

Rhythmus-Analyse mit KI-Systemen

Musik zeichnet sich neben harmonischer und melodischer auch durch rhythmische Vielfalt aus. Letztere lässt sich jedoch zumeist auf ein kontinuierliches zeitliches Raster reduzieren auf dem Ereignisse platziert werden können, der sogenannten Microtime. Ausgehend von den Beats, den Hauptschlägen, stellt sich dann meist nur die Frage ob diese in 3 Schläge oder in 2 bzw damit verbunden in 4 Schläge unterteilt werden, woraus sich das Microtime-Raster ergibt. KI-Audioanalyse wird in vielfältigen Bereichen genutzt. So stellt sich die Frage, ob und wie mithilfe dieser Analysemethoden auch die Microtime erkannt werden kann. Es handelt sich also um ein Problem mit den zwei Kategorien „2“ und „3“.

Zur Durchführung des Projekts wurde ein Datensatz zusammengestellt, der überwiegend aus Pop- und Jazzmusik bestand. Vereinzelt gab es auch Beispiele aus Filmmusik, Klassik, Rock und elektronischer Musik. Dieser Datensatz wurde in Rhythmen basierend auf der Microtime 3 oder 2 eingeteilt. Damit ergaben sich 82 Beispiele für „2“ und 35 für „3“, in Summe also 117 Beispiele. Die Audiodateien wurden dann mittels Python und den Bibliotheken Librosa, sowie Scikit Learn (sklearn) verarbeitet, analysiert und kategorisiert. Dabei wurden sie zunächst in 80 % Trainingsdaten und 20 % Testdaten aufgeteilt und darauf folgend in 10 Sekunden lange Abschnitte zerteilt. Damit ergaben sich 2372 Trainingsbeispiele und 647 Testbeispiele. Mittels Librosas *Harmonic-percussive source separation (librosa.hpss)* wurden harmonische von perkussiven Spektralanteilen getrennt. Ziel dabei war, eine klarere rhythmische Information für die Weiterverarbeitung zu erhalten. Mit der Funktion *librosa.beat* konnte das Tempo und die Beatlänge bestimmt werden und mit *librosa.onset* wurde ein Vektor mit Zeitpunkten des Beginns musikalischer Ereignisse erstellt. Dies ist etwa in Abbildung 1 mit einem Spektrum dargestellt.

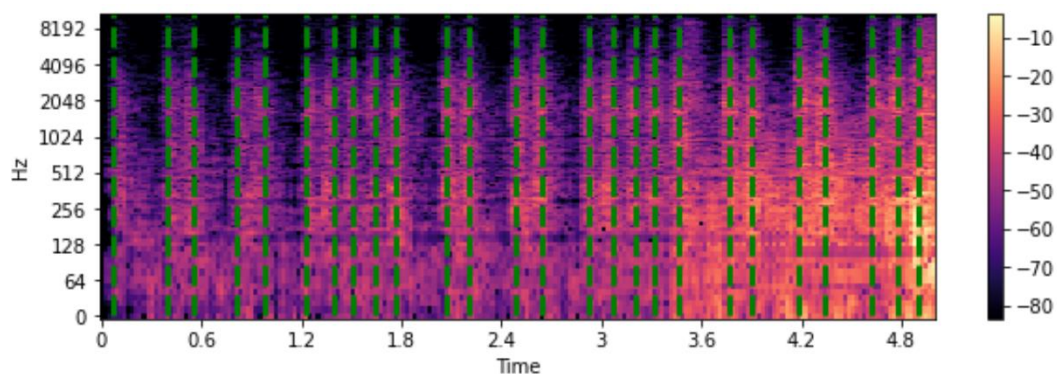


Abbildung 1: Spektrum des Audiobeispiels "He's a pirate" von Klaus Badelt mit detektierten Ereignissen mittels *librosa.onset* (grüne Linien).

Mittels Autokorrelation kann die Selbstähnlichkeit bei Verschiebung um ein bestimmtes Zeitintervall berechnet werden. Diese Ähnlichkeit sollte etwas bei einem Stück aus „3“ höher sein, wenn es um ein Intervall verschoben wird, das einem Drittel der Beatlänge entspricht, als bei der Hälfte der Beatlänge für ein Stück aus „2“ – und umgekehrt. Damit wurde mit *librosa.autocorrelate* die

Korrelationskurve berechnet und die Werte an den Stellen Beatlänge geteilt durch 3 sowie geteilt durch 2 entnommen. Eine solche Kurve ist in Abbildung 2 dargestellt mit den entsprechenden Stellen markiert.

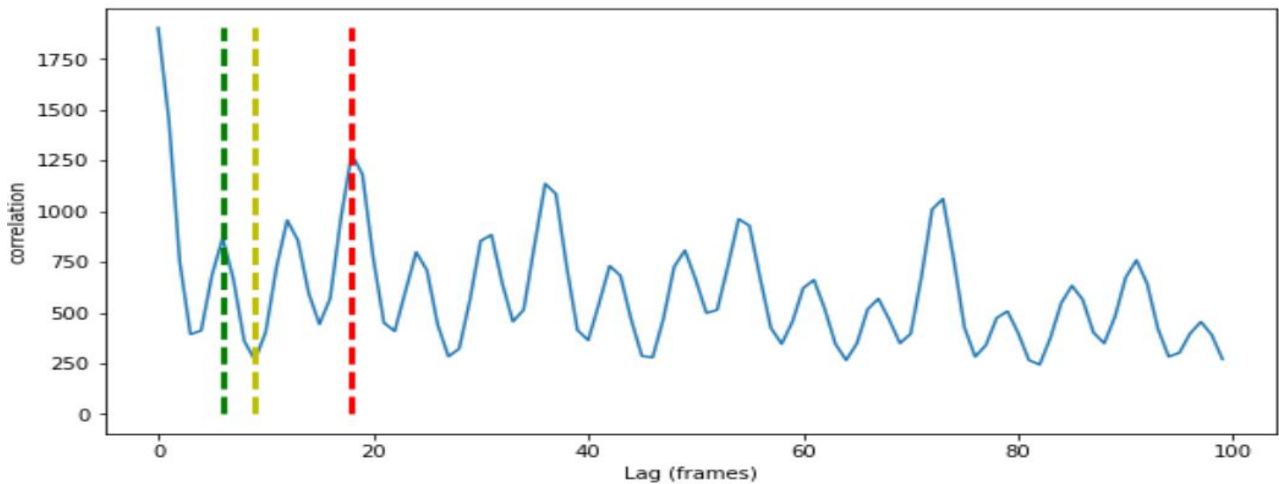


Abbildung 2: Korrelationskurve (blau) des Audiobeispiels "He's a pirate" von Klaus Badelt mit den markierten Stellen der Beatlänge (rot), der halben Beatlänge (gelb) und einem Drittel der Beatlänge (grün). Das Stück gehört zur Kategorie „3“, entsprechend ist der Korrelationswert bei einem Drittel der Beatlänge höher als bei der halben Beatlänge.

Die beiden Korrelationswerte wurden anschließend als Merkmale für die Kategorisierung durch den KI-Algorithmus verwendet. Die Datenpunkte sind in Abbildung 3 dargestellt. Mehrere Algorithmen wurden verglichen, wobei dies im esentlichen immer zu zwei Ergebnissen führte. Hier soll ein linearer Ansatz mit dem *LinearSVC* und eine flexiblere Kategorisierung mit dem *GaussianProcessClassifier* (*GPC*) verglichen werden. Die statistischen Ergebnisse sind in Tabelle 1 zusammen gefasst. Der *GPC* hat nahezu alle Datenpunkte der größeren Kategorie „2“ zugeordnet. Während *LinearSVC* balanciert war und

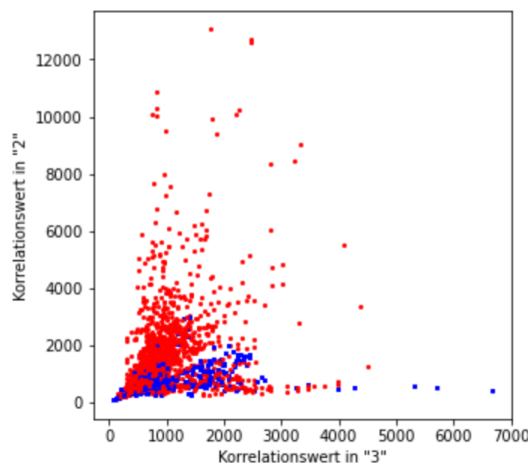


Abbildung 3: Korrelationswerte von „2“ (rot) und „3“ (blau)

„3“ besser erkannte, es jedoch bei „2“ zu einem sehr schlechten Ergebnis kommt.

Keiner der beiden Klassifikatoren konnte die Microtime gut unterscheiden und erkennen. Der Lineare Ansatz scheint eine Trennung der beiden zu ermöglichen, jedoch mit sehr großen Fehlern, insbesondere durch die ungleich verteilten Daten. Die vorangegangene Datenaufbereitung mit

librosa.onset und *librosa.autocorrelate* führte insbesondere bei Jazzbeispielen zu sehr ungenauen und oft auch falschen Ergebnissen. Auch aus musikalischer Sicht ist eine Trennung der Microtimes oft nur bedingt möglich. Einzelne musikalische Elemente lassen sich nicht immer auf ein konstantes mathematisches Raster in 2, 4 oder 3 genau zuordnen. Dieser fließende Übergang zwischen den Microtimes ist essentiell für ein natürliches Groovegefühl.

Die Klassifikation der Microtime mit KI erscheint dennoch mit einem balancierten größeren Datensatz und dem beschränken auf bestimmte Genre mit klaren „mathematischen“ Rhythmus möglich.

	GPC	LinearSVC
accuracy	0.70	0.56
Sensitivity für „2“	0.93	0.48
Sensitivity für „3“	0.23	0.74

Tabelle 1: Statistische Ergebnisse der Klassifikatoren