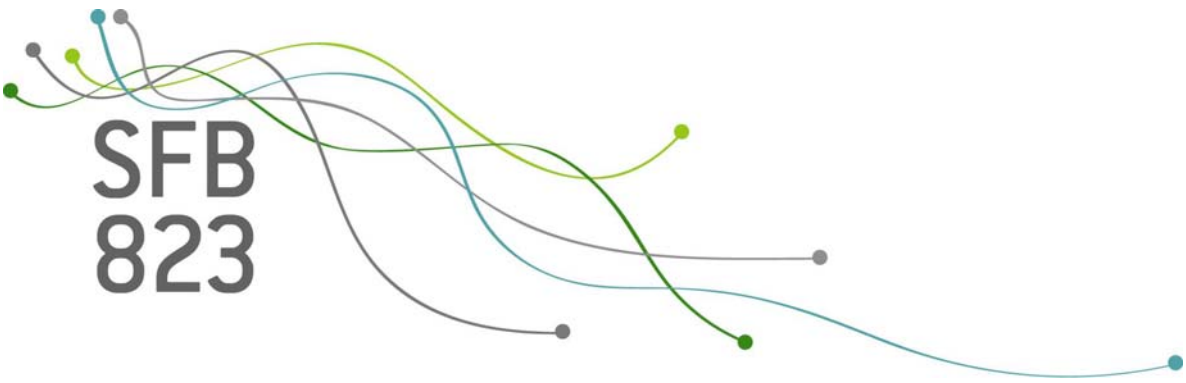
SFB 823

Discussion Paper

Einfluss der Musikinstrumente auf die Güte der Einsatzzeiterkennung

Nadja Bauer, Julia Schiffner, Claus Weihs

Nr. 10/2012



Einfluss der Musikinstrumente auf die Güte der Einsatzzeiterkennung

Nadja Bauer, Julia Schiffner, Claus Weihs

Lehrstuhl Computergestützte Statistik
Fakultät Statistik
TU Dortmund
E-mail: {bauer,schiffner,weihs}@statistik.tu-dortmund.de

Abstract

Erkennung der Toneinsätze in Musikaufnahmen ist der erste und sehr wichtige Schritt bei der Musiktranskription. Es existieren bereits sehr viele Algorithmen, die diesem Ziel dienen. Bei den meisten davon werden die Algorithmusparameter mittels genetischer Verfahren so optimiert, dass diese Algorithmen für alle Musikinstrumente durchschnittlich die besten Erkennungsraten liefert. Dabei sind die Klangeigenschaften von Instrumenten verschiedener Instrumentenarten sehr unterschiedlich, so dass es sinnvoll ist, optimale Parametereinstellungen in Abhängigkeit von Instrumentenklassen zu bestimmen. Bei Musikstücken, die von mehreren Instrumenten gespielt werden, ist dieses Problem allerdings komplizierter.

Ziel dieses Berichtes ist, einen einfachen Algorithmus zur Einsatzzeiterkennung auf Tonabfolgen verschiedener Musikinstrumente zu testen, um zunächst grobe Zusammenhänge zwischen Einstellungen von Algorithmusparametern und Instrumentenart zu bestimmen. Wegen großer Unterschiede zwischen echten und synthetischen Musiktönen bezüglich ihrer Klangeigenschaften wird großer Wert auf die Erzeugung von Tonabfolgen mittels echter Tonaufnahmen gelegt. Dazu wird in dieser Arbeit ein Verfahren vorgestellt.

1 Einleitung

Diese Arbeit ist eine Vorstudie für das Projekt „Versuchsplanung in der Signalanalyse“, das im Rahmen des SFB 823, Teilprojekt C2, unterstützt wird. Ziel des Projektes „Versuchsplanung in der Signalanalyse“ ist es, Algorithmen zur Audiosignalverarbeitung mittels Methoden der Versuchsplanung zu optimieren (vgl. dazu Bauer et al., 2011).

Im zweiten Kapitel wird der benutzte Algorithmus zur Einsatzzeiterkennung kurz eingeführt. In Kapitel 3 wird zum einen ein neues Verfahren zur Erstellung der Musikstücke mittels Echttonaufnahmen vorgestellt. Zum anderen wird erklärt, nach welchem Prinzip die experimentellen Tonabfolgen generiert werden.

Im vierten Kapitel wird der erste Teil der Vorstudie ausgeführt, mit dem Ziel, zwei Verfahren der Einsatzzeiterkennung für unterschiedliche Musikinstrumente zu testen und dabei den Einfluss einiger freier Parameter sowie des Instrumentes auf die Einsatzzeiterkennungsrate zu beobachten. Im zweiten Teil der Vorstudie werden

unterschiedliche Kombinatonsmöglichkeiten von diesen zwei Verfahren in Bezug auf die Erkennungsrate bei verschiedenen Musikinstrumenten bewertet (s. Kapitel 5). Das sechste Kapitel fasst die Arbeit zusammen, wobei anhand der Ergebnisse der durchgeführten Vorstudie die endgültigen Einstellungen des Algorithmus zur Einsatzzeiterkennung festgelegt werden.

2 Algorithmus der Einsatzzeiterkennung: kurze Vorstellung

Der benutzte Algorithmus basiert auf zwei in Bauer et al. (2010) vorgeschlagenen Verfahren: Bei dem ersten Verfahren wird die Steigung der Amplitude und bei dem zweiten Verfahren die Veränderung der spektralen Struktur eines Audiosignals als Indikatoren für Toneinsätze verwendet. Der Parameter ***M*** gibt an, welcher der beiden Indikatoren benutzt wird. Das Audiosignal wird gefenstert, wobei die Länge (***L***) des Fensters und der prozentuale Anteil der Überlappung (***U***) als zwei Parameter in den Algorithmus eingehen. In jedem Fenster werden zwei Merkmale erhoben: Differenz zwischen Amplitudenmaxima (***M1***) bzw. Korrelationskoeffizient zwischen den Spektren (***M2***) des aktuellen und des vorherigen Fensters. Die Vektoren für ***M1*** bzw. ***M2*** werden dann jeweils auf das Intervall [0,1] skaliert.

Der vierte Parameter ist der Schwellenwert (***S***) aus dem Intervall [0,1]. Allen Fenstern, in denen ***M1*** (bzw. ***M2***) den Schwellenwert übersteigt, wird 1, sonst 0 zugeordnet. Sollten Sequenzen von 1-Werten vorkommen (d.h. mehrere Nachbarfenster haben den Wert 1), wird in demjenigen Fenster ein Toneinsatz vermerkt, der den maximalen Wert von ***M1*** (bzw. ***M2***) aufweist (lokale Maxima werden gesucht). Der Vektor der Toneinsätze ***E*** enthält dann nur für diejenige Fenster eine 1, in denen die Toneinsätze geschätzt wurden (s. Beispiel in Tabelle 1).

Tabelle 1 Berechnung des Vektors ***E*** für den Schwellenwert ***S***=0.3

Fensternummer	1	2	3	4	5	6	7
M1-Wert	0.34	0.67	0.13	0.09	0.56	0.67	0.98
(***M1*** $>$ ***S***)?	1	1	0	0	1	1	1
E	0	1	0	0	0	0	1

Anschließend werden die genauen Stellen der Toneinsätze bestimmt, wobei Fenster, in denen der Vektor ***E*** den Wert 1 hat, betrachtet werden. Hier werden zwei Methoden der Einsatzzeitschätzung vorgeschlagen und verglichen. Bei der ersten Methode gelten die Endzeitpunkte der Fenster mit dem Wert ***E***=1 als die Toneinsätze. Als Motivation, warum das Ende der jeweiligen Fenster genommen wird, gilt, dass die Einsätze natürlicherweise etwas später erkannt werden, als sie stattfinden (wie z.B. beim menschlichen Ohr). Der Toneinsatz gilt in dieser Arbeit als richtig erkannt, falls er höchstens 50 ms (2205 Samples bei einer Samplingrate von 44100 Hz) von dem wahren Einsatzzeitpunkt entfernt ist (s. Dixon, 2006). Diese

Tatsache motiviert die zweite Methode der Toneinsatzschätzung: Toneinsätze sind die Mittelzeitpunkte der Fenster mit dem Wert $\boldsymbol{E}$=1. Es wird erwartet, dass diese Methode bei großer Fensterlänge eine bessere Erkennungsgüte liefert als die erste Methode.

Die Erkennungsgüte wird dann mit Hilfe des F-Werts berechnet, der sowohl die Anzahl der richtig erkannten Einsätze als auch die Anzahlen von fälschlicherweise und nicht erkannten Einsätzen berücksichtigt (s. Dixon, 2006). Der F-Wert liegt immer zwischen 0 und 1, wobei 1 einer perfekten Erkennungsrate entspricht.

3 Musikdatenbank

3.1 Erzeugung der WAVE-Files aus MIDI-Files mittels echter Töne

Wenn Algorithmen der Audiosignalverarbeitung optimiert werden sollen, stellt sich immer die Frage einer geeigneten Musikdatenbank. Diese Datenbank soll Aufnahmen enthalten, zu denen die Information über die interessierenden Ereignisse (wie z.B. Einsatzzeiten, Musikinstrumente, Tonhöhen usw.) vorhanden ist. Eine Möglichkeit ist, MIDI-Files, die diese Information enthalten, mittels frei verfügbarer Software zu WAVE-Files zu konvertieren (z.B. MIDI to WAVE Maker[1]). Der Vorteil ist, dass die Arbeit mit MIDI-Daten viel Flexibilität bietet: man kann einige Musikinstrumente ohne Mühe entfernen oder zu anderen Instrumenten umwandeln lassen, Tempo und Lautstärke beliebig einstellen usw. Der Nachteil ist aber, dass bei der Konvertierung mittels dieser Software lediglich synthetische Töne benutzt werden, die sich in vielen Eigenschaften von den echten Tönen unterscheiden.

Eine andere Möglichkeit besteht darin, Bibliotheken echter Musikaufnahmen zu finden, für welche die interessierenden Merkmale manuell notiert wurden (z.B manuelle Erkennung der Toneinsatzzeiten). In diesem Fall ist einerseits damit zu rechnen, dass solche Bibliotheken nicht in großem Umfang vorhanden sind, und andererseits muss auf die oben erwähnte Flexibilität verzichtet werden. Einen optimalen Ausweg würde eine Software bieten, die MIDI-Files zu WAVE-Files mittels echter Töne konvertiert.

Um so ein Programm zusammenzustellen, werden Aufnahmen von echten Tönen als Grundlage gebraucht. Das erste große Problem ist dann die Gewinnung der Töne der gewünschten Länge aus den Originaltonaufnahmen. Wenn man nämlich einen Ton (besonders bei Blasinstrumenten) an einer beliebigen Stelle abschneidet, ruft dies unerwünschte Audiogeräusche hervor. Der Ton muss also (wie auch bei echten Instrumenten) abklingen und darf nicht abgeschnitten werden.

Das zweite Problem ist die Zusammensetzung der Einzeltöne zu einer Tonabfolge, insbesondere die Bindung der Töne. Sollten die Töne in einer Tonabfolge

[1] http://www.computerbild.de/download/MIDI-To-WAV-Maker-913121.html, Stand: 01.01.2012

ohne akustische Unterbrechung erklingen, müsste berücksichtigt werden, aus welchem Tonabschnitt der Originaltonaufnahme der jeweilige Ton geschnitten werden soll. Weiter sollten diese Tonabschnitte so miteinander verknüpft werden, dass die Schnittstellen von keinen unerwünschten Audiogeräuschen begleitet werden.

Außer diesen beiden genannten Problemen gibt es eine Reihe von anderen Ursachen, die zwangsläufig dazu führen, dass ein generiertes Musikstück niemals einer echten Aufnahme ähnelt. Da das Ziel dieser Arbeit nicht die Erzeugung von ästhetisch klingenden Musikstücken ist, sondern von solchen, die für das Trainieren der Algorithmen zur Audiosignalverarbeitung geeignet sind, wird hier ein Vorgehen implementiert, das diesem Ziel dient.

Echte Tonaufnahmen wurden der kommerziellen Tonbibliothek *RWC music database* entnommen (Goto et al., 2003). Diese Datenbank bietet Töne verschiedener Musikinstrumente in mehreren Ausführungen an: zu jeder Tonstärke (Forte, Mezzoforte und Piano) gibt es jeweils mehrere Spielweisen (z.B. Normal, Staccato und Vibrato). Außerdem sind für jedes Musikinstrument die Töne von mindestens zwei Exemplaren dieses Instruments aufgenommen worden.

Für diese Arbeit wurden folgende Musikinstrumente berücksichtigt: Klavier, Gitarre, Geige, Flöte, Klarinette und Trompete (s. Abschnitt 3.2).

Realisierung

Hier werden einige Lösungen zu den entstandenen Problemen bei der Realisierung der Software zur Erzeugung der WAVE-Files mittels echter Tonaufnahmen diskutiert.

Bezüglich des Abschneidens der Töne werden verschiedene Vorgehensweisen verfolgt. Der Grund dafür ist, dass die Saiten- bzw. Zupfinstrumente in der Regel langsamer als z.B. die Blasinstrumente abklingen. Im Laufe der Experimente haben sich zwei Abschneidetechniken herauskristallisiert: *Typ 1* und *Typ 2*. Bei der *Typ 1*-Technik wird die Amplitude des Tonsignals ab dem gewünschtem Endzeitpunkt mit der Funktion $2^{-1.5 \cdot 10^{-5} x}$ gefaltet und bei der *Typ 2*-Technik mit der Funktion $\sqrt[500]{\chi_2(x)}$, wobei $\chi_2(x)$ die Dichte der Chi-Quadrat-Verteilung mit dem Freiheitsgrad 2 ist. In beiden Fällen darf der Ton maximal 3 in der MIDI-Datei vorgeschriebene Tonlängen (nach dem Endzeitpunkt) abklingen. Die Technik *Typ 1* wird für Klavier und Gitarre und die Technik *Typ 2* für Trompete, Flöte, Klarinette und Geige verwendet.

Weiterhin wird unterschieden, wie lange ein Ton klingen soll. Bei einer Tondauer von weniger als 150 Millisekunden wird der Ton aus der Staccato-Bibliothek genommen, sonst aus der Normal-Bibliothek. Der Nutzer muss sich allerdings vorab für die Tonstärke entscheiden, d.h. ob die Töne aus der Mezzoforte-, Forte- oder Piano-Bibliothek genommen werden. Allerdings wird die Lautstärke bei der Konvertierung in Abhängigkeit von MIDI-File-Eingaben variiert, indem die Tonamplitude entsprechend umskaliert wird.

Bei Geige und Gitarre tritt das Problem auf, dass manche Töne auf mehreren Saiten gespielt werden können. Geige hat beispielsweise vier Saiten: G3[2], D4, A4 und E5, wobei G3 die tiefste und E5 die höchste Saite ist. Sollte ein Ton auf zwei Saiten spielbar sein, wird dieser von der höheren Saite genommen. Beispielsweise wird für den Ton D4, der sowohl auf der Saite G3 als auch auf der Saite D4 vorkommt, die Seite D4 ausgewählt. Für Gitarre ist das Vorgehen analog.

Zu bemerken ist, dass es unmöglich ist, Tonbindungen (bzw. Legato) automatisch zu modellieren. Auch bei professioneller kommerzieller Software für die Musikerzeugung - Vienna Symphonic Library [3] - muss der Anwender bei jedem einzelnen Ton selbst entscheiden, aus welcher Tonbibliothek dieser kommen soll (es gibt dann zusätzlich Bibliotheken von Legato-Tönen). Auf Legato-Modellierung wird aus diesem Grund bei der vorliegenden Realisierung ganz verzichtet.

Das Programm zur Konvertierung von MIDI-Files in WAVE-Files mittels echter Töne wurde mit Hilfe der Programmiersprache R (R Development Core Team, 2008) implementiert.

3.2 Erzeugung der Tonabfolgen

Da das Ziel dieser Vorstudie ist, Einflüsse verschiedener Instrumente auf die Toneinsatzerkennungsrate zu ermitteln, wird angestrebt, dass sich die Musikstücke lediglich in der Instrumentenbesetzung unterscheiden. Für diesen Zweck sollten solche Musikstücke gefunden werden, die theoretisch von allen interessierenden Instrumenten gespielt werden können. In Tabelle 2 sind die betrachteten Instrumente sowie deren Tonumfänge als Notenintervalle bzw. als zugehörige MIDI-Codierung vorgestellt. Dabei beziehen sich die Angaben zu den Tonumfängen lediglich auf die vorhandenen Daten. Zum Teil können mit einzelnen Instrumenten tiefere oder höhere Töne erzeugt werden.

Tabelle 2 Instrumente und in der Datenbank vorhandene Töne

Instrument	Tonumfang als Noten	Tonumfang als MIDI-Codierung
Flöte	C4-C7	60-96
Geige	G3-E5	55-100
Gitarre	E2-E5	40-76
Klarinette	D3-F6	50-89
Klavier	A0-C8	21-108
Trompete	E3-D6	52-86

Wie aus Tabelle 2 errechnet werden kann, besteht der gemeinsame Tonumfang aus lediglich 17 Tönen: C4 bis E5. Da es sehr unwahrscheinlich ist, für diesen To-

[2] Wir benutzen hier die englische Notation für die Tonhöhe

[3] http://vsl.co.at/, Stand 01.01.2012.

numfang mehrere Musikstücke zu finden, werden Tonabfolgen zufällig generiert. Die Simulationseinstellungen sind Tabelle 3 zu entnehmen. Ausprägungen der Variablen *Lautstärke* und *Note* sind auf die MIDI-Kodierung abgestimmt. Bei der ersten Tondauereinstellung werden langsame Abfolgen simuliert mit durchschnittlich 2 Schlägen pro Sekunde (oder 120 bpm[4]) und bei der zweiten schnelle Abfolgen mit 5 Schlägen pro Sekunde (oder 300 bpm). Es wurden insgesamt pro Musikinstrument und Tondauereinstellung 5 Tonabfolgen mit 100 Toneinsätzen zufällig generiert. Bei der Tondauereinstellung mit 120 bmp dauern die Tonabfolgen durchschnittlich 50 Sekunden und bei der schnelleren Tondauereinstellung mit 300 bpm durchschnittlich 20 Sekunden.

Tabelle 3 Einstellungen für die zufällige Generierung der Tonabfolgen

Variable	Verteilung
Lautstärke	Gleichverteilung auf [70,90]
Note	Gleichverteilung auf [60,76]
Tondauer	1. absolute Werte einer Normalverteilung mit $\mu = 0.5$, $\sigma = 0.2$ 2. absolute Werte einer Normalverteilung mit $\mu = 0.2$, $\sigma = 0.1$

4 Erster Teil der Vorstudie

4.1 Zusammenfassung der Einflussgrößen

Bevor mit der Auswertung begonnen wird, werden die betrachteten Einflussgrößen mit ihren Einstellungen aufgelistet. Dabei handelt es sich nur um einen groben Versuchsplan, der lediglich auf die Richtung des Zusammenhanges zwischen diesen Einflussgrößen und der Erkennungsrate hindeuten soll, aber auf keinen Fall ein Optimum liefern kann. Die Einflussgrößen sind also die Folgenden:

1. Musikinstrument (Geige, Flöte, Gitarre, Klarinette, Klavier und Trompete),
2. Tondauer (120 bpm und 300 bpm),
3. Fensterlänge $\boldsymbol{L}$ (512, 1024, 2048 und 4096 Samples),
4. Überlappung $\boldsymbol{U}$ (0% und 50%),
5. Merkmal (***M1***: *Amplitude*, ***M2***: *Spektrum*),
6. Zeitberechnungsmethode (*Fensterende*, *Fenstermitte*) .

Es werden alle möglichen Kombinationen dieser 6 Einflussgrößen auf 100 Tonabfolgen getestet. Um den Parameter $\boldsymbol{S}$ (Schwellenwert) nicht berücksichtigen zu

[4] bpm (*beats per minute*): Schläge pro Minute

müssen, werden für jedes Experiment $\boldsymbol{S}$-Werte aus dem Intervall von 0.1 bis 0.9 (mit dem Schritt 0.1) ausprobiert und nur das beste Ergebnis zurückgegeben. Die Verteilung der zugehörigen F-Werte über die Tonabfolgen für die jeweiligen Einstellungen wird mit Hilfe von Boxplots veranschaulicht.

4.2 Auswertung

In diesem Abschnitt werden Ergebnisse der Simulation dargestellt. Es werden hier nur die Abbildungen vorgestellt, die für die Analyse wichtig sind und neue Information liefern. Alle anderen Abbildungen befinden sich im Anhang.

Klavier und Gitarre

Als erstes werden die Ergebnisse für Klavier-Stücke ausgewertet. In Abbildungen 1 bzw. 2 sind die Verteilung der F-Werte für die langsamen Stücke (mit 120 bpm) und Zeitberechnungsmethode *Fenstermitte* bzw. *Fensterende* dargestellt. Folgendes ist dabei zu betrachten:

1. Die amplitudenbasierte Methode ist sowohl ohne als auch mit Überlappung besser als die spektralbasierte Methode.
2. Für $\boldsymbol{L}$=512 und $\boldsymbol{L}$=1024 verschlechtert die Überlappung die Ergebnisse und für $\boldsymbol{L}$=2048 und $\boldsymbol{L}$=4096 verbessert sie sie.
3. Die Zeitschätzungsmethode *Fensterende* weist bessere Ergebnisse für $\boldsymbol{L}$=512 bis $\boldsymbol{L}$=2048 mit geringerer Varianz auf. Dagegen schneidet die Zeitschätzungsmethode *Fenstermitte* für $\boldsymbol{L}$=4096 besser ab.
4. Für alle Fensterlängen kann man eine Parameterkombination finden, die akzeptable Ergebnisse liefert (F-Wert größer als 0.9).

Weiter ist in Abbildung 3 die Verteilung der F-Werte für die schnellen Klavier-Stücke zu sehen (mit 300 bpm und Zeitberechnungsmethode *Fenstermitte*). Dabei kann keine Tendenz beobachtet werden (sowohl für die Zeitberechnungsmethode *Fenstermitte* als auch für *Fensterende* (s. Abbildung A1)).

Deutlich zu erkennen ist aber, dass die Güte der Einsatzzeiterkennung von dem Tempo des Musikstückes abhängt. Langsame Tonabfolgen (mit 120 bpm) (Abbildungen 1, 2) weisen deutlich bessere F-Werte auf als schnelle Tonabfolgen (mit 300 bpm) (Abbildungen 3, A1). Dies gilt für alle nachfolgenden Instrumente und wird nicht weiter erwähnt.

Die beschriebenen Besonderheiten von Klavier-Stücken stimmen mit denen der Gitarre-Stücke überein (vgl. Abbildungen A2, A3, A4, A5). Dies verdeutlicht die Nähe der beiden Instrumente zueinander. Klavier und Gitarre gehören beide zu der Klasse der Saiteninstrumente. Die Gitarre ist ein Zupfinstrument und bei Klavier werden die Saiten angeschlagen.

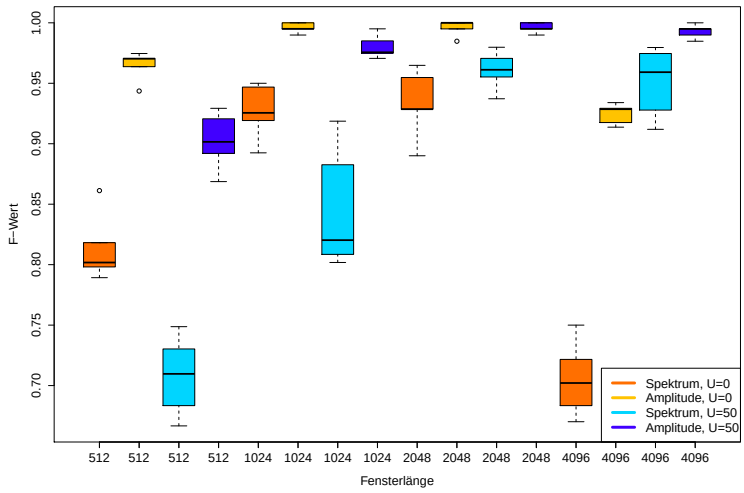


Abb. 1 Verteilung der F-Werte für Klavier, Tempo: 120 bpm, Zeitberechnungs: *Fenstermitte*

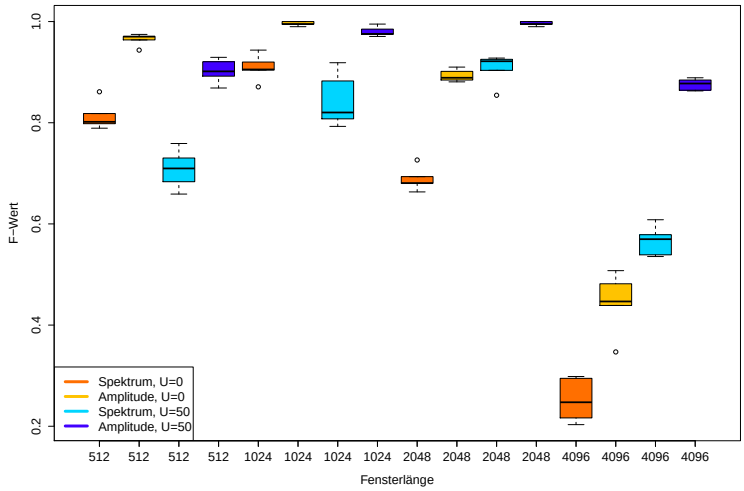


Abb. 2 Verteilung der F-Werte für Klavier, Tempo: 120 bpm, Zeitberechnung: *Fensterende*

Flöte, Klarinette und Trompete

In Abbildung 4 sind die Ergebnisse für die Flöte veranschaulicht (Tempo: 120 bmp, Zeitberechnungsmethode *Fenstermitte*).

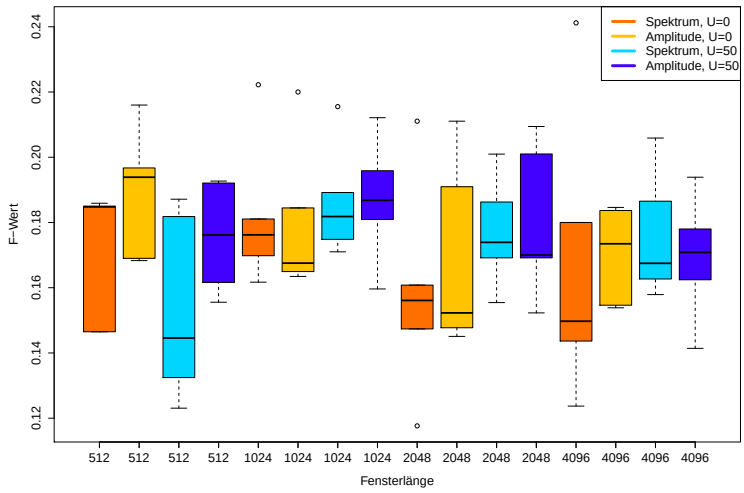


Abb. 3 Verteilung der F-Werte für Klavier, Tempo: 300 bpm, Zeitberechnung: *Fenstermitte*

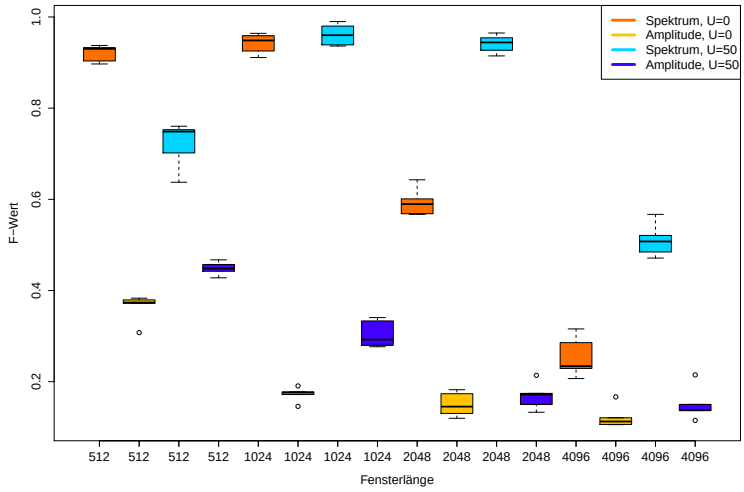


Abb. 4 Verteilung der F-Werte für Flöte, Tempo: 120 bpm, Zeitberechnung: *Fenstermitte*

Dabei ist für die langsamen Tonabfolgen Folgendes zu sehen (vgl. dazu auch Abbildung A6):

1. Die spektralbasierte Methode ist sowohl ohne als auch mit Überlappung besser als die amplitudenbasierte Methode (mit wenigen Ausnahmen).

2. Für $\boldsymbol{L}$=512 verschlechtert die Überlappung die Ergebnisse, und für $\boldsymbol{L}$=1024 bis $\boldsymbol{L}$=4096 verbessert sie sie.
3. Die Zeitschätzungsmethode *Fenstermitte* scheint für alle Fensterlängen bessere Ergebnisse zu liefern als die Methode *Maximum*.
4. Nur für die Fensterlängen $\boldsymbol{L}$=512 und $\boldsymbol{L}$=1024 (und bei der Zeitberechnungsmethode *Fenstermitte* auch für $\boldsymbol{L}$=2048) kann man eine Parameterkombination finden, die akzeptable Ergebnisse liefert.

Für schnelle Tonabfolgen ist eine unerwartete Tendenz zu beobachten: amplitudenbasierte Einsatzzeiterkennung zeigt bessere Erkennungsraten als die spektralbasierte und zwar mit und ohne Überlappung (vgl. Abildungen 5, A7). Für Fensterlängen $\boldsymbol{L}$=1024 und $\boldsymbol{L}$=2048 sind im Vergleich zu anderen Längen-Einstellungen relativ gute Ergebnisse zu vermerken. Zwischen den Zeitschätzungsmethoden *Fenstermitte* und *Fensterende* scheint kein großer Unterschied vorzuliegen.

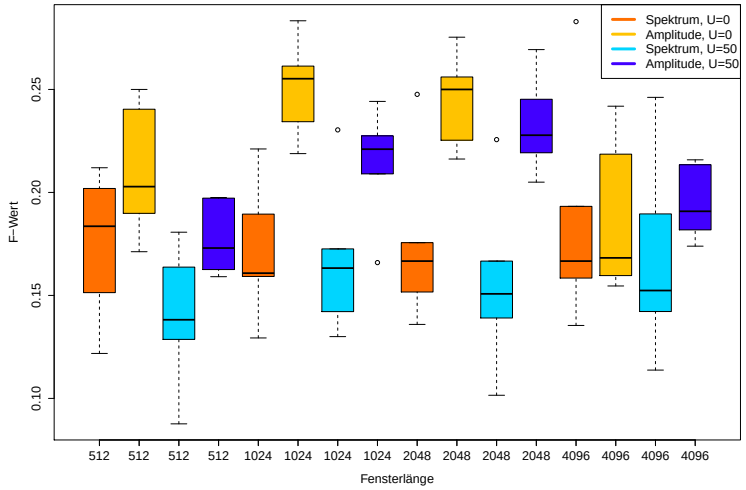


Abb. 5 Verteilung der F-Werte für Klavier, Tempo: 300 bpm, Zeitberechnungsmethode: *Fensterende*

In Bezug auf die langsamen Tonabfolgen stimmen die oben diskutierten Ergebnisse für die Musikinstrumente Trompete und Klarinette mit denen für Flöte überein (vgl. Abbildungen A8, A9, A12, A15). Bei den schnellen Tonabfolgen dagegen ist bei Trompete eine ähnliche Tendenz wie bei Flöte zu betrachten (die amplitudenbasierte Einsatzzeiterkennung ist besser als die spektralbasierte), während bei Klarinette das umgekehrte (und somit erwartete) Verhalten vorliegt (vgl. Abbildungen A10, A11, A14, A15).

Die Ähnlichkeiten zwischen Flöte, Klarinette und Trompete erklären sich durch ihre Zugehörigkeit zu derselben Instrumentenfamilie der Blasinstrumente.

Geige

Die Geige gehört zu der Instrumentenklasse der Streichinstrumente. In Abbildung 6 ist die Verteilung der F-Werte für die langsamen Geigen-Stücke veranschaulicht (Zeitschätzungsmethode *Fensterende*). Dabei kann Folgendes beobachtet werden (vgl. auch Abbildung A16):

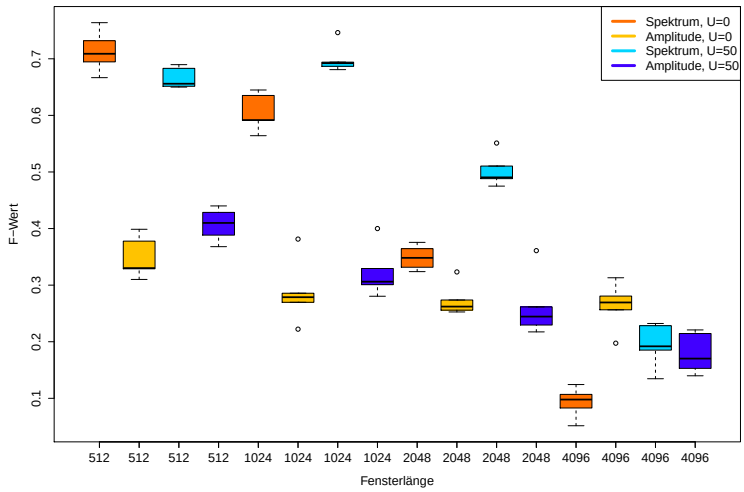


Abb. 6 Verteilung der F-Werte für Geige, Tempo: 120 bpm, Zeitberechnungsmethode: *Fenstermitte*

1. Die spektralbasierte Methode ist im Bereich der guten Erkennungsraten sowohl ohne als auch mit Überlappung besser als die amplitudenbasierte Methode.
2. Für $\boldsymbol{L}$=512 verschlechtert die Überlappung die Ergebnisse und für $\boldsymbol{L}$=1024 bis $\boldsymbol{L}$=4096 verbessert sie sie.
3. Nur für die Fensterlängen $\boldsymbol{L}$=512 und $\boldsymbol{L}$=1024 kann man eine Parameterkombination finden, die akzeptable Ergebnisse liefert.

Bei den schnellen Tonabfolgen ist zu erkennen, dass die amplitudenbasierte Einsatzzeiterkennung bessere Erkennungsraten zeigt als die spektralbasierte und zwar mit und ohne Überlappung (vgl. Abbildungen A17, A18)

Die beschriebenen Besonderheiten von Geige-Stücke ähneln den Besonderheiten der Blasinstrumente.

Fazit

Es werden sechs Musikinstrumente betrachtet, wobei diese bezüglich der Ergebnisse in zwei Klassen aufgeteilt werden können. Klavier und Gitarre werden demnach einer Klasse mit folgenden Besonderheiten zugeordnet: Amplitudenbasierte Einsatzzeiterkennung ist besser als die spektralbasierte, und für alle Fensterlängen können Parametereinstellungen gefunden werden, die akzeptable Ergebnisse lieferten. Alle drei Blasinstrumente (Flöte, Klarinette und Trompete) sowie Geige werden der zweiten Klasse zugeordnet: Spektralbasierte Einsatzzeiterkennung ist besser als die amplitudenbasierte und nur für kurze Fensterlängen ($\boldsymbol{L}$=512 und $\boldsymbol{L}$=1024) können Parametereinstellungen gefunden werden, die akzeptable Ergebnisse liefern.

5 Zweiter Teil der Vorstudie: Kombination der Merkmale *M1* und *M2*

5.1 Definition der Kombinationsmethoden

Im ersten Teil der Vorstudie wurden die Merkmale ***M1*** (amplitudenbasierte Einsatzzeiterkennung) und ***M2*** (spektralbasierte Einsatzzeiterkennung) unabhängig voneinander betrachtet. Nun werden Möglichkeiten einer Kombination untersucht. Hier werden allerdings nicht alle Fensterlängen, sondern nur die kurzen Fenster ($\boldsymbol{L}$=512 und $\boldsymbol{L}$=1024) berücksichtigt. Weiterhin werden nur die langsamen Stücke (mit 120 bpm) sowie nur die Zeitberechnungsmethode *Fenstermitte* berücksichtigt. Zur Erinnerung: Die Merkmalsvektoren ***M1*** und ***M2*** werden jeweils auf das Intervall [0,1] skaliert (s. Kapitel 2). Die vier vorgeschlagenen Merkmalskombinationen werden in Tabelle 4 vorgestellt.

Für zwei Kombinationsmethoden - *Additiv* und *Multiplikativ* - wird ein weiterer Parameter $\boldsymbol{A}$ auf dem Intervall [0,1] definiert, der den Anteil des jeweiligen Merkmals an der gesamten Summe bzw. das Produkt regelt. Somit wird diese Summe bzw. Produkt auch im Intervall [0,1] liegen. Das Gleiche gilt auch für die Kombinationsmethoden *Maximum* und *Minimum*. Das sich ergebende Merkmal ***Komb*** wird im Weiteren mit dem Schwellenwert $\boldsymbol{S}$ verglichen, und der Einsatzzeitvektor $\boldsymbol{E}$ wird genau so bestimmt, wie im Kapitel 2 bereits veranschaulicht.

Tabelle 4 Ansätze zur Merkmalskombination

Methode	Zusammensetzung
Additiv	$\boldsymbol{Komb} = \mathbf{A} \cdot \boldsymbol{M1} + (1 - \boldsymbol{A}) \cdot \boldsymbol{M2}$
Multiplikativ	$\boldsymbol{Komb} = \boldsymbol{M1}^{\mathbf{A}} \cdot \boldsymbol{M2}^{1-\boldsymbol{A}}$
Maximum	$\boldsymbol{Komb} = max(\boldsymbol{M1}, \boldsymbol{M2})$
Minimum	$\boldsymbol{Komb} = min(\boldsymbol{M1}, \boldsymbol{M2})$

Genauso wie im ersten Teil der Vorstudie wurde hier der Parameter ***S*** im Intervall [0.1,0.9] in 0.1 Schritten variiert und für jede Tonabfolge wurde nur die beste Erkennungsrate berücksichtigt. Jede Parametereinstellung (wie Fensterlänge, Überlappung und Merkmalskombination) wird auf 5 langsamen Tonabfolgen ausgewertet (s. Abschnitt 3.2) und die zugehörigen Verteilungen der *F*-Werte werden mittels Boxplots im folgenden Abschnitt veranschaulicht. Für die Kombinationsmethoden *Additiv* und *Multiplikativ* wurde lediglich der Wert ***A***=0.5 ausprobiert.

5.2 Auswertung

In Abbildungen 7 sind die Erkennungsraten der verschiedenen Kombinationsmethoden am Beispiel von Klavierstücken veranschaulicht. Abbildungen für die anderen Instrumente befinden sich im Anhang (Abbildungen A20, A21, A22, A23). Dabei ist zu sehen, dass die Verteilungen der *F*-Werte von Klavier und Gitarre wieder sehr ähnlich sind und sich von denen der Blasinstrumente und Geige unterscheiden. Auffällig ist aber für alle Instrumente, dass die Kombinationsmethode *Multiplikativ* mit Ausnahme eines Musikstückes sehr schlecht abgeschnitten hat.

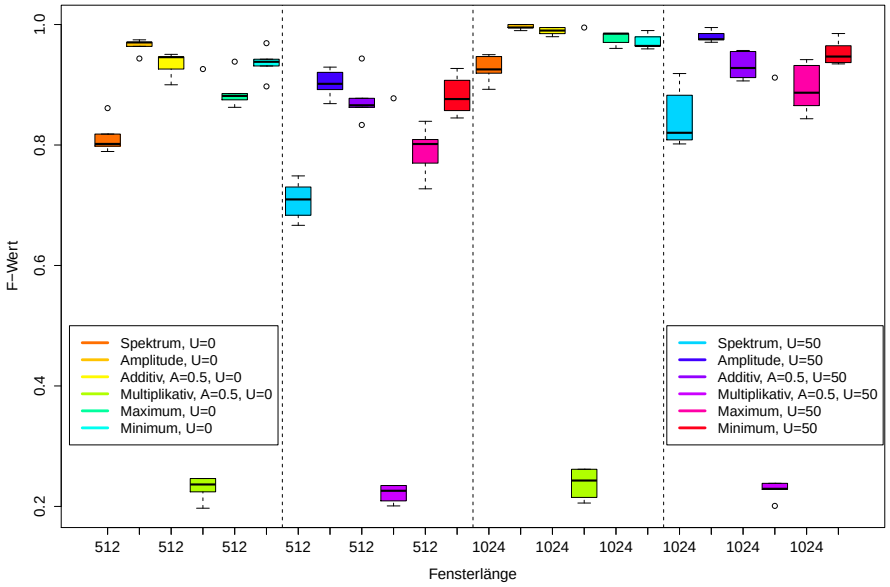


Abb. 7 Verteilung der *F*-Werte für Klavier bei verschiedenen Kombinationsmethoden, Tempo: 120 bpm, Zeitberechnung: *Fenstermitte*

Weiter gilt für alle Musikinstrumente: Keine der benutzten Kombinationsmethoden schlägt die beste Methode für das jeweilige Instrument (amplituden- bzw. spektralbasierte Methode). Die Methoden *Additiv* und *Maximum* sind durchgehend besser als die Methoden *Multiplikativ* und *Minimum*. Die Schlussfolgerungen aus diesen Beobachtungen werden im nächsten Kapitel zusammengefasst.

6 Zusammenfassung

In dieser Vorstudie wurde der Einfluss von verschiedenen Faktoren auf die Erkennungsgüte eines Algorithmus zur Einsatzzeiterkennung bei Audiosignalen untersucht. Das Hauptziel dieser Vorstudie ist, systematische Variationen einiger dieser Faktoren zu analysieren, um sich dann für deren beste Einstellung zu entscheiden. Somit soll die Parameterdimension des benutzten Algorithmus zur Einsatzzeiterkennung reduziert werden, um ihn weiter in einer ausführlicheren Studie zu optimieren. Zu solchen Parametern gehören Zeitberechnungsmethoden und Merkmalskombinationsmethoden.

Ein weiteres Ziel der Vorstudie ist, erste Erkenntnisse über den Einfluss von Musikinstrumenten auf die Erkennungsgüte des Verfahrens zu bekommen. Zu diesem Zweck werden Tonabfolgen generiert, die sich lediglich durch die Instrumentbesetzung unterscheiden. Die Besonderheit von diesen Tonabfolgen ist, dass sie mittels echter Töne erstellt wurden.

Im ersten Teil der Vorstudie wurde festgestellt, dass für Klavier und Gitarre die amplitudenbasierte Einsatzzeiterkennung besonders gute Ergebnisse liefert, wobei für die vier anderen Instrumente (Flöte, Klarinette, Trompete und Geige) die spektralbasierte Methode besser abgeschnitten hat. Ein weiterer Schluss ist, dass die Zeitberechnungsmethode *Fenstermitte* im Allgemeinen bessere Erkennungsraten als die Methode *Fensterende* liefert, weshalb die erste Methode für den endgültigen Algorithmus übernommen wird. Es werden vier Fensterlänge-Einstellungen untersucht und zwar 512, 1024, 2048 und 4096 Samples. Allerdings stellt sich heraus, dass Blasinstrumente und Geige für die Fensterlänge 4096 Samples keine akzeptablen Ergebnisse liefern. Also wird diese Einstellung für die weitere Studie nicht berücksichtigt. Außerdem konnte beobachtet werden, dass das Tempo einer Tonabfolge einen entscheidenden Einfluss auf die Ergebnisse hat: für schnelle Stücke (mit 300 Schlägen pro Minute) können keine akzeptablen Erkennungsraten verzeichnet werden, wohingegen für die langsamen Stücke (mit 120 Schlägen pro Minute) teilweise Erkennungsraten in der Nähe von 1 (Optimum) vorkommen.

Im zweiten Teil der Vorstudie werden verschiedene Merkmalskombinationsmethoden untersucht, wobei zwei Methoden besonders positiv auffallen: *Additiv* und *Maximum*. Da die Methode *Additiv* von einem weiteren Parameter abhängt (dem Anteil des jeweiligen Merkmals an der Gesamtsumme), wird sie für die weitere Studie bevorzugt. Durch die systematische Variation dieses Parameters ist eine Verbesserung der Erkennungsgüte möglich.

Danksagung

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

Literaturverzeichnis

1. **Bauer**, N., Schiffner, J., Weihs, C. (2010): Einsatzzeiterkennung bei polyphonen Musikzeitreihen. SFB 823 discussion paper 22/2010.
http://www.statistik.tu-dortmund.de/sfb823-dp2010.html

2. **Bauer**, N., Schiffner, J., Weihs, C. (2011): *Comparison of classical and sequential design of experiments in note onset detection*. Accepted for Proceedings of the 35th Annual Conference of the German Classification Society (GfKl).

3. **Dixon**, S. (2006): Onset detection revisited. In Proceedings of the 9th Int. Conference on Digital Audio Effects (DAFx'06), Montreal, Canada, 133–137.

4. **Goto**, M., Hashiguchi, H., Nishimura, T., Oka, R. (2003): RWC music database: Music genre database and musical instrument sound database. In: ISMIR 2003 Proceedings, pp. 229-230.

5. **R Development Core Team** (2008): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL: http://www.R-project.org. ISBN 3-900051-07-0.

A1 Anhang

Abbildungen für den ersten Teil der Studie

Klavier

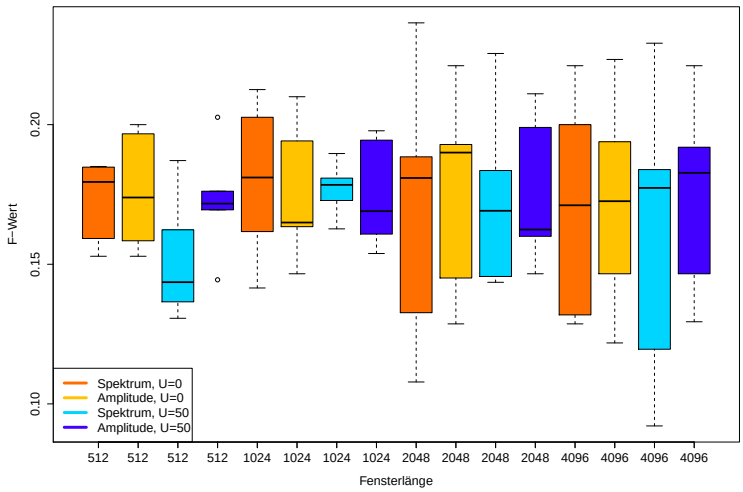


Abb. A1 Verteilung der F-Werte für Klavier, Tempo: 300 bpm, Zeitberechnung: *Fensterende*

Gitarre

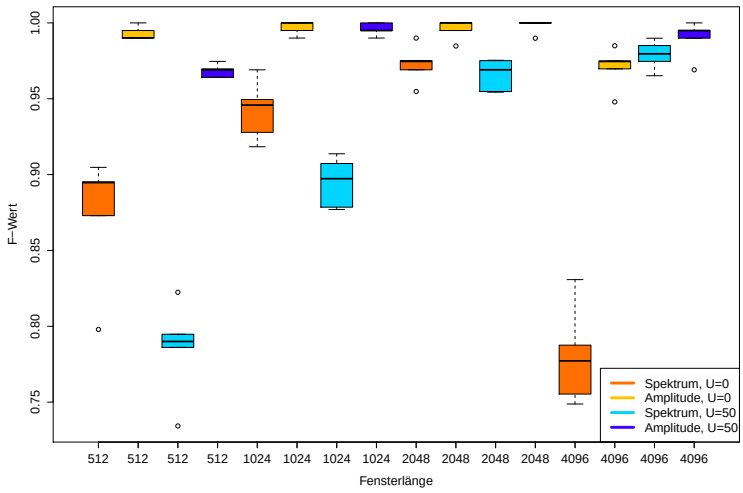


Abb. A2 Verteilung der F-Werte für Gitarre, Tempo: 120 bpm, Zeitberechnung: *Fenstermitte*

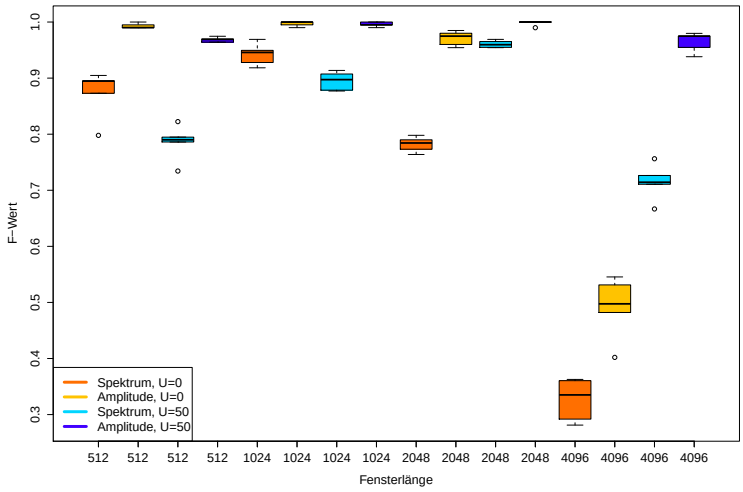


Abb. A3 Verteilung der F-Werte für Gitarre, Tempo: 120 bpm, Zeitberechnung: *Fensterende*

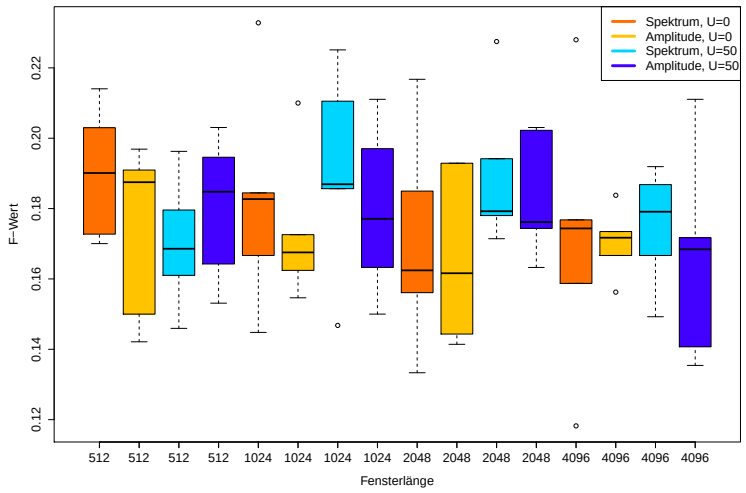


Abb. A4 Verteilung der F-Werte für Gitarre, Tempo: 300 bpm, Zeitberechnung: *Fenstermitte*

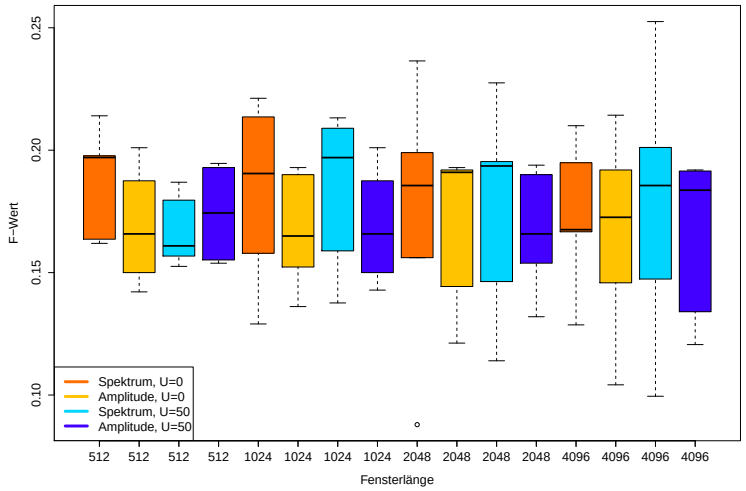


Abb. A5 Verteilung der F-Werte für Gitarre, Tempo: 300 bpm, Zeitberechnung: *Fensterende*

Flöte

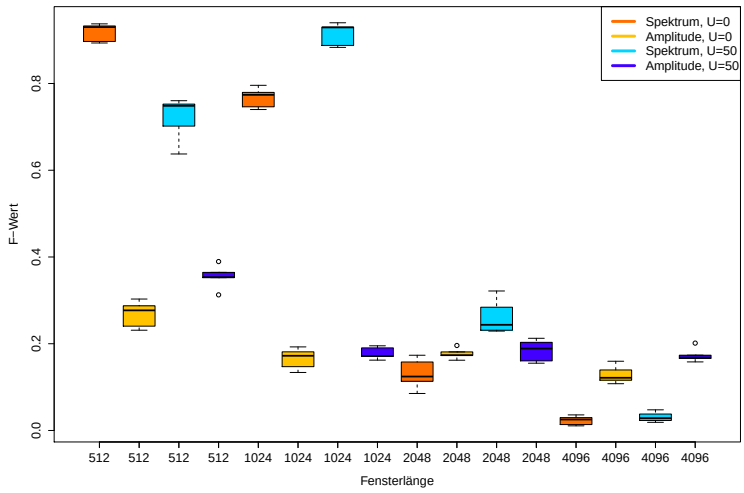


Abb. A6 Verteilung der F-Werte für Flöte, Tempo: 120 bpm, Zeitberechnungsmethode: *Fensterende*

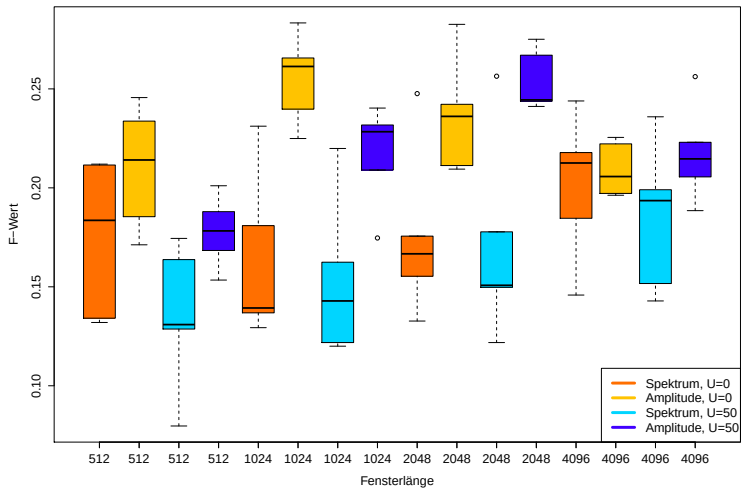


Abb. A7 Verteilung der F-Werte für Flöte, Tempo: 300 bpm, Zeitberechnungsmethode: *Fensterende*

Klarinette

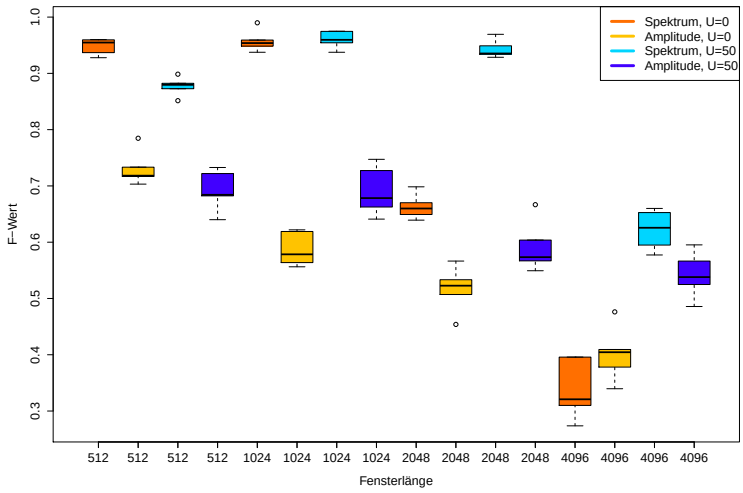


Abb. A8 Verteilung der *F*-Werte für Klarinette, Tempo: 120 bpm, Zeitberechnungsmethode: *Fenstermitte*

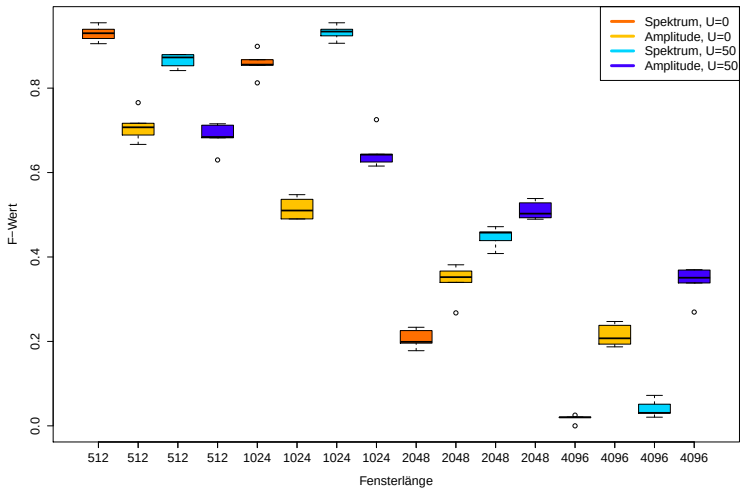


Abb. A9 Verteilung der *F*-Werte für Klarinette, Tempo: 120 bpm, Zeitberechnungsmethode: *Fensterende*

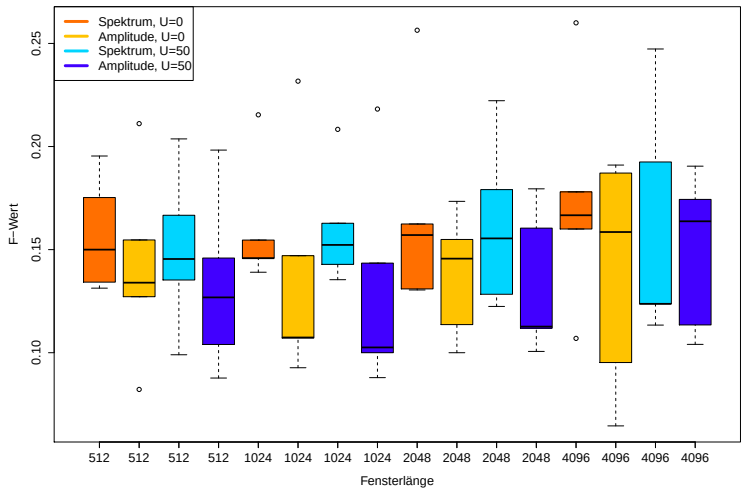


Abb. A10 Verteilung der F-Werte für Klarinette, Tempo: 300 bpm, Zeitberechnungsmethode: *Fenstermitte*

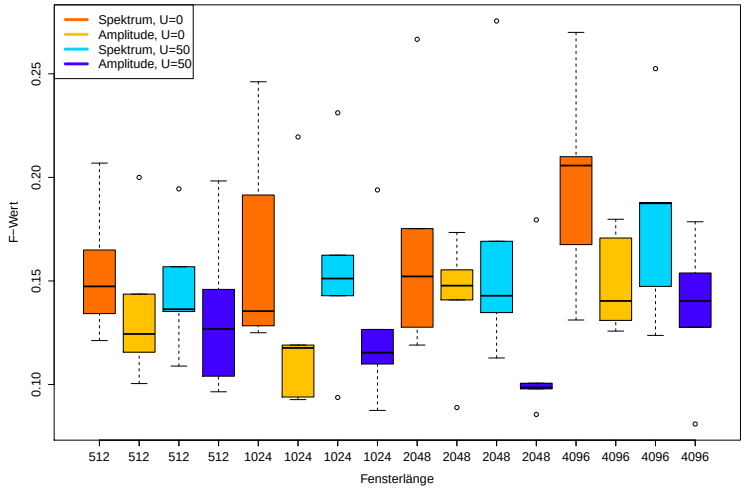


Abb. A11 Verteilung der F-Werte für Klarinette, Tempo: 300 bpm, Zeitberechnungsmethode: *Fensterende*

Trompete

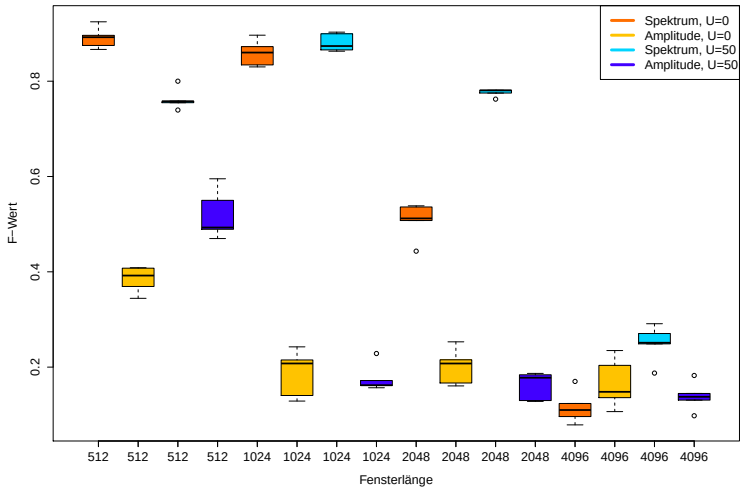


Abb. A12 Verteilung der *F*-Werte für Trompete, Tempo: 120 bpm, Zeitberechnungsmethode: *Fenstermitte*

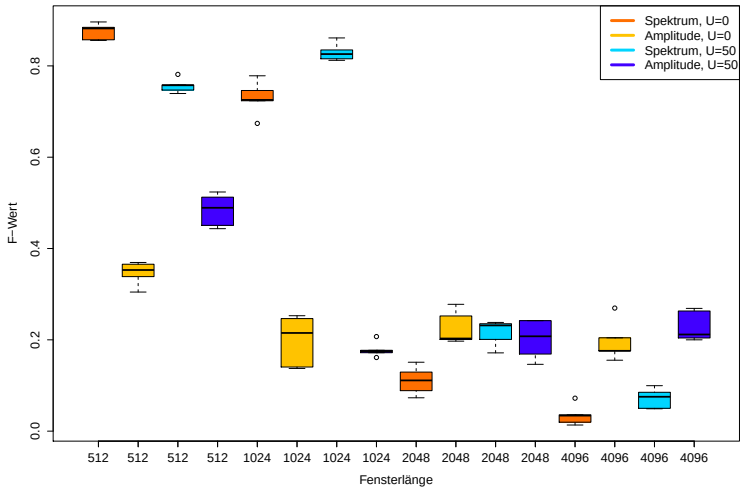


Abb. A13 Verteilung der *F*-Werte für Trompete, Tempo: 120 bpm, Zeitberechnungsmethode: *Fensterende*

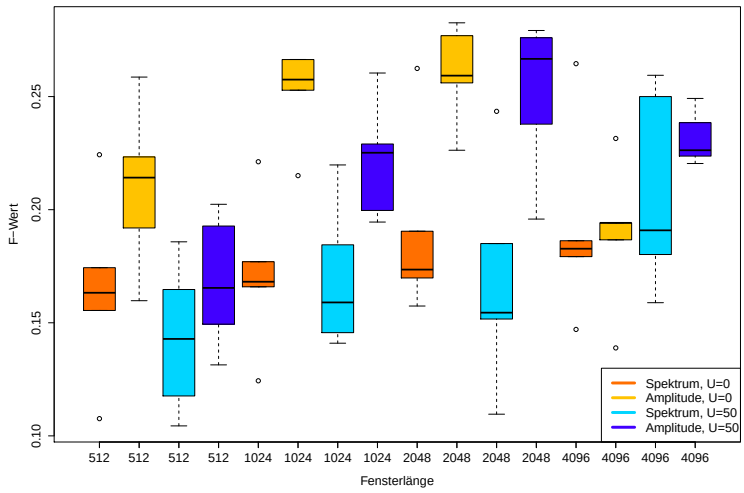


Abb. A14 Verteilung der *F*-Werte für Trompete, Tempo: 300 bpm, Zeitberechnungsmethode: *Fenstermitte*

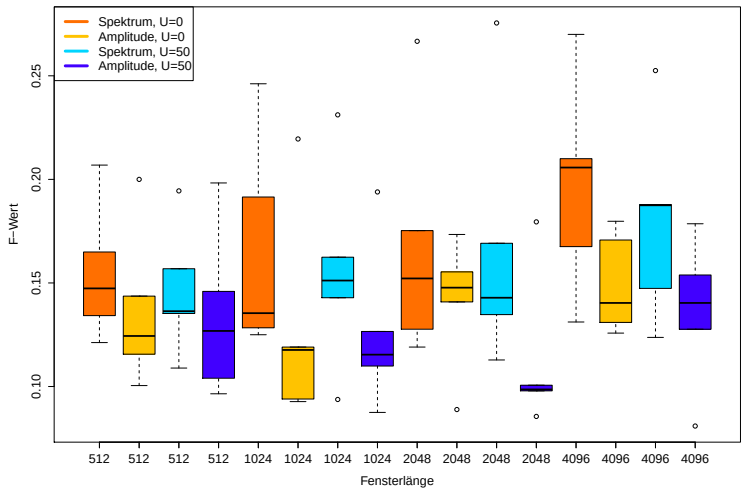


Abb. A15 Verteilung der *F*-Werte für Trompete, Tempo: 300 bpm, Zeitberechnungsmethode: *Fensterende*

Geige

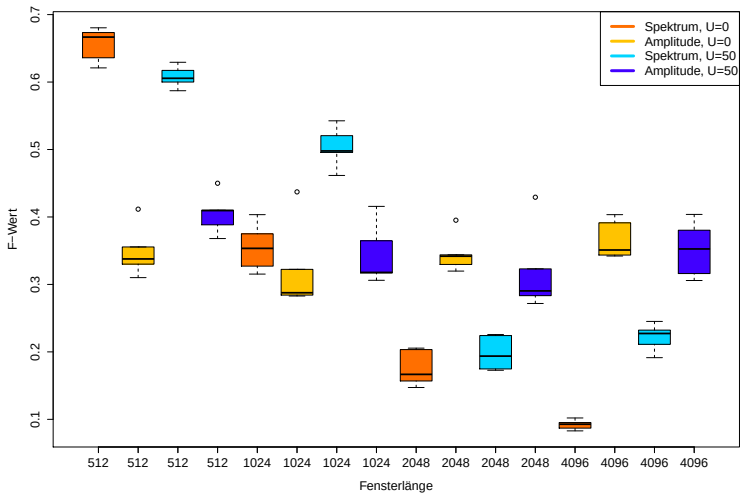


Abb. A16 Verteilung der F-Werte für Geige, Tempo: 120 bpm, Zeitberechnungsmethode: *Fensterende*

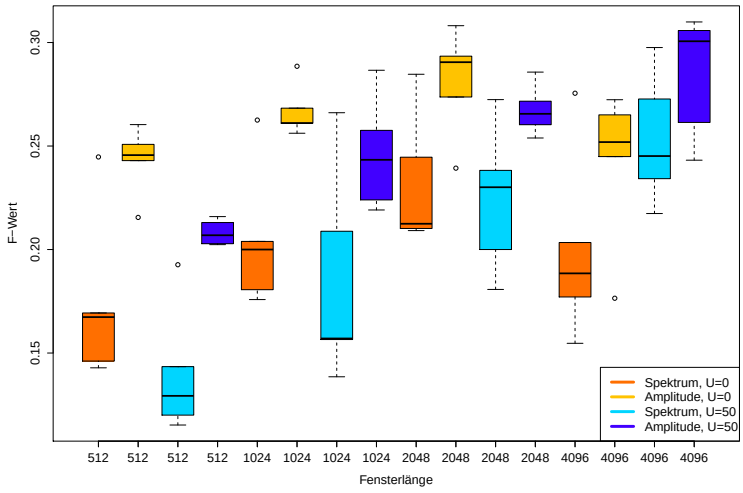


Abb. A17 Verteilung der F-Werte für Geige, Tempo: 300 bpm, Zeitberechnungsmethode: *Fenstermitte*

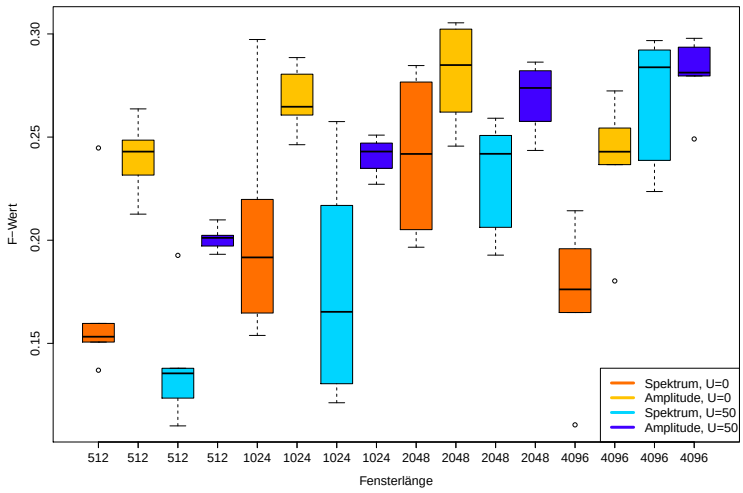


Abb. A18 Verteilung der F-Werte für Geige, Tempo: 300 bpm, Zeitberechnungsmethode: *Fensterende*

Abbildungen für dem zweiten Teil der Studie

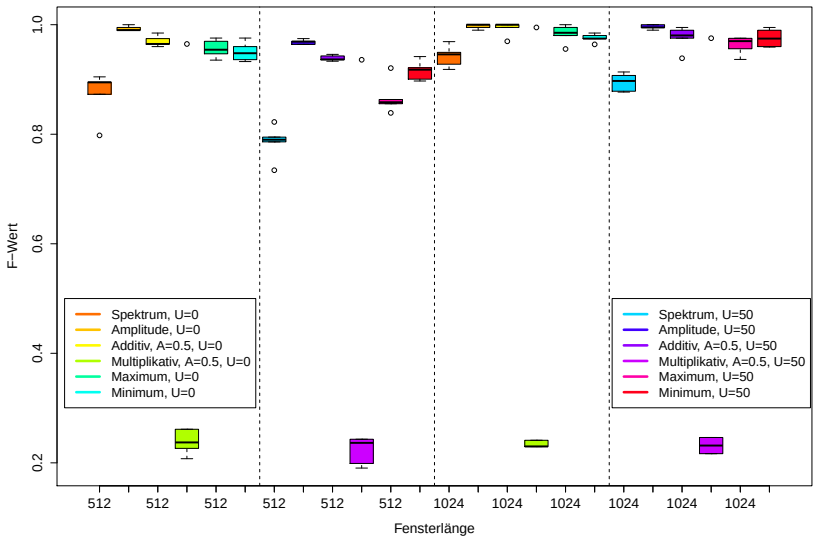


Abb. A19 Verteilung der F-Werte für Gitarre bei verschiedenen Kombinationsmethoden, Tempo: 120 bpm,Zeitberechnung: *Fenstermitte*

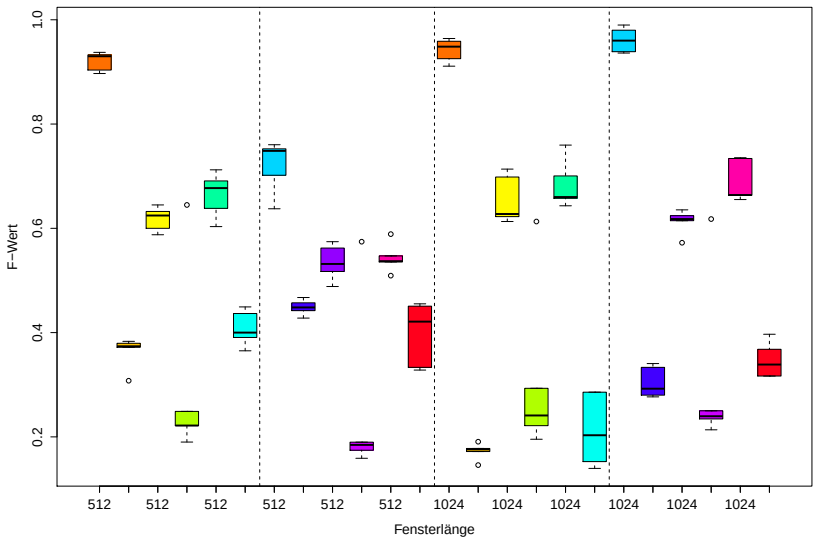


Abb. A20 Verteilung der *F*-Werte für Flöte bei verschiedenen Kombinationsmethoden, Tempo: 120 bpm, Zeitberechnung: *Fenstermitte*

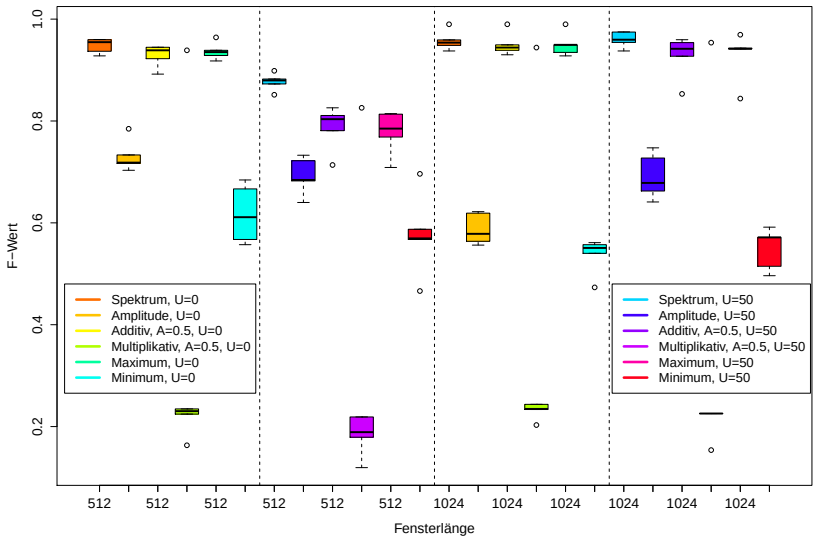


Abb. A21 Verteilung der *F*-Werte für Klarinette bei verschiedenen Kombinationsmethoden , Tempo: 120 bpm, Zeitberechnung: *Fenstermitte*

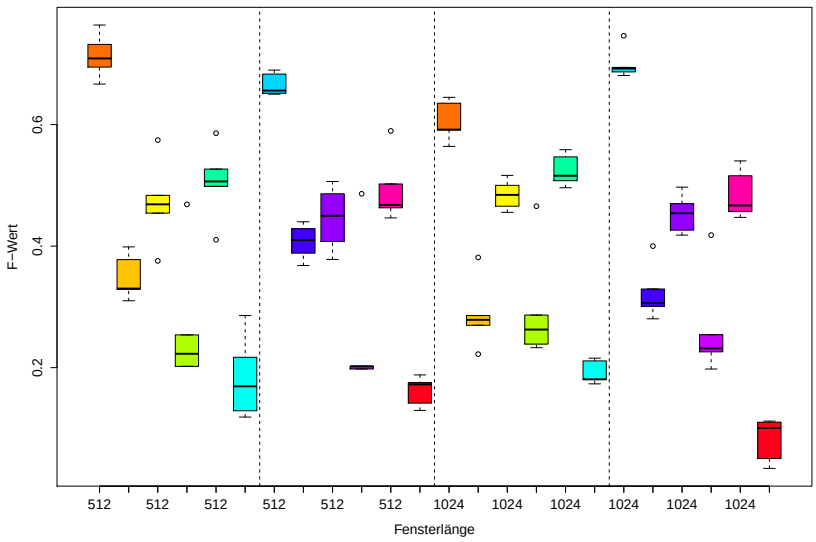


Abb. A22 Verteilung der F-Werte für Trompete bei verschiedenen Kombinationsmethoden, Tempo: 120 bpm, Zeitberechnung: *Fenstermitte*

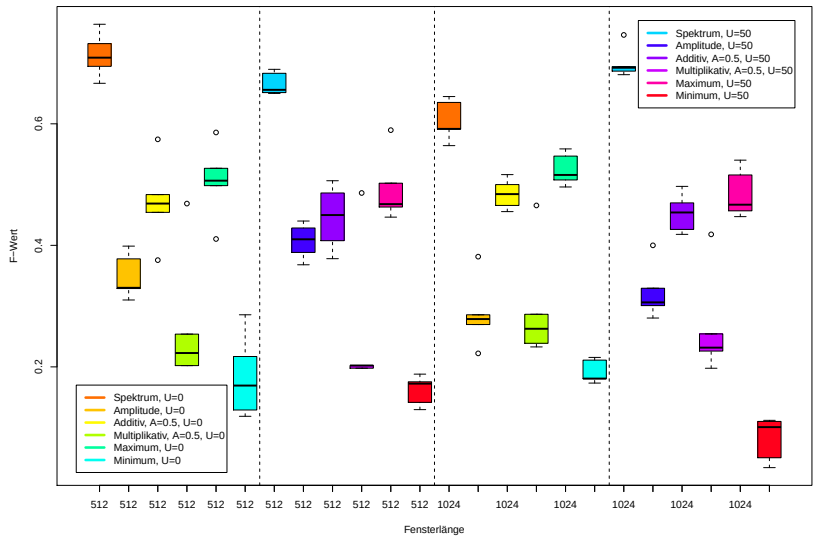


Abb. A23 Verteilung der F-Werte für Geige bei verschiedenen Kombinationsmethoden, Tempo: 120 bpm, Zeitberechnung: *Fenstermitte*