

Onbeat oder offbeat? Überlegungen zur symbolischen Darstellung von Musik am Beispiel der metrischen Quantisierung

Klaus Frieler¹ und Martin Pfeleiderer²

Abstract: Musik als psychologisches und kulturelles Phänomen stellt besondere Anforderungen sowohl an ihre symbolische Repräsentation als auch an ihre Analyse. Ausgehend von generellen Überlegungen zu Musik und Musikanalyse möchte der Beitrag am Beispiel der metrischen Quantisierung für die Problematik der symbolischen Musikrepräsentation sensibilisieren. Metrische Quantisierung ist nötig, denn musikalische Ereignisse werden in der Regel innerhalb eines zyklisch-metrischen Rahmens wahrgenommen und zu Zwecken der Analyse oder Reproduktion in diesem Raster dargestellt. Mitunter sind jedoch gerade jene rhythmischen Eigenheiten von Interesse, die tendenziell durch ein Quantisierungsrastrer fallen wie z. B. Mikrotiming oder Synkopierungen (Offbeat-Akzente), die für Groove- und Swing-Phänomene relevant sind. Angesichts dieser Problematik wird ein Quantisierungsalgorithmus (FlexQ) vorgestellt, der im Rahmen des Jazzomat Research Project entwickelt wurde. Sein Potenzial wird anhand von zwei Forschungsbeispielen (Synkopen/Offbeats und Swing-Ratio-Messungen) verdeutlicht.

Keywords: Musik, Musikanalyse, Computational Musicology, symbolische Musikrepräsentation, Quantisierung, Rhythmus, Metrum, Mikrotiming, Swing-Ratio, Groove, Jazz

1 Einleitung

In unserem Beitrag möchten wir Probleme bei der Darstellung und Analyse rhythmischer Informationen und deren metrische Quantisierung diskutieren sowie einen im Rahmen des Jazzomat Research Project gewählten Lösungsansatz vorstellen. Das Jazzomat Research Project widmet sich der computergestützten Analyse von Jazzimprovisationen auf der Grundlage eines großen Korpus von transkribierten Jazzsoli aus den 1920er bis 2000er Jahren. Im Rahmen des von der DFG von 2012 bis 2017 geförderten Forschungsprojektes wurde die Weimar Jazz Database (WJD) mit einem eigenen Datenformat und 456 qualitativ hochwertigen Transkriptionen sowie die Analysesoftware MeloSpySuite/GUI zur Extraktion von musikalischen Merkmalen und zur Patternsuche entwickelt; Datenbank und Software sind frei verfügbar und sollen anderen Forschern Perspektiven für eigene computergestützte Untersuchungen eröffnen.

¹ Hochschule für Musik Franz Liszt Weimar, Institut für Musikwissenschaft Weimar-Jena, Hochschulzentrum am Horn, Carl-Alexander-Platz 1, 99425 Weimar / Max-Planck-Institut für empirische Ästhetik, Frankfurt/M., kf@omniversum.de

² Hochschule für Musik Franz Liszt Weimar, Institut für Musikwissenschaft Weimar-Jena, Hochschulzentrum am Horn, Carl-Alexander-Platz 1, 99425 Weimar, martin.pfeleiderer@hfm-weimar.de

Mit unseren Überlegungen wollen wir exemplarisch in Besonderheiten des Forschungsfelds der computergestützten Musikanalyse einführen. Diese unterscheidet sich von einer auf digitalen Noteneditionen fokussierten Digital Musicology und einer vorwiegend auf der Audioebene arbeitenden Music Information Retrieval dadurch, dass sie Musik mit Hilfe computerlesbarer, vorwiegend symbolischer Musikformate und entsprechender Analysesoftware Musik untersucht. In unserem Beitrag möchten wir einige Erfordernisse und Besonderheiten von symbolischen Musikdatenformaten diskutieren, die nicht nur im Rahmen des Jazzomat Research Project, sondern generell für die computergestützte Musikforschung relevant sind. Beginnen möchten wir jedoch mit einigen generellen Überlegungen zu Musik und deren Analyse.

2 Computergestützte Musikanalyse

2.1 Was ist Musik?

Diese Frage mag überraschen, denn wir alle sind mit Musik vertraut, sind von Musik umgeben und genießen sie sowohl im Alltag als auch in außertäglichen Situationen wie Konzerten oder Festivals. Die vielleicht einfachste Antwort lautet: Musik ist das, was von Menschen als Musik wahrgenommen und so bezeichnet wird. Musik ist also zunächst ein psychologisches und soziales Phänomen, sie wird praktiziert, erlebt und in sozialen Diskursen verhandelt. Doch wie lässt sich der ontologische und epistemologische Status von Musik genauer fassen?

Der Musikphilosoph Daniel Feige hat vorgeschlagen, drei Seinsweisen von Musik zu unterscheiden: Erstens Musik als im Notentext fixiertes Werk, zweitens Musik als Aufführung oder Live-Darbietung und drittens Musik als Klangdokument [Fe17]. Natürlich hängen diese Seinsweisen eng miteinander zusammen: Musikwerke werden im Konzert aufgeführt und auf Tonträgern veröffentlicht. Umgekehrt beziehen sich Live-Darbietungen oftmals auf bereits existierende Musikwerke oder, wie im Pop und Rock, auf Tonträger, die sie in der performativen Situation interpretieren, also erweitern und verändern. Ungeachtet dieser Übergänge und Wechselbezüge, unterscheiden sich Musikwerke, Live-Darbietungen und Musikaufnahmen in ihrer Seinsweise jedoch grundlegend voneinander.

Während es für die Analyse von Musikwerken als Notentexten innerhalb der Musikwissenschaft und Musiktheorie eine lange Tradition und zahlreiche methodologische Ansätze gibt, sind Methoden für die Analyse von Tonträgern erst im 20. Jahrhundert entstanden, zunächst innerhalb der Musikethnologie, später dann in der Jazz- und Popmusikforschung. Am schwierigsten – aber womöglich auch am spannendsten – ist wohl die Analyse von Aufführungen, da in performativen Situationen weitere Aspekte wichtig sind: Die Interaktion der Musiker mit ihren Körpern und Instrumenten, mit ihren Mitmusikern, mit dem Raum und seiner Akustik sowie mit dem Publikum eröffnet zahlreiche situative Rückkopplungsmöglichkeiten, die vom Musiker spontane Reaktionsweisen

fordern und so zur Einmaligkeit und Ereignishaftigkeit von Musik und ihrem Erleben beitragen. Neben kleineren Gestaltungsfreiheiten wie der Wahl der Dynamik, des Tempos oder mikrorhythmischer Feinheiten, die auch bei der Interpretation klassischer Musik gegeben sind, bieten viele Musikrichtungen den Musikern zudem eine besondere Möglichkeit, Musik im Augenblick zu erschaffen: zu improvisieren. Dies geschieht natürlich nicht voraussetzungslos, sondern auf der Grundlage von zumeist langwierigen Lernprozessen und oftmals in Bezug auf Aufführungstraditionen, wie sie etwa im Jazz sowohl mündlich innerhalb der Musiker-Community und in Institutionen wie der Jam Session als auch medial durch Tondokumente und schließlich im Unterricht vermittelt werden. Die komplexen performativen Qualitäten und Gestaltungsspielräume gilt es im Hinterkopf zu behalten, wenn aus forschungspragmatischen Gründen von Musikdarbietungen nur Aufnahmen oder deren Transkriptionen Gegenstand der Analyse sind.

2.2 Musikalische Analyse

Analysiert man einen Gegenstand, so zergliedert man ihn, um anschließend seine Komponenten zu beschreiben und so mehr über seine inneren Zusammenhänge, seine Funktionsweise und seine besonderen Qualitäten zu erfahren. In den Musikwissenschaften steht dahinter oftmals die Annahme, dass durch ein vertieftes Verständnis von Musik auch deren ästhetische Erfahrung erweitert und bereichert werden kann. Darüber hinaus verfolgt die Musikanalyse den Zweck, die individuelle ästhetische Musikerfahrung explizit zu machen, um sie anderen Menschen vermitteln zu können.

In ihren Anfängen im 19. Jahrhundert folgte die Musikanalyse zunächst der Absicht, mit wissenschaftlichen Methoden den ästhetischen Wert von bestimmten Musikwerken sowie durch analytische Vergleiche verschiedener Epochen- oder Personalstile eine musikalische Stilgeschichte zu begründen. Oftmals geht die Stilanalyse Hand in Hand mit einer Rekonstruktion und Deutung der historisch und kulturell verankerten Gebrauchsweisen und Bedeutungen, die mit der jeweiligen Musik verbunden waren oder sind. Darüber hinaus kann Musikanalyse einer Beschreibung und Erklärung sowohl der musikalischen Wahrnehmung als auch der Herstellung von Musik durch Komposition, Aufführung (Performance) und Improvisation dienen, was nicht zuletzt Auswirkungen auf deren didaktische Vermittlung hat.

Mag es zum Zwecke einer ästhetischen Beurteilung des Wertes einzelner Musikwerke noch genügen, diese Werke isoliert zu analysieren, so ist es für stilgeschichtliche und vergleichende Untersuchungen wünschenswert, auf einer möglichst großen Datengrundlage zu arbeiten. Dies ist bei den herkömmlichen Analyseverfahren jedoch sehr mühselig und zeitaufwändig. Methoden der computergestützten Musikanalyse ermöglichen nun eine automatisierbare Analyse größerer Musikkorpora – so etwa seit den 1990er Jahren mit dem *Humdrum Toolkit* von David Huron und Mitarbeitern [Hu99] oder seit 2006 mit *music21* von Michael Cuthbert und Christopher Ariza [Cu10]. Der Musikhistoriker Nicholas Cook sah bereits im Jahr 2004 in diesen Entwicklungen “[...] a significant opportunity for disciplinary renewal: [...] there is potential for musicology to be pursued

as a more data-rich discipline than has generally been the case up to now, and this in turn entails a re-evaluation of the comparative method.” [Co04, S. 103] Cook erläutert seine Einschätzung anhand des Humdrum Toolkits und entsprechender Studien.

In der Tat sind seither zahlreiche computergestützte Korpusstudien durchgeführt und veröffentlicht worden, die im Gegensatz zum „close reading“ einzelner Werke dem Ansatz eines „distant reading“ größerer Korpora im Sinne des Literaturwissenschaftlers Franco Moretti [Mo09, Mo16] folgen.³ Diese Erneuerungspotenziale beziehen sich unseres Erachtens nicht nur auf eine Verbreiterung der Datengrundlage, sondern auch auf die Analyseterminologie. Musikforscher sind nämlich gezwungen, Begriffe etwa aus dem Bereich des Rhythmus (z. B. Synkope, Offbeat, Takt und Metrum) oder der Tonhöhen-gestaltung (Diatonik, Modalität etc.) so eindeutig und präzise zu definieren, dass sie von Musikinformatikern für Algorithmen operationalisiert werden können. Neben diesen terminologischen Fragen liegen Herausforderungen in den Besonderheiten von Musikdaten, in denen musikalische Gegebenheiten, die etwa mit dem zyklischen Aufbau von Metren oder den zyklischen Oktavwiederholungen in Tonsystemen zusammenhängen, angemessen zu repräsentieren sind.

2.3 Besonderheiten symbolischer Musikdaten

Obwohl es in den verschiedenen Kodierungsformen gewisse immer wiederkehrende Konzepte gibt, besteht bislang kein Konsens darüber, wie Musik optimal repräsentiert werden soll. Im Gegensatz zur normierten Darstellung von Texten durch vordefinierte Alphabete wurden verschiedene symbolische Musikformate entwickelt, die jeweils einige Aspekte in den Vordergrund stellen, andere dagegen vernachlässigen.

Audioformate (z.B. WAV, AIFF, MP3) erlauben es, Schallwellen zu repräsentieren. Sie bieten jedoch keine direkte Möglichkeit, musikalische Konzepte wie Melodie, Harmonie oder Rhythmus darzustellen. Noten dagegen kodieren explizit viele strukturelle Merkmale (Tonhöhen, Rhythmus, Metrum), und erlauben so Musikern, unter Rückgriff auf gelernte expressive Aufführungskompetenzen, ein Musikstück darzubieten. Sie enthalten jedoch in der Regel nur marginale Angaben zu expressiven Aspekten wie Dynamik, Klangfarbe und Mikrorhythmik. Das MIDI-Format ist zwar expressiver als ein Notentext, da es erlaubt, aufführungswichtige Momente wie Mikrotiming und Lautstärke darzustellen, hat aber nur eingeschränkte Möglichkeiten, abstraktere musikanalytische Aspekte, wie etwa Harmonien, Phrasenbau, Form usw. adäquat darzustellen.

Welches symbolische Musikrepräsentationssystem in welcher Situation zum Einsatz kommt, hängt stark von den dabei verfolgten Intentionen ab. MusicXML wurde primär zu Darstellung von Noten entwickelt, MIDI zu Echtzeitsteuerung von elektronischen Musikinstrumenten. Man kann symbolische Musikrepräsentationssysteme auch in deskriptive und präskriptive Formate unterteilen. Deskriptive Formate dienen der beschrei-

³ Vgl. etwa die entsprechenden Themenhefte der Zeitschrift *Music Perception*, 31/1 (2013) und 31/3 (2014).

benden Darstellung von Musik und daher vorwiegend analytischen Zwecken, während präskriptive Formate primär zur (Re-)Produktion von Musik gedacht sind. Allerdings sind präskriptive Notationsweisen in der Regel bereits ausdrucksstark genug, um auch zur Beschreibung bestimmter musikalischer Dimensionen wie der Melodik oder Harmonik dienen zu können. Der amerikanische Musikwissenschaftler Charles Seeger, der die Unterscheidung zwischen präskriptiver und deskriptiver Notation 1958 vorgeschlagen hat [See58], favorisiert zum Zwecke der musikalischen Analyse eine Kombination verschiedener grafischer Repräsentationsweisen, etwa der herkömmlichen Notenschrift und einer Art Pianorollennotation sowie Spektraldarstellungen und Intensitätskurven. Nur durch diese Kombination könne es, so Seeger, gelingen, möglichst viele der musikalischen Dimensionen, die für die musikalische Analyse von Interesse sind oder sein können, adäquat zu notieren. Seeger erkannte zugleich, dass Audio-Formate (wie die damaligen Schallplattenrillen) zwar einen maximalen Informationsgehalt besitzen, jedoch nicht direkt für die Analyse geeignet sind.

Computergestützte Musikanalyse ist primär an deskriptiven Musikrepräsentationssystemen in digitaler Form interessiert. Die entscheidende Frage lautet dabei, welche musikalischen Dimensionen auf welche Weise computerlesbar dargestellt und gespeichert werden. Da man nicht vorhersehen kann, welche musikanalytischen Fragestellungen sich in Zukunft ergeben werden, ist bei der Konzeption von Datenformaten prinzipiell ein maximaler Informationsgehalt anzustreben. Das würde dafür sprechen, von audiobasierten Repräsentationen auszugehen, und weite Teile des Music Information Retrievals folgen tatsächlich diesem Ansatz. Audiodaten besitzen zwar einen maximalen Informationsgehalt, sind aber, wie schon Charles Seeger erkannte, analytisch sehr schwach, da man jede Merkmalsdimension, z. B. Tempo, Metrum, Tonhöhe, Melodie, Harmonie, Instrumentation, Form etc., erst aus der Audiodatei extrahieren muss. Diese Aufgabe ist jedoch für die meisten musikalischen Dimensionen bislang nicht mit hinreichender Präzision gelöst. Zwar kann man für ausgewählte analytische Fragestellungen mit den heute vorhandenen Algorithmen zu sinnvollen Antworten kommen, allerdings nur mit Hilfe von statistischen Verfahren und großen Datensätzen. Wenn man etwa die Tempoverteilung in elektronischer Tanzmusik untersuchen möchte (vgl. [Moe02]), kann man mit einem State-of-the-Art-Tempo-Extraktionsverfahren und einem hinreichend großen Korpus zu durchaus validen Ergebnissen kommen. Für andere Fragestellungen, etwa zur mikrorhythmischen Gestaltung im Jazz (vgl. Abschnitt 3.3), ist dies aber nur eingeschränkt möglich, denn dazu bräuchte man nicht nur eine sehr gute Grundschlagsermittlung (Beattracking) im Jazz (wofür bislang kein tauglicher Algorithmus entwickelt wurde), sondern auch eine möglichst fehlerfreie und zeitlich exakte Extraktion des Spiels der Solisten aus einer polyphonen Jazzaufnahme.

Womöglich ist es nur eine Frage der Zeit, bis MIR-Algorithmen gut genug sind, um die genannten Extraktionsaufgaben mit hinreichender Präzision zu lösen. Dennoch bleibt dann die Frage offen, wie der Output dieser Algorithmen repräsentiert werden soll. Bislang wird sowohl das konkrete Ausgabeformat (Text/CSV, XML, JSON etc.) als auch das ihm zugrunde liegende Datenmodell mehr oder minder *ad hoc* von den Entwicklern

der Algorithmen definiert. Die seit einigen Jahren stattfindenden MIREX-Wettbewerbe⁴ wie auch die Entwicklung von Musikontologien [Ra12] habe hier zwar zu einer Standardisierung zumindest für bestimmte Aufgaben beigetragen. Dennoch ist kein allgemein verbindlicher Standard in Sicht. Wichtiger als das konkrete Format ist allerdings die Frage nach einer adäquaten Darstellung der grundlegenden Dimensionen von Musik – adäquat sowohl hinsichtlich des Wahrnehmens und Erlebens von Musik als auch der musikalischen Intentionen von Musikern und Komponisten.

Wir wollen diese Überlegungen an einem Aspekt genauer erläutern, der zunächst simpel erscheint und auch Laien intuitiv zugänglich ist, aber doch viele interessante Fragen und Probleme aufwirft: der metrischen Quantisierung des musikalischen Rhythmus.

3 Metrum, Quantisierung, Synkopen, Swing-Ratio

3.1 Metrum und Quantisierung

Musik ist in der Regel metrisch gebunden. Metrum hat zwei Komponenten, zum einen den Beat oder Grundschatlag, zum anderen den Takt. Die meisten Menschen sind in der Lage, den Grundschatlag (soweit vorhanden) in der Musik wahrzunehmen und mental so fortzuschreiben, dass sie sich zur Musik durch Mitklopfen, Mitwippen oder Tanzen synchronisieren können. Der Grundschatlag kann als eine Folge von Zeitpunkten, mit (mehr oder minder) konstanten Abstand (dem Inter-Beat-Intervall) verstanden werden. Das Tempo eines Musikstücks ergibt sich aus dem Kehrwert des zeitlichen Abstandes zwischen zwei Grundschatlägen und wird in Beats per Minute (bpm) gemessen. 120 bpm bedeuten also zwei Grundschatläge pro Sekunde (2 Hz), was einem Inter-Beat-Intervall von 500 ms entspricht.

Der Takt kann als eine Gruppierung von Grundschatlägen verstanden werden, die zu einer Einheit zusammengefasst werden. Oft sind weder Grundschatlag noch Takt explizit in der Musik vorhanden, sondern werden vielmehr beim Hörer (und bei den Musikern) aufgrund statistischer Regelmäßigkeiten im Schallsignal (bottom up) sowie schematisierter Erwartungen (top down) induziert [Fr82]. Wahrnehmungspsychologische Untersuchungen haben gezeigt, dass Menschen einen einmal induzierten Grundschatlag und Takt auch dann weiterführen, wenn die entsprechenden Klangereignisse für eine Zeit lang aussetzen [Re05, Se00]. Das in vielen Musikgenres am weitesten verbreitete Metrum ist der 4/4-Takt, d. h. vier Beats stellen eine Takteinheit dar. Zwischen den Schlägen eines 4/4-Taktes gibt es in der Regel ein phänomenologisches Ungleichgewicht, so werden die erste und die dritte Zählzeit als „wichtiger“ oder „betonter“ wahrgenommen (Akzentstufentakt). Durch die Gruppierung von Grundschatlägen zu noch größeren Einheiten entsteht zudem eine Art metrische Hierarchie. Eine (regelmäßige) Gruppierung von Takten in größere Einheiten (z. B. zwei, vier oder acht Takte) wird als Hypermetrum bezeichnet

⁴ http://www.music-ir.org/mirex/wiki/MIREX_HOME

und ist in der westlichen Musik weit verbreitet. Die metrische Hierarchie wird in der Regel durch die Teilung der Inter-Beat-Intervalle in Subbeat-Divisionen auch nach unten ausgedehnt werden. So kann ein Inter-Beat-Intervall in gleiche Hälften, aber auch in drei Teile (Triolen) oder in ungleiche Hälften wie im Jazz (Swing-Achtel, vgl. 3.5) geteilt werden. Diese Untereinheiten können zudem weiter unterteilt werden. Wie baut sich nun aber ganz konkret die Hierarchie unterhalb der Grundschiebeebene auf, wenn es unendlich viele, unendliche feine Möglichkeiten der Unterteilung gibt?

Das Problem der Subbeat-Division zeigt sich ganz praktisch bei der Quantisierung. Der Begriff „Quantisierung“ bezeichnet in der MIDI-Technik den Vorgang, ein von Menschen eingespieltes Musikstück, rhythmisch zu „begradigen“, d. h. in ein metrisches Gitter einzufügen. Bekanntlich zeichnen sich Menschen in der Regel durch ein schwankendes Timing aus, das oft mit vom Sequenzer gesteuerten Schlagzeugpatterns oder Keyboard-Arpeggien schlecht harmoniert oder einfach als „untight“ (unpräzise) empfunden wird. Deswegen gibt es Optionen, manuell eingespielte MIDI-Tracks zu quantisieren, d. h. die Einsatzzeitpunkte algorithmisch auf ein vorgegebenes Raster ideeller Zeitpunkte zu verschieben, z. B. auf ein 16tel-Raster, wie es viele frühe Drumcomputer benutzen. Der verwendete Algorithmus ist recht einfach. Bei gegebener Gitterkonstante, also einer bestimmten Unterteilung des Beats, kann man für jedes Ereignis arithmetisch den nächstliegenden Gitterpunkt bestimmen. Die Wahl der Gitterkonstante wird dabei in der Regel dem Nutzer überlassen. Da ein zu feines Gitter keinen hörbaren Quantisierungseffekt hat und ein zu grobes Raster manchmal zu „unmusikalischen“ oder zu falschen Resultaten führt, gibt es allerdings bestimmte Heuristiken. Will man etwa sowohl 32tel-Noten als auch 16tel-Triolen darstellen, was für Popmusik hinreichend genau ist, so braucht man mindestens 24 Subbeat-Divisionen, da acht 32tel- und sechs 16tel-Triolen in eine Viertelnote passen, und das kleinste gemeinsame Vielfache davon $\text{kgV}(8, 6) = 24$ ist. Bei einem zu groben Raster können zeitlich separate Ereignisse unter Umständen auf demselben Gitterpunkt landen, mit der Folge, dass eine monophone Melodie mehrstimmig wird. Wer schon einmal eine manuell eingespielte MIDI-Datei in einem der gängigen Notensatzprogramme geöffnet hat, wird dieses Problem kennen. Zwar sind feinere Raster grundsätzlich realitätsnäher als grobe, allerdings benötigt man zumindest für Notendarstellungen eine definierte kleinste Unterteilung des Beats. Auch für die Transkriptionen von Jazzsoli ist daher eine Quantisierung erforderlich. Anders als bei der MIDI-Quantisierung impliziert der im Jazzomat Research Project gewählte Weg jedoch keine reale Verschiebung der Einsatzzeitpunkte, sondern ist eine metrische Annotation, d. h. eine Zuordnung von Tonereignissen zu metrischen Gitterpunkten. Aufgrund der oben genannten Mängel der einfachen Gitterquantisierung wurde dafür ein neuer Algorithmus entworfen.

3.2 Flexible Quantisierung (FlexQ)

Die Transkriptionen der Weimar Jazz Database bestehen aus Sonic-Visualiser-Projektdateien, die als zentrales Element einen Noten-Layer sowie verschiedenen Annotationslayer enthalten, vor allem einen manuell geklopften Grundschiebe mit Angaben

zum Metrum (Taktart, Lage der Taktanfänge) und daran gekoppelten Annotationen der Formteile und Akkorde. Die Aufgabe des Algorithmus besteht nun darin, den Tonereignissen mit Hilfe des Grundschlags und der Taktangaben eine vollständige metrische Annotation zuzuordnen. Die Grundschlagannotation liefert dabei zusammen mit den Taktangaben Taktnummern und Beatpositionen, sodass nur noch die Subbeat-Positionen zu bestimmen sind.

Angesichts der geschilderten Problematik einer zu starren Gitterquantisierung wurde ein einfacher Gedanke zum Ausgangspunkt genommen, der auch in der herkömmlichen westlichen Notenschrift benutzt wird: Finde für jeden Beat anhand der gerade erklingenden Musik auf lokaler Ebene eine optimale Unterteilung. Damit ergibt sich ein Optimierungsproblem, das durch eine Rastersuche über mögliche Unterteilungen zu lösen ist, wobei als Randbedingung zusätzlich die strikte Einstimmigkeit einzuhalten ist, also einem bestimmten Gitterpunkt maximal ein Ton zugeordnet werden darf. Dazu wurden vier heuristische Bedingungen aufgestellt, die es gleichzeitig zu erfüllen gilt.

1. Die lokalen Subbeat-Gitter sollen möglichst einfach sein, d. h. die Unterteilung soll eine möglichst kleine Zahl sein („keine Unterteilung“ ist dabei als Sonderfall erlaubt).
2. Der mittlere Abstand der realen Einsatzzeitpunkte von den quantisierten Gitterpunkten soll möglichst gering ist.
3. Abgesehen von 2er- und 3er-Teilung sind Primzahlunterteilungen in der Musik eher selten anzutreffen. Unterteilungen in 5, 7, 11, 13 etc. Gitterpunkte sind somit tendenziell zu vermeiden.
4. Die Quantisierungsverschiebungen sollen möglichst gleichmäßig sein, was man als möglichst geringe Standardabweichung der Verschiebungen operationalisieren kann. Der Hintergrund dieser Regel ist, dass man manchmal zwar einen niedrigen Quantisierungsfehler durch wenige sehr große und wenige sehr kleine Verschiebungen erreicht, man aber annehmen kann, dass ein Spieler eher gleichmäßig verteilte Abweichungen von idealen Unterteilungen produziert und nicht einige Töne präzise und andere sehr unpräzise spielt.

Der FlexQ-Algorithmus iteriert über Paare aufeinander folgender Grundschläge, wobei für die Menge aller Einsatzpunkte aller Tonerereignisse innerhalb dieses Grundschlagpaares die optimale Unterteilung gesucht wird. Dabei werden alle Gitterunterteilungen bis zu einer gewissen Obergrenze getestet. Die Obergrenze der Unterteilung ergibt sich aus der maximalen Geschwindigkeit menschlicher Bewegungsabläufe bzw. der wahrnehmbaren Ereignistrennschärfe, die beide in der Größenordnung von ca. 20–30 ms liegen. Die gefundene minimale Unterteilung wird dann zur metrischen Annotation genutzt. Die auf diese Weise generierten Ergebnisse sind durchaus zufriedenstellend und sowohl für detaillierte Analysen als auch für Notendarstellungen geeignet.

3.3 Beispiel 1: Jazzomat vs. externe Transkription

In der Weimar Jazz Database ist ein kurzes Solo von Charlie Parker über das Stück „Ornithology“ (1946, Dial 1002) enthalten. Eine Transkription davon findet sich u. a. im sogenannten „Omnibook“ [Pa87], einer Sammlung der Transkriptionen von 56 Soli Charlie Parkers, die vor allem für praktizierende Jazzmusiker gedacht sind. In Abb. 1A findet sich die Notendarstellung der Takte 5-7 des Solos aus dem Omnibook, in Abb. 1B die entsprechende Stelle aus der Weimar Jazz Database (WJD), die mit dem FlexQ-Algorithmus quantisiert wurde, in Abb. 1C ein Screenshot der entsprechenden Passage in der Sonic-Visualiser-Projektdatei. Wie man sieht, weichen die Notendarstellungen der beiden Transkriptionen deutlich voneinander ab. Die Tonhöhen stimmen im Wesentlichen überein, es finden sich aber drei Unterschiede (in Takt 6 auf der 1+ und 4, in Takt 7 auf der 2+). Der Vergleich mit der Aufnahme zeigt, dass die Töne in der WJD richtig sind.

Was die metrische Annotation angeht, gibt es bei insgesamt 20 Tönen nur sechs Übereinstimmungen zwischen Omnibook- und WJD-Version. Die Übereinstimmungen sind alle auf vollen Zählzeiten (Onbeat) zu finden, die Abweichungen finden sich alle in der Subbeat-Ebene (Offbeats). Geht man davon aus, dass die Grundschräge und Töne hinreichend korrekt annotiert sind, was der Ohrenschein bestätigt, so ist die metrische Annotation der WJD Version in der Tat präziser. Das kann man gut am zweiten Ton illustrieren (c^ˆ), der deutlich nach der zweiten Zählzeit erklingt und deswegen auf der zweiten 16tel im zweiten Beat zu liegen kommt. Die folgenden drei Töne sind bezogen auf den Beat alle etwas zu spät, die sich anschließenden zwei Töne erklingen dann etwas früher, so dass auf der ersten Zählzeit in Takt 6 beide Transkriptionen wieder übereinstimmen.

Auf der ersten und der dritten Zählzeit des Taktes 6 sowie auf der dritten Zählzeit von Takt 7 finden sich zudem typische Beispiele für den Einfluss von Konventionen auf die (präskriptive) Notation. Während in der Omnibook-Version gerade (binäre) Achtel notiert sind, annotiert FlexQ triolische Swing-Achtelnoten (lang-kurz). In der Tat ist es bei Jazztranskriptionen verbreitet, Swing-Achtel nicht triolisch, sondern binär zu schreiben, um die Notation zu vereinfachen, wobei impliziert wird, dass die Achteln triolisch (lang-kurz) gespielt werden. Auch FlexQ folgt dieser Konvention, wenn die Swing-Ratio, also das Verhältnis der längeren ersten zur kürzeren zweiten Achtel, näher an 1:1 als an 2:1 liegt (vgl. Abschnitt 3.4). Obwohl die metrische Annotation der WJD-Version in diesem Fall exakter als die präskriptiv gedachte Transkription des Omnibooks ist, so kann man doch behaupten, dass in gewisser Hinsicht die Rhythmik der Omnibook-Transkription besser ist, denn sie ist nicht nur einfacher (und damit leichter lesbar und zum Nachspielen besser geeignet), sondern entspricht wahrscheinlich eher sowohl der Intention Parkers als auch dem, wie Hörer diese Stelle wahrnehmen.

Abb. 1. Verschiedene Darstellungen der Takte 5-7 aus Charlie Parkers Solo über „Ornithology“ (1946): A. Notendarstellung auf Basis der Transkription aus dem Omnibook (eigene Reproduktion). B. Notendarstellung mit dem FlexQ-Algorithmus auf Basis der Weimar Jazz Database. C. Screenshot der Sonic-Visualiser-Projektdatei; die horizontalen Balken sind die annotierten Notenergebnisse, die vertikalen Striche die manuell annotierten Zählzeiten.

3.4 Beispiel 2: Synkopen und metrische Komplexität

Unter einer Synkope verstehen wir ein Klangereignis auf einer Subbeat-Ebene, dem kein Ereignis auf der nächsthöheren Ebene folgt [Pf06]. In der Regel handelt es sich bei der nächsthöheren Ebene um die Ebene des Grundschlags, sodass der Ton vor dem (fehlenden) Grundschlag oder „Onbeat“ auch als „Offbeat“ bezeichnet wird. Die zweite Achtelnote auf der zweiten Zählzeit in Takt 5 in der Omnibook-Version (s. Abb. 1A) ist demnach keine Synkope, da ihr ein Onbeat-Ereignis auf der dritten Zählzeit (der nächsthöheren metrischen Ebene) folgt. In der WJD-Darstellung findet sich dieser Ton aber auf der vierten 16tel des Beats und das nächste Ereignis erklingt auf der zweiten 16tel der dritten Zählzeit, damit handelt sich um eine Synkope. Da der zweite Ton ebenfalls eine Synkope ist, hat man hier insgesamt eine Kette von vier 16tel-Synkopen. In der Tat findet sich in der Omnibook-Version im ganzen Ausschnitt nur eine einzige Synkope (auf dem letzten Ton), wohingegen in der WJD-Version insgesamt acht Synkopen zu finden sind – ein erheblicher Unterschied.

Generell ist Synkopizität ein wichtiges Phänomen in Jazz und populärer Musik. Synkopizität arbeitet gewissermaßen einem etablierten Metrum entgegen, da Zeitpunkte betont werden, die nicht mit denen der zentralen metrischen Ebene des Grundschlags übereinstimmen. Nehmen sie nicht überhand und etablieren ein alternatives Metrum, dann trägt der geschickte Einsatz von Synkopen in der Regel zum Bewegungskarakter der Musik und mitunter zu deren „Groove“-Charakter entscheidend bei. Daher ist es durchaus von Interesse, den Anteil von Synkopen in Jazzsoli zu bestimmen. Zugleich ist der Anteil an Synkopen ein Maß für die metrische Komplexität einer Melodie.

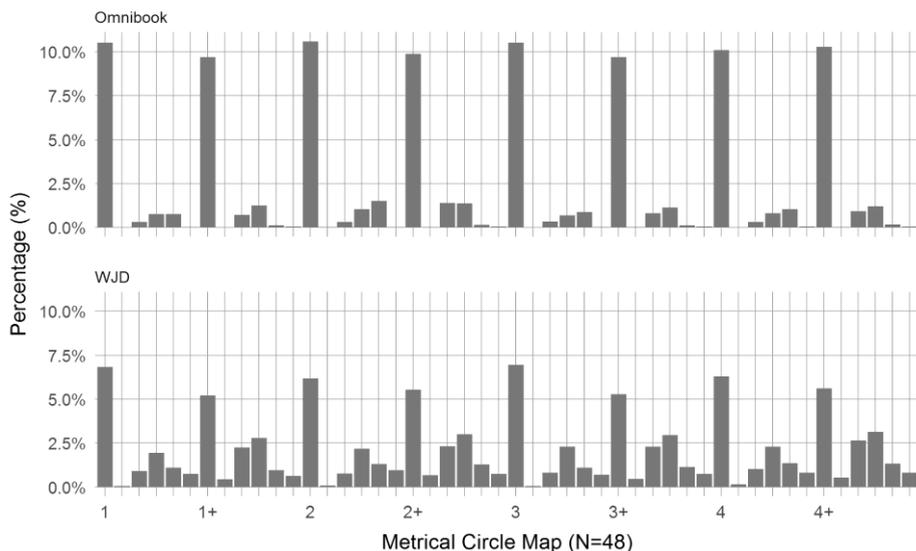


Abb. 2. Verteilung der MCM48 (Frieler, 2007) für das Omnibook-Korpus (N=56) und das Charlie-Parker-Subkorpus (N=17) der WJD. In beiden Korpora sieht man eine recht gleichmäßige Verteilung der Achtelpositionen im 4/4-Takt (alle Soli sind 4/4).

Die metrische Annotation in der WJD erlaubt es, für jeden Ton recht einfach zu entscheiden, ob er eine Synkope ist oder nicht. Der Anteil der Synkopen (*syncopicity*) innerhalb eines Stückes oder eines Musikrepertoires ist ein mögliches Maß für dessen metrische Komplexität. Auf der Basis der FlexQ-Quantisierung kann man weitere einfache Indikatoren für metrische Komplexität definieren. Einem der Indikatoren liegt die Überlegung zu Grunde, dass eine möglichst gleichmäßige Unterteilung von Inter-Beat-Intervallen einfacher ist als ständig wechselnde die Grundsclagunterteilungen. Das kann als mittlere Anzahl von Unterteilungsänderungen (*metric division complexity*) gemessen werden. Ein weiterer Indikator für metrische Komplexität beruht auf der Idee, dass vollbesetzte lokale Subbeat-Gitter einfacher sind als solche mit Lücken. So ist die 16tel-Notenkette `|xxxx|xxxx|xxxx|xxxx|` im 4/4-Takt einfacher, da redundanter als ein komplexer Rhythmus wie etwa `|xoox|ooxo|oxoo|xoox|` oder auch `|xox|xox|xox|xox|`. Wenn man den Anteil unbesetzter Gitterpunkte berechnet, erhält man die sog. *metric compression complexity*. Ein viertes Maß für metrische Komplexität kann über die metrische Kreisabbildung (Metrical Circle Map, MCM, Frieler 2007) definiert werden. Dabei wird jeder Takt in N gleiche Teile unterteilt. In Abb. 2 findet man die MCM (mit N=48) für das Omnibook-Korpus und das Charlie-Parker-Subkorpus der WJD. Man sieht dort sehr gut, wie die flexible Quantisierung in der WJD zu einer flacheren Verteilung führt. Die (Sample-)Entropie der MCM-Verteilung lässt sich ebenfalls als Komplexitätsmaß auffassen.

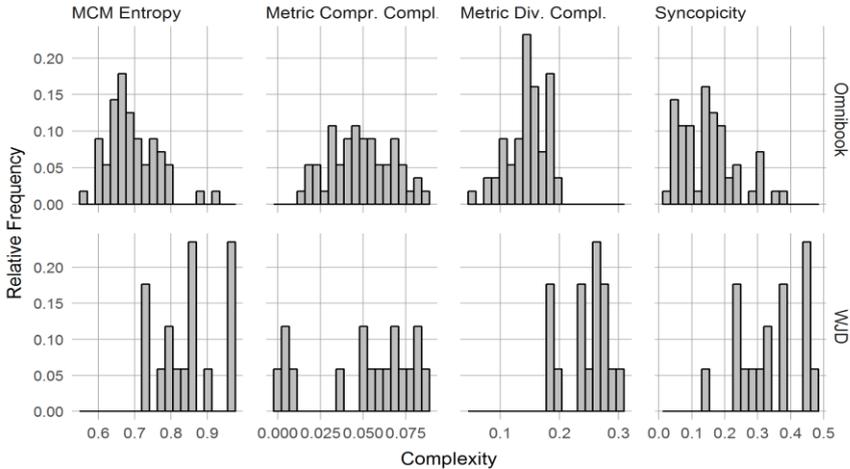


Abb. 3. Vier metrische Komplexitätsmaße für das Omnibook-Korpus (N=46) und das Charlie-Parker-Subkorpus (N=17) der WJD. Für eine Erklärung der Maße s. Text.

In Abb. 3 findet man die Verteilung der vier genannten metrischen Komplexitätsmaße für das Omnibook-Korpus und das Charlie-Parker-Subkorpus der WJD. Die Werte für das Omnibook sind generell niedriger, einzige Ausnahme bildet die *metric compression complexity*, da diese nicht so sehr von der Unterteilung der Beats abhängt. Der Mittelwert der Synkopizität liegt für das Omnibook bei 15%, und für die Parker-Soli der WJD bei 34%, ist also mehr als doppelt so hoch.

3.5 Beispiel 3: Swing Ratio.

Als ein Spezifikum des Jazz gilt das Phänomen des „swing“, das oftmals mit dem fließenden Bewegungscharakter und der Häufigkeit und Platzierung von Synkopen, vor allem jedoch mit den Swing-Achteln (lang-kurz) in Zusammenhang gebracht wird. Das Verhältnis von langer zu kurzer Achtelnote nennt man die Swing-Ratio. Verschiedene Messungen haben gezeigt, dass es bei verschiedenen Spielern und bei verschiedenen Tempi eine relativ große Bandbreite der Swing-Ratio gibt [Pfl16, S. 262-271]. Anders als in herkömmlichen Transkriptionen ermöglichen es die Zeitinformationen in der WJD, die Swing-Ratios genau zu messen und damit differenziert zu analysieren. Dazu müssen zunächst alle Achtelketten identifiziert werden, die aufgrund des FlexQ-Algorithmus auch als Triolen oder mit noch kleineren Notenwerten annotiert sein können. Dieser Umstand wird bei der Berechnung der Swing-Ratios mit Hilfe der Analysesoftware MeloSpyGUI berücksichtigt. In Abb. 4 finden sich Boxplots der Swing-Ratio-Verteilung in den Charlie-Parker-Soli der WJD: Man sieht eine recht hohe Streuung der Werte innerhalb der einzelnen Soli wie auch über alle Soli hinweg. Der Median der Swing-Ratio über alle Soli liegt bei 1.53 : 1 bzw. ca. 3 : 2, also gerade in der Mitte zwischen binären (1 : 1) und triolischen (2 : 1) Swing-Achteln.

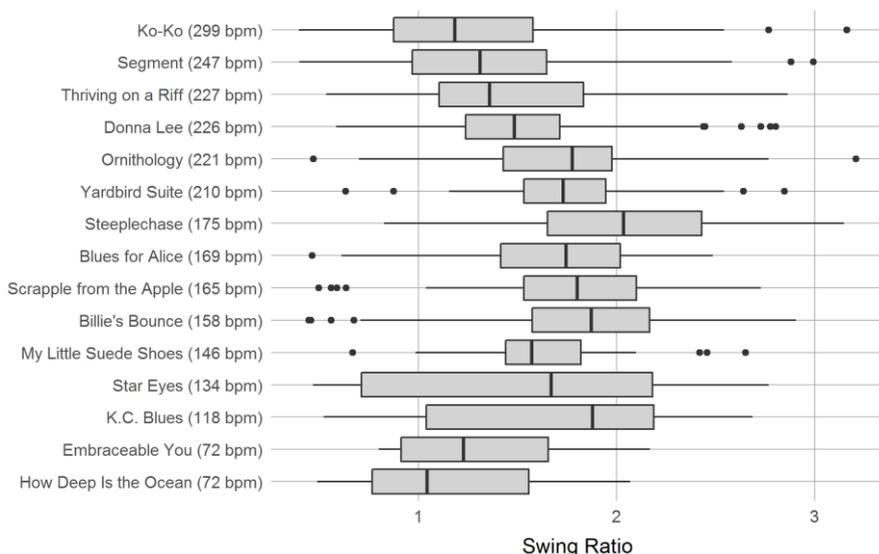


Abb. 4. Boxplots der Swing-Ratios der Charlie-Parker-Soli in der WJD. Die Soli sind nach Tempo (in Klammern hinter dem Titel) absteigend angeordnet. Zwei Balladen hatten nur jeweils einen Wert und wurden weggelassen.

4 Fazit und Ausblick

Eine zentrale Herausforderung der computergestützten Musikanalyse ist es, adäquate Repräsentationsformen, Datenformate und Datenkonzepte für Musik zu entwickeln, die einen hohen Informationsgehalt besitzen und zugleich hinreichend flexibel sind, um die Grundlage für die Untersuchung verschiedener analytischer Fragestellungen zu bilden. Dies betrifft ebenso die automatisierten Audio-Extraktions- und Transkriptionsverfahren des Music Information Retrieval. Die herkömmliche Notenschrift ist hier nur ein erster Anlaufpunkt, der wohl für viele strukturelle Aspekte der Musik – und insbesondere für Musikwerke der europäischen Tradition – bereits ausreichend ist. Allerdings lassen sich viele expressive und performative Aspekte der Musik, wie z. B. Swing und Groove, aber auch Vibrato, Tonhöhengleiten, Dynamik, Klangegebung usw., mit herkömmlichen Darstellungsformen kaum oder gar nicht adäquat erfassen.

Musik ist ein mehrdimensionales Phänomen. Aufgrund ihrer verschiedenen Dimensionen sind beim Erleben von Musik je nach Aufmerksamkeitsfokus mehrere (aber nicht beliebige) Wahrnehmungsweisen möglich. So kann sich beispielsweise das Interesse und die Aufmerksamkeit der Hörer auf strukturelle Aspekte der rhythmisch-metrischen Gestaltung oder aber auf den Bewegungscharakter, der durch Synkopen oder mikrorhythmische Details erzeugt wird, richten – oder auf beides zugleich. Die Mehrdimensionali-

tät von Musik hat auch Auswirkungen auf die Güte der symbolischen Repräsentationsformen, die sich nicht an einer einzig gültigen Ground Truth ausrichten lässt, sondern vielmehr von Kriterien wie Reichhaltigkeit, Informationsgehalt und Flexibilität abhängt. Konkret: Dass sich der mit FlexQ quantisierte Notentext bisweilen schlechter lesen und nachspielen lässt, ist kein Argument gegen die Validität dieser Quantisierungsform. Vielmehr spricht für die Güte des Algorithmus, dass sich auf der Grundlage der von ihm generierten Repräsentationsformen mehrere unterschiedliche analytische Fragestellungen sinnvoll untersuchen lassen – was für die Omnibook-Transkriptionen nur bedingt zutrifft, da sie weder eine angemessene Bestimmung der Synkopizität und metrische Komplexität von Charlie-Parker-Soli zulassen noch Informationen über deren mikro-rhythmische Gestaltung bieten. Ziel ist somit die Entwicklung eines flexiblen mehrschichtigen Repräsentationssystems, das möglichst viele expressiv-performative und syntaktisch-strukturelle Dimensionen der Musik erfasst und einer Analyse zugänglich macht. Neben der hier thematisierten rhythmisch-metrischen Dimension gibt es weitere musikalische Dimensionen, bei denen die Entwicklung computerlesbarer Repräsentationsformen noch in der Kinderschuhen steckt, so etwa die klangliche Gestaltung oder die interaktiven Aspekte einer musikalischen Darbietung. So stellen etwa Beobachtungen zur rhythmisch-metrischen Gestaltung von R&B-Grooves das herkömmliche Konzept des Grundschlags als einer Folge von *Zeitpunkten* in Frage; vielmehr gibt es gute Argumente dafür, die einzelnen Grundschläge als „bins“ mit einer gewissen zeitlichen Ausdehnung zu konzipieren [Da10].

Eines der Ziele von Musikanalyse ist die Erweiterung des Verständnisses von Musik – und zwar aller Dimensionen, die für das ästhetische Erleben von Musik relevant sind. Dies schließt auch Erfahrungen mit ein, die in der herkömmlichen, auf europäische Musikwerke ausgerichteten musiktheoretischen Terminologie bislang stark vernachlässigt wurden, wie etwa der spezifische Bewegungscharakter im Jazz und anderen populären Musikstilen, die mit Begriffen wie „fließend“, „drive“ oder „swing“ nur sehr vage beschrieben werden. Dies hat durchaus Konsequenzen für die Vermittlung von Musik – man denke nur an die pädagogische Vermittlung einer swingenden Spielweise, die lange als Spielen von triolischen Achtelketten (Swing-Ratio 2 : 1) gelehrt wurde, obschon die Bandbreite der Swing-Ratios, wie sie von Jazzmusikern tatsächlich gespielt wird, viel größer ist und vermutlich von mehreren Faktoren abhängt [Bu02].

Durch eine engere Zusammenarbeit zwischen Musikforschern und Musikinformatikern sollten einerseits Algorithmen an bestimmte Forschungsfragen und die Eigenheiten der untersuchten Musik angepasst werden, andererseits jedoch umfassende Repräsentationssysteme von Musik entwickelt werden, die eine computergestützte Analyse möglichst vieler Aspekte und Dimensionen von Musik erlauben. Voraussetzung für diese Kooperation ist es, dass sich Musikinformatiker mit Konzepten und Begriffen der Musikforschung auseinandersetzen. Um eine gemeinsame Sprache zu finden, müssen sich jedoch zugleich die Musikforscher mit Konzepten und Begriffen der Informatik und Statistik, der musikalischen Akustik und digitalen Signalverarbeitung vertraut machen.

Literaturverzeichnis

- [Bu02] Busse, W. G.: Toward Objective Measurement and Evaluation of Jazz Piano Performance Via MIDI-Based Groove Quantize Templates. *Music Perception*, 19, S. 443-461, 2002.
- [Co04] Cook, N.: Computational and Comparative Musicology. In (Clarke, E.; Cook, N. Hrsg.) *Empirical Musicology. Aims, Methods, Prospects*, Oxford, S. 103–126, 2004.
- [Cu10] Cuthbert, M.S., Ariza, C.: music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data, in: *Proceedings of the International Symposium on Music Information Retrieval 11 (ISMIR11)*, S. 637–642, 2010.
- [Da10] Danielsen, A.: Here, There and Everywhere: Three Accounts of Pulse in D’Angelo’s ‘Left and Right’. In (Danielsen, A. Hrsg): *Musical Rhythm in the Age of Digital Reproduction*, Farnham, S. 19-35, 2010.
- [Fe17] Feige, D.: Zur Ontologie der Popmusik. In: *Musik & Ästhetik*, Bd. 21, S. 40-54, 2017.
- [Fr82] Fraise, P.: Rhythm and Tempo, In (Deutsch, D. Hrsg.): *The Psychology of Music*. Academic Press, New York, S. 149–180, 1982.
- [Hu99] Huron, D.: *The Humdrum Toolkit: Software for Music Research*, 1999.
- [Mo02] Moelants, D.: Preferred tempo reconsidered. In (Stevens, C., Burnham, D., McPherson, G.E., Schubert, E., Renwick, J. Hrsg.): *Proceedings of the 7th International Conference for Music Perception and Cognition*. Presented, Sidney, S. 580–583, 2002.
- [Mo09] Moretti, F.: *Kurven, Karten, Stammbäume. Abstrakte Modelle für die Literaturgeschichte*. Suhrkamp, Frankfurt 2009.
- [Mo16] Moretti, F.: *Distant Reading*. Konstanz University Press, Konstanz 2016.
- [Pa87] Parker, C.: *Charlie Parker Omnibook: For C Instruments (Treble Clef)*. Atlantic Music, Hollywood, 1987.
- [Pf06] Pfeleiderer, M.: *Rhythmus. Psychologische, theoretische und stilanalytische Aspekte populärer Musik*. Transcript, Bielefeld, 2006.
- [Ra12] Raimond, Y., Sandler, M.: Evaluation of the Music Ontology Framework. In: *Proceedings of the 9th International Conference on The Semantic Web: Research and Applications, ESWC’12*. Springer-Verlag, Berlin, Heidelberg, S. 255–269. 2012.
- [Re05] Repp, B.: Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review* 12, 969–992, 2005.
- [Se00] Semjen, A., Schulze, H.H., Vorberg, D.: Timing precision in continuation and synchronization tapping. *Psychol Res* 63, 137–147, 2000.
- [Se58] Seeger, C.: Prescriptive and descriptive music writing. *The Musical Quarterly*, 44, 2, S. 184-195, 1958