend, sondern die Vergleiche der untersuchten Instrumente untereinander. Falls die Standardabweichung $\sigma_s \quad 0$, so handelt es sich um einen linearen Schwinger, je höher dieser Wert, desto weniger gut, kann das nichtlineare Modell durch ein linearisiertes Modell dargestellt werden.

5.7.1 *Sinussignal*

Abbildung 41 zeigt das EDR eines reinen Sinussignals mit einer Frequenz von $f \quad 440\text{Hz}$. Der rote schmale Bereich zeigt das Amplitudenmaximum im EDR, also den energiereichsten Bereich. Die blauen Bereiche bedeuten, dass die Amplituden dort sehr schwach ausgeprägt sind. Der grüne Bereich zeigt Amplituden, hervorgerufen durch Nebenkeulen, welche durch die Fensterung entstanden sind. Die detektierte modale Frequenz liegt bei $f \quad 40,0848$. Also beträgt die relative Abweichung zum Originalsignal $\Delta f/f \approx 0.2\%$.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 41: EDR Sinus

Der Vergleich der Hüllkurven in Abbildung 42 zeigt, dass die beiden Kurven sehr gut übereinstimmen.

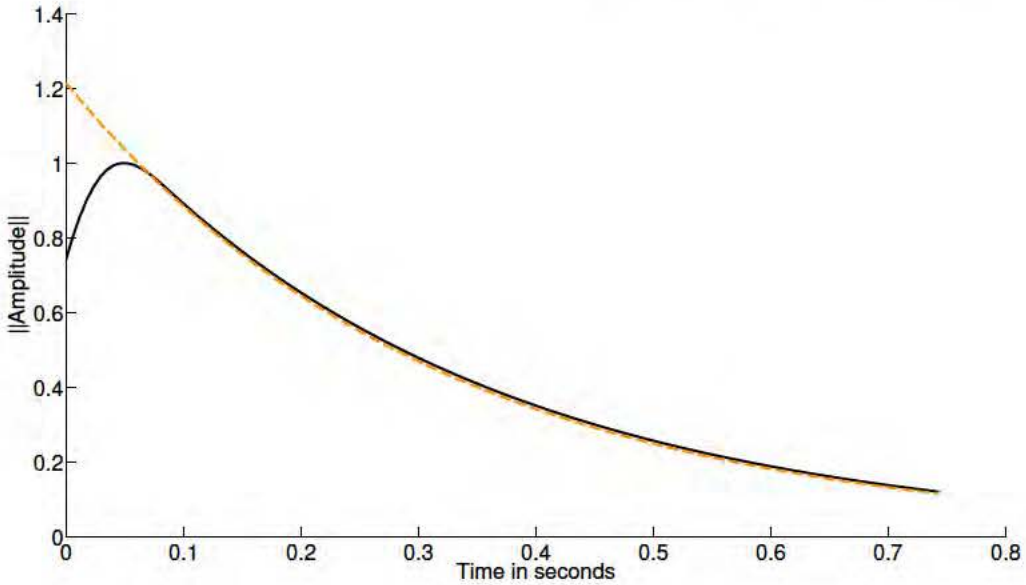
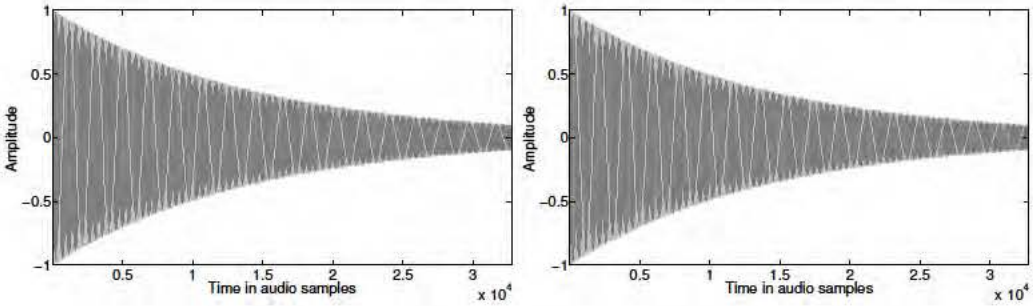


Abbildung 42: Hüllkurvenannäherung bei reinem Sinus
— geschätzte Kurve, – realer Verlauf

Der Vergleich des Originalsignals mit dem synthetisierten Signal zeigt ebenfalls eine sehr hohe Übereinstimmung. Die beiden Signale sind phasengleich und auch die Amplituden stimmen überein.



a) Originalsignal b) Synthesesignal

Abbildung 43: Signalvergleich Sinus

Die vorherige gute Übereinstimmung spiegelt sich auch im Korrelationsfaktor wider. Die parametriesierte Synthese erreicht mit einem Korrelationsfaktor von $\rho_{O,S} = 0,9956$ sogar einen leicht höheren Wert als das additive Synthese mit einem Wert von $\rho_{O,A} = 0,9921$. Dies liegt an der exponentiellen Regression, mit der die Amplitude A bestimmt wird. In Abbildung 42 kann man gut sehen, dass der Kurvenverlauf der STFT am Anfang des Signals ansteigt und erst bei einer Zeit von $t \approx 0.1s$ einen exponentiell fallenden Verlauf annimmt. Nach der Modenentfernung beträgt die Anzahl der Moden $N_{Modes} = 1$. Die Differenz der Energien ist mit $\Delta_s = 0.0030$ ebenfalls sehr gering und der Wert $\sigma_s = 0$ zeigt, dass es sich um einen linearen Signalverlauf handelt.

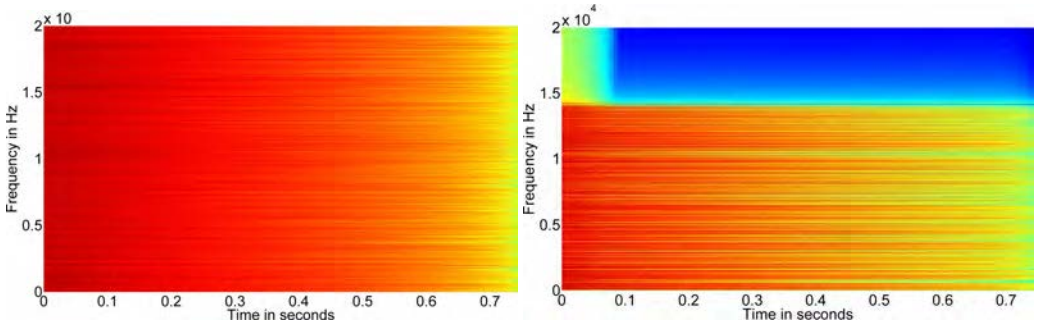
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
pure_sine_440	0,9921	0,9956
N_{modes}	Δ_s	σ_s
1	0.0030	0

Tabelle 2: Kennwerte Sinussignal

Die Analyse und Resynthese eines exponentiell abfallenden Sinussignal zeigt, dass Signale, welche dem behandelten Klangmodell entsprechen, sehr gut parametrisiert werden.

5.7.2 *Weißes Rauschen*

Das EDR des eines Signals mit weißem Rauschen in Abbildung 44 zeigt wie erwartet einen gleich verteilten Amplitudengang im gesamten Frequenzbereich. Das EDR des synthetisierten Signals zeigt deutlich den Einfluss der Modenstutzung im Bereich von $f > 15$kHz. Die detektierten Moden in diesem Bereich wurden entfernt.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 44: EDR weißes Rauschen

Die Amplitude einer jeden Frequenz des weißen Rauschens ist per Definition zufällig. Abbildung 45 bestätigt Dies. Die Annäherung an den Kurvenverlauf ist dementsprechend schlecht.

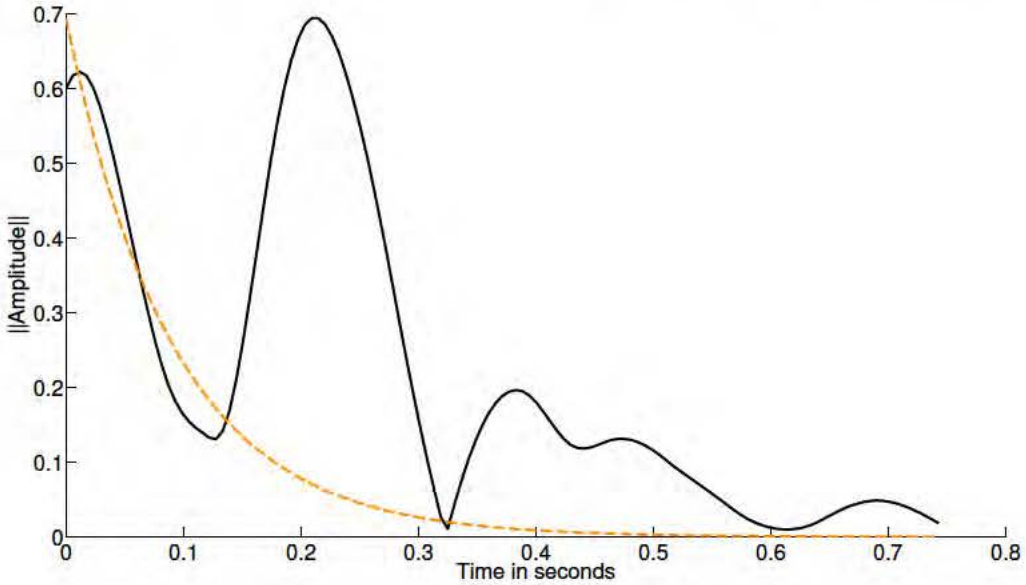
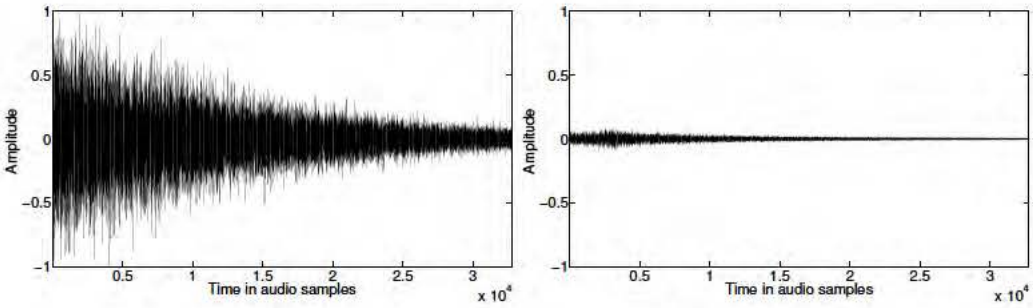


Abbildung 45: Hüllkurvenannäherung bei Tom
— geschätzte Kurve, – realer Verlauf

Der Vergleich der Resynthese mit dem Originalsignal zeigt, dass die Resynthese wesentlich kleinere Amplituden aufweist (siehe Abbildung 46). Dies hängt damit zusammen, dass die Amplitude im ersten Zeitschritt extrem groß ist. Die Normierung führt dazu, dass die Amplituden für die restlichen Werte sehr klein sind.



a) Originalsignal b) Synthesesignal

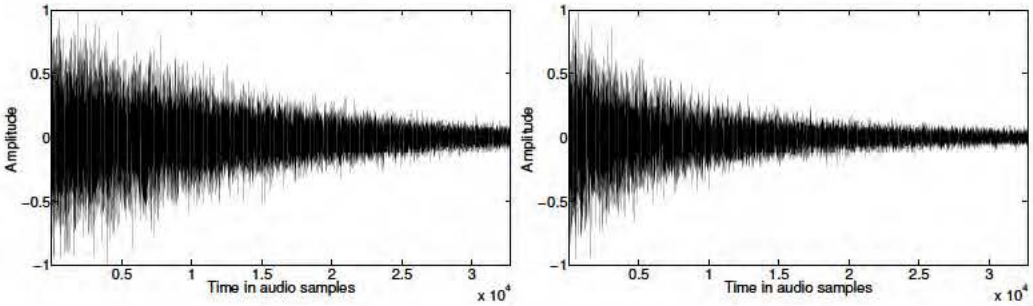
Abbildung 46: Signalvergleich weißes Rauschen

Die vorherigen Ergebnisse geben auch die Kennwerte wieder.

AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
white_noise2	0,0001	−0,0145
N_{modes}	Δ_s	σ_s
923	0.0272	9.4285

Tabelle 3: Kennwerte weißes Rauschen

Unter Einbeziehung der Phaseninformationen (siehe Abbildung 47) kann ein wesentlich besseres Ergebnis im Hinblick auf die Amplituden erreicht werden.



a) Originalsignal b) Synthesesignal

Abbildung 47: Signalvergleich weißes Rauschen mit Phasenbezug

Hervorgerufen wird der Effekt des großen ersten Wertes durch die Summierung von vielen Sinuskurven unterschiedlicher Frequenz. Sind dabei alle Frequenzanteile gleichphasig, beziehungsweise ohne Phase, so werden alle Amplituden im ersten Zeitschritt aufsummiert, da alle Frequenzanteile im Zeitschritt $k = 1$ einen positiven Wert aufweisen. Dies ist durch das *Nyquist-Shannon*-Abtasttheorem beweisbar.
Mathematisch kann das Problem folgendermaßen erklärt werden: Das synthetisierte diskrete Signal $s(k)$ setzt sich folgendermaßen zusammen:

$$s(k) = \sum_{n=1}^{N} A_n e^{-k/(\tau_n \cdot F_S)} \sin(2\pi f_n k/F_s). \tag{134}$$

Der problematische große Wert von $s(k)$ liegt immer bei $k = 1$. Für Frequenzen $f_n < F_s/2$ gilt für den Sinusterm in Gleichung 134:

$$\sin(2\pi f_n/F_s) = \sin(\pi\kappa) = \zeta_n > 0; \quad 0 \leqslant \kappa = 2f_n/F_s \leqslant 1. \tag{135}$$

Für $k = 1$ und für ein nicht allzu kleines τ_n ist:

$$e^{-1/(\tau_n \cdot F_S)} \approx 1. \tag{136}$$

Somit kann man für ein Signal mit großem $N \to \infty$ folgendes schreiben:

$$s(1) = \lim_{N\to\infty} \sum_{n=1}^{N} A_n \zeta_n \to \infty. \tag{137}$$

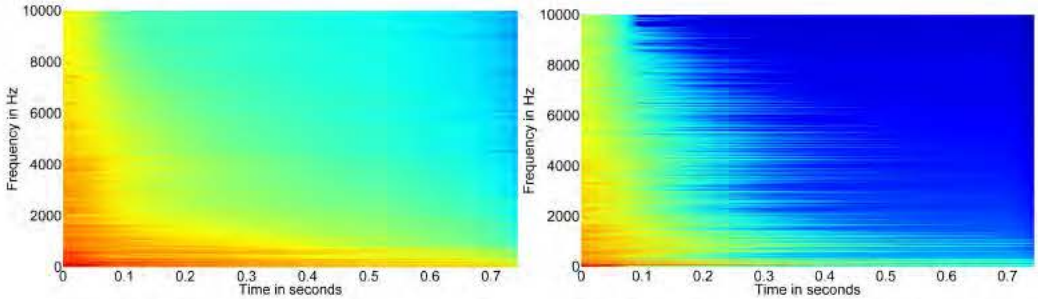
Aber auch bei kleinem N kann es zu diesem Effekt kommen. Existieren viele Moden bei einer Frequenz $f_n \approx F_s/4$, so gilt für den Sinusterm in Gleichung (134):

$$\sin(2\pi f_n/F_s) = \sin(\kappa\pi/2) = \zeta_n \approx 1 \quad ; \quad \kappa = 4f_n/F_s \approx 1. \tag{138}$$

Somit addieren sich alle Amplituden in Gleichung (137) auf. Es kann also zu sehr großen Werten zu Anfang des Signals kommen, wenn viele Moden betrachtet werden oder wenn die Frequenzen der Moden im Bereich von $f \approx F_s/4$ liegen. Mit Hilfe der Phaseninformation kann dieses Problem gelöst werden, da der Sinusterm in Gleichung (134) einen Wert $\zeta_n \in [-1; 1]$ annehmen kann.

5.7.3 *Tom*

Das EDR einer Tom 48 zeigt, dass die energiereichen Moden im niederfrequenten Bereich liegen, insgesamt hat die Tom viele Moden im gesamten Frequenzbereich. Die Energie fällt aber mit steigender Frequenz schnell.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 48: EDR Tom

Die Abbildung 49 zeigt den Amplitudenverlauf der energiereichsten Mode mit einer Frequenz von $f = 60\text{Hz}$. Die angenäherte und die original Kurve decken sich gut, obwohl der Verlauf der gemessenen Kurve keinen glatten Verlauf hat, sondern zeitlich schwankt.

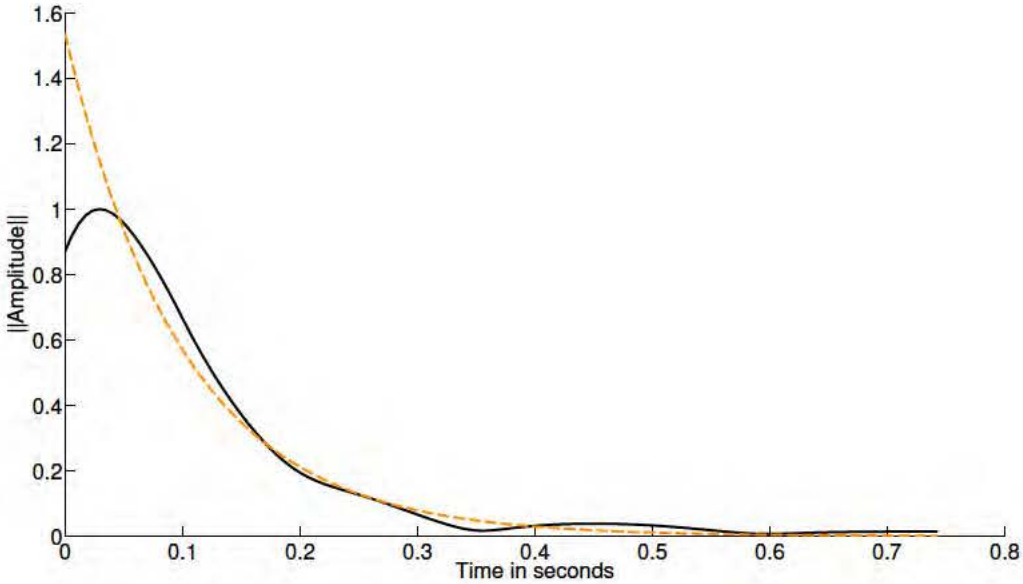
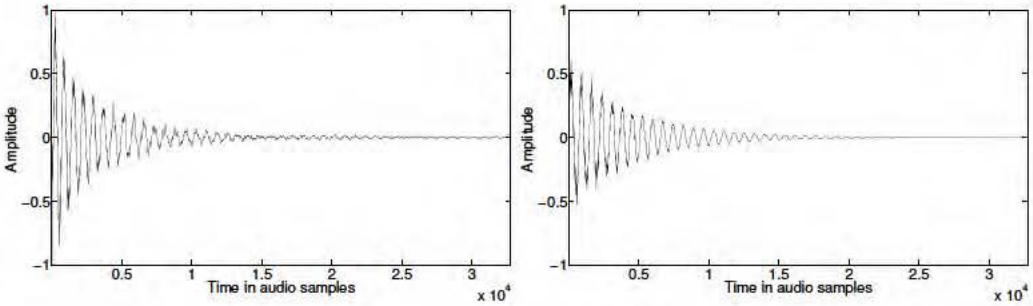


Abbildung 49: Hüllkurvenannäherung bei Tom
— geschätzte Kurve, – realer Verlauf

Die zeitlichen Signale des Originalsignals und des Synthesesignals sind in Abbildung 50 zu sehen. Die Signale weisen ebenfalls große Ähnlichkeiten auf. Lediglich die Amplituden der beiden Signale unterscheiden sich im Bereich der ersten 10^3 Samples.



a) Originalsignal b) Synthesesignal

Abbildung 50: Signalvergleich Tom
—— geschätzte Kurve, — realer Verlauf

Die relativ gute Übereinstimmung wird auch durch die Korrelation (Tabelle 4) wiedergegeben. Die Korrelation zwischen dem additiven Signal und dem Originalsignal erreicht einem ähnlich hohen Wert wie die Korrelation zwischen dem synthetisierten Signal und dem Originalsignal. Insgesamt wurden 501 hörbare Moden detektiert. Die Energiedifferenz der beiden Signal ist mit $\Delta_s = 0.0125$ ebenfalls sehr gering. Die große Standardabweichung $\sigma_s = 4.7267$ zeigt aber, dass es sich bei dem Klangbeispiel um einen nichtlinearen Schwinger handelt.

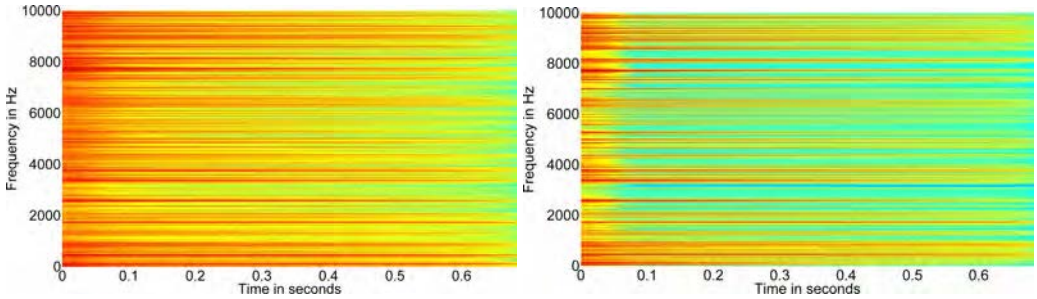
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
41439_sandyrb_srbm − tom − 003	0,5613	0,5853
N_{modes}	Δ_s	σ_s
501	0.0125	4.7267

Tabelle 4: Kennwerte Tom

Zusammengefasst kann gesagt werden, dass der Klang einer Tom gut durch das modale Modell repräsentiert werden kann. Die Klangtreue des synthetisierten Signals ist aufgrund der Nichtlinearität jedoch nicht gut.

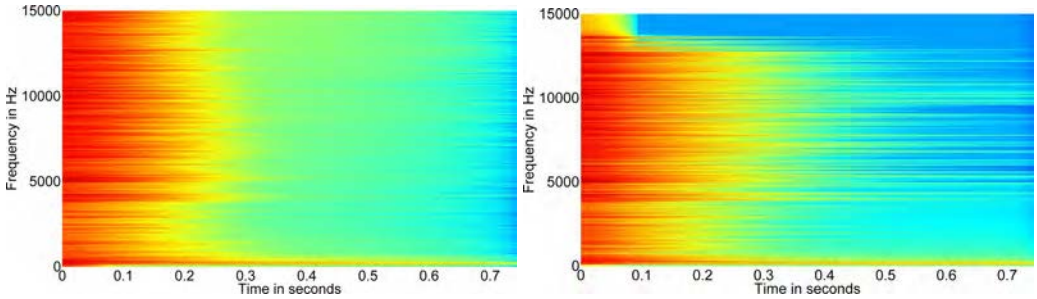
5.7.4 *Becken*

Becken Instrumente entsprechen, abgesehen von der Hi-Hat, im Grunde des in dieser Arbeit behandelten Modells eines Soliden Körpers, welches eine quasi impulsartige Anregung erfährt. Eine Problematik bei diesem Instrument liegt in der großen Anzahl von Eigenmoden, die sich teilweise in einem hohen Grundrauschen widerspiegeln. Außerdem bewegen sich die meisten Becken nach der Anregung aufgrund der Lagerung im Ganzen, was zu einem schwankenden Pegel bei der Aufnahme führt. Abhängig von der Anregung des Becken kann es außerdem zu einer starken Verformung des Beckenkörpers kommen, was zu einem starken nichtlinearen Schwingungsverhalten führen kann. Diese Punkte sind eine mögliche Erklärung für Moden, welche sich teilweise nicht durch eine exponentiell abfallende Amplitude erklären lassen (siehe dazu Abbildung 53).



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 51: EDR Ride Becken



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 52: EDR Hi-Hat

Aus dem Bereich der Beckeninstrumente wurden Hi-Hats, Crash-Becken und Ride-Becken analysiert. Die Ergebnisse der Hi-Hats und der Crash waren insgesamt nicht zufriedenstellend, was auch auf ihr hohen Anteil an Grundrauschen zurückzuführen ist. Dies wird auch ersichtlich, wenn man die beiden EDRs des Ride-Beckens (Abbildung 51 a)) und der Hi-Hat (Abbildung 52 a)) vergleicht. Die Hi-Hat besitzt sehr viele Moden, welche sich über das gesamte Frequenz Spektrum erstrecken und den rauschartigen Charakter erzeugen. Das Ride Becken dagegen, hat weniger Moden, welche sich klarer voneinander abgrenzen. Auf Grund der schlechten Ergebnisse bei Hi-Hats und Crash-Becken wird im folgenden auf die Analyse- und Syntheseergebnisse von Ride-Becken eingegangen.

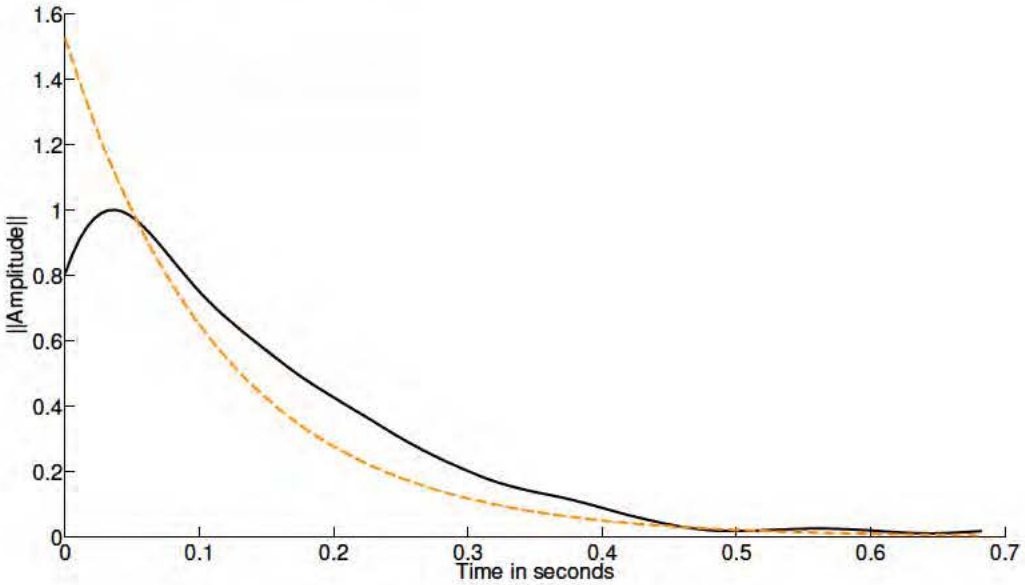
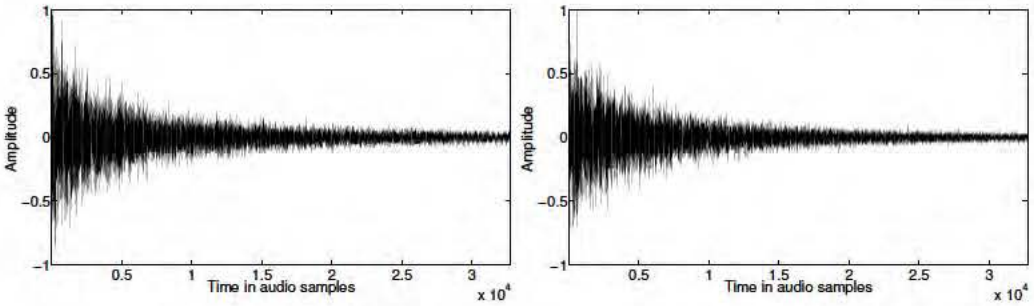


Abbildung 53: Hüllkurvenannäherung bei Becken
— geschätzte Kurve, — realer Verlauf

Der Amplitudenverlauf des untersuchten Becken stimmt tendenziell mit der angenäherten exponentiellen Kurve überein (siehe Abbildung 53). Die energiereichste Mode liegt bei einer Frequenz von $f = 7740$Hz.
Der Vergleich der beiden Signal 54, zeigt, dass die Amplituden des synthetisierten Klanges etwas kleiner sind als die des Originalsignals. Tendenziell stimmen sie aber gut überein. Aufgrund der vielen Moden ist eine qualitative Aussage mit Hilfe des Signalvergleichs aber nicht möglich.



a) Originalsignal b) Synthesesignal

Abbildung 54: Signalvergleich Becken

Der relativ geringe Korrelationsfaktor $\rho_{O,S} = 0.2359$ zeigt, dass die Synthese nicht gut mit dem Original übereinstimmt. Die Anzahl der Moden nach Stutzen ist zu gering, um den Klang gut wiedergeben zu können. Die Abweichung der Energie Δ_s und die Standardabweichung σ_s sind überraschenderweise relativ gering.

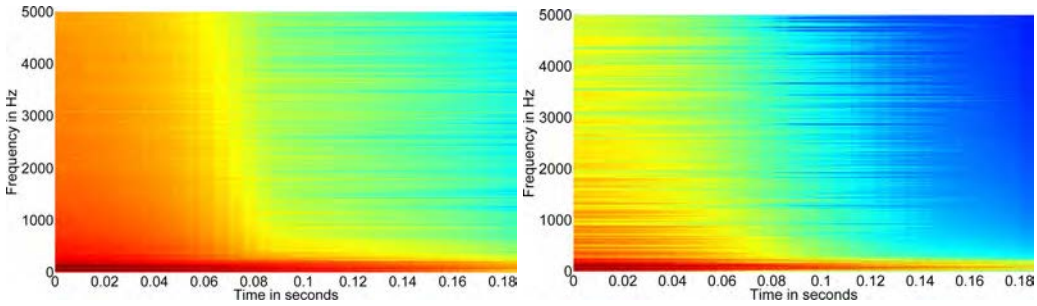
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
154713_mistorcaveman_cymbal_6_1_2	$-0,0852$	0.2359
N_{modes}	Δ_s	σ_s
187	0.0097	1.7333

Tabelle 5: Kennwerte Becken

Zusammengefasst kann gesagt werden, dass das lineare Verhalten eines Becken gut durch das modale Modell angenähert werden könnte. Jedoch nur, wenn die Modenstutzung deaktiviert ist. Ansonsten werden zu viele Moden im hohen Frequenzbereich gestutzt.

5.7.5 *Bassdrum*

Die Bassdrum ist ähnlich aufgebaut wie eine Tom. Unterschiede bestehen in der größeren Membranen, was sich in einer tieferen ersten Resonanzfrequenz widerspiegelt. Weiter zeichnen sich viele Bassdrums bei kraftvoller Anregung durch ein stark nichtlineares Schwingverhalten aus. Dies wird durch das hier behandelte lineare Klangmodell nicht erfasst. Anhand der dargestellten EDRs in Abbildung 55 ist zu sehen, dass sich fast die gesamte Energie des Signals in den niedrigen Frequenzen < 500Hz befindet. Die Energie der Moden nimmt mit steigender Frequenz sehr schnell ab. Anhand des Amplitudenverlaufs der



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 55: EDR Bassdrum

ersten Mode bei einer Frequenz von $f \quad 71$Hz (siehe Abbildung 56) ist gut zu erkennen, das der Amplituden verlauf stark von einer exponentiell fallenden Kurve abweicht, so dass die Annäherung ungenügend ist.

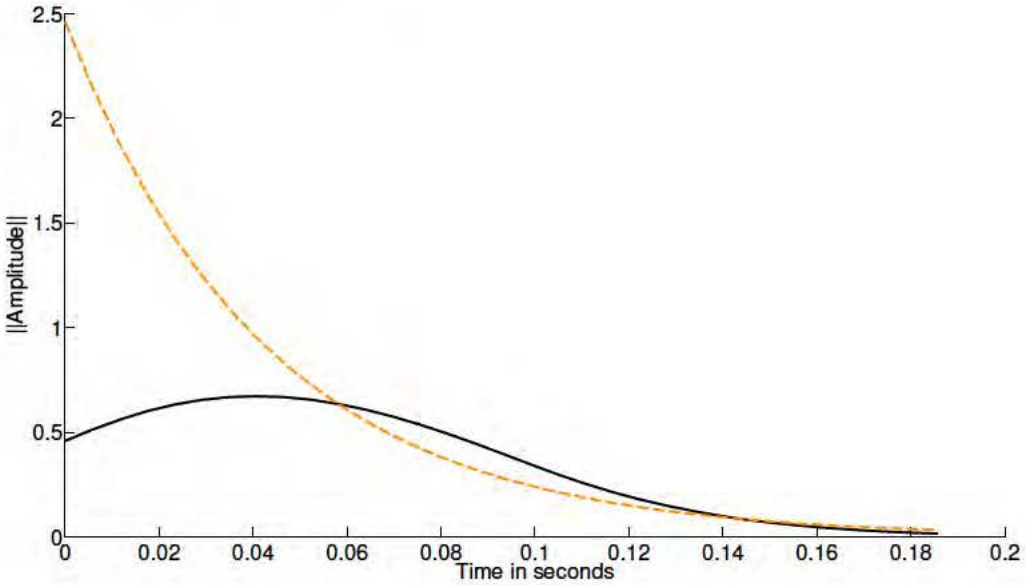
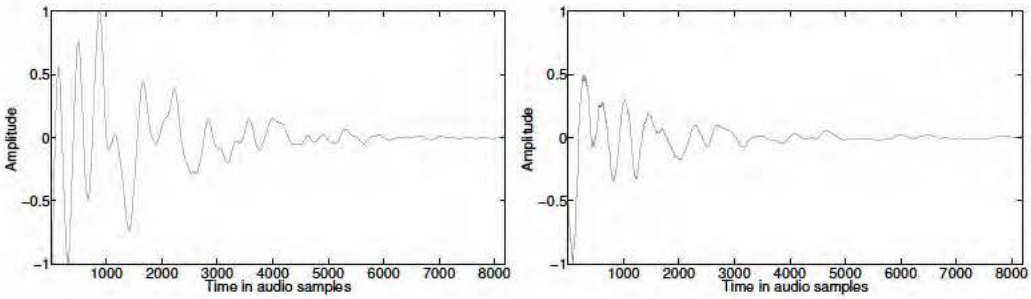


Abbildung 56: Hüllkurvenannäherung bei Bassdrum
— geschätzte Kurve, — realer Verlauf

Der Vergleich der beiden Signale in Abbildung 57 zeigt, das deutliche Unterschiede zwischen dem Original- und dem Synthesesignal bestehen. Die Amplituden des Synthesesignals fallen schneller ab. Zudem ist es gegenüber dem Originalsignal phasenverschoben. Die beiden Signale scheinen insgesamt nicht zu korrelieren.



a) Originalsignal b) Synthesesignal

Abbildung 57: Signalvergleich Bassdrum

Der geringe und negative Korrelationsfaktor zeigt die oben beschriebenen Unterschiede deutlich auf. Weiter deutet die große Standardabweichung $\Delta_s = 9.5138$ auf ein großes Maß an Nichtlinearität des untersuchten Instruments hin.

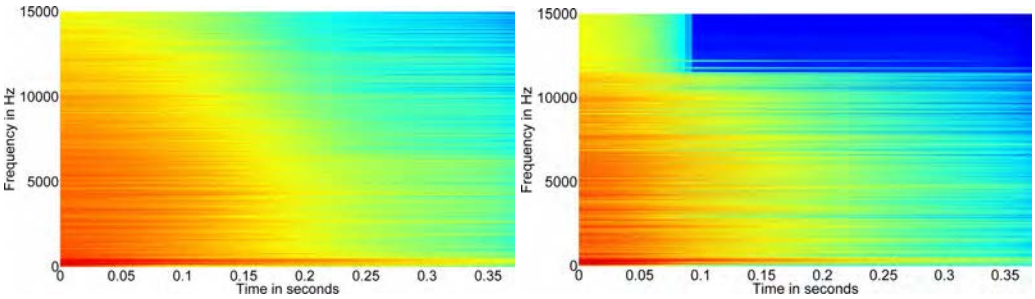
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
82313_kevoy_kick−2	−0,2382	−0,1456
N_{modes}	Δ_s	σ_s
422	0.0598	9.5138

Tabelle 6: Kennwerte Bassdrum

Insgesamt wird die analysierte Bassdrum nicht gut durch das verwendete lineare Modell widergespiegelt. Dennoch konnte ein Klang erzeugt werden, der Eigenschaften einer linearen Bassdrum aufweist.

5.7.6 *Snaredrum*

Die Resynthese der Snaredrum erzeugte keine zufriedenstellenden Ergebnisse. Die Hauptursache liegt im Unterschied der Klangerzeugung bei einer Snaredrum, zu der in Kapitel 3 beschriebenen Theorie der Klangerzeugung, welche in dieser Arbeit vorausgesetzt wird. Bei einer Snaredrum umfasst die Klangerzeugung im wesentlichen zwei Schritte. Die Anregung der Schlagmembran und die dadurch ausgelöste Vibration des Schnarrteppichs unterhalb der Schlagmembran, welche wiederum eine Anregung der Schlagmembran zur Folge hat. Diese rückgekoppelte Anregeform wird durch das in dieser Arbeit verwendete Klangmodell nicht hinreichend modelliert, so dass die erzeugten Klänge insbesondere in der Abklingphase ein unterschiedliches Klangbild besitzen. Dies wird bei Betrachtung der energiereichsten Mode in Abbildung 59 deutlich. Ähnlich wie Beckeninstrumente weisen Snaredrums einen hohen Rauschanteil auf. Dies wird in Abbildung 58 deutlich. Die Snaredrum hat sehr viele Moden, dessen Energien relativ langsam mit steigender Frequenz abfallen. In Abbildung 59



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 58: EDR Snaredrum

ist der Amplitudenverlauf des Original- und Synthesesignals dargestellt. Nach einem schnellen Abfall der Amplitude kurz nach der Anregung, hat die Amplitude in der Abklingphase keinen glatten fallenden Verlauf, sondern ein unregelmäßiges Anstiegs- und Abstiegsverhalten. Dieses wird eventuell durch die Rückkopplung des Schnarrteppichs, oder die Nichtlinearität der schwingenden Membran ausgelöst.

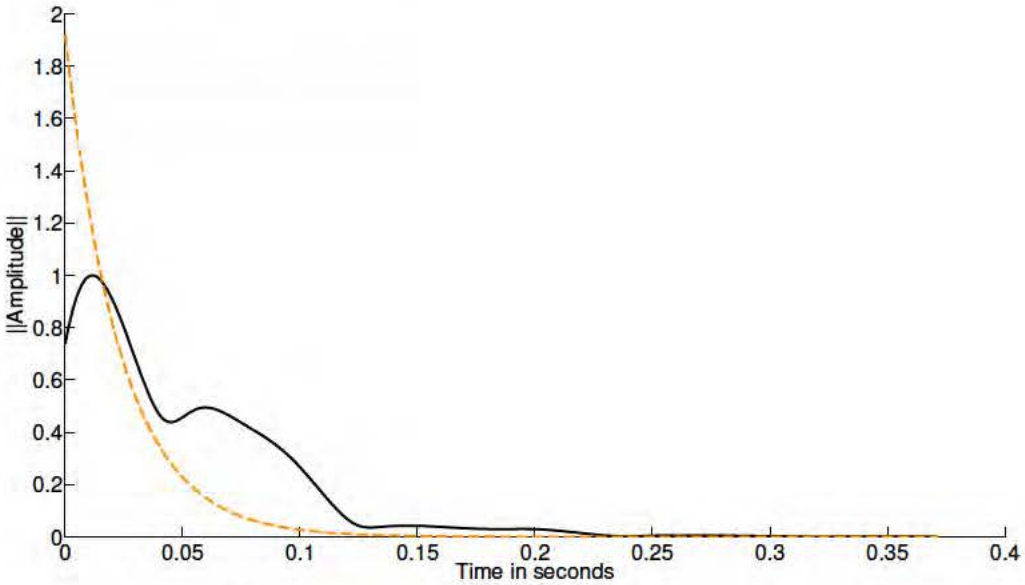
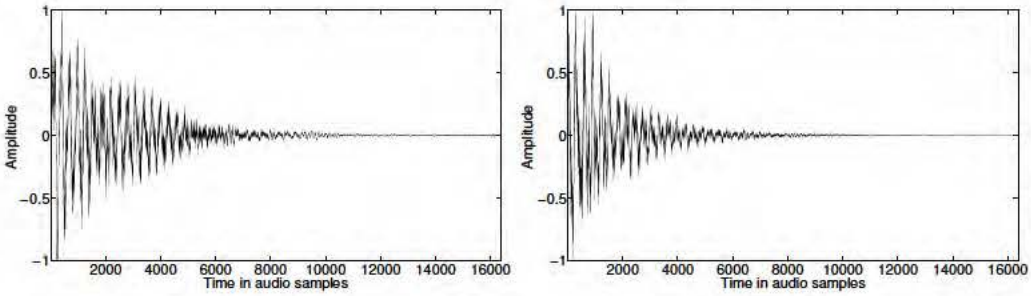


Abbildung 59: Hüllkurvenannäherung einer Snaredrum
— geschätzte Kurve, — realer Verlauf

Die beiden Zeitsignal überlagern sich auf den ersten Blick gut. Der Klang einer Snaredrum ist jedoch hauptsächlich durch Moden im mittleren und hohen Frequenzbereich charakterisiert. Diese Frequenzanteile sind im Signalvergleich in Abbildung 60 nicht sehr gut sichtbar. Die Amplituden der höheren Frequenzen sind, verglichen mit der Amplitude der Grundmode, sehr gering.



a) Originalsignal b) Synthesesignal

Abbildung 60: Signalvergleich Snaredrum

Das die in Abbildung 60 gezeigten Signale nicht sehr stark korrelieren wird anhand des geringen Korrelationsfaktors mit einem Wert von $\rho_{O,S} = 0,1341$ deutlich . Die große Standardabweichung von $\sigma_s = 5.0965$ zeigt, dass es sich um einen nichtlinearen Schwinger handelt. Das sich die Signale bei oberflächlicher Betrachtung ähneln wird durch die geringe Energiedifferenz $\Delta_s = 0.0224$ widergespiegelt.

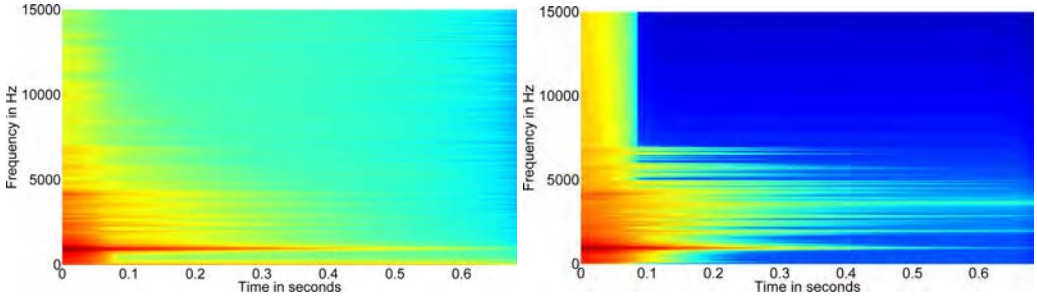
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
82238_kevoy_snare − drum	0,1595	0,1341
N_{modes}	Δ_s	σ_s
524	0.0224	5.0965

Tabelle 7: Kennwerte Snaredrum

Eine weiteres Problem bei der Resynthese liegt in der großen Anzahl an Moden einer Snaredrum, diese sind mit der verwendeten Analysemethode nicht erfassbar. Die große Anzahl an Moden eignet sich zudem aufgrund der Rechenleistung heutiger Endverbrauchergeräte nicht für eine Echtzeit-Synthese. Es ist davon auszugehen, dass diese Klänge in Zukunft, unter der Voraussetzung weiter wachsender Rechnerkapazitäten, ebenfalls in Echtzeit synthetisierbar sind.

5.7.7 *Clave*

Die Clave besteht aus zwei Holzstäben und wird als Rhythmuselement in der Musik eingesetzt. Der Klang entsteht durch das Zusammenschlagen der beiden Stäbe. Somit wird der Klang durch zwei Schwinger produziert. Das EDR zeigt, dass es eine dominierende Mode gibt. Die Amplituden und Abklingzeiten der Obertöne fallen mit der Frequenz sehr schnell ab.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 61: EDR Clave

Der Amplitudenverlauf in Abbildung 62 weist keinen exponentiellen Verlauf auf. Die angenäherte Kurve weicht daher stark von der Originalkurve ab. Möglicherweise hängt dies damit zusammen, dass sich der Klang der Clave aus zwei Schwingern zusammensetzt, welche unterschiedliche Material und Geometrieeigenschaften besitzen. Außerdem sind die beiden Stäbe der Clave während des Schwingvorgangs unterschiedlich „gelagert “, was zu unterschiedlichen Eigenfrequenzen führt.

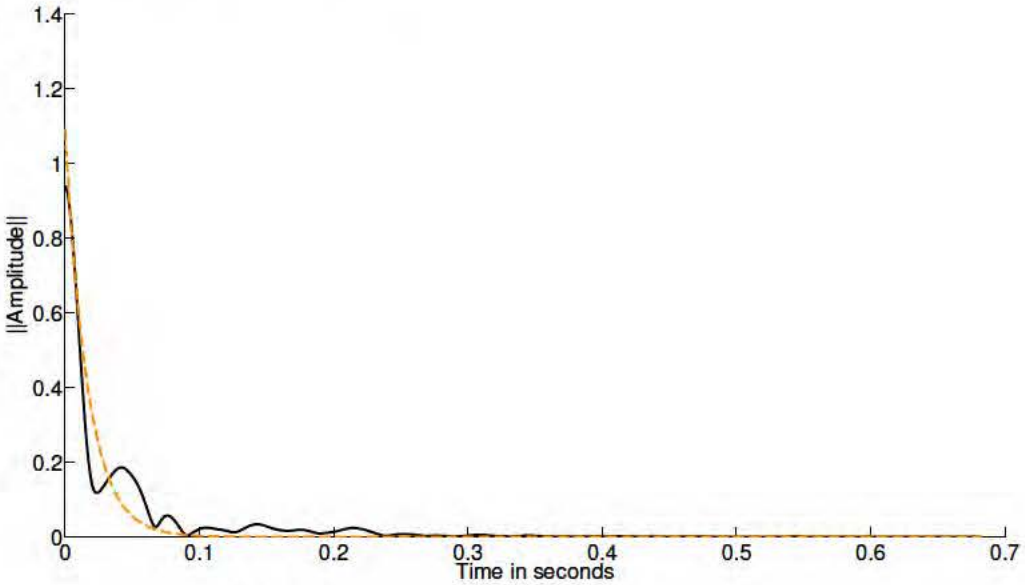
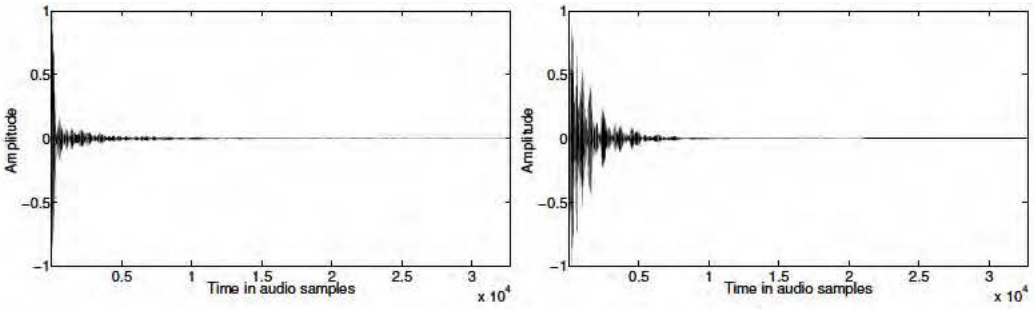


Abbildung 62: Hüllkurvenannäherung Clave
— geschätzte Kurve, – realer Verlauf

Das synthetisierte Signal weißt Charakteristiken des Originalsignals auf. Jedoch unterscheiden sich die Amplituden stark voneinander.



a) Originalsignal b) Synthesesignal

Abbildung 63: Signalvergleich Clave

Der Unterschied im Amplitudenverlauf wird auch bei Betrachtung des geringen Korrelationsfaktors deutlich. Der geringe Energieunterschied spiegelt, die oberflächlich betrachtete Ähnlichkeit der Signalverläufe wieder. Weiter scheint der Schwingungsverlauf der Clave aufgrund der hohen Standardabweichung von $\sigma_s = 21.7651$ stark nichtlinear zu sein. Wahrscheinlich hängt die hohe Standardabweichung jedoch mit der Klangerzeugung einer Clave zusammen. Der Klang wird wie oben beschrieben dadurch erzeugt, dass zwei Holzbalken aufeinander geschlagen werden. Dabei erzeugen beide Balken einen Klang. Dabei können zwei sich unterscheidende Klänge entstehen, welche sich zudem Überlagern.

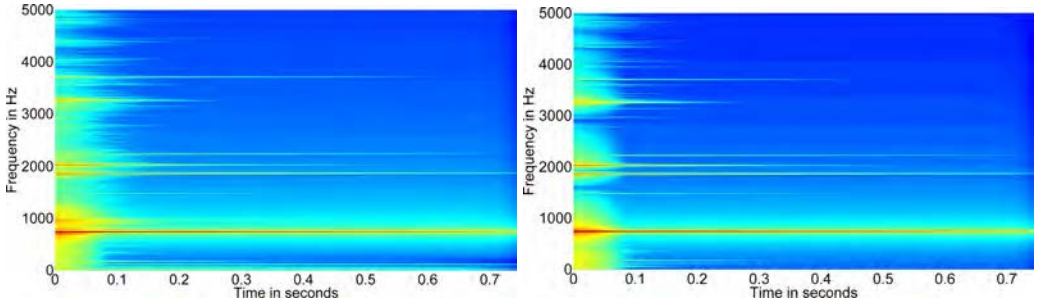
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
132417__sajmund_ _percussion − clave − like − hit	0,1515	−0,3185
N_{modes}	Δ_s	σ_s
298	0.0064	21.7651

Tabelle 8: Kennwerte Clave

Insgesamt ist der synthetisierte Klang einer Clave sehr gut. Er hat zudem eine hohe Übereinstimmung mit dem Originalklang. Da scheinbar nichtlineare Schwingverhalten hat keinen negativen Einfluss auf die lineare Resynthese.

5.7.8 *Glocke*

Glocken bestehen aus einem soliden Körper und werden meist impulsartig angeregt. Daher haben sie eine gute Übereinstimmung mit den hier behandelten Ansätzen zur Klangsynthese. Dies spiegelt sich auch in den Ergebnissen wieder. Das in Abbildung 64 gezeigte EDR einer untersuchten Glocke zeichnet sich durch klar definierte ersten Moden aus. Es existieren außerdem nur sehr wenig Moden. Diese sind auf einen begrenzten Spektralbereich bis zu einer Frequenz von f 500Hz verteilt.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 64: EDR Glocke

Abbildung 68 zeigt, dass sich die Annahme einer exponentiell abfallenden Amplitude auch bestätigt. Die angenäherte Kurve kommt der realen Kurve sehr nahe. Dies spiegelt sich auch im Klang wider. Der Klang der synthetisierten Glocke wird bis auf den Klang in den ersten Millisekunden sehr gut wiedergegeben.

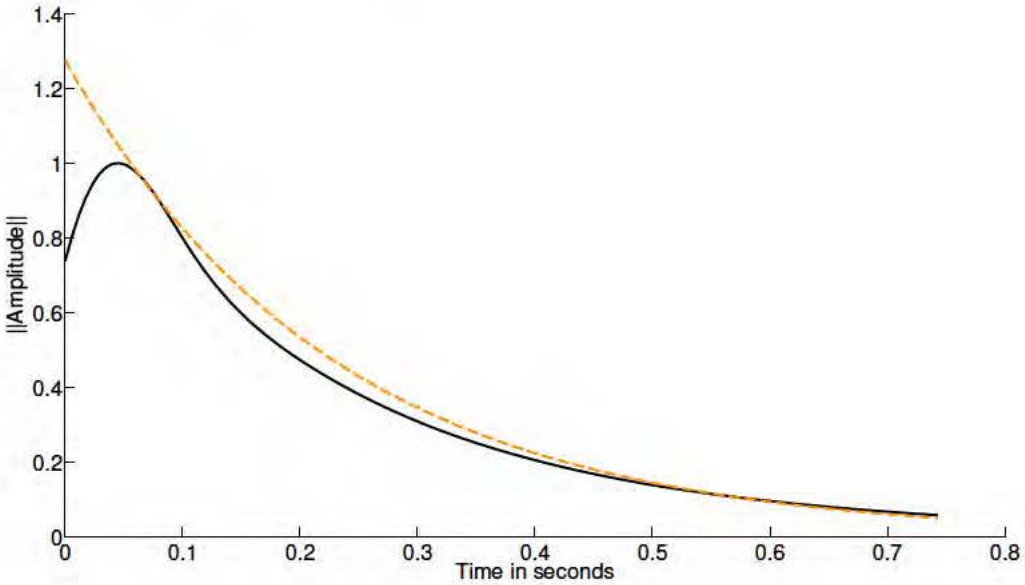
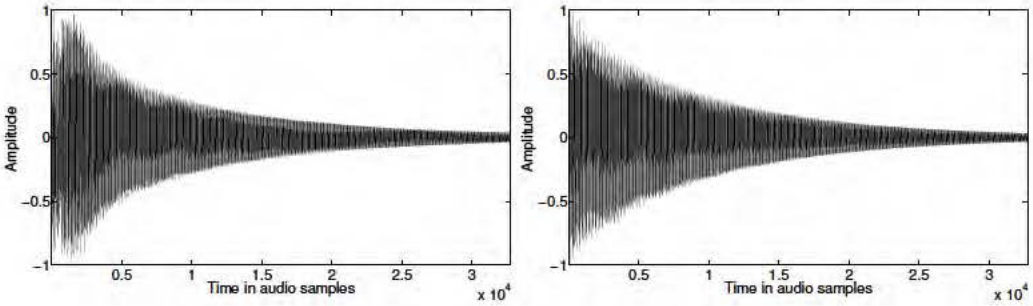


Abbildung 65: Hüllkurvenannäherung bei Glocke
— geschätzte Kurve, — realer Verlauf

Der Vergleich des Originalsignals und des Synthesesignals in Abbildung 66 zeigt, dass sich die Kurven für den Bereich nach den ersten 2000 Samples sehr gut annähern.



a) Originalsignal b) Synthesesignal

Abbildung 66: Signalvergleich Glocke

Die guten Ergebnisse bei Betrachtung der EDRs und der Signalverläufe spiegelt sich auch bei den numerischen Werten wider. So ist der Korrelationskoeffizient mit einem Wert von $\rho_{O,S} = 0,6121$ nach dem synthetisierten Sinussignal der zweithöchste der hier vorgestellten Klänge. Das die Standardabweichung nicht Null ist hängt damit zusammen, dass die Glocke durch die starke Anregung wahrscheinlich in den ersten Millisekunden ein nichtlineares Schwingungsverhalten aufweist.

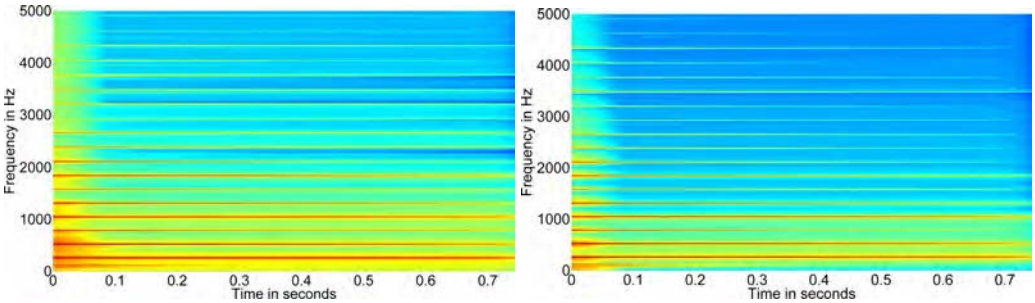
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
10447_batchku_gbell − 4 − 2 − pp	0,5219	0,6121
N_{modes}	Δ_s	σ_s
153	0.0555	1.2816

Tabelle 9: Kennwerte Glocke

Die Glocke gehört also zu den Klängen, welche sich sehr gut resynthetisieren lassen und dem beschriebenen Klangmodell sehr nahe kommen.

5.7.9 *Klaviersaite*

Die Klaviersaite ist in der Regel sehr stark gespannt, so dass sie bei der Anregung weniger verformt wird als zum Beispiel eine Gitarrensaite. Daher eignet sich ein Klavier theoretisch sehr gut für die lineare Modalsynthese. Die beiden EDRs in Abbildung 69 zeigen so auch eine große Übereinstimmung.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 67: EDR Klavier

Der Amplitudenverlauf der Originals kann gut angenähert werden. Der Originalverlauf scheint aber keinen exakten exponentiellen Verlauf zu haben, was möglicherweise mit den Resonanzeigenschaften des Klavierkörpers zusammenhängt.

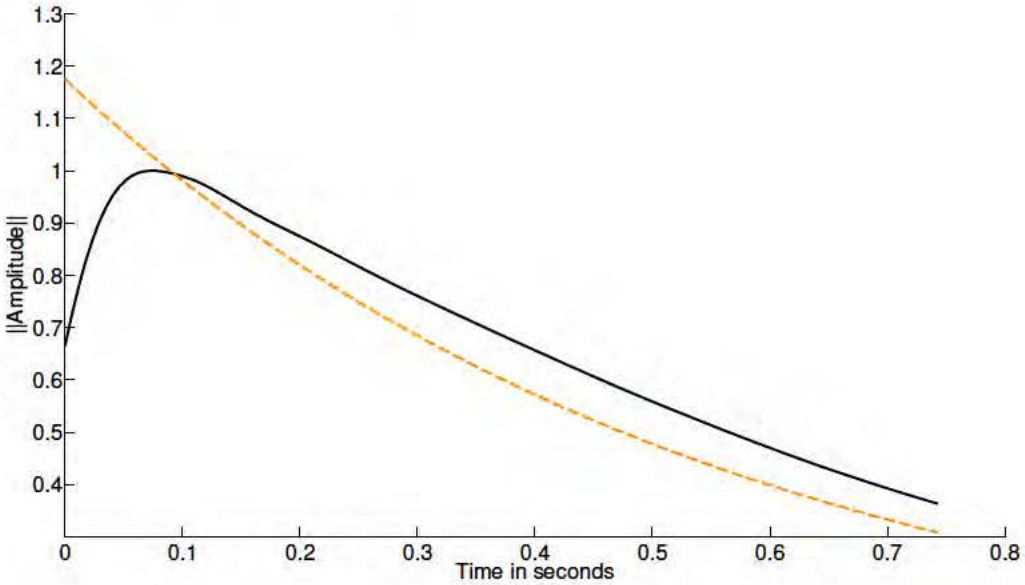
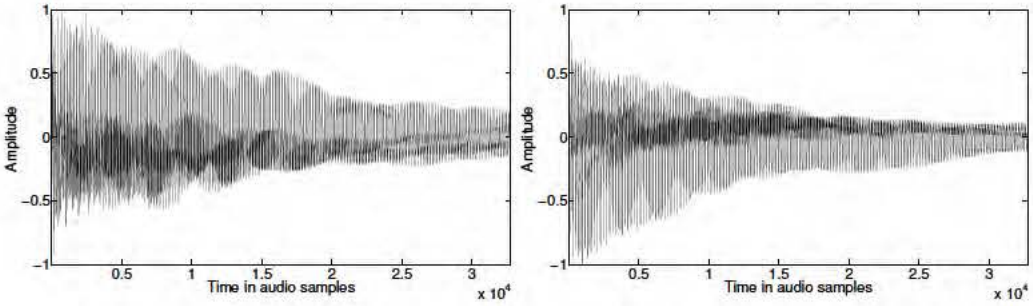


Abbildung 68: Hüllkurvenannäherung bei Glocke
— geschätzte Kurve, — realer Verlauf

Die Zeitsignale stimmen in diesem Beispiel nicht sehr gut überein. Das Synthesesignal scheint gegenüber dem Originalsignal an der X-Achse gespiegelt zu sein.



a) Originalsignal b) Synthesesignal

Abbildung 69: Signalvergleich Klavier

Die Standardabweichung $\sigma_s = 0$ zeigt, das die Klaviersaite linear schwingt. Die großen Unterschiede im Signalverlauf spiegeln sich in dem geringen Korrelationsfaktor von $\rho_{O,S} = 0.1853$ wider. Das Signal kann mit wenigen Moden ($N_{modes} = 31$) resynthetisiert werden.

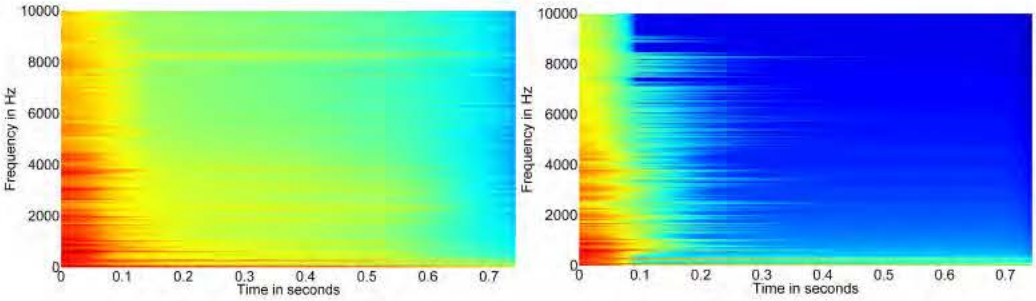
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
10447_batchku_gbell−4−2−pp	0.1563	0.1853
N_{modes}	Δ_s	σ_s
31	0.0711	0

Tabelle 10: Kennwerte Glocke

Die Klaviersaite gehört wie die Glocke also zu den Klängen, welche sich ebenfalls sehr gut resynthetisieren lassen und dem vorgestellten Klangmodell sehr nahe kommen.

5.7.10 *Marimba*

Die Marimba besitzt viele Moden oberhalb der Grundmode. Diese sind jedoch auf einen kleinen Spektralbereich beschränkt. Oberhalb von 10kHz treten praktisch keine signifikanten Moden mehr auf. Amplituden und Abklingzeiten fallen schnell mit wachsender Frequenz.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 70: EDR Marimba

Der Amplitudenverlauf der stärksten Mode kann nicht sehr gut durch einen exponentiellen Verlauf angenähert werden. Jedoch folgt die Annäherung dem Amplitudenverlauf tendenziell. Die globale Amplitude des Amplitudenverlaufs wird rasch nach der Anregung erreicht.

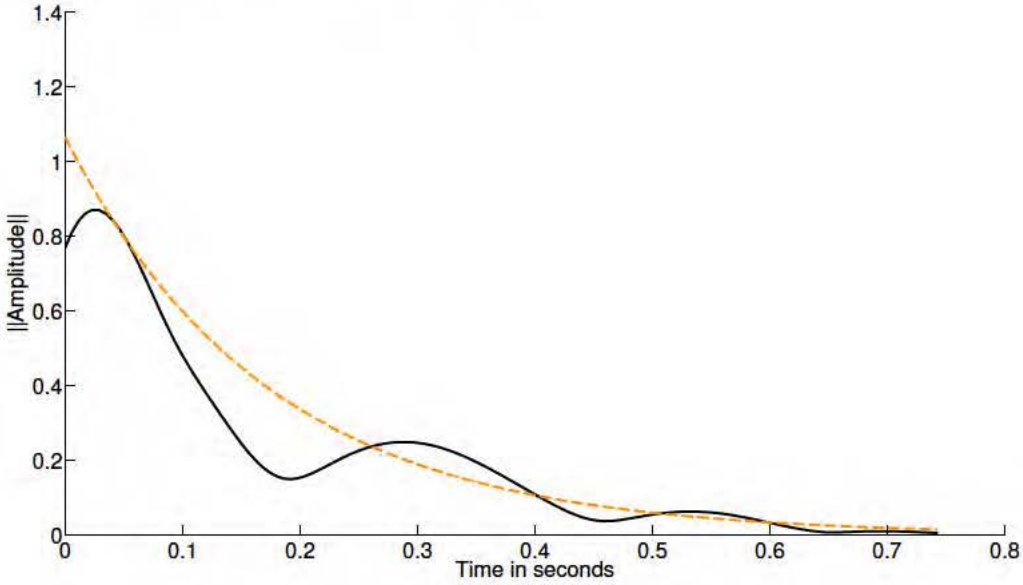
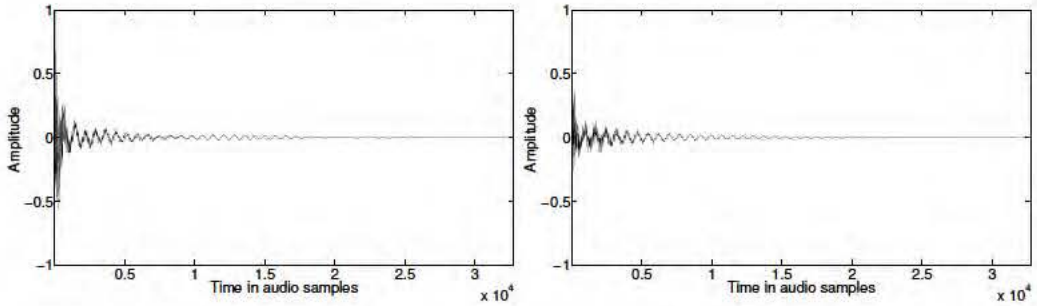


Abbildung 71: Hüllkurvenannäherung bei Marimba
— geschätzte Kurve, – realer Verlauf

Die Übereinstimmung der Signale ist tendenziell gut. Jedoch unterscheiden sich die Amplituden zu Beginn des Signals.



a) Originalsignal b) Synthesesignal

Abbildung 72: Signalvergleich Marimba

Der Korrelationsfaktor zeigt, dass das die Synthese gewisse Ähnlichkeiten mit dem Originalsignal aufweist. Jedoch das Synthesesignal gegenüber dem Originalsignal Phasenverschoben. Die sehr große Standardabweichung $\sigma_s = 5.0871$ weist auf ein nichtlineares Verhalten hin, welches eventuell durch den Resonanzkörper unterhalb der Stäbe einer Marimba hervorgerufen wird.

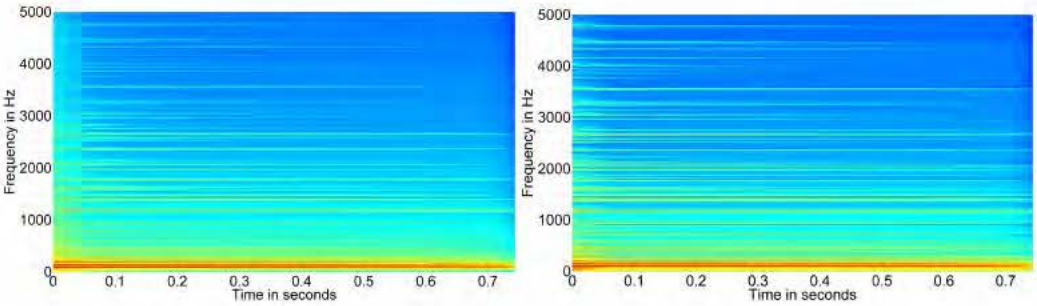
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
185833_lloydevans09 _marimba-note-struck	$-0,3851$	$-0,3844$
N_{modes}	Δ_s	σ_s
279	0.0018	5.0871

Tabelle 11: Kennwerte Marimba

Klanglich ist das Ergebnis der Synthese zufriedenstellend.

5.7.11 *Akustikgitarre*

Das EDR der Saite einer Akustikgitarre in Abbildung 73 a) zeigt, dass es wenige Moden im tiefen Frequenzbereich gibt. Vereinzelte Moden sind auch im mittleren Frequenzbereich zu sehen. Die Übereinstimmung der gezeigten EDRs in Abbildung 73 ist hoch.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 73: EDR Akustikgitarre

Der Amplitudenverlauf der stärksten Mode 74 in Abbildung zeigt, dass die globale Amplitude nicht direkt nach der Anregung erreicht wird. Dies hängt in diesem Fall damit zusammen, dass die lineare Beschreibung der Gitarrensaite bei der anfänglich starken Auslenkung nicht mehr gültig ist. Etwa ab der Signalhälfte fällt die Amplitude stetig. Die angenäherte Kurve liegt aus diesem Grund weit unterhalb der Originalkurve.

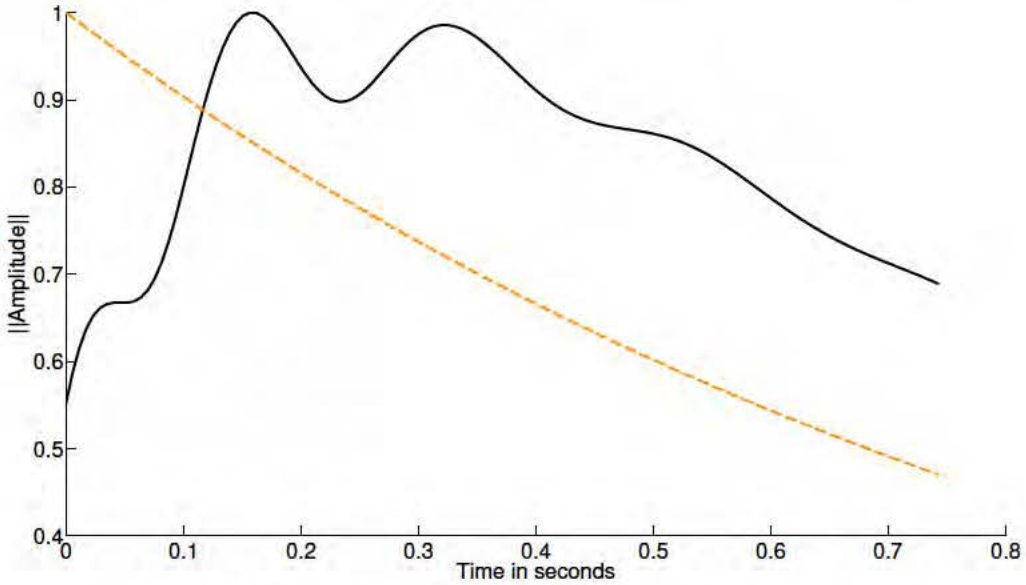
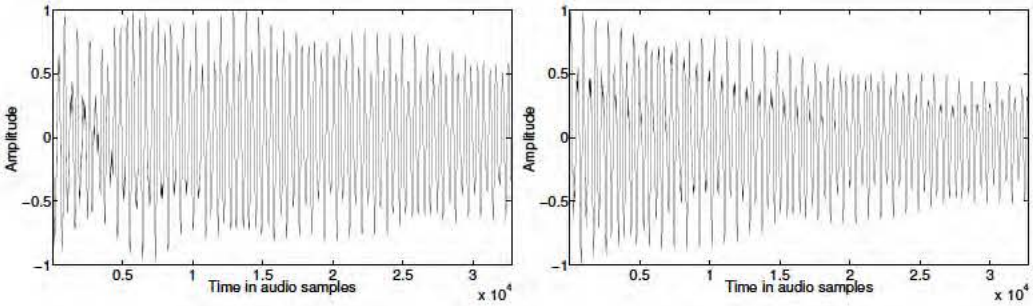


Abbildung 74: Hüllkurvenannäherung bei Akustik Gittare
— geschätzte Kurve, – realer Verlauf

Die Zeitsignale stimmen gut überein. Nur zu Beginn und am Ende des Signals unterscheiden sich die Amplituden des Synthese- und Originalsignals.



a) Originalsignal b) Synthesesignal

Abbildung 75: Signalvergleich Akustikgitarre

Die Korrelation der beiden Signal $\rho_{O,S} = -0,4511$ ist hoch, jedoch deutet der negative Wert auf eine Phasenverschiebung hin. Die Große Standardabweichung mit einem Wert $\sigma_s = 5.1403$ zeigt, dass die Gitarrensaite kurz nach der Anregung ein nichtlineares Schwingungsverhalten aufweist.

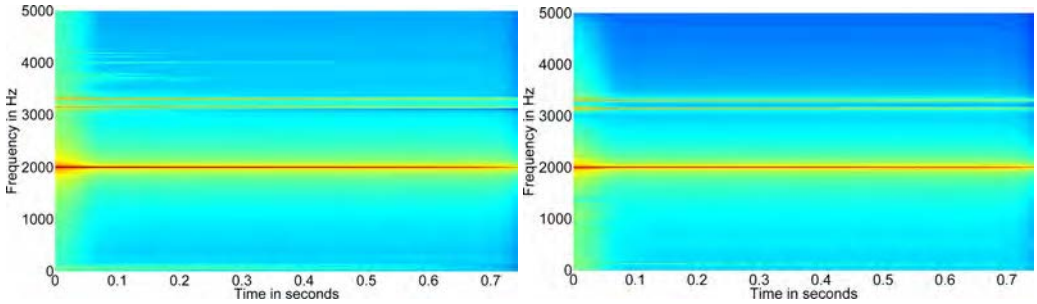
AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
187697_afrodrumming_acoustic − guitar	−0,4642	−0,4511
N_{modes}	Δ_s	σ_s
103	0.0224	5.1403

Tabelle 12: Kennwerte Akustikgitarre

Insgesamt lässt sich der lineare Klang einer Gitarrensaite jedoch sehr gut annähern. Jedoch sollte für die zukünftige Modellbildung der Schwingung einer Gitarrensaite der nichtlineare Anteil berücksichtigt werden.

5.7.12 *Kantilan*

Das EDR eines Kantilan in Abbildung 76 zeigt eine ausgeprägte erste Mode und wenige Obertöne . Oberhalb von 5kHz gibt es praktisch kein Moden mehr. Das EDR des synthetisierten Klanges scheint fast identische mit dem EDR des Originalsignals zu sein.



a) EDR Originalsignal b) EDR Synthesesignal

Abbildung 76: EDR Kantilan

Der Amplitudenverlauf kann tendenziell gut angenähert werden. Jedoch kann die Amplitude nicht exakt angenähert werden, da die globale Amplitude in Abbildung 77 bei $t > 0.1s$ liegt und somit keine exponentielle Extrapolation durchgeführt wird. Weiter wird die Schwebung durch das exponentielle Modell in erster Betrachtung nicht erfasst.

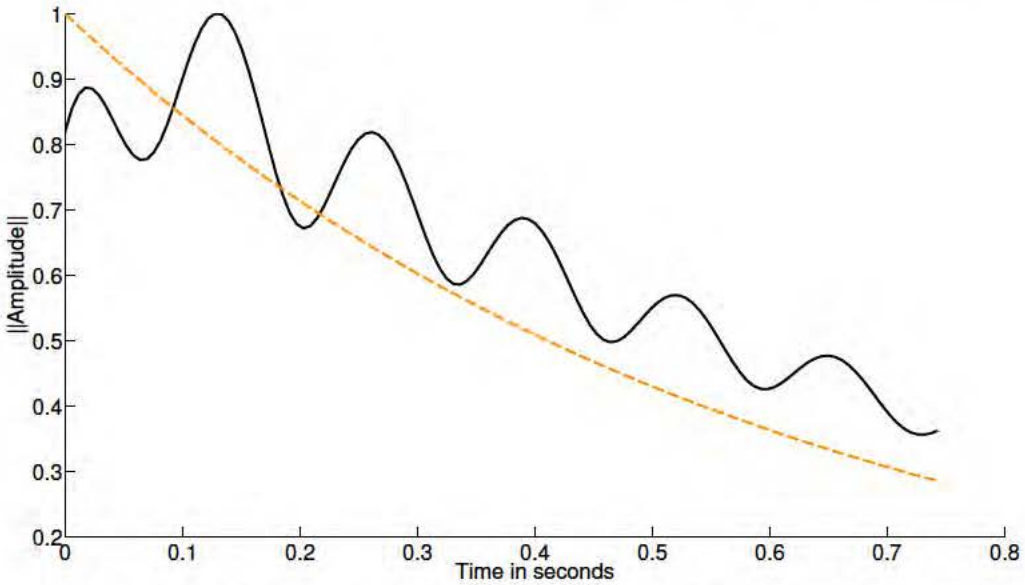
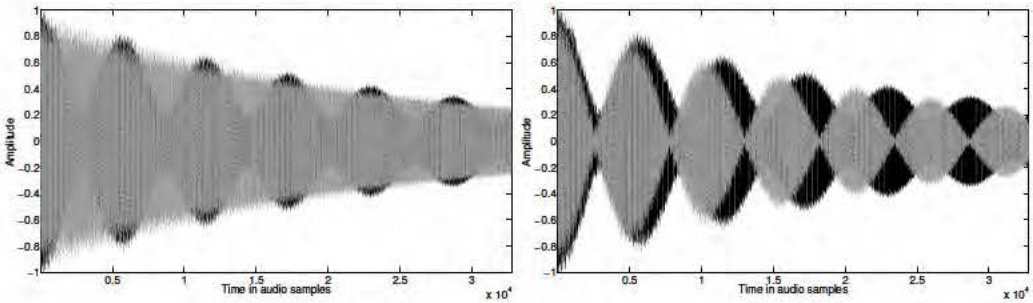


Abbildung 77: Hüllkurvenannäherung Kantilan
— geschätzte Kurve, – realer Verlauf

Da die, die Schwebung verursachenden Frequenzen durch die spektrale Auflösung nicht erfasst werden, unterscheiden sich das Original- und Syntheseignal stark. Ist die Detektierung von Schwebungen aktiviert, kann das Signal zum Teil reproduziert werden.



a) Signalvergleich ohne Schwebungen
–– geschätzte Kurve, – realer Verlauf

b) Signalvergleich mit Schwebungen
–– geschätzte Kurve, – realer Verlauf

Abbildung 78: Signalvergleich Kantilan

Der Korrelationsfaktor $\rho_{O,A}$ und $\rho_{O,S}$ zeigt, dass die Korrelation mit dem additiven Signal besser als mit dem synthetisierten Signal ist. Dies liegt hauptsächlich an der nicht erfassten Schwebung, welche durch das einfach exponentielle Modell nicht erfasst wird. Die Energiedifferenz ist mit $\Delta_s = 0.0947$ aufgrund der nicht vorhandenen Schwebung ebenfalls sehr groß.

AUDIODATEI	$\rho_{O,A}$	$\rho_{O,S}$
$24621__anamorphosis__gmb - kantilan - 10$	$-0,3541$	$0,1081$
N_{modes}	Δ_s	σ_s
73	0.0947	1.8760

Tabelle 13: Kennwerte Kantilan

Die Kantilan kann unter Berücksichtigung der Schwebungen gut synthetisiert werden, obwohl die Schwebefrequenz und Amplitude nicht exakt angenähert werden kann.

Teil VI

FAZIT UND AUSBLICK

6 DISKUSSION UND AUSBLICK

Hauptbestandteil der Arbeit war die Entwicklung des *MATLAB* Programms *Mode-A-Lizer*. Das Programm beinhaltet eine Benutzeroberfläche, um dem Nutzer die Möglichkeit zu geben, Analyseparameter zu ändern. Außerdem können die Ergebnisse visualisiert werden. Das Programm ermöglicht die automatische Analyse von Klangdateien und den Export der Ergebnisse als *Synth-A-Modeler* Modell. Es wurden insgesamt 120 Klangdateien analysiert und als *Synth-A-Modeler* Modelle gespeichert. Zusätzlich wurden 10 Modelle auf Grundlage von analytischen Resonanzmodellen erstellt.

Im folgendem Abschnitt werden die Ergebnisse der Signalanalyse und Resynthese zusammengefasst und diskutiert. Weiter werden Möglichkeiten präsentiert, um das Analyseverfahren und die Ergebnisse zu verbessern.

6.1 DISKUSSION

Die Spektralanalyse erzielte bei Signalen mit wenigen Moden gute Ergebnisse. Der zeitliche und spektrale Verlauf konnte in diesen Fällen ausreichend aufgelöst werden. Bei Signalen mit sehr vielen Moden oder rauschartigen Signalen, war die spektrale Auflösung nicht ausreichend. Weiter besteht ein generelles Problem in der Analyse der ersten Millisekunden einer Datei. Diese konnten auch mit der angewendeten Frequenzanalyse nicht erfasst werden.
Die Modendetektierung erfasst alle Frequenzen, welche sich aus dem Rauschboden abheben. Dies ist teilweise problematisch, da es sich bei diesen Freuquenzen nicht unbedingt um Moden handelt. Bei den erfassten Frequenzen kann es sich zum Beispiel auch um Störfrequenzen handeln, welche durch Hall oder durch ein Netzrauschen hervorgerufen werden. Außerdem kann es sich auch um Nebenkeulen handeln, welche aufgrund der Fouriertransformation entstehen. Dies führt bei der Synthese teilweise zu schlechten Ergebnissen oder zu Modellen mit unnötig vielen Moden, welche den Rechenaufwand erhöhen. Diese Effekte können bei nicht hörbaren Frequenzbestandteilen durch die Modenstutzung mit Hilfe der vorgestellten psychoakustischen Verfahren verhindert werden. Andererseits werden aber auch spektrale Bestandteile erkannt, welche zwar keine Moden sind, aber den Klang des Instrumentes für eine bestimmte Anregung prägen. Diese Bestandteile werden auch als Residuen bezeichnet [20]. Sie werden in dieser Arbeit in den Synthesen integriert.
Die Ermittlung der Abklingzeiten und Amplituden liefert für Signale, welche dem in dieser Arbeit verwendeten Klangmodell entsprechen gute Ergebnisse, so dass die synthetisierten Signale den Originalsignalen sehr nahe kommen. Die Herausforderung bestand darin Signale, welche nicht direkt durch exponentiell abklingende Amplituden beschreibbar sind zu analysieren. Mit Hilfe der implementierten Algorithmen konnten auch für diese Klänge Ergebnisse erzielt werden, welche klangliche Ähnlichkeiten mit den ursprünglichen Tö-

nen aufwiesen.
Die vom Nutzer selbstständig durchzuführende Modenstutzung im Anschluss der Analyse gibt dem Nutzer sowohl die Möglichkeit Moden im Hinblick auf deren Einfluss auf den Klang zu stutzen, als auch den Klang nach seinen Wünschen zu verändern. Es ist möglich, eine große Anzahl an Moden zu stutzen und somit den Rechenaufwand bei der Synthese zu verkleinern, ohne die Klangcharakteristik maßgeblich zu verändern.
Bei der Analyse der Dateien hat sich gezeigt, dass vor allem Audiosignale von Instrumenten, welche aus einem soliden Körper bestehen, sich nicht im Ganzen bewegen, und somit nicht richtungsabhängig abstrahlen, gut resynthetisiert werden können. Dazu gehören vor allem Glocken, Metallophone und Gongs. Instrumente aus Metall mit einer gewissen Strukturstärke eignen sich generell besser, da die Anstiegszeit bei diesen Instrumenten in der Regel kürzer ist als bei Instrumenten mit einer Fellmembran oder hölzernen Instrumenten.
Instrumente mit einem hohen Rauschanteil, lassen sich nicht gut durch das verwendete Klangmodell darstellen. Zu den Instrumenten mit einem hohen Rauschanteil gehören vor allem Snare-Drums und Instrumente aus der Gruppe der Becken. Tabelle 14 fasst die Ergebnisse der Resynthese zusammen. Dabei wurden die Bewertungen in zwei Kategorien eingeteilt. Zum einen wurde die Qualität des Klanges unabhängig vom Originalklang bewertet. Zum anderen wurde die Klangqualität hinsichtlich der Klangtreue bezogen auf den Originalklang bewertet. Die Bewertung wurde in vier Kategorien, von sehr gut über gut und nicht gut bis schlecht eingeordnet. Weiter wird gekennzeichnet, ob es sich bei den untersuchten Signalen um einen linearen oder nicht linearen Schwingvorgang handelt.

INSTRUMENT	KLANGQUALITÄT	KLANGTREUE	[L]INEAR/ [N]ICHT LINEAR
Tom	gut	nicht gut	n
Hi-Hat	schlecht	schlecht	n
Becken	nicht gut	schlecht	n
Bassdrum	nicht gut	schlecht	n
Snaredrum	schlecht	schlecht	n
Clave	gut	gut	l/n
Glocken	sehr gut	gut	l/n
Marimba	gut	gut	l/n
Gittarensaite	sehr gut	gut	l/n
Klaviersaite	sehr gut	sehr gut	l
Kantilan	sehr gut	gut	l

Tabelle 14: Zusammenfassung Resynthese

Nicht lineare Modelle konnten im Laufe der Arbeit nicht zufriedenstellend analysiert und resynthetisiert werden. Dies liegt in erster Linie daran, dass kein sinnvolles Modell für den Zusammenhang zwischen Signalamplitude $s(t)$ und

vorherrschender Modenfrequenz $\omega(t)$ erstellt werden konnte. Weiter konnten aufgrund der Modenauswahl nur die ungefähren Eigenfrequenzen eines nicht linearen Schwingers ermittelt werden.
Die Parametrisierung des frequenzabhängigen Dämpfungsfaktors konnte für viele Klänge zufriedenstellend realisiert werden. Die linearisierte Dämpfungstheorie stellt zwar eine starke Vereinfachung dar, dennoch konnten damit analytische Klangmodelle erstellt werden, welche Klangcharakteristiken des untersuchten Materials aufweisen.
Die analytischen erstellten physikalischen Modelle sind insgesamt, verglichen mit den Resynthetisierten, klanglich weniger komplex. Der Klang eines realen Objekts ist nicht vergleichbar mit den in dieser Arbeit entwickelten Modellen gleicher Geometrie. Dies hat mehrere Gründe. Die Abstrahlung der schwingender Strukturen wurde in dieser Arbeit nicht behandelt. Bei realen Schwingern, zum Beispiel einem Zylinder, ist die Abstrahlung der axialen Moden gegenüber der Abstrahlung der Ringmoden wesentlich größer. Weiter konnte keine zufriedenstellende Methode entwickelt werden, welche ein realistisches Dämpfungs- und Amplitudenmodell abbildet. Ein weiterer Aspekt ist der Einfluss der Residuen. Dieses wird in den analytischen Modellen nicht berücksichtigt, beziehungsweise wurde kein Modell für die Simulation des Residuums erstellt.

6.2 AUSBLICK

Um die Ergebnisse der Resynthese zu verbessern, ist es denkbar, in Zukunft qualitativ höherwertige Signale zu verwenden. Alle in dieser Arbeit verwendeten Klangdateien sind lizenzfreie Dateien. Die Qualität der Audiosignale ist teilweise sehr schlecht. Außerdem ist nicht bekannt, ob die Klangdateien nach der Aufnahme weiter bearbeitet wurden. Vielversprechender wäre die Analyse professionell aufgenommer Klangdateien. Außerdem wäre die Verwendung von Signalen mit einer höheren Abtastrate sinnvoll, denn dadurch kann die spektrale und zeitliche Auflösung erhöht werden. Huynh schlägt in [56] alternative zur Verwendung der herkömmlichen Kurzzeit-Fourier-Transformation, die Verwendung sogenannter Multitapers vor, mit deren Hilfe laut Huynh kleinere Fenster bei gleicher spektraler Auflösung genutzt werden können.
Um in Zukunft die automatische Analyse zu verbessern, damit auch Nutzer ohne Fachkenntnisse optimale Ergebnisse erzielen können, wäre eine Voranalyse der Dateien denkbar. So ist es Momentan noch notwendig, bei rauschbehafteten Signalen die Analyseparameter manuell einzustellen. Wäre der Rauschgehalt des Signals vor der Analyse bekannt, so wäre eine automatische, optimierte Einstellung der Analyseparameter möglich.
Die Voranalyse könnte auch helfen, Dateien hinsichtlich ihrer Charakteristiken, wie zum Beispiel der Abklingzeit, präziser zu analysieren. So haben Gongs teilweise sehr lange Abklingzeiten, hier wäre also ein längerer Analysezeitraum und eine angepasste, weniger „sensilble“ Abklingzeitenbestimmmung von Vorteil. Bei rauschbehafteten Klängen oder Klängen mit sehr vielen Moden hat sich außerdem gezeigt, dass die Berücksichtigung der Phase für die Synthese unumgänglich ist.

Die Modendetektierung könnte in Zukunft folgendermaßen in zwei Schritten erfolgen. Im ersten Schritt würden nur die „echten “ Moden detektiert werden,

um dann im zweiten Schritt die residualen Anteile zu bestimmen. Die „echten “ Moden könnten dann dafür genutzt werden, um Materialparameter zu bestimmen. Die Residualanteile könnten ebenfalls als modales Modell gespeichert werden und als Klangcharakteristik für ein Material und eine Art der Anregung dienen.
Die Extraktion bestimmter Materialparameter, wie der Dämpfung, sollte präzisiert und um nichtlineare Modelle erweitert werden, um auch komplexere Strukturen abbilden zu können.
Nichtlineare Signale konnten in dieser Arbeit nur unzureichend resynthetisiert werden. Daher muss die Analyse in Zukunft an diese Modelle angepasst werden. Eine Möglichkeit der Resynthese bestünde in der vorherigen Erstellung eines nichtlinearen analytischen Modells mit bekannten Parametern. Die spätere Analyse einer Klangdatei eines sehr änlichen Objekts könnte Aufschlüsse über die notwendigen Schritte der Analyse und Resynthese geben.
Die Analyse nichtlinearer Schwinger müsste weiter um eine zeitabhängige Modendetektierung erweitert werden. Dies bedeutet, dass in jedem Zeitschritt das Spektrum auf Moden untersucht werden müsste und diese anschließend einander zugeordnet werden müssten.
In dieser Arbeit wurde angenommen, dass die untersuchten Objekte durch einen Impuls angeregt wurden. Für eine genauere Betrachtung müsste auch diese Annahme präzisiert werden. So sollte in Zukunft die Art Anregung bekannt sein, um den frequenzabhängigen Einfluss der Anregung auf die Amplitude für die spätere Synthese herausrechnen zu können und so das Synthesemodell anregungsunabhängig zu gestalten.

Teil VII

ANHANG

A

ANHANG

A.1 DATEN CD

Die Daten CD beinhaltet folgen Daten:

- Schriftliche Anfertigung der Masterarbeit als PDF Datei *MasterThesis_ PascalKaap*
- Die in der Arbeit verwendete Literatur im Ordner *\Literatur*
- *Citavi* Datei mit weiteren Informationen über die Literatur im Ordner *\Citavi*
- Das *MATLAB* Programm *Mode-A-Lizer* im Ordner *\Mode-A-Lizer* mit der in *MATLAB* ausführbaren Datei *main.m*, über welche das Programm gestartet werden kann.
- Unverarbeitet Audiodateien im Ordner *\Mode-A-Lizer\data\input_ data*
- Die Audiodateien der Synthetisierten Klänge im Ordner *\Mode-A-Lizer\data\output_ data \wav_ data\DATEINAME*
  - Syntheseklänge in der Verarbeitungslänge *addSynth_ DATEINAME*
  - Syntheseklänge in der Originallänge *origSizeAddSynth_ DATEINAME*
  - Syntheseklänge in der Verarbeitungslänge nacht Postprocessing *addSynthFiltered_ DATEINAME*
  - nichtlinearer Syntheseklänge in der Verarbeitungslänge *addSynthNonLin_ DATEINAME*
  - minimalphasiger Originalklang in Originallänge *minPhase_ DATEINAME*
  - minimalphasiger Originalklang in Verarbeitungslänge *mono_ DATEINAME*
- Die *Synth-A-Modeler* Dateien im Ordner *\Mode-A-Lizer\data\output_ data \mdl\DATEINAME*
- Sämtliche generierte Grafiken in *\Mode-A-Lizer\data\output_ data \gfx\DATEINAME*

A.2 EINTEILUNG IN GEHÖRRICHTIGE FREQUENZGRUPPEN

zBark	f_u/Hz	f_o/Hz	Δf_G/Hz	f_m/Hz
0	0	100	100	50
1	100	200	100	150
2	200	300	100	250
3	300	400	100	350
4	400	510	110	450
5	510	630	120	570
6	630	770	140	700
7	770	920	150	840
8	920	1080	160	1000
9	1080	1270	190	1170
10	1270	1480	210	1370
11	1480	1720	240	1600
12	1720	2000	280	1850
13	2000	2320	320	2150
14	2320	2700	380	2500
15	2700	3150	450	2900
16	3150	3700	550	3400
17	3700	4400	700	4000
18	4400	5300	900	4800
19	5300	6400	1100	5800
20	6400	7700	1300	7000
21	7700	9500	1800	8500
22	9500	12000	2500	10500
23	12000	15500	3500	13500
24	15500			

Tabelle 15: Einteilung in gehörrichtige Frequenzgruppen [41]

A.3 BERECHNUNGSVORSCHRIFTEN ANALYTISCHER RESONANZFREQUENZEN

Mit Ausnahme von Gleichung 152 wurden sämtliche Berechnungsvorschriften aus Fletcher [29] entnommen.

A.3.1 *Resonanzfrequenzen Stab*

Für den Stab der an beiden Enden frei ist gilt folgende Gleichung:

$$f_n \quad \frac{\pi K}{8L^2}\sqrt{\frac{E}{\rho}}\left[3.011^2, 5^2, 7^2, \ldots, (2n+1)^2\right]. \tag{139}$$

Für den Stab welcher an einer Seite fest eingespannt und an der anderen Seite frei ist gilt:

$$f_n \quad \frac{\pi K}{8L^2}\sqrt{\frac{E}{\rho}}\left[1.194, 2.988^2, 5^2, \ldots, (2n-1)^2\right]. \tag{140}$$

Für den an beiden Seiten einfach gelagerten Stab gilt:

$$f_n \quad \frac{\pi K}{8L^2}\sqrt{\frac{E}{\rho}}m^2 \quad ; m \quad 1, 2, 3 \ldots. \tag{141}$$

Dabei ist E das E-Modul, L die Länge des Stabes, ρ die Dichte des Stabes und K der Trägheitsradius für den gilt:

$$K \quad \frac{1}{S}\int z^2 dS, \tag{142}$$

mit dem Abstand zur neutralen Faser z und der Fläche S.

A.3.2 *Resonanzfrequenzen vorgespannte Saite*

Für die erste Mode einer Saite gilt:

$$f_1^\circ \quad \sqrt{\frac{T}{\mu}}\frac{1}{2L} \tag{143}$$

wobei μ die Masse pro Länge und T die Spannung der Saite darstellt.
Für die aufgesteckte Saite gilt:

$$f_n \quad nf_1^\circ(1+Bn^2)^{1/2}. \tag{144}$$

Für die eingespannte Saite gilt:

$$f_n \approx nf_1^\circ(1+Bn^2)^{1/2}[1+(2/\pi)B^{1/2}+(4/\pi^2)B], \tag{145}$$

mit $B \quad \pi^2 ESK^2/TL^2$, wobei T die Spannung ist.

A.3.3 *Resonanzfrequenzen vorgespannte rechteckige Platte*

Für die vorgespannte rechteckige Platte gilt:

$$f_m n \quad \frac{1}{2}\sqrt{\frac{T}{\sigma}}\sqrt{\frac{m^2}{L_x^2}+\frac{n^2}{L_y^2}} \quad ; m, n \quad 1, 2, \ldots. \tag{146}$$

mit der flächenspezifischen Dichte σ.
Für die einfach gelagerte rechteckige Platte gilt:

$$f_m n \quad 0.453 c_L h\left[\left(\frac{m+1}{L_x}\right)^2+\left(\frac{n+1}{L_y}\right)^2\right] \quad ; m, n \quad 0, 1, 2, \ldots. \tag{147}$$

A.3.4 *Resonanzfrequenzen runde Platte*

Die Resonanzfrequenzen der runden eingespannte Platte können die Resonanzfrequenzen folgender Tabelle entnommen werden.

$f_{01} = 0.4694 c_L h/a^2$	$f_{11} = 2.08 f_{01}$	$f_{21} = 3.41 f_{01}$	$f_{31} = 5.00 f_{01}$	$f_{41} = 6.82 f_{01}$
$f_{02} = 3.89 f_{01}$	$f_{12} = 5.95 f_{01}$	$f_{22} = 8.28 f_{01}$	$f_{32} = 10.87 f_{01}$	$f_{42} = 13.71 f_{01}$
$f_{03} = 8.72 f_{01}$	$f_{13} = 11.75 f_{01}$	$f_{23} = 15.06 f_{01}$	$f_{33} = 18.63 f_{01}$	$f_{43} = 22.47 f_{01}$

Tabelle 16: Resonanzfrequenzen eingespannte Rundplatte [29]

Es folgen die Resonanzfrequenzen der runden Platte mit freien Enden.

–	–	$f_{20} = 0.2412 c_L h/a^2$	$f_{30} = 2.328 f_{20}$	$f_{40} = 4.11 f_{20}$	$f_{50} = 6.30 f_{20}$
$f_{01} = 1.73 f_{20}$	$f_{11} = 3.91 f_{20}$	$f_{21} = 6.71 f_{20}$	$f_{31} = 10.07 f_{20}$	$f_{41} = 13.92 f_{20}$	$f_{51} = 18.24 f_{20}$
$f_{02} = 7.34 f_{20}$	$f_{12} = 11.40 f_{20}$	$f_{22} = 15.97 f_{20}$	$f_{32} = 21.19 f_{20}$	$f_{42} = 27.18 f_{20}$	$f_{52} = 33.31 f_{20}$

Tabelle 17: Resonanzfrequenzen Rundplatte mit freien Enden[29]

Und die Resonanzfrequenzen der runden einfach gelagerten Platte.

$f_{01} = 0.2287 c_L h/a^2$	$f_{11} = 2.80 f_{01}$	$f_{21} = 5.15 f_{01}$
$f_{02} = 5.98 f_{01}$	$f_{12} = 9.75 f_{01}$	$f_{22} = 14.09 f_{01}$
$f_{03} = 14.91 f_{01}$	$f_{13} = 20.66 f_{01}$	$f_{23} = 26.99 f_{01}$

Tabelle 18: Resonanzfrequenzen einfach gelagerte Rundplatte [29]

Dabei ist a der Radius und $c_L \quad \sqrt{\frac{E}{\rho(1-\nu^2)}}$ die Longitudinalwellenlänge mit der Dichte ρ und dem *POISSON*- Modul ν.

A.3.5 *Resonanzfrequenz elliptische Platte*

Für die Resonanzfrequenzen der elliptischen Platte gilt:

$$f \approx \frac{0.291 c_L h}{a^2} \sqrt{1 + \frac{3}{2}\left(\frac{a}{b}\right)^2 + \left(\frac{a}{b}\right)^4}. \tag{148}$$

Mit der Hauptachse a und der Nebenachse b.

A.3.6 *Resonanzfrequenz Holzstrukturen*

Da Holz anders als zum Beispiel Metall, Richtungsabhängige molekulare Strukturen aufweist, sind die Materialparameter ebenfalls von der Koordinatenrichtung abhängig. Daher gilt für die hölzerne rechteckige Platte:

$$f_{mn} \quad 0.453 \left[c_x \left(\frac{m+1}{L_x}\right)^2 + c_y \left(\frac{n+1}{L_y}\right)^2 \right] \quad ; m, n \quad 0, 1, 2, \ldots. \tag{149}$$

Mit den richtungsabhängigen Longitudialwellenausbreitungsgeschwindigkeiten:

$$c_x \quad \sqrt{E_x/\rho(1-\nu_{xy}\nu_{yx})} \quad \text{und} \quad c_y \quad \sqrt{E_y/\rho(1-\nu_{xy}\nu_{yx})}. \tag{150}$$

Für den schmalen an beiden Enden freien Holzbalken gilt:

$$f_n \quad \frac{0.113h}{L^2}\sqrt{c_x c_y}\sqrt{\frac{E}{\rho}}\left[3.0112^2, 5^2, 7^2, \ldots, (2n+1)^2\right]. \tag{151}$$

A.3.7 *Resonanzfrequenz Zylinderschale*

Nach Soedel [49] ergibt sich für die Frequenzen eines Zylinders folgende Gleichung:

$$\omega_{n,m} \quad \frac{1}{a}\sqrt{\frac{(m\pi/L)^4}{[(m\pi a/L)^2+n^2]^2}+\frac{(h/a)^2}{12(1-\mu^2)}[\left(\frac{m\pi a}{L}\right)^2+n^2]^2}\sqrt{\frac{E}{\rho}} \tag{152}$$

A.3.8 *Resonanzfrequenz Rundmembran*

Für die Resonanzfrequenz der Rundmembran gilt folgende Tabelle.

$f_{01} = \frac{2.405}{2\pi a}\sqrt{\frac{T}{\sigma}}$	$f_{11} = 1.594 f_{01}$	$f_{11} = 2.136 f_{01}$	$f_{11} = 2.296 f_{01}$	$f_{11} = 2.653 f_{01}$	$f_{11} = 2.918 f_{01}$
$f_{11} = 3.156 f_{01}$	$f_{11} = 3.501 f_{01}$	$f_{11} = 3.600 f_{01}$	$f_{11} = 3.652 f_{01}$	$f_{11} = 4.060 f_{01}$	$f_{11} = 4.154 f_{01}$

Tabelle 19: Rundmembran [29]

Mit der flächenspezifischen Dichte σ.

A.4 USER INTERFACE GUIDE FOR MODE-A-LIZER

In the following section a guide for the use of Mode-A-Lizer is presented. The user interface is divided into tabs that are arranged chronologically (see figure 79). If the user wishes to develop a model manually, he can perform the individual steps for each tab from left to right. The user interface has a total of five tabs. The first step is for audiopreprocessing, further a spectral analysis is performed. Then the modes of the signal are detected. The next step is to determine the amplitude and the decay time. In the last step the user can continue to trim modes based on psychoacoustic models. All tabs are arranged as follows: Controls and additional information are listed on the left, while graphical analysis results are shown on the right.

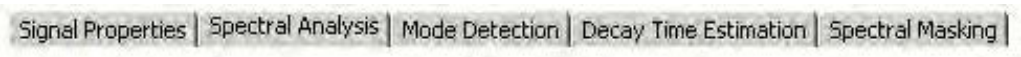


Abbildung 79: Tab Overview

A.4.1 *Audiopreprocessing*

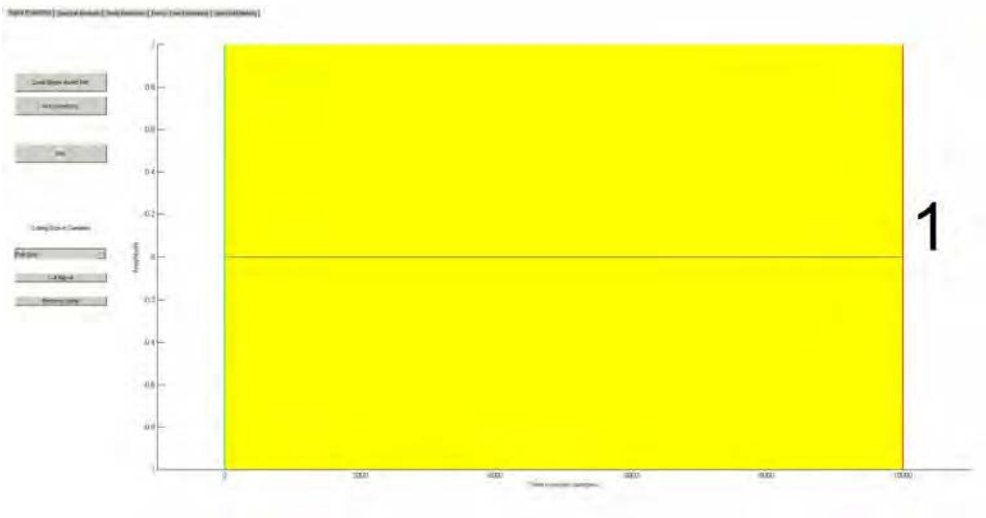


Abbildung 80: Tab Audiopreprocessing

In the first tab „Audiopreprocessing“ the user has the possibility to load an audio file using the button „Load Single Audio File“ (see figure 81 (1)). The file is automatically phase corrected upon loading. The file name is shown in the top window bar. The signal graph is then plotted in the window (see figure 80 (1)). If the user wishes to conduct an automated analysis, he has the option to load multiple audio files by clicking „Auto Analysis“ (see figure 81 (2)). The analysis and synthesis is then carried out with the default settings. All options of the following tabs can be changed in advance and will be taken into account during the automatic analysis. If the user selects the manual analysis of the audio file, so he can play the loaded file using the buttons play „Play“ (see figure 81 (3)).

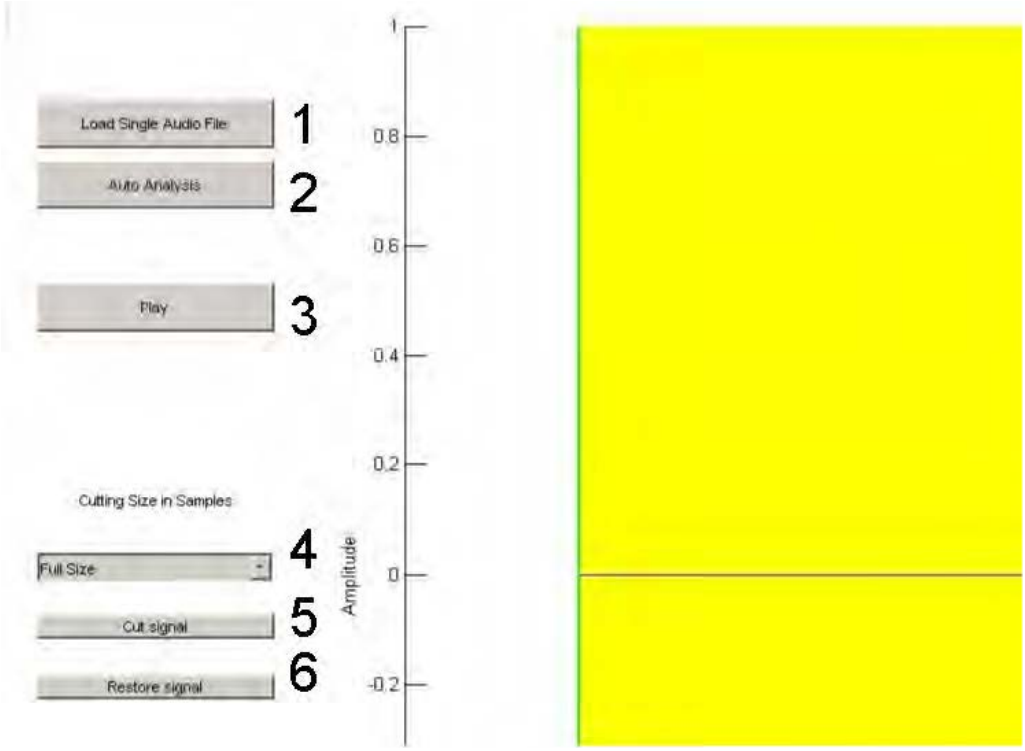


Abbildung 81: UI Controls Audiopreprocessing

In the signal window the user has the option to trim the file with the help of the right and left slider (see figure 82 ① ②). The moveable sliders allow selecting certain signal sections and choosing a particular part of the signal for processing. Using the pop-up menus „Cutting in Samples“ (see figure81 ④), the signal length can be cut into predefined lengths. If the signal is less than 32768 samples, it is recommended to cut the signal to 16384 samples. This step is optional, since the signal length is automatically reduced to a maximum of 32768 samples and minimum of 8192 samples during the „Spectral Analyses“ step depending on the original signal length, if no trim was performed before. The user can manually cut the signal by first selecting a signal length and pressing the button „Cut Signal“ (see figure 81 ⑤to cut the signal. Using the button „Restore signal“ (see figure 81 ⑥), the original signal is restored.

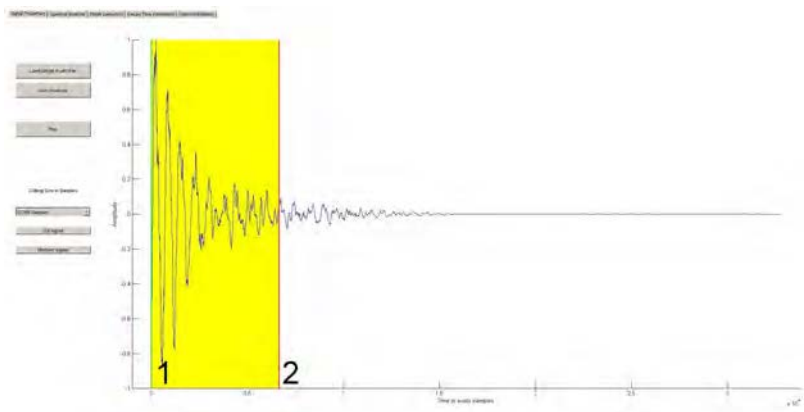


Abbildung 82: Slider Audiopreprocessing

A.4.2 *Spectral Analysis*

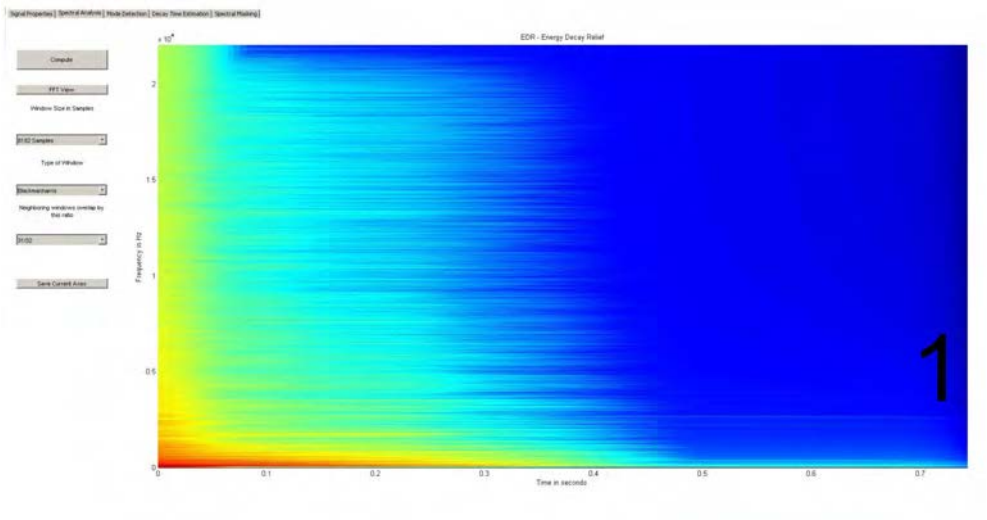


Abbildung 83: Tab Spectral Analysis

The following tab „Spectral Analysis“ is used for the spectral analysis of the audio signal. The default values are generally well suited, so an inexperienced user may not need to change them. By pressing the buttton „Calculate“ (see figure 84 ①) the spectral analysis is calculated. Once done, the Energy Decay Relief is displayed on the screen (see figure 83 ①). Red areas represent high amplitude values while blue areas represent low amplitude values. The time course is shown on the X-axis, frequency on the Y-axis.
The user can press the button „FFT View“ (see figure 84 ②), to show a spectogramm instead of the Energy Decay Relief. Using the pop-up menu „Window Size in Samples“ (see figure 84 ③), the user can resize the window. If a smaller window is chosen, the spectral resolution decreases as the temporal resolution increases. If the window length is increased, the opposite effect occurs. The window function can be changed via the pop-up menu „type of Window“ (see figure 84 ④). For this work, it makes sense to make use of the Hann window, which is also set by default. If signals with less than 500 modes are processed, the use of the Blackmann-Harris window is recommended. Using the pop-up menus „Neighboring windows overlap by this ratio“ (see figure 84 ⑤), the size of the overlap can be changed. For a large ratio, such as 1/2, the temporal resolution is very low, but the computing time and volume of data are much lower. The adjusted value of 31/32 is a good combination of computation time and temporal resolution. Once the desired spectral analysis has been carried out, the user can switch to the next tab. Saving and exporting of the EDR or FFT window is possible by pressing the button „Save Current Axes“ (see figure 84 ⑥).

A.4.3 *Mode Detection*

The tab „Mode Detection“ is used to determine the modes or resonance frequencies of the analyzed signal. Using the information from the EDR, the frequencies with the highest energies can be selected as mode frequencies. At the

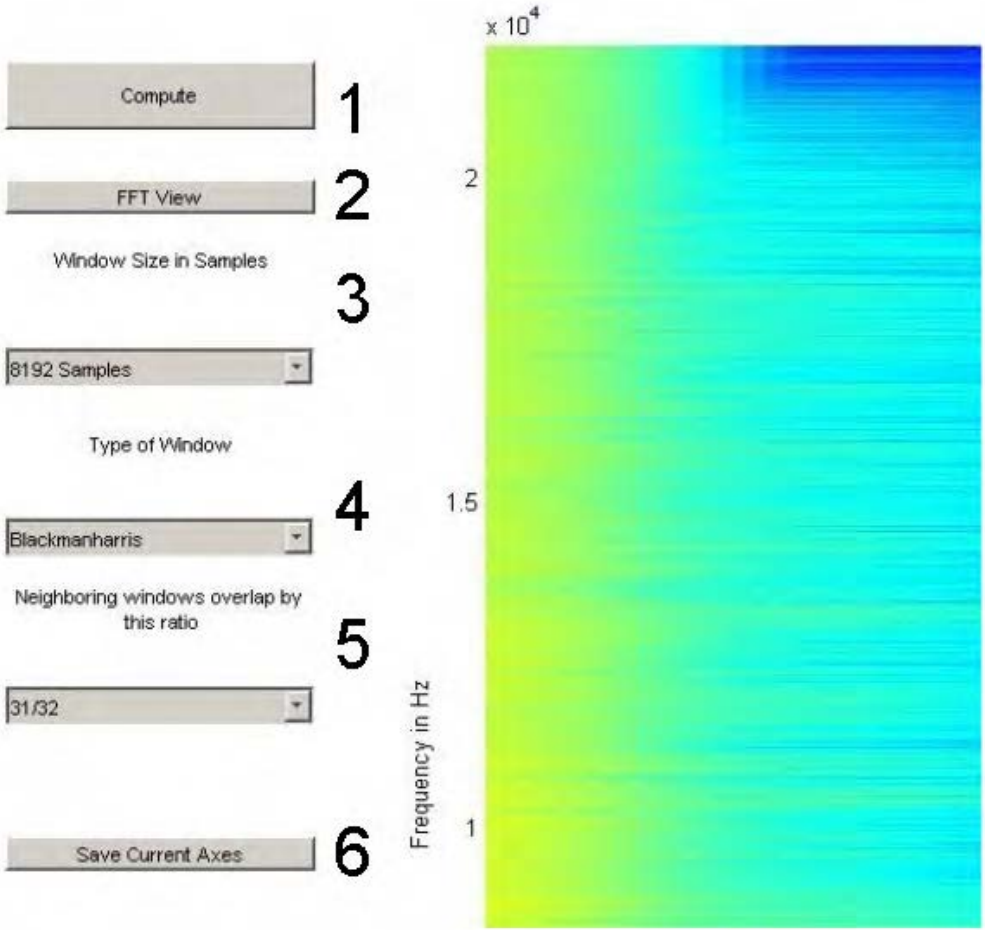


Abbildung 84: UI Controls Spectral Analysis

top of the graphics window (see figure 85 ①), the EDR of the original signal can be seen. In order to compare the original signal with the additive synthesized signal, an EDR of this synthesized signal is displayed in the lower window (see figure 85 ②).

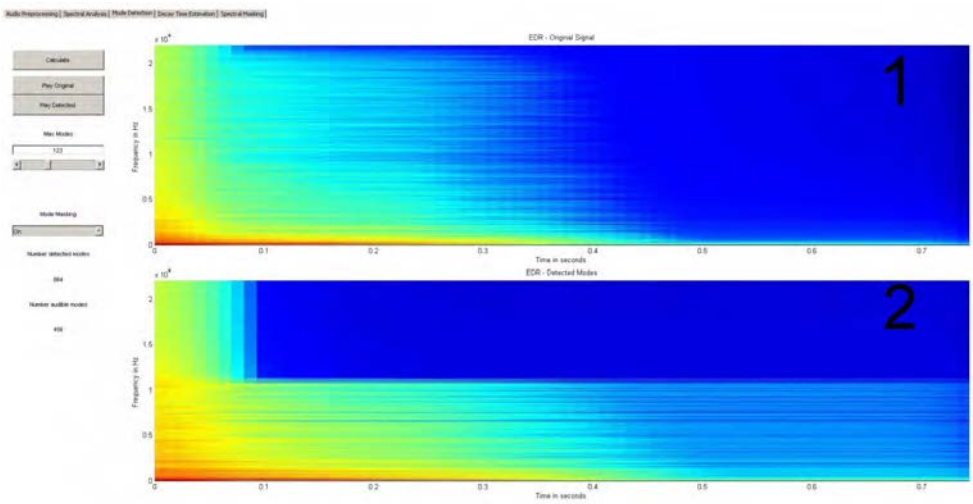


Abbildung 85: Tab Mode Detection

After pressing the button „Calculate“ (see figure 86 ①) the mode detection starts. Subsequently, the EDR of the detected signal appears in the lower graphic window. The original signal can be played back by pressing the button „Play Original“ (see figure 86 ②). The detected signal is played back by pressing the „Play Detected“ (see figure 86 ③) button. The repeated plays of the signal gives a first impression of the spectral distribution. The user can limit the number of modes by changing the number in the text box „Max Modes“, or via the slider below(see figure 86 ④). By default, the modes are trimmed based on psychoacoustic models. It is deactivatable via „Mode Masking“ (see figure 86 ⑤) popup menu. If a parameter is changed, the button „Calculate“ must be pressed again for these changes to be considered. The number of detected modes is displayed in the text field „Number of detected modes“ (see figure 86 ⑥). If the modes have been trimmed, the number of audible modes is displayed in the text box „Number audible modes“ (see figure 86 ⑦). If no trimming has been performed, this text field displays the same number, as in the upper one.

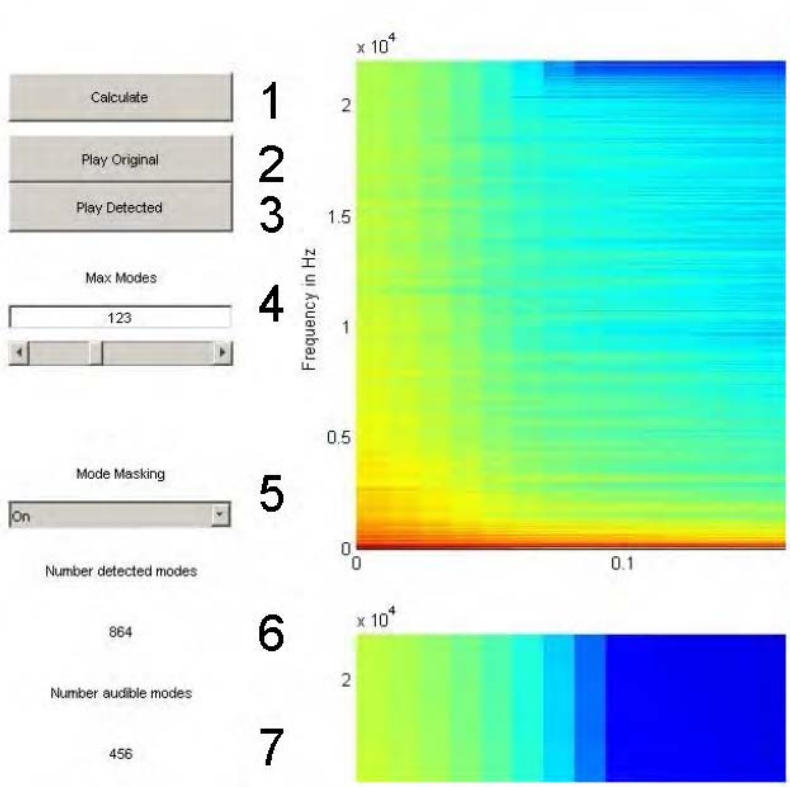


Abbildung 86: Mode Detection UI Controls

A.4.4 *Mode Properties Detection*

In the tab „Mode Properties Detection“ the amplitude and the decay time of the detected modes are determined. The top window shows the logarithmic amplitude curve of the EDR, shown as solid lines (see figure 87 ①). The approximated curves, which have been calculated by linear regression, are shown in the same color, but as dashed lines.In the lower window, the solid lines represent the amplitude curves which have been calculated by the short-time Fourier

transformation(see figure 87 ②). The approximate curves which are determined by the calculated decay time and amplitude, are shown in the same color by dashed lines.

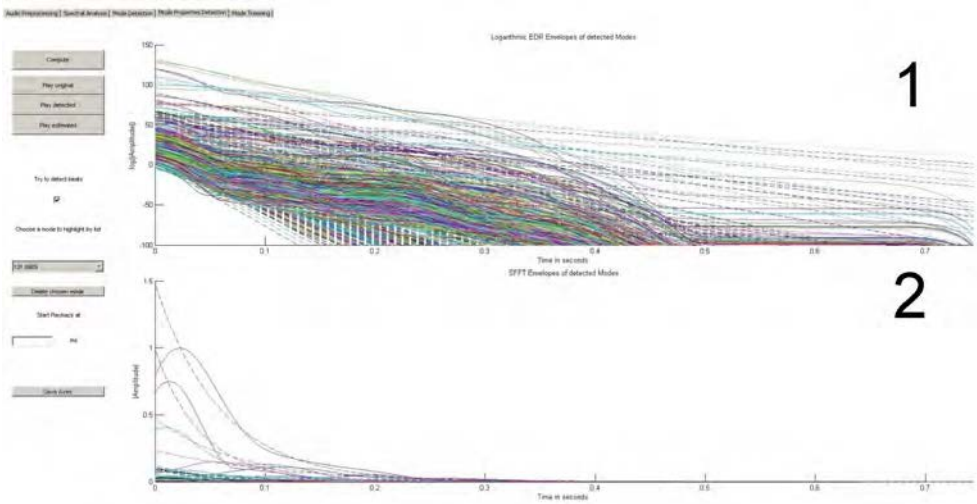


Abbildung 87: Tab Mode Properties Detection

By pressing the „Calculate“ button (see figure 88 ①) the modal properties are calculated. With the key „Play Original“ (see figure 88 ②), the original signal can be played with the key „Play detected“ (see figure 88 ③), the signal of the detected modes can be played. By pressing the key „Play estimated “ (see figure 88 ④) the signal of the detected modes include the calculated decay times and amplitudes is played. If the checkbox „Try to detect beats“ (see figure 88 ⑤) is activated, the algorithm tries to find beats and tries to synthesize them with the help of the model discussed above. This option is disabled by default. By clicking on an amplitude curve, the concerned mode is highlighted in green color. The modes can be selected in both, the upper and the lower window. Further, an amplitude curve can be selected via the pop up menu „Choose a mode to highlight by list“(see figure88 ⑥), too. The selected mode can be removed from the model by pressing the button „delete Chosen mode“ (see figure 88 ⑦). Deleting a Mode can be helpfull, if you want to remove noise for example. If it is desired to start the playback at a certain time, the time can be entered in text box „start playback at“ (see figure 88 ⑧). By default the playback starts at 0ms. The selected amplitude curves can be saved by pressing the „save Axe“ button (see figure 88 ⑨).

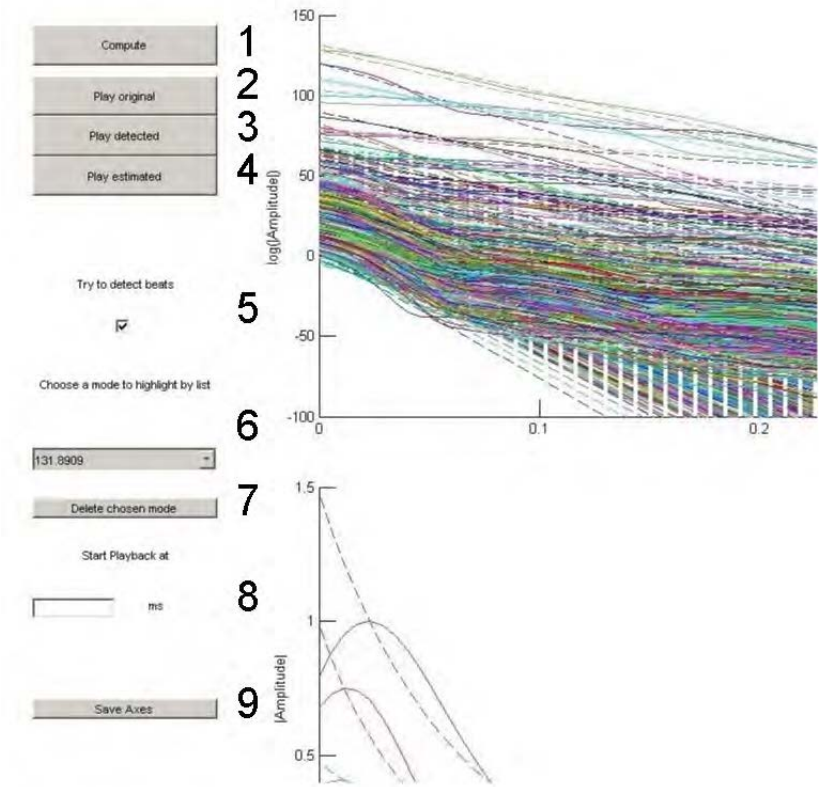


Abbildung 88: Mode Properties Detection UI Controls

A.4.5 *Mode Trimming*

In the final step, further modes can be trimmed.

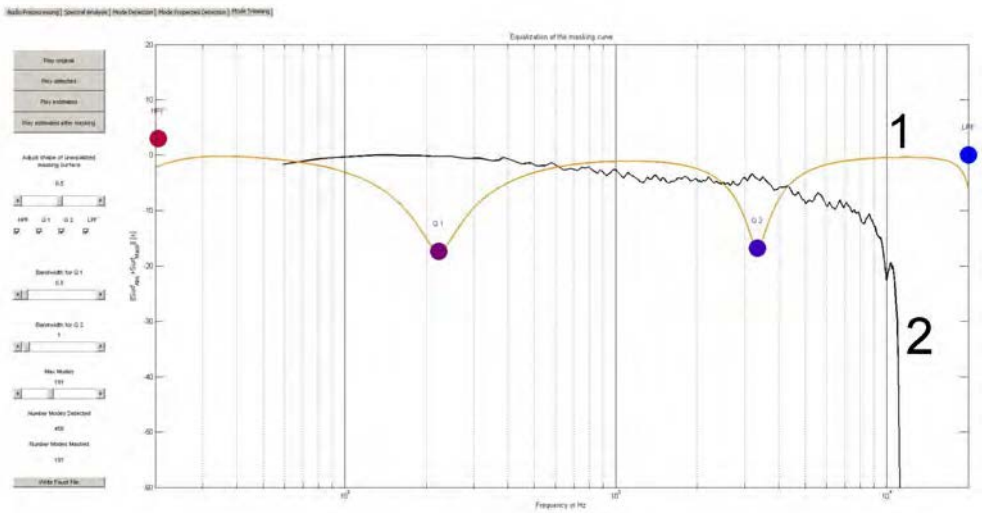


Abbildung 89: Tab Postprocessing

In the window two curves are shown (see figure 89). An orange freely adjustable filter curve (see figure 89 ①), which is used for frequency-dependent truncation of modes. The four marker „HPF“, „Q1“, „Q2“ and „LPF“, are used to adjust the curve. They can be moved freely across the window surface. The marker „LPF“ (see figure 90 ①) can be used to adjust the cutoff frequency of

a low pass filter. The marker „Q1“ and „Q2“ (see figure 90 (2) (3)) are the filter coefficients of a peak filter. The point „HPF“ (see figure 90 (4)) the cutoff frequency of a high pass filter. In each case, the Y-axis controls the gain parameter of the node. The black curve (see figure 89 (2)) is calculated from the area between amplitude level curve and the absolute hearing threshold level or masking level.

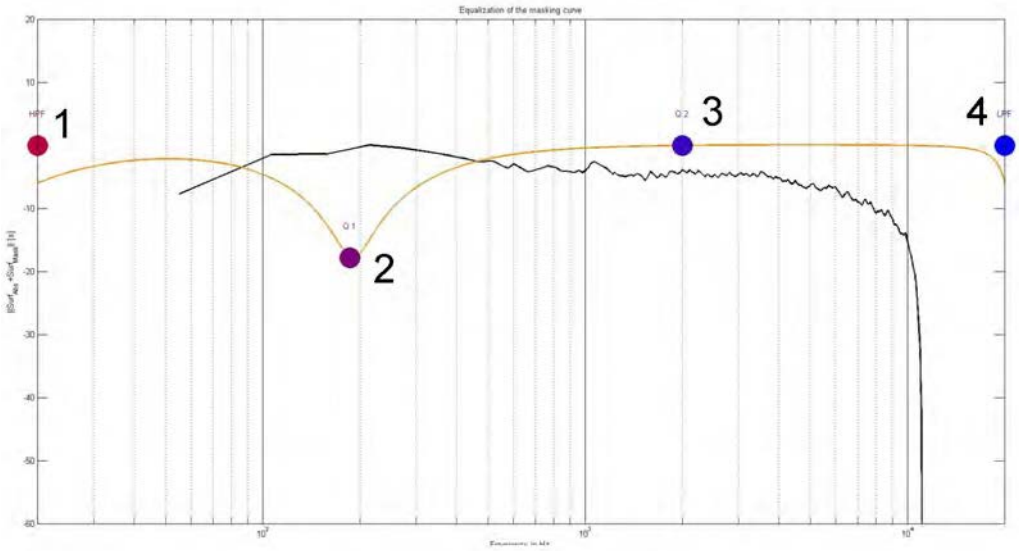


Abbildung 90: Postprocessing Axes

The first three buttons to play the audio files, have the same function as in the previous tab (see figure 90 (1) (2) (3)). The button „Play estimated after masking“ (see figure 91 (4)) plays the signal after filtering according to the currently set parameter values. With the help of the text box and the slider „Max Number Modes“ (see figure 91 (1)) the desired maximum number of modes can be set. Underneath the slider „Adjust shape of unequalized masking surface “(see figure 91 (1)) can be used to create the black curve. This is calculated as described above from the sum of area between between amplitude level curve and the absolute hearing threshold level or the masking level. With the slider you can adjust the ratio of the sum of the two surfaces. If the slider value is set to 0.5, the sum is calculated from equal parts of the two surfaces. If it is set to 0, the sum is calculated from of the area between the amplitude curve and masking level, at 1 the sum is calculated from the area between amplitude level curve and the absolute hearing thresshhold level.
The checkboxes „HPF“, „Q1“, „Q2“ and „LPF“ activate or deactivate the filter node(see figure 91 (1)). All filters are active by default. The two sliders below adjust the bandwidth of the peaking EQ filters „Q1“ and „Q2“ (see figure 91 (1)). In the text field „Number audible mode“(see figure 91 (1)) the audible modes are displayed, the modes, which remain after trimming are displayed in the text box „Number modes after trimming“ (see figure 91 (1)). With the last button „Write Faust File“ (see figure 91 (1)) the automatically generated *Synth-A-Modeler* file with the model created in the tab „Mode Properties“ will be overwritten with the model created in this tab.

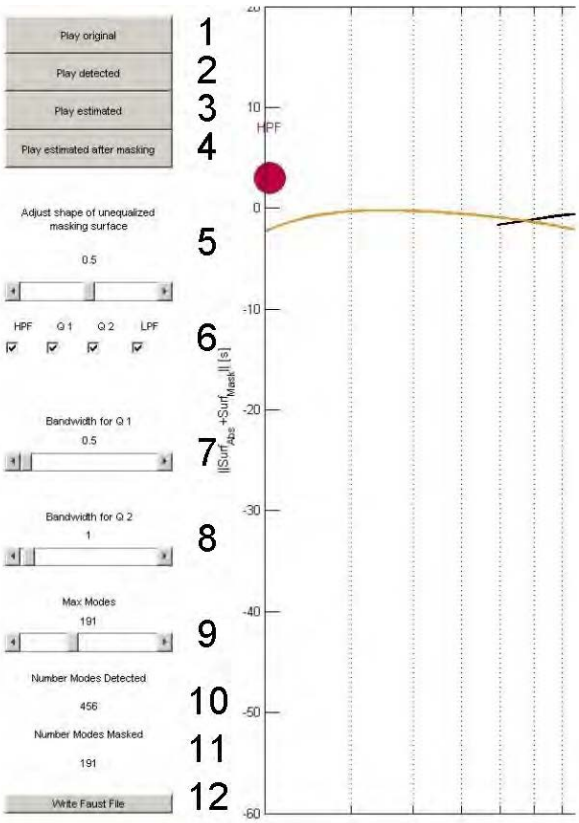


Abbildung 91: Postprocessing UI Controls

A.4.6 *Supplementary notes*

If a model does not meet the expectations, certain parameter changes in the program can significantly improve the result. If Signals with many of noiselike components are processed (E.g. hi-hats and snares), spectral masking should be disabled. Subsequently, problematic modes can be removed via the „Mode Mrimming“ tab. If beats are audible in the original signal, the user can activate the checkbox „Try to detect beats“. In case a constant beat frequency and amplitude is present, the resynthesized signal will be accurate.

LITERATURVERZEICHNIS

[1] L. Rayleigh, *The Theory of Sound, Vol. I.* MacMillan and Company Ltd., London, 1894.

[2] C. Cadoz, A. Luciani, J.-L. Florens, *et al.*, "Synthèse musicale par simulation des mécanismes instrumentaux, transducteurs gestuels rétroactifs pour l'étude du jeu instrumental," *Revue d'acoustique*, no. 59, pp. 279–292, 1981.

[3] J. O. Smith, "Physical modeling using digital waveguides," *Computer Music Journal*, vol. 16, no. 4, pp. 74–91, 1992.

[4] J. O. Smith, "Physical modeling synthesis update," *Computer Music Journal*, vol. 20, no. 2, pp. 44–56, 1996.

[5] J. O. Smith, *Physical Audio Signal Processing*. http://ccrma.stanford.edu/~jos/pasp/, accessed <23.01.14, 09:00>. online book.

[6] S. Hou, "Review of modal synthesis techniques and a new approach," *Shock and vibration bulletin*, vol. 40, no. 4, pp. 25–39, 1969.

[7] R. M. Hintz, "Analytical methods in component modal synthesis," *AIAA Journal*, vol. 13, no. 8, pp. 1007–1016, 1975.

[8] Z. Zhao-chang, "Dynamic analysis of nonlinear systems by modal synthesis techniques," *Applied Mathematics and Mechanics*, vol. 4, no. 4, pp. 611–623, 1983.

[9] V. Valimaki, T. Tolonen, and M. Karjalainen, "Plucked-string synthesis algorithms with tension modulation nonlinearity," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 2, pp. 977–980, IEEE, 1999.

[10] T. Tolonen, C. Erkut, V. Välimäki, and M. Karjalainen, "Simulation of plucked strings exhibiting tension modulation driving force," in *Proc. Int. Computer Music Conf.(Beijing, China, October 1999)*, pp. 5–8, 1999.

[11] M. Karjalainen, "New techniques and effects in model-based sound synthesis," in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, pp. 9–11, 1999.

[12] C. Erkut, M. Karjalainen, P. Huang, and V. Välimäki, "Acoustical analysis and model-based sound synthesis of the kantele," *The Journal of the Acoustical Society of America*, vol. 112, p. 1681, 2002.

[13] R. Marogna and F. Avanzini, "Physically-based synthesis of nonlinear circular membranes," in *Proc. Int. Conf. Digital Audio Effects (DAFx-09)*, pp. 373–379, 2009.

[14] S. Bilbao, "Modal type synthesis techniques for nonlinear strings with an energy conservation property," in *Proc. Int. Conf. Digital Audio Effects (DAFx-04)*, pp. 119–124, 2004.

[15] N. Ellis, J. Bensoam, and R. Caussé, "Modalys demonstration," in *Proc. of the 2005 International Computer Music Conference (ICMC05)*, pp. 101–102, 2005.

[16] C. Bruyns, "Modal synthesis for arbitrarily shaped objects," *Computer Music Journal*, vol. 30, no. 3, pp. 22–37, 2006.

[17] S. Bilbao, *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics*. Wiley, Dec. 2009.

[18] D. B. Lloyd, N. Raghuvanshi, and N. K. Govindaraju, "Sound synthesis for impact sounds in video games," in *Symposium on Interactive 3D Graphics and Games*, pp. PAGE–7, ACM, 2011.

[19] Z. Ren, H. Yeh, R. Klatzky, and M. C. Lin, "Auditory perception of geometry-invariant material properties," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 19, no. 4, pp. 557–566, 2013.

[20] Z. Ren, H. Yeh, and M. C. Lin, "Example-guided physically based modal sound synthesis," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 1, p. 1, 2013.

[21] F. Avanzini and R. Marogna, "A modular physically based approach to the sound synthesis of membrane percussion instruments," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 4, pp. 891–902, 2010.

[22] E. Berdahl and J. Smith, "An introduction to the synth-a-modeler compiler: Modular and open-source sound synthesis using physical models," in *Proceedings of the Linux Audio Conference*, 2012.

[23] H. Irretier, "Schwingungstechnik," *Uni Kassel Institut für Mechanik*, 2005.

[24] W. Nolting, "Grundkurs: Theoretische physik, 1 klassische mechanik, 2 analytische mechanik," 1993.

[25] M. Knaebel, H. Jäger, and R. Mastel, *Technische Schwingungslehre*. Springer DE, 2009.

[26] C. Beerens, "Zur modellierung nichtlinearer därnpfungs—phänomene in der strukturrnechanik," 1994.

[27] A. K. Chopra, *Dynamics of structures: Theory and applications to earthquake engineering*. Prentice Hall Saddle River^ eNY NY, 2001.

[28] K. W. Möser, M., *Körperschall*. Springer, 2009.

[29] N. H. Fletcher and T. D. Rossing, *The physics of musical instruments*. Springer Verlag, 1998.

[30] K. M. Chew, R. Sudirman, N. Seman, and C. Y. Yong, "Relaxation frequency and relaxation time estimation for phantom modeling by proposed fitting linear models," in *Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on*, pp. 911–915, IEEE, 2012.

[31] N. H. Fletcher, "Analysis of the design and performance of harpsichords," *Acustica*, vol. 37, no. 2, pp. 139–147, 1977.

[32] E. Kahle and J. Woodhouse, "The influence of cell geometry on the elasticity of softwood," *Journal of materials science*, vol. 29, no. 5, pp. 1250–1259, 1994.

[33] M. McIntyre and J. Woodhouse, "On measuring the elastic and damping constants of orthotropic sheet materials," *Acta Metallurgica*, vol. 36, no. 6, pp. 1397–1416, 1988.

[34] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. FreeTech Books, 2003.

[35] J.-M. Jot, L. Cerveau, and O. Warusfel, "Analysis and synthesis of room reverberation based on a statistical time-frequency model," in *Audio Engineering Society Convention 103*, Audio Engineering Society, 1997.

[36] M. Gasior and J. Gonzalez, "Improving fft frequency measurement resolution by parabolic and gaussian interpolation," tech. rep., CERN-AB-Note-2004-021, 2004.

[37] J. O. Smith, *Spectral audio signal processing*. W3K Pub., 2007.

[38] J. O. Smith, *Introduction to Digital Filters with Audio Applications*. http://-ccrma.stanford.edu/~jos/filters/, 20.02.2014 - 14.00. online book.

[39] R. FELDTKELLBR and I. E. ZWICKER, "Das ohr als nachrichtenempfänger," 1956.

[40] H. Fastl and E. Zwicker, *Psychoacoustics: facts and models*, vol. 22. Springer, 2007.

[41] U. Zölzer, *Digital audio signal processing*. Wiley. com, 2008.

[42] J. W. Beauchamp, "Analysis and synthesis of musical instrument sounds," in *Analysis, Synthesis, and Perception of musical sounds*, pp. 1–89, Springer, 2007.

[43] M. Möser, *Technische Akustik*. Springer, 2009.

[44] M. Möser, "Theoretische akustik," *Vorlesungsscript, Institut für Technische Akustik, Technische Universität Berlin*, 1991.

[45] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

[46] L. Fahrmeir, T. Kneib, and S. M. Lang, *Regression*. Springer DE, 2007.

[47] C. Gerthsen and D. Meschede, *Gerthsen physik*. Springer DE, 2010.

[48] M. Schmidt, G. Fung, and R. Rosales, "Fast optimization methods for l1 regularization: A comparative study and two new approaches," in *Machine Learning: ECML 2007*, pp. 286–297, Springer, 2007.

[49] W. Soedel, *Vibrations of shells and plates*, vol. 177. CRC Press, 2004.

[50] D. A. Florencio, "On the use of asymmetric windows for reducing the time delay in real-time spectral analysis," in *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on*, pp. 3261–3264, IEEE, 1991.

[51] D. A. Florencio, "Investigating the use of asymmetric windows in celp vocoders," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, vol. 2, pp. 427–430, IEEE, 1993.

[52] R. Rozman and D. M. Kodek, "Using asymmetric windows in automatic speech recognition," *Speech communication*, vol. 49, no. 4, pp. 268–276, 2007.

[53] J. A. Morales-Cordovilla, V. Sánchez, A. M. Gómez, and A. M. Peinado, "On the use of asymmetric windows for robust speech recognition," *Circuits, Systems, and Signal Processing*, vol. 31, no. 2, pp. 727–736, 2012.

[54] M. Zivanovic and A. Carlosena, "On asymmetric analysis windows for detection of closely spaced signal components," *Mechanical systems and signal processing*, vol. 20, no. 3, pp. 702–717, 2006.

[55] A. Volker, "Signalverarbeitung eine einführung," *Vorlesungsscript,Heinrich-Heine Universität Düsseldorf*, 2010.

[56] D. Q. Huynh, "Frequency estimation of musical signals using STFT and multitapers," in *Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on*, pp. 34–39, IEEE, 2009.

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt hab;

die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungskommission vorgelegt und auch nicht veröffentlicht.

Berlin, Februar 2014

Pascal Kaap