Technische Universität Berlin

Institut für Sprache und Kommunikation
Fachbereich Audiokommunikation

Generalisierte, grundtonabhängige Modelle für quasi-harmonische Instrumente

Magisterarbeit

4. Juni 2010

Vorgelegt von:

Henrik Hahn

[redacted]

Geburtsdatum : 29.01.1980

Erstkorrektor : Prof. Dr. Stefan Weinzierl
Zweitgutachter : Dr. Ing. Axel Röbel (IRCAM)

Eidesstattliche Erklärung

Die selbständige und eigenhändige Ausfertigung versichert an Eides statt
Berlin, den 4. Juni 2010

_____ Unterschrift

Kurzfassung

Das Ziel dieser Arbeit ist die Entwicklung eines Quellenmodells zur Repräsentation der Timbre quasi-harmonischer Instrumente. Das Modell soll dazu die zeitveränderliche, spektrale Verteilung der Energie der Klänge eines Instruments repräsentieren und reproduzieren können. Der Ansatz umfasst dabei ein statistisches, parametrisches Quelle-Filter-Modell sowie Methoden zum Schätzen der Parameter anhand einer Trainingsdatenbasis. Die Quelle stellt dabei das klangerzeugende und das Filter das klangfärbende Element dar. Der Ansatz wird anhand einer Reihe ausgewählter Instrumente evaluiert, wozu sowohl die Prototypen und ihre statistischen Eigenschaften gezeigt werden als auch eine Auswahl auf Basis der Prototypen reproduzierter Klänge.

Quellenmodelle dienen dazu, in Form von a priori Wissen für signaladaptive Verfahren zur Instrumentenklassifikation oder Quellentrennung, die Genauigkeit der Klassifikation oder Qualität der Trennung zu verbessern.

Inhaltsverzeichnis

Eidesstattliche Erklärung **i**

Kurzfassung **ii**

1 Einführung **1**

2 Grundlagen **4**

2.1 Akustische Wahrnehmung 4

2.1.1 Lautstärkewahrnehmung 4

2.1.2 Tonhöhenwahrnehmung 5

2.2 Akustik der Musikinstrumente 5

2.2.1 Die spektrale Dimension 6

2.2.2 Die Dynamikdimension 6

2.2.3 Die zeitliche Dimension 7

2.3 Diskrete Signalrepräsentationen 8

2.3.1 Der Zeitbereich 8

2.3.2 Der Frequenzbereich 9

2.3.3 Die Kurzzeit-Spektralanalyse 11

2.3.4 Das Sinusoidalmodell 13

2.4 Maschinelles Lernen 15

2.4.1 Paradigmen des Lernens 15

2.4.2 Kosten und Gradientenabstieg 16

2.4.3 Stochastische Approximation 18

2.4.4 Anpassung der Schrittweite 19

2.4.5 Modellselektion 19

3 Stand der Forschung **22**
3.1 Modell von Juan José Burred 23
3.2 Modell von Anssi Klapuri 24

4 Das Modell **27**
4.1 Das Signalmodell 27
4.1.1 Approximation der Frequenzwerte der Partialwellen . . 28
4.1.2 Skalierung der Amplituden 28
4.1.3 Teilung der spektralen Hüllkurve 29
4.2 Das Quelle-Filter-Modell 33
4.2.1 Die Quelle 35
4.2.2 Das Filter 37
4.2.3 Die Quelle-Filter-Summen 39
4.2.4 Projektion der Eingangsdaten 40
4.2.5 Prädiktion 41
4.2.6 Synthese 41
4.2.7 Modellvarianz 44
4.2.8 Modellkomplexität 44
4.2.9 Mathematische Notationskonventionen 45
4.3 Schätzen der Modellparameter 45
4.3.1 Kostenfunktion 46
4.3.2 Gradientenfunktionen 47
4.3.3 Algorithmen 49

5 Implementierung **52**
5.1 Aufbau der Software 54
5.1.1 Der Import-Prozess 57
5.1.2 Der Analyse-Prozess 58
5.1.3 Der Trainings-Prozess 59
5.1.4 Der Statistik-Prozess 60
5.1.5 Der Synthese-Prozess 61
5.2 Verwendung der Software 62
5.2.1 Definition der Datenbibliothek 63

5.2.2 Definition der Parameter 63
5.2.3 Programmaufruf 64

6 Evaluation **65**
6.1 Die Datenbank 65
6.2 Analyse der Daten 67
6.3 Modelltraining 68

7 Ergebnisse **72**
7.1 Klarinette 75
7.1.1 Analyse der Daten 75
7.1.2 Prototypen 76
7.1.3 Synthese 78
7.2 Flügel 79
7.2.1 Analyse der Daten 79
7.2.2 Prototypen 80
7.2.3 Synthese 81
7.3 Violine 82
7.3.1 Analyse der Daten 82
7.3.2 Prototypen 83
7.3.3 Synthese 85

8 Zusammenfassung **87**

A Instrumente Teil 1 - Ergänzung **90**

B Instrumente Teil 2 **96**
B.1 Alt Saxophon 96
B.2 Oboe 101
B.3 Trompete 105
B.4 Violoncello 109

C DVD **113**

Tabellenverzeichnis **116**

Abbildungsverzeichnis **119**

Literaturverzeichnis **123**

Online-Quellen **124**

Kapitel 1

Einführung

Aktuelle Entwicklungen im Bereich der digitalen Audiosignalverarbeitung verwenden vermehrt Verfahren aus den Bereichen des maschinellen Lernens und der Mustererkennung, um „intelligente“ Algorithmen zu entwickeln. Hierzu gehören Anwendungen, die semantische Informationen aus Signalen extrahieren, signaladaptive Bearbeitungsschritte vollführen oder aber auch die auditive Wahrnehmung imitieren können. Zu letzterem gehört die Nachbildung menschlicher Fähigkeiten, wie der, komplexe akustische Situationen zu analysieren und einzelne Bestandteile eines Klanggemisches ihren sie erzeugenden Ursachen zuordnen zu können. Ein Beispiel für dieses psychoakustische Phänomen ist der ursprünglich von Colin Cherry beschriebene Cocktail-Party-Effect [6]:

> One of our most important faculties is our ability to listen to, and follow, one speaker in the presence of others [12]. [1]
>
> — Colin Cherry (1957)

Ein menschlicher Hörer besitzt demnach die Fähigkeit, selektiv, in der gleichzeitigen Gegenwart weiterer, einer einzelnen Quelle seiner auditorischen Wahrnehmung zu folgen und sie zu verstehen. Störende Quellen können dabei sowohl weitere Sprecher als auch Hintergrundgeräusche oder Musik sein. Sie

[1] Eine unserer wichtigsten Fähigkeiten ist unser Vermögen, einem Sprecher in der Gegenwart anderer zuzuhören und ihm dabei zu folgen. (Übersetzung des Autors)

bleibt zudem auch erhalten, wenn die Energie der Störquellen ähnlich der ist, der der Hörer folgt. Diese Fähigkeit zur Wahrnehmung einzelner auditorischer Objekte von sich überlagernden, akustischen Quellen wird von dem Psychologen Albert S. Bregman als *Auditory Scene Analysis* bezeichnet. Insbesondere Musiksignale eröffnen hierauf eine interessante Sichtweise, da Musik aus der Überlagerung vielfältiger auditorischer Objekte entsteht. Kunio Kashino hat hierfür den Begriff der *Music Scene Analysis* geprägt [15]. Die Grundlage für eine Vielzahl von rechnergestützten Modellbildungen dieser kognitiven Fähigkeit, die zusammenfassend als Quellentrennung bezeichnet werden, stellen die Arbeiten Bregmans dar [2]. Neben rein neurophysiologisch motivierten Ansätzen [30] [25] existieren weitere mathematisch statistische Ansätze [14] [5] [17], aber auch Hybride aus beiden, um die individuellen Quellen eines Klanggemischs detektieren und extrahieren zu können. Solche Ansätze verwenden zumeist sehr allgemeine Annahmen über die Art des Klanggemischs und die Struktur der Quellen und gelten daher als „blind“. Allerdings kann zur Quellentrennung von Musiksignalen das Einbringen von latentem Vorwissen, über die in einem Signal zu erwartende Struktur der Quellen, die Qualität der Trennung verbessern [3]. Solche a priori Informationen können aus statistischen Quellenmodellen bestehen, die die spezifischen *Timbre* der Musikinstrumente abbilden. Das *Timbre* bezeichnet die wie folgt definierte perzeptive Qualität eines Klangs:

> That attribute of auditory sensation in terms of which a listener can judge that two sounds are similarly presented and having the same loudness and pitch are dissimilar [19]. [2]
>
> — *American Standards Association* (1960)

Das Timbre ist demnach eine objektiv nur schwer beschreibbare Größe, allerdings kann der zeitveränderlichen, spektralen Hüllkurve eines Instrumentenklangs eine zentrale Bedeutung in der auditorischen Wahrnehmung zugeschrieben werden [3]. Verfahren zur Quellentrennung unter der Verwendung

[2] Die Eigenschaft einer auditorischen Erregung, anhand derer ein Hörer urteilen kann, ob zwei ähnlich präsentierte Klänge, mit gleicher Lautheit und Tonhöhe, voneinander verschieden sind. (Übersetzung des Autors)

von Quellenmodellen, die die spezifischen Timbre der zu erwartenden einzelnen Quellen in einem Klanggemisch abbilden, werden als *Semi-Blind* bezeichnet und sind ein aktueller Schwerpunkt in der Audiotechnologieforschung.

Die Entwicklung eines spezifischen Quellenmodells zur Repräsentation der Timbre quasi-harmonischer Instrumente ist Gegenstand dieser Magisterarbeit. Das Ziel ist hierbei, ein kompaktes, statistisches Modell zur Darstellung der zeitveränderlichen, spektralen Eigenschaften eines quasi-harmonischen Instruments zu entwerfen und anhand ausgewählter Instrumente zu evaluieren. Neben der Verwendung eines solchen Modells zur Quellentrennung ist der Einsatz in Algorithmen zur Instrumentenerkennung, bzw. -klassifikation denkbar, oder auch zur automatischen Transkription polyphoner, polyinstrumentaler Musiksignale.

Kapitel 2

Grundlagen

In diesem Kapitel werden die theoretischen Hintergründe, physikalischer, mathematischer, wie auch signal- und lerntheoretischer Natur, die die Basis dieser Arbeit darstellen, dargelegt. Hierzu werden die in der Arbeit verwendeten Verfahren, Methoden und Algorithmen in Zusammenhang mit ihren jeweiligen Notationskonventionen vorgestellt.

2.1 Akustische Wahrnehmung

Die für die akustische Wahrnehmung maßgeblichen physikalischen Merkmale sind die *Lautstärke* und die *Frequenz*. Die der Lautstärke zu Grunde liegende physikalische Feldgröße ist der Schalldruck p mit der Einheit $\mathrm{N/m^2}$, welcher in Form eines Wechseldrucks dem atmosphärischen Ruhedruck überlagert ist. Die Frequenz f beschreibt die Anzahl der Periodendauern der Druckschwankungen pro Sekunde und besitzt die Einheit Hertz (Hz) [20].

2.1.1 Lautstärkewahrnehmung

Wie für die meisten Wahrnehmungsgrößen gilt auch für die empfundene Lautstärke das Weber-Fechner-Gesetz, aus dem hervorgeht, dass die subjektive Stärke eines Sinneseindrucks proportional dem Logarithmus der objektiven physikalischen Größe ist. Dies ermöglicht dem menschlichen Gehör, Schalldrücke in einem Bereich von etwa $2{\cdot}10^{-5}$ bis $2{\cdot}10^{2}$ $\mathrm{N/m^2}$ wahrzunehmen.

2.1.2 Tonhöhenwahrnehmung

Der durch das menschliche Gehör wahrnehmbare Frequenzbereich liegt circa zwischen 20Hz und 20kHz [31] und ist insbesondere altersbedingten Schwankungen unterworfen. Dabei gilt, dass die Wahrnehmung der Höhe eines Tons nicht in einer linearen Beziehung zur physikalischen Größe der Frequenz steht. Vielmehr wird die Tonhöhendifferenz zwischen 100 und 125Hz, sowie zwischen 1000 und 1250Hz als gleichwertig empfunden. Ebenso wird ein Ton dann als doppelt so hoch zu einem Bezugston empfunden, wenn seine Frequenz dem doppelten des Frequenzwertes des Bezugstones entspricht. Somit wird das Tonintervall zwischen 100 und 200Hz ebenso als Tonhöhenverdopplung bewertet, wie das Intervall zwischen 200 und 400Hz. Die wahrgenommene Tonhöhendifferenz entspricht somit nicht der Frequenzdifferenz, sondern dem Frequenzverhältnis.

2.2 Akustik der Musikinstrumente

Der Begriff des Musikinstruments bezeichnet ein Objekt, welches als akustische Quelle der gezielten Erzeugung eines Schallereignisses dient. Die verschiedenen Instrumente weisen dabei jeweils eine eigene, individuelle Klangcharakteristik auf, für deren Entstehung die synergetische Kombination eines schwingenden Anregesystems, eines Resonanzsystems und das richtungsabhängige Abstrahlverhalten maßgeblich zeichnet. Das schwingende Anregesystem einerseits überträgt die Schwingungen an die umgebende Luft und repräsentiert somit den Tonerzeuger, während das Resonanzsystem einzelne Komponenten, der an die Luft übertragenen Schwingungen, verstärkt oder dämpft. Je nach Instrumentenklasse kann es sich bei dem Anregesystem beispielsweise um eine zwischen zwei Punkten fixierte Saite (Chordophone), oder aber auch um eine schwingende Luftsäule (Aerophone) handeln. Zusätzlich wird die Wahrnehmung eines Klangs durch psychoakustische Parameter, wie auch durch Raumwirkungsaspekte beeinflusst [18]. Die Klangfarbe ist demnach eine mehrdimensionale Größe, die sowohl physikalische als auch psychoakustische Parameter vereint. Sie wird als *Timbre* bezeichnet. Reduziert

man die Klangeigenschaften jedoch auf ihre messbaren, physikalischen Komponenten und vernachlässigt somit den Einfluss der Wahrnehmungsgrößen und Raumwirkungsfaktoren, so lassen sich drei strukturelle Dimensionen beschreiben, die sowohl das Erkennen eines Instruments ermöglichen, als auch spieltechnische Details und somit die Intention des Instrumentalisten offenbaren.

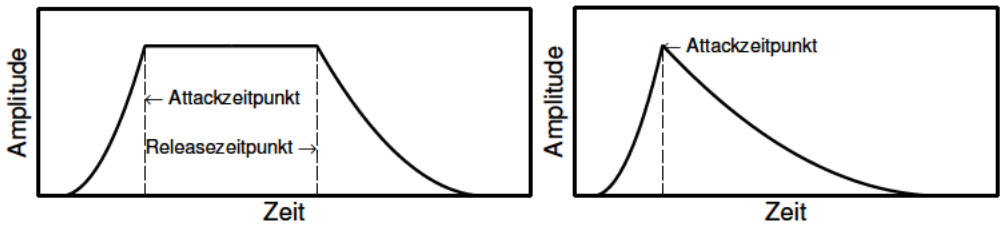
2.2.1 Die spektrale Dimension

Das Spektrum eines Klangs gibt Auskunft über die Verteilung der einzelnen Frequenzen der abgestrahlten Schallenergie. Wird ein schwingungsfähiges System angeregt, so führt dies zu periodischen Schwingungen mit einer mehr oder minder ausgeprägten, harmonischen Oberwellenstruktur. Das heißt, das System schwingt nicht nur in einer festen Frequenz, sondern zusätzlich in allen ganzzahligen Vielfachen dieser Grundfrequenz. Das spezifische Verhältnis der Energien der Oberwellen bezüglich ihrer Grundfrequenz ist dabei ein Alleinstellungsmerkmal eines jeden Musikinstruments. Statt der Bezeichnungen Grund- und Oberton wird im Rahmen dieser Arbeit der Begriff der *Partialwelle* verwendet, wobei die erste Partialwelle dem Grundton entspricht und alle höherwertigen Partialwellen k dem $(k-1)$-ten Oberton.

2.2.2 Die Dynamikdimension

Einen Einfluss auf die Klangfarbe hat auch die Lautstärke des Klangs. So verstärkt eine laute Spielweise insbesondere die höheren Frequenzen. Dieser Effekt lässt sich durch den Tonansatz noch verstärken.

In der Musik der westlichen Hemisphäre werden zur Beschreibung einer lauten oder leisen Spielart die dem Italienischen entstammenden Begriffe *forte (f)* für laute und *piano (p)* für leise bzw. *mezzoforte (mf)* für eine mittellaute Spielweise verwendet. Die weiter gesteigerten und in der Musiktheorie gebräuchlichen Begriffe wie *fortissimo (ff)* oder *pianissimo (pp)* etc. werden in dieser Arbeit keine weitere Betrachtung finden.



(a) kontinuierlich (b) impulsartig

Abbildung 2.1: Exemplarische Amplitudenhüllkurven

2.2.3 Die zeitliche Dimension

Die zeitliche Abfolge der Spektraldimension ist ebenfalls ein wichtiger Informationsträger und besitzt einen signifikanten Einfluss auf die Erkennbarkeit eines Instruments. Die größte Bedeutung weist dabei der so genannte Einschwingvorgang auf, währenddessen dem schwingungsfähigen System von außen Energie zugeführt und das System reagiert mit einem Ausgleichsvorgang. In diesem Zeitfenster findet eine spektrale Verbreiterung um die sich einschwingenden Frequenzen statt [23]. Zu unterscheiden sind dabei eine *kontinuierliche* Anregung bei der sich nach Ablauf des Ausgleichsvorgangs ein quasi-stationärer Zustand einstellt und eine *impulsartige* Energiezufuhr, bei der sich kein stationärer Zustand herausbildet, sondern das System direkt vom Ein- in einen Ausschwingvorgang übergeht. Im Ausschwingvorgang werden die Energien in den Resonanzen des Systems in Abhängigkeit der jeweiligen Dämpfung schneller oder langsamer verbraucht. Im Gegensatz zum Einschwingvorgang bilden sich allerdings keine neuen Klang- oder Geräuschanteile aus.

Die jeweilige Dauer dieser Vorgänge wird als *Einschwing*-, respektive *Ausschwingzeit* bezeichnet. Zusätzlich gebräuchlich sind die Begriffe des *Attacks* und *Release*. Wobei der *Attack* dem Ende der Einschwingzeit gleichzusetzen ist und umgekehrt der *Release* dem Beginn der Ausschwingzeit. In Abbildung 2.1 sind exemplarische Amplitudenhüllkurven sowohl einer kontinuierlich angeregten Schwingung als auch einer impulsartigen Anregung dargestellt.

2.3 Diskrete Signalrepräsentationen

Für eine rechnergestützte Verarbeitung der Klänge von Musikinstrumenten ist es notwendig, die in ihrer Natur sowohl zeit- als auch amplitudenkontinuierlichen Signale in eine Form mit diskreten Zeit- und quantisierten Amplitudeneinheiten zu überführen. Dies geschieht mit Hilfe einer Analog-Digital-Wandlung an deren Eingang ein kontinuierliches, analoges Signal angelegt wird und als dessen Resultat ein zeitdiskretes, digitales Signal ausgegeben wird.

Für die zeitliche Diskretisierung gilt das Nyquist'sche Abtasttheorem, dargestellt in Gleichung 2.1, welches eine mindestens doppelt so hohe Abtastrate verlangt, wie die höchste zu rekonstruierende Frequenz (Nyquist-Frequenz)[21].

$$f_s > 2 \cdot f_{ny} \tag{2.1}$$

$$\begin{aligned} f_s &: \text{Abtastrate} \\ f_{ny} &: \text{Nyquist-Frequenz} \end{aligned}$$

Die Qualität der Abtastung der Amplitude misst sich an dem durch die Quantisierung entstehendem Fehlersignal zwischen dem abgetastetem und dem Originalsignal. Ein Maß für die in Kauf genommene Verschlechterung des Signals ist das Signal-Rausch-Verhältnis (*Signal-Noise-Ratio* (SNR)) aus Gleichung 2.2.

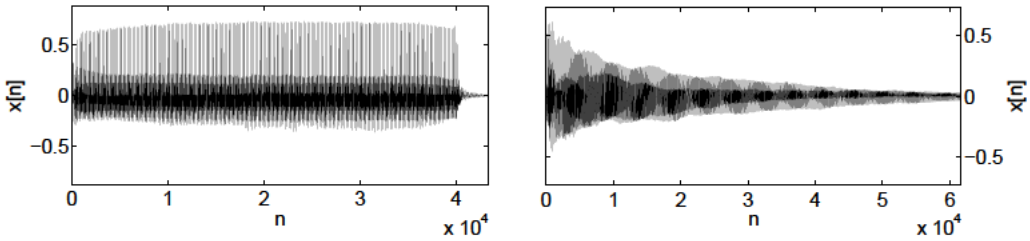
$$SNR = 6.02 \cdot bit \quad [dB] \tag{2.2}$$

$$bit \quad : \quad \text{Bitrate der Quantisierung}$$

2.3.1 Der Zeitbereich

Direkt aus der Analog-Digital-Wandlung resultiert die Zeitbereichsrepräsentation, die sich in Form einer Folge von Amplitudenwerten $x[n]$ über der Samplezeit n darstellen lässt. Abbildung 2.2 zeigt die Amplitudenfolgen für eine kontinuierliche Anregung (Bild 2.2a) und eine impulshafte Anregung

(Bild 2.2b). Beide Signale wurden mit einer Abtastrate $f_s = 44100Hz$ und einer Bitrate von $bit = 16$ digitalisiert.



(a) Alt-Saxophon (A3) (b) Piano (A3)

Abbildung 2.2: Amplitudenfolgen

2.3.2 Der Frequenzbereich

Eine für die Signalverarbeitung wichtige Repräsentation von Signalen ist die Darstellung in Form einer Linearkombination komplexer Exponentialfunktionen bzw. Sinusschwingungen. Diese Darstellungsform wird als Frequenzbereich oder Spektrum bezeichnet und basiert auf dem Theorem der Fourierreihe, nach welchem sich jede komplexe, periodische Wellenform durch die Summe einzelner Sinusschwingungen darstellen lässt. Für zeitdiskrete Eingangsfolgen wird das Spektrum mit Hilfe der diskreten Fourier-Transformation (DFT) in Gleichung 2.3 berechnet.

$$X[b] = \sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)bn}, \quad b = 0, 1, \ldots, N-1 \tag{2.3}$$

N : Länge der Signalfolge
b : DFT Abtastwert (bin)

Für die Umwandlung der Stützstellen der DFT in Frequenzwerte gilt die Beziehung in Gleichung 2.4, wobei sich die Berechnung auf reellwertige Signale beschränkt, deren DFT um den Punkt $N/2$ symmetrisch ist – dieser entspricht der Nyquist-Frequenz.

$$f[b] = \frac{b}{N} f_s, \quad b = 0, 1, \ldots, N/2 \tag{2.4}$$

Die Energie E eines Signalausschnitts $x[n]$ der Länge N und die Energie seiner DFT folgt der Beziehung in Gleichung 2.5.

$$E = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2 = \frac{1}{N} \sum_{b=0}^{N-1} |X[b]|^2 \tag{2.5}$$

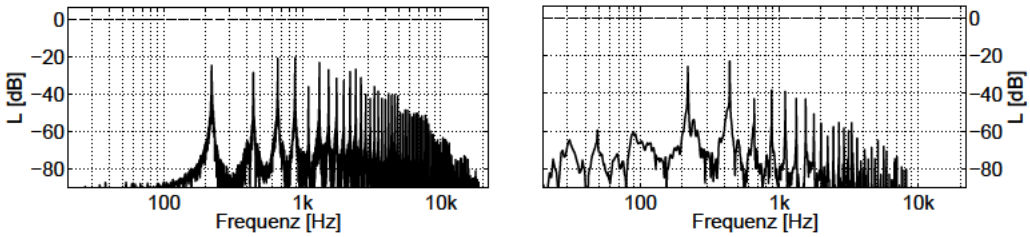
Aus diesem als *Parseval'sches Theorem* bezeichnetem Zusammenhang lässt sich die Funktion $\frac{1}{N}|X[b]|^2$ als Energie-, respektive Leistungsdichtespektrum interpretieren.

Um die Werte der DFT in eine Repräsentation zu überführen, die der logarithmischen Lautstärkewahrnehmung des Gehörs angepasst ist, werden die jeweiligen Werte in Dezibel (dB) angegeben. In Gleichung 2.6 wird die Interpretation einer Amplitudengröße (links) und ihrer Leistungsgröße (rechts) angegeben.

$$L = 20 \cdot \log_{10} \left(\frac{|X[b]|}{X_{\text{norm}}} \right) = 10 \cdot \log_{10} \left(\frac{|X[b]|}{X_{\text{norm}}} \right)^2, \quad [\text{dB}] \tag{2.6}$$

$$X_{\text{norm}} \quad : \quad \text{Normierung}$$

Abbildung 2.3 zeigt die logarithmierten Spektren der in Abbildung 2.2 dargestellten Folgen.



(a) Alt-Saxophon (A3) (b) Piano (A3)

Abbildung 2.3: Spektren

Wie sich in Gleichung 2.3 sehen lässt, werden zur Berechnung eines DFT-Werts N Multiplikationen benötigt, was für alle b einen Gesamtaufwand von $\mathcal{O}(n^2)$ verursacht. Es existieren allerdings leistungsfähige Algorithmen, die als Schnelle-Fourier-Transformation (*Fast-Fourier-Transformation* (FFT)) bezeichnet werden und deren Aufwand auf $\mathcal{O}(nlogn)$ beschränkt ist [21]. Für die häufig verwendeten Radix-k Implementierungen gilt jedoch die Einschränkung einer auf k^x beschränkten Anzahl an Werten b für die DFT.

2.3.3 Die Kurzzeit-Spektralanalyse

Eine Beschränkung der DFT liegt darin, dass sie keine Aussage über das zeitliche Auftreten der einzelnen Frequenzkomponenten trifft. Da es sich bei Audiosignalen aber zumeist um nicht-stationäre Signale handelt, ist die Einführung einer zeit-veränderlichen Fourier-Repräsentation sinnvoll [22].
Gleichung 2.7 zeigt das als Kurzzeit-Spektralanalyse, bzw. -transformation (*Short-Time-Fourier-Transformation* (STFT)) bezeichnete Verfahren [21].

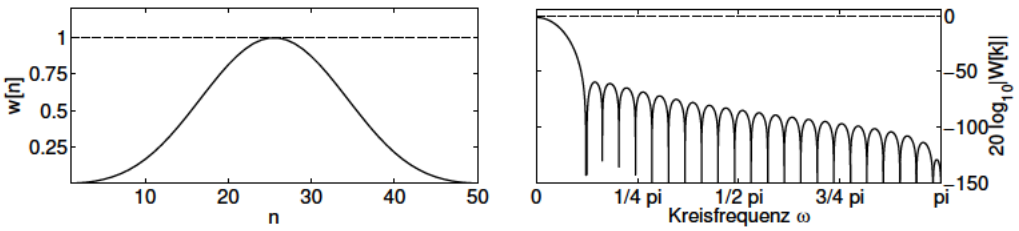
$$X_r[b] = \sum_{n=0}^{N-1} x(rH+n)w(n)e^{-j(2\pi/M)bn}, \quad b = 0, \ldots, M-1 \tag{2.7}$$

r	:	Blockindex
M	:	Länge des DFT-Fensters
H	:	Schrittweite
$w(n)$	:	Fensterfunktion

Dabei wird das Spektrum jeweils nur über einem Ausschnitt des Signals berechnet, welcher den Zeitindex r erhält. Die aufeinanderfolgenden Segmente werden dann in einer Schrittweite mit der Länge H dem Signal entnommen und zusätzlich mit einer Fensterfunktion $w(n)$ multipliziert. Das Resultat der Analyse $X_r[b]$ erhält dabei den Begriff des Spektrogramms. Für diese so genannte *Fensterung* werden in der Literatur verschiedene Funktionen vorgeschlagen, die einen in der Näherung gaussartigen Verlauf aufweisen. Eine häufig verwendete Funktion ist das in Abbildung 2.4 dargestellte *Blackman*-Fenster, dessen Verlauf sich aus Gleichung 2.8 ergibt. Die gebräuchlichen

Fensterfunktionen werden im allgemeinen an Hand ihres Spektrums unterschieden, wobei insbesondere die Breite der Hauptkeule und die Absenkung der Nebenkeulen betrachtet werden.

$$w[n] = 0.42 + 0.5 \cdot \cos\left(\frac{2n\pi}{N}\right) + 0.08 \cdot \cos\left(\frac{4n\pi}{N}\right) \tag{2.8}$$

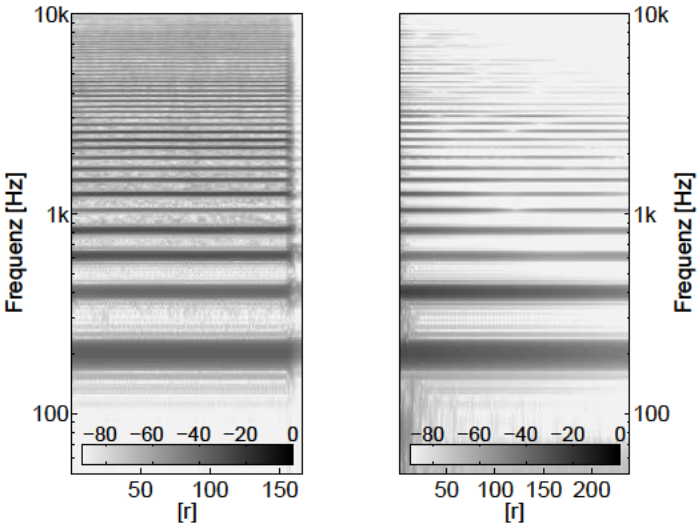


(a) Funktion (b) Spektrum

Abbildung 2.4: *Blackman*-Fenster

Das Spektrogramm $X_r[b]$ ist somit eine Funktion von r und b. Für ein festes r entspricht sie den Eigenschaften der normalen DFT der Sequenz $x(rH + n)w(n)$ und für ein festes b kann X_r als eine Funktion über der Zeit r interpretiert werden.

Entscheidend für die Aussagekraft des Spektrogramms ist die Parametrisierung von N, M und H und die Wahl der Fensterfunktion. Die Länge N des Signalausschnitts muss so gewählt sein, dass mindestens drei bis vier Perioden der tiefsten im Signal enthaltenen Frequenz erfasst werden. Zusätzlich wird über die Länge des Analysefensters die Frequenzauflösung eingestellt, wobei sie mindestens der Länge des Ausschnitts entsprechen muss. Wird eine Länge $M > N$ gewählt, werden für die Berechnung der DFT dem Signalausschnitt Nullen angefügt (*Zero-Padding*). Mit Hilfe des Parameters H wird die zeitliche Auflösung festgelegt. Abbildung 2.5 zeigt die Spektrogramme der in Abbildung 2.2 vorgestellten Amplitudenfolgen mit den Parametern: $N = 2005$, $M = 8192$ und $H = 251$ unter Verwendung des *Blackman*-Fensters.



(a) Alt Saxophon (A3) (b) Piano (A3)

Abbildung 2.5: Spektrogramme

2.3.4 Das Sinusoidalmodell

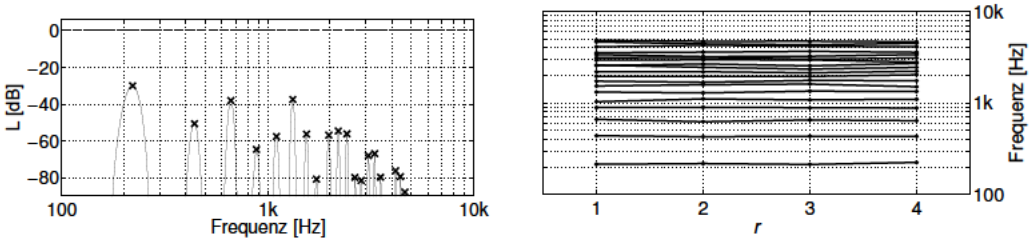
Der Ansatz des Sinusoidalmodells (SMS) basiert auf der Annahme, einen Klang aus stabilen Sinusoiden, den so genannten Partialwellen, und einem residualem Rauschen modellieren zu können [27]. Die dabei zugrunde liegende Modellvorstellung über die Erzeugung eines Klangs eines musikalischen Instruments oder physikalischen Systems beinhaltet somit die Existenz einer deterministischen und einer stochastischen Komponente. Die deterministische repräsentiert die Hauptmoden des schwingenden Systems und wird daher durch die Summe einzelner Sinusoide, die als Funktion langsam veränderlicher Amplituden und Frequenzen beschrieben werden, charakterisiert. Die stochastische Komponente repräsentiert Signalanteile, die in ihrer Natur nicht sinusoidal sind und zum Beispiel durch den Anregemechanismus oder den über eine Saite streichenden Bogen erzeugt werden. Sie wird durch ihr Leistungsdichtespektrum charakterisiert. Die zeit-diskrete Synthesegleichung in Form einer Linearkombination der Sinusoiden bzw. Partialwellen plus additivem Rauschen bzw. Residualsignal ist in Gleichung 2.9 dargestellt.

$$x[r] = \tilde{x}[r] + \epsilon[r] = \sum_{k=1}^{K} a[k,r]\cos(\phi[k,r]) + \epsilon[r] \tag{2.9}$$

$a[k,r]$	:	Amplitude der Partialwelle k in Block r
$\phi[k,r]$	:	Phasenwinkel der k-ten Partialwelle in Block r
$\epsilon[r]$	:	Residuales Rauschen in Block r

Das Verfahren zur Konstruktion des deterministischen plus stochastischen Modells basiert auf der Kurzzeitspektralanalyse des Eingangssignals. Die deterministische Komponente wird dabei erzeugt, indem, wie in Bild 2.6a dargestellt, die signifikanten Amplitudenmaxima in den einzelnen Betragsspektren $|X_r[b]|$ detektiert werden, wobei jedes detektierte Maximum einer Partialwelle k mit der Amplitude $|X_r[b]| = a[k,r]$ zugeordnet wird [37]. Eine verbesserte Variante zur Maximadetektion stellt die quadratische Interpolation über das detektierte Bin und seine zwei direkten Nachbarn dar [1], da die reale Frequenz eines Sinusoids nicht zwangsläufig auf eine Stützstelle der DFT entfallen muss.

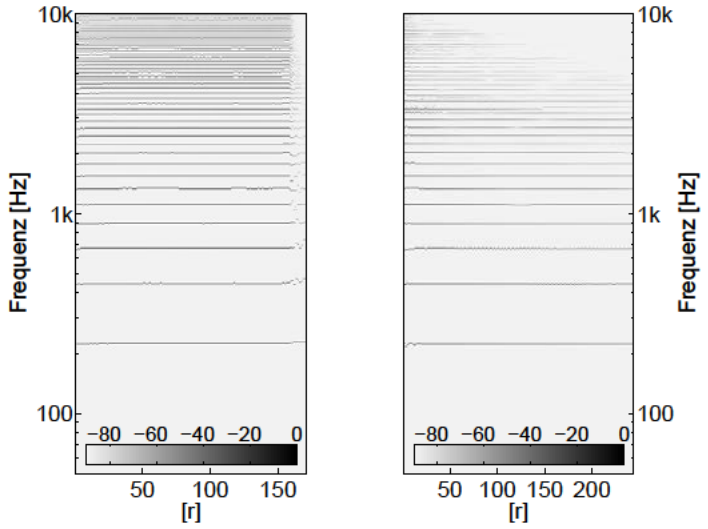
In einem zweiten Schritt werden benachbarte Amplitudenmaxima über die zeitlich aufeinanderfolgenden Segmente r zu, in Bild 2.6b dargestellten, Trajektorien verknüpft. Die stochastische Komponente wird im Anschluss



(a) Maxima Detektion (b) Maxima Trajektorien

Abbildung 2.6: Die 2 zentralen Schritte zur Erzeugung des Sinusoidalmodells

aus der Subtraktion des durch additive Synthese der Sinusoide erzeugten deterministischen Signals vom Eingangssignal gewonnen. Abbildung 2.7 zeigt die Verläufe der Partialwellen der in Absatz 2.3.1 vorgestellten Signalfolgen.



(a) Alt-Saxophon (A3) (b) Piano (A3)

Abbildung 2.7: Sinusoidalmodelle

2.4 Maschinelles Lernen

Lernen in dem Sinne, wie es in dieser Arbeit Verwendung findet, bezeichnet einen Prozess, bei dem die freien Parameter eines Modells, auf Grund einer Stimulation durch eine das Modell einbettende Umgebung, angepasst werden[11]. Der Prozess soll die freien Parameter dahingehend schätzen, dass das Modell allgemeine Gesetzmäßigkeiten aus den Daten der Umgebung erkennt („lernt“) und somit eine Vorhersage für bislang unbekannte Daten möglich wird. Diese aus Beispieldaten gewonnene Generalisierungsfähigkeit ist das Ziel des maschinellen Lernens[8] und wird daher in Anlehnung an die Psychologie auch induktives Lernen genannt.

2.4.1 Paradigmen des Lernens

Die das Modell einbettenden Umgebungen lassen sich als Lernparadigmen bezeichnen, da sie die verschiedenen Ansätze des maschinellen Lernens anhand ihres jeweiligen Verwendungszwecks bzw. der Form der vorliegenden Trainingsbeispiele einteilen. So stellen die verschiedenen Paradigmen jeweils

grundlegende Varianten des Lernens dar. Zwei der im Rahmen dieser Arbeit zentrale Paradigmen sind die folgenden:

Unüberwachtes Lernen (Unsupervised Learning)

Der Lernprozess erzeugt eine Repräsentation, die die Trainingsdatenvektoren $\mathbf{x}^{(\alpha)}, \quad \alpha = 1 \ldots \mathcal{A}$ ($\mathcal{A}$: Anzahl an Trainingsbeispielen) möglichst optimal beschreibt. *Clustering*-Verfahren versuchen die Daten auf Grund verschiedener Muster in Kategorien einzuordnen und Verfahren wie die Hauptkomponentenanalyse (PCA) versuchen die Daten in eine dimensionsreduzierte Darstellung bei maximaler Informationserhaltung zu transformieren.

Überwachtes Lernen (Supervised Learning)

Hierbei wird der Lernprozess von einem Lehrer begleitet, welcher zu jedem Trainingsbeispiel $\mathbf{x}^{(\alpha)}$ ein Kategorien-*Label* $y_T^{(\alpha)}$ verwaltet. Die durch das Kategorien-*Label* repräsentierte, gewünschte Modellausgabe stellt somit die optimale Ausgabe dar und die Schätzung der Parameter geschieht durch den kombinierten Einfluss aus einem Trainingsdatenvektor $\mathbf{x}^{(\alpha)}$ und seinem *Label* $y_T^{(\alpha)}$. Beispiele des überwachten Lernens sind Neuronale Netze mit *Backpropagation*-Algorithmus oder *Support-Vector*-Maschinen (SVM).

2.4.2 Kosten und Gradientenabstieg

Für das überwachte Lernen lässt sich zeigen, dass durch den Vergleich zwischen den, durch die Kategorien-*Label* $y_T^{(\alpha)}, \quad \alpha = 1 \ldots \mathcal{A}$ vorgegebenen, gewünschten Modellausgaben und den zu den Trainingsbeispielen tatsächlich erzeugten Modellausgaben eine Fehler- bzw. Kostenfunktion $C(\mathbf{w})$ in Abhängigkeit der freien Parameter des Modells definiert werden kann (Gleichung 2.11).

$$C(\mathbf{w}) = \frac{1}{P}\sum_{\alpha=1}^{\mathcal{A}} c^{(\alpha)}(\mathbf{w}) \tag{2.10}$$

$$c^{(\alpha)}(\mathbf{w}) = \frac{1}{d}|y_T^{(\alpha)} - y(\mathbf{x}^{(\alpha)}, \mathbf{w})|^d \tag{2.11}$$

$\mathbf{w}$	:	Freie Modellparameter
$y(\mathbf{x}^{(\alpha)}, \mathbf{w})$	:	Modellausgabe

Diese Funktion kann als eine multi-dimensionale Fehleroberfläche visualisiert werden, mit den freien Parametern des Modells als Koordinaten. Abbildung 2.8 zeigt eine exemplarische Fehleroberfläche über 2 freie Parameter w_1 und w_2, wobei jede berechnete Modellausgabe somit einen Punkt auf dieser Oberfläche darstellt.

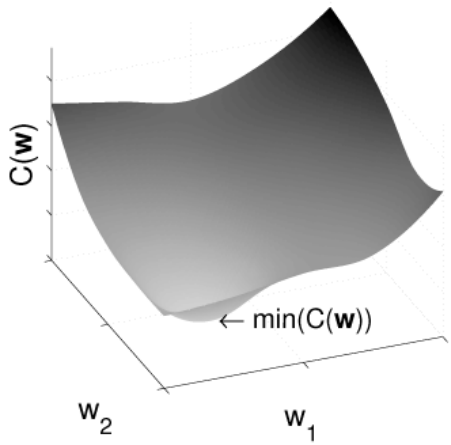


Abbildung 2.8: Fehleroberfläche

Für eine Optimierung des Verhaltens des Modells in Bezug auf seinen Fehler, respektive die Kosten, werden die freien Parameter so angepasst, dass ein globales oder lokales Minimum der Fehleroberfläche erreicht wird. Unter der Voraussetzung, dass die Fehlerfunktion stetig abgeleitet werden kann, wird dies im Rahmen eines iterativen Prozesses über den in Gleichung 2.13 dargestellten Gradienten und die in Gleichung 2.14 gezeigte *Update*-Regel erzielt. In Abbildung 2.9 ist verdeutlicht, wie die freien Parameter entlang der Rich-

tung des negativen Gradienten adaptiert werden und somit der Modellfehler optimiert wird.

$$C(\mathbf{w}) \stackrel{!}{=} \min \tag{2.12}$$

$$\Delta \mathbf{w}_t = -\eta \frac{\partial C(\mathbf{w}_t)}{\partial \mathbf{w}_t} \tag{2.13}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \Delta \mathbf{w}_t \tag{2.14}$$

t : Zeitindex des iterativen Prozesses

η : Lernschrittweite

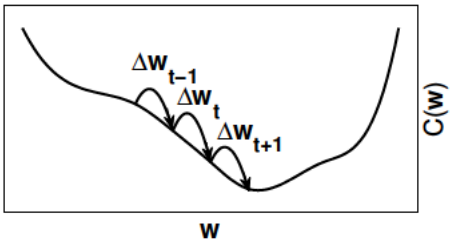


Abbildung 2.9: Gradientenabstieg

Eine solche fest vorgeschriebene Menge an wohldefinierten Regeln wird als Lernalgorithmus bezeichnet.

2.4.3 Stochastische Approximation

Der in Absatz 2.4.2 vorgestellte Lernalgorithmus wird in seiner Form auch *Batch*- oder *Offline*-Algorithmus genannt, da er die Existenz aller Trainingsdaten voraussetzt. Wird das Lernen allerdings bereits während der Datenerhebung ausgeführt, müssen die Kosten und Gradienten für jedes einzelne Trainingsbeispiel erhoben werden, was einer lokalen Approximation der Kosten entspricht. Diese Variante des Lernens wird *Online*-Lernen genannt und kann sich auch bei Vorhandensein aller Trainingsdaten als sinnvoll erweisen.

$$\Delta \mathbf{w}_t = -\eta \frac{\partial c^\alpha(\mathbf{w}_t)}{\partial \mathbf{w}_t} \tag{2.15}$$

Der in Gleichung 2.15 gezeigte, geänderte Gradient und die unverändert gültige *Update*-Regel in Gleichung 2.14 führen ebenfalls zur Konvergenz in einem lokalen oder globalen Minimum der Fehlerfunktion.

2.4.4 Anpassung der Schrittweite

Die Schrittweite η hat einen entscheidenden Einfluss auf das Konvergenzverhalten des Lernalgorithmus. Während ein zu groß eingestellter Wert zu einer Divergenz des Systems führen kann, bedeutet ein zu kleiner Wert, dass das Lernen unnötig verlangsamt wird. Zusätzlich gilt, dass insbesondere im Fall des *Online*-Lernens die Schrittweite im Verlauf des Trainings verringert werden muss, um eine Konvergenz des Algorithmus zu garantieren.

2.4.5 Modellselektion

Eine der Schlüsselannahmen im Bereich des maschinellen Lernens ist die Annahme von verrauschten Daten. Das bedeutet, dass die Daten nicht exakt den Zusammenhang widerspiegeln, aus dem heraus sie erzeugt wurden, sondern mehr oder weniger stark davon abweichen. Diese Abweichungen sollen jedoch von dem zu trainierenden Modell nicht mit abgebildet werden. Daraus folgt, dass für eine gute Generalisierungseigenschaft eines Modells nicht die Minimierung der Kosten das eigentliche Ziel darstellt, sondern die Minimierung der zu erwartenden Kosten für unbekannte Daten. Diese werden Generalisierungsfehler bzw. Risiko R genannt. Um eine Abschätzung des Risikos zu erhalten, wird die zur Verfügung stehende Datenmenge D in eine Trainingsmenge D^C und eine Testdatenmenge D^R geteilt, wobei die Parameterschätzung anhand der Trainingsdaten erfolgt und die Abschätzung des Risikos mit Hilfe der Testdaten. Da dieses als Trainings-/Testdatenmethode bezeichnete Verfahren jedoch eine Überanpassung an die Testdaten nicht ausschließt, wird als eine Verallgemeinerung das *J-fold*-Kreuzvalidierungsverfahren ein-

gesetzt. Abbildung 2.10 zeigt im oberen Abschnitt die Aufteilung der ge-

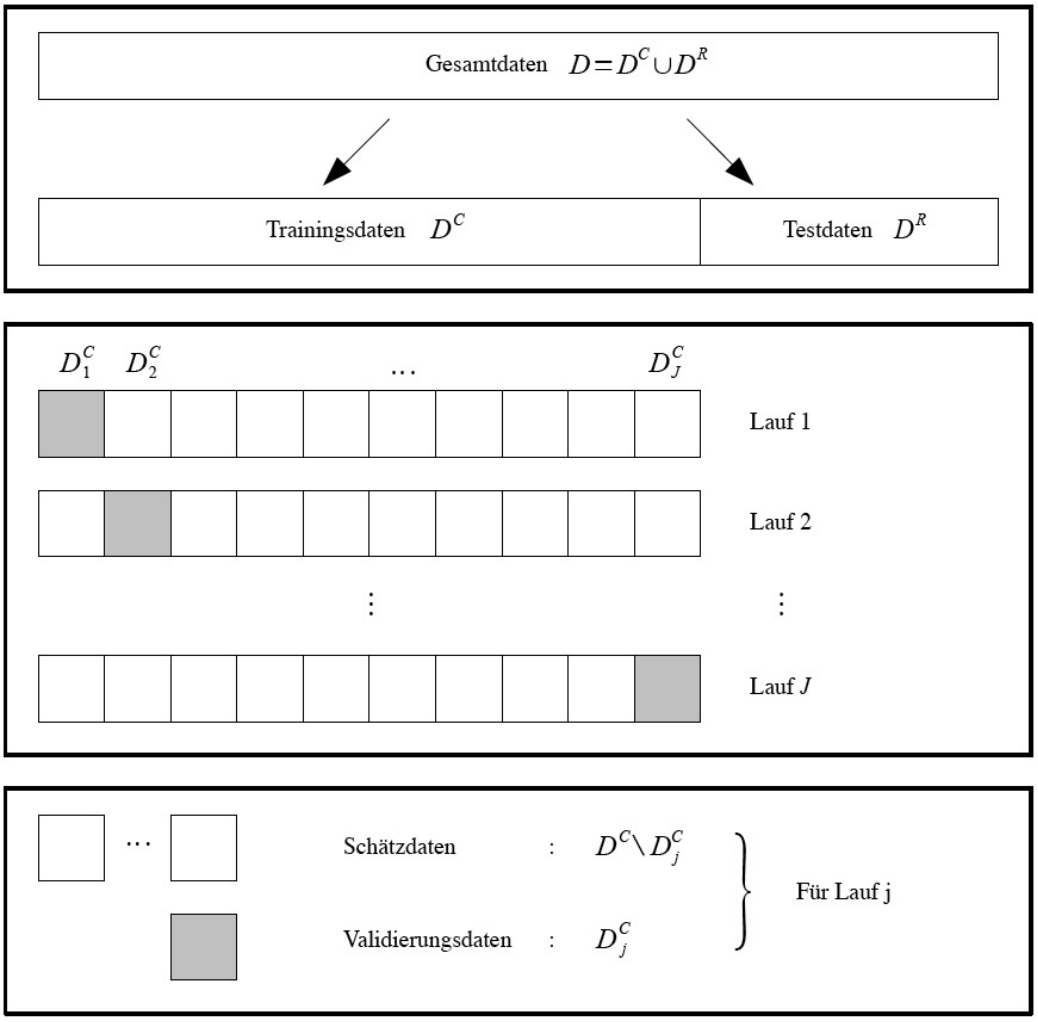


Abbildung 2.10: 10-*fold*-Kreuzvalidierung

samten zur Verfügung stehenden Datenmenge in eine Trainings- und Testdatenmenge, während im mittleren und unteren Teil die Verwendung der Trainingsdaten für eine *J-fold*-Kreuzvalidierung mit $J = 10$ dargestellt ist. Hierbei wird die Trainingsdatenmenge D^C in J disjunkte Untermengen der Form $D^C = D_1^C \cup D_2^C \cup \ldots \cup D_J^C$ separiert. Bei einer *J-fold*-Kreuzvalidierung werden demnach J Modelle auf jeweils variierenden Untermengen D_j^C der Trainingsdaten geschätzt, wobei zur Parameterschätzung $J-1$ Untermengen verwendet werden und die jeweils ausgelassene Untermenge zur Validitäts-

schätzung herangezogen wird. Das Modell mit den geringsten Kosten für die Validierungsmenge gilt nach Abschluss des Verfahrens als aussichtsreichster Kandidat, dessen Risiko wiederum auf der Testdatenmenge D^R geschätzt wird. Die Kosten des selektierten Modells werden dann jedoch als Mittelwert der Kosten aller J Modelle auf den Validierungsmengen angegeben [11].

Kapitel 3

Stand der Forschung

Bevor mit Hilfe der in Kapitel 2 vorgestellten theoretischen Grundlagen das Modell, welches der zentrale Gegenstand dieser Arbeit ist, vorgestellt wird, werden im aktuellen Kapitel zwei bestehende Ansätze für eine Modellbildung quasi-harmonischer Instrumente gezeigt, die Verfahren des maschinellen Lernens verwenden. Im Anschluss an die kurzen Beschreibungen werden die Beschränkungen des jeweiligen Ansatzes erläutert, die von dem in den folgenden Kapiteln vorgestellten Modell aufgehoben werden sollen.

Beide Ansätze zielen darauf hin ab, das jeweilige Timbre musikalischer Instrumente zu modellieren und für eine Verwendung in einer Quellentrennung oder Instrumentenklassifikation zur Verfügung zu stellen. Dabei wird in beiden Ansätzen zur Beschreibung des jeweiligen Timbre der Instrumentenklänge, ihre über die Zeit veränderliche, spektrale Verteilung der Energie als maßgebliche Größe zu Grunde gelegt. Darüber hinaus wird in den Ansätzen angenommen, dass der aus den einzelnen Partialwellen des Sinusoidalmodells bestehende deterministische Signalanteil, die wesentlichste Signaleigenschaft zur Diskrimination der einzelnen Timbre musikalischer Instrumente darstellt.

3.1 Modell von Juan José Burred

Für seinen Modellansatz formuliert Juan José Burred drei zentrale Kriterien, anhand derer die Modellbildung ausgerichtet ist [3]:

- Allgmeingültigkeit (*Generality*)
- Kompaktheit(*Compactness*)
- Genauigkeit (*Accuracy*)

Allgemeingültigkeit beschreibt hierbei die Fähigkeit des Modells, die verschiedenen klanglichen Qualitäten ein und desselben Instrumententyps zu erfassen, um die während der Modellierungsphase noch unbekannten Signale später entsprechend erkennen zu können. Dies umfasst neben dem gesamten Tonhöhenumfang eines jeden Instruments auch die verschiedenen klanglichen Nuancen, die von unterschiedlichen Instrumenten des gleichen Typs oder durch verschiedene Instrumentalisten erzeugt werden.

Der Begriff der Kompaktheit bezeichnet die Anforderung an die Modellbildung, neben einer recheneffizienten Lösung, möglichst nur die wesentlichen Merkmale der spezifischen Timbre der Instrumente abzubilden.

Damit ein solches Modell im Kontext eines Verfahrens zur Quellentrennung zur Maskierung der spektralen Verteilung eines Signalgemischs verwendet werden kann, fordert Burred, dass das Modell eine möglichst hohe Genauigkeit in der Abbildung der spezifischen Timbre der Instrumente ermöglicht, da anderenfalls Abweichungen im Maskierungsprozess zu Artefakten führen würden, die die perzeptive Qualität einer solchen Separation verschlechtern.

Burred vereint diese Anforderungen, indem er den Modellierungsprozess in zwei voneinander getrennte Verarbeitungsstufen einteilt. Eine erste so genannte Repräsentationsphase (*Representation Stage*) und eine zweite die so genannte Prototypisierung (*Prototyping Stage*).

In der Repräsentationsphase werden aus einer Trainingsdatenbasis, bestehend aus den Einzelklängen aller zu modellierenden Instrumente, die Sinusoidalmodelle gewonnen, die über der Zeit konkateniert und in einer Gesamtdatenmatrix zusammengefasst werden. Durch die Verwendung einer Trainings-

datenbank, bestehend aus realen Instrumentenklängen, wird somit dem Kriterium der Allgemeingültigkeit Rechnung getragen, da hierfür entsprechende Beispiele verwendet werden können, die die oben genannten Kriterien erfüllen. Im Anschluss wird die Gesamtdatenmatrix mit Hilfe einer Hauptkomponentenanalyse in einen Timbre-Raum mit geringerer Dimensionalität überführt, wobei die Dimensionreduktion über den Rekonstruktionsfehler eingestellt wird. Dies ermöglicht die Darstellung der Einzelklänge der Eingangsdaten in kompakterer Form bei gleichzeitiger Kontrolle der Genauigkeit und erfüllt demnach zwei der oben genannten Kriterien.

In der Phase der Prototypisierung wird dann zu jedem Instrument, auf Basis seiner in den Timbre-Raum projizierten Einzelklänge, eine aus einem Gaussprozess bestehende Trajektorie konstruiert, die die Entwicklung des Timbre über der Zeit modelliert. Diese besteht aus den über der Zeit parametrisierten, empirischem Mittelwert und Varianz der Trainingsdaten.

Angewandt auf eine Instrumentenklassifikation für isolierte Einzelklänge von fünf Instrumenten (Flügel, Klarinette, Oboe, Violine und Trompete) erreicht Burred mit seinem Modell eine Klassifikationsgenauigkeit von bis zu 94.9% bei 2.1% Standardabweichung [4].

Die Beschränkung des Modells von Juan José Burred liegt in der Tatsache, dass die Abhängigkeit des Timbre der Instrumente von der Grundfrequenz implizit in das Gesamtmodell eingebettet ist und die Prototypen daher die Klangfarbenunterschiede für verschiedene Grundfrequenzen über ihren Mittelwert und ihre Varianz repräsentieren. Eine Aufgabe für die in den folgenden Kapiteln angestrebte Modellbildung ist daher, diese Abhängigkeit der zeitveränderlichen, spektralen Verteilung der Energie der Instrumentenklänge von der Grundfrequenz explizit in das Modell einzubinden.

3.2 Modell von Anssi Klapuri

Der von Anssi Klapuri entwickelte Modellansatz [16] verwendet ein Quelle-Filter-Dämpfungs-System (*Source-Filter-Decay*) zur verallgemeinerten Repräsentation der Timbre quasi-harmonischer Instrumente. Alle drei Bestandteile sind dabei in ein lineares Gesamtmodell eingebunden und bilden für

einen festen Parametersatz einen Prototypen für ein einzelnes Instrument.

Die Quelle symbolisiert ein über dem Partialwellenindex parametrisiertes Anregesignal, welches die Amplituden der Partialwellen des Sinusoidalmodells erzeugt. Das Filter bildet den Resonanzkörper eines Instruments ab und beinhaltet daher eine Frequenzantwort, anhand derer die Amplituden der Partialwellen entsprechend ihrer Frequenz gedämpft oder verstärkt werden. Das abschließende Dämpfungsglied des Ansatzes modelliert eine frequenzabhängige Abnahme der Amplitude über einen konstanten Zeitabschnitt. Somit beinhaltet das Modell bereits eine explizite Abhängigkeit des Timbre von der Grundfrequenz, da sowohl das Filter als auch das Dämpfungsglied, die aus der Grundfrequenz resultierenden Frequenzen der Partialwellen auswerten.

Die drei Bestandteile des Modells bestehen dabei jeweils aus einer Menge gewichteter Basisfunktionen, wobei als Basisfunktion der Quelle eine cosinusförmige Funktion gewählt wurde und für die Filter- bzw. Dämpfungskomponente Dreiecksfunktionen mit frequenzabhängigen Bandbreiten. Die Gewichtungsparameter der Komponenten werden anhand einer Trainingsdatenbasis, bestehend aus den Sinusoidalmodellen der Einzelklänge eines jeden Instruments, mittels eines Minimale-Quadrate-Schätzer (*Least-Squares-Estimator*) ermittelt.

Tuomas Virtanen hat hierzu in [29] ein Verfahren zur Verwendung des Quelle-Filter-Dämpfungs-Systems in einer nicht negativen Matrixfaktorisierung zur Quellentrennung vorgestellt und in [13] wird der Einsatz des Modells zur Instrumentenerkennung mit vorgeschalteter Quellentrennung geschildert.

Im Gegensatz zu dem von Burred vorgeschlagenen Modell, besitzt der Ansatz von Klapuri bereits eine explizite Abhängigkeit der Timbre von der Grundfrequenz. Allerdings weist auch dieser Modellansatz einige Nachteile auf. So besitzt das Modell keine Möglichkeit zur Abbildung des zur Diskrimination der Timbre hilfreichen Ein- und Ausschwingverhaltens, da mit Hilfe des Dämpfungsglieds nur ein frequenzabhängiges Abklingverhalten über eine konstante Zeiteinheit abgebildet werden kann. Darüber hinaus werden sowohl das Resonanzfilter, als auch das Dämpfungsfilter über eine Superposition stückweise linearer Funktionen konstruiert, wodurch eine Beschreibung der jeweiligen Frequenzantwort in Form einer stetig differenzierbaren Funk-

tion verhindert wird.

Kapitel 4

Das Modell

Im aktuellen Kapitel wird das Instrumentenmodell vorgestellt, das den zentralen Bestandteil dieser Arbeit darstellt. Hierbei wird in Abschnitt 4.2 neben einer detaillierten Schilderung des Modellaufbaus auch seine Verwendung im Kontext der Vorhersage von zeit-veränderlichen Amplitudenhüllkurven dargelegt. Darüber hinaus werden seine statistischen Eigenschaften und eine Analyse der Modellkomplexität gegeben. Abschnitt 4.3 zeigt im Anschluss, wie die Modellparameter anhand einer Trainingsdatenbasis auf Grundlage eines Gradientenabstiegsverfahrens geschätzt werden können. Vorab wird in Abschnitt 4.1 jedoch eine Beschreibung über das dem Instrumentenmodell zugrunde liegende Signalmodell gegeben.

4.1 Das Signalmodell

Basierend auf der Annahme, dass die in Abschnitt 2.2 vorgestellten strukturellen Dimensionen der Klangcharakteristika musikalischer Instrumente durch die Zeitverläufe der Amplituden der einzelnen Partialwellen bestimmt sind, dient das in Absatz 2.3.4 gezeigte Sinusoidalmodell als Grundlage der Betrachtungen über ein parametrisches Modell zu ihrer Repräsentation. Daraus folgt, dass die Klangcharakteristika nur für die harmonischen Partialwellen der Signale erfasst und abgebildet werden sollen und daher ausschließlich die Amplituden der Partialwellen $a[k, r]$ und die jeweilige Grundfrequenz f_0 der

Signale betrachtet werden.

4.1.1 Approximation der Frequenzwerte der Partialwellen

Wie die Amplitudenwerte einer jeden Partialwelle, so sind im Sinusoidalmodell auch deren jeweilige Frequenzwerte f über k und r parametrisiert. Allerdings kann, auf Grund der Tatsache, dass sowohl das Modell ausschließlich Einzeltöne vorhersagen soll, als auch dass die Eingangsdaten in Form separater Einzeltöne vorliegen, von einem quasi-stationären Verlauf der Frequenzwerte f über die Blockindizes r ausgegangen werden. Darüber hinaus ist die Modellbildung auf quasi-harmonische Instrumente beschränkt, sodass die Frequenzwerte der Partialwellen ferner als ganzzahlige Vielfache ihrer Grundfrequenz approximiert werden können. Diese Approximation ist in Gleichung 4.1 dargestellt und reduziert signifikant die anschließende Modellbildung in Bezug auf ihren Speicheraufwand.[1]

$$f(k) = f_0 \cdot k \quad , \quad k = 1...K \tag{4.1}$$

Mit der Grundfrequenz f_0 und der Gesamtanzahl an im Signal vorliegenden Partialwellen K ergibt sich $f(k)$ zu einer über r konstanten Folge von Frequenzwerten für die einzelnen Amplitudenwerte $a[k, r]$.

4.1.2 Skalierung der Amplituden

Auf Grund der Tatsache, dass weder die Amplitudenwerte zeit-diskreter Signale $x[n]$, noch die Amplitudenwerte $a[k, r]$ der Partialwellen ohne entsprechenden Referenzwert einen Rückschluss auf den tatsächlich vorherrschenden physikalischen Schalldruck p zum Zeitpunkt der Aufnahme zulassen, werden die Amplitudenwerte der Partialwellen anhand der relativen Energie des ap-

[1] Ich verwende die in der Literatur gebräuchliche Bezeichnung f_0 für die Grundfrequenz, auch wenn es dadurch an dieser Stelle zu der etwas verwirrenden Konsequenz von $f_0 = f(1)$ führt. Dies begründet sich allerdings durch die in der Literatur häufig verwendeten Begriffe von Grund- und Oberton und der von mir verwendeten Bezeichnung der Partialwellen. Siehe dazu Absatz 2.2.1

proximierten Signals $\tilde{x}[n]$ skaliert. Die Signalenergie wird hierfür über den einzelnen Zeitblöcken aus den Amplitudenwerten aller Partialwellen berechnet (Gleichung 4.2). Im Anschluss werden die über die maximale Signalenergie skalierten Amplitudenwerte der Partialwellen in Pegelwerte überführt (Gleichung 4.3).

$$E[r] = \sum_{k=1}^{K} \big(a[k,r]\big)^2 \tag{4.2}$$

$$A[k,r] = 10 \cdot \log_{10}\left(\frac{a[k,r]^2}{\max_r(E[r])}\right) \tag{4.3}$$

Dies hat zur Folge, dass im Rahmen der Modellbildung keine Aussage über die Klangeigenschaften zu einer spezifischen Lautstärke getroffen werden kann, sondern die spektrale Hüllkurve ausschließlich in Abhängigkeit der relativen Signalenergie abgebildet werden kann. Die in Absatz 2.2.2 beschriebene Dynamikdimension kann demnach ohne Referenzwert weder aus vorliegenden Signalen gewonnen noch von einem Modell abgebildet werden. Auch Aufnahmen in den verschiedenen Dynamikstufen *m, f, p* etc. werden skaliert, indem jeweils ihre summierte Signalenergie normiert wird.

4.1.3 Teilung der spektralen Hüllkurve

Da die Amplitudenfunktionen über den Blockindizes r nicht-stationär und somit veränderlich sein können, kann die Form der spektralen Hüllkurve über die Zeit variieren. Diese Veränderlichkeit gilt, wie in Absatz 2.2.3 beschrieben, insbesondere während der Ein- und Ausschwingdauern eines Instrumentenklangs. Da jedoch die tatsächlichen Dauern des Attack/Release aber auch die des Gesamtsignals stark variieren können, ist eine Aussage anhand der Zeitindizes ungeeignet. Daher sei angenommen, dass die Variationen der spektralen Hüllkurve mit der relativen Signalenergie korrelieren. Somit gilt die Annahme, dass die spektrale Hüllkurve für einen spezifischen Wert der relativen Energie konstant ist und für verschiedene Werte verschiedene Hüllkurven erlaubt sind. Wie bereits die Amplitudenfunktionen der Partialwellen wird

auch die relative Signalenergie in Pegelwerte umgewandelt (Gleichung 4.4).

$$L[r] = 10 \cdot \log_{10} \left(\frac{E[r]}{\max_r(E[r]} \right) \tag{4.4}$$

Hierbei muss nun in Betracht gezogen werden, dass Zeitindizes, deren Pegelwerte signifikant unterhalb des Maximalwerts von 0dB liegen, entweder dem Ein- oder dem Ausschwingvorgang zuzurechnen sind, die spektralen Hüllkurven für diese Bereiche jedoch voneinander verschieden sein können. Daher sollen die Amplitudenhüllkurven in zwei Bereiche partitioniert und in der Modellbildung getrennt voneinander betrachtet werden. Hierzu müssen jedoch sowohl der *Attack*- als auch der *Release*-Zeitpunkt ermittelt werden. Dazu wird eine Schwellwertmethode verwendet, bei der die Zeitpunkte anhand eines Über- bzw. Unterschreitens des Schwellwertes in der Pegelfunktion $L[r]$ detektiert werden, wobei zwischen einer kontinuierlichen oder impulshaften Form der Anregung des Signals unterschieden werden muss.

Teilung bei kontinuierlicher Anregung

Für den Fall einer kontinuierlichen Anregung gehen wir, wie in Absatz 2.2.3 beschrieben, von einer Einschwingzeit, einem stationären Zustand und einer Ausschwingphase aus und verwenden einen Schwellwert γ unterhalb des relativen Signalmaximums von 0dB. Wie in Gleichung 4.5 und 4.6 gezeigt, werden die Zeitpunkte des *Attack*- respektive *Release*-Zeitpunkts an der ersten bzw. letzten Überschreitung des Schwellwertes innerhalb der Pegelfunktion $L[r]$ detektiert.

$$r_A = \text{first}(r_s), \quad r_s : L[r] > \gamma \tag{4.5}$$

$$r_R = \text{last}(r_s), \quad r_s : L[r] > \gamma \tag{4.6}$$

In Abbildung 4.1 ist darüber hinaus veranschaulicht, inwiefern sich mit der Bestimmung der *Attack*- und *Release*-Zeitpunkte die Abschnitte für die Ein- und Ausschwingphase sowie den quasi-stationären Bereich eines kontinuierlich angeregten Signals feststellen lassen. Anhand der so detektierten Signalbereiche müssen nun die Abschnitte bestimmt werden, in die das Signal

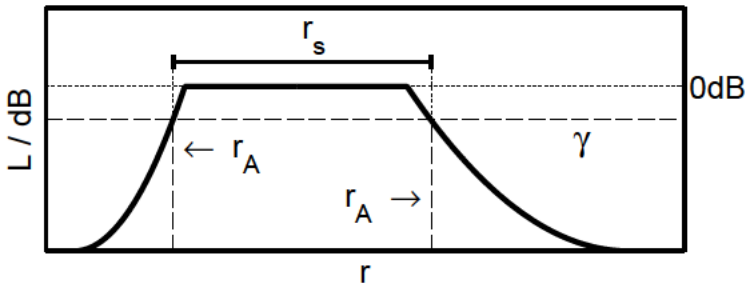
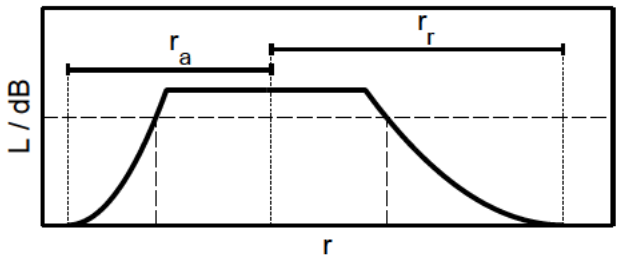


Abbildung 4.1: Bestimmung der *Attack/Release*-Zeitpunkte bei kontinuierlicher Anregung

partitioniert werden soll und die für die Modellbildung getrennt betrachtet werden. Für die Teilung der Amplitudenfunktionen der Partialwellen für kontinuierlich angeregte Signale sollen im Rahmen dieser Magisterarbeit zwei Verfahren evaluiert werden.

Angrenzende Partitionierung: Im ersten Verfahren werden kontinuierlich angeregte Signale an Hand des zeitlichen Mittelpunktes zwischen n_A und n_R in einen *Attack/Sustain*- und einen *Sustain/Release*-Bereich geteilt. Dies ist in Abbildung 4.2 veranschaulicht und zeigt die Definition der Zeitindizes n_a und n_r an Hand derer die Amplitudenfunktionen der Partialwellen dem jeweiligen Bereich zugeordnet werden. Die Bereiche stoßen direkt aneinan-

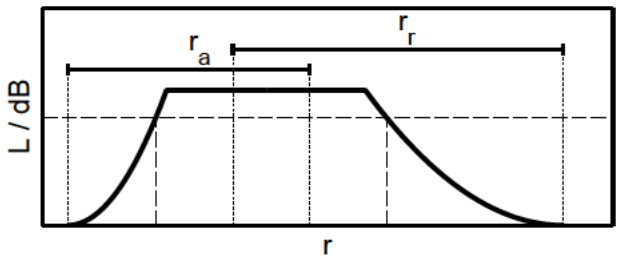


$$r_a \quad : \quad r \leq \frac{1}{2}(r_A + r_R)$$
$$r_r \quad : \quad r > \frac{1}{2}(r_A + r_R)$$

Abbildung 4.2: Angrenzende Bereichsgrenzen

der an und die Zeitindizes r_a und r_r entstammen daher zweier disjunkter Mengen.

Überlappende Partitionierung: Eine Alternative zum obigen Teilungsverfahren verwendet überlappende Bereiche, wobei die Grenzen so gesetzt werden, dass sich die einzelne Abschnitte ausschließlich in dem als quasi-stationär angenommenen Bereich überlappen. Abbildung 4.3 zeigt das Verfahren, wobei in diesem Fall die Indizes r_a und r_r nicht disjunkter Mengen entstammen, sondern den überlappenden Bereich an Hand ihrer Schnittmenge definieren.



$$r_a \quad : \quad r \leq \frac{1}{3}(r_A + 2r_R)$$
$$r_r \quad : \quad r \geq \frac{1}{3}(2r_A + r_R)$$

Abbildung 4.3: Überlappende Bereichsgrenzen

Teilung bei impulshafter Anregung

Eine impulshafte Anregung stellt gewissermaßen einen Sonderfall der kontinuierlichen Anregung mit unendlich kurzem quasi-stationären Bereich dar. Daher muss in diesem Fall bei einer Partitionierung nur auf die Einteilung in einen *Attack*- und einen *Release*-Bereich geachtet werden, da davon ausgegangen werden kann, dass kein *Sustain*-Bereich im Signal vorhanden ist. Die Bestimmung von n_A und n_R kann somit ebenfalls mit Gleichung 4.5 respektive 4.6 erfolgen wobei der Schwellwert γ auf 0dB gesetzt werden muss. Dies hat zur Folge, dass $n_A = n_R$.

Ohne einen quasi-stationären Abschnitt lässt sich nur eine Einteilung in Form der angrenzenden Partitionierung vornehmen, da eine überlappende zur Folge hätte, dass in beiden Bereichen sowohl Signalanteile des *Attack* als auch des *Release* erfasst würden.

Angrenzende Partitionierung: Sie entspricht der gleichnamigen Methode für kontinuierlich angeregte Signale, kann allerdings, wie Abbildung 4.5

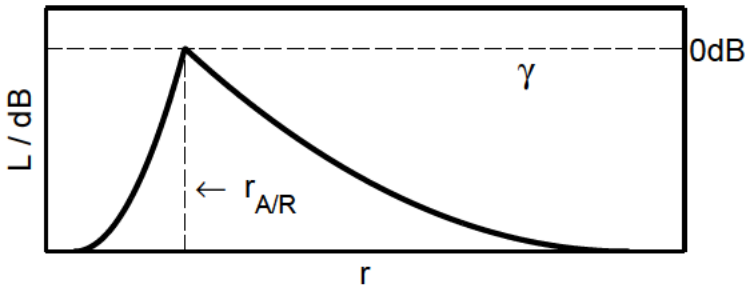
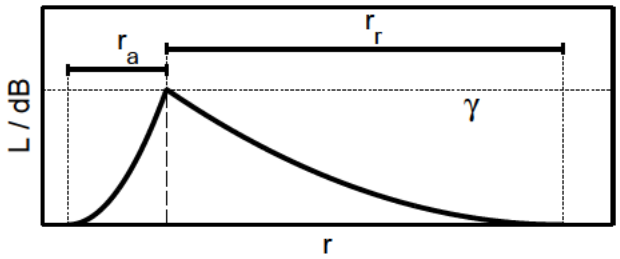


Abbildung 4.4: Bestimmung des *Attack/Release* bei impulshafter Anregung

zeigt, mathematisch einfacher dargestellt werden.



$$r_a \quad : \quad r \leq r_{A/R}$$
$$r_r \quad : \quad r > r_{A/R}$$

Abbildung 4.5: Bereichsgrenzen bei impulshafter Anregung

Aus der Teilung der Partialwellenfunktionen für jeden Einzelton der Eingangsdaten folgen je zwei getrennte Amplitudenfunktionen $A[k, r_a]$, $A[k, r_r]$, wobei die Folge an Frequenzwerten für die Partialwellen $f(k)$ für beide Amplitudenfunktionen ihre Gültigkeit behält. Die jeweilige Anzahl an Zeitblöcken für die zwei Bereiche wird im Folgenden mit N_A für den *Attack/Sustain*- bzw. *Attack*-Bereich und mit N_R für den *Sustain/Release*- respektive *Release*-Bereich angegeben.

4.2 Das Quelle-Filter-Modell

Die zentrale Anforderung an ein parametrisches Modell zur Repräsentation der Klangcharakteristika eines musikalischen Instruments besteht darin, die in Abschnitt 4.1 besprochenen Amplitudenfunktionen der Partialwellen abbilden und wiedergeben zu können. Das Modell soll demnach mit geeigneten

Parametern befähigt sein, die zeitveränderliche, spektrale Hüllkurve für eine spezifische Grundfrequenz eines Instruments vorhersagen zu können. Eine solche Instanz des Modells mit den geschätzten Parametern für ein Instrument wird im Folgenden als Prototyp bezeichnet.

Neben der zuvor genannten Anforderungen an die Modellbildung, ist ein zentraler Anspruch, dies in einer möglichst kompakten Repräsentation zu erzielen. Insbesondere die Abhängigkeit der spektralen Hüllkurve von der Grundfrequenz soll nicht implizit im Modell enthalten sein, sondern ausdrücklich in Form einer funktionalen Abhängigkeit abgebildet werden.

Die Grundidee des Modellansatzes basiert daher auf der Annahme, dass sich die Klangcharakteristika musikalischer Instrumente einteilen in jene, die mit der Grundfrequenz korrelieren und solche, die unabhängig von der Grundfrequenz die Klangfarbe beeinflussen. Eigenschaften, die mit der Grundfrequenz korrelieren, beziehen sich beispielsweise auf Klangcharakteristika von Instrumenten, bei denen die ungeradzahligen Partialwellen stärker ausgeprägt sind als deren geradzahlige Nachbarn. Diese Eigenschaften lassen sich daher besser als Funktion des Partialwellenindex k beschreiben als über tatsächliche Frequenzwerte. Im Gegensatz dazu beschreiben Resonanzen und Formanten Eigenschaften, die unabhängig von der Grundfrequenz den Klang beeinflussen und daher explizit über Frequenzwerte beschrieben werden müssen. Um diese Anforderungen und Ansätze zu vereinen wird ein physikalisch motiviertes, parametrisches Quelle-Filter-Modell eingesetzt.

Quelle-Filter-Modelle werden bereits seit vielen Jahren erfolgreich in der Sprachverarbeitung, insbesondere in der Sprachsynthese, verwendet [9], aber auch bei der Analyse und Beschreibung von Instrumentenklängen [28]. Im Rahmen dieser Arbeit stellen sowohl die Quelle als auch das Filter mathematische Abstraktionen separater, physikalischer Vorgänge dar. Entsprechend der Unterscheidung in f_0-korrelierte und von f_0 unabhängige Eigenschaften, werden Erstere mit Hilfe der Quelle als eine Funktion des Partialwellenindex ausgedrückt, während die von f_0 unabhängigen Eigenschaften mit Hilfe des über die Frequenz parametrisierten Filters beschrieben werden.

4.2.1 Die Quelle

Die Quelle repräsentiert das klangerzeugende Element eines Instruments. Dabei kann es sich zum Beispiel um eine schwingende Saite oder um eine schwingende Luftsäule handeln. Da die Quelle über dem Partialwellenindex k parametrisiert wird, kann sie als ein Oszillator betrachtet werden, der jede einzelne Partialwelle mit einer spezifischen Amplitude erzeugt. Somit ist das Anregesignal unabhängig von der Grundfrequenz und die Amplituden definieren sich über ihr Verhältnis untereinander. Darüber hinaus sei angenommen, dass, entgegen des in Abschnitt 3.2 geschilderten Ansatzes, auch die Zeitveränderlichkeit der spektralen Hüllkurve mit der Grundfrequenz korreliert. Da, wie bereits in Absatz 4.1.3 beschrieben, angenommen wird, dass die Zeitveränderlichkeit mit der relativen Signalenergie korreliert, wird der Oszillator zusätzlich über dem Pegel L der relativen Signalenergie parametrisiert. Die zeitliche Entwicklung eines Klangs soll somit nicht explizit im Modell abgebildet, sondern über die relative Energieentwicklung eingestellt werden. Ausgehend von der Annahme, dass die Entwicklung der Partialwellenfunktionen über die Energie stetig und nicht sprunghaft stattfindet, wird ihre Entwicklung mit einer kontinuierlichen Parametertrajektorie realisiert. Dazu werden stückweise Polynome eingesetzt, wobei für die Polynome so genannte *basic-splines* (B-Splines) [24] verwendet werden.

$$\sum_{l} b_l B_l(t) \tag{4.7}$$

$B_l(t)$	:	B-Spline Funktion
b_l	:	Gewichtungsparameter des l-ten B-Spline Polynoms

Die lineare Superposition gewichteter B-Splines (Gleichung 4.7) erzeugt maximal glatte Trajektorien, wobei die B-Splines vollständig definiert werden durch die Größe der Segmente und ihrer Ordnung o. Die Ordnung beschreibt dabei die Anzahl der Segmente, die von je einer B-Spline Funktion überspannt werden. Ihre jeweilige Polynomordnung entspricht daher $o-1$, wodurch wiederum ihr Grad an Glattheit bestimmt wird. Auf Grund der Zwangskonvergenz von B-Spline Polynomen auf den Wert 0 an den definierten Seg-

mentgrenzen, werden zusätzlich Segmente mit der Größe 0 an den Rändern des Definitionsbereichs eingeführt, um Trajektorien mit von 0 verschiedenen Werten an den Extremwerten des Definitionsbereichs modellieren zu können. Abbildung 4.6 zeigt exemplarisch den Verlauf von B-Spline Funktionen der Ordnung 3 über einen Energiebereich in Pegeldarstellung von -90 bis 0dB. Der gesamte Bereich ist dabei in 5 Segmente geteilt. An den Rändern sind jeweils zwei Segmente mit der Größe 0 angefügt, was zu einer Gesamtanzahl von 7 B-Spline Polynomen führt. Basierend auf dieser Definition gewichteter,

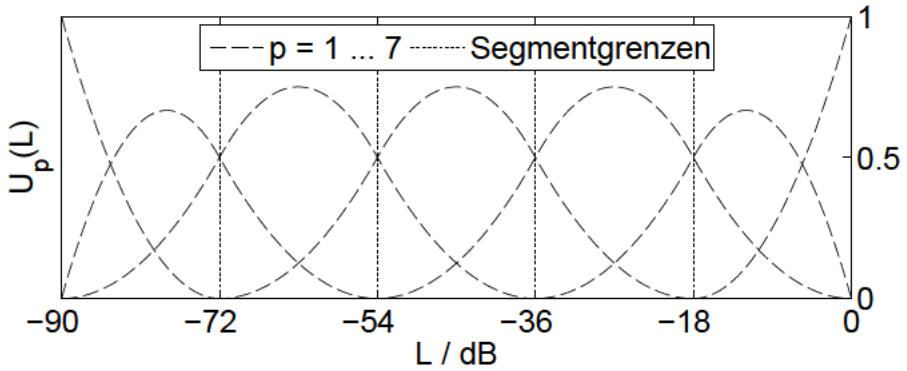


Abbildung 4.6: B-Spline Funktionen U_w über dem Pegel der relativen Energie L

stückweiser Polynome kann somit die Trajektorie einer einzelnen Partialwelle für einen spezifischen Bereich des relativen Signalpegels anhand einer Folge von Gewichtungsparametern der B-Splines modelliert werden. Um darüber hinaus eine spektrale Hüllkurve, bestehend aus einer Vielzahl von Partialwellen zu modellieren, benötigt man je eine Folge von Gewichtungsparametern zu je einer Partialwelle. Da jedoch von sich unterscheidenden spektralen Hüllkurven für die Bereiche des *Attack* und *Release* ausgegangen wird, werden zwei getrennte Oszillatoren mit jeweils eigenen Gewichtungsparametern definiert.

$$\mathrm{O}_A \quad : \quad O(k,L)_A = \sum_p^P g_{k,p}^A U_p(L) \tag{4.8}$$

$$\mathrm{O}_R \quad : \quad O(k,L)_R = \sum_p^P g_{k,p}^R U_p(L) \tag{4.9}$$

Das in Gleichung 4.8 und 4.9 gezeigte Oszillatormodell besteht somit aus zwei voneinander unabhängigen Anregeoszillatoren mit den über k und p indizierten Gewichten g. Je nachdem, ob es dabei sich um eine kontinuierliche oder eine impulshafte Anregung handelt, symbolisieren sie so entweder den *Attack-Sustain*- oder *Attack*-Bereich (O_A) oder den Bereich des *Sustain-Release* oder *Release* (O_R). Die einzelnen B-Spline Funktionen U_p sind dabei für alle Partialwellen identisch und daher ausschließlich über dem relativen Energiepegel L des Gesamtsignals parametrisiert.

4.2.2 Das Filter

Das Filter stellt das klangmodulierende Element eines Instruments dar und dämpft oder verstärkt einzelne Komponenten der, von der Quelle erzeugten, Partialwellenfunktionen $O(k,L)$ in Abhängigkeit ihrer Frequenz $f(k)$ und ist daher von der eigentlichen Grundfrequenz f_0 unabhängig. Diese Eigenschaft kann im allgemeinen dem Korpus eines Instruments zugewiesen werden. Auf Grund der Verstärkungseigenschaft ist auch der Begriff des Resonanzfilters respektive Resonanzkörpers gebräuchlich. Allgemein können dem Filter alle Bestandteile eines Instruments zugerechnet werden, die nicht direkt dem klangerzeugenden Element zugeordnet werden. Zudem sei angenommen, dass das Filter zeit-invariant ist. Dieser Ansatz vereinfacht die Modellbildung erheblich und ermöglicht eine Betrachtung des Filters unabhängig vom Energieverlauf und somit konstant für die zwei getrennten Anregeoszillatoren.

Da die Klänge von Instrumenten ausschließlich über die Amplitudenfunktionen der Partialwellen beschrieben werden, beschränkt sich die Menge an Informationen über die Frequenzantwort des Filters auf die jeweiligen Fre-

quenzwerte $f(k)$ der einzelnen Partialwellen. Da jedoch, wie bereits für die Trajektorien der Energieverläufe der Oszillatoren, ein kontinuierlicher Verlauf der Frequenzantwort angenommen werden kann, wird auch das Filter mit Hilfe von B-Spline Polynomen modelliert. Im Gegenzug zu den Oszillatoren müssen für das Filter jedoch die Segmentgrenzen in der spektralen Domäne gesetzt werden. Für Resonatorspektren seien markante und deutlich ausgeprägte Resonanzen im unteren Frequenzbereich angenommen, wohingegen für höhere Frequenzen weniger hervorstechende dafür jedoch sehr dichte Resonanzen erwartet werden. Daher werden die Segmentgrenzen frequenzabhängig als Vielfache von Oktavbandbreiten bestimmt, beginnend mit der je Instrument tiefsten spielbaren Grundfrequenz bis zu einem durch die Nyquist-Frequenz f_{ny} vorgegebenem Frequenzmaximum. Folglich werden durch das Filter Resonanzen und Formanten bei tiefen Frequenzen mit einer höheren Genauigkeit abgebildet, während Resonanzen für höhere Frequenzen stärker gemittelt repräsentiert werden. Abbildung 4.7 zeigt an einem Beispiel B-Spline Funktionen mit der Ordnung $o = 4$ über drei Oktaven und einer Segmentbreite von 1/3 Oktave. Somit lässt sich das Filter,

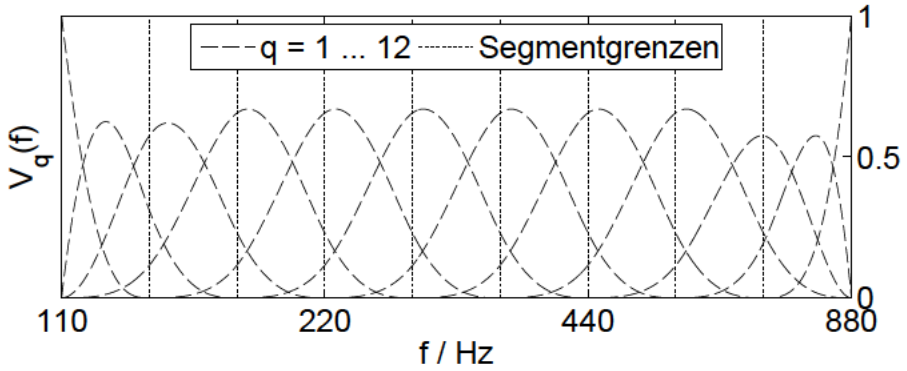


Abbildung 4.7: B-Spline Kurven V_q über der Frequenz f

wie in Gleichung 4.10 gezeigt, als eine gewichtete, lineare Superposition von B-Spline-Funktionen darstellen, die über der Frequenz parametrisiert sind.

$$\mathrm{F} \quad : \quad F(f) = \sum_{q}^{Q} z_q V_q(f) \tag{4.10}$$

Die Parameter z_q beschreiben demnach eine die jeweilige B-Spline Funktion gewichtende Sequenz der Länge Q. Die tatsächliche Länge dieser Sequenz und somit Anzahl an B-Spline Funktionen ergibt sich aus der notwendigen Menge an Segmenten mit der jeweiligen Bandbreite, um den gesamten Frequenzbereich von der tiefsten spielbaren Note bis zum Frequenzmaximum zu füllen.

4.2.3 Die Quelle-Filter-Summen

Auf Grund der Tatsache, dass ausschließlich logarithmierte Werte für die Amplitudenfunktionen der Partialwellen als auch für die Frequenzantwort des Filter verwendet werden, kann das Quelle-Filter-System in Form eines linearen Modells als einfache Summe aus Quelle und Filter beschrieben werden, wobei zwei getrennte Amplitudenfunktionen für die Teilbereiche des *Attack/Sustain* und *Sustain/Release* definiert werden.

$$\begin{aligned} \hat{A}(k, L, f_0)_A &= O(k, L)_A + F(f(k)) \qquad (4.11) \\ &= \sum_p^P g_{k,p}^A U_p(L) + \sum_q^Q z_q V_q(f_0 \cdot k) \\ \hat{A}(k, L, f_0)_R &= O(k, L)_R + F(f(k)) \qquad (4.12) \\ &= \sum_p^P g_{k,p}^R U_p(L) + \sum_q^Q z_q V_q(f_0 \cdot k) \end{aligned}$$

Das Modell schätzt demnach die Amplitude $\hat{A}$ einer Partialwelle k zu einem konkreten Wert L der relativen Gesamtenergie und der aus einem gegebenen f_0 resultierenden Frequenz $f(k)$ der Partialwelle. Dies jedoch getrennt für den Bereich des *Attack-Sustain* bzw. *Attack* und für den Bereich des *Sustain-Release* respektive *Release*.

Damit ausgehend von der Schätzung der Amplituden der Partialwellen zu einem spezifischen Wert der Gesamtenergie die zeitveränderliche, spektrale Hüllkurve eines Signals unbekannter Herkunft prädiktiert werden kann, wird demnach neben der Grundfrequenz f_0 und der Anzahl der Partialwellen K auch der Verlauf der relativen Energie in Form der Pegelfunktion $L[r]$

benötigt. Diese Daten müssen daher zuvor aus jedem einzelnen Datum der Eingangsdatenbasis extrahiert werden und stellen damit die Grundlage einer jeden Prädiktion des Modells dar.

4.2.4 Projektion der Eingangsdaten

Ausgehend von einem festen Satz an vordefinierten B-Spline-Funktionen U_p und V_q können die Abbildungsmatrizen für den Energieverlauf $L[r]$ als auch der Frequenzfolge $f(k)$ bestimmt werden. Dies entspricht einer Projektion der beobachteten Amplitudenfunktionen $A[k, r_{a/r}]$ vom Raum der Eingangsdaten in den Modellraum.

Abbildung für die Anregeoszillatoren:

Gleichung 4.8 und 4.9 zeigen, dass die B-Spline-Polynome U_p der Anregeoszillatoren über dem relativen Pegel L parametrisiert sind. Für eine Projektion der Eingangsdaten in den Raum der B-Spline Funktionen müssen also die Werte der B-Splines in Abhängigkeit des relativen Pegelverlaufs $L[r]$ der Eingangsdaten berechnet werden. Da jedoch, wie in Absatz 4.1.3 beschrieben, eine Teilung der Eingangsdaten in die Abschnitte r_a und r_r vorgenommen wird, müssen demnach die Werte der B-Spline Funktionen sowohl für den Pegelverlauf $L[r_a]$ des *Attack/Sustain* Bereichs als auch für den Verlauf $L[r_r]$ des *Sustain/Release*-Bereichs getrennt berechnet werden. Dementsprechend erhält man für die B-Spline Funktionen im ersten Fall die Abbildung $U_p(L[r_a])$ und im zweiten Fall die Abbildung $U_p(L[r_r])$. Beide Abbildungen ergeben eine Wertematrix in den Dimensionen $[P \times N_A]$ bzw. $[P \times N_R]$.

Abbildung für die Filterfunktion:

Das Filter ist in Gleichung 4.10 als Superposition von B-Spline Polynomen V_q über der Frequenz f beschrieben worden. Da jedes Datum der Eingangsdaten, eine bezüglich seiner Grundfrequenz f_0 spezifische Folge von Frequenzwerten $f(k)$ besitzt, muss diese auf die B-Spline Funktionen V_q abgebildet werden. Diese ist konstant über alle Zeitblöcke r und damit auch für alle Energiewerte. Somit ergibt sich $V_q(f(k))$ als zeit-invariante Abbildung in

den Dimensionen $[Q \times K]^2$, die sowohl für den *Attack/Sustain* als auch den Bereich des *Sustain/Release* gültig ist,

4.2.5 Prädiktion

Mit bekannten Abbildungsmatrizen $U_p(L[r_a])$, $U_p(L[r_r])$ und $V_q(f(k))$ für ein spezifisches Eingangsdatum bezüglich eines Prototypen, können anhand der Gewichte $g^A_{k,p}$, $g^R_{k,p}$ und z_q des Modells die spektralen Hüllkurven für den *Attack/ Sustain* als auch den Bereich des *Sustain/Release* geschätzt werden. Hierzu werden die Modellsummengleichungen 4.13 und 4.14 um die funktionalen Abhängigkeiten der relativen Signalpegel $L[r_a]$ und $L[r_r]$ erweitert.

$$\begin{aligned} \hat{A}(k, L[r_a], f_0)_A &= O(k, L[r_a])_A + F(f(k)) \qquad (4.13) \\ &= \sum_p^P g^A_{k,p} U_p(L[r_a]) + \sum_q^Q z_q V_q(f_0 \cdot k) \\ \hat{A}(k, L[r_r], f_0)_R &= O(k, L[r_r])_R + F(f(k)) \qquad (4.14) \\ &= \sum_p^P g^R_{k,p} U_p(L[r_r]) + \sum_q^Q z_q V_q(f_0 \cdot k) \end{aligned}$$

Die prädiktierten, spektralen Hüllkurven konstruieren sich demnach über der Anzahl der Partialwellen K und den Blocklängen N_A für den *Attack/ Sustain* respektive N_R für den *Sustain/Release* Bereich.

4.2.6 Synthese

Um mit Hilfe der Synthesegleichung 2.9 für Sinusoidalmodelle aus den Prädiktionen der Amplitudenverläufe der Partialwellen ein Audiosignal $\hat{x}[r]$ zu erzeugen, müssen die Amplituden für die Bereiche des *Attack/Sustain* und *Sustain/Release* miteinander verknüpft werden, so dass eine das Gesamtsi-

[2]Hier zeigt sich, wieso die Annahme des quasi-stationären Verhaltens der Partialwellen bezüglich ihrer Frequenz die Modellbildung vereinfacht, da sich im umgekehrten Fall die Abbildungsmatrix zu einem Abbildungstensor mit den Dimensionen $[Q \times K \times N_A]$, bzw. $[Q \times K \times N_R]$ erweitern würde

gnal repräsentierende Amplitudenfolge $\hat{A}[k,r]$ erzeugt wird. Wie bereits bei der Einteilung der Originalsignale, muss hierbei allerdings wieder zwischen den Formen der Anregung des Signals unterschieden werden. Die Frequenzen der einzelnen Partialwellen $f(k)$ müssen dabei nicht gesondert betrachtet werden, da sie, als quasi-stationär und quasi-harmonisch angenommen, bereits zur Prädiktion vorliegen.

Konnektion bei kontinuierlicher Anregung

Bei der Partitionierung von Signalen mit kontinuierlicher Anregung wird zwischen zwei Varianten unterschieden. Dem muss daher auch für die Verbindung der prädiktierten Amplitudenfolgen Rechnung getragen werden.

Angrenzende Konnektion Die Verbindung der Amplitudenfolgen im Fall der angrenzenden Partitionierung stellt den trivialen Fall dar, da hier die Zeitindizes r_a und r_r zweier disjunkter Mengen entstammen, die darüber hinaus keine Lücke aufweisen. Gleichung 4.15 zeigt wie die resultierende Amplitudenfolge $\hat{A}[k,r]$ durch die direkte Aneinanderreihungen der Teilfolgen $\hat{A}(k, L[r], f_0)_A$ und $\hat{A}(k, L[r], f_0)_R$ konstruiert wird.

$$\hat{A}[k,r] = \begin{cases} \hat{A}(k, L[r], f_0)_A & \text{für } r \in r_a \\ \hat{A}(k, L[r], f_0)_R & \text{für } r \in r_r \end{cases} \tag{4.15}$$

Überlappende Konnektion Für den Fall, dass die Originalsignale mit überlappenden Blockindizes r_a und r_r partitioniert wurden, wird für den gesamten überlappenden Bereich der Teilfolgen ein lineares Ein- und Ausblenden der Amplituden der Partialwellen vorgenommen. In Gleichung 4.16 wird dies auf der logarithmischen Dezibel-Skala mit Hilfe der Gewichte λ_1

und λ_2 ausgedrückt.

$$\hat{A}[k,r] = \begin{cases} \hat{A}(k, L[r], f_0)_A & \text{für } r \in r_a \text{ und } r \notin r_r \\ \lambda_1 \cdot \hat{A}(k, L[r], f_0)_A + \lambda_2 \cdot \hat{A}(k, L[r], f_0)_R & \text{für } r \in r_a \text{ und } r \in r_r \\ \hat{A}(k, L[r], f_0)_R & \text{für } r \in r_r \text{ und } r \notin r_a \end{cases} \tag{4.16}$$

Abbildung 4.8 zeigt wie die Gesamtsignale der Einzelanregungen von *Attack/Sustain* und *Sustain/Release* auf der logarithmischen Skala ineinander übergeblendet werden.

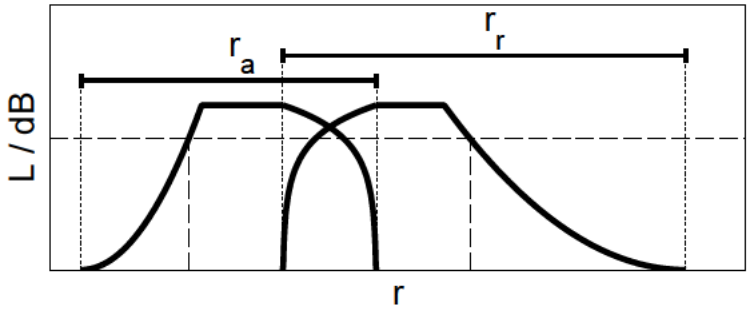


Abbildung 4.8: Konnektion der Anregungen bei überlappenden Bereichen

Konnektion bei impulshafter Anregung

Da bei der Einteilung der Signale im Fall einer impulshaften Anregung nur eine Form verwendet wurde, muss auch für die Verbindung der prädiktierten Amplitudenfolge nur diese eine Variante der Teilung betrachtet werden.

Angrenzende Konnektion Es gelten hier alle Aussagen von 4.2.6 inklusive der Gleichung 4.15, da sich die Art und Weise der Teilung der Signale ausschließlich anhand des Schwellwerts γ unterscheidet, der hier keine Rolle mehr spielt.

4.2.7 Modellvarianz

Da die prädiktierte, zeitveränderliche, spektrale Hüllkurve $\hat{A}[k,r]$ den Erwartungswert für den Verlauf einer solchen Hüllkurve eines Instrumentenklangs darstellt, kann, wie in Gleichung 4.17 gezeigt, die der realen Verteilung entstammende Stichprobe von beobachteten Daten $A[k,r]$ in Form eines linearen Modells als Summe aus dem Erwartungswert und einem additivem Rauschen $\mathcal{R}$ beschrieben werden.

$$A[k,r] = \hat{A}[k,r] + \mathcal{R}[k,r] \tag{4.17}$$

Unter der Annahme, dass es sich bei dem additiven Rauschen um eine gaussverteilte und vollständig unabhängige Größe handelt, kann die Varianz des Modells aus dem Rauschen geschätzt werden. Hierfür werden die Amplitudenwerte zuvor in die lineare Darstellung transformiert. Die Modellvarianz kann somit geschätzt werden anhand des Mittelwerts über den beobachteten Amplitudenverläufen $a[k,r]_\alpha$ der Trainingsdaten $\mathcal{A}$ und den aus der Pegelfunktion des Amplitudenverlaufs geschätzten Prädiktionen $\hat{a}[k,r]$ eines Prototypen.

$$\sigma^2(\mathcal{R}) = \frac{1}{\mathcal{A}} \sum_{\alpha=1}^{\mathcal{A}} \frac{1}{KN} \left|\left| a[k,r]_\alpha - \hat{a}[k,r]_\alpha \right|\right|_2^2 \tag{4.18}$$

Somit wird der Prototyp eines einzelnen Instruments beschrieben durch seine Gewichte $g^A_{k,p}$, $g^R_{k,p}$, z_q und seine Varianz σ^2. Das Modell enthält demnach neben der Schätzung der zeitveränderlichen Amplitudenhüllkurve auch ein statistisches Maß über die zu erwartende Abweichung von der Schätzung.

4.2.8 Modellkomplexität

Damit mit Hilfe des Modells ein Prototyp an Hand einer Trainingsdatenbasis geschätzt werden kann, müssen zuvor die B-Spline Funktionen für die Anregeoszillatoren und für das Filter fest definiert werden. Diese Definitionen beschreiben, in welcher Form und wie detailliert das Modell an die Daten angepasst wird und beinhaltet somit implizit die Modellkomplexität. Für sie

gilt, je exakter und detailreicher die Daten abgebildet werden können, desto größer die Komplexität des Modells. Für das Quelle-Filter-Modell bedeutet dies, dass die Modellkapazität eine mehrdimensionale Größe ist, die von den Definitionen für U_p und V_q abhängt. Aus der in Absatz 4.2.1 beschriebenen Definition der B-Spline Funktionen folgt daher, dass mit steigender Anzahl der Segmente, respektive kleiner werdenden Längen, die Modellkomplexität steigt, während sie mit ansteigender Ordnung der Polynome wiederum fällt. Da es sich bei einer solchen Schätzung jedoch um ein statistisches Verfahren mit verrauschten Daten handelt, bedeutet eine höhere Komplexität nicht zwangsläufig eine bessere Generalisierungsfähigkeit des Modells. Da die Fähigkeit der Generalisierung jedoch die eigentliche Güte des Modells darstellt, gilt es im Rahmen des Schätzverfahrens den optimalen Kompromiss aus hinreichend exakter Modellierung ohne Überanpassung zu erzielen.

4.2.9 Mathematische Notationskonventionen

Für die weitere mathematische Betrachtungsweise wird eine Vektor-/ Matrixnotation verwendet. Die dazu nötigen Konventionen zeigt Tabelle 4.1. Hierbei sei angemerkt, dass die Gewichte $\mathbf{z}_A$ und $\mathbf{z}_R$ über alle N_A, bzw. N_R konstant sind, da das Filter als zeit-invariant angenommen wurde.

4.3 Schätzen der Modellparameter

Damit die Parameter des Modell bestmöglich geschätzt werden können, müssen Aufnahmen der Klänge von Instrumenten bereitgestellt werden, die alle Eigenschaften der Klangcharakteristika beinhalten. Da insbesondere die Klangcharakteristika mit der jeweils gespielten Tonhöhe variieren, ist es darüber hinaus erforderlich, dass für die Trainingsdatenbasis alle Aufnahmen als separate Einzeltöne in der Form $\mathbf{A}_A$ und $\mathbf{A}_R$ und $f(k)$ vorliegen müssen.

Für die Bildung eines einzelnen Prototypen müssen unter der Bedingung eines festen Satzes an vordefinierten B-Spline-Funktionen U_p und V_q ausschließlich die freien Parameter $\mathbf{G}_A$, $\mathbf{G}_R$, und $\mathbf{z}$ geschätzt werden. Zum

Bezeichner der freien Parameter							
$\mathbf{G}_A$	$[K \times P]$	:	$g^A_{k,p}$	$\mathbf{G}_R$	$[K \times P]$	:	$g^R_{k,p}$
$\mathbf{z}_A$	$[Q \times N_A]$	:	z_q	$\mathbf{z}_R$	$[Q \times N_R]$	:	z_q
Bezeichner für die Datenvariablen							
$\mathbf{A}_A$	$[K \times N_A]$	:	$A[k, r_a]$	$\mathbf{A}_R$	$[K \times N_R]$	:	$A[k, r_r]$
$\mathbf{A}$	$[K \times N]$	:	$A[k, r]$				
Bezeichner für die Abbildungsmatrizen							
$\mathbf{U}_A$	$[P \times N_A]$	:	$U_p(L[r_a])$	$\mathbf{U}_R$	$[P \times N_R]$	:	$U_p(L[r_r])$
$\mathbf{V}$	$[Q \times K]$	:	$V_q(f(k))$				
Bezeichner für die Modellvorhersagen							
$\hat{\mathbf{A}}_A$	$[K \times N_A]$	:	$\hat{A}_A(k, L[r_a], f_0)$	$\hat{\mathbf{A}}_R$	$[K \times N_R]$	:	$\hat{A}_R(k, L[r_r], f_0)$
$\hat{\mathbf{A}}$	$[K \times N]$	:	$\hat{A}[k, r]$				

Tabelle 4.1: Matrix- und Vektor-Konventionen

Schätzen der Parameter für die Anregeoszillatoren und das Filter soll mit Hilfe der Trainingsdaten ein Gradientenabstiegsverfahren verwendet werden. Hierzu wird eine Kostenfunktion benötigt. Sie stellt ein Maß für die Abweichung des Modells von den Daten dar. Mit Hilfe des aus der 1. Ableitung gewonnenen Gradienten werden die Modellparameter so adaptiert, dass der Modellfehler, respektive die Kosten, minimiert werden.

4.3.1 Kostenfunktion

Für das Modell wird eine quadratische Kostenfunktion verwendet. Dies hat zweierlei Gründe. Einerseits ist die Bildung der Ableitung nicht sonderlich schwierig und andererseits gilt für quadratische Kosten in Verbindung mit einem linearen Modell, dass die gesamte Fehleroberfläche nur ein globales Minimum und keine lokalen Minima aufweist. Dies erleichtert signifikant

die anschließende Parameterschätzung. Gleichung 4.19 zeigt die quadratische Kostenfunktion, wobei die Kosten summiert über die zwei Oszillatoren und in Abhängigkeit der Segmentierung der Daten berechnet werden. Es handelt sich also um die lokalen Kosten für ein einzelnes Trainingsbeispiel α. Zur Berechnung der globalen Kosten bleibt die in Gleichung 2.10 vorgestellte Definition gültig.

$$\begin{aligned} c^{\alpha} &= \frac{1}{2K}\Bigg(\frac{1}{2N_A}\Big|\Big| \quad \hat{\mathbf{A}}_A \quad - \quad \mathbf{A}_A \quad \Big|\Big|_2^2 \\ &\qquad + \frac{1}{2N_R}\Big|\Big| \quad \hat{\mathbf{A}}_R \quad - \quad \mathbf{A}_R \quad \Big|\Big|_2^2 \Bigg) \\ &= \frac{1}{2K}\Bigg(\frac{1}{2N_A}\Big|\Big|\big(\mathbf{G}_A\mathbf{U}_A + \mathbf{V}^T\mathbf{z}_A\big) - \mathbf{A}_A\Big|\Big|_2^2 \\ &\qquad + \frac{1}{2N_R}\Big|\Big|\big(\mathbf{G}_R\mathbf{U}_R + \mathbf{V}^T\mathbf{z}_R\big) - \mathbf{A}_R\Big|\Big|_2^2 \Bigg) \end{aligned} \tag{4.19}$$

Mit $||\cdot||_2^2$ wird das Quadrat der Frobenius-Norm bezeichnet, welche die Summe der Quadrate über alle Einträge des inneren Ausdrucks meint.

4.3.2 Gradientenfunktionen

Die freien Parameter des Modells entsprechen, wie bereits in Absatz 4.2.2 geschildert, den B-Spline Koeffizienten $\mathbf{G}_A$ und $\mathbf{G}_R$ der Oszillatoren, sowie den B-Spline Koeffizienten $\mathbf{z}$ des Filters. Für die Parameterschätzung muss die Kostenfunktion nach den genannten Parametern abgeleitet werden. Die Bildung der Gradienten erfolgt in allen drei Fällen durch ein einmaliges Anwenden der Kettenregel und für die Oszillatoranregungen durch Streichen des jeweils konstanten Summanden. Es ergeben sich für die Anregeoszillatoren

die Gradientengleichungen 4.20 und 4.21.

$$\begin{aligned}\frac{\partial c^\alpha}{\partial \mathbf{G}_A} &= \frac{1}{N_A}\big(\quad \hat{\mathbf{A}}_A \quad - \quad \mathbf{A}_A \big)\mathbf{U}_A^T \\ &= \frac{1}{N_A}\big((\mathbf{G}_A\mathbf{U}_A + \mathbf{V}^T\mathbf{z}) - \quad \mathbf{A}_A \big)\mathbf{U}_A^T\end{aligned} \tag{4.20}$$

$$\begin{aligned}\frac{\partial c^\alpha}{\partial \mathbf{G}_R} &= \frac{1}{N_R}\big(\quad \hat{\mathbf{A}}_R \quad - \quad \mathbf{A}_R \big)\mathbf{U}_R^T \\ &= \frac{1}{N_R}\big((\mathbf{G}_R\mathbf{U}_R + \mathbf{V}^T\mathbf{z}) - \quad \mathbf{A}_R \big)\mathbf{U}_R^T\end{aligned} \tag{4.21}$$

Da die Gradienten für alle Sequenzen über K berechnet werden, ist wie in den Gleichungen zu sehen, die Skalierung mittels $1/K$ nicht mehr notwendig, ebenso wie die in der Kostenfunktion notwendige Skalierung mit dem Faktor $1/2$, die aus der Summe über den Oszillatoren herrührt.

Für den Gradienten der B-Spline-Koeffizienten $\mathbf{z}$ des Filters ergibt sich die aus beiden Summanden der Kostenfunktion bestehende Gleichung 4.22.

$$\begin{aligned}\frac{\partial c^\alpha}{\partial \mathbf{z}} &= \frac{1}{2K}\Big(\quad \mathbf{V}\frac{1}{N_A}\sum_{r_a}^{N_A}\big\{ \quad \hat{\mathbf{A}}_A \quad - \quad \mathbf{A}_A\big\} \\ &\qquad + \mathbf{V}\frac{1}{N_R}\sum_{r_r}^{N_R}\big\{ \quad \hat{\mathbf{A}}_R \quad - \quad \mathbf{A}_R\big\}\Big) \\ &= \frac{1}{2K}\Big(\quad \mathbf{V}\frac{1}{N_A}\sum_{r_a}^{N_A}\big\{(\mathbf{G}_A\mathbf{U}_A + \mathbf{V}^T\mathbf{z}) - \quad \mathbf{A}_A\big\} \\ &\qquad + \mathbf{V}\frac{1}{N_R}\sum_{r_r}^{N_R}\big\{(\mathbf{G}_R\mathbf{U}_R + \mathbf{V}^T\mathbf{z}) - \quad \mathbf{A}_R\big\}\Big)\end{aligned} \tag{4.22}$$

Hier bleibt die Mittelwertbildung über beide Oszillatoren nach wie vor notwendig, da $\mathbf{z}$ zeit-invariant und damit für beide Anregungeoszillatoren konstant ist. Dies hat gleichzeitig zur Folge, dass $\mathbf{z}$ nur für eine Sequenz der Länge Q bestimmt werden muss, da die Koeffizienten konstant über alle Zeitblöcke N_A und N_R sind.

4.3.3 Algorithmen

Anhand der vorgestellten Gradienten werden die Parameter in einem iterativen Verfahren angepasst und somit der Modellfehler optimiert. Wie in Abschnitt 2.4 beschrieben, können die Gradienten sowohl für ein *Online*- als auch für ein *Offline*-Verfahren eingesetzt werden, wobei sich die Verfahren dabei hauptsächlich in der *Update*-Regel zur Adaption der Gewichte $\mathbf{G}_A$, $\mathbf{G}_R$ und $\mathbf{z}$ unterscheiden. Darüber hinaus lassen sich auch einige algorithmenspezifische Optimierungen mit Hilfe verschiedener Start- und Abbruchbedingungen einstellen.

***Offline* Algorithmus:** Für den Offline-Lernalgorithmus wird je Epoche des Trainings der globale Gradient für alle freien Modellparameter anhand der gemittelten Summe der lokalen Gradienten bestimmt. Die in den Gleichungen 4.23, 4.24 und 4.25 gezeigten Parameteraktualisierungen werden mit der Schrittweite η skaliert und stellen einen einzelnen Schritt der Iteration bzw. Epoche dar.

$$\Delta\mathbf{G}_A = \eta\frac{\partial C}{\partial\mathbf{G}_A} = -\eta\frac{1}{P}\sum_{\alpha=1}^{P}\frac{\partial c^\alpha}{\partial\mathbf{G}_A} \tag{4.23}$$

$$\Delta\mathbf{G}_R = \eta\frac{\partial C}{\partial\mathbf{G}_R} = -\eta\frac{1}{P}\sum_{\alpha=1}^{P}\frac{\partial c^\alpha}{\partial\mathbf{G}_R} \tag{4.24}$$

$$\Delta\mathbf{z} = \eta\frac{\partial C}{\partial\mathbf{z}} = -\eta\frac{1}{P}\sum_{\alpha=1}^{P}\frac{\partial c^\alpha}{\partial\mathbf{z}} \tag{4.25}$$

Abbildung 4.9 zeigt den Algorithmus der *Offline*-Methode in Pseudocode mit den dazugehörigen *Update*-Regeln in den Zeilen 5 bis 7. Nach einer Epoche wird vom Algorithmus eine Anpassung der Lernschrittweite in Abhängigkeit von der Entwicklung des globalen Fehlers vorgenommen. Vor dem Eintritt in die nächste Epoche wird das Konvergenzkriterium überprüft und die Anpassung der Schrittweite erfolgt wie in den Zeilen 9 bis 13 gezeigt, wobei die Schrittweite bei kleiner werdendem Fehler geringfügig vergrößert und bei einer Vergrößerung des Fehlers halbiert wird.

```
1   t      ← 1
2   C(1)   ← ∞
3   η      ← 0.5
4   while !converged
5         G_A   ←  G_A   +  ΔG_A
6         G_R   ←  G_R   +  ΔG_R
7         z     ←  z     +  Δz
8       t ← t + 1
9       if C(t) < C(t − 1)
10            η ← η · 1.1
11      else
12            η ← η · 0.5
13      end
14  end
```

Abbildung 4.9: Pseudocode des Offline-Algorithmus

***Online* Algorithmus:** Bei der in Absatz 2.4.3 beschriebenen *Online*-Methode wird der Gradient lokal für jedes einzelne Trainingsbeispiel berechnet und die freien Parameter jeweils angepasst. Eine Trainingsepoche entspricht erneut einem kompletten Durchlauf aller Trainingsbeispiele in einer immer wieder neuen, zufälligen Reihenfolge, wobei die freien Parameter des Modells für jedes einzelne Trainingsbeispiel angepasst werden.

$$\Delta \mathbf{G}_A = -\eta \frac{\partial c^\alpha}{\partial \mathbf{G}_A} \tag{4.26}$$

$$\Delta \mathbf{G}_R = -\eta \frac{\partial c^\alpha}{\partial \mathbf{G}_R} \tag{4.27}$$

$$\Delta \mathbf{z} = -\eta \frac{\partial c^\alpha}{\partial \mathbf{z}} \tag{4.28}$$

Die *Update*-Regeln zur Anpassung der Gewichte sind in den Gleichungen 4.26, 4.27 und 4.28 dargestellt. Die Gewichtsänderungen werden demnach für alle Gewichte anhand des Gradienten des lokalen Fehlers c ermittelt und wieder mit der Schrittweite η skaliert.

Abbildung 4.10 zeigt das iterative Verfahren, wobei die Unterschiede zum *Offline*-Algorithmus in den veränderten Aktualisierungsregeln und Startbe-

dingungen liegen.

```
t      ← 1
C(1)   ← ∞
η      ← 0.05
while !converged
    for each sample i (randomized)
        G_A  ←  G_A  +  ΔG_A
        G_R  ←  G_R  +  ΔG_R
        z    ←  z    +  Δz
    end
    t ← t + 1
    if C(t) < C(t − 1)
        η ← η · 1.1
    else
        η ← η · 0.5
    end
end
```

Abbildung 4.10: Pseudocode des Online-Algorithmus

Kapitel 5

Implementierung

Die Implementierung des Modells erfolgte mit Hilfe der Simulationssoftware Matlab® 2007b. Bei dieser von der Firma *The Mathworks*™ entwickelten Software handelt es sich um eine höhere Programmiersprache mit integrierter Laufzeitumgebung, deren Anwendungsfeld im Bereich ingenieur- und naturwissenschaftlicher Algorithmenentwicklung, Datenanalyse und Simulationen sowie deren Evaluation und Visualisierung liegt. Im Vergleich zu anderen Hochsprachen liefert Matlab® sowohl eine einfache Syntax für mathematische Berechnungen in Vektor-, bzw. Matrixform, als auch eine Vielzahl an Programmbibliotheken für mathematisch-technische Algorithmen und Verfahren. Zudem bietet die Software eine leicht zu benutzende Schnittstelle zur Visualisierung komplexer Daten in Form von Diagrammen in zwei oder drei Dimensionen. Darüber hinaus lässt sich der Funktionsumfang von Matlab® durch so genannte *Toolboxes* erweitern und durch verschiedene Schnittstellen ist es möglich externe Programme und Bibliotheken einzubinden. Ein weiterer Vorteil liegt in dem hohen Verbreitungsgrad der Software im Bereich der Ingenieurs- und Naturwissenschaften, wodurch Matlab® einen de facto Standard darstellt. Da für die Umsetzung des Modells auf eine Vielzahl der genannten Funktionalitäten zurückgegriffen werden muss, stellt Matlab® die favorisierte Softwarelösung dar.

Allerdings bietet Matlab® in Version 2007b noch keine native Unterstützung von objektorientierter Programmierung an, weswegen in den meisten

Programmbausteinen ein einfacher Mechanismus verwendet wird, um zumindest die Konzepte der Kapselung und Persistenz zu realisieren. Hierzu wird in der als Konstruktor dienenden Funktion eine Struktur verwendet, die beliebige Strukturvariablen enthalten kann und alle Funktionen, die auf diese Variablen zugreifen sollen, werden im lokalen Gültigkeitsbereich des Konstruktors definiert. Alle Funktionen die von außen erreichbar sein sollen werden dann in Form von Funktionszeigern (*Callbacks*) als eigene Strukturvariablen gesetzt. Abbildung 5.1 zeigt den Mechanismus zum Erzeugen von Kapselung

```
1  function Object = TSomeObject()
2
3      Object = struct('val', [], 'get', @get, 'set', @set);
4
5      function v = get()
6
7          v = Object.val;
8
9      end
10
11     function set(v)
12
13         Object.val = v;
14
15     end
16
17 end
```

Abbildung 5.1: Quellcode Beispiel für objektbasierte Modellierung in Matlab® 2007b

und Objektpersistenz im Matlab® Syntax. Abbildung 5.2 verdeutlicht die Verwendung eines solchen Objekts auf der Matlab®-Konsole. Allerdings gibt

```
1  » Obj = TSomeObject();
2  » Obj.set(5.0);
3  » Obj.get();
4
5  ans
6
7      5.0
```

Abbildung 5.2: Beispielnutzung von Strukturobjekten mit *Callbacks*

es zwei Einschränkungen zu beachten. Einerseits sind kaskadierte Aufrufe

wie `Obj.getAnotherObject().set(v)` nicht möglich, andererseits kann auf Variablen, deren Wert mit Hilfe eines *Callbacks*, somit einer lokalen Funktion, verändert wurde, nicht mehr direkt von außen mit Hilfe von `Obj.var` zugegriffen werden, da die neue Referenz nur lokal bekannt ist.

Die Verwendung dieses Mechanismus hat sich im gesamten Entwicklungsprozess der Implementierung bewährt und ermöglicht eine effiziente und vor allem strukturierte Implementierung der einzelnen Programmbausteine. Allen Matlab®-Funktionen, die eine solche Struktur mit eingebetteten lokalen Funktionen zurückgeben, wird zur besseren Erkennung ein T vorangestellt. Da Matlab® in seiner neuesten Version jedoch eine native Unterstützung der objektorientierten Programmierung anbietet, ist somit ein umfangreiches *Refactoring* der bestehenden Quellcodebasis zu einer vollständigen und nativen Objektorientierung schnell und einfach möglich.

5.1 Aufbau der Software

Der Entwurf der Implementierung basiert auf der grundlegenden Anforderung, sämtliche Berechnungen und Verarbeitungsschritte in einem einzelnen Programm vollständig automatisiert ausführen zu können. Somit soll sichergestellt sein, dass bei den zu erwartenden rechenintensiven Aufgaben keine manuellen Eingaben den Verarbeitungsfluss unterbrechen und daher aufhalten können. Um ein solches Höchstmaß an Automatisierung zu erzielen, werden die einzelnen Arbeitsschritte in separate Verarbeitungsprozesse gegliedert und die jeweils notwendige Datenverwaltung, als auch die Verwaltung aller Parameter von der Software übernommen. Somit benötigt die Applikation keine grafische Benutzeroberfläche und kann vollständig über die Kommandozeile bedient werden, wobei die separaten Verarbeitungsprozesse in Form von Kommandozeilenparametern angesteuert und die benötigten Parameter und Datenspezifikationen mit Hilfe von Konfigurationsdateien festgelegt werden. In Abbildung 5.3 ist der schematische Aufbau der Applikation gezeigt, wobei insbesondere die Ein-/Ausgabe Relationen als auch die separaten Verarbeitungsprozesse verdeutlicht sind. Als Programmeingaben sind die Übergabeparameter während des Programmaufrufs (`Argumente`) als auch die

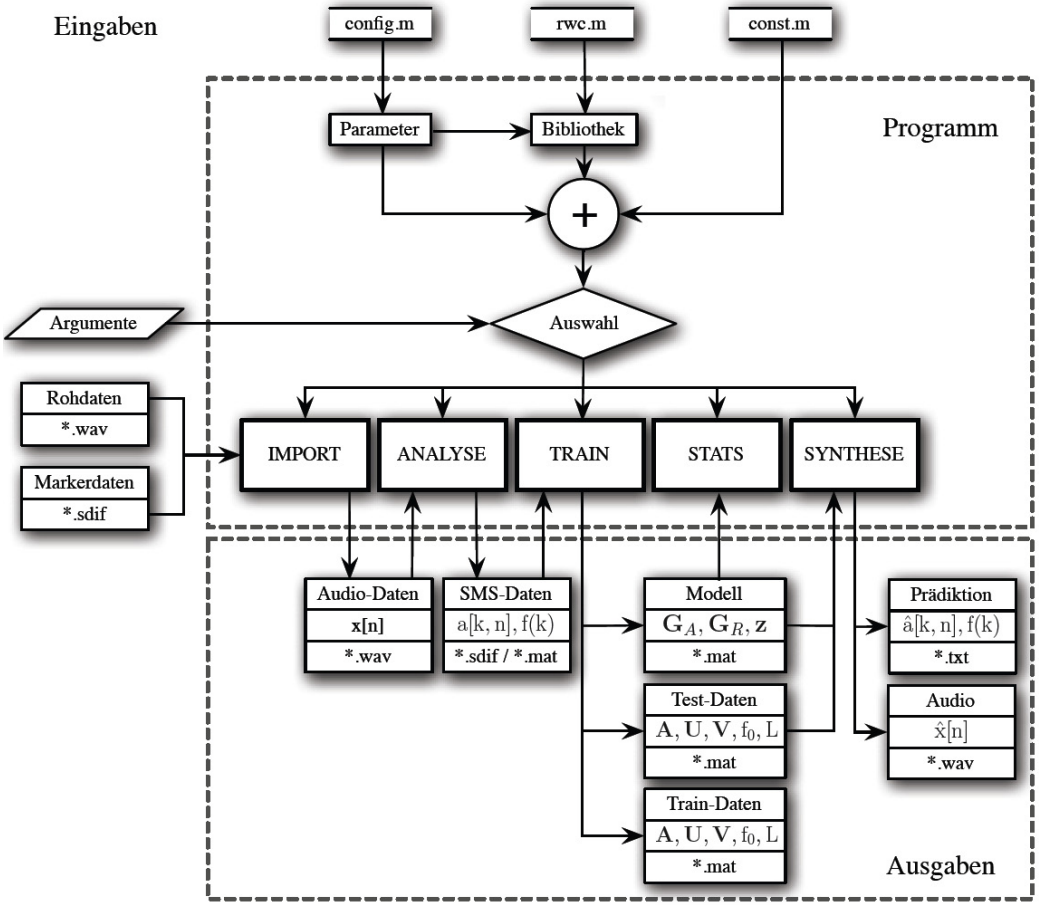


Abbildung 5.3: Gesamtschema des Programms

drei Konfigurationsdateien `config.m`, `rwc.m` und `const.m` spezifiziert. Zudem sind die Rohaudiodaten in Form von `*.wav` Dateien mit ihren dazugehörigen Markierungsdaten im `*.sdif` Format auf der Eingabeseite verzeichnet, da diese nicht vom Programm erzeugt werden. Anhand dieser Eingaben entscheidet die Applikation selbstständig welche Verarbeitungsschritte mit den entsprechenden Parametern aufgerufen werden. Die fünf verschiedenen Verarbeitungsprozesse: *Import, Analyse, Train, Stats* und *Synthese* können somit separat und unabhängig voneinander angesteuert und gestartet werden, kommunizieren allerdings über die Ausgabeebene miteinander, da die Prozesse zumeist auf Daten zurück greifen, die von einem anderen Verarbeitungsschritt vorher erzeugt worden sein müssen. Daraus folgt, dass die Prozesse

konsekutiv gestartet werden müssen um zu garantieren, dass benötigte Daten zum Startzeitpunkt eines Prozesses tatsächlich vorliegen. Zudem speichern die Verarbeitungsschritte ihre jeweiligen Zwischenergebnisse in zusätzlichen `*.mat`-Dateien, sodass bei einem Abbruch und anschließendem Neustart der Applikation, die Berechnungen ungefähr an dem Punkt fortgesetzt werden können, an dem die sie unterbrochen wurde.

Alle Prozesse greifen in gleicher Form auf die verschiedenen Inhalte der Konfigurationsdateien zurück, in denen alle Programmparameter sowie Einstellungen festgelegt werden. Die `rwc.m` beinhaltet alle Definitionen bezüglich der verwendeten Datenbank. Dies umfasst neben den relativen Pfadangaben aller Ein- und Ausgabedaten in den verschiedenen Formaten auch die Konventionen zur Generierung der Dateinamen der Ergebnisdaten und die Definitionen der in der Datenbank vorhandenen Instrumente. Die Instrumentendefinitionen beinhalten neben den eindeutigen Bezeichnern vor allem die Beschreibungen für die Form des Amplitudenverlaufs (impulshaft oder kontinuierlich angeregt) in Abhängigkeit der in der Datenbank abgelegten Spieltechniken (Artikulationen). In der `config.m` hingegen werden Parameter definiert, die in Form von Variablen in die verschiedenen Berechnungen einfließen. Das schließt insbesondere Variablen ein, für die im Verlauf der Modellevaluation verschiedene Werte gesetzt werden sollen. Deshalb können für alle Parameter in der `config.m` mehrere Werte in Form von Feldern angelegt werden. Die daraus resultierenden Varianten von Parameterkonfigurationen werden dann automatisiert beim Laden der Parameter berechnet. Das Gegenstück hierzu findet sich in der `const.m` Konfigurationsdatei in die nur Parameter aufgenommen sind, deren Werte im gesamten Evaluationsprozess konstant gehalten werden.

In den meisten Prozessen werden zur Kontrolle neben den Ausgaben auf die Konsole auch Vektorgrafiken im `*.pdf`-Format exportiert in denen die einzelnen Ergebnisse präsentiert werden. Hierfür wird die freie Bibliothek `export_fig` [38] verwendet, die eine Reihe von Vorteilen im Vergleich zu der Matlab® internen Exportfunktion für Grafiken besitzt.

5.1.1 Der Import-Prozess

Da die Rohdaten aus Audiodateien bestehen, die zu je einem Instrument, je einer Artikulationsart und je Spiellautstärke (p, m, f) alle möglichen Tonhöhen enthalten, müssen diese für die weitere Verarbeitung geschnitten und mit ihrer jeweiligen Tonhöhe markiert werden. Dies ist die Aufgabe des in

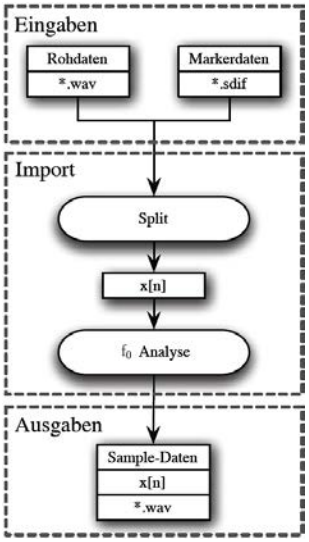


Abbildung 5.4: Schema des Import-Prozesses

Abbildung 5.4 dargestellten Import-Prozesses. Hierzu wird neben den Rohdaten im `*.wav`-Format auch jeweils eine Markierungsdatei im `*.sdif`-Format benötigt, in der die einzelnen On- und Offsets der Einzelklänge verzeichnet sind. Die On- und Offsets markieren die Punkte, an denen die Rohdaten geschnitten werden und zwischen denen die Einzelklänge extrahiert werden. Zur automatischen Erkennung der Grundfrequenz wird der YIN Algorithmus [7] verwendet, der bereits in einer Implementierung des Autors für Matlab® verfügbar ist [33]. Da eine solche Grundfrequenzschätzung jedoch keine vollständige Sicherheit der Erkennung gewährleisten kann, muss dieser Prozess manuell überprüft werden. Dies stellt damit den einzigen Verarbeitungsschritt dar, der einen manuellen Eingriff erfordert.

5.1.2 Der Analyse-Prozess

Die Berechnung der Sinusoidalmodelle der Eingangsdaten findet im Analyse-Prozess statt. Wie in Abbildung 5.5 dargestellt, wird hierzu zu jedem Einzelklang aus der Datenbank und jeder Parameterkonfiguration eine solche Analyse durchgeführt, wobei für die eigentliche Analyse die Kommandozeilenapplikation PM2 [32] in Version 1.6.2 des IRCAM Instituts verwendet wird. Da das Analyseergebnis in Form einer `*.sdif`-Datei [26] ausgegeben wird, wird dieses mit Hilfe einer SDIF *Reader*-Funktion (v0.3.0) [36] in das Matlab® eigene `*.mat`-Format konvertiert. Sollten aus einem früheren Pro-

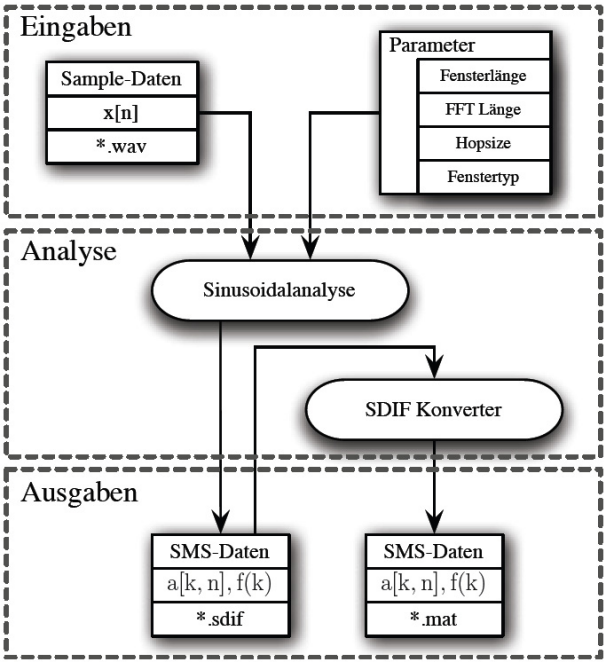


Abbildung 5.5: Schema des Analyse-Prozesses

grammlauf bereits fertig berechnete Dateien vorliegen, so werden die entsprechenden Berechnungen übersprungen und nur für die Eingangsdaten vorgenommen, für die noch keine Ergebnisse in Form von `*.sdif`- oder `*.mat`-Dateien vorliegen.

5.1.3 Der Trainings-Prozess

Der Trainings-Prozess stellt das Herzstück des gesamten Programms dar und ist daher auch deutlich komplexer und variabler aufgebaut. Ziel des Prozesses ist die Selektion des Modells, dessen geschätzte Parameter $\mathbf{G}_A$, $\mathbf{G}_R$ und $\mathbf{z}$ die geringsten Kosten auf dem Validierungsdatensatz besitzen. Abbildung 5.6 zeigt den schematischen Aufbau der wichtigsten Komponenten. Auf der Eingabeseite finden sich die konfigurierten Bibliotheksinformationen

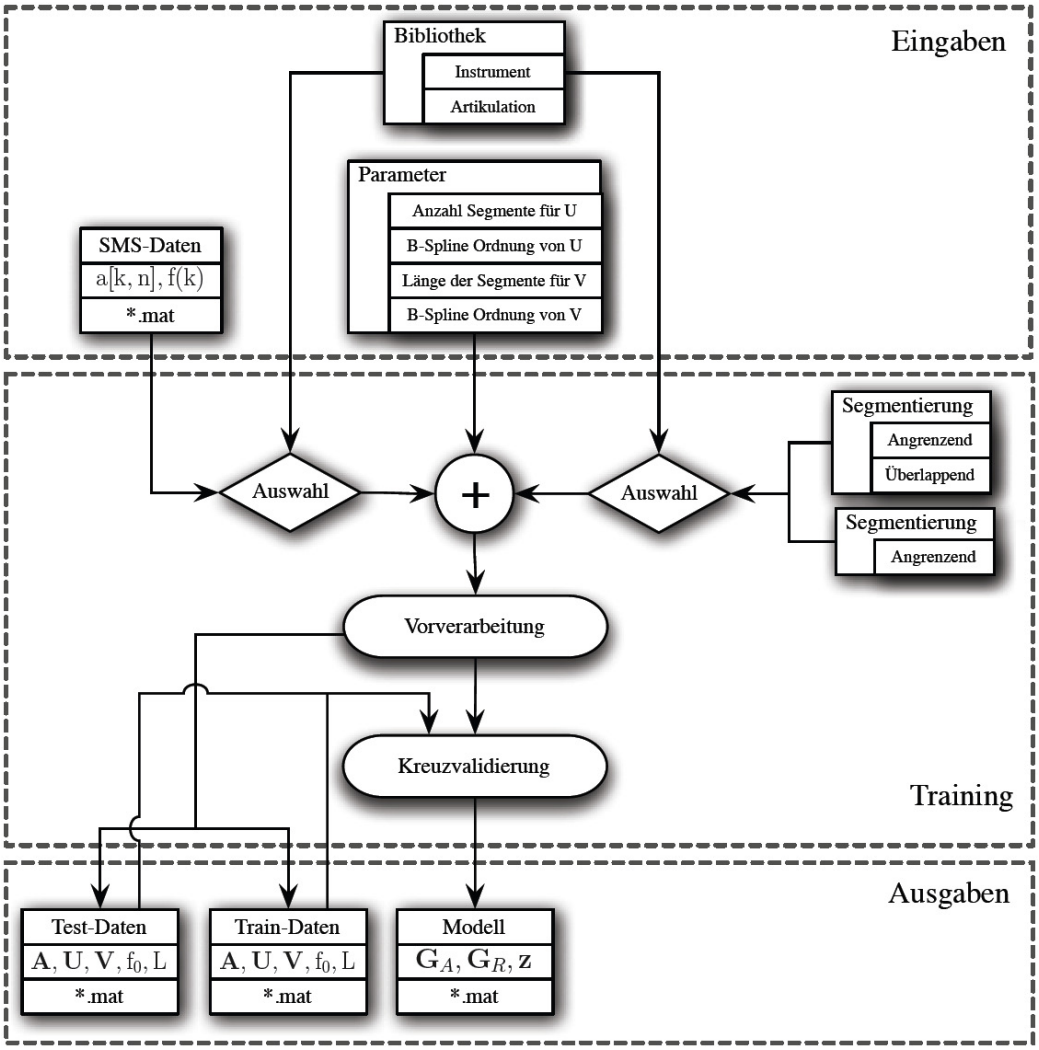


Abbildung 5.6: Schema des Trainings-Prozesses

anhand derer die Trainingsbeispiele $\tilde{x}$ für jeweils eine Artikulationsform ei-

nes Instruments selektiert und mit Hilfe derer die entsprechenden Varianten zur Segmentierung gewählt werden. Darüber hinaus werden auch alle zuvor definierten Parameterkombinationen an die Vorverarbeitung übergegeben. Die Vorverarbeitung erstellt darauf basierend die Trainings- und Testdaten $\mathbf{A}$ zusammen und berechnet die, während des Trainingsprozesses konstanten Matrizen $\mathbf{U}_A$, $\mathbf{U}_R$ und $\mathbf{V}$, sowie die Indizierungen n_a und n_r entsprechend der übergebenen Parameter und legt diese in Form einzelner `*.mat`-Dateien ab. Somit kann für verschiedene Parametrisierungen auf bereits berechnete Daten zugegriffen und die Vorverarbeitung daher wesentlich beschleunigt werden. An die Vorverarbeitung der Daten schließt sich die Kreuzvalidierungsprozedur an. Während der Kreuzvalidierung werden J Modelle trainiert und das Modell mit dem geringsten Fehler auf dem Validierungsdatensatz wird in einer eigenen `*.mat`-Datei abgelegt. Zusätzlich wird zu jeder Kreuzvalidierungsprozedur ein Ergebnisprotokoll ebenfalls in Form einer `*.mat`-Datei geschrieben, die neben einem Verweis auf das selektierte Modell und sein Risiko, die Kosten aller trainierten Modelle enthält. Diese Protokolldatei dient auch dazu, im Falle einer Unterbrechung des Programmablaufs, die Berechnungen zu überspringen, die bereits durchgeführt worden sind.

5.1.4 Der Statistik-Prozess

Der Statistik-Prozess dient der Auswertung der verschiedenen Parametrisierungen für jeweils eine Artikulationsform eines Instruments und die Aufbereitung der Ergebnisse in Form einer ANSI[1] kodierten Tabelle. Hierzu werden alle Ergebnisprotokolle der Kreuzvalidierungsprozeduren geladen und die dazugehörigen mittleren Kosten und das Risiko zu jedem Instrument und jeder Segmentierung in einem eigenen Ergebnisprotokoll gespeichert und auf der Konsole ausgegeben. Wird dieser Verarbeitungsschritt erneut aufgerufen werden ausschließlich die eigenen Protokolle geladen und die gespeicherten Daten ausgegeben.

[1] *American National Standards Institute*

5.1.5 Der Synthese-Prozess

Der Synthese-Prozess stellt den finalen Teil der Applikation dar und dient dazu, sowohl eine objektive als auch eine subjektive Bewertung der trainierten Modelle vornehmen zu können. Diesbezüglich werden die trainierten Modelle verwendet, um die Amplitudenverläufe der Partialwellen eines Instruments zu einer bestimmten Grundfrequenz vorherzusagen. Auf der Ein-

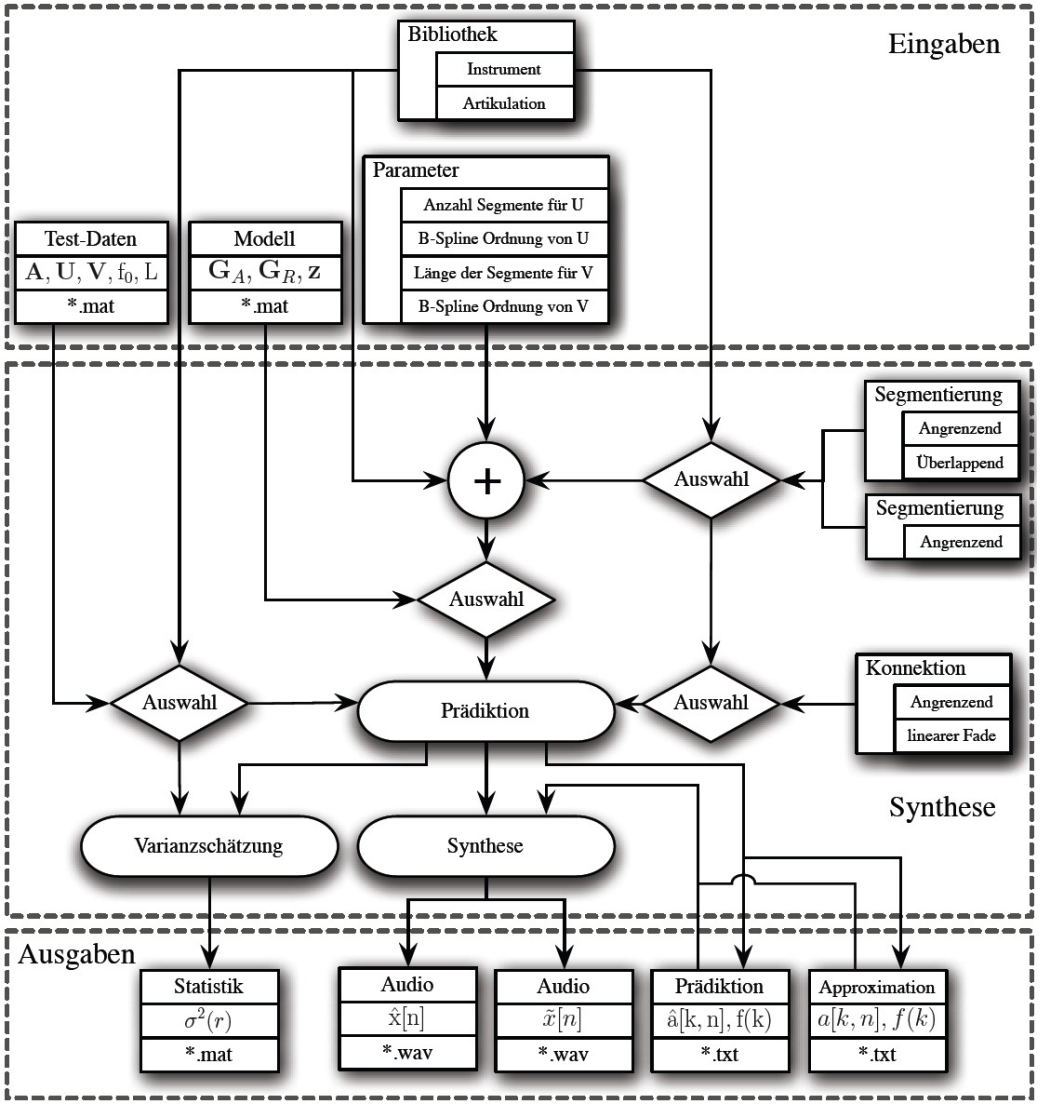


Abbildung 5.7: Schema des Synthese-Prozesses

gabeseite befinden sich erneut sowohl die Bibliotheksspezifikationen als auch

die Parametrisierungen der Modelle. Anhand dieser Informationen können alle zuvor durch die Kreuzvalidierungsprozedur selektierten Modelle separat geladen und zur Prädiktion automatisiert verwandt werden. Hierfür werden die Gesamtpegelverläufe und dazugehörigen Grundfrequenzen der, zu dem jeweiligen Instrument und seiner Artikulation gehörenden, Testdaten benötigt. Sie müssen daher ebenfalls geladen werden. Auf Grund der verschiedenen Segmentierungstechniken wird abschließend noch die geeignete Methode zum Verknüpfen der Oszillatoranregungen gewählt. Hieraus werden die Vorhersagen $\hat{a}[k,n]$ für die Partialwellenverläufe berechnet, die zur Schätzung der Modellvarianz σ^2 herangezogen werden und in Form einer Textdatei gespeichert werden. Das Format zur Speicherung der Partialwellenfunktionen in einer Textdatei orientiert sich hierbei an dem von der Applikation PM2 benötigten Format, um daraus abschließend ein Audiosignal $\hat{x}[n]$ zu generieren. Neben den geschätzten Partialwellenfunktionen aus den trainierten Modellen wird zusätzlich noch der approximierte Partialwellenverlauf $a[k,n]$ der Testdaten zuerst in eine Textdatei geschrieben aus dem anschließend das Vergleichssignal $\tilde{x}[n]$ mit Hilfe von PM2 generiert wird. Anhand der generierten Audiodaten wird nach einem Neustart des Programms auch wieder entschieden, welche Berechnungen bereits durchgeführt worden sind und welche noch ausgeführt werden müssen.

5.2 Verwendung der Software

Auf Grund der vergleichsweise hohen Komplexität der Software und der nahezu vollständigen Automatisierung, müssen alle Definitionen und Abläufe vorab festgelegt werden. Dies geschieht mit Hilfe der Konfigurationsdateien. Damit diese und alle anderen Bestandteile des Programms auch von Matlab® gefunden werden können, ist es notwendig Matlab® aus dem Projektverzeichnis über die Kommandozeile zu starten, wodurch automatisch das Script `startup.m` gestartet wird. Alternativ kann das Script auch manuell gestartet werden. Da die Applikation intern ausschließlich relative Pfadangaben verwendet, werden mit Hilfe des Start-Scripts alle Matlab®-Pfade zu den Quellcode-Verzeichnissen gesetzt. Somit kann das gesamte Programm unab-

hängig von seinem tatsächlichem Speicherort agieren und daher direkt auf jedem System ausgeführt werden.

5.2.1 Definition der Datenbibliothek

Die Definition, der im Rahmen dieser Arbeit verwendeten Datenbibliothek, wird mit Hilfe der Datei `rwc.m` vorgenommen. In dieser Datei werden drei separate Inhalte beschrieben:

- Die Pfadangaben zu allen Speicherorten der von dem Programm zu verarbeitenden und zu erzeugenden Daten.
- Die Konventionen zur Erzeugung der Dateinamen für zu speichernde Daten.
- Eine Beschreibung der in der Bibliothek bereitgestellten Instrumentendaten.

Das heißt es werden alle Pfade für zu ladende und zu speichernde Daten zusammen mit den Regeln zur Erzeugung der Dateinamen zentral in einer Datei verwaltet. Die Beschreibung der bereitgestellten Instrumente durch eindeutige Bezeichner für ihre Klassenzugehörigkeit, ihren Namen und ihre möglichen Artikulationsformen ist ebenfalls über Dateipfade kodiert und wird daher ebenfalls an dieser Stelle definiert. Da sich allerdings aus den Artikulationsarten der einzelnen Instrumentendaten verschiedene Methoden für den Prozess der Segmentierung ergeben, dient die Definition der Instrumente zusätzlich bei der Vorverarbeitung der Daten dazu, die geeignete Methode zu selektieren. So erfordert beispielsweise eine „normale“Spielweise beim Piano eine andere Segmentierung als eine ebensolche bei der Violine (vgl. Absatz 4.2.3).

5.2.2 Definition der Parameter

Bei der Parameterdefinition gilt es zu unterscheiden, welche Parameter im Programmablauf tatsächlich variabel sein sollen und welche einen für alle

Berechnungen konstanten Wert erhalten sollen. Alle Konstanten werden in der `const.m` innerhalb einer einfachen Matlab®-Struktur definiert und nur in den Verarbeitungsschritten geladen, in denen sie tatsächlich verwendet werden. Parameter, die im Programmablauf verschiedene Werte annehmen können, werden in der `config.m` definiert. Diese werden jeweils einem Verarbeitungsschritt zugewiesen und erhalten darüber hinaus einen eindeutigen Bezeichner in Form einer Zeichenkette. Da für die Parameter auch mehrere Werte über Felder definiert werden können, werden im Programmablauf die Pfade und Dateinamen der zu einer Parameterkombination gehörenden Ergebnisse, anhand der Bezeichner und der definierten Werte gesetzt. Die Konventionen hierzu befinden sich in der Bibliotheksdatei `rwc.m`.

5.2.3 Programmaufruf

Der Aufruf der Applikation gestaltet sich auf Grund der vorherigen Definition aller notwendigen Parameter und Daten äußerst einfach. Im Projektverzeichnis kann, nachdem das `startup.m` Skript ausgeführt worden ist, das Programm direkt mit dem Aufruf `main` gestartet werden. Da auf diese Weise jedoch alle Verarbeitungsschritte in Reihe ausgeführt werden, ist es ratsam, die einzelnen Schritte direkt anzuwählen. Jeder Verarbeitungsschritt wird dabei über die in Tabelle 5.1 dargestellten Kurzbezeichner angewählt.

`import`	:	Der Import-Prozess
`analyse`	:	Der Analyse-Prozess
`train`	:	Der Trainings-Prozess
`stats`	:	Der Statistik-Prozess
`synth`	:	Der Synthese-Prozess

Tabelle 5.1: Prozessbezeichner

Es können auch mehrere Schritte gleichzeitig angegeben werden. Ein Aufruf des Trainingsprozesses und seiner anschließenden Auswertung sieht dann folgendermaßen aus:

```
>> main train stats
```

Kapitel 6

Evaluation

Im Folgenden Kapitel wird dargelegt, anhand welcher Daten und Parameter der Modellansatz evaluiert wird.

6.1 Die Datenbank

Da die Parameterschätzung für das Quelle-Filter-Modell auf dem Ansatz des induktiven Lernens, dem Lernen aus Beispielen, basiert, muss in einem ersten Schritt eine geeignete Datenbasis ausgewählt werden, anhand derer die Parameter des Modells geschätzt werden sollen. Um die Parameter für einen Prototypen eines Instruments zu schätzen, werden ausreichend Audio-Aufnahmen benötigt, die alle spezifischen Charakteristika umfassen. Auf Grund der statistischen Eigenschaften eines solch parametrischen Schätzverfahrens, als auch der Annahme von stochastischen Daten, ist es für eine gute Generalisierungsfähigkeit darüber hinaus erforderlich, dass die Daten ausreichend redundant vorliegen, damit der Einfluss, der durch die menschliche Spielweise induzierten, Varianz eines Instrumentenklangs durch die statistische Näherung auf ein Minimum reduziert wird. Dies meint jedoch nicht verschiedene Spieltechniken, sondern die, sich durch leicht veränderte Spielweise ergebenden, klanglichen Nuancen ein und derselben Spieltechnik. Idealerweise sollten die einzelnen Varianten der Klangbeispiele durch verschiedene Instrumentalisten und verschiedene Instrumente des gleichen Typs einge-

spielt worden sein. Die Menge aller Audioaufnahmen eines Instruments stellt die Eingangs- bzw. Trainingsdatenbasis zur Schätzung eines Prototypen dar.

Damit weitere störende Effekte weitestgehend ausgeschlossen werden können, muss eine solche Datenbibliothek monophone, rauscharme und nachhallfreie Aufnahmen aller spielbaren Tonhöhen eines Instruments beinhalten. Für die akademische Verwendung gebräuchlich sind die Music Instrument Samples (MIS) Bibliothek der University of Iowa [34], die McGill University Master Samples (MUMS) Bibliothek [35] und die RWC Music Instrument Database (RWC MID) [10]. Da nur die RWC MID Bibliothek tatsächlich alle spielbaren Töne zu jedem Instrument beinhaltet und zudem die benötigte Redundanz durch je drei aufgenommene Varianten eines Instruments bietet, wird zur Evaluation des Modellansatzes ausschließlich die RWC MID verwendet, wobei aus Zeitgründen nur einige wenige Instrumente, aus der insgesamt 50 Instrumente umfassenden Datenbank, evaluiert werden. Alle Daten der RWC MID sind mit $f_s = 44.1$kHz und $bit = 16$ digitalisiert worden. Die ausgewähl-

Instrument	$\mathcal{A}$	Form der Anregung
Alt Saxophon	297	kontinuierlich
Klarinette	360	kontinuierlich
Oboe	198	kontinuierlich
Trompete	209	kontinuierlich
Flügel	792	impulshaft
Violine	576	kontinuierlich
Violoncello	565	kontinuierlich

Tabelle 6.1: Verwendete Instrumente und deren vorhandene Anzahl an Einzelaufnahmen $\mathcal{A}$, sowie ihre jeweilige Form der Anregung

ten Instrumente sind in Tabelle 6.1 verzeichnet zusammen mit den Angaben über der Anzahl der vorhandenen Einzeltöne und der Form der Signalanregung für die verwendete Spieltechnik. Diese setzen sich zusammen aus der Menge der spielbaren Tonhöhen, den drei Dynamiken piano, mezzoforte und forte und der Verwendung dreier Varianten eines jeden Instruments. Eine

Ausnahme bildet die Oboe, zu der nur 2 Varianten zur Verfügung standen.

6.2 Analyse der Daten

Für die Parameterschätzung der Modelle müssen, für die als Signalfolgen vorliegenden Einzelaufnahmen der Instrumentenklänge, die Sinusoidalmodelle berechnet werden. Dies stellt einen Vorverarbeitungsschritt dar, der für alle Aufnahmen einmal durchgeführt werden muss und dessen Ausgangsdaten in der weiteren Betrachtung keine Verwendung mehr finden. Um aus den Signalfolgen $x[n]$ die über dem Partialwellenindex parametrisierten Amplitudenfolgen $a[k, n]$ zu berechnen, werden die in Tabelle 6.2 dargestellten Parameter verwendet. Die ersten vier Parameter entsprechen der Parame-

Länge des Signalausschnitts N	:	$5/f_0$
Länge des Analyse Fensters M	:	$2^{nextpow2(N)*4}$
Schrittweite H	:	0.25
Fensterung $w(n)$	:	*Blackman*
Bias-Korrektur	:	Abe-Smith

Tabelle 6.2: Analyse Parameter

trisierung der Kurzzeitspektralanalyse, die die Grundlage für die Schätzung des Sinusoidalmodells darstellt. Hierfür wird eine adaptive Fensterlänge verwendet, um dem Signal angepasst sicher zu stellen, dass in einem Zeitfenster ungefähr fünf Perioden der Grundfrequenz enthalten sind und ein weitestgehend exakter Wert für die Amplitude der Grundfrequenz berechnet wird, ohne für die meisten Daten unnötig lange Fenster verwenden zu müssen. Die Länge des Analysefensters der FFT orientiert sich an der Länge des Signalausschnitts und ist somit ebenfalls signaladaptiv. Dies gilt auch für die Schrittweite, dessen Länge von 25% des Zeitfensters einen Kompromiss zwischen Rechenaufwand und gewünschter zeitlicher Modellierung darstellt. Die Fensterung mit Hilfe eines *Blackman*-Fensters bietet im Vergleich zum *Hann*-, oder *Hamming*-Fenster eine stärkere Nebenkeulenunterdrückung bei einer

Verbreiterung der Hauptkeule. Da ausschließlich monophone Klänge analysiert werden, ist dies das zu bevorzugende Fenster. Zur Bias Korrektur bei der Detektion der Amplitudenmaxima in den einzelnen Spektren wird eine quadratische Interpolation (QFFT) nach Abe und Smith [1] verwendet. Da die Sinusoidalmodelle vollständig von dem Programm PM2 berechnet werden, werden die Parameter nur über die Kommandozeile übergeben.

6.3 Modelltraining

Für das Training der Modelle werden sowohl global einstellbare Konstanten als auch variable Parameter verwendet um das Modell für verschiedene Parameterkombinationen zu evaluieren. Für die Vorverarbeitung der Sinu-

Untere Pegelgrenze	:	-90dB
Attack/Release Schwellwert γ	:	-6dB (kontinuierliche Anregung)
		0dB (impulshafte Anregung)

Tabelle 6.3: Vorverarbeitungsparameter

soidalmodelle werden die in Tabelle 6.3 dargestellten Werte für die untere Pegelgrenze und die Schwellwerte für die Segmentierung der Daten verwendet. Die untere Pegelgrenze dient dazu log(0) zu verhindern, während die Schwellwerte für die zwei Varianten der Signalanregung voneinander unabhängig betrachtet werden müssen.

Im Anschluss an die Vorverarbeitung werden die vorhandenen Daten in disjunkte Trainings- und Testdatenmengen separiert. Für die Kreuzvalidierung werden die Trainingsdaten darüber hinaus in weitere Untermengen geteilt. Wie in Absatz 2.4.5 beschrieben, setzt eine 10-fach Kreuzvalidierung demnach eine Unterteilung der Trainingsdaten in zehn disjunkte Untermengen voraus. Des Weiteren wird zu einer unteren Grenze für die Anzahl an Trainingsepochen auch eine Obergrenze angegeben, um einerseits zu verhindern, dass durch die stochastische Natur des Gradientenabstiegverfahrens das Training zu früh abgebrochen wird und andererseits bei einem problematischen Konvergenzverhalten eine Endlositeration des Trainingsalgorithmus

auftritt. Zudem hat es sich als sinnvoll erwiesen für die adaptive Schrittweitenanpassung ebenfalls Grenzen anzugeben. Das Abbruchkriterium wird adaptiv zu den Kosten gesetzt, die vor dem Training auf den initialisierten Modellparametern ermittelt werden. Bei den variablen Parametern in Tabel-

Datenseparation		
Trainings- und Testdaten	:	80% - 20%
Kreuzvalidierung	:	10-fach
Trainingseinstellungen		
minimale Anzahl Epochen	:	15
maximale Anzahl Epochen	:	50
Startschrittweite	:	0,05
maximale Schrittweite	:	1
minimale Schrittweite	:	10^{-5}
Abbruchkriterium		
minimale Kostenänderung	:	$C(0) \cdot 10^{-5}$

Tabelle 6.4: Konstante Trainingsparameter

le 6.5 handelt es sich, um die eigentlich zu evaluierenden Parameter, da mit ihrer Hilfe verschiedene Modellkonfigurationen respektive Modellkomplexitäten, auf ihr Verhalten bezüglich der Daten bewertet werden sollen. Hierbei handelt es sich um die in Kapitel 4 vorgestellten Parameter für die einzelnen Oszillatormodelle und das Filtermodell, genauer gesagt, um die Anzahl, bzw. Länge der Segmente für die B-Spline Funktionen, sowie ihrer jeweiligen Ordnungen. Mit Hilfe der vier zu evaluierenden Werte für die Anzahl der B-Spline Segmente der Oszillatoren soll ermittelt werden, wie veränderlich die Amplituden der Partialwellen über der Gesamtenergie modelliert werden müssen. Je mehr Segmente dabei eingesetzt werden, umso variabler kann sich

der Amplitudenverlauf über der Energie entwickeln. Für das Filter gilt hingegen, je länger die Segmente, umso weniger Segmente und damit Koeffizienten verteilen sich über den gesamten Frequenzbereich. Das bedeutet, dass die Filterfunktion bei kleineren Segmentlängen einen variableren Verlauf annehmen kann als bei größeren Segmentlängen. Für die Wahl der Ordnungen der B-Spline Funktionen gilt sowohl für die Oszillatoren als auch für das Filter, dass einerseits B-Splines der Ordnung 3 als auch der Ordnung 4 untersucht werden sollen. Diese Ordnungen entsprechen quadratischen bzw. kubischen Polynomen für die einzelnen B-Splines.

Anzahl der B-Spline Segmente für die Oszillatoren				
3	5	10	20	
Länge der B-Spline Segmente für das Filter in 1/12 Oktavbreiten				
$\frac{1}{2}$	1	3	6	12
Ordnungen der B-Spline-Funktionen				
3		4		

Tabelle 6.5: Variable Trainingsparameter

Im Rahmen der Evaluation sollen alle Kombinationen der in Tabelle 6.5 gezeigten Werte für die variablen Parameter getestet und miteinander verglichen werden, wobei die Werte für die zwei Anregeoszillatoren paarweise konstant gesetzt werden. Da die Polynomordnungen sich sowohl auf U_p als auch V_q beziehen, folgt daraus, dass insgesamt $4 \cdot 2 \cdot 5 \cdot 2 = 80$ Modellkonfigurationen evaluiert werden müssen. Darüber hinaus gilt für Instrumentendaten mit kontinuierlich angeregten Signalen, dass diese Konfigurationen sowohl für die angrenzende als auch die überlappende Segmentierung untersucht werden müssen. Bedenkt man, dass zu jeder Modellkonfiguration 10 Modelle auf Grund der Kreuzvalidierung trainiert werden müssen, ergeben sich 1600, bzw. 800 zu trainierende Modelle je Instrument.

Die Berechnung der Modellvarianzen erfolgt auf Grund der begrenzten Rechenkapazitäten anhand der im Vergleich zur Trainingsdatenmenge wesentlichen kleineren Testdaten. Dies kann allerdings als unproblematisch angesehen werden, da die Datenmengen hinreichend umfangreich sind und beide Datenmengen der gleichen statistischen Verteilung entstammen.

Kapitel 7

Ergebnisse

Im aktuellen Kapitel werden einige ausgewählte Ergebnisse aus den Parameterschätzungen für die Prototypen gezeigt. Diese werden jeweils für Prototypen mit minimaler und maximaler Modellkomplexität für die Klarinette, die Violine und den Flügel vorgestellt, wobei für die Klarinette und die Violine die Ergebnisse für die Varianten der angrenzenden und überlappenden Signalpartitionierungen präsentiert werden. Darüber hinaus werden jeweils an einem Beispiel die Synthesen der Partialwellenfunktionen anhand der Pegelfunktion eines Testdatums gezeigt und die Syntheseergebnisse miteinander verglichen.

Die Ergebnisse des Alt-Saxophons, der Oboe, der Trompete und des Violoncellos werden nicht ausführlich besprochen und befinden sich daher in Appendix B.

Vor der Präsentation der Modellprototypen erfolgt zu den drei genannten Instrumenten eine kurze Analyse der Verteilung der Eingangsdaten im Modellraum, sowohl für die zur Parameterschätzung verwendeten Trainingsdaten, als auch für die zur Risiko- und Varianzschätzung verwendeten Testdaten. In Absatz 4.2.4 wurde gezeigt, wie die Trainings- bzw. Testdaten in den Modellraum projiziert werden. Das heißt, die Daten $a[k, r]$ wurden anhand ihrer Pegelfunktion $L[r]$ und der Frequenzfolge $f(k)$ mit Hilfe der B-Spline Funktionen U_p und V_q in den Modellraum abgebildet. Da diese Abbildung der Daten jedoch nicht zu einer Gleichverteilung von $U_p(L[r])$ und $V_q(f(k))$

über L respektive f führen, soll mit Hilfe ihrer tatsächlichen Verteilungen gezeigt werden, für welche Wertebereiche die freien Parameter $g^A_{k,p}$, $g^R_{k,p}$ und v_q überhaupt geschätzt werden bzw. welche Bereiche gut oder weniger gut geschätzt und generalisiert werden können[1]. Die Berechnung der Verteilungen $P(L[r_a])$, $P(L[r_r])$ sowie $P(f(k))$ erfolgt jeweils mit Hilfe der Häufigkeiten spezifischer Pegelwerte bzw. Frequenzen innerhalb der für die B-Spline Segmentierung verwendeten Pegel- bzw. Frequenzbereiche und einer Normierung auf eine Gesamtsumme von 1. Wie bereits in den Abbildungen 4.6 und 4.7, ist zur Demonstration der Gesamtpegelbereich in 5 Segmente eingeteilt und der Frequenzbereich in Segmente mit $\frac{1}{3}$ Oktavbreite. Zusätzlich wird in den Grafiken jeweils noch die Gesamtanzahl der Zeitblöcke r und Partialwellen K summiert über alle Daten der Trainings- und Testdatenmengen angegeben, da, entgegen der reinen Angabe der Anzahl an Beispielsamples, diese Zahlen der tatsächlichen, zur Parameterschätzung verwendeten, Datenmenge entsprechen und diese auch Grundlage für die Darstellung über die Verteilung der Daten im Modellraum darstellen.

In den Abschnitten zu den Prototypen der einzelnen Instrumente werden je Partitionierung zwei Prototypen präsentiert, wobei jeweils einer mit der kleinsten und einer mit der größten Modellkomplexität gezeigt wird. Die entsprechenden Konfigurationen sind in Tabelle 7.1 gezeigt. Die vorgestellten

	Anzahl Segmente für U	Länge der Segmente für V	Ordnung
Minimal	3	12	4
Maximal	20	$\frac{1}{2}$	3

Tabelle 7.1: Extremwerte der Modellkomplexitäten

Prototypen sind auf Basis des Kreuzvalidierungsverfahrens mit den geringsten Kosten C auf dem jeweiligen Trainingsdatensatz unter Verwendung des *Online*-Verfahrens selektiert worden. Sie werden daher über ihre, anhand der gelernten, freien Parameter $g^A_{k,p}$, $g^R_{k,p}$ und z_q geschätzten Kosten C und

[1] Eine Gleichverteilung der Daten im Modellraum wäre optimal zur Schätzung der Parameter, da es andererseits zu Skalierungsproblemen bei den Gradienten kommen kann.

ihrem auf dem Testdatensatz ermitteltem Risiko R, sowie ihrer Varianz σ^2 charakterisiert. Hierbei stellen die Werte für C und R abstrakte Werte der Kostenfunktion dar und sind demnach kein qualitatives Maß. Sie hängen ausschließlich von den Eingangsdaten und den Modellparametrisierungen ab und eignen sich daher nur zu einem relativen Vergleich zwischen verschiedenen Prototypen für ein einzelnes Instrument. Die Varianz σ^2 hingegen repräsentiert den Erwartungswert der Abweichung der Modellprädiktionen $\hat{A}[k,r]$ von den Daten $A[k,r]$ als Mittelwert über alle Partialwellen k und Zeitblöcke r.

Auf die Verwendung des *Offline*-Verfahrens zur Parameterschätzung wurde verzichtet, da der Rechenaufwand hierfür um ein Vielfaches größer als der des *Online*-Verfahrens ist. Auf Grund der begrenzten Rechenkapazitäten musste daher auf seine Verwendung verzichtet werden.

Die grafische Darstellung der einzelnen Prototypen erfolgt getrennt anhand der Oszillatorfunktionen $O(k,L)_A$ und $O(k,L)_R$ sowie der Filterfunktion $F(f)$, wobei die Oszillatorfunktionen für diskrete Werte von L über k dargestellt werden. Somit werden die Anregeoszillationen in Form von Partialwellenfunktionen für verschiedene, relative Gesamtpegel L^σ dargestellt. Es werden allerdings nur die ersten 16 Partialwellen abgebildet, da bei höheren Indizes zumeist nur ein annähernd linearer Abfall zu beobachten ist und der Grossteil der spezifischen Klanginformation somit in den unteren Partialwellen kodiert ist. Folglich lassen sich die f_0-korrelierten Klangcharakteristika in den Oszillatorfunktionen ablesen, während die f_0-unabhängigen Eigenschaften in der Filterfunktion dargestellt sind.

Die Modellprädiktionen respektive Synthesen werden abschließend zu jedem Instrument exemplarisch für ein Beispiel der Testdaten gezeigt, wobei insbesondere ein Vergleich zwischen dem Testdatum und den Prädiktionen der Prototypen mit minimaler bzw. maximaler Modellkomplexität für je eine Form der Partitionierung hergestellt wird. In den Grafiken sind daher die Amplitudenverläufe $A[k,r]$ des Testdatums und die prädiktierten Verläufe $\hat{A}[k,r]$ für die Partialwellen $k = 1 \ldots 4$ abgebildet. Darüber hinaus ist jeweils der Pegel des Gesamtenergieverlaufs dargestellt, wobei der Pegel des Energieverlaufs $L[r]$ des Testdatums Grundlage für die jeweiligen Prädiktionen ist

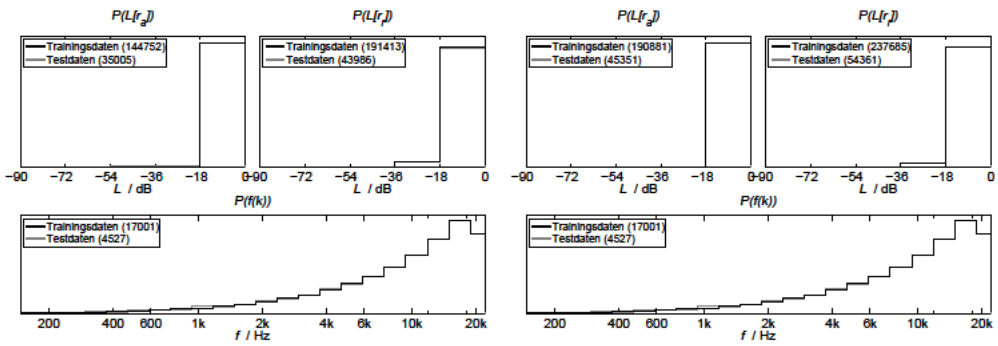
und die Gesamtenergieverläufe der prädiktierten Amplitudenfunktionen sich im optimalen Fall nicht davon unterscheiden sollten.

Die Kosten, Risiken und Varianzen für alle weiteren Modellkomplexitäten der hier vorgestellten drei Instrumente sind in Appendix A dargestellt und die Ergebnisse inklusive Grafiken für die hier nicht gezeigten Instrumente befinden sich in Appendix B.

7.1 Klarinette

7.1.1 Analyse der Daten

In den Abbildungen 7.1a und 7.1b ist erkennbar, dass sowohl die Trainings- als auch die Testdaten für den größten Teil aller Zeitblöcke r_a und r_r einen Signalpegel L zwischen -18 und 0dB aufweisen. Somit kann bereits gefolgert werden, dass für Pegelwerte, die deutlich unter -18dB liegen, sowohl die Parameterschätzung problematisch ist, als auch die Validierung der Prototypen. Darüber hinaus ist in den oben genannten Abbildungen zu sehen, dass die Form der Partitionierung der Zeitblöcke (Angrenzend oder Überlappend) bei der Wahl von 5 Segmenten über den gesamten Pegelbereich kaum einen Einfluss auf die Verteilung der Eingangsdaten nimmt und somit auch die Parameterschätzung nur geringfügig beeinflussen wird. Eine ähnliche Problematik



(a) Angrenzende Partitionierung (b) Überlappende Partitionierung

Abbildung 7.1: Verteilung der Eingangsdaten der Klarinette im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite

ist bei der Verteilung der Frequenzen der Partialwellen zu beobachten, wo in den unteren Frequenzbereichen sehr viel weniger Daten zur Parameterschätzung zur Verfügung stehen, als für höhere Frequenzen. Dieser Effekt wird insbesondere dadurch erzeugt, dass die Segmentierung des Frequenzbereichs frequenzabhängig erfolgt (siehe Absatz 4.2.2), aber auch durch die Tatsache, dass die Einzelklänge verschiedene Grundfrequenzen aufweisen.

7.1.2 Prototypen

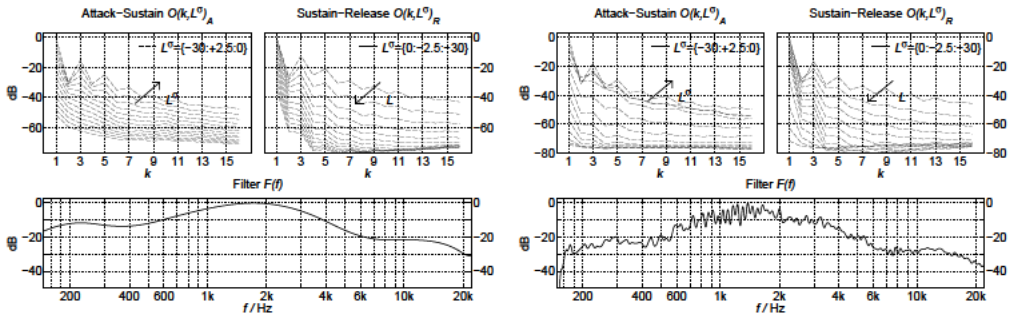
Die Ergebnisse für die Kosten, Risiken und Varianz der mit Hilfe der Kreuzvalidierung selektierten Prototypen sind in Tabelle 7.2 dargestellt. Die Ergeb-

Angrenzend	C	R	σ^2 [dB]
minimale Modellkomplexität	37.3799	39.4335	-25.0768
maximale Modellkomplexität	37.2099	39.5238	-25.5096
Überlappend			
minimale Modellkomplexität	38.0314	40.1409	-25.3109
maximale Modellkomplexität	37.7447	39.9304	-25.6001

Tabelle 7.2: Kosten, Risiko und Varianz für Prototypen der Klarinette

nisse für die vier Prototypen weisen jeweils von einander kaum unterschiedliche Werte auf, was eine abschließende Bewertung über den jeweils besser angepassten Prototypen erschwert. Während alle Prototypen erwartungsgemäß Werte für das Risiko aufweisen, die leicht über den Kosten liegen, ist die Varianz für die Modelle mit maximaler Modellkomplexität kleiner als für die Modelle mit minimaler Komplexität. Dies widerspricht der Erwartungserhaltung, nach der der umgekehrte Fall zutreffen müsste.

In den Abbildungen 7.2a und 7.2b für die Prototypen bei angrenzender Partitionierung der Eingangsdaten sind die Unterschiede für die Formen der Modellkomplexität deutlich an den Verläufen der Kurven zu erkennen. Während der Prototyp mit der minimalen Komplexität eine sehr glatte Filterkur-

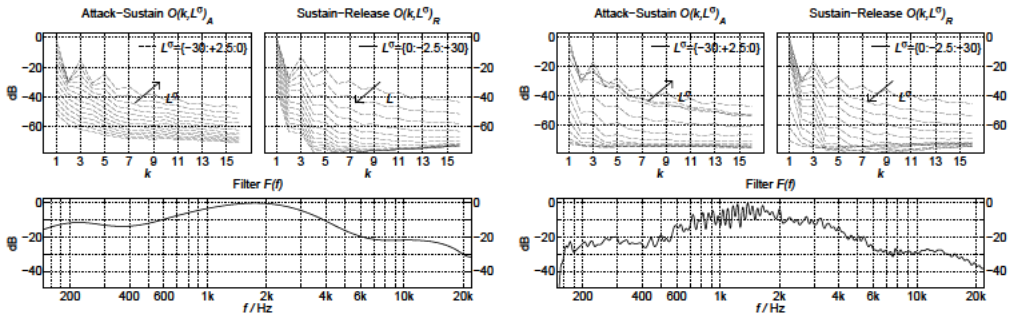


(a) Minimal (b) Maximal

Abbildung 7.2: Prototypen der Klarinette mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung

ve und nur geringfügig variierende Partialwellenhüllkurven aufweist, sind in den Abbildungen für den Prototyp mit maximaler Modellkomplexität eine Vielzahl an einzelnen Resonanzen in der Filterkurve erkennbar. Die Hüllkurven der Partialwellen hingegen weisen nur im Falle des *Attack/Sustain* signifikante Unterschiede auf und sehr geringe Unterschiede in den Kurven des *Sustain/Release*. Dies ist in erster Linie auf die Verteilung $P(L[r_a])$ der Daten zurückzuführen. Wie in 7.1a gezeigt worden ist, werden die Eingangsdaten vorrangig auf hohe Pegelwerte abgebildet, was die Parameterschätzung mit einem komplexen Modell erschwert und dazu führt, dass der Prototyp mit der größeren Modellkomplexität die Partialwellenhüllkurven für tiefere Gesamtpegelwerte weniger gut schätzen kann. Die in 7.3a und 7.3b abgebildeten Prototypen für die überlappende Partitionierung der Eingangsdaten weisen kaum einen erkennbaren Unterschied zu den Prototypen mit angrenzender Partitionierung auf. Dieses Ergebnis deckt sich mit der in Absatz 7.1.1 diskutierten Vermutung und den Ergebnissen in Tabelle 7.2.

Die Tatsache, dass, sowohl bei angrenzender, als auch bei überlappender Partitionierung, annähernd identische Filterfunktionen und Partialwellenhüllkurven geschätzt wurden, stützt die Annahme, dass das Modell sowohl f_0-unabhängige als auch f_0-korrelierte Charakteristika der Klänge der Klarinette extrahiert und voneinander getrennt hat. Allerdings lässt insbesondere die Tatsache, dass die Varianz der Prototypen mit steigender Modellkomple-



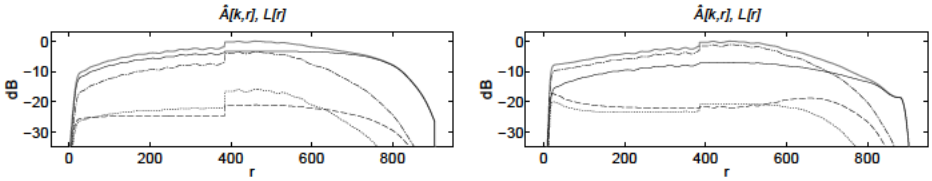
(a) Minimal (b) Maximal

Abbildung 7.3: Prototypen der Klarinette mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung

xität nicht ansteigt und die Kosten nur geringfügig kleiner werden, vermuten, dass die Parameterschätzung noch nicht bis zum Optimum der Kostenfunktion vorgedrungen ist. Das heißt, dass das Abbruchkriterium wahrscheinlich noch nicht bestmöglich gewählt ist.

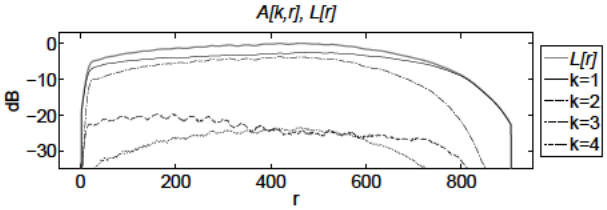
7.1.3 Synthese

In Bild 7.4c ist für ein Bb5 der Klarinette ein realer Amplitudenverlauf $A[k, r]$ der ersten 4 Partialwellen und der Gesamtpegelverlauf $L[r]$ gezeigt. Basierend auf diesem Pegelverlauf und der zur Note Bb5 gehörenden Grundfrequenz werden die prädiktierten Partialwellenfunktionen der einzelnen Prototypen in den Bildern 7.4a, 7.4b, 7.4d und 7.4e gezeigt. Alle vier Prototypen schätzen zum Referenzsignal subjektiv vergleichbar verlaufende Partialwellenfunktionen, wobei jedoch insbesondere die Unstetigkeit der Partialwellenfunktionen für die Prototypen mit angrenzender Partitionierung auffällt. Diese begründet sich in der in 4.2.6 vorgestellten Konnektion ohne Überlappung.

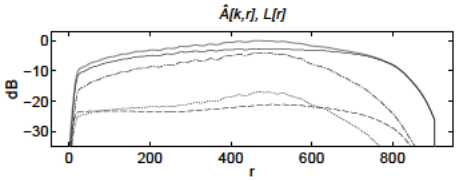


(a) Angrenzend, Minimal

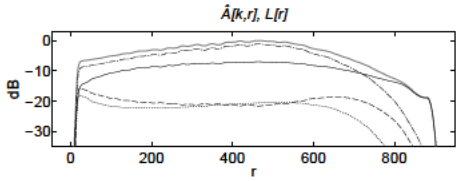
(b) Angrenzend, Maximal



(c) Realer Beispielklang als Referenzsignal



(d) Überlappend, Minimal



(e) Überlappend, Maximal

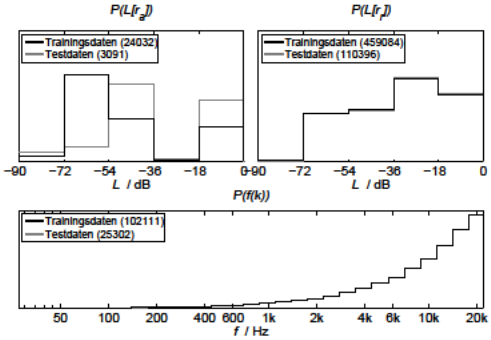
Abbildung 7.4: Synthesebeispiel eines Bb4 der Klarinette

7.2 Flügel

7.2.1 Analyse der Daten

Aufgrund der Impulshaftigkeit der Klänge des Flügels wird nur die angrenzende Partitionierung der Partialwellen betrachtet. Wie bei der Verteilung von $P(L[r_a])$ in Abbildung 7.5 zu erkennen, führt der sprunghafte Verlauf des Signals dazu, dass deutlich weniger Zeitblöcke r_a dem *Attack*-Bereich zugeordnet werden können als dem *Release* r_r. Der *Attack*-Bereich scheint zudem stark verrauscht, da hier insbesondere die beiden Teildatenmengen unterschiedliche Verteilungen aufweisen. Hingegen weist der *Release*-Bereich eine annähernde Gleichverteilung der Signalpegel auf, was sich im kontinuierlichen Energieabfall eines von einem Flügel gespielten Klangs begründet. Daher ist zu erwarten, dass für die Parameterschätzung des *Release*-Oszillators

ein vergleichsweise großer Energiebereich abgedeckt werden kann, wohingegen der Bereich des Einschwingvorgangs von starkem Rauschen beeinflusst wird und die Parameterschätzung somit ebenfalls einen verrauschten Charakter aufweisen kann. Für die Verteilung der Frequenzen der Partialwellen gilt



(a) Angrenzende Partitionierung

Abbildung 7.5: Verteilung der Eingangsdaten des Flügels im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite

die gleiche Problematik wie für die Klarinette in Absatz 7.1.1, wobei sich auf Grund des größeren Tonumfangs des Pianos die Datenmenge für die unteren Oktavlagen im Vergleich zu den höheren Lagen noch stärker verringert.

7.2.2 Prototypen

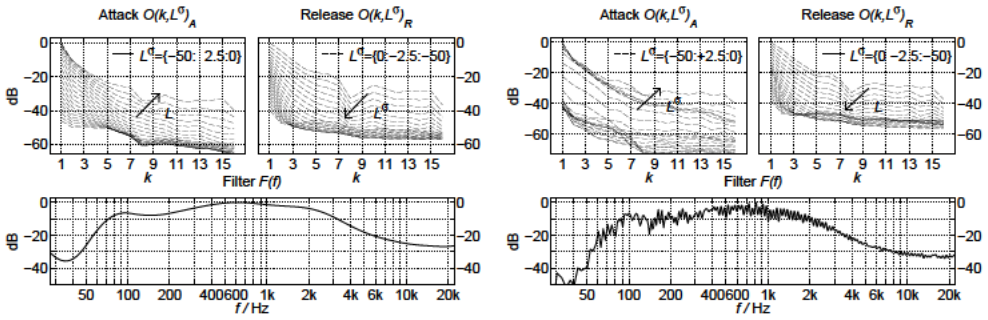
Auffallend an den Ergebnissen für die Prototypen des Flügels in Tabelle 7.3 sind die entgegen der Erwartung höheren Kosten im Vergleich zum Risiko und die nur geringfügig verschiedenen Ergebnisse für die verschiedenen Modellkomplexitäten, auch wenn die Varianz für den Prototyp mit der maximalen Komplexität über der des Prototypen mit der minimalen Komplexität liegt. Auch in den Abbildungen 7.6a und 7.6b zu den Prototypen des Flügels sind die verschiedenen Modellkomplexitäten, insbesondere in den Verläufen der Filterkurven, deutlich zu erkennen, da im Gegenzug zum Prototyp mit der minimalen Komplexität, der Prototyp mit der maximalen eine Vielzahl an einzelnen Resonanzen detektiert.

In den Partialwellenfunktionen für den *Attack*-Bereich des Prototyps mit der maximalen Komplexität in Bild 7.6a scheint, auf Grund der hohen Kom-

Angrenzend	C	R	σ^2
minimale Modellkomplexität	42.0474	40.5797	-34.3413
maximale Modellkomplexität	41.9821	40.0923	-33.8821

Tabelle 7.3: Kosten, Risiko und Varianz für die Prototypen des Flügel

plexität, wie bereits in Absatz 7.2.1 vermutet, in erster Linie das, durch die Impulshaftigkeit des Einschwingvorgangs verursachte, rauschhafte Verhalten abgebildet zu sein, welches von dem Prototyp mit der minimalen Komplexität scheinbar besser generalisiert wird. Der Ausschwingvorgang hingegen wird von beiden Prototypen auf Grund der annähernden Gleichverteilung der Daten über einen großen Pegelbereich ähnlich abgebildet.



(a) Minimal (b) Maximal

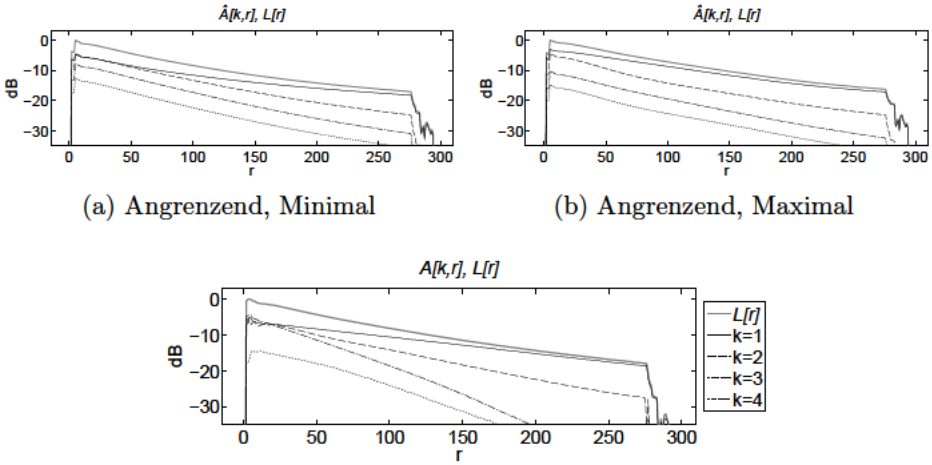
Abbildung 7.6: Prototypen des Flügels mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung

Wie bereits bei den Ergebnissen zu der Klarinette, scheint die Parameterschätzung noch zu keinem absoluten Optimum der Kostenfunktion gelangt zu sein, auch wenn die Abbildungen der Prototypen bereits deutlich f_0-korrelierte und f_0-unabhängige Klangcharakteristika des Flügels aufzeigen.

7.2.3 Synthese

Abbildung 7.7 zeigt ein Beispiel prädiktierter Partialwellenfunktionen an

Hand eines gespielten G1. Hierbei fällt auf, dass beide Prototypen sehr ähnliche Amplitudenverläufe der Partialwellen prädiktieren, die aber vor allem im letzten Signaldrittel erkennbar vom Referenzsignal abweichen.



(a) Angrenzend, Minimal

(b) Angrenzend, Maximal

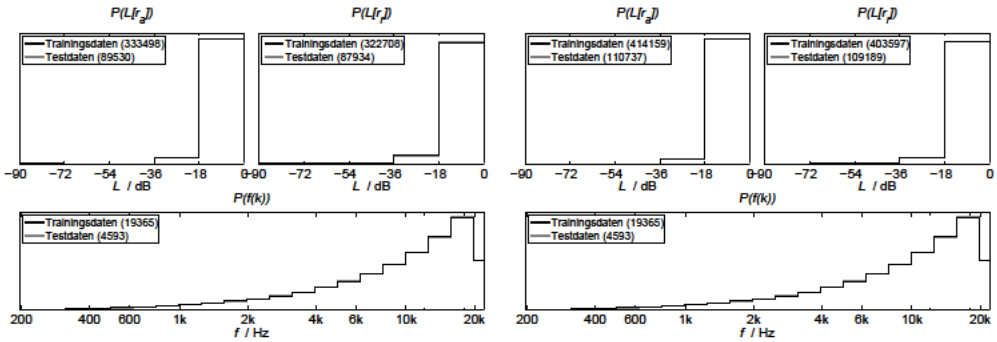
(c) Realer Beispielklang als Referenzsignal

Abbildung 7.7: Synthesebeispiel eines F3 des Flügels

7.3 Violine

7.3.1 Analyse der Daten

Wie bereits für die Klarinette, ist in den Abbildungen 7.8a und 7.8b kein wesentlicher Unterschied in den Verteilungen für die verschiedenen Formen der Partitionierungen zu erkennen. Erneut konzentriert sich die Verteilung nahezu vollständig auf den Energiebereich um -18 bis 0dB mit den bereits in Absatz 7.1.1 geschilderten Folgen. Ebenso gilt für die Verteilung der Frequenzen der Partialwellen, dass die in den Modellraum abgebildeten Daten, einerseits durch die auf Grund der variierenden Grundtonhöhe entstehende Konzentration von Frequenzen bei höheren Werten und andererseits durch die frequenzabhängige Segmentierung, eine schiefe Verteilung aufweisen. Somit stehen zur Parameterschätzung für tiefe Frequenzen wesentlich weniger Daten zur Verfügung, als für hohe Frequenzen.



(a) Angrenzende Partitionierung (b) Überlappende Partitionierung

Abbildung 7.8: Verteilung der Eingangsdaten der Violine im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite

7.3.2 Prototypen

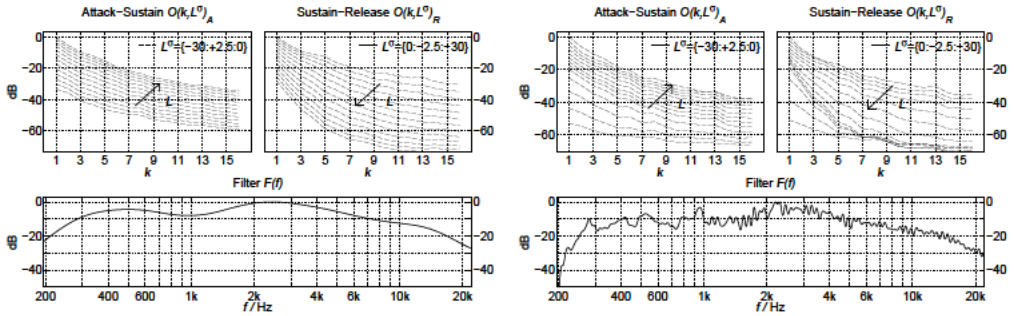
Die in Tabelle 7.4 dargestellten Ergebnisse für die Prototypen der Violine entsprechen der Erwartungshaltung in dem Sinne, dass für alle Prototypen die Werte für die Risiken größer sind als deren Kosten und die Varianzen für die Modelle mit maximaler Modellkomplexität größer sind, als die der Modelle mit der minimalen Komplexität. Auch die Tatsache, dass die Ergebnisse für die zwei Varianten der Partitionierung der Signale nur geringe Unterschiede aufweisen, deckt sich mit der im vorigen Abschnitt beschriebenen Annahme über die Ähnlichkeit der Verteilungen der Daten. Die Abbildungen zu

Angrenzend	C	R	σ^2
minimale Modellkomplexität	28.5574	28.707	-24.8699
maximale Modellkomplexität	27.9608	28.2569	-26.7404
Überlappend			
minimale Modellkomplexität	28.3208	28.5869	-24.877
maximale Modellkomplexität	27.7078	28.0671	-26.5902

Tabelle 7.4: Kosten, Risiko und Varianz für die Protoypen der Violine

den Prototypen der Violine zeigen sowohl für die angrenzende (7.9) als auch

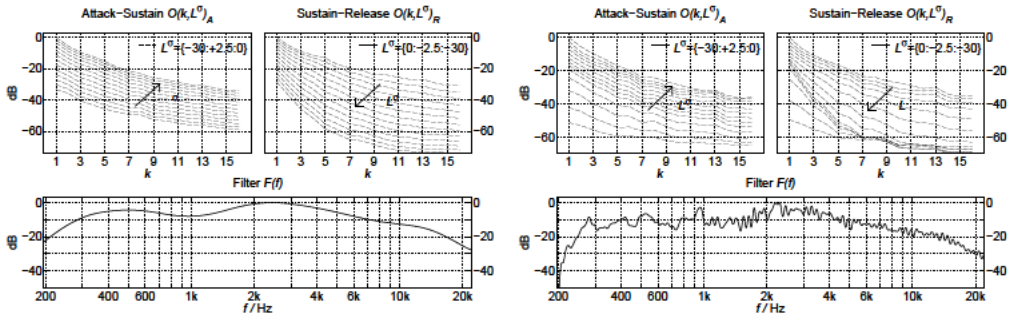
für die überlappende Partitionierung (7.10) nahezu identische Verläufe der Filter- und der Partialwellenfunktionen. Somit unterscheiden sich die Prototypen vorrangig anhand der Modellkomplexität und nicht auf Grund der Partitionierung der Daten.



(a) Minimal (b) Maximal

Abbildung 7.9: Prototypen der Violine mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung

In den Filterfunktionen bei maximaler Modellkomplexität sind eine Reihe signifikanter Resonanzen zu erkennen, die, unter Betrachtung der Ergebnisse für die Kosten und Risiken, einen scheinbar bedeutenden Einfluss auf die Qualität der Schätzung haben. Darüber hinaus kann angenommen werden, dass die variableren Verläufe der Partialwellenhüllkurven der Modelle mit maximaler Komplexität eher den realen Verläufen entsprechen.



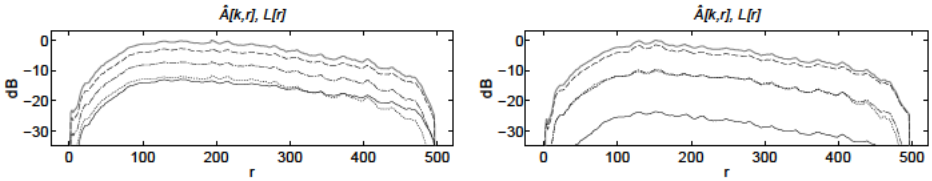
(a) Minimal (b) Maximal

Abbildung 7.10: Prototypen der Violine mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung

Die Parameterschätzung der Violine zeigt im Gegensatz zu denen der Klarinette und der des Flügels ein der Erwartungshaltung entsprechendes Ergebnis. Neben der Annahme, dass somit die Optimierung der Kostenfunktion für die Prototypen der Violine ein globales Minimum erreicht hat, muss jedoch auf Grund der Ähnlichkeit der absoluten Werte auch in Betracht gezogen werden, das es sich hierbei schlicht um eine statistische Abweichung handeln könnte, die aus einer günstigen Wahl der Trainings- und Testdaten herrührt. Ein Vergleich mit der im Appendix A gezeigten Tabelle mit den Ergebnissen aller Modellkomplexitäten stützt diese Vermutung.

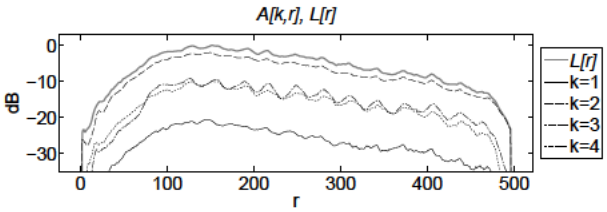
7.3.3 Synthese

Abbildung 7.11 zeigt die Prädiktionen der Prototypen der Violine für ein Testdatum eines gespielten Ab3. Deutlich erkennbar sind die voneinander stark abweichenden Prädiktionen der Modelle mit minimaler und maximaler Modellkomplexität, wobei letztere die wesentlich exaktere Vorhersage über den Verlauf der Partialwellen erzeugt. Dieses Ergebnis deckt sich mit den Erkenntnissen aus den Werten für die Kosten und das Risiko der geschätzten Prototypen, nach denen die Modelle mit der maximalen Modellkomplexität einen, im Vergleich zur minimalen Komplexität, geringeren Fehler bei der Prädiktion ermöglichen.

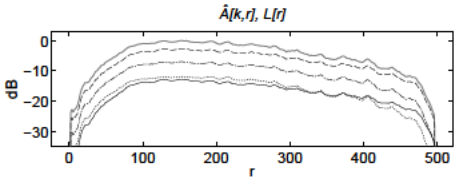


(a) Angrenzend, Minimal

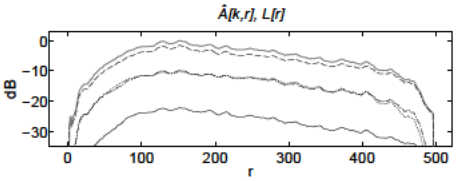
(b) Angrenzend, Maximal



(c) Realer Beispielklang als Referenzsignal



(d) Überlappend, Minimal



(e) Überlappend, Maximal

Abbildung 7.11: Synthesebeispiel eines Ab3 der Violine

Kapitel 8

Zusammenfassung

Ziel dieser Arbeit war die Entwicklung eines spezifischen Quellenmodells zur Repräsentation der Timbre quasi-harmonischer Instrumente. Hierfür wurde ein auf einem Quelle-Filter-Modell basierender Ansatz zur Darstellung der zeitveränderlichen, spektralen Hüllkurve eines Instruments entworfen. Im Vorfeld des Entwurfs wurden zwei bestehende Modellansätze zusammen mit ihren Einschränkungen erläutert, die somit die Motivation für den in dieser Arbeit vorgestellten Ansatz geliefert haben. So bestand an den präsentierten Entwurf sowohl der Anspruch, die Timbre über ihre Abhängigkeit von der Grundfrequenz zu konstruieren, als auch die spektrale Hüllkurve als Funktion der Gesamtenergie zu modellieren. Somit wurde einerseits die Möglichkeit geschaffen, stark mit der Grundfrequenz variierende Timbre abbilden zu können und andererseits die *Attack-, Sustain-* und *Release*-Bereiche eines Instrumentenklangs differenziert beschreiben zu können, wobei zwei verschiedene Techniken zur Unterscheidung der jeweiligen Bereiche vorgestellt wurden. Zur Konstruktion der Komponenten des Quelle-Filter-Modells wurden gewichtete B-Spline Polynome verwendet, deren Superposition maximal glatte Trajektorien erzeugt. Somit wurde das Modell befähigt, auf Basis einer Grundfrequenz und eines Gesamtenergieverlaufs, die zeitveränderliche, spektrale Hüllkurve in Form einer nicht-linearen, kontinuierlichen Funktion zu schätzen.

Weiter wurde gezeigt, wie die Gewichtungsparameter der B-Spline Funk-

tionen mit Hilfe zweier Verfahren anhand einer Trainingsdatenbasis gelernt werden können und wie das zu erwartende Risiko unter Einsatz einer Testdatenbasis geschätzt werden kann. Unter der Annahme, dass beide Datensätze der gleichen Verteilung entstammen, wurde zudem die Modellvarianz auf dem Testdatensatz ermittelt, wodurch das Modell einen vollständigen, statistischen Ansatz darstellt, bestehend aus dem, den mathematischen Erwartungswert repräsentierenden, geschätzten Timbre und der Varianz der Verteilung der zugrunde liegenden Daten.

Der Ansatz wurde in der Simulationssoftware Matlab® implementiert und für eine Vielzahl von Modellparametrisierungen für folgende Instrumente evaluiert: Klarinette, Oboe, Alt-Saxophon, Trompete, Flügel, Violine und Violoncello.

Es konnte gezeigt werden, dass die erzeugten Prototypen sowohl von f_0 abhängige als auch von der Grundfrequenz unabhängige Charakteristika abbilden und in Abhängigkeit der Modellkomplexität mit variabler Genauigkeit repräsentieren. Die Ergebnisse blieben dabei auch bei einer Zufallsinitialisierung wiederholbar. Dies lässt den Schluss zu, dass das Modell tatsächlich in der Lage ist, die instrumentenspezifischen Timbre zu extrahieren und dabei in ihre f_0-korrelierten und von f_0 unabhängigen Bestandteile zu zerlegen. Dabei konnte auch gezeigt werden, dass die von den Prototypen prädiktierten Amplitudenverläufe mit den realen Verläufen subjektiv vergleichbar sind.

Weiter wurde gezeigt, dass die verschiedenen Techniken für die Partitionierung der Eingangsdaten keinen wesentlichen Einfluss auf die anschließende Schätzung der Prototypen haben und erst bei der Konnektion der Signale der einzelne Oszillatoren zu teils recht starken Unstetigkeiten in den Parametertrajektorien führen. Insbesondere im Hinblick auf eine zur Auralisation gedachte Synthese der Modellprädiktionen, ist daher die überlappende Partitionierung der angrenzenden vorzuziehen, da, durch die somit notwenig werdende überlappende Konnektion, Unstetigkeiten im resynthetisierten Signal ausgeschlossen werden können.

Allerdings haben sich, trotz der teilweise deutlich unterscheidbaren Prädiktionen der Prototypen unterschiedlicher Modellkomplexität, keine signifikanten Tendenzen der Kosten, Risiken und Varianzen ergeben. Dies nährt

die These, dass die Optimierung der Kostenfunktion während der Parameterschätzung das globale Optimum noch nicht hinreichend erreicht hat. Hierfür können verschiedene Ursachen verantwortlich gemacht werden. Möglich ist ein Skalierungsproblem der Gradienten, da die Trainingsdaten nicht gleichverteilt sind im Modellraum. Dies könnte zu Konvergenzproblemen der Oszialltorfunktionen für niedrige Energiepegel führen, aber auch in den unteren Frequenzbereichen der Filterfunktion. Eine andere Möglichkeit kann im Abbruchkriterium sowie der Schrittweitenanpassung begründet sein. Das heißt, dass entweder der Abbruch des Schätzverfahrens zu früh erfolgt oder die Schrittweite nicht optimal angepasst wird. Neben der weiteren Optimierung des bestehenden Verfahrens und der Suche nach den optimalen Einstellungen für die Parameterschätzung, scheint die Verwendung eines *Conjugate Gradient* Verfahrens die aussichtsreichste Lösung, aussagekräftigere Ergebnisse zu erhalten.

Darüber hinaus ist die zentrale Frage nach der Anwendbarkeit des Modells im Kontext eines Verfahrens zur Instrumentenklassifikation, Transkription oder Quellentrennung bislang unbeantwortet. Im Anschluss, an eine Verbesserung des Lernverfahrens zur Parameterschätzung, wäre demnach das Potential des Ansatzes in den genannten Anwendungen zu evaluieren und mit bestehenden Modellen zu vergleichen. Dies kann und sollte Gegenstand weiterer Forschungsarbeiten sein, da die bislang ermittelten Ergebnisse des vorgestellten Modellansatzes in eine viel versprechende Richtung weisen.

Anhang A

Instrumente Teil 1

In Appendix A befinden sich die Ergebnisse für die Prototypen der Klarinette, des Flügels und der Violine für alle Modellkomplexitäten. In den Tabellen sind jeweils horizontal die Kosten, Risiken und die Varianzen zu je einem Modell aufgetragen. Darüber hinaus werden ebenfalls in der horizontalen Achse die Segmentlängen für das Filter angegeben. Die übereinander liegenden X-Achsen stellen die B-Spline Parametrisierung für die verschiedenen Ordnungen dar. Auf der Y-Achse ist die Parametrisierung der B-Spline Funktionen des Anregeoszillators mehrfach aufgetragen. Je einmal für eine Kombination seiner Ordnung mit der des Filters.

Innerhalb der Tabelle ist das jeweilige Modell mit der minimalen Komplexitä hellgrau unterlegt, während das Modell mit der maximalen Komplexität grau hervorgehoben wird.

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	37.9416	37.7974	37.593	37 539	37.6807	39.8688	39.8339	39.7587	39 5819	39.556	-25.1808	-25.1381	-25.0978	-25.303	-25 5722
3	5	37.2981	37.1445	36.9306	36.9689	37.1025	39.2063	39.2878	39.1852	38.9235	39.071	-25.4794	-25.3109	-25.1218	-25.4426	-25.4618
	10	37.2702	37.0776	36.7604	36.7375	36.8744	40.0509	39.6632	39.0468	38.8639	38.791	-25.6252	-25 538	-25.4226	-25.6058	-25.8161
	20	37.2099	36.9151	36.6254	36.6852	36.921	39 5238	39.1371	38.6022	38.469	38.5981	-25.5096	-25.4142	-25.3756	-25.541	-25.7313
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	38.0707	37.8705	37 5383	37 581	37.7354	40.6022	40.3039	39.6051	39.7434	39.9635	-25.037	-25.028	-25.0984	-25.4232	-25.3881
3	5	37.4266	37.2842	36.979	37.0405	37.1358	40.0323	39.6264	39.2167	39.1093	39.0647	-25.3093	-25.2101	-25.1632	-25.5421	-25 5772
	10	37.3947	37.1606	36.8321	36.8149	36.9613	40.4606	39.7677	39.2416	38.949	39.0388	-25.5708	-25.432	-25.3428	-25.7346	-25.7872
	20	37.1716	36.9595	36.6587	36.669	36.8309	39.7888	39.2297	38.7435	38.4971	38.559	-25.488	-25.3227	-25.2968	-25.6577	-25.7057
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	37.4121	37.3255	37.0711	37.136	37.2313	39.4733	39 5399	39.2701	39.1632	39.1161	-24.989	-24.7876	-24.8057	-25.0349	-25.4336
4	5	37.1687	37.0449	36.803	36.8093	36.9204	39.4619	39.3902	38.9889	38.8953	38.8272	-25.6059	-25.4169	-25.3214	-25.4449	-25.7152
	10	37.1508	37.0042	36.7139	36.7408	36.8729	39.3206	39.2085	38.8096	38.6069	38.604	-25.4404	-25.3257	-25.2071	-25.4406	-25.6737
	20	37.0122	36.7908	36 5793	36.7273	36.8408	39.2155	38.7763	38.4362	38.3942	38.5811	-25.4015	-25.2334	-25.2147	-25.4036	-25 5692
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	37.5472	37.4329	37.137	37.2035	37.3799	39.9643	39.6829	39.3343	39.2383	39.4335	-24.8985	-24.8005	-24.7778	-25.1637	-25.0768
4	5	37.3368	37.1061	36.7979	36.869	36.9868	40.01	39.4236	39.1124	38.8629	39.0199	-25.5457	-25.3826	-25.2478	-25.72	-25.6541
	10	37.2	37.0926	36.7295	36.7298	36.9119	39.4704	39.65	38.8522	38.7326	38.9174	-25.3895	-25.2378	-25.2252	-25.5187	-25 534
	20	37.1774	36.914	36.6093	36.7416	36.815	39.4649	38.8452	38 569	38.3779	38.4763	-25.3281	-25.1626	-25.1058	-25.5174	-25 5535

Tabelle A.1: Trainingsergebnisse der Klarinette bei angrenzender Partitionierung

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	38 504	38.4568	38.1228	38.1105	38.2713	41.0499	40.8201	40.2532	40.4773	40.487	-25.1251	-25.1509	-25.2879	-25.3172	-25 538
3	5	37.975	37.8473	37 5871	37 5986	37.7311	40.4577	40.3448	39.7301	39.7178	39.826	-25.3813	-25.2103	-25.3889	-25.4255	-25.6519
	10	38.0442	37.7186	37.4375	37.4503	37 5512	40 5654	40.1679	39.7562	39.6287	39.682	-25.654	-25 5362	-25.4459	-25.5914	-25.819
	20	37.7447	37.6023	37.2554	37.2556	37.4609	39.9304	39.6035	39.2727	39.1434	39.3005	-25.6001	-25.4303	-25.378	-25.5394	-25.7208
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	38.6991	38.6126	38.1163	38.1813	38.3367	41.23	41.3527	40.4974	40.315	40.7629	-25.1715	-25.0236	-25.1244	-25.5632	-25.4582
3	5	38.1672	37.9393	37.6215	37.6548	37.8194	40.7964	40.2354	40.0233	39.7549	39.8185	-25.3674	-25.2853	-25.1835	-25.6303	-25.6537
	10	38.0693	37.9333	37.4585	37 5386	37.6576	40.4836	40.233	39.9273	39.679	39.7935	-25.5798	-25.4318	-25.3218	-25.7269	-25.7721
	20	37.9107	37.6268	37.3476	37.3135	37.4486	40 5329	40.0616	39.3881	39.1921	39.2579	-25.5464	-25.3438	-25.3271	-25.6766	-25.7306
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	38.2021	38.0265	37.8152	37.8263	37.8912	40.6317	39.8659	40.0627	40.0423	40.1094	-24.8908	-25.125	-24.9273	-25.1056	-25.3774
4	5	37.9683	37.7068	37.4839	37 5014	37 5965	40.6105	39.9504	39.6676	39.6629	39.6116	-25.5552	-25.4332	-25.3921	-25.5317	-25.7573
	10	37.9211	37.6393	37.3833	37.4269	37 5494	40.6012	39.9549	39 5002	39.3573	39.5487	-25.4922	-25.3094	-25.2299	-25.4084	-25 5959
	20	37.7046	37.3984	37.2124	37.3436	37 5185	40.0284	39.3484	39.1701	39.1484	39.3751	-25.4331	-25.2353	-25.2305	-25.3844	-25 5721
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	38.2521	38.0739	37.7952	37.7766	38.0314	41.2142	40.4202	40.0467	39.9338	40.1409	-24.7698	-24.9194	-24.933	-25.312	-25.3109
4	5	37.9335	37.8438	37 5039	37 5076	37.6954	40.8781	40.3246	39.8601	39.6393	39.6857	-25.5358	-25.349	-25.2618	-25.6495	-25.7411
	10	38.0523	37.7459	37.4603	37.4949	37 5743	40.871	39.8428	39.8054	39 5012	39.4423	-25.4708	-25.2531	-25.1655	-25.5187	-25 5789
	20	37.8735	37.5547	37.1895	37.259	37.4275	40.0912	39.6171	39.2384	39.1474	39.2231	-25.355	-25.1763	-25.1748	-25.5103	-25 5386

Tabelle A.2: Trainingsergebnisse der Klarinette bei überlappender Partitionierung

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	41.9408	42.1739	42.1587	42.1279	42.0928	40.9307	41.1653	40.9935	40.8109	41.0417	-35.2277	-33.9591	-32.7339	-33.0752	-33.2936
3	5	41.7566	42.0136	41.8325	41.7881	41.8155	40.4588	40.602	40.4413	40.3595	40.5147	-35.4055	-33.8841	-33.4402	-33.5776	-33.823
	10	41.613	41.857	41.7114	41.4823	41.4677	40.0068	40.0499	39.7242	39.666	39.6996	-33.3282	-31.6949	-31 5726	-31.8909	-32.0601
	20	41.9821	41.9973	41.7302	41 5553	41 5212	40.0923	40.2034	39.9887	39.8803	39.9535	-33.8821	-32.6375	-30.9768	-30.9273	-30.9987
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	41.9191	42.323	42.3177	42.057	42.2035	40.8119	41.4244	41.2193	41.0817	40.8807	-33.9747	-34.3527	-33.1601	-33.4515	-33 5239
3	5	41.7804	42.0841	41.9858	41.8628	41.9013	40 5586	40.7687	40 5938	40.6924	40.4885	-35.1576	-34.2879	-33 5348	-34.1123	-33.9907
	10	41.6843	41.9613	41.7091	41 5598	41.6388	39.9598	40.342	39.8242	39.7933	40.0491	-32.6949	-32.9017	-31.6701	-32.2339	-32 586
	20	41.8768	42.1877	41.796	41 5577	41 591	40.2791	40.1522	39.9004	39.9682	39.8213	-34.2446	-32.2107	-31.6301	-31.1118	-31.7942
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	41.7014	41.9722	41.9653	41.9922	42.0057	40.4692	40.7883	40.7375	40 5909	40.61	-33.9846	-34.2821	-33.8016	-33.8781	-33.95
4	5	41.5438	41.8349	41.7106	41.6755	41.6553	40.3014	40.4396	40.2935	40.2719	40.0453	-33.8076	-32.7897	-31.9951	-32.2382	-32.4321
	10	41.4798	41.7304	41 5148	41.4021	41.3718	39.9392	40.1112	39.8105	39.6099	39.7201	-33.9867	-32.7238	-30.7836	-31.4407	-31.0725
	20	41.8201	41.8301	41.6167	41.4698	41.4559	40.1356	40.0494	39.8501	40.0119	40.0905	-33.6411	-31.235	-31.2572	-31.8385	-32.0187
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	41.6619	42.059	42.0539	41.9417	42.0474	40 5937	41.1056	40.8538	40 5309	40.5797	-34.9654	-34.8025	-33.7752	-33.9852	-34.3413
4	5	41.5278	41.9461	41.8417	41.6271	41.6489	40.2839	40.7177	40.6427	40.4152	40.173	-33.4568	-32.9811	-32.2777	-32.4391	-32.834
	10	41.478	41.8095	41 5918	41.3601	41.4805	40.0039	40.0363	39.8186	39.6303	39.6849	-33.8036	-31.7623	-31.2735	-31.9312	-32.0606
	20	41.7125	41.9984	41.6205	41.4963	41 5485	39.9434	40.2238	39.9914	39.8464	39.865	-32.5557	-32.0956	-31.6958	-31.7737	-32.2104

Tabelle A.3: Trainingsergebnisse des Flügels bei angrenzender Partitionierung

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.6253	28.5671	28.6966	28.7806	28.9214	29.0349	28.8111	28.984	29.0766	29.1382	-26.0248	-26.0391	-25.6514	-25.6093	-25 5393
3	5	28.1922	28.1159	28.2113	28.2656	28.4146	28.3608	28.1917	28.2947	28.3509	28.4775	-26.5052	-26.2963	-25.8501	-25.7668	-25.6806
	10	28.025	27.9344	28.0706	28.132	28.3	28.2983	28.1481	28.3747	28.423	28.6039	-26.3343	-26.1615	-25.9049	-25.7709	-25.6241
	20	27.9608	27.8964	28.0881	28.1671	28.3625	28.2569	28.1674	28.499	28 517	28.8818	-26.7404	-26.4721	-26.1159	-26.0096	-25.8607
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.6019	28.5426	28.6815	28.7599	28.9376	28.928	28.9229	29.005	29.0331	29.1947	-25.9785	-25.9797	-25 5963	-25.4775	-25.3476
3	5	28.2042	28.1399	28.238	28.2649	28.4491	28.2954	28.2856	28.3194	28.352	28.4715	-26.3916	-26.3113	-25.7625	-25.7168	-25 5502
	10	28.0112	27.9549	28.0726	28.1171	28.3052	28.3013	28.2383	28.3203	28.3759	28.5205	-26.3634	-26.3165	-25.8571	-25.6938	-25 5019
	20	27.9113	27.8815	28.0537	28.1443	28.3637	28.1559	28.2024	28.3946	28 5653	28.7462	-26.5519	-26 544	-26.101	-25.8925	-25.7558
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.2905	28.2218	28.3518	28.4054	28 5418	28 5487	28.397	28 5523	28.4916	28.6836	-25.7982	-25.4902	-25.0215	-25.4639	-25.1378
4	5	28.0523	27.9771	28.104	28.1843	28.3316	28.301	28.1495	28.3144	28.3981	28.5553	-26.7904	-26 5549	-26.007	-25.8274	-25.7567
	10	27.9433	27.8888	28.0513	28.1142	28.3001	28.207	28.111	28.3455	28.4024	28.6357	-26.1888	-26.0155	-25.7868	-25.7904	-25.6247
	20	27.9383	27.879	28.0773	28.1867	28.3565	28.2381	28.1644	28 5023	28.6761	28.7819	-26.6579	-26.465	-26.1411	-26.0003	-25.9101
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.2642	28.2262	28.3761	28.3948	28 5574	28 5085	28.4505	28.6496	28 5466	28.707	-25.7607	-25.7188	-24.6413	-25.1933	-24.8699
4	5	28.0337	27.9856	28.1041	28.1523	28.3387	28.3066	28.234	28.3203	28.3833	28.5049	-26.7306	-26 5134	-25.8521	-25.8422	-25.7043
	10	27.9388	27.8776	28.0267	28.0941	28.2906	28.1402	28.1358	28.2996	28.4133	28.532	-26.1207	-25.9881	-25.7423	-25.6304	-25.4564
	20	27.892	27.8661	28.0481	28.1657	28.3498	28.1279	28.1856	28.4152	28 5233	28.8013	-26.5338	-26.4921	-26.1079	-25.9163	-25.7464

Tabelle A.4: Trainingsergebnisse der Violine bei angrenzender Partitionierung

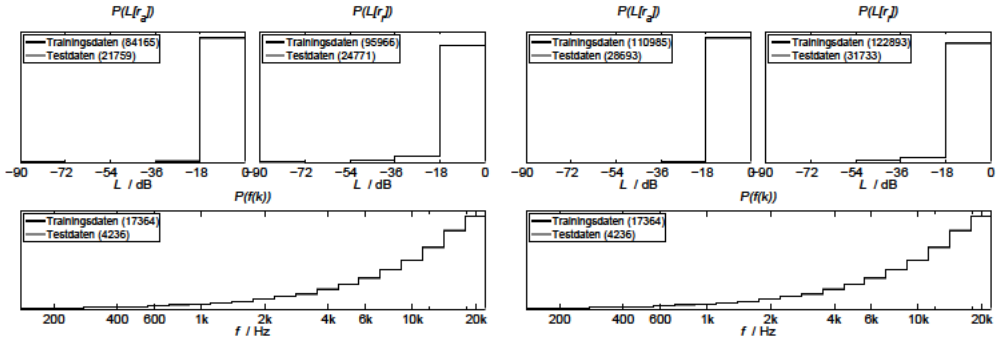
Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.3865	28.3377	28.4749	28 5657	28.7046	28.8562	28.7523	28.915	28.9802	29.1355	-26.0141	-25.9665	-25 5972	-25.5916	-25.4704
3	5	27.9454	27.8643	27.971	28.0308	28.1858	28.2575	28.0784	28.1889	28.2321	28.3814	-26.4185	-26.1001	-25.6389	-25.7626	-25 5073
	10	27.7672	27.6749	27.8132	27.8777	28.0626	28.1395	28.0217	28.1984	28.2744	28.4592	-26.3059	-26.2707	-25.8586	-25.7425	-25.601
	20	27.7078	27.6241	27.8334	27.9379	28.1222	28.0671	27.96	28.2965	28.4733	28.6433	-26.5902	-26.3573	-26.0899	-25.9833	-25.8314
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.3517	28.3182	28.4656	28 5201	28.7269	28.7814	28.7735	28.8635	29.0111	29.0771	-25.9049	-25.9311	-25 5909	-25.4119	-25.3388
3	5	27.9348	27.8692	27.9762	28.0224	28.1976	28.1769	28.1574	28.224	28.2957	28.3587	-26.2871	-26.1791	-25.4612	-25.2917	-25 5049
	10	27.7452	27.6915	27.8119	27.8721	28.0579	28.1755	28.0679	28.1675	28.2528	28.4144	-26.4996	-26.1264	-25.8167	-25.6733	-25 5055
	20	27.6511	27.6203	27.8197	27.89	28.1104	28.0109	28.0232	28.2229	28.3649	28.5596	-26.5224	-26.4384	-26.0625	-25.8789	-25.7312
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.0227	27.9781	28.1035	28.1748	28.3256	28.3783	28.2763	28.4383	28.4873	28.5546	-25.7017	-25.4006	-24.9783	-25.047	-25.1491
4	5	27.8155	27.7294	27.8444	27.9543	28.0992	28.1733	28.0152	28.1977	28.2855	28.4311	-26.74	-26.405	-25.7885	-25.6847	-25.6511
	10	27.7063	27.6448	27.7944	27.8822	28.0624	28.0829	27.9464	28.1962	28.3044	28.4811	-26.332	-25.9975	-25.7894	-25.7121	-25.609
	20	27.6796	27.6365	27.8583	27.9412	28.1275	28.0595	28.0308	28.3833	28 5531	28.6327	-26.6446	-26.4696	-26.1194	-25.9768	-25.8483
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	27.9903	27.9806	28.1166	28.1643	28.3208	28.3708	28.3289	28.4288	28 5217	28.5869	-25.5912	-25.4135	-24.958	-24.8502	-24.877
4	5	27.7964	27.7311	27.865	27.9152	28.1138	28.104	28.0793	28.1784	28.2583	28.3828	-26.3672	-26.3853	-25.9129	-25.7349	-25 5709
	10	27.6734	27.6204	27.7711	27.8661	28.0445	28.0216	28.0322	28.1087	28.2697	28.4027	-26.083	-26.1734	-25.7829	-25.6179	-25.4355
	20	27.624	27.6029	27.8087	27.9202	28.12	27.9796	28.0475	28.2827	28.4552	28.5593	-26.4628	-26.4831	-26.07	-25.8575	-25.7576

Tabelle A.5: Trainingsergebnisse der Violine bei überlappender Partitionierung

Anhang B

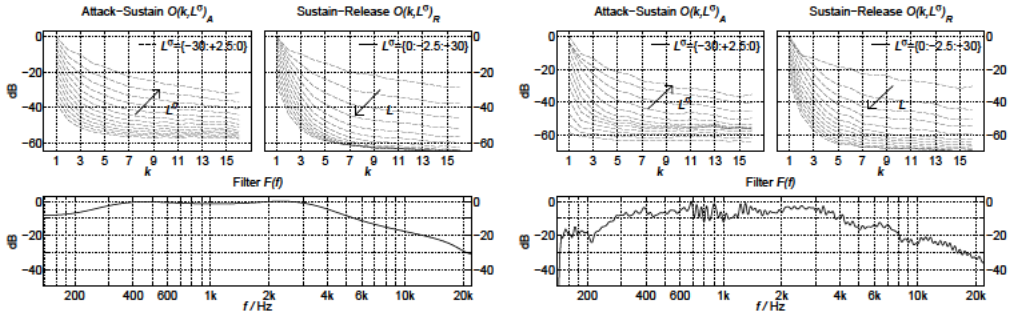
Instrumente Teil 2

B.1 Alt Saxophon



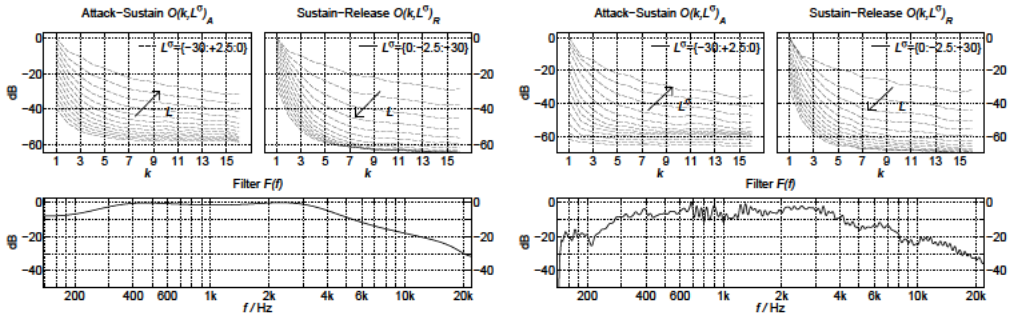
(a) Angrenzende Partitionierung (b) Überlappende Partitionierung

Abbildung B.1: Verteilung der Eingangsdaten des Alt-Saxophons im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite



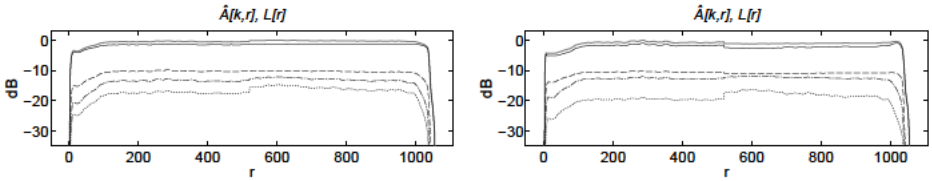
(a) Minimal (b) Maximal

Abbildung B.2: Prototypen der Alt-Saxophons mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung



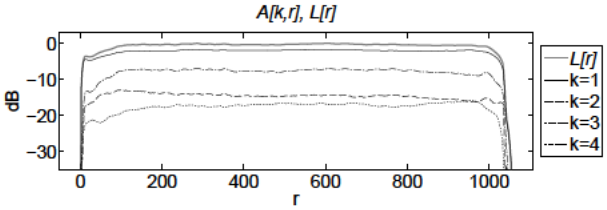
(a) Minimal (b) Maximal

Abbildung B.3: Prototypen der Alt-Saxophons mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung

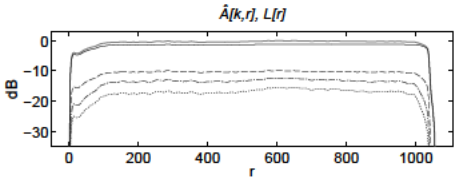


(a) Angrenzend, Minimal

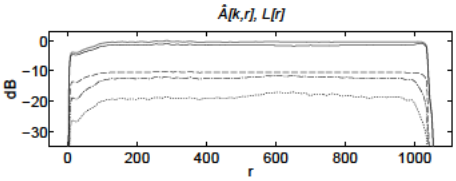
(b) Angrenzend, Maximal



(c) Realer Beispielklang als Referenzsignal



(d) Überlappend, Minimal



(e) Überlappend, Maximal

Abbildung B.4: Synthesebeispiel eines Ab4 des Alt-Saxophons

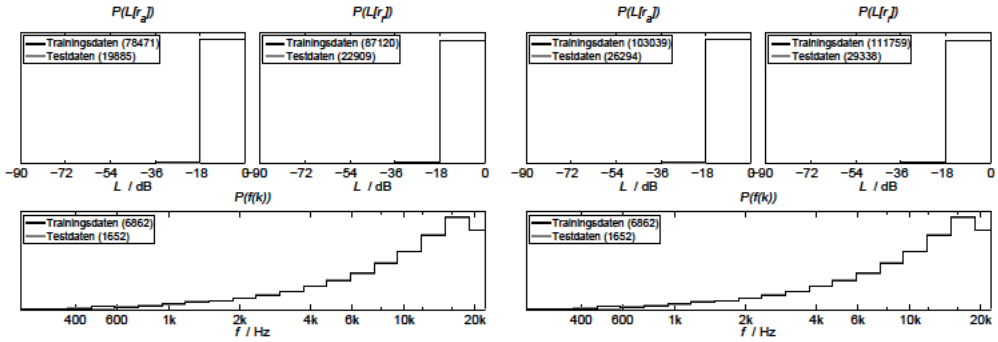
Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	34.8758	34.7431	34 5964	35.2878	35.432	31.9727	31.9166	31.682	32.4029	32.3794	-25.8074	-26.041	-26.1348	-26.261	-26.3472
3	5	34.4321	34.2543	34.0507	34.7666	34.8971	31.9043	32.0251	31 569	32 5001	32.4841	-25.6529	-25.8341	-25.9081	-26.1781	-26.2524
	10	34.0723	33.8082	33.722	34.3545	34 5507	32 568	32.1169	32.1642	32.6757	32.7991	-25.5375	-25.827	-25.7017	-25.8598	-25.9149
	20	33.849	33.6525	33 5273	34.2402	34.3958	32 5602	32.4661	32.0359	32.7649	32.8205	-25.7934	-25.8786	-25.965	-25.9829	-26.0935
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	34.9126	34.7837	34 5842	34.8542	35 5448	32.0513	31.9173	31.6144	31.9356	32.6376	-25.9028	-26.0687	-26.2059	-26.3228	-26.4141
3	5	34.3634	34.2551	34.0409	34.3184	34.94	31.9293	31.8636	31 5604	31.9358	32.5415	-25.768	-25.8911	-26.0078	-26.1288	-26.2168
	10	34.1873	33.8629	33.6901	33.9632	34 5491	32.8784	32.2598	32.0698	32.3837	32.9439	-25.5007	-25.8178	-25.8406	-25.9139	-25.9561
	20	33.9369	33.6631	33 5148	33.7254	34.3999	32.8087	32.4608	32.0738	32.2882	32.9816	-25.8063	-25.9557	-26.0378	-26.1455	-26.074
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	34.3857	34.2659	34.088	34.7743	34.9263	32.2247	32.0833	31.7377	32.4699	32.5261	-25.4798	-25.8085	-25.9282	-26.0963	-26.2341
4	5	34.2144	33.9802	33.8107	34.4958	34.656	32 5123	32.2553	31.9728	32.6264	32.7115	-25.2712	-25 5541	-25.6352	-25.7685	-25.9397
	10	33.9076	33.7242	33 5393	34.262	34.3948	32.6383	32.4535	32.034	32.7546	32.854	-25.6672	-25.8179	-25.8898	-25.93	-26.0129
	20	33.827	33.5927	33 5609	34.2101	34.3892	32.4481	32.3396	32.0494	32.6148	32.7765	-25.7638	-25.9139	-25.9156	-26.0902	-26.1122
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	34.4511	34.336	34.0918	34.288	34.9852	31.9796	32.1282	31.7083	31.862	32.6714	-25.7118	-25.8328	-25.9702	-26.0256	-26.2072
4	5	34.2013	34.0128	33.8476	34.0707	34.7076	32.4603	32.1887	32.0372	32.2514	32.9401	-25.4767	-25.6726	-25.7299	-25.8444	-25.9538
	10	33.9518	33.6754	33 5583	33.8024	34.4181	32 5226	32.0752	32.0942	32.388	32.9261	-25.8266	-25.9713	-25.9513	-26.0276	-26.0791
	20	33.8551	33.6089	33.4666	33.7916	34.393	32 5592	32.1214	31.9462	32.3126	32.8465	-25.8387	-26.0304	-26.0744	-26.1075	-26.1135

Tabelle B.1: Trainingsergebnisse des Alt-Saxophons bei angrenzender Partitionierung

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	35.3383	35.1883	35.1132	35.797	35.9274	32.3493	31.8175	32.0071	32.6877	32.6586	-25.7931	-26.067	-26.1748	-26.3233	-26.396
3	5	34.8906	34.7289	34 5925	35.2995	35.4479	32.1169	31.9049	31.9342	32 5788	32.7225	-25.6215	-25.8709	-25.9767	-26.1253	-26.2945
	10	34.7023	34.4624	34.3502	35.0048	35.2229	32 5697	32.7024	32.2985	32.942	33.0577	-25.5851	-25 5305	-25.6217	-25.7365	-25.8323
	20	34.6501	34.4164	34.345	34.9332	35.2052	32.9845	32.8457	32.406	32.9856	33.1477	-25.5263	-25.6089	-25.7032	-25.8373	-25.9142
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	35.4942	35.2979	35.1159	35.3183	36.0678	32.3978	32.3644	32.062	32.1516	32.9718	-25.8785	-26.0691	-26.2752	-26.3552	-26.4405
3	5	34.9163	34.7082	34 5952	34.8579	35.4761	32.1836	31.8611	31.9088	32.3042	32.7887	-25.7256	-25.9092	-26.0222	-26.2052	-26.2376
	10	34.7524	34.5225	34.3414	34.6254	35.2277	32.9297	32.6314	32.2998	32.7196	33.1755	-25.5029	-25.6486	-25.7375	-25.8053	-25.8549
	20	34.6467	34.4911	34.2646	34 5189	35.1415	32.9335	32.9788	32.3646	32 5717	33.1757	-25.6817	-25.6623	-25.8518	-25.978	-25.9366
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	34.9745	34.7728	34.6948	35.3526	35.4762	32.3001	32.1619	32.0318	32.7269	32.7246	-25.5749	-25.8455	-25.9863	-26.1553	-26.2247
4	5	34.6736	34.5816	34.4278	35.0696	35.2301	32 5477	32.487	32.2132	32.831	32.9096	-25.3444	-25 5022	-25.6137	-25.7347	-25.8532
	10	34.6255	34.3961	34.3002	34.9697	35.1444	32 5949	32.4539	32.3475	32.9642	33.1035	-25.6753	-25.7773	-25.7011	-25.8267	-25.9356
	20	34.5142	34.3768	34.3057	35.0219	35.2171	32 5518	32.4053	32.326	32.9409	33.1465	-25.7365	-25.8335	-25.7249	-25.8663	-25.9658
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	34.9895	34.7982	34.687	34.9208	35 552	32.4969	32.2128	32.0814	32.4287	32.9136	-25.6451	-25.8429	-26.0722	-26.2327	-26.2571
4	5	34.8509	34.5684	34.4461	34.6943	35.3106	32.8464	32.6271	32.2525	32.643	33.0589	-25.3136	-25 5278	-25.718	-25.9033	-25.8734
	10	34.7347	34.481	34.2705	34 5255	35.142	32.9	32.8111	32.2957	32.6135	33.1934	-25.5847	-25.637	-25.8044	-25.9151	-25.9624
	20	34.6667	34.3703	34.244	34 5321	35.1471	32.823	32 5099	32.296	32 5233	33.0626	-25.7099	-25.8229	-25.8587	-25.9772	-25.9921

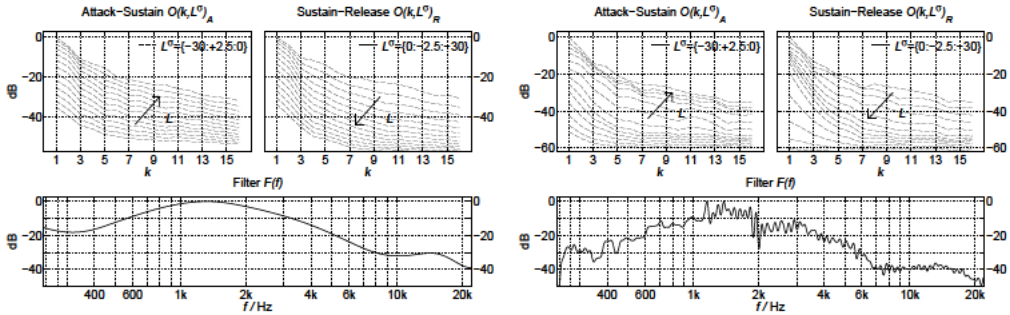
Tabelle B.2: Trainingsergebnisse des Alt-Saxophons bei überlappender Partitionierung

B.2 Oboe



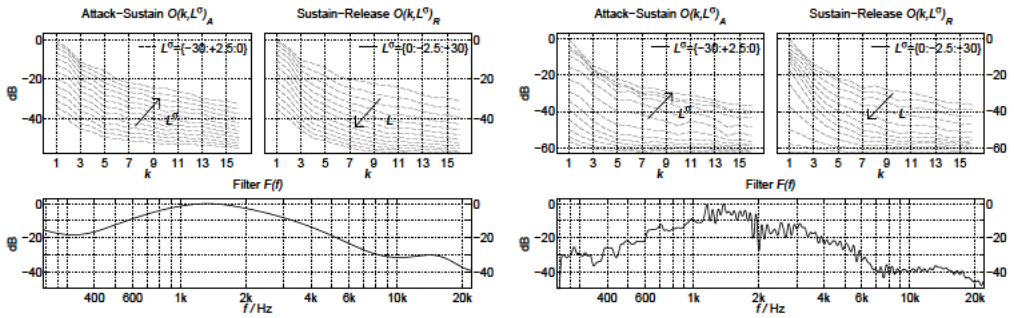
(a) Angrenzende Partitionierung (b) Überlappende Partitionierung

Abbildung B.5: Verteilung der Eingangsdaten der Oboe im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite



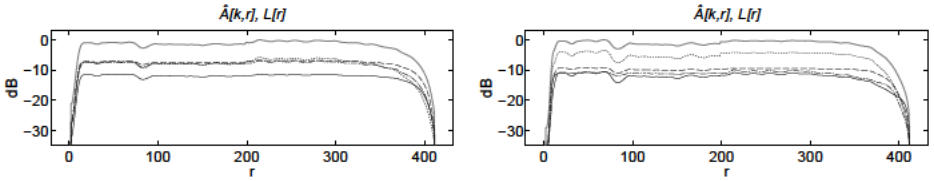
(a) Minimal (b) Maximal

Abbildung B.6: Prototypen der Oboe mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung

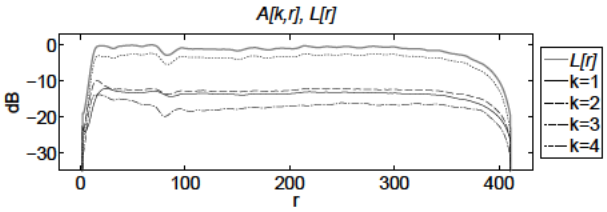


(a) Minimal (b) Maximal

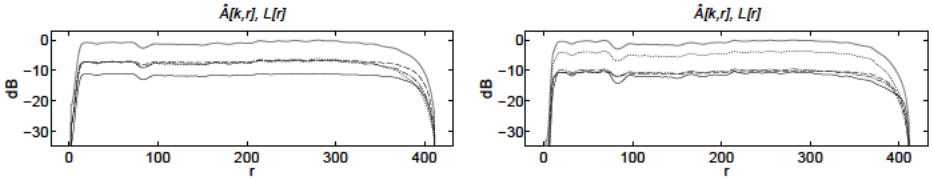
Abbildung B.7: Prototypen der Oboe mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung



(a) Angrenzend, Minimal (b) Angrenzend, Maximal



(c) Realer Beispielklang als Referenzsignal



(d) Überlappend, Minimal (e) Überlappend, Maximal

Abbildung B.8: Synthesebeispiel eines D4 der Oboe

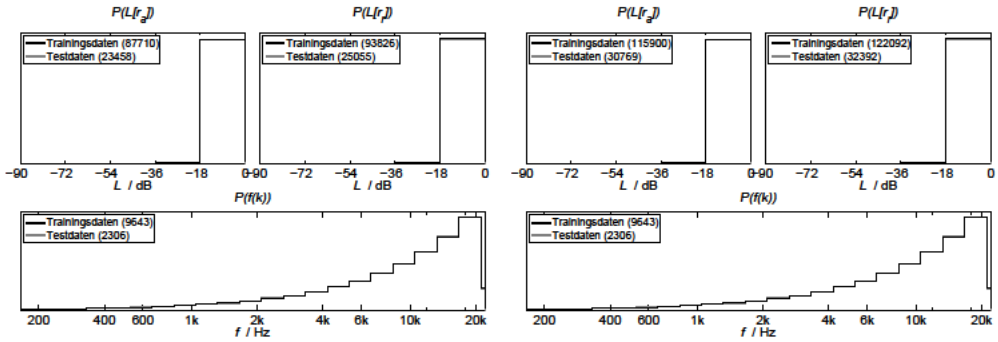
Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28 567	28.4406	28.2486	28.4043	29.6881	31.238	31.2349	31.2608	31.3392	32.5908	-25.0051	-25.0018	-24.7104	-24.9772	-24.4843
3	5	28 522	28.4009	28.2036	28.3556	29.6105	30.9747	31.1643	31.2292	31.2334	32.5066	-24.6063	-24.6002	-24.1822	-24.3098	-23.9226
	10	28.5205	28.3514	28.2221	28.3425	29.6353	30.934	30.7701	30.8389	30.8198	32.1127	-24.9661	-24.768	-24.1935	-24.4178	-23.8631
	20	28.4747	28.3591	28.183	28.3937	29.7206	30.7522	30.6018	30.7713	30.818	32.1015	-25.0251	-24.7159	-24.4363	-24.5651	-24.1576
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.6774	28.5001	28.2985	28.4565	28.9938	31.0789	31.2721	31.3586	31.4228	31.9086	-24.8276	-24.8707	-24.8264	-24.9215	-24.694
3	5	28.7438	28.4779	28.2451	28.3816	28.9858	31.1109	31.3897	31.1037	31.1971	31.9717	-24.5614	-24 5065	-24.2205	-24.2573	-24.2913
	10	28.6815	28.4533	28.2079	28.3637	28.919	30.9696	30.9997	30.8362	30.8649	31.5865	-24.8211	-24.7374	-24.4624	-24.4557	-24.3948
	20	28.6975	28.3911	28.1814	28.3305	28.8221	30.8193	30.8493	30.7436	30.7847	31.3503	-24.7556	-24.7559	-24 5323	-24.4968	-24 5849
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.4445	28.3867	28.1977	28.3525	29.6074	30.9832	31.2522	31.2174	31.3071	32.5791	-24.6857	-24 5278	-24.1794	-24.4055	-24.0138
4	5	28.5564	28.3945	28.1856	28.3184	29.6087	30.875	30.9664	30.9777	30.9861	32.4542	-24.611	-24 5132	-23.7268	-23.9842	-23.4353
	10	28.4381	28.3614	28.1409	28.3352	29 5728	30.7866	30.7657	30.7571	30.8068	32.06	-24.9331	-24.7221	-24.264	-24.4843	-24.067
	20	28.3249	28.2996	28.1092	28.3786	29.6534	30 5049	30 5536	30.7411	30.8506	32.1486	-24.757	-24 5621	-24.3693	-24.44	-24.0632
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.6413	28.4137	28.257	28.3666	28.9889	31.0977	31.3466	31.2966	31.3314	31.9995	-24.3598	-24.4299	-24.297	-24.2943	-24.2664
4	5	28.6719	28.5087	28.264	28.3571	28.941	30.9917	31.225	31.0802	31.0148	31.6668	-24.5659	-24.3625	-23.7305	-23.945	-24.0198
	10	28.6148	28.407	28.143	28.3164	28.8506	31.0402	30.895	30.7947	30.8664	31.4665	-24.8392	-24.7342	-24.4774	-24.4015	-24.4911
	20	28.5631	28.3427	28.1259	28.2865	28.7826	30.6793	30.8314	30.7254	30.8085	31.359	-24.6505	-24.6803	-24.455	-24.3938	-24.4881

Tabelle B.3: Trainingsergebnisse der Oboe bei angrenzender Partitionierung

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.8832	28.7272	28.56	28.7087	30.0062	31.2312	31.2516	31.3798	31.3245	32.6125	-25.0469	-25.0219	-24.8161	-24.8442	-24.4453
3	5	28.8672	28.7369	28 5454	28.6785	29.9205	31.0051	31.223	31.2848	31.2951	32.5463	-24.5496	-24 5319	-24.2144	-24.2975	-23.9424
	10	28.9058	28.7829	28 5567	28.6977	29.9726	30.7496	30.7872	30.9044	30.8872	32.1543	-24.9359	-24.6987	-24.1065	-24.2492	-23.8722
	20	28.9606	28.7577	28 5796	28.7834	30.1425	30.7809	30.7226	30.7986	30.9364	32.2118	-24.9172	-24.7305	-24.3841	-24.52	-24.154
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.9804	28.784	28 5767	28.7566	29.3446	31.2417	31.4938	31.3565	31.4224	32.0825	-24.9287	-25.0775	-24.7809	-24.8547	-24.8193
3	5	29.0349	28.8313	28 5856	28.7169	29.2466	31.273	31.3403	31.1755	31.2298	31.9789	-24.4184	-24.468	-24.1662	-24.2264	-24.2867
	10	29.0522	28.9721	28 5865	28.6837	29.2744	31.0116	30.9775	30.8768	30.9875	31.5959	-24.8365	-24.6707	-24.3462	-24.2585	-24.4497
	20	29.0466	28.7861	28 5507	28.6817	29.2203	30.9154	30.7768	30.8305	30.8773	31 57	-24.8031	-24.6968	-24 5239	-24.4465	-24.6085
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.8415	28.6605	28 5153	28.6619	29.9924	31.0214	31.2253	31.2629	31.3518	32.6375	-24.4801	-24 5685	-24.2423	-24.4182	-24.1336
4	5	28.881	28.7444	28 5437	28.643	29.9315	30.8447	30.8262	31.0506	31.0209	32.391	-24.53	-24.2929	-23.6519	-23.9214	-23.4472
	10	28.8372	28.6883	28 5046	28.6864	29.9578	30.6899	30.6551	30.7972	30.833	32.0795	-24.9	-24.6905	-24.1972	-24.4452	-24.1013
	20	28.8413	28.6971	28 5385	28.7622	30.1012	30.6886	30.6551	30.753	30.8975	32.1888	-24.9174	-24.6997	-24.3772	-24.443	-24.0833
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	28.9389	28.7286	28 5547	28.6603	29.3011	31.0385	31.3723	31.3213	31.3887	32.1338	-24.3665	-24.4873	-24.2838	-24.3291	-24.3238
4	5	29.053	28.8633	28.6021	28.6828	29.2215	30.967	31.1834	31.057	31.1013	31.8254	-24.4153	-24.0502	-23.8355	-23.8298	-23.8972
	10	28.9342	28.739	28 5315	28.6239	29.161	30.8629	30.9027	30.8488	30.869	31.5138	-24.7926	-24.7159	-24.3737	-24.3125	-24.4974
	20	29.0409	28.8042	28.523	28.6864	29.1735	30.7217	30.7769	30.8025	30.9125	31.4537	-24.7926	-24.7159	-24.3737	-24.3125	-24.4974

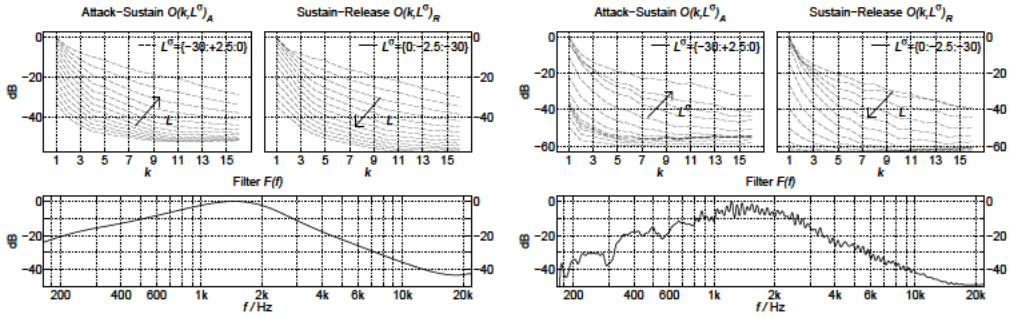
Tabelle B.4: Trainingsergebnisse der Oboe bei überlappender Partitionierung

B.3 Trompete



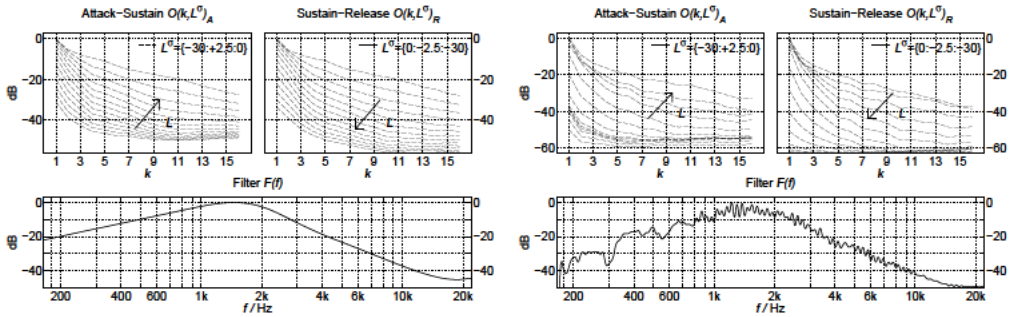
(a) Angrenzende Partitionierung (b) Überlappende Partitionierung

Abbildung B.9: Verteilung der Eingangsdaten der Trompete im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite



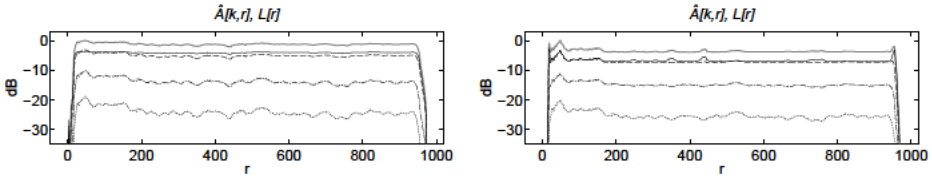
(a) Minimal (b) Maximal

Abbildung B.10: Prototypen der Trompete mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung

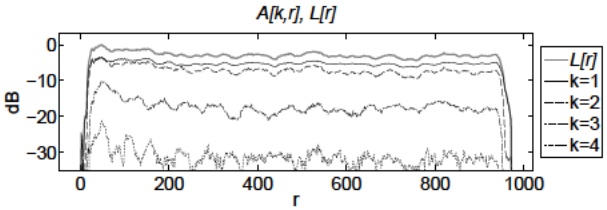


(a) Minimal (b) Maximal

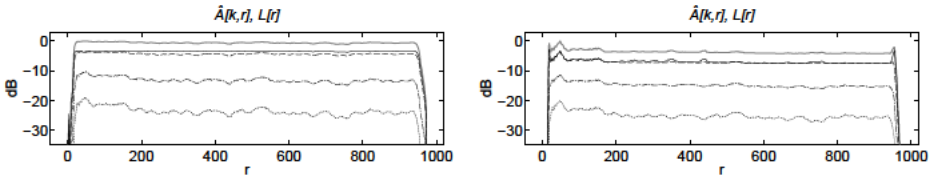
Abbildung B.11: Prototypen der Trompete mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung



(a) Angrenzend, Minimal (b) Angrenzend, Maximal



(c) Realer Beispielklang als Referenzsignal



(d) Überlappend, Minimal (e) Überlappend, Maximal

Abbildung B.12: Synthesebeispiel eines Gb5 der Trompete

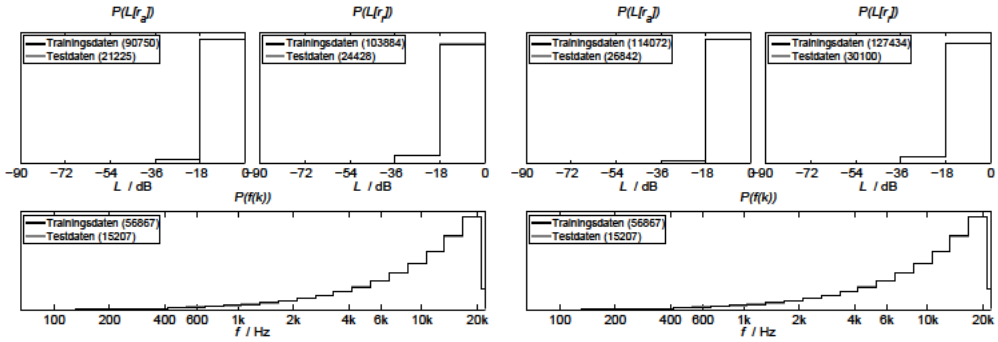
Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	30.6598	30.2241	29.8385	29.8092	29.7631	29 5174	28.9261	28 5657	28.6584	28.5541	-25.1599	-25.4382	-25.6471	-25.8221	-25.8477
3	5	30.5808	30.0564	29.6563	29 5998	29 5764	29.2723	28.6788	28.3531	28.415	28.3703	-24.5949	-24.747	-24.6883	-24.9294	-24.6679
	10	30.4223	29.9656	29.4992	29.4524	29.4054	29.2167	28.7471	28.4202	28.3547	28.3912	-25.1037	-25.2141	-25.189	-25.1637	-24.7948
	20	30.3047	29.8791	29.349	29.2893	29.3134	29.3551	28.6745	28.4409	28.3536	28.5626	-25.4216	-25.8641	-25 5449	-25.4472	-25.2102
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	30.7788	30.4034	29.9865	29.9237	30.0515	29 5098	28.8469	28.6792	28 5968	28.4628	-25.177	-25.4282	-25 5152	-25.789	-25.6058
3	5	30.638	30.189	29.7308	29.6923	29.7997	29.3839	28.7622	28.468	28.407	28.292	-24.642	-24.8083	-24 5573	-24.8735	-24.6553
	10	30.5239	30.1335	29.613	29 5888	29.4949	29.3354	28.6481	28.4468	28.4029	28.4025	-25.1687	-25.4982	-25.174	-25.1615	-25.0156
	20	30.4027	30.0001	29.4662	29.4613	29.3932	29.1281	28 5402	28 5548	28 5487	28.4565	-25.8033	-25.9141	-25.432	-25.4535	-25.3984
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	30.5795	30.0833	29.6929	29.6688	29.6816	29.3446	28.7581	28.4146	28.4041	28.4056	-24.7023	-24.913	-24.7984	-25.0626	-25.1543
4	5	30.283	29.9638	29 5881	29.4843	29.4641	29.3316	28.7632	28.3676	28.386	28.3578	-24.5983	-24.6964	-24 5849	-24.5659	-24 5404
	10	30.2367	29.7683	29.3694	29.3295	29.3385	28.6782	28.6074	28.399	28.3567	28.4194	-26.0012	-25.7161	-25.4404	-25.368	-25.3114
	20	30.2603	29.7027	29.3241	29.2725	29.2832	29 5623	28.6673	28.4288	28.4356	28.4004	-25.3574	-25.8557	-25 5156	-25.3478	-25.1889
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	30.6518	30.2094	29.776	29.7475	29.7147	29.3592	28.7489	28 523	28.4367	28.284	-24.6497	-24.8396	-24.6874	-24.8098	-25.1414
4	5	30.5287	30.306	29.6561	29 5943	29.6616	29.2251	28.6885	28 5333	28.3292	28.3276	-24.6396	-24.7087	-24.4786	-24.5818	-24 5493
	10	30.2976	29.9379	29.4115	29.4527	29.4586	29.1485	28.6272	28.4449	28.3427	28.3095	-25.6015	-25 5142	-25.2906	-25.3958	-25.4152
	20	30.3732	29.8525	29.3538	29.3787	29.3376	29.1129	28 5759	28.4884	28.3888	28.3699	-25.6995	-25.8779	-25.4167	-25.2701	-25.2579

Tabelle B.5: Trainingsergebnisse der Trompete bei angrenzender Partitionierung

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	30.7701	30.3379	30.0316	29.8736	29.8426	29.4121	28.9548	28.6189	28.6864	28.6754	-24.8429	-25.3405	-25 5578	-25.8312	-25.7603
3	5	30.7406	30.2957	29.8014	29.781	29.8284	29.4403	28.8683	28 506	28 5051	28.4818	-24.5392	-24.6558	-24.6406	-24.6729	-24.8369
	10	30.6964	30.2235	29.8534	29.7778	29.7135	29 5054	28.9648	28.7145	28.6563	28.5674	-25.1167	-24.8853	-24.7518	-24.7879	-24.8541
	20	30.6967	30.2303	29.6872	29.6453	29.644	29.457	28.949	28 5841	28.6002	28.6753	-25.4788	-25.8123	-25.7283	-25.6093	-25.4125
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	30.8906	30.5317	30.0672	30.0866	29.9789	29.4657	28.9527	28.7029	28.6301	28.5504	-25.0765	-25.48	-25.4782	-25.896	-25.6509
3	5	30.908	30.4839	29.8785	29.9531	29.8793	29.4547	28.8872	28 5844	28 5156	28.3857	-24.5805	-24.7621	-24 5065	-24.7169	-24.6456
	10	30.891	30.5276	30.02	29.9431	29.8396	29 5205	28.8401	28.7117	28.6345	28.5197	-25.1675	-25.4591	-24.9188	-24.8676	-24.8939
	20	30.8355	30.3909	29.7802	29.7546	29.7595	29.431	28.95	28.757	28.6502	28.5167	-25.7579	-25.8959	-25 5903	-25.5732	-25 5532
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	30 591	30.1684	29.8826	29.7957	29.8358	29.3865	28.8695	28 5577	28 5835	28.5672	-24.5353	-24.7476	-25.0801	-24.9575	-24.8585
4	5	30.7237	30.299	29.7996	29.6959	29.7206	29.4717	28.8876	28 5079	28 5093	28.4948	-24.6192	-24.7287	-24 5828	-24.536	-24.4878
	10	30.5109	29.998	29.7402	29.6476	29.6075	29 5018	28.8614	28 5924	28 5385	28.6124	-25.3157	-25 5511	-25.4144	-25.3571	-25.1867
	20	30.6336	30.1095	29.6631	29.6211	29.6553	29.4106	28.8378	28.6507	28 5457	28.6117	-25.2377	-25.8569	-25.7051	-25.5589	-25.3522
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	30.8595	30.3639	30.0113	29.915	29.8475	29 516	28.8167	28.6443	28 5496	28.5087	-24.6511	-24.8199	-25.0866	-24.7991	-24.6726
4	5	30.8685	30.5143	29.9215	29.8176	29.8914	29 5881	28.9902	28.6364	28 5415	28.4451	-24.6232	-24.7072	-24.4802	-24.5718	-24 5509
	10	30.6485	30.3235	29.7974	29.7135	29.7357	29.3914	28.8178	28.6175	28.4805	28.483	-25.4474	-25.7199	-25.3265	-25.3949	-25.1887
	20	30.8444	30.2498	29.7708	29.6845	29.6828	29 5376	28.7975	28 568	28.6531	28.529	-25.4539	-25.8658	-25 5953	-25.5313	-25 5425

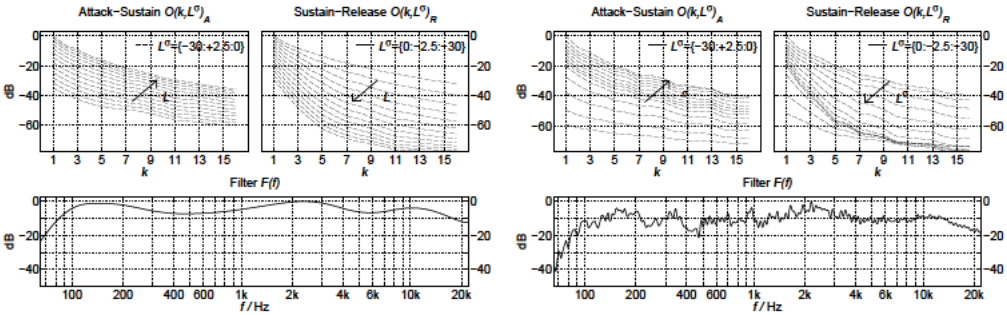
Tabelle B.6: Trainingsergebnisse der Trompete bei überlappender Partitionierung

B.4 Violoncello



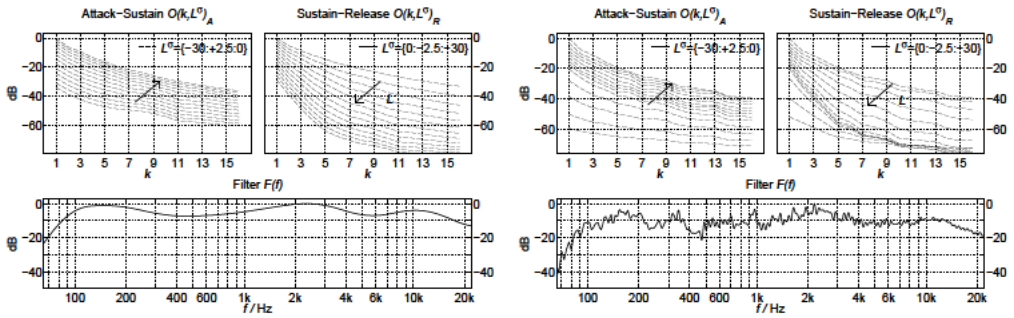
(a) Angrenzende Partitionierung (b) Überlappende Partitionierung

Abbildung B.13: Verteilung der Eingangsdaten des Violoncellos im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite



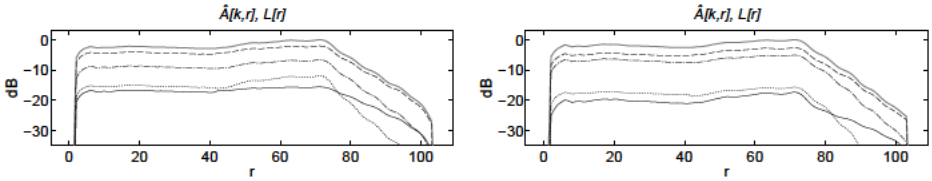
(a) Minimal (b) Maximal

Abbildung B.14: Prototypen des Violoncellos mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung

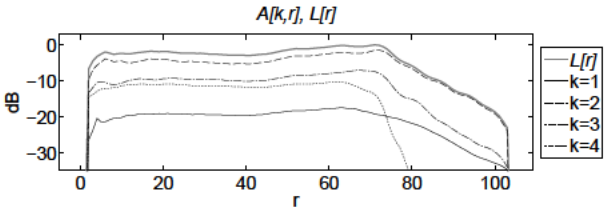


(a) Minimal (b) Maximal

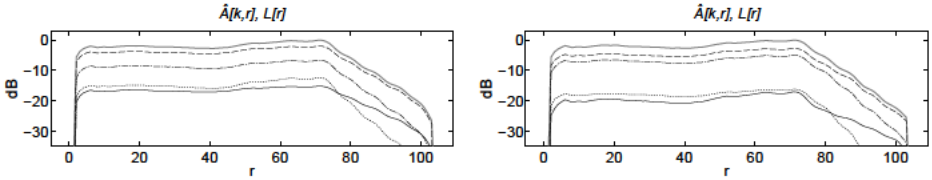
Abbildung B.15: Prototypen der Violoncellos mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung



(a) Angrenzend, Minimal (b) Angrenzend, Maximal



(c) Realer Beispielklang als Referenzsignal



(d) Überlappend, Minimal (e) Überlappend, Maximal

Abbildung B.16: Synthesebeispiel eines Db2 des Violoncellos

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	20.454	20.4762	20.6129	20.7176	20.9895	20.623	20.675	20.7983	20.93	21.1686	-29.1795	-29.1493	-29.0382	-28.8195	-28.6381
3	5	20.3779	20.3473	20 5102	20.633	20.9042	20 5426	20.4828	20.7155	20.8894	21.097	-29.2783	-29.4769	-29.0231	-28.7729	-28.6505
	10	20.1231	20.1261	20.3155	20.4671	20.7865	20.2754	20.2776	20.4649	20.6548	20.968	-29.3416	-29.3786	-29.0027	-28.7795	-28 5313
	20	20.1064	20.1505	20.4178	20 5902	20.9333	20.2365	20.2906	20.6386	20.794	21.0948	-29.4409	-29.4047	-29.044	-28.8544	-28.6047
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	20.4657	20.4595	20.6108	20.7176	20.8455	20.63	20.6331	20.7791	20.9048	21.0277	-29.2236	-29.2491	-28.9401	-28.4568	-28.4768
3	5	20.3546	20.3496	20 5088	20.6154	20.7456	20.4761	20 5026	20.6674	20.81	20.9106	-29.4736	-29.4317	-29.066	-28.532	-28 5661
	10	20.1362	20.1283	20.3128	20.4261	20 5765	20.2913	20.2828	20 5419	20.6367	20.7373	-29.3836	-29.3511	-28.9245	-28.4305	-28.4242
	20	20.1097	20.1382	20.3876	20 5254	20.7062	20.1975	20.283	20.6177	20.6519	20.8698	-29.4458	-29.4008	-28.9392	-28.5332	-28 5007
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	20.4378	20.4339	20 5958	20.7013	20.997	20.6153	20 5634	20.7531	20.8842	21.1715	-28.9177	-29.1912	-28.8071	-28.6595	-28.3513
4	5	20.1802	20.175	20.3224	20.4612	20.7373	20.2901	20.349	20.4811	20.6957	20.911	-29.5338	-29 5221	-29.2143	-28.9285	-28.7258
	10	20.0977	20.1086	20.2991	20.4563	20.7787	20.2244	20.2423	20.4689	20.6535	20.9282	-29.3457	-29.2675	-28.9508	-28.7818	-28 5422
	20	20.1058	20.1518	20.4044	20.6216	20.9522	20.2451	20.3046	20 5629	20.7956	21.1705	-29.4108	-29.3409	-28.9951	-28.8539	-28.6295
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	20.4284	20.422	20 5921	20.7032	20.8278	20 5121	20 5301	20.7703	20.8804	20.9912	-29.3238	-29.3283	-28.6116	-28.1956	-28.227
4	5	20.1903	20.1628	20.3272	20.4505	20 5783	20.302	20.343	20 5117	20.6539	20.7569	-29.5627	-29.4975	-29.1103	-28.606	-28.6422
	10	20.1123	20.0922	20.2915	20.4072	20 5737	20.2066	20.2258	20.4634	20.6087	20.7097	-29.3524	-29.2626	-28.8487	-28.4016	-28.4592
	20	20.0948	20.1297	20.389	20 5317	20.7297	20.2481	20.285	20.6229	20.7269	20.9162	-29.4107	-29.3393	-28.879	-28.5428	-28 5198

Tabelle B.7: Trainingsergebnisse des Violoncellos bei angrenzender Partitionierung

Order		C					R					σ^2 / dB				
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	20.2387	20.2401	20.3983	20 5095	20.7926	20 5541	20 5955	20.7465	20.8578	21.1217	-29.1435	-29.0908	-29.0095	-28.8234	-28.6108
3	5	20.1114	20.1088	20.2783	20.4043	20.6836	20.4045	20.4273	20 5998	20.8045	21.0223	-29.4233	-29.388	-29.1162	-28.7897	-28.6503
	10	19.8783	19.8626	20.0625	20.2106	20 5312	20.1727	20.1818	20.4875	20.6729	20.9614	-29.3918	-29.4534	-29.096	-28.8384	-28 5882
	20	19.841	19.8733	20.1321	20.3445	20.7035	20.1159	20.1542	20 5246	20.8124	21.1731	-29.4735	-29.4482	-29.1024	-28.8901	-28.636
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	20.2383	20.2305	20.3951	20 5096	20.6449	20 5785	20 577	20.727	20.8558	20.9442	-29.1619	-29.182	-28.8876	-28.3997	-28.4838
3	5	20.1132	20.1107	20.2773	20.3846	20 5219	20.3964	20.4333	20.6012	20.732	20.9258	-29.4575	-29.369	-29.0078	-28.5316	-28.4934
	10	19.8774	19.8665	20.0674	20.1864	20.3348	20.1831	20.1597	20.4421	20.6592	20.7317	-29.4292	-29.3961	-28.9606	-28.4964	-28.491
	20	19.8405	19.8697	20.1366	20.2647	20.4639	20.1025	20.2384	20 5034	20.7029	20.8219	-29.4998	-29.4328	-28.9869	-28.5467	-28 5679
3		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	20.2006	20.192	20.3633	20.4863	20.7785	20.6708	20.4887	20.6462	20.8497	21.0638	-28.3568	-29.1122	-28.9141	-28.5494	-28.4577
4	5	19.922	19.9069	20.0752	20.2098	20 5044	20.2461	20.2243	20.3856	20.6498	20.9298	-29.4753	-29 5769	-29.2335	-28.9478	-28.7026
	10	19.8517	19.8489	20.0628	20.2204	20 5385	20.1367	20.2023	20.4729	20.6818	20.9705	-29.3821	-29.4121	-29.0037	-28.8067	-28 5502
	20	19.8249	19.8814	20.1628	20.3736	20.6945	20.1129	20.2211	20.4898	20.7998	21.0834	-29.455	-29.4143	-29.0855	-28.8957	-28.6703
4		0 5	1	3	6	12	0 5	1	3	6	12	0 5	1	3	6	12
	3	20.2036	20.1927	20.3681	20.4754	20.6212	20 5063	20.4838	20.6694	20.8038	20.9017	-29.0697	-29.0775	-28.7229	-28.1761	-28.3017
4	5	19.9074	19.9023	20.0762	20.1999	20.3286	20.3027	20.236	20.4089	20 5585	20.7504	-29.4126	-29 536	-29.1357	-28.6243	-28.6226
	10	19.8407	19.845	20.0361	20.1699	20.3175	20.0939	20.1532	20 503	20 5731	20.6965	-29.3955	-29.3385	-28.8814	-28.464	-28.4833
	20	19.8257	19.8766	20.1405	20.2919	20.4658	20.0904	20.2076	20 562	20.7829	20.832	-29.4714	-29.3914	-28.9457	-28.5535	-28 5934

Tabelle B.8: Trainingsergebnisse des Violoncellos bei überlappender Partitionierung

Anhang C

DVD

Auf der beigefügten DVD befinden sich folgende zusätzlichen Daten:

DATA/RWC/GRAPHICS/TRAIN/MODEL

Enthält die Grafiken zu den Prototypen in allen Modellkomplexitäten.

DATA/RWC/GRAPHICS/SYNTH/PREDICTED

Enthält die Grafiken mit den prädiktierten Verläufen für die ersten vier Partialwellen.

DATA/RWC/GRAPHICS/SYNTH/REFERENCE

Enthält die Grafiken mit den Testdaten-Verläufen der ersten vier Partialwellen.

DATA/RWC/GRAPHICS/DATASTATS_MIN

Enthält die Grafiken für die in den Modellraum projizierten Eingangsdaten bei minimaler Modellkomplexität.

DATA/RWC/GRAPHICS/DATASTATS_MAX

Enthält die Grafiken für die in den Modellraum projizierten Eingangsdaten bei maximaler Modellkomplexität.

DATA/RWC/SYNTH/PREDICTED

Die von den Modellen mit minimaler und maximaler Komplexität erzeugten Samples im *.wav Format.

DATA/RWC/SYNTH/REFERENCE

Die aus den Sinusoidalmodellen der Testdaten resynthetisierten Audiosamples im *.wav Format.

DATA/RWC/TRAIN/... *.MAT

Enthält alle mat-Dateien der trainierten Modelle, inklusive der vorverarbeiteten Trainings- und Testdaten.

DOCS/SYNTH.PDF

Ein Dokument mit allen Synthesegrafiken für Modelle mit minimaler und maximaler Komplexität.

DOCS/THESIS.PDF

Dieses Dokument.

SOURCE

Der Matlab®-Quellcode der Implementierung.

Tabellenverzeichnis

4.1 Matrix- und Vektor-Konventionen 46

5.1 Prozessbezeichner 64

6.1 Verwendete Instrumente 66
6.2 Analyse Parameter 67
6.3 Vorverarbeitungsparameter 68
6.4 Konstante Trainingsparameter 69
6.5 Variable Trainingsparameter 70

7.1 Extremwerte der Modellkomplexitäten 73
7.2 Klarinette (Kosten, Risiko und Varianz) 76
7.3 Flügel (Kosten, Risiko und Varianz) 81
7.4 Violine (Kosten, Risiko und Varianz) 83

A.1 Ergebnisse der Klarinette (Angrenzend) 91
A.2 Ergebnisse der Klarinette (Überlappend) 92
A.3 Ergebnisse des Flügels (Angrenzend) 93
A.4 Ergebnisse der Violine (Angrenzend) 94
A.5 Ergebnisse der Violine (Überlappend) 95

B.1 Ergebnisse des Alt-Saxophons (Angrenzend) 99
B.2 Ergebnisse des Alt-Saxophons (Überlappend) 100
B.3 Ergebnisse der Oboe (Angrenzend) 103
B.4 Ergebnisse der Oboe (Überlappend) 104
B.5 Ergebnisse der Trompete (Angrenzend) 107

B.6 Ergebnisse der Trompete (Überlappend) 108
B.7 Ergebnisse des Violoncellos (Angrenzend) 111
B.8 Ergebnisse des Violoncellos (Überlappend) 112

Abbildungsverzeichnis

2.1 Exemplarische Amplitudenhüllkurven 7
2.2 Amplitudenfolgen 9
2.3 Spektren 10
2.4 *Blackman*-Fenster 12
2.5 Spektrogramme 13
2.6 Die 2 zentralen Schritte zur Erzeugung des Sinusoidalmodells . 14
2.7 Sinusoidalmodelle 15
2.8 Fehleroberfläche 17
2.9 Gradientenabstieg 18
2.10 10-*fold*-Kreuzvalidierung 20

4.1 Bestimmung von *Attack/Release* bei kontinuierlicher Anregung 31
4.2 Angrenzende Bereichsgrenzen 31
4.3 Überlappende Bereichsgrenzen 32
4.4 Bestimmung des *Attack/Release* bei impulshafter Anregung . . 33
4.5 Bereichsgrenzen bei impulshafter Anregung 33
4.6 B-Spline Funktionen U_w über dem Pegel der relativen Energie L 36
4.7 B-Spline Kurven V_q über der Frequenz f 38
4.8 Konnektion der Anregungen bei überlappenden Bereichen . . . 43
4.9 Pseudocode des Offline-Algorithmus 50
4.10 Pseudocode des Online-Algorithmus 51

5.1 Quellcode Beispiel für objektbasierte Modellierung 53
5.2 Beispielnutzung von Strukturobjekten mit *Callbacks* 53
5.3 Gesamtschema des Programms 55

5.4 Schema des Import-Prozesses 57
5.5 Schema des Analyse-Prozesses 58
5.6 Schema des Trainings-Prozesses 59
5.7 Schema des Synthese-Prozesses 61

7.1 Verteilung der Eingangsdaten der Klarinette im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite 75
7.2 Prototypen der Klarinette mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung 77
7.3 Prototypen der Klarinette mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung 78
7.4 Synthesebeispiel eines Bb4 der Klarinette 79
7.5 Verteilung der Eingangsdaten des Flügels im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite 80
7.6 Prototypen des Flügels mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung 81
7.7 Synthesebeispiel eines F3 des Flügels 82
7.8 Verteilung der Eingangsdaten der Violine im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite 83
7.9 Prototypen der Violine mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung 84
7.10 Prototypen der Violine mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung 84
7.11 Synthesebeispiel eines Ab3 der Violine 86

B.1 Verteilung der Eingangsdaten des Alt-Saxophons im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite . 96
B.2 Prototypen der Alt-Saxophons mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung 97
B.3 Prototypen der Alt-Saxophons mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung 97
B.4 Synthesebeispiel eines Ab4 des Alt-Saxophons 98
B.5 Verteilung der Eingangsdaten der Oboe im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite 101

B.6 Prototypen der Oboe mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung 101
B.7 Prototypen der Oboe mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung 102
B.8 Synthesebeispiel eines D4 der Oboe 102
B.9 Verteilung der Eingangsdaten der Trompete im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite 105
B.10 Prototypen der Trompete mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung 105
B.11 Prototypen der Trompete mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung 106
B.12 Synthesebeispiel eines Gb5 der Trompete 106
B.13 Verteilung der Eingangsdaten des Violoncellos im Modellraum für 5 Oszillator- und Filtersegmente mit $\frac{1}{3}$ Oktavbreite 109
B.14 Prototypen des Violoncellos mit minimaler und maximaler Modellkomplexität bei angrenzender Partitionierung 109
B.15 Prototypen der Violoncellos mit minimaler und maximaler Modellkomplexität bei überlappender Partitionierung 110
B.16 Synthesebeispiel eines Db2 des Violoncellos 110

Literaturverzeichnis

[1] Mototsugu Abe and Julius O. Smith III. Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of fft magnitude peaks. In *Audio Engineering Society (AES) Convention*, volume 117, october 2004.

[2] Albert S. Bregman. *Auditory Scene Analysis - The Perceptual Organization of Sound*. MIT Press, third edition, 2001.

[3] Juan José Burred. *From Sparse Models to Timbre Learning: New Methods for Musical Source Separation*. PhD thesis, Technische Universität, September 2008.

[4] Juan Jose Burred, Axel Röbel, and Thomas Sikora. Dynamic spectral envelope modeling for timbre analysis of musical instrument sounds. *IEEE Transactions on Audio, Speech and Language Processing*, 18(3), March 2010.

[5] M. Casey and A. Westner. Separation of mixed audio sources by independent subspace analysis. In *Proc. International Computer Music Conference (ICMC)*, 2000.

[6] E.C. Cherry. Some experiments on the recognition of speech, with one and two ears. *The Journal of the Acoustical Society of America*, 25(5):975–979, 1953.

[7] Alain Cheveigné and Hideki Kawahara. YIN, A fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 4(111):1917–1930, April 2002.

[8] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2 edition, November 2000.

[9] Stephan Euler. *Grundkurs Spracherkennung*. Frdr. Vieweg & Sohn Verlag, Wiesbaden, 2006.

[10] Masataka Goto and Takuichi Nishimura. Rwc music database: Music genre database and musical instrument sound database. In *in ISMIR*, pages 229–230, 2003.

[11] Simon Haykin. *Neural Networks: A Comprehensive Foundation (2nd Edition)*. Prentice Hall, 2 edition, July 1998.

[12] Simon Haykin and Zhe Chen. The Cocktail Party Problem. *Neural Computation*, 17:1875–1902, 2005.

[13] Toni Heittola, Anssi Klapuri, and Tuomas Virtanen. Musical instrument recognition in polyphonic audio using source-filter model for sound separation. In *10th International Society for Music Information Retrieval Conference (ISMIR)*, pages 327–332, 2009.

[14] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis*. John Wiley & Sons, Inc, June 2001.

[15] Kunio Kashino. *Auditory Scene Analysis in Music Signals*, chapter 10, pages 299–302. Signal Processing Methods for Music Transcription (Anssi Klapuri (eds)). Springer, 2006.

[16] A. Klapuri. Analysis of musical instrument sounds by source-filter-decay model. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 1, pages I–53–I–56, 2007.

[17] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.

[18] Jürgen Meyer. *Akustik der Musikinstrumente*, chapter 4, pages 123–180. Handbuch der Audiotechnik (Stefan Weinzierl (Hrsg)). Springer-Verlag, Berlin Heidelberg, 2008.

[19] Brian C. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, fifth edition, April 2003.

[20] Michael Möser. *Technische Akustik*. Springer-Verlag Berlin Heidelberg, 2003.

[21] Alan V. Oppenheim et al. *Zeitdiskrete Signalverarbeitung*. Pearson Studium, 2. edition, 2005.

[22] Lawrence R. Rabiner and Ronald W. Schafer. *Digital Processing of Speech Signals*. Prentice Hall, Signal Processing Series, 1978.

[23] Christoph Reuter. *Der Einschwingvorgang nichtperkussiver Musikinstrumente*. Peter Lang - Europäischer Verlag der Wissenschaften, 1995.

[24] Axel Röbel. Adaptive additive modeling with continuous parameter trajectories. *IEEE Transactions on Audio, Speech and Language Processing*, 14(4):1440–1453, July 2006.

[25] David F. Rosenthal and Hiroshi G. Okuno (eds). *Computational Auditory Scene Analysis*. Lawrence Erlbaum Ass0ociates, 1998.

[26] Diemo Schwarz and Matthew Wright. Extensions and applications of the sdif sound description interchange format. In *International Computer Music Conference (ICMC2000)*, 2000.

[27] Xavier Serra. *Musical Sound Modeling with Sinusoids Plus Noise*, chapter 3, pages 91–122. Musical Signal Processing (Roads, Curtis and Pope, Steven Travis and Piccialli, Aldo and de Poli, Giovanni (eds)). Swets & Zeitlinger B. V., 1997.

[28] Wayne Slawson. *The Color of Sound: A Theoretical Study in Musical Timbre*, volume 3, pages 132 – 141. University of California Press on behalf of the Society for Music Theory, 1981.

[29] Tuomas Virtanen and Anssi Klapuri. Analysis of polyphonic audio using source-filter model and non-negative matrix factorization. In *Advances in Models for Acoustic Processing, Neural Information Processing Systems Workshop (AMAC)*, 2006.

[30] D.L. Wang and G.J. Brown (eds). *Computational Auditory Scene Analysis: Principles, algorithms and applications*. IEEE Press/Wiley Interscience, 2006.

[31] E. Zwicker and H. Fastl. *Psychoacoustics - Facts and Models*. Springer-Verlag Berlin Heidelberg, third edition, 2007.

Online Quellen

[32] Analysis/Synthesis Team at IRCAM. SuperVP, PM2 and Psola analyse. http://forumnet.ircam.fr/702.html, 2007. Zuletzt abgerufen: 12.04.2010.

[33] Alain Cheveigné. Der Yin Algorithmus. http://audition.ens.fr/adc/sw/yin.zip. Zuletzt abgerufen: 12.04.2010.

[34] Lawrence Fritts. University of Iowa - Music Instrument Samples. http://theremin.music.uiowa.edu/MIS.html, 2002. Zuletzt abgerufen: 22.04.2010.

[35] Frank Opolko and Joel Wapnick. McGill University Master Samples. http://www.music.mcgill.ca/resources/mums/html/index.htm, 2002. Zuletzt abgerufen: 22.04.2010.

[36] Diemo Schwarz. SDIF Sound Description Interchange Format. http://www.ircam.fr/sdif/download, 2003. Zuletzt abgerufen: 12.04.2010.

[37] Julius O. Smith III. Spectral Audio Signal Processing. http://www.dsprelated.com/dspbooks/sasp/, 2007. Zuletzt abgerufen: 12.04.2010.

[38] Oliver Woodford. export_fig. http://www.mathworks.com/matlabcentral/fileexchange/23629-exportfig, 2010. Zuletzt abgerufen: 12.04.2010.