

Zu den Anforderungen einer musikalischen Stilometrie

Kepper, Johannes

kepper@edirom.de

Universität Paderborn, Deutschland

In der Literaturwissenschaft ist die Methode der Stilometrie ein etabliertes Verfahren, um Fragestellungen verschiedener Art zu bearbeiten. Die Zuordnung anonym bzw. pseudonym überlieferter Werke zu bekannten Autoren aufgrund persönlicher stilistischer Gewohnheiten steht dabei oft im Vordergrund, aber auch die Untersuchung stilistischer Merkmale selbst ist Gegenstand von Stilometrie. Mit den Möglichkeiten der Digital Humanities lassen sich nicht nur größere Texte, sondern sogar ganze Korpora automatisiert auswerten und im Hinblick auf unterschiedlichste Textmerkmale untereinander vergleichen. Eine nach wie vor bestehende Hürde stellt dabei die Verfügbarkeit zuverlässigen Ausgangsmaterials dar, also die digitale Volltexterschließung des zu untersuchenden Materials. Zwar gibt es inzwischen mehrere Repositorien, die nennenswerte Textkorpora bereitstellen,¹ eine vollständige Abdeckung auch nur einzelner Epochen der Literaturgeschichte ist jedoch trotz der stetig verbesserten Erkennungsraten im Bereich der *Optical Character Recognition* (OCR) noch immer in weiter Ferne.

Stilkritik ist auch in der Musikwissenschaft eine übliche Methode. Hier wie dort geht es zunächst darum, autor- bzw. komponistenspezifische Eigen- und Gewohnheiten durch den Abgleich mit Vergleichswerken und -komponisten herauszuarbeiten. Während aber Stilometrie im literaturwissenschaftlich geprägten Bereich der Digital Humanities dankbar und schnell aufgegriffen und als eigener digitaler Forschungsbereich etabliert wurde, bleibt dieses Feld im musikwissenschaftlichen Teil der Digital Humanities bislang weitgehend unbespielt – eine digitale musikalische Stilometrie findet zumindest in der eingangs erwähnten Breite der Fragestellungen nicht statt. Verschiedene musikbezogene Disziplinen – in erster Linie Musikinformatik und Music Information Retrieval (MIR) – forschen zwar durchaus an Fragestellungen, die der Stilometrie inhaltlich nahestehen, setzen diese jedoch nicht im Sinne einer musikwissenschaftlicher Digital Humanities um.²

Der vorliegende Beitrag soll versuchen, die methodischen und gegenstandsbezogenen Gemeinsamkeiten und Unterschiede zwischen (textbezogener) Stilometrie und Music Information Retrieval herauszuarbeiten. Davon ausgehend sollen Perspektiven entwickelt werden, welchen Anforderungen eine erfolgversprechende musikalische Stilometrie gerecht werden müsste. Dabei soll es nicht um statistische bzw. algorithmische Optimierungen etwa im Hinblick auf einzusetzende Distanzmaße (Evert 2017)

gehen, sondern vielmehr um grundsätzliche Überlegungen zu den Anforderungen aus musikwissenschaftlicher Sicht.

Musik folgt grundsätzlich Regeln, die sich zwar von Epoche zu Epoche unterscheiden, und deren Durchbrechen oft als Innovation wahrgenommen und damit erwünscht ist, die sich aber in einer Musiktheorie beschreiben lassen, und deren Umsetzung Raum zur Entwicklung eines eigenen Stils bietet. Die grundsätzliche Formalisierbarkeit charakteristischer Merkmale bildet die Grundlage für Forschungen im Bereich der Komponisten-Attribution. Eine wichtige Arbeit in diesem Feld, die sich schon in ihrem Titel *On musical stylometry: A pattern recognition approach* ausdrücklich auf Stilometrie bezieht, wurde 2005 von Eric Backer und Peter van Kranenburg vorgelegt (Backer 2005). Darin extrahieren Backer und van Kranenburg zwanzig verschiedene Features aus dreissig Fugen von Komponisten, um dann eine weitere Fuge, die unklaren Ursprungs ist und in der Literatur allen drei Komponisten zugeordnet wird, auf ihre stilistische Nähe zu den anderen Werken zu untersuchen. Tatsächlich stellen sie eine große stilistische Nähe zu einem der Komponisten fest und können so glaubhaft für dessen Autorschaft argumentieren. Allerdings lässt sich ihre Arbeit nicht ohne weiteres übertragen. Die von ihnen gewählten Features sind sehr deutlich auf die musikalische Faktur des untersuchten Repertoires – barocke Fugen – abgestellt. Die Aussagekraft der bei Backer und van Kranenburg deutlich im Vordergrund stehenden Harmonie- und Stimmführungsbezogenen Features (Backer 2005: 304) ist jedoch zunächst nur für das von ihnen untersuchte Material als gegeben anzusehen. Tatsächlich beobachten sie in einem vorgeschalteten Experiment, dass aus einer heterogeneren Mischung von Komponisten – J.S. Bach, G.P. Telemann, G.F. Händel, W.A. Mozart und J. Haydn – deutlich schlechtere Ergebnisse resultieren, als wenn sie sich auf die Barock-Komponisten Bach, Telemann und Händel beschränken. Die gewählten Features können damit zur Beschreibung eines musikalischen Stils nicht als allgemeingültig betrachtet werden.

In der Musikwissenschaft wird musikalischer Stil auf verschiedenen Ebenen untersucht: Stil in Bezug auf Epochen ebenso wie auf Gattungen, harmonischer, melodischer und rhythmischer Stil, idiomatische, d.h. instrumentenspezifische Notationsweisen und Spielfiguren usw. (Pascall 2001). Eine Möglichkeit, dem zu begegnen, wäre die Berücksichtigung zahlreicher weiterer Features, wie es etwa in der Software *jMIR* bzw. dessen Komponente *jSymbolic* geschieht. Deren Zielsetzung "is to extract statistical information from musical data", um "research in the fields of music information retrieval (MIR), musicology and music theory"³ zu ermöglichen. Zu diesem Zweck werden insgesamt 246 verschiedene Features extrahiert, die in ihrer Breite deutlich generischer angelegt sind als bei Backer und van Kranenburg, daher auch über Gattungs- und Epochengrenzen hinweg verschiedene Stile besser abbilden und so insgesamt "robustere" Ergebnisse produzieren sollten. Sie decken unterschiedliche Aspekte eines Notensatzes ab, darunter Tonhöhen (41 Features),

Melodielinien (25), Akkorde (35), Rhythmus (95), Instrumentation (20).⁴ Damit wird auf den ersten Blick bereits eine hohe Bandbreite musikalischer Eigenschaften berücksichtigt, die allerdings im Vorfeld (automatisiert) extrahiert werden müssen, und die wiederum im Einzelnen nicht immer von gleichbleibender Aussagekraft sind, sondern je nach Stilbegriff und untersuchtem Material gezielt ausgewählt werden müssen, um aussagekräftige Ergebnisse zu erzielen, indem "irrelevant features that would introduce unproductive noise into classification systems" (McKay 2010: 197) vermieden werden.

Spätestens an dieser Stelle wird deutlich, dass der Datenaufbereitung im Fall von Musik eine besondere Bedeutung zukommt. In der (literaturwissenschaftlichen) Stilometrie werden demgegenüber i.d.R. vollständige Texte ausgewertet, die auf den ersten Blick nicht speziell aufbereitet werden. Allerdings finden auch hier durchaus vergleichbare Vorbereitungen statt, die allerdings im Fall des verbreiteten Programms *Stylo*⁵ in den üblichen Arbeitsablauf direkt integriert sind. Üblicherweise basieren stilometrische Untersuchungen mit *Stylo* auf der Untersuchung der *most frequent words* (MFW), also der Verteilung der am häufigsten genutzten Wörter, da diese "kaum bewusst manipuliert" werden⁶ und so ein gutes Kriterium zur Autorattribution sind. Hierzu werden also die Texte in ihren Schreibweisen normalisiert und dann anhand ihrer Frequenzen die Verteilung der *n* häufigsten Worte bestimmt. Die eigentliche stilometrische Untersuchung besteht nun darin, die resultierenden Histogramme statistisch abzugleichen. Dieses Verfahren ist durchaus vergleichbar mit der Untersuchung extrahierter Features im Bereich MIR. Es erscheint daher naheliegend, die Eignung von *Stylo* zur stilometrischen Untersuchung musikalischer Daten zu überprüfen. Ein geeignetes Datenset findet sich im *Haydn/Mozart String Quartet Quiz*.⁷ Bei diesem Quiz werden dem Benutzer zufällig Streichquartette von Mozart oder Haydn vorgespielt, die dieser den Komponisten zuordnen soll. Die durchschnittliche Trefferquote für die 287 Sätze, die im Korpus enthalten sind, liegt bei annähernd 50.000 gehörten Sätzen bei ungefähr 58%, gemittelt über Hörer aller Grade an musikalischem Vorwissen.⁸ Diese geringe Quote, die nur unwesentlich besser als eine statistische Normalverteilung ist, verdeutlicht die stilistische Nähe der beiden Komponisten: Es ist auch für menschliche Hörer alles andere als einfach, die Werke beider Komponisten zu unterscheiden. Damit stellt dieser Korpus einen interessanten Testfall dar, da die Hürde für eine nach menschlichen Maßstäben "erfolgreichere" Klassifizierung nicht allzu hoch liegt, während die Schwierigkeit gleichzeitig recht hoch ist und damit genug Raum zur Optimierung bietet.

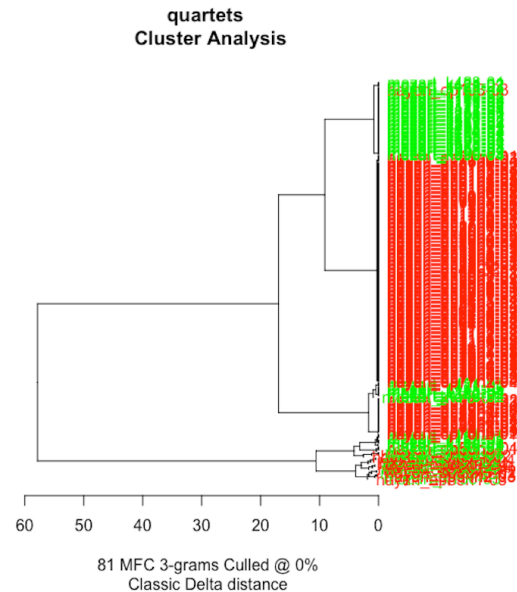


Abbildung 1: Auswertung der Daten des Haydn/Mozart-Quiz, basierend auf MusicXML. Grüne Einträge verweisen auf Mozart, rot steht für Haydn.

Das Datenmaterial findet sich gleich mehrfach im Netz. Es basiert auf Codierungen, die in den 1990er Jahren am *Center for Computer Assisted Research in the Humanities* (CCARH) der Stanford University entstanden sind, und die in verschiedenen Datenformaten von <http://kern.ccarh.org/> heruntergeladen werden können. Um ein auch strukturell möglichst einfaches Format nutzen zu können, wurden die im MusicXML-Format geladenen Daten per Skript ins abc-Format konvertiert.⁹ Dieses plain-text-Format nutzt eine sehr einfache, nach Stimmen geordnete Syntax, um Melodielinien zu codieren. Mit entsprechenden Parametern¹⁰ wird hier eine Erkennungsrate erzielt, die mit ~62% etwa im Bereich der menschlichen Benutzer des Mozart-Haydn-Quiz liegt. Eine etwas bessere Unterscheidung (~69%) gelingt mit *Stylos classify()*-Funktion, wenn dort als Algorithmus *SVM (Support Vector Machines)*¹¹ ausgewählt werden. Die genutzten Wörter sind dabei in jedem Fall deutlich aussagekräftiger, wie das Beispiel "D E F G F E D C B A" zeigt. Trotz der besseren Ergebnisse zeigt sich, dass die Betrachtung nur eines einzelnen Aspekts der Notation – hier der Abfolge von Tonhöhen – noch keine ausreichende Aussagekraft bietet, um darauf stilometrische Analysen aufzubauen. Musik ist "mehrdimensional", d.h. mehrere Stimmen bzw. Instrumente verlaufen zeitgleich. Während es textbasierte Datenformate gibt, die sich im Fall von einstimmiger Musik recht plausibel als Text behandeln lassen, lässt sich mehrstimmige Musik nur mit einem Fokus entweder auf Harmonik (d.h. als Fortschreibung von Klängen) oder auf Melodien (d.h. als Abfolge einzelner Stimmen, deren Gleichzeitigkeit in den Daten nicht ersichtlich wird) in einem Datenformat darstellen. Das explizite Aufbereiten

der Daten – die Extraktion von Features – ist also unumgänglich, um sämtliche Facetten von Musik in den Blick nehmen und im Rahmen einer stilometrischen Analyse berücksichtigen zu können.

Vor diesem Hintergrund erscheint es naheliegend, zu *jSymbolic* zurückzukehren, da dessen oben erwähnte Features ein weitaus breiteres Spektrum der musikalischen Mehrdimensionalität abbilden. *jSymbolic* arbeitet i.d.R. mit MIDI-Daten – es gibt zwar durchaus die Möglichkeit, MEI-Daten zu verarbeiten, allerdings werden lediglich zwei Features (Häufigkeit von Bindebögen bzw. Vorschlagsnoten) aus diesen Daten extrahiert, die bei einem MIDI-Input nicht zur Verfügung stehen. *jSymbolic* übernimmt dabei nur die Extraktion und Aufbereitung der Features, die eigentliche statistische Auswertung erfolgt durch *Weka*, eine Java-basierte, unter GPL lizenzierte Software zum maschinellen Lernen.¹² Diese bietet eine zu *Stylo* vergleichbare Benutzeroberfläche und erlaubt es u.a., die extrahierten Features ebenfalls mit dem SVM-Algorithmus auszuwerten. Bei Übernahme der im *jSymbolic*-Tutorial beschriebenen Standard-Einstellungen¹³ wird bereits eine korrekte Klassifizierung von knapp 82% erreicht, d.h. 233 von 286 Dateien werden korrekt zugeordnet. Dieses Ergebnis ist überaus beachtlich, insbesondere da keinerlei Optimierungen vorgenommen wurden, und auch auf genau dieses Datenset hin optimierte Lösungen aktuell nur auf Raten von gut 85% kommen (Kempfert 2019: 13).

Damit scheinen die von *jSymbolic* extrahierten Features eine gute Ausgangslage zu bieten, um stilometrische Fragestellungen im Bereich Musik zu bearbeiten – sei es mithilfe von *Stylo*, *Weka*, oder anderen entsprechenden Werkzeugen. Allerdings stellt sich die Frage, ob der geschilderte Workflow bereits das volle Potential des zur Verfügung stehenden Datenmaterials ausschöpft: MIDI taugt als Datengrundlage nur sehr bedingt, wenn es um Musiknotation geht. Einer der offensichtlichsten Gründe ist dabei die Art, Tonhöhen quasi als durchnummerierte Klaviertasten zu codieren, da in diesem System kein Unterschied zwischen z.B. den Tönen Dis und Es besteht. Will man nun das Intervall etwa von einem C zu diesem Ton bestimmen, so ergibt sich in einem Fall eine Sekunde, im anderen ein Terzabstand. Auch hier gilt wieder, dass das Ignorieren dieser sog. *Enharmonik* nicht zwangsläufig und in jedem Fall zu schlechteren Ergebnissen führen muss, allerdings sind derartige Informationen für eine "händische" Klassifizierung überaus hilfreich, nicht zuletzt zur Epochenzuordnung. An dieser Stelle merkt man *jSymbolic* an, dass es im Rahmen von *jMIR* nur eine Komponente darstellt, und es (neben weiteren) mit *jAudio* eine Entsprechung zur Feature-Extraktion aus Audiodaten gibt – in denen dann naturgemäß keine Informationen zur Enharmonik enthalten sein können. *jSymbolic* schöpft hier also aus Gründen der Merkmalsparität der extrahierten Features nicht das volle Potential der Datengrundlage codierter Partituren aus.¹⁴ Da in den letzten Jahren immer mehr Editionsprojekte hochwertige Datenkorpora

erarbeiten und bereitstellen, bietet sich hier eine Perspektive, wie stilometrische Fragestellungen auch im Bereich Musik sinnvoll projiziert werden können: Noch vor der Optimierung der statistischen Methoden für das zu untersuchende Material muss dessen Aufbereitung für eine solche Auswertung optimiert werden. Dazu gibt es umfangreiche Vorarbeiten, insbesondere im Bereich des Music Information Retrieval, die aber stellenweise aus musikwissenschaftlicher Sicht noch defizitär sind und das volle Potential der zur Verfügung stehenden Daten bislang nicht ausschöpfen.

Fußnoten

1. Vgl. etwa die über TextGrid frei verfügbaren Bestände von <http://zeno.org>.
2. Als Beispiel sei hier Cilibrasi, Rudi / Vitányi, Paul / de Wolf, Ronald (2004): "Algorithmic clustering of music based on string compression", in: *Computer Music Journal*, 28 (4), 49–67 genannt. Dieser Beitrag diskutiert auf rein statistischer Ebene mögliche Verwandtschaftsbeziehungen zwischen verschiedenen Musikcodierungen, ohne dabei jedoch deren musikwissenschaftliche Bedingungen bzw. Implikationen angemessen zu thematisieren.
3. Vgl. <http://jmir.sourceforge.net/> (letzter Zugriff 06.01.2020).
4. Vgl. http://jmir.sourceforge.net/manuals/jSymbolic_manual/home.html (letzter Zugriff 06.01.2020).
5. Vgl. <https://github.com/computationalstylistics/stylo> (letzter Zugriff 06.01.2020).
6. Jan Horstmann (2018, § 11): „Stilometrie“. In: *forTEXT. Literatur digital erforschen*. URL: <https://fortext.net/routinen/methoden/stilometrie> (letzter Zugriff 06.01.2020). Das Zitat wird dort Fotis Jannidis und Gerhard Lauer: „Burrows's Delta and Its Use in German Literary History“. In: Matt Erlin und Lynne Tatlock (Hrsg.): *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*. Rochester, New York: Camden House, 29–54, S. 180 zugeschrieben, findet sich dort aber zumindest nicht in dieser Form.
7. Vgl. <http://qq.themefinder.org/> (letzter Zugriff 06.01.2020; das Quiz selbst war zu diesem Zeitpunkt unter verschiedenen aktuellen Browsern nicht funktionsfähig).
8. Die Benutzer des Quiz müssen sich durch einige Fragen selbst auf einer Skala von 0 ("novice") bis 10 ("expert") einordnen. Die Erkennungsrate schwankt laut einer Auswertung im Frühjahr 2006 zwischen ~51% (Level 1) und ~88% (Level 8). Vgl. <http://qq.themefinder.org/cgi-bin/mhstyle?submit=statistics> (letzter Zugriff 06.01.2020).
9. Zur Konvertierung vgl. <https://wim.vree.org/svgParse/xml2abc.html>, zum Datenformat vgl. <http://abcnotation.com> (letzter Zugriff jeweils 06.01.2020). Tatsächlich lassen sich auch die MusicXML-Daten direkt in *Stylo* auswerten. Die dabei zu erreichende

Erkennungsrate von ~75% ist aber irreführend: Die von *Stylo* angelegte Wortliste offenbart, dass vor allem Metadaten zur Besetzung, die in den beiden Datensätzen unterschiedlich angelegt sind, ausschlaggebend sind, und musikalische Inhalte im Grunde nicht berücksichtigt werden.

10. Auswertung von n-grams mit n=10 auf Basis vollständiger Wörter, 100 MFW, kein Culling, Classic Delta.

11. Vgl. https://de.wikipedia.org/wiki/Support_Vector_Machine (letzter Zugriff 06.01.2020).

12. Vgl. <http://www.cs.waikato.ac.nz/ml/weka/> (letzter Zugriff 06.01.2020).

13. Vgl. http://jmir.sourceforge.net/manuals/jSymbolic_tutorial/usingweka.html (letzter Zugriff 06.01.2020).

14. Gleiches gilt etwa für die Dissertation von Christof Weiß ("Computational methods for tonality-based style analysis of classical music audio recordings", Ilmenau 2017,), die zwar einen überaus wichtigen Beitrag zur automatischen Klassifizierung von Musik anhand stilistischer Merkmale liefert, dabei aber ausschließlich auf Audiodaten operiert.

Bibliographie

Backer, Eric / van Kranenburg, Peter (2005): "On musical stylometry – a pattern recognition approach", in: *Pattern Recognition Letters* 26, 299–309.

Evert, Stefan / Proisl, Thomas / Jannidis, Fotis / Reger, Isabella / Pielström, Steffen / Schöch, Christof / Vitt, Thorsten (2017): "Understanding and explaining Delta measures for authorship attribution", in: *Digital Scholarship in the Humanities* 32, ii4–ii16. <https://doi.org/10.1093/llc/fqx023>

Horstmann, Jan (2018): "Stilometrie", in: *forTEXT. Literatur digital erforschen*. URL: <https://fortext.net/routinen/methoden/stilometrie>.

McKay, Cory (2010): "Automatic music classification with jMIR" (PhD). McGill University, Montréal.

Kempfert, Katherine C. / Wong, Samuel W. K. (2019): "Where Does Haydn End and Mozart Begin? Composer Classification of String Quartets", arXiv:1809.05075 [stat].

Pascall, Robert (2001): "Style", in: *The New Grove Dictionary of Music and Musicians*. Oxford. 24, 638–642.