

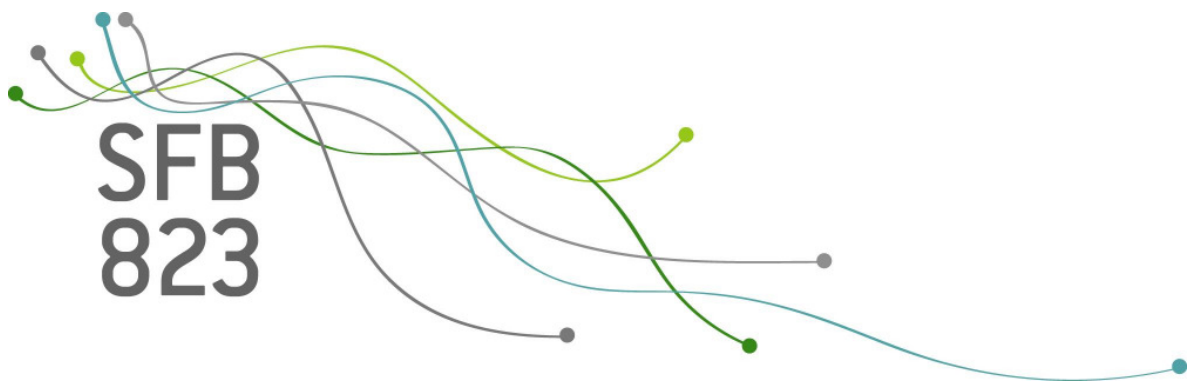
SFB
823

Einsatzzeiterkennung bei polyphonen Musikzeitreihen

Nadja Bauer, Julia Schiffner, Claus Weihs

Nr. 22/2010

Discussion Paper



Einsatzzeiterkennung bei polyphonen Musikzeitreihen

Technischer Bericht

TU Dortmund
Fakultät Statistik
Lehrstuhl für Computergestützte Statistik

Nadja Bauer, Julia Schiffner, Claus Weihs

7. Juni 2010

1 Einleitung

Die Detektion der Einsatzzeitpunkte von Tönen ist von zentraler Bedeutung bei der automatischen Musiktranskription. Für diese Problemstellung wurden bereits mehrere Methoden vorgeschlagen, wobei die in Echtzeit erfolgende Detektion der Toneinsätze bei polyphonen Aufzeichnungen einen aktuellen Forschungsgegenstand im Bereich der Signalanalyse darstellt. Die Problematik dieser Aufgabe besteht einerseits darin, einen Algorithmus zu entwickeln, der gleich gute Ergebnisse für unterschiedliche Arten von Musikinstrumenten liefert (beispielsweise Schlag- oder Blasinstrumente). Andererseits stehen im Moment nur wenige polyphone Musikstücke zur Verfügung, bei denen die wahren Einsatzzeiten bekannt sind, so dass die Güte der Algorithmen verglichen bzw. optimale Parametereinstellungen gefunden werden können. Dabei sind Aufzeichnungen echter Musikinstrumente von Interesse, die sich in ihren Eigenschaften wesentlich von synthetisch generierten Tonsequenzen unterscheiden. Ein großer Schritt in diesem Gebiet wäre zum einen die Zusammenstellung einer Datenbank, die für jedes Musikinstrument mehrere monophone und polyphone Aufzeichnungen mit bekannten Einsatzzeiten enthält, wobei diese Aufzeichnungen in ihren Eigenschaften so unterschiedlich wie möglich sein sollten. Dafür wäre zunächst notwendig, relevante Audiosignaleigenschaften zu definieren wie z.B. Tempo, Lautstärke, Signal-Rausch-Verhältnis, Tonreihenfolge und dann den gesamten Raum unterschiedlicher Tonsequenzen mit möglichst wenigen Punkten abzudecken. Zum anderen könnten auf diesen Daten die optimalen Parameter unterschiedlicher Algorithmen zur Einsatzzeitdetektion mittels statistischer Versuchsplanung geschätzt werden. Um einen optimalen (im Sinne eines geeigneten Optimalitätskriteriums) Versuchsplan aufzustellen, soll zunächst ein funktionaler Zusammenhang zwischen der Zielgröße (Einsatzzeit) und den Einflussfaktoren (Algorithmusparameter) für jeden zu testenden Algorithmus gefunden werden.

In Hinblick auf diese Zielsetzung wurden im Rahmen der Projektarbeit (Teilprojekt C2 des SFBs 823) als Einarbeitung in den Untersuchungsgegenstand zwei Algorithmen zur Einsatzzeiterkennung (speziell für Klavieraufzeichnungen) entwickelt und ein Versuch der automatischen Einsatzzeitdetektion für ein klassisches Klaviermusikstück vorgenommen. Im zweiten Abschnitt wird zunächst der Stand der Forschung zur Einsatzzeitdetektion beschrieben und die beiden entwickelten Algorithmen vorgestellt. Der dritte Abschnitt befasst sich mit der Anwendungsaufgabe. Abschließend sind im vierten Abschnitt eine kurze Zusammenfassung des Berichts sowie ein Ausblick auf weitere Ziele der zukünftigen Arbeit gegeben.

2 Automatische Detektion der Einsatzzeiten bei Musikzeitreihen

2.1 Stand der Forschung

Bei den existierenden Ansätzen zur Einsatzzeitdetektion werden sehr unterschiedliche Vorgehensweisen vorgeschlagen. Bei einigen werden zu eingehenden Signalsequenzen mittels Klassifikatoren oder Wahrscheinlichkeitsmodellen ([11],[3]) die gespielten Noten zugewiesen, wobei Information über die spektralen Eigenschaften einzelner Noten verwendet wird. Andere Algorithmen erkennen die Toneinsätze unabhängig von der Kenntnis der aktuell gespielten Noten (ein kurzer Überblick ist bei [2] gegeben), wobei zu dieser Kategorie die meisten Verfahren gehören. So wird im ersten Schritt dieser Verfahren optional das eingehende Signal transformiert (*Pre-processing*) und dann eine Detektorfunktion definiert (*Reduction*): Sie basiert entweder auf der zeitlichen ([1]) oder auf der spektralen ([6],[9],[10]) Darstellung des Audiosignals, in einigen Fällen wird eine Annahme über die Verteilung der Amplitude innerhalb der Sampleblöcke gemacht ([1]) bzw. das Audiosignal als AR-Modell

modelliert ([9]). Im weiteren Schritt werden unter lokalen Maxima (*Peaks*) der Detektorfunktion diejenigen ausgewählt, die als neue Toneinsätze gelten, dazu werden z.B. dynamische oder fest vorgegebene Schwellenwerte verwendet (*Peak-picking*). Bei [10] wird ein statistisches Modell zur Klassifikation der gefundenen Peaks aufgestellt.

Es existieren auch zahlreiche Programme, die unter anderem der Noteneinsatzerkennung dienen, darunter z.B. die kommerzielle Software Melodyne¹, „Matlab Code that implements Piano Note Detection“² bzw. die MATLAB MirToolbox³. Diese Programme liefern für monophone Aufzeichnungen teilweise gute Ergebnisse.

Wesentlich anspruchsvoller als bei monophonen Aufzeichnungen ist die Erkennung der Einsatzzeiten im polyphonen Fall. Da meistens die automatische Transkription der Musikstücke von Interesse ist, d.h. die Umwandlung des Audiosignals in den ursprünglichen Notentext, sollen sowohl die Einsatzzeitpunkte der einzelnen Töne bestimmt als auch ihre Grundfrequenzen innerhalb dieser Zeitabschnitte geschätzt werden. Letzteres ist in der Literatur unter *Multipitch-estimation* bekannt und stellt ein wichtiges Forschungsgebiet in der Signalanalyse dar. Ein Überblick über einige existierende Ansätze ist z.B. bei [4] gegeben. Von Interesse sind ebenfalls die Veröffentlichungen von [8], [7] sowie [5]. Allerdings wurden die meisten entwickelten Algorithmen anhand von synthetisch generierten Audiosignalen getestet und liefern für die „echten“ polyphonen Klavieraufzeichnungen nicht akzeptable Ergebnisse. Zu bemerken ist, dass eine verlässliche Grundfrequenzschätzung bei Polyphonie für die Erkennung der Noteneinsätze von entscheidender Bedeutung ist.

¹<http://www.celemony.com>, Stand: 01.01.2010.

²<http://cnx.org/content/m14196/latest>, Stand: 01.01.2010.

³<https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox>.

2.2 Ein auf der zeitlichen Darstellung des Signals basierender Algorithmus

Die vorgeschlagene Methode beruht auf der Annahme, dass der Einsatz eines neuen Klaviertons durch einen deutlichen Amplitudenanstieg gekennzeichnet ist und eignet sich besonders gut für Klavierstücke. Die Analyse erfolgt hier lediglich auf dem Low-Level-Niveau: Es werden nur die Amplitudenschwankungen des eingehenden Audiosignals im Zeitbereich berücksichtigt. Die grundlegende Idee wird im Folgenden vorgestellt: Ausgehend von einem Startpunkt werden fortschreitend globale Maxima innerhalb von Sampleblöcken einer festen Länge L bestimmt. Falls der Absolutbetrag von Maximum eines Blocks um einen Parameter S gegenüber dem Maximum des vorherigen Blockes größer ist, wird in diesem Block der neue Toneinsatz registriert. Formal lässt sich diese Methode wie folgt aufschreiben:

$$Einsatz[T] = \text{sign}(\max(x[t])|_{t=(T-1)\cdot L}^{T\cdot L} - S \cdot \max(x[t])|_{t=(T-2)\cdot L}^{(T-1)\cdot L}).$$

Dabei stehen $t = 1, \dots, N$ für den Sampleindex und $T = 2, \dots, \lfloor \frac{N}{L} \rfloor$ für den Blockindex. $x[t]$ ist die Amplitude des eingehenden Audiosignals der Samplelänge N . Funktion $\text{sign}(x)$ nimmt Werte $+1$, falls $x > 0$ ist und -1 sonst. Der Toneinsatz wird also in dem T -ten Block gefunden, falls $Einsatz[T] = 1$ gilt. In dem so ermittelten Block erfolgt dann eine genauere Analyse des Amplitudenverhaltens und die Bestimmung der Einsatzzeit des neuen Tons. Weiterhin werden Pausen als die Zeitabschnitte erkannt, in denen die absoluten Amplituden der Zeitreihe einen vorgegeben Wert A nicht überschreiten. Die Parameter L , S und A können beliebig gewählt werden, für die analysierten Aufnahmen hat allerdings folgende Kombination besonders verlässliche Ergebnisse geliefert: $L = 512$ Samples, $S = 1.4$ und $A = 80$ Samples. Bei den getesteten monophonen Klavieraufzeichnungen lieferte der entwickelte Algorithmus genauere Einsatzzeiten der neuen Töne als beispielsweise die vergleichbare Funktion `mironsets()` aus der MATLAB MirToolbox.

Abbildung 1 stellt die mittels des entwickelten Algorithmus berechneten Einsatzzeiten für die monophone Klavieraufzeichnung des Liedes *Halleluja*⁴ dar. Dabei ist zu bemerken, dass alle Toneinsätze richtig gefunden worden sind.

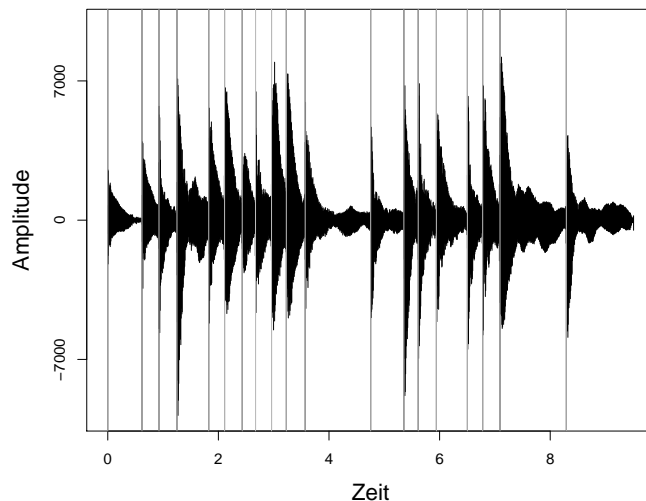


Abbildung 1: Mit Hilfe des entwickelten Algorithmus erkannte Einsatzzeiten (vertikale Linien) für die monophone Klavieraufzeichnung des Liedes *Halleluja* (erste 9.5 Sekunden).

2.3 Ein auf der spektralen Darstellung des Signals basierender Algorithmus

Der in diesem Abschnitt beschriebene Algorithmus basiert auf der Annahme, dass sich die spektralen Strukturen zweier unterschiedlicher Töne (bzw. Tonkombinationen bei Polyphonie) voneinander unterscheiden. Es sollte also möglich sein, eine Änderung der spektralen Struktur bei einem neuen Toneinsatz festzustellen. Hier

⁴*Halleluja, Hallelu, Halleluja* (Text: Winfried Offele (1975); Melodie: aus England, 19. Jh.) In: *Halleluja - Lieder auf dem Weg des Glaubens* (1995); Hrsg.: Bistum Essen, Bischöfliches Seelsorgeamt, Ferdinand Kamp GmbH a& Co. KG, Bochum

wird eine sehr einfache und intuitive Methode vorgeschlagen: Teile das Signal der Samplelänge N in die Blöcke der Samplelänge L und bestimme mittels der schnellen Fourier-Transformation die spektrale Darstellung innerhalb der jeweiligen Blöcke (zu dem T -ten Block gehört ein Vektor des Spektrums $S[T]$ der Länge $L/2$). Zu bemerken ist, dass umso mehr Frequenzen erfasst werden können, je größer die Fensterlänge L ist, allerdings darf L nicht zu groß sein, damit bei sehr schnellem Tempo Toneinsätze nicht übersprungen werden. In der Literatur wird oft $L = 2048$ Samples (46 ms bei einer Samplingrate von 44100 Hz) gesetzt ([6]). Die Detektorfunktion wird in diesem Fall als die Korrelation der benachbarten Vektoren der spektralen Darstellung definiert:

$$D[T] = \text{Cor}(S[T], S[T - 1]), T = 2, \dots, \left\lfloor \frac{N}{L} \right\rfloor.$$

T ist hier der Blockindex. Innerhalb eines Tones bzw. einer Tonkombination sollten dann die Werte von D groß sein im Gegensatz zu den Zeitpunkten, wo ein neuer Toneinsatz stattfindet. Anschließend soll ein Schwellenwert δ definiert werden, ab welchem ein lokales Minimum als neuer Toneinsatz klassifiziert wird. Hier wurde nur ein konstanter Schwellenwert ausprobiert im Gegensatz zu den in Literatur üblichen dynamischen Schwellenwerten.

Abbildung 2 zeigt die Anwendung des oben beschriebenen Verfahrens auf (ab der zweiten Hälfte) polyphonischem Klaviersignal, das mittels synthetischer Töne aus einer MIDI-Datei erzeugt wurde. Bei den Einstellungen $L = 2048$ Samples und Schwellenwert $\delta = 0.85$ ist nur einer von 46 gegebenen Einsätzen nicht erkannt worden, wobei die mittlere Abweichung der geschätzten Zeiten von den wahren Einsatzzeiten ca. 21.5 ms beträgt. In der Literatur wird üblicherweise ein geschätzter Toneinsatz als richtig erkannt, falls der Zeitunterschied zu dem wahren Toneinsatz weniger als 50 ms beträgt ([2]).

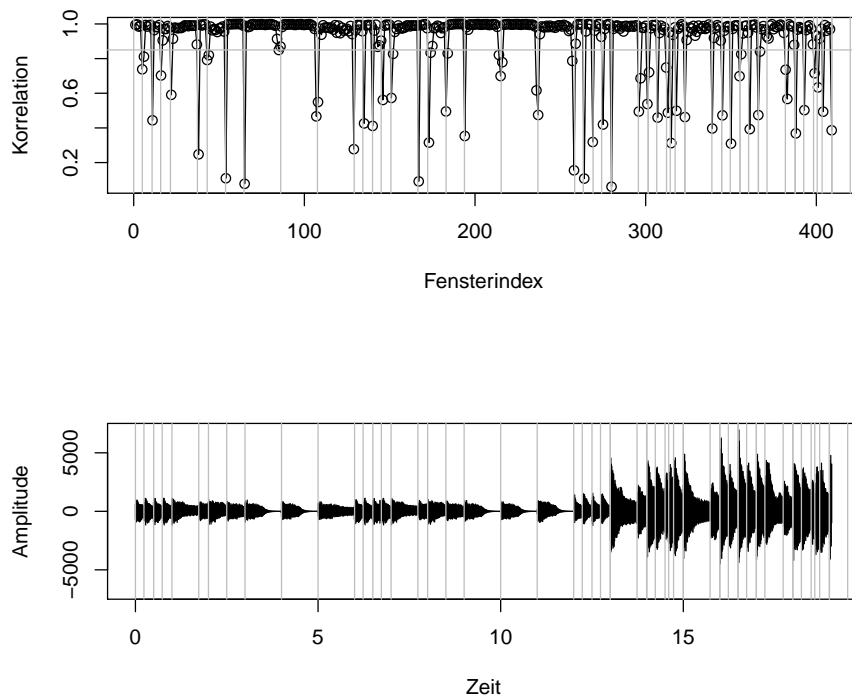


Abbildung 2: Korrelation der Spektren von benachbarten Fenstern für eine durch synthetische Töne generierte Klavieraufzeichnung (oben). Vertikale Linien entsprechen den wahren Einsatzzeiten und horizontale Linie entspricht dem Schwellenwert $\delta = 0.85$. Unten ist die Zeitliche Darstellung des Audiosignales gegeben.

In Abbildung 3 sind dagegen die Ergebnisse des Algorithmus für das Klavierstück *Halleluja* vorgestellt, wobei hier eine Besonderheit zu sehen ist: Bei lang klingenden echten Klaviertönen verändert sich deren spektrale Struktur sehr stark (vermutlich wegen der Obertöne), so dass diese Veränderungen fälschlicherweise als neue Töne erkannt werden (bei den ersten 9.5 Sekunden des Liedes wurden sieben Einsätze fälschlicherweise gefunden). Dies ist ein Beispiel dafür, dass der vorgeschlagene Algorithmus nicht immer gut funktioniert. Die Anwendung eines vergleichbaren Algorithmus (*Spectral Flux*, s. [6]) zeigte wesentlich bessere Ergebnisse: Nur zwei falsch erkannte Toneinsätze. Als Verbesserungsvorschlag für den entwickelten Algorithmus

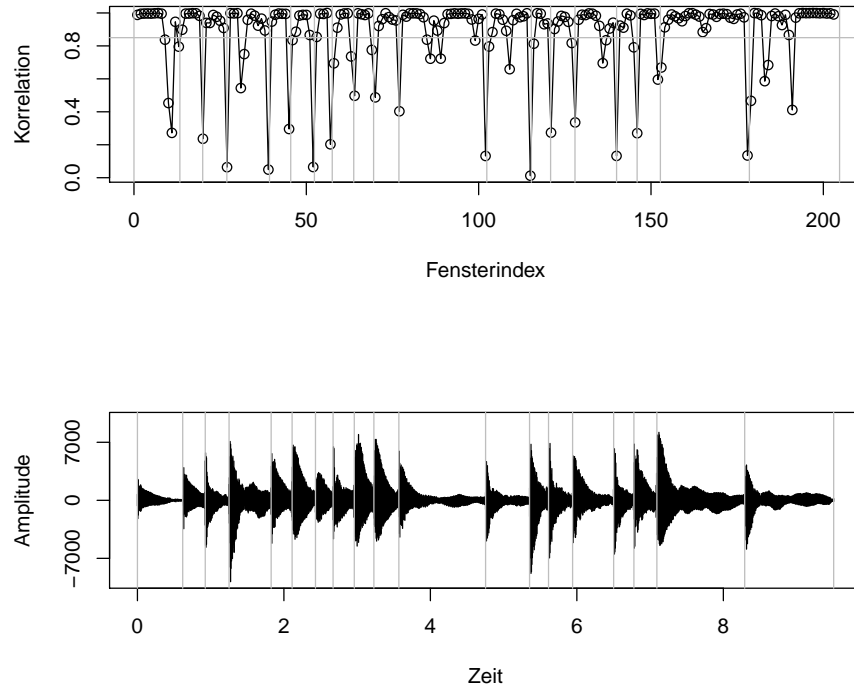


Abbildung 3: Korrelation der Spektren von benachbarten Fenstern für die Klavieraufzeichnung des Liedes *Halleluja* (oben). Vertikale Linien entsprechen den wahren Einsatzzeiten. Die horizontale Linie entspricht dem Schwellenwert $\delta = 0.85$. Unten ist die zeitliche Darstellung des Musikstückes gegeben.

kommt die Bildung überlappender Fenster in Frage (z.B. mit 10 ms Überlappung) bzw. die Definition eines dynamischen Schwellenwertes (anstatt des benutzten konstanten Schwellenwertes).

3 Anwendung der Einsatzzeitanalyse

In diesem Abschnitt werden Vorschläge zur Lösung des folgenden Problems vorgestellt: Analyse der Unterschiede bzw. Gemeinsamkeiten in dem Stil bei Interpretation der Schubert-Sonate (B-Dur, D 960). Folgende Einspielungen stehen dabei zur

Verfügung:

- Alfred Brendel (AB), Aufnahmejahr 1972,
- Sviatoslav Richter (SR), Aufnahmejahr 1972,
- Maria Judina (MJ), Konzertmitschnitt vom 13. August 1974.

Um die Spielweise der drei Interpreten miteinander vergleichen zu können, müssen ihre Einspielungen zuerst genau analysiert werden. Dies erfordert die Erkennung der Zeitpunkte aller Klaviertoneinsätze. Zurzeit existiert weder eine kommerzielle noch eine Open Source Software, die diese Aufgabe korrekt lösen könnte. Im ersten Schritt der Analyse wurden MIDI-Daten für das o.g. Musikstück⁵ mithilfe der Software Anvil Studio⁶ in Ereignislisten umgewandelt, die unter anderem Informationen darüber enthalten, welche Noten zu welchen Zeitpunkten gespielt werden (*Onset* bzw. *Offset*-Zeiten). Diese Ereignislisten sollen als a-priori Wissen für die Einsatzzeitanalyse benutzt werden. Folgendes Vorgehen wurde dabei als Erweiterung des auf Amplitudenschwankungen basierenden Algorithmus (s. Abschnitt 2.2) um das a-priori Wissen umgesetzt:

- Die erwarteten Einsatzzeiten (in Sekunden), die mithilfe der MIDI-Datei ermittelt wurden, wurden im Vektor *aprEinsatz* gespeichert.
- Das a-priori Intervall, in dem der *ite* Einsatz gesucht wurde, wurde bestimmt durch:

$$\begin{aligned} - \text{aprIntervallAnfang}[i] &= \text{aprEinsatz}[i] - p1 \cdot (\text{aprEinsatz}[i+1] - \text{aprEinsatz}[i]), \\ - \text{aprIntervallEnde}[i] &= \text{aprEinsatz}[i] + p2 \cdot (\text{aprEinsatz}[i+1] - \text{aprEinsatz}[i]). \end{aligned}$$

⁵http://www.piano-midi.de/schub_d.htm, Stand: 01.01.2010.

⁶<http://www.anvilstudio.com>, Stand: 01.01.2010.

- In dem a-priori Intervall wurde der Toneinsatz als Zeitpunkt des stärksten Amplitudenanstieges geschätzt ($geschätztEinsatz[i]$).
- $Differenz[i] = geschätztEinsatz[i] - aprEinsatz[i]$.
- Der Vektor der a-priori Einsatzzeiten wurde um die oben berechnete Differenz verändert:
 - $aprEinsatz = aprEinsatz + Differenz[i]$.

Problematisch dabei ist die Wahl der Koeffizienten $p1$ und $p2$, die widerspiegeln sollen, ob der Interpret schneller bzw. langsamer als die MIDI-Aufzeichnung spielt. Zum Vergleich spielt Alfred Brendel den ersten Satz (Molto Moderato) der Sonate in 14:35 Minuten, Sviatoslav Richter in 24:31 und Maria Judina in 21:59 Minuten. Die Dauer der MIDI-Aufzeichnung beträgt ungefähr 18 Minuten. Ein weiteres Problem ist, dass nicht kontrolliert werden kann, ob der im jeweiligen a-priori Intervall geschätzte Noteneinsatz korrekt ist, da folgende Situationen theoretisch auftreten können: In dem a-priori Intervall befinden sich tatsächlich mehrere Toneinsätze (d.h. der Interpret spielt schneller als erwartet) oder in dem a-priori Intervall befindet sich kein neuer Toneinsatz (d.h. der Interpret spielt langsamer als erwartet).

Weiterhin wurde versucht, Einsatzzeiten mittels des zweiten entwickelten Algorithmus zu bestimmen und das Ergebnis mit einem vergleichbaren Algorithmus aus der Literatur (*Spectral Flux*, s. [6]) zu vergleichen. Als Vergleichskriterium gilt der sogenannte F -Wert, der sich folgendermaßen berechnen lässt:

$$F = \frac{2c}{2c + f^+ + f^-},$$

wobei c , f^+ und f^- entsprechend die Anzahl der richtigen, falsch erkannten und nicht erkannten Einsätze sind. Für die von Alfred Bredel gespielten ersten sieben Takte der Sonate wurden die wahren Einsatzzeiten manuell bestimmt. Die beste Parameterkombination (laut einer Gitteroptimierung) für den Algorithmus *Spectral*

Flux ergab für diese Aufzeichnung den F -Wert von 0.74. Der entwickelte spektralbasierte Algorithmus lieferte für Parameter $L = 2048$ Samples und $\delta = 0.85$ den F -Wert von 0.78 und der oben beschriebene zeitbasierte Algorithmus den F -Wert von 0.76. Abbildung 4 zeigt die Ergebnisse des entwickelten spektralbasierten Algorithmus, wobei die wahren manuell bestimmten Einsatzzeiten mit vertikalen Linien markiert sind.

Anhand dieser Darstellung ist deutlich zu erkennen, dass die neuen Toneinsätze oft mit keinem bzw. nur einem sehr schwachen Amplitudeneinstieg verbunden sind. Auch sind dabei keine ausgeprägten Änderungen in der spektralen Struktur des Signals zu beobachten. Zu bemerken ist weiterhin, dass die Toneinsätze an diesen kritischen Stellen nur bei sehr starker Konzentration während des Hörens zu erkennen sind.

Die korrekte Bestimmung der Einsatzzeiten ist die Grundlage für weitere Vergleiche der Einspielungen der verschiedenen Komponisten. Da die wahren Toneinsätze in den ersten sieben Takten der drei Aufnahmen manuell bestimmt wurden, kann anhand dieser Daten eine vorläufige Analyse durchgeführt werden und zwar werden diese Einsatzzeiten mittels unterschiedlicher Distanzmaße miteinander verglichen (s. Tabelle 1). Dabei ist die Minkowski-Metrik für zwei m -dimensionale Vektoren \mathbf{x} und \mathbf{y} folgendermaßen definiert:

$$MM(\mathbf{x}, \mathbf{y}, p) = \left(\sum_{i=1}^m |x_i - y_i|^p \right)^{\frac{1}{p}}.$$

Die Minkowski-Metrik mit $p = 2$ wird oft als die Euklidische Distanz bezeichnet. Die Mahalanobis-Distanz berechnet sich durch

$$MD(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T C^{-1} (\mathbf{x} - \mathbf{y})},$$

wobei C die Kovarianzmatrix der Vektoren \mathbf{x} und \mathbf{y} ist.

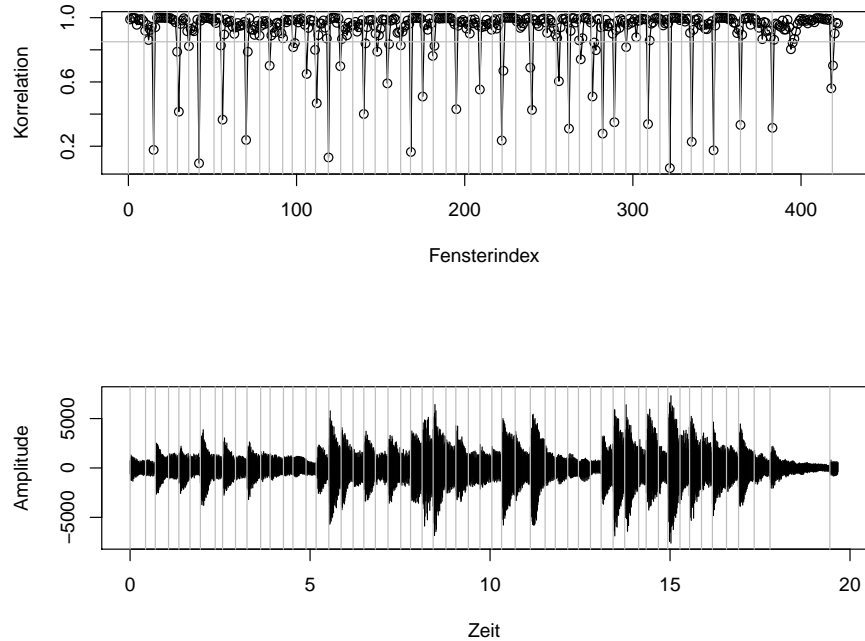


Abbildung 4: Die ersten sieben Takte der Schubert-Sonate in der Interpretation von Sviatoslav Richter überlappt mit den zugehörigen wahren Einsatzzeiten (unten). Oben ist die Detektorfunktion des in Abschnitt 2.3 vorgestellten Algorithmus gegeben (die horizontale Linie entspricht dem Schwellenwert).

Aus Tabelle 1 kann abgelesen werden, dass die Einsatzzeiten der ersten sieben Takte von Alfred Brendel den a-priori Zeiten (MIDI-Zeiten) am ähnlichsten und die von Maria Judina am unähnlichsten sind. Am ähnlichsten spielen dann Alfred Brendel und Sviatoslav Richter und die größten Distanzen weisen die Einsatzzeiten von Alfred Brendel und Maria Judina auf.

Zu bemerken ist, dass der drittkleinste Wert der Euklidischen Distanz (35.884) dem viertkleinsten der Mahalanobis-Distanz (291.878) entspricht (und umgekehrt), d.h. dass unterschiedliche Distanzmaße teilweise unterschiedliche Ergebnisse liefern.

<i>Verhältnis</i>	<i>Euklidische Distanz</i>	<i>Mahalanobis-Distanz</i>
MIDI zu AB	11.771	58.174
MIDI zu SR	37.452	213.777
MIDI zu MJ	73.142	475.084
AB zu SR	25.701	159.111
AB zu MJ	61.409	432.543
SR zu MJ	35.884	291.878

Tabelle 1: Distanzmaße zwischen den Einsatzzeiten der drei Interpreten sowie den MIDI-Zeiten für die ersten sieben Takte der Schubert-Sonate.

4 Zusammenfassung

In diesem Bericht wurde das Thema der Noteneinsatzerkennung bei Musikaufzeichnungen diskutiert. Zuerst wurde der Stand der Forschung zu diesem Problem diskutiert und anschließend zwei Algorithmen zur Einsatzzeitdetektion als Einarbeitung in den Untersuchungsgegenstand entwickelt. Anschließend wurde eine Anwendung vorgestellt, die auf der Analyse der Einsatzzeiten bei polyphonen Klavieraufnahmen basiert. Dabei wurden unterschiedliche Ansätze zur Lösung des Problems getestet und verglichen. Wegen der nicht zufriedenstellenden Ergebnisse ist es notwendig, die Parameter der gegebenen Algorithmen besser auf das untersuchte Klavierstück anpassen zu können. Diese Problematik kann zu folgender Zielsetzung verallgemeinert werden: Bestimmung der optimalen Parameterkombination für die jeweiligen Algorithmen in Abhängigkeit von den spielenden Instrumenten bzw. Musikarten. Diese Aufgabe ist mittels statistischer Versuchsplanung realisierbar. Die Intention der weiteren Arbeit ist daher zunächst, für einen geeigneten Algorithmus zur Einsatzzeitdetektion ein statistisches Modell zur Vorhersage der Einsatzzeit in Abhängigkeit von den Algorithmusparametern aufzustellen und dann diese Parameter mit Hilfe eines geeigneten Versuchsplan mit möglichst wenigen Versuchen zu optimieren.

Danksagung

Diese Arbeit ist unterstützt worden von der DFG (SFB 823, Statistik nichtlinearer dynamischer Prozesse, Teilprojekt C2: Optimale Versuchsplanung für dynamische statistische Modelle)

Literatur

- [1] Abdallah S. A., Plumbley M. D. (2003): *Probability as metadata: event detection in music using ICA as a conditional density model*. In Proc. 4th Int. Symp. Independent Component Analysis and Signal Separation (ICA2003), Nara, Japan, pp. 233-238.
- [2] Bello J., Daudet L., Abdallah S., Duxbury C., Davies M., Sandler M.(2005): *A tutorial on onset detection in musical signals*. IEEE Trans. Speech and Audio Proc., vol. 13, no. 5, pp. 1035-1047.
- [3] Boogaart, C. G. v. d., Lienhart, R. (2009): *Note onset detection for the transcription of polyphonic piano music*. IEEE International Conference on Multimedia and Expo 2009 (ICME09), New York, USA, pp. 446-449.
- [4] de Cheveigné, A. (2006): *Multiple f_0 estimation*. In Computational Auditory Scene Analysis: Principles, Algorithms and Applications, D.-L. Wang and G. J. Brown, Eds. IEEE Press, Wiley.
- [5] Christensen, M. G., Jakobsson, A. (2009): *Multi-pitch estimation*. Morgan & Claypool Publishers.
- [6] Dixon, S. (2006): *Onset detection revisited*. In Proc. DAFx-06, Sept. 18-20, pp. 133-137.
- [7] Hoyer, P. O. (2004): *Non-negative matrix factorization with sparseness constraints*. Journal of Machine Learning Research 5, pp. 1457-1469.
- [8] Klapuri, A. (2006): *Multiple fundamental frequency estimation by summing harmonic amplitudes*. 7th International Conference on Music Information Retrieval (ISMIR-06), Victoria, Canada.
- [9] Lee, W.-C., Jay Kuo, C.-C. (2006): *Musical onset detection based on adaptive linear prediction*. Proc. International Conference on Multimedia and Expo (ICME06), Toronto, Canada.

- [10] Röbel, A. (2005): *Onset detection in polyphonic signals by means of transient peak classification*. In MIREX Online Proceedings (ISMIR 2005), London, Great Britain.
- [11] Thornburg, H., Leistikow, R. J., Berger, J. (2007): *Melody extraction and musical onset detection via probabilistic models of framewise STFT peak data*. IEEE Trans. Audio Speech Language Process., vol. 15, no. 4, pp. 1257-1272.

