Audioschnitt in digitalen Noten

Simon Waloschek[1]

Abstract: Trotz der inzwischen vollständig digitalen Aufnahme und Nachbearbeitung von Musik wird speziell bei Produktionen klassischer Musik in der Regel zusätzlich 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. Es 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.

Keywords: Interaktive Partituren, Musikproduktion, Audio to Score Alignment.

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[2] als auch ausführenden 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 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] Hochschule für Musik Detmold, Center of Music and Film Informatics, Hornsche Str. 44, 32756 Detmold, s.waloschek@cemfi.de

[2] In diesem Paper wird aus Gründen der leichteren Lesbarkeit auf das konsequente Gendern verzichtet. Im Zweifel sind jeweils sowohl die weibliche als auch die männliche Form gemeint.

doi:10.18420/in2017_13

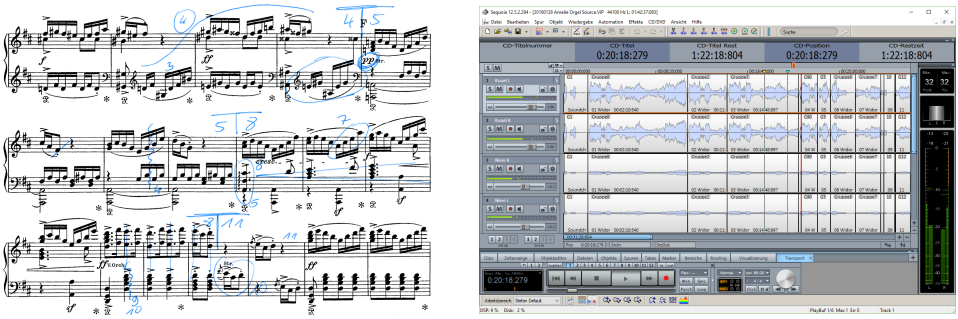

Abb. 1: Beispielhafte Annotationen eines Tonmeisters im Notentext (**links**) gegenüber der Wellenformdarstellung des aufgezeichneten Audiomaterials in einem Audiobearbeitungsprogramm (**rechts**).

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 Forschungsarbeit 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.

2 Produktion klassischer Musik

Dieser Abschnitt 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.

Vergleichende Literatur [Ea06, Mo92] unterteilt Musikproduktionen stets in drei große Phasen: *Vorbereitung*, *Aufnahme* und *Nachbearbeitung*. Jede dieser Phasen hat unterschiedliche Ansprüche an die genutzten Programme, den Notentext sowie die Arbeitsumgebung. Häufig situations- und projektabhängig angepasst, sind die nachfolgend beleuchteten Schritte einer Klassikproduktion nicht als starres Konstrukt zu verstehen. Wirtschaftliche Betrachtungen werden außen vor gelassen.

2.1 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. Potentiell für die Musizierenden herausfordernde Passagen werden identifiziert und ein Aufnahmeplan erstellt mit dem Ziel, die zur Verfügung stehende Aufnahmezeit

möglichst optimal ausnutzten sollte. Bereits in dieser Phase wurden erste Anmerkungen in die Noten eingetragen, wie z.B. mögliche Stellen für nachträgliche Schnitte.

2.2 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 entwickelten Audioeditoren. Aufgenommene Takes werden hier direkt horizontal aneinandergereiht (siehe Abbildung 1, rechts).

Im Laufe der Produktion werden verschiedene Annotationen direkt im Notentext eingetragen. Damit werden zwei unterschiedliche Ziele verfolgt: Zunächst dienen Sie als „Gedächtnisstütze“, um nach dem Ende des gerade aufgezeichneten Takes alle wichtigen Anmerkungen kommunizieren zu können. Nach der Aufnahme dienen diese Annotationen der Orientierung bei der Bearbeitung, da nicht nur take-spezifische musikalische Kommentare verzeichnet werden, 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.

Einzelne Passagen werden häufig mehrfach gespielt, wodurch die Fülle an Einzeichnungen zunimmt. Dies erschwert den Überblick und macht eine nachträgliche Zuordnung einzelner Annotationen zu spezifischen Takes zusehends schwieriger. Zusätzlich zu den Annotationen wird häufig parallel eine *Take-Liste* geführt, die für jeden Take die gespielten Takte verzeichnet.

2.3 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 [Ea06].

Der Schnitt gliedert sich in in zwei Teilaufgaben. Zunächst werden im gesammelten Audiomaterial die Teile herausgesucht, die für das Endergebnis genutzt werden sollen. Im Idealfall unterstützen die entsprechenden Annotationen im Notentext sowie ein bereits

während der Aufnahme angefertigter Schnittplan diesen Suchprozess. Die selektierten Teile müssen dann „nach Gehör“ derart aneinandergeschitten werden, dass nicht mehr hörbar ist, dass es sich um unterschiedliche Takes handelt.

Für den eigentlichen Schnittprozess bedient man sich oft 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.

Beide Schnitt-Strategien haben gemein, dass ein ständiger Wechsel zwischen dem annotierten Notenmaterial und der Wellenformdarstellung in der DAW (siehe Abbildung 1, rechts) nötig ist. Die Entscheidung, welche Takes bzw. welche Teil der Takes aneinandergeschnitten werden sollen, ist 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 sind daher eine zeitsparende Hilfe.

Die letzten Schritte in der Produktion von Instrumentalmusik sind das *Mixing* und *Mastering*. Mixing bezeichnet dabei Klanganpassungen und -korrekturen nach dem Schnitt. Das finale Mastering hingegen 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. Diese Prozesse sind weitestgehend unabhängig vom Notentext und betrachten lediglich klangliche Aspekte und werden daher an dieser Stelle nicht weiter adressiert.

3 Related Work

Die dieser Forschungsarbeit zugrunde liegende Idee der Synchronisation mehrerer Takes innerhalb einer Aufnahme unter Einbeziehung von Noten haben Dannenberg et al. bereits 2003 skizziert [DH03]. Konkrete Implementierungen wurden jedoch nicht diskutiert, sodass es lediglich bei einem Mock-up geblieben ist (siehe Abbildung 2).

In [MC11a] wird ein ähnliches Konzept verfolgt, jedoch mit einem anderen Synchronisierungsalgorithmus auf Basis von Partikelfiltern [MC11b].

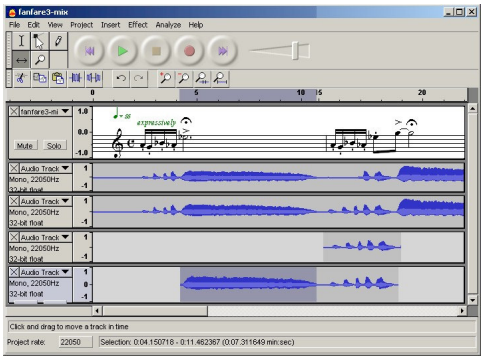

Abb. 2: Mock-up einer DAW mit synchronisierten Takes unterhalb einer Notendarstellung. (Quelle: [DH03, S. 6])

Eine erste praktische Nutzung der Audiosynchronisierung erfolgte 2009 mit der Markteinführung des *Multisynchronschnitts* [HT06] in der Audio-Software *MAGIX Sequoia 11*. 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.

Abb. 3: Multisynchronschnitt in der Software Sequoia (Quelle: `www.magix-audio.com/de/sequoia/klassikproduktion`, abgerufen am 30.04.2017)

Eine entscheidende Neuerung beim Multisynchronschnitt war die Abstraktion mehrspuriger Aufnahmen zu einer einzelnen Spur (siehe Abbildung 3). 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 Interaktive Partituren

Ausgehend von der im vorhergehenden Abschnitt skizzierten Idee der Abstraktion aller Mikrofonspuren eines Takes zu einem einzigen Objekt (Multisynchronschnitt), wurde ein Visualisierungs- und Bedienungskonzept entwickelt. Im Fokus stand die Integration von Noten, die zum aufgenommen Audiomaterial synchronisiert sein sollen. Das Konzept orientiert sich an ersten Überlegungen in [Wa16b, Wa16a].

4.1 Benutzeroberfläche und Visualisierung

Die Software-Arbeitsumgebungen während der Musikproduktion sind das Ergebnis eines über mehrere Jahrzehnte gewachsenen Bedienkonzepts. Grundsätzliche Änderung setzen sich nur schwerlich durch, da die kognitive Hürde bei Inkonsistenzen zu etablierte Bedienungen vergleichsweise hoch ist, wie auch entsprechende Studien gezeigt haben [BWH16, PD10]. 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, siehe Abbildung 4.

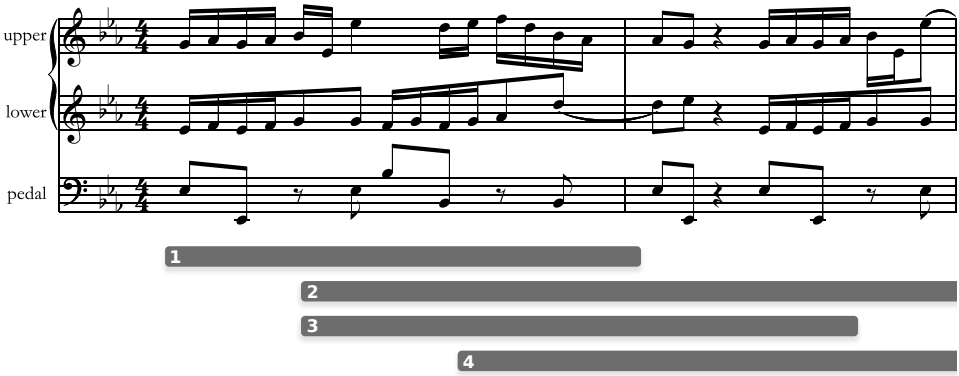

Abb. 4: Konzeptentwurf der Benutzeroberfläche. Dunkle Balken markieren dabei einzelne Takes, die unterhalb der Noten dargestellt werden.

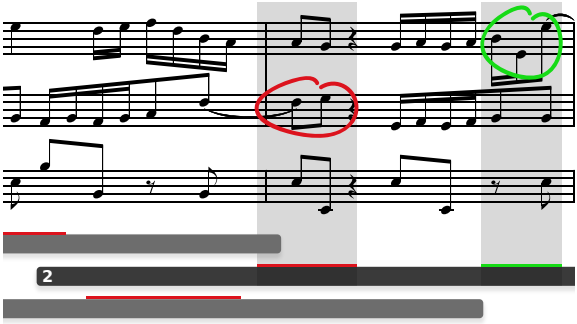

Abb. 5: Annotationen werden direkt im digitalen Notentext angezeigt und ihre Zuordnung zum jeweiligen Take hervorgehoben. Farbbalken signalisieren das Vorhandensein auch ausgeblendeter Einzeichnungen.

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. 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.

4.2 Bedienkonzept

Das vorgestellte Konzept setzt voraus, dass Annotationen digital erfasst werden können. Produkte wie z.B. das *Surface Book* von Microsoft 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 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 eineindeutig 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.

4.3 Implementierung

Die praktische Implementierung des Prototypen ist in der Programmiersprache *Python* durchgeführt worden. 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 6 fasst die dafür erforderlichen Schritte zusammen und bietet einen einleitenden Überblick.

Da im Folgenden immer wieder auf Abbildung 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.

Zunächst stehen die Noten als MEI-kodierte Datei [Ro02] zur Verfügung 1. MEI ist ein XML-basiertes Notenkodierungsformat, das stetig wachsende Verwendung im musikhistorischen Kontext findet und dazu genutzt wird, sowohl gedruckte als auch handgeschriebene Noten zu beschreiben. Der erste Schritt besteht nun in der Generierung

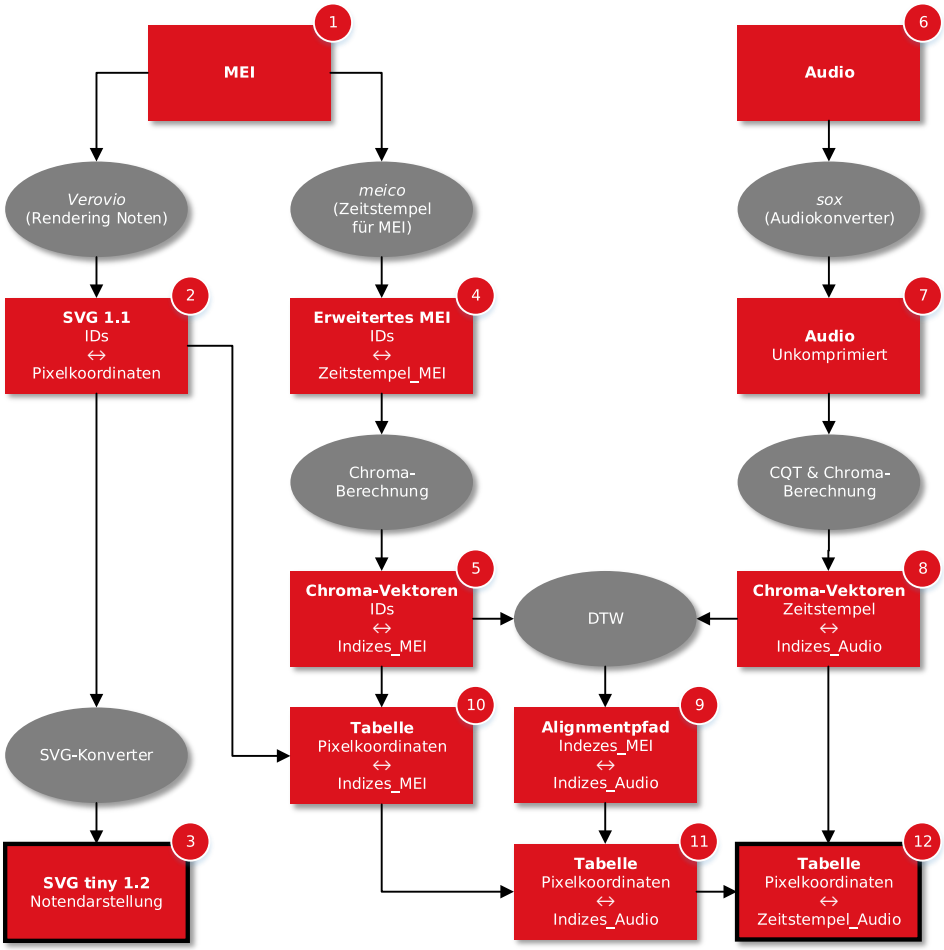

Abb. 6: Zusammenspiel der einzelnen Komponenten zum Synchronisieren von Noten und Audio. In Rechtecken mit „↔“ markierte Kommentare geben an, dass zwischen diesen Daten eine Verknüpfung besteht.

der Notendarstellung. Hierfür wird das Programm *Verovio* [PZR14] 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 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).

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 Python-Anbindung und umfangreiche Funktionen zur Darstellung von Pixel- und Vektorgrafiken.

Das Alignment des Notentextes mit den Audioaufnahmen wird mithilfe eines klassischen *Dynamic Time Warping*-Verfahrens (siehe [Mu07]) durchgeführt. Dafür ist es jedoch notwendig, dass für die einzelnen Noten Zeitpunkte definiert sein müssen. 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*[3]. 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 Zeitstempel 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).

Vor der Berechnung des Alignments muss ggf. das Audiomaterial dekodiert werden, da Audiodateien in verschiedenen Formaten (z.B. *MP3*, *FLAC*) mit unterschiedlichen Abtast- und Bit-Raten vorliegen können. Daher wird bei der Implementierung der Konverter *sox* [4] 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 des Alignment-Pfads kann mit der Python-Library *librosa* [Mc15] durchgeführt werden.

Der resultierende Pfad bildet Zeitpunkte innerhalb der Audiodatei auf Noten in der MEI-Kodierung ab (9). An dieser Stelle wird die Kenntnis der Pixelkoordinaten der einzelnen MEI-IDs (2) genutzt und so die Zuordnung von Koordinaten zu MEI-Noten (10) gebildet. Mithilfe des gerade berechneten Alignmentpfads lässt sich nun von Pixelkoordinaten der Notendarstellung auf die zeitliche Position der Audio-Dateien (12) schließen.

5 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.

[3] `https://github.com/cemfi/meico/`, abgerufen am 10.02.2017
[4] `http://sox.sourceforge.net/`, abgerufen am 14.05.2017

5.1 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 7. 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. Die Töne wurden auf einer zusätzlichen Spur erstellt, die anschließend in der DAW ausgeblendet wurde, sodass keine Auffälligkeiten optisch ersichtlich waren.

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 [Ha17] in das MEI-Format übersetzt. In einem Papierausdruck des Notensatzes wurden die 10 Stellen markiert, an denen der Signalton 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:

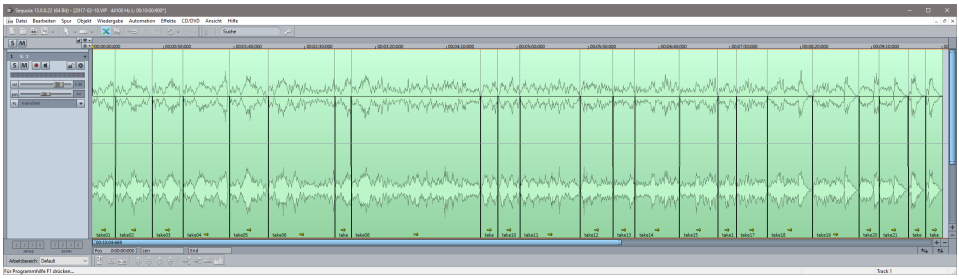

Abb. 7: Hintereinander gerenderte Takes der Testaufnahme in Sequoia.

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.

5.2 Ergebnisse

Die gestoppten Zeiten der Teilnehmer sind in Tabelle 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 konnten beim Nutzen des Prototypen, also bei allen Teilnehmenden eine eindeutige Verbesserung des Workflows zu beobachten war.

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

Tab. 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)

Die Evaluation des Prototypen hat die Erwartungen an die dadurch gewonnene Zeitersparnis erfüllt. Obwohl der abschließende Nutzertest kein gänzlich realistisches Szenario aufgespannt hat, wurde dennoch gezeigt, dass die zeitaufwendigsten Vorgänge – die gezielte Suche in Aufnahmen aufgrund von Noteneintragungen 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.

Literaturverzeichnis

[BWH16] Berndt, Axel; Waloschek, Simon; Hadjakos, Aristotelis: Hand Gestures in Music Production. 2016.

[DH03] Dannenberg, Roger B.; Hu, Ning: Polyphonic Audio Matching for Score Following and Intelligent Audio Editors. In: Proceedings of the International Computer Music Conference. 2003.

[Ea06] Eargle, J.: Handbook of Recording Engineering. Springer US, 2006.

[Ha17] Sibelius to MEI Plugin, https://github.com/music-encoding/sibmei, Stand: 2017-04-20.

[HT06] System and method for synchronized multi-track editing, US Patent 8,464,154.

[MC11a] Montecchio, Nicola; Cont, Arshia: Accelerating the Mixing Phase in Studio Recording Productions by Automatic Audio Alignment. In: International Symposium on Music Information Retrieval. 2011.

[MC11b] Montecchio, Nicola; Cont, Arshia: 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.

[Mc15] McFee, Brian; Raffel, Colin; Liang, Dawen; Ellis, Daniel PW; McVicar, Matt; Battenberg, Eric; Nieto, Oriol: librosa: Audio and music signal analysis in python. In: Proceedings of the 14th Python in Science Conference. 2015.

[Mo92] Moylan, W.: The Art of Recording: The Creative Resources of Music Production and Audio. Van Nostrand Reinhold, 1992.

[Mu07] Mueller, Meinard: Information Retrieval for Music and Motion. Springer, 2007.

[PD10] Preim, B.; Dachselt, R.: Interaktive Systeme – Band 1: Grundlagen, Graphical User Interfaces, Informationsvisualisierung. eXamen.press. Springer Berlin Heidelberg, 2010.

[PZR14] Pugin, L.; Zitellini, R.; Roland, P.: Verovio: A Library For Engraving MEI Music Notation Into SVG. In: Proceedings of the International Society for Music Information Retrieval Conference. 2014.

[Ro02] Roland, Perry: The Music Encoding Initiative (MEI). In: Proceedings of the 1st International Conference on Musical Applications Using XML. 2002.

[Wa16a] Waloschek, Simon; Berndt, Axel; Bohl, Benjamin W.; Hadjakos, Aristotelis: Accelerating the editing phase in music productions using interactive scores. In: Proceedings of the 2nd AES Workshop on Intelligent Music Production. 2016.

[Wa16b] Waloschek, Simon; Berndt, Axel; Bohl, Benjamin W.; Hadjakos, Aristotelis: Interactive Scores in Classical Music Production. In: Proceedings of the 17th International Society for Music Information Retrieval Conference. 2016.