

Masterarbeit

# **INTERAKTIVE PARTITUREN IN DER MUSIKPRODUKTION**

Simon Waloschek

Mat.-Nr.: 15343086

Betreut durch

Prof. Dr.-Ing. Aristotelis Hadjakos

und

Dr.-Ing. Axel Berndt

Eingereicht am 23. Februar 2017



# Zusammenfassung

Die vorliegende Masterarbeit beschäftigt sich mit der Integration digitaler Noten in den Produktionsprozess klassischer Musik. Trotz der inzwischen vollständig digitalen Aufnahme und Nachbearbeitung von Musik wird in der Regel mit gedruckten Notenmaterialien gearbeitet. Dieser Bruch in der Medienverwendung führt bei der Nachbearbeitung des Audiomaterials zu Zeitverlusten, da stetig zwischen Computerbildschirm und Papiernoten gewechselt werden muss. Im Rahmen dieser Arbeit wurde ein Lösungsansatz entwickelt, der diesen Medienbruch umgeht. Aufgenommenes Audiomaterial und die Darstellung digitaler Noten auf dem Computer werden dabei synchronisiert. In der Praxis wird dadurch die Navigation innerhalb der Aufnahmeprojekte effektiv beschleunigt. Eine Vorstudie konnte zeigen, dass die Zeit für das Auffinden von Audio-Positionen anhand von Noteneintragungen – eine während des Schnitts häufig wiederkehrende Aufgabe – um durchschnittlich 79% gesenkt wurde.



# Inhaltsverzeichnis

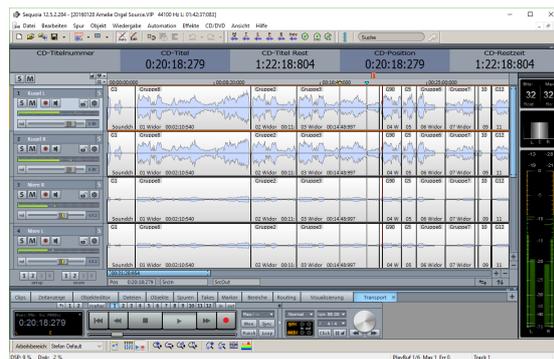
<b>1. Einleitung</b>	<b>1</b>
<b>2. Grundlagen</b>	<b>5</b>
2.1. Repräsentationsformen von Musik . . . . .	6
2.1.1. Audio-Aufnahme . . . . .	6
2.1.2. Symbolische Kodierung . . . . .	7
2.1.3. Grafische Notation (Noten) . . . . .	8
2.1.4. Transformationen . . . . .	10
2.2. Audio-to-Score Alignment . . . . .	11
2.2.1. CQT - Constant Q Transformation . . . . .	12
2.2.2. Chroma-Vektoren . . . . .	13
2.2.3. Dynamic Time Warping . . . . .	14
2.3. Einordnung in diese Arbeit . . . . .	17
<b>3. Produktion klassischer Musik</b>	<b>19</b>
3.1. Produktionsablauf . . . . .	20
3.1.1. Vorbereitung . . . . .	20
3.1.2. Aufnahme . . . . .	21
3.1.3. Nachbearbeitung . . . . .	23
3.2. Fazit . . . . .	26
<b>4. Navigationshilfen in DAWs</b>	<b>27</b>
4.1. Alternative Audiodarstellungen . . . . .	28
4.1.1. Erweiterte Wellenformen . . . . .	28
4.1.2. Strukturorientierte Ansätze . . . . .	30
4.1.3. Spektrogramme . . . . .	32
4.2. Synchronisation mehrerer Takes . . . . .	33
4.3. Einfluss auf diese Arbeit . . . . .	34
<b>5. Interaktive Partituren</b>	<b>37</b>
5.1. Nutzerbefragung . . . . .	38

5.2. Navigationskonzept . . . . .	40
5.2.1. Oberflächengestaltung . . . . .	41
5.2.2. Bedienkonzept . . . . .	42
5.3. Implementierung . . . . .	44
5.3.1. Notendarstellung . . . . .	45
5.3.2. Berechnung der Chroma-Features . . . . .	47
5.3.3. Alignment . . . . .	49
5.3.4. User Interface . . . . .	50
5.4. Evaluation . . . . .	52
5.4.1. Testdesign und Durchführung . . . . .	52
5.4.2. Ergebnisse . . . . .	54
<b>6. Fazit</b>	<b>57</b>
<b>A. Anhang</b>	<b>59</b>

# 1

## Einleitung

Die Produktion klassischer Musik wird heutzutage weitestgehend digital durchgeführt. Sämtliche Produktionsschritte – von der Aufnahme über die Nachbearbeitung bis hin zur Erstellung des Abspielmediums – nutzen Computer als Hilfs- und Arbeitswerkzeug. Basis dieser Musik ist jedoch weiterhin gedruckter Notentext, der sowohl Musikern<sup>1</sup> als auch ausführende Produzenten bzw. Tonmeistern als Interpretations- und Kommunikationsgrundlage dient. Trotz der voranschreitenden Verbreitung von Tablet-Computern und mobilen Endgeräten mit adäquaten Bildschirmgrößen, haben sich diese bislang nicht als modernes Pendant klassischer Partituren etablieren können. Besonders im digitalen Nachbearbeitungsprozess innerhalb der Musikproduktion haben sich im Kontrast zu Partituren Darstellungsweisen behauptet, die keinen direkten



**Abbildung 1.1.:** Beispielhafte Annotationen eines Tonmeisters im Notentext (**links**) gegenüber der Wellenformdarstellung des aufgezeichneten Audiomaterials in einem Audibearbeitungsprogramm (**rechts**).

<sup>1</sup>In dieser Arbeit wird aus Gründen der leichteren Lesbarkeit auf das konsequente Gendering verzichtet. Im Zweifel sind jeweils sowohl die weibliche als auch die männliche Form gemeint.

Bezug zum Notentext haben. Der Fokus liegt hier stark auf den akustisch-physikalischen Eigenschaften des Audiomaterials und erfordert daher zum effizienten Arbeiten ein gewisses Maß an Erfahrung. Viele Kommentare und Informationen, die für die Nachbearbeitung essentiell sind, werden jedoch während der Aufnahme weiterhin handschriftlich im Notentext eingetragen, sodass letztendlich beide Darstellungen der Musik für die Bearbeitung genutzt werden (siehe Abbildung 1.1).

Dieser Medienbruch kostet im Arbeitsprozess viel Zeit, da häufig zwischen den Ansichten gewechselt werden muss, ohne dass diese in irgendeiner Weise miteinander verknüpft sind. Im Rahmen dieser Masterarbeit wird daher evaluiert, ob die Nutzung interaktiver Partituren die Nachbearbeitung beschleunigen kann. Hauptaugenmerk liegt auf der Zusammenführung der in Audioprogrammen gängigen Darstellungsweisen einerseits und digitalem Notentext andererseits, um so die Navigation innerhalb des aufgenommenen Audiomaterials zu erleichtern. Grundlage für die Erarbeitung eines entsprechenden Konzepts bildet eine Analyse des Produktionsprozesses klassischer Musik.

## **Gliederung**

Die Masterarbeit gliedert sich wie folgt: Im Grundlagenkapitel 2 werden zunächst die wichtigsten Repräsentationsformen von Musik, wie beispielsweise Audioaufnahmen oder Notentext, eingeführt. Begrifflichkeiten, die zur Diskussion über aufgenommene Musik und Notentexte unabdingbar sind, werden darin ebenso erläutert wie die Beziehungen zwischen diesen unterschiedlichen Formaten. Da im Laufe der Arbeit digitale Noten mit Einspielungen dieser verknüpft werden sollen, werden darüber hinaus die mathematischen Grundlagen zur Analyse der Audioaufnahmen gelegt. Ein kompakter Überblick über ein gängiges Verfahren zur Synchronisierung von Noten und Aufnahmen, das in späteren Teilen Anwendung findet, schließt das Kapitel ab.

Kapitel 3 widmet sich einem kompakten Überblick über den Aufnahme- und Nachbearbeitungsprozess klassischer Musik. In ihm wird also der Kontext der Arbeit erschlossen und die Notwendigkeit einer Verknüpfung zwischen Noten und Aufnahmen beleuchtet. Gleichzeitig bietet das Kapitel einen interessanten Einblick in eine Arbeitsweise, die gerne „hinter verschlossenen Türen“ gehalten wird.

Forschungsprojekte und -gebiete, die inhaltlich an diese Masterarbeit grenzen, werden in Kapitel 4 diskutiert. Als klassisches „Related-Work-Kapitel“ beinhaltet es unterschiedliche Ansätze zur Nutzung synchronisierter Darstellungen in Audioeditoren, aber auch generelle Ideen zur weiteren Vereinfachung der Navigation innerhalb aufgenommener Musik.

---

Den Kern bildet in Kapitel 5 schließlich die konzeptionelle Beschreibung und Implementierung eines neuen Navigationsinterfaces. Es stellt sowohl Noten als auch Audioaufnahmen auf eine Weise verknüpft dar, welche den in der Einleitung beschriebene Medienbruch vermeidet. Eine Vorstudie zur potentiellen Nützlichkeit eines solchen Interfaces sowie die anschließende Evaluation rahmen das Kapitel ein.

Im abschließenden Résumé der Arbeit werden Ansatzpunkte zur Weiterentwicklung skizziert.





# 2

## Grundlagen

<b>2.1. Repräsentationsformen von Musik</b> . . . . .	<b>6</b>
2.1.1. Audio-Aufnahme . . . . .	6
2.1.2. Symbolische Kodierung . . . . .	7
2.1.3. Grafische Notation (Noten) . . . . .	8
2.1.4. Transformationen . . . . .	10
<b>2.2. Audio-to-Score Alignment</b> . . . . .	<b>11</b>
2.2.1. CQT - Constant Q Transformation . . . . .	12
2.2.2. Chroma-Vektoren . . . . .	13
2.2.3. Dynamic Time Warping . . . . .	14
<b>2.3. Einordnung in diese Arbeit</b> . . . . .	<b>17</b>

---

Im Folgenden werden die für das Verständnis der vorliegenden Arbeit nötigen Grundlagen dargelegt. Zunächst wird auf verschiedene Möglichkeiten zur Darstellung von Musik eingegangen. Beziehungen zwischen diesen Darstellungen und die spezifischen Einsatzbereiche werden erklärt.

Der zweite Teil des Kapitels widmet sich der Synchronisation von Noten und Audio. Zu diesem Zweck werden nicht allgemein bekannte mathematische Konstrukte hergeleitet, soweit sie zum Verständnis der ebenfalls eingeführten Synchronisationsalgorithmen notwendig sind.

Um dem Leser das Zusammenspiel der einzelnen Themen zu verdeutlichen, schließt das Kapitel mit einer Einordnung der vorgestellten Grundlagen in den Kontext dieser Arbeit. Es sei angemerkt, dass Teile der einzelnen Themen im Detail recht komplex sind und daher nicht in ihrer Gesamtheit ausschöpfend behandelt werden können. Verweise auf weiterführende Informationen werden daher begleitend angegeben.

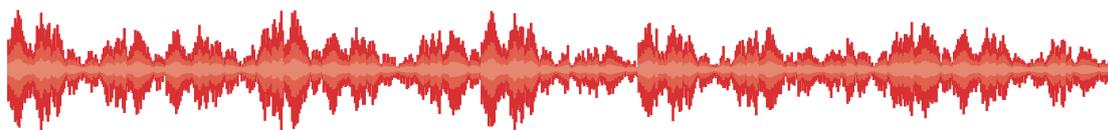
## 2.1. Repräsentationsformen von Musik

Musik ist eine Kunstgattung, die nüchtern betrachtet aus strukturierten Schallereignissen besteht. Um diese Strukturen festzuhalten, haben sich verschiedene sowohl physikalische als auch digitale Repräsentationsformen bzw. -formate entwickelt. Je nach Zweck unterscheiden sich diese Formen stark voneinander und bilden daher stets nur Teilaspekte der gesamten Musik ab. Dieser Abschnitt ist eine Einführung in die gängigsten Repräsentationsformen, geordnet nach dem Grad ihrer Abstraktion. Er konzentriert sich dabei auf jene, die im Verlauf dieser Arbeit genutzt werden und stellt keinesfalls eine vollständige Übersicht dar. Zusätzlich wird auf mögliche Transformationen zwischen diesen Formaten eingegangen.

### 2.1.1. Audio-Aufnahme

Aus physikalischer Sicht besteht jeder Ton aus periodischen Änderungen des Luftdrucks. Die durch Schwingungen von Objekten – z.B. einer Gitarrensaite oder einer angeschlagenen Stimmgabel – induzierten Druckschwankungen breiten sich als überlagerte Wellen (Schall) im Raum aus. Diese Wellen werden definiert durch ihre *Amplitude*, also dem physikalischen Äquivalent zur Lautstärke, sowie durch ihre *Frequenz*, die wir als unterschiedliche Tonhöhen wahrnehmen können. Klänge mit nur einer Frequenz (Sinus-Ton) sind jedoch nur ein theoretisches Konstrukt und kommen in der Natur nicht vor; Töne von akustischen Instrumenten bestehen neben der vornehmlich gehörten und tonhöhenbestimmenden *Grundfrequenz* aus einer Reihe von *Obertönen*, deren Frequenzen den unterschiedlich gewichteten Vielfachen der Grundfrequenz entsprechen. Diese charakteristische Obertonstruktur bestimmt den Klang eines Instruments, sein *Timbre*. Auch nichtperiodische Anteile des Schalls (z.B. Rauschen) zählen hierzu.

Bei der Aufnahme von Musik werden diese überlagerten Wellen, also die Druckschwankungen in der Luft über die Zeit, mit Mikrofonen erfasst. Der momentane Druck bzw. die Druckveränderung wird in schneller Folge abgetastet und gespeichert. Die so aufgezeichneten Messwerte lassen sich direkt als *Wellenform* mit fortlaufender horizontaler Zeitachse (siehe Abbildung 2.1) darstellen. Diese eignet sich besonders zur schnellen Erfassung der zeitlichen Veränderung der Hüllkurve, also der (physikalischen) Lautstärke des Materials.



**Abbildung 2.1.:** Wellenformdarstellung einer Audio-Datei.

### 2.1.2. Symbolische Kodierung

Um Musik auf symbolischer Ebene prozessierbar zu machen, benötigt es eine abstraktere Form der Kodierung. Dazu werden musikalische Attribute wie beispielweise Tonhöhe und -dauer von Noten in definierten Strukturen festgehalten. Zwei dieser Kodierungsformate sind aufgrund ihres Einsatzzweckes von besonderem Interesse:

**MIDI - Musical Instrument Digital Interface** Ursprünglich zur Verbindung von musikalischen Instrumenten – z.B. Keyboards und Synthesizern – entworfen, haben sich das MIDI-Protokoll und -Format seit ihrer Einführung im Jahre 1982 zu einem universellen Austauschformat für Steuer- und Noteninformationen entwickelt. Eine in MIDI kodierte Musiksequenz besteht im Wesentlichen aus einer Abfolge von *Note-on-* und *Note-off-Ereignissen* mit definierten Zeitstempeln. Diese bestimmen jeweils den Start und das Ende einer gespielten Note. Zusätzlich legen die *Note-on-Ereignisse* zum Kodieren der Dynamik ganzzahlige Lautstärkewerte zwischen 0 und 127 fest. Vortragsanweisungen wie z.B. in Abbildung 2.2 können in MIDI daher nur abgebildet werden, wenn sie sich auf diese Parameter reduzieren lassen. Sie können daher nicht wieder ohne Informationsverlust in eine Notendarstellung überführt werden. Somit dient MIDI als reine Musiksequenzkodierung, die – außer einer optionalen Instrumentenbezeichnung als Metadatum – keine Aussagen über den Klangcharakter liefert.

Durch seine simple und offene Struktur ist MIDI sehr weit verbreitet und wird insbesondere im Umfeld der Audioproduktion häufig als Format für virtuelle Instrumente und Synthesizer genutzt. Aktuelle Notationsprogramme bieten durchweg eine MIDI-Unterstützung an. Eine detaillierte Beschreibung des Formats ist zu finden bei Selfridge-Field [43] bzw. in der offiziellen Spezifikation [27].

**MEI - Music Encoding Initiative** Das XML-basierte Kodierungsformat MEI [39] ist ein Open-Source-Ansatz, der aus den Bedürfnissen der musikhistorischen Forschung heraus entstanden ist. Es hat den Anspruch, musikalische Dokumente – Drucke ebenso wie Manuskripte – möglichst genau abzubilden. Neben zahlreichen bibliothekarisch relevanten Metadaten ist auch das Speichern des Layouts von handgeschriebenen oder gedruckten Noten möglich. Dies gestattet die genaue digitale Wiedergabe der ursprünglichen Drucklegung.

Zu diesem Zweck erweitert MEI das verbreitete *MusicXML-Format* [24], das von vielen Notationsprogrammen zum Speichern von symbolischem Notentext genutzt wird. MusicXML beschränkt sich auf den reinen Noteninhalt und einfachste Formatierungsanweisungen wie beispielsweise Systemumbrüche und ist daher in seiner Verwendbarkeit über einfache Drucke hinaus limitiert.

MEI findet stetig wachsende Verwendung in der musikwissenschaftlichen Editions- und Verarbeitungs-forschung. Der derzeitige Mangel an nötigen Be- und Verarbeitungswerk-zeugen hindert das Format bislang jedoch an der Erschließung angrenzender Forschungs- und Arbeitsbereiche wie z.B. die Musikproduktion.

Im Gegensatz zum MIDI-Format legt MEI die Interpretation der Musik nicht zwangsläufig fest, sondern beschreibt vorrangig die vom Komponisten definierten Spielanweisungen sowie den vom Verleger gesetzten Druck. Da im weiteren Verlauf der Arbeit häufig die digitale Darstellung von Noten und somit eine mög-lichst getreue Wiedergabe des Layouts gefordert wird, eignet sich MEI als Kodie-rungsformat besser als MusicXML oder gar MIDI.

### 2.1.3. Grafische Notation (Noten)

Die moderne Notenschrift (engl. *Common Practice Notation*, CPN) ist das heutzutage gängige System zur grafischen Darstellung westlicher Instrumental- und Vokalmusik. Es definiert eine feste Menge musikalischer Symbole und textueller Anweisungen. Ziel dieser Notation ist im weitesten Sinne, dem Musiker Spielanweisungen für Übe- bzw. Konzertsituation zur Verfügung zu stellen, die eine klangliche Aufführung eines Wer-kes gemäß der Vorstellung des Komponisten ermöglichen. Sie erhebt dabei jedoch nicht den Anspruch einer absoluten musikalischen „Wahrheit“, sondern beschränkt

The image shows a musical score for a string quintet, page 24. It consists of five staves. The top staff is marked with a 'C' in a box and the tempo 'un poco mosso'. The score includes various musical notations such as notes, rests, and dynamic markings. The dynamics 'dim. molto' and 'pp' (pianissimo) are used throughout. The score is written in a key signature of two sharps (D major or F# minor) and a 4/4 time signature.

**Abbildung 2.2.:** Beispiel für moderne Notenschrift (Streichquintett, Auszug).

Das Diagramm zeigt eine musikalische Partitur mit sechs Instrumenten: Klarinette, Klavier, Violine I, Violine II, Viola und Violoncello. Die Partitur ist in vier Hauptbereiche unterteilt, die durch farbige Linien markiert sind:

- System:** Ein roter Rahmen umschließt die gesamte Partiturzeile für ein Instrument.
- Akkolade:** Eine gelbe geschweifte Klammer fasst mehrere Notenzeilen eines Instruments zusammen.
- Notenzeile:** Eine grüne Linie markiert eine einzelne Notenzeile.
- Gruppe:** Ein brauner Rahmen umschließt die Partiturzeilen aller Instrumente in einer Gruppe.
- Takt:** Eine violette Klammer markiert einen Takt in einer Notenzeile.
- Metrum:** Ein blauer Rahmen markiert das Metrum (Taktart) einer Notenzeile.
- Schlüssel:** Ein grauer Rahmen markiert den musikalischen Schlüssel einer Notenzeile.

**Abbildung 2.3.:** Bezeichnungskonventionen für Strukturelemente moderner Noten.

sich naturgemäß auf Vortragsanweisungen wie z.B. Dynamik, Tempo und Artikulation (siehe Abbildung 2.2). Dies lässt dem aufführenden Musiker weitreichenden Spielraum für seine persönliche Interpretation. Gesetzte Noten dienen somit als primäres Format für den musikalisch-fachlichen Austausch.

Abbildung 2.3 zeigt die überordnenden Strukturbestandteile einer modernen Partitur und definiert die entsprechende Nomenklatur, die im weiteren Verlauf dieser Arbeit Anwendung findet. Jedes Instrument besitzt mindestens eine *Notenzeile*. Mehrere Zeilen werden mithilfe einer geschweiften Klammer zu *Akkoladen* zusammengefasst. Um die einzelnen Instrumentenklassen (Holzblasinstrumente, Streicher etc.) besser optisch voneinander abzusetzen, werden diese außerdem häufig zu (*Stimm-*)*Gruppen* verbunden. Eine gesamte Partiturzeile mit allen Instrumenten wird *System* genannt. Prinzipiell kann die horizontale Achse als rechtsläufige Zeitachse betrachtet werden. Auf ihr wird der rhythmische Verlauf festgelegt, weshalb übereinander notierte Noten eines Systems gleichzeitig erklingen. Die vertikale Achse hingegen bestimmt für jede Notenzeile separat die Tonhöhe.

Vorsicht ist geboten bei den Begriffen „System“ und „Zeile“; diese werden – selbst von professionellen Musikern – häufig fälschlicherweise synonym verwendet. Für eine umfassende Einführung in die Notenschrift und ihre historische Entwicklung wird an dieser Stelle auf die Notationskunde von Schmid [42] als weiterführende Literatur verwiesen.

#### 2.1.4. Transformationen

Nicht immer steht die für einen bestimmten Zweck benötigte Repräsentationsform eines Werkes zur Verfügung. So können beispielsweise gesetzte Noten eines Werkes erforderlich sein, das aber nur als Aufnahme zur Verfügung steht. In solchen Fällen ist eine Transformation nötig. Folgende Möglichkeiten stehen hierbei zur Verfügung:

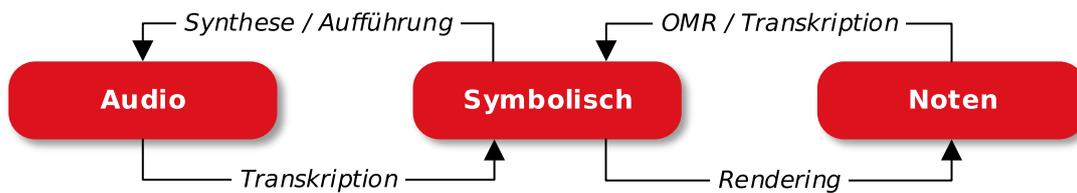
**Synthese / Aufführung** Soll aus einer symbolischen Kodierung eine anhörbare Aufnahme erstellt werden, stehen zwei Wege zur Verfügung: Eine Möglichkeit ist, die Musik am Computer mithilfe von *Virtuellen Instrumenten* zu generieren. Je nach Anspruch an die Gefälligkeit der resultierenden Aufnahme kann dies entweder vollautomatisch oder in „Handarbeit“ von Musikproduzenten geschehen.

Die andere – deutlich aufwendigere – Möglichkeit ist die Aufnahme des Werkes mit echten Musikern. Hierbei muss im Regelfall auch Notenmaterial zur Verfügung gestellt werden. Trotz seiner Komplexität ist dieser Weg üblich bei der Einspielung neu komponierter Musik: Die vom Komponisten häufig digital vorgelegte symbolische Kodierung wird zunächst professionell in Noten gesetzt, um die Musiker anschließend aus diesen spielen zu lassen.

**Transkription** Der Rückweg, also die Umwandlung einer gespielten Interpretation in Symbolform, bedarf heutzutage noch intensive Nacharbeit durch Menschen. Auch hier kann wieder auf automatisierte Systeme zurückgegriffen werden, die allerdings nur in einfachsten Fällen verwertbare Resultate liefern. Automatische Transkriptionssoftwares entwickeln sich zwar stetig weiter, differenzieren aber z.B. zwischen unterschiedlichen Instrumenten nur sehr fehlerhaft. Dieser Schritt wird daher bis dato mithilfe eines geübten Gehörs üblicherweise „per Hand“ durchgeführt und die Ergebnisse handschriftlich auf Papier oder direkt in Notensatzprogrammen festgehalten.

**Optical Music Recognition (OMR)** Um aus gedruckten Noten für den Computer verständliche Symbole zu extrahieren, gibt es zahlreiche Ansätze, die zur Gruppe der optischen Notenerkennung gehören. Auch hier gilt, dass die Erkennungsrate in den vergangenen Jahren zwar stetig verbessert wurde, aber dennoch aufwendige händische Nacharbeit benötigt. Eine Übersicht über den aktuellen Stand der OMR-Forschung und ihre Herausforderungen bieten Rebelo u. a. [37].

**Rendering** Liegt die symbolische Kodierung bereits vor, kann sie – sofern das Format unterstützt wird – mit Notensatzprogrammen in Notentext umgewandelt werden. Einfache automatische Ergebnisse sind hier bereits auf qualitativ hohem Niveau. Professionelle Notensetzer können aber auch hier (noch) nicht vollständig ersetzt werden.



**Abbildung 2.4.:** Hauptformen digitaler Musikrepräsentationen und ihre Beziehungen (nach [12]).

Abbildung 2.4 fasst die entsprechenden Beziehung zwischen den Hauptformen digitaler Musikrepräsentationen nochmals zusammen.

In dieser Masterarbeit wird die Brücke zwischen Audio und Noten geschlagen. Da beide Repräsentationen über die symbolische Kodierung miteinander verbunden sind, wird im Rahmen dieser Arbeit vom Vorhandensein einer symbolischen Kodierung ausgegangen. Dies vereinfacht den algorithmischen Aufwand, der zur Verknüpfung notwendig ist, da hauptsächlich die automatische Synthese zur Generierung hörbarer Audiodateien sowie das ebenfalls automatische Rendering zur Darstellung von Notentext auf dem Bildschirm verwendet werden können.

## 2.2. Audio-to-Score Alignment

Die Synchronisation von Notentext und klingender Musik wird in der Fachliteratur als *Score Following* oder auch *Audio-to-Score Alignment* bezeichnet. Bis heute sind eine Vielzahl unterschiedlicher Lösungen für diese Problemstellung entwickelt worden. Eine Zusammenfassung des aktuellen Forschungsstandes und der noch offenen Probleme auf diesem Gebiet bieten Thomas u. a. [45].

Allen modernen Verfahren ist gemein, dass sie zunächst sowohl die aufgenommene Musik als auch die kodierten Noten auf die selbe Repräsentationsebene überführen. Da die Transkription einer Aufnahme in eine Musikkodierung (wie in Abschnitt 2.1.4 beschrieben) sehr aufwendig sein kann, nutzen neuere Methoden bevorzugt die Synthese der kodierten Noten in Audio-Dateien. Die Aufgabe des Audio-to-Score Alignments wird dadurch auf das Problem des Audio-to-Audio Alignments reduziert, also der Synchronisation zweier Audiodateien. Berücksichtigt wird dabei die zeitliche Streckung bzw. Stauchung in Bezug auf die jeweils andere Datei. Zu diesem Zweck wird die symbolische Kodierung der Noten durch Synthese in eine mit Zeitmarkern annotierte Audiodatei transformiert.

### 2.2.1. CQT - Constant Q Transformation

Der folgende Abschnitt setzt die Kenntnis der diskreten *Fourier-Transformation* voraus. Eine praxisnahe Einführung dazu bieten Loy und Chowning [21].

Zur Analyse von Audioaufnahmen kann das Frequenzspektrum betrachtet werden. Für westliche Musik ist die Betrachtung des gesamten Spektrums jedoch nicht zielführend, da nicht direkt von den einzelnen Spektrallinien auf Töne geschlossen werden kann. Zu diesem Zweck wurde die *Constant Q Transform (CQT)* [6] entwickelt. Nachfolgend wird kurz auf die Herleitung der CQT eingegangen. Dieser Abschnitt orientiert sich an [4].

Die CQT ist sehr eng verwandt mit der Fourier-Transformation. Ähnliche wie diese, besteht die CQT aus einer Filterbank, hat jedoch eine logarithmische Verteilung der Mittenfrequenzen:

$$f_k = f_0 \cdot 2^{\frac{k}{b}}, k \in \mathbb{N}$$

$f_0$  entspricht hierbei der Mittenfrequenz des tiefsten betrachteten Tones,  $b$  der Anzahl der Filter pro Oktave. Um die einzelnen Filter lückenlos aneinander anzureihen, wird die Bandbreite des  $k$ -ten Filters festgelegt als

$$\Delta_k^{cqt} = f_{k+1} - f_k = f_k \left( 2^{\frac{1}{b}} - 1 \right)$$

Dies führt zu einem fixen Verhältnis von Mittenfrequenz zu Auflösung, der namensgebenden konstanten Güte  $Q$ :

$$Q = \frac{f_k}{\Delta_k^{cqt}} = \left( 2^{\frac{1}{b}} - 1 \right)^{-1}$$

Mit steigender Frequenz erhöht sich also auch die zeitliche Auflösung, ähnlich der Funktionsweise des menschlichen Gehörs.

Bei geeigneter Wahl der minimalen Mittenfrequenz  $f_0$  sowie  $b$  entsprechen die Mittenfrequenzen musikalischen Noten: Für  $f_0$  als Frequenz der Midinote 0 und  $b = 12$  entspricht  $f_k$  der Mittenfrequenz der MIDI-Note  $k$ . Diese Eigenschaften prädestinieren die CQT zur Anwendung in der Musikanalyse. Während ältere Literatur sich vielfach noch auf die Arbeit von Müller [31] stützt, die die Verwendung elliptischer Filter propagiert, verwenden moderne Frameworks zur Musikanalyse wie z.B. *librosa* [25] häufig nur noch die CQT.

### Berechnung der CQT

Zur Herleitung der CQT für eine Sequenz  $x[n]$  gehen wir von der bekannten Formel des diskreten Fourier-Filters an der Stelle  $z$  aus:

$$\sum_{n=0}^{N-1} x[n] \cdot e^{-2\pi i n z / N}$$

Jede Komponente der CQT wird nun als solch ein Filter berechnet. Hierfür müssen passende Werte für  $z$  sowie die Fensterlänge  $N$  gefunden werden, um die vorher genannten Bedingungen zu erfüllen. Unter gegebener Abtastrate  $f_s$  ist die Bandbreite des Filters bei der Fourier-Transformation jeweils  $\Delta_z^{ft} = f_s / N$  und somit unabhängig von  $z$ . Die gewünschte Bandbreite  $\Delta_k^{cqt} = f_k / Q$  kann somit leicht erreicht werden durch die Wahl der Fensterlänge:

$$N_k = \frac{f_s}{\Delta_k^{cqt}} = Q \frac{f_s}{f_k}$$

Das Verhältnis von Frequenz zu Auflösung entspricht hier  $f_z / \Delta_z^{ft} = z$ . Um eine gleichbleibende Güte der Filter zu gewährleisten, wird festgelegt  $z := Q$ . Zur Vermeidung des Leck-Effekts wird schließlich noch eine Fensterfunktion  $\omega_N[n]$  genutzt.

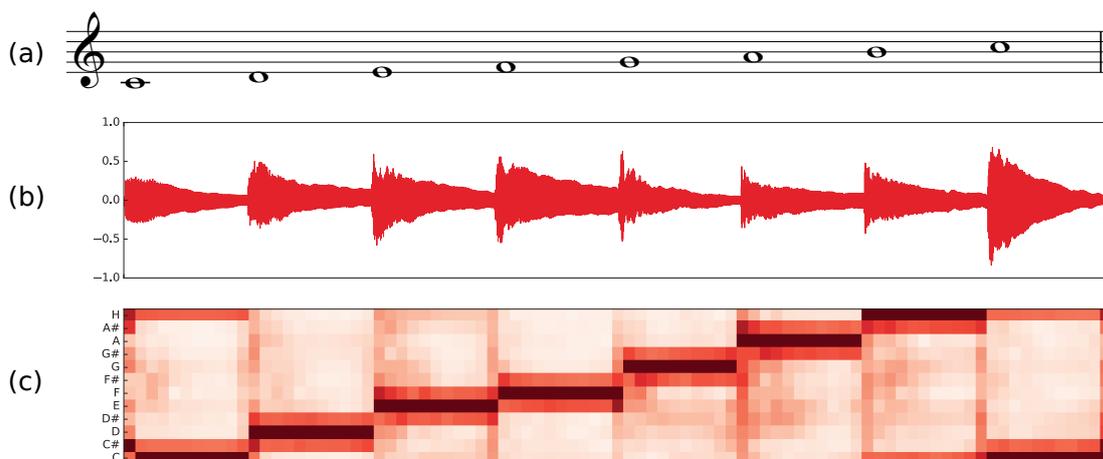
Zusammengefasst ergibt sich für die Berechnung des  $k$ -ten Filters:

$$X^{cq}[k] := \frac{1}{N_k} \sum_{n=0}^{N-1} x[n] \omega_{N_k}[n] \cdot e^{-2\pi i n \cdot Q / N_k} \text{ mit } N_k := \left\lceil Q \frac{f_s}{f_k} \right\rceil$$

#### 2.2.2. Chroma-Vektoren

Bei der weiteren Verarbeitung der durch die CQT gewonnenen Daten ist es für das Alignment zweier Aufnahmen nötig, den Einfluss aufnahmeinheränter Klangcharakteristika auf ein Mindestmaß zu reduzieren, da sonst z.B. unterschiedliche Instrumentalbesetzungen das Alignment erschweren. Hierfür wird zunächst mithilfe der CQT für jede MIDI-Note die Lautstärkeintensität ermittelt. Diese Werte werden dann zu *Chromas* zusammengefasst. Jedes Chroma repräsentiert die Energie einer bestimmten *Tonklasse* – z.B. *C* – im betrachteten Zeitfenster. Eine Tonklasse fasst dabei alle *Cs* ungeachtet ihrer Oktavlage im betrachteten Zeitfenster zusammen. Enharmonische Verwechslungen werden ignoriert, sodass beispielsweise  $D^\flat$  und  $C^\sharp$  dem selben Chroma angehören. Fasst man alle Chromas (12 Halbtöne) zusammen, erhält man für das betrachtete Zeitfenster einen 12-dimensionalen *Chroma-Vektor*:

$$v = (C, C^\sharp, D, D^\sharp, E, F, F^\sharp, G, G^\sharp, A, A^\sharp, H)$$



**Abbildung 2.5.:** Noten einer C-Dur-Skala (a), die Wellenform einer Klavieraufnahme dieser Noten (b) sowie das entsprechende Chromagramm (c).

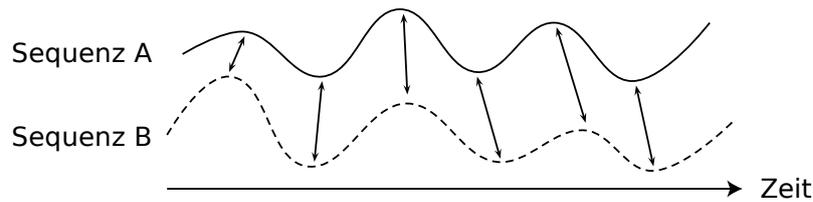
Das zu analysierende Zeitfenster jedes Chroma-Vektors ist zwar frei wählbar, in der Praxis haben sich jedoch Werte zwischen 20 und 50 ms je Block bewährt, vgl. [31]. Abbildung 2.5 veranschaulicht in einem *Chromagramm*, dass Chroma-Vektoren zwar ein gewisses Maß an Rauschen beinhalten, den melodischen und harmonischen Verlauf jedoch trotzdem deutlich abbilden können. Chromagramme dienen der Darstellung einer Folge von Chroma-Vektoren über die Zeit. Die Intensitäten der einzelnen Chromas werden hierbei farbkodiert. Zur Verbesserung des Rauschabstands können auf die so gewonnenen Daten noch Gate-Filter, Gewichtungen und Ähnliches angewendet werden, wie in [31, S. 60 – 64] beschrieben.

Ein weiterer Effekt ist der schwindende Einfluss der Klangeigenschaften eines Instruments: Obgleich jedes Instrument seine eigene charakteristische Obertonstruktur besitzt, relativieren sich diese Unterschiede bei der Reduktion auf Chroma-Vektoren wieder zu großen Teilen. So erhält man man selbst für zwei sehr unterschiedliche Aufnahmen – auch mit abweichender Instrumentalbesetzung – vergleichbare Werte.

Eine detaillierte Beschreibung zu Chroma-Vektoren und ihrer Anwendung findet man bei Hu, Dannenberg und Tzanetakis [18] sowie Müller [31, 32].

### 2.2.3. Dynamic Time Warping

Zum Vergleichen des zeitlichen Verlaufs zweier Aufnahmen, kann das *Dynamic Time Warping* (DTW) verwendet werden. Dieser Algorithmus findet häufig bei Zeitreihen Anwendung, also auf der Zeitachse äquidistant abgetasteten Signalen, die sich in ihrer



**Abbildung 2.6.:** Abbildung zweier Sequenzen aufeinander. (Nach [31, S. 70])

Geschwindigkeit unterscheiden können. Ziel ist das Finden einer optimalen (im Allgemeinen nichtlinearen) Abbildung dieser Sequenzen aufeinander, siehe Abbildung 2.6.

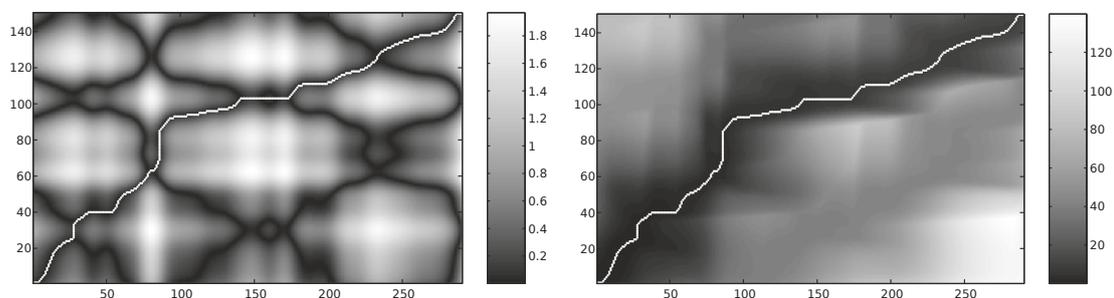
Im Musikkontext werden diese Sequenzen ( $\mathcal{F}$ ) aus Chroma-Vektoren gebildet. Zum Vergleich zieht man die Chroma-Vektoren beider zu vergleichenden Aufnahmen heran und berechnet paarweise die Differenz zwischen diesen. Es wird also eine lokale Kostenfunktion  $c$  aufgestellt:

$$c : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}^{12}$$

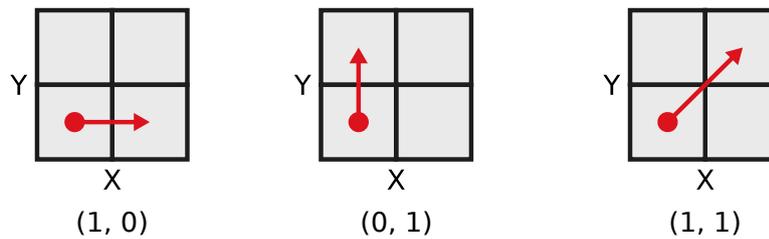
Als Maß wird bei musikalischem Inhalt in der Regel der euklidische Abstand genutzt:

$$c(x, y) = \sqrt{\sum_{k=1}^{12} (x_k - y_k)^2}$$

Andere Maße sind möglich und werden in [31] beschrieben, haben in der Praxis jedoch keine nennenswerten Vorteile und werden daher kaum verwendet. Die so gewonnenen Werte können anschaulich in einer *Kostenmatrix* (Abbildung 2.7, links) dargestellt werden, welche die paarweisen Distanzen farbkodiert.



**Abbildung 2.7.:** Kostenmatrix zweier Sequenzen (**links**) und die entsprechende akkumulierte Kostenmatrix (**rechts**). Der optimale Pfad ist jeweils weiß eingezeichnet. Geringe Kosten werden dunkler dargestellt. (Quelle: [31, S. 74])



**Abbildung 2.8.:** Mögliche Schritte innerhalb der Kostenmatrix.

Ziel ist nun, einen optimalen Pfad durch diese Matrix zu legen. Optimal bedeutet, dass die Summe der mit dem Pfad durchschrittenen paarweisen Kosten minimal ist. Zur Berechnung dieses Pfads wird zunächst festgelegt, welche Schritte innerhalb der Matrix zulässig sind. Da Musik prinzipiell nur „vorwärts“ gespielt wird, kann ausgeschlossen werden, dass beim Vergleichen zweier Sequenzen rückwärtige Schritte notwendig sind. Lässt man zusätzlich keine Sprünge innerhalb der Musik bzw. der Sequenzen zu, bleiben so drei mögliche Schritte, um sich innerhalb der aus den Sequenzen  $X$  und  $Y$  gebildeten Kostenmatrix zu bewegen (siehe Abbildung 2.8):

- $(1, 0)$  falls Sequenz  $Y$  schneller voranschreitet als Sequenz  $X$
- $(0, 1)$  falls Sequenz  $X$  schneller voranschreitet als Sequenz  $Y$
- $(1, 1)$  falls beide Sequenzen an der betrachteten Stelle mit gleicher Geschwindigkeit voranschreiten

Mit diesen Einschränkungen in der Schrittweite lässt sich eine *akkumulierte Kostenmatrix*  $D$  (siehe Abbildung 2.7, rechts) errechnen, die für jeden Eintrag der Matrix die bis dahin aufsummierten minimalen Kosten enthält. Für  $D \in \mathbb{R}^{M \times N}$  gilt demnach:

- $D(n, 1) = \sum_{k=1}^n c(x_k, y_1)$  für  $n \in [1, N]$
- $D(1, m) = \sum_{k=1}^m c(x_1, y_k)$  für  $m \in [1, M]$
- $D(n, m) = \min [D(n - 1, m), D(n, m - 1), D(n - 1, m - 1)] + c(x_n, y_m)$

Aus dieser Matrix wird mittels dynamischer Programmierung der optimale Pfad  $p$  berechnet. Um sicherzustellen, dass die Matrix vollständig durchlaufen wird, beginnt man dabei am Ende beider Sequenzen. Ausgehend von  $p_{\text{Ende}} = (M, N)$  wird gerechnet bis  $p_{\text{Beginn}} = (1, 1)$ :

$$p_{k-1} = \begin{cases} (1, m - 1) & \text{für } n = 1 \\ (n - 1, 1) & \text{für } m = 1 \\ \arg \min [D(n - 1, m), D(n, m - 1), D(n - 1, m - 1)] & \text{sonst} \end{cases}$$

$p$  enthält nun die zusammengehörigen Paare beider Sequenzen. Angenommen, dass aus einer Notenkodierung die Chroma-Sequenz  $X$  berechnet wurde, kann nun für jede Note, deren Index  $n$  innerhalb der Sequenz bekannt ist, der entsprechende Index in der Sequenz  $Y$  (Audio) in  $p$  nachgeschlagen werden und umgekehrt.

## 2.3. Einordnung in diese Arbeit

An dieser Stelle wird kurz rekapituliert, wie die einzelnen Bausteine der betrachteten Themen in diesem Kapitel ineinandergreifen.

Wie eingangs beschrieben, ist das Ziel dieser Arbeit, das häufige Wechseln zwischen gedruckten Noten und dem Audioeditor am Computer zu unterbinden. Hierfür ist zunächst die Darstellung der Noten in digitaler Form auf dem Bildschirm nötig. Als Quelle für diese Darstellung eignet sich das vorgestellte MEI-Format, da es gegenüber MIDI neben den reinen Noten auch Layout-Anweisungen enthalten kann.

Um nun die Navigation in den digitalen Noten und den dazu korrespondierenden Audioaufnahmen zu erleichtern, werden beide Repräsentationsarten synchronisiert. Dafür werden beide zunächst auf eine gemeinsame, vergleichbare Form gebracht. Da bereits Algorithmen zur Synchronisierung zweier Chroma-Sequenzen entwickelt wurden, wird zur Nutzung dieser das Audiomaterial mithilfe der Constant-Q-Transformation in Chroma-Features unterteilt. Für die als MEI vorliegende Musikkodierung wird dies im weiteren Verlauf der Arbeit auf ähnliche Weise durchgeführt.

Durch das Dynamic Time Warping werden die Beziehungen zwischen den Chroma-Sequenzen schließlich berechnet, sodass bekannt ist, welche Stellen im Audiomaterial zu welchen Position im Notentext gehören und vice versa.





# 3

## Produktion klassischer Musik

<b>3.1. Produktionsablauf</b> . . . . .	<b>20</b>
3.1.1. Vorbereitung . . . . .	20
3.1.2. Aufnahme . . . . .	21
3.1.3. Nachbearbeitung . . . . .	23
<b>3.2. Fazit</b> . . . . .	<b>26</b>

---

Dieses Kapitel gibt einen einführenden Überblick über den Entstehungsprozess von Instrumental- und Gesangsaufnahmen mit dem Ziel, die Rolle von Noten in der Produktion klassischer Musik einzuordnen und ihre Relevanz hervorzuheben. Um dabei einen möglichst authentisches und zugleich aktuelles Bild darzustellen, wurden im Vorfeld im Rahmen dreier Klassikproduktionen kontextuelle Interviews mit den verantwortlichen Tonmeistern durchgeführt. Gleichzeitig wurde so das Problem umgangen, dass der Produktionsprozess gerade im Bezug auf den Umgang mit Noten in der Vergleichsliteratur nur sehr rudimentär betrachtet wird.

Die Umgebung, in der eine Tätigkeit ausgeführt wird oder ein Produkt genutzt wird, hat oft einen großen Einfluss auf die Durchführung bzw. Usability. Kontextuelle Interviews versuchen an diesem Punkt anzudocken und den Interviewpartner in seinem tatsächlichen Nutzungskontext zu beobachten und durch gezielte Fragen die Bedürfnisse und Anforderungen des Nutzers unter Realbedingungen zu erheben. Details zum Konzept der kontextuellen Interviews finden sich unter [17].

Im Rahmen dieser Arbeit wurden drei Tonmeister bei Aufnahme und Nachbearbeitung klassischer Werke mit folgenden Besetzungen begleitet:

- Klavier Solo
- Klavier + Gesang
- Streichquartett

Vergleichende Literatur [10, 30] unterteilt Musikproduktionen stets in drei große Phasen: *Vorbereitung*, *Aufnahme* und *Nachbearbeitung*. Alle an den Interviews teilnehmenden Tonmeister strukturierten ihre Produktion auf ebendiese Weise. Jede dieser Phasen hat unterschiedliche Ansprüche an die genutzten Programme, den Notentext sowie die Arbeitsumgebung. Daher wurde jeder Tonmeister in allen drei Phasen begleitet. Die Sitzungen hatten den Fokus auf der Arbeit der Tonmeister mit gedrucktem Notenmaterial während der Aufnahme und im nachfolgenden Schnitt.

### **3.1. Produktionsablauf**

Häufig situations- und projektabhängig angepasst, sind die nachfolgend beleuchteten Schritte einer Klassikproduktion nicht als starres Konstrukt zu verstehen. Wirtschaftliche Betrachtungen werden im Rahmen dieser Arbeit außen vor gelassen. Die Ausführungen basieren auf den durchgeführten Interviews und werden an ausgewiesenen Stellen durch weiterführende Literatur unterstützt.

#### **3.1.1. Vorbereitung**

Ziel der Vorbereitung ist vorrangig die Erstellung eines Produktionsplanes in Übereinstimmung mit dem künstlerischen Leiter bzw. Dirigenten. Aus Sicht des Aufnahmeleiters bzw. Tonmeisters bedeutet dies vor allem die Auseinandersetzung mit dem aufzunehmenden Werk. Die drei Tonmeister studierten dafür die Noten und hörten sich bereits verfügbare Aufnahmen an. Potentiell für die Musizierenden herausfordernde Passagen wurden bei zwei Produktionen identifiziert und ein Aufnahmeplan erstellt mit dem Ziel, die zur Verfügung stehende Aufnahmezeit möglichst optimal auszunutzen sollte. Bereits in dieser Phase wurden erste Anmerkungen in die Noten eingetragen, wie z.B. mögliche Stellen für nachträgliche Schnitte.

Zugunsten einer unkomplizierten Kommunikation mit den Musikern war es von großem Vorteil, dass Musiker und Aufnahmeleiter das gleiche Notenmaterial in Bezug auf das Layout besaßen. Seiten-, Taktzahlen etc. waren identisch, damit beide Parteien explizite Stellen innerhalb der Musik mithilfe von gesprochenen Referenzpunkten wie z.B. „Seite sieben, zweites System, letzter Takt“ auffinden konnten.

Da die Raumakustik entscheidenden Einfluss auf den Gesamtklang hat, ist auch die Wahl des Aufnahmeortes im Vorfeld der Aufnahme von großer Relevanz. Ausgehend vom gewählten Raum haben zwei der Tonmeister einen *Mikrofonierungsplan* erstellt, der die Positionen und weitere klangbestimmende Parameter (z.B. *Richtcharakteristik*) der

verwendeten Mikrofone skizziert hatte. Alle Teilnehmer haben ein *Hauptmikrofonsystem* [34] aufgestellt bzw. gehängt, das genutzt wird, um ein möglichst authentisches Klangbild der Musik im Raum aufzufangen. Es hat daher den größten Einfluss auf die Hörwirkung. Diffus oder zu leise klingende Instrumente wurden zum Teil durch *Stützmikrofone* gezielt hörbarer gestaltet. Weitere Informationen zum Thema Mikrofonierung sind unter [14, 34] erhältlich.

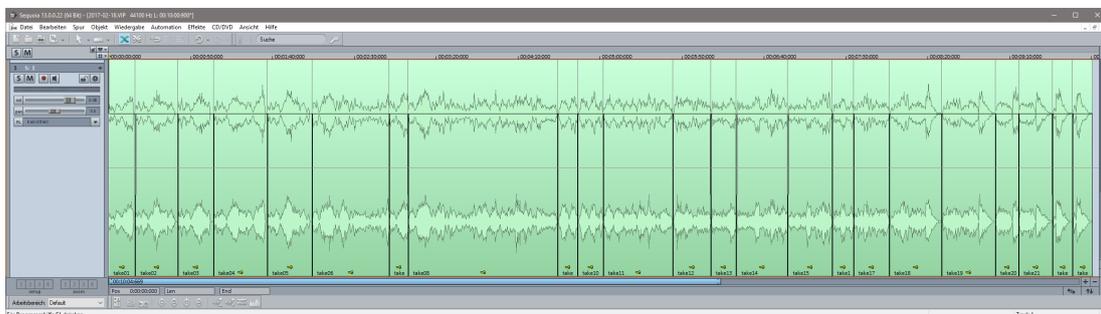
Neben der Hardware wurde schließlich noch die Software entsprechend konfiguriert. Die Mikrofonsignale wurden hierbei jeweils einzeln aufgezeichnet, um für den späteren Mix-Schritt (siehe Abschnitt 3.1.3) die größtmögliche Flexibilität zu erhalten.

### 3.1.2. Aufnahme

Während der Aufnahme kommuniziert der Tonmeister entweder direkt mit allen Musikern oder mit dem musikalischen Leiter bzw. Dirigenten mit dem Ziel, ein sowohl nach technischen als auch musikalisch-ästhetischen Maßstäben befriedigendes Endergebnis zu produzieren. Dies bedeutet in der Regel, dass anspruchsvolle Passagen unter Berücksichtigung der Kritik des Aufnahmeleiters häufiger gespielt und aufgezeichnet werden. Ein Werk muss also nicht zwangsläufig „am Stück“ aufgenommen werden, sondern kann in kürzere Abschnitte (*Takes*) aufgeteilt werden, die schließlich zu einer Gesamtaufnahme zusammengefügt werden.

Die eigentlich Aufnahme geschieht am Computer mithilfe von *Digital Audio Workstations* (DAWs), speziell für den Zweck der Musikproduktion entwickelte Audioeditoren. Aufgenommene Takes werden hier direkt horizontal aneinandergereiht (siehe Abbildung 3.1).

Alle Interview-Teilnehmer teilten die jeweiligen Musikstücke in Abschnitte ein, zwischen denen ein Schnitt vergleichsweise leicht vorgenommen werden konnte. Unbefriedigende Passagen wurden stets mehrfach wiederholt. Es wurde jeweils mithilfe eines



**Abbildung 3.1.:** Mehrere Takes werden in DAWs üblicherweise hintereinander dargestellt.

kleinen Lautsprechers im Aufnahmerraum direkt mit dem Musiker bzw. den Musikern gesprochen.

Bereits während der Aufnahme wurde eine vorläufige Mischung aus den aufgezeichneten Mikrofonspuren erstellt und verwendet. Dies diente zum Einen der Überprüfung, ob die Mikrofonierung die für das Klangideal nötigen Signale bereitstellten. Zum Anderen vermittelte es ein angenehmeres und realistischeres Klangbild für den produzierenden Tonmeister.

### **Eintragungen im Notentext**

Im Laufe der Produktion wurden verschiedene Annotationen direkt im Notentext eingetragen. Damit wurden zwei unterschiedliche Ziele verfolgt: Zunächst dienten Sie als „Gedächtnisstütze“, um nach dem Ende des gerade aufgezeichneten Takes alle wichtigen Anmerkungen kommunizieren zu können. Nach der Aufnahme dienten diese Annotationen der Orientierung bei der Bearbeitung, da nicht nur take-spezifische musikalische Kommentare verzeichnet wurden, sondern auch Schnittpositionen und -anweisungen. Auf diese Art kann bei konsequenter Arbeitsweise parallel zum Aufnahmeprozess bereits ein vorläufiger *Schnittplan* erstellt werden, aus dem ersichtlich ist, welche Takes aneinandergeschnitten werden sollen.

Da die Musiker während des Eintragens weiterspielten, ist es notwendig, dass der Fokus des Aufnahmeleiters nicht zu sehr vom aktiven Zuhören gelenkt werden. Die Eintragungen wurden daher zügig vorgenommen und bestanden deshalb zumeist aus sehr einfachen Elementen. Wie eine Studie zu persönlichen Interpretationsanweisungen in Noten [51] gezeigt hat, orientiert sich die von Musikern und Tonmeistern mit zunehmender Erfahrung entwickelte persönliche „Handschrift“ zum Annotieren von Noten vorwiegend an der Symbolik der modernen Notenschrift. Obwohl in der Praxis daher häufig ähnliche Elemente benutzt werden (siehe Abbildung 3.2), gibt es bislang keinen allgemeingültigen Notationsstandard, der sich in der klassischen Musikproduktion durchgesetzt hat. Dadurch können die Bedeutungen für außenstehende schwierig nachzuvollziehen sein. Auch die Kennzeichnung, ob ein Eintrag als positiv oder negativ ausgezeichnet ist, ist nicht immer eindeutig zu lesen. In den begleiteten Produktionen wurden meistens Symbole wie „+“ und „-“ hierfür verwendet, in einigen Fällen jedoch auch gänzlich auf derartige Einzeichnungen verzichtet.

Einige der aufgenommenen Passagen wurde mehrfach gespielt, wodurch die Fülle an Einzeichnungen zunahm. Dies erschwerte den Überblick und machte eine nachträgliche Zuordnung einzelner Annotationen zu spezifischen Takes zusehends schwieriger.

**Abbildung 3.2.:** Beispielhafte Eintragungen in Noten während einer Aufnahme: Ende Take 4, Beginn Take 5 (a), Note zu laut oder zu kurz gespielt (b), Passage in Take 7 nicht befriedigend (c), Akkorde in Take 9 und 10 nicht gemeinsam erklingen (d).

Zusätzlich zu den Annotationen wurde von zwei Tonmeistern parallel eine *Take-Liste* geführt, die für jeden Take die gespielten Takte verzeichnete.

### 3.1.3. Nachbearbeitung

Ist das Werk vollständig aufgenommen, folgen der Zusammenschnitt der aufgenommenen Passagen zu einer Gesamtaufnahme und die klangliche Optimierung, in der Vergleichsliteratur auch *Post-processing* genannt [10]. Diese Phase nahm bei allen drei Tonmeister die meiste Zeit in Anspruch.

#### Schnitt

Der Schnitt<sup>1</sup> gliederte sich in allen Fällen in zwei Teilaufgaben. Zunächst wurden im gesammelten Audiomaterial die Teile herausgesucht, die für das Endergebnis genutzt

<sup>1</sup>Der Begriff *Schnitt* wird in der englischen Literatur als *Editing* bezeichnet. Entgegen der eigentlich direkten Übersetzung, beinhaltet das *Editing* somit keine weiteren Nachbearbeitungsprozesse.

werden sollten. Im Idealfall unterstützten die entsprechenden Annotationen im Notentext sowie ein bereits während der Aufnahme angefertigter Schnittplan diesen Suchprozess. Ein Tonmeister hatte auf einen derartigen Plan verzichtet und verbrachte deutlich mehr Zeit damit, das aufgenommene Material beim Schnitt akustisch zu sichten. Die selektierten Teile mussten dann „nach Gehör“ derart aneinandergeschitten werden, dass nicht mehr hörbar war, dass es sich um unterschiedliche Takes gehandelt hat.

Häufig erfolgt der Schnitt in mehreren Iterationen. Ist eine erste Version angefertigt, wird sie in Rücksprache mit dem Dirigenten bzw. den Musikern weiter editiert. Sind beide Seiten sowohl aus technischer als auch musikalischer Sicht mit dem Ergebnis zufrieden, ist der Schnitt abgeschlossen. Je nach Interpretation, Verwertung, Anspruch und Erfahrung der Beteiligten kann dieser Prozess langwierig sein und ist daher in der Regel der zeitintensivste Teil der Musikproduktion, nachzulesen unter [41]. Da es sich bei den begleiteten Produktionen jeweils um studentische Produktionen handelte und der wirtschaftliche Faktor nicht berücksichtigt werden musste, kam nur bei einer der drei Produktion seitens des Künstlers der Wunsch auf, Schnitte nachträglich zu korrigieren.

Für den eigentlichen Schnittprozess bediente man sich jeweils einer der beiden folgenden Strategien:

**Ersetzungen im Ganztake** Hat man einen insgesamt zufriedenstellenden *Ganztake*, also eine ununterbrochene Aufnahme des ganzen Werkes, kann man diesen als Grundlage verwenden und ggf. unzulängliche Passagen durch andere Takes ersetzen. Vorteil dieser Methode ist, dass die Musik ihren sogenannten „Bogen“ nicht verliert, die musikalisch-ästhetische Idee also überwiegend unangetastet bleibt.

**Inkrementelles Zusammenschneiden** Steht kein adäquater Ganztake zur Verfügung, muss aus den einzelnen Passagen inkrementell eine vollständige Version geschnitten werden. Hierbei muss vermehrt auf Differenzen zwischen den Takes (unterschiedliches Tempo, Lautstärke etc.) geachtet werden, um die Illusion einer durchgängigen Interpretation, die musikalisch plausibel wirkt, zu schaffen.

Als DAW wurde *MAGIX Sequoia* [22] verwendet. Diese unterstützt den sogenannten *4-Punkt-Schnitt* (siehe Abbildung 3.3): Mit *In-* und *Out-Points* werden sowohl im Quelltake als auch in der Zielspur jeweils die Start- und Endpunkte im Audiomaterial markiert, die entnommen bzw. ersetzt werden sollen. In der Zielspur nachfolgendes Material wird dabei automatisch je nach Länge des Quellausschnitts entsprechend zeitlich



**Abbildung 3.3.:** 4-Punkt-Schnitt in Sequoia<sup>2</sup>. Im linken Fenster wird ein Ausschnitt gewählt (grün) und in die rot markierte Fläche im rechten Fenster eingefügt.

verschoben. Die Tonmeister verwendeten durchweg diese Funktion mit der Begründung, dass diese Arbeitsweise deutlich komfortabler und schneller sei als das manuelle Ausschneiden und Einfügen einzelner Passagen.

Beide Schnitt-Strategien hatten gemein, dass ein ständiger Wechsel zwischen dem annotierten Notenmaterial und der Wellenformdarstellung in der DAW (siehe Abbildung 3.3, rechts) nötig war. Die Entscheidung, welche Takes bzw. welche Teil der Takes aneinandergeschnitten werden sollten, war oft eine Frage des ständigen Hörvergleichs unter Zuhilfenahme der Noten und der darin verzeichneten Take-Nummern. Sauber geführte Auszeichnung der Start- und Endpositionen der einzelnen Takes in den Noten waren daher eine zeitsparende Hilfe.

### Mix & Master

Die letzten Schritte in der Produktion von Instrumentalmusik sind das *Mixing* und *Mastering*. Mixing bezeichnet dabei Klanganpassungen und -korrekturen nach dem Schnitt. Darunter fällt z.B. das Sicherstellen der richtigen Balance in Bezug auf die Lautstärken der einzelnen Instrumente. Dynamik und Position im Stereofeld (engl. *panorama*) werden gezielt manipuliert bzw. festgelegt sowie Filter und Effekte wie beispielsweise Hall werden hinzugefügt, um eine bestimmte Klangästhetik zu erzielen. Anders als bei Aufzeichnungen im Bereich der Unterhaltungsmusik, deren Instrumente und Klänge oftmals keinen akustischen Ursprung haben, wird vorrangig versucht, ein möglichst real klingendes Abbild der Konzert- oder Aufnahmesituation herzustellen. Teilweise wird auch in diesem Schritt wieder das Notenmaterial referenziert, wenn z.B. einzelne Instrumente nur stellenweise hervorgehoben werden sollen und daher die Lautstärken einzelner Mikrofonspuren temporär erhöht werden müssen.

<sup>2</sup>Quelle: <http://www.magix-audio.com/de/sequoia/klassikproduktion/>, abgerufen am 12.02.2017

Das finale Mastering soll sicherstellen, dass der Gesamtklang nicht nur auf den Lautsprechern des Tonmeisters bzw. Produzenten adäquat reproduziert werden kann, sondern auf einer Vielzahl von unterschiedlichen Lautsprechern, Kopfhörern und sonstigen Endgeräten (z.B. Smartphones) überzeugt. Dieser Schritt ist gänzlich unabhängig vom Notentext und betrachtet lediglich klangliche Aspekte, wird also im Rahmen dieser Arbeit nicht weiter adressiert und wurde in den kontextuellen Interviews auch nicht berücksichtigt.

### **3.2. Fazit**

Die Beobachtung des Produktionsprozesses klassischer Musik hat aufgezeigt, dass die Navigation innerhalb der DAWs beim Schnitt recht aufwendig ist. Es ist für den Benutzer der Software optisch nicht unmittelbar ersichtlich, in welchem Teil der Musik er sich gerade befindet. Die Umriss der Wellenform geben zwar für geübte Benutzer gute Hinweise, sind aber ohne ständiges „Hineinhören“ in die Aufnahme keine verlässliche Navigationsbasis. Häufiges Wechseln zwischen dem Computerbildschirm und den annotierten Noten kostet zusätzlich Zeit.

Auf Basis dieser Erkenntnisse scheint es hilfreich, die Navigation in der Schnitt-Software zu überarbeiten. Idealerweise sollte die Visualisierung der Aufnahmen direkten Aufschluss über den musikalischen Inhalt geben.



# 4

## Navigationshilfen in DAWs

<b>4.1. Alternative Audiodarstellungen</b> . . . . .	<b>28</b>
4.1.1. Erweiterte Wellenformen . . . . .	28
4.1.2. Strukturorientierte Ansätze . . . . .	30
4.1.3. Spektrogramme . . . . .	32
<b>4.2. Synchronisation mehrerer Takes</b> . . . . .	<b>33</b>
<b>4.3. Einfluss auf diese Arbeit</b> . . . . .	<b>34</b>

---

Die grundsätzliche Idee, die Navigation in Audiomaterial innerhalb von DAWs durch Visualisierungen und Algorithmen zu unterstützen, wird bereits seit Jahrzehnten wissenschaftlich bearbeitet. 1982 beschrieben Chafe, Mont-Reynaud und Rush [7] die potentiellen Möglichkeiten, die sich dadurch eröffnen. Zu diesem Zeitpunkt befand sich die Musikaufnahme auf digitale Medien noch in den Kinderschuhen<sup>1</sup>. Der Artikel nahm für sich in Anspruch, bis dato „unvorstellbare“ Anweisungen wie beispielsweise „start at the 17th measure“ („Starte ab dem 17. Takt“) vorstellbar oder in Zukunft gar praktisch nutzbar zu machen. Wenngleich solche Aussagen aus heutiger Perspektive zum Schmunzeln verleiten mögen, haben tatsächlich praktikable Konzepte nur zögerlich Eingang in moderne Audio-Software gefunden. Nichtsdestotrotz gibt es bis dato eine Vielzahl an Forschungsprojekten, die sich mit den gleichen Fragestellungen und Lösungsansätzen konfrontiert sehen. Diese Kapitel bietet hierüber einen Überblick.

---

<sup>1</sup>Zur historischen Einordnung: Die CD wurde offiziell 1981 auf der Funkaustellung in Berlin der Öffentlichkeit vorgestellt. Im selben Jahr präsentierten EMT sowie Studer ihre ersten digitalen Bandmaschinen. MIDI wurde ein Jahr später, also im Jahr des Artikels spezifiziert.

## 4.1. Alternative Audiodarstellungen

Reine Wellenformdarstellungen, wie sie in heutigen DAWs üblich sind, ermöglichen zwar eine vergleichsweise intuitive, aber keineswegs optimale Navigation in Audio-material. So sind zum Beispiel wiederkehrende Strukturen nicht zweifelsfrei ersicht-lich. Ebenso wenig lassen Wellenformen Rückschlüsse über die Energieverteilung in einzelnen Frequenzbändern zu. Auf der Suche nach geeigneteren Visualisierungen ha-ben sich seit der flächendeckenden Einführung der digitalen Audiotbearbeitung unter-schiedliche Ansätze herausgebildet.

### 4.1.1. Erweiterte Wellenformen

Eine Möglichkeit, den Informationsgehalt von Wellenformen zu erhöhen, besteht dar-in, die Einhüllende entsprechend der Energie einzelner Frequenzbänder einzufär-ben (siehe Abbildung 4.1). Rice beschreibt in [38] eine Möglichkeit der Übertragung der Daten des Frequenzspektrums auf den bei Computerbildschirmen üblichen RGB-Farbraum. Das Audiomaterial wird hierfür zunächst in kürzere Abschnitte unterteilt und einer Fourieranalyse unterzogen, um ein Energiedichtespektrum zu erhalten. Die-ses Spektrum wird in drei Frequenzbänder unterteilt (Höhen, Mitten und Bässe), de-ren individuelle Gesamtenergien schließlich direkt auf die RGB-Farben Rot, Grün und Blau übersetzt werden. Daraus ergeben sich wiederum drei perzeptive Faktoren, die einen schnellen Überblick ermöglichen:

**Farbton** Eine rötliche Farbe zeigt einen hohen Anteil niederfrequenter, tiefer Töne an. Grün hingegen deutet auf vorwiegend mittlere, blau auf hohe Frequenzen hin.

**Helligkeit** Die resultierende Helligkeit ist ein Maß für das Klangvolumen bzw. die Lautstärke innerhalb des Abschnitts. Helle Passagen erklingen lauter als dunkle.

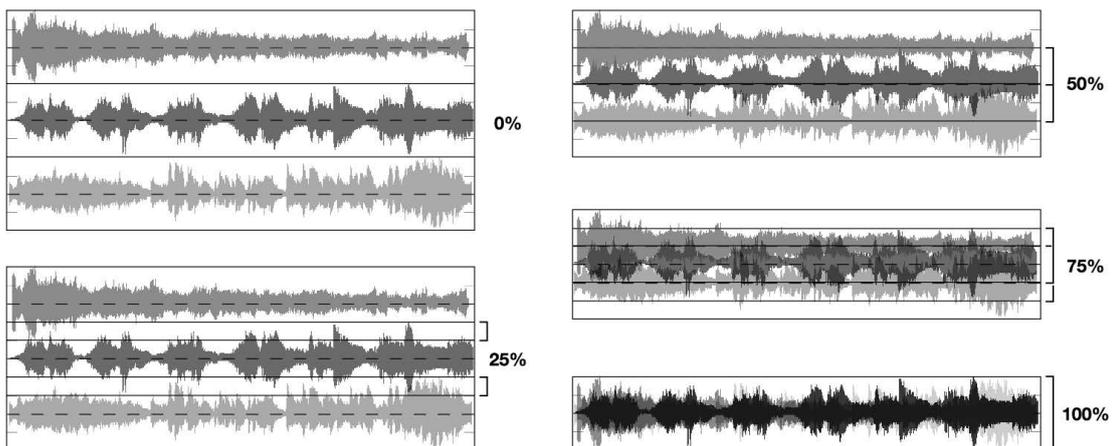
**Sättigung** Intensive Farben bedeuten einen hohen Energieanteil in einem heraus-tretenden Frequenzband. Ist die Musik im betrachteten Zeitfenster ausgewogen über die Frequenzen, ergeben sich für alle drei Farbkomponenten ähnliche Werte und münden so in einer gräulichen Farbe.



**Abbildung 4.1.:** Einfärbung von Soundeffekten. Von rechts nach links: Explosion (**dunkel**), Schrei (**rot**), Sirene (**alternierende Farben**) und weißes Rauschen (**grau**). (Quelle: [38, S. 3])

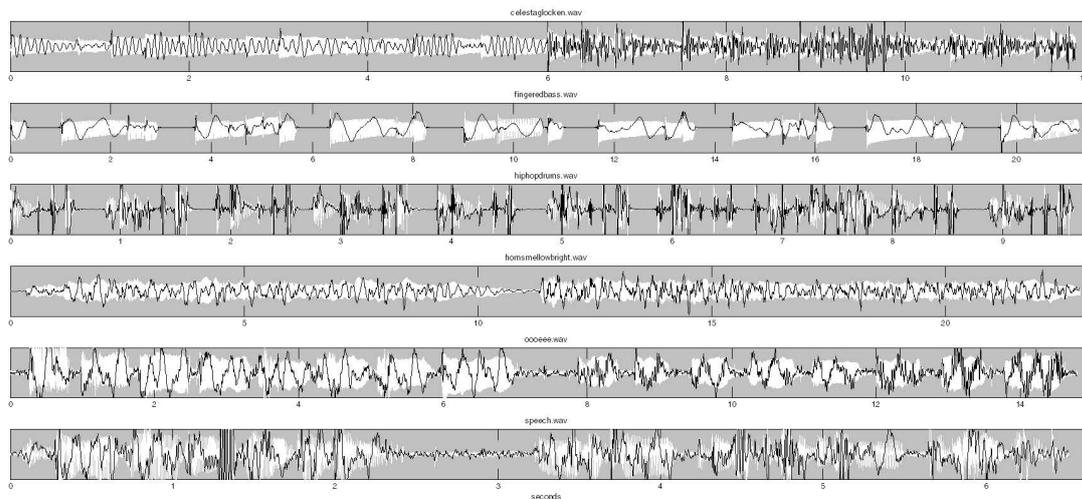
Einige quelloffene Wiedergabeprogramme wie z.B. *Amarok*, *Exaile* und *Clementine* benutzen dieses Verfahren, um eine sogenannte *moodbar* [52] zu generieren mit dem Hauptziel, Strukturelemente wie z.B. Strophe oder Refrain sowie eine allgemeine emotionale Stimmung der Musik ersichtlich zu machen.

Das ursprüngliche Konzept der Wellenform kann durch Überlagerungen mehrerer Tracks übereinander [13] erweitert werden (siehe Abbildung 4.2). Die einzelnen Wellenformen können dabei durch Einfärbungen oder Schattierungen besser voneinander abgegrenzt werden. Sinnvoll ist diese Visualisierung erst beim Arbeiten mit mehreren Audiospuren unterschiedlicher Instrumente, wie es z.B. in der Pop-Musik üblich ist. Ihr Vorteil fußt auf dem deutlich geringen Platzverbrauch auf Computerbildschirmen, sodass eine größere Zahl an Audiospuren gleichzeitig angezeigt werden kann. Da sowohl bei Produktionen klassischer Musik als auch im Unterhaltungsbereich heutzutage mitunter mehrere Dutzend Spuren aufgenommen werden, erleichtert eine kompaktere Darstellung die Navigation.



**Abbildung 4.2.:** Transparente Überlagerungen mehrerer Wellenformen mit unterschiedlichen Abständen. (Quelle: [13, S. 7])

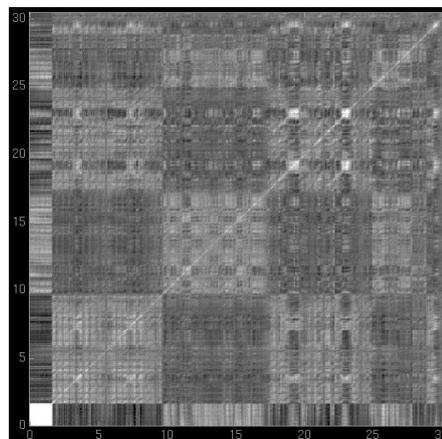
Eine weitere Möglichkeit der Überlagerung ist die Darstellung des Verlaufs einer einzelnen Periode innerhalb der Wellenform [20], ähnlich einem Oszilloskop. Abbildung 4.3 verdeutlicht diese Idee. Sowohl die Schwingungen (schwarz) als auch die Einhüllende (weiß) werden dabei mit einer Potenzfunktion komprimiert, um extreme Ausschläge innerhalb eines betrachtbaren Rahmens zu halten. Vorteil dieser Visualisierung ist die schnell erfassbare Klangstruktur: Tiefe, gleichmäßige Töne werden als weit auseinandergezogene Wellen dargestellt, bei höheren Tönen werden diese Wellen immer weiter zusammengestaucht. Komplexere Klänge erscheinen auch optisch deutlich komplexer.



**Abbildung 4.3.:** Darstellung einzelner Schwingungsperioden innerhalb der komprimierten Wellenformen. (Quelle: [20, S. 6])

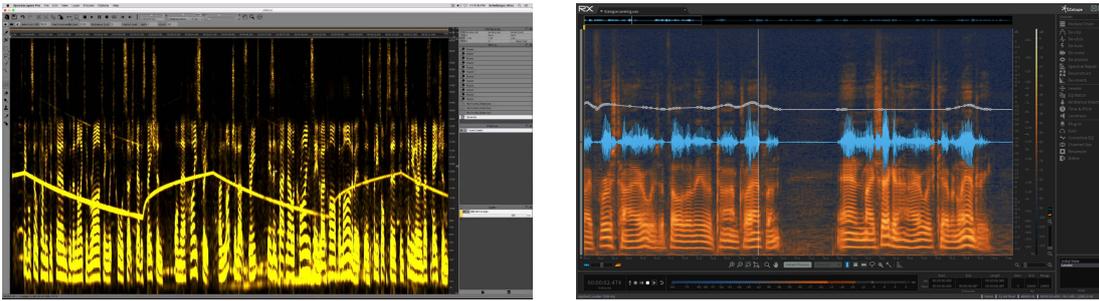
#### 4.1.2. Strukturorientierte Ansätze

Wellenformen sind gut geeignet, um Einblicke in lokale Eigenschaften einer Audio-datei zu erhalten, nicht jedoch für eine strukturelle Übersicht. Ausgehend von den in Kapitel 2 beschriebenen Chroma-Vektoren kann man hierfür eine *Selbstähnlichkeitsmatrix* [11] (siehe Abbildung 4.4) erstellen, die – analog zur Ähnlichkeitsmatrix (siehe Abbildung 2.7) – Distanzen zwischen den Vektoren darstellt. Der Unterschied besteht darin, dass für beide Eingangssequenzen die selbe Chroma-Vektorenssequenz genutzt wird. So entsteht eine an der Hauptdiagonalen symmetrische, quadratische Matrix, auf der sich wiederkehrende Elemente abzeichnen. Durch weitere grafische Filter wie z.B.



**Abbildung 4.4.:** Selbstähnlichkeitsmatrix des Rondos aus dem Hornkonzert Nr. 4, KV 495 von Mozart. (Quelle: [11, S. 5])

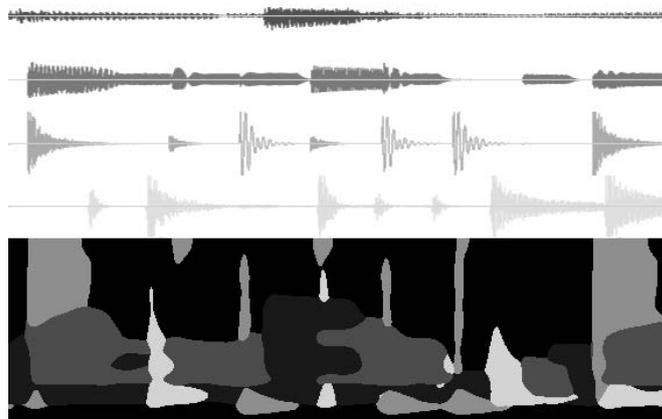




**Abbildung 4.6.:** Spektrogramme in den Programmen *MAGIX SpectraLayers 4 Pro* (links) und *iZotope RX 5 Audio Editor* (rechts).

### 4.1.3. Spektrogramme

Besteht der Wunsch, detaillierte Aussagen über das Frequenzverhalten einer Aufnahme zu erhalten, ist das direkte Darstellen des aus der Fourieranalyse resultierenden Energiedichtespektrums möglich. Programme wie beispielsweise *MAGIX SpectraLayers 4 Pro* [23] und *iZotope RX 5 Audio Editor* [19] (siehe Abbildung 4.6) erlauben neben der Visualisierung sogar gezielte Eingriffe in das Spektrum (vergleichbar mit Bildbearbeitungsprogrammen), sodass z.B. Störgeräusche effektiv entfernt werden können. *RX 5* überlagert das Spektrum zusätzlich optional mit der Wellenform, um die Vorteile beider Darstellungsweisen parallel nutzen zu können. Derartige Darstellung lassen mit etwas Erfahrung Rückschlüsse auf die gespielten Noten zu. Moderne DAWs haben häufig eine entsprechende Visualisierungsfunktion integriert. Einsatzzweck ist dabei hauptsächlich die Restauration (Entfernung von Störgeräuschen und Fehlern) von Aufnahmen.



**Abbildung 4.7.:** Dominante Anteile 4 einzelner Spuren (oben) im summierten Frequenzspektrum (unten). (Quelle: [13, S. 9])

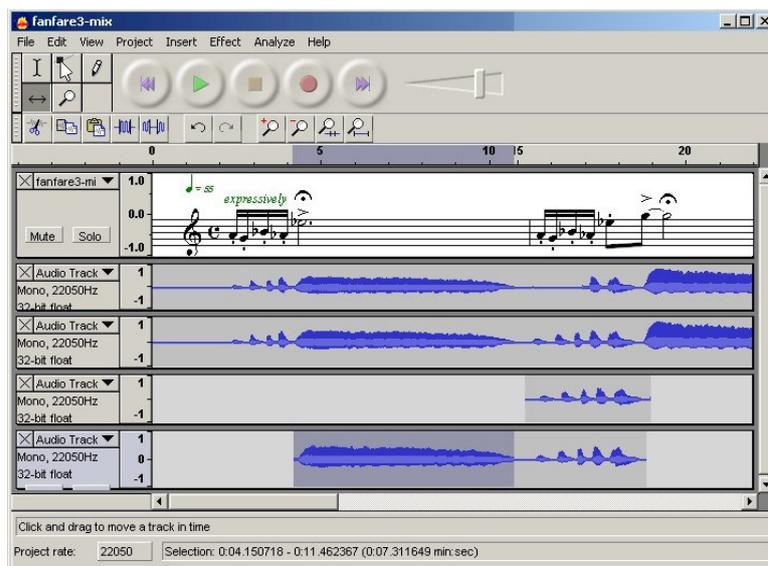
Gohlke u. a. [13] versuchen, die gleichzeitige Darstellung von Wellenform und Spektrogramm weiter zu integrieren. Beim Zusammenmischen (Summieren) mehrerer Tracks werden im resultierenden Frequenzspektrum die energetischen Einflüsse der einzelnen Instrumente farblich oder durch Schattierungen hervorgehoben (siehe Abbildung 4.7).

Auf diese Weise wird schnell ersichtlich, welche Instrumente sich einen Frequenzbereich teilen. In der Pop-Musik versucht man in der Regel, solche Überlappungen zu Gunsten eines transparenten Mixes zu vermeiden.

## 4.2. Synchronisation mehrerer Takes

Die dieser Masterarbeit zugrunde liegende Idee der Synchronisation mehrerer Takes innerhalb einer Aufnahme unter Einbeziehung von Noten haben Dannenberg und Hu bereits 2003 skizziert [9]. Konkrete Implementierungen wurden jedoch nicht diskutiert, sodass es lediglich bei einem Mock-up geblieben ist (siehe Abbildung 4.8). Das Paper geht jedoch von der Nutzung eines DTW-Algorithmus' aus (vgl. Abschnitt 2.2.3).

In [29] wird ein ähnliches Konzept verfolgt, jedoch mit einem neuartigen Synchronisierungsalgorithmus auf Basis von Partikelfiltern [28]. Der genutzte Algorithmus scheint sehr vielversprechend, wurde jedoch aufgrund einer beruflichen Neuorientierung des Autors nicht weiter verfolgt.



**Abbildung 4.8.:** Mock-up einer DAW mit synchronisierten Takes unterhalb einer Notendarstellung. (Quelle: [9, S. 6])



**Abbildung 4.9.:** Multisynchronschnitt in der Software Sequoia<sup>2</sup>.

Eine erste praktische Nutzung der Audiosynchronisierung erfolgte 2009 mit der Markteinführung des *Multisynchronschnitts* [16] in der Audio-Software MAGIX Sequoia 11 [22]. Erstmals war es möglich, nach einer Aufnahme die einzelnen Takes vollautomatisch analysieren und synchronisieren zu lassen, um eine Ansicht zu erhalten, in der Takes mit gleichem musikalischen Inhalt untereinander dargestellt wurden. Problematisch war die Analyse jedoch bei Musik mit wiederholenden Passagen oder häufig wiederkehrenden musikalischen Motiven, da hier oft Fehler bei der korrekten Zuordnung aufgetreten sind. Eine nachträgliche manuelle Korrektur nimmt wiederum zusätzlich Zeit in Anspruch. Da eine entsprechende Notendarstellung nicht integriert war, wurde hierdurch allein der direkte Vergleichsprozess zwischen einer Vielzahl von Takes beschleunigt bzw. vereinfacht, nicht jedoch die globale Navigation innerhalb des aufgenommenen Materials.

Eine entscheidende Neuerung beim Multisynchronschnitt war die Abstraktion mehrspuriger Aufnahmen zu einer einzelnen Spur (siehe Abbildung 4.9). Im rechten Teil des Bildes sieht man die gewöhnliche Anordnung in einer DAW, sodass z.B. im „grünen“ Take alle dazugehörigen Spuren untereinander gerendert werden. Nach dem Analyseprozess werden jedoch all diese Spuren als einzelnes Objekt dargestellt und mit den darunter angezeigten Takes synchronisiert. Dafür ist es unter anderem notwendig, die Wellenform optisch zu strecken bzw. stauchen – Die Zeitachse wird so nichtlinear.

### 4.3. Einfluss auf diese Arbeit

Die Entwicklung eines Programms zur notengestützten Navigation in Audio-Dateien bedient sich im weiteren Verlauf der Arbeit an einigen Elementen der vorgestellten Forschungsarbeiten. Grundlage bildet die Idee des Multisynchronschnitts, bei dem jedem Take optisch ein einzelnes Objekt zugewiesen wird. Dies erleichtert die Übersicht über die Takes erheblich. Erweitert wird dieses Konzept durch die Darstellung von Noten direkt oberhalb der Takes wie in Abbildung 4.8.

<sup>2</sup>Quelle: [www.magix-audio.com/de/sequoia/klassikproduktion](http://www.magix-audio.com/de/sequoia/klassikproduktion), abgerufen am 08.02.2017

Das Verkleinern der Abstände zwischen den Spuren wie in Abbildung 4.2 wird konsequent weitergedacht, sodass auf die Darstellung von Wellenformen gänzlich verzichtet wird und die einzelnen Spuren als schmale Balken angezeigt werden.

Im Rahmen der Optimierungen der Eingangsdaten für das DTW werden während der Entwicklung häufig Selbstähnlichkeitsmatrizen (siehe Abbildung 4.4) zu Rate gezogen. Diese bilden nicht nur die Struktur einer Aufnahme ab, sondern lassen auch Rückschlüsse auf die Qualität der Chroma-Features zu. Die Spektraldarstellung (Abbildung 4.6) wird im weiteren Verlauf bei der Analyse von Audiodateien benutzt, deren Inhalt für die DTW-Berechnung herausfordernd ist.





# 5

## Interaktive Partituren

<b>5.1. Nutzerbefragung</b> . . . . .	<b>38</b>
<b>5.2. Navigationskonzept</b> . . . . .	<b>40</b>
5.2.1. Oberflächengestaltung . . . . .	41
5.2.2. Bedienkonzept . . . . .	42
<b>5.3. Implementierung</b> . . . . .	<b>44</b>
5.3.1. Notendarstellung . . . . .	45
5.3.2. Berechnung der Chroma-Features . . . . .	47
5.3.3. Alignment . . . . .	49
5.3.4. User Interface . . . . .	50
<b>5.4. Evaluation</b> . . . . .	<b>52</b>
5.4.1. Testdesign und Durchführung . . . . .	52
5.4.2. Ergebnisse . . . . .	54

---

Die Entwicklung einer notenbasierten Navigationsoberfläche für klassische Musikproduktionen bildet den Kern dieser Arbeit. Dieses Kapitel beleuchtet den gesamten Entstehungsprozess der Software. Eine Voruntersuchung zur Notwendigkeit und den Potentialen einer solchen Software legt den Grundstein für weitere Überlegungen. Die Konzeptentwicklung für die Oberfläche und Bedienung findet direkte Verwendung in der praktischen Implementierung. Essentielle Bestandteile der Software sowie kritische Abhängigkeiten von externen Programmen werden vertieft betrachtet. Eine Evaluation des Mehrwerts im Vergleich zu heutigen DAWs schließt das Kapitel ab.

## 5.1. Nutzerbefragung

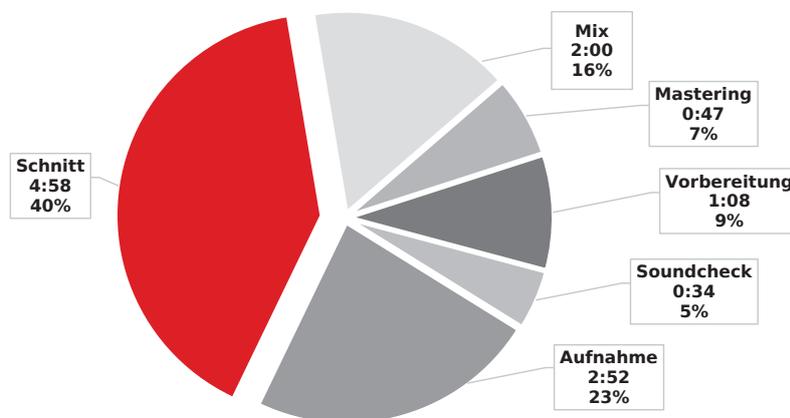
Im Vorfeld der Softwareentwicklung wurde abgeschätzt, ob die Entwicklung einer notenbasierten Navigation tatsächlich einen praktischen Mehrwert bzw. einen schnelleren Workflow bieten könnte. Um grobe Richtwerte für die Dauern einzelner Prozesse innerhalb der Musikproduktion zu erhalten, wurde eine Online-Befragung unter 16 Tonmeisterstudenten der Hochschule für Musik Detmold durchgeführt. Die Teilnehmer hatten im Durchschnitt mindestens 4,2 Jahre Erfahrung in der Produktion klassischer Musik.

Zur Abschätzung des zeitlichen Anteil des Schnitts an der Gesamtdauer einer Produktion, wurde zunächst folgendes Szenario aufgespannt:

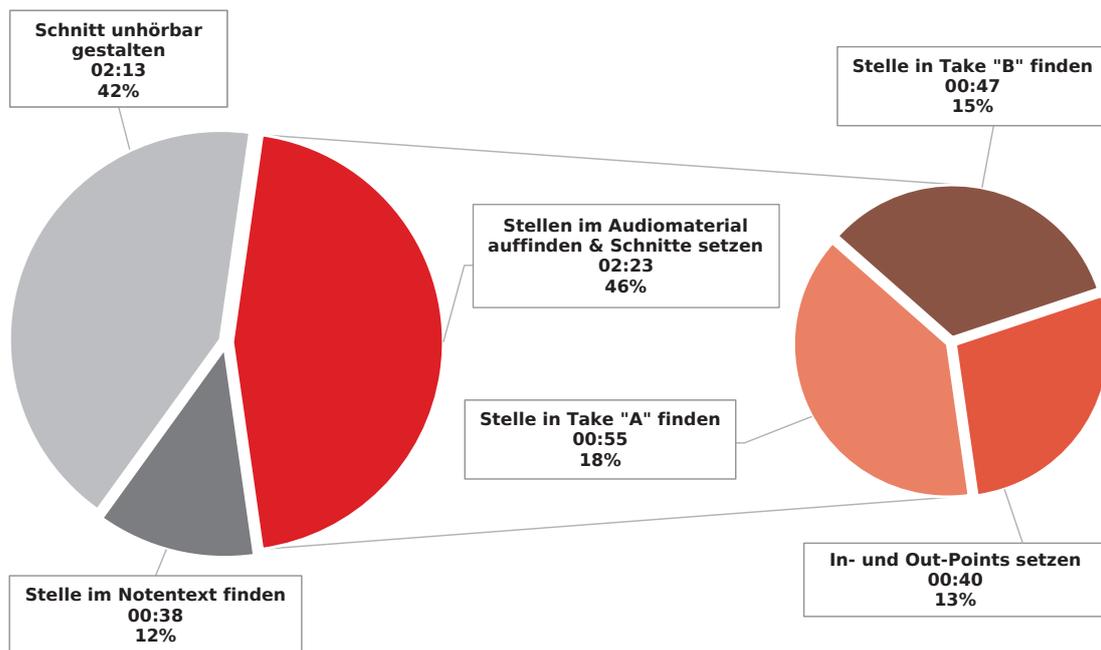
„Ein klassisches Musikstück soll aufgenommen werden, welches fertig geschnitten eine Spieldauer von 10 Minuten hat.“

Gefragt wurde, wie viel Zeit durchschnittlich auf folgende Prozesse entfiel: **Vorbereitung** (Notentext studieren, evtl. Aufnahmen anhören etc.), **Soundcheck**, **Aufnahme**, **Schnitt**, **Mixing** und **Mastering**. Nicht berücksichtigt werden sollte der technischer Aufbau sowie Sonderprozesse wie z.B. Korrekturaufnahmen und Pausen während der Produktion.

Abbildung 5.1 fasst die Ergebnisse zusammen. Die Zeiten für die einzelnen Prozesse wurden zunächst über alle Teilnehmer gemittelt und anschließend in prozentuale Anteile umgerechnet. Demnach entfallen rund 40% der Gesamtzeit (12 Stunden, 19 Minuten) auf den Schnitt. Diese Werte schwanken in der Praxis zwar erheblich in Abhängigkeit von den künstlerischen Fertigkeiten der Musiker, der Komplexität des Stückes



**Abbildung 5.1.:** Gemittelte zeitliche Anteile unterschiedlicher Prozesse innerhalb der klassischen Musikproduktion. (Zeitangaben in H:MM)



**Abbildung 5.2.:** Zeitliche Anteile einzelner Tätigkeiten innerhalb des Schnitts. Farblich hervorgehobene Anteile bergen Optimierungspotential im Rahmen dieser Arbeit. (Zeitangaben in MM:SS)

und der Besetzung, zeigen jedoch deutlich, dass der Schnitt den Großteil der Nachbearbeitung stellt.

In einem zweiten Schritt wurde evaluiert, wie sich der Schnitt zeitlich in weitere Unterroutrinen unterteilt. Folgende Fragestellung wurde zugrunde gelegt:

„Ein Schnitt soll ausgeführt werden. Die bereits erfolgte Take-Auswahl kann dem Schnittplan entnommen werden, Kandidat A und B sollen eine Überblendung erhalten. Wieviel Zeit wird für die einzelnen Schritte benötigt?“

Derartige Schnitte gliedern sich in folgende Tätigkeiten auf: **Stelle im Notentext finden**, entsprechende **Stelle in Take A finden**, parallele **Stelle in Take B finden**, **Schnittpunkte setzen** und schließlich den **Schnitt unhörbar gestalten** (vgl. Abschnitt 3.1.3).

Die Hauptanteile des Schnitts entfallen auf die Suche nach den entsprechenden Stellen im Audiomaterial und den Grobschnitt (46%) sowie die Optimierung des Schnitts (42%), siehe Abbildung 5.2. Ersteres lässt sich weiter aufschlüsseln in die Suche innerhalb des ersten betrachteten Takes und konsekutive Stellen: Ist nach der durchgeführten Suche für den ersten Take bereits bekannt, wie die Umrisse der Wellenform an der betrachteten Stelle aussehen, fallen Suchprozesse für parallele Stellen in anderen Takes etwas leichter und sind somit schneller durchzuführen. Die Suchvorgänge sowie

das Setzen der In- und Out-Points würden durch eine synchronisierte Notendarstellung obsolet werden und könnten den Zeitaufwand für einen Schnitt somit beinahe halbieren.

Ein optionales Kommentarfeld für konstruktive Kritik am dargelegten Schnittprozess wurde in zwei Fällen genutzt, um auf Misstände beim Multisynchrnschnitt in Sequoia hinzuweisen:

- „Oft fehlende Tools für echten Multisynchrnschnitt; Vorbereitung für den Schnitt könnte verbessert werden; Takeauswahl könnte für bessere Vergleichstools für die Takes verbessert werden“
- „Das automatische Analysieren könnte verbessert werden (Sequoia) - überhaupt eingeführt werden (Pyramix), sodass das sortieren der Takes (was eine unglaubliche Zeitersparnis nach sich zieht) schneller passiert (soweit wie möglich automatisiert)“

Diese Kritik zeigt, dass eine Synchronisierung der einzelnen Takes zwar wünschenswert ist, jedoch in der verfügbaren Software bislang keine befriedigenden Ergebnisse liefert.

Im Rahmen der Umfrage wurde untersucht, ob die Verknüpfung von Noten und Musikaufnahmen zeitliches Einsparpotential in der Navigation während des Schnitts birgt. Die Ergebnisse der Umfrage zeigen, dass der ständige Wechsel zwischen gedruckten Noten und der DAW und die damit verbundene Suche im Audiomaterial die zeitintensivsten Schritte sind. Beide Schwachstellen könnten durch eine alternative Benutzeroberfläche mit automatischer Synchronisierung eliminiert werden.

## 5.2. Navigationskonzept

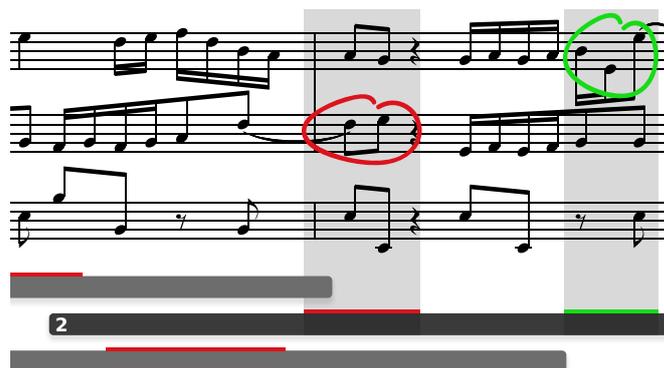
Ausgehend von der in Abschnitt 4.2 skizzierten Idee der Abstraktion aller Mikrofonspuren eines Takes zu einem einzigen Objekt (Multisynchrnschnitt), wurde ein Visualisierungs- und Bedienungskonzept entwickelt. Im Fokus stand die Integration von Noten, die zum aufgenommenen Audiomaterial synchronisiert sein sollen. Das im Folgenden dargestellte Konzept wurde vom Autor dieser Arbeit bereits mehrfach international publiziert [5, 46, 47] und entsprechend unter Einbeziehung des Fachpublikums diskutiert. Teile des Konzepts (integrierte Schnittfunktionalität, Annotationsmanagement) sprengen den Rahmen einer Masterarbeit und werden daher nicht weiter ausgeführt.



**Abbildung 5.3.:** Konzeptentwurf der Benutzeroberfläche. Dunkle Balken markieren dabei einzelne Takes, die unterhalb der Noten dargestellt werden.

### 5.2.1. Oberflächengestaltung

Die Software-Arbeitsumgebungen während der Musikproduktion sind das Ergebnis eines über mehrere Jahrzehnte gewachsenen Bedienkonzepts. Grundsätzliche Änderungen setzen sich nur schwerlich durch, da die kognitive Hürde bei Inkonsistenzen zu etablierte Bedienungen vergleichsweise hoch ist, wie auch entsprechende Studien gezeigt haben [3, 35]. Möchte man neue Ideen in diese Arbeitsumgebung einfließen lassen, ist es daher ratsam, sich an den etablierten Elementen zu orientieren. Ein solches wichtiges Element ist die (theoretisch) endlose Zeitachse in DAWs. Auf ihr wird die aufgenommene Musik chronologisch geordnet dargestellt. Ähnliches ist auch bei Noten möglich, sofern man alle Systemumbrüche entfernt und das gesamte Werk bzw. Stück in einem einzigen, langen System zusammenfasst. Von dieser Grundidee ausgehend lassen sich nun inhaltlich parallele Klangereignisse – also Takes, die den gleichen musikalischen Inhalt haben – untereinander darstellen, in etwa wie unterschiedliche Mikrofonspuren eines konkreten Takes in der DAW. Um die Übersicht zu wahren, können mehrspurige Aufnahmen, wie beim Multisynchronschnitt, optisch auf eine einzige Spur reduziert werden. Synchronisiert man nun das Audiomaterial dieser Takes mit dem Notentext, lassen sich die Positionen der Takes in Bezug auf den Notentext ermitteln. Im User-Interface-Konzept werden die Takes als Balken an den korrespondierenden Positionen unterhalb der Noten dargestellt, da der Mehrwert einer Wellenformdarstellung nicht mehr gegeben ist, siehe Abbildung 5.3. Um eine spätere Zuordnung zu eventuellen Take-Listen zu ermöglichen, werden die Balken jeweils zusätzlich mit einer Take-Nummer versehen.



**Abbildung 5.4.:** Annotationen werden direkt im digitalen Notentext angezeigt und ihre Zuordnung zum jeweiligen Take hervorgehoben. Farbbalken signalisieren das Vorhandensein auch ausgeblendeter Einzeichnungen.

Ein Nebeneffekt dieser Darstellungsweise ist der Verlust der Linearität der horizontalen Achse: Dadurch, dass die einzelnen Takes in unterschiedlichen Tempi gespielt sein können, jedoch dank der Abhängigkeit von den Noten mit der gleichen Balkenlänge gerendert werden, kann keine genaue Aussage mehr über die tatsächliche Spieldauer eines Takes gemacht werden.

Während einer Produktion werden vom Tonmeister fortwährend Annotationen in den Notentext eingetragen. Werden lediglich gedruckte Noten verwendet, füllt sich der Notentext im Laufe der Aufnahme zunehmend mit derartigen Kommentaren. Liegen diese Annotationen digital vor, können sie ebenfalls auf dem Bildschirm dargestellt werden. Dies eröffnet die Möglichkeit, gezielt Eintragungen ein- und auszublenden und die Zuordnung zu bestimmten Takes hervorzuheben, siehe Abbildung 5.4. Kleine Farbmarkierungen direkt oberhalb der Take-Balken signalisieren, ob an der entsprechenden Stelle Anmerkungen vorliegen, die bei Bedarf nach oben „ausgeklappt“ werden können.

Häufig wird bei Eintragungen differenziert, ob sie als positive oder negative Kritik zu verstehen sind. Während der laufenden Produktion verwenden viele Tonmeister intuitiv zusätzliche Symbole wie z.B. „+“ und „-“. Durch die Farbkodierung von Annotationen (z.B. grün für „positiv“ bzw. rot für „negativ“) könnte man diese Differenzierung prägnanter gestalten und auf zusätzliche Symbole verzichten.

### 5.2.2. Bedienkonzept

Das vorgestellte Konzept setzt voraus, dass Annotationen digital erfasst werden können. Technologien wie *Anoto Paper* [1] würden es theoretisch erlauben, mit einem Stift auf echtem Notenpapier zu schreiben und die Einzeichnungen in einem zweiten Schritt

zu digitalisieren. Allerdings ist die folgende optische Notenerkennung auf dem aktuellen Stand der Forschung noch bei Weitem zu fehleranfällig, um zuverlässig genutzt werden zu können. Die Verknüpfung von Eintragungen und Positionen im Audio wäre damit nicht jederzeit gegeben. Zudem entsteht das in Abschnitt 3 beschriebene Problem der übervollen Seiten und der damit einhergehenden schlechteren Lesbarkeit bei steigender Anzahl der Annotationen. Der Artikel „Optical music recognition: state-of-the-art and open issues“ [37] widmet sich den Problemen und Grenzen moderner optischer Notenerkennungsverfahren.

Eine Alternative stellen aktuelle Notebooks und Tablets mit Stifteingabe dar. Produkte wie z.B. das *Surface Book* von Microsoft [26] sowie einige Smartphones und Tablets des Herstellers Samsung erlauben die gleichzeitige Bedienung der Geräte mit Touchscreen und entsprechenden Digitizer-Stiften. Somit könnte während der Aufnahme direkt auf dem Display geschrieben werden. Der Stift des Surface Books (siehe Abbildung 5.5) besitzt eine Taste, die während des Schreibens mit dem Daumen gedrückt werden kann. Auf diese Weise ist ein schnelles Umschalten zwischen zwei Modi denkbar, sodass einfach und zügig zwischen positiven und negativen Eintragungen unterschieden werden könnte.

Die Bedienung der Oberfläche nutzt zusätzlich zum Stift auch die Mauseingabe bzw. die Touchscreen-Funktionalität des Surface Books: Die Navigation entlang der Zeitachse (scrollen nach links bzw. rechts) kann durch „ziehen“ des Notentextes mit der Maus erfolgen oder aber durch „wischen“ mit dem Finger über den Bildschirm. Ist die gesuchte Stelle im Notentext gefunden, können die darunter platzierten Takes abgespielt werden, indem mit der Maus oder dem Stift auf den gewünschten Take geklickt wird. Solange die Maustaste gedrückt ist oder der Stift nicht abgesetzt wird, wird der gewählte Take ab der Cursor-Position abgespielt. Eine vertikale Linie zeigt die aktuelle Position im Take und in den Noten an und wandert während des Abspielens entsprechend nach rechts mit. Als Orientierungshilfe wird sowohl unter dem Notentext als auch unter den Takes ein Scrollbalken angezeigt, der die momentane Position widerspiegelt.

Ein Problem des Multisynchronschnitts ist die unzuverlässige Synchronisation von Audioaufnahmen bei sich wiederholenden Passagen. Ist der aufgenommene Take nicht



**Abbildung 5.5.:** Der Stift des Surface Books von Microsoft. Gut sichtbar ist die langgezogene Seitentaste.

eindeutig einem Abschnitt im Notentext zuordenbar, versagt der Algorithmus der Sequoia DAW oftmals. Um diese Mehrdeutigkeiten aufzulösen bzw. zu verhindern, könnten bei jedem Take während der Aufnahme Beginn und Ende der aktuell gespielten Passage in den digitalen Noten grob markiert werden. Derartige Einzeichnungen hätten als Annotationen für den Tonmeister keinen weiteren Mehrwert und müssten daher nicht sichtbar sein.

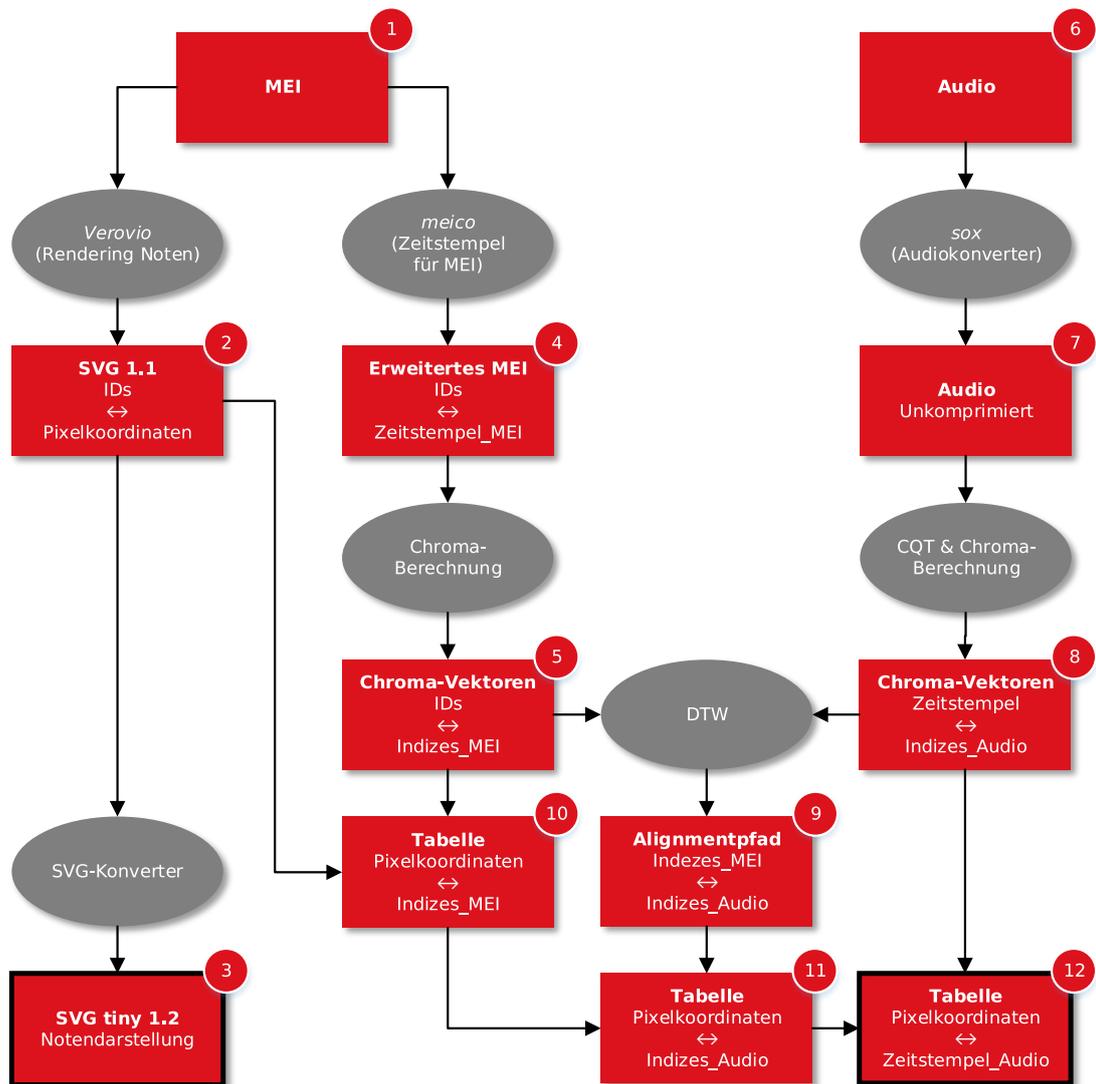
### 5.3. Implementierung

Die praktische Implementierung des Prototypen ist in der Programmiersprache *Python* durchgeführt worden. Grund hierfür ist, dass die wissenschaftliche Gemeinschaft, die sich mit den Problemen des *Music Information Retrieval* (MIR) beschäftigt, in jüngster Zeit eine Vielzahl von Werkzeugen und Programmbibliotheken für gängige Aufgabenstellungen aus dem Fachgebiet in dieser Sprache entwickelt hat. Zudem ist Python dank umfangreicher und optimierter Mathematik-Bibliotheken (*numpy*, *scipy*, *matplotlib* etc.) universell für mathematisch-wissenschaftliche Berechnungen nutzbar.

Nicht Teil der Implementierung ist die Funktionalität der Audioaufnahme. Dieser Part kann unter realen Arbeitsbedingungen aufgrund unterschiedlicher Treiberarchitekturen und Audio-Interfaces sehr komplex werden, ohne dabei den Erkenntnisgewinn in dieser Arbeit nennenswert zu erhöhen. Im weiteren Verlauf wird also davon ausgegangen, dass etwaiges Audiomaterial bereits vorliegt. Dies gilt folglich auch für die während der Produktion gesetzten Marker zur Markierung des Beginns und des Endes eines Takes im Notentext.

Der Aufbau der Software ist modular gestaltet. Nachfolgend werden mehrere externe Tools und Bibliotheken verwendet, um den digitalen Notentext mit Audioaufnahmen zu synchronisieren. Reduziert man dieses Alignment auf rein quantitative Werte, besteht die Aufgabe letztlich darin, Pixelkoordinaten einer Grafik (angezeigte Noten) mit bestimmten Zeitpositionen innerhalb von Audio-Dateien zu verknüpfen. Abbildung 5.6 fasst die dafür erforderlichen Schritte zusammen und bietet einen einleitenden Überblick.

Da im Folgenden immer wieder auf Abbildung 5.6 Bezug genommen wird, werden die einzelnen Zwischenstände im Text mit Zahlen wie z.B. „**3**“ markiert. Diese Zahlen finden sich auch in der Abbildung wieder. Um ein ständiges Wechseln zwischen der Abbildung und dem Fließtext zu verhindern, ist diese Übersicht nochmals im Anhang zum Ausklappen abgedruckt.



**Abbildung 5.6.:** Zusammenspiel der einzelnen Komponenten zum Synchronisieren von Noten und Audio. In Rechtecken mit „↔“ markierte Kommentare geben an, dass zwischen diesen Daten eine Verknüpfung besteht.

### 5.3.1. Notendarstellung

Zunächst stehen die Noten als MEI-kodierte Datei zur Verfügung (1). Der erste Schritt besteht nun in der Generierung der Notendarstellung. Hierfür wird das Programm *Verovio* [36] genutzt. Dabei handelt es sich um eine Open Source Library, die aus MEI-Dateien Vektorgrafiken im SVG-Format erstellt. Wichtig hierbei zu erwähnen ist, dass

```

1 <rest xml:id="m-45" dur="8" dur.ges="128p" />
2 <beam xml:id="m-48" >
3   <note xml:id="m-47" dur="8" dur.ges="128p" oct="4" pname="g" stem.dir="up" />
4   <note xml:id="m-50" dur="8" dur.ges="128p" oct="5" pname="e" stem.dir="up" />
5 </beam>

```

```

1 <g class="rest" id="m-45" >
2   <use xlink:href="#E4E6" x="5838" y="1215" height="720px" width="720px" />
3 </g>
4 <g class="beam" id="m-48" >
5   <path d="M6504 1553 L6504 561" style="stroke-width: 20;" />
6   <path d="M6954 1103 L6954 515" style="stroke-width: 20;" />
7   <g class="note" id="m-47" >
8     <use xlink:href="#E0A4" x="6288" y="1575" height="720px" width="720px" />
9   </g>
10  <g class="note" id="m-50" >
11    <use xlink:href="#E0A4" x="6738" y="1125" height="720px" width="720px" />
12  </g>
13 </g>

```

**Code 1:** Ausschnitt einer MEI-Datei (**oben**) und die daraus von Verovio generierte SVG-Ausgabe (**unten**). IDs der einzelnen XML-Elemente werden übernommen.

alle IDs, die in MEI für Noten, Pausen etc. genutzt werden, ebenfalls in die generierte SVG-Datei übernommen werden. Dies ermöglicht eine Zuordnung jedes gerenderten Objekts zum entsprechenden Element in der MEI-Datei und umgekehrt (2) (siehe Code 1).

Obwohl Verovio sich selbst als „portable“ – also plattformübergreifend – bewirbt, traten im Laufe der Entwicklung zahlreiche Probleme unter Windows auf. Es waren für die Betriebssystemkompatibilität zum Teil umfangreiche Änderungen, Ergänzungen sowie Fehlerkorrekturen am Programmcode notwendig<sup>1</sup>, die ganz im Sinne der Open Source Community in das öffentliche Repository des Projekts zurückgeflossen sind.

Zur tatsächlichen Anzeige der durch Verovio erstellten Vektorgrafiken wird das Framework *Qt* verwendet. *Qt* ist eine plattformübergreifende Bibliothek zur Entwicklung von grafischen Oberflächen für Computer und mobile Systeme. Es verfügt über eine (von Dritten) entwickelte Python-Anbindung und umfangreiche Funktionen zur Darstellung von Pixel- und Vektorgrafiken. Im Zusammenspiel mit *Qt* ergaben sich hinsichtlich der Notendarstellung allerdings weitere Schwierigkeiten. Obwohl *Qt* offiziell

<sup>1</sup>Änderungen, die im Rahmen dieser Arbeit getätigt wurden, sind unter <https://github.com/rism-ch/verovio/pull/282> sowie <https://github.com/rism-ch/verovio/pull/301> nachzulesen (besucht am 20.02.2017).

das Rendering von SVG-Daten unterstützt, beschränkt sich die Formatunterstützung lediglich auf eine Teilmenge (genannt *SVG tiny*) der gesamten Formatspezifikation. Verovio hingegen nutzt durchgängig Befehle, die in dieser Teilmenge nicht enthalten sind. Zunächst wurde versucht, die SVG-Dateien durch externe Kommandozeilenprogramme in pixelbasierte Grafiken umzuwandeln und diese darzustellen. Hierbei traten jedoch Formatierungsfehler auf, sodass die Positionen der einzelnen SVG-Elemente im Bild nicht mehr mit ihren ursprünglichen Koordinaten übereinstimmten. Daher wurde schließlich ein Hilfsprogramm entwickelt<sup>2</sup>, das die Verovio-Ausgabe entsprechend in SVG tiny konvertiert (3).

Durch die Konvertierung mit Verovio stehen schließlich zu jeder ID der ursprünglichen MEI-Datei die Pixelkoordinaten zur Verfügung. So kann für jede Note, Pause etc. eindeutig die Position im Notenbild bestimmt werden, aber auch umgekehrt von Pixelkoordinaten auf ein bestimmtes Notenelement geschlossen werden (2).

### 5.3.2. Berechnung der Chroma-Features

Das Alignment des Notentextes mit den Audioaufnahmen soll mithilfe des in Abschnitt 2.2.3 eingeführten DTW-Verfahrens erfolgen. Dafür ist es jedoch notwendig, dass sowohl für die Noten als auch die Aufnahmen Chroma-Features berechnet werden. Für das MEI-Format ergibt sich an dieser Stelle das Problem, dass MEI keine Interpretation definiert und daher keine absoluten Zeiten für die einzelnen Noten verfügbar sind. Abhilfe schafft an dieser Stelle das Programm *meico* [2]. Der Hauptzweck von *meico* ist die Generierung abspielbarer MIDI-Dateien aus MEI-Kodierungen. Als „Nebenprodukt“ kann hierbei jedoch auch eine MEI-Datei ausgegeben werden, die den einzelnen Noten und Pausen der Originalkodierung die für MIDI kalkulierten Zeitsempel anfügt. Wie auch Verovio, achtet *meico* dabei auf eine persistente Nutzung der originalen IDs aller Elemente. So ist schließlich der Bezug zwischen Noten und ihrem jeweiligen Spielzeitpunkt hergestellt (4).

Zur Berechnung der Chroma-Features der MEI-Datei wird der in Code 2 dargestellte Pseudo-Code auf der durch *meico* erweiterten MEI-Datei verwendet. Im Code wird die zeitliche Auflösung der Chroma-Matrix so gewählt, dass die kürzeste Note einem einzelnen Chroma-Feature zugeordnet werden kann. Auf diese Weise wird sichergestellt, dass keine Note durch Rundungen „verschluckt“ wird. Nun ist jeder MEI-ID ein bestimmter Index innerhalb der Chroma-Matrix zugewiesen und vice versa (5).

Für die Audioaufnahmen (6) werden die Chroma-Features mithilfe der Constant-Q-Transformation (siehe Abschnitt 2.2.1) gebildet. Hierbei stellt sich die Frage, wie groß

<sup>2</sup>Hier veröffentlicht: <https://gist.github.com/sonovice/> (besucht am 21.02.2017).

```
1 mei = "Meico-MEI-Datei"
2 d_min = "Dauer kürzeste Note oder Pause"
3 d_gesamt = "Gesamtspieldauer der Noten"
4
5 # Berechne Gesamtzahl der Chroma-Features
6 # und initialisiere Chroma-Matrix
7 n_chromas = d_gesamt / d_min
8 chroma_matrix = matrix(12, n_chromas)
9
10 # Durchlaufe alle Noten in der MEI-Datei
11 # und berechne Chroma-Features
12 for note in mei:
13     # Berechne Indizes der Noten in der Chroma-Matrix
14     start_index = abrunden(note.datum / d_min)
15     stop_index = aufrunden((note.datum + note.dauer) / d_min)
16
17     tonklasse = note.tonhöhe % 12
18
19     # Trage Chroma-Features in Matrix ein
20     for chroma_feature in range(start_index, stop_index):
21         chroma_matrix[tonklasse, chroma_feature] += 1
```

**Code 2:** Pseudo-Code zur Berechnung einer Chroma-Matrix aus einer MEI-Datei mit durch meico generierten Zeitstempeln.

das Fenster für die Berechnung der einzelnen Chroma-Features gewählt werden sollte. In der einschlägigen Literatur zu diesem Thema [31] werden je Feature meist zwischen 20 und 50 ms benutzt. Ein solcher Richtwert ist auch sinnvoll, sofern man annehmen kann, dass sich die Längen der beiden zu vergleichenden Aufnahmen ähneln. In einem solchen Fall wäre die berechnete Kostenmatrix annähernd quadratisch. Dies hat zur Konsequenz, dass die Schrittfolge (1, 1) beim Durchschreiten der Matrix prozentual häufiger genutzt werden kann, als es bei nicht-quadratischen Dimensionen möglich wäre. Durch derartige diagonale Schritte gelangt man insgesamt schneller an die gegenüberliegende „Ecke“ einer Matrix. Die Anzahl der Paarelemente des optimalen Pfads wird also durch geschicktes Festlegen der zeitlichen Auflösung der Chroma-Sequenzen minimiert. Für die Fenstergröße im vorliegenden Fall bedeutet dies, dass ein Audio-Take nach Möglichkeit in genau so viele Chroma-Vektoren unterteilt werden sollte, wie in der „MEI-Chroma-Sequenz“ **5** für den äquivalenten Ausschnitt genutzt werden. Zur Bestimmung dieser Ausschnitte werden die eingangs erwähnten Start- und Endmarker der Audio-Takes verwendet. Diese beinhalten den ersten bzw. letzten Index des aufgenommenen Ausschnitts innerhalb der MEI-Chroma-Sequenz. Im Rahmen der prototypischen Entwicklung wurden diese Marker manuell ermittelt. In einer marktreifen Version der Software würden während der Aufnahme entsprechende Markierungen im Notentext vorgenommen werden (vgl. Abschnitt 5.2.2). Dank der

zur Verfügung stehenden Zuordnung von Pixelkoordinaten zu MEI-Chroma-Indizes könnten die Markerwerte leicht bestimmt werden.

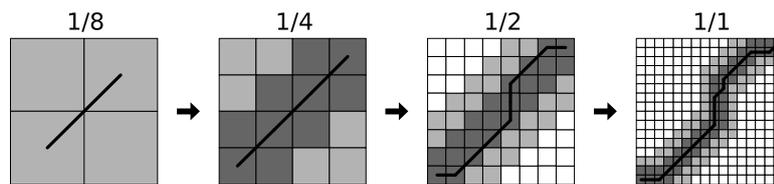
Vor der Berechnung der Chroma-Vektoren muss ggf. das Audiomaterial dekodiert werden, da Audiodateien in verschiedenen Formaten (z.B. *MP3*, *FLAC*, *WAVE*) mit unterschiedlichen Abtast- und Bit-Raten vorliegen können. Daher wird bei der Implementierung der quelloffene Konverter *sox* [8] verwendet. Beim Laden der Audiodateien wird durch ihn automatisch jegliches Material dekodiert, im Bedarfsfall auf die bei CDs übliche Rate von 44 100 Hz neu abgetastet und als unkomprimierte *WAVE*-Dateien temporär gespeichert (7). Die tatsächliche Berechnung der Chroma-Vektoren kann mit der Python-Library *librosa* [25] durchgeführt werden. *Librosa* beinhaltet einen hocheffizienten Algorithmus zur Berechnung der CQT und liefert im Ergebnis eine Chroma-Sequenz, deren Datenstruktur problemlos mit der MEI-Chroma-Sequenz vergleichbar ist (8).

### 5.3.3. Alignment

Da nun sowohl für die Audioaufnahmen, als auch für die MEI-Kodierung Chroma-Sequenzen zur Verfügung stehen, kann der in Abschnitt 2.2.3 beschriebene DTW-Algorithmus angewendet werden. Die resultierenden Pfade bilden jeweils die Indizes von Audio-Chroma-Sequenzen auf die entsprechenden Ausschnitte der MEI-Chroma-Sequenz ab (9). An dieser Stelle wird die Kenntnis der Pixelkoordinaten der einzelnen MEI-IDs (2) genutzt und so die Zuordnung von Koordinaten zu MEI-Chroma-Indizes (10) gebildet. Mithilfe des gerade berechneten Alignmentpfads lässt sich nun von Pixelkoordinaten der Notendarstellung auf Indizes innerhalb der Audio-Chroma-Sequenzen schließen (11) und damit direkt von den Koordinaten auf die zeitliche Position der Audio-Dateien (12).

Für Chroma-Sequenzen der Länge  $M$  und  $N$  beträgt die Komplexität des DTW-Algorithmus'  $\mathcal{O}(MN)$ , da sämtliche Kombinationen der beiden Eingangssequenzen miteinander verrechnet werden müssen. Aufgrund der speziellen Wahl der Fenstergröße bei der Berechnung der Audio-Chroma-Features, lässt sich dies zu  $\mathcal{O}(N^2)$  vereinfachen. In der Praxis bedeutet dies, dass der Berechnungsaufwand mit wachsender Take-Länge quadratisch steigt. Für Takes mit mehreren Minuten Länge, wie sie in der Klassik durchaus vorkommen können, nimmt dieser Prozess auch mit schnellen Computern zu viel Zeit in Anspruch.

Der bis dahin genutzte DTW-Algorithmus ist in der Bearbeitungszeit dieser Arbeit schrittweise in *librosa* von der Open Source Community verbessert worden. Die aus



**Abbildung 5.7.:** Unterschiedliche Auflösungen, die während des Durchlaufs des FastDTW-Algorithmus' durchlaufen werden. (nach [40, S. 4])

der Komplexität erwachsenen Performanceprobleme konnten jedoch nicht abschließend gelöst werden. Daher wurde vom Autor als Alternative das sogenannte *FastDTW* [40] implementiert. FastDTW ist eine Näherung des DTW-Algorithmus', die trotz einer Laufzeit von  $\mathcal{O}(N)$  vergleichbare Ergebnisse liefert.

Die Idee hinter FastDTW ist, die Auflösung der betrachteten Kostenmatrix rekursiv zu halbieren. Für die größte Auflösung und damit den einfachsten Fall wird das DTW vollständig berechnet. Der errechnete Pfad wird auf die nächsthöhere Auflösung projiziert. Diese Projektion stellt nun den Suchraum für ein weiteres DTW dar, vergrößert um einen wählbaren Radius  $r$ . Der Vorgang wird so oft wiederholt, bis die ursprüngliche Auflösung erreicht wird (siehe Code 3). Abbildung 5.7 veranschaulicht die verschiedenen Stufen während der Berechnung. Die Genauigkeit von FastDTW steigt mit wachsender Suchdistanz  $r$  und erreicht bei Werten von  $r = 30$  durchschnittliche Fehler von unter 1% (vgl. [40, S. 7]). Für die Belange dieser Arbeit hat sich diese Genauigkeit als ausreichend herausgestellt, da die dadurch potentiell auftretenden Abweichungen somit weit unterhalb der Spieldauer der kürzesten Note der MEI-Kodierung lagen.

### 5.3.4. User Interface

Um Annotationen direkt auf dem Bildschirm in die Noten einzeichnen zu können, war eine Unterstützung des Stiftes des Microsoft Surface Books nötig. Hier zeigten sich Kompatibilitätsprobleme mit dem Betriebssystem (Windows 10), da Qt zwar eine sehr umfangreiche und komfortable Unterstützung für Stifteingaben bereithält, Microsoft jedoch zusammen mit der Markteinführung des Surface Books die Treiberarchitektur in Windows 10 grundlegend verändert hat. Zahlreiche Programme und Bibliothek sind dadurch inkompatibel geworden, unter anderem Qt. Um diese Hürde zu überwinden, musste ein weiteres Tool entwickelt werden, das die Stift-Events der neuen Treiberarchitektur auf das alte System „übersetzt“. Aus rechtlichen Gründen darf dieses Programm leider nicht veröffentlicht werden.

Die seitlich am Stift angebrachte Taste wird im Prototypen genutzt, um zwischen zwei verschiedenen Annotationsfarben zu wechseln. Sobald der Stift den Bildschirm zum

```

1 x = "Chroma-Sequenz 1"
2 y = "Chroma-Sequenz 2"
3 radius = "Suchdistanz zum projizierten Pfad beim Verfeinern der Auflösung"
4
5 def FastDTW(x, y, fenster):
6     # Kleinste Auflösung festlegen
7     min_fenster = radius + 2
8
9     if (x.length <= min_fenster) or (y.length <= min_fenster):
10        # Führe vollständiges DTW für kleinsten Fall aus
11        return DTW(x, y)
12
13    else:
14        # Halbiere die Auflösung des betrachteten Pfades
15        x_neu = x.halbiere()
16        y_neu = y.halbiere()
17
18        # Führe Funktion rekursiv aus mit geringerer Auflösung
19        pfad_grob = FastDTW(x_neu, y_neu, radius)
20
21        # Projiziere den Pfad der größeren Auflösung auf die aktuelle
22        # Auflösung und erweitere das Suchfenster um 'radius'
23        such_fenster = erweitere_such_fenster(pfad_grob, x, y, radius)
24
25        # Führe DTW auf 'such_fenster' aus
26        return DTW(x, y, such_fenster) # Komplexität  $\mathcal{O}(N)$ 

```

**Code 3:** Pseudocode des FastDTW-Algorithmus. (nach [40, S. 4])

Schreiben berührt, wird der Status der Taste ein Mal ausgewertet und der Pfad der Eintragung entsprechend eingefärbt, unabhängig von weiteren Tastenereignissen. So wird sichergestellt, dass innerhalb einer abgeschlossenen Annotation nicht mehrfach die Farbe gewechselt wird.

Alle mit dem Notentext synchronisierten Aufnahme-Takes werden wie im vorgestellten Bedienkonzept (vgl. Abschnitt 5.2.1) direkt unter den Noten als horizontale Balken angezeigt (siehe Abbildung 5.8). Klickt man mit der Maus oder dem Stift auf solch einen Take-Balken, wird der entsprechende Take ab der entsprechenden Stelle im Notentext abgespielt. Zur Erhöhung der Genauigkeit des Startpunktes der Wiedergabe wird beim Klicken zwischen zwei Noten bzw. Pausen der Zeitstempel für den gewünschten Take an der Stelle interpoliert. Ähnlich wird auch bei der Anzeige des vertikalen Positionsbalken bei der Wiedergabe vorgegangen. Hier wird – diesmal ausgehend von der aktuellen Zeit des gerade abgespielten Takes – ebenfalls interpoliert und die Pixelkoordinaten entsprechend fein aufgelöst berechnet.



**Abbildung 5.8.:** Fertiger Prototyp mit einer Beispielaufnahme und Annotationen für Demonstrationszwecke.

## 5.4. Evaluation

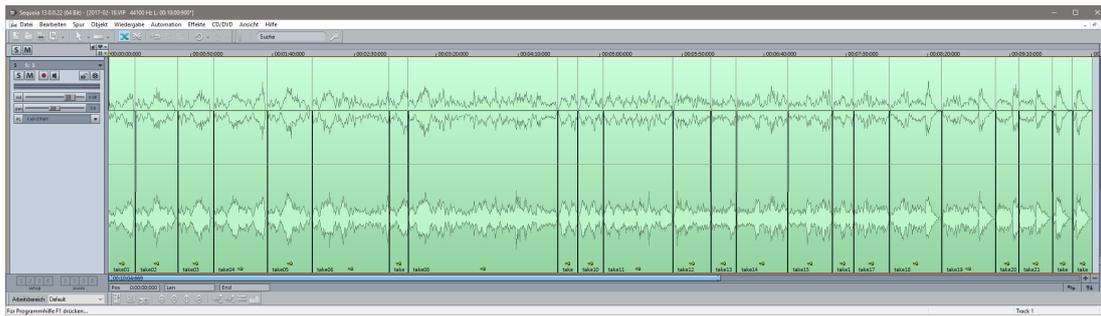
Im Anschluss an die Implementierung der neuen Navigationsoberfläche sollte diese getestet werden. Interessant war vor allem die Frage, ob das neue Konzept eine schnellere Navigation in einer Aufnahme mit mehreren Takes zulassen würde, als es mit herkömmlichen DAWs möglich ist. Hierzu wurde ein Test konzipiert, der aufzeigen sollte, wie schnell bestimmte Positionen in Takes aufgefunden und abgespielt werden können, wenn die äquivalente Position in den Noten bereits bekannt ist.

### 5.4.1. Testdesign und Durchführung

Die Probanden des Tests sollten sowohl in einer DAW als auch im Prototypen Stellen im Audiomaterial finden, an denen künstlich ein Signalton eingewoben wurde. Die Positionen der Töne wurden durch entsprechende Markierungen im Notentext kenntlich gemacht.

Zum direkten Vergleichen der entwickelten Software mit modernen DAWs, wurde eine Aufnahme mit mehreren Takes benötigt. Es wurde das Präludium in Es-Dur aus dem *Wohltemperierten Klavier II* von Johann Sebastian Bach (BWV 876) vom Autor selbst eingespielt. Als Resultat lag eine Aufnahme in 23 Takes vor, die zum Teil absichtlich mit tonalen Fehlern gespickt waren, um zusätzlich die Robustheit des Alignment-Algorithmus' zu demonstrieren.

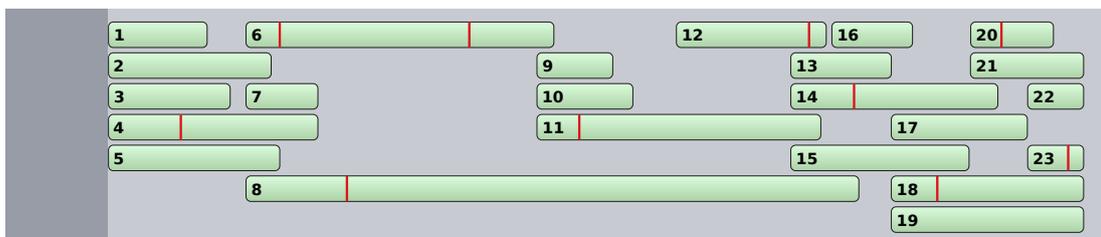
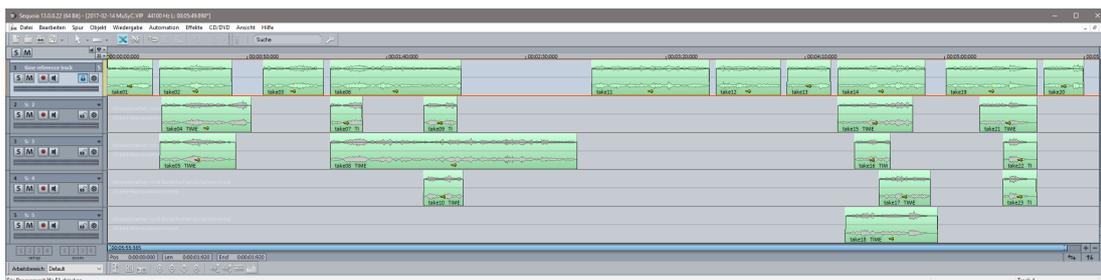
Nach der Einspielung lagen die Takes hintereinander gerendert in der DAW vor, siehe Abbildung 5.9. Es wurden willkürlich 10 Stellen innerhalb dieser Takes gewählt, an denen ein gut hörbarer Signalton eingearbeitet wurde. Es wurde ein Sinuston mit einer Frequenz von 1760 kHz und einer Länge von 200 ms genutzt. In Abbildung 5.10 (unten)



**Abbildung 5.9.:** Hintereinander gerenderte Takes der Testaufnahme in Sequoia.

sind die Positionen der Signaltöne in den einzelnen Takes rot markiert. Die Töne wurden auf einer zusätzlichen Spur erstellt, die anschließend in der DAW ausgeblendet wurde, sodass keine Auffälligkeiten optisch ersichtlich waren.

Im Interesse eines fairen Vergleichs mit den Synchronisierungsfähigkeiten moderner DAWs wurde anschließend versucht, den Multisynchronschnitt von Sequoia zu nutzen. Der Algorithmus zum Alignment der Takes versagte jedoch bei der verwendeten Testaufnahme. Wie in Abbildung 5.10 (oben) zu sehen ist, war Sequoia nicht in der Lage, eine durchgängige Aufnahme aus den Takes zu bilden. Korrekt wäre die darunter skizzierte Anordnung gewesen, wie sie auch vom entwickelten Prototypen in dieser Arbeit berechnet wird.



**Abbildung 5.10.:** Erfolgreiche Analyse des Multisynchronschnitts in Sequoia (oben) im Vergleich zum korrekten Alignment der Takes der Testaufnahme (unten).

Zur Nutzung der in dieser Arbeit entwickelten Software wurde eine entsprechende Notenkodierung des Präludiums benötigt. Dazu wurde der Notentext zunächst in der Notationssoftware *Sibelius* gesetzt und anschließend mit einem frei verfügbaren Exporter [15] in das MEI-Format übersetzt. Dieser Schritt verlief nicht fehlerfrei und musste daher von Hand nachgebessert werden. In einem Papierausdruck des Notensatzes wurden die 10 Stellen markiert, an denen der Signaltone in den Takes untergebracht wurde. Nicht ersichtlich war, zu welchem Take die Markierungen gehören.

Anschließend wurde die MEI-Datei zusammen mit den Takes in die entwickelte Software geladen. Die Takes wurden dabei im Vorfeld per Hand mit entsprechenden Start- und Endmarkern annotiert, sodass das Alignment korrekt durchgeführt werden konnte. Die 10 Positionen der Signaltöne wurden direkt in die Noten auf dem Bildschirm eingezeichnet.

Alle Teilnehmer sollten folgende zwei Aufgaben ausführen:

**Navigation in DAW** Die Probanden wurden gebeten, die Signaltöne innerhalb der hintereinander dargestellten Takes in Sequoia zu finden. Als Hilfsmittel wurden ihnen der Notenausdruck mit den Einzeichnungen zur Verfügung gestellt. Zusätzlich erhielten sie eine Tabelle, in der notiert war, welche Takte in den einzelnen Takes aufgenommen wurden. Dieses Szenario entspricht der typischen Ausgangslage nach einer fertigen Aufnahme.

**Navigation in Prototyp** Den Teilnehmern wurde kurz das Konzept des Prototypen erläutert. Anschließend sollten sie die obige Aufgabe mit der neuen Software lösen. Die Eintragungen in den Noten waren entsprechend digital verfügbar, eine separate Take-Liste wurde nicht angefertigt.

Um die Einflüsse der Reihenfolge der durchgeführten Aufgaben auf das Testergebnis zu minimieren, wurden drei der sechs Teilnehmer gebeten, erst mit dem Prototypen und anschließend in der DAW zu arbeiten. Die verbleibenden Teilnehmer taten dies in umgekehrter Reihenfolge. Im Anschluss wurde jeder Proband in einem kurzen Gespräch um Kritik zur Oberfläche und Bedienung gebeten.

#### 5.4.2. Ergebnisse

Die gestoppten Zeiten der Teilnehmer sind in Tabelle 5.1 zusammengefasst. Im Durchschnitt dauerte die Suche in der DAW mithilfe der gedruckten Noten etwa 8 Minuten. Die gleiche Aufgabe konnte mit dem Prototypen in durchschnittlich 1:40 Minuten gelöst werden, benötigte also im Vergleich nur etwa 21% der Zeit. Auffallend war, dass ausnahmslos alle sechs Probanden mindestens 76% der Zeit für die Aufgabe einsparen

Teilnehmer	#1	#2	#3	#4	#5	#6	Ø
Zeit mit DAW	7:31	7:58	8:44	7:30	8:02	8:16	<b>8:00</b>
Zeit mit Prototyp	1:25	1:19	2:06	1:46	1:33	1:51	<b>1:40</b>
Zeitersparnis	81.2%	83.5%	76.0%	76.4%	80.7%	77.6%	<b>79.2%</b>

**Tabelle 5.1.:** Zeiten der einzelnen Teilnehmer für die Suche der in den Noten markierten Stellen im Audiomaterial. Die Teilnehmer #1, #2 und #3 nutzten zuerst den Prototypen und danach die DAW, die anderen entsprechend umgekehrt. (Zeitangaben in M:SS)

konnten beim Nutzen des Prototypen, also bei allen Teilnehmenden eine eindeutige Verbesserung des Workflows zu beobachten war.

### Kritik

Die Resonanz der Teilnehmer war nach dem Test durchweg positiv. Konstruktive Kritik gab es zu zwei Punkten, die inhaltlich korrelieren:

**Unübersichtlichkeit** Zwei der Probanden sagten aus, dass die Darstellung der Musik in einer Endloszeile den Überblick erschweren würde, da man unter Umständen die Orientierung im Gesamtwerk verlieren könne. Abhilfe würden einfache Bezugspunkte wie z.B. Taktnummern schaffen. Es könne auch auf eine Endloszeile verzichtet werden und Systemumbrüche und Seitenumbrüche wie bei gedruckten Noten dargestellt werden.

**Zu häufiges Scrollen** Ein Proband befand, dass das Scrollen bzw. „Wischen“ durch den Notentext bei längerem Einsatz der Software ermüdend wirken könnte. Der zu jedem Zeitpunkt angezeigte Notenausschnitt solle „irgendwie“ vergrößert werden.

### Diskussion

Die Evaluation des Prototypen hat die Erwartungen an die dadurch gewonnene Zeitersparnis im Vergleich zu den Ergebnissen der durchgeführten Vorstudie nicht nur erfüllt, sondern sogar übertroffen. Obwohl der abschließende Nutzertest kein gänzlich realistisches Szenario aufgespannt hat, wurde dennoch gezeigt, dass die zeitaufwendigsten Vorgänge – die gezielte Suche in Aufnahmen aufgrund von Noteneinträgen und der Vergleich mehrerer Takes – erheblich beschleunigt werden konnten.

Eine Testgruppe mit sechs Probanden erfüllt nicht die Ansprüche an eine vollständige Nutzerstudie. Obwohl daher eine umfangreichere Evolution noch aussteht, zeichnet

sich bei der Benutzung des Prototypen dennoch eindeutig eine zeitsparende Verbesserung des Workflows beim Schnitt ab.



# 6

## Fazit

Im Rahmen dieser Arbeit wurde ein Konzept zur Navigation innerhalb von Audioaufnahmen bei klassischen Musikproduktionen erarbeitet und prototypisch implementiert. Neuartig war die Verknüpfung zwischen mehreren Aufnahme-Takes mit dem dazugehörigen Notentext. Auf diese Weise wurde der Wechsel zwischen gedruckten Noten und dem Computerbildschirm, wie er beim Schnitt klassischer Musik fortwährend auftritt, eliminiert. Um eine derartige Synchronisation zu ermöglichen, wurden etablierte Algorithmen aus dem Fachgebiet des Music Information Retrieval genutzt. Die Verfahren wurden für den betrachteten Anwendungszweck angepasst bzw. optimiert. Der ständige Kontakt zu Tonmeistern sowie der fachliche Diskurs auf internationalen Konferenzen stellten sicher, dass die Entwicklungen sich an praxisnahen Problemen orientierten.

Erste Nutzerstudien mit der entwickelten Software machten deutlich, dass die durchgeführte Verknüpfung von Noten und Audio beträchtliches zeitliches Einsparpotenzial birgt, das besonders während des Schnitts aufgenommener Musik zum Tragen kommen kann. Die Dauer zum Suchen von im Notentext markierten Positionen in Audio-Dateien konnte im Test um durchschnittlich 79% gesenkt werden.

## Ausblick

Die Entwicklung des Prototypen lässt eine ganze Reihe von Problemen unangetastet, die einen Produktiveinsatz der Software in der vorliegenden Form unmöglich machen:

**Verfügbarkeit digitaler Noten** Viele Musikverlage bieten ihre Produkte nach wie vor nur als Printmedien an. Auch wenn mittlerweile größere Verlage eine Umstellung ihres Katalogs für die digitale Darstellung planen, wird die Verfügbarkeit digitaler Notenkodierung in den nächsten Jahren ein Problem bleiben.

**Aufnahmefunktion** Bei der Implementierung des Prototyps wurde bewusst auf die Entwicklung einer Aufnahmefunktion verzichtet. Um jedoch die Annotationen der Stifteingabe automatisch den korrekten Takes zuordnen zu können, müsste diese Funktion noch nachgereicht werden.

**Alignment-Algorithmen** Der verwendete DTW-Algorithmus ist mit hoher Wahrscheinlichkeit noch verbesserungsfähig. Eine ausführliche Evaluation mit unterschiedlichem Klangmaterial steht noch aus. Neuere Forschungen beschäftigen sich mit der Berechnung der DTW über mehrere Aufnahmen hinweg [48, 49] und könnten ebenfalls einbezogen werden. Der in Abschnitt 4.2 erwähnte Ansatz mittels Partikelfiltern scheint in seiner Grundidee eine interessante Alternative zu sein.

Die Arbeitsumgebungen in der Musikproduktion haben sich durch die Einführung der Digitaltechnik in den letzten 50 Jahren sehr stark gewandelt. Während Speichermedien und Klangtechnologien sich seither stetig weiterentwickeln, fehlt neuen Bedienkonzepten die nötige Akzeptanz. Hier ist Fingerspitzengefühl gefragt. Tonmeister und Musikproduzenten müssen bei der Diskussion möglicher Neuentwicklungen stärker zu Wort kommen. Konzepte, die versuchen, einen über Jahrzehnte gewachsenen Workflow zu ersetzen, haben meist nur wenig Aussicht auf Erfolg. Es ist daher ratsam, die vorhandenen Konzepte zu studieren und mit neuen Ideen zu bereichern.

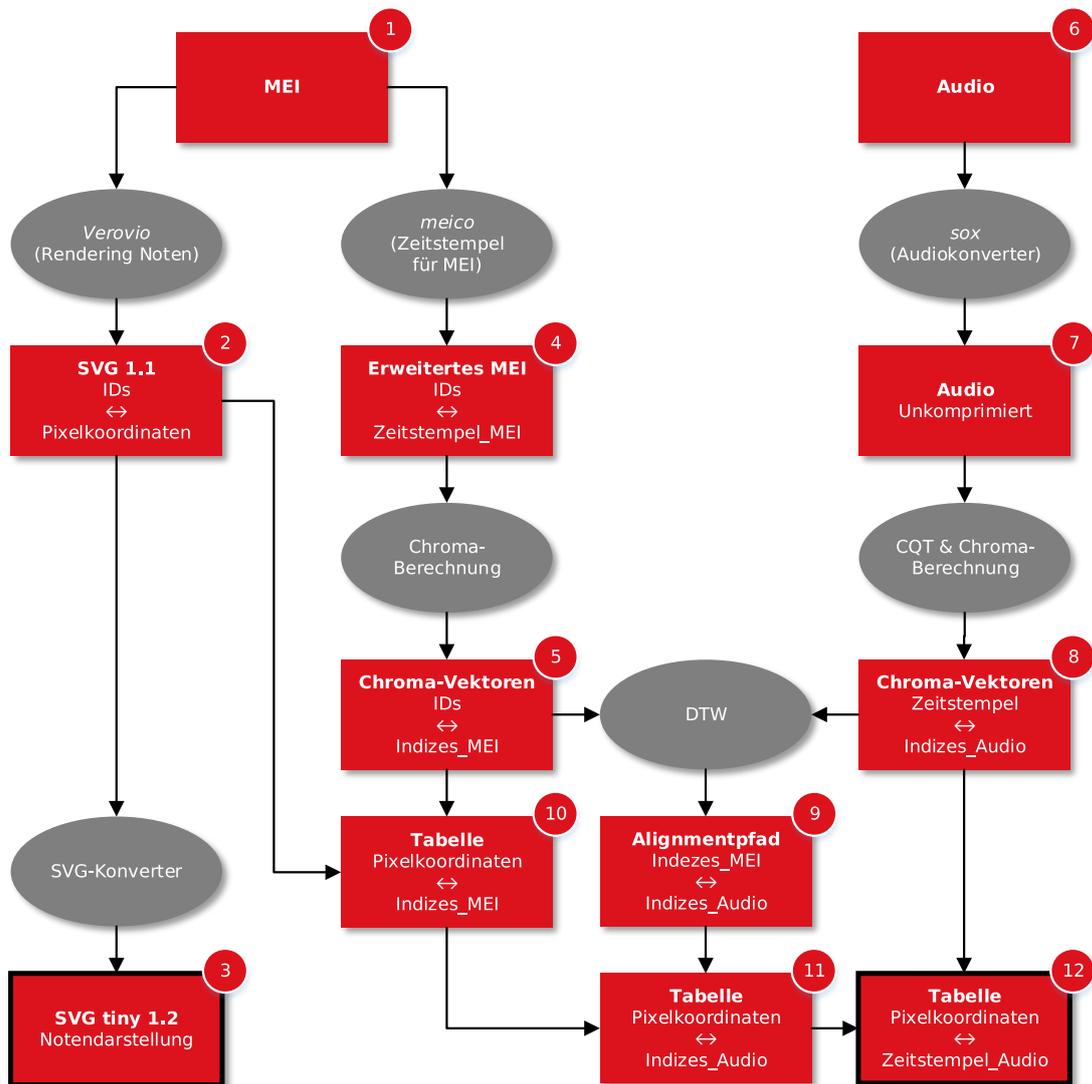
In diesem Sinne ist die vorliegende Arbeit ein erster (praktischer) Schritt, der das Potential digitaler Noten in einem Teilbereich der Musikproduktion demonstriert. Es bleibt zu hoffen, dass künftige Entwicklungen neuer Nutzungsschnittstellen den Tonmeister und seine Belange und Gewohnheiten in den Vordergrund rücken, sodass Ideen nicht an den gefestigten Arbeitsstrukturen der Musikbranche abprallen.



# A

## Anhang

Der Anhang beschränkt sich auf die Abbildung A.1. Diese ist als „Ausklapp-Grafik“ gedacht, um während des Lesens des Implementationsabschnitts 5.3 stets einen Gesamtüberblick der beteiligten Komponenten vor Augen zu haben.



**Abbildung A.1.:** Zusammenspiel der einzelnen Komponenten zum Synchronisieren von Noten und Audio. In Rechtecken mit „↔“ markierte Kommentare geben an, dass zwischen diesen Daten eine Verknüpfung besteht.

# Literatur

- [1] Anoto. *Development Guide for Service Enabled by Anoto Functionality*. Techn. Ber. Anoto AB, 2002.
- [2] Berndt, Axel. *meico – MEI Converter*. Version 0.2.11. URL: <https://github.com/cemfi/meico/> (besucht am 10.02.2017).
- [3] A. Berndt, S. Waloschek und A. Hadjakos. „Hand Gestures in Music Production“. In: Utrecht, The Netherlands, Sep. 2016.
- [4] B. Blankertz. *The Constant Q Transform*. 1999. URL: <http://wwwmath.uni-muenster.de/logik/Personen/blankertz/constQ/constQ.html> (besucht am 18.01.2017).
- [5] B. W. Bohl u. a. „Dem Igel Sitte lehren... Musikedition: von der digitalen Verfügbarkeit zur aktiven Nutzung“. In: *„Ei, dem alten Herrn zoll' ich Achtung gern'“ – Festschrift für Joachim Veit zum 60. Geburtstag*. Hrsg. von K. Richts und P. Stadler. Munich, Germany: Allitera Verlag, Jan. 2016. Kap. 12, S. 141–163.
- [6] J. Brown. *Calculation of a constant Q spectral transform*. In: *The Journal of the Acoustical Society of America* 89.1 (1991), S. 425–434.
- [7] C. Chafe, B. Mont-Reynaud und L. Rush. *Toward an Intelligent Editor of Digital Audio: Recognition of Musical Constructs*. In: *Computer Music Journal* 6.1 (1982), S. 30–41.
- [8] Chris Bagwell and Rob Sykes and Pascal Giard. *SoX - Sound eXchange*. URL: <http://sox.sourceforge.net/> (besucht am 15.02.2017).
- [9] R. B. Dannenberg und N. Hu. „Polyphonic Audio Matching for Score Following and Intelligent Audio Editors“. In: *Proceedings of the International Computer Music Conference*. 2003.
- [10] J. Eargle. *Handbook of Recording Engineering*. Springer US, 2006.
- [11] J. Foote. „Visualizing Music and Audio Using Self-similarity“. In: *Proceedings of the 7th ACM International Conference on Multimedia (Part 1)*. 1999.
- [12] C. Fremerey. *Automatic Organization of Digital Music Documents – Sheet Music and Audio*. Dissertation. 2010.

- [13] K. Gohlke u. a. „Track Displays in DAW Software: Beyond Waveform Views“. In: *Audio Engineering Society Convention 128*. Mai 2010.
- [14] T. Görne. *Tontechnik*. Hanser, 2008.
- [15] A. Hankinson. *Sibelius to MEI Plugin*. URL: <https://github.com/music-encoding/sibmei> (besucht am 18.02.2017).
- [16] T. Herberger und T. Tost. *System and method for synchronized multi-track editing*. US Patent 8,464,154. Juni 2013. URL: <https://www.google.com/patents/US8464154> (besucht am 15.01.2017).
- [17] K. Holtzblatt, J. Wendell und S. Wood. *Rapid Contextual Design: A How-to Guide to Key Techniques for User-centered Design*. Interactive Technologies Series. Elsevier/Morgan Kaufmann, 2005.
- [18] N. Hu, R. B. Dannenberg und G. Tzanetakis. „Polyphonic audio matching and alignment for music retrieval“. In: *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 2003.
- [19] iZotope, Inc. *RX 5 Audio Editor*. URL: <https://www.izotope.com/en/products/repair-and-edit/rx.html> (besucht am 01.02.2017).
- [20] J. Loviscach. „The Quintessence of a Waveform: Focus and Context for Audio Track Displays“. In: *Audio Engineering Society Convention 130*. Mai 2011.
- [21] G. Loy und J. Chowning. *Musimathics, Volume 2: The Mathematical Foundations of Music*. The MIT Press, 2007.
- [22] Magix Software GmbH. *Sequoia 13*. URL: <http://www.magix-audio.com/de/sequoia/> (besucht am 17.01.2017).
- [23] Magix Software GmbH. *SpectralLayers Pro 4*. URL: <http://www.magix-audio.com/gb/spectralayers-pro/> (besucht am 17.01.2017).
- [24] MakeMusic, Inc. *MusicXML 3.0 Documentation*. 2016. URL: <http://usermanuals.musicxml.com/MusicXML/MusicXML.htm> (besucht am 21.01.2017).
- [25] B. McFee u. a. „librosa: Audio and music signal analysis in python“. In: *Proceedings of the 14th Python in Science Conference*. 2015.
- [26] Microsoft Corporation. *Surface Book - The Ultimate i5 or i7 powered Laptop Computer*. URL: <https://www.microsoft.com/en-us/surface/devices/surface-book/overview/> (besucht am 14.02.2017).
- [27] MIDI Manufacturers Association. *The Complete MIDI 1.0 Detailed Specification*. Version 96.1, 2<sup>nd</sup> edition. 2001.

- 
- [28] N. Montecchio und A. Cont. „A unified approach to real time audio-to-score and audio-to-audio alignment using sequential Montecarlo inference techniques“. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2011.
- [29] N. Montecchio und A. Cont. „Accelerating the Mixing Phase in Studio Recording Productions by Automatic Audio Alignment“. In: *International Symposium on Music Information Retrieval*. Okt. 2011.
- [30] W. Moylan. *The Art of Recording: The Creative Resources of Music Production and Audio*. Van Nostrand Reinhold, 1992.
- [31] M. Müller. *Information Retrieval for Music and Motion*. Springer, 2007.
- [32] M. Müller. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, 2015.
- [33] J. Paulus, M. Müller und A. Klapuri. „Audio-based music structure analysis“. In: *Proceedings of the 11th International Society for Music Information Retrieval Conference*. 2010.
- [34] N. Pawera. *Mikrofonpraxis: Tipps und Tricks für Bühne und Studio; Technik, Akustik und Aufnahmepraxis für Instrumente und Gesang*. 4. Aufl. PPV-Medien, 2004.
- [35] B. Preim und R. Dachsel. *Interaktive Systeme – Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung*. eXamen.press. Springer Berlin Heidelberg, 2010.
- [36] L. Pugin, R. Zitellini und P. Roland. „Verovio: A Library For Engraving MEI Music Notation Into SVG“. In: *Proceedings of the International Society for Music Information Retrieval Conference*. 2014.
- [37] A. Rebelo u. a. *Optical music recognition: state-of-the-art and open issues*. In: *International Journal of Multimedia Information Retrieval* 1.3 (2012), S. 173–190.
- [38] S. V. Rice. „Frequency-Based Coloring of the Waveform Display to Facilitate Audio Editing and Retrieval“. In: *Audio Engineering Society Convention 119*. Okt. 2005.
- [39] P. Roland. „The Music Encoding Initiative (MEI)“. In: *Proceedings of the 1st International Conference on Musical Applications Using XML*. 2002.
- [40] S. Salvador und P. Chan. *FastDTW : Toward Accurate Dynamic Time Warping in Linear Time and Space*. In: 11.5 (2007).
- [41] S. Savage. *The Art of Digital Audio Recording*. Oxford University Press, USA, 2011.
- [42] M. Schmid. *Notationskunde: Schrift und Komposition 900-1900*. Bärenreiter Studienbücher Musik. Bärenreiter, 2012.
- [43] E. Selfridge-Field. *Beyond MIDI: The Handbook of Musical Codes*. MIT Press, 1997.

- [44] R. Spence. *Information Visualization: An Introduction*. 3rd. Springer, 2014.
- [45] V. Thomas u. a. *Linking sheet music and audio-Challenges and new approaches*. In: *Dagstuhl Follow-Ups 3* (2012).
- [46] S. Waloschek u. a. „Accelerating the editing phase in music productions using interactive scores“. In: *Proceedings of the 2nd AES Workshop on Intelligent Music Production*. London, UK, Juli 2016.
- [47] S. Waloschek u. a. „Interactive Scores in Classical Music Production“. In: *Proceedings of the 17th International Society for Music Information Retrieval Conference*. Aug. 2016.
- [48] S. Wang, S. Ewert und S. Dixon. „Robust Joint Alignment of Multiple Versions of a Piece of Music“. In: *Proceedings of the 15th International Society for Music Information Retrieval Conference*. 2014.
- [49] S. Wang, S. Ewert und S. Dixon. *Robust and Efficient Joint Alignment of Multiple Musical Performances*. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24.11 (Nov. 2016), S. 2132–2145.
- [50] M. Wattenberg. „Arc diagrams: Visualizing structure in strings“. In: *IEEE Symposium on Information Visualization 2002*. 2002.
- [51] M. A. Winget. *Annotation of Musical Scores: Interaction and Use Behaviours of Performing Musicians*. Dissertation. 2006.
- [52] G. Wood und S. O’Keefe. „On Techniques for Content-Based Visual Annotation to Aid Intra-Track Music Navigation.“ In: *ISMIR*. 2005.
- [53] H.-H. Wu und J. P. Bello. „Audio-based music visualization for music structure analysis“. In: *Proceedings of Sound and Music Computing Conference*. 2010.

# **Selbstständigkeitserklärung**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt, nicht anderweitig zu Prüfungszwecken vorgelegt und keine anderen als die angegebenen Hilfsmittel verwendet habe. Sämtliche wissentlich verwendeten Textausschnitte, Zitate oder Inhalte anderer Verfasser sind ausdrücklich als solche gekennzeichnet.

Detmold, den 23. Februar 2017

---

Simon Waloschek